BIRKHÄUSER

Mathematical Modeling of Biological Systems, Volume I

Cellular Biophysics, Regulatory Networks, Development, Biomedicine, and Data Analysis

Andreas Deutsch
Lutz Brusch
Helen Byrne
Gerda de Vries
Hanspeter Herzel

Editors

Modeling and Simulation in Science, Engineering and Technology

Series Editor
Nicola Bellomo
Politecnico di Torino
Italy

Advisory Editorial Board

Mathematical Modeling of Biological Systems, Volume I

Cellular Biophysics, Regulatory Networks,
Development, Biomedicine, and
Data Analysis

Andreas Deutsch
Lutz Brusch
Helen Byrne
Gerda de Vries
Hanspeter Herzel
Editors

Andreas Deutsch
Center for Information Services
   and High Performance Computing
Technische Universität Dresden
01062 Dresden
Germany

Lutz Brusch
Center for Information Services
   and High Performance Computing
Technische Universität Dresden
01062 Dresden
Germany

Helen Byrne
Centre for Mathematical Medicine
School of Mathematical Sciences
University of Nottingham
Nottingham NG7 2RD
U.K.

Gerda de Vries
Department of Mathematical and
   Statistical Sciences
University of Alberta
Edmonton, AB T6G 2G1
Canada

Hanspeter Herzel
Institute of Theoretical Biology
Humboldt-Universität zu Berlin
Invalidenstraße 43
10115 Berlin
Germany

Preface

This edited volume contains a selection of chapters that are an outgrowth of the European Conference on Mathematical and Theoretical Biology (ECMTB05, Dresden, Germany, July 2005). The peer-reviewed contributions show that mathematical and computational approaches are absolutely essential for solving central problems in the life sciences, ranging from the organizational level of individual cells to the dynamics of whole populations.

The contributions indicate that theoretical and mathematical biology is a diverse and interdisciplinary field, ranging from experimental research linked to mathematical modeling to the development of more abstract mathematical frameworks in which observations about the real world can be interpreted, and with which new hypotheses for testing can be generated. Today, much attention is also paid to the development of efficient algorithms for complex computation and visualisation, notably in molecular biology and genetics. The field of theoretical and mathematical biology and medicine has profound connections to many current problems of great relevance to society. The medical, industrial, and social interests in its development are in fact indisputable. Insights and predictions from mathematical modeling are used increasingly in decision support in medicine (e.g., immunology and spread of infectious diseases, cancer research, cardiovascular research, neurological research, optimisation of medical treatments, imaging), environmental and nature management, climate problems, agriculture, and management of natural resources. Rapid developments in areas such as biotechnology (e.g., genome projects, genetic modification, tissue engineering) continue to add new focal points of activity to the field. The contributions of this volume capture some of these developments.

The volume is divided into five parts—cellular biophysics, regulatory networks, development, biomedical applications, and data analysis and model validation.

Part I deals with cellular biophysics and contains six chapters.

Kovalenko and Riznichenko consider multiparticle simulations of photosynthetic electron transport processes. In particular, a 3D model of cyclic electron transport is developed and applied to a study of fast and slow components of the reaction center of a photosystem 1 pigment-protein complex. It is demonstrated that the slow phase of

this process is diffusion-controlled and determined by the diffusion of reduced plasto-quinone and plastocyanin molecules from the granal to stromal areas of the thylakoid membrane.

Knoke, et al. study the selective regulation of protein activity by complex Ca^{2+} oscillations. Calcium oscillations play an essential role in intracellular signal transduction. A particular question is how two or more classes of proteins can be specifically regulated at the same time. The question is general and concerns the problem of how one second messenger can transmit more than one signal simultaneously (bow-tie structure of signalling). To investigate whether a complex Ca^{2+} signal like bursting, a succession of low-peak and high-peak oscillatory phases, could selectively activate different proteins, several bursting patterns with simplified square pulses were applied in a theoretical model. The results indicate that bursting Ca^{2+} oscillations allow a differential regulation of two different calcium-binding proteins, and hence, perform the desired function.

Gamba, et al. focus on phase separation in eukaryotic directional sensing. Many eukaryotic cell types share the ability to migrate directionally in response to external chemoattractant gradients. The binding of chemoattractants to specific receptors leads to a wide range of biochemical responses that become highly localized as cells polarize and migrate by chemotaxis. This ability is central in the development of complex organisms, and is the result of millions of years of evolution. Cells exposed to shallow gradients in chemoattractant concentration respond with strongly asymmetric accumulation of several factors, including the phosphoinositides PIP_3 and PIP_2, the PI 3-kinase PI3K and phosphatase PTEN. An early symmetry-breaking stage is believed to trigger effector pathways leading to cell movement. Although many signaling factors implied in directional sensing have been recently discovered, the physical mechanism of signal amplification is not yet well understood. The authors propose that directional sensing is the consequence of a phase ordering process mediated by phosphoinositide diffusion and driven by the distribution of chemotactic signals. By studying a realistic reaction-diffusion lattice model that describes PI3K and PTEN enzymatic activity, recruitment to the plasmamembrane, and diffusion of their phosphoinositide products, it is shown that the effective enzyme-enzyme interaction induced by catalysis and diffusion introduces an instability of the system towards phase separation for realistic values of physical parameters. In this framework, large reversible amplification of shallow chemotactic gradients, selective localization of chemical factors, macroscopic response timescales, and spontaneous polarization arise naturally.

Brusch, et al. consider the formation of spatial protein domains of small guanosine triphosphatases (GTPases) on membranes. In particular, several mechanisms for spatial domain formation of GTPases on cellular membranes are discussed. Furthermore, a kinetic model of the basic guanine-nucleotide cycle common to all GTPases is developed and coupled along a one-dimensional axis by diffusion of inactive and activated GTPases. It is asked, whether a parameter set exists such that domain formation is possible by Turing's mechanism, i.e., purely by reactions and diffusion, and shown that the Turing instability does not occur in this model for any parameter combination. But as revealed by stability and bifurcation analysis, domain formation is reproduced after augmenting the model with combinations of two spatial interaction mechanisms:

(1) attraction; and (2) adhesion among active GTPases. These interactions can be mediated by effector proteins that bind active GTPases. The model predicts domains to disintegrate if effector binding is inhibited.

Tracqui, et al. discuss in vitro tubulogenesis of endothelial cells. The formation of new blood vessels *in vivo* is a multistep process in which sprouting endothelial cells (ECs) form tubes with lumen, these tubes being additionally organized as capillary networks. In vitro models of tubulogenesis have been developed to investigate this highly regulated multifactorial process, with special attention paid to the determinant role of mechanical interactions between ECs and the extracellular matrix (ECM). In agreement with experimental results obtained when culturing endothelial EAhy926 cells on fibrin gels, the authors define theoretical thresholds between cellular traction and active cell migration along ECM strain fields above which tubulogenesis is induced. In addition, it is illustrated how mechanical factors may provide long-range positional information signals leading to localized network formation. This provides an alternative view to the classical approach of morphogenesis based on gradients of diffusible morphogens.

Time distributions in biocatalytic systems are considered by *Kühl and Jobmann*. Formal kinetic methods to analyze biocatalytic systems are traditionally based on the law of mass action. This law involves the assumption that each molecular state has an exponentially distributed lifetime. The authors regard this assumption as unduly restrictive and propose a more general, service theory-based approach (termed mass service kinetics or briefly service kinetics). In service-theoretic terms, biocatalysts are servers and their ligands are customers. The time intervals between arrivals of ligand molecules at special service loci (active or binding sites) as well as the service periods at these loci need not be exponentially distributed; rather, they may adopt any distribution (e.g., Erlangian, hyperexponential, variomorphic). The authors exemplify the impact of nonexponential time distributions on a performance measure of wide interest: the steady-state throughput. Specifically, it is shown that nonexponential interarrival times convert hyperbolic mass action systems (whose characteristic is a hyperbolic velocity-concentration or dose-response curve) into nonhyperbolic mass service systems, and that type and extent of their nonhyperbolicity are determined by type and parameters of the interarrival time distribution. A major conclusion is that it is a questionable practice to routinely and exclusively use mass action kinetics for the interpretation and performance evaluation of biocatalytic systems.

Part II deals with regulatory networks and comprises five chapters.

Booth, et al. analyze a stochastic model of gene regulation using the chemical master equation. This equation in combination with chemical rate equations is employed as a tool to study Markovian models of genetic regulatory networks in prokaryotes. States of the master equation represent the binding and unbinding of protein complexes to DNA, resulting in a gene being expressed in a cell or not, while protein and substrate concentrations are represented by continuum variables which evolve via differential equations. The model is applied to a moderately complex biological system, the switching mechanism of the bacteriophage λ driven by competition between production of CI and Cro proteins. Numerical simulations of the model successfully move

between lysogenic and lytic states as the host bacterium is stressed by the application of ultraviolet light.

Ropers, et al. consider piecewise-linear models of genetic regulatory networks and analyze the carbon starvation response in *Escherichia coli*. The growth adaptation of *Escherichia coli* to the availability of the carbon source is controlled by a complex genetic regulatory network whose functioning is still very little understood. Using a qualitative method based on piecewise-linear differential equations, which is able to overcome the current lack of quantitative data on kinetic parameters and molecular concentrations, the authors model the carbon starvation response network and simulate the response of *E. coli* cells to carbon deprivation. This allows one to identify essential features of the transition between the exponential and the stationary phase and to make new predictions on the qualitative system behavior, following a carbon upshift.

Elo and Aittokallio present an attempt to predict gene expression by combining information from expression and promoter profiles. Gene expression microarrays have become a popular high-throughput technique in functional genomics. By enabling the monitoring of thousands of genes simultaneously, this technique holds enormous potential to extend our understanding of various biological processes. The large amount of data poses, however, a challenge when interpreting the results. Moreover, microarray data often contain frequent missing values, which may drastically affect the performance of different data analysis methods. Therefore, it is essential to effectively exploit additional biological information when analyzing and interpreting the data. In the present study, the authors investigate the relationship between gene expression profile and promoter sequence profile in the context of missing value imputation. In particular, it is demonstrated that the selection of predictive genes for expression value estimation can be considerably improved by the incorporation of transcription factor binding information.

Centler, et al. focus on chemical organization in the central sugar metabolism of *Escherichia coli*. The theory of chemical organizations is employed as a novel method to analyze biological network models. The method allows one to decompose a chemical reaction network into subnetworks that are (algebraically) closed and self-maintaining. Such subnetworks are termed *organizations*. Although only stoichiometry is considered to compute organizations, the analysis allows one to narrow down the potential dynamic behavior of the network: organizations represent potential steady-state compositions of the system. When applied to a model of sugar metabolism in *E. coli* including gene expression, signal transduction, and enzymatic activities, some organizations are found to coincide with inducible biochemical pathways.

Noé and Smith present transition networks. A transition network (TN) is a graph-theoretical concept describing the transitions between (meta)stable states of dynamical systems. The authors review methods to generate and analyze TNs for molecular systems. The appropriate identification of states and transitions from the potential energy surface of the molecule is discussed. Furthermore, a formalism transforming a TN on a static energy surface into a time-dependent dynamic TN is described that yields the population probabilities for each system state and the inter-state transition rates. Three analysis methods that help in understanding the dynamics of the molecular system based on the TN are discussed: (1) Disconnectivity graphs allow important features

of the energy surface captured in a static TN to be visualized; (2) Graph-theoretical methods enable the computation of the best transition paths between two predefined states of the TN; and (3) Statistical methods from complex network analysis identify important features of the TN topology.

Part III focuses on development and consists of five chapters.

Sekimura, et al. consider pigmentation pattern formation in butterfly wings, one of the most spectacular and vivid examples of pattern formation in biology. The authors devote their attention to the mechanisms for generating global patterns with a focus on the relationship between pattern forming mechanisms for the fore- and hind-wing patterns. Through mathematical modeling and computational analysis of *Papilio dardanus* and *polytes*, the results indicate that the patterns formed on the fore-wing need not correlate to those of hind-wing patterns in the sense that the formation mechanism is the same for both patterns. The independence of pattern formation mechanisms means that the coordination of unified patterns of fore- and hind-wing is accidental. This is remarkable, because owing to Oudemans's principle, patterns appearing on the exposed surface of fore- and hind-wing at the natural resting position are often integrated to form a composite and unified adaptive pattern with their surrounding environment.

Christley, et al. introduce an agent-based model for developmental pattern formation with multiscale dynamics and varying cell geometry. Cells of the embryonic vertebrate limb in high-density culture undergo chondrogenic pattern formation, which results in the formation of regularly-spaced "islands" of cartilage analogous to the cartilage primordia of the developing limb skeleton. The authors describe a discrete, multiscale agent-based stochastic model, which is based on an extended cell representation coupled with biologically motivated reaction-diffusion processes and cell-matrix adhesion, for studying the behavior of limb bud precartilage mesenchymal cells. The model is calibrated using experimental data and the sensitivity of key parameters is studied.

Starruß, et al. address bacterial swarming driven by rod shape. Swarming pattern formation of self-propelled entities is a prominent example of collective behavior in biology. The authors show that the rod shape of self-propelled individuals is able to drive swarm formation without any kind of signaling. The proposed mechanism is purely mechanical and is evidenced through two different mathematical approaches: an on-lattice and an off-lattice individual-based model. The intensities of swarm formation obtained in both approaches uncover that the length-width aspect ratio controls swarm formation, and that there is an optimal aspect ratio that favors swarming.

King and Franks consider stability properties of some tissue-growth models. In particular, free-boundary problems associated with biological tissue growing under conditions of nutrient limitation are formulated. Analysis by linear-stability methods, clarifying the models' stability properties, is then described.

Madzvamuse introduces a modified first-order backward Euler finite difference scheme to solve advection-reaction-diffusion systems on fixed and continuously deforming domains. This scheme is compared to the second-order semi-implicit backward finite differentiation formula, and it is concluded that for the type of equations considered, the first-order scheme has a larger region of stability for the time-step than

that of the second-order scheme (at least by a factor of ten), and therefore the first-order scheme becomes a natural choice when solving advection-reaction-diffusion systems on growing domains.

Part IV deals with biomedical applications and consists of twelve chapters.

Iomin considers fractional transport of cancer cells due to self-entrapping by fission. In particular, a simple mathematical model is proposed to study the influence of cell fission on transport. The model describes fractional tumor development, which is a one-dimensional continuous-time random walk (CTRW). Furthermore, an answer to the question of how malignant neoplasm cells can appear at an arbitrary distance from the primary tumor is proposed. The model may provide a possible explanation for diffusive cancers as well. In addition, a chemotherapy influence on the CTRW is studied by an observation of stationary solutions.

Panovska, et al. address mathematical modeling of vascular tumor growth and implications for therapy. The authors discuss the results of a mathematical model that incorporates many processes associated with tumor growth. The deterministic model, a system of coupled nonlinear partial differential equations, is a combination of two previous models that describe the tumor-host interactions in the initial stages of growth and the tumor angiogenic process. Numerical simulations show that the model captures both the avascular and vascular growth phases. Furthermore, a number of characteristic features of vascular tumor growth are recovered, such as the rate of tumor growth and the invasion speed. It is also shown how the model can be used to investigate the effects of different anti-cancer therapies.

Stein, et al. present a stochastic model of glioblastoma invasion. Glioblastoma is the most malignant form of brain cancer. It is extremely invasive; the mechanisms that govern invasion are not well understood. To better understand the process of invasion, the authors conducted an *in vitro* experiment in which a 3D tumor spheroid is implanted into a collagen gel. The paths of individual invasive cells were tracked. These cells were modeled as radially biased, persistent random walkers. The radial velocity bias was found to be 19.6 μm/hr.

A model for the morphology of the tumor vasculature is introduced by *Bartha and Rieger*. The model is based on the molecular interactions between a growing tumor and a dynamically evolving blood vessel network, and describes the transformation of the regular vasculature in normal tissues into a highly inhomogeneous tumor specific capillary network. The emerging morphology, characterized by the compartmentalization of the tumor into several regions differing in vessel density, diameter, and degree of tumor necrosis, is in accordance with experimental data for human melanoma. Vessel collapse, due to a combination of severely reduced blood flow and solid stress exerted by the tumor, leads to a correlated percolation process that is driven towards criticality by the mechanism of hydrodynamic vessel stabilization.

Clairambault, et al. present a mathematical model of the cell cycle and its circadian control. The following question is addressed: Can one sustain, on the basis of mathematical models, that for cancer cells, the loss of control by a circadian rhythm favors a faster population growth? This question, which comes from the observation that tumor growth in mice is enhanced by experimental disruption of the circadian

rhythm, may be tackled by mathematical modeling of the cell cycle. For this purpose an age-structured population model is considered with control of death (apoptosis) rates and phase transitions, and two eigenvalues: one for periodic control coefficients (via a variant of Floquet theory in infinite dimension) and one for constant coefficients (taken as the time average of the periodic case). It is shown by a direct proof that, surprisingly enough considering the above-mentioned observation, the periodic eigenvalue is always greater than the steady-state eigenvalue when the sole apoptosis rate is considered. It is also demonstrated by numerical simulations when transition rates between the phases of the cell cycle are taken into account, that, without further hypotheses, no natural hierarchy between the two eigenvalues exists. This at least shows that, if such models are to take account the above-mentioned observation, control of death rates inside phases is not sufficient, and that transition rates between phases are a key target in proliferation control.

Moroz and Wimpenny consider a bone turnover cycle model with a torus-like steady state. A quantitative understanding of the bone remodeling process is of considerable biomedical and practical biotechnological interest to support the application of layer manufacturing techniques to produce scaffolds for surgical applications. Osteoclasts and osteoblasts play a principal role in different models of the bone multicellular unit operating in bone and display a rich spectrum of behaviors. The goal of the authors is to show that it is possible to capture the cyclic dynamics of operating cells. The central idea of the mathematical model is that the regulatory nature of osteocytes is the basis of the cyclic behavior associated with the system (remodeling process) as a whole. Simulations show that for a particular range of constants, the model exhibits a torus-like quasi-steady state. Further investigation of these simulations indicates the existence of a surface in the osteoclasts-osteoblasts-osteocytes-bone space, which could be interpreted as a conservative value confirming the substrate-energy regenerative capability of the bone remodeling system. It is suggested that the nature of this recovering potential is directed against both mechanical and biochemical damage to the bone.

Plank, et al. address the modeling of the early stages of atherosclerosis. Atherosclerotic lesions are predominantly localised to arterial bifurcations and bends, and are highly correlated with areas of low wall shear stress (WSS), but the underlying reason for this localisation is not fully understood. A key role is played by endothelial cells, which regulate the transport of materials from the bloodstream to the artery wall and secrete vasoactive agents that modulate vascular tone. A mathematical model is presented, exploring the link between arterial geometry, WSS, and factors related to atherogenesis. The model simulates the cellular response to the fluid shear stress on the cell membrane and the binding of ligands to cell surface receptors. This is used to calculate the rate of production of nitric oxide (NO), which is a potent vasodilator and anti-atherogenic factor. It is hypothesised that the section of endothelium adjacent to a region of recirculating flow is most at risk of developing atherosclerotic plaque, due to reduced bioavailability of NO.

Trenado and Strauss consider magnetic nanoparticles for in vivo applications. In particular, in vivo applications of biocompatible magnetic nanoparticles in a carrier liquid controlled by an external magnetic field from outside the body have recently

been proposed for specific drug delivery, such as in locoregional cancer therapies or occlusion aneurysms. Such particles can also be used as guided contrast agents in myocardial imaging after myocardial infarction. However, the choice of the optimal clinical setting still remains a challenge for each of the mentioned applications. The authors introduce a numerical heterogeneous multiscale model that can be used for the optimal a priori determination of the free parameters and might help to overcome this problem.

Cherniha, et al. address fluid transport in peritoneal dialysis. In particular, a mathematical model incorporating water flow between the dialysis fluid in the peritoneal cavity, blood flow through the capillary wall, and homogeneous interstitium driven by high hydrostatic and osmotic pressure of dialysis fluid is formulated. The model is based on nonlinear equations of reaction-diffusion-convection type. Numerical simulations provide the distribution profiles for hydrostatic pressure, glucose concentration, and water flux in the tissue for different times from the infusion of dialysis fluid into the peritoneal cavity for different transport parameters that represent clinical treatments of peritoneal dialysis.

Sibona, et al. discuss the relevance of intracellular replication to the evolution of Chagas' disease. In particular, a model is introduced for the interaction between the parasite *Trypanosoma cruzi* and the immune system in Chagas' disease by separately describing the intracellular and extracellular parasite stages. The solution of the case where two antibody species are active is worked out in detail, and a diagram showing the differents outcomes of the model is presented. The predictions accurately reproduce experimental data on the infection evolution during the acute phase of the disease and lead to an estimate of the damage generated by direct parasite action.

Gerisch and Geris introduce a finite-volume spatial discretisation scheme for taxis-diffusion-reaction systems with axi-symmetry. In particular, the numerical simulation of a time-dependent taxis-diffusion-reaction model of fracture healing in mice using the method of lines is considered. The partial differential equation problem has an axi-symmetric structure, and this is employed to properly reduce the model to an equivalent problem in 2D space leading subsequently to an efficient spatial discretisation. Special care is given to respect conservation of mass and the nonnegativity of the solution. The numerical simulation results are contrasted to those obtained from a simplistic reduction of the axi-symmetric model to 2D space (at the same computational cost). Quantitative and qualitative differences are observed.

The information content of clinical time series is analyzed towards the development of a neonatal disease severity score system by *Menconi, et al*. In particular, a score is introduced to classify the severity of patients by analysing the information content of clinical time series.

Part V focuses on data analysis and model validation and is comprised of four chapters.

The statistical analysis and physical modeling of oligonucleotide microarrays is introduced by *Burden, et al*. Inference of regulatory networks from microarray data relies on expression measures to identify gene activity patterns. However, currently existing expression measures are not the direct measurements of mRNA concentra-

tion one would ideally need for an accurate determination of gene regulation. If the development of expression measures is to advance to the point where absolute target concentrations can be estimated, it is essential to have an understanding of physical processes leading to observed microarray data. The authors survey here the performance of existing expression measures for oligonucleotide microarrays and describe recent progress in developing physical dynamic adsorption models relating measured fluorescent dye intensities to underlying target mRNA concentration.

Bortfeldt, et al. discuss the validation of human alternative splice forms using the EASED platform and multiple splice site discriminating features. The authors have shown for a data set of computationally predicted alternative splice sites how inherent information can be utilized to validate the predictions by applying statistics on different features typical for splice sites. As a promising splice-site feature, the frequencies of binding motifs in the context of exonic and intronic splice-site flanks and between the alternative and reference splice sites have been investigated. It is shown that both partitions of splice sites can statistically be separated, not only by their distance to the splice signal consensus, but also via frequencies of splice regulatory proteins (SRp) binding motifs in the splice-site environment.

Polańska, et al. consider the Gaussian mixture decomposition of time-course DNA microarray data. Especially, the decomposition approach to the analysis of large gene expression profile data sets is presented, and the problem of analysis of transient time-course data of expression profiles is addressed. The assumption that co-expression of genes can be related to their belonging to the same Gaussian component is accepted, and it is assumed that parameters of Gaussian components, means and variances, can differ between time instants. However, the gene composition of components is unchanged between time instants. For such problem formulation the appropriate version of the expectation maximization algorithm is derived as well as recursions for the estimation of model parameters. The derived method is applied to the data on gene expression profiles of human K562 erythroleukemic cells, and the obtained gene clustering is discussed.

Simek and Jarząb discuss SVD analysis of gene expression data. The analysis of gene expression profiles of cells and tissues, performed by DNA microarray technology, strongly relies on proper bioinformatical methods of data analysis. Due to a large number of analyzed variables (genes) and a usually low number of cases (arrays) in one experiment, limited by high cost of the technology, the biological reasoning is difficult without previous analysis, leading to the reduction of the problem's dimensionality. A wide variety of methods has been developed, with the most useful, from the biological point of view, methods of supervised gene selection with estimation of false discovery rate. However, supervised gene selection is not always satisfying for the user of microarray technology, as the complexity of biological systems analyzed by microarrays rarely can be explained by one variable. Among unsupervised methods of analysis, hierarchical clustering and PCA have gained wide biological application. In the authors' opinion, Singular Value Decomposition (SVD) analysis, which is similar to PCA, has additional advantages very essential for the interpretation of biological data. The authors show how to apply the SVD to unsupervised analysis of transcriptome data, obtained by oligonucleotide microarrays.

Finally, the volume owes its existence to the support of many colleagues. First of all, thanks go to the authors of the various contributions. We would also like to express our gratitude to the members of the ECMTB05 scientific committee and to a significant number of other colleagues for providing reviews and suggestions. ECMTB05 and these peer-reviewed proceedings have only become possible thanks to the strong institutional support provided by the Centre for Information Services and High Performance Computing (Technical University of Dresden). Particular thanks go to Wolfgang E. Nagel, the head of this Centre, and many colleagues at the Centre, particularly Niloy Ganguly, Christian Hoffmann, Samatha Kottha, Claudia Schmidt, Jörn Starruß, and Sabine Vollheim. Finally, we would like to thank Tom Grasso from Birkhäuser for making this project possible.

Dresden, January 2007
Andreas Deutsch (on behalf of the volume editors)

Contents

Part I

Cellular Biophysics

1

Multiparticle Direct Simulation of Photosynthetic Electron Transport Processes

Ilya B. Kovalenko and Galina Yu. Riznichenko

Department of Biology, Lomonosov Moscow State University, Leninskie Gory, Moscow, 119992, Russia; kovalenko78@mail.ru

Summary. In our previous study [3] we described the method for a direct three-dimensional (3D) computer simulation of ferredoxin-dependent cyclic electron transport around the photosystem 1 pigment-protein complex. Simulations showed that the spatial organization of the system plays a significant role in shaping the kinetics of the redox turnover of P700 (the reaction center of a photosystem 1 pigment-protein complex). In this paper we develop the direct 3D model of cyclic electron transport and apply it to study the nature of fast and slow components of the P700$^+$ dark reduction process. We demonstrate that the slow phase of this process is diffusion controlled and is determined by the diffusion of reduced plastoquinone and plastocyanin molecules from the granal to the stromal areas of the thylakoid membrane.

Key words: Photosynthesis, cyclic electron flow, Brownian diffusion.

1.1 Introduction

The photosynthetic electron transport chain of thylakoid in green plants and algae involves the pigment-protein complexes photosystem 1 (PS1) and photosystem 2 (PS2). The two photosystems are connected by a series of electron carriers that include plastoquinone (PQ), the cytochrome b_6/f complex (cyt b_6/f) and plastocyanin (Pc). Plastoquinone molecules diffuse in the thylakoid membrane. Mobile electron carriers Pc and ferredoxin (Fd) are small proteins that diffuse in the lumen (internal space between thylakoid membranes) and stroma (surrounding fluid medium), respectively.

Under illumination PS1 catalyzes the process of plastocyanin oxidation on the luminal side of the thylakoid membrane and ferredoxin reduction on its stromal side (Fig. 1.1, [1]). These reactions are followed by oxidation of Fd and reduction of plastoquinone (PQ) pool. Since Fd molecules are localized within the stroma and PQ is a hydrophobic carrier residing in the lipid layer of the membrane, these events are likely to be mediated by a protein, exposed to the stroma with Fd-PQ-oxidoreductase (FQR) activity. The subsequent oxidation of PQ involves the cytochrome b_6/f complex and results in the reduction of Pc, which is localized in the lumen.

Experimentally [3, 10] the kinetics of a light-induced electron spin resonance (ESR) I signal was studied in the time span 0.1–10 s. This ESR I signal represents
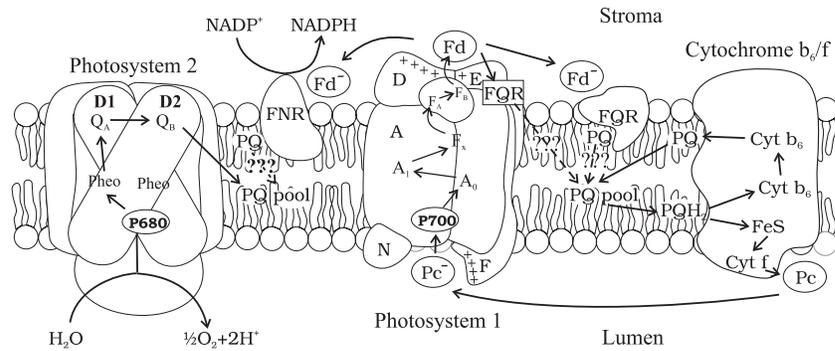
Fig. 1.1. Organization of cyclic electron transport in chloroplasts. Shown are thylakoid membrane and components of electron transport chain: complexes PS1, PS2, FQR, FNR and cytochrome b_6/f complex and also mobile electron carriers plastocyanin (Pc), ferredoxin (Fd) and plastoquinone (PQ). Question marks indicate where the mechanism of electron transfer is still unclear [9].

redox changes of PS1 pigment P700. A typical example of the experimental kinetics of the ESR I signal is shown in Fig. 1.2.

In our previous work [3] we formulated a kinetic model with 26 ordinary differential equations for studying the mechanisms of dark P700 reduction kinetics at different concentrations of added ferredoxin. We were interested in the nature of the slow component of the signal. We used a bi-exponential fit to represent the results of numerical simulations. The numerical simulations showed that the fast component (characteristic
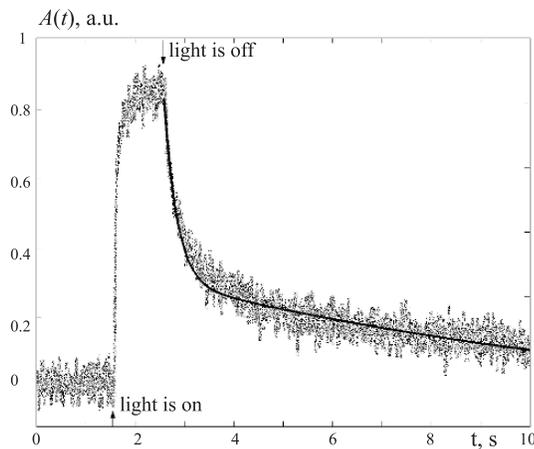
Fig. 1.2. Temporal evolution of the photoinduced ESR I signal from cation radical P700. Solid line is a bi-exponential fit to the experimental curve: $A(t) = A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t)$, where A_1 and A_2 are the amplitudes of the fast and slow components, respectively; k_1 and k_2 are their time constants.

time is about 0.2 s) represents cyclic electron transport. The rate of this fast phase was determined by the electron transfer rates of individual steps of cyclic electron transfer, the slowest of which was the oxidation of the plastoquinol molecule by cytochrome complex.

The nature of the slow phase (characteristic time is several seconds) was still unclear. As suggested by Scheller [6], the slow phase of P700 reduction reflects the ability of $P700^+$ to extract electrons from surrounding molecules, because this reaction was always present, even in the presence of oxygen. The slow phase of the reduction process could be described in the model by incorporating a large nonspecific electron pool, from which electrons required for the completion of $P700^+$ reduction may be taken.

As we will show below, the nature of the fast and slow components of the P700 reduction signal may be explained by means of 3D direct multiparticle simulation of cyclic electron transport in heterogeneous membrane systems, without hypothesis about the existence of the pool of nonspecific electron acceptors and donors. The results of 3D simulation visually display the role of spatial organization of the system in forming the kinetics of the P700 reduction signal.

1.2 Direct 3D Model

Recent data from electron and atomic-force microscopy reveal details of thylakoid membrane organization. We know [1] about the molecular structure of the protein complexes and mobile electron carriers as well as the architecture of the thylakoid membrane. Despite the advances in the study of the structure and function of individual components of the photosynthetic electron transport chain, there are still difficulties in understanding the coupling mechanisms between separate processes and the regulation of the entire system.

Experimental data on the spatial organization of the thylakoid membrane, kinetic data about the rate constants of single reactions, and the hypothesis about the mechanisms of regulation can be integrated in a direct 3D computer simulation model. The building of such a model became possible recently due to affordable powerful computer resources and the development of object-oriented programming methods and visualization.

Recently similar simulation methods of biochemical reactions were developed by S. Andrews and D. Bray [4] and J. Stiles and T. Bartol [5]. These methods allow simulation of biochemical reaction networks with spatial resolution and single molecule detail. The method from [4] was applied to the simulation of signal transduction in *Escherichia coli* chemotaxis [11], the method from [5] to the simulation of signaling in neuromuscular junctions.

In our previous study [3] we described the method for the direct 3D computer simulation of photosynthetic electron transport processes. This method was used to build a direct 3D model of ferredoxin-dependent cyclic electron transport around PS1. The model represented a 3D scene consisting of three compartments (thylakoid membrane,
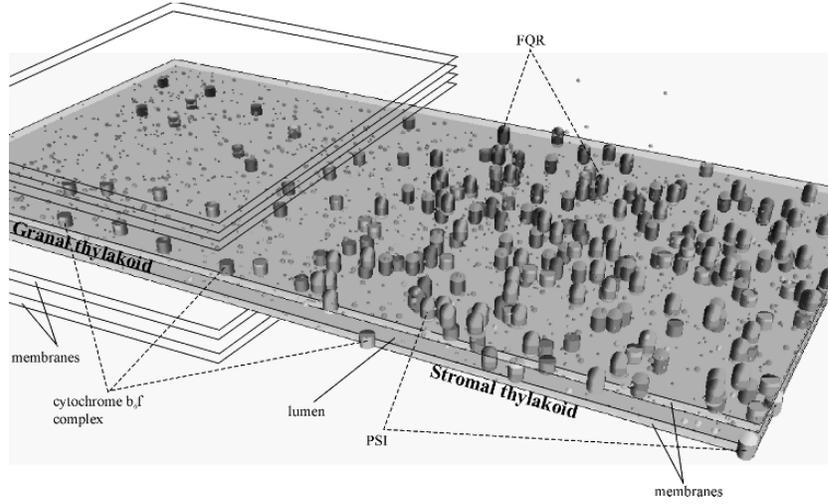
Fig. 1.3. Visualization of the 3D scene of the multiparticle model of cyclic electron transport. PS2 complexes are not shown, although they were simulated. One can see granal and stromal parts of thylakoid membrane, luminal and stromal spaces.

luminal space, and stroma) with protein complexes (PS1, PS2, cyt $b_6 f$, FQR) embedded in the membrane and mobile electron carriers (Pc, Fd, PQ), each in its own compartment.

In this study we further develop the direct 3D model of cyclic electron transport. The model represents two areas of thylakoid membrane, the granal area and the stromal area, so the model is spatially heterogenous (Fig. 1.3). Different types of complexes are located in different areas. PS1 is mostly found in the stromal area and PS2 in the granal area. Cyclic electron transport is likely to occur in stromal membrane areas [7].

In the direct 3D model movements of Pc, Fd, PQ in corresponding compartments (lumen, stroma, membrane) are simulated by the mathematical formalism of Brownian motion. We use the Langevin equation for the description of Brownian diffusion processes:

$$\xi \frac{dx}{dt} = f(t), \tag{1.1}$$

where ξ is the friction coefficient, and $f(t)$ is a random force. The random force has a normal distribution with mean 0 and variance $2kT\xi$ (where k is the Boltzmann constant, and T is temperature).

The mechanism of electron transfer is the following. If a mobile carrier moving by Brownian diffusion (chaotically) approaches a protein complex by a distance shorter than the effective radius of their interaction, the carrier docks to the complex with some probability. The probability and effective radius of interaction are parameters of the model (different for different types of complexes and mobile carriers). We can use kinetic data to estimate the effective radius of interaction and the probability.
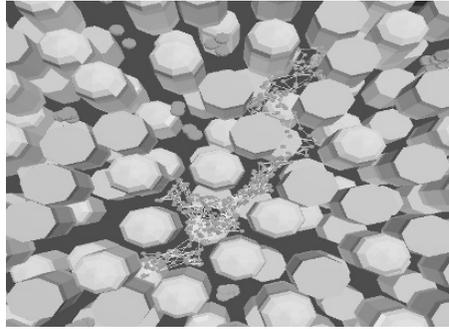
Fig. 1.4. A model trajectory of a PQ particle in a membrane with complexes PS1 and cyt b_6/f.

The concentrations and the sizes of protein complexes were taken from [2,7]. The PS1 particle density in a membrane was taken as 8.47×10^{-4} particles nm^{-2}, that of cyt $b_6 f$ was 3.5×10^{-4} particles nm^{-2}, and PS2 2.2×10^{-4} particles nm^{-2} [7]. The number of FQR complexes was assumed to be equal to the number of PS1 complexes. The PS1 size in the lateral plane was taken as 13 nm (with LHCI), cyt $b_6 f$ was 9 nm, and PS2 13 nm [2]. PS2 complexes are not shown in Fig. 1.5 although their presence was taken into account in simulations.

In the native thylakoid membrane and in the luminal space, free diffusion of the mobile carriers PQ and Pc is impossible because the membrane and the luminal space are narrow and full of the protein complexes protruding through the membrane. We compared the PQ diffusion coefficient in the membrane full of PS1 and cyt b_6/f complexes and the diffusion coefficient in a membrane without complexes. It turned out that if 1/3 of the membrane area is occupied with transmembrane complexes, then the PQ diffusion coefficient is ten times lower than in a case of free diffusion, which is in agreement with experiments [8]. The visualization of PQ diffusion trajectories shows the formation of PQ diffusion domains in a thylakoid membrane (Fig. 1.4).

Pc and Fd diffusion coefficients were taken as 10^{-10} m^2 s^{-1}, although the actual Pc diffusion coefficient was lower due to nonfree (restricted) diffusion in the lumen. The PQ diffusion coefficient was taken as 10^{-11} m^2 s^{-1}.

For estimation of the direct 3D model parameters (docking probabilities) we have simulated the processes of interaction of mobile carriers and complexes for particles in a solution (for example, PS1 and Pc particles or cyt b_6/f and plastocyanin particles).

1.3 Results and Discussion

We used a direct 3D model for the numerical simulations of cyclic electron flow around PS1. The time step was taken as 100 ns. At the initial moment of time all the P700 and Pc were reduced. In simulation the light was turned on for 1.5 s (saturating illumination). Then the P700 redox turnover was observed. During the illumination the PQ pool was partly reduced in the stromal part of the membrane. Reduced molecules of
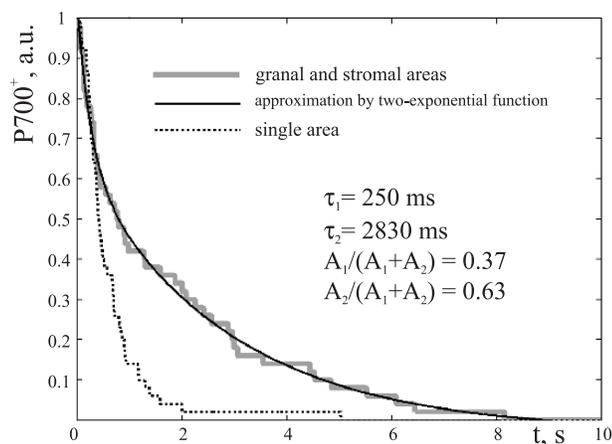
Fig. 1.5. Results of multiparticle simulation of dark $P700^+$ reduction. Thick gray line is a $P700^+$ reduction curve in the presence of the two areas of the thylakoid membrane (granal and stromal areas). Dotted line represents homogenous distribution of all complexes in the single area. Solid thin line is a bi-exponential fit to the experimental curve: $A(t) = A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t)$, where A_1 and A_2 are the amplitudes of the fast and slow components, respectively; k_1 and k_2 are their time constants.

PQ distributed evenly between the stromal and granal parts of the membrane. After switching the light off PQ molecules reduced cytochrome b_6/f complex and plastocyanin in both the granal and stromal areas of the membrane. Then Pc diffused to PS1 particles and reduced them. In the stromal part of the membrane this dark P700 reduction was fast (characteristic time 200 ms), because in the stromal area the average distance between PS1 and cyt b_6/f is short (20 nm). This process corresponds to the fast phase of the P700 reduction curve (Fig. 1.5).

Plastocyanin and plastoquinone molecules located in the granal areas diffused longer distances to reach PS1 particles since PS1 particles are located only in the stromal areas. This corresponds to the slow phase of the P700 reduction curve (Fig. 1.5).

The multiparticle simulations showed that the slow phase of the kinetics of photooxidized $P700^+$ dark reduction at cyclic electron flow around PS1 is diffusion controlled and is determined by diffusion of reduced PQ and Pc molecules from the granal to stromal areas of the thylakoid membrane, whereas the fast component represents cyclic Fd-mediated electron transport.

The kinetics of $P700^+$ dark reduction is determined not only by the concentrations and redox states of reagents, but also by the spatial distribution of the reacting molecules, the geometry of the system and the rate of mobile carrier diffusion processes.

1.4 Conclusions

The simulation method presented here adequately describes the electron transfer processes in a spatially heterogeneous membrane of a chloroplast thylakoid. This method can be applied for the description of the functioning of a large number of macromolecules which interact in the heterogeneous interior of subcellular systems.

Acknowledgments

References

1. Malkin, R., Niyogi, K.: Photosynthesis. In: Buchanan, B., Gruissem, W., Jones, R. (eds) Biochemistry and Molecular Biology of Plants. Kluwer Academic Publishers (2000).
2. Staehelin, L.A., van der Staay, G.W.M.: Structure, Composition, Functional Organization and Dynamic Properties of Thylakoid Membranes. In: Ort, D.R., Yocum, Ch.F. (eds) Oxygenic Photosynthesis. The Light Reactions. Kluwer Academic Publishers (1996).
3. Kovalenko, I.B., Riznichenko, G.Yu.: Theoretical and experimental study of cyclic electron transport around photosystem 1. Biophysics, **48**, 614–623 (2003).
4. Andrews, S.S., Bray, D.: Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. Physical Biology, **1**, 137–151 (2004).
5. Stiles, J.R., Bartol, T.M.: Monte Carlo methods for simulating realistic synaptic microphysiology using MCell. In: Computational Neuroscience: Realistic Modeling for Experimentalists (ed.) E. De. Schutter. Boca Raton, FL: CRC Press (2001).
6. Scheller, H.V.: In vitro cyclic electron transport in barley thylakoids follows two independent pathways. Plant Physiol., **110**, 187–194 (1996).
7. Albertsson, P.-A.: A quantitative model of the domain structure of the photosynthetic membrane. TRENDS in Plant Science, **6**, 349–354 (2001).
8. Blackwell, M., Gibas, C., Gygax, S., Roman, D., Wagner, B.: The plastoquinone diffusion coefficient in chloroplasts and its mechanistic implications. Biochim. Biophys. Acta, **1183**, 533–543 (1994).
9. Bendall, D.S., Manasse, R.S.: Cyclic photophosphorylation and electron transport. Biochim. Biophys. Acta, **1229**, 23–38 (1995).
10. Krendeleva, T.E., Kukarskih, G.P., Timofeev, K.N., Ivanov, B.N., Rubin, A.B.: Ferredoxin-dependent cyclic electron transport in isolated thylakoids involves Fd-Nadph reductase. Dokl. Ross. Akad. Nauk., **379**, 1–4 (2001).
11. Lipkow, K., Andrews, S.S., Bray, D.: Simulated diffusion of phosphorylated CheY through the cytoplasm of *Escherichia coli*. J. Bacteriol., **187**(1), 45–53 (2005).

Selective Regulation of Protein Activity by Complex Ca^{2+} Oscillations: A Theoretical Study

Beate Knoke,[1] Marko Marhl,[2] and Stefan Schuster[1]

[1] Dept. of Bioinformatics, Faculty of Biology and Pharmaceutics, Friedrich-Schiller University of Jena, Ernst-Abbe-Platz 2, D-07743 Jena, Germany; knoke@minet.uni-jena.de, schuster@minet.uni-jena.de

[2] Dept. of Physics, Faculty of Education, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia; marko.marhl@uni-mb.si

Summary. Calcium oscillations play an important role in intracellular signal transduction. As a second messenger, Ca^{2+} represents a link between several input signals and several target processes in the cell. Whereas the frequency of simple Ca^{2+} oscillations enables a selective activation of a specific protein and herewith a particular process, the question arises of how at the same time two or more classes of proteins can be specifically regulated. The question is general and concerns the problem of how one second messenger can transmit more than one signal simultaneously (bow-tie structure of signalling). To investigate whether a complex Ca^{2+} signal like bursting, a succession of low-peak and high-peak oscillatory phases, could selectively activate different proteins, several bursting patterns with simplified square pulses were applied in a theoretical model. The results indicate that bursting Ca^{2+} oscillations allow a differential regulation of two different calcium-binding proteins, and hence, perform the desired function.

Key words: Bow-tie structure of signalling, calcium oscillations, bursting, decoding.

2.1 Introduction

Calcium ions regulate a variety of cellular processes, like muscle contraction, fertilization and liver metabolism [1, 2]. After stimulating a cell by an agonist the concentration of free cytosolic Ca^{2+} very often changes periodically in time. This phenomenon is known as Ca^{2+} oscillation [1]. These oscillations have been the subject of intense modelling studies [3, 4]. The information they transmit is mainly encoded in frequency [5–8], but the amplitude and temporal pattern also play a role [9–11].

Usually, the oscillatory Ca^{2+} signal results in a stationary effect, for example, fertilizing oocytes, generating an endocrine signal or enhancing the transcription of a gene. A central role in the decoding of Ca^{2+} signals is played by calmodulin in many cells, cf. [12]. By binding Ca^{2+}, calmodulin can activate other proteins, e.g., Ca^{2+}/calmodulin-dependent protein kinase type II (CaM kinase II, EC 2.7.1.123), cf. [13] and myosin light-chain kinase (EC 2.7.1.117), cf. [14]. There are also proteins

that are activated by Ca^{2+} without involvement of calmodulin, e.g., protein kinase C (EC 2.7.1.37), cf. [15].

Besides simple spike-like, regular Ca^{2+} oscillations, experimental data on Ca^{2+} dynamics also show more complex oscillatory patterns [7, 16, 17]. A succession of low-peak and high-peak oscillatory phases, known as bursting, is a common pattern. Bursting has been investigated in modelling studies of transmembrane potential oscillations in nerve cells [5, 18] and of Ca^{2+} oscillations [7, 19–22]. In electric bursting in neurons, often both the active and the quiescent phases involve several spikes. In Ca^{2+} bursting, in contrast, the active phase consists of only one large spike.

Larsen and Kummer [23] and Rozi and Jia [24] were the first to simulate the decoding of complex Ca^{2+} oscillations, on the basis of the models proposed by Kummer et al. [7] and Borghans et al. [19], respectively. Larsen and co-workers [23,25] showed that information could be encoded in the shape and complexity of Ca^{2+} oscillations. Considering cooperative, activatory binding of Ca^{2+} to two different effector enzymes, it was demonstrated that cooperativity enables enzymes to decode different Ca^{2+} dynamics into different enzyme activity.

Many signal transduction systems as well as metabolic systems consist of a structure where several inputs can influence several targets via only one or a few intermediary components. This architecture is called the bow-tie structure [26,27]. The question arises of how such an architecture can operate [25,28], and if multiple signals can be transmitted and decoded not only successively, but also simultaneously.

Here, we investigate how periodic bursting may transmit two independent signals simultaneously, like the selective activation of two Ca^{2+}-binding proteins. In particular a separate activation by spikes and secondary peaks is of interest. To explore which characteristics of the complex signal could be responsible for an independent regulation of low-peak and spike-activated targets, we analyse frequency decoding by taking into account that regular bursting oscillations are characterized by two inherent frequencies of spikes and secondary peaks. To separate the questions of generating and decoding of bursting Ca^{2+} oscillations, the oscillations were simulated by artificially generated square-shaped patterns. Such square-shaped pulses have also been used in experiments [6] and in simulations [29,30]. In other studies, artificially generated sinusoidal patterns have been considered [31].

2.2 Model Description

In the model two Ca^{2+}-binding proteins are taken into account. Both proteins are considered to be activated cooperatively. An example is provided by calmodulin, cf. [12]. Regarding that calmodulin is usually activated by 4 Ca^{2+} ions, we considered the first protein being activated by four Ca^{2+} ions ($y = 4$). The second protein is assumed to contain additional inhibitory Ca^{2+} binding sites, resulting in a bell-shaped activation curve, inhibited at higher Ca^{2+} concentrations. A bell-shaped Ca^{2+} dependency is reported, for example, for the interaction of Ca^{2+}/calmodulin with the edema factor (EF) factor (EC 4.6.1.1), an adenylate cyclase toxin secreted by *Bacillus anthracis* [32,33]. Whereas calmodulin with two bound Ca^{2+} ions activates EF, Ca^{2+} is also assumed to

$$x\ Ca^{2+}$$

$$Prot_2 \longleftarrow\!\!\!\!\longrightarrow Prot_2Ca_x$$

$$K_2$$

$$x\ Ca^{2+} \qquad\qquad x\ Ca^{2+}$$

$$K_I \qquad\qquad\qquad K_I$$

$$Prot_2Ca_{xI} \qquad\qquad Prot_2Ca_xCa_{xI}$$
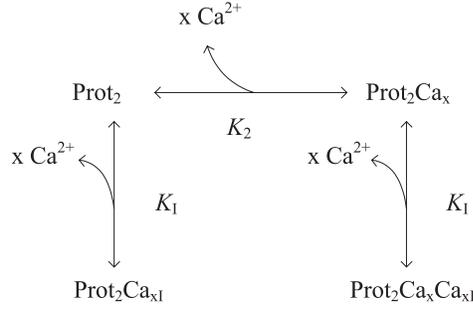
Fig. 2.1. Reaction scheme of Ca^{2+} binding to the protein with activatory and inhibitory Ca^{2+} binding sites ($Prot_2$). For the other protein ($Prot_1$) the inhibitory binding reactions are absent. K_2 and K_I are the reaction-associated dissociation constants.

interfere directly with the binding of Mg^{2+} to the catalytic metal binding site of EF, thereby inhibiting catalysis [32]. Therefore, we calculated with two activating and two inhibiting Ca^{2+} ions for protein 2 ($x = 2$).

For our model, the reaction scheme of Ca^{2+} binding to the protein with activatory and inhibitory Ca^{2+} binding sites ($Prot_2$) is presented in Fig. 2.1. The binding scheme for the other protein ($Prot_1$) is the same, however, the inhibitory binding reactions are lacking.

The inhibition type considered is noncompetitive, so the binding affinity of Ca^{2+} to the inhibitory site is independent of whether or not the activatory site is occupied. The two Ca^{2+}-binding proteins are considered as signalling proteins, which are supposed to have high rates of Ca^{2+} binding and dissociation [34,35], occur in low concentrations and not shape the Ca^{2+} signal. Therefore, we keep the total protein concentration so small that at most 10% of Ca^{2+} can be bound to proteins, assuming that the sequestration of Ca^{2+} by the two proteins can be neglected in the Ca^{2+} balance. Thus, no conservation relation for the amount of Ca^{2+} was included. However, a conservation relation holds for each protein. For protein 2, it reads

$$Prot_{2T} = Prot_2 + Prot_2Ca_{xI} + Prot_2Ca_x + Prot_2Ca_xCa_{xI}, \qquad (2.1)$$

where $Prot_{2T}$ denotes the total concentration of protein 2. Considering the mass action laws for the dissociation constants K_2 and K_I, the amount of active protein is given by the following rapid-equilibrium approximation:

$$Prot_2Ca_x = \frac{Prot_{2T} \times Ca^x}{(K_2 + Ca^x) \times \left(1 + \frac{Ca^x}{K_I}\right)}. \qquad (2.2)$$

For the solely Ca^{2+}-activated protein, the inhibition factor drops out, resulting in the well-known Hill equation for cooperative binding, cf. [36]:

$$Prot_1Ca_y = \frac{Prot_{1T} \times Ca^y}{K_1 + Ca^y}. \qquad (2.3)$$
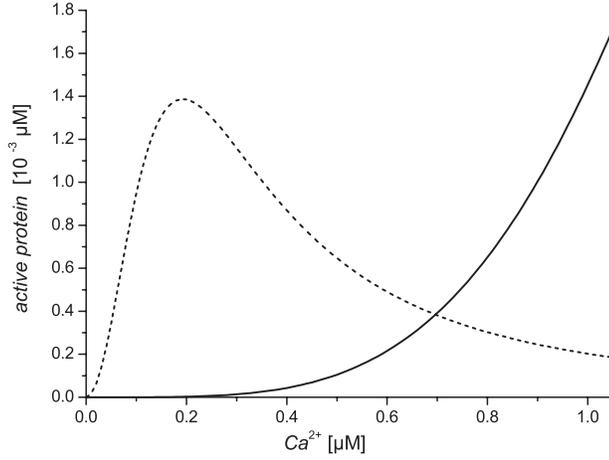
Fig. 2.2. Binding curves of the two proteins (Prot$_1$, solid line; Prot$_2$, dashed line). Protein activation is calculated by rapid-equilibrium approximation at constant Ca^{2+} and the following parameter values: K$_1$ = 5.88 μM^y, K$_2$ = 6.25$\times$10^{-2} μM^x, K$_I$ = 2.2$\times$10^{-2} μM^x. Total protein concentrations are Prot$_{1T}$ = 10^{-2} μM, Prot$_{2T}$ = 10^{-2} μM. All values remain the same in all calculations.

Fig. 2.2 shows the binding curves for both classes of proteins according to Eqs. (2.2) and (2.3).

The parameters concerning the Ca^{2+} association, dissociation, and inhibition of Ca^{2+} binding to proteins have been chosen such that the maxima of the binding curves (Fig. 2.2) are sufficiently separated.

In general, when the rapid-equilibrium conditions are not fulfilled, the protein activation is calculated by differential equations. If we take square-shaped Ca^{2+} pulses, where the Ca^{2+} concentration during each plateau is constant, the active protein concentrations can be calculated as follows:

$$\frac{\mathrm{d}\,Prot_2Ca_x}{\mathrm{d}t} = k_{\mathrm{on},2} \times Prot_2 \times Ca^x - k_{\mathrm{off},2} \times Prot_2Ca_x$$
$$- k_{\mathrm{on},I} \times Prot_2Ca_x \times Ca^x + k_{\mathrm{off},I} \times Prot_2Ca_xCa_{xI} \qquad (2.4)$$

$$\frac{\mathrm{d}\,Prot_2Ca_{xI}}{\mathrm{d}t} = k_{\mathrm{on},I} \times Prot_2 \times Ca^x - k_{\mathrm{off},I} \times Prot_2Ca_{xI} \qquad (2.5)$$

$$\frac{\mathrm{d}\,Prot_2Ca_xCa_{xI}}{\mathrm{d}t} = k_{\mathrm{on},I} \times Prot_2Ca_x \times Ca^x - k_{\mathrm{off},I} \times Prot_2Ca_xCa_{xI} \qquad (2.6)$$

$$\frac{\mathrm{d}\,Prot_1Ca_y}{\mathrm{d}t} = k_{\mathrm{on},1} \times Prot_1 \times Ca^y - k_{\mathrm{off},1} \times Prot_1Ca_y \qquad (2.7)$$

The bursting Ca^{2+} oscillation is simulated by a square pulse signal with spikes of two different heights (Fig. 2.3).

The bursting signal is characterized by the baseline, h_0, the height of the high and low spikes, h_1 and h_2, respectively, the duration of peaks and interpeak intervals, and
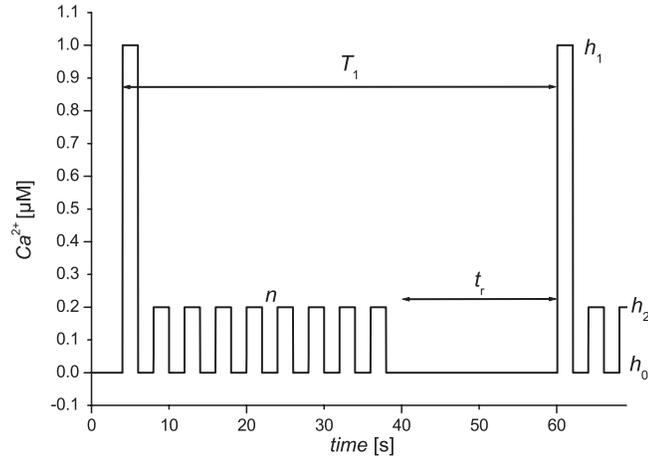
Fig. 2.3. Bursting Ca^{2+} oscillation used in all calculations. $h_0 = 0$ μM, $h_1 = 1.0$ μM, $h_2 = 0.2$ μM. T_1, t_r, and n are varied.

the number of low peaks occurring between two high peaks, n. According to experimental results [7,16,17], the number of high peaks per burst is set equal to one, and the refractory period between the main patterns is taken into account (t_r). The frequency (f_1) of high-peak oscillations is defined as the reciprocal value of the period (T_1):

$$f_1 = \frac{1}{T_1}. \tag{2.8}$$

For the low-peak oscillations, an averaged, effective frequency, $f_2{}^*$, is defined, counting the number of low peaks per period:

$$f_2^* = \frac{n}{T_1}. \tag{2.9}$$

2.3 Computational Information

The calculations for the rapid-equilibrium approximations were performed with the program MS Excel. The differential equations were solved numerically by using the software Madonna (University of Berkeley, CA) with the Rosenbrock (Stiff) integration method.

2.4 Results

Different bursting signals with varying parameter values are applied to the two model proteins described above. The rapid-equilibrium approximation (Eqs. (2.2) and (2.3)) is used for the first series of simulations, and in the second series of simulations, differential equations are used (Eqs. (2.4)–(2.7)). The binding curves for the two classes
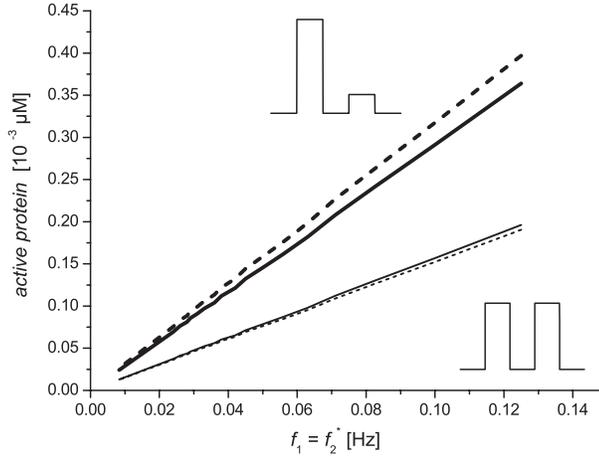
Fig. 2.4. Activation of $Prot_1$ (thick solid line) and $Prot_2$ (thick dashed line) vs. frequency $f_1 = f_2{}^*$. A bursting Ca^{2+} signal with $n = 1$ is used. The frequencies f_1 and $f_2{}^*$ are varied by different durations of the refractory period t_r. Thin solid and dashed lines represent the activation of $Prot_1$ and $Prot_2$, respectively, by simple spiking oscillations with the amplitude $0.7\ \mu M$ (see inset) vs. the frequency. For all calculations the rapid-equilibrium approximation is used. Parameter values are given in Fig. 2.2.

of proteins (Fig. 2.2) indicate that a selective regulation of proteins 1 and 2 is possible. Sole activation of protein 1 can be achieved by a signal with an amplitude corresponding to a high activation of protein 1 in a concentration range where protein 2 is already inhibited (see binding curves in Fig. 2.2). As the Ca^{2+} level in the cytosol is limited by about $1\ \mu M$ (cf. [5]), we chose this value for the amplitude of the high peaks, h_1. On the contrary, only protein 2 has a substantially elevated activity at $h_2 = 0.2\ \mu M$. In an oscillatory signal, the level of this activation can be regulated by changing the frequency of the corresponding constituent of the oscillation (low peaks vs. high peaks). For example, one protein can be gradually activated whereas the other protein remains nearly inactive if only one of the two frequencies is increased, keeping the other one constant and small. A gradual activation of protein 1, whereas protein 2 remains in a nearly inactive state, can be achieved by increasing the frequency f_1 (by shortening the period time T_1) and taking $n = 0$, thereby keeping $f_2{}^*$ constant. An independent regulation of protein 2 is achieved by a bursting oscillation upon increasing the number of low peaks, n, at constant period T_1, thus reducing more and more the refractory time t_r. Such a variation, which was observed experimentally [17], implies a variation of the average frequency of low peaks, $f_2{}^*$, while keeping frequency f_1 constant. An example of both effects of selective protein activation are depicted in [37].

To investigate whether one signal can gradually activate both proteins, we have analysed a bursting pattern with $n = 1$ (i.e., a 1:1 ratio of high and low peaks) and shortened the refractory period t_r. Thereby both frequencies f_1 and $f_2{}^*$ (which are equal in this case) are concomitantly increased (Fig. 2.4, thick lines). A simultaneous activation of both proteins is achieved. To compare the efficiency of the regulation
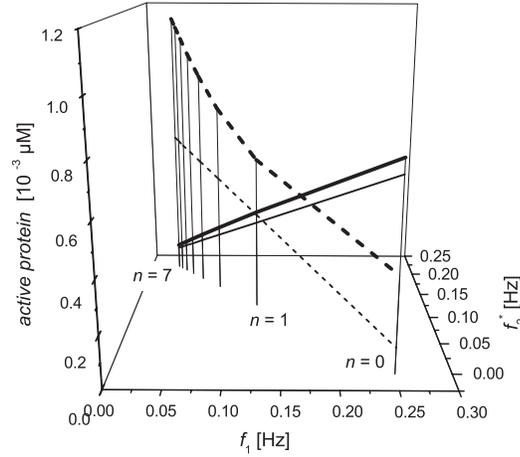
Fig. 2.5. Opposite regulation of Prot$_1$ (solid lines) and Prot$_2$ (dashed lines) by varying the frequency ratio, $n = f_2^*/f_1$, of the low and high peaks in the signal with variable period time T_1 without refractory time. Each vertical line (serving for 3D visualisation) corresponds to one value of n. The results were obtained by numerically integrating the differential equations (2.4)–(2.7) (thick lines) and rapid-equilibrium approximation (thin lines). For the dynamic simulations, the following kinetic constants are used: $k_{on,1} = 3\times10^{-3}$ s^{-1} μM^{-y}, $k_{off,1} = 0.01764$ s^{-1}, $k_{on,2} = 0.6$ s^{-1} μM^{-x}, $k_{off,2} = 0.0375$ s^{-1}, $k_{on,I} = 0.4$ s^{-1} μM^{-x}, $k_{off,I} = 8.8\times10^{-3}$ s^{-1}.

by bursting oscillations with a regulation by simple spiking oscillations, in Fig. 2.4, protein activation by an increasing frequency of spiking oscillations is also plotted (thin lines). The amplitude of simple spiking oscillations was set to 0.7 μM, corresponding to the intersection point of the two binding curves in Fig. 2.2. Note that the average Ca^{2+} level is then even higher than the Ca^{2+} level in the bursting signal. A simultaneous activation of both proteins is achieved more efficiently by bursting than by simple oscillations. This is understandable, as the high and low peaks in a bursting pattern correspond to the activation maxima of proteins 1 and 2, respectively. On the contrary, the peaks in simple spiking oscillations of 0.7 μM cannot coincide to both maxima simultaneously.

A simultaneous and selective up- and down-regulation of the two proteins can be achieved by increasing the number of low peaks n, thus prolonging the period time T_1 ($t_r = 0$). Thereby, the frequency, f_1, of high spikes decreases, whereas the frequency f_2^* increases, activating protein 2 and concomitantly deactivating protein 1 (see Fig. 2.5, thin lines). Moreover, as can be guessed from Fig. 2.5, the relationships between the average active protein concentrations and f_1 and f_2^* are linear. For a derivation see [37].

The experimental data show that for many Ca^{2+}-binding proteins, the average residence time of Ca^{2+} bound to the protein can take values from several microseconds to several seconds [34]. In the microseconds range, the kinetics is so fast that the rapid-equilibrium approximation can be justified while it may not be in the seconds range, depending on the oscillation period. For fast Ca^{2+}-binding signalling proteins, a rapid-

equilibrium approach was also considered in [35]. We now consider the case where the rate constants of binding and dissociation are not high enough to justify this approximation. In that case, the differential equations (2.4)–(2.7) should be used. Then, the time course of protein activity is on a nearly constant level after an initial transient. This is due to the slower dynamics of binding and dissociation in the differential equations. In particular, exponential decay of protein activity in the interspike intervals causes the smoothing effect. To see the effect of the dynamics, the protein activation curves obtained by both methods of calculation are compared in Fig. 2.5. Especially protein 2 is more efficiently activated by the slow kinetics than it would be activated by a fast kinetics which could be simulated by the rapid-equilibrium approximation.

2.5 Discussion

A mathematical model for the decoding of regular bursting Ca^{2+} oscillations has been proposed, based on the binding of cytosolic Ca^{2+} to two distinct proteins, both cooperatively binding Ca^{2+} at activatory sites with different binding constants and numbers of Ca^{2+} ions bound. Extending the work by Larsen and Kummer [23, 25], we include the assumption that one of the two proteins can cooperatively bind Ca^{2+}, in addition, at inhibitory sites. A biphasic regulation of protein activation at low Ca^{2+} levels and inhibition at high levels is known for the IP_3 receptor channel in the endoplasmic reticulum membrane [38, 39]. Its activity curve has a bell shape and is, therefore, similar to the curve shown in Fig. 2.2. The IP_3 receptor is composed of four subunits, each containing one activatory and one inhibitory binding site for Ca^{2+} [19]. Although not likely to function as a decoder of Ca^{2+} oscillations, our model is inspired by the IP_3 receptor properties of activatory and inhibitory Ca^{2+}-binding. Shen et al. [32] showed for the Ca^{2+}/calmodulin-activated edema factor (EF) a bell-shaped Ca^{2+} dependency where EF inhibition is presumably due to direct Ca^{2+} interference at its catalytic metal binding site. Both effects occur at physiological Ca^{2+} concentrations. As two Ca^{2+} ions are sufficient for EF activation via calmodulin, EF is considerably activated before maximum activation is attained for most endogenous cellular calmodulin targets [40].

A reciprocal activation and inhibition oftwo Ca^{2+}-calmodulin-dependent target enzymes was shown by Cho et al. [41] and Lee et al. [42–44]. The enzymes nitric oxide synthase (NOS, EC 1.14.13.39) and NAD kinase (EC 2.7.1.23) were differentially activated by two soybean calmodulin isoforms SCaM-1 and SCaM-4: Plants contain several, partly divergent, CaM isoforms, some of them having different capabilites to activate target enzymes. While neuronal NOS (nNOS) is strongly activated by SCaM-4, its activation by SCaM-1 is only weak. A competitive inhibition of SCaM-4-activated NOS was observed by increasing the concentration of the weakly activating SCaM-1 isoform [41].

In contrast to the activation scheme of nNOS, plant NAD kinase is activated by the highly conserved SCaM-1, but not by the divergent soybean CaM isoform, SCaM-4 [42]. Furthermore, Lee et al. [43, 44] indicate SCaM-4 acting as a competitive antagonist of NAD kinase. Therefore, SCaM-1 activates NAD kinase and competitively inhibits NOS while SCaM-4 activates NOS and competitively inhibits CaN ([41,44]).

Although these experiments were conducted with neuronal NOS, both plant and neuronal nitric oxide synthase are activated by Ca^{2+}-dependent CaM [45,46]. Weissmann et al. [47] indicate that four calcium ions have to be bound to CaM to activate neuronal NOS. The activation of a plant-specific NOS enzyme by SCaM-1 and -4 has not yet been studied to our knowledge.

Nitric oxide synthase catalyzes the production of nitric oxide (NO), an important second messenger: In plants, pathogen infection induces a Ca^{2+}-dependent activation of NOS, resulting in NO-mediated defense gene expression and programmed cell death [45,46,48,49]. Cytosolic Ca^{2+} elevations are one of the earliest events in pathogen-challenged cells [50,51]. NAD kinase catalyses the phosphorylation of NAD to NADP [43], which may indirectly contribute to the production of reactive oxygen species (ROS), involved in Ca^{2+}-mediated plant disease response [50]. Whereas the expression of some defense-related genes can be mediated solely by NO [49], induction of host cell death requires synergistic action of both NO and ROS [48,52]. The findings support a model for specific and concerted roles of Ca^{2+}-activated CaM isoforms in plant defense response against pathogens, in which some CaM isoforms mediate ROS increases, whereas other CaM isoforms activate defense gene expression [53].

Elaborating on results by Larsen and Kummer [25], here we provide theoretical evidence that bursting Ca^{2+} oscillations can perform the function of simultaneous transmission of two signals, which enables differential regulation of two proteins, and hence selective regulation of two cellular processes. We show that the selective activation of proteins can be achieved by adjusting the two inherent frequencies of the investigated bursting pattern, which are connected to the relative occurrence of the low and high peaks. These frequencies can be regulated independently or in a correlated way, depending on how the number of low peaks and/or period time are changed. So, the two proteins can even be regulated in the opposite way.

Frequency encoding is considered to be more robust to noise than amplitude encoding [5,54]. In the case of bursting, however, no sharp distinction can be made between encoding by frequency and amplitude. A change in the frequency ratio of high and low peaks might also be regarded as a change in the amplitudes.

In summary, the key result of this study is that a selective regulation of different cellular processes is possible by bursting Ca^{2+} signals. This is in support of the "bow-tie" concept of signalling [26]. Recently, also another possibility of a selective regulation of cellular processes by Ca^{2+} signals has been demonstrated via protein cascades or frequencies of time-limited oscillations: A bell-shaped and separate activation of two Ca^{2+}-binding proteins with distinct velocities of Ca^{2+}-binding and dissociation is possible by applying different frequencies of the calcium signal, modelled by differential equations. Essential for this phenomenon is a limited number of calcium spikes [55,56]. It is hoped that future experimental studies will allow us to check the physiological relevance of these theoretical predictions.

Acknowledgments

References

1. Berridge, M.J., Bootman, M.D., Lipp, P.: Calcium–a life and death signal. Nature, **395**, 645–648 (1998).
2. Dupont, G., Swillens, S., Clair, C., Tordjmann, T., Combettes, L.: Hierarchical organization of calcium signals in hepatocytes: from experiments to models. Biochim. Biophys. Acta, **1498**, 134–152 (2000).
3. Falcke, M.: Reading the patterns in living cells—the physics of Ca^{2+} signaling. Adv. Phys., **53**, 255–440 (2004).
4. Schuster, S., Marhl, M., Höfer, T.: Modelling of simple and complex calcium oscillations. From single-cell responses to intercellular signalling. FEBS J., **269**, 1333–1355 (2002).
5. Goldbeter, A.: Biochemical Oscillations and Cellular Rhythms. Cambridge University Press, Cambridge (1996).
6. De Koninck, P., Schulman, H.: Sensitivity of CaM kinase II to the frequency of Ca^{2+} oscillations. Science, **279**, 227–230 (1998).
7. Kummer, U., Olsen, L.F., Dixon, C.J., Green, A.K., Bornberg-Bauer, E., Baier, G.: Switching from simple to complex oscillations in calcium signaling. Biophys. J., **79**, 1188–1195 (2000).
8. Dupont, G., Houart, G., De Koninck, P.: Sensitivity of CaM kinase II to the frequency of Ca^{2+} oscillations: a simple model. Cell Calcium, **34**, 485–497 (2003).
9. Dolmetsch, R.E., Lewis, R.S., Goodnow, C.C., Healy, J.I.: Differential activation of transcription factors induced by Ca^{2+} response amplitude and duration. Nature, **386**, 855–858 (1997).
10. Prank, K., Gabbiani, F., Brabant, G.: Coding efficiency and information rates in transmembrane signaling. Biosystems, **55**, 15–22 (2000).
11. Tompa, P., Toth-Boconadi, R., Friedrich, P.: Frequency decoding of fast calcium oscillations by calpain. Cell Calcium, **29**, 161–170 (2001).
12. Van Eldik, L.J., Watterson, D.M.: Calmodulin and Signal Transduction. Academic Press, London (1998).
13. Colbran, R.J.: Targeting of calcium/calmodulin-dependent protein kinase II. Biochem. J., **378**, 1–16 (2004).
14. Stull, J.T., Tansey, M.G., Tang, D.C., Word, R.A., Kamm, K.E.: Phosphorylation of myosin light chain kinase: a cellular mechanism for Ca^{2+} desensitization. Mol. Cell. Biochem., **127–128**, 229–237 (1993).
15. Webb, B.L.J., Hirst, S.J., Giembycz, M.A.: Protein kinase C isoenzymes: a review of their structure, regulation and role in regulating airways smooth muscle tone and mitogenesis. Brit. J. Pharmacol., **130**, 1433–1452 (2000).
16. Dixon, C.J., Woods, N.M., Cuthbertson, K.S.R., Cobbold, P.H.: Evidence for two Ca^{2+}-mobilizing purinoceptors on rat hepatocytes. Biochem. J., **269**, 499–502 (1990).
17. Dixon, C.J.: Evidence that 2-methylthioATP and 2-methylthioADP are both agonists at the rat hepatocyte $P2Y_1$ receptor. Brit. J. Pharmacol., **130**, 664–668 (2000).

18. Izhikevich, E.M.: Neural excitability, spiking and bursting. Int. J. Bifurcat. Chaos, **10**, 1171–1266 (2000).
19. Borghans, J.A.M., Dupont, G., Goldbeter, A.: Complex intracellular calcium oscillations. A theoretical exploration of possible mechanisms. Biophys. Chem., **66**, 25–41 (1997).
20. Houart, G., Dupont, G., Goldbeter, A.: Bursting, chaos and birhythmicity originating from self-modulation of the inositol 1,4,5,-trisphosphate signal in a model for intracellular Ca^{2+} oscillations. Bull. Math. Biol., **61**, 507–530 (1999).
21. Marhl, M., Haberichter, T., Brumen, M., Heinrich, R.: Complex calcium oscillations and the role of mitochondria and cytosolic proteins. BioSystems, **57**, 75–86 (2000).
22. Perc, M., Marhl, M.: Different types of bursting calcium oscillations in non-excitable cells. Chaos Sol. Fract., **18**, 759–773 (2003).
23. Larsen, A.Z., Kummer, U.: Information processing in calcium signal transduction. In: Falcke, M., Malchow, D. (Eds.) Understanding Calcium Dynamics: Experiments and Theory, Lecture Notes in Physics, Vol. 623. Springer, Berlin, pp. 153–178 (2003).
24. Rozi, A., Jia, Y.: A theoretical study of effects of cytosolic Ca^{2+} oscillations on activation of glycogen phosphorylase. Biophys. Chem., **106**, 193–202 (2003).
25. Larsen, A.Z., Olsen, L.F., Kummer, U.: On the encoding and decoding of calcium signals in hepatocytes. Biophys. Chem., **107**, 83–99 (2004).
26. Csete, M., Doyle, J.: Bow ties, metabolism and disease. Trends Biotechnol., **22**, 446–450 (2004).
27. Ma, H.W., Zeng, A.P.: The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics, **19**, 1423–1430 (2003).
28. Somsen, O.J, Siderius, M., Bauer, F.F., Snoep, J.L., Westerhoff, H.V.: Selectivity in overlapping MAP kinase cascades. J. Theor. Biol., **218**, 343–354 (2002).
29. Li, Y.-X., Goldbeter, A.: Pulsatile signaling in intercellular communication. Biophys. J., **61**, 161–171 (1992).
30. Salazar, C., Politi, A., Höfer, T.: Decoding of calcium oscillations by phosphorylation cycles. Genome Informatics Series, **15**, 50–51 (2004).
31. Gall, D., Baus, E., Dupont, G.: Activation of the liver glycogen phosphorylase by Ca^{2+} oscillations: a theoretical study. J. Theor. Biol., **207**, 445–454 (2000).
32. Shen, Y., Lee, Y.S., Soelaiman, S., Bergson, P., Lu, D., Chen, A., Beckingham, K., Grabarek, Z., Mrksich, M., Tang, W.-J.: Physiological calcium concentrations regulate calmodulin binding and catalysis of adenylyl cyclase exotoxins. EMBO J., **21**, 6721–6732 (2002).
33. Soelaiman, S., Wei, B.Q., Bergson, P., Lee, Y.-S., Shen, Y., Mrksich, M., Shoichet, B.K., Tang, W.-J.: Structure-based inhibitor discovery against adenylyl cyclase toxins from pathogenic bacteria that cause anthrax and whooping cough. J. Biol. Chem., **278**, 25990–25997 (2003).
34. Falke, J.J., Drake, S.K., Hazard A.L., Peersen, O.B.: Molecular tuning of ion binding to calcium signaling proteins. Quart. Rev. Biophys., **27**, 219–290 (1994).
35. Marhl, M., Schuster, S., Brumen, M., Heinrich, R.: Modelling oscillations of calcium and endoplasmatic reticulum transmembrane potential. Role of the signalling and buffering proteins and of the size of the Ca^{2+} sequestering ER subcompartments. Bioelectrochem. Bioenerg., **46**, 79–90 (1998).
36. Segel, I.H.: Enzyme Kinetics. Wiley, New York (1993).
37. Schuster, S., Knoke, B., Marhl, M.: Differential regulation of proteins by bursting calcium oscillations—a theoretical study. BioSystems, **81**, 49–63 (2005).
38. Bezprozvanny, I., Watras, J., Ehrlich, B.E.: Bell-shaped calcium-response curves of Ins(1,4,5)P$_3$- and calcium-gated channels from endoplasmic reticulum of cerebellum. Nature, **351**, 751–754 (1991).

39. De Young, G.W., Keizer, J.: A single-pool inositol 1,4,5-trisphosphate-receptor-based model for agonist-stimulated oscillations in Ca^{2+} concentration. Proc. Nat. Acad. Sci. USA, **89**, 9895–9899 (1992).

40. Ulmer, T.S., Soelaiman, S., Li, S., Klee, C.B., Tang, W.-J., Bax, A.: Calcium dependence of the interaction between calmodulin and anthrax edema factor. J. Biol. Chem., **278**, 29261–29266 (2003).

41. Cho, M.J., Vaghy, P.L., Kondo, R., Lee, S.H., Davis, J.P., Rehl, R., Heo, W.D., Johnson, J.D.: Reciprocal regulation of mammalian nitric oxide synthase and calcineurin by plant calmodulin isoforms. Biochemistry, **37**, 15593–15597 (1998).

42. Lee, S.H., Kim, J.C., Lee, M.S., Heo, W.D., Seo, H.Y., Yoon, H.W., Hong, J.C., Lee, S.Y., Bahk, J.D., Hwang, I., Cho, M.J.: Identification of a novel divergent calmodulin isoform from soybean which has differential ability to activate calmodulin-dependent enzymes. J. Biol. Chem., **270**, 21806–21812 (1995).

43. Lee, S.H., Seo, H.Y., Kim, J.C., Heo, W.D., Chung, W.S., Lee, K.J., Kim, M.C., Cheong, Y.H., Choi, J.Y., Lim, C.O., Cho, M.J.: Differential activation of NAD kinase by plant calmodulin isoforms. J. Biological Chemistry, **272**, 9252–9259 (1997).

44. Lee, S.H., Johnson, J.D., Walsh, M.P., van Lierop, J.E., Sutherland, C., Xu, A., Snedden, W.A., Kosk-Kosicka, D., Fromm, H., Narayanan, N., Cho, M.J.: Differential regulation of Ca^{2+}/calmodulin-dependent enzymes by plant calmodulin isoforms and free Ca^{2+} concentration. J. Biol. Chem., **350**, 299–306 (2000).

45. Crawford, N.M., Guo, F.-Q.: New insights into nitric oxide metabolism and regulatory functions. Trends Plant Sci., **10**, 195–200 (2005).

46. del Rio, L.A., Corpasa, F.J., Barroso, J.B.: Nitric oxide and nitric oxide synthase activity in plants. Phytochemistry, **65**, 783–792 (2004).

47. Weissman, B.A., Jones, C.L., Liu, Q., Gross, S.S.: Activation and inactivation of neuronal nitric oxide synthase: characterization of Ca^{2+}-dependent [^{125}I]Calmodulin binding. Eur. J. Pharmacol., **435**, 9–18 (2002).

48. Delledonne, M., Xia, Y., Dixon, R.A., Lamb, C.: Nitric oxide functions as a signal in plant disease resistance. Nature, **394**, 585–588 (1998).

49. Durner, J., Wendehenne, D., Klessig, D.F.: Defense gene induction in tobacco by nitric oxide, cyclic GMP, and cyclic ADP-ribose. Proc. Nat. Acad. Sci. USA, **95**, 10328–10333 (1998).

50. Bouche, N., Yellin, A., Snedden, W.A., Fromm, H.: Plant-specific calmodulin-binding proteins. Annu. Rev. Plant Biol., **56**, 435–466 (2005).

51. White, P.J., Broadley, M.R.: Calcium in plants. Annals of Botany, **92**, 487–511 (2003).

52. Nimchuk, Z., Eulgem, T., Holt, B.F., Dangl, J.L.: Recognition and response in the plant immune system. Annu. Rev. Genet., **37**, 579–609 (2003).

53. Heo, W.D., Lee, S.H., Kim, M.C., Kim, J.C., Chung, W.S., Chun, H.J., Lee, K.J., Park, C.Y., Park, H.C., Choi, J.Y., Cho, M.J.: Involvement of specific calmodulin isoforms in salicylic acid-independent activation of plant disease resistance responses. Proc. Nat. Acad. Sci. USA, **96**, 766–771 (1999).

54. Berridge, M.J.: The AM and FM of calcium signalling. Nature, **386**, 759–760 (1997).

55. Marhl, M., Perc, M., Schuster, S.: Selective regulation of cellular processes via protein cascades acting as band-pass filters for time-limited oscillations. FEBS Lett., **579**, 5461–5465 (2005).

56. Marhl, M., Perc, M., Schuster, S.: A minimal model for decoding of time-limited Ca^{2+} oscillations. Biophys. Chem., **120**, 161–167 (2006).

3

Phase Separation in Eukaryotic Directional Sensing

Andrea Gamba,[1] Antonio de Candia,[2] Stefano Di Talia,[3] Antonio Coniglio,[2] Federico Bussolino,[4] and Guido Serini[4]

[1] Department of Mathematics and CNISM, Politecnico di Torino, and INFN–Unit of Turin, 10129 Torino, Italy; andrea.gamba@polito.it
[2] Department of Physical Sciences, University of Naples "Federico II" and INFM–Unit of Naples, 80126, Naples, Italy
[3] Laboratory of Mathematical Physics, The Rockefeller University, New York, NY 10021, USA
[4] Department of Oncological Sciences and Division of Molecular Angiogenesis, IRCC, Institute for Cancer Research and Treatment, University of Torino School of Medicine, 10060 Candiolo (TO), Italy; guido.serini@ircc.it

Summary. Many eukaryotic cell types share the ability to migrate directionally in response to external chemoattractant gradients. The binding of chemoattractants to specific receptors leads to a wide range of biochemical responses that become highly localized as cells polarize and migrate by chemotaxis. This ability is central in the development of complex organisms, and is the result of a billion years of evolution. Cells exposed to shallow gradients in chemoattractant concentration respond with strongly asymmetric accumulation of several factors, including the phosphoinositides PIP_3 and PIP_2, the PI 3-kinase PI3K, and phosphatase PTEN. This early symmetry-breaking stage is believed to trigger effector pathways leading to cell movement. Although many signaling factors implied in directional sensing have been recently discovered the physical mechanism of signal amplification is not yet well understood. We propose that directional sensing is the consequence of a phase ordering process mediated by phosphoinositide diffusion and driven by the distribution of the chemotactic signal. By studying a realistic reaction-diffusion lattice model that describes PI3K and PTEN enzymatic activity, recruitment to the plasma membrane, and diffusion of their phosphoinositide products, we have shown that the effective enzyme-enzyme interaction induced by catalysis and diffusion introduces an instability of the system towards phase separation for realistic values of physical parameters. In this framework, large reversible amplification of shallow chemotactic gradients, selective localization of chemical factors, macroscopic response timescales, and spontaneous polarization arise naturally.

Key words: Directional sensing, first-order phase transitions.

3.1 Introduction

A wide variety of eukaryotic cells exhibit the capacity to respond and migrate directionally in response to external gradients. This behavior is essential for a variety of
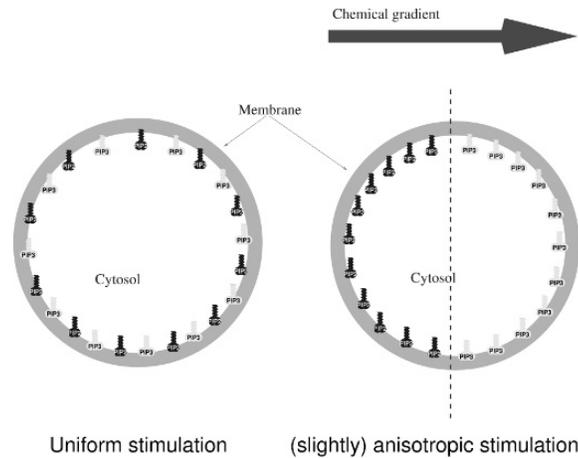
Fig. 3.1. Under slightly anisotropic chemoattractant stimulation a strong and sharp separation in PIP_2- and PIP_3-rich phases is observed at the level of the cell plasma membrane [13], realizing a powerful and efficient amplification of the external chemotactic signal.

processes including angiogenesis, nerve growth, wound healing, and embryogenesis. Perhaps the most distinguished chemotactic response is exemplified by neutrophils as they navigate to sites of inflammation. When exposed to an attractant gradient, these cells quickly orient themselves and move using anterior pseudopod extension together with posterior contraction and retraction. This highly regulated amoeboid motion can be achieved in the presence of very shallow attractant gradients. The signaling factors responsible for this complex behavior are now beginning to emerge. The general picture emerging from the analysis of chemotaxis in different eukaryotic cell types indicates that, in the process of directional sensing, a shallow extracellular gradient of chemoattractant is translated into an equally shallow gradient of receptor activation [13] that in turn induces the recruitment of the cytosolic enzyme phosphatidylinositol 3-kinase (PI3K) to the plasma membrane, where it phosphorylates PIP_2 into PIP_3.

However, phosphoinositide distribution does not simply mirror the receptor activation gradient, but rather a strong and sharp separation in PIP_2- and PIP_3-rich phases arises [13], realizing a powerful and efficient amplification of the external chemotactic signal (Fig. 3.1). PIP_3 acts as a docking site for effector proteins that induce cell polarization [3], and eventually cell motion [10] (Fig. 3.2). Cell polarization can be decoupled from directional sensing by the use of inhibitors of actin polymerization so that cells are immobilized, but respond with the same signal amplification of untreated cells [8]. The action of PI3K is counteracted by the phosphatase PTEN that dephosphorylates PIP_3 into PIP_2 [13]. PTEN localization at the cell membrane depends upon the binding to PIP_2 of its first 16 N-terminal amino acids [7].
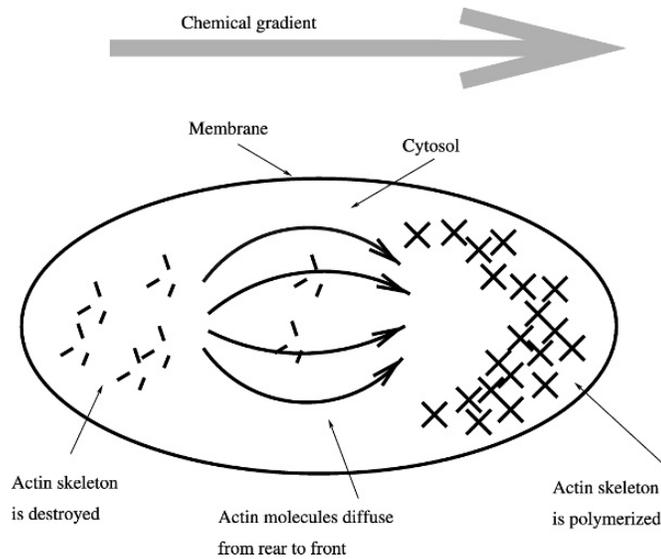
Fig. 3.2. One of the basic features of cell movement is the treadmill-like effect realized by systems of actin polymers. Actin forms an extensive network of proteins at the leading edge of cells. Actin polymerization allows the cell to move forward. PIP_3 plasma membrane accumulation induces actin polymerization in the cell anterior, while the actin cytoskeleton is depolymerized in the posterior, thus inducing cell motion. PIP_3 plasma membrane accumulation is induced by anisotropies in the concentration of extracellular attractant factors as low as 2% from rear to front.

3.2 A Phase Separation Process

In physical terms, the process of directional sensing shows the characteristic phenomenology of phase separation [11]. However, it is not clear which mechanism could be responsible for it. In known physical models, such as binary alloys, phase separation is the consequence of some kind of interaction among the constituents of a system, which can favor their segregation in separated phases [12]. However, one can show [5] that, even in the absence of direct enzyme-enzyme or phosphoinositide-phosphoinositide interactions, catalysis and phosphoinositide diffusion mediate an effective interaction among enzymes, which is sufficient to drive the system towards phase separation. To this aim, we have simulated the kinetics of the network of chemical reactions that represents the ubiquitous biochemical backbone of the directional sensing module. Since the chemical system is characterized by extremely low concentrations of chemical factors and evolution takes place out of equilibrium, we used a stochastic approach [2,6]. Indeed, rare, large fluctuations are likely to be relevant for kinetics in the presence of unstable or metastable states. Simulated reactions and diffusion processes taking place in the inner face of the cell plasma membrane are

1. $PI3K(cytosol)+Rec(i) \rightleftharpoons PI3K \cdot Rec(i)$
2. $PTEN(cytosol)+PIP_2(i) \rightleftharpoons PTEN \cdot PIP_2(i)$
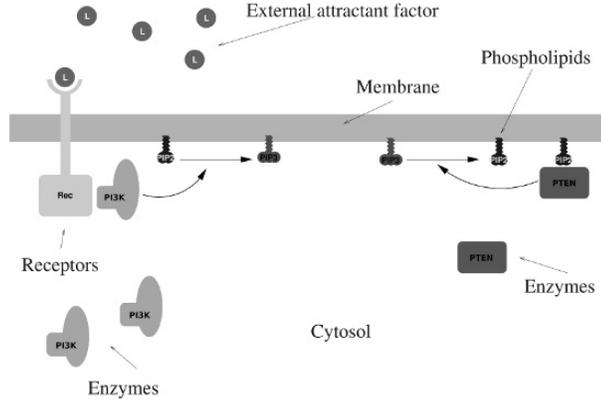
Fig. 3.3. Biochemical scheme of the simulated reaction network.

3. PI3K·Rec(i)+PIP$_2$(i) $\rightarrow$ PI3K·Rec(i)+PIP$_3$(i)
4. PTEN·PIP$_2$(i)+PIP$_3$(i) $\rightarrow$ PTEN·PIP$_2$(i)+PIP$_2$(i)
5. PIP$_2$(i) $\rightarrow$ PIP$_2$(j)
6. PIP$_3$(i) $\rightarrow$ PIP$_3$(j),

where index i represents a generic plasma membrane site and j one of its nearest neighbors (see also Fig. 3.3). The probability of performing a simulated reaction on a given site is proportional to realistic kinetic reaction rates and local reactant concentrations (Tables 3.1, 3.2).

The plasma membrane is represented as a spherical surface of radius $R = 10$ μm partitioned in $N_s = 10242$ computational sites. The cell cytosol is represented as an unstructured reservoir containing a variable number of PI3K and PTEN enzymes, which can bind and unbind to the cell membrane according to the rules described in Table 3.1. Chemical factors localized in the cytosol are indicated in Table 3.1 with the corresponding subscript, while factors attached to the membrane are indicated with a subscript representing the membrane site where they are localized. PIP$_2$ and PIP$_3$ molecules are assumed to freely diffuse on the cell membrane with the diffusion coefficient D specified in Table 3.2. The surface diffusivity of PI3K and PTEN molecules bound to phosphoinositides is neglected, since it is expected to be much less than the diffusivity of free phosphoinositides. The reaction-diffusion kinetics is simulated according to Gillespie's method [6], generalized to the case of an anisotropic environment.

A convenient order parameter measuring the degree of phase separation of the phosphoinositide mixture is Binder's cumulant [1]

$$g = \frac{1}{2}\left(3 - \frac{\langle(\varphi - \langle\varphi\rangle)^4\rangle}{\langle(\varphi - \langle\varphi\rangle)^2\rangle^2}\right)$$

where $\varphi = \varphi_i = [\text{PIP}_3]_i - [\text{PIP}_2]_i$ is a difference of local concentrations on site i and $\langle\cdots\rangle$ denotes average over many different random realizations.

Table 3.1. Probabilities of chemical reactions and diffusion processes. Let X·Y denote the bound state of species X and Y, [X] the global concentration of species X in the whole cell, $[X]_{\text{cyto}}$ the cytosolic concentration, and $[X]_i$ the local concentration on plasma membrane site i. The rate for a given reaction on site i is denoted by f_i, V is the cell volume, $\sum'$ denotes sum over nearest neighbors, and $(x)_+ = x$ for positive x and 0 otherwise. Time is advanced as a Poisson process of intensity equal to the reciprocal of the sum of the frequencies for all the processes. The simulations were performed using the values for kinetic rates and Michaelis–Menten constants given in Table 3.2.

Reaction	f_i
PI3K(cytosol)+Rec(i) $\rightarrow$ PI3K·Rec(i)	$\frac{V}{N_s} k_{\text{ass}}^{\text{Rec}} [\text{Rec}]_i [\text{PI3K}]_{\text{cyto}}$
PI3K(cytosol)+Rec(i) $\leftarrow$ PI3K·Rec(i)	$\frac{1}{N_s} k_{\text{diss}}^{\text{Rec}} [\text{Rec} \cdot \text{PI3K}]_i$
PTEN(cytosol)+PIP$_2$(i) $\rightarrow$ PTEN·PIP$_2$(i)	$\frac{V}{N_s} k_{\text{ass}}^{\text{PIP}_2} [\text{PIP}_2]_i [\text{PTEN}]_{\text{cyto}}$
PTEN(cytosol)+PIP$_2$(i) $\leftarrow$ PTEN·PIP$_2$(i)	$\frac{1}{N_s} k_{\text{diss}}^{\text{PIP}_2} [\text{PIP}_2 \cdot \text{PTEN}]_i$
PI3K·Rec(i)+PIP$_2$(i) $\rightarrow$ PI3K·Rec(i)+PIP$_3$(i)	$k_{\text{cat}}^{\text{PI3K}} \frac{[\text{Rec·PI3K}]_i [\text{PIP}_2]_i}{K_{\text{M}}^{\text{PI3K}}+[\text{PIP}_2]_i}$
PTEN·PIP$_2$(i)+PIP$_3$(i) $\rightarrow$ PTEN·PIP$_2$(i)+PIP$_2$(i)	$k_{\text{cat}}^{\text{PTEN}} \frac{[\text{Rec·PTEN}]_i [\text{PIP}_3]_i}{K_{\text{M}}^{\text{PTEN}}+[\text{PIP}_3]_i}$
PIP$_2$(i)$\rightarrow$PIP$_2$(j)	$\frac{D}{\sqrt{3}S_{\text{site}}} \sum' \left([\text{PIP}_2]_i - [\text{PIP}_2]_j\right)_+$
PIP$_3$(i)$\rightarrow$PIP$_3$(j)	$\frac{D}{\sqrt{3}S_{\text{site}}} \sum' \left([\text{PIP}_3]_i - [\text{PIP}_3]_j\right)_+$

Table 3.2. Physical and kinetic parameters used in the simulations

Parameter	Value	Parameter	Value
R	10.00 µm	$k_{\text{cat}}^{\text{PI3K}}$	1.00 s^{-1}
[Rec]	0.00–50.00 nM	$k_{\text{cat}}^{\text{PTEN}}$	0.50 s^{-1}
[PI3K]	50.00 nM	$K_{\text{M}}^{\text{PI3K}}$	200.00 nM
[PTEN]	50.00 nM	$K_{\text{M}}^{\text{PTEN}}$	200.00 nM
[PIP$_2$]	500.00 nM	$k_{\text{ass}}^{\text{Rec}}$	50.00 (s µM)$^{-1}$
D	0.10–1.00 µm^2/s	$k_{\text{ass}}^{\text{PIP}_2}$	50.00 (s µM)$^{-1}$
$k_{\text{diss}}^{\text{Rec}}$	0.10 s^{-1}	$k_{\text{diss}}^{\text{PIP}_2}$	0.10 s^{-1}

Phosphoinositide separation into distinct phosphoinositide clusters corresponds to g of order 1. Spontaneous phase symmetry breaking leads to the formation of PIP_2-, PIP_3-rich clusters of different size. Cluster sizes can be characterized by harmonic analysis. For each realization, the fluctuations $\delta\varphi = \varphi - \langle\varphi\rangle$ of the φ field can be expanded in spherical harmonics. Let us consider the two-point correlation functions $\langle\delta\varphi(\mathbf{u})\delta\varphi(\mathbf{u}')\rangle = \sum_{l=1}^{+\infty} C_l P_l(\mathbf{u}\cdot\mathbf{u}')$ where P_l are Legendre polynomials. When most of the weight is concentrated on the lth harmonic component, average phosphoinositide clusters extend over the characteristic length $\pi R/2l$.

3.3 Dynamic phase diagram

We have run many random realizations of the system for different (ρ, D) pairs, where ρ is the surface concentration of activated receptors and D is phosphoinositide diffusivity. For each random realization we started from a stationary homogeneous PTEN, PIP_2 distribution. At time $t = 0$ receptor activation was switched on; either activated receptors were isotropically distributed or the isotropic distribution was perturbed with a linear term producing a 5% difference in activated receptor density between the North and the South poles. In the isotropic case, we found that in a wide region of parameter space the chemical network presents an instability with respect to phase separation, i.e., the homogeneous phosphoinositide mixture realized soon after receptor activation is unstable and tends to decay into spatially separated PIP_2- and PIP_3-rich phases.

Characteristic times for phase separation vary from the order of a minute to that of an hour, depending on receptor activation.

The dynamic behavior and stationary state of the system strongly depend on the values of two key parameters: the concentration ρ of activated receptors and the diffusivity D. In the case of anisotropic stimulation, orientation of PIP_2 and PIP_3 patches clearly correlates with the signal anisotropy (see Fig. 3.4) In the anisotropic case, phase separation takes place in a larger region of parameter space and in times that can be shorter by one order of magnitude.

Average phase separation times as functions of receptor activation $\rho = [\text{Rec}]$ and diffusivity D are plotted for isotropic activation in Fig. 3.5a and for 5% anisotropic activation in Fig 3.5c. Light areas correspond to non phase-separating systems. In the dark areas phase separation takes place in less than 5 minutes of simulated time, while close to the boundary of the broken symmetry region phase separation can take times of the order of an hour.

Average cluster sizes at stationarity are plotted in Figs. 3.5b,d. In the light region, cluster sizes are of the order of the size of the system, corresponding to the formation of pairs of complementary PIP_2 and PIP_3 patches (Fig. 3.4).

For diffusivities smaller than 0.1 $\mu m^2/s$ the diffusion-mediated interaction is unable to establish correlations on lengths of the order of the size of the system and one observes the formation of clusters of separated phases of size much smaller than the size of the system.
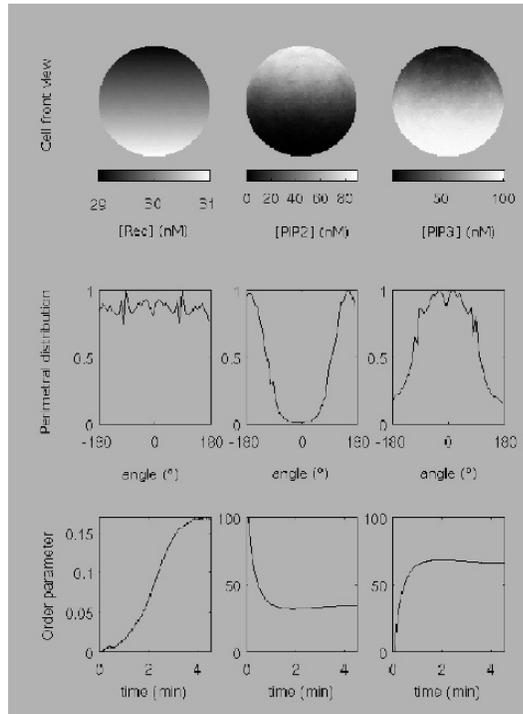
Fig. 3.4. Phase separation in the presence of 5% anisotropic receptor activation switched on as described in the text. The 5% activation gradient pointed in the upward vertical direction. First row: cell front view. Second row: concentrations measured along the cell perimeter and normalized with their maximum value. Third row: time evolution of Binder's parameter g. First column: receptor activation. Second column: PIP_2 concentration. Third column: PIP_3 concentration.

For diffusivities larger than 2 $\mu m^2/s$ the tendency to phase separation is contrasted by the disordering action of phosphoinositide diffusion. Average phase separation times for the anisotropic case are plotted in Fig. 3.5c.

By comparing the isotropic and the anisotropic case it appears that there is a large region of parameter space where phase separation is not observed with isotropic stimulation, while a 5% anisotropic modulation of activated receptor density triggers a fast phase separation process. Cluster sizes are on average larger in the anisotropic case than in the isotropic case.

The transition from a phase-separating to a phase-mixing regime results from a competition between the ordering effect of the interactions and the disordering effect of molecular diffusivity. The frontier between these two regimes varies continuously as a function of parameters. Importantly, we found that the overall phase separation picture is robust with respect to parameter perturbations, since it persists even for concentrations and reaction rates differing from those of Table 3.2 by one order of magnitude.
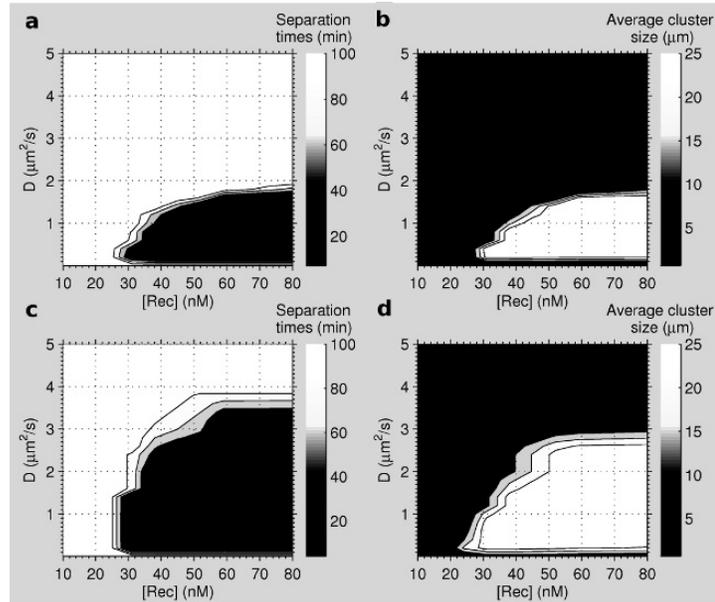
Fig. 3.5. Dynamic phase diagram. Average phase separation times and average cluster sizes are shown using a grayscale as functions of receptor activation [Rec] and diffusivity D, for isotropic and 5% anisotropic activation. In the isotropic case, panels show: (**a**) Average phase separation time, (**b**) average cluster size as a function of [Rec] and D. In the anisotropic case, panels show: (**c**) Average phase separation time, (**d**) average cluster size. For anisotropic activation phase separation is faster, takes place in a larger region of parameter space, and is correlated with the anisotropy direction.

It is also worth noticing that in both isotropic and anisotropic conditions signal amplification is completely reversible. Switching off receptor activation abolishes phase separation, delocalizes PI3K from the plasma membrane to the cytosol, and brings the system back to the quiescent state.

Physically, the mechanism leading to cluster formation can be understood as follows. Receptor activation shifts the chemical potential for PI3K, which is thus recruited to the plasma membrane. PI3K catalytic activity produces PIP_3 molecules from the initial PIP_2 sea. Initially, the two phosphoinositide species are well mixed. Fluctuations in PIP_2 and PIP_3 concentrations are however enhanced by preferential binding of PTEN to its own diffusing phosphoinositide product, PIP_2. Binding of a PTEN molecule to a cell membrane site induces a localized transformation of PIP_3 into PIP_2, resulting in a higher probability of binding other PTEN molecules at neighboring sites. This positive feedback loop not only amplifies the inhibitory PTEN signal, but via phosphoinositide diffusion it also establishes spatio-temporal correlations that enhance the probability of observing PTEN enzymes at neighboring sites as well. If strong enough, this diffusion-induced interaction drives the system towards spontaneous phase separation. The time needed by the system to fall into the more stable, phase-separated

phase can however be a long one if the symmetric, unbroken phase is metastable. In that case, a small anisotropic perturbation in the pattern of receptor activation can be enormously amplified by the system instability.

3.4 Conclusions

Our results provide a simple physical cue to the enigmatic behavior observed in eukaryotic cells. There is a large region of parameter space where the cell can be insensitive to uniform stimulation over very long times, but responsive to slight anisotropies in receptor activation in times of the order of minutes. Accordingly, simulating shallow gradients of chemoattractant we observed PIP_3 patches accumulating with high probability on the side of the plasma membrane with a higher concentration of activated receptors, thus resulting in a large amplification of the chemotactic signal. Moreover, we identified an intermediate region of parameters, where phase separation under isotropic stimulation is observed on average in a long but finite time. In this case, one would predict that on long timescales cells undergo spontaneous polarization in random directions, and that the number of polarized cells grows with time. Intriguingly, this peculiar motile behavior is known as chemokinesis and is observed in cell motility experiments when cells are exposed to chemoattractants in the absence of a gradient [9].

In summary, the phase separation scenario provides a simple and unified framework for different aspects of directed cell motility, such as large amplification of slight signal anisotropies, insensitivity to uniform stimulation, appearance of isolated and transient phosphoinositide patches, and stochastic cell polarization.

It provides a link between known microscopic and macroscopic timescales. Finally, it unifies apparently conflicting aspects which previous modeling efforts could not satisfactorily reconcile [4], such as insensitivity to absolute stimulation values, large amplification of shallow chemotactic gradients, reversibility of phase separation, robustness with respect to parameter perturbations, stochastic character of cell response, and use of realistic biochemical parameters and space-time scales.

References

1. Binder, K.: Theory of first-order phase transitions. Rep. Prog. Phys., **50**, 783–859 (1987).
2. Binder, K., Heermann, D.W.: Monte carlo simulations in statistical physics. An Introduction. 4th edition. Springer, Berlin (2002).
3. Cullen, P., Cozier, G., Banting, G., Mellor, H.: Modular phosphoinositide-binding domains– their role in signalling and membrane trafficking. Curr. Biol., **11**, R882–R893 (2001).
4. Devreotes, P., Janetopoulos, C.: Eukaryotic chemotaxis: distinctions between directional sensing and polarization. J. Biol. Chem., **278**, 20445–20448 (2003).
5. Gamba, A., de Candia, A., Di Talia, S., *et al*.: Diffusion limited phase separation in eukaryotic chemotaxis. Proc. Nat. Acad. Sci. U.S.A., **102**, 16927–16932 (2005).
6. Gillespie, D.: Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem., **81**, 2340–2361 (1977).

 7. Iijima M., Huang E.Y., Luo H.R., Vazquez, F., and Devreotes, P.N.: Novel mechanism of PTEN regulation by its phosphatidylinositol 4,5-bisphosphate binding motif is critical for chemotaxis. J. Biol. Chem., **16**, 16606–16613 (2004).

 8. Janetopoulos, C., Ma, L., Devreotes, P., Iglesias, P.: Chemoattractant-induced phosphatidylinositol 3,4,5-trisphosphate accumulation is spatially amplified and adapts, independent of the actin cytoskeleton. Proc. Natl. Acad. Sci. U.S.A., **101**, 8951–8956 (2004).

 9. Lauffenburger, D.A., Horwitz, A.F.: Cell migration: a physically integrated molecular process. Cell, **84**, 359–369 (1996).

10. Ridley, A., Schwartz, M., Burridge, K., Firtel, M.H., Ginsberg, G., Borisy, G., Parsons, J.T., and Horwitz, A.R.: Cell migration: integrating signals from front to back. Science, **302**, 1704–1709 (2003).

11. Seul, M., Andelman, D.: Domain shapes and patterns: the phenomenology of modulated phases. Science, **267**, 476–483 (1995).

12. Stanley, H.: Introduction to phase transitions and critical phenomena. Oxford University Press, London (1987).

13. Van Haastert, P., Devreotes, P.: Chemotaxis: signalling the way forward. Nat. Rev. Mol. Cell Biol., **5(8)**, 626–624 (2004).

Protein Domains of GTPases on Membranes: Do They Rely on Turing's Mechanism?

Lutz Brusch,[1] Perla Del Conte-Zerial,[1,2] Yannis Kalaidzidis,[2,3] Jochen Rink,[2] Bianca Habermann,[2] Marino Zerial,[2] and Andreas Deutsch[1]

[1] Center for Information Services and High Performance Computing, Technische Universität Dresden, 01062 Dresden, Germany; lutz.brusch@tu-dresden.de

[2] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden, Germany

[3] A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow 119899 Russia

Summary. We evaluate different mechanisms for spatial domain formation of guanosine triphosphatases (GTPases) on cellular membranes. A kinetic model of the basic guanine-nucleotide cycle common to all GTPases is developed and coupled along a one-dimensional axis by diffusion of inactive and activated GTPases. We ask whether a parameter set exists such that domain formation is possible by Turing's mechanism, i.e., purely by reactions and diffusion, and show that the Turing instability does not occur in this model for any parameter combination. But, as revealed by stability and bifurcation analysis, domain formation is reproduced after augmenting the model with combinations of two spatial interaction mechanisms: 1. attraction and 2. adhesion among active GTPases. These interactions can be mediated by effector proteins that bind active GTPases, and the model therefore predicts domains to disintegrate if effector binding is inhibited.

Key words: GTPase, membrane domain, Turing pattern, bifurcation analysis.

4.1 Introduction

Small GTPases are proteins of small mass (20–25 kDa) that regulate essential cell functions, such as growth, movement, differentiation and intracellular vesicle transport by switching between two conformations. In its On-state, a small GTPase is bound to guanosine triphosphate (GTP) which can be hydrolyzed by the protein's GTPase activity and thereby switch the GTPase into its GDP-bound Off-state. Replacing GDP for GTP requires a guanine-nucleotide exchange factor (GEF) and activates the GTPase again. For mammalian cells, small GTPases are grouped into 5 subfamilies based on similarities in sequence, structure and function: Ras, Rho, Rab, Arf, Ran. In their active form, many GTPases are localized to the plasma membrane and/or internal membranes, respectively, where they coordinate and functionally utilize membrane

domains [1]. In particular, about 50 members of the Rab subfamily organize individual membrane domains on intracellular vesicles [2,3].

Here we ask whether the guanine-nucleotide cycle of GTPases described above together with lateral diffusion of the GTPases in the membrane are sufficient to account for the ability of some GTPases, notably those of the Rab subfamily, to form spatial patterns of alternating high and low GTPase densities. Moreover, given the similarities among the GTPase superfamily, could the activity of regulators (like GEFs) of the guanine-nucleotide cycle explain why some GTPases form domains and others don't?

The formation of stationary spatial domains by means of reaction and diffusion had been proposed in a seminal paper by Alan Turing [4]. He had shown by means of linear stability analysis that (bio-)chemical reactions among at least two species of which at least one diffuses are sufficient to form domains given that the kinetic terms and diffusion constants obey well-defined relations [4]. Gierer and Meinhardt developed a closely related theory focusing on local activation and lateral inhibition [5]. Since then many examples of Turing patterns have been studied in mathematical models with an emphasis on morphogenesis in tissues [6,7]. Experimentally Turing's mechanism was first proven in a purely chemical medium, the CIMA reaction [8,9], and has been used to explain domain formation at the scale of the plasma membrane of single cells, e.g., domains of proton pumps [10], filopodia in neurons [11] and bipolarity formation [12]. Moreover, biophysical mechanisms have been studied that can pattern the lipid composition of a membrane due to electrostatic interactions with GMC proteins [13,14], membrane curvature [15] or varying lipid composition [16].

We tackle the question for GTPase domain formation by means of linear stability and bifurcation analysis of the derived coupled nonlinear partial differential equations for the densities of inactive and active GTPases along a one-dimensional axis across the membrane. We show that, given the developed reaction-diffusion model, Turing's mechanism is not responsible for GTPase domains and that the homogeneous distribution of GTPases always remains stable. We then propose and analyze active coupling mechanisms as a necessary requirement for domain formation. Similar mechanisms have been proposed to account for waves of MinD membrane density during center finding for cell division in *E.coli* [17].

4.2 Derivation of the Model

We develop a model of the guanine-nucleotide cycle of GTPases with lateral diffusion and spatial interactions among active GTPases. The model describes the density of inactive GTPases on a line across the membrane by $u(x,t)$ and the density of active GTPases by $v(x,t)$.

Here we consider an infinitely extended system $x \in R$. Fig. 4.1 (a) shows the reaction cycle and Fig. 4.1 (b) depicts the spatial coupling mechanisms.

4.2.1 Kinetic Model of Guanine-Nucleotide Cycle

As depicted in Fig. 4.1 (a), we consider the following reactions common to most GTPases: cytosolic GDP-bound GTPases are recruited onto the membrane with constant
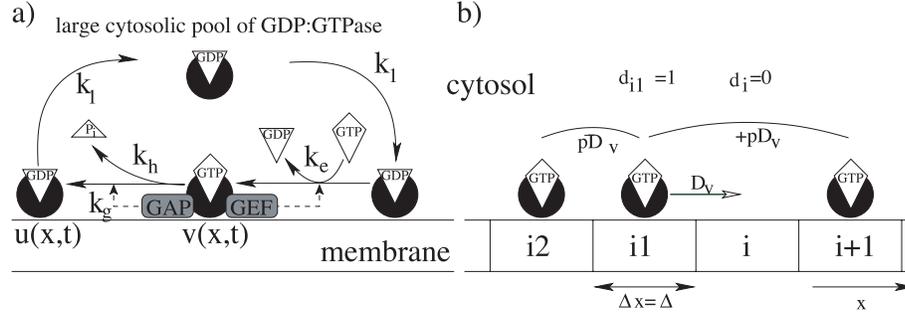
Fig. 4.1. Scheme of the modeled reaction cycle of GTPases (black circles) and their spatial interactions. (a) The reaction cycle converts GDP-bound (inactive) and GTP-bound (active) forms of the GTPase. First-order rate constants are assumed to describe the individual reactions except for k_e which denotes the maximum rate of the GEF-catalyzed exchange of GDP for GTP and a GAP-catalyzed contribution k_g to the hydrolysis of GTP. (b) In the discrete version of the model, occupation numbers $d_i \in \{0, 1\}$ on a discrete lattice of sites $\ldots, i-1, i, i+1, \ldots$ are used to describe the spatial distribution of GTPases. Inactive GTPases are assumed to diffuse randomly regardless of the configuration in their neighborhood since they do not recruit effectors. For active GTPases we introduce neighborhood-dependent offsets p (attraction) and $\bar{p}$ (adhesion) to the diffusion constant that mimic interactions mediated by bound effectors, e.g., the probability for a move from site $i-1$ to site i has three contributions, free diffusion and the two offsets represented by the arcs.

flux k_1 which assumes a large (therefore constant) cytosolic pool. For many GTPases this process requires the unbinding from GDP dissociation inhibitors (GDIs) that accompany the GTPases in cytosol; hence this unbinding shall also be included in the flux k_1. Moreover, we include a maximum membrane occupation by modulating the constant recruitment flux k_1 with a limiting factor of the density of free membrane sites $k_1(1 - u(x, t) - v(x, t))$ implying that u and v are measured as fractions of the maximum density which we scaled to 1.

GDP-bound GTPases may be extracted from the membrane at a rate k_{-1} by binding of GDI. These two processes regulate the amount of membrane-bound inactive GTPases via the balance equation (4.1) for $u(x, t)$.

On the membrane, GTPases can be activated by exchange of their GDP for GTP. This process is modeled as an autocatalytic reaction with saturation (Michaelis–Menten constant k_M). The amount of GTP in the cytosol is assumed to be constant (usually 10 times larger than that of GDP) and is included in the maximum exchange rate k_e. The autocatalysis stems from the requirement for guanine-nucleotide exchange factors (GEFs) that catalyze the exchange reaction and as in the case of Rab GTPases are themselves recruited by the active form of the GTPase. The active GTPase may be deactivated by hydrolysis of its GTP. The hydrolysis is modeled with a constant intrinsic rate k_h which may be increased up to $k_h(1 + k_g)$ by the action of GTPase activating proteins (GAPs) that can be recruited by the active form of some types of GTPases. Effector recruitment is assumed to saturate with the same constant k_M both for GEFs and GAPs. For GTPases that do not themselves recruit GAPs we set $k_g = 0$

and include a constant GAP activity in k_h. Hence, nucleotide exchange and hydrolysis convert the two membrane-bound forms of the GTPase and contribute inversely to both balance equations via $g(u, v)$.

The inactive form of the GTPase is assumed to freely diffuse on the membrane with diffusion constant D_u in the transport term $D_u(\partial^2 u/\partial x^2)$, whereas for the active form we need to derive a separate transport term $-\partial J(u, v)/\partial x$ that reduces to free diffusion with constant D_v in the limit of vanishing strength of the spatial coupling. Hence, the model is given by two coupled nonlinear partial differential equations (PDEs) for $x \in R, t \in R^+, u, v \in [0, 1]$ and random initial conditions:

$$
\begin{aligned}
\frac{\partial u}{\partial t} &= f(u, v) + D_u \frac{\partial^2 u}{\partial x^2} \\
\frac{\partial v}{\partial t} &= g(u, v) - \frac{\partial J(u, v)}{\partial x} \\
f(u, v) &= k_1(1 - u - v) - k_{-1}u - g(u, v) \\
g(u, v) &= k_e \frac{v}{k_M + v}u - k_h \left(1 + k_g \frac{v}{k_M + v}\right)v.
\end{aligned}
\tag{4.1}
$$

4.2.2 Model of Spatial Interactions

Here we model the lateral transport of active GTPases by a parameterized sum of three contributions: 1. diffusion of, 2. attraction of and 3. adhesion between active GTPases. These interactions are motivated by the ability of active GTPases in general to recruit effector proteins that serve biological functions and that mediate interactions between GTPase-effector complexes which are here collectively described by $v(x, t)$. Inactive GTPases that do not recruit effectors are assumed to diffuse freely without the mentioned extensions.

In order to obtain an extended PDE for $v(x, t)$ with spatial coupling terms that depend only on a few parameters directly related to the microscopic mechanisms, we first formulate a discrete microscopic model and then perform a continuum limit (these calculations are shown in Appendix 4.4). Alternatively, the coupling terms in the PDE could be derived from a macroscopic model of the free energy as outlined by Cahn and Hilliard [18] and recently applied to protein-lipid interactions [13]. The strength of the derivation presented here lies in the compatibility of the microscopic and macroscopic descriptions which enables further studies of the microscopic mechanisms, e.g., by means of cellular automata [22], at parameter values that are identical in the macroscopic description.

A single active GTPase without neighboring molecules performs a random walk of steps Δ (of the size of a GTPase-effector complex) during time intervals Δt with probability $D_v(\Delta t/\Delta^2)$ as shown in Fig. 4.1 (b). If there are other GTPases in its neighborhood then the considered active GTPase preferably jumps towards other active GTPases with an enhanced probability $D_v(1 + p)(\Delta t/\Delta^2)$. If it has a direct neighbor then it departs only with a decreased probability $D_v(1 - \bar{p})(\Delta t/\Delta^2)$. The parameter p lumps all attractive (medium range) forces, whereas the parameter $\bar{p}$ describes adhesion forces upon contact. The attractive forces can also contribute to the adhesive

interaction upon contact and the parameter choice for $\bar{p}$ can thus include an offset proportional to p.

This microscopic model has been introduced and analyzed by Kruse for the case $\bar{p} = 0$ and was applied to model MinD oligomerization for center finding dynamics in *Escherichia coli* [17]. Here we include nearest neighbor adhesion by extending the model to $\bar{p} > 0$ and couple it to the guanine-nucleotide cycle of GTPases. The flux term $J(v)$ is calculated in Appendix 4.4.

4.2.3 Complete Model

We insert the spatial coupling term $J(v)$ from (4.13) into (4.1) and in the following first analyze the resulting complete model for the case of simple diffusive coupling ($p = \bar{p} = 0$) and subsequently analyze with attractive ($p \neq 0$) and adhesive ($\bar{p} \neq 0$) interactions:

$$u_t = k_1(1 - u - v) - \left(k_{-1} + k_e \frac{v}{k_M + v}\right) u + k_h \left(1 + k_g \frac{v}{k_M + v}\right) v + D_u u_{xx}$$

$$
\begin{aligned}
v_t = {} & k_e \frac{v}{k_M + v} u - k_h \left(1 + k_g \frac{v}{k_M + v}\right) v + D_v v_{xx} \\
& + D_v p \left[(-2v + 3v^2) v_{xx} + (-2 + 6v) v_x^2 \right. \\
& \qquad\qquad \left. + \Delta^2 \left\{\left(-\frac{7}{6} v + \frac{5}{4} v^2\right) v_{xxxx} + \left(-\frac{2}{3} + 4v\right) v_x v_{xxx} + \left(\frac{1}{2} + \frac{3}{2} v\right) v_{xx}^2\right\}\right] \\
& - D_v \bar{p} \left[(4v - 3v^2) v_{xx} + (4 - 6v) v_x^2 \right. \\
& \qquad\qquad \left. + \Delta^2 \left\{\left(\frac{4}{3} v - \frac{5}{4} v^2\right) v_{xxxx} + \left(\frac{10}{3} - 4v\right) v_x v_{xxx} + \left(2 - \frac{3}{2} v\right) v_{xx}^2\right\}\right].
\end{aligned}
$$

$$(4.2)$$

4.3 Results

We perform a linear stability analysis around the homogeneous steady state and corroborate these results by nonlinear bifurcation analysis to show that the basic reaction-diffusion mechanism does not account for domain formation of GTPases whereas additional spatial interactions do.

4.3.1 Homogeneous Steady State

Homogeneous steady states are solutions of constant densities u^*, v^* independent of space and time. Hence, the derivative terms in (4.1) vanish and u^*, v^* are determined

as the roots of the algebraic equations $0 = f(u^*, v^*), 0 = g(u^*, v^*)$. We restrict our analysis of spatial domain formation to that part of the parameter space that accounts for a finite positive density of active GTPases, i.e.,

$$\frac{k_e}{h_h} \frac{k_1}{k_1 + k_{-1}} > k_M.$$

(4.3)

Hence, the ratio of gain to loss rates needs to exceed the saturation k_M of the autocatalytic activation. In the considered parameter range exactly one homogeneous steady state with finite positive u^* and v^* exists and reads

$$u^* = \frac{k_1}{k_1 + k_{-1}} \left[1 - \frac{1 - \frac{(k_1 + k_{-1})k_h}{k_1 k_e}}{1 - \frac{(k_1 + k_{-1})k_h}{k_1 k_e}(1 + k_g)} k_M \right]$$

$$v^* = \frac{1 - \frac{(k_1 + k_{-1})k_h}{k_1 k_e}}{1 - \frac{(k_1 + k_{-1})k_h}{k_1 k_e}(1 + k_g)} k_M \quad .$$

(4.4)

4.3.2 GTPase Domain Formation Due To Turing's Mechanism?

Here we ask whether in principle a combination of parameter values could be found such that Turing's mechanism (as outlined in the introduction) leads to domain formation given only lateral diffusion ($p = \bar{p} = 0$) and our general assumptions about the GTPase cycle. The formation of spatial domains of alternating high and low densities of GTPases can occur when the homogeneous steady state is unstable against spatially heterogeneous perturbations $\tilde{u}(x, t), \tilde{v}(x, t)$. For small perturbations this instability can be tested in a linearization of (4.1) after expanding the nonlinearities $f(u, v), g(u, v)$ with $u(x, t) = u^* + \tilde{u}(x, t), v(x, t) = v^* + \tilde{v}(x, t)$ into a Taylor series. The linear system of PDEs is solved with an expansion of the perturbations into spatial Fourier modes $\sim e^{ikx}$ with wavenumbers k and an exponential time dependence $\sim e^{\lambda(k)t}$:

$$\begin{pmatrix} u(x, t) \\ v(x, t) \end{pmatrix} = \begin{pmatrix} u^* \\ v^* \end{pmatrix} + \int dk \sum_{j=1,2} \begin{pmatrix} u_j \\ v_j \end{pmatrix} e^{\lambda_j(k)t} e^{ikx} \quad .$$

(4.5)

Inserting into (4.1) and evaluating the derivatives yields the eigenvalue problem

$$\lambda_{1,2}(k) \begin{pmatrix} u_{1,2} \\ v_{1,2} \end{pmatrix} = \left[\left(\begin{matrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{matrix} \right) \Bigg|_{u^*, v^*} + \begin{pmatrix} -D_u k^2 & 0 \\ 0 & -D_v k^2 \end{pmatrix} \right] \begin{pmatrix} u_{1,2} \\ v_{1,2} \end{pmatrix}$$

(4.6)

with two eigenvalues $\lambda_{1,2}(k)$, eigenvectors $\begin{pmatrix} u_{1,2} \\ v_{1,2} \end{pmatrix}$ and the Jacobian matrix (first on the right-hand side) to be evaluated at the steady state solution, see Appendix 4.4. In the following, derivatives with respect to u or v always imply evaluation at (u^*, v^*) without explicit mentioning. Moreover, we denote such derivatives by a corresponding index u or v, respectively.

The eigenvalues can be determined from the characteristic polynomial

$$0 = \lambda_{1,2}(k)^2 + \lambda_{1,2}(k)[-f_u + k^2 D_u - g_v + k^2 D_v]$$
$$+ [(f_u - k^2 D_u)(g_v - k^2 D_v) - f_v g_u] \tag{4.7}$$

and the homogeneous steady state is found to be stable if the real parts of both $\lambda_{1,2}(k)$ are negative for all $k \in R$; hence any perturbation decays like $\sim e^{\Re[\lambda(k)]t + i\Im[\lambda(k)]t}$.

First, we verify that the homogeneous steady state (4.4) is stable against spatially homogeneous perturbations ($k = 0$) which requires $f_u + g_v < 0$ and $f_u g_v - f_v g_u > 0$. In Appendix 4.4 we calculate the derivatives and find that both $f_u < 0$ (4.17) and $g_v < 0$ (4.18); hence the first inequality holds. Also the second condition $f_u g_v - f_v g_u = (-g_u - k_1 - k_{-1})g_v - (-g_v - k_1)g_u = (k_1 + k_{-1})(-g_v) + k_1 g_u > 0$ is fulfilled.

Now we search for $\lambda_{1,2}(k)$ with positive real part at $k > 0$ for which a necessary condition is $D_v f_u + D_u g_v > 0$, or else the functions $\Re[\lambda_{1,2}(k)]$ monotonously decrease with k. As we calculate in Appendix 4.4, $f_u < 0$ (4.17) and $g_v < 0$ (4.18) for any arbitrary combination of the parameter values. Hence, the condition $D_v f_u + D_u g_v > 0$ is violated for any choice of the kinetic parameter values or the diffusion constants. We repeated these calculations for arbitrary but monotonous kinetic terms and obtained the same $f_u < 0$, $g_v < 0$ around the steady state with high values of u, v. On the contrary, for all models that possess a Turing instability there is always one of the derivatives f_u and g_v that has positive values whereas the other one has negative values at the steady state solution. The closest Turing-like variant of our model is a substrate-depletion system that possesses $g_v > 0$. However, the appropriately adjusted kinetic terms that can yield $g_v > 0$ also yield an instability for homogeneous perturbations, which would be biologically unrealistic. Maintaining the design of an activation(via GEF)-inactivation(hydrolysis) cycle of the GTPases, the only possibility of obtaining a Turing instability is then to increase the number of species but immediate candidates for these additional species are lacking in our GTPase system. As an example for such a strategy, a 4-variable model of similar activation-inactivation cycles was used to describe the Min oscillations in *E. coli* [19] and this also constitutes an alternative model to the spatial interactions assumed by Kruse [17].

For the general reaction cycle of a single GTPase with two membrane-bound species and a constant cytoplasmic pool (4.1), we conclude that pure lateral diffusion of GTPases with any combination of diffusion constants does not provide an explanation for domain formation of GTPases on membranes as it is observed for instance for members of the Rab subfamily. In the following we show that the proposed additional coupling mechanisms (attraction and adhesion) readily cause a linear instability of the homogeneous state and that stationary heterogeneous domain solutions exist.

4.3.3 Contribution of Spatial Interactions

First, we analyze the balance of diffusion and spatial interactions in the absence of reactions, i.e., $v_t = -J_x$ with J from (4.13) and proceed as in the previous Sect. 4.3.2. Any homogeneous v^* with $0 \leq v^* \leq 1$ is now a steady state as the interactions are
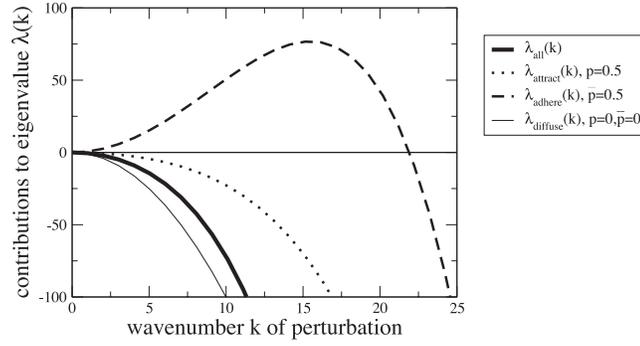
Fig. 4.2. Dependence of the eigenvalue λ on the wavenumber k of perturbations. The thick solid curve shows the full result (4.8) for $v^* = 0.8$, $D = 1$, $p = 0.5$, $\bar{p} = 0.5$, $\Delta = 0.1$ whereas the other curves denote only single terms of (4.8), see legend.

mass conserving. The stability property of v^* is then given by a single real eigenvalue per wavenumber:

$$\lambda(k) = -Dk^2 + Dp\left[(2 - 3v^*)v^*k^2 - \Delta^2\left(\frac{7}{6} - \frac{5}{4}v^*\right)v^*k^4\right]$$
$$+ D\bar{p}\left[(4 - 3v^*)v^*k^2 - \Delta^2\left(\frac{4}{3} - \frac{5}{4}v^*\right)v^*k^4\right]. \tag{4.8}$$

Fig. 4.2 shows $\lambda(k)$ from Eq. 4.8 and individual contributions thereof for a fixed choice of the parameters. The homogeneous state v^* is unstable if any k exists such that $\lambda(k)$ is positive. The diffusion equation alone ($p = \bar{p} = 0$) yields stability for any choice of v^*, e.g., the thin solid curve. The attractive medium-range interactions alone (no diffusion and $\bar{p} = 0$) cause an instability for $v^* < 2/3$ as was found in [17].

The adhesive contact interactions alone (no diffusion and $p = 0$) cause instability for any density v^*. However, for the choice of parameters in Fig. 4.2, the potential instability due to adhesive interactions is overcompensated by the stabilizing contributions of the other two processes.

4.3.4 GTPase Domain Formation Due To Spatial Interactions

In this section the linear stability analysis is performed for the complete two-species model of Eq. 4.2. Again using the ansatz (4.5) in the linearization of (4.2) and evaluating the higher-order spatial derivatives yields an eigenvalue problem similar to (4.6) but for the matrix

$$\begin{bmatrix} f_u - Dk^2 & f_v \\ g_u & g_v - Dk^2 + Dp\left[(2 - 3v^*)v^*k^2 - \Delta^2\left(\frac{7}{6} - \frac{5}{4}v^*\right)v^*k^4\right] \\ & + D\bar{p}\left[(4 - 3v^*)v^*k^2 - \Delta^2\left(\frac{4}{3} - \frac{5}{4}v^*\right)v^*k^4\right] \end{bmatrix}. \tag{4.9}$$
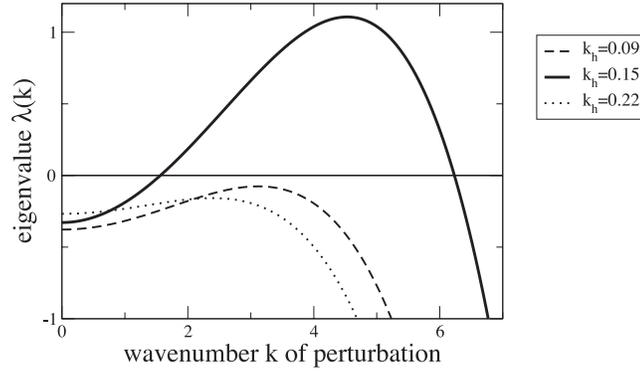
Fig. 4.3. Dependence of the largest eigenvalue $\lambda_1(k)$ of perturbations with wavenumber k around the homogeneous steady state (u^*, v^*) for different k_h as denoted in the legend. Other parameters are $k_1 = k_{-1} = k_e = k_M = k_g = 10$, $D = 1$, $\Delta = 0.1$, $p = 0.5$, $\bar{p} = 0.8$.

The results of the linear stability analysis are shown in Fig. 4.3 for a chosen set of kinetic parameters $k_1 = k_{-1} = k_e = k_M = k_g = 10$, $D = 1$, $\Delta = 0.1$, $p = 0.5$, $\bar{p} = 0.8$ and three different values for k_h. The choice $p = 0.5$ describes a situation where out of 5 steps a free molecule of active GTPase will move 2 times away from a nearby aggregate and move 3 times towards it, hence a directional bias of 20%. The parameter $\bar{p} = 0.8$ denotes a scenario in which the probability for a single molecule to detach from the edge of and move away from an aggregate is only 20% of the probability for a movement of a free molecule.

In Fig. 4.3 each of the curves contains contributions from the spatial interaction processes (4.13), from the diffusion of inactive GTPases and from the local kinetics (4.1). The latter two processes are stabilizing for all conditions. Only for a limited range of parameter values does there exist positive $\lambda(k)$ which enables domain formation. For the present choice of parameters this window spans $0.092 < k_h < 0.211$.

4.3.5 GTPase Domain Solutions

So far, we have analyzed the linear dynamics of solutions close to the homogeneous steady states. New coexisting solutions with a characteristic spatial profile emerge at the detected instabilities of the homogeneous solutions. Here we compute these new solutions from the full nonlinear model Eq. 4.2. The spatial profiles of the emerging stationary solutions $u^*(x)$, $v^*(x)$ satisfy a system of 6 first-order ordinary differential equations derived from (4.2) by setting time derivatives to zero and reordering the two right-hand sides as equations for the highest-order spatial derivative (second and fourth, respectively), then rewriting these as a first-order system. Such a transformed system can be efficiently computed by means of continuation methods as implemented in the software package AUTO [20]. Typical domain solutions are shown in Fig. 4.4. The bifurcations where the domain solutions emerge (here, e.g., $k = \pi$) correspond to
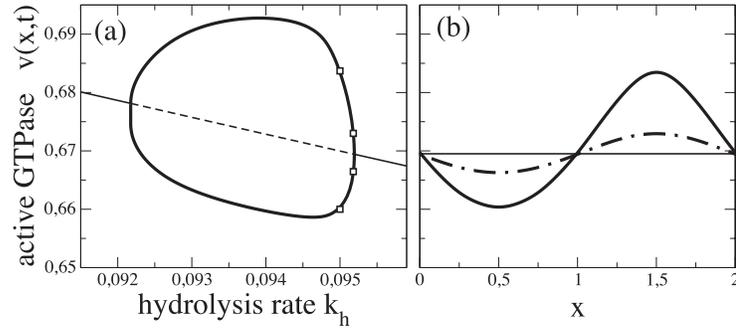
Fig. 4.4. (a) Bifurcation diagram of the homogeneous state (thin curve) and maxima and minima of the domain solutions (thick), (un-)stable solutions are denoted by full (dashed) curves. (b) Two particular solutions of domains with spatial areas of low and high GTPase density. Pairs of symbols in (a) denote the maximum and minimum amplitude of the two solutions in (b) that correspond to two different values of k_h as denoted by the pairs of symbols. Other parameters are the same as in Fig. 4.3.

those parameters k_h at which $\lambda(k) = 0$. Moreover, the bifurcations are supercritical, i.e., domain solutions only exist where the homogeneous state is unstable.

4.4 Discussion

Many GTPases are recruited to membranes and domain formation is observed for some of them, including Rab GTPases. We investigated the question of whether a general kinetic model of the guanine-nucleotide cycle of GTPases together with lateral diffusion may account for domain formation on membranes. A linear stability analysis showed that no domains form for any choice of parameters under these assumptions. The extension of the model by attractive and adhesive interactions among active GTPases provided the possibility of domain formation as was shown by stability and bifurcation analysis for selected examples.

This model considers a one-dimensional axis across a cellular membrane but the results are readily extended to the two-dimensional plane if we can neglect fluctuations in the angular direction. Previously, two-dimensional lattice-gas automata with adhesive interactions have been shown to possess a reciprocal dependence of the critical adhesiveness versus density [21, 22] in agreement with our instability condition $\bar{p} > 1/[v^*(4 - 3v^*)]$ derived from (4.8) for $p = 0$. The same dependence is also known from simpler continuum models with interactions.

Microscopically the recruitment of effector proteins by active GTPases could mediate the spatial interactions that are required for domain formation. Domain formation could therefore be controlled by modifications of the biophysical properties of effector proteins. Moreover, if effector binding is inhibited this model predicts the membrane domains to disintegrate towards the homogeneous distribution. We suggest performing

such experiments by using small inhibitory molecules or mutations of effector binding domains.

Acknowledgments

Appendix A: Mean-Field Approximation for Spatial Interactions

We formulate a discrete model (see Fig. 4.1 (b)) by collecting all events that change the occupation number $d_i \in \{0, 1\}$ of active GTPases at an arbitrarily chosen site i due to jumps from or to neighboring sites:

$$
\begin{aligned}
\frac{\Delta d_i}{\Delta t} = {} & \frac{D_v}{\Delta^2} \left[1 + pd_{i+1} - \bar{p}d_{i-2} \right] (1 - d_i)d_{i-1} && \text{jump } i-1 \to i \\
& + \frac{D_v}{\Delta^2} \left[1 + pd_{i-1} - \bar{p}d_{i+2} \right] (1 - d_i)d_{i+1} && \text{jump} \qquad i \leftarrow i+1 \\
& - \frac{D_v}{\Delta^2} \left[1 + pd_{i-2} - \bar{p}d_{i+1} \right] (1 - d_{i-1})d_i && \text{jump } i-1 \leftarrow i \\
& - \frac{D_v}{\Delta^2} \left[1 + pd_{i+2} - \bar{p}d_{i-1} \right] (1 - d_{i+1})d_i && \text{jump} \qquad i \to i+1,
\end{aligned}
\tag{4.10}
$$

which after rearrangement yields the discrete model

$$
\begin{aligned}
\frac{\Delta d_i}{\Delta t} = {} & \frac{D_v}{\Delta^2} \left[d_{i-1} + d_{i+1} - 2d_i \right] \\
& + \frac{D_v p}{\Delta^2} \big[2d_{i-1}d_{i+1} - d_i(d_{i-2} + d_{i+2}) - 2d_{i-1}d_i d_{i+1} \\
& \qquad\quad + d_i(d_{i-1}d_{i-2} + d_{i+1}d_{i+2}) \big] \\
& - \frac{D_v \bar{p}}{\Delta^2} \big[d_{i-1}d_{i-2} + d_{i+1}d_{i+2} - d_i(d_{i-1} + d_{i+1}) + 2d_{i-1}d_i d_{i+1} \\
& \qquad\quad - d_i(d_{i-1}d_{i-2} + d_{i+1}d_{i+2}) \big].
\end{aligned}
\tag{4.11}
$$

Next the continuum limit is performed to derive a continuous model from the discrete model. This continuous model is deterministic and facilitates effective linear stability and bifurcation analyses. The occupation numbers d_i are replaced by the local density $v(x, t)$ and the left-hand side of Eq. 4.11 becomes the time derivative $\Delta d_i / \Delta t \to (\partial / \partial t)v = v_t$. The linear terms approximate the second spatial derivative and can be replaced by the diffusion operator $(d_{i-1} + d_{i+1} - 2d_i)/\Delta^2 \to (\partial^2/\partial x^2)v = v_{xx}$. Factors of the product terms need to be expanded in the Taylor series

$$d_{i+j} \rightarrow v(x + j\Delta, t) = v + \frac{(j\Delta)^1}{1!} v_x + \frac{(j\Delta)^2}{2!} v_{xx} + \frac{(j\Delta)^3}{3!} v_{xxx}$$
$$+ \frac{(j\Delta)^4}{4!} v_{xxxx} + \dots \tag{4.12}$$

and terms up to second order in Δ are summed. Hence, the transport terms for $v(x, t)$ in the model are given by

$$J = -D_v v_x + D_v p \frac{d}{dx} \left[(1 - v)v^2 + \Delta^2 \left\{ \left(\frac{7}{6} - \frac{5}{4}v \right) vv_{xx} - \left(\frac{5}{6} - \frac{1}{2}v \right) v_x^2 \right\} \right]$$

$$+ D_v \bar{p} \frac{d}{dx} \left[(2 - v)v^2 + \Delta^2 \left\{ \left(\frac{4}{3} - \frac{5}{4}v \right) vv_{xx} - \left(-\frac{1}{3} - \frac{1}{2}v \right) v_x^2 \right\} \right]. \tag{4.13}$$

The linear part ($p = \bar{p} = 0$) represents the diffusion equation and the nonlinear terms denote the coupling to the local neighborhood. The transport terms are compactly written in terms of the flux J with $v_t = -J_x$ since the considered interaction processes conserve the total amount of GTPases.

Appendix B: Signs of the Elements of the Jacobian Matrix

Here we analytically calculate the entries of the Jacobian matrix, i.e., derivatives of the kinetic terms with respect to the variables, and evaluate the signs of these derivatives at the stationary state (u^*, v^*). Our calculation is valid for arbitrary values of any parameter and does not depend on specific values, as long as they allow a finite membrane density of GTPases (4.3). We consider the kinetic part of the model (4.1) with

$$f(u, v) = k_1(1 - u - v) - k_{-1}u - g(u, v) \tag{4.14}$$

$$g(u, v) = k_e \frac{v}{k_M + v} u - k_h \left(1 + k_g \frac{v}{k_M + v} \right) v \tag{4.15}$$

and note that $f_u = -k_1 - k_{-1} - g_u$ and $f_v = -k_1 - g_v$. One finds

$$g_u = k_e \frac{v^*}{k_M + v^*} > 0 \tag{4.16}$$

and hence

$$f_u < 0. \tag{4.17}$$

The calculation of g_v proceeds by means of logarithmic derivatives and using the abbreviation $h(v) = v/(k_M + v)$:

$$g_v = \frac{\partial}{\partial v}[k_e h(v)u - k_h v - k_h k_g h(v)v]$$

$$= k_e h(v^*)u^* \frac{\frac{\partial h(v)}{\partial v}}{h(v^*)} - k_h - k_h k_g h(v^*)v^* \left(\frac{\frac{\partial h(v)}{\partial v}}{h(v^*)} + \frac{1}{v^*} \right).$$

We use the steady state condition $g(u^*, v^*) = 0$ to replace $k_e h(v^*)u^*$ in the first term by $k_h v^* + k_h k_g h(v^*)v^*$ and cancel one term, then insert $(\partial h(v)/\partial v)/h(v^*) = 1/v^* - 1/(k_M + v^*)$:

$$
\begin{aligned}
g_v &= k_h v^* \frac{\frac{\partial h(v)}{\partial v}}{h(v^*)} - k_h - k_h k_g h(v^*) \\
&= k_h v^* \left(\frac{1}{v^*} - \frac{1}{k_M + v^*} \right) - k_h - k_h k_g \frac{v^*}{k_M + v^*} \\
&= -k_h (1 + k_g) \frac{v^*}{k_M + v^*} \quad < 0.
\end{aligned}
\tag{4.18}
$$

For $k_1 > k_h (1 + k_g)$ we obtain

$$
f_v < 0,
\tag{4.19}
$$

whereas for $k_1 < k_h (1 + k_g)$ the sign of f_v may turn positive depending on the values of the other parameters.

References

1. Balch, W., Der, C., Hall, A.: GTPases Regulating Membrane Dynamics. Academic Press, London (2005)
2. Zerial, M., McBride, H.: Rab proteins as membrane organizers. Nature Rev. Mol. Cell Biol., **2**, 107–117 (2001).
3. Miaczynska, M., Zerial, M.: Mosaic organization of the endocytic pathway. Exp. Cell Res., **272**, 8–14 (2002).
4. Turing, A.M.: The chemical basis of morphogenesis. Phil. Trans. R. Soc. Lond. B. Biol. Sci., **237**, 37–72 (1952).
5. Gierer A., Meinhardt H.: A theory of biological pattern formation. Kybernetik, **12**, 30–39 (1972).
6. Meinhardt, H., Gierer, A.: Pattern formation by local self-activation and lateral inhibition. Bioessays, **22**, 753–760 (2000).
7. Murray, J.D.: Mathematical Biology II: Spatial Models and Biomedical Applications. Springer, Berlin (2002).
8. Castets, V.V., Dulos, E., Boissonade, J., De Kepper, P.: Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern. Phys. Rev. Lett., **64**, 2953–2956 (1990).
9. Dulos, E., Boissonade, J., Perraud, J.J., Rudovics, B., De Kepper, P.: Chemical morphogenesis: Turing patterns in an experimental chemical system. Acta Biotheor., **44**, 249–261 (1996).
10. Fisahn, J., Lucas, W.J.: Spatial organization of transport domains and subdomain formation in the plasma membrane of *Chara corallina*, J. Membr. Biol., **147**, 275–281 (1995).
11. Hely, T.A., Van Ooyen, A., Willshaw, D.J.: A simulation of growth cone filopodia dynamics based on Turing morphogenesis patterns. In: Paton, R.C., Holcombe, M. (eds.) Information Processing in Cells and Tissues. Plenum, New York (1998).
12. Hunding, A.: Microtubule dynamics may embody a stationary bipolarity forming mechanism related to the prokaryotic division site mechanism. J. Biol. Phys., **30**, 325–344 (2004).

13. John, K., Bär, M.: Alternative mechanisms of structuring biomembranes: self-assembly versus self-organization. Phys. Rev. Lett., **95**, 198101 (2005).
14. John, K., Bär, M.: Travelling lipid domains in a dynamic model for protein-induced pattern formation in biomembranes. Phys. Biol., **2**, 123–132 (2005).
15. Gebhardt, C., Gruler, H., Sackmann, E.: On domain structure and local curvature in lipid bilayers and biological membranes. Z. Naturforsch., **32**, 581–596 (1977).
16. Lipowsky, R., Sackmann, E. (eds.): The Structure and Dynamics of Membranes. Elsevier, Amsterdam (1995)
17. Kruse, K.: A dynamic model for determining the middle of *Escherichia coli*. Biophys. J., **82**, 618–627 (2002).
18. Cahn, J.W., Hilliard, J.E.: Free energy of a nonuniform system I. Interfacial free energy. J. Chem. Phys., **28**, 258–267 (1958).
19. Howard, M., Rutenberg, A.D., de Vet, S.: Dynamic compartmentalization of bacteria: accurate division in *E. coli*, Phys. Rev. Lett., **87**, 278102 (2001).
20. Doedel, E., Keller, H.B., Kernevez, J.P.: Numerical analysis and control of bifurcation problems. Int. J. Bif. Chaos., **1**, 493–520 and 745–772 (1991).
21. Bussemaker, H.J.: Analysis of a pattern-forming lattice-gas automaton: Mean-field theory and beyond. Phys. Rev. E., **53**, 1644–1661 (1996).
22. Deutsch, A., Dormann, S.: Cellular Automation Modeling of Biological Pattern Formation. Birkhäuser, Boston (2006).

5

In Vitro Tubulogenesis of Endothelial Cells: Analysis of a Bifurcation Process Controlled by a Mechanical Switch

Philippe Tracqui, Patrick Namy, and Jacques Ohayon

Laboratoire TIMC-IMAG, Equipe DynaCell, UMR CNRS 5525
Institut d'Ingénierie et de l'Information de Santé
Faculté de Médecine de Grenoble, 38706 La Tronche Cedex, France
Philippe.Tracqui@imag.fr

Summary. The formation of new blood vessels *in vivo* is a multistep process in which sprouting endothelial cells (ECs) form tubes with lumina, these tubes being additionally organized as capillary networks. *In vitro* models of tubulogenesis have been developed to investigate this highly regulated multifactorial process, with special attention paid to the determinant role of mechanical interactions between ECs and the extracellular matrix (ECM). In agreement with experimental results obtained when culturing endothelial EAhy926 cells on fibrin gels, we defined theoretical thresholds between cellular traction and active cell migration along ECM strain fields above which tubulogenesis is induced. We additionally illustrated how mechanical factors may provide long-ranged positional information signals leading to localized network formation, thus providing an alternative view to the classical approach of morphogenesis based on gradients of diffusible morphogens.

Key words: Mechanical signalling, *in vitro* angiogenesis, extracellular matrix rheology, positional information, network morphogenesis.

5.1 Introduction

The formation of new capillaries from pre-existing vasculature, i.e., *angiogenesis*, is a biological process of major importance in several physio-pathological contexts, including wound healing and tumour invasion. Under physiological conditions, angiogenesis is a highly regulated phenomenon controlled by different angiogenic stimulators and inhibitors, which are not only soluble factors (various peptides and growth factors, like vascular endothelial growth factor (VEGF) [9]), but also nondiffusible components of the extracellular matrix (ECM) [10]. Interestingly, a large amount of experimental data indicate that ECMs not only display adhesive ligands important for anchorage-dependent cells, but also present a wide range of mechanical and structural properties which influence many cell functions such as migration, proliferation, proteolytic activity and signal transduction. For example, using extracellular substrates of different

mechanical stiffness shows that cells organize their cytoskeleton and adhesive contacts differently on soft and stiff surfaces [12, 17].

In the context of angiogenesis, significant insights into the molecular and cellular biology of this process have come from *in vitro* assays using cultured endothelial cells on ECMs with different rigidities [24]. Indeed, these assays mimic quite well the early morphogenesis of cellular networks, or tubulogenesis, a process during which endothelial cells self-organise around a meshwork of growing areas devoid of ECMs, the lacunae. Further mechanical distortion of these lacunae by cellular forces gives rise to a network of cellular cords with a tubular morphology, i.e., capillary-like networks (CLNs) [26], [22].

In order to design pro- or anti-angiogenic therapeutic strategies, it is essential to understand and to predict modifications of the tubulogenesis scenario which may occur when cell micro-environment modifications are induced by both intrinsic ECM mechanical properties and ECM deformations induced by cellular traction forces. In this paper, this analysis has been conducted by considering the amplitude of cell traction forces as a bifurcation parameter.

5.2 Influence of Mechanical Factors on Tubulogenesis Regulation

In [22], [23] Vailhé et al. demonstrated that the promotion or inhibition of CLNs depends on the fibrin gel concentrations, and thus on the fibrin gel rigidity. When fibrin concentration increases from 0.5 mg/ml to 2 mg/ml, the initial number of lacunae decrease and do not evolve to CLNs. The data points in Fig. 5.1 summarize these experimental results, together with a fit of these data by a Hill function of the form

$$Ind(C) = Ind0 - \frac{\alpha C^n}{(K^n + C^n)}, \tag{5.1}$$

where C is the fibrin gel concentration (mg/ml) and $Ind(C)$ the ratio of the surface occupied by the CLN with respect to the total biogel surface. Using a floating gel technique [2], we were able to derive a phenomenological relationship for the Young's modulus $E(C)$ of the fibrin biogel of the form $E(C) = a \cdot C^P$, with E in kPa and $a = 2.04 \pm 0.27$ kPa/(mg/ml)^p, $p = 0.97 \pm 0.14$ [3]. The x-axis in Fig. 5.1 has been rescaled accordingly, giving a range of elasticity modulus from 1 kPa to 4 kPa. Thus, Fig. 5.1 can be considered as a bifurcation diagram defining a switch-like threshold of Young's modulus values below which instabilities and CLN pattern formation will occur. One goal of the modelling approach developed below is thus to provide an integrated description and understanding of this phenomenon.

5.3 Model Formulation

5.3.1 Biomechanical Components of the Model

Even if it is known that ECMs and biogels exhibit complex rheological properties, we only retained as a first approximation a linear viscoelastic stress/strain relationship to
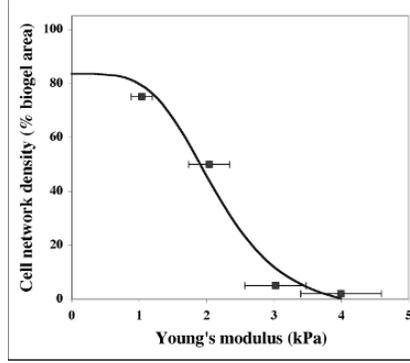
Fig. 5.1. Explicit description of the dependence of CLN formation with regard to fibrin gel concentrations by the function $Ind(C)$ given in the text ($Ind0 = 0.835$; $a = 0.9$; $K = 1.06$, $n = 4$). Squares correspond to the experimental data points of [22] for C values between 0.5 mg/ml and 2.0 mg/ml. Error bars correspond to SD on the estimated Young's modulus. CLN pattern formation only occurs for sufficiently soft fibrin gels ($E < 3$ kPa).

model the biogel mechanical response to cell traction, where the viscous and elastic stresses add linearly. However, since biogels are mostly a network of reticulated fibres, we also consider that long-range elastic resistive stresses may be generated across the biogel [7]. The viscoelastic stress $\boldsymbol{\sigma}_{ECM}$ generated within the fibrin gel as the result of the cell traction stress $\boldsymbol{\sigma}_{cell}$ is thus expressed as the sum of the following elastic and viscous tensor stresses:

$$\boldsymbol{\sigma}_{elast} = 2\mu(\boldsymbol{\epsilon} - \beta_1 \nabla^2 \boldsymbol{\epsilon}) + (\lambda(\theta - \beta_2 \nabla^2 \theta)\mathbf{I}) \tag{5.2}$$

$$\boldsymbol{\sigma}_{visq} = \left(\mu_1 \frac{\partial \boldsymbol{\epsilon}}{\partial t} + \mu_2 \frac{\partial \theta}{\partial t} \mathbf{I} \right), \tag{5.3}$$

where the positive coefficients β_1 and β_2 are the long-range elasticity coefficients, λ and μ are the Lamé coefficients, while μ_1 and μ_2 are the shear and bulk viscosities, respectively. $\boldsymbol{\epsilon}$ is the strain tensor, θ the dilation and $\mathbf{I}$ the unit tensor. In the light of experimental data, the active stress tensor $\boldsymbol{\sigma}_{cell}$ is assumed (i) to be proportional to the ligands or biogel density and to plateau at high cell density because of cell traction inhibition, i.e.,

$$\boldsymbol{\sigma}_{cell} = \tau \rho n (N_2 - n)\mathbf{I}, \tag{5.4}$$

where the parameter τ monitors the cellular traction amplitude, while the positive constant N_2 controls the inhibition of cell traction as cell density increases. These parameters can be identified by considering experimental cell traction curves obtained for a given biogel stiffness (Fig. 5.2).

The biogel attachment to the dish has been modelled as an elastic force $\mathbf{R}$ that decreases as the gel thickness increases. Introducing the restoring parameter s, we assumed that
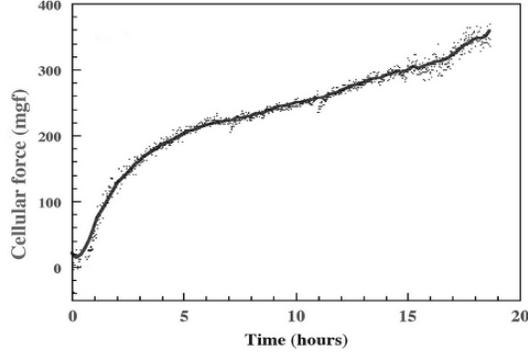
Fig. 5.2. Experimental record of the cellular traction force progressively developed after spreading by EAhy926 endothelial cells seeded at a density of 2×10^6 cells on a fibrin gel of concentration 1 mg/ml.

$$\mathbf{R} = s\mathbf{u}/\rho, \qquad (5.5)$$

where the displacement $\mathbf{u}(\mathbf{r}, t)$ of the cell-biogel composite is given as the solution of the force balance equation:

$$\nabla \cdot [\boldsymbol{\sigma}_{ECM} + \boldsymbol{\sigma}_{\text{cell}}] = \mathbf{R}. \qquad (5.6)$$

5.3.2 Cellular and Biochemical Components of the Model

In addition to passive translocation and random cell motion by cellular diffusion, we assumed an active cell migration up to a biogel density gradient, which includes both haptotactic cell migration (monitored by an adhesivity gradient) and durotaxis (monitored by a rigidity gradient). The evolution of the spatio-temporal cell density is thus given by the conservation equation:

$$\frac{\partial n}{\partial t} = -\nabla \cdot [\mathbf{J}_n + \mathbf{J}_d + \mathbf{J}_h] \qquad (5.7)$$

$$\mathbf{J}_d = -\nabla \cdot [\mathbf{D}(\boldsymbol{\epsilon})n]; \qquad \mathbf{J}_n = n\frac{\partial \mathbf{u}}{\partial t}; \qquad \mathbf{J}_h = hn\nabla\rho, \qquad (5.8)$$

where h is the active migration coefficient and $\mathbf{J}_n$ the flux of cells convected at velocity $\partial \mathbf{u}/\partial t$. The influence on cellular diffusion of the time-varying anisotropy of the biogel due to the strain field [11] has been included by considering a strain-biased diffusion tensor $\mathbf{D}(\boldsymbol{\epsilon})$ in the expression of the cellular diffusive flux $\mathbf{J}_d$ [6].

A similar conservation equation can be derived for the biogel density $\rho(r, t)$. The biogel proteolysis has been neglected due to the experimental protocol we considered (inhibition of plasmin enzymatic action by aprotinin) and only biogel convection by mechanical effects has been considered.

After variable normalisation, the nonlinear partial differential system associated to this mechanocellular model of tubulogenesis is given by [16]

$$\frac{\partial n}{\partial t} + \nabla \cdot \left[-\nabla \cdot (\mathbf{D}(\boldsymbol{\epsilon})n) + hn\nabla\rho + n\frac{\partial \mathbf{u}}{\partial t} \right] = 0$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \left[\rho \frac{\partial \mathbf{u}}{\partial t} \right] = 0 \tag{5.9}$$

$$\nabla \cdot \left[\left\{ 2\mu \left(\boldsymbol{\epsilon} - \beta_1 \nabla^2 \boldsymbol{\epsilon} \right) + \lambda(\theta - \beta_2 \nabla^2 \theta)\mathbf{I} \right\} \right.$$

$$\left. + \left\{ \mu_1 \frac{\partial \boldsymbol{\epsilon}}{\partial t} + \mu_2 \frac{\partial \theta}{\partial t} \mathbf{I} \right\} + \left\{ \tau\rho n(N_2 - n)\mathbf{I} \right\} \right] = s\mathbf{u}/\rho.$$

We assumed as initial spatial conditions a homogeneous normalised value $\rho(x) = 1$ of the ECM density. Initially, the ECM is in an unstressed state ($\boldsymbol{\epsilon}(x) = \mathbf{0}$) and cells are randomly seeded, within a range $n_0(x) = 1 \pm 0.1$. No-flux boundary conditions were considered for the cells and the ECM at the boundaries of the two-dimensional (2D) domain, together with zero displacement conditions. The corresponding nonlinear partial differential system was numerically solved on a 2D square by a finite-element method (COMSOL Multiphysics software).

5.4 Results

5.4.1 Theoretical Basis for Understanding CLN Formation Process

Considering the cell traction force coefficient amplitude τ as a bifurcation parameter, the linear stability analysis performed in a neighbourhood of the normalised homogeneous steady state establishes that non-homogeneous steady states of cell and biogel density will emerge beyond a critical value τ_c of the cell traction amplitude given by [16]

$$\tau_c = \frac{(\lambda + 2\mu) + 2\sqrt{s(\lambda\beta_2 + 2\mu\beta_1)}}{\frac{h}{D_0}(N_2 - 2) + 0.5(3N_2 - 4)}. \tag{5.10}$$

This analytical expression highlights the antagonistic effects on one side, the stabilizing influence of the biogel mechanical stiffness (Lamé coefficients λ and μ, long-range elasticity parameters β_1 and β_2), on the other side, the heterogeneous cell aggregation driven by the ratio h/D_0 between active cell migration and passive cellular diffusion. Thus, the coupling between cellular traction and active cell migration behaves as a nonlinear *strain-induced strain increase* (SISI) autocatalytic process.

5.4.2 Simulation of CLN Morphogenesis

According to EC type and fibrin concentration, significant differences in time scales have been reported for the development of *in vitro* tubulogenesis [26], [23]. For
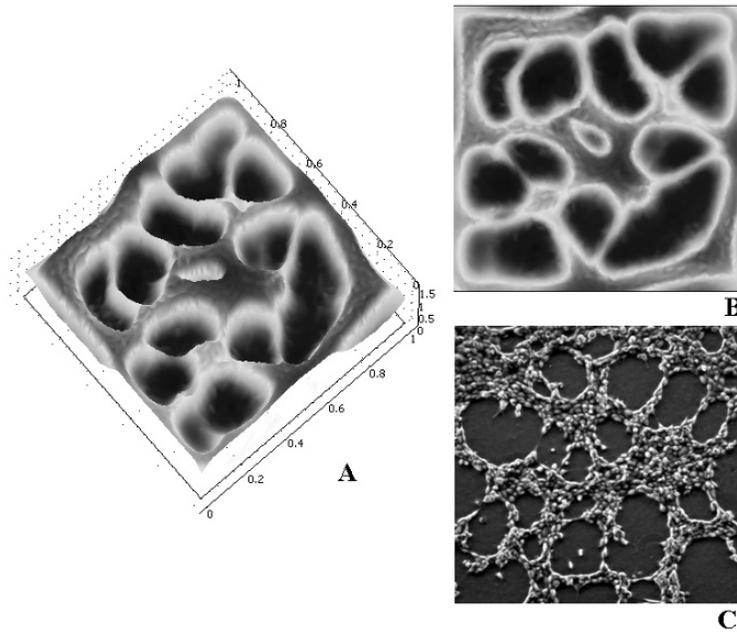
Fig. 5.3. Simulated CLN formation within viscoelastic fibrin biogels. The 3D profile of the biogel density, which correlates linearly with the biogel thickness, shows lacunae with different sizes and shapes, separated by thin cords of extracellular material (Fig. 5.3A). The associated upper view of the CLN pattern (Fig. 5.3B) corresponds very satisfactorily to the experimental observations (Fig. 5.3C).

HUVEC cultured on fibrin gel, lacunae appear typically in less than 2 hours, while CLN formation is achieved after 24 hours, with lacunae size ranging from 0.2 mm to 1.5 mm. On the basis of these experiments, we simulated and analysed the dynamical behaviour of the mechanocellular model for different fibrin gel concentrations. Fig. 5.3 illustrates the morphogenesis of CLN for a cell traction amplitude τ_1 above the critical bifurcation value τ_c defined previously. The initially smooth gel density progressively deforms up to the formation of pronounced depressions within the biogel (Fig. 5.3A), these lacunae being separated by more or less thin cords of extracellular material, giving rise to a morphogenetic pattern which looks very much like the CLN patterns observed during the early phases of the CLN morphogenesis process (Fig. 5.3B–5.3C).

When focusing more accurately on the time course of the CLN formation, we can observe both an increase of the mean lacuna size and a decrease in the number of lacunae (Fig. 5.4A–5.4B). The simulated CLN morphogenesis reproduces these experimental observations, as illustrated in Fig. 5.4C–5.4D: lacunae morphogenesis takes place within twelve hours, and leads to successive network topologies very similar to the experimental observations. The formation of CLN is associated with an intense mechanical remodelling of the biogel, as illustrated by the 2D map of the stress field
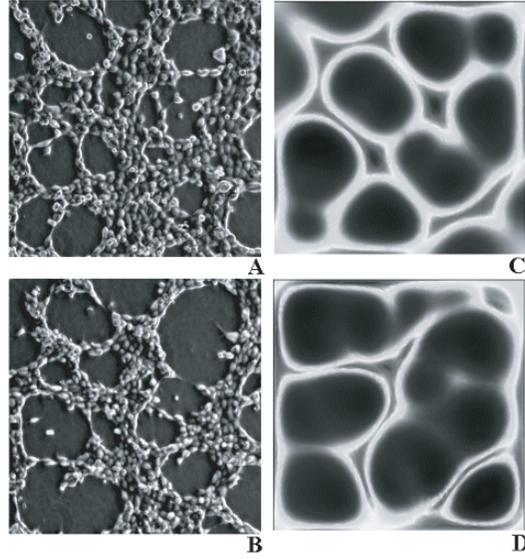
Fig. 5.4. Morphologies of CLN at two successive times. Mechanical stresses generated by cellular forces induce the enlargement of the lacunae (Fig. 5.4A–5.4B). The simulated morphogenesis (Fig. 5.4C–5.4D) reproduces quite well the progressive modification of the CLN observed experimentally by phase-contrast videomicroscopy: cellular cords may break, thus decreasing the number of lacunae.

which progressively develops within the gel. In Fig. 5.5, we computed the von Mises stresses σ_{VM}:

$$\sigma_{VM} = \sqrt{\frac{3}{2} \sum_{i,j} s_{ij}^2}, \tag{5.11}$$

where s_{ij} are the components of the deviatoric stress tensor. Assuming that the vertical stress components are negligible ($\sigma_{3i} = \sigma_{i3} = 0$), we have

$$s_{11} = \frac{E}{3(1 - v^2)} \left((2 - v)\epsilon_{11} + (2v - 1)\epsilon_{22}\right) \tag{5.12}$$

$$s_{22} = \frac{E}{3(1 - v^2)} \left((2v - 1)\epsilon_{11} + (2 - v)\epsilon_{22}\right) \tag{5.13}$$

$$s_{12} = \frac{E}{(1 + v)}\epsilon_{12}. \tag{5.14}$$

In Fig. 5.5, isolevels of cell densities have been superimposed on the von Mises stresses and plotted at two successive times (left, $t = 1.3$ h, right, $t = 9.1$ h) during the tubulogenesis process. Regions of maximal stresses co-localize with cords of compressed extracellular material and of cell aggregates and tubes. Stresses induced by stretching in the bottom of the lacunae are also visible on these stress maps.
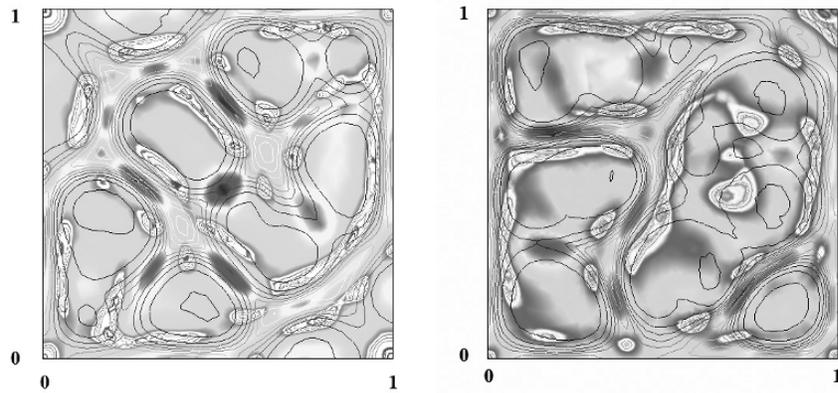
Fig. 5.5. 2D von Mises stress maps associated with the simulated CLN morphologies shown in Figs. 5.4C and 5.4D, respectively. Grey color code ranges from 0.6 (white areas) to 60 Pa (black areas). Superimposing isolevels of cell density clearly shows cell aggregation along the lines of maximal stresses.

5.4.3 Localized Formation of Cellular Networks: ECM as a Mechanical Support for Positional Information

The concept of positional information is commonly used to explain how cell differentiation during organism and organ development is controlled by the position of the cell. Following Turing's original proposition, such spatial information can result from gradients of diffusible molecules, called morphogens, secreted from a spatially localized source. Even if several morphogens have been identified, the formation, readability and robustness of morphogen gradients is still poorly understood and a subject of recent theoretical development [4], [5]. We illustrate here how the ECM geometrical and mechanical properties may provide robust and long-ranged positional information for the emergent cellular process by considering CLN formation over an ECM with varying thickness (Fig. 5.6A). According to our model formulation, this spatial variation of ECM thickness modulates the elastic restoring body force R, whose amplitude decreases with increasing ECM thickness. The simulated cellular pattern is shown in Fig. 5.6B. One can observe a localized formation of CLNs in the central part of the biogel, i.e., in a region where this substrate is thick enough. In other words, the gradient of the biogel thickness provides a positional information which, combined with appropriate force balance between cell traction forces and biogel stiffness, leads to the emergence of CLNs only at specific locations. Such a theoretically predicted tubulogenesis is supported by experiments exhibiting the influence of gradients of ECM thickness on the formation of cellular networks. Indeed, [25] reported that bovine aortic cells cultured on a matrix with increasing thickness gradient self-organised as CLNs only in the thickest region of this extracellular substrate.
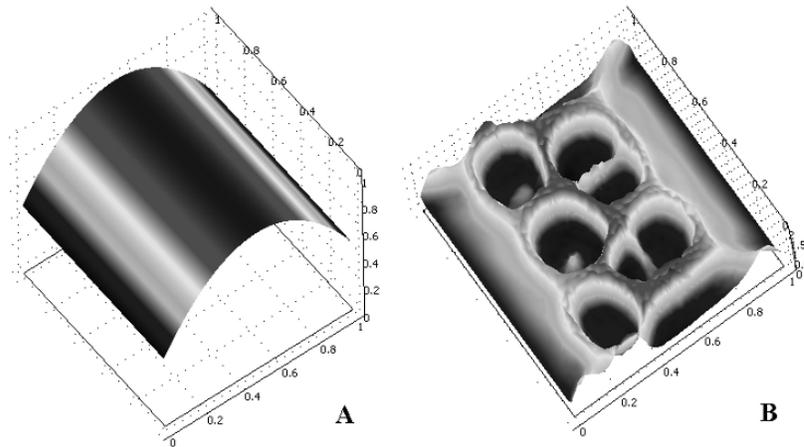
Fig. 5.6. Localised morphogenesis of CLN driven by an initial gradient of biogel thickness. The bell shape variation of the biogel thickness (Fig. 5.6A) acts as a distributed source of positional information which induces CLN formation only in the central part of the biogel.

5.5 Discussion

An increasing body of experimental data indicates that local changes in the cellular force balance may mediate pattern formation during tissue morphogenesis, as especially well demonstrated by *in vitro* angiogenesis assays [10]. In this context, the pioneering theoretical framework developed by J.D. Murray and G. Oster starting in the 1980s [14], [15] appears clearly as a highly valuable and experimentally based approach to deal with the fundamental question of how self-organised cellular networks emerge from collective mechanical cell-matrix interactions [13], [21], [20]. However, other modelling approaches have been developed during recent years (reviewed in [1]), based for example on the percolative transition concept [8]. This latter approach was successfully used in the analysis of *in vitro* angiogenesis experiments conducted on thin Matrigel film (thickness of 44 μm) [8]. In these experimental conditions, the observed multicellular network morphogenesis is mainly driven by cell-cell adhesion and chemotactic migration induced by the autocrine cell response to the gradient of VEGF. In our experiments, different processes drive the network morphogenesis, since CLN formation will not occur on thin fibrin gels, as observed experimentally and pointed out theoretically in Section 5.4.3. On the contrary, our experimental protocol uses biogel 1 mm thick on which the leading role of cellular forces may be observed, since then they are able to significantly deform the fibrin gel.

In order to understand how the morphogenesis of vascular-like networks is controlled by cellular force balance, we analysed the dynamical properties of a theoretical model coupling cell motility behaviour to the mechanical status of the surrounding extracellular matrix. Changes of biogel stiffness act as a mechanical switch which triggers CLN morphogenesis. In addition our model allows us to compute the mechanical status of the ECM during CLN morphogenesis. Indeed, the tension lines created by

the cells as well as the time-varying strain and stress maps cannot be obtained directly in the time course of the experiments. Such information can only be obtained indirectly, by solving displacement-force inverse problems using elastography methods. However, these data on stress and strain evolution are crucial for our understanding of both *in vitro* and *in vivo* angiogenesis since, as stated in the Introduction, mechanical stresses and strains are increasingly recognised as effective regulators of a wide spectrum of cell functions. Nevertheless, this argument also underlines the limitations of our modelling approach in its present state. First, the CLN morphogenesis is analysed within a linearised formulation of the stress-strain relationships. Second, we do not take into account the feedback of the biogel mechanical status on cell metabolic activity, including proteolytic activity and secretion of diffusible chemoattractant factors. Since the experimental counterpart we considered is based on *in vitro* models using fibrin gels prepared from purified fibrinogen, thus devoid of other angiogenic factors, we did not take into account the cell response to soluble stimuli. But thanks to the mechanochemical framework used here, this later aspect, considered in recent theoretical models of *in vitro* angiogenesis [18], [19], may be introduced as an additional modulator of the CLN formation process described here. On the other hand, modelling how cells translate the variation of ECM rigidity into downstream responses remains far less obvious. This pleads for the development of multiscaled models, which should be able to integrate the effect of micro and nano stress- or strain-activated processes, ranging from mechanical control of cell adhesion to the up or down regulation of mechanosensitive gene expression.

Acknowledgment

References

1. Ambrosi, D., Bussolino, F., Preziosi, L.: A review of vasculogenesis models. J. Theor. Med., **6**, 1–19 (2005).
2. Benkherourou, M., Rochas, C., Tracqui, P., Tranqui, L., Gumery, P.Y.: Standardization of a method for characterizing low-concentration biogels: elastic properties of low-concentration agarose gels. J. Biomech. Eng., **121**, 184–187 (1999).
3. Benkherourou, M., Gumery, P.Y., Tranqui, L. Tracqui, P.: Quantification and macroscopic modeling of the nonlinear viscoelastic behavior of strained gels with varying fibrin concentrations. IEEE Trans. Biomed. Eng., **47**, 1465–1475 (2000).
4. Bollenbach, T., Kruse, K., Pantazis, P., Gonzalez-Gaitan, M., Jülicher, F.: Robust formation of morphogen gradients. Phys. Rev. Lett., **94**, 0181031–0181034 (2005).
5. Cinquin, O: Fast-tracking morphogen diffusion. J. Theor. Biol., **238**, 532–540 (2005).
6. Cook, J.: Mathematical models for dermal wound healing: wound contraction and scar formation. Ph.D. thesis, 98–133. University of Washington, Seattle (1995).
7. Cruywagen, G.C., Maini, P.K., Murray, J.D.: Biological pattern formation on two-dimensional spatial domains: A nonlinear bifurcation analysis. SIAM J. Appl. Math., **57**, 1485–1509 (1997).

8. Gamba, A., Ambrosi, D., Coniglio, A., de Candia, A., Di Talia, S. Giraudo, E., Serini, G., Preziosi, L., Bussolino, F.: Percolation, morphogenesis and burgers dynamics in blood vessels formation. Phys. Rev. Lett., **90**, 118101-1–118101-4 (2003).

9. Gerhardt, H., Golding, M., Fruttiger, M., Ruhrberg, C., Lundkvist, A., Abramsson, A., Jeltsch, M., Mitchell, C., Alitalo, K., Shima, D., Betsholtz, C.: VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. J. Cell Biol., **161**, 1163–1177 (2003).

10. Ingber, D.E.: Mechanical signaling and the cellular response to extracellular matrix in angiogenesis and cardiovascular physiology. Circ. Res., **91**, 877–887 (2002).

11. Korff, T., Augustin, H.G.: Tensional forces in fibrillar extracellular matrices control directional capillary sprouting. J. Cell Sci., **112**, 3249–3258 (1999).

12. Lo, C.M., Wang, H.B., Dembo, M., Wang, Y.L.: Cell movement is guided by the rigidity of the substrate. Biophys. J., **79**, 144–152 (2000).

13. Manoussaki, D., Lubkin, S.R., Vernon, R., Murray, J. D.: A mechanical model for the formation of vascular networks in vitro. Acta Biotheor., **44**, 271–282 (1996).

14. Murray, J.D., Oster, G.F.: Cell traction models for generating pattern and form in morphogenesis. J. Math. Biol., **19**, 265–279 (1984).

15. Murray, J.D.: Mathematical Biology, Vol. I, II. Springer, New York (2003).

16. Namy, P., Ohayon, J., Tracqui, P.: Critical conditions for pattern formation and in vitro tubulogenesis driven by cellular traction fields. J. Theor. Biol., **227**, 103–120 (2004).

17. Pelham Jr., R.J., Wang, Y.: Cell locomotion and focal adhesions are regulated by substrate flexibility. Proc. Natl. Acad. Sci. USA, **94**, 13661–13665 (1997).

18. Serini, G., Ambrosi, D., Giraudo, E., Gamba, A., Preziosi, L., Bussolino, F.: Modeling the early stages of vascular network assembly. EMBO J., **22**, 1771–1779 (2003).

19. Tosin, A., Ambrosi, D., Preziosi, L.: Mechanics and chemotaxis in the morphogenesis of vascular networks. Bull. Math. Biol., **68**, 1819–1836 (2006).

20. Tracqui, P.: Mechanical instabilities as a central issue for in silico analysis of cell dynamics. Proceedings IEEE, **94**, 710–724 (2006).

21. Tranqui, L., Tracqui, P.: Mechanical signalling and angiogenesis. The integration of cell-extracellular matrix couplings. C. R. Acad. Sci., **323**, 31–47 (2000).

22. Vailhé, B., Ronot, X., Tracqui, P., Usson, Y., Tranqui, L.: In vitro angiogenesis is modulated by the mechanical properties of fibrin gels and is related to avb3 integrin localization. In Vitro Cell. Dev. Biol. Anim., **33**, 763–773 (1997).

23. Vailhé, B., Lecomte, M., Wiernsperger, N., Tranqui, L.: The formation of tubular structures by endothelial cells is under the control of fibrinolysis and mechanical factors. Angiogenesis, **2**, 331–344 (1998).

24. Vailhé, B., Vittet, D., Feige, J.-J.: In vitro models of vasculogenesis and angiogenesis. Lab. Invest. **81**, 439–452 (2001).

25. Vernon, R.B., Angello, J.C., Iruela-Arispe, M.L., Lane, T.F., Sage, E.H.: Reorganization of basement membrane matrices by cellular traction promotes the formation of cellular networks in vitro. Lab. Invest. **66**, 536–547 (1992).

26. Vernon, R.B., Lara, S.L., Drake, C.J., Iruela-Arispe, M.L., Angello, J.C., Little, C.D., Wight, T.N., Sage, E.H.: Organized type I collagen influences endothelial patterns during "spontaneous angiogenesis in vitro": planar cultures as models of vascular development. In Vitro Cell. Dev. Biol. **31**, 120–131 (1995).

Nonexponential Time Distributions in Biocatalytic Systems: Mass Service Replacing Mass Action

Peter W. Kühl[1] and Manfred Jobmann[2]

[1] Institute of Theoretical Biology, Schaulistr. 2, CH-4142 Münchenstein BL, Switzerland
Peter-W.Kuehl@unibas.ch
[2] Department of Informatics, Technische Universität München, Boltzmannstr. 3, D-85748
Garching, Germany; jobmann@in.tum.de

Summary. Formal kinetic methods to analyze biocatalytic systems are traditionally based on the law of mass action. This law involves the assumption that each molecular state has an exponentially distributed lifetime. We regard this assumption as unduly restrictive and advocate a more general, service theory based approach (termed mass service kinetics or, briefly, service kinetics). In service-theoretic terms biocatalysts are servers and their ligands are customers. The time intervals between arrivals of ligand molecules at special service loci (active or binding sites) as well as the service periods at these loci need not be exponentially distributed; rather, they may adopt any distribution (e.g., Erlangian, hyperexponential, variomorphic). We exemplify the impact of nonexponential time distributions on a performance measure of wide interest: the steady-state throughput. For its computation we use matrix-analytic methods. Specifically, we show that nonexponential interarrival times convert hyperbolic mass action systems (whose characteristic is a hyperbolic velocity-concentration or dose-response curve) into nonhyperbolic mass service systems and that the type and extent of their nonhyperbolicity are determined by the type and parameters of the interarrival time distribution. Furthermore, we analyze the combined effect of a non-Poissonian arrival process and a waiting site near the catalyst's active site on the throughput of the system. A major conclusion of our and other studies is that it is a questionable practice to routinely and exclusively use mass action kinetics for the interpretation and performance evaluation of biocatalytic systems.

Key words: Nonhyperbolic velocity-concentration curves, queueing theory, interarrival time, waiting space, steady-state throughput.

6.1 Introduction

The kinetic behaviour of biocatalytic systems (i.e., systems containing biological macromolecules that catalyze chemical transformations or nutrient and metabolite transport or signal transduction) is traditionally described by mass action kinetics. The latter presupposes the validity of the law of mass action. Though usually formulated nonstochastically (by ordinary differential equations or algebraic expressions), the law of mass action clearly has stochastic roots and must be interpretable as the result of probabilistic phenomena at the molecular level by virtue of the fact that matter

is not a continuum but consists of discrete units (atoms, molecules) with an intrinsically random behaviour. C. M. Goldberg (1836–1902) and P. Waage (1833–1900), the originators of the law of mass action, were already aware of this, and since then numerous authors, e.g., Rényi [1], have investigated both the kinetic and equilibrium versions of the law of mass action in terms of probability and stochastic processes. It is beyond the scope of this article to compile and review the pertinent literature. We rather confine ourselves to mention two fundamental stochastic properties of the law of mass action: (i) ergodicity (= equality of time and ensemble averages) and (ii) Markovity (= lack of memory, complete forgetfulness). If a system is nonergodic and/or non-Markovian, it is by definition not a mass action system. Deviations from ergodicity, reported in experimental studies on conformational transitions in biological macromolecules (e.g., [2,3]), are not a topic of the present paper. We rather focus on violation of the second above-mentioned property, the Markov property, by allowing for nonexponential interevent times at the molecular level. How can one cope with biocatalytic systems containing non-Markovian elements, and to what extent does the performance of such systems differ from that of conventional mass action systems? These are the main issues we address here.

6.2 The Service-Theoretic Approach

We propose to apply—instead of the traditional theory of mass action—the "theory of mass service" (also known under the names queueing theory and service theory) to biochemical kinetics for a number of conceptual and methodological reasons: (i) Biocatalysts are stochastic "servers" and their ligands are "customers" which are served at special "service loci" (active or binding sites). Various modes of operation (waiting or rejection, first come-first served, service interruptions, priority rules, etc.), being well known in man-made service systems, may also be encountered in biocatalytic systems. Furthermore, biocatalysts can build up chains and networks endowed with structural, dynamic and regulatory properties that are largely analogous to those of production lines or queueing networks of anthropic origin. (ii) For the quantitative analysis and performance evaluation of service systems and queueing networks, probability theorists have developed a rich repertoire of mathematical tools and methods. At least some of these are also expected to be applicable and useful in the nanoworld of enzymes, transporters and signal transducers. (iii) Nonexponential time distributions, being foreign to mass action systems, are quite common in mass service systems. It is primarily for this reason that we advocate the application of the service-theoretic approach to biocatalytic systems and the build-up of a more general type of kinetics, termed—in contradistinction to mass action kinetics—*mass service kinetics* or briefly *service kinetics*. Biokinetics is thus put on a broader fundament, gains in flexibility and versatility and may lead to new interpretations of old phenomena. In the next section we use a very simple enzymic model in order to give an idea of the potentialities of the service-kinetic approach in a biochemical setting.

6.3 The Van Slyke–Cullen Scheme Modelled as a Service-Kinetic System

The simplest possible scheme of an enzymic reaction is

$$E + S \rightarrow ES \rightarrow E + P$$

where E, S, ES and P stand for enzyme, substrate, enzyme-substrate complex and product, respectively. This scheme has been named after Van Slyke and Cullen [4] (VC) since these authors ignored in their mathematical analysis the (commonly included) reverse reaction ES $\rightarrow$ E + S. We, too, ignore this reaction since its inclusion usually affects only the scale but not the shape of the v(S) curve (v = reaction velocity; S = substrate concentration). We use the VC scheme as the basic paradigm of a biocatalytic system.

The mass-action kinetic analysis of the VC scheme leads to only one type of v(S) curve: the rectangular hyperbola. An entirely different situation is encountered when one subjects the VC scheme to a (mass-) service kinetic analysis: besides the familiar hyperbolic v(S) curve a great variety of nonhyperbolic v(S) curves can be obtained. This diversity is achieved by allowance for (i) nonexponential time distributions and/or (ii) a waiting room for substrate molecules which sit and wait there for clearance of the active site. In Figs. 6.1–6.3 some illustrative examples of nonhyperbolic v(S) curves as consequences of (i) and/or (ii) are shown.

Let us have a closer look at the various situations giving rise to the v(S) curves of Figs. 6.1–6.3.

Figure 6.1. When the interarrival times of the substrate molecules at the enzyme's active centre are nonexponential, the resulting nonhyperbolic v(S) curves are—compared with the classical hyperbola—either "lifted" or "depressed," depending on whether the coefficient of variation of the interarrival time distribution is smaller or larger than 1 (see curves (a), (b), and (e), respectively). A special situation is represented by curve (d): here the stochastic arrival pattern varies with the arrival intensity (i.e., with the substrate concentration); specifically, its coefficient of variation gradually decreases with increasing substrate concentration in such a way that a sigmoidal v(S) curve is generated. We regard arrival patterns with an intensity-dependent coefficient of variation (called "variomorphic" in our earlier work [5]) not as an exotic rarity but as the rule in biocatalytic systems whose ligand arrival process is nonexponential. A more detailed description of nonexponential and variomorphic arrival processes and the resulting v(S) or dose-response curves can be found in [5].

Figure 6.2. When there exist near the enzyme's active centre waiting spaces from which substrate molecules move to the active centre as soon as the latter is cleared, one usually obtains lifted nonhyperbolic v(S) curves. The extent of lifting and nonhyperbolic "deformation" depends on (i) the number of available waiting spaces, (ii) the time needed for the substrate's transit from the waiting space to the active centre, and—except for the limiting case of an infinite number of waiting spaces—(iii) the probability distributions of interarrival and service times. In Fig. 6.2 we assumed that the mentioned transit time is negligibly small and both interarrival and service times
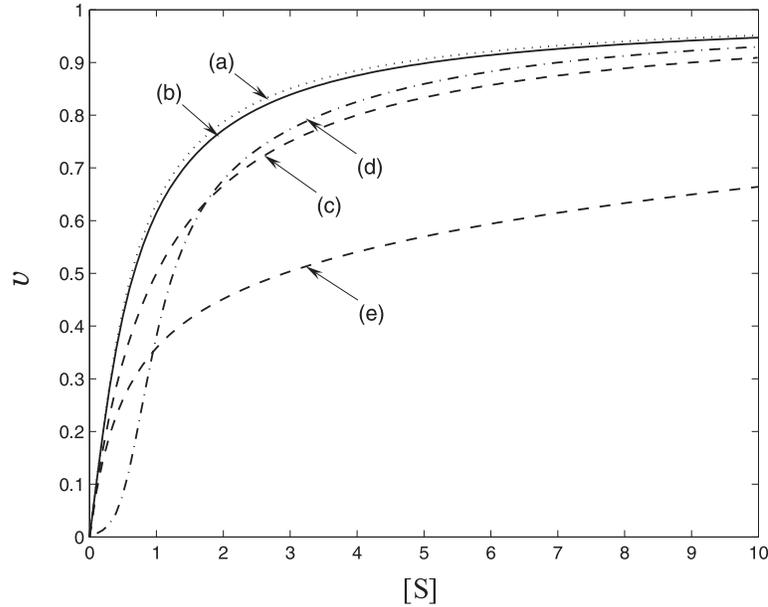
Fig. 6.1. The effect of nonexponential interarrival time distributions on the throughput in a VC system without waiting facilities. [S] and v stand for substrate concentration and reaction velocity, respectively. [S] is given in units of the Michaelis constant K_m and v as a fraction of the maximal throughput. The lowercase letters attached to the curves indicate various interarrival time distributions: (a) = deterministic, (b) = Erlangian of order 10, (c) = exponential, (d) = variomorphic with a coefficient of variation decreasing nonlinearly from 14.1 at [S] = 0 to about 0.7 at [S] = 2.2 and (e) = hyperexponential of order 2 with a coefficient of variation of 4. The computational techniques (based on matrix-analytic methods) used here as well as in Figs. 6.2 and 6.3 for generating the v(S) curves are described in [5]. For readers familiar with the symbolic representation of service systems (explained in [5]) we add that curves (a), (b), (c), (d) and (e) represent the throughputs in the service systems $G/M/1/0$ and $G^{\text{vario}}/M/1/0$ with (a) $G = D$, (b) $G = E_{10}$, (c) $G = M$, (d) $G^{\text{vario}} = PH_4^{\text{vario}}$ and (e) $G = H_2$, respectively.

are exponentially distributed. Under these conditions allowance for 1, 2, 3, 10, 100 or infinitely many waiting spaces yields the v(S) curves shown. The larger the number of waiting spaces, the more the v(S) curves approach the form of a ramp with a sharp transition at [S] = K_m, the Michaelis constant. The latter type of kinetics is phenomenologically equivalent to the Blackman kinetics [6,7] in biochemistry and physiology and the Holling type 1 "functional response" [8,9] in ecology. The waiting room effect on the throughput in a purely exponential enzyme system was reported earlier by Trenkenshu [10,11].

Figure 6.3. Here we illustrate the combined effect of two service-kinetic possibilities: a nonexponential arrival process and a waiting room with space for just one substrate molecule. Analogously to the purely exponential system of Fig. 6.2, in this system allowance for one waiting space suffices to considerably increase the system's
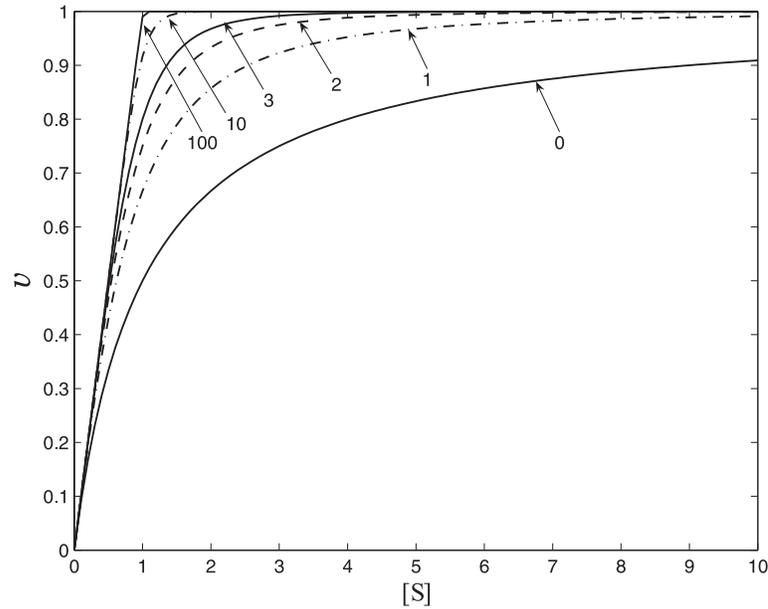
Fig. 6.2. The effect of waiting spaces on the throughput in a purely exponential VC system. The numbers attached to the curves indicate the number K of waiting spaces near the enzyme's active centre. [S] and v are normalized as in Fig. 6.1. For further details see the text. For readers familiar with the symbolic representation of service systems (explained in [5]) we add that the curves labelled 0 to 100 represent the throughputs in the service system $M/M/1/K$ with $K = 0$, $1, 2, 3, 10$, and 100, respectively. Note that $K = 100$ already approximates very well the case of $K = \infty$.

throughput (see the $v(S)$ curves in pairs (b), (d) and (e) of Fig. 6.3). However, this increase is not equal for systems with exponential and nonexponential interarrival time distributions—neither absolutely nor as percentage values, regardless of being measured at selected substrate concentrations, e.g., at $[S] = K_m$, or when summed up over the whole substrate concentration range $(0, \infty)$.

The $v(S)$ curves shown in Figs. 6.1–6.3 all originate from a VC scheme with an exponential service time distribution. What do $v(S)$ curves look like when (case 1) solely the service time or (case 2) both the service and interarrival times are nonexponentially distributed? In answering this question we have to differentiate between VC systems with waiting facilities and those without them.

Answer in case 1. In a VC system possessing no waiting room one always obtains one and the same curve shape: the classical hyperbola. In other words, this system is insensitive to the stochastic character of the service process (provided the system is at steady state as we tacitly assume throughout this chapter). However, this insensitivity is abolished when the enzyme molecule can offer a finite number of waiting spaces to substrate molecules arriving at an occupied active centre. In this case the VC system's
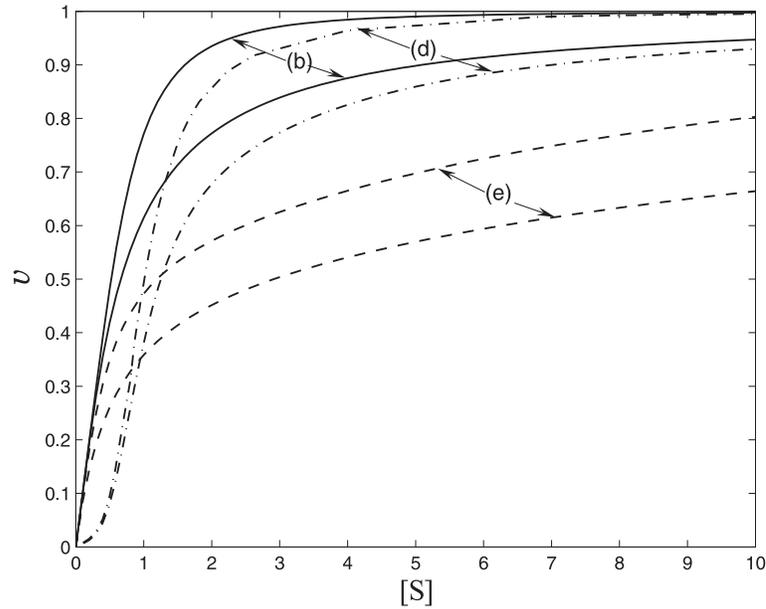
Fig. 6.3. The effect of nonexponential interarrival time distributions and one waiting space on the throughput in the VC system. [S] and v are normalized as in Fig. 6.1. In pairs (b), (d) and (e) the interarrival time distributions are identical with those used in Fig. 6.1 for generating the curves (b), (d) and (e), respectively. The upper (lower) curve in each pair of curves represents the throughput in a VC system with one (no) waiting space. When expressed symbolically [5], the upper and lower curves represent the throughputs in the service systems (b) $E_{10}/M/1/1$ and $E_{10}/M/1/0$, (d) $PH_4^{\mathrm{vario}}/M/1/1$ and $PH_4^{\mathrm{vario}}/M/1/0$ and (e) $H_2/M/1/1$ and $H_2/M/1/0$.

throughput is better (i.e., the $v(\mathrm{S})$ curves are more lifted), the smaller the coefficient of variation of the service time (data not shown).

Answer in case 2. In a VC system devoid of waiting facilities the $v(\mathrm{S})$ curves often do not differ much when both the interarrival time and the service time are nonexponential. However, special conditions—e.g., arrival and service processes being both Erlangian of high order—can give rise to quite unusual (undulatory or sawtooth-like) $v(\mathrm{S})$ curves which markedly differ from those in systems with an identical nonexponential arrival process but an exponential server; for illustrative examples see [5]. Finally, when comparing the position and shape of $v(\mathrm{S})$ curves obtainable in VC systems equipped with a finite number of waiting spaces, one again finds differences between doubly (arrival *and* service) and singly (arrival *or* service) nonexponential systems. These differences become smaller with an increasing number of waiting spaces and totally vanish when this number goes to infinity. In the latter case the $v(\mathrm{S})$ curve adopts the ramp shape shown in Fig. 6.2 and it is irrelevant whether the interarrival and/or service time distributions are exponential or nonexponential (case of total insensitivity).

6.4 General Conclusions and Open Questions

Using the VC scheme as the basic paradigm of a biocatalytic system, we showed in the preceding section that a mass service theory inspired approach opens up prospects and possibilities that are closed to the traditional mass action approach. In particular we demonstrated that allowance for nonexponential time distributions and incorporation of waiting facilities for arriving substrate molecules yield a wide spectrum of $v(S)$ curves of which the classical hyperbola is merely a special case occurring under special, quite restrictive conditions. The mass service approach is of course also applicable to reaction schemes more complex than the VC scheme treated above. If in such schemes conventional mass action kinetics already leads to nonhyperbolic $v(S)$ or dose-response curves, the inclusion of mass-service specific elements may either reinforce or attenuate the system's deviations from the hyperbolic response. However complicated a reaction scheme may be, it is clear that service-theoretic concepts and methods are apt to augment our understanding of biocatalytic systems. Time-honoured mass action kinetics should therefore be supplemented (if not replaced) by mass service kinetics. Other authors [12–14], using concepts and methodological approaches different from ours (e.g., a deterministic fractal approach [12] or Monte Carlo simulation algorithms [13,14]), have also emphasized that conventional mass action kinetics is often inadequate to describe biocatalytic systems, especially under in vivo conditions.

Finally, we would like to touch upon the question: What physical mechanisms can be envisaged to give rise to nonexponential time distributions at the molecular level? As far as the interarrival time distribution is concerned, we consider the following possibilities: (i) Substrate molecules are often products of a neighbouring enzyme in a metabolic chain. If these product molecules are released in a nonexponential manner and are—due to channelling—not (or not fully) thermally randomized, their arrival pattern at the next enzyme is also nonexponential. (ii) Coherent spiking of enzymic reactions in small volumes, as described by Mikhailov and Hess [15], may more or less derandomize the stream of arriving ligand molecules. (iii) The intracellular milieu which is characterized by extensive compartmentalization, macromolecular crowdedness, lacunarity and a non-Euclidian geometry may generate unusual, e.g., power-tailed, arrival patterns. (iv) The matrix process [16], i.e., the intramatrix migration of ligand molecules between the first (or last) contact point on the biocatalyst's surface and the active centre, may not only change the intensity with which ligand streams arrive at the active centre but may also modify their stochastic pattern by various derandomizing mechanisms (e.g., overflow, regularizing filtering, alternative gated pathways). The direction and degree of "nonexponentialization" may often be dependent on the ligand concentration (e.g., exponential at low [S] and increasingly hypo- or hyperexponential at rising [S]) and thus give rise to what are called variomorphic arrival patterns (see [5] and Sect. 6.3 above). Concerning the service time distribution, we mention two further feasible mechanisms of nonexponentialization: (v) The biocatalyst has to undergo a number of sequential conformational transitions or has to perform a number of sequential chemical or physical operations before the catalytic act (formation or fission of a covalent bond, emission of a signal, transportation of a

nutrient across a membrane, etc.) can occur. Sequential multistage processes of this kind make the service time hypoexponential (Erlangian) though each individual stage is exponentially distributed. (vi) The biocatalyst's conformational state required for the catalytic act may be reached by two or more parallel single- or multiple-step pathways which leads to a service time with hyperexponential or hyper-Erlangian distribution.

The above-listed mechanisms of nonexponentialization in biocatalytic systems are largely hypothetical, and direct experimental evidence supporting or invalidating them is scarce so far. However, it is expected that the rapidly advancing experimental techniques in single-molecule biochemistry (e.g., sophisticated optical techniques [17, 18] with high time resolution) will provide unambiguous data concerning nonexponential time distributions and enable open questions to be answered or reformulated.

References

1. Rényi, A.: A discussion of chemical reactions using the theory of stochastic processes. In: Turán, P. (ed) Selected Papers of Alfréd Rényi, Volume 1, 1948–1956, pp. 367–380. Akadémiai Kiadó, Budapest (1976).
2. Shyamsunder, E.: Broken ergodicity in myoglobin. Ph.D. Thesis, University of Illinois at Urbana-Champaign (1986).
3. Wennmalm, S., Edman, L., Rigler, R.: Non-ergodic behaviour in conformational transitions of single DNA molecules. Chem. Phys., **247**, 61–67 (1999).
4. Van Slyke, D.D., Cullen, G.E.: The mode of action of urease and of enzymes in general. J. Biol. Chem., **19**, 141–180 (1914).
5. Kühl, P.W., Jobmann, M.: Receptor-agonist interactions in service-theoretic perspective. Effects of molecular timing on the shape of dose-response curves. J. Recept. Sign. Transd., **26**, 1–34 (2006).
6. Blackman, F.F.: Optima and limiting factors. Ann. Bot., **19**, 281–295 (1905).
7. Dabes, J.N., Finn, R.K., Wilke, C.R.: Equations for substrate-limited growth: the case for Blackman kinetics. Biotechnol. Bioeng., **25**, 1159–1177 (1973).
8. Holling, C.S.: The components of predation as revealed by a small-mammal predation of the European pine sawfly. Canad. Entomol., **91**, 293–320 (1959).
9. Taylor, R.J.: Predation. Chapman & Hall, New York and London (1984).
10. Trenkenshu, R.P.: Effect of metabolic bottleneck organization on kinetics of enzyme substrate conversion. Mol. Biol., **22**, 1170–1177 (1988).
11. Trenkenshu, R.P.: Application of the queueing theory to biokinetics. In: Shokin, Y.I. (ed) Evolution Modeling and Kinetics: Collection of Scientific Works, pp. 125–160. Nauka, Novosibirsk (1992) (in Russian).
12. Savageau, M.A.: Michaelis–Menten mechanism reconsidered: implications of fractal kinetics. J. Theor. Biol., **176**, 115–124 (1995).
13. Berry, H.: Monte Carlo simulations of enzyme reactions in two dimensions: fractal kinetics and spatial segregation. Biophys. J., **83**, 1891–1901 (2002).
14. Schnell, S., Turner, T.E.: Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. Progr. Biophys. Mol. Biol., **85**, 235–260 (2004).
15. Mikhailov, A., Hess, B.: Microscopic self-organization of enzymic reactions in small volumes. J. Phys. Chem., **100**, 19059–19065 (1996).

16. Stetzkowski, F., Banerjee R., Marden, M.C., Beece, D.K., Bowne, S.F., Doster, W., Eisenstein, L., Frauenfelder, H., Reinisch, L., Shyamsunder, E., Jung, C.: Dynamics of dioxygen and carbon-monoxide binding to soybean leghemoglobin. J. Biol. Chem., **260**, 8803–8809 (1985).
17. Min, W., English, B.P., Luo, G., Cherayil, B.J., Kou, S.C., Xie, X.S.: Fluctuating enzymes: lessons from single-molecule studies. Acc. Chem. Res., **38**, 923–931 (2005).
18. Tinnefeld, P., Sauer, M.: Branching out of single-molecule fluorescence spectroscopy: challenges for chemistry and influence on biology. Angew. Chem. Int. Ed., **44**, 2642–2671 (2005).

Part II

Regulatory Networks

7

A Stochastic Model of Gene Regulation Using the Chemical Master Equation

Hilary S. Booth,[1] Conrad J. Burden,[2,3] Markus Hegland,[4,5] and Lucia Santoso[2,5]

[1] This paper is dedicated to Dr. Hilary Booth who died unexpectedly on May 26, 2005.
[2] Centre for Bioinformation Science, Mathematical Sciences Institute, Australian National University, Canberra, A.C.T. 0200, Australia; conrad.burden@anu.edu.au
[3] John Curtin School of Medical Research, Australian National University, Canberra, A.C.T. 0200, Australia
[4] Centre for Mathematics and Its Applications, Mathematical Sciences Institute, Australian National University, Canberra, A.C.T. 0200, Australia
[5] ARC Centre in Bioinformatics, The University of Queensland, St. Lucia, QLD. 4072, Australia

Summary. The chemical master equation in combination with chemical rate equations is used as a tool to study Markovian models of genetic regulatory networks in prokaryotes. States of the master equation represent the binding and unbinding of protein complexes to DNA, resulting in a gene being expressed or not expressed in a cell, while protein and substrate concentrations are represented by continuum variables which evolve via differential equations.

The model is applied to a moderately complex biological system, the switching mechanism of the Bacteriophage λ driven by competition between production of CI and Cro proteins. Numerical simulations of the model successfully move between lysogenic and lytic states as the host bacterium is stressed by the application of ultraviolet light.

Key words: Gene regulation, chemical master equation, bacteriophage lambda.

7.1 Modelling Gene Regulatory Networks

One of the major challenges in bioinformatics is to determine how genes are regulated and how their products interact within cellular networks. In a complex cell, gene products and external factors regulate the genes that are expressed in that cell. Some gene products promote or repress other genes, usually depending on the concentration of the gene product. In some cases two genes compete for expression, resulting in a population of cells distributed between the competing states. The expression levels of other genes are affected by environmental parameters such as temperature or UV light, sometimes *via* the degradation rate of the protein that regulates the gene expression, e.g., the *cI* and *cro* genes of the *Bacteriophage* λ [1].

The challenge is to model the complex genetic componentry that enables a cell to switch genes "on" and "off" as required [2, 3]. Small models have been developed to

describe gene promotion and repression [4], and these components can be combined into more complex models [5–7]. In some cases, artificial genetic machines based upon well-understood genetic components have been constructed and their behaviour has been analysed in a more controlled environment [8,9]. Some biologically stable states and bistable systems can be modelled using stochastic differential equations that describe the protein levels as functions of the gene states [5,10].

This chapter discusses the mathematical modelling of gene regulation in the *Bacteriophage* λ. Reflecting the stochastic nature of gene regulation, the chemical master equation is used as a tool to study Markovian models of networks of gene states between which probabilistic transitions occur. These states represent the binding or unbinding of protein complexes to DNA, resulting in a gene being expressed or not expressed in the cell. In Section 7.2 the stochastic master equation for an arbitrary network is given. We also describe a generic hybrid model which combines deterministic chemical rate equations with a stochastic model of gene switching.

In Section 7.3 our hybrid model is applied to a simplified version of the switching mechanism of the phage λ as it moves stochastically between its alternate lifestyles (lysis or lysogeny). A stochastic simulation in Section 7.4 shows the phage λ evolving through a set of states. At each time step the virtual phage λ chooses a new state and, depending upon that state, CI and Cro proteins are produced. These proteins compete with each other to reach one of the two alternative lifestyles of the phage λ. The application of UV light to the host bacterium is also simulated and its effect is shown on the outcome of the competition between CI and Cro.

7.2 Stochastic Master Equation Model of Gene Regulation

The stochastic formulation of gene regulation is based on an assumption that the underlying chemical processes are Markovian. An efficient tool for dealing with Markovian processes is the stochastic chemical master equation [11]. In this formalism, a regulatory network is typically represented in a state space, elements of which describe the states or abundances of a finite number of chemical species which may be made up of any combination of genes, RNA, proteins or substrates. The gene states, for example, may be defined as gene on/off, i.e., the gene is/is not being expressed in the cell. Alternatively, we may wish to specify that a protein or enzyme is attached/not attached to a promoter or operator site, and also specify the absolute protein levels measured in numbers of molecules or in concentrations.

Consider a system that can be in any one of a finite number of states $s = 1, \ldots, N$, and capable of making transitions $i = 1, \ldots, N_T$ between states. Since the system is Markovian, the probability of making a transition at any given time depends only on the state of the system at that time, and not on its history. We represent the system by a directed graph with N vertices and N_T edges. Associated with each transition is a *propensity* $\alpha_i > 0$. If i is the transition from state s to state r, the probability of making the transition i in the time interval $[t, t + dt)$, conditional on being in state s at time t, is $\alpha_i dt$. Given an initial probability distribution among the states of $\mathbf{p}(0) = (p_1(0), \ldots, p_N(0))^{\mathrm{T}}$, with $p_s > 0$ and $\sum_s p_s = 1$, the system evolves in time to a

distribution $\mathbf{p}(t) = (p_1(t), \dots, p_N(t))^{\mathrm{T}}$ via the stochastic master equation

$$\frac{d\mathbf{p}}{dt} = A\mathbf{p}(t), \tag{7.1}$$

where the matrix A is given in terms of propensities. If α_{sr} is the propensity associated with the edge running from vertex s to vertex r, then

$$A = \sum_{s,r} \alpha_{sr} \Pi_{sr}, \tag{7.2}$$

where the matrices Π_{sr} are defined by

$$\Pi_{sr} = \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & -1 & & & \\ & & \vdots & \ddots & & \\ & & +1 & & \ddots & \\ & & & & & 0 \end{pmatrix}, \tag{7.3}$$

with -1 in the sth row and column to account for the flow of probability away from the sth state, $+1$ in the rth row and the sth column to account for the flow of probability into the rth state, and all other entries zero. Because each column of Π_{sr} sums to zero, so does each column of A. One easily checks that this is equivalent to saying that total probability is conserved: $d\left(\sum_s p_s(t)\right)/dt = 0$. For a specified initial condition, and with constant propensities, the solution to Eq. (7.1) is given formally by

$$\mathbf{p}(t) = e^{tA}\mathbf{p}(0). \tag{7.4}$$

For any realistic system, the dimensionality of the state space soon becomes prohibitively large if the concentration of all chemical species (proteins and enzymes for instance) and the binding of protein complexes to operator sites are to be described. To see this, consider that every possible combination of concentrations and operator binding states is represented by a different basis vector of the space in which the probability vector $\mathbf{p}(t)$ evolves. To reduce the problem to manageable proportions, we propose here a hybrid stochastic/deterministic model in which the evolution of continuum concentrations $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))$ of chemical species is modelled using deterministic chemical rate equations, and the switching between operator binding states $s = 1, \dots, N$ is modelled using the stochastic master equation. In general, the chemical rate equations will be differential equations of the form

$$\frac{dx_i}{dt} = f_i(\mathbf{x}, s), \tag{7.5}$$

and the propensities $\alpha_{rs}(\mathbf{x})$ will depend on chemical concentrations, since the probability that a given protein complex will bind to a particular operator site will depend on the concentration of that protein complex.

In principle, the system evolves over *infinitesimal* time steps dt along a path $(\mathbf{x}(t), s(t))$, where

$$x_i(t + dt) = x_i(t) + f_i(\mathbf{x}(t), s(t)) \, dt, \tag{7.6}$$

and $s(t + dt)$ is drawn from the probability distribution resulting from the action of the matrix $I + A(s)dt$ on the distribution concentrated at the state $s(t)$.

In practice, our numerical simulations use small *finite* time steps Δt over which the protein concentrations and protein binding state are evolved. The choice of Δt is crucial to the success of the model. Gene regulatory systems in prokaryotes typically have two intrinsic time scales: a slow scale set by chemical rate constants over which protein and substrate concentrations change, and a fast time scale over which gene switching driven by the binding of proteins to operators occurs. By choosing Δt intermediate between these scales, the essential dynamics can be efficiently captured. In the simulations that follow, protein concentrations are evolved using a Runge–Kutta algorithm with the state s set to its value at the beginning of the time interval Δt. At the end of each time step, a new operator binding state $s(t + \Delta t)$ is chosen from a distribution obtained by evolving Eq. (7.4) over the finite Δt using propensities set to their values at the beginning of the time step and a very accurate approximation to the matrix exponential times the distribution $p_r(t) = \delta_{rs(t)}$ concentrated at the state $s(t)$. We note that this model is not suitable for modelling regulation in eukaryotes, in which transcription can be regulated over much slower time scales.

7.3 The Switching Mechanism of the Bacteriophage λ

An example of a genetically regulated biological switching system is the Bacteriophage λ. The phage λ invades a bacteria cell (*E. coli*), after which it can enter into one of two alternative lifestyles called *lysogeny* and *lysis* [1]. The lysogeny stage is a dormant stage in which the phage inserts its DNA into the host's DNA and passively reproduces with the host. When the host becomes stressed, the phage is more likely to go into lysis, in which case it reproduces more phages, kills the host and spreads to other bacteria cells. The decision between lysis and lysogeny can be thought of as a switching mechanism. The stochastic switch is based upon a competition between two genes in the phage's DNA: *cro* and *cI*.[6] If the production of Cro proteins dominates, the system enters the lytic stage. If the production of CI proteins dominates, the system remains in the lysogenic stage.

Fig. 7.1 shows the section of phage λ DNA containing the *cro* and *cI* genes. The intervening region contains three operators O_{R1}, O_{R2} and O_{R3} which control the promoters P_R and P_{RM}. The gene products readily dimerise to form complexes Cro_2 and CI_2 which can bind to any of the three intervening operator sites. CI_2 dimers bind preferentially to O_{R1} and O_{R2}, but can also bind to O_{R3}, with a lower probability. Cro_2 dimers, on the other hand, bind preferentially to O_{R3}, but can also bind to O_{R1} and

---

[6] Capitalised names such as CI and Cro are used to represent proteins, while italicized names, such as *cI* and *cro* are used to represent the genes which encode for them.
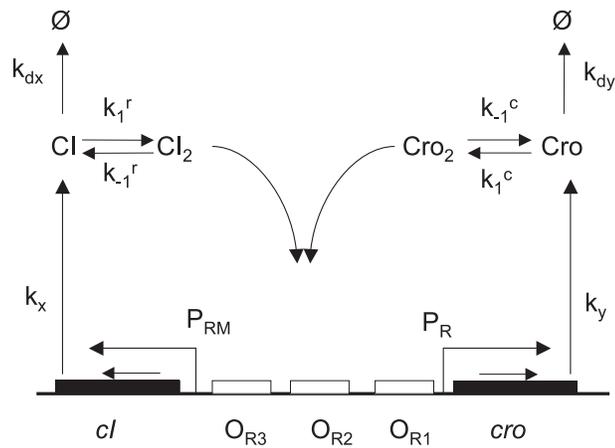
Fig. 7.1. The switching mechanism of the Bacteriophage λ in which *cI* and *cro* genes compete stochastically for expression.

O_{R2}. If either O_{R1} or O_{R2} is occupied, RNA polymerase is prevented from binding to the *cro* promoter P_R, and production of Cro protein ceases. Likewise, if O_{R3} is occupied, the promoter P_{RM} is blocked and production of CI protein ceases. Thus, at low concentrations both genes enhance their own production by inhibiting the production of the competing protein, and at high concentrations repress their own production. Also represented in Fig. 7.1 are decays of CI and Cro monomers.

The full story of the lysis/lysogeny decision mechanism is considerably more complex than the simplified version given here. See [1] for a full biological description, and [12] for a comprehensive stochastic model that includes the other proteins (CII, CIII and N) involved in the process, as well as the production of mRNA, its translation into proteins, the degradation rates of all chemical species and cell division.

The hybrid model we describe below differs from the previous model of Reinitz and Vaisnys [13] in our use of the stochastic master equation to introduce noise into the system via the switching between operator binding states. In contrast, Reinitz and Vaisnys assume operator binding states to be in chemical equilibrium, from which they estimate a time averaged protein synthesis rate. This leads to a deterministic set of coupled differential equations. There exist models which include noise into the differential equations by arbitrarily introducing additive or multiplicative stochastic terms to protein synthesis or decay rates [10,14]. However, we are unaware of any hybrid treatment in which substrate concentrations are determined by chemical rate equations, but the operator binding is determined using the stochastic master equation. The highly sophisticated model of Arkin et al. [5] is a simulation of the stochastic master equation using the Gillespie algorithm [15] to model all aspects of protein and RNA synthesis, including translation and transcription of genes, and decay in the phage λ system. Such models are a valuable and necessary contribution but are unfortunately computationally expensive and not easily adaptable to more complex regulatory networks. The model we propose here is designed to be computationally efficient and easily adapt-

able while still retaining the important characteristics of a moderately complex genetic system.

7.3.1 The Model

Following the conventions of Ref. [13], we indicate the operator binding states of the system by a discrete variable $s = \{1, \ldots , 40\}$. This variable represents the 40 ways in which either CI_2 or Cro_2 dimers can bind to O_{R1}, O_{R2} and O_{R3}, and RNA polymerase can bind to the promoters P_R and P_{RM} subject to the constraints that P_R cannot be occupied if either O_{R1} or O_{R2} is occupied and P_{RM} cannot be occupied if O_{R3} is occupied. RNA polymerase is bound to P_R only for any of the states $s \in \{28, 29, 30\}$, to P_{RM} only for $s \in \{31, \ldots , 39\}$, and to both for the state $s = 40$. The protein and dimer concentrations

$$x = [CI], \quad y = [Cro], \quad u = [CI_2] \quad \text{and} \quad v = [Cro_2] \tag{7.7}$$

evolve via the chemical rate equations

$$
\begin{aligned}
\frac{dx}{dt} &= k_x I_x(s(t)) - k_{dx}x - 2k_1^r x^2 + 2k_{-1}^r u, \\
\frac{dy}{dt} &= k_y I_y(s(t)) - k_{dy}y - 2k_1^c y^2 + 2k_{-1}^c v, \\
\frac{du}{dt} &= k_1^r x^2 - k_{-1}^r u, \\
\frac{dv}{dt} &= k_1^c y^2 - k_{-1}^c v,
\end{aligned}
\tag{7.8}
$$

where $I_x(s)$ and $I_y(s)$ are indicator functions equal to 1 if $s \in \mathcal{S}_x = \{31, 32, \ldots , 40\}$ or $s \in \mathcal{S}_y = \{28, 29, 30, 40\}$) respectively, and 0 otherwise.[7]

The state $s(t + \Delta t)$ is drawn randomly from a distribution

$$\mathbf{p}(t + \Delta t) = e^{A(u(t),v(t))\Delta t}\mathbf{p}(t), \tag{7.9}$$

where $A(u, v)$ is a dimer concentration dependent matrix of propensities taking the form of Eq. (7.2) and $p_r(t) = \delta_{rs(t)}$ is the probability distribution concentrated at the state $s(t)$.

7.3.2 Parameter Values: Chemical Rate Constants

In Table 7.1 are listed our values for the chemical rate constants occurring in Eqs. (7.8). Protein monomer production is a complex process involving transcription of DNA to mRNA which is then translated to proteins, which we model by effective rate constants k_x and k_y. Our model is similar to the phage λ switch model of Reinitz and

---

[7] This model can be further refined by replacing each indicator function by an improved estimate of the expected occupancy fraction $\sum_{s \in \mathcal{S}_x} \int_t^{t+\Delta t} p_s(\tau)d\tau / \Delta\tau$, and similarly for y.

Table 7.1. Chemical rate constants for the hybrid model of the phage λ switch.

Parameter	Value	Source
k_x	1.1 nM s^{-1}	Table 2 of Ref. [13]
k_y	0.5 nM s^{-1}	Adjusted
k_{dx}	7.0×10^{-4} s^{-1}	Table 3 of Ref. [5]
k_{dy}	2.5×10^{-3} s^{-1}	Table 3 of Ref. [5]
k_1^r	9.0×10^{-2} nM^{-1} s^{-1}	Table 2 of Ref. [16] and k_{-1}^r
k_{-1}^r	0.5 s^{-1}	Table 3 of Ref. [5]
k_1^c	1.5×10^{-3} nM^{-1} s^{-1}	Table 1 of Ref. [17] and k_{-2}^r
k_{-1}^c	0.5 s^{-1}	Table 3 of Ref. [5]

Vaisnys [13], and we use their value for the CI production rate k_x. However, we find that their value of 0.078 nM s^{-1} for the Cro monomer production does not produce sufficient quantities of Cro proteins to establish a lytic phase within our simulations. Instead, we have chosen a higher value of 0.5 nM s^{-1}, which is of the same order as the CI production rate. For the protein monomer decay rates we take values of the more sophisticated stochastic simulation model of Arkin et al. [5].

Dimerisation equilibrium constants are well established for CI from the work of Burz et al. [16], who give $k_1^r / k_{-1}^r = 1.8 \times 10^8$ M^{-1}, and for Cro from the work of Darling et al. [17], who give $k_1^c / k_{-1}^c = 3.07(\pm 1.08) \times 10^6$ M^{-1}. For the backward reaction rates we use the Arkin [5] value of $k_{-1}^r = k_{-1}^s = 0.5$ s^{-1} and infer forward rates from the equilibrium constants.

7.3.3 Parameter Values: Propensities

A series of papers by Ackers and co-workers beginning with Ref. [18] and culminating in Darling et al. [19] measure equilibrium binding free energies for the 40 states representing the possible binding configurations of CI and Cro dimers to operators and RNA polymerase to the promoters P_R and P_{RM}. As we explain below, these can give us ratios of forward to backward reaction rates for binding, though not a complete specification of the rates themselves.

Following Ref. [19], the equilibrium probablility of configuration s at temperature T is given in terms of the dimensionless binding free energy $\Delta G_s / (RT)$ by the Gibbs distribution

$$f_s = \frac{C_s \exp(-\Delta G_s / RT) u^{i(s)} v^{j(s)} [\text{RNAP}]^{k(s)}}{\sum_{r=1}^{40} C_r \exp(-\Delta G_r / RT) u^{i(r)} v^{j(r)} [\text{RNAP}]^{k(r)}}, \tag{7.10}$$

where u and v are the CI and Cro dimer concentrations, [RNAP] is the concentration of RNA polymerase and $i(s)$, $j(s)$ and $k(s)$ are the number of CI dimer, Cro dimer and

RNAP molecules bound to the DNA in state s. For each state we have included a pre-exponential factor C_s of dimension (concentration)$^{-[i(s)+j(s)+k(s)]}$ to ensure that each term in the sum in the denominator is of the same dimension. This factor is not present in the papers of Ackers and co-workers, nor in the work of Refs. [13] and [10] who use the same formula. Reading between the lines, it appears that all previous authors have implicitly used $C_s = (1\text{ M})^{-[i(s)+j(s)+k(s)]}$, and we shall use the same assumption here.

We are interested in estimating the propensities occurring in the discrete state master equation (7.9). It is sufficient to include only transitions in which a single dimer or polymerase attaches to or detaches from the DNA, as the probability of more than one such event occurring in time dt is of $O(dt^2)$. Thus, every possible transition is described by a process of the form

$$R + O \underset{k_{\text{b}}}{\overset{k_{\text{f}}}{\rightleftharpoons}} RO, \tag{7.11}$$

where R represents a dimer or polymerase and O the operator or promoter to which it binds. In the case of the phage λ switch with its 40 binding states there are 82 such possible processes. Consider two states r and s which differ only in that s is the state r with one extra dimer or polymerase attached. If we denote the propensity for the attachment transition $r \rightarrow s$ as α_{rs} and that for the detachment transition $s \rightarrow r$ as β_{sr}, then in terms of the chemical reaction rates defined above we have

$$\alpha_{rs} = k_{\text{f}}[R] = C_{\text{f}}e^{-E_{\text{f}}/(RT)}[R],$$
$$\beta_{sr} = k_{\text{b}} = C_{\text{b}}e^{-E_{\text{b}}/(RT)}, \tag{7.12}$$

where $[R] = u$, v or [RNAP]. Here we have assumed the usual Arrhenius form for chemical rate constants in terms of activation energies E_{f} and E_{b}. Note that both α_{rs} and β_{sr} have the correct dimensions of (time)$^{-1}$. Consistency with Eq. (7.10) in the equilibrium distribution gives

$$\frac{\alpha_{rs}}{\beta_{sr}} = Ce^{-\Delta G_{sr}/(RT)}[R], \tag{7.13}$$

where $\Delta G_{sr} = E_{\text{f}} - E_{\text{b}} = \Delta G_s - \Delta G_r$, and $C = C_{\text{f}}/C_{\text{b}} = C_s/C_r = 1\text{ M}^{-1}$.

Note that we do not know either α_{rs} or β_{sr} individually, so some estimate of one or the other must be made. Aurell et al. [7] argue that, under the assumption of diffusion-limited aggregation, the association time (i.e., the inverse of the forward propensity) of a single molecule can be estimated from the Smoluchowski equation: $\alpha_{rs}^{-1} = V/(4\pi DL)$. Assuming diffusion-limited aggregation into a region of length $L = 5$ nm, a bacterial volume of $V = 2 \times 10^{-15}$ l and a diffusion coefficient of $D = 5\ \mu\text{m}^2\text{s}^{-1}$ give an aggregation time of ~ 6 s at a concentration of 0.8 nM, or a forward reaction rate of $k_{\text{f}} = \alpha_{rs}/[R] \approx 0.5\text{ nM}^{-1}\text{s}^{-1}$. This figure is slightly higher than the diffusion-limited prediction of Berg et al. [20]. However, Berg et al. go on to say that *measured* values of protein binding rates can exceed $k_{\text{f}} = 10\text{ nM}^{-1}\text{s}^{-1}$, and explain these higher values as the effect of a two-step process in which the protein first

binds to a nearby non-specific site, and then either slides along the DNA or is directly transferred from one segment of DNA to another. Using these values as a guide, for all forward propensities we have used $\alpha_{rs} = 1 \text{ nM}^{-1}\text{s}^{-1} \times [R]$, where $[R]$ is the concentration of the aggregating molecule. Backward propensities β_{sr} are then inferred from Eq. (7.13) using binding energies ΔG from Table 4 of Ref. [19] and a temperature of $T = 37°C$, or $RT = 0.616$ kcal.

The RNA polymerase concentration is taken to be $[\text{RNAP}] = 30$ nM [5].

7.4 Numerical Simulation

Fig. 7.2 shows the results of our simulation of the phage λ switch. The top panel shows the binding of RNA polymerase to the promoters P_{RM} and P_R and binding of CI_2 and Cro_2 dimers to the three intervening operators. The bottom panel shows the concentrations of the proteins and dimers.

For the period $t = 0$ to $t = 1000$ seconds, parameter values are those listed in the previous section, and the algorithm described at the end of Section 7.2 evolves over time steps of 0.2 s. Over this period CI_2 is bound to O_{R1} and O_{R2} for most of the time and Cro production is repressed. There is binding of Cro_2 to O_{R3}, but not enough to prevent production of CI. In the lower panel we see steady concentrations of CI_2 and Cro_2 dimers, with values typical of the lysogenic state [5].
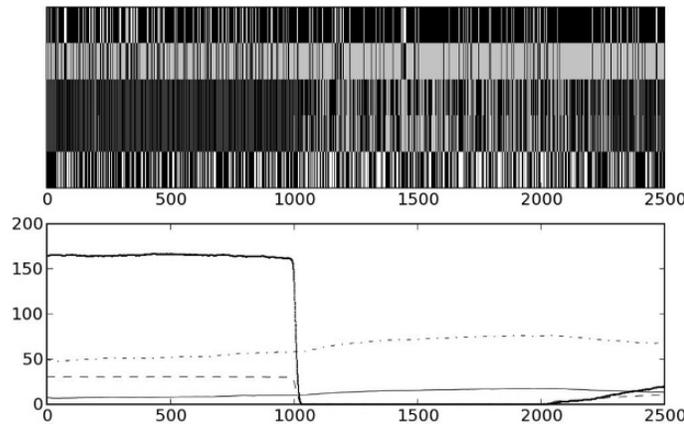
Fig. 7.2. Simulation of the evolution of the phage λ switch. Top panel: The five bands represent (from top to bottom) $P_{RM}, O_{R3}, O_{R2}, O_{R1}$ and P_R sites with time in seconds running along the horizontal axis. RNA polymerase (white) binds to promoters P_{RM} and P_R. CI_2 (dark grey) and Cro_2 (light grey) dimers bind to operators O_{R3}, O_{R2} and O_{R1}. Unoccupied sites are shown in black. Bottom panel: Concentration in nM of CI (dashed line) and Cro (dash-dot line) monomers and CI_2 (thick line) and Cro_2 (thin line) dimers. The horizontal axis is the time in seconds. The effect of UV light on CI monomers is simulated by increasing the degradation rate k_{dx} to 1 s^{-1} between $t = 1000$ s and $t = 2000$ s.

When phage λ infected bacteria are exposed to UV light, a protein called RecA is activated. This protein has the effect of cleaving the CI_2 dimers, rendering them unable to bind to operators [1]. We have simulated this by increasing the CI degradation rate k_{dx} from its previous value of 7.0×10^{-4} s^{-1} to 1 s^{-1} over the period from $t = 1000$ s to $t = 2000$ s. As the CI_2 concentration drops, O_{R1} and O_{R2} are freed up. There is some binding of Cro_2 to these operators, but not enough to prevent an increase in production of Cro, causing the system to move towards the lysis. As the UV light is switched off at $t = 2000$ s the system begins to return to lysogeny.

7.5 Conclusion

The ultimate aim of mathematical models of regulatory networks is to be predictive rather than descriptive. Analyses of stability, bistability and robustness may be possible once one has a sound model of the system, usually based upon stochastic processes and differential equations. The stochastic master equation can be used to model the evolving probability distribution of a system over the entire state space, or it can be used to simulate typical trajectories of single cells through state space. The eventual aim is to provide models that can lead back into experiment by predicting the proportions of sub-populations which have followed particular paths, by predicting the upper and lower limits of unknown pathway parameters and rates, by modelling the behaviour of systems under perturbation or by providing the quantitative reasoning behind existing biological systems.

In this chapter we have presented a step on the road to this eventual aim: a hybrid model of gene regulation in which gene switching is dealt with probabilistically using the stochastic master equation while protein and substrate concentrations are dealt with deterministically using classical chemical rate equations. The model deals with moderately complex biological systems economically by reducing the dimension of the phase space to manageable proportions without losing the probabilistic nature of the system.

We have applied the model to perhaps the best-known simple but realistic genetic regulatory system: the switching mechanism between lysogeny and lysis of the Bacteriophage λ. We have tried to stick as closely as possible to empirically measured physical input parameters with minimal adjustment. The model moves between lysogeny and lysis over realistic time scales as the host bacterium is stressed by the application of ultraviolet light.

This model is not yet at the stage of being predictive. Further refinements of our generic hybrid model are possible, and the next stage will be to implement them. Gene transcription and translation is currently modelled as a single-step process with a single rate parameter. We next intend to include mRNA as an intermediate step and introduce a time delay to account for the transcription step.

References

1. Ptashne, M.: A Genetic Switch: Phage λ and Higher Organisms (second edition). Cell Press, Cambridge, Massachusetts (1992).
2. de Jong, H.: Modeling and simulation of genetic regulatory systems: a literature review. J. Comp. Biol., **9**, 67–103 (2002).
3. Booth, H.S., Burden, C.J., Hegland, M., Santoso, L.: Markov process modelling of gene regulation. Austr. Math. Soc. Gazette, **32**, 31–41 (2005).
4. Hasty, J., McMillen, D., Isaacs, F., Collins, J.J.: Computational studies of gene regulatory networks: *in numero* molecular biology. Nature, **2**, 268–278 (2001).
5. Arkin, A., Ross, J., McAdams H.H.: Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-Infected *Escherichia coli* Cells. Genetics, **149**, 1633–1648 (1998).
6. Kepler, T.B., Elston T.C.: Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys. J., **81**, 3116–3136 (2001).
7. Aurell, E., Brown, S., Johanson, J., Sneppen, K.: Stability puzzles in phage λ. Phys. Rev. E., **65**, 051914 (2002).
8. Judd E.M., Laub, M.T., McAdams H.H.: Toggles and oscillators: new genetic circuit designs. BioEssays, **22**, 507–509 (2000).
9. Elowitz M.B., Lelbler S.: A synthetic ocsillatory network of transcriptional regulators. Nature, **403**, 335–338 (2000).
10. Tian, T., Burrage, K.: Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage λ. J. Theor. Biol., **227**, 229–237 (2004).
11. van Kampen, N.G.: Stochastic Processes in Physics and Chemistry. North-Holland, New York, USA (1981).
12. McAdams, H.H., Arkin, A.: Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. USA, **94**, 814–819 (1997).
13. Reinitz, J., Vaisnys, J.R.: Theoretical and experimental analysis of the phage lambda genetic switch missing levels of co-operativity. J. Theor. Biol., **145**, 295–318 (1990).
14. Hasty, J., Isaacs, F., Dolnik, M., McMillen, D., Collins, J.J.: Designer gene networks: Towards fundamental cellular control. Chaos, **11**, 207–220 (2001).
15. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys., **22** 403–434 (1976).
16. Burz, D.S., Beckett, D., Benson, N., Ackers, G.K.: Self assembly of bacteriophage λ CI repressor: Effects of single site mutations on monomer-dimer equilibrium. Biochemistry, **33**, 8399–8405 (1994).
17. Darling, P.J., Holt, J.M., Ackers, G.K.: Coupled energetics of λ *cro* repressor self-assembly and site-specific DNA operator binding I: Analysis of *cro* dimerisation from nanomolar to micromolar concentrations. Biochemistry, **33**, 11500–11507 (2000).
18. Ackers, G.K., Johnson, A.D., Shea, M.A.: Quantitative model for gene regulation by λ phage repressor. Proc. Nat. Acad. Sci., **79**, 1129–1133 (1982).
19. Darling, P.J., Holt, J.M., Ackers, G.K.: Coupled energetics of λ *cro* repressor self-assembly and site-specific DNA operator binding II: Cooperative interactions of *cro* dimers. J. Mol. Biol., **139**, 163–194 (2000).
20. Berg, O.G., Winter, R.B., von hippel, P.H.: How do genome-regulatory proteins locate their DNA target sites? Trends Biochem., **7**, 52–55 (1982).

8

Piecewise-Linear Models of Genetic Regulatory Networks: Analysis of the Carbon Starvation Response in *Escherichia coli*

Delphine Ropers,[1] Hidde de Jong,[1] Jean-Luc Gouzé,[2] Michel Page,[1,3] Dominique Schneider,[4] and Johannes Geiselmann[4]

[1] Institut National de Recherche en Informatique et en Automatique (INRIA), Unité de recherche Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France; {Delphine.Ropers,Hidde.de-Jong,Michel.Page}@inrialpes.fr
[2] Institut National de Recherche en Informatique et en Automatique (INRIA), Unité de recherche Sophia-Antipolis, Sophia-Antipolis, France; Jean-Luc.Gouze@sophia.inria.fr
[3] Ecole Supérieure des Affaires, Université Pierre Mendès France, Grenoble, France
[4] Laboratoire Adaptation et Pathogénie des Microorganismes (CNRS UMR 5163), Université Joseph Fourier, France; {Dominique.Schneider,Hans.Geiselmann}@ujf-grenoble.fr

Summary. The adaptation of the growth of *Escherichia coli* to the availability of a carbon source is controlled by a complex genetic regulatory network whose functioning is still little understood. Using a qualitative method based on piecewise-linear differential equations, which is able to overcome the current lack of quantitative data on kinetic parameters and molecular concentrations, we model the carbon starvation response network and simulate the response of *E. coli* cells to carbon deprivation. This allows us to identify essential features of the transition between exponential and stationary phase and to make new predictions on the qualitative system behavior, following a carbon upshift.

Key words: Qualitative modeling and simulation, piecewise-linear differential equations, genetic regulatory networks, carbon starvation response, *Escherichia coli*.

8.1 Introduction

In the case of nutritional stress, like the deprivation of an essential carbon source, an exponentially growing *Escherichia coli* population may enter a non-growth state, called stationary phase. During the growth-phase transition, individual bacteria undergo numerous physiological changes, concerning among other things their morphology and metabolism, as well as gene expression [13]. On the molecular level, the growth-phase transition is controlled by a large and complex genetic regulatory network integrating various environmental signals [31]. The backbone of the network is composed of a class of pleiotropic transcription factors, called global regulators, and their interactions. In a combinatorial fashion, they mediate the activation or repression of a large number of genes in response to changes in environmental conditions.

Understanding the molecular basis of the transition between the exponential and stationary phase of *E. coli* cells has been the focus of extensive studies for decades [12]. However, most studies have focused on only one or a few components of the network, and currently the functioning of the network as a whole remains little understood. These circumstances have motivated the analysis of the network of global regulators in order to understand how the switch between the growth phases of *E. coli* arises from the interaction between these transcription factors. As a first step, we have decided to focus on the particular part of the network responding to a well-studied stress, namely carbon starvation.

Computer modeling and simulation tools may help to analyze the behavior of this kind of complex system [21]. However, except for well-studied processes (i.e., the carbon uptake and metabolism [19]), numerical values for the parameters characterizing the interactions and the molecular concentrations are most often absent. This makes it difficult to apply traditional methods for the dynamical modeling of genetic regulatory networks. To overcome these constraints, we have used a qualitative modeling and simulation method based on piecewise-linear (PL) differential equations, supported by the publicly-available computer tool *Genetic Network Analyzer (GNA)* [3,16,23]. The method allows a coarse-grained, qualitative analysis of the network dynamics to be carried out. Instead of numerical values for the parameters, the method uses inequality constraints that can be inferred from the experimental literature. In addition, this makes it possible to analyze the stability of the attractors of the system [4].

We have constructed a model including key proteins and their interactions involved in the carbon starvation response. Using GNA and its recently added, new functionalities [23], we simulated the response of *E. coli* cells to carbon source availability. This has allowed us to identify essential features of the transition between exponential and stationary phases, and to make new predictions on the qualitative system behavior following a carbon upshift.

In the next section, we present the network controlling the carbon starvation response of *E. coli*. The qualitative modeling and simulation approach that we have used is summarized in Sect. 8.3, as well as its application to the carbon starvation response network in Sect. 8.4. Some results of qualitative simulation are presented in Sect. 8.5, followed by a discussion of their biological implications in the final section.

8.2 The Carbon Starvation Response Network

The aim of this chapter is to give a minimal representation of the adaptation of the growth of *E. coli* to carbon source availability. To this end, a first, simple network of the carbon starvation response has been built on the basis of literature data and information available in public databases [17,28]. It is composed of six genes that are believed to play a key role in the process: *crp*, *cya*, *fis*, *gyrAB*, *topA*, and *rrn* (Fig. 8.1). Notice that not all global regulators thought to be relevant have been included in the analysis in this paper. Notably, ppGpp and RpoS have been excluded for the moment (but see Sect. 8.6). The genes and their interactions in Fig. 8.1 make up four modules with specific functions, whose interconnections control the propagation of the carbon
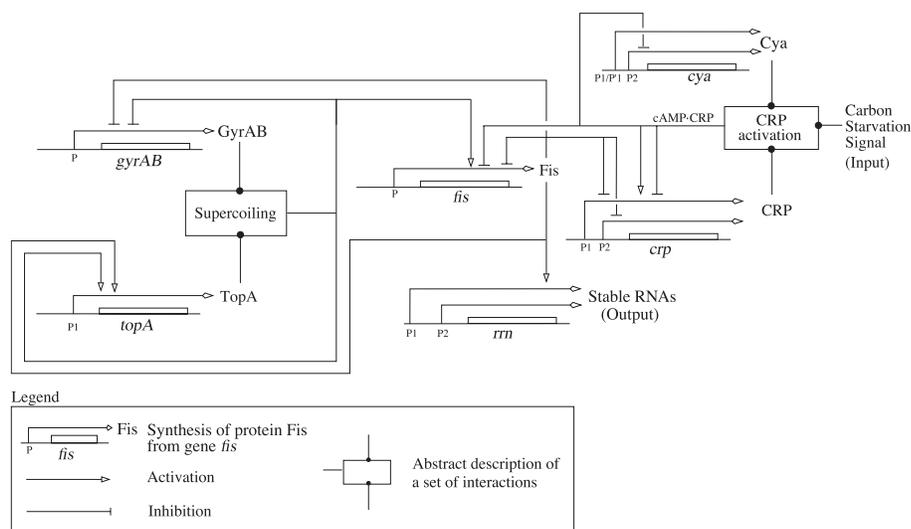
Fig. 8.1. Network of key genes, proteins, and regulatory interactions involved in the carbon starvation response network in *E. coli*. The contents of the boxes CRP activation and Supercoiling are detailed in [26]. The graphical conventions [18] are explained in the legend.

starvation signal through the network, leading to the adaptation of the growth of *E. coli*.

The first network module concerns the input of the system. A signal-transduction pathway transmits to the network the information that an essential carbon source is lacking, by activating the adenylate cyclase enzyme (Cya). This allows the enzyme to efficiently produce a metabolite, cAMP [27]. The second network module is composed of genes *fis*, *crp*, and *cya*. The global regulator CRP is the target of the signal-transduction pathway: it is activated by its interaction with the cAMP molecule. The cAMP·CRP complex is able to regulate the expression of a large set of genes directly involved in the response to a lack of carbon source (i.e., synthesis of catabolic enzymes, changes of cellular morphology), as well as the expression of genes coding for global regulators, in particular *fis* and *crp*. The protein Fis is an important regulator of genes involved in the cellular metabolism [8] and in addition, it controls the expression of *crp*, as well as its own expression. The third module, composed of the genes *gyrAB*,[5] *topA*, and *fis*, controls the DNA topology, another important modulator of gene expression [6,11]. The level of DNA supercoiling is mainly the result of a balance between the activity of GyrAB, which supercoils the DNA structure, and TopA, which relaxes the DNA. In *E. coli* cells, the DNA topology is tightly regulated, since an increase of negative supercoiling upregulates *topA* expression and downregulates *gyrAB* expression. Fis also participates in this homeostatic control, since it controls the expression of *gyrAB* and *topA*, while the expression of its own gene is regulated by

---

[5] GyrAB is considered here as the product of a single gene, because the *gyrA* and *gyrB* genes are regulated in a similar manner.

the DNA supercoiling level. Finally, the last module controls cellular growth and thus represents the output of the system. It is composed of genes encoding Fis and stable RNAs. In *E. coli*, these RNAs are produced from seven operons similarly organized, each being composed of three genes for ribosomal RNAs and one gene for transfer RNA. The level of stable RNAs is a reliable indicator of cellular growth. Indeed, exponentially growing cells need large amounts of stable RNAs, whereas low amounts are sufficient in stationary phase. The expression of the stable RNA genes is stimulated by Fis [24].

The carbon starvation response network in Fig. 8.1 is composed of numerous interactions between its components. In particular, the connections between different modules involve a positive feedback loop—the mutual inhibition of *fis* and *crp*—and a negative feedback loop—the homeostatic control of the DNA topology. This makes it difficult to understand the dynamics of the system in an intuitive manner. Mathematical modeling and computer simulation could be used to address questions on the dynamics of the carbon starvation response network. Thus far, only parts of the network have been the focus of modeling studies (for some examples, see [5, 20]). An analysis of the dynamics of the entire carbon starvation response network has never been carried out, and is also difficult to achieve as a consequence of the lack of quantitative data on most of the network components. In order to deal with the current lack of data, we have used a qualitative method, briefly outlined in the next section.

8.3 Qualitative Modeling and Simulation of Genetic Regulatory Networks

The qualitative modeling and simulation method is fully described in [16]. In this section, we illustrate the method through an example of a two-gene network, composed of the genes *crp* and *fis* (Fig. 8.2(a)), which is a (simplified) part of the carbon starvation response network in Fig. 8.1. The CRP protein inhibits *fis* expression above a certain threshold concentration, while the Fis protein inhibits the expression of *crp* above one threshold concentration, and inhibits its own expression above a second, higher threshold concentration [2,9,22,25]. The degradation of the proteins is not regulated.

(a)   (b)   (c)
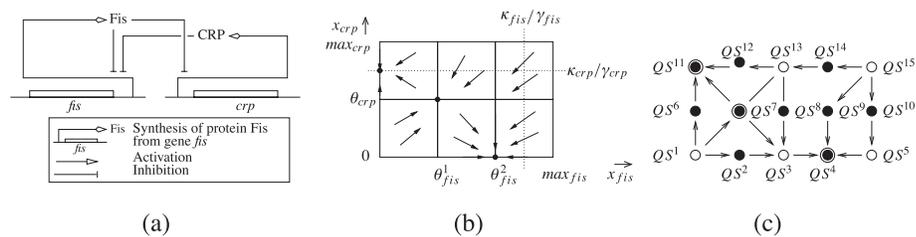
Fig. 8.2. (a) Simple genetic regulatory network composed of the genes *fis* and *crp*. (b) Sketch of the dynamics in each domain of the phase space for the two-gene network. Dots represent the equilibrium points of the system. (c) State transition graph for the two-gene network. QS denotes a qualitative state. The qualitative equilibrium states are circled.

The dynamics of such a genetic regulatory network can be modeled by a class of piecewise-linear (PL) differential equations originally proposed by Glass and Kauffman [7]. The example network gives rise to the following model:

$$\dot{x}_{fis} = \kappa_{fis} \, s^-(x_{crp}, \theta_{crp}) \, s^-(x_{fis}, \theta_{fis}^2) \, - \gamma_{fis} \, x_{fis} \, , \qquad (8.1)$$

$$\dot{x}_{crp} = \kappa_{crp} \, s^-(x_{fis}, \theta_{fis}^1) - \gamma_{crp} \, x_{crp} \, , \qquad (8.2)$$

where x_{fis} and x_{crp} denote the concentrations of the proteins Fis and CRP, $\dot{x}_{fis}$ and $\dot{x}_{crp}$ their time derivatives, θ_{fis}^1, θ_{fis}^2, and θ_{crp} threshold concentrations, κ_{fis} and κ_{crp} (positive) synthesis parameters, and γ_{fis} and γ_{crp} (positive) degradation parameters. The step function $s^-(x, \theta)$ evaluates to 1 if $x < \theta$, and to 0 if $x > \theta$. Step functions are approximations of the sigmoidal, Hill functions often characterizing gene regulation, preserving their non-linear, switch-like character.

Equations (8.1) and (8.2) describe the rate of change of protein concentrations. For instance, (8.1) states that protein Fis is expressed at a rate κ_{fis}, if the concentration of protein CRP is below the threshold θ_{crp} and the concentration of protein Fis below the threshold θ_{fis}^2. The protein is degraded at a rate proportional to its own expression.

Mathematical analysis of the model reveals that, for each of the two variables, the knowledge of the relative order of the threshold parameter(s), as well as the order of the quotient of the synthesis and degradation parameters with respect to the thresholds, is sufficient to sketch the solution trajectories in the phase space. This result has been shown generalizable to the whole class of PL models considered here [16]. In the example, we have specified the following constraints (Fig. 8.2(b)):

$$0 < \theta_{fis}^1 < \theta_{fis}^2 < max_{fis} \, , \quad \theta_{fis}^2 < \kappa_{fis}/\gamma_{fis} < max_{fis} \, , \qquad (8.3)$$

$$0 < \theta_{crp} < max_{crp} \, , \quad \theta_{crp} < \kappa_{crp}/\gamma_{crp} < max_{crp} \, . \qquad (8.4)$$

In practice, the use of inequality constraints implies that we do not need numerical values for the threshold and rate parameters in the differential equations, which are generally not available. In contrast, the inequality constraints can be inferred from the experimental literature. For instance, the constraints on the quotients $\kappa_{fis}/\gamma_{fis}$ and $\kappa_{crp}/\gamma_{crp}$ define the strength of gene expression on the scale of ordered threshold concentrations.

The parameter inequalities divide the phase space into domains where the system behaves in a qualitatively distinct way. Fig. 8.2(b) represents the phase space corresponding to the two-gene network and the system dynamics in the phase space.[6] The system possesses three equilibrium points. Using the partition of the phase space into domains and the qualitative characterization of the dynamics in each of the domains, we can discretize the continuous dynamics. In the resulting abstract description, a qualitative state represents a domain in the phase space and its associated dynamical properties. The transition between two qualitative states corresponds to solutions leaving one domain in the phase space and reaching another. The set of all qualitative states and

---

[6] In this simple representation of the method, we omit the problems raised by the discontinuities in the right-hand side of the differential equations, which are explained in [10,16].

their transitions define the state transition graph. The state transition graph corresponding to the two-gene network is represented in Fig. 8.2(c). It contains three qualitative equilibrium states, the stability of which can be determined by further analysis [4].

For many purposes, it is sufficient to know which qualitative states are reachable from a given initial qualitative state, that is, which qualitative behaviors the system can exhibit when it is initially in that state. Simple rules have been formulated for the symbolic computation of the reachable part of a state transition graph from a qualitative PL model and an initial domain [15]. They have been implemented in the publicly-available computer tool, *Genetic Network Analyzer (GNA)*. In order to analyze large and complex state transition graphs in detail, GNA allows the simulation results to be exported to model-checking tools [3]. Recently, additional rules have been implemented in the software, allowing the search of attractors in the state transition graph and the determination of their stability [23]. GNA has been used before to analyze other bacterial genetic regulatory networks, e.g., controlling the initiation of sporulation in *Bacillus subtilis* [14] and quorum sensing in *Pseudomonas aeruginosa* [30].

8.4 Modeling of the Carbon Starvation Response Network

The lack of quantitative data on the threshold and rate parameters has motivated the modeling of the carbon starvation response network in *E. coli* by means of the method summarized in the previous section. The graphical representation of the network presented in Fig. 8.1 has been translated into a PL differential equation model supplemented by qualitative constraints on the parameters. The resulting model consists of seven variables, one protein/RNA concentration variable for each of the six genes and one input variable representing the presence or absence of a carbon starvation signal [26]. In addition, 17 threshold parameters and 17 rate parameters have been defined, bounded by 48 threshold and parameter inequalities, the choice of which is largely determined by the available biological data.

As an illustration, we briefly present how we have obtained the PL differential equation and the parameter inequality constraints for the state variable x_{crp}. The state equation for x_{crp} in the two-gene network example in Sect. 8.3 describes only part of the regulatory logic for the gene *crp*. As can be seen in Fig. 8.1, the regulation of *crp* is much more complex. The *crp* gene is expressed from two promoters, P1 and P2, that are inhibited by protein Fis [27]. In addition, the cAMP·CRP complex is able to both repress and activate the promoter P1 [27], through a regulatory mechanism that is still unclear. In order to simplify, we omit the negative control of *crp* P1, because this mechanism only plays a role during the exponential growth phase, when the concentration of CRP is low (see [26] for a more detailed discussion).

We denote by κ_{crp}^1 the background synthesis rate from promoter P1 during exponential growth. In addition, κ_{crp}^2 denotes the synthesis rate induced by derepression of this promoter, and κ_{crp}^3 the synthesis rate from promoter P2. With a degradation rate equal to $\gamma_{crp}\, x_{crp}$, we obtain the following state equation for x_{crp}:

$$\dot{x}_{crp} = \kappa_{crp}^1 + \kappa_{crp}^2\, s^-(x_{fis}, \theta_{fis}^2)\, s^+(x_{crp}, \theta_{crp}^1)\, s^+(x_{cya}, \theta_{cya}^1)\, s^+(u_s, \theta_s)$$

$$+ \kappa_{crp}^3 \, s^-(x_{fis}, \theta_{fis}^1) - \gamma_{crp} \, x_{crp} \, . \qquad (8.5)$$

As shown in [26], the product of step functions $(s^+(x_{crp}, \theta_{crp}^1) \, s^+(x_{cya}, \theta_{cya}^1) \, s^+(u_s, \theta_s))$ describes stimulation of *crp* expression by the cAMP·CRP complex, where u_s is an input variable representing the carbon starvation signal. It is constant over the time interval of interest ($\dot{u}_s = 0$). The cAMP·CRP-mediated activation of *crp* occurs in the presence of a carbon starvation signal ($u_s > \theta_s$, that is, $s^+(u_s, \theta_s) = 1$) and sufficiently high concentrations of proteins CRP and Cya ($s^+(x_{crp}, \theta_{crp}^1) = 1$ and $s^+(x_{cya}, \theta_{cya}^1) = 1$). The thresholds in the step-function expressions for Fis-mediated repression ($s^-(x_{fis}, \theta_{fis}^1)$ and $s^-(x_{fis}, \theta_{fis}^2)$) reflect the fact that Fis has different affinities for the two *crp* promoter regions [9].

Different concentration thresholds of CRP are required in cells to allow production of different levels of the cAMP·CRP complex. Indeed, a low concentration of the complex is sufficient to stimulate *crp* expression in the presence of the carbon starvation signal. Above a second, higher threshold concentration, the complex controls the expression of a large set of genes involved in the adaptation of *E. coli* cells to the entry into stationary phase. Finally, above its highest threshold concentration, the complex inhibits its further accumulation, through repression of *cya* [27]. This gives the following threshold: inequalities

$$0 < \theta_{crp}^1 < \theta_{crp}^2 < \theta_{crp}^3 < max_{crp} \, . \qquad (8.6)$$

The quotient of the synthesis and degradation rates derived from (8.5) can be ordered with respect to the threshold parameters. Because *crp* P1 is a house-keeping promoter [9], we set $\kappa_{crp}^1/\gamma_{crp} > \theta_{crp}^1$ and $(\kappa_{crp}^1 + \kappa_{crp}^2)/\gamma_{crp} > \theta_{crp}^1$. If this were not the case, CRP would not be able to reach a basal concentration at which it can stimulate its own expression when the carbon starvation signal is switched on. The fact that *crp* P2 is a strong promoter [9] implies that $(\kappa_{crp}^1 + \kappa_{crp}^3)/\gamma_{crp} > \theta_{crp}^3$ and $(\kappa_{crp}^1 + \kappa_{crp}^2 + \kappa_{crp}^3)/\gamma_{crp} > \theta_{crp}^3$, because this activation of P2 is required for CRP to reach a level at which regulation of *cya* occurs. We thus arrive at the following equilibrium inequalities

$$\theta_{crp}^1 < \frac{\kappa_{crp}^1}{\gamma_{crp}} < \theta_{crp}^2 \, , \quad \theta_{crp}^1 < \frac{\kappa_{crp}^1 + \kappa_{crp}^2}{\gamma_{crp}} < \theta_{crp}^2 \, , \qquad (8.7)$$

$$\theta_{crp}^3 < \frac{\kappa_{crp}^1 + \kappa_{crp}^3}{\gamma_{crp}} < max_{crp} \, , \quad \theta_{crp}^3 < \frac{\kappa_{crp}^1 + \kappa_{crp}^2 + \kappa_{crp}^3}{\gamma_{crp}} < max_{crp} \, .$$

8.5 Simulation of Carbon Starvation Response

Using the PL model of the carbon starvation response network, we analyze the attractors of the system and simulate the response of *E. coli* to the depletion or sudden availability of an essential carbon source in the growth medium. The simulations lead

to qualitative behavioral predictions that can be compared with experimental observations reported in the literature. All simulations and attractor analysis described below have been carried out using the computer tool GNA.

The attractor analysis has allowed the identification of two stable, qualitative equilibrium states in less than 1 s on a PC (2.4 GHz, 512 Mb). The first state, described in (8.8), is characterized by the presence of a carbon starvation signal and a low level of stable RNAs:

$$
\begin{aligned}
\theta_{crp}^3 < x_{crp} \le max_{crp} \ , \quad x_{cya} = \theta_{cya}^3 \ , \\
0 \le x_{fis} < \theta_{fis}^1 \ , \quad x_{gyrAB} = \theta_{gyrAB}^2 \ , \\
0 \le x_{topA} < \theta_{topA}^1 \ , \quad 0 \le x_{rrn} < \theta_{rrn} \ , \\
\theta_s < u_s < max_s \ .
\end{aligned}
\tag{8.8}
$$

The level of stable RNAs is a reliable indicator of cellular growth (Sect. 8.2). Since the level of RNAs is low in the qualitative equilibrium state reached by the system in the presence of a carbon starvation signal, we conclude that this state is representative for a stationary-phase cell.

The second qualitative equilibrium state, described in (8.9), is characterized by the absence of a carbon starvation signal and a high level of stable RNAs, which is representative for an exponential-phase cell:

$$
\begin{aligned}
\theta_{crp}^1 < x_{crp} < \theta_{crp}^2 \ , \quad \theta_{cya}^3 < x_{cya} \le max_{cya} \ , \\
x_{fis} = \theta_{fis}^4 \ , \quad x_{gyrAB} = \theta_{gyrAB}^1 \ , \\
0 \le x_{topA} < \theta_{topA}^1 \ , \quad \theta_{rrn} < x_{rrn} \le max_{rrn} \ , \\
0 \le u_s < \theta_s \ .
\end{aligned}
\tag{8.9}
$$

We conclude from this analysis that, depending on the presence or absence of an essential carbon source in the growth medium, E. coli cells enter exponential phase or stationary phase, respectively. In order to understand how E. coli cells perform their growth transition, we next simulate their response to the presence or absence of a carbon source.

Starting from the initial conditions representing exponential growth in (8.9), the system is perturbated with a carbon starvation signal. This gives rise to a state transition graph of 65 states, including a single qualitative equilibrium state corresponding to (8.8), computed in less than 1 s. Fig. 8.3 shows the temporal evolution of selected protein concentrations during the transition from exponential to stationary phase in a representative path in the state transition graph. The first event after receiving the carbon starvation signal is the decrease of the Fis concentration, followed by the decrease of the stable RNA concentration. The next event concerns the increase of the level of CRP. In parallel, the concentration of GyrAB increases, whereas the concentration of TopA remains constant.

The process driving the cell's growth arrest can be explained by relating the qualitative behavior to the carbon starvation response network in Fig. 8.1. During exponential
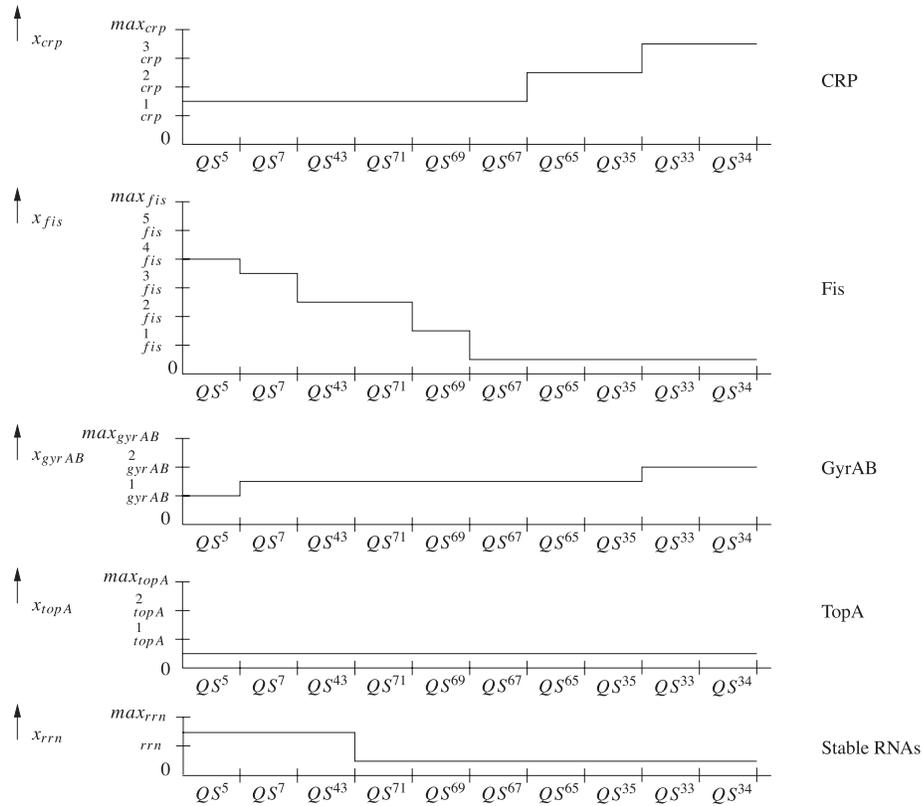
Fig. 8.3. Temporal evolution of selected protein concentrations in a typical qualitative behavior in the state transition graph generated from the *E. coli* carbon starvation response network. The behavior represents the molecular events accompanying the transition from exponential to stationary phase.

phase, the adenylate cyclase is present, but inactive. When the carbon starvation signal is switched on, it activates the protein and thus enables it to produce the cAMP metabolite. The small molecule binds to CRP, which is not yet abundant, thus giving rise to a low concentration of the cAMP·CRP complex. The level of cAMP·CRP is nevertheless high enough to start repressing the expression of *fis*. This stimulates further accumulation of CRP, and thus further repression of Fis, through the derepression of the Fis-controlled promoters of *crp*. The decrease of the Fis concentration causes the downregulation of the expression of the *rrn* genes. As a consequence, the level of the stable RNAs decreases and the cell enters stationary phase. We conclude from our model that a positive feedback loop, the mutual inhibition of Fis and CRP, plays a determining role in the transition from exponential to stationary phase.

The level of DNA supercoiling is determined by the relative concentration of the enzymes GyrAB and TopA (Sect. 8.2). The predicted qualitative evolution of the con-

centrations of the two proteins in Fig. 8.3 implies that the DNA supercoiling level increases at the onset of stationary phase. However, this is not what has been observed experimentally. On the contrary, the DNA supercoiling level has been shown to decrease when *E. coli* cells enter stationary phase [1]. This inconsistency suggests that our picture of the carbon starvation response network is incomplete, in the sense that the network in Fig. 8.1 may need to be extended with interactions not yet identified or with regulators not yet considered.

In addition to simulating the transition from exponential to stationary phase, we have also studied the reentry into exponential phase after a carbon upshift, that is, when stationary-phase cells have been put into a fresh medium. We used the same model as above, but started the simulation from the qualitative state characterizing stationary-phase conditions in (8.8) (except for the adjustment of the GyrAB concentration so as to satisfy the observed low level of DNA supercoiling), and with the carbon starvation signal switched off [26]. Unexpectedly, the cells were predicted to resume growth through damped oscillations in some of the protein concentrations. The negative feedback involving the homeostatic control of the DNA supercoiling level in the carbon starvation response network has been shown to play a key role in the process [26]. However, no experimental data are currently available to confirm or disconfirm this surprising prediction, which calls for experimental verification.

8.6 Discussion

We have presented a modeling and simulation of the genetic regulatory network controlling the carbon starvation response in *E. coli*. We have first defined the carbon starvation response network by identifying key global regulators involved in the process (Sect. 8.2). Since quantitative information on the parameters is lacking in most cases, we have chosen a qualitative modeling and simulation method to analyze the network (Sect. 8.3). The method is based on a class of PL differential equations that use step functions in order to describe the regulatory mechanisms. Instead of giving numerical values to the parameters and initial conditions, the PL model has been supplemented with inequality constraints. By integrating the available experimental data on the regulatory mechanisms underlying the interactions, we have developed a model of six PL differential equations describing the carbon starvation response network (Sect. 8.4). Instead of assigning numerical values to the kinetic parameters, about fifty constraints in the form of algebraic inequalities have been obtained from the experimental literature.

Using this model, we have analyzed the attractors of the system and simulated the carbon starvation response of *E. coli* cells using the computer tool GNA (Sect. 8.5). The model predicted that a positive feedback mechanism, the mutual inhibition of *fis* and *crp*, plays a key role in the transition between exponential phase and stationary phase. This kind of control mechanism is known to play a key role in developmental processes [29]. In the presence of a carbon starvation signal, it causes a switch from a state with a high Fis concentration and a low CRP concentration to a state with a low Fis concentration and a high CRP concentration. The positive feedback circuit thus enables the cell to leave the exponential phase in the absence of a carbon source and

enter stationary phase, while it also enables the cell to reenter exponential phase in case of carbon recovery.

The model has also yielded the prediction that a negative feedback loop involving the homeostatic control of the DNA supercoiling level is responsible for the occurrence of damped oscillations of certain protein concentrations after a carbon upshift. These damped oscillations may speed up cellular response times, thus allowing a rapid adaptation of the cell to the sudden availability of a carbon source [26]. These predictions, as well as other model predictions on the expression of key network genes, have never been observed experimentally. This has motivated us to start experiments aiming to validate the model predictions. At present, we are working to obtain data on the temporal evolution of the expression of key genes in the network by means of reporter gene assays. Comparison of the observed and predicted expression profiles may lead to a refinement of the model, which will initiate further experimental validation.

Certain predictions of the model are inconsistent with the biological observations, as noted for the DNA supercoiling level in Sect. 8.5. The inconsistencies point to limits of the present model of the carbon starvation response network and, more importantly, suggest possible extensions. Indeed, the carbon starvation response network is part of a larger genetic regulatory network comprising other global regulators such as RpoS, ppGpp, IHF, FNR, Lrp, HNS, and ArcA, in addition to DNA supercoiling, cAMP·CRP, and Fis [12]. This network senses and responds to a variety of stresses acting, often simultaneously, on the bacteria. Our aim is to model the whole network of global regulators to understand how it coordinates the different stress responses and allows *E. coli* to adapt its lifestyle to a wide range of environmental conditions.

Acknowledgments

References

1. Balke, V., Gralla, J.: Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. J. Bacteriol., **169**, 4499–506 (1987).
2. Ball, C., Osuna, R., Ferguson, K., Johnson, R.: Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. J. Bacteriol., **174**, 8043–56 (1992).
3. Batt, G., Ropers, D., de Jong, H., Geiselmann, J, Mateescu, R., page, M., Schneider, D.: Validation of qualitative models of genetic regulatory networks by model checking: analysis of the nutritional stress response in *Escherichia coli*. Bioinformatics, **21**, i19–28 (2005).
4. Casey, R., de Jong, H., Gouzé, J.-L.: Piecewise-linear models of genetic regulatory networks: equilibria and their stability. J. Math. Biol., **52**, 27–56 (2006).
5. Dennis, P., Ehrenberg, M., Bremer, H.: Control of rRNA synthesis in *Escherichia coli*: a systems biology approach. Microbiol Mol. Biol. Rev., **68**, 639–68 (2004).

6. Drlica, K.: Bacterial topoisomerases and the control of DNA supercoiling. Trends Genet., **6**, 433–7 (1990).

7. Glass, L., Kauffman, S.: The logical analysis of continuous, non-linear biochemical control networks. J. Theor. Biol., **39**, 103–29 (1973).

8. Gonzalez-Gil, G., Bringmann, P., Kahmann, R.: Fis is a regulator of metabolism in *Escherichia coli*. Mol. Microbiol., **22**, 21–9 (1996).

9. Gonzalez-Gil, G., Kahmann, R., Muskhelishvili, G.: Regulation of *crp* transcription by oscillation between distinct nucleoprotein complexes. EMBO J., **17**, 2877–85 (1998).

10. Gouzé, J.-L., Sari, T.: A class of piecewise linear differential equations arising in biological models. Dynam. Syst., **17**, 299–316 (2003).

11. Hatfield, G., Benham, C.: DNA topology-mediated control of global gene expression in *Escherichia coli*. Annu. Rev. Genet., **36**, 175–203 (2002).

12. Hengge-Aronis, R.: The general stress response in *Escherichia coli*. In: Storz, G., Hengge-Aronis, R. (eds.) Bacterial Stress Responses, 161–77. ASM Press (2000).

13. Huisman, G., Siegele D.A., Zambrano, M.M., Kolter, R.: Morphological and physiological changes during stationary phase. In: Neidhardt et al., F. (ed.) *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 1672–82. ASM Press (1996).

14. de Jong, H., Geiselmann, J., Batt, G., Hernandez, C., Page, M.: Qualitative simulation of the initiation of sporulation in *Bacillus subtilis*. Bull. Math. Biol., **66**, 261–99 (2004).

15. de Jong, H., Geiselmann, J., Hernandez, C., Page, M.: Genetic network analyzer: qualitative simulation of genetic regulatory networks. Bioinformatics, **19**, 336–44 (2003).

16. de Jong, H., Gouzé, J.-L., Hernandez, C., Page, M., Sari, T., Geiselmann, J.: Qualitative simulation of genetic regulatory networks using piecewise-linear models. Bull. Math. Biol., **66**, 301–40 (2004).

17. Keseler, I.M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I.T., Peralta-Gil, M., Karp, P.D.: EcoCyc: a comprehensive database resource for *Escherichia coli*. Nucleic Acids Res., **33**, D334–7 (2005).

18. Kohn, K.: Molecular interaction maps as information organizers and simulation guides. Chaos, **11**, 84–97 (2001).

19. Kremling, A., Bettenbrock, K., Laube, B., Jahreis, K., Lengeler, J.W., Gilles, E.D.: The organization of metabolic reaction networks: III. Application for diauxic growth on glucose and lactose. Metab. Eng., **3**, 362–79 (2001).

20. Kremling, A., Gilles, E.: The organization of metabolic reaction networks: II. Signal processing in hierarchical structured functional units. Metab. Eng., **3**, 138–50 (2001).

21. McAdams, H., Arkin, A.: Simulation of prokaryotic genetic circuits. Annu. Rev. Biophys. Biomol. Struct., **27**, 199–224 (1998).

22. Ninnemann, O., Koch, C., Kahmann, R.: The *E. coli fis* promoter is subject to stringent control and autoregulation. EMBO J., **11**, 1075–83 (1992).

23. Page, M., de Jong, H.: Search of steady states of piecewise-linear differential models of generic regulatory networks. Submitted.

24. Paul, B., Ross, W., Gaal, T., Gourse, R.: rRNA transcription in *Escherichia coli*. Annu. Rev. Genet., **38**, 749–770 (2004).

25. Pratt, T.S, Steiner, T., Feldman, L.S., Walker, K.A., Osuna, R.: Deletion analysis of the *fis* promoter region in *Escherichia coli*: antagonistic effects of integration host factor and Fis. J. Bacteriol., **179**, 6367–77 (1997).

26. Ropers, D., de Jong, H., Page, M., Schneider, D., Geiselmann, J.: Qualitative simulation of the carbon starvation response in *Escherichia coli*. *BioSystems*, in press (2006).

27. Saier, Jr, M.J., Ramseier, T.M., Reizer, J.: Regulation of carbon utilization. In: Neidhardt, F.C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff,

W.S., Riley, M., Schaechter, M., Umbarger, H.E. (eds.) *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 1325–43. ASM Press, Washington D.C. (1996).

28. Salgado, H., Gama-Castro, S., Martinez-Antonio, A., Diaz-Peredo, E., Sánchez-Solano, F., Peralta-Gil, M., Garcia-Alonso, D., Jiménz-Jacinto, V., et al.: RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli K-12*. Nucleic Acids Res., **1**, D303–6 (2004).
29. Thomas, R., d'Ari, R.: *Biological Feedback*. CRC Press (1990).
30. Usseglio Viretta, A., Fussenegger, M.: Modeling the quorum sensing regulatory network of human-pathogenic *Pseudomonas aeruginosa*. Biotechnol. Prog., **20**, 670–8 (2004).
31. Wick, L., Egli, T.: Molecular components of physiological stress responses in *Escherichia coli*. Adv. Biochem. Eng. Biotechnol., **89**, 1–45 (2004).

Predicting Gene Expression from Combined Expression and Promoter Profile Similarity with Application to Missing Value Imputation

Laura L. Elo[1,2], Johannes Tuikkala[3], Olli S. Nevalainen[3], and Tero Aittokallio[1,2]

[1] Department of Mathematics, FI-20014 University of Turku, Finland; laliel@utu.fi
[2] Turku Centre for Biotechnology, P.O. Box 123, FI-20521, Turku, Finland
[3] Department of Information Technology, FI-20014 University of Turku, Finland

Summary. Gene expression microarrays have become a popular high-throughput technique in functional genomics. By enabling the monitoring of thousands of genes simultaneously, this technique holds enormous potential to extend our understanding of various biological processes. However, the large amount of data poses a challenge when interpreting the results. Moreover, microarray data often contain frequent missing values, which may drastically affect the performance of different data analysis methods. Therefore, it is essential to effectively exploit additional biological information when analyzing and interpreting the data. In the present study, we investigate the relationship between gene expression profile and promoter sequence profile in the context of missing value imputation. In particular, we demonstrate that the selection of predictive genes for expression value estimation can be considerably improved by the incorporation of transcription factor binding information.

Key words: Microarrays, gene expression, promoter sequence, imputation.

9.1 Introduction

DNA microarrays provide a rapid means for monitoring the expression of several thousands of genes in a given biological sample. Microarray experiments have been successfully applied in many fields of biomedical research, especially in cancer diagnosis, prognosis and treatment. Expression patterns observed across different phenotypes, time points, or in response to environmental changes contribute to our understanding of the cellular processes involved in the regulation of gene expression. While providing a promising opportunity, however, the large amount of data presents a challenge when interpreting the microarray results. Therefore, instead of using the expression measurements only, it is also essential to exploit other biological information, such as DNA or protein sequences, published literature and functional taxonomies [2].

Studies on combining DNA sequence data with gene expression data are motivated by the hypothesis that genes with similar expression profiles are likely to be regulated via the same mechanisms and therefore might share common regulatory elements in

their upstream promoter sequences. Several authors have provided indirect evidence of such co-regulation in yeast by showing that genes that cluster together based on their expression profiles share common motifs in their promoter regions [5, 12, 13]. Recently, Allocco et al. [1] used data from a genome-wide in vivo binding analysis of known yeast transcription factors (TFs) and demonstrated explicitly that genes with strongly correlated expression profiles are more likely to have their promoter regions bound by a common TF.

Additional biological information is traditionally employed subsequent to gene expression analysis only, e.g., by finding overrepresented sequence motifs or enriched functional classes in gene clusters defined from the expression data [13]. Even the more advanced data integration methods typically follow sequential or iterative procedures [3, 10]. These types of analyses, however, have a limited capacity to overcome the technical and biological noise originating from the individual data sources. There are only a few methods that aim at using heterogeneous biological data in a fully integrated manner [7, 8]. Hanisch et al. [8] cluster genes based on combined information from expression profiles and biological networks, whereas Glenisson et al. [7] integrate expression data with literature-extracted information to reveal biologically meaningful gene clusters. However, neither of these studies explores the true contributions of the individual data sources; instead their weights are selected rather arbitrarily before the clustering.

In the present study, we investigate the link between co-expression and co-regulation in the context of missing value imputation. We first introduce a combined distance function which integrates TF binding information with expression profile distances. We then use this new similarity measure when selecting the best genes for missing value estimation. Instead of fixing the relative contributions of the data types beforehand, we systematically investigate their effects on the imputation accuracy with a procedure that estimates their relative importance directly from the data. To evaluate the benefits gained from using the combined distance, we compare the imputation accuracy of the TF-guided method to that obtained using expression data alone in different situations.

9.2 Materials and Methods

9.2.1 Distance Measures

Gene expression data from a series of m microarray experiments (arrays) for n genes can be represented as an expression matrix $\mathbf{X} = \{x_{ij}\}_{i=1,j=1}^{n,m}$, where the entry x_{ij} is the log-transformed relative expression level of gene i in experiment j. Each row in the matrix corresponds to *the expression profile* of a particular gene. We characterize *the expression distance* between genes g and i with the Euclidean distance

$$d_E(g,i) = \sqrt{\sum_{j=1}^{m}(x_{gj} - x_{ij})^2}. \tag{9.1}$$

The TF binding activity of a set of p distinct TFs with respect to the promoter regions of the n genes is defined as a binary TF binding matrix $\mathbf{Y} = \{y_{il}\}_{i=1,l=1}^{n,p}$, where the entry y_{il} indicates the binding (1) or non-binding (0) relationship of TF l to gene i. Each row in the matrix represents *the promoter profile* of a particular gene. We determine *the promoter distance* between genes g and i using the Jaccard distance

$$d_P(g, i) = 1 - \frac{\sum_{l=1}^{p} \min(y_{gl}, y_{il})}{\sum_{l=1}^{p} \max(y_{gl}, y_{il})}, \tag{9.2}$$

where the numerator $\sum \min(y_{gl}, y_{il})$ is the number of common TFs (intersection) and the denominator $\sum \max(y_{gl}, y_{il})$ is the total number of different TFs (union) in the two promoters. If the two promoter profiles contain only zeros, it is reasonable to define $d_P(g, i) = 0$. In our application, however, we restricted to the genes that contained at least one TF.

To obtain the overall similarity between the expression and promoter profiles of genes g and i, we define *the combined distance* of the expression and promoter distances as

$$d_C(g, i) = d_E(g, i) + a d_P(g, i), \tag{9.3}$$

where a is a non-negative weight factor that balances the relative contributions of the two components. The larger the value of a, the larger the contribution of the promoter distance.

9.2.2 Imputation Algorithms

We applied the weighted K nearest neighbor imputation (KNN) [14], which is perhaps the most popular approach for missing value estimation in microarray data. Let $Q = \{j_1, \dots, j_q\}$ denote the set of q experiments, in which the gene g has a missing value. The KNN imputation starts by determining for each experiment $j \in Q$ a separate neighborhood set $N_j^{(K)}(g)$ of K genes that are closest to the target gene g and that do not have a missing value in the experiment j. Missing values in other experiments in the neighbors are allowed. In conventional KNN imputation, the neighboring genes are selected according to the smallest expression distances d_E, whereas in TF-guided KNN imputation (referred to as TFKNN), the K nearest genes are identified using the combined distance d_C instead. In both cases, the expression distance d_E is calculated over the non-missing entries of $\mathbf{X}$ only. The missing value x_{gj} is then estimated by the weighted average

$$\hat{x}_{gj} = \frac{\sum_{i \in N_j^{(K)}(g)} x_{ij}/d(g, i)}{\sum_{i \in N_j^{(K)}(g)} 1/d(g, i)}, \tag{9.4}$$

where d is the given distance measure (d_E in KNN and d_C in TFKNN).

9.2.3 Testing Procedure

The microarray expression matrices used for the evaluation of the imputation algorithms were first pre-processed by removing all the rows (genes) containing at least one missing value. New incomplete expression matrices were then generated from these complete matrices by randomly setting 1, 5, 10, 15 and 20 percent of their values as missing. The imputation accuracy was assessed by calculating the normalized root mean squared error (NRMSE) between the original and imputed values of the missing entries defined as

$$\sqrt{\frac{\sum_{i,j}(x_{ij} - \hat{x}_{ij})^2}{\sum_{i,j} x_{ij}^2}}, \tag{9.5}$$

where the summations are calculated over the missing entries (i, j). The imputation was repeated 50 times for each test data set with different missing value rates.

9.2.4 Selection of the Weight Factor *a*

To estimate the weight factor a automatically from the data, we followed the idea of Kim et al. [9] suggested for the selection of the model parameter K. In an incomplete data matrix to be imputed, we first selected a non-missing entry and marked it as missing. We then estimated this artificial missing value using different values of a. Finally, we compared these estimates to the known true value. This procedure was repeated for 100 random non-missing entries from each experiment. In these estimations, no information about the actual missing values to be imputed was utilized. The a-value that produced the smallest overall NRMSE for the artificial missing values was selected for the actual imputation.

The motivation of this procedure was that the a-value that shows good performance in estimating an additional missing value is assumed to be close to the optimal a-value. To assess the performance of the automatic a-value estimator, we also repeated the actual imputation of the missing values used for testing the imputation accuracy with different values of a and recorded these minima as lower bounds for the automatic procedure. It should be noted that in practice one cannot optimize the weight factor based on the actual missing values, whereas the automatic procedure proposed here is also practicable in real situations, where the true values of missing data are not known. However, as the estimation of the training and testing missing values may use common genes, the overall result might be somewhat biased.

9.2.5 Test Material

We used three publicly available yeast cDNA microarray data sets. The first data set (diauxic) was from a temporal gene expression study during the diauxic shift from anaerobic to aerobic metabolism in yeast ($m = 7$) [5]. The second data set (alpha) was from a time series study monitoring dynamic changes in gene expression during the

alpha-factor synchronized yeast cell cycle ($m = 18$) [12]. The third data set (carbon) was a non-time series subset of experiments from a study monitoring the response of yeast cells to alternative carbon sources ($m = 6$) [6].

The TF binding matrix was constructed using the data from the genome-wide chromatin immunoprecipitation (ChIP) experiment of Lee et al. [11], which identified the binding of $p = 106$ known yeast TFs to their respective target promoters. Following Lee et al., we considered that a TF binds to the promoter region of a gene if the reported binding p-value was less than 0.001. We restricted our analysis to those genes that contained at least one TF according to the binding data and had no missing values in their expression profiles. This resulted in $n = 2104$, 1645 and 2142 genes in the diauxic, alpha and carbon data sets, respectively.

9.3 Experimental Results

In each test material, both the KNN and TFKNN imputation methods performed well with 20 neighbors. This K-value is in line with the results of Troyanskaya et al. [14] and was used in all the results below.

Fig. 9.1a illustrates the accuracy of the imputation methods at different missing value percentages in the diauxic and carbon data. As expected, the imputation accuracy degraded as the proportion of missing values increased. In all cases, the TF-guided KNN method was at least as accurate as the KNN method based on the expression data only. The benefit from using the TF binding information was emphasized especially at large missing value percentages. In general, the automatic a-value estimator (TFKNNe) yielded error values close to the minimum errors (TFKNNm), indicating a good performance of the a priori weight selection procedure.

To assess how the number of arrays (experiments) affects the imputation accuracy, we randomly sampled smaller subsets of sizes 4, 6, 8 and 10 from the alpha data with 10% and 20% missing value rates. Fig. 9.1b shows that the imputation accuracy decreased as the number of arrays in the data decreased. The benefit gained from the TF-guided gene selection was largest when only a small number of arrays was available.

Fig. 9.2 supports these conclusions in terms of the estimated a-values in the three data sets. The relative importance of the TF binding data increased when the percentage of missing values increased or the number of arrays decreased.

9.4 Conclusions

The imputation framework was selected in this study because it provides a direct way to evaluate the performance of the different distance measures. Moreover, missing value imputation remains a challenging practical problem in microarray analysis. It has been observed that missing observations can drastically reduce the performance of different data analysis methods such as clustering [4]. When the number of genes with missing values is large, it is not possible to simply remove these genes; instead

the missing values need to be imputed as accurately as possible before the actual data analysis. The present results indicate that, in general, TF binding information enhances the prediction of expression values, especially when the number of arrays is small or the proportion of missing values is large.

A proper combination of the individual distance functions is crucial when designing the integrated gene selection measure. We observed that the weight a was highly dependent on the data set used. Hence, we propose a method to select a automatically from the data instead of using a fixed a for all data sets. We have also tried other ways to combine the information sources, including the weighted sum of the individual ranks and a linear regression applied to a training data set, but these did not provide better results than the simple linear combination of the expression and promoter distances.
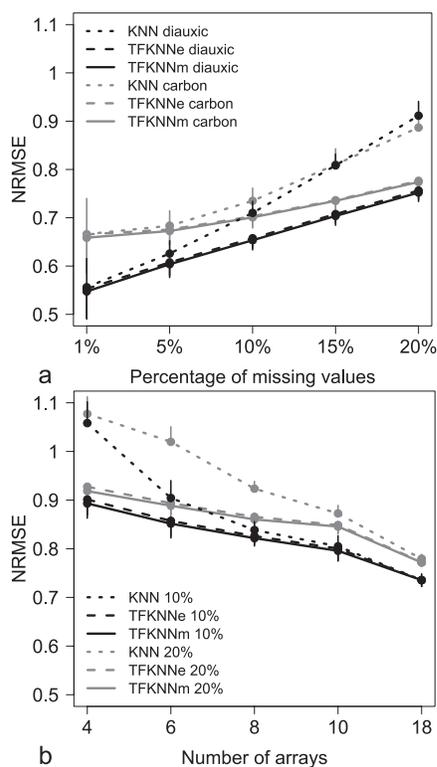

Fig. 9.1. The average normalized root mean squared error (NRMSE) of the imputation methods and their standard deviations (error bars). The results are shown **(a)** at different missing value percentages in the diauxic and carbon data, and **(b)** at different numbers of arrays in the alpha data with 10% and 20% missing value rates. KNN refers to the conventional K nearest neighbor imputation and TFKNN to the TF-guided KNN imputation. TFKNNe corresponds to the automatic selection of the weight factor a, whereas TFKNNm shows the minimum errors obtained when a was selected on the basis of the true values of missing data.
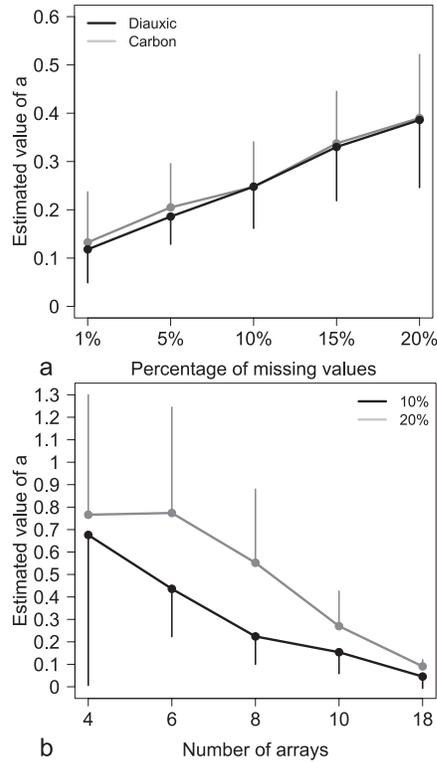
Fig. 9.2. The average values of the weight factor a and their standard deviations (error bars) **(a)** at different missing value percentages in the diauxic and carbon data, and **(b)** at different numbers of arrays in the alpha data with 10% and 20% missing value rates.

We have also tested alternative expression and promoter similarities, such as Pearson correlation and Hamming distance, respectively.

In addition to the KNN-based imputation, we have investigated the effect of TF-guided gene selection on the local least squares (LLS) imputation [9]. In general, the more advanced LLS estimation outperformed the KNN estimation, but the benefit gained from the TF binding information was less evident. This occurred mainly because the LLS imputation performs best with relatively large K-values (typically $K > 100$), reducing the importance of accurate gene selection. Moreover, the LLS formulation for the neighboring genes is based on the expression data only. Our future goal is to improve the accuracy of the LLS method by also using the promoter information in weighing the estimation step, similarly to what was done in (9.4).

Instead of utilizing the different data sources sequentially, we incorporated the additional biological knowledge into the combined distance function. With coherent information sources, this should lead to increased stability in the selection of similar genes [8]. Although we applied the new distance measure in the context of imputation, a similar approach could also be used when clustering genes. Moreover, the method

provides a simple way to investigate the regulatory mechanisms of specific genes. The general idea of data integration is not limited to TF binding data only; other external information on the relations of genes could also be used. The eventual goal is to integrate such complementary information on the complex networks that regulate biological processes at multiple levels.

References

1. Allocco, D.J., Kohane, I.S., Butte, A.J.: Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics, **5**, 18 (2004).
2. Altman, R.B., Raychaudhuri, S.: Whole-genome expression analysis: challenges beyond clustering. Curr. Opin. Struct. Biol., **11**, 340–347 (2001).
3. Bar-Joseph, Z., Gerber, G.K., Lee, T.I., Rinaldi, N.J., You, J.Y., Robert, F., et al.: Computational discovery of gene modules and regulatory networks. Nat. Biotechnol., **21**, 1337–1342 (2003).
4. De Brevern, A.G., Hazout, S., Malpertuy, A.: Influence of microarrays experiments missing values on the stability of gene groups by hierarchical clustering. BMC Bioinformatics, **5**, 114 (2004).
5. DeRisi, J.L., Iyer, V.R., Brown, P.O.: Exploring the metabolic and genetic control of gene expression on a genomic scale. Science, **278**, 680–686 (1997).
6. Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., et al.: Genomic expression program in the response of yeast cells to environmental changes. Mol. Biol. Cell, **11**, 4241–4257 (2000).
7. Glenisson, P., Mathys, J., De Moor, B.: Meta-clustering of gene expression data and literature-based information. SIGKDD Explorations, **5**, 101–112 (2003).
8. Hanisch, D., Zien, A., Zimmer, R., Lengauer, T.: Co-clustering of biological networks and gene expression data. Bioinformatics, **18**, S145–S154 (2002).
9. Kim, H., Golub, G.H., Park, H.: Missing value estimation for DNA microarray gene expression data: local least squares imputation. Bioinformatics, **21**, 187–198 (2005).
10. Lapidot, M., Pilpel, Y.: Comprehensive quantitative analyses of the effects of promoter sequence elements on mRNA transcription. Nucleic Acids Res., **31**, 3824–3828 (2003).
11. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., et al.: Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science, **298**, 799–804 (2002).
12. Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., et al.: Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol. Biol. Cell, **9**, 3273–3297 (1998).
13. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., Church, G.M.: Systematic determination of genetic network architecture. Nat. Genet., **22**, 281–285 (1999).
14. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., et al.: Missing value estimation methods for DNA microarrays. Bioinformatics, **17**, 520–525 (2001).

Chemical Organizations in the Central Sugar Metabolism of *Escherichia coli*

Florian Centler, Pietro Speroni di Fenizio, Naoki Matsumaru, and Peter Dittrich

Bio Systems Analysis Group, Jena Centre for Bioinformatics (JCB) and
Department of Mathematics and Computer Science, Friedrich-Schiller-University Jena
D-07743 Jena, Germany; http://www.minet.uni-jena.de/csb/; centler@minet.uni-jena.de

Summary. The theory of chemical organizations is employed as a novel method to analyze and understand biological network models. The method allows us to decompose a chemical reaction network into sub-networks that are (algebraically) closed and self-maintaining. Such sub-networks are termed *organizations*. Although only stoichiometry is considered to compute organizations, the analysis allows us to narrow down the potential dynamic behavior of the network: organizations represent potential steady state compositions of the system. When applied to a model of sugar metabolism in *E. coli* including gene expression, signal transduction, and enzymatic activities, some organizations are found to coincide with inducible biochemical pathways.

Key words: Chemical organization theory, network analysis, stoichiometry, systems biology, sugar metabolism, *Escherichia coli*.

10.1 Introduction

A living cell consists of a tremendous number of components that interact in complicated ways sustaining the processes of life [7]. In order to understand cells, these interactions are commonly portrayed as networks on different levels [2]. Gene regulatory networks describe how genes are regulated, metabolic networks detail how substrates are transformed into products by proteins acting as enzymes, and signal transduction networks focus on how external stimuli are sensed and transduced leading to a change in gene expression. With more and more detailed knowledge on the various molecular interactions, the constructed networks modeling cellular processes grow steadily in size and complexity. Novel methods have to be developed to analyze and study them. For example, methods originating from graph theory have been succesfully applied to study cellular networks [1]. Other methods concentrate on feasible steady state flux distributions in metabolic networks [10]. In this chapter, we employ the theory of chemical organizations [3] as a novel tool to analyze intracellular reaction networks. The network is decomposed into sub-networks that are (algebraically) closed and self-maintaining, revealing the internal structure of the network. Applying

the method to a well-established model of *E. coli* sugar metabolism reveals an organizational structure in accordance with biological knowledge. Although the analysis does not lead to novel biological insights in this case, it highlights the potential and the limits of this approach. This chapter exemplifies how organization theory can contribute to a systems-level understanding of large-scale models of biological systems, in turn contributing to the emerging field of systems biology.

The outline of the chapter is as follows. The theory of chemical organizations is introduced in Sect. 10.2. The method is then applied to the sugar metabolism network of *E. coli*, and the results are presented in Sect. 10.3. The discussion follows in Sect. 10.4, and we conclude in Sect. 10.5.

10.2 Theory of Chemical Organizations

The theory of chemical organizations [3] extends ideas by Fontana and Buss [4]. It provides a new method to analyze complex general reaction networks. Since the static part of the theory, which is used here, is based solely on network structure and stoichiometric information, no kinetic data is required. The main objective is to determine combinations of network species that are more likely to be present over a long period of (simulation-) time than others. More precisely, the given reaction network is decomposed into sets of molecular species that form algebraically closed and self-maintaining sub-networks. Such species sets are called organizations. The first property—closure—ensures that, given the molecular species of an organization, there is no reaction within the reaction network that could create a species not yet present in the organization from the organization species set. The second property—self-maintenance—guarantees that every molecular species that is used up within an organization can be reproduced by reactions among species of that organization: considering only the reaction network made up by the species contained in the organization, a flux vector[1] can be found, such that all species of the organization are produced at nonnegative rates from within the organization, facilitating maintenance of the organization. Formal definitions of these concepts are given in Sect. 10.2.1. Using this approach, the network is analyzed on a more abstract level than by investigating its state space. In the classic systems approach, the concentrations of all system variables determine the state of the system; however, here, the system state is characterized by a set of species being present. The theory of chemical organizations delivers a set of organizations, representing all self-maintaining and closed sub-networks of the system. It is shown by Dittrich and Speroni di Fenizio [3] that, assuming that the dynamics is modeled using ordinary differential equations, all steady states of the system are instances of organizations, i.e., the species with concentrations greater than zero in a particular steady state are exactly the species contained in a corresponding organization. But not all organizations harbor steady states. For example, an internal cycle not depending on input can fulfill the properties of closure and self-maintenance, yet it

---

[1] For a reaction system of n reactions, the flux vector $\mathbf{v} \in \mathbb{R}_+^n$ assigns to each reaction a nonnegative value that represents the reaction's turnover rate.

is thermodynamically infeasible. Furthermore, organizations can contain species with positive production rates. Since organizations may share the same species, the set of organizations together with the set inclusion $\subseteq$ form a partially ordered set that can be visualized in a Hasse diagram providing a hierarchical view on the network under consideration (see Fig. 10.1 for examples). The organizations are vertically arranged according to their size, with organizations containing few molecular species at the bottom. Two organizations are connected by a line if the upper organization contains all species of the lower organization and there exists no other organization between them. The label of an organization in the Hasse diagram contains a list of species contained in the organization. To keep the labels short, only those species are listed that are not already contained in organizations to which a downlink exists. Hence to get the complete list of molecular species of an organization, it is required to collect the molecular species contained in organizations to which a downlink exists plus the species denoted in the organization label.

10.2.1 Formal Definition of Central Concepts

Algebraic chemistry [3] Let $\mathcal{M}$ be a set of elements (called species, molecular species, or just molecules). $\mathcal{P}_M(\mathcal{M})$ denotes the set of all multisets with elements from $\mathcal{M}$. A multiset differs from a set in that it can contain the same element more than once. Reactions occurring among the species $\mathcal{M}$ can then be defined by a relation $\mathcal{R} \subseteq \mathcal{P}_M(\mathcal{M}) \times \mathcal{P}_M(\mathcal{M})$. We call the pair $\langle \mathcal{M}, \mathcal{R} \rangle$ an *algebraic chemistry*.

Closed set [4] A set of species $S \subseteq \mathcal{M}$ is *closed*, if for all reactions $(A \to B) \in \mathcal{R}$ with $A \in \mathcal{P}_M(S)$, also $B \in \mathcal{P}_M(S)$. In other words: if the educts of a reaction are contained in S, then its products must also be in S. There is no reaction that could create any new species not yet in S from species contained in S.

Self-maintaining set [3] Given an algebraic chemistry $\langle \mathcal{M}, \mathcal{R} \rangle$ with $m = |\mathcal{M}|$ species and $n = |\mathcal{R}|$ reactions, its dynamics can be described by $\dot{\mathbf{c}} = \mathbf{M}\mathbf{v}$ with concentration vector $\mathbf{c} \in \mathbb{R}_+^m$, stoichiometric matrix $\mathbf{M}$, and flux vector $\mathbf{v} \in \mathbb{R}_+^n$. A set of species $S \subseteq \mathcal{M}$ is called *self-maintaining* if a flux vector $\mathbf{v}$ exists, so that the following three conditions are fulfilled:
(1) For every reaction $(A \to B) \in \mathcal{R}$ with $A \in \mathcal{P}_M(S)$, its corresponding flux is $v_{A \to B} > 0$.
(2) For every reaction $(A \to B) \in \mathcal{R}$ with $A \notin \mathcal{P}_M(S)$, its corresponding flux is $v_{A \to B} = 0$.
(3) For every species $i \in S$, its concentration change is nonnegative: $\dot{c}_i \geq 0$.
In other words: if we consider only the sub-network made up by the species of S and additionally the species that can be created from S (but are not in S) (conditions (1) and (2)), we can find a positive flux vector, such that no species of S decays (condition (3)). Note that the steady state condition with $\dot{c}_i = 0$ for all species $i \in \mathcal{M}$ is a special case of condition (3).

Organization [3,4] A set of species $S \subseteq \mathcal{M}$ that is closed and self-maintaining is called an *organization*.

10.3 Application to a Model of Regulated Sugar Metabolism in *E. coli*

In order to demonstrate the feasibility of organization theory as a tool to analyze intracellular reaction networks, we apply it to a relatively small network model encompassing the well-studied sugar metabolism of *E. coli*. If several sugars are available in the growth medium, *E. coli* first exclusively metabolizes its preferred carbon source, glucose. Only after depletion of glucose will the bacterium begin to utilize other available sugars. This diauxic growth phenomenon has been extensively studied in experiments and by mathematical modeling [6,11,12], leading to a good understanding of the molecular mechanisms at work. The two main mechanisms facilitating the switch-like behavior are inducer exclusion and catabolite repression. See the referenced literature for details of these mechanisms. Extending models by Kremling et al. [6] and Wang et al. [12], Puchalka and Kierzek constructed a reaction network modeling the sugar metabolism of *E. coli* including gene expression, signal transduction, and transport and enzymatic activities [9]. We take this network as an example to demonstrate how the theory of organizations can be applied to intracellular networks. First, the network is adapted as described in the next section. Then, organizations are analyzed for several scenarios representing bacterial growth on different sugar sources.

10.3.1 Reaction Network

The original network by Puchalka and Kierzek consists of 92 substances reacting with each other in 120 reactions. The model contains reactions modeling transcription and translation of 21 genes. The uptake and utilization of external glucose, lactose, and glycerol is included in the model as well as catabolic repression and inducer exclusion, allowing the model to exhibit diauxic growth. Each reaction of the network consists of (up to) three different types of species: educts, products, and modifiers. If a reaction occurs, the educt species are transformed into the product species while the modifiers are not affected. Modifier species only change the reaction rate. Two types of modifiers are used in the model: enzymes, which are required for a reaction to take place, and effectors, which increase the reaction rate acting as an activator, or decrease the reaction rate acting as an inhibitor or repressor. Since algebraic chemistries do not contain modifiers, we have to handle them separately for our analysis as follows. If a reaction does not have modifiers, we take the reaction exactly as it is. In the presence of modifiers, we inspect the reaction rate equation. If the modifier species concentration has to be greater than zero for the reaction rate to become greater than zero, we add the modifier species on both educt and product side of the reaction. This is the typical case for enzymes. Only in their presence can the reaction in question be performed. If the reaction rate can be greater than zero even in the absence of the modifier species, we simply ignore them, as they are not necessary for the reaction to take place. They merely increase (or decrease) the reaction rate, acting as nonessential activators (resp. repressors or inhibitors). It is important to note that all inhibitory or negative interactions are ignored by this procedure.

The handling of modifiers as described above cannot be applied to reactions modeling gene expression. Negative interactions can be ignored as before, but activators need special treatment. The model contains five transcription reactions that have activating and/or repressing effectors. With activator concentrations being zero, the transcription reaction rates in the original model are computed to be positive. This corresponds to a basal transcription rate of a gene: even if activators are not present, RNA polymerase occasionally binds to the promoter and transcription is initiated, leading to a basal concentration of the respective protein. Applying the procedure as described above to these reactions (i.e., ignoring all activators) would lead to an unconditional transcription of all genes, giving rise to a basal concentration of the corresponding gene products. But as shown below for the transcription of the *lac* genes, basal concentration of proteins might not be sufficient to perform certain metabolic tasks. Consequently, a protein having only basal concentration should be regarded as not being present in our analysis. Only if activators are present, increasing the transcription rate so that protein concentrations reach levels that are significantly above basal level—effectively switching the gene on—should the corresponding protein be regarded as being present. Activators and inducers for gene transcription should therefore be modelled as necessary catalysts in gene transcription reactions. The five transcription reactions having effectors are discussed separately as follows.

Transcription of crp: *effectors Crp, cAMP.* Crp is activated by the binding of cAMP. The activated Crp-cAMP complex negatively regulates the transcription of *crp*. It was also shown that with further increasing concentration of Crp-cAMP this inhibition is overcome and an upregulation occurs [5]. The inhibition is ignored and since the activation only occurs at high concentrations, it is ignored as well (since the reaction can take place in the absence of the effector species). Hence the effectors Crp and cAMP are ignored for this reaction.

Transcription of cya: *effectors Crp, cAMP.* Crp-cAMP downregulates transcription of *cya*. Being an inhibition, the effector species Crp and cAMP are ignored for this reaction.

Transcription of lacZY, glpFK, and glpD: *effectors Crp, cAMP, LacI/GlpR, and Allo/ G3P.* These genes code for enzymes necessary for lactose and glycerol uptake and utilization. The transcription regulation is similar for both. Two mechanisms are at work for transcription regulation of *lacZY* (*glpFK*, *glpD*). First, repressor LacI (GlpR) represses transcription. If inducer Allo (G3P) is present, it binds to LacI (GlpR) and by this inactivates the repressor. Second, Crp-cAMP complex acts as an activator. Both mechanisms are modeled in one reaction equation in the model. We ignore the inhibiting effect of effector species LacI (GlpR). Instead, by adding the inducer Allo (G3P) on both educt and product side of the reaction, we require the inducer to be present for transcription. This is in accordance with biological knowledge: only in the presence of the inducer are the corresponding gene products synthesized at above basal concentration levels. Mutants not being able to synthesize Crp or cAMP were found unable to grow on several carbon sources [8]. Therefore, we conclude that the presence of Crp and cAMP is also required to synthesize enzymes necessary for carbon uptake and

utilization in sufficient concentrations. Accordingly, effectors Crp and cAMP are also added on both educt and product side of the reactions.

The original model contains six reactions that are reversible. We add an explicit back reaction for each of them in our model. Cell growth and cell division are accounted for in the original model by dividing all species concentrations by two on cell division, except for the DNA species. Hence, we add decay reactions for all non-DNA species that do not already decay in the original model. The remaining species that do not decay are: all 21 promoter species, RNAP, Tscription, Glcex, Lacex, and Glyex. Several species are not produced from within the original network model. Among them are all 21 promoter species, ATP, ADP, and AMP. We assume that they are present in the cell at all times by providing them as external input. We add a reaction of the form $\emptyset \rightarrow$ inputspecies for each of them. Additionally, RNAP is provided as input. Finally, our network model consists of 92 species and 168 reactions. See the Appendices for a complete list of species and reactions. Glucose, lactose, and glycerol in the growth medium are represented by the species Glcex, Lacex, and Glyex. By adding additional input reactions for these species, growth on different sugar sources can be modeled.

10.3.2 Hierarchies of Organizations

We compute the hierarchy of organizations of the network for five different scenarios. The scenarios only differ in which external sugars are supplied as input, resembling bacterial growth on different sugar sources. First, no external sugars are supplied at all. Then, one of the three sugars glucose, lactose, or glycerol is consecutively supplied as the exclusive carbon source. And finally, all three sugars are provided simultaneously. Supplying a sugar source is accomplished simply by adding an input reaction of the form $\emptyset \rightarrow$ externalsugar to the reaction network. Changing the reaction network also changes the hierarchy of organizations. The resulting hierarchies are depicted in Fig. 10.1. They all consist of four organizations. The labels within organizations refer to sets of species as detailed in Table 10.1. The network model covers the transformation of external sugar into pyruvate, which is then fed into further metabolic processes not considered by the model. These follow-up processes enabling cellular survival are represented by pseudo species Metabolism. Species set Metabolites contains all relevant species of this pathway, and its presence in an organization hence represents a cell being able to maintain its metabolism and survive.

Starvation. No external sugars are supplied as input. The resulting hierarchy of organizations is depicted in Fig. 10.1(a). The smallest organization Org. 1 contains all input species (21 promoter species, ATP, ADP, AMP, and RNAP). In the presence of the promoters and RNA polymerase, all unregulated genes are transcribed and translated, so that all mRNA and protein species of all 18 unregulated genes are also contained in the smallest organization (cf. Genes+Enzymes, Table 10.1). Organizations Org. 2 and Org. 3 contain all species from Org. 1 and additionally Glyex and Lacex, respectively. This seems surprising since these species are not supplied as input in this scenario. But recall that an organization is a set of species that is algebraically closed and self-maintaining. Although the species Glyex and Lacex are not supplied as input,
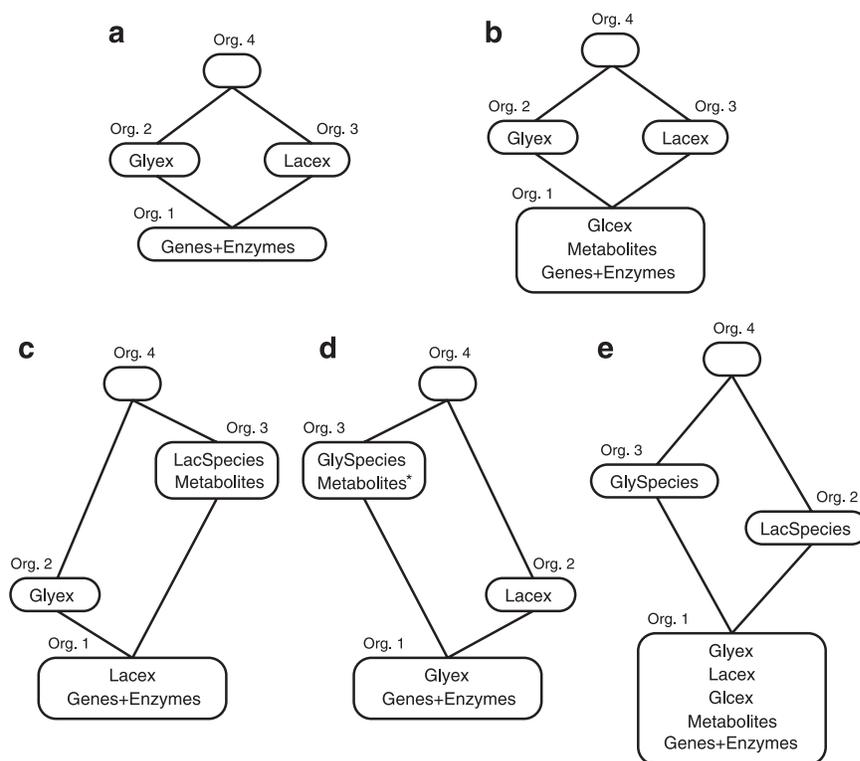
Fig. 10.1. Hierarchies of organizations of the *E. coli* network for five scenarios differing in supplied external sugars, resembling growth on different carbon sources. Organizations consist of the species sets contained in their lower organization(s) plus the species set(s) denoted in their label. Species set labels are detailed in Table 10.1. (**a**) starvation; (**b**) growth on glucose only; (**c**) growth on lactose only; (**d**) growth on glycerol only; (**e**) growth on glucose, lactose, and glycerol. See text for details.

they are still a regular part of the reaction network. Inspecting the networks making up Org. 2 and Org. 3, we find that Glyex and Lacex do not participate in any reaction there. They are isolated nodes in the reaction network. As such, they do not decay, neither are they produced, fulfilling the requirements of closure and self-maintenance. The two organizations represent a state in which a fixed amount of Glyex, respectively Lacex, entered the system "by accident" and the uptake systems are not induced. In this case, the concentration of the external sugars will not change. Only after the uptake systems have been induced will the external sugars be used up completely and the system falls back to Org. 1. The largest organization Org. 4 combines Org. 2 and Org. 3. All species of the smallest organization and Glyex and Lacex are contained. In this scenario, we find no organization containing the metabolites of the network. This indicates that with no external sugar source, the network cannot sustain its metabolism, i.e., the cell is starving.

Table 10.1. Sets of species as used in Fig. 10.1.

Genes		
+Enzymes	:=	{PromCrp, PromCya, PromEIIA, PromEIIBC, PromEI, PromFbp, PromFda, PromGap, Prom GlcT, PromGlk, PromGlpD, PromGlpFK, PromGlpR, PromGpm, PromHPr, PromLacI, PromLacZY, PromPfk, PromPgi, PromPyk, PromTpi, RNAP, Tscription, CrpmRNA, CyamRNA, EIIAmRNA, EIIBCmRNA, EImRNA, FbpmRNA, FdamRNA, GapmRNA, GlcTmRNA, GlkmRNA, GlpRmRNA, GpmmRNA, HPrmRNA, LacImRNA, PfkmRNA, PgimRNA, PykmRNA, TpimRNA, Crp, Cya, EIIA, EIIBC, EI, Fbp, Fda, Gap, GlcT, Glk, GlpR, Gpm, HPr, LacI, Pfk, Pgi, Pyk, Tpi, AMP, ATP, ADP, cAMP}
Metabolites	:=	{Glc, Glc6P, Fru6P, FBP, DHAP, T3P, 3PG, PEP, Pyr, Metabolism, EIIAP, HPrP}
Metabolites*	:=	Metabolites\{Glc}
Glcex	:=	{Glcex}
Lacex	:=	{Lacex}
Glyex	:=	{Glyex}
LacSpecies	:=	{Lac, Allo, LacZYmRNA, LacZYmRNA1, LacZ, LacY}
GlySpecies	:=	{Gly, G3P, GlpDmRNA, GlpFKmRNA, GlpFKmRNA1, GlpD, GlpF, GlpK}

Growth on glucose. After adding the reaction $\emptyset \rightarrow$ Glcex, the hierarchy of organizations again contains four organizations as shown in Fig. 10.1(b). The smallest organization Org. 1 contains the same species as in the first scenario and additionally Glcex. With Glcex present, all metabolites can be created and maintained. Consequently, all these species are part of the smallest organization, too. With species set Metabolites present in the smallest organization, the cell can maintain its metabolism when external glucose is supplied. The remaining part of the organization hierarchy is equivalent to the first scenario without any sugar input.

Growth on lactose. When lactose is supplied as the exclusive external sugar source, the resulting hierarchy of organizations again contains four organizations as depicted in Fig. 10.1(c). The smallest organization contains all unregulated genes and enzymes and additionally Lacex. In Org. 2, only Glyex is added as in the previous cases. Organization Org. 3 contains the species of the smallest organization, all species necessary for taking up and metabolizing external lactose, and the species belonging to the metabolism. Being an organization, the network composed of all these species is algebraically closed and self-maintaining, representing a cell that has switched its *lac* genes on and utilizes external lactose. Fig. 10.2(a) details schematically how Org. 1 is expanded to form Org. 3. Once inducer allolactose is present, the *lac* genes are switched on and LacY and LacZ are synthesized. LacY facilitates the uptake of external lactose while LacZ transforms intracellular lactose and allolactose to glucose and
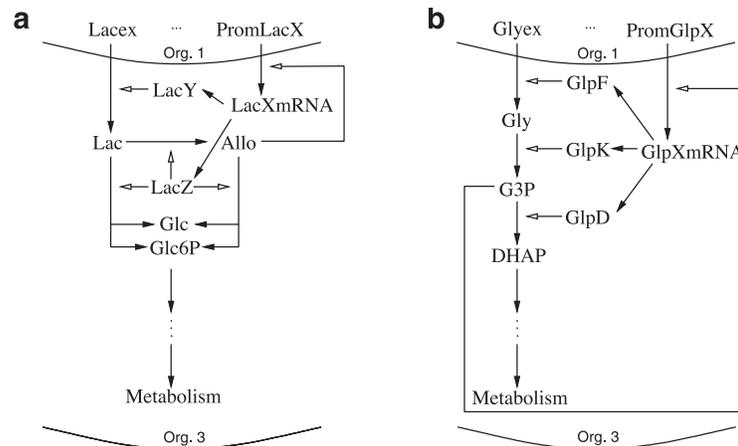
Fig. 10.2. Induction of sugar uptake systems. When lactose or glycerol is the exclusive carbon source, organization Org. 1 corresponds to the state in which the respective uptake systems are not activated and the bacterium is starving (upper part). In organization Org. 3, the systems are induced and the external sugar is utilized. A schematic sketch of the reaction network of organization Org. 3 responsible for utilization of (**a**) external lactose and (**b**) external glycerol is shown. Open arrows point from species acting as catalysts to the reactions that are catalyzed. See text for details.

glucose-6-phosphate. Additionally, LacZ transforms lactose to allolactose, closing the positive feedback loop. Glucose then enters the metabolic pathway leading to pyruvate and further metabolic processes. Adding Glyex to Org. 3 results in the largest organization Org. 4. This scenario shows that bacterial growth is possible on lactose as the only carbon source after induction of the lactose uptake system (in Org. 3 and 4).

Growth on glycerol. Now glycerol is provided as the exclusive carbon source. The resulting hierarchy of organizations is visualized in Fig. 10.1(d). The result is equivalent to the lactose scenario. The smallest organization Org. 1 contains the unconditionally transcribed genes and resulting enzymes, and external glycerol. Organization Org. 3 additionally contains the molecular species necessary for utilizing external glycerol and the metabolism species. Fig. 10.2(b) shows how this organization is formed by expanding Org. 1. Once inducer G3P is present, the genes corresponding to glycerol utilization are switched on and GlpF, GlpK, and GlpD are synthesized. GlpF then enables uptake of external glycerol, GlpK transforms internal glycerol to G3P closing the positive feedback loop, and GlpD transforms G3P to DHAP, which in turn fuels the pathway ending in pyruvate and further metabolic processes. Adding Lacex to this organization leads to the largest organization Org. 4. Again we find that once the uptake system for the external sugar is induced, the cell can maintain its metabolism in Org. 3 and 4.

Growth on all sugars. In the last scenario, all three external sugars are supplied as input simultaneously. Figure 10.1(e) depicts the resulting hierarchy of organizations. With external glucose being input, the smallest organization resembles the smallest

organization of the glucose scenario, with external lactose und glycerol added. Glucose alone is sufficient for growth, hence the smallest organization already represents a state in which the cell grows (on glucose). The two organizations above the smallest one contain the species necessary for utilizing lactose (Org. 2) and glycerol (Org. 3). They represent states in which the cell metabolizes lactose, respectively glycerol, in addition to glucose. The largest organization Org. 4 finally merges Org. 2 and 3, containing all species of the model. Here, all three sugars are metabolized simultaneously. From a biological point of view, only organization 1 is meaningful since the uptake of lactose and glycerol is repressed in the presence of glucose. The existence of the remaining organizations will be discussed in the next section.

10.4 Discussion

In all five analyzed scenarios the hierarchy of organizations consists of four organizations, representing four potential steady state species compositions of the system. Some organizations just contain a lower organization and a new species that does not interact with the species of the lower organization (e.g., Org. 2 and 3 in the starvation scenario and in the glucose scenario). In other cases, exactly those species performing a specific cellular function make up the difference between an organization and its lower neighbor (cf. Org. 2 and 3 in the scenario with all sugars supplied). In these cases a modularity of the analyzed network model is uncovered by organization theory. In this example, the uncovered modules correspond to the inducible uptake systems for lactose and glycerol. Only those organizations that contain the metabolic species correspond to system states facilitating bacterial growth. As expected, such an organization is not found in the scenario without any supplied sugar. For glucose as the exclusive carbon source, all organizations contain the metabolites. For lactose and glycerol, only those organizations contain the metabolites that also contain the species of the respective uptake systems. This result confirms that glucose can be unconditionally utilized, while lactose and glycerol can only be utilized after their respective uptake systems have been induced. The diauxic growth behavior of *E. coli* is not revealed by the hierarchy of organizations. In the scenario with three sugars supplied as input, organizations are found that correspond to states where glucose and other sugars are utilized simultaneously. First, this highlights the fact that organizations only represent potential steady states of the system. Further kinetic information is required to determine whether an organization indeed contains steady states or not. And second, inhibitory interactions play a crucial role in diauxic growth, but had to be ignored in the conversion of the original network model. Since inhibitory interactions in the original network only decrease reaction rates, they in principle cannot be captured by the theory of organizations in which only the presence or absence of molecular species is considered.

10.5 Conclusion

We have demonstrated how the theory of chemical organizations can be employed to uncover modularity in intracellular reaction network models. The theory operates on a high level of abstraction, as only the presence or absence of species is considered compared to the continuous state space considered in classic approaches. Consequently, concentration-dependent interactions (e.g., nonessential activation of enzymes or inhibitory interactions) cannot be taken into account. Nevertheless, profound results can be obtained. Organizations represent potential steady state species compositions of the model. The hierarchy of organizations, reflecting the structure of the network model, provides a new perspective on the system and its potential dynamic behavior. The movement of the system through state space can be mapped to a movement in the space of its organizations [3], leading to a reduction in dimensionality. Organizations, being closed and self-maintaining sub-networks, can be separately analyzed using classic methods. Especially for large networks, analyzing small sub-networks is more feasible than studying the whole network at once. With species in organizations typically having more interactions among each other than with outside species, organizations can also be used for network visualization. By grouping species belonging to one organization closely together, a clearer graphical representation of the whole network can be achieved. Since only stoichiometry is required for the analysis, the method can be applied to a broad range of network models ranging from chemical and biochemical networks to social networks. The results presented in this chapter suggest that the theory of organizations will be a helpful tool for studying and understanding large-scale intracellular network models.
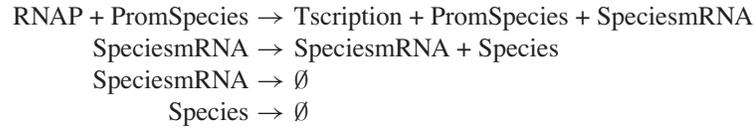
Acknowledgments

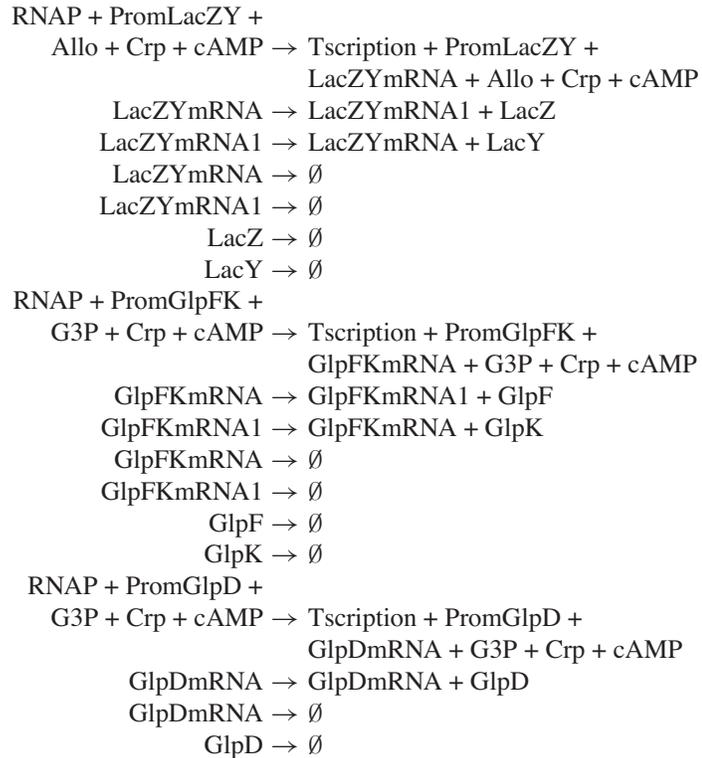## Appendix A: List of Species

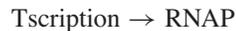| Species Names	Substances
ATP, ADP, AMP, cAMP	ATP, ADP, AMP, and cyclic AMP
RNAP, Tscription	RNA polymerase and RNAP bound to DNA
Crp, PromCrp, CrpmRNA	catabolite repressor protein, gene, and mRNA
Cya, PromCya, CyamRNA	adenylate cyclase, gene, and mRNA
EIIA, PromEIIA, EIIAmRNA	PTS system enzyme IIA^{Glc}, gene, and mRNA
EIIAP	phosphorylated PTS system enzyme IIA^{Glc}
EIIBC, PromEIIBC, EIIBCmRNA	PTS system enzyme IIBC^{Glc}, gene, and mRNA
EI, PromEI, EImRNA	PTS system enzyme I, gene, and mRNA
Fbp, PromFbp, FbpmRNA	fructose bisphosphatase, gene, and mRNA
Fda, PromFda, FdamRNA	fructose bisphosphate aldolase, gene, and mRNA
Gap, PromGap, GapmRNA	glyceraldehyde-3-phosphate dehydrogenase, gene, and mRNA
GlcT, PromGlcT, GlcTmRNA	glucose transporter, gene, and mRNA
Glk, PromGlk, GlkmRNA	glucokinase, gene, and mRNA
GlpD, PromGlpD, GlpDmRNA	glycerol-3-phosphate dehydrogenase, gene, and mRNA
GlpFKmRNA, GlpFKmRNA1	glpFK operon mRNA
GlpR, PromGlpR, GlpRmRNA	glp regulon repressor, gene, and mRNA
Gpm, PromGpm, GpmmRNA	phosphoglycerate mutase, gene, and mRNA
HPr, PromHPr, HPrmRNA	PTS system HPr protein, gene, and mRNA
HPrP	phosphorylated PTS system HPr protein
LacI, PromLacI, LacImRNA	lac operon repressor, gene, and mRNA
LacZYmRNA, LacZYmRNA1	lac operon mRNA
Pfk, PromPfk, PfkmRNA	phosphofructokinase, gene, and mRNA
Pgi, PromPgi, PgimRNA	phosphoglucose isomerase, gene, and mRNA
Pyk, PromPyk, PykmRNA	pyruvate kinase, gene, and mRNA
Tpi, PromTpi, TpimRNA	triose phosphate isomerase, gene, and mRNA
PromGlpFK, GlpF, GlpK	glpFK operon, glycerol faciliator, and kinase
PromLacZY, LacZ, LacY	lac operon, β-galactosidase, and lactose permease
Glcex, Glyex, Lacex	extracellular glucose, glycerol, and lactose
Glc, Gly, Lac	intracellular glucose, glycerol, and lactose
Allo	Allolactose
Glc6P	glucose-6-phoshpate
G3P	glycerol-3-phosphate
Fru6P	fructose-6-phosphate
FBP	fructose-1,6-bisphosphate
DHAP	dihydroxyacetone phosphate
T3P	glyceraldehyde-3-phosphate
3PG	3-phospho-glycerate
PEP	phosphoenolopyruvate
Pyr	pyruvate
Metabolism	further metabolic processes

Appendix B: Reaction Network

1. Synthesis and decay is identical for species Crp, Cya, EIIA, EIIBC, EI, Fbp, Fda, Gap, GlcT, Glk, GlpR, Gpm, HPr, LacI, Pfk, Pgi, Pyk, and Tpi:
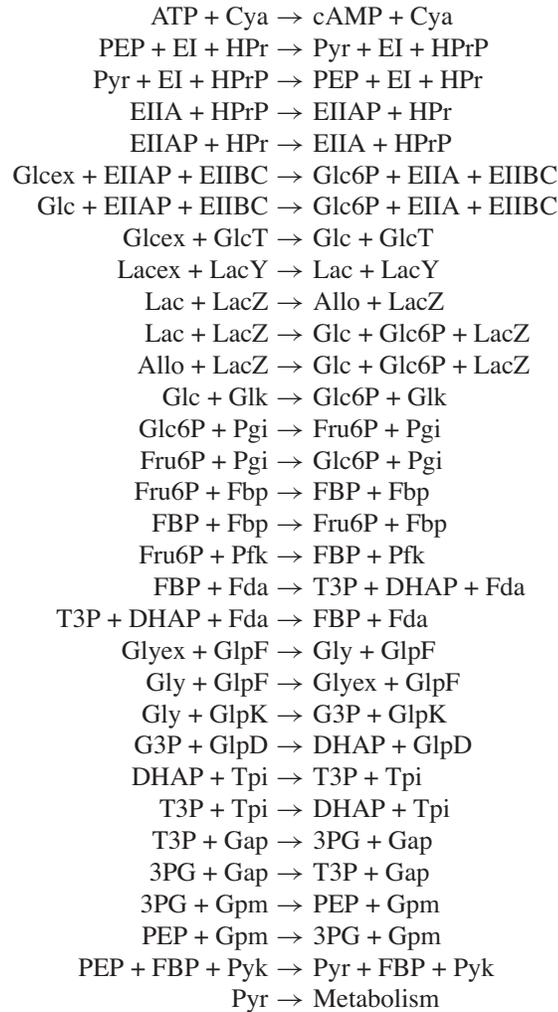
$$RNAP + PromSpecies \rightarrow Tscription + PromSpecies + SpeciesmRNA$$
$$SpeciesmRNA \rightarrow SpeciesmRNA + Species$$
$$SpeciesmRNA \rightarrow \emptyset$$
$$Species \rightarrow \emptyset$$

2. Synthesis and decay of inducible species LacZY, GlpFK, and GlpD:

$$RNAP + PromLacZY +$$
$$Allo + Crp + cAMP \rightarrow Tscription + PromLacZY +$$
$$LacZYmRNA + Allo + Crp + cAMP$$
$$LacZYmRNA \rightarrow LacZYmRNA1 + LacZ$$
$$LacZYmRNA1 \rightarrow LacZYmRNA + LacY$$
$$LacZYmRNA \rightarrow \emptyset$$
$$LacZYmRNA1 \rightarrow \emptyset$$
$$LacZ \rightarrow \emptyset$$
$$LacY \rightarrow \emptyset$$
$$RNAP + PromGlpFK +$$
$$G3P + Crp + cAMP \rightarrow Tscription + PromGlpFK +$$
$$GlpFKmRNA + G3P + Crp + cAMP$$
$$GlpFKmRNA \rightarrow GlpFKmRNA1 + GlpF$$
$$GlpFKmRNA1 \rightarrow GlpFKmRNA + GlpK$$
$$GlpFKmRNA \rightarrow \emptyset$$
$$GlpFKmRNA1 \rightarrow \emptyset$$
$$GlpF \rightarrow \emptyset$$
$$GlpK \rightarrow \emptyset$$
$$RNAP + PromGlpD +$$
$$G3P + Crp + cAMP \rightarrow Tscription + PromGlpD +$$
$$GlpDmRNA + G3P + Crp + cAMP$$
$$GlpDmRNA \rightarrow GlpDmRNA + GlpD$$
$$GlpDmRNA \rightarrow \emptyset$$
$$GlpD \rightarrow \emptyset$$

3. Unbinding of RNAP:

$$Tscription \rightarrow RNAP$$

4. Signal transduction, transport, and metabolic reactions:

$$\text{ATP} + \text{Cya} \rightarrow \text{cAMP} + \text{Cya}$$
$$\text{PEP} + \text{EI} + \text{HPr} \rightarrow \text{Pyr} + \text{EI} + \text{HPrP}$$
$$\text{Pyr} + \text{EI} + \text{HPrP} \rightarrow \text{PEP} + \text{EI} + \text{HPr}$$
$$\text{EIIA} + \text{HPrP} \rightarrow \text{EIIAP} + \text{HPr}$$
$$\text{EIIAP} + \text{HPr} \rightarrow \text{EIIA} + \text{HPrP}$$
$$\text{Glcex} + \text{EIIAP} + \text{EIIBC} \rightarrow \text{Glc6P} + \text{EIIA} + \text{EIIBC}$$
$$\text{Glc} + \text{EIIAP} + \text{EIIBC} \rightarrow \text{Glc6P} + \text{EIIA} + \text{EIIBC}$$
$$\text{Glcex} + \text{GlcT} \rightarrow \text{Glc} + \text{GlcT}$$
$$\text{Lacex} + \text{LacY} \rightarrow \text{Lac} + \text{LacY}$$
$$\text{Lac} + \text{LacZ} \rightarrow \text{Allo} + \text{LacZ}$$
$$\text{Lac} + \text{LacZ} \rightarrow \text{Glc} + \text{Glc6P} + \text{LacZ}$$
$$\text{Allo} + \text{LacZ} \rightarrow \text{Glc} + \text{Glc6P} + \text{LacZ}$$
$$\text{Glc} + \text{Glk} \rightarrow \text{Glc6P} + \text{Glk}$$
$$\text{Glc6P} + \text{Pgi} \rightarrow \text{Fru6P} + \text{Pgi}$$
$$\text{Fru6P} + \text{Pgi} \rightarrow \text{Glc6P} + \text{Pgi}$$
$$\text{Fru6P} + \text{Fbp} \rightarrow \text{FBP} + \text{Fbp}$$
$$\text{FBP} + \text{Fbp} \rightarrow \text{Fru6P} + \text{Fbp}$$
$$\text{Fru6P} + \text{Pfk} \rightarrow \text{FBP} + \text{Pfk}$$
$$\text{FBP} + \text{Fda} \rightarrow \text{T3P} + \text{DHAP} + \text{Fda}$$
$$\text{T3P} + \text{DHAP} + \text{Fda} \rightarrow \text{FBP} + \text{Fda}$$
$$\text{Glyex} + \text{GlpF} \rightarrow \text{Gly} + \text{GlpF}$$
$$\text{Gly} + \text{GlpF} \rightarrow \text{Glyex} + \text{GlpF}$$
$$\text{Gly} + \text{GlpK} \rightarrow \text{G3P} + \text{GlpK}$$
$$\text{G3P} + \text{GlpD} \rightarrow \text{DHAP} + \text{GlpD}$$
$$\text{DHAP} + \text{Tpi} \rightarrow \text{T3P} + \text{Tpi}$$
$$\text{T3P} + \text{Tpi} \rightarrow \text{DHAP} + \text{Tpi}$$
$$\text{T3P} + \text{Gap} \rightarrow \text{3PG} + \text{Gap}$$
$$\text{3PG} + \text{Gap} \rightarrow \text{T3P} + \text{Gap}$$
$$\text{3PG} + \text{Gpm} \rightarrow \text{PEP} + \text{Gpm}$$
$$\text{PEP} + \text{Gpm} \rightarrow \text{3PG} + \text{Gpm}$$
$$\text{PEP} + \text{FBP} + \text{Pyk} \rightarrow \text{Pyr} + \text{FBP} + \text{Pyk}$$
$$\text{Pyr} \rightarrow \text{Metabolism}$$

5. Decay reactions for species ATP, ADP, AMP, cAMP, EIIAP, HPrP, Glc, Gly, Lac, Allo, Glc6P, G3P, Fru6P, FBP, DHAP, T3P, 3PG, PEP, Pyr, and Metabolism have the form:

$$\text{Species} \rightarrow \emptyset$$

6. Input reactions for ATP, ADP, AMP, RNAP, PromCrp, PromCya, PromEIIA, PromEIIBC, PromEI, PromFbp, PromFda, PromGap, PromGlcT, PromGlk, PromGlpD, PromGlpR, PromGpm, PromHPr, PromLacI, PromPfk, PromPgi, PromPyk, PromTpi, PromGlpFK, and PromLacZY have the form:

$$\emptyset \rightarrow \text{Species}$$

References

1. Barabási, A.-L., Oltvai, Z.N.: Network biology: Understanding the cell's functional organization. Nat. Rev. Genet., **5**, 101–113, (2004).
2. Cardelli, L.: Abstract machines of systems biology. In: Priami, C., Merelli, E., Gonzalez, P., Omicini, A. (eds) Transactions on Computational Systems Biology III, vol.3737 of LNCS. Springer, Berlin (2005), 145–168.
3. Dittrich, P., Speroni di Fenizio, P.: Chemical organization theory. Bull. Math. Biol., 2007 (in press).
4. Fontana, W., Buss, L.W.: "The arrival of the fittest": Toward a theory of biological organization. Bull. Math. Biol., **56**, 1–64 (1994).
5. Hanamura, A., Aiba, H.: A new aspect of transcriptional control of the *Escherichia coli crp* gene: Positive autoregulation. Mol. Microbiol., **17**, 2489–2497 (1992).
6. Kremling, A., Bettenbrock, K., Laube, B., Jahreis, K., Lengeler, J.W., Gilles, E.D.: The organization of metabolic reaction networks, iii. application for diauxic growth on glucose and lactose. Metab. Eng., **3**, 362–379 (2001).
7. Monk, N.A.: Unravelling nature's networks. Biochem. Soc. Trans., **31**, 1457–1461 (2003).
8. Postma, P.W., Lengeler, J.W., Jacobson, G.R.: Phosphoenolpyruvate: Carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev., **57**, 543–594 (1993).
9. Puchalka, J., and Kierzek, A.M.: Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. Biophys. J., **86**, 1357–1372 (2004).
10. Schuster, S., Dandekar, T., Fell, D.A.: Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. Trends Biotechnol., **17**, 53–60 (1999).
11. Thattai, M., Shraiman, B.I.: Metabolic switching in the sugar phosphotransferase system of *Escherichia coli*. Biophys. J., **85**, 744–754 (2003).
12. Wang, J., Gilles, E.D., Lengeler, J.W., Jahreis, K.: Modeling of inducer exclusion and catabolite repression based on pts-dependent sucrose and non-pts-dependent glycerol transport systems in *Escherichia coli* k-12 and its experimental verification. J. Biotechnol., **92**, 133–158 (2001).

11

Transition Networks: A Unifying Theme for Molecular Simulation and Computer Science

Frank Noé and Jeremy C. Smith

Interdisziplinäres Zentrum für wissenschaftliches Rechnen (IWR), Universität Heidelberg, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany; Frank.Noe@iwr.uni-heidelberg.de; Jeremy.Smith@iwr.uni-heidelberg.de

Summary. A transition network (TN) is a graph-theoretical concept describing the transitions between (meta)stable states of dynamical systems. Here, we review methods to generate and analyze a TN for molecular systems. The appropriate identification of states and transitions from the potential energy surface of the molecule is discussed. We describe a formalism transforming a TN on a static energy surface into a time-dependent dynamic TN that yields the population probabilities for each system state and the interstate transition rates. Three analysis methods that help in understanding the dynamics of the molecular system based on the TN are discussed: (1) Disconnectivity graphs allow important features of the energy surface captured in a static TN to be visualized, (2) graph-theoretical methods enable the computation of the best transition paths between two predefined states of the TN, and (3) statistical methods from complex network analysis identify important features of the TN topology. A broad review of the literature is given, and some open research directions are discussed.

Key words: AU: Please provide keywords.

11.1 Introduction

Complex dynamical systems with many degrees of freedom are ubiquitous. Examples include climate systems, stock markets and condensed-phase molecular systems, among which biomolecules such as polypeptides, nucleic acids or proteins are of particular interest. The immense number of possible states and state transitions pose a challenge to the simulation of these systems [1–3]. However, the qualitative and quantitative analysis of transitions between stable states is at the heart of understanding their dynamics [4–7]. Here, we review some of the state-of-the-art methods related to *transition networks*, which pursue exactly this goal. Transition networks (TNs) can in principle be used to model the kinetic behavior of any dynamical system that can be appropriately described by a (possibly large) number of states and interstate transition rules. However, we will concentrate on molecular systems in this chapter.

Molecular dynamical systems are often modeled using a potential energy function $E_{\text{pot}}(\mathbf{x}) : \mathbb{R}^D \to \mathbb{R}$, which depends on a system configuration, or state vector $\mathbf{x}$

(here: the atomic coordinates, D is the number of degrees of freedom of the system, in general each atom has three degrees of freedom). Dynamical trajectories typically reside most of the time within the energy basins of $E_{pot}(\mathbf{x})$ and occasionally jump to neighboring basins [8]. The dynamics of the system can be simulated by numerically integrating the equations of motion involved. For stability, the integration time step must not exceed a value that depends on the fastest motions in the system and is often many orders of magnitude below the timescale during which the transitions of interest occur [9]. Larger steps are possible in Monte Carlo simulations, but lead to a considerable reduction of the acceptance ratio [10]. These difficulties often lead to an insufficient number (if any) of occurrences of the transitions being investigated. Despite considerable progress in enhancing sampling methods [9, 11], this *sampling problem* is still the main obstacle to using direct simulation methods for the characterization of rare transitions.

An alternative approach to exploring $E_{pot}(\mathbf{x})$ directly is to "map" its interesting features into a TN. Formally, a TN is a weighted graph, $\mathcal{G}$, whose definition involves a list of vertices, $\mathcal{V}$, representing the stable, low-energy states of the molecule, and a list of edges, $\mathcal{E}$, specifying between which pairs of states direct sub-transitions are considered. The construction of TNs is documented in a large number of studies [1,5–7,12–29].

Here, we distinguish between *static* and *dynamic* TNs. Static TNs describe features of the potential energy surface $E_{pot}(\mathbf{x})$ and their network weights correspond to energies of stable states and transition states. Dynamic TNs are time dependent and give, for each time t, the residence probabilities of the TN vertices and the transition rates associated with the TN edges.

The general goal of molecular simulation is to compute some system properties which are, in general, global (i.e., they arise from the collective interplay of the microscopic interaction rules), such as the most dominant pathway for a transition between two defined system states or the mean time required for this transition. Fig. 11.1 illustrates how the present TN approach is related to other approaches of molecular simulation. The common procedure for this is to compute the system dynamics (e.g., classical) based on the potential energy surface, giving rise to a time series (a trajectory through configurational space) and distributions (e.g., a configurational state density). Statistical mechanics is used to calculate the desired global system properties [30]. In the approach described here, one samples the potential energy surface and represents its features in a *static* TN. From this, a *dynamic* TN is generated using either equilibrium statistical mechanics (equilibrium case) or a master-equation approach (nonequilibrium case). Using graph theory paired with statistical mechanics allows global system properties to be derived from the dynamic TN. A dynamic TN can also be derived from time series obtained by computing the system dynamics [31,32].

This chapter is organized as follows: Sect. 11.2 describes how static TN energies can be obtained. How to obtain the residence probabilities and transition rates involved in a dynamic TN is described in Sect. 11.3. Sect. 11.4 concentrates on the analysis of either a static or dynamic TN.
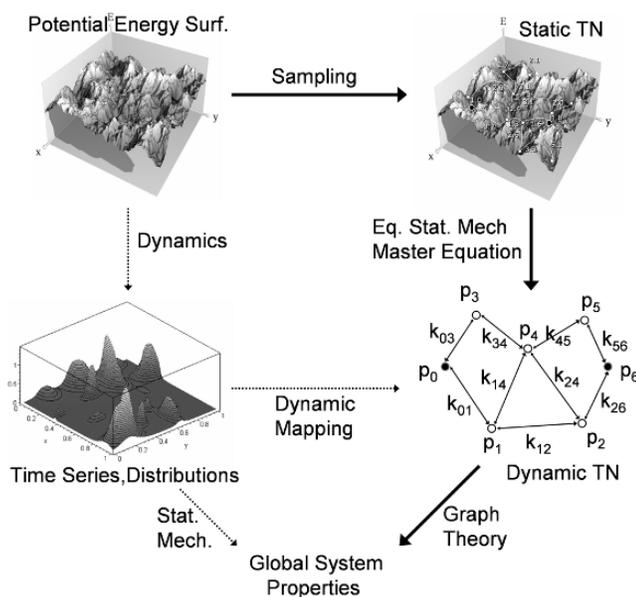
Fig. 11.1. Illustration of the relationship of the TN approach described in this chapter (solid arrows) to other molecular simulation approaches (dashed arrows).

11.2 Static Transition Networks

A static TN is formally defined by the weighted graph $\mathcal{G} = (\mathcal{V}, \mathbf{X}^S, \mathbf{E}^S, \mathcal{E}, \mathbf{X}^{TS}, \mathbf{E}^{TS})$. $\mathcal{V} = (1, \ldots, |\mathcal{V}|)$ is the list of vertices, representing the stable system states. $\mathbf{X}^S = (\mathbf{x}_1^S, \ldots, \mathbf{x}_{|\mathcal{V}|}^S)$ are the corresponding configuration vectors, and $\mathbf{E}^S = (E_1^S, \ldots, E_{|\mathcal{V}|}^S)$ are corresponding state energies. The list of edges, $\mathcal{E} = ((u, v), \ldots, (w, y))$, specifies between which pairs of vertices a direct transition is considered. $\mathbf{X}^{TS} = (\mathbf{x}_{uv}^{TS}, \ldots, \mathbf{x}_{wy}^{TS})$ are the configuration vectors of the corresponding transition states and $\mathbf{E}^{TS} = (E_{uv}^{TS}, \ldots, E_{wy}^{TS})$ are the associated transition state energies.

Each TN vertex, v, corresponds to a region R_v of the configurational space, containing a group of geometrically similar molecular configurations. What is appropriate as a definition of "group" depends on the application. For the present discussion, each given vertex v corresponds to an *attraction basin*, i.e., the set of configurations that can be mapped to the same local minimum $\mathbf{x}_v^S$ by a direct minimization [1,4]. Each vertex, v, is associated with a state energy E_v^S.

Each given edge, (u, v), representing the sub-transitions between a pair of neighboring vertices, is associated with the configuration of the transition state structure $\mathbf{x}_{uv}^{TS}$. $\mathbf{x}_{uv}^{TS}$ can, e.g., be taken as the rate-limiting first-order saddle point on the minimum energy path (MEP) between $\mathbf{x}_u^S$ and $\mathbf{x}_v^S$ (see Sect. 11.2.1). The energy E_{uv}^{TS} is the transition state energy associated with the edge, so that $\Delta E_{uv}^{TS} = E_{uv}^{TS} - E_u$ yields the energy barrier for the transition $u \to v$.
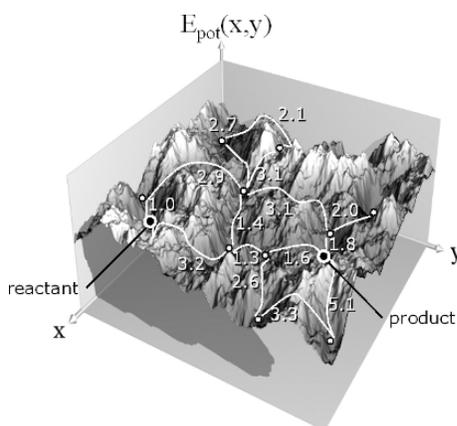
Fig. 11.2. Static TN on a schematic two-dimensional energy surface. The network vertices (white bullets) correspond to low-energy intermediates between the reactant and product end states of the transition (black bullets). The network edges (white lines) correspond to sub-transitions between the vertices and are associated with the rate-limiting barrier energies along the sub-transitions (white numbers).

Fig. 11.2 shows a schematic representation of a static TN.

Ideally, E_v^S corresponds to the free energy of region R, $E_v^S = \Delta G_v = G_v - G_0$, relative to some arbitrary reference energy G_0. The edge energy E_{uv}^{TS} should likewise correspond to the relative free energy of the transition state $E_{uv}^{TS} = \Delta G_{uv} = G_{uv} - G_0$. According to the first law of thermodynamics, free energy differences can be expressed as

$$\Delta G = \Delta E_{\text{pot}} + \Delta E_{\text{kin}} + \Delta(pV) - T\Delta S, \tag{11.1}$$

where E_{pot} is the potential energy, E_{kin} is the kinetic energy, p is the pressure, V is the volume, T is the temperature and S is the entropy. In liquid and solid systems and at low pressure, the pressure-volume product is nearly constant ($\Delta(pV) \approx 0$) [33]. Also, if the temperature and the number of particles are constant (canonical ensemble), $\Delta E_{\text{kin}} = 0$, on average. Thus,

$$\Delta G \approx \Delta E_{\text{pot}} - T\Delta S. \tag{11.2}$$

Accurate free energies are required for the static TN to derive reliable dynamic TN weights from these. Given current methodological and computational shortcomings, however, the calculation of reliable free energies is difficult, often even impossible. We therefore need to consider several levels of approximation:

1. Constant-entropy approximation: Entropic changes are assumed to be negligible ($\Delta S \approx 0$), and thus $\Delta G \approx \Delta E_{\text{pot}}$.
2. Harmonic approximation: A harmonic expansion around the minima and transition states is used to estimate free energy differences: $\Delta G \approx \Delta U - T\Delta S_{\text{harm}}$.
3. Free energy differences ΔG are computed.

11.2.1 Constant-Entropy Approximation

In the constant-entropy approximation, we assume that the regions of configurational space corresponding to the different vertices are of approximately similar size and shape. We furthermore assume the transitions between them to lead through narrow reaction channels such that the transition pathways are well defined. Following these assumptions, energy differences are dominated by enthalpic contributions, while the entropic contributions are comparatively small. Here, we set $\Delta S = 0$ in Eq. 11.2, which gives $\Delta G \approx \Delta E_{pot}$. The approximation is useful for systems with few degrees of freedom, having well-defined structures separated by high energy barriers, or which are studied at low temperatures. Biomolecules in physiological conditions (aqueous solvent, $T > 300$ K) do not in general satisfy these conditions. For the dynamics of biomolecules, this approximation disables any quantitative accuracy on the level of a dynamic TN. The benefit of this approximation, however, is that the associated theory is very mature and it is always feasible to obtain potential energy differences ΔE_{pot} even for very large and complex systems. Even if the constant-entropy approximation is used for the analyzed system (protein or peptide), the important free energy contributions from the bulk solvent can be accounted for by incorporating continuum solvent methods (such as Poisson–Boltzmann [34] or generalized Born [35–37]), so that $E_{pot}(\mathbf{x})$ actually becomes a mixed potential/free energy function.

Determination of the vertex energies E_u is relatively easy as it simply requires a local optimization of $E_{pot}(\mathbf{x}_u)$ starting from some initial point $\mathbf{x}_{u,0}$. The selection of these initial points is sometimes *not* trivial, depending on the nature of the molecular system and the process analyzed. Strategies to efficiently generate an ensemble of initial points for complex rearrangements in proteins are given in Ref. [38].

The edge energy E_{uv} is given the value of the rate-limiting saddle point of the reaction channel connecting u and v. For this, we define a pathway of "least effort," i.e., one that can be accessed with a minimum amount of energy. Such a *minimum energy path* (MEP) is a continuous path $\mathbf{z}(\lambda)$ connecting $\mathbf{x}_u$ and $\mathbf{x}_v$ ($\mathbf{z}(0) = \mathbf{x}_u$, $\mathbf{z}(1) = \mathbf{x}_v$, $\lambda \in [0, 1]$), satisfying the following criteria:

1. $\nabla E_{pot}(\mathbf{z}(\lambda))|_{\perp} = 0 \, \forall \, \lambda \in [0, 1]$, i.e., the gradient orthogonal to the path tangent is zero everywhere along the path.
2. $\mathbf{H}(\mathbf{z}(\lambda))|_{\perp}$ is positive definite $\forall \, \lambda \in [0, 1]$, i.e., the Hessian matrix at each path point, formulated in the subspace orthogonal to the path tangent, has only positive eigenvalues. Therefore, all path points have a minimum of potential energy in all directions except the path tangent.

All local energy maxima along the MEP are first-order saddle points on $E_{pot}(\mathbf{x})$ [39]. The highest-energy saddle point gives the transition state structure and the edge energy E_{uv}.

The MEP can be computed with the *conjugate peak refinement (CPR)* method [40]. In contrast to other MEP methods, such as *self-penalty walk* [41], *nudged elastic band (NEB)* [42] or the *string method* [43], CPR automatically finds all saddle points along the path to a desired accuracy. The algorithm does not evaluate second derivatives, but uses only the energy (which must be continuous) and its gradient.

11.2.2 Harmonic Approximation

In the vicinity of the stationary points (minima and saddle points), the energy surface can be approximated by a quadratic expansion. This assumes that the anharmonic portions of the energy surface (if any) are not sampled. The assumption is reasonable for local rearrangements in many solid state systems such as crystals, for most systems in the gas phase and for temperatures below the dynamical transition temperature that occurs in the range of 180 to 220 K [44]. For small peptides where all basins can be mapped, it has been shown that free energies calculated from harmonic approximations to the potential in each basin are able to reproduce some thermodynamic properties [25,26] even at higher temperatures, but a detailed evaluation remains to be performed. It is commonly accepted that proteins in aqueous solvent at physiological temperatures exhibit highly anharmonic behavior (Sect. 11.2.1) [45].

   The application of the harmonic approximation is straightforward. Given a stationary point (minimum or saddle point), the vibrational frequencies can be obtained as the positive eigenvalues of a normal-mode analysis [46]. As the main contribution to entropy is given by the lowest-frequency motions, a full diagonalization of the Hessian matrix is often not necessary, so that a harmonic analysis can be conducted for very large systems [47].

   Given the vibrational frequencies, or eigenvalues of the mass-weighted Hessian, v_i, corresponding to the eigenvector representing the vibrational motion i, the (classical) vibrational entropy can be computed [33] as

$$S_{\mathrm{harm}} \approx k_B \ln \prod_{i=1}^{D} \frac{k_B T}{h v_i}, \tag{11.3}$$

where k_B and h are the Boltzmann and Planck constants, respectively, and T is the temperature. The vertex energies E_u, in the harmonic approximation are given by substituting S_{harm} into Eq. (11.2):

$$E_u = E_{\mathrm{pot}}(\mathbf{x}_u) - k_B T \ln \prod_{i=1}^{D} \frac{k_B T}{h v_{u,i}}.$$

To obtain the edge energy we compute the harmonic expansion at the rate-limiting saddle point of the MEP (see Sect. 11.2.1), where only the $D-1$ positive eigenvalues are considered. We obtain

$$E_{uv} = E_{\mathrm{pot}}(\mathbf{x}_{ub}) - k_B T \ln \prod_{i=2}^{D} \frac{k_B T}{h v_{uv,i}}.$$

11.2.3 Free Energies

Both approximations given above are generally not quantitatively valid for biomolecules in aqueous solution at physiological temperature. The dynamics of these systems often involves considerable changes in entropy and is not restricted to the harmonic

regime near the energy minima. A rigorous treatment requires that the vertex energy E_u is determined as the free energy of the vertex region R_u, and the edge energy E_{uv} as the free energy of the boundary region between R_u and R_v, R_{uv}.

Free energy calculation methods, such as free energy perturbation or thermodynamic integration [30], attempt to compute free energy differences $\Delta G_{uv} = G_v - G_u$ between two thermodynamic states of the system by slowly changing one state into the other. Free energy barriers of transitions $\Delta G_{uv} = G_{uv} - G_u$ can be obtained with the umbrella sampling method [48,49]. As any vertex u can serve as a reference point with $G_u = 0$, the static TN energies can be determined from free energy differences.

Two practical problems exist with this approach. First, it must be assured that free energy calculations are confined to the regions R_u, R_v and to the reaction channel between them. Furthermore, free energy calculations on large systems such as proteins typically face the problem that convergence of the entropic contribution to free energy is very difficult to achieve. These problems are a major limitation in the applicability of TN at present.

11.3 Dynamic Transition Networks

Dynamic TNs describe the thermodynamics and the kinetics of the system. Formally, a dynamic TN is a weighted graph with time-dependent weights $\mathcal{G}_{\mathrm{dyn}}(t) = (\mathcal{V}, \mathbf{X}^S, \mathbf{p}(t), \mathcal{E}, \mathbf{X}^{TS}, \mathbf{K}(t))$. The vertex and edge list, $\mathcal{V}$ and $\mathcal{E}$, and the state and transition state configurations, $\mathbf{X}^S$ and $\mathbf{X}^{TS}$, have already been defined in Sect. 11.2. The list of vertex weights, $\mathbf{p}(t) = (p_0(t), \ldots, p_{|\mathcal{V}|}(t))$, is of size $|\mathcal{V}|$ and assigns to each vertex a probability of finding the system in this state at time t. The list of edge weights, $\mathbf{K}(t) = (K_{uv}(t), \ldots, K_{wy}(t))$ specifies a transition rate, or flux, i.e., the number of individual transitions per time unit, for each edge at time t.

Given that the individual sub-transitions are sufficiently simple (i.e., they are characterized by a single dominant barrier) and the energy barrier is significantly higher than the thermal energy, $k_B T$ (approximately 0.6 kcal/mol at $T = 300$ K), the rate $K_{uv}(t)$ can be expressed by a time-independent rate constant k_{uv} and the probability $p_u(t)$ [50]:

$$K_{uv}(t) = k_{uv} p_u(t). \tag{11.4}$$

11.3.1 Rate Constants k_{uv} Obtained with Transition State Theory

The rate constant, k_{uv}, captures the kinetic properties of the transition $u \rightarrow v$, such as the height and the form of the energy barrier. Here we summarize the main results from transition state theory (TST) which can be used to determine k_{uv}. TST applies when the energy barrier between states u and v is significantly higher than the thermal energy $k_B T$ and the motions orthogonal to the direction of the reaction $u \rightarrow v$ equilibrate much faster than the motion along this direction. A detailed treatment of TST is given in [51]. Using what is called the phenomenological form of the TST rate constant law, we obtain

$$k_{uv} = \frac{k_B T}{h} \exp\left(-\frac{E_{uv}^{TS} - E_u^S}{k_B T}\right). \tag{11.5}$$

11.3.2 Time Evolution of the Dynamic TN $\mathcal{G}_{\mathbf{dyn}}(t)$

When all rate constants k_{uv} are given and an initial vertex probability distribution, $p(t = 0)$, is specified, the time evolution of the dynamic TN $\mathcal{G}_{\mathrm{dyn}}(t)$ can be specified for all times t by using a dynamic model which propagates the vertex probabilities over time. A common way to do this is to employ the master equation [52]. The master equation is a loss-gain equation which describes the time evolution for the probability $p_u(t)$ of finding the system in state u:

$$\frac{dp_u(t)}{dt} = \sum_v [k_{vu} p_v - k_{uv} p_u], \tag{11.6}$$

where the sum runs over all vertices v which are connected with vertex u. By integrating Eq. (11.6) for all vertices over time, one obtains the kinetics of relaxing the vertex probabilities from the initial distribution towards the equilibrium distribution. The master equation has been used to obtain the kinetics in various reaction and folding studies [6, 13, 16, 19–23, 25, 26, 28, 29].

As illustrated in Fig. 11.1, the dynamic TN $\mathcal{G}_{\mathrm{dyn}}(t)$ can also be obtained without any static TN. In Ref. [32], long-time molecular dynamics trajectories of a 20-residue peptide were mapped on a predefined lattice of backbone dihedral angle ranges. A TN vertex, u, was defined for each lattice cell that was visited during the dynamics and obtained a corresponding statistical weight, p_u. A TN edge, (u, v), was defined if the transition $u \to v$ or $u \to v$ occurred during the dynamical trajectories. Another approach based on molecular dynamics trajectories is followed in Ref. [31], where a hidden Markov model is generated from the trajectory. The model represents a discrete set of metastable system states along with a transition matrix which yields the pairwise transition probabilities between them. The route followed in both [31, 32] has the advantage that an appropriate definition of system states and the connectivity that is dynamically relevant can be identified from the dynamics. The difficulty with this method is that an accurate TN can only be generated from a very long dynamical trajectory, in which both the state probabilities and the transition probabilities have converged. This is often difficult to obtain for large systems.

11.3.3 The Equilibrium Transition Network $\mathcal{G}_{\mathbf{dyn}}^{\mathbf{eq}}$

Of particular interest is the equilibrium dynamic TN, $\mathcal{G}_{\mathrm{dyn}}^{\mathrm{eq}} = \mathcal{G}_{\mathrm{dyn}}(t \to \infty)$, or the stationary case. In equilibrium, both the vertex and edge weights are stable (they do not change with time), and the edge weights are symmetric: $K_{uv} = K_{vu}$ for all edges (u, v). For this case, the dynamic TN weights, p_u^{eq} and K_{uv}^{eq}, can be directly computed from the static TN energies.

The equilibrium probability, p_u^{eq}, equals the fraction of the partition function associated with u [33]:

$$p_u^{\text{eq}} = \frac{Z_u}{Z} = \frac{\exp\left(-\frac{E_u}{k_B T}\right)}{\sum_{v \in \mathcal{V}} \exp\left(-\frac{E_v}{k_B T}\right)}. \tag{11.7}$$

The equilibrium rate, K_{uv}^{eq}, can be computed by combining Eq. (11.5) with Eqs. (11.7) and (11.4):

$$K_{uv}^{\text{eq}} = \frac{k_B T}{hZ} \exp\left(\frac{-E_u}{k_B T}\right) \exp\left(-\frac{E_{uv} - E_u}{k_B T}\right) = \frac{k_B T}{hZ} \exp\left(-\frac{E_{uv}}{k_B T}\right). \tag{11.8}$$

The quality of both p_u^{eq} and K_{uv}^{eq} depends on which level of accuracy is used to compute the static TN energies (Sect. 11.2).

11.4 Transition Network Analysis

11.4.1 Topographic Information via Disconnectivity Graphs

A TN contains information on the topography of the energy surface. A common way to visualize this information is the disconnectivity graph [12, 16] or decoy tree [53]. Fig. 11.3 shows some archetypal energy landscapes with their corresponding disconnectivity graphs. Each node in the graph represents a super-basin of the energy surface, defined as the set of local minima connected by barriers not higher than the energy level at which the node is situated. Thus, the single node at the top of the graph represents the full energy surface and gives the energy at which all its individual basins are connected. At lower energy levels, the graph splits into more local basins. The lines stretching downward from the nodes extend to the energy of individual conformations. The horizontal dimension is not quantitative (it does not correspond to any conformational coordinate) and is chosen so as to yield a good visualization of the connectivity.

A system which rapidly and reliably relaxes into a well-defined minimum (such as fast-folding peptides or proteins) is characterized by a steep and smooth funnel on
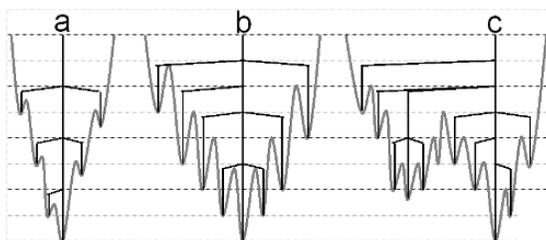
Fig. 11.3. Archetypal energy surfaces (gray) and the corresponding disconnectivity graphs (black). The energy increases from below to above. (a) A steep folding funnel with a well-defined energy minimum. (b) A less steep folding funnel with considerable roughness on the energy surface, leading to slower folding times. (c) An energy surface with two major competing energy minima.

the energy surface and a narrow and deep disconnectivity graph (Fig. 11.3a). A slow-down of the relaxation can be caused by a widening of the funnel, which is shown by a larger amount of side branches in the disconnectivity graph, or by an increased amount of roughness on the energy surface, shown by deeper side branches (Fig. 11.3b). As shown in Refs. [17, 20], peptides with conformational constraints like charged ter-mini and cyclic closure of the backbone have energy surfaces with multiple competing basins (Fig. 11.3c). Such a topography is also expected for energy surfaces of proteins undergoing conformational change between metastable states, such as the quaternary rearrangement of the hemoglobin tetramer upon oxygen binding [54], the lever-arm motion in myosin during muscle contraction [55] and the molecular switch in Ras p21 that signals cell division [56–58].

In a disconnectivity graph, the lowest-energy node which connects a pair of confor-mations gives an upper bound to the rate-limiting energy barriers of a transition path-way between these conformations. The full information on the lowest-energy transition pathways, however, can be obtained by a graph-theoretical analysis of the TN.

11.4.2 Best Transition Paths via Graph Theory

In order to employ graph-theoretical algorithms to compute best paths, we first need edge costs which are additive. In contrast to energies or rates, the mean passage time τ is an additive quantity. For a given transition, the mean passage time is simply the inverse of the rate:

$$\tau = K^{-1}. \tag{11.9}$$

Using the equilibrium rate law in the form of Eq. 11.8, and setting the constant hZ/k_BT to unity, we can define the edge costs c_{uv} as being equal to the inverse Boltz-mann weight of the edge energies:

$$c_{uv} = \exp\left(\frac{E_{uv}}{k_BT}\right). \tag{11.10}$$

The best path connecting vertices $v_1 = v_R$ and $v_m = v_P$ *via* a series of m vertices, $P = (v_1, v_2, \ldots, v_m)$, traveling over edges $((v_1, v_2), \ldots, (v_{m-1}, v_m))$, is defined here as that which minimizes the cumulative edge costs

$$C(P) = \sum_{k=1}^{m-1} c_{v_k v_{k+1}}. \tag{11.11}$$

This definition of a best path is similar to the notion of the path with the "maximum flux" or "minimum resistance," as given in Refs. [59, 60]. As for the exponential de-pendence of the cost on the energy (see Eq. (11.10)), the best path tends to be one that minimizes the highest energy barrier along it. To determine the best path in practice, the edge energy vector $\mathbf{E}^{TS}$ is transformed into a cost vector $\mathbf{c}$ using Eq. (11.10). $\mathbf{c}$ has size $|\mathcal{E}|$ and assigns a cost c_{uv} to each edge (u, v) in $\mathcal{E}$. Subsequently, the Dijkstra algorithm [61] is used to identify a best path between the two end states through the

weighted network defined by $(\mathcal{V}, \mathcal{E}, \mathbf{c})$. This path minimizes the path cost $C(P)$ given in Eq. 11.11.

This best path furnishes a preliminary understanding of the transition [58], and it may be used as a guess for a reaction coordinate for free energy calculations [45], or as a starting point for discrete path sampling [62]. However, it dominates the transition only if the barriers of alternative pathways are considerably higher. To obtain an idea of the multitude of accessible pathways, it is useful to determine the set of k different pathways, $(P_1, P_2, \dots, P_k)$, with the best (*i.e.*, lowest) costs ($C_1 \leq C_2 \leq \dots \leq C_k$) [63]. Best transition pathways in this sense have been computed for transitions in *n*-butane [64], for transitions between the α_L, β and α_R states in polyalanine [65] and for the molecular switch in the Ras p21 protein [38].

11.4.3 Topological Properties via Complex Network Analysis

As a last point, we address the statistical analysis of the TN topology, i.e., the way its conformational sub-states are connected. Although seemingly obvious, this kind of information is of particular interest as it allows the TN to be put into the more general context of the field of complex networks network analysis, in which there has been a surge of interest during recent years [66].

Of particular interest here is the vertex degree $\rho(u)$, i.e., the number of neighbors for vertex u, as well as the degree distribution of the network, $p(\rho)$, which gives the probability of finding a vertex with ρ neighbors. The degree distribution is interesting because networks can be characterized by the functional form of this distribution. While random networks [67] are characterized by a Poisson degree distribution, many real-world networks have a scale-free degree distribution [66,68–71]. A distribution is called scale-free if it follows a power law, i.e.,

$$p(\rho) \propto \rho^{-\gamma},$$

which decays much slower than the Poisson distribution (double exponential decay beyond the maximum) and also slower than a single exponential decay. This means that there is a relatively high chance to find several *hubs*, *i.e.*, highly connected vertices in the network. TNs for both atomic clusters [72] and a peptide [32] have been shown to have a scale-free degree distribution. In both cases it turned out that the lowest-energy structures tend to correspond to vertices with a high degree value. It may thus be a general property of molecular energy surfaces that the low energy minima are densely connected to neighboring conformational sub-states.

Another interesting network property which seems to be almost ubiquitous is the small-world property [73]. This property is related to the average path length, i.e., the length of the shortest path between a pair of vertices, averaged over all pairs of vertices in the network. Originally, small-world behavior was defined such that on growing the network by adding new vertices to it, the average path length grows as $\propto \ln(|\mathcal{V}|)$ [74]. Since then, the term was used in a more qualitative sense, meaning that even for a very large network, any pair of vertices can be connected by a small number of elementary transitions. Both TNs for atomic clusters [72] and a peptide

[32] have been shown to have the small-world property. Unlike the original model of small-world networks [73], this property cannot be explained by the presence of edges of arbitrary length, because in TNs, direct transitions can only occur between conformations which are neighbors in conformational space. The small-world property of TN may thus be a result of the high dimensionality of conformational space and the presence of a few highly connected conformations which can be accessed from many vertices in the network. However, a satisfactory answer to this question remains to be given.

11.5 Applications

For small reaction systems it has long been common to compute TNs between minima obtained from initial structures modeled through "chemical intuition." Another class of systems where TNs have been widely used is that of atom clusters, see, e.g., [14, 15,27,75,76] and references therein. In clusters, the number of stationary points may be immense, even for a relatively small number of particles. This is due to the lack of constraints (e.g., covalent bonds) which would limit the ability to recombine and thus limit combinatorial explosion. Nevertheless, clusters relax to their global minimum at low temperatures relatively fast because this global minimum is well connected with the rest of the network and can be reached from any other minimum by a short sequence of barriers [72]. For the same reason, it is feasible to identify the global minimum in simulations using a network-based search.

The generation of TNs for polypeptides is arguably more demanding because of the heterogeneous structure of these systems. Consequently, the different applications on biomolecules exhibit a wide spectrum of simulation techniques used to generate the different conformations which serve as TN vertices. The first comprehensive TN of a biomolecule, a tetraalanine, was presented in Ref. [12]. Here, minimum energy paths were used to identify saddle points and intermediate minima between pairs of already-available minima until no new minima could be identified. In later studies of the same system [13,16], its relaxation dynamics was studied using master equations, and grouping of minima into fast-equilibrating clusters was considered. The effects of conformational constraints on the energy landscape and folding kinetics of a peptide were studied in detail [17,21,77]. There, the TN vertices were generated by minimizing structures from high-temperature molecular dynamics simulations. Refs. [19,23] present very large TNs of polyalanines which were generated by an iterative search following the eigenvalues starting from a set of initially modeled minima (EV following). These studies focused on a network-based minimization of the systems under consideration and the topography of the energy surface depending on the solvent environment.

Recent publications have reported on free energy TNs. In [26] discrete path sampling (DPS) was used to identify minima and saddle points for Met-enkephalin and their free energies were estimated using a harmonic expansion around the stationary points. A similar study of the much larger GB1 hairpin system was reported in [28]. A similar-sized system, the second β-hairpin of protein G, was studied in Ref. [78]

Table 11.1. Comparison of some landmark TN calculations.

System	Atoms	Sampling	Ener[a]	$|\mathcal{V}|$[b]	$|\mathcal{E}|$[c]	Year	Refs
IAN	26	Reaction paths	pot	138	490	1986	[12,13]
LJ cluster	19	MD	pot	299	461	1995	[6]
Ala$_6$	64	MD	pot	280	≈ 20000	1998	[17,20,21]
Model protein	46	EV-following	pot	500	805	1999	[18]
Morse cluster	14	EV-following	pot	12760	54439	1999	[75]
Ala$_{16}$	≈ 170	EV-following	pot	≈ 10000	≈ 15000	2002	[23]
MET-enkephalin	≈ 50	DPS	free	4100	8186	2003	[26]
GB1 hairpin	≈ 250	DPS	free	n/a	n/a	2004	[28]
Protein G β hairpin	≈ 250	MD	free	35377	83331	2004	[78]
Ras p21	≈ 1800	Conformational	pot	6242	47404	2006	[38]

[a]Type of energies given for the network vertices and edges (potential or free).

[b]Number of vertices.

[c]Number of edges.

by a free energy TN generated from microsecond MD simulations. A disconnectivity graph of the TN helped to identify multiple non-native folding funnels which were not visible in previous energy surface representations for the same system.

In [18], the first TN of a model protein was presented, using eigenvector following to identify the stationary points. The study examined the topology of the energy surface and identified possible sources of frustration in the folding funnel. The TN for the largest system considered yet was computed for the Ras p21 protein [38]. This network was used to identify the mechanism for the conformational switch of the protein.

Table 11.1 compares some major works involving TN calculations.

11.6 Conclusions and Outlook

The concept of transition networks (TNs), which has attracted considerable interest in recent years, sheds new light on some traditional problems of molecular simulation. One particularly attractive property of TNs is that they form a platform that establishes a communication between physico-chemical theories on the thermodynamics and kinetics of molecules on one side and computational disciplines such as graph theory and complex network theory on the other.

The TN methodology has the potential to be applied to complex structural changes in many peptides and proteins whose functional timescale and complexity precludes the use of direct simulation. Pioneering works in this direction are the folding of the GB1 hairpin [28] and the TN for the molecular switch in Ras p21 [38].

At present, a major limitation for TNs of such large systems is that their analysis is qualitative rather than quantitative in nature. This is because the static TN is based on either potential energies or free energies obtained through a harmonic approximation (which is likely to be poor if only a subset of the energy basins can be captured in

the TN). The reliable computation of free energies for sub-states of complex systems such as proteins is an open problem of great interest in the physical chemistry community [30]. Here, the combination of free-energy calculation methods from statistical mechanics with the TN concept may help to obtain accurate free energies for a small number of interesting sub-states that can be defined and selected with the help of the TN.

Finally, TNs may shed new light on the particular physical properties underlying the dynamics of complex systems such as peptides or proteins. A concept that has received considerable attention is that of the essential subspace of a biomolecule [79]. The essential subspace is the part of state space that is accessible to the molecule at a given temperature. Available knowledge on the form, size and connectivity of that space has been obtained either indirectly from interpreting experimental data, such as relaxation times [80], or by analysis of molecular dynamics simulations [81], the latter being limited by the sampling problem. TN approaches have the potential to fully explore the essential subspace, so the analysis of TN is promising to answer exciting questions concerning the form, topology and kinetics of the essential subspace. These answers will help us not only to understand the physics of proteins but also to enhance our ability to model and predict complex system dynamics in general.

References

1. Stillinger, F.H., Weber, T.H.: Hidden structure in liquids. Phys. Rev. A, **25**, 978–989 (1982).
2. Elber, R., Karplus, M.: Multiple conformational states of proteins: A molecular dynamics study of myoglobin. Science, **235**, 318–321 (1987).
3. Miller, M.A., Doye, J.P.K., Wales, D.J.: Energy landscapes of model polyalanines. J. Chem. Phys., **117**, 1363–1376 (2002).
4. Stillinger, F.H., Weber, T.A.: Packing structures and transitions in liquids and solids. Science, **228**, 983–989 (1984).
5. Stillinger, F.H.: A topographic view of supercooled liquids and glass formation. Science, **267**, 1935–1939 (1995).
6. Berry, R.S., Breitengraser-Kunz, R.: Topography and dynamics of multidimensional interatomic potential surfaces. Phys. Rev. Lett., **74**, 3951–3954 (1995).
7. Brooks III, C.L., Onuchic, J.N., Wales, D.J.: Taking a walk on a landscape. Science, **293**, 612–613, (2001).
8. Schütte, C., Huisinga, W.: Biomolecular conformations can be identified as metastable sets of molecular dynamics. In: Ciaret, P.G., Lions, J.L. (eds) Handbook of Numerical Analysis. Computational Chemistry, 699–744, North-Holland, Amsterdam, (2003).
9. Schlick, T., Barth, E., Mandziuk, M.: Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation Annu. Rev. Biophys. Biomol. Struct., bf 26, 181–222 (1997).
10. Frenkel, D.: Introduction to monte carlo methods. In: Attig, N., Binder, K., Grubmüller, H., Kremer, K. (eds) Computational Soft Matter: From Synthetic Polymers to Proteins. **23**, 29–59, NIC Series, Jülich, Germany, (2004).
11. Schlick, T., Skeel, R.D., Brunger, A.T., Kalé, L.V., Board Jr., J.A., Hermans, J., Schulten, K.: Algorithmic challenges in computational molecular biophysics. J. Comp. Phys., **151**, 9–48 (1999).

12. Czerminski, R., Elber, R.: Reaction path study of conformational transitions and helix formation in a tetrapeptide. Proc. Nat. Acad. Sci. USA, **86**, 6963–6967 (1989).
13. Czerminski, R., Elber, R.: Reaction path study of conformational transitions in flexible systems: Application to peptides. J. Chem. Phys., **92**, 5580–5601 (1990).
14. Wales, D.J.: Structure, dynamics, and thermodynamics of clusters: Tales from topographic potential surfaces. Science, **271**, 925–933 (1996).
15. Ball, K.D., Berry, R.S., Kunz, R.E., Li, F.Y., Proykova, A., Wales, D.J.: From topographies to dynamics on multidimensional potential energy surfaces of atomic clusters. Science, **271**, 963–967 (1996).
16. Becker, O.M., Karplus, M.: The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. J. Chem. Phys., **106**, 1495–1517 (1996).
17. Levy, Y., Becker, O.M.: Effect of conformational constraints on the topography of complex potential energy surfaces. Phys. Rev. Lett., **81**, 1126–1132 (1998).
18. Miller, M.A., Wales, D.J.: Energy landscape of a model protein. J. Chem. Phys., **111**, 6610–6616 (1999).
19. Mortenson, P.N., Wales, D.J.: Energy landscapes, global optimization and dynamics of the polyalanine Ac(ala)$_8$NHMe. Journal of Chemical Physics, **114**, 6443–6453 (2001).
20. Levy, Y., Becker, O.M.: Energy landscapes of conformationally constrained peptides. J. Chem. Phys., **114**, 993–1009 (2001).
21. Levy, Y., Jortner, J., Becker, O.M.: Dynamics of hierarchical folding on energy landscapes of hexapeptides. J. Chem. Phys., **115**, 10533–10547 (2001).
22. Levy, Y., Jortner, J., Becker, O.M.: Solvent effects on the energy landscapes and folding kinetics of polyalanines. Proc. Nat. Acad. Sci. USA, **98**, 2188–2193 (2001).
23. Mortenson, P.N., Evans, D.A., Wales, D.J.: Energy landscapes of model polyalanines. J. Chem. Phys., **117**, 1363–1376 (2002).
24. Levy, Y., Becker, O.M.: Conformational polymorphism of wild-type and mutant prion proteins: energy landscape analysis. Proteins, **47**, 458–468 (2002).
25. Evans, D.A., Wales, D.J.: Free energy landscapes of model peptides and proteins. J. Chem. Phys., **118**, 3891–3897 (2003).
26. Evans, D.A., Wales, D.J.: The free energy landscape and dynamics of met-enkephalin. J. Chem. Phys., **119**, 9947–9955 (2003).
27. Wales, D.J., Doye, J.P.K.: Stationary points and dynamics in high-dimensional systems. J. Chem. Phys., **119**, 12409–12416 (2003).
28. Evans, D.A., Wales, D.J.: Folding of the gb1 hairpin peptide from discrete path sampling. J. Chem. Phys., **121**, 1080–1090 (2004).
29. Despa, F., Wales, D.J., Berry, R.S.: Archetypal energy landscapes: Dynamical diagnosis. J. Chem. Phys., **122**, 024103 (2005).
30. Kollmann, P.: Free energy calculations: Applications to chemical and biochemical phenomena. Chem. Rev., **93**, 2395–2417 (1993).
31. Horenko, I., Dittmer, E., Fischer, A., Schütte, C.: Automated model reduction for complex systems exhibiting metastability. SIAM Mult. Model. Sim., **5**, 802–827 (2006).
32. Rao, F., Caflisch, A.: The protein folding network. J. Mol. Bio., **342**, 299–306 (2004).
33. Moore, W.J., Hummel, D.O.: Physikalische Chemie, 3rd ed. Walter de Gruyter, Berlin, (1983).
34. Im, W., Beglow, D., Roux, B.: Continuum solvation model: Electrostatic forces from numerical solutions to the Poisson–Boltzmann equation. Comp. Phys. Comm., **111**, 59–75 (1998).
35. Schaefer, M., Karplus, M.: A comprehensive analytical treatment of continuum electrostatics. J. Chem. Phys., **100**, 1578–1599 (1996).

36. Bashford, D., Case, D.A.: Generalized Born models of macromolecular solvation effects. Ann. Rev. Phys. Chem., **51**, 129–152 (2000).
37. Im, W., Lee, M.S., Brooks III, C.L.: Generalized Born model with a simple smoothing function. J. Comp. Chem., **24**, 1691–1702 (2003).
38. Noé, F., Krachtus, D., Smith, C.J., Fischer, S.: Transition networks for the comprehensive characterization of complex conformational change in proteins. J. Chem. Theo. Comput., **2**, 840–857 (2006).
39. Henkelman, G., Jóhannesson, G., Jónsson, H.: Progress on Theoretical Chemistry and Physics, ch. Methods for Finding Saddle Points and Minimum Energy Paths, 269–300, Kluwer Academic Publishers, Dordrecht, (2000).
40. Fischer, S., Karplus, M.: Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. Chem. Phys. Lett., **194**, 252–261 (1992).
41. Czerminski, R., Elber, R.: Self avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. Int. J. Quant. Chem., **24**, 167 (1990).
42. Jónsson, H., Mills, G., Jacobsen, K.W.: Classical and Quantum Dynamics in Condensed Phase Simulations, ch. Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions, 385–404, World Scientific, Singapore, (1998).
43. Weinan, E., Ren, W., Vanden-Eijnden, E.: String method for the study of rare events. Phys. Rev. B, **66**, 052301 (2002).
44. Tournier, A., Smith, J.: Principal components of the protein dynamical transition. Phys Rev. Lett., **91**, 208106 (2003).
45. Elber, R.: Recent Developments in Theoretical Studies of Proteins, ch. Reaction Path Studies of Biological Molecules, World Scientific, Singapore (1996).
46. Wilson, E., Decius, J., Cross, P.: Molecular Vibrations, McGraw-Hill, New York (1955).
47. Perahia, D., Mouawad, L.: Computation of low-frequency normal modes in macromolecules: Improvements to the method of diagonalization in a mixed basis and application to hemoglobin. Comput. Chem., **19**, 241–246 (1995).
48. Torrie, G.M., Valleau, J.P.: Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J. Comp. Phys., **23**, 187–199 (1977).
49. Rajamani, R., Naidoo, K.J., Gao, J.: Implementation of an adaptive umbrella sampling method for the calculation of multidimensional potential of mean force of chemical reactions in solution. Proteins, **24**, 1775–1781 (2003).
50. Karplus, M.: Aspects of protein reaction dynamics: Deviations from simple behavior. J. Phys. Chem. B, **104**, 11–27 (2000).
51. Hänggi, P., Talkner, P., Borkovec, M.: Reaction rate theory: Fifty years after kramers. Rev. Mod. Phys., **62**, 251–342 (1990).
52. van Kampen, N.G.: Stochastic Processes in Physics and Chemistry, North-Holland, Amsterdam, (1992).
53. Herges, T., Wenzel, W.: Free-energy landscape of the villin headpiece in an all-atom force field. Structure, **13**, 1–8 (2005).
54. Olsen, K., Fischer, S., Karplus, M.: A continuous path for the $T \rightarrow R$ allosteric transition of hemoglobin. Biophys. J., **78**, 394A (2000).
55. Fischer, S., Windshuegel, B., Horak, D., Holmes, K.C., Smith, J.C.: Structural mechanism of the recovery stroke in the myosin molecular motor. Proc. Nat. Acad. Sci. USA, **102**, 6873–6878 (2005).
56. Coleman, M.L., Marshall, C.J., Olson, M.F.: Ras and Rho GTPases in G1-phase cell-cycle regulation. Nat. Rev. Mol. Cell Bio., **62**, 851–891, (1993).
57. Vojtek, A.B., Der, C.J.: Increasing complexity of the Ras signaling pathway. J. Biol. Chem., **32**, 19925–19928 (1998).

58. Noé, F., Ille, F., Smith, J.C., Fischer, S.: Automated computation of low-energy pathways for complex rearrangements in proteins: Application to the conformational switch of ras p21. Proteins, **59**, 534–544 (2005).

59. Berkowitz, M., Morgan, J.D., McCammon, J.A., Northrup, S.H.: Diffusion-controlled reactions: A variational formula for the optimum reaction coordinate. J. Chem. Phys., **79**, 5563–5565 (1983).

60. Huo, S., Straub, J.E.: The maxflux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body systems at finite temperature. J. Chem. Phys., **107**, 5000–5006 (1997).

61. Dijkstra, E.: A note on two problems in connexion with graphs. Num. Math., **1**, 269–271 (1959).

62. Wales, D.J.: Discrete path sampling. Mol. Phys., **100**, 3285–3305 (2002).

63. Eppstein, D.: Finding the k shortest paths. Proc. 35th Symp. Found. Comp. Sci., 154–165 (1994).

64. Preis, R., Dellnitz, M., Hessel, M., Schütte, C., Meerbach, E.: Dominant paths between almost invariant sets of dynamical systems. Preprint 154 (2004).

65. Noé, F., Oswald, M., Reinelt, G., Fischer, S., Smith, J.C.: Computing best transition pathways in high-dimensional dynamical systems. SIAM Mult. Model. Sim., **5**, 393–419 (2006).

66. Newman, M.E.J.: The Structure and function of complex networks. SIAM Review, **45**, 167–256 (2003).

67. Karoński, M., Ruciński, A.: The origins of the theory of random graphs. In: Graham, R.L., Nešetřil, J. (eds.) The mathematics of Paul Erdös, vol. 13 of *Algorithms and Combinatorics*, 311–336, Springer, Berlin, (1997).

68. Albert, R., Barabasi, A.L., Jeong, H.: Diameter of the world wide web. Nature, **401**, 130–131 (1999).

69. Faloutsos, P.F.M, Faloutsos, C.: On power-law relationships of the internet topology. Comp. Commun. Rev., **29**, 251–262 (1999).

70. Redner, S.: How popular is your paper? an empirical study of the citation distribution. Eur. Phys. J. B, **4**, 131–134 (1998).

71. Jeong, H., Tomber, B., Albert, R., Oltvai, Z.N., Barabasi, A.-L.: The large-scale organization of metabolic networks. Nature, **407**, 651 (2000).

72. Doye, J.P.K.: Network topology of a potential energy landscape: A static scale-free network. Phys. Rev. Lett., **88**, 238701 (2002).

73. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'smallworld' networks. Nature, **393**, 440–442 (1998).

74. Strogatz, S.: Exploring complex networks. Nature, **410**, 268–276 (2001).

75. Miller, M.A., Doye, J.P.K., Wales, D.J.: Structural relaxation in morse clusters: Energy landscapes. J. Chem. Phys., **110**, 328–334 (1999).

76. Ball, K.D., Berry, R.S.: Dynamics on statistical samples of potential energy surfaces. J. Chem. Phys., **111**, 2060–2070 (1999).

77. Levy, Y., Jortner, J., Becker, O.M.: Dynamics of hierarchical folding on energy landscapes of hexapeptides. J. Chem. Phys., **115**, 10533–10547 (2001).

78. Krivov, S.V., Karplus, M.: Hidden complexity of free energy surfaces for peptide (protein) folding. Proc. Nat. Acad. Sci. USA, **101**, 14766–14770 (2004).

79. Amadei, A., Linssen, A.B., Berendsen, H.J.C.: Essential dynamics of proteins. Proteins, **17**, 412–225 (1993).

80. Frauenfelder, H., Sligar, S.G., Wolynes, P.G.: The energy landscapes and motions of proteins.. Science, **254**, 1598–1603 (1991).

81. Sullivan, D.C., Kuntz, I.D.: Conformation spaces of proteins. Proteins, **42**, 495–511 (2001).

Development

Pigmentation Pattern Formation in Butterfly Wings: Global Patterns on Fore- and Hindwing

Toshio Sekimura,[1] Anotida Madzvamuse,[2] and Philip K. Maini[3]

[1] Department of Biological Chemistry, College of Bioscience and Biotechnology, Chubu University, Kasugai, Aichi, 487-8501, Japan; sekimura@isc.chubu.ac.jp

[2] Department of Mathematics, University of Sussex, Mantell Building, Brighton, BN1 9RF, England, UK; a.madzvamuse@sussex.ac.uk

[3] Centre for Mathematical Biology, Mathematical Institute, University of Oxford, 24-29 St Giles, Oxford OX1 3LB, UK; maini@math.ox.ac.uk

Summary. Pigmentation patterns in butterfly wings are one of the most spectacular and vivid examples of pattern formation in biology. In this chapter, we devote our attention to the mechanisms for generating global patterns. We focus on the relationship between pattern forming mechanisms for the fore- and hindwing patterns. Through mathematical modeling and computational analysis of *Papilio dardanus* and *polytes*, our results indicate that the patterns formed on the forewing need not correlate to those of the hindwing in the sense that the formation mechanism is the same for both patterns. The independence of pattern formation mechanisms means that the coordination of unified patterns of fore- and hindwings is accidental. This is remarkable, because from Oudemans's principle [10], patterns appearing on the exposed surface of fore- and hindwing at the natural resting position are often integrated to form a composite and unified adaptive pattern with their surrounding environment.

Key words: Color pattern formation, global patterns, butterfly wing, *Papilio dardanus*, *Papilio polytes*, reaction-diffusion, Gierer–Meinhardt.

12.1 Introduction

One of the most striking phenomena about wing color patterns is the close match between patterns of fore- and hindwing when wings are held at their resting position. This phenomenon is known as the Oudemans principle [10]. The overall unified pattern of both wings is usually explained through their functional significance, i.e., the adaptation to the environment [8]. We have investigated color pattern formation of butterfly wings from a mathematical modeling point of view [7,9,14]. In this paper, using *Papilio dardanus* and *polytes* as examples, we discuss the close match between fore- and hindwing patterns from a mathematical modeling point of view. In the following sections, we discuss butterfly wing color patterns focusing on: (i) current research activities on wing color patterns, (ii) ground plan of global patterns, (iii) global pattern of fore- and hindwing, (iv) *Papilio dardanus* and *polytes* as model butterflies, and

Fig. 12.1. Local and global patterns: local eyespot patterns of the forewing of *Mycalesis gotama* (left) and global patterns of fore- and hindwings of *Luehdorfia japonica* (right).

(v) mathematical modeling and numerical simulations on geometrically accurate wing shapes. In Section 12.5 we conclude with a comparison of the pattern on the hindwing with the forewing pattern.

12.2 Global and Local Patterns

Butterfly wings are composed of two monolayers (upper and lower) of epidermal cells which are separated by an extracellular space. After pupation, a fraction of the epidermal cells on each wing surface differentiates into scale cells, which then arrange their positions to form parallel rows in the proximal-distal direction of the wing [12, 13]. The colors on wings are due to the colors of scale cells that cover the entire wing surface. Currently, there exist two different research directions on color pattern formation (Fig. 12.1). The first one is that of localized patterns such as eyespot patterns. The second one is of global patterns which cover the whole dorsal or ventral wing monolayer. The best-understood mechanism of color pattern formation is that of local eyespot patterns in which the spatial patterns of expressions of the gene *Distal-less* and several other genes have been detected and examined [1, 2, 4, 5]. Little is known about genes for global patterns except for a few cases such as the butterfly *Papilio dardanus* [3]. Our mathematical model is a reaction-diffusion model for global pattern formation, which is analyzed mathematically and simulated computationally. Our research interest in this chapter is concerned with global patterns of the fore- and hindwing which we discuss below.

12.2.1 Ground Plan of Global Pattern

Global pigmentation patterns of butterfly wings are very complicated in structure and are sometimes used for identification of butterfly species. When looking at global wing patterns it is difficult sometimes to understand the rules governing diverse wing patterns. However, because of the pioneering work of Schwanwitsch [11] and Süffert [16] on the nymphalid ground plan, the complicated patterns on the wing can be understood

as a composite of a small number of pattern elements. The ground plan is not really a pattern existing in nature, but a hypothetical one from which a large number of real wing patterns on the Nymphalidae butterflies could be generated by some organizing principles such as dislocation of pattern elements along the veins [8].

12.2.2 Global Patterns of Fore- and Hindwings

A close match between global patterns on fore- and hindwings when these are held at rest occurs on the dorsal sides of many species. It is usually assumed that dorsal patterns have evolved independently from ventral patterns in the vast majority of species and are adapted for a wide variety of functions. Oudemans pointed out that when a butterfly settles to assume its natural resting position, patterns appearing on the exposed surface of the fore- and hindwings, head, thorax, abdomen, and even the legs are very often integrated to form a composite but unified pattern. This phenomenon is generally known as Oudemans's principle [10]. The integrated pattern of both the fore- and hindwing is often suggested as evidence of a unified adaptive pattern just like we see in *Kallima inachus* [8]. However, at the same time we find a mismatch between patterns of fore- and hindwing on the ventral wing surface. For example, the integrated pattern of living Lepidoptera fitted to Oudemans's principle seems to break this match when insects feed [15]. We are thus faced with the dilemma of having to interpret the adaptive significance of the consistency and diversity of these patterns.

12.3 *Papilio dardanus* and *Papilio polytes*

In this chapter, we choose the Papilionidae butterflies *Papilio dardanus* and *Papilio polytes* as model butterflies for mathematical analysis and computer simulation. We briefly describe these two butterflies in the following subsections.

12.3.1 *Papilio dardanus*

A species of *Papilio dardanus* is widely distributed across sub-Saharan Africa and is well known for the spectacular phenotypic polymorphism in females. The females have evolved more than a dozen different wing color patterns of which several mimic different species of unpalatable danaids, other butterflies and moths (Fig. 12.2). The males, on the other hand, are monomorphic and strikingly different from the females, exhibiting a characteristic yellow and black color pattern and tailed hindwing. The female wing patterns look very complicated in their appearance and at first glance it seems difficult to find an underlying logical relationship between the patterns even in the single species. However, Nijhout [8] proposed the idea that the black color pattern elements in the wing constitute the principal pattern elements, even though the background color attracts our attention most. The elements differ in size depending on the mimetic form and this can have dramatic effects on the overall appearance of the pattern. For this reason, our problem then simplifies to presenting a mechanism that can account for only the black pattern elements (see Fig. 12.3).
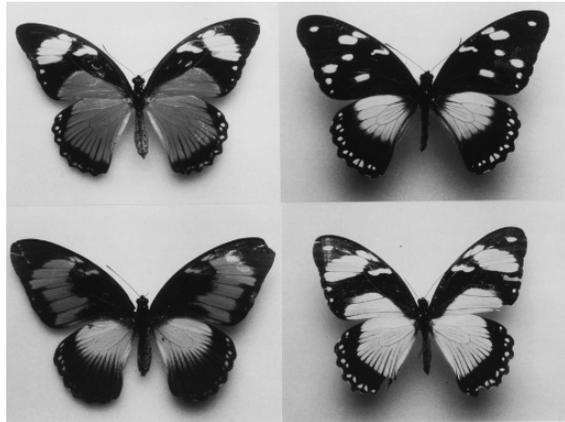
Fig. 12.2. Polymorphism in mimetic females of *Papilio dardanus: trophonius* (top left), *cenea* (top right), *planemoides* (bottom left), and *hippocoonides* (bottom right).

12.3.2 *Papilio polytes*

A species of *Papilio polytes* is widely distributed across India and Southeast Asia including the southern islands of Japan. *Papilio polytes* has monomorphic males and several female forms. The male-like female is nonmimetic and resembles the male. Other female forms are mimetic and mimic different species of unpalatable *Aristolochia*-feeding swallowtail butterflies.

The forewing pattern of the male has a white band along the distal wing margin in the entire black-colored wing. The hindwing pattern has a white band passing through
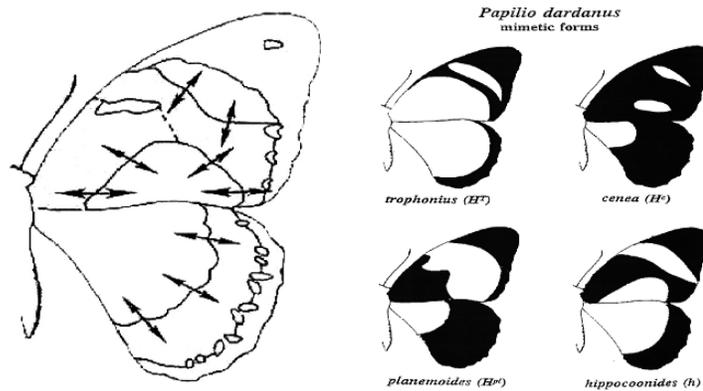
Fig. 12.3. Ground plan and the black pattern elements in mimetic forms of *Papilio dardanus* (reproduced with permission of Fred Nijhout, Duke University).

Fig. 12.4. Male and two different female forms of *Papilio polytes* with a typical unpalatable swallowtail *Pachliopta aristolochiae*: male-like female of *P. polytes* (top left), a mimetic form of *P. polytes* (top right), a male of *P. polytes* (bottom left), and a model *Pachliopta aristolochiae* (bottom right).

the middle of the wing in the anterior-posterior direction which appears to link continuously to the forewing white band when both wings are held at rest. On the other hand, a mimetic form shown in Fig. 12.4 has a different and characteristic hindwing pattern in which the white band is shortened in the central area of the wing and red-colored spots appear along the wing margin just like the model butterfly *Pachliopta aristolochiae* (see Fig. 12.4).

12.4 A Reaction-Diffusion Model and Numerical Results

The model is based on the idea that a system of reacting and diffusing chemicals could evolve from an initially uniform spatial distribution to concentration profiles that vary spatially by what is called the diffusion-driven instability [17]. We solve the non-dimensionalised reaction-diffusion system with Gierer–Meinhardt [6] reaction kinetics

$$u_t = \gamma \left(a - b\,u + \frac{u^2}{v\,(1 + k\,u^2)}\right) + \nabla^2 u, \quad v_t = \gamma \left(u^2 - v\right) + d\,\nabla^2 v$$

using the finite element method [14] on fixed two-dimensional wing domains. Here $u(\underline{x}, t)$ and $v(\underline{x}, t)$ represent chemical (morphogen) concentrations at spatial position $\underline{x}$ and time t. a, b, d, k, and γ are positive parameters.

In all simulations we fix the parameter values of $a = 0.1, b = 1.0, d = 70.8473$, $k = 0.5$, and $\gamma = 619.45$ to isolate the specific (3,0) mode pattern. Initial conditions are taken as small random perturbations around the uniform steady state and mixed
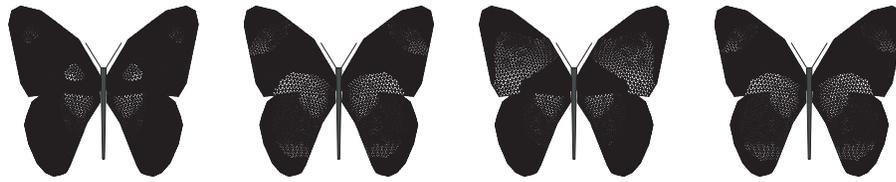
Fig. 12.5. Numerical results illustrating global patterns for cenea, hippocoonides, planemoides, and trophonius [14].

boundary conditions applied. Shading is according to a gradient threshold function of the form of a plane $\alpha y + \beta x + c_0$ where α or β or both are non-zero and c_0 is a non-negative constant (see [14] for further details).

Figs. 12.5 and 12.6 illustrate computational results corresponding to global patterns of *Papilio dardanus* and *polytes*, respectively. For *Papilio dardanus* these correspond to observed patterns of the butterflies cenea, hippocoonides, planemoides, and trophonious. In all our simulations small changes in the gradient threshold give rise to a variety of different observed patterns. We have observed in simulations that wing geometry, model parameter values, gradient threshold, and boundary conditions are key factors in obtaining the global patterns illustrated.

12.5 Conclusions and Discussion

We showed that a Turing model [17] could account for the global pigmentation patterns on butterfly wings by solving the model equations on geometrically accurate adult wing shapes of *Papilio dardanus* and *polytes*. Our results suggest that the global wing coloration is essentially due to underlying stripe-like patterns of some pigment-inducing morphogen [14]. Computations on the fore- and hindwing shapes were carried out separately. Except for a small change in a parameter value of the threshold function, we used the same parameter values to obtain both fore- and hindwing patterns. This means that from a mathematical modeling point of view, global pigmentation patterns of the fore- and hindwing are independent in the sense that they are
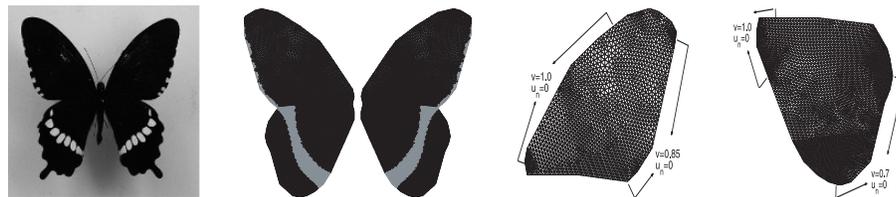
Fig. 12.6. Numerical solutions of the model equations computed on a geometrically accurate fore- and hindwing of *Papilio polytes*. Parameter values for the gradient threshold are $a_0 = -0.00111$, $b_0 = -0.005$, and $c_0 = 0.56$ (forewing) and $c_0 = 0.645$ (hindwing).

produced or controlled independently by the same mechanism. This result suggests that the close match between global pigmentation patterns on both wings is logically reasonable and somewhat accidental, but need not be explained by an unknown adaptive or functional viewpoint. This might solve the dilemma noted in Section 12.2.2, namely, that of the close match and mismatch between fore- and hindwing patterns.

Acknowledgments

References

1. Beldade, P., Brakefield, P.M.: The genetics and evo-devo of butterfly wing patterns. Nature Rev Genet, **3**, 442–452 (2002).
2. Brakefield, P.M.: The evolution-development interface and advances with eyespot patterns in Bicyclus butterflies. Heredity, **80**, 265–272 (1998).
3. Clarke, C.A., Sheppard, P.M.: Interactions between major genes and polygenes in the determination of the mimetic pattern of *Papilio dardanus*. Evolution, **17**, 404–413 (1963).
4. Dilão, R., Sainhas, J.: Modelling butterfly eyespot patterns. Proc. Roy. Soc. Lond., B **271**, 1565–1569 (2004).
5. French, V., Brakefield, P.M.: Pattern formation: A focus on Notch in butterfly wings. Curr. Biol., **14**, R663–R665 (2004).
6. Gierer, A., Meinhardt, H.: A theory of biological pattern formation. Kybernetik, **12**, 30–39 (1972).
7. Madzvamuse, A., Maini, P.K., Wathen, A.J., Sekimura, T.: A predictive model for color pattern formation in the butterfly wing of *Papilio dardanus*. Hiroshima Math. J., **32**, No.2, 325–336 (2002).
8. Nijhout, H.F.: The development and evolution of butterfly wing patterns. Smithonian Institution Press, Washington and London (1991)
9. Nijhout, H.F., Maini, P.K., Madzvamuse, A., Wathen, J.W., Sekimura, T.: Pigmentation pattern formation in butterflies: experiments and models. C. R. Biologies, **326**, 717–727 (2003).
10. Oudemans, J.T.: Etudes sur la position de repos chez les Lepidopteres. Verhandelingen der Koning-klijke Akademie van Wetenschappen, **10**, 1–90 (1903)
11. Schwanwitsch, B.N.: On the ground plan forewing-pattern in nymphalids and certain other families of rhopalocerous Lepidoptera. Proc. Zool. Soc. Lond., B **34**, 509–528 (1924).
12. Sekimura, T., Maini, P.K., Nardi, J.B., Zhu, M., Murray, J.D.: Pattern formation in lepidopteran wings. Comments Theor. Biol., **5**, No. 2–4, 69–87 (1998).
13. Sekimura, T., Zhu, M., Cook, J., Maini, P.K., Murray, J.D.: Pattern formation of scale cells in lepidoptera by differential origin-dependent cell adhesion. Bull. Math. Biol., **61**, 807–827 (1999).
14. Sekimura, T., Madzvamuse, A., Wathen, A.J., Maini, P.K.: A model for colour pattern formation in the butterfly wing of *Papilio dardanus*. Proc. R. Soc. Lond., B **267**, 851–859 (2000)

15. Sibatani, A.: Oudemans' principle and its extension in pattern formation on the wing of Lepidoptera (Insecta). J. Lib. Arts, Kansai Med. Uni., **11**, 1–10 (1987).
16. Süffert, F.: Zur vergleichende analyse der schmetterlingszeichung. Biologisches Zentralblatt, **47**, 385–413 (1927).
17. Turing, A.M.: The chemical basis of morphogenesis. Phil. Trans. Roy. Soc. Lond., B **237**, 37–72 (1952).

13

Agent-Based Model for Developmental Pattern Formation with Multiscale Dynamics and Varying Cell Geometry

Scott Christley,[1] Stuart A. Newman,[2] and Mark S. Alber[3]

[1] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA; schristl@nd.edu
[2] Department of Cell Biology and Anatomy, New York Medical College, Valhalla, NY 10595, USA; newman@nymc.edu
[3] Department of Mathematics and Center for the Study of Biocomplexity, University of Notre Dame, Notre Dame, IN 46556, USA; malber@nd.edu

Summary. Cells of the embryonic vertebrate limb in high-density culture undergo chondrogenic pattern formation, which results in the formation of regularly spaced "islands" of cartilage analogous to the cartilage primordia of the developing limb skeleton. In this chapter a discrete, multiscale agent-based stochastic model is described, which is based on an extended cell representation coupled with biologically motivated reaction-diffusion processes and cell-matrix adhesion, for studying the behavior of limb bud precartilage mesenchymal cells. The model is calibrated using experimental data, and the sensitivity of key parameters is studied.

Key words: Stochastic system, Monte Carlo simulations, developmental biology, limb development.

13.1 Introduction

Modeling pattern formation and morphogenesis are fundamental problems in developmental biology. Models of developmental processes fall into two categories: continuous models that use families of differential or integro-differential equations to describe "fields" of interaction, and discrete models in which space, time, or state may be discrete. Models may be deterministic or stochastic.

In biological applications, continuous models have been used to describe oceanic microbial cycles [2], microbial growth dynamics [32], the spread of species through an ecosystem [33], and biofilm formation [38]. Other examples can be found in Jones and Sleeandhoopesman [17] and Murray [28].

Discrete models describe individual (autonomous) behaviors. They are often applied to microscale events where a small number of elements can have a large (and stochastic) effect on a modeled system. For example, while many periodic growth patterns can be modeled using continuous methods, such patterns which depend sensitively on substrate concentration are best modeled with discrete methods. Discrete

methods applied to biological systems include cellular automata [10], lattice-gas-based cellular automata [6, 35], the cellular Potts model [3, 13], and agent-based modeling [37].

Developmental morphogenesis is the molding of living tissues during development, regeneration, wound healing, and disease. It is a complex phenomenon involving gene regulation, changes in cell shape, cell-cell interactions, and cell division, growth, and migration. Representing cell shape realistically is an important problem in modeling morphogenesis.

In silico experiments are becoming an important option, in addition to in vivo and in vitro experiments, for studying pattern formation in biological systems [16, 22]. In this chapter we describe a discrete, multiscale agent-based stochastic model, based on extended cell representation coupled with the biologically motivated reaction-diffusion processes and cell-matrix adhesion, for studying the behavior of limb bud precartilage mesenchymal cells. Those cells in micromass culture undergo chondrogenic pattern formation, which results in the formation of regularly spaced "islands" of cartilage analogous to the cartilage primordia of the developing limb skeleton.

In a previous study [18], we presented a biological lattice gas model that remains the most successful computational model to date for pattern formation in the limb cell micromass system. Nevertheless, that model diverged from biological reality in several important respects:

- The physical representation of entities of such disparate scales as cells and morphogens on a common grid is physically unrealistic.
- Cells move less than a cell diameter over the time period of condensation formation and thus should not accumulate in condensations by leaving the surrounding regions devoid of cells.
- Cells should not halt within condensations, but rather should have a finite probability of moving out of a condensation.

These issues raise the question of whether the simulation of realistic pattern formation would still occur if the artifactual aspects of the earlier model were replaced with more realistic assumptions. We therefore designed a more sophisticated model that overcomes each of these deficiencies. The cells in the new model are extended, multipixel objects that can change shape and move; cells are separated by less than a cell diameter, condense without denuding the regions surrounding condensation centers, and are not trapped once entering a condensation. Moreover, two different grids of different mesh size are used for cell and molecular dynamics, while chemical reactions, molecular diffusion, and cellular diffusion operate on different temporal scales.

13.2 Biological Background

Skeletal pattern formation in the developing vertebrate limb depends on interactions of precartilage mesenchymal cells with factors that control the spatiotemporal differentiation of cartilage. The most basic skeletogenic processes involve the spatial separation of precartilage mesenchyme into chondrogenic and nonchondrogenic domains
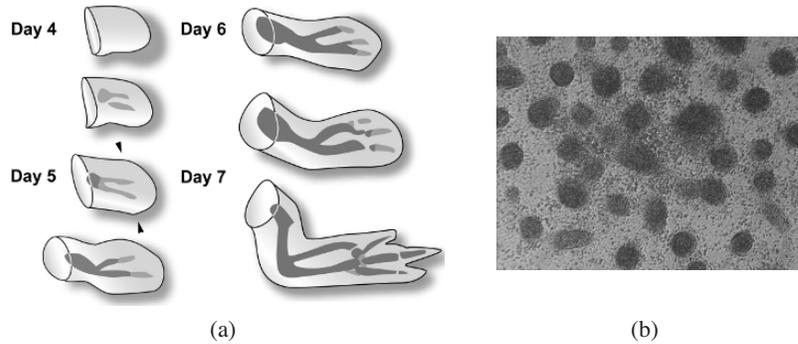
Fig. 13.1. A (left panel) Progress of limb skeletal development in chicken forelimb (wing) between 3 and 7 days of embryogenesis. Gray represents precartilage condensation and black represents definitive cartilage. The developing limb, or limb bud, is paddle-shaped, being flatter in the back-to-front (dorsoventral) dimension than in the thumb-to-little finger (anteroposterior) dimension, or the shoulder-to-finger tips (proximodistal) direction in which it mainly grows. The cartlilages that prefigure the bones first arise as stripe-like (e.g., long bones, digits) or spot-like (e.g., wrist bones shown here, or ankle bones in the hindlimb) mesenchymal condensations. The apical zone of the 5-day chicken wing bud (indicated by the arrowheads) or leg bud provide a source of not-yet-condensed mesenchymal cells that when grown in high-density "micromass" culture will form precartilage condensations. B (right panel) Discrete spot-like precartilage condensations that have formed after 72 hours in a micromass culture of 5-day leg bud apical zone limb mesenchymal cells, visualized by Hoffman Modulation Contrast optics. The cells in these cultures are initially plated as a densely packed monolayer and rearrange over short distances in the 2D plane of the culture during the indicated period. Those that enter into condensations, however, round up into the third dimension. This quasi-3D aspect of the culture system is incorporated into our model (see text). Actual size of the microscopic field is 1×1.4 mm, and each condensation contains approximately 100 cells. The spatial scale of the spot-like condensations in the developing limb and in the micromass cultures are comparable. Panel A is adapted, with changes, from [11].

(Fig. 13.1). In vitro as well as in vivo (see reviews [14, 31]), morphogens of the TGF-β family induce precartilage condensation by a process that involves the upregulation of the adhesive extracellular glycoprotein fibronectin [20]. Mesenchymal cells accumulate in regions of increased cell-matrix adhesive interactions [7, 8, 12] and then acquire epithelioid properties by upregulation of cell-cell adhesion molecules such as N-CAM [39] or cadherin-11 [21]. Cartilage differentiation or chondrogenesis follows at the sites of condensation in vitro and in vivo.

We have suggested [29, 30] that interactions between diffusible activators and inhibitors of chondrogenesis can explain the approximately periodic patterns of chondrogenesis in the developing limb and in micromass cultures. Results of Miura et al. [23–25] provide strong evidence for such a reaction-diffusion mechanism in vitro. In particular, TGF-β2 acts as an activator by positively regulating its own production, as well as precartilage condensation [23]. In addition, Moftah et al. [27] found that activation of fibroblast growth factor receptor 2 (FGFR2), which appears on cells at

sites of incipient condensation, suppresses condensation in surrounding mesenchyme by eliciting production of an inhibitor of chondrogenesis. While the molecular identity of the inhibitor is unknown, it acts laterally, spreading by an unknown mechanism from its sites of production [27]. Finally, a recent analysis of the limb skeleton of the *Doublefoot* mouse mutant [26] demonstrated that the unusual shape of the digits could be accounted for by the assumption that chondrogenic patterning in vivo is governed by a reaction-diffusion system on a growing domain.

13.3 Computational Model

The spatial environment that cells and molecules occupy is modeled on a two-dimensional plane. The implementation provides support for multiple superimposed discrete grids of various spatial scales. In our current model, we utilize two scales, one for the cellular level and another finer resolution scale for the molecular level. The coarsest resolution spatial scale is considered to be the base spatial scale, which is the cellular level for our model, and all other grids are an integer ratio size of that base grid. The base spatial grid can be defined as a square or rectangular grid of any height and width, and all of the grids overlay one another and cover the same physical area. Our model supports both periodic and no-flux boundary conditions.

13.3.1 Cell Representation

Each cell is a discrete agent represented as a set of seven contiguous pixels operating on the base spatial grid as shown in Fig. 13.2(a). We chose the simplest multipixel representation of limb mesenchymal cells subject to the following biological constraints: (i) these cells have essentially isotropic geometry, that is they do not elongate in the direction of migration but rather probe their environment by extending short randomly placed projections; (ii) the cell nucleus is also isotropic but is relatively unchanging in shape and comprises more than half the cell volume; (iii) cells in fibronectin-rich, condensing areas of the micromass round up such that their cross section in the plane of the culture is significantly reduced by 20–30 percent [5]. We maintain four pixels in a two-by-two square (kernel) configuration that represents the portion of the cell that contains the nucleus and allow the remaining pixels to occupy the border region around the nucleus. Cells that round up shrink their spatial area to five pixels as shown in Fig. 13.2(c).

Cell diffusion is implemented as a random walk. If the cell moves, then all of its seven (or five) pixels move one pixel in the appropriate direction. Cells can also fluctuate in shape, yet such fluctuations maintain a structural representation of the central region containing the nucleus by preserving intact a two-by-two square block of pixels. Therefore, shape fluctuations are restricted to the motion of the three (or one) border pixels around the nucleus which either move to new border pixels or displace nucleus pixels; Fig. 13.2(b) gives an example of both types of fluctuations for a cell changing shape. The rate and probability that cells move and change shape is parameterized separately from molecules so that they can be calibrated to the scale of actual biological cells.

Analysis of cell movement and shape within precartilage condensations [5] indicates that cells have a smaller contact area with substratum close to the condensation center and that the rate of movement of those cells increases. We model this behavior by shrinking the area of cells associated with levels of fibronectin above a threshold value from seven pixels to five pixels; and we increase the rate at which those cells can move and change shape. Likewise, once a cell ventures onto fibronectin it has the tendency to remain there with a low probability of leaving the condensation.

13.3.2 Molecular Model and Reaction-Diffusion Mechanism

The reaction-diffusion mechanism introduced by Turing [36] is based upon interactions of slow diffusing, positive feedback activator molecules and fast diffusing, negative feedback inhibitor molecules. This continuum mechanism was shown to be capable of self-organizing spatiotemporal patterns of chemical concentration. In our cellular automaton representation of the Turing mechanism, a discrete number of activator and inhibitor molecules occupies each pixel on the grid and each molecule is considered to have a spatial representation of just one pixel.

We model the reaction dynamics of the activator and inhibitor molecules at each pixel as follows: let U_t and V_t be the concentration of the activator and inhibitor, respectively, at time t and let ϕ_t be an indicator function for the existence of a cell at time t,

$$\Delta U_t = \min\{MAX_U, (k_1 U(t) + B_U)\phi_t + k_2 V_t\} \tag{13.1}$$

$$\Delta V_t = \min\{MAX_V, k_3 U_t \phi_t + k_4 V_t\}. \tag{13.2}$$

(a)   (b)

(c)
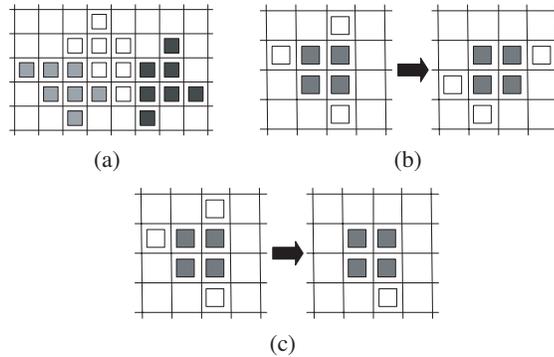
Fig. 13.2. (a) Three cells on the spatial grid each occupying seven pixels. (b) Cell changes shape. The region of the cell that contains the nucleus, indicated by the four gray pixels, is structurally maintained; two border pixels move to new locations, and one border pixel (upper right) displaces a nucleus pixel which gets shifted to the right. (c) Cell rounding up on fibronectin. The surface area with fibronectin is reduced with two border pixels moving into a quasi third dimension above the cell.

Equation (13.1) shows the change over time for each pixel on the grid of the activator morphogen concentration based upon a proportion (as defined by chemical reaction rates) of the current activator and inhibitor concentrations. Equation (13.2) shows the corresponding change over time for each pixel on the grid of the inhibitor morphogen. The activator morphogen is considered a positive self-regulating molecule and a positive regulator of the inhibitor; thus, chemical rate parameters k_1 and k_3 both have positive values. The inhibitor morphogen is considered a negative regulator of activator that decays over time; thus, chemical rate parameters k_2 and k_4 are both negative values.

In our model, production of the activator and inhibitor molecules, as represented by the parameters k_1 and k_3, can only occur in the presence of a cell; however, the decay of activator and inhibitor, as represented by the parameters k_2 and k_4, is considered to occur independently of cell presence. Cells are initially randomly distributed on the grid and secrete a small basal amount (B_U) of activator morphogen which provides the initial concentration of activator; cells continue this basal production throughout the simulation, and the inhibitor concentration starts at zero.

We consider the chemical rate parameters to be averages of discrete cell behavior of the underlying stochastic gene expression for the two morphogens, so they can assume real number values. We consider cells to respond to small amounts of morphogens, so molecules are represented as discrete entities instead of averages, which provides a more accurate representation of the biology. Therefore, the morphogen concentrations (U_t, V_t) are whole numbers, and changes in the concentrations at a time step are rounded to the nearest integer and prevented from going negative.

In any physicochemical reaction there is a limitation on how much reagent a single cell can realistically produce during any period of time. For this reason, our model provides separate parameters (MAX_U, MAX_V) for the maximum amount of activator and inhibitor that can be produced during a single reaction step. The maximums are imposed on individual pixels of the molecular grid rather than across the entire cell to represent polarization of limb mesenchymal cells [15]; this allows for small morphogen gradients to be present across the spatial extent of an individual cell through spatially polarized secretion of morphogens. The peaks of activator concentration produced by the reaction-diffusion dynamics define a large prepattern equal in spatial area to the fibronectin patches, containing around one hundred cells within a single patch. Thus, polarization plays a role for the cells on the border region of the patch, while cells in the patch interior perceive a relatively constant morphogen concentration across their entire spatial extent.

Molecular diffusion from any pixel can occur randomly toward any of the four neighboring pixels (up, down, left, right). The diffusion rate D is scaled into a probability factor $0 < p < 1$ and a time step n such that $D = pn$. The probability determines the chance that a molecule will diffuse, and the time step indicates how many opportunities a molecule has to diffuse for a single simulation iteration; if the molecule diffuses then one of the four neighboring pixels is picked with equal probability. The chemical reaction operates at a much slower rate than molecular diffusion, so the time scales are separated with diffusion calculated at a small time step and the reaction calculated at a longer time step. Algorithm 1 shows how the diffusion is performed at a

---

Algorithm 1 calculateReactionDiffusion

---

  for each simulation iteration **do**
     Calculate chemical reaction for each pixel on grid.
     for $i = 1$ to n **do**
        Calculate activator and inhibitor diffusion for each pixel on grid.
     end for
  end for

---

finer time scale.

13.3.3 Fibronectin Production

Fibronectin is a nondiffusing, extracellular matrix molecule that forms the template for precartilage condensations. As the concentration levels of the activator morphogen increases in the presence of a cell, that cell produces fibronectin mRNA which can then be translated into actual fibronectin protein molecules. We support a simple threshold level such that once the sum of activator concentration across the entire spatial area of a cell exceeds that threshold value, the cell differentiates into a fibronectin-producing cell. Because we do not directly describe the level of fibronectin mRNA within the cell, the trigger for cell differentiation is separated from the actual production of fibronectin, and a model parameter defines the delay between cell differentiation and secretion of fibronectin.

When a cell produces fibronectin, a single multimolecular unit is secreted with random probability for each of the pixels on the molecular grid in the cell's spatial extent, and each molecule is allowed to perform an initial small diffusion of at most one pixel [1]. Production of fibronectin molecules continues until a maximum concentration level is reached at a pixel, although cells may still continue to produce fibronectin on pixels that have not yet reached the maximum. The production rate of fibronectin, the duration of such production, and the maximum amount of fibronectin allowed per pixel can be adjusted with model parameters.

13.4 Model Calibration

In attempting to calibrate our model parameters with known empirical parameters we desire to correlate the spatial and temporal patterns produced by computer simulation results with in vivo and in vitro experiments. For spatial patterns, we consider the size, shape, and distribution of the fibronectin-rich spatial domains; for temporal patterns, we consider the reaction rates of activator and inhibitor production, the diffusion rates of both cells and molecules, the onset of fibronectin production, the production rate of fibronectin, and the shape and movement fluctuations of cells on fibronectin. The actual values for the set of key parameters used in the simulation and their corresponding physical measurements, if known, are shown in Table 13.1.

Table 13.1. Calibrated simulation parameters to known physical values.

Parameter	Physical Value	Simulation Value
Cell diameter/area	10 μm / 79 μm^2	7 pixels
Cell spatial grid		240 $\times$ 240 pixels
Molecular spatial grid		480 $\times$ 480 pixels
Spatial ratio cells : molecules	10000 : 1	28 pixels : 1 pixel
Reaction temporal scale	70.9 msec	1 reaction
Diffusion temporal scale ($n = 200$)	0.3544 msec	1 diffusion step
Basal activator production (B_U)	unknown	28
Activator self-regulation (k_1)	unknown	0.3356
Activator regulation of inhibitor (k_2)	unknown	0.16
Inhibitor regulation of activator (k_3)	unknown	-1.1
Inhibitor decay (k_4)	unknown	-0.4615
Maximum activator produced (MAX_U)	unknown	8000
Maximum inhibitor produced (MAX_V)	unknown	8000
Activator diffusion rate (D_U)	10 μm^2/sec	50 pixels/iteration
Inhibitor diffusion rate (D_V)	unknown	196 pixels/iteration
Cell diffusion rate	0.42 μm^2/min	1 pixel/60 iterations
Cell diffusion rate on fibronectin	0.62 μm^2/min	1 pixel/40 iterations

13.4.1 Diffusion Rates

Diffusion rates for the activator and inhibitor play a vital role in defining the wavelength of the Turing patterns produced by the reaction-diffusion dynamics that ultimately determine the size and distribution of the fibronectin-rich patches. Lander et al. [19] calculate the effective diffusion coefficient for a molecule the size and shape of Dpp to be 10 μm^2 s^{-1}. TGF-β which we assume to be our activator morphogen, based upon cell-culture experiments [20, 25], is a morphogen of the same molecular class as Dpp; therefore, we take the diffusion coefficient of Dpp to be our activator diffusion rate. The inhibitor morphogen, whose identity is unknown but is induced at sites of incipient condensation by ectodermally produced fibroblast growth factors (FGFs) [27], must diffuse at a faster rate than the activator morphogen for patterns to be produced. We found that an inhibitor diffusion rate approximately four times faster than the activator diffusion rate is sufficient to produce stable patterns.

The diffusion rate for cells is considerably slower than for the activator and inhibitor molecules. Cui [5] used phase-constrast microscopy and video-based cell tracking to measure the movements of cells during developing chicken-limb precartilage mesenchyme over the time period of condensation formation. He calculated an average diffusion coefficient of 0.5058 μm^2/min, and his data is consistent with cells

moving slightly faster in condensations (see also [9]). We qualitatively support this by making cells associated with fibronectin diffuse faster than cells not associated with fibronectin (see Table 13.1).

13.4.2 Reaction Rates

The key requirement is that the parameters for the reaction rates fall within the morphogenetic region in order for Turing-type patterns to be produced. Within this morphogenetic region, two types of behavior can be observed for the morphogen concentrations: steady-state equilibrium with stable patterns and oscillatory behavior with transient patterns. The oscillatory behavior is induced by the maximums on production; otherwise a steady state would be attained but at an unrealistic concentration level. As these reaction rates are unknown, we take the approach of picking parameters that correspond to our understanding of qualitative cell-culture behavior.

13.4.3 Onset and Rate of Fibronectin Production

It was previously shown that transient exposure of precartilage mesenchymal cells in vitro to TGF-β for 3–6 hours a day after cultures are established is sufficient to induce the production of precocious condensations by those cells: the onset of precartilage condensations begins about 32 hours after plating and condensations reach their final size at 72 hours [20]. In this study cells responded very quickly to exposure to TGF-β by producing fibronectin mRNA even though synthesis of the fibronectin protein molecule does not occur until later. Furthermore, production of FGFR2 protein (the mediator of FGF induction of the inhibitor at condensation sites) is no longer detectable after 48 hours in similar cultures [27]. These results suggest that a reaction-diffusion mechanism does not have to maintain stable peaks throughout the period of complete formation of precartilage condensation, but may in fact only require transient peaks that induce transcription of fibronectin mRNA. This could provide a prepattern for condensations occurring later in time.

Based upon these experimental observations, we pick reaction parameters giving oscillatory behavior as a model for the transient nature of the activator and inhibitor morphogens. As experimental results do not give conclusive evidence for how long the morphogens are active in the system and the various morphogens may have other roles beyond induction of fibronectin production, we do not attempt to match the total time length of activity of the morphogens in our simulation. Instead we calibrate with the primary knowledge that cells respond quickly to TGF-β, so we trigger cell differentiation early in our simulation when the sum of activator concentration across the entire cell reaches a threshold value of 5,000 molecules.

13.5 Results

Fig. 13.3(a) shows the typical output of the fibronectin patches that are produced for a simulation run; visual inspection of the patterns matches well with the cell culture experiments as typified by Fig. 13.1.

(a)   (b)
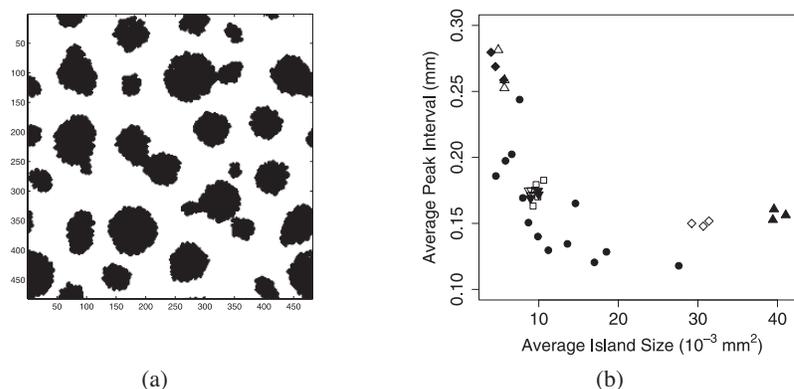
Fig. 13.3. A (left panel) Fibronectin-rich patches produced by simulation using parameter values in Table 13.1. B (right panel) Average peak interval versus average island size for 13 experimental (●) and 5 simulation (□) points using parameter values in Table 13.1 with different random initial conditions. Simulation points for variations in some of the key parameters are shown: +5% (◇) and −5% (◆) for activator self-regulation (k_1), +5% (△) and −5% (▲) for activator regulation of inhibitor (k_3), +5% (▽) and −5% (▼) for inhibitor regulation of activator (k_2). All simulations were run for 3000 iterations with periodic boundary conditions.

We measured the average peak interval using the peak length method of [23] and the average island size for the fibronectin patches for a set of simulations. We then compared those results against measurements for a set of experimental images like Fig. 13.1. Fig. 13.3(b) shows a comparison between the experimental and simulation measurements. We explored the robustness of the parameter set by varying key parameters independently (±5%). Minor variation of the inhibitor strength on activator (k_2) produced little change in the resulting fibronectin patches, while variations of activator self-regulation (k_1) and regulation of inhibitor (k_3) produced results at the extremes of the experimental data.

13.6 Conclusion

Development of a discrete, stochastic model that incorporates aspects of agent-based modeling and multiple physical and temporal scales has provided us with the ability to focus on specific biological phenomena while abstracting away biological details that may not be central to the process. The agent-based representation of mesenchymal cells allows for a simple implementation of more biologically realistic cells as extended, multipixel objects with variable cell geometry, while facilitating a computationally efficient simulation. Also, the agent-based representation of cells provides fine control over the specification of how cells perceive morphogen concentrations especially as a trigger for differentiation, how cells produce and secrete morphogens and fibronectin, and how cells behave on fibronectin including the shrinking of area and

tendency to remain confined with the corresponding domains. The multiscale nature of our model incorporated additional biological realism through the separation of physical scales by modeling cells and morphogens on grids of differing resolution and the separation of time scales for chemical reactions, morphogen diffusion, cell diffusion, and the production of fibronectin.

We have shown that our more sophisticated model produces fibronectin-rich patches consistent in size, shape, and distribution to condensations in micromass culture experiments while overcoming the deficiencies of our earlier model. Specifically, cells maintained a space-filling, nonoverlapping, high-density culture throughout the spatial domain. Likewise, cells maintained movement within condensations without being trapped once entering a condensation, and variations of key chemical rates indicate the robustness of the parameter values for the reaction-diffusion dynamics. Model calibration is important to keep the mechanics of the model closely tied to known experimental values. The fact that our use of biologically motivated values for activator and cell diffusion, variable cell geometry, and authentic details of molecular production and cell behavior gave rise to realistic patterns is a strong indication of the validity of our model. Once tied to experimental values, other parameters in the simulation provide hypotheses for biological rates that can be experimentally tested.

The reaction-diffusion dynamics is unconventional in that molecules are represented as discrete entities rather than as continuous concentrations. The two regimes of behavior that we have uncovered, steady-state and oscillatory, for the temporal dynamics of the pattern-forming morphogen concentrations, are explored further in [4], as are transitions between spot-like and stripe-like patterns in this system.

Acknowledgments

References

1. Alber, M., Kiskowski, M., Glazier, J., Jiang, Y.: On cellular automaton approaches to modeling biological cells. In: Rosenthal, J., Gilliam, D. (eds.) Mathematical Systems Theory in Biology, Communication, and Finance, volume 134, 1–39. Springer-Verlag, New York (2003).
2. Belov, A.P., Giles, J.D.: Dynamical model of buoyant cyanobacteria. Hydrobiologia, **349**, 87–97 (1997).
3. Chaturvedi, R., Huang, C., Kazmierczak, B., Schneider, T., Izaguirre, J.A., Glimm, T., Hentschel, H.G.E., Glazier, J.A., Newman, S.A., Alber, M.S.: On multiscale approaches to three-dimensional modelling of morphogenesis. J. R. Soc. Interface, **2**, 237–253 (2005).
4. Christley, S., Alber, M.S., Newman, S.A.: Patterns of mesenchymal condensation in a multiscale, discrete stochastic model. PLoS Comput. Biol., in press.
5. Cui, C.: Dynamics of cell movement and tissue motion in gastrulation and micromass cell culture. Ph.D. thesis, Indiana University, Bloomington, IN, (2005).

6. Deutsch, A.: Orientation-induced pattern formation: swarm dynamics in a lattice-gas automaton model. Int. J. Bifurc. Chaos, **6**, 1735–1752 (1996).

7. Downie, S., Newman, S.: Morphogenetic differences between fore and hind limb precartilage mesenchyme: Relation to mechanisms of skeletal pattern formation. Dev. Biol., **162**, 195–208 (1994).

8. Downie, S., Newman, S.: Different roles for fibronectin in the generation of fore and hind limb precartilage condensations. Dev. Biol., **172**, 519–530 (1995).

9. Ede, D., Flint, O., Wilby, O., Colquhoun, P.: The development of precartilage condensations in limb bud mesenchyme in vivo and in vitro. In: Ede, D., Hinchliffe, J., Balls, M. (eds.) Vertebrate Limb and Somite Morphogenesis, 161–179. Cambridge University Press, Cambridge (1977).

10. Ermentrout, G.B., Edelstein-Keshet, L.: Cellular automata approaches to biological modeling. J. Theor. Biol., **160**, 97–133 (1993).

11. Forgacs, G., Newman, S.: Biological Physics of the Developing Embryo. Cambridge University Press (2005).

12. Frenz, D., Jaikaria, N., Newman, S.: The mechanism of precartilage mesenchymal condensation: A major role for interaction of the cell surface with the amino-terminal heparin-binding domain of fibronectin. Dev. Biol., **136**, 97–103 (1989).

13. Glazier, J.A., Graner, F.: Simulation of the differential adhesion driven rearrangement of biological cells. Phys. Rev. E, **47**, 2128–2154 (1993).

14. Hall, B.K., Miyake, T.: All for one and one for all: Condensations and the initiation of skeletal development. BioEssays, **22**, 138–147 (2000).

15. Holmes, L.B., Trelstad, R.L.: Cell polarity in precartilage mouse limb mesenchyme cells. Dev. Biol., **78**, 511–520 (1980).

16. Izaguirre, J.A., Chaturvedi, R., Huang, C., Cickovski, T., Coffland, J., Thomas, G., Forgacs, G., Alber, M., Hentschel, G., Newman, S.A., Glazier, J.A.: COMPUCELL, a multi-model framework for simulation of morphogenesis. Bioinformatics, **20**, 1129–1137 (2004).

17. Jones, D.S., Sleeandhoopesman, B.D.: Differential Equations and Mathematical Biology. CRC Press, Boca Raton, FL, (2003).

18. Kiskowski, M., Alber, M., Thomas, G., Glazier, J.A. Bronstein, N.B., Pu, J., Newman, S.A.: Interplay between activator-inhibitor coupling and cell-matrix adhesion in a cellular automaton model for chondrogenic patterning. Dev. Biol., **271**, 372–387 (2004).

19. Lander, A., Nie, Q., Wan, F.: Do morphogen gradients arise by diffusion? Dev. Cell, **2**, 785–796 (2002).

20. Leonard, C., Fuld, H., Frenz, D., Downie, S.A., Massagué, J., Newman, S.A.: Role of transforming growth factor-β in chondrogenic pattern formation in the embryonic limb: Stimulation of mesenchymal condensation and fibronectin gene expression by exogenenous TGF-β and evidence for endogenous TGF-β-like activity. Dev. Biol., **145**, 99–109 (1991).

21. Luo, Y., Kostetskii, I., Radice, G.: N-cadherin is not essential for limb mesenchymal chondrogenesis. Dev. Dyn., **232**, 336–344 (2005).

22. Marée, A., Hogeweg, P.: How amoeboids self-organize into a fruiting body: Multicellular coordination in *Dictyostelium discoideum*. Proc. Nat. Acad. Sci. USA, **98**, 3879–3883 (2001).

23. Miura, T., Komori, M., Shiota, K.: A novel method for analysis of the periodicity of chondrogenic patterns in limb bud cell culture: Correlation of in vitro pattern formation with theoretical models. Anat. Embryol., **201**, 419–428 (2000).

24. Miura, T., Shiota, K.: Extracellular matrix environment influences chondrogenic pattern formation in limb bud micromass culture: Experimental verification of theoretical models. Anal. Rec., **258**, 100–107 (2000).

25. Miura, T., Shiota, K.: TGFβ2 acts as an "activator" molecule in reaction-diffusion model and is involved in cell sorting phenomenon in mouse limb micromass culture. Dev. Dyn., **217**, 241–249 (2000).

26. Miura, T., Shiota, K., Morriss-Kay, G., Maini, P.K.: Mixed-mode pattern in *Doublefoot* mutant mouse limb-Turing reaction-diffusion model on a growing domain during limb development. J. Theor. Biol. (in press) doi:10.1016/j.jtbi.2005.10.016, (2005).

27. Moftah, M., Downie, S., Bronstein, N.B, Mezemtseva, N., Pu, J., Maher, P.A., Newman, S.A.: Ectodermal FGFs induce perinodular inhibition of limb chondrogenesis in vitro and in vivo via FGF receptor 2. Dev. Biol., **249**, 270–282 (2002).

28. Murray, J.: Mathematical Biology. Springer-Verlag, New York (2003).

29. Newman, S.: Sticky fingers: Hox genes and cell adhesion in vertebrate development. BioEssays, **18**, 171–174 (1996).

30. Newman, S., Frisch, H.: Dynamics of skeletal pattern formation in developing chick limb. Science, **205**, 662–668 (1979).

31. Newman, S., Tomasek, J.: Morphogenesis of connective tissues. In: Comper, W.D. (ed.) Extracellular Matrix Molecular Components and Interactions, volume 2, 335–369. Harwood Academic Publishers, Amsterdam (1996).

32. Panikov, N.S.: Microbial Growth Kinetics. Springer, New York, (1995).

33. Scott, E.M., Rattray, E.A.S., Prosser, J.I., Killham, K., Glover, L.A., Lynch, J.M., Bazin, M.J.: A mathematical model for dispersal of bacterial inoculants colonizing the wheat rhizosphere. Soil Biol. Biochem., **27**, 1307–1318 (1995).

34. Sozinova, O., Jiang, Y., Kaiser, D., Alber, M.: A three-dimensional model of fruiting body formation. Proc. Natl. Acad. Sci. USA, **103**(46), 17255–17259 (2006).

35. Sozinova, O., Jiang, Y., Kaiser, D., Alber, M.: A three-dimensional model of myxobacterial aggregation by contact-mediated interactions. Proc. Nat. Acad. Sci. USA, **102**, 11308–11312 (2005).

36. Turing, A.: The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, **237**, 37–72 (1952).

37. Walker, D.C., Hill, G., Wood, S.M., Smallwood, R.H., Southgate, J.: Agent-based computational modeling of wounded epithelial cell monolayers. IEEE Transactions on Nanobioscience, **3**, 153–163 (2004).

38. Wanner, O.: Modelling of biofilms. Biofouling, **10**, 31–41 (1996).

39. Widelitz, R., Jiang, T., Murray, B., Chuong, C.: Adhesion molecules in skeletogenesis: II. Neural cell adhesion molecules mediate precartilaginous mesenchymal condensations and enhance chondrogenesis. J. Cell. Physiol., **156**, 399–411 (1993).

Bacterial Swarming Driven by Rod Shape

Jörn Starruß,[1,3] Fernando Peruani,[1,3] Markus Bär,[2] and Andreas Deutsch[1]

[1] Center for Information Services and High Performance Computing, Technische Universität Dresden, Dresden, Germany; joern.starruss@tu-dresden.de; fernando.peruani@tu-dresden.de
[2] Physikalisch-Technische Bundesanstalt, Berlin, Germany
[3] The two first authors have contributed equally to this paper.

Summary. Swarming pattern formation of self-propelled entities is a prominent example of collective behavior in biology. Here we focus on bacterial swarming and show that the rod shape of self-propelled individuals is able to drive swarm formation without any kind of signaling.

The underlying mechanism we propose is purely mechanical and is evidenced through two different mathematical approaches: an on-lattice and an off-lattice individual-based model. The intensities of swarm formation we obtain in both approaches uncover that the length-width aspect ratio controls swarm formation. Moreover we show that there is an optimal aspect ratio that favors swarming.

Key words: Swarming, rod-shaped bacteria, cellular Potts model, cellular automaton, individual-based modelsmotion.

14.1 Introduction

Pattern formation in densely packed biological populations is an ubiquitous phenomenon based on the interaction of the constituent entities. Here we focus on swarming patterns which arise on the microscopic scale of gliding bacteria. Intuitively, one associates some kind of communication or rather signaling between the entities that align their direction of motion and move in coherent swarms. However, inspired by cell swarms in rod-shaped gliding bacteria like myxobacteria, we study a new mechanism for swarm formation in which no communication mechanism is involved.

Myxobacteria are model organisms for self-organized pattern formation; they can organize their movement without using diffusive chemical signals. The morphogenesis of complex cell patterns in the life cycle of myxobacteria (e.g., ripples, fruiting bodies) has attracted remarkable interest in developmental biology [5,9]. Within the myxobacterial life cycle swarms evolve in various phases: in the growth phase flares leave the colony to reach out for new nutrients, and during the early fruiting body formation cell clusters move collectively into aggregation centers. A "swarm" here describes aligned individuals tight-knit in a cluster, moving collectively into the same direction.

We want to find out which mechanism can drive swarm formation in the absence of diffusive signaling. Which general conditions are required for this mechanism to work?

Though the formation of swarms can be induced by a variety of local rules [1, 2], the swarming pattern itself cannot elucidate the mechanisms responsible for its formation. For this, the underlying mechanisms of agent interaction have to be analyzed.

Experimentalists as well as modelers have suggested various mechanisms responsible for swarm formation in myxobacteria. One very early suggestion was based on the slime secreted by myxobacteria while gliding on solid surfaces. Other cells are reported to follow such slime trails. Those cells were supposed to increase the attraction of the slime trail by enlarging it [10]. The stochastic cellular automaton model designed by Stevens tests whether this mechanism can explain swarm formation [18]. Although small cell tracks arise in this model, larger stable patterns like aggregates are possible solely by introducing an additional diffusing chemoattractant.

Jelsbak and Søgaard-Andersen discussed a chain formation mechanism, where two cells can attach "head" to "tail," causing swarm-like patterns [8]. The adhesive effect of extracellular appendages (i.e., pili and fibrils) is known, but the mechanism has not been tested with a model so far.

Here we introduce a new hypothesis, how rod-shaped entities (e.g., cells) can form swarms simply by mechanical interaction. Explicitly we exclude any intercellular diffusive signaling and we consider interaction among cells due only to volume exclusion. We propose that swarm formation arises from the collision of the individuals, which induces local alignment. Contrary to the alignment in liquid crystals, active asymmetric particles align their orientations and their directions of motion; consequently, they are able to move together.

We tested this hypothesis through simulations of on-lattice and off-lattice models, assuming flexible or rigid rods, respectively. Recently a model for myxobacteria dealing with similar assumptions was proposed [7]. However, in this approach physical cell interaction is represented by a local ad hoc averaging of cell orientation. In contrast, we model interaction based on first principles, i.e., volume exclusion. With our models we studied agent interaction and analyzed the impact of the anisotropy of the individuals on the swarm formation competence.

This chapter is organized as follows: In Section 14.2 the models are defined. The results of model simulations are discussed in Section 14.3, and the conclusions are presented in Section 14.4.

14.2 Model Definitions

14.2.1 On-Lattice Model

Contrary to cellular automata, models dealing with point-like objects, the cellular Potts model (CPM) allows the explicit consideration of different object shapes [3]. Our first model is based on the cellular Potts model which has been previously used for modeling differential adhesion-driven morphogenesis and fruiting body formation in *Dictyostelium discoideum*, among others [6, 13]. In order to mimic a rod-shaped cell, including its stiffness, we introduce the concept of a segmented cell. Cell segments

are formed out of lattice nodes, like the cells in the CPM. A chain of such cell segments represents the elongated cell. Using the mechanism of differential adhesion [6] we favor a connected cell by reducing the interaction energy between neighboring cell segments, resulting in a sorting of the segments in a row. Depending on other energetic model parameters, such a cell can remain permanently connected. Energetic constraints penalize the deviation of the model cell from the rod shape (by means of curvature and segment distance).

The state σ of each node $\mathbf{i} \in \mathbb{L}$ of the model space $\mathbb{L} \subset \mathbb{Z}^2$ is defined as

$$\sigma(\mathbf{i}) = (\mu, \nu) \in \mathbb{S} = \{(0,0), (k,l), k \in \{1, \dots, n\}, l \in \{1, \dots, s\}\}, \qquad (14.1)$$

where n is the number of cells and s the number of segments per cell. In this sense $\sigma_1 = (\mu_1, \nu_1)$ belongs to segment number ν_1 of cell μ_1 except for $\mu_1 = \nu_1 = 0$ where σ_1 is part of the medium. An orientation of the cell was introduced by denoting segments with $\nu = 1$ "head," the first in moving direction; consequently, segments with $\nu = s$ are denoted "tail." A center of mass $\mathbf{S}_{\mu,\nu}$ can be defined for each cell segment by averaging the positions of the nodes occupied by the segment.

We extended the Hamiltonian of the CPM by constraints for the cell deformation on the basis of the segments' centers of mass $\mathbf{S}$ (third and fourth terms in (14.2), also see Fig. 14.1):

$$\mathcal{H} = \sum_{\langle \mathbf{i},\mathbf{j} \rangle \text{ neighbors}} \frac{1}{2} J_{\sigma(\mathbf{i}),\sigma(\mathbf{j})} + \lambda \sum_{\mu=1}^{n} \sum_{\nu=1}^{s} \left(a_{\mu,\nu} - A_S \right)^2$$

$$+ \zeta \sum_{\mu=1}^{n} \sum_{\nu=1}^{s-1} \left(\left| \mathbf{S}_{\mu,\nu} - \mathbf{S}_{\mu,\nu+1} \right| - D_S \right)^2$$

$$+ \xi \sum_{\mu=1}^{n} \sum_{\nu=1}^{s-2} \left(1/R_{\text{curve}} \left(\mathbf{S}_{\mu,\nu}, \mathbf{S}_{\mu,\nu+1}, \mathbf{S}_{\mu,\nu+2} \right) \right)^2. \qquad (14.2)$$
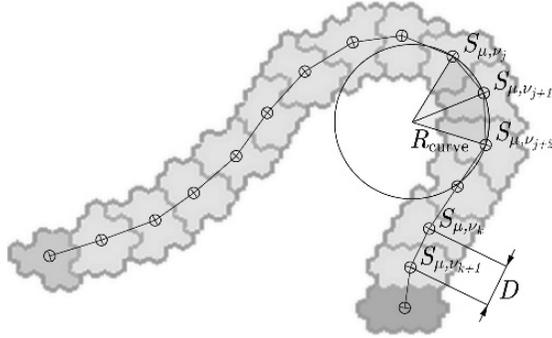
Fig. 14.1. Scheme of calculating cell deformation (R_{curve}, D) based on the segments' centers of mass $\mathbf{S}_{\mu,\nu}$.

The definition of the Hamiltonian (14.2), i.e., the system's free energy, includes an interaction energy $J_{\sigma(\mathbf{i}),\sigma(\mathbf{j})}$ between neighboring nodes[4] $(\mathbf{i}, \mathbf{j})$, depending on the state of the nodes σ. The energy between identical states is zero by default. Low energies between the states of different cells, i.e., lower than between cell and medium, cause the cells to adhere as has been shown by Graner and Glazier [6]. Here, a thin film of medium between myxobacterial cells was modeled through a short-range repulsion between model cells. For this, the interaction energy between cells was chosen 1.5 times as high as the cell-medium interaction energy.

The parameters λ, ξ, ζ represent the sensitivity of the three energetic constraints. The second term of (14.2) defines the area constraint for the cell segment area $a_{\mu,\nu}$ (A_S target area). The third term penalizes the deviation of the distance of neighboring cell segments from an optimal distance D_S. The last term defines a curvature energy based on the curvature radius R_{curve} of three consecutive cell segments. For details see Starruß et al. [17].

Active motion of the rods is modeled by an energetic advance Δd of updates shifting the cell into its moving direction $\widetilde{\theta}$, which is defined by the cell orientation. We assume an update to shift the center of mass of the two involved states[5] into a direction $\mathbf{c}$, which is identical to the copy direction. The energetic advance Δd compares both moving directions to $\mathbf{c}$ via a scalar product and weights the sum of both products with a propulsion parameter. Numerical simulations of the model were performed on a periodic hexagonal lattice using the Metropolis kinetics [14] as follows: the state of a randomly selected node is copied to one of its first neighbors (randomly selected); then the total change in free energy including the energy for motion is calculated ($\Delta\mathcal{H}' = \Delta\mathcal{H} + \Delta d$); the copy operation is accepted with probability p (14.3) and refused otherwise. The time scale is defined as a Monte Carlo step (MCS) and includes a certain number of such updates depending on the lattice dimensions. Model parameters were determined using experimental data of single cell experiments [16].

$$p(\Delta\mathcal{H}') = \begin{cases} 1 & \text{if } \Delta\mathcal{H}' < 0 \\ e^{-\Delta\mathcal{H}'/kT} & \text{if } \Delta\mathcal{H}' \geq 0 \end{cases}. \tag{14.3}$$

14.2.2 Off-Lattice IPS Model

In this subsection we introduce an interacting particle system model that could be thought of as a further abstraction of the previous model. From now on we refer to cells as particles. By comparing both models, we will be able to analyze the possible universality of the observed phenomena.

There are significant differences from the previous model: (i) the shape of particles is fixed, i.e., they cannot bend, (ii) particles are soft objects, i.e., they can be compressed but they cannot be deformed, (iii) particles move off-lattice, and (iv) we

---

[4] The neighborhood we used includes first- and second-nearest neighbors on a hexagonal lattice.

[5] An update is a copy operation involving two states. One state is copied from node $\mathbf{i}$ to $\mathbf{i}'$, while the initial state of $\mathbf{i}'$ is substituted.

distinguish a completely deterministic and, by turning on noise terms, a stochastic model version. However, contrary to the on-lattice model, the off-lattice model is not intrinsically stochastic.

We start with a brief description of the model. We consider rod-like particles moving on a plane. Each particle is equipped with a self-propelling force acting along the long particle axis. We also assume that particles are submerged in a viscous medium. Therefore, velocity and angular velocity are assumed to be always proportional to the force and torque correspondingly. The anisotropy of the particles requires the use of three different friction coefficients [4,11,12] which correspond to the resistance exerted by the medium when particles move either along their long or short axis, or even when they rotate.[6] The state of a particle is determined exclusively by the position of its center of mass and the orientation of its long axis. This is due to the fact that we assume that bacteria are objects driven by Newtonian forces, and that the viscosity of the medium is such that we are able to neglect inertial terms.[7] In consequence, the evolution of a particle is simply given by the equations for the velocity of the center of mass and the angular velocity of the long axis. The equations of motion can be expressed in the following form:

$$(v_\parallel, v_\perp) = \left(\frac{1}{\zeta_\parallel}(F - \frac{\partial U}{\partial x_\parallel}), -\frac{1}{\zeta_\perp}\frac{\partial U}{\partial x_\perp} \right) \tag{14.4}$$

$$\frac{d\theta}{dt} = -\frac{1}{\zeta_\theta}\frac{\partial U}{\partial \theta}, \tag{14.5}$$

where $v_\parallel$, $v_\perp$ refer to the velocities along the long axis and short axis of the particle, respectively, ζ_i indicates the corresponding friction coefficients (ζ_θ is related to the friction torque), U refers to the interaction potential, and F is the magnitude of the self-propelling force. Note that equation (14.4) is written in the coordinate system of the particle, which is specified by the orientation of the long axis of the particle (indicated by $\parallel$-symbol) whose dynamics is given by equation (14.5). So, equations (14.4) and (14.5) are coupled. Also, observe that F acts only along the long axis, and that $\zeta_\parallel < \zeta_\perp$ for elongated objects.

Finally, particle interaction is modeled by a potential which penalizes particle overlapping in the following way (see Fig. 14.2):

$$U(x, \theta, x', \theta') = C \left(\frac{1}{(\gamma - a(x, \theta, x', \theta'))^\beta} - \frac{1}{\gamma^\beta} \right), \tag{14.6}$$

where $a(x, \theta, x', \theta')$ is the overlapping area, γ is a parameter that can be associated to the maximum compressibility, β controls the stiffness of the particle up to the maximum compressibility, and C is an arbitrary constant.

We can also test the robustness of the model against fluctuations by adding noise terms $R_\perp(t)$, $R_\parallel(t)$ in (14.4) and $R_\theta(t)$ in (14.5).

---

[6] Assuming hydrodynamical interactions with the medium, we can make use of an explicit expression for the friction coefficients that are functions of particle length, width, and medium viscosity [4,11,12].

[7] In the over-damped case, second derivatives are assumed to vanish.
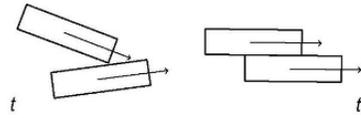
Fig. 14.2. The scheme illustrates that the potential (bottom figures) as well as the force and torque depend on how particles are colliding (top figures). Arrows (top figures) indicate the tentative movement of one of the particles while the other one, at the origin of the coordinate system, is considered to stay at rest. Bottom figures show how the potential responds depending on the collision configuration.

Simulations are performed placing N particles at random inside a box of surface V with periodic boundary conditions. For more details about this model see Peruani et al. [15].

14.3 Results

Starting from a random initial condition and provided bacteria are sufficiently elongated, the system evolves, in both models, towards a steady state in which the microorganisms move in swarms. The length-to-width aspect ratio of bacteria κ turns out to be a key parameter that controls the level of clustering in the system for a given density of cells.[8] In what follows we discuss in more detail how the "microscopic" rules in both models lead to the emergent "macroscopic" patterns observed in the simulations.

Fig. 14.3 illustrates how in both models the local interaction between two individuals can cause local alignment. The active directed movement of the two cells plus volume exclusion force bacteria to become locally aligned and to point in the same direction. This local arrangement of cells lasts for some characteristic time which depends mainly on the magnitude of the active force and the length of the particles. During this period a two-bacteria cluster can eventually incorporate a third bacterium upon a similar collision process. By repetition of this process an initially small cluster can evolve into a big swarm. We stress that in contrast to Brownian particles, i.e., particles moving passively in a stochastic manner, the center of mass of the swarms also moves at roughly the same characteristic speed of the particles. The aligned orientation of the particles inside swarms implicitly results in aligned moving directions too. Fig. 14.4 demonstrates that these simple interactions cause self-organized migration of bacteria in swarms. Fig. 14.5 shows that particles cannot form moving clusters at low density and short aspect ratio.

To characterize these emergent patterns we measure the level of clustering in the system by use of the normalized mean maximum cluster size ψ. The choice of ψ to quantify clustering effects is arbitrary and other quantities could have been chosen as well. However, since ψ allows us to characterize the clustering tendency, as well as to compare both models, it constitutes a reasonable option. For a more detailed analysis

---

[8] κ is defined as $\kappa = L/W$, where L is the length and W the width of the bacterium.

(a) Alignment in the on-lattice model   (b) Alignment in the off-lattice model
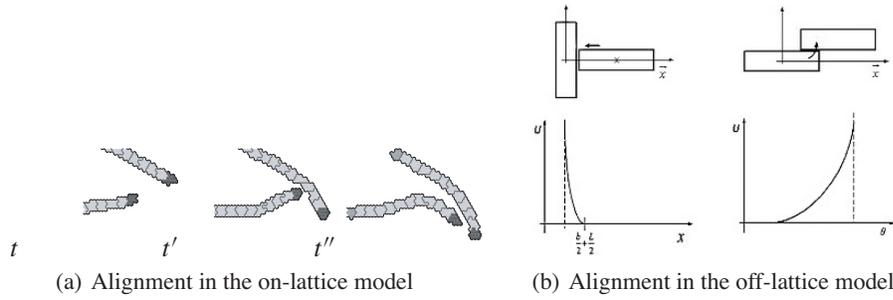
Fig. 14.3. Mechanical collision scheme. Interaction between two entities can lead to alignment in both models, while the resulting moving direction can be different in the models.

of the clustering dynamics and characterization, we refer the reader to [15]. $\psi = 1$ indicates that all cells in the system form a huge cluster which contains all particles in

(a) initial and final state in off-lattice model

(b) initial and final state in on-lattice model
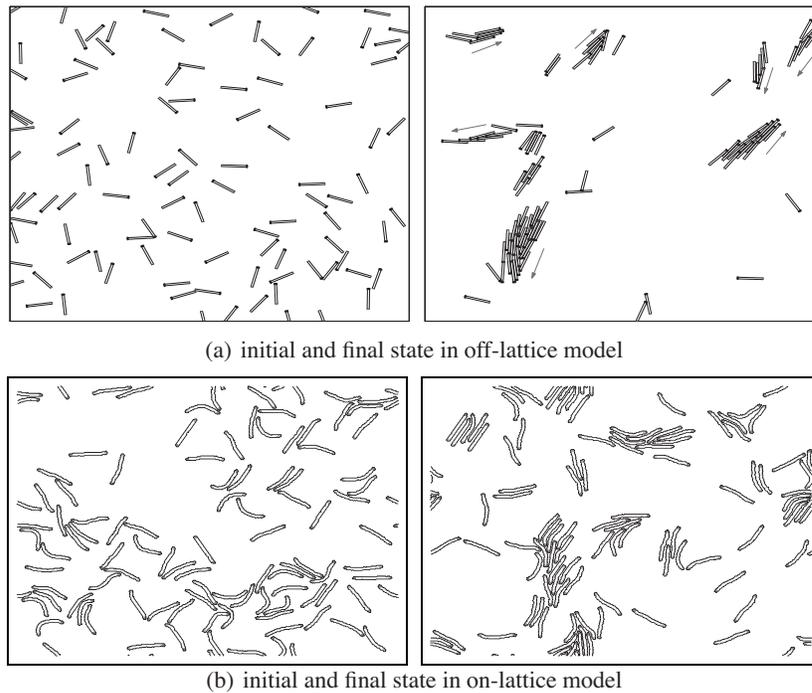
Fig. 14.4. Snapshots of collective movement in both models at low density. Starting from a random initial condition, in both models swarms evolve after a short time of simulation. Particles' length-to-width aspect ratio $\kappa \approx 10$. "Heads" are marked with black dots.
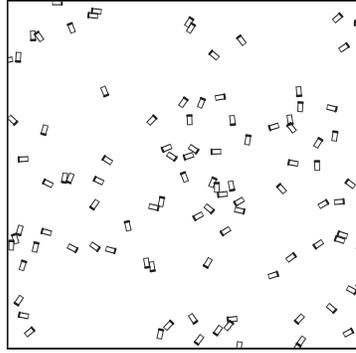
Fig. 14.5. No swarm formation is observed for aspect ratios κ lower than 3 (off-lattice model, $\kappa = 2$). "Heads" of particles are marked with black dots.

the system, while $\psi \longrightarrow 0$ indicates "no clustering" in the system,[9] i.e., all particles are isolated. ψ is calculated in the following way. For each time step we identify all clusters in the system, and we determine the size of the biggest one. Then we simply average these quantities to obtain ψ, i.e., we perform a time average. Therefore, the key point in the calculation of ψ is the identification of clusters.

In the off-lattice model we consider that two individuals are neighbors, and that consequently they also belong to the same cluster, if they are in physical contact. We refer to the mean maximum cluster size ψ calculated in this way as Ψ. On the other hand, for the on-lattice model we consider two bacteria as neighbors if they fulfill the following condition:

$$D_{max} > \min \left\{ |\mathbf{S}_{\alpha,1} - \mathbf{S}_{\beta,1}|, |\mathbf{S}_{\alpha,1} - \mathbf{S}_{\beta,2}|, \dots, \right.$$
$$\left. |\mathbf{S}_{\alpha,1} - \mathbf{S}_{\beta,s}|, |\mathbf{S}_{\alpha,2} - \mathbf{S}_{\beta,1}|, \dots, |\mathbf{S}_{\alpha,s} - \mathbf{S}_{\beta,s}| \right\} - W \qquad (14.7)$$
$$\varphi_{max} > \measuredangle \left(\widetilde{\theta}_{\alpha,1}, \widetilde{\theta}_{\beta,1} \right), \qquad (14.8)$$

where D_{max} and φ_{max} are arbitrary constants which are related to the maximum distance and the maximum relative angle of the moving directions between two cells considered as neighbors (see Section 14.2.1). The reason for this condition is twofold. On one hand, we have to deal with the discreteness of the space, the lattice, and on the other hand, in the on-lattice model there are always nodes representing the medium between any two cells. The condition on $\widetilde{\theta}$, the direction of motion of the bacterium, allows us to determine whether the distance between the cells might be shortened in the next integration step. When both conditions, (14.7) and (14.8), are fulfilled at the same time we say that the two particles are neighbors. We call the directed mean maximum cluster size calculated in this way, $\overline{\Psi}$.

---

[9] To be more specific, $\psi \longrightarrow 1/N$, which means that only for $N \longrightarrow \infty$ does this quantity go to 0.

(a) On-lattice model

(b) Off-lattice model
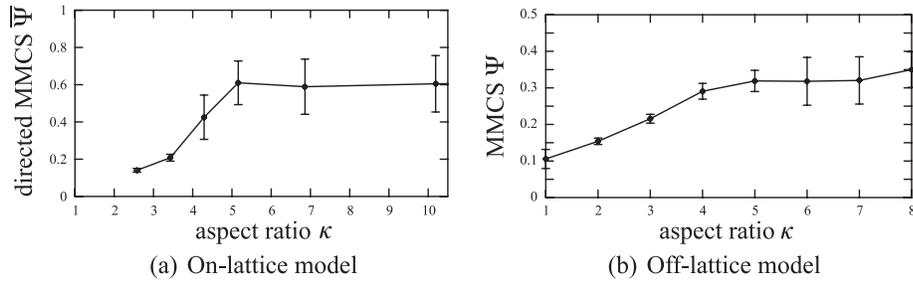
Fig. 14.6. The temporal mean of the size of the largest cluster, measured by $\overline{\Psi}$ in a) and by Ψ in b), is shown in dependence of the particle's length-to-width aspect ratio κ. A transition of non-swarming to swarming behavior can be seen for $\kappa \approx 4$, while the intensity of swarm formation saturates with higher κ. This saturation is counterintuitive, since longer cells could be expected to have a higher chance to collide and align. Error bars indicate the standard deviation of 12 simulations per data point.

Through Ψ and $\overline{\Psi}$ we observe that in both models the mean maximum cluster size exhibits qualitatively the same behavior with respect to the aspect ratio κ.[10] Fig. 14.6 shows that ψ has two regimes: a fast increase for low values of κ followed by a "saturation" regime for large values of κ. Notice that the saturation is not due to the simple fact that ψ assumes its maximum possible value, but to a nontrivial relation between κ and the clustering effects. For a given cell density[11] and a fixed "noise amplitude"[12] and for each value of κ an equilibrium between creation and disintegration of clusters is reached. This equilibrium, as is shown in Fig. 14.6, is strongly influenced by κ.

All this indicates, on one hand, that κ regulates clustering effects. On the other hand, it shows that there is a minimum κ for which the maximum level of clustering is reached (see also Fig. 14.5).

14.4 Conclusions

In the first model, the on-lattice model, we have assumed that the entities representing bacteria have a preferred volume, length, and width, and like to stay unbent. Interactions between different bacteria are based on volume exclusion, i.e., a node belonging to a particular bacterium cannot be occupied by any other bacterium. The resulting patterns resemble those of real bacteria [8], and swarm formation is obtained for a set of parameters. For instance, it is observed that the bacterial clustering tendency increases

---

[10] We stress that when κ was varied, the relation between the area covered by cells and the total area of the box in which cells move was kept constant.

[11] The relevant quantity in fact is the area covered by cells with respect to the total surface, which is proportional to the number of particles.

[12] In the on-lattice model the noise amplitude is intrinsic and can be affected by all parameters.
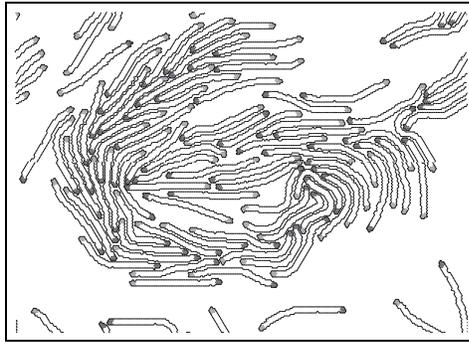
Fig. 14.7. Vortex pattern formation in the on-lattice model. The collision of swarms can result in the formation of such a "recursively bound" swarm.

with the length, but only up to a certain critical length, after which the clustering saturates or even decreases.[13] This was particularly evidenced through the response of the mean maximum cluster size to the length-to-width aspect ratio. But which model ingredients are responsible for such behavior? To which extent does this depend on particle shape? Is this an artifact of the discrete cellular Potts model?

To answer this question we have introduced the second model. We have represented bacteria by particles which have a certain shape with a well-defined aspect ratio. Particles are equipped with a self-propelling force, move off-lattice, and interact when overlapping, i.e., volume exclusion is also present. Surprisingly, we observed that particle clustering is governed by particle shape in a very similar way to what we observed in the cellular Potts model. The mean maximum cluster size exhibits the two regimes which were also reported in the on-lattice model (See Figs. 14.6 a and b). This means that by replacing the elastic active objects used in the on-lattice model with non-deformable self-propelled objects which move off-lattice, and with replacing the energy-based interactions with simple Newtonian force balance laws, we obtain basically the same behavior. We can go even further in the abstraction and show that the main characteristics of the clustering process can be captured by a mean field approach (introduced in [15]), which indicates that the onset of swarming is determined by a κ_c which depends on the relation between the area covered by bacteria and the total surface of the box in which bacteria move. Conclusion: Neither the bending property of particles nor the particular choice of the automaton Hamiltonian is responsible for swarm formation, but instead the simple concept of volume exclusion in a system of asymmetric active objects.

From the intensities of swarm formation measured in simulations of the two models, we conclude that, regardless of the model, the asymmetric shape of self-propelled individuals interacting by volume exclusion can lead to swarm formation. These find-

---

[13] This was tested by fixing the value of parameters in the swarming range and varying only the length of particles, but keeping the fraction of nodes occupied by particles and total nodes of the system constant.

ings are particularly relevant for gliding bacteria where the bacterial shape may play a fundamental role in the pattern formation. For instance, myxobacteria are able to move in swarms, though they have a cell-to-cell signaling mechanism which only controls the speed and reversal frequency of the motion direction of the cells. Inside swarms myxobacteria are aligned in arrangements that resemble those shown in Fig. 14.4. Moreover, *Myxococcus xanthus* cells exhibit an aspect ratio around 5, corresponding to the value where the plateau in Fig. 14.6 is reached.

All this suggests that in collective movement of myxobacteria the shape of the cells may play a fundamental role. To test this idea, experiments with myxobacteria mutants, which exhibit a much simpler behavior than the wild-type myxobacteria, and closer to our hypothesis, could be performed. For instance, there are ongoing experiments with mutants which have no adhesion and no cell reversal.

Beyond these considerations, the models described here exhibit a richer repertoire of patterns which includes streams and vortex formation (Fig. 14.7). However, the characterization of those patterns is the subject of future research.

Acknowledgments

References

1. Bussemarker, J.H., Deutsch, A., Geigant, E.: Mean-field analysis of a dynamical phase transition in a cellular automaton model for collective motion. Phys. Rev. Lett., **78**, 5018–5021 (1997).
2. Czirók, A., Vicsek, T.: Collective behavior of interacting self-propelled particles. Physica A, **281**, 17–29 (2000).
3. Deutsch, A., Dormann, S.: Cellular Automaton Modeling of Biological Pattern Formation—Characterization, Applications, and Analysis. Birkhäuser, Boston (2005).
4. Doi, M., Edwards, S.F.: The Theory of Polymer Dynamics. Oxford University Press, London (1986).
5. Dworkin, M.: Recent advances in the social and developmental biology of the myxobacteria. Microbiol. Rev., **60**, 70–102 (1996).
6. Graner, F., Glazier, J.A.: Simulation of biological cell sorting using a two-dimensional extended Potts model. Phys. Rev. Lett., **69**, 2013–2016 (1992).
7. Igoshin, O.A., Welch, R., Kaiser, D., Oster, G.: Waves and aggregation patterns in myxobacteria. Proc. Natl. Acad. Sci. U.S.A., **101**, 4256–4261 (2004).
8. Jelsbak, L., Søgaard-Andersen, L.: Pattern formation: fruiting body morphogenesis in *Myxococcus xanthus*. Curr. Opin. Microbiol., **3**, 637–642 (2000).
9. Kaiser, D., Welch, R.: Dynamics of fruiting body morphogenesis. J. Bacteriol., **186**, 919–927 (2004).
10. Koch, A., White, D.L.: The social lifestyle of myxobacteria. Bioessays, **20**, 1030–1028 (1998).

11. Levine, A., Liverpool, T., MacKintosh, F.: Dynamics of rigid and flexible extended bodies in viscous films and membranes. Phys. Rev. Lett., **93**, 038102 (2004).
12. Levine, A., Liverpool, T., MacKintosh, F.: Mobility of extended bodies in viscous films and membranes. Phys. Rev. E, **69**, 021503 (2004).
13. Mareé, S.: From pattern formation to morphogenesis: multicellular coordination in *Dictyostelium discoideum*. Ph.D. thesis, University Utrecht (2002).
14. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculation by fast computing machines. J. Chem. Phys., **21**, 1087–1092 (1953).
15. Peruani, F., Deutsch, A., Bär, M.: Nonequilibrium clustering of self-propelled rods. Phys. Rev.E, **74**, 030904(R) (2006).
16. Spormann, A.M., Kaiser, D.: Gliding movements in *Myxococcus xanthus*. J. Bacteriol., **177**, 5846–5852 (1995).
17. Starruß, J.,Søgaard-Andersen, L., Bley, T., Deutsch, A.: A new mechanism for collective migration in *Myxococcus xanthus*. J. Stat. Phys., (2007).
18. Stevens, A.: A stochastic cellular automaton modeling gliding and aggregation of myxobacteria. SIAM J. Appl. Math., **61**, 172–182 (2000).

15

Stability Properties of Some Tissue-Growth Models

John R. King[1] and Susan J. Franks[2]

[1] Centre for Mathematical Medicine, School of Mathematical Sciences, University of
   Nottingham, Nottingham NG7 2RD, UK; john.king@nottingham.ac.uk
[2] Health and Safety Laboratory, Harpur Hill, Buxton SK17 9JN, UK; susan.franks@hsl.gov.uk

Summary. Free boundary problems associated with biological tissue growing under conditions
of nutrient limitation are formulated. Analysis by linear-stability methods, clarifying the models'
stability properties, is then described.

Key words: Stability, tissue growth, nutrient transport, growth-induced stresses.

15.1 Introduction

We are concerned here with the stability properties (particularly those associated with
the interplay between nutrient transport and growth-induced stresses) in some simple
models of tissue growth. We shall launch straight into the dimensionless model formu-
lation, referring to Franks and King [3] and King and Franks [5] (see also Franks [2])
for background and physical meaning of the parameters and references, deferring dis-
cussion of the associated phenomena until they arise.

We define n to be the volume fraction of the growing tissue, and ρ that of the
surroundings, with c being the concentration of a generic nutrient (required for cell
division in particular) and $\mathbf{v}$ being the internal velocity field generated by growth. We
then set

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{v}) = \hat{K}(c)n, \qquad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho\mathbf{v}) = 0,$$

$$\nabla^2 c = \hat{k}(c)n, \qquad n + \rho = 1, \tag{15.1}$$

where nutrient transport can be taken to be quasi-steady and we shall for simplicity
take the mitotic rate $\hat{K}(c)$ and the nutrient consumption rate $\hat{k}(c)$ to be linear, i.e.,
henceforth

$$\hat{K}(c) = Kc, \qquad \hat{k}(c) = kc \tag{15.2}$$

for constants K and k. It is worth noting that equation (15.1) can be rewritten as

$$\frac{\partial n}{\partial t} + \mathbf{v} \cdot \nabla n = K(c)n(1-n),$$

where the saturating (logistic, "Fisher-like") nonlinearity on the right-hand side is an immediate consequence of the way the internal velocity field incorporates the constraints on volume, without needing any additional terms to limit growth and to prevent unphysical interpenetration of different tissue types. It is noteworthy that the additional $-n^2$ term should thus not be included to mimic space limitations; moreover, it is not an appropriate representation of nutrient limitations (pertinent modelling being described in Ward and King [6]).

In one dimension (or with radial symmetry) the system (15.1) represents four equations in four unknowns, n, ρ, $|\mathbf{v}|$ and c. However, in the multidimensional cases with which we are concerned here, additional equations (constitutive assumptions and sometimes momentum equations) are needed to determine the vector $\mathbf{v}$. We shall concern ourselves exclusively with the sharp interface limit in which either $n = 1$ or $\rho = 1$, these two regions being separated by an interface $x = s(y, t)$ (for brevity we confine ourselves here to the two-dimensional case). Then

$$\nabla \cdot \mathbf{v} = Kc, \qquad\qquad \nabla \cdot \mathbf{v} = 0,$$
$$\nabla^2 c = kc, \qquad\qquad \nabla^2 c = 0, \qquad\qquad (15.3)$$

where the left-hand column holds in the growing tissue ($x < s(y, t)$) and the right-hand one in the surroundings ($x > s(y, t)$).

We shall pursue three constitutive assumptions, the first being an artificial one which establishes a benchmark against which the others can be assessed. We shall consider perturbations to one-dimensional growth, whereby (setting $\mathbf{v} = (u, v)$ in (x, y) coordinates) the conditions

$$\text{at } x = 0 \quad u = 0, \qquad \frac{\partial c}{\partial x} = 0 \qquad\qquad (15.4)$$

$$\text{at } x = s \quad [c]_-^+ = \left[\frac{\partial c}{\partial x}\right]_-^+ = 0, \qquad \frac{\partial s}{\partial t} = \left(u - v\frac{\partial s}{\partial y}\right)\Big|_+ = \left(u - v\frac{\partial s}{\partial y}\right)\Big|_- \quad (15.5)$$

will hold in all cases where continuity of nutrient concentration, flux of nutrient and normal velocity on the boundary are assumed. The base solution is

$$c = c_0(x, t), \qquad \mathbf{v} = (u_0(x, t), 0), \qquad s = s_0(t) \qquad\qquad (15.6)$$

so that

$$c_0 = \alpha_0(t)\frac{\cosh(\sqrt{k}x)}{\cosh(\sqrt{k}s_0)}, \quad x < s_0,$$
$$c_0 = \alpha_0(t) + \sqrt{k}\alpha_0(t)(x - s_0)\tanh(\sqrt{k}s_0), \quad x > s_0 \qquad (15.7)$$

and

$$u_0 = \frac{K}{k}\frac{\partial c_0}{\partial x}, \qquad \frac{ds_0}{dt} = \frac{K}{\sqrt{k}}\alpha_0(t)\tanh(\sqrt{k}s_0). \qquad\qquad (15.8)$$

The quantity α_0 is determined by the far-field condition on c, representative examples including

$$c = c_L \text{ at } x = L$$

(in which case the problem formulation can make sense only while $s_0 < L$ and we need to scale $c_L \propto L$ to get a meaningful problem for $L \to \infty$) and the flux prescription

$$\frac{\partial c}{\partial x} \to 1 \text{ as } x \to +\infty.$$

By keeping α_0 ($= c_0|_{x=s_0}$) free we shall be able to address simultaneously a range of infinite domain problems, thus providing an economy of calculation that can be beneficial in a wide variety of linear stability problems.

Our three constitutive assumptions are as follows.

1. One-dimensional growth, $v = 0$.
2. Darcy flow, applicable to growth within a porous scaffold. Here

$$\mathbf{v} = -\kappa(n)\nabla p$$

and we set $\kappa = \kappa_-$ for $n = 1$ and $\kappa = \kappa_+$ for $n = 0$ (κ is inversely proportional to the viscosity of the material), and we append to (15.5) the pressure continuity condition

$$\text{at } x = s \quad [p]_-^+ = 0. \tag{15.9}$$

3. Stokes flow, relevant to growth unconstrained by a matrix. In deriving these equations we have assumed Navier–Stokes equations for an incompressible fluid (note that new material is being produced due to cell division, though) with negligible inertia and no body forces; we note that the creation of new material leads to a term of bulk-viscosity type, no such term being present in conventional incompressible flow. It is most convenient for what follows to set this up in terms of the Airy stress function A, whereby

$$\frac{\partial^2 A}{\partial y^2} = -\left(p - \frac{2}{3}\mu(n)\nabla \cdot \mathbf{v}\right) + 2\mu(n)\frac{\partial u}{\partial x},$$

$$-\frac{\partial^2 A}{\partial x \partial y} = \mu(n)\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right),$$

$$\frac{\partial^2 A}{\partial x^2} = -\left(p - \frac{2}{3}\mu(n)\nabla \cdot \mathbf{v}\right) + 2\mu(n)\frac{\partial v}{\partial y}, \tag{15.10}$$

wherein we set $\mu = \mu_-$ for $n = 1$ and $\mu = \mu_+$ for $n = 0$. It is also convenient to define a stream function ψ via

$$u = \frac{\partial \psi}{\partial y} + \frac{K}{k}\frac{\partial c}{\partial x}, \qquad v = -\frac{\partial \psi}{\partial x} + \frac{K}{k}\frac{\partial c}{\partial y}. \tag{15.11}$$

The higher order of this formulation requires that we append more additional conditions to (15.4) and (15.5), namely (assuming symmetry about $x = 0$)

$$\text{at } x = 0 \qquad \frac{\partial v}{\partial x} = 0,$$

$$\text{at } x = s \qquad [A]_-^+ = \left[\frac{\partial A}{\partial x}\right]_-^+ = [\psi]_-^+ = \left[\frac{\partial \psi}{\partial x}\right]_-^+ = 0, \qquad (15.12)$$

the simplicity of the interface stress-continuity conditions here (cf. King [4]) providing a strong motivation for formulating the problem in terms of A.

We are now in a position to describe the stability properties of each of these cases in turn. Setting (as we shall do throughout) $c \sim c_0 + C$, $s \sim s_0 + S$ and

$$C(x, y, t) = \begin{cases} \alpha_-(t)\dfrac{\cosh\left(\sqrt{k + \lambda^2}x\right)}{\cosh\left(\sqrt{k + \lambda^2}s_0\right)} \cos(\lambda y) & \text{for } x < s_0, \\ \alpha_+(t)e^{-\lambda(x - s_0)} \cos(\lambda y) & \text{for } x > s_0, \end{cases} \qquad (15.13)$$

$$S(y, t) = \Sigma(t) \cos(\lambda y), \qquad (15.14)$$

where $\lambda > 0$ is the wavenumber of the perturbation and we are imposing on C as $x \to +\infty$ only that there be no exponential growth (we can of course without loss of generality consider only the cosine Fourier components).

15.2 One-Dimensional Growth

Here

$$u = K \int_0^x c(x', y, t)\, dx' \text{ for } x < s, \qquad \frac{\partial s}{\partial t} = K \int_0^s c\, dx. \qquad (15.15)$$

Omitting the algebra, we then ultimately find

$$\frac{d\Sigma}{dt} = K\alpha_0 \left(1 - \frac{k}{k + \lambda^2 + \lambda\sqrt{k + \lambda^2} \coth\left(\sqrt{k + \lambda^2}s_0\right)}\right) \Sigma. \qquad (15.16)$$

Given that $\alpha_0 > 0$, the stability properties can be readily inferred from (15.16). In particular, for $k = 0$, $c \equiv 1$, say, we have

$$\frac{d\Sigma}{dt} = K\alpha_0\Sigma, \qquad (15.17)$$

which is a growth-driven instability (cf. [3]) whereby regions in which s is largest grow fastest due to their having more tissue dividing behind $x < s$. Nutrient consumption k mitigates this effect (leading to the second term on the right-hand side of (15.16) being negative) since this competitive advantage is lost when the cells that lie significantly behind the interface do not have access to the nutrient; this is in sharp contrast to the

classical diffusion-limited fingering instability to which we shall return. In particular, as $s_0 \to \infty$ we have

$$\frac{d\Sigma}{dt} \sim K\alpha_0 \frac{\lambda(\lambda + \sqrt{k + \lambda^2})}{k + \lambda^2 + \lambda\sqrt{k + \lambda^2}}\Sigma, \tag{15.18}$$

so that perturbations grow for any $\lambda > 0$, but typically more slowly than the tissue itself; perturbations grow fastest in the limit in which the wavenumber $\lambda \to \infty$, since cells in a short-wavelength protuberance have access to relatively more nutrient than those in a long-wavelength one.

15.3 Darcy Flow

In view of space constraints we simply summarise the final result in the remaining cases. We find for Darcy flow that

$$\frac{d\Sigma}{dt} = \frac{K\lambda\kappa_+\alpha_0}{\kappa_- \tanh(\lambda s_0) + \kappa_+}\left(\frac{1 + \tanh(\lambda s_0)}{\lambda + \sqrt{k + \lambda^2}\tanh(\sqrt{k + \lambda^2}s_0)}\right.$$
$$\left. + \left(\frac{1}{\kappa_+} - \frac{1}{\kappa_-}\right)\frac{\kappa_- \tanh(\lambda s_0)\tanh(\sqrt{k}s_0)}{\sqrt{k}}\right)\Sigma. \tag{15.19}$$

Equation (15.19) requires considerable unravelling. First, in the nutrient-rich case $k = 0, c \equiv 1$ we have

$$\frac{d\Sigma}{dt} \sim \frac{K\kappa_+}{\kappa_- + \kappa_+}\left(\left(\frac{1}{\kappa_+} - \frac{1}{\kappa_-}\right)\lambda\kappa_- s_0 + 1\right)\Sigma \tag{15.20}$$

as $s_0 \to \infty$. Hence if $\kappa_+ < \kappa_-$, meaning the growing tissue is less viscous than the surroundings, perturbations grow faster than exponentially; this corresponds to the well-known Saffman–Taylor instability mechanism whereby the expanding tissue tends to follow the path of least resistance, namely through any existing tissue excrescence (thereby enhancing the latter). Moreover, the shorter the wavelength (i.e., the larger the value of λ), the more rapidly the instability grows. In contrast, if $\kappa_- < \kappa_+$, perturbations are strongly suppressed for large s_0. As $k \to \infty$ we instead have (for any s_0, κ_+, κ_-) the significant simplification

$$\frac{d\Sigma}{dt} \sim \frac{K\lambda\alpha_0}{\sqrt{k}}\Sigma. \tag{15.21}$$

For $s_0 \to \infty$ with $k = O(1)$ we have

$$\frac{d\Sigma}{dt} \sim \frac{K\lambda\kappa_+\alpha_0}{\kappa_- + \kappa_+}\left(\frac{2}{\lambda + \sqrt{k + \lambda^2}} + \left(\frac{1}{\kappa_+} - \frac{1}{\kappa_-}\right)\frac{\kappa_-}{\sqrt{k}}\right)\Sigma. \tag{15.22}$$

The first term in brackets on the right-hand side of (15.22) is necessarily positive and corresponds to the nutrient-limited instability mechanism whereby a protruding finger of tissue gains access to more nutrient than the tissue behind it and hence grows yet faster (this mechanism is unmasked most explicitly in (15.21)). The second (viscosity-contrast) term has a similar interpretation to that described above for $K = 0$ and dominates for large λ. An interesting case arises when $\kappa_- < \kappa_+$ since there is then a competition between the stabilizing effects of the second term and the destabilizing ones of the first; the relative (but not absolute) effects of the latter are most pronounced in the long-wavelength limit $\lambda \to 0$ when the bracketed term becomes $(\kappa_- + \kappa_+)/\kappa_+$, implying that there is always a range of unstable wavenumbers. This limit is of course closely related to (15.21).

15.4 Stokes Flow

For Stokes flow we find after considerable algebra that perturbations satisfy

$$\frac{d\Sigma}{dt} = \frac{2K\lambda\mu_-\alpha_0 e^{2\lambda s_0}((\mu_+ + \mu_-)e^{2\lambda s_0} + (\mu_+ - \mu_-))}{(\lambda + \sqrt{k+\lambda^2}\tanh(\sqrt{k+\lambda^2}s_0))\Pi}\Sigma, \quad (15.23)$$

where $\Pi = (\mu_+ + \mu_-)^2 e^{4\lambda s_0} - 4(\mu_+^2 - \mu_-^2)\lambda s_0 e^{2\lambda s_0} - (\mu_+ - \mu_-)^2$.

The linear-stability result (15.23) also warrants a certain amount of disentangling. In the limit $s_0 \to +\infty$ we have

$$\frac{d\Sigma}{dt} \sim \frac{2K\lambda\mu_-\alpha_0}{(\mu_+ + \mu_-)(\lambda + \sqrt{k+\lambda^2})}\Sigma, \quad (15.24)$$

implying that a planar front is unstable, whatever the viscosity contrast. The exception is $\mu_- = 0$ when it is neutrally stable—in this case the velocity in $x > s$ is uniform and one dimensional ($v = 0$, with $u(t)$ determined by the total rate of cell division in $x < s$), so the nonlinear case is tractable. This universal instability is in marked contrast to the Darcy case: in the absence of a scaffold (which in some sense shields the two fluids from one another), the "path-of-least-resistance" argument is irrelevant, with a finger of the growing material being able to advance only by pushing aside the surroundings (and with this resistance becoming more significant as μ_- decreases).

In the nutrient-rich case, $k = 0, \alpha_0 = 1$, the expression (15.24) implies

$$\frac{d\Sigma}{dt} \sim \frac{K\mu_-}{\mu_+ + \mu_-}\Sigma \quad (15.25)$$

and, on comparison with (15.17), we see that the growth-driven instability is mitigated somewhat by the multidimensional flow pattern, except in the limit case $\mu_+ = 0$ (when for $k = 0$ the velocity field in $x < s$ becomes one dimensional in the limit $s_0 \to \infty$, corresponding exactly to the assumptions that led to (15.17)). As $k \to +\infty$, the asymptotic limit of (15.24) reads

$$\frac{d\Sigma}{dt} \sim \frac{2K\lambda\mu_-\alpha_0}{(\mu_+ + \mu_-)\sqrt{k}}\Sigma, \quad (15.26)$$

which should be compared with (15.21); the correspondence is exact when $\mu_+ = \mu_-$ and this can be viewed as expressing the nutrient-limitation mechanism of instability in its cleanest form. The instability is enhanced when $\mu_- > \mu_+$, the extreme case being $\mu_+ = 0$ when

$$\frac{d\Sigma}{dt} \sim \frac{2K\lambda\alpha_0}{\sqrt{k}}\Sigma. \tag{15.27}$$

In the multiplicative factor $(1+(\mu_--\mu_+)/(\mu_-+\mu_+))$ that converts (15.21) to (15.26), the second term can be interpreted as being associated with a buckling instability in which the compressive stresses generated by cell division in the thin proliferating rim cause it to buckle (cf. [1]); an analysis of this phenomenon in the framework of the nonlinear model will be given elsewhere. In contrast, if $\mu_- < \mu_+$, then the more viscous surrounding material seeks to straighten out the advancing interface, with the result that (15.26) implies less rapid growth of perturbations than (15.21).

15.5 Discussion

We have presented what we believe to be by some margin the most comprehensive linear-stability analysis to date of this class of tissue-growth model. In summary, the following distinct instability mechanisms have been identified.

1. *Growth-driven (cf. [3])*, whereby thicker tissue grows faster because there is more dividing tissue behind it.
2. *Diffusion-limited*, associated in this classical mechanism by a faster-growing region gaining access to more nutrient.
3. *Viscosity-contrast-driven (Darcy)*, which occurs when the growing tissue is less viscous than its surroundings and growth tends to follow the path of least resistance.
4. *Buckling due to growth-induced compressive stresses (Stokes)*, which is most pronounced when the growing tissue is more viscous than the surroundings.

We have concerned ourselves here with linear-stability properties. The dependence of growth rates on the wavenumber λ provides some clues regarding nonlinear behaviour, and the strongly nonlinear behaviour is amenable to thin-film modelling approaches (cf. [3]). It is worth mentioning that in some contexts (e.g., 1) nonlinear effects can enhance the growth rate of perturbations, whereas in others (e.g., 3, in which the growth within (the linearly favoured) narrow fingers tends to thicken them) they reduce them. Finally we note that, in the framework of Stokes-flow moving-boundary problems, the buckling instability can be equated with negative surface tension and raises intriguing issues of well-posedness, regularisation and finger selection akin to those which arise in the Hele–Shaw problem without surface tension. Thus, the area raises a considerable number of open problems, both mathematical and biological.

Insight may be gained from *in vitro* studies, for instance, monolayer cultures or multicell spheroids, in which avacular tumours are grown in nutrient-rich media. *In*

vivo studies might include examination of ductal carcinoma *in situ* in which the tumour is confined within a membrane that is impenetrable to blood vessels but permeable to vital nutrients. Such studies directed at determination of the associated parameter values would be invaluable in allowing such models to be applied in a quantitative and predictive fashion.

Acknowledgments

References

1. Drasdo, D.: Buckling instabilities of one-layered growing tissues. Phys. Rev. Lett., **84**, 4244–4247 (2000).
2. Franks, S.J.: Mathematical modelling of tumour growth and stability. Ph.D. Thesis, University of Nottingham (2001).
3. Franks, S.J., King, J.R.: Interactions between a uniformly proliferating tumour and its surroundings: uniform material properties. Math. Med. Biol., **20**, 47–89 (2003).
4. King, J.R.: Mathematical aspects of semiconductor process modelling. Ph.D. Thesis, University of Oxford (1986).
5. King, J.R., Franks, S.J.: Mathematical analysis of some multi-dimensional tissue growth models. Euro. J. Appl. Math. **15**, 273–295 (2004).
6. Ward, J.P., King, J.R.: Mathematical modelling of avascular-tumour growth. J. Math. Appl. Med. Biol. **14**, 39–69 (1997).

A Modified Backward Euler Scheme for Advection-Reaction-Diffusion Systems

Anotida Madzvamuse

Department of Mathematics, University of Sussex, Mantell Building, Brighton, BN1 9RF, England, UK; a.madzvamuse@sussex.ac.uk

Summary. We present a modified first-order backward Euler finite difference scheme to solve advection-reaction-diffusion systems on fixed and continuously deforming domains. We compare our scheme to the second-order semi-implicit backward finite differentiation formula and conclude that for the type of equations considered, the first-order scheme has a larger region of stability for the time step than that of the second-order scheme (at least by a factor of ten) and therefore the first-order scheme becomes a natural choice when solving advection-reaction-diffusion systems on growing domains.

Key words: Advection-reaction-diffusion, moving grid finite elements, finite difference, pattern formation, Schnakenberg model.

16.1 Introduction

A number of models used in biology, ecology and biochemistry comprise reaction of "species" in the presence of diffusion: hence reaction-diffusion systems arise [14]. In biology and bio-medicine reaction-diffusion systems are used frequently to model the emergence of pattern formation, wound healing, cancer and angiogenesis. Most of these equations comprise nonlinear reaction terms which makes it difficult to find solutions in closed form, and therefore numerical solutions are sought. A variety of time-stepping methods have been used to approximate reaction-diffusion models on fixed domains. These include explicit, semi-implicit, and implicit-explicit (IMEX) schemes [4, 11]. It has been shown that first-order time-stepping schemes impose severe restrictions on the time step and that other schemes are not recommended because they give weak damping to high frequency spatial modes [4, 11]. The second-order semi-implicit backward finite differentiation formula (2-SBDF) scheme was found to be the most appropriate and therefore recommended when solving reaction-diffusion systems since the scheme strongly damps high frequency errors [11]. We take this method as the benchmark and compare results to those obtained by using the modified first-order backward Euler finite difference scheme (1-SBEM). The novelty and key difference of our method from other first-order schemes is that we use an implicit

scheme to approximate the linear reaction kinetics, growth and diffusion terms and a semi-implicit scheme for the nonlinear terms, as shown in Section 16.2.1.

16.2 Model Equations

In this chapter we focus on the nondimensionalised advection-reaction-diffusion system on a continuously deforming domain $\Omega(t)$ with Schnakenberg [13] reaction kinetics given by

$$u_t + \nabla.(\mathbf{a}\,u) = \gamma\left(0.1 - u + u^2\,v\right) + \nabla^2 u, \tag{16.1}$$

$$v_t + \nabla.(\mathbf{a}\,v) = \gamma\left(0.9 - u^2\,v\right) + 10\,\nabla^2 v. \tag{16.2}$$

Here $\Omega(t)$ is a time-dependent domain, and $u(x(t), y(t), t)$ and $v(x(t), y(t), t)$ are the two chemical concentrations under investigation. The term $\mathbf{a}$ represents the field flow velocity. γ is a scaling parameter [10]. In all our simulations zero-flux (homogeneous Neumann) boundary conditions are used. These are appropriate for the case of an impermeable growing membrane, for example, or where we wish to explore self-organising processes on moving boundaries. Other types of boundary conditions can be used [4]. Initial conditions are prescribed as small random perturbations about the uniform steady state of the corresponding reaction system. It must be noted that complex dynamics such as periodic motion, quasi-periodicity or chaos are not examined here. Equations (16.1) and (16.2) are more difficult to integrate than the standard reaction-diffusion systems because of the advection terms.

16.2.1 A Modified Backward Euler Finite Difference Scheme

Following [4,5] we integrate in time using a modified backward Euler finite difference scheme by approximating implicitly the growth, linear and diffusion terms in equations (16.1) and (16.2) and semi-implicitly the nonlinear reaction terms as follows:

$$\frac{u^{m+1} - u^m}{\Delta t} + \nabla \cdot (\mathbf{a}u^{m+1}) = \gamma\left(0.1 - u^{m+1} + u^m u^{m+1} v^m\right) + D_u \nabla^2 u^{m+1},$$

$$\frac{v^{m+1} - v^m}{\Delta t} + \nabla \cdot (\mathbf{a}\,v^{m+1}) = \gamma\left(0.9 - u^m\,u^m\,v^{m+1}\right) + D_v\,\nabla^2 v^{m+1}. \tag{16.3}$$

For simplicity we assume that domain growth is uniform and isotropic and that it takes place slowly. For example, on fixed domains the term u^2 can be approximated by $u^m u^{m+1}$ where we assume that two successive approximate solutions u^m and u^{m+1} at successive time steps become closer and closer with time. On continuously deforming domains, domain growth takes place slowly and therefore this assumption still holds. This is the key novelty of our scheme. Observe that we exploit as much as we can the full implicitness of the scheme (see [4,5] for further details).

16.2.2 The Moving Grid Finite Element Method

We formulate the moving grid finite element method (MGFEM) (space integration) applied to equations (16.3). Following [4, 6] let $w \in H^1(\Omega(t))$ be a test function. Multiplying equations (16.3) by w leads to finding $u, \ v \in H^1(\Omega(t))$ such that

$$\left(\frac{u^{m+1} - u^m}{\Delta t}, w \right) + \left(\dot{x} \, u_x^{m+1} + \dot{y} \, u_y^{m+1}, w \right) + \left(\nabla . \mathbf{a} \, u_h^{m+1}, w \right)$$

$$= (\gamma \, 0.1, w) - \left(\gamma \, u^{m+1}, w \right) + \left(\gamma \, u^m \, u^{m+1} \, v^m, w \right) - D_u \left(\nabla u^{m+1}, \nabla w \right),$$

$$\left(\frac{v^{m+1} - v^m}{\Delta t}, w \right) + \left(\dot{x} \, v_x^{m+1} + \dot{y} \, v_y^{m+1}, w \right) + \left(\nabla . \mathbf{a} \, v_h^{m+1}, w \right)$$

$$= (\gamma \, 0.9, w) - \left(\gamma \, u^m \, u^m \, v^{m+1}, w \right) - D_v \left(\nabla v^{m+1}, \nabla w \right), \tag{16.4}$$

for all $w \in H^1(\Omega(t))$ where $(u, w) = \int \int_{\Omega(t)} u \, w \, d\mathbf{x}$ is the L_2-inner product. Note that we have assumed zero-flux (homogeneous Neumann) boundary conditions and applied Green's theorem. Observe that $\mathbf{a} = (\dot{x}, \dot{y})^T$ represents the grid velocity and the terms $\nabla \cdot (\mathbf{a} \, u^{m+1})$ and $\nabla \cdot (\mathbf{a} \, v^{m+1})$ have been expanded appropriately as shown above. Let u_h^m (or similarly v_h^m) be an MGFEM approximation to u (or v) defined by $u_h^m = \sum_{j=1}^n U_j^m(t) \, \phi_j(\mathbf{x}, \xi(t))$, where $U_j^m(t)$ are the solution values at the nodal positions, $\phi_j(\mathbf{x}, \xi(t))$ are the basis functions with compact support and $\xi(t)$ represents the finite element moving grid. The time derivative of u (similarly v) in equation (16.1) (similarly (16.2)) can be expressed in two dimensions [1,3] as

$$\frac{\partial u_h^m}{\partial t} = \sum_{j=1}^n \left(\frac{dU_j^m}{dt} - \dot{x}_j \, u_{h_x}^m - \dot{y}_j \, u_{h_y}^m \right) \phi_j(\mathbf{x}, \xi(t))$$

and therefore the time discretisation needs to be readjusted to take into account the extra terms from the MGFEM. Without loss of generality, taking the test function to be $w = \phi_i, i = 1, \ldots, n$, integrating over the whole domain gives rise to the following system of linear algebraic equations in compact form (see [4,5,7] for specific details):

$$M \frac{\mathbf{U}^{m+1} - \mathbf{U}^m}{\Delta t} - (P + Q) \, \mathbf{U}^{m+1} + \left(\bar{U}_x + \bar{U}_y \right) \mathbf{U}^{m+1} + \nabla \cdot \mathbf{a} \, M \, \mathbf{U}^{m+1}$$

$$= \gamma \left[0.1 \, \mathbf{F} - M \, \mathbf{U}^{m+1} + C(\mathbf{U}^m, \mathbf{V}^m) \mathbf{U}^{m+1} \right] - D_u \, K \, \mathbf{U}^{m+1}, \tag{16.5}$$

$$M \frac{\mathbf{V}^{m+1} - \mathbf{V}^m}{\Delta t} - (P + Q) \, \mathbf{V}^{m+1} + \left(\bar{V}_x + \bar{V}_y \right) \mathbf{V}^{m+1} + \nabla \cdot \mathbf{a} \, M \, \mathbf{V}^{m+1}$$

$$= \gamma \left[0.9 \, \mathbf{F} - C(\mathbf{U}^m, \mathbf{U}^m) \mathbf{V}^{m+1} \right] - D_v \, K \, \mathbf{V}^{m+1}, \tag{16.6}$$

where $\mathbf{U}$ and $\mathbf{V}$ represent the vectors of the solutions at the nodal grid points. M is the mass matrix, K is the stiffness matrix, $\mathbf{F}$ is the force vector, $P, Q, \bar{U}_x, \bar{U}_y, \bar{V}_x, \bar{V}_y$ are resultant matrices from domain growth and $C(\mathbf{U}, \mathbf{V})$ is the linearised matrix

corresponding to the term $u^2 v$ [4]. In all our simulations $\nabla \cdot \mathbf{a}$ is calculated from plausible growth functions or those derived from biological experiments and therefore is a known quantity [7]. We can write in matrix form

$$A_u(\mathbf{U}^m, \mathbf{V}^m)\, \mathbf{U}^{m+1} = b_u(\mathbf{U}^m) \quad \text{and} \quad A_v(\mathbf{U}^m, \mathbf{V}^m)\, \mathbf{V}^{m+1} = b_v(\mathbf{V}^m),$$

where the matrices $A_u(\mathbf{U}^m, \mathbf{V}^m)$ and $A_v(\mathbf{U}^m, \mathbf{V}^m)$ are functions of $\mathbf{U}^m$ and $\mathbf{V}^m$ and the right-hand side vectors are given by

$$b_u(\mathbf{U}^m) = M\, \mathbf{U}^m + \gamma\, a\, \Delta t\, \mathbf{F} \quad \text{and} \quad b_v(\mathbf{V}^m) = M\, \mathbf{V}^m + \gamma\, b\, \Delta t\, \mathbf{F}.$$

The matrices A_u and A_v are symmetric and positive definite [4]. In one dimension, these are tri-diagonal matrices which are solved efficiently and fast by using the Thomas algorithm [9]. In two dimensions we use a preconditioned generalised minimum residual method [12].

16.3 Numerical Simulations

We illustrate the stability regions for the 1-SBEM scheme on one-dimensional domains only and compare our results to those obtained by use of the 2-SBDF scheme [4, 11]. Two-dimensional results can be found in [4]. The criterion for convergence to the spatially inhomogeneous steady state and hence a stopping criterion (for the $\mathbf{U}$ component solution, for example) is given by $\sqrt{\sum |\mathbf{U}^{m+1} - \mathbf{U}^m|^2 / \sum |\mathbf{U}^{m+1}|^2} \leq \epsilon$. We call this the relative error or accuracy.

Example 1 Fixed domains: Let us compute solutions on a fixed interval $[0, 1]$ with homogeneous Neumann (zero-flux) boundary conditions at both ends. We fix the scaling parameter $\gamma = 1000$ and vary the time step of both the 1-SBEM and 2-SBDF schemes until they fail to converge. Table 16.3 shows the regions of stability and accuracy for the two schemes. The 1-SBEM scheme is stable and converges to a meaningful result for time-step sizes given by approximately $\Delta t < 6.2 \times 10^{-4}$ while the 2-SBDF scheme has a region of stability given by approximately $\Delta t < 5.71 \times 10^{-5}$. The 1-SBEM scheme allows for larger time steps (by a factor of at least ten) as compared to those required for the 2-SBDF scheme. This is true for all the computations that we have carried out [4]. However, it must be observed that the 2-SBDF scheme outperforms the 1-SBEM scheme in terms of precision as shown: the 2-SBDF scheme is second-order accurate in time while the 1-SBEM is first-order accurate. Although the 1-SBEM scheme has a larger region of stability, in terms of precision it cannot outperform the secord-order scheme as expected. Ruuth [11] demonstrated theoretically and computationally that first-order explicit methods require finer time steps than those required by the 2-SBDF or other higher-order semi-implicit methods. However, this is not the case with our first-order scheme. On growing domains, the fact that the 2-SBDF scheme has a restrictive region of stability makes it computationally prohibitive and expensive for carrying out realistic simulations [4]. Therefore, the 1-SBEM becomes a natural choice.

Table 16.1. The regions of stability and accuracy for the 1-SBEM and 2-SBDF schemes. Computations were carried out to final time $t_F = 2.5$.

1-SBEM		2-SBDF	
Time-step Δt	Accuracy	Time-step Δt	Accuracy
5×10^{-6}	1.575×10^{-16}	5×10^{-6}	1.5×10^{-18}
5×10^{-5}	2.1281×10^{-15}	5×10^{-5}	1.506×10^{-17}
5×10^{-4}	8.185×10^{-14}	5.5×10^{-5}	3.3415×10^{-17}
6×10^{-4}	4.073×10^{-13}	5.6×10^{-5}	3.7562×10^{-17}
6.1×10^{-4}	4.5952×10^{-13}	5.7×10^{-5}	2.8×10^{-17}
6.2×10^{-4}	1.2464×10^{-12}	5.71×10^{-5}	5.72×10^{-17}
6.3×10^{-4}	0.231741	5.8×10^{-5}	0.183827

Example 2 Growing domains: For illustrative purposes we solve the model equations on an exponentially growing interval $x(t) = X(0) e^{\rho t}$—with homogeneous Neumann (zero-flux) boundary conditions—until the interval reaches twice its original size, where the initial interval is $X(0) = [-1, 1]$ and $\rho = 10^{-3}$ is the growth rate. Note that the domain speed is given by $\dot{x}(t) = \rho X(0) e^{\rho t}$. The 1-SBEM scheme converges for $\Delta t = 5 \times 10^{-3}$ while the 2-SBDF scheme requires $\Delta t = 5 \times 10^{-4}$. The 1-SBEM and 2-SBDF schemes produce qualitatively similar results even though the two schemes use different time steps, as shown in Fig. 16.1 for $\gamma = 10, 29$ and 1000. Observe that the transient solutions are different for the two schemes (because of the different time steps), however, the solutions on the final interval are identical and this is due to the moving boundaries.

16.4 Conclusion

We have carried out numerous numerical computations in two dimensions comparing the two schemes and have found out that, even in two dimensions, the region of stability of the time step for the 1-SBEM scheme is at least ten times larger than that of the 2-SBDF scheme [4]. This is true on both fixed and continuously deforming two-dimensional domains. The 2-SBDF scheme applied to advection-reaction-diffusion systems on continuously deforming two-dimensional domains requires Δt to be in the order of 10^{-6}, which is extremely restrictive in terms of numerical simulations. In this case the number of time steps required to grow a unit square domain, for example, to say twice its original size is 500 times more than that required by the 1-SBEM scheme [4]. The fact that our scheme allows for larger time steps makes it more suitable when solving advection-reaction-diffusion problems on continuously deforming domains. In this scenario, the number of time steps required by the 2-SBDF scheme is at least ten times more than that required by the 1-SBEM scheme. In most biologi-

(a)   (b)   (c)

(d)   (e)   (f)

(g)   (h)   (i)
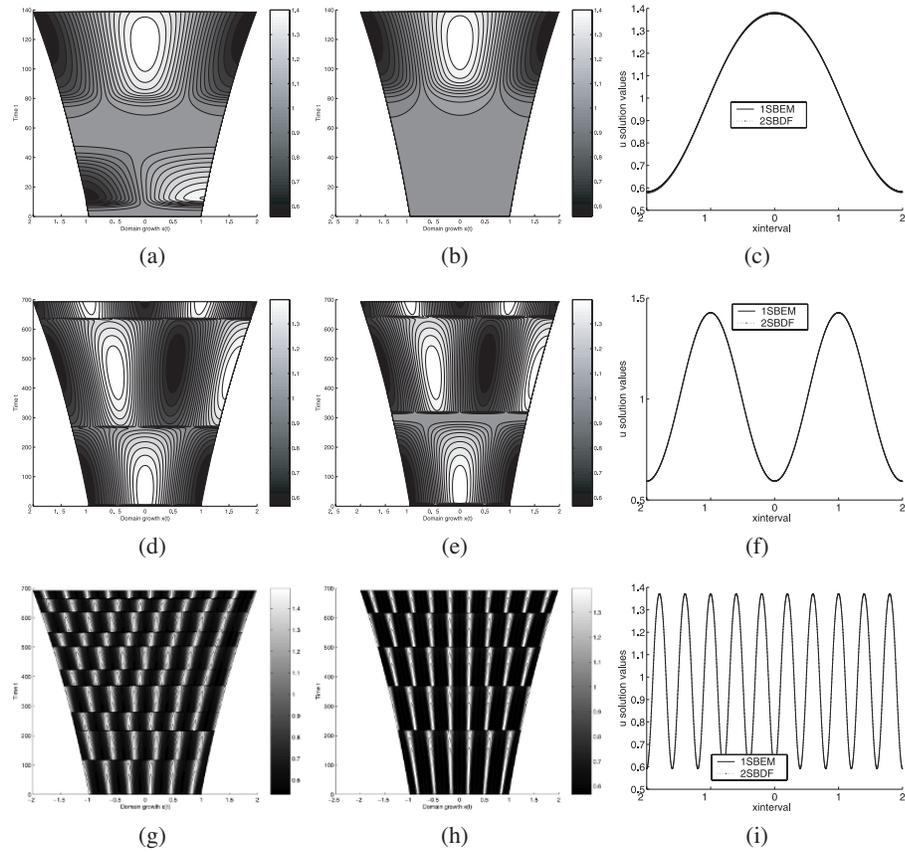
Fig. 16.1. Moving grid finite element results computed by the 1-SBEM and 2-SBDF schemes for $\gamma = 10$ ((a)–(c)), $\gamma = 29$ ((d)–(f)) and $\gamma = 1000$ ((g)–(i)). Solutions (c), (f) and (i) are plots of the u values when the interval reaches domain length $[-2, \, 2]$.

cal problems, domain growth takes place on a very slow time scale, hence the growth rate is very small. In multiple dimensions it becomes computationally prohibitive and too expensive to use the 2-SBDF scheme since it requires time steps of the order of millions when domain growth is involved [4]. However, in the case of highly oscillatory solutions, first-order schemes have been shown [4,11] to fail to resolve the highest growing mode and our scheme is no different. In this case, the 2-SBDF outperforms the first-order scheme as expected [4,11]. On fixed domains, the choice of which scheme to use depends on the nature of the problem: if stationary solutions are sought, then the 1-SBEM scheme is recommended. However, in cases of highly oscillatory solutions, the 2-SBDF scheme must be used, since the 1-SBEM fails to damp the high frequency errors associated with highly oscillatory solutions [11]. We only plotted results for the u solution; those of v are 180 degrees out of phase to those of u. We have applied the 1-SBEM scheme to other advection-reaction-diffusion systems with different reaction

kinetics [2,5,6]. The end product of this research has culminated in the development of a package for MGFEM applied to generalised advection-reaction-diffusion systems on fixed and continuously deforming domains. This package is freely available and downloadable from the website: http://www.maths.sussex.ac.uk/~anotida/software.php.

Acknowledgments

References

1. Carlson, N.N., Miller, K.: Design and application of a gradient-weighted moving finite element code II: In two dimensions. SIAM J. Sci. Comput., **19**, 766–798 (1998).
2. Crampin, E.J., Gaffney, E.A., Maini, P.K.: Reaction and diffusion on growing domains: Scenarios for robust pattern formation. Bull. Math. Biol., **61**, 1093–1120 (2002).
3. Jimack, P.K., Wathen, A.J.: Temporal derivatives in the finite element method on continuously deforming grids. SIAM J. Numer. Anal., **28**, 990–1003 (1991).
4. Madzvamuse, A.: Time-stepping schemes for moving grid finite elements applied to reaction-diffusion systems on fixed and growing domains. J. Comp. Phys., **24**, 239–263 (2005).
5. Madzvamuse, A., Maini, P.K., Wathen, A.J.: A moving grid finite element method for the simulation of pattern generation Turing models on growing domains. J. Sci. Comp., **24**, 247–262 (2005).
6. Madzvamuse, A., Maini, P.K., Wathen, A.J.: A moving grid finite element method applied to a model biological pattern generator. J. Comp. Phys., **190**, 478–500 (2003).
7. Madzvamuse, A., Thomas, R.D.K., Maini, P.K., Wathen, A.J.: A numerical approach to the study of spatial pattern formation in the Ligaments of Arcoid Bivalves. Bull. Math. Bio., **64**, 501–530 (2002).
8. Miller, K., Miller, R.N.: Moving finite elements. Part I. SIAM J. of Num. Anal., 1019–1032, **18** (1981).
9. Morton, K.W., Mayers, D.F.: Numerical Solution of Partial Differential Equations. Cambridge University Press (1994).
10. Murray, J.D.: Mathematical Biology. II: Spatial Models and Biomedical Applications. Springer-Verlag, Heidelberg, NY (2002).
11. Ruuth, S.: Implicit-explicit methods for reaction-diffusion problems in pattern-formation. J. Math. Biol., **34**, 148–176 (1995).
12. Saad, Y.: Iterative Methods for Sparse Linear Systems. SIAM, 2nd Edition, (2003).
13. Schnakenberg, J.: Simple chemical reaction systems with limit cycle behaviour. J. Theor. Biol., **81**, 389–400 (1979).
14. Turing, A.M.: The chemical basis of morphogenesis. Phil. Trans. Roy. Soc. Lond., B **237**, 37–72 (1952).

Part IV

Biomedical Applications

Fractional Transport of Cancer Cells Due to Self-Entrapment by Fission

Alexander Iomin

Department of Physics, Technion, Haifa 32000, Israel; iomin@physics.technion.ac.il

Summary. A simple mathematical model is proposed to study the influence of cell fission on transport. The model describes fractional tumor development, which is a one-dimensional continuous time random walk (CTRW). An answer to the question of how the malignant neoplasm cells appear at an arbitrary distance from the primary tumor is proposed. The model is a possible consideration for diffusive cancers as well. A chemotherapy influence on the CTRW is studied by an observation of stationary solutions.

Key words: Tumor, fractional transport, cell fission, self-entrapment, comb model.

17.1 Introduction

Cancer is a complex disease which leads to the uncontrolled growth of abnormal cells, destruction of normal tissues, and invasion of vital organs. The malignant neoplasm cells spread through vascular or lymphatic vessels which disseminates the disease with further lesion of vital organs. Mathematical modeling of tumor development is mainly aimed at diagnostics and treatments of cancers, since it may lead to the reduction of expensive experiments *in vivo*. Recent reviews describe different aspects of the modeling of tumors including solid tumors [14, 20] interacting with the immune system [3], diffusive models related, e.g., to brain tumors [21], the process of tumor-induced vascularization [11, 14, 15] and lymphangiogenesis [11], the fractal geometry of the pathological architecture of tumors [1, 5], and chemotherapy strategies [3, 21].

There are different stages of tumor development with varying duration, starting from genetic changes on the cell level and finishing with detachment of metastasis and invasion. The transport of tumor cells and their proliferation are the main contributors to the malignant neoplasm dissemination (see, e.g., [3, 21]). Interplay between these two main processes of cell proliferation and transport leads to an essential complication in the mathematical modeling of the tumor growth [3, 15]. This evolution related to the collective or macroscopic behavior of cells is described (in many cases) by kinetic cellular theory [18] (see also, e.g., [3, 21]). For example, different aspects of solid tumor development including kinetic cellular theory have been described in [3]. It is

worth mentioning that at a certain stage of tumor development the processes of angiogenesis and lymphangiogenesis are the most important biological factors for the transport and invasion of cancer cells and metastasis [11, 14, 15]. In this connection, the fractional transport can be an important consequence of these biological phenomena. The most relevant approach for the present consideration is a phenomenological approach to diffusive cancer which is exemplified by glioma modeling [21]. In this approach, the macroscopic behavior of tumor cells is described by the diffusion equation

$$\frac{\partial P}{\partial t} + \nabla \cdot \mathbf{j} = \mathcal{C}P - G(t)P\,, \tag{17.1}$$

where $P \equiv P(\mathbf{x}, t)$ is the tumor cell density or probability distribution function (PDF) at the space-time location $(\mathbf{x}, t)$, and $\mathcal{C}$ is a net proliferation rate, while $G(t)$ is the loss of tumor cells due to chemotherapy. The divergence term $\nabla \cdot \mathbf{j} = \nabla \cdot [D(\mathbf{x})\nabla P(\mathbf{x})]$ implies inhomogeneous diffusion with the diffusion coefficient $D = D(\mathbf{x})$, which is determined by the inhomogeneous properties of tissues. The initial condition $P(\mathbf{x}, t = 0)$ defines the initial spatial distribution. The Dirichlet boundary condition $P = 0$ is imposed, indicating the absence of cancer cells outside the brain. Since the divergence term is a particular case of a continuous time random walk (CTRW), some generalization of cell transport can be performed as well [7,8].

In this chapter, we focus primarily on the influence of cell fission on the transport properties of cells. We propose a simple mathematical model of a CTRW by virtue of two time scales of tumor development [7,8]. We consider a collective behavior of cells, paying particular attention to the influence of tumor cell fission on transport. This leads to an essential decrease in cell motility during fission time or self-entrapment that is determined by the interaction of cells with their environment. The simplest realization of the model is a modification of the comb structure [7,22]. It was shown, for this model, that tumor development corresponds to fractional transport. We note that the mathematical apparatus of the fractional CTRW is well established for many applications in physics (see, e.g., [6,16]). Using this simplified approach of fractional transport, a possible answer to the question of how the neoplasm cells appear arbitrarily far from the main (primary) tumor in the case of a solid tumor was proposed. The model can be considered as a possible approach to diffusive cancers as well.

17.2 Fractional Mechanism of Tumor Development: The CTRW Equation

We consider a simplified scheme of cell dissemination through the vessel network in the following two steps. The first step is a biological process of cell fission. The duration of this stage is $\mathcal{T}_f$. The second process is cell transport itself with duration $\mathcal{T}_t$. Therefore, the cell dissemination is approximately characterized by the fission time $\mathcal{T}_f$ and the transport time $\mathcal{T}_t$. During the time scale $\mathcal{T}_f$ the cells interact strongly and motility of the cells is small, and we suppose that there is no transport (approximately). The

duration of $\mathcal{T}_f$ could be arbitrarily large. During the second time $\mathcal{T}_t$ interaction between the cells is weak and motility of the cells is determined by the velocity V of either vascular or lymphatic flow through the vessel network. It is convenient to introduce a "jump" length X_t as a distance which a cell travels during the time $\mathcal{T}_t$ $X_t = V\mathcal{T}_t$. Hence, the cells form an initial packet of free spreading particles, and the contribution of cell dissemination to the tumor development process consists of the following time consequences: $\mathcal{T}_f(1)\mathcal{T}_t(2)\mathcal{T}_f(3)\dots$. There are different realizations of this chain of times, due to different durations of $\mathcal{T}_f(i)$ and $\mathcal{T}_t(i)$, where $i = 1, 2, \dots$. Therefore, one concludes that transport is characterized by random values $\mathcal{T}(i)$ which are waiting (or self–entrapping) times between any two successive jumps of random length $X(i)$. This phenomenon is known as a continuous time random walk (CTRW) [17]. It arises as a result of a sequence of independent identically distributed random waiting times $\mathcal{T}(i)$, each having the same PDF $w(t), t > 0$ with a mean characteristic time T and a sequence of independent identically distributed random jumps, $x = X(i)$, each having the same PDF $\lambda(x)$ with the jump length variance σ^2. It is worth mentioning that a cell carries its own trap, by which it is set apart from transport. This process of self-entrapment is the main difference between this and the standard CTRW, where traps are external with respect to the transporting particles.

Now we introduce the PDF $P(x, t)$ of the particle at the space-time location (x, t). Suppose that $P(x, t)$, $w(t)$, $\lambda(x)$ are well-behaved functions, such that the Fourier–Laplace $(\hat{F}\hat{L})$ transforms can be applied $\tilde{w}(s) = \hat{L}[w(t)]$; $\tilde{\lambda}(k) = \hat{F}[\lambda(x)]$; $\tilde{P}(k, s) = \hat{F}\hat{L}[P(x, t)]$. Then, due to the probabilistic description that defines an appropriate relation between these three PDFs, $P(x, t)$, $w(t)$, $\lambda(x)$ in the Fourier–Laplace space (see, e.g., [4,16]), one obtains the Montroll–Weiss equation [4,16]:

$$\tilde{P}(k, s) = \frac{1 - \tilde{w}(s)}{s} \cdot \frac{1}{1 - \tilde{w}(s)\tilde{\lambda}(k)} \,, \tag{17.2}$$

where the initial condition $P(x, 0) = \delta(x)$ was used as well. This result, described by Eq. (17.2), establishes a link between the tumor development and the CTRW process. There are different realizations of $w(t)$ and $\lambda(x)$ as a result of different relations between $\mathcal{T}_f$ and $\mathcal{T}_t$. For example, when $\mathcal{T}_f \sim \mathcal{T}_t$, and both T and σ^2 are finite, the solution of Eq. (17.2) is $\tilde{P}(k, s) = 1/(s + Dk^2)$, which corresponds to normal diffusion with diffusion coefficient $D = \sigma^2/T$. In this case, the relevant description of the CTRW is in the framework of Eq. (17.1). Here we consider another relation between $\mathcal{T}_f$ and $\mathcal{T}_t$: namely, $\mathcal{T}_f \gg \mathcal{T}_t$. This case corresponds to the CTRW with subdiffusion. It differs essentially from Eq. (17.1) and will be considered in this chapter.

We consider a situation when σ^2 is finite, while T diverges and is described by a long-tailed waiting time PDF with the asymptotic behavior $w(t) \sim A_\alpha (T/t)^{1+\alpha}$, $0 < \alpha < 1$. The asymptotic behavior of the Laplace transform is $\tilde{w}(s) \sim 1 - (sT)^\alpha$ [12]. Therefore, Eq. (17.2) reads

$$\tilde{P}(k, s) = 1/(p + K_\alpha s^{1-\alpha} k^2) \,, \tag{17.3}$$

where the generalized diffusion constant is now $K_\alpha = \sigma^2/T^\alpha$. The mean squared displacement (MSD) is calculated from (17.3) via the following relation (see, e.g.,

[16]):

$$\langle x^2(t)\rangle = \hat{L}^{-1}\lim_{k\to 0}[-(d^2/dk^2)\tilde{P}(k,s)],$$

where $\hat{L}^{-1}$ denotes the Laplace inversion. It results in the subdiffusive MSD

$$\langle x^2(t)\rangle = 2K_\alpha t^\alpha/\Gamma(1+\alpha), \tag{17.4}$$

where $\Gamma(z)$ is a gamma function [9] and $0 < \alpha < 1$.

To obtain the fractional or CTRW equation which produces subdiffusion solution (17.3), we use here the Riemann–Liouville fractional derivatives (see, for example, [13,16] and Appendix)

$$\left[\frac{\partial^\alpha}{\partial t^\alpha}f(t)\right]_{RL} \equiv D_{RL}^\alpha f(t).$$

Employing the Laplace transform for fractional integrals [16,19]

$$\hat{L}\left[D_{RL}^{-\alpha}f(t)\right] = s^{-\alpha}\tilde{f}(p), \tag{17.5}$$

we present the CTRW equation (17.3) in the following form:

$$\frac{\partial P}{\partial t} + D_{RL}^{1-\alpha}\hat{\mathcal{L}}_{FP}(x)P = 0, \tag{17.6}$$

where the Fokker–Planck operator is diffusion $K_\alpha(\partial^2/\partial x^2)$. In what follows we extend our consideration to a more general form of inhomogeneous diffusion,

$$\hat{\mathcal{L}}_{FP}(x) = -\frac{\partial}{\partial x}\tilde{D}(x)\frac{\partial}{\partial x}. \tag{17.7}$$

For $\alpha = 1/2$ the analysis is simplified, and traps can be modeled by normal diffusion in the additional y direction. In this case the fractional equation (17.6) corresponds to a comb model [2,22].

17.3 Comb-Like Model with Proliferation

Fractional transport of cells, namely subdiffusion, can be described in the framework of the comb model (or CTRW structure) [22]. The comb model is an example of sub-diffusive one-dimensional (1D) media where CTRW takes place along the x structure axis. Diffusion in the y direction plays the role of traps with the PDF of delay times of the form $w(t) \sim 1/(1+t/T)^{3/2}$. A special behavior of diffusion on the comb structure is that the displacement in the x direction is possible only along the structure axis (x axis at $y = 0$). Therefore, cell motility is highly inhomogeneous in the x direction, while the diffusion coefficient in the transversal y direction is a constant $D_{yy} = D_0$. A random walk on the comb structure, which is relevant to Eqs. (17.6) and (17.7), is described by the distribution function $P_1 = P_1(x, y, t)$ and the current

$$\mathbf{j} = \left(-\delta(y)D(x)\frac{\partial P_1}{\partial x}, -D_0\frac{\partial P_1}{\partial y} \right),$$

where

$$P(x,t) = \int_{-\infty}^{\infty} P_1(x,y,t)dy. \tag{17.8}$$

The Liouville equation

$$\frac{\partial P_1}{\partial t} + \nabla \cdot \mathbf{j} = C(P_1) \tag{17.9}$$

yields the following Fokker–Planck equation:

$$\frac{\partial P_1}{\partial t} + \delta(y)\mathcal{L}_{FP}(x)P_1 - D_0\frac{\partial^2 P_1}{\partial y^2} = C(P_1), \tag{17.10}$$

with the initial conditions $P_1(x,y,0) = \delta(x)\delta(y)$ and the boundary conditions on the infinities $P_1(t) = P_1'(t) = 0$. The primes denote spatial derivatives. The relation between K_α in (17.3), (17.4) and D, D_0 can be found from MSD calculation; for example, when $\mathcal{L}_{FP}(x) = D(\partial^2/\partial x^2)$, then $K_\alpha = D/2\sqrt{D_0}$, and for $D = K_\alpha$, $D_0 = 1/4$.

In the rest of this section we consider a possible mechanism of tumor cell proliferation. The term $C(P_1)$ in Eqs. (17.9) and (17.10) determines the change in the total number of transporting cells due to proliferation at rate $\mathcal{C}$. According to the CTRW model, the transporting cells along the x axis do not proliferate. It means that cells proliferate only if they have a non-zero y coordinate. We introduce the proliferation rate as a change in the total number of cells with time,

$$dN(t)/dt = \mathcal{C}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dydx\,[P_1(x,y,t) - \delta(y)P_1(x,y,t)]. \tag{17.11}$$

Therefore, $C(P_1) = \mathcal{C}(1 - \delta(y))P_1$.

The first term in the right-hand side of Eq. (17.10) is eliminated by substitution $P_1 = e^{\mathcal{C}t}F_1$. Carrying out the Laplace transform $\tilde{F}_1(s,x,y) = \hat{L}[F_1(x,y,t)]$ and looking for the solution in the form $\tilde{F}_1 = e^{-\sqrt{s/D_0}|y|}f(x,s)$, it is straightforward to see that

$$F_1(x,y,t) = \hat{L}^{-1}\left[f(x,s)\exp(-\sqrt{s/D_0}|y|)\right]. \tag{17.12}$$

Integrating Eq. (17.10) with respect to the variable y (see Eq. (17.8)), one obtains the following equation for F in the Laplace space $\tilde{F}(s) = \hat{L}[F(t)]$:

$$s\tilde{F} + \hat{\mathcal{L}}_{FP}f = \delta(x) - \mathcal{C}f. \tag{17.13}$$

Integrating Eq. (17.12) over y, we obtain a relation between the PDFs of the total number of cells F and transporting number of cells f in the Laplace space when $\mathcal{C} = 0$:

$$f \equiv \tilde{F}_1(x, y = 0, s) = (1/2)\sqrt{s/D_0}\,\tilde{F}(x, s)\,.$$

Substitution of this relation in Eq. (17.13) yields, after the Laplace inversion, the Fokker–Planck equation for the distribution $F = e^{-\mathcal{C}t}P$. To this end, Eq. (17.13) is multiplied by $\sqrt{s}$ and then by Eq. (17.27) the inverse Laplace transform yields the following equation for $P = e^{\mathcal{C}t}F$:

$$e^{\mathcal{C}t}D_C^{1/2}e^{-\mathcal{C}t}P + \hat{\mathcal{L}}_{FP}P = -\mathcal{C}P\,, \qquad (17.14)$$

where we used a fractional derivative in the Caputo form [13, 19] (see Appendix)

$$I^{1-\alpha}[\partial F(t)/\partial t] = D_C^\alpha F(t)$$

with the following property of the Laplace transform:

$$\hat{L}\left[D_C^\alpha F(t)\right] = s^\alpha \tilde{F}(s) - \delta(x)s^{\alpha-1}$$

and $D_C^\alpha[1] = 0$. The condition $D_0 = 1/4$ is used as well. This equation describes fractional transport of cells with fission. As shown here, the cell fission is a source of the fractional time derivatives. This equation can be extended for an arbitrary fractional exponent $0 < \alpha < 1$, when fission time is determined by a waiting time PDF $w(t) \sim 1/(1 + t/T)^{1+\alpha}$.

17.4 Stationary Solutions Due to Chemotherapy

Application of chemotherapy changes the tumor development. As seen from Eqs. (17.10) and (17.1), when $G(t) = \mathcal{C}$, chemotherapy leads to a decrease in the number of tumor cells and, correspondingly, eliminates tumor development. In reality, a cancer cell is unstable and can mutate, developing a clone which resists the chemotherapeutical influence (see, e.g., Ref. [21] and references therein). Therefore, mathematical modeling of chemotherapy optimization is an important problem of cancer modeling [3, 21]. In the framework of the simple model of Eq. (17.14), we show that the chemotherapy term can be present in the form of Eq. (17.1), namely: $G_1(x, t) = G(t)P$. We find a condition when chemotherapy compensates cell proliferation and leads to a stationary solution which describes time independent localization of cancer development.

In the general case, tumor development is a non-equilibrium process of fractional cell diffusion with fission and chemotherapy, which is described by the following generalization of Eq. (17.14):

$$e^{\mathcal{C}t}D_C^\alpha e^{-\mathcal{C}t}P + \hat{\mathcal{L}}_{FP}(x)P = -\mathcal{C}P - G_1(x, t)\,, \qquad (17.15)$$

where $G_1(x, t)$ is a chemotherapy term. Let the chemotherapy influence lead to the compensation of tumor development including cell fission due to the following condition:

$$e^{\mathcal{C}t} D_C^{\alpha} e^{-\mathcal{C}t} P = -G_1(x, t) . \tag{17.16}$$

Thus, Eqs. (17.15) and (17.16) describe a stationary process, where $P_{st} = P_{st}(x)$ is a time independent function

$$e^{\mathcal{C}t} D_C^{\alpha} e^{-\mathcal{C}t} P_{st} = -\mathcal{C} P_{st} e^{\mathcal{C}t} D^{\alpha-1} e^{-\mathcal{C}t} = -G(t) P_{st} , \tag{17.17}$$

where $e^{\mathcal{C}t} D^{\alpha-1} e^{-\mathcal{C}t} P_{st} = -G(t)$ and $D^{\alpha-1} \equiv I^{1-\alpha}$. Thus, $G_1(x, t) = G(t) P_{st}(x)$. One should bear in mind that $G(t)$ is a known function of an external control, which determines, together with $\mathcal{C}$, the finite relaxation. Taking this into account, we obtain the following equation for the stationary distribution:

$$\hat{\mathcal{L}}_{FP}(x) P_{st} = -\mathcal{C} P_{st} . \tag{17.18}$$

When the Fokker–Planck operator describes normal diffusion, namely

$$\hat{\mathcal{L}}_{FP}(x) = -D \frac{\partial^2}{\partial x^2} ,$$

the stationary solution corresponds to the exponentially localized tumor cell dissemination $P_{st}(x) = e^{-x/X_L}$ with an effective localization length $X_L = \sqrt{D/\mathcal{C}}$. This is in agreement with earlier works on glioma development (see [21] and references therein). When $\mathcal{C} \gg D$ the PDF corresponds to a solid tumor with a well-defined boundary. Conversely, when $\mathcal{C} \ll D$ this solution corresponds to a larger range of cancer invasion with exponentially localized distribution.

When the tumor development is described by inhomogeneous diffusion with $\tilde{D}(x) = Dx^2$, the solution of Eq. (17.18) corresponds to the power law distribution $P_{st} = A|x|^l$, where $l = l_{\pm} = -1/2 \pm \sqrt{1/4 + \mathcal{C}/D}$. For $|x| < 1$, $l = l_+$ and for $|x| > 1$, $l = l_-$. This distribution corresponds to the long-range invasion of the diffusion cancer and metastasis. In this case the chemotherapy term of the form $G_1(x, t) = G(t) P_{st}(x)$ is not sufficient to localize tumor development.

Nevertheless, this unrestricted tumor spreading can be localized due to the control the chemotherapy action has on the tumor development. To obtain localization of a stationary distribution, additional chemotherapy action can be applied. For instance, if $G_1(x, t) = G(t) P_{st}(x) + gx^2 P_{st}(x)$, the equation for the stationary distribution is

$$D \frac{\partial}{\partial x} x^2 \frac{\partial}{\partial x} P_{st} - (\mathcal{C} + gx^2) P_{st} = 0 . \tag{17.19}$$

This equation is a standard equation for the Bessel functions [9]. Taking into account the boundary conditions $P_{st}(\pm\infty) = 0$, we obtain the solution in the form of the modified Bessel functions of the third kind $I_\nu(z)$ and $K_\nu(z)$ [9], where $K_\nu(z)$ decays exponentially for $|z| \gg 1$. Thus, we have the following solution:

$$P_{st}(x) = x^{-1/2} Z_\nu \left(i \frac{g}{D} x \right) , \tag{17.20}$$

where $\nu = \sqrt{1 + 4\mathcal{C}/D}$ and $Z_\nu(z) = I_\nu(z)$ for $|z| < 1$, while for $|z| > 1$ $Z_\nu(z) = K_\nu(z)$. The asymptotic behavior of the modified Bessel function

$$K_\nu \left(\frac{g}{D} x \right) \sim \exp[-\sqrt{g/D}|x|] \tag{17.21}$$

ensures exponential localization of the stationary solution of the tumor development.

17.5 Conclusion

The present study is mainly focused on the influence of cell proliferation on transport properties through vessel networks (either vascular or lymphatic). Two main stages have been taken into account: cell fission with the self-entrapping time $\mathcal{T}_f$ between cell transport with durations $\mathcal{T}_t$, correspondingly. Using these two time scales, we were able to reduce a description of tumor development to a CTRW process. A simple mathematical model, which is a toy model of cancer development, is constructed, using heuristic arguments on the relation between tumor development and the CTRW. In this case a fractional tumor development becomes a well-defined problem since a mathematical apparatus of CTRW is well established (see, e.g., [4,6,16,19]. The constructed model is a modification of a comb structure [2,22]. Using this simplified approach to the fractional transport of tumor cells, we can answer the question of how the malignant neoplasm cells spread for both solid tumors and diffusive cancers. We presented analytical solutions of the problem. To this end, 1D transport was considered. A generalization of the analytical consideration of the 3D case is straightforward, when either an interaction between the degrees of freedom is absent or the problem has the property of spherical symmetry and is independent of space angles, and transport is only considered along the radius. This problem is important for solid tumor formation [20].

An important feature of this consideration of cell transport in the framework of the CTRW model is an essential enhancement of anomalous transport due to proliferation. Moreover, it is a dominant process which could be eliminated by chemotherapy. In the present analysis we obtained stationary solutions for the transporting cells, which corresponds to the localization of metastasis due to chemotherapy. A specific property of the stationary solutions is that they appear as a result of complicated interplay between fractional cell transport and time dependent chemotherapy action.

We admit that the present consideration in the framework of the toy model does not take cell size into account. Therefore, cells are considered as point-like particles, whereas for practical applications, this approach needs cell-size corrections. It is worth mentioning that the fractional transport does not contradict the processes of angiogenesis or lymphangiogenesis which are the most important factors of cancer cell transport and invasion [11, 14, 15] at a certain stage of tumor development. Moreover, these factors can be important for the fractional cell transport as well. Nevertheless, in the framework of the 1D toy model presented here, these biological phenomena cannot be considered. This question on the influence of, e.g., vascularization on the fractional cell transport needs separate consideration and will be studied in the future.

17.6 Appendix: Fractional Integro-Differentiation

A basic introduction to the fractional calculus can be found, e.g., in Ref. [19]. Fractional integration of the order of α is defined by the operator

$$I_a^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_a^t f(\tau)(t-\tau)^{\alpha-1} d\tau, \quad (\alpha > 0).$$ (17.22)

There is no constraint on the limit a. In our consideration, $a = 0$ since this is a natural limit for the time. The fractional derivative is defined as an inverse operator to $I^\alpha \equiv I_0^\alpha$:

$$\frac{d^\alpha}{dt^\alpha} = I^{-\alpha} = D^\alpha, \quad I^\alpha = \frac{d^{-\alpha}}{dt^{-\alpha}} = D^{-\alpha}.$$

Its explicit form is convolution

$$D^\alpha = \frac{1}{\Gamma(-\alpha)} \int_0^t \frac{f(\tau)}{(t-\tau)^{\alpha+1}} d\tau. \tag{17.23}$$

For arbitrary $\alpha > 0$ this integral is, in general, divergent. As a regularization of the divergent integral, the following two alternative definitions for D^α exist [13]:

$$D_{RL}^\alpha f(t) = D^n I^{n-\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \frac{d^n}{dt^n} \int_0^t \frac{f(\tau)}{(t-\tau)^{\alpha+1-n}} d\tau, \tag{17.24}$$

$$D_C^\alpha f(t) = I^{n-\alpha} D^n f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t \frac{f^{(n)}(\tau)}{(t-\tau)^{\alpha+1-n}} d\tau, \tag{17.25}$$

where $n - 1 < \alpha < n, n = 1, 2, \ldots$. Eq. (17.24) is the Riemann–Liouville derivative, while Eq. (17.25) is the fractional derivative in the Caputo form [13, 19]. Performing integration by parts in Eq. (17.24) and then applying Leibniz's rule for the derivative of an integral and repeating this procedure n times, we obtain

$$D_{RL}^\alpha f(t) = D_C^\alpha f(t) + \sum_{k=0}^{n-1} f^{(k)}(0^+) \frac{t^{k-\alpha}}{\Gamma(k-\alpha+1)}. \tag{17.26}$$

The Laplace transform can be obtained for Eq. (17.25). If $\hat{L} f(t) = \tilde{f}(s)$, then

$$\hat{L}[D_C^\alpha f(t)] = s^\alpha \tilde{f}(s) - \sum_{k=0}^{n-1} f^{(k)}(0^+) s^{\alpha-1-k}. \tag{17.27}$$

We also note that

$$D_{RL}^\alpha[1] = \frac{t^{-\alpha}}{\Gamma(1-\alpha)}, \quad D_C^\alpha[1] = 0. \tag{17.28}$$

The following fractional derivatives are helpful for the present analysis:

$$D_{RL}^\alpha t^\beta = \frac{t^{\beta-\alpha} \Gamma(\beta+1)}{\Gamma(\beta+1-\alpha)}, \tag{17.29}$$

where $\beta > -1$ and $\alpha > 0$. The fractional derivative from an exponential function can be simply calculated as well by virtue of the Mittag–Leffler function (see, e.g., [19]):

$$E_{\gamma,\delta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\gamma k+\delta)}. \tag{17.30}$$

Therefore, we have the following expression:

$$D_{RL}^{\alpha} e^{\lambda t} = t^{\alpha} E_{1,1-\alpha}(\lambda t) . \qquad (17.31)$$

In our consideration $\alpha < 1$ (and $n = 1$), and it follows from (17.26) and (17.30) that

$$D_C^{\alpha} e^{\lambda t} = D_{RL}^{\alpha} e^{\lambda t} - \frac{t^{-\alpha}}{\Gamma(1 - \alpha)} . \qquad (17.32)$$

This yields an explicit expression for the chemotherapy term in Eq. (17.17).

References

1. Bahish, J.W., Jain, R.K.: Fractals and cancer. Cancer Research, **60**, 3683–3688 (2000).
2. Baskin, E.M., Iomin, A.: Superdiffusion on a comb structure. Phys. Rev. Lett., **93**, 120603 (2004).
3. Bellomo, N., Preziosi, L.: Modelling and mathematical problems related to tumor evolution and its interaction with the immune system. Math. Comput. Modelling, **32**, 413–452 (2000).
4. Gorenflo, R., Mainardi, F.: Fractional diffusion process: probability distributions and continuous time random walk. In: Processes With Long Range Correlations, pp. 148–166. Springer-Verlag, Berlin, (2003).
5. Gottlieb, M.E.: Vascular networks: fractal anatomies from non-linear physiologies. IEEE Eng. Med. Bio. Mag., **13**, 2196 (1991).
6. Hilfer, R. (ed): Fractional Calculus in Physics. World Scientific, Singapore, (2000).
7. Iomin, A.: Superdiffusion of cancer on a comb structure. J. Phys.: Conference Series, **7**, 57–67 (2005).
8. Iomin, A.: Fractional transport of tumor cells. WSEAS Trans. Biol. Biomed., **2**, 82–86 (2005)
9. Janke, E., Emde, F., Lösh, F.: Tables of Higher Functions. McGraw-Hill, New York, (1960).
10. Kamke, E.: Differentialgleichungen: Lösungsmethoden und Lösungen. Leipzig, (1959).
11. Keleg, S., Büchler, P., Ludwig, R., Büchler, M.W., Friess, H.: Invasion and metastasis in pancreatic cancer. Molecular Cancer, **2:14**, (2003) .
12. Klafter, J., Blumen, A., Slesinger, M.F.: Stochastic pathway to anomalous diffusion. Phys. Rev. A, **35**, 3081–3085 (1987).
13. Mainardi, F.: Fractional relaxation-oscillation and fractional diffusion-wave phenomena. Chaos Solitons Fractals, **7**, 1461–1477 (1996).
14. Mantzaris, N.V., Webb, S., Othmer, H.G.J.: Mathematical modeling of tumor-induced angiogenesis. Math. Biol., **49**, 111–187 (2004).
15. McDougall, S.R., Anderson, A.R.A., Chaplain, M.A.J., Sherratt, J.A.: Mathematical modelling of flow through vascular networks: implications for tumour-induced angiogenesis and chemotherapy strategies. Bull. Math. Biol. **64**, 673–702 (2002).
16. Metzler, R., Klafter, J.: The random walk's guide to anomalous diffusion: a fractional dynamics approach. Phys. Rep., **339**, 1–77 (2000).
17. Montroll, E.W., Weiss, G.H.: Random walks on lattices, II. J. Math. Phys., **6**, 167–181 (1965).
18. Murray, J.D.: Mathematical Biology, Springer, Heidelberg, (1993).
19. Podlubny, I.: Fractional Differential Equations, Academic Press, San Diego, (1999).

20. Sherratt, J.A., Perumpanani, A.J., Owen, M.R.: Pattern formation. In: cancer, in On Growth and Form, Editors: Chaplain, M.A.J., Singh, G.D., McLachlan, J.C., pp. 47–73. Wiley, Chichester, (1999).
21. Swanson, K.R., Bridge, S., Murray, J.D., Alvord, Jr, E.S.: Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion. J. Neurol. Sci., **216**, 1–10 (2003).
22. Weiss, G.H., Havlin, S.: Some properties of a random-walk on a comb structure. Physica A, **134**, 474–482 (1984).

Mathematical Modelling of Vascular Tumour Growth and Implications for Therapy

Jasmina Panovska,[1] Helen M. Byrne,[2] and Philip K. Maini[3]

[1] Chemical Engineering, Heriot-Watt University, Riccarton Campus, Edinburgh, EH14 4AS;
J.Panovska@hw.ac.uk

[2] Centre for Mathematical Medicine, Division of Applied Mathematics, School of
Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD;
Helen.Byrne@maths.nottingham.ac.uk

[3] Centre for Mathematical Biology, Mathematical Institute, Oxford University, Oxford,
OX1 3LB; maini@maths.ox.ac.uk

Summary. In this chapter we briefly discuss the results of a mathematical model formulated in [22] that incorporates many processes associated with tumour growth. The deterministic model, a system of coupled non-linear partial differential equations, is a combination of two previous models that describe the tumour-host interactions in the initial stages of growth [11] and the tumour angiogenic process [6]. Combining these models enables us to investigate combination therapies that target different aspects of tumour growth. Numerical simulations show that the model captures both the avascular and vascular growth phases. Furthermore, we recover a number of characteristic features of vascular tumour growth such as the rate of growth of the tumour and invasion speed. We also show how our model can be used to investigate the effect of different anti-cancer therapies.

Key words: Vascular tumours, angiogenesis, hypoxia, anti-cancer therapy.

18.1 Introduction

Tumour growth is a complex process that involves a sequence of well-orchestrated events. These characterise the initial avascular phase of growth, the angiogenesis that enables the tumour to become vascularised and the vascular phase of growth. During the early stages of growth, oxygen is delivered to the tumour cells via diffusion from nearby blood vessels and the tumour cells proliferate rapidly and consume more oxygen than the host cells [7]. Due to the diffusion-limited supply of oxygen such growth is limited in size [31]. To grow larger the tumour must undergo a cascade of processes that include the secretion of tumour angiogenic factors, such as vascular endothelial growth factor (VEGF). VEGF stimulates the formation of a tumour-specific vascular network from the host vessels. Upon successful vascularisation oxygen is rapidly supplied to the tumour and it can grow larger.

Within the last three decades a number of mathematical models for tumour growth have been developed as part of the quest to understand tumour growth dynamics. Most of these models focus on one particular aspect, for example, avascular growth (e.g., [27, 32]), tissue-tissue interactions (e.g., [11, 28]), angiogenesis (e.g., [1, 6, 21]) or vascular tumour growth (e.g., [4, 5, 14, 18]). However, if we wish to compare and contrast the effectiveness of different treatment protocols via mathematical modelling, we need a model that integrates several key processes that occur during tumour growth. A first attempt at deriving such a model was made by de Angelis and Preziosi [8]. They developed a model to describe the evolution of tumour growth from the avascular stage to the vascular stage through the angiogenic process. The model was able to predict the formation of necrotic regions, the control of mitosis by the presence of an inhibitory factor, the angiogenesis process with proliferation of capillaries just outside the tumour surface and the regression of the tumour and the angiogenic capillaries when angiogenesis was controlled or inhibited. Here we briefly describe an extended model to the one in [8] by including the density of the healthy host cells in the system, and we also model two distinct components of the vasculature, distinguishing between capillary tips and blood vessels. We refer the reader to [22] for full details. In Section 18.2, we present the model equations. In Section 18.3 we illustrate the types of behaviour that the model yields when formulated on a one-dimensional spatial domain. The potential use of our model for testing anti-tumour drug protocols is illustrated in Section 18.4. We present our conclusions and comment on future research directions in Section 18.5.

18.2 Model Formulation

The model we develop comprises a system of non-linear partial differential equations and aims to reproduce the animal chamber experiments of Gimbrone et al. [12] and Muthukkaruppan et al. [17]. Thus we consider a small solid tumour implanted in the cornea of a test animal close to the limbal vessels. Angiogenesis is quick (14–21 days) and tumour growth evolves continuously from the avascular phase, through angiogenesis to the vascular phase.

A novel feature of our model, compared to previous models, is that tumour-host interactions are active in the region while the new capillary network is forming during angiogenesis. This enables us to investigate how the coupling of tumour-host dynamics and angiogenesis influences tumour growth. To our knowledge this has not been considered in existing models, which have tended to focus on a single specific aspect of tumour development.

Tumour growth, via invasion of the surrounding host cells, and angiogenesis are multidimensional processes. By averaging the dependent variables in a plane perpendicular to the direction of motion of the vascular front it is possible to restrict attention to one spatial dimension. This direction is chosen to be parallel to the line connecting the limbus, situated at $x = 0$ and where the nearest host blood vessels are found, to the tumour centre at $x = 1$ (in dimensionless terms). We introduce independent variables t and x representing, respectively, time and spatial position in a direction parallel to that of tip growth.

Within our modelling framework we consider two types of dependent variables: those that contribute to the tumour volume and those of negligible volume. In the former category we include the healthy (host) cell density $n_1(x, t)$, the tumour cell density $n_2(x, t)$ and two components for the vasculature, namely the capillary tip density $n_3(x, t)$ and the density of the blood vessels $b(x, t)$. The nutrient, which in our case is oxygen, concentration $a(x, t)$ and TAF, which in our model is VEGF, concentration $c(x, t)$ are assumed to be of negligible volume. To formulate the model equations we combine models by Byrne and Chaplain [6] and Gatenby and Gawlinski [11]. Following [6], the deterministic modelling of the vasculature-TAF interactions is based on the fungal growth model of [9], the two processes sharing many common features, including branching, anastomosis and migration [12, 17]. Based on experimental results by Sholley et al. [29], we assume that migration of the capillary tips up the gradient of VEGF concentration is the key mechanism during angiogenesis and that proliferation of the cells at the capillary tip stimulated by VEGF is secondary, and as a result, less significant. This makes our model assumptions biologically different from the modelling assumptions from [6].

Following [11] we model the tumour-host interactions via non-linear reaction-diffusion equations for the cell density. However, we assume that the normal tissue is immobile and neglect its random motion coefficient. As in [11] we assume that the tumour is unable to spread unless the surrounding healthy tissue has been diminished from its carrying capacity by, for example, increased acidity leading to death of normal cells. Thus we consider the expansion of the tumour into the adjacent tissue to depend on its composition and we model the random motion coefficient for the tumour cell density to be dependent on the density of the surrounding normal cells. Unlike [11] the equations we use for the tumour and the host cell density are coupled via the oxygen equation, rather than hydrogen ion density H^+ as a measure of the pH and acidity of the region. Oxygen is blood-borne and it controls cell proliferation and oxygen-deprived (hypoxic) death. These assumptions are based on experimental observations of tumour-host interactions in the presence and absence of oxygen [7]. The novel aspect of our model is the coupling of the equations for the cell densities and oxygen concentration with equations for the vasculature-TAF interactions. We explore the fact that when oxygen concentration in the region lowers, tumour cells (and to a lesser extent normal cells) secrete VEGF [13]. We incorporate this by assuming that VEGF is produced by the tumour and the normal cells under hypoxia. This coupling distinguishes our model from [11], from [6] and from previous models such as [8].

Combining the above ideas we arrive at the following non-dimensionalised system of equations (see [22] for details):

$$\frac{\partial n_1}{\partial t} = \underbrace{\frac{r_1 \rho_1 a}{1 + \rho_1 a} n_1}_{\text{proliferation}} - \underbrace{r_1 n_1^2}_{\text{crowding}} - \underbrace{\frac{R_1 n_1}{1 + \rho_1 a}}_{\text{hypoxic death}} - \underbrace{c_1 n_1 n_2}_{\text{competition}} \tag{18.1}$$

$$\frac{\partial n_2}{\partial t} = \underbrace{\frac{\partial}{\partial x}\left(d_{n_2}(1-n_1)\frac{\partial n_2}{\partial x}\right)}_{\text{random motion}} + \underbrace{\frac{r_2\rho_2 a}{1+\rho_2 a}n_2}_{\text{proliferation}} - \underbrace{r_2 n_2^2}_{\text{crowding}} - \underbrace{\frac{R_2 n_2}{1+\rho_2 a}}_{\text{hypoxic death}} - \underbrace{c_2 n_1 n_2}_{\text{competition}}$$

(18.2)

$$\frac{\partial a}{\partial t} = \underbrace{d_a\frac{\partial^2 a}{\partial x^2}}_{\text{diffusion}} + \underbrace{hb(1-a)}_{\substack{\text{delivery by}\\\text{blood vessels}}} - \underbrace{\frac{\lambda_1 r_1 a n_1}{1+\rho_1 a}}_{\substack{\text{consumption by}\\\text{host cells}}} - \underbrace{\frac{\lambda_2 r_2 a n_2}{1+\rho_2 a}}_{\substack{\text{consumption}\\\text{by tumour cells}}}$$

(18.3)

$$\frac{\partial c}{\partial t} = \underbrace{d_c\frac{\partial^2 c}{\partial x^2}}_{\text{diffusion}} + \underbrace{\frac{r_3}{1+\rho_1 a}n_1}_{\substack{\text{secretion by}\\\text{host cells}}} + \underbrace{\frac{r_4}{1+\rho_2 a}n_2}_{\substack{\text{secretion by}\\\text{tumour cells}}} - \underbrace{p_1 bc}_{\substack{\text{removal by}\\\text{vessels}}} - \underbrace{\gamma c}_{\substack{\text{natural}\\\text{loss}}}$$

(18.4)

$$\frac{\partial n_3}{\partial t} = \underbrace{d_{n_3}\frac{\partial^2 n_3}{\partial x^2}}_{\substack{\text{random}\\\text{motion}}} - \underbrace{\frac{\partial}{\partial x}\left(\frac{\psi c}{1+\eta c}n_3\frac{\partial c}{\partial x}\right)}_{\text{chemotaxis}} + \underbrace{p_2 bc}_{\text{branching}} - \underbrace{\beta_1 n_3 b}_{\text{anastomosis}}$$

(18.5)

$$\frac{\partial b}{\partial t} = \underbrace{d_{n_3}\frac{\partial n_3}{\partial x}}_{\substack{\text{random}\\\text{motion}}} - \underbrace{\frac{\psi c}{1+\eta c}n_3\frac{\partial c}{\partial x}}_{\substack{\text{snail-trail}\\\text{production}}} + \underbrace{s_1 b(1-b)}_{\substack{\text{vessel}\\\text{remodelling}}} - \underbrace{\delta b}_{\substack{\text{natural}\\\text{loss}}}.$$

(18.6)

The corresponding initial and boundary conditions are

$$\frac{\partial n_2}{\partial x}(0,t) = 0, \quad \frac{\partial a}{\partial x}(0,t) = -hb(1-a), \quad \frac{\partial c}{\partial x}(0,t) = p_3 bc, \quad n_3(0,t) = e^{-kt}, \quad (18.7)$$

$$\frac{\partial n_2}{\partial x}(1,t) = 0, \quad \frac{\partial a}{\partial x}(1,t) = 0, \quad \frac{\partial c}{\partial x}(1,t) = 0, \quad \frac{\partial n_3}{\partial x}(1,t) = 0, \quad (18.8)$$

$$n_1(x,0) = 1 - n_2(x,0), \quad n_2(x,0) = \frac{1}{1+\exp(-\epsilon_2(x-a_2))}, \quad a(x,0) = 1, \quad (18.9)$$

$$c(x,0) = 0, \quad b(x,0) = \frac{1}{1+\exp(\epsilon_4(x-a_4))}, \quad n_3(x,0) = \frac{1}{1+\exp(\epsilon_3(x-a_3))}. \quad (18.10)$$

We make the simple modelling assumption that captures the effect of a one-off formation of capillary tips at the limbus as the condition $n_3(0,t) = e^{-kt}$, where k represents the rate of tip decrease at the limbus. The blood vessels at the limbus supply the region with oxygen and also remove the excess VEGF. For the tumour density we impose a no-flux boundary condition at $x = 0$. We assume symmetry of the tumour about its centre and hence impose no flux boundary conditions for n_2, a, c and n_3 at $x = 1$. We assume that initially some tumour cells are located at $x = 1$, the rest of the domain is filled with normal cells and that the vasculature is only present near the limbus. Initially the region is well oxygenated and no VEGF is present.

18.3 Model Simulations

We investigate, using numerical computation, the behaviour of the model in various parameter regimes. We use the NAG library routine DO3PCF which discretises the system of equations using finite differences and solves the resulting system of ordinary differential equations using backward differentiation [26]. We find that, by changing parameter values, the model can simulate a growing tumour before and after vascularisation, as well as the clearance of the tumour due to interactions with the host tissue. Qualitatively we can capture avascular tumour growth with invasion of the host cells, the migration of the neovasculature during the angiogenic process and also vascular tumour growth characterised by the tumour growing larger and invading the host cells more rapidly than its avascular counterpart.

In Fig. 18.1(a)–(e) we present numerical solutions of the equations (18.1)–(18.10) for different parameter values. We observe avascular tumour growth and tumour invasion of the host cells (see Fig. 18.1(a)); successful angiogenesis and tumour invasion of the host cells (see Fig. 18.1(b)); avascular tumour growth and tumour coexistence with the host cells (see Fig, 18.1(c)); successful angiogenesis and tumour-host coexistence (see Fig. 18.1(d)); and tumour regression during avascular growth only (see Fig. 18.1(e)). We note that changes in key model parameters (competition parameters c_1, c_2 and oxygen consumption λ_2 as well as the chemotactic parameters η, ψ) allow us to switch from one type of behaviour to another. We illustrate this in the bifurcation diagrams in Fig. 18.2(a)–(c) where parameter space is divided into distinct regions depending on the outcomes of the simulations.

Our results suggest that the success of the angiogenic process depends on the strength of tumour-host competition: only when the tumour can compete with the host cells will angiogenesis be completed (regions M and P in Fig. 18.2(a)–(b)). Thus we predict that angiogenesis must follow invasive avascular tumour growth and it is not possible for a tumour that initially regresses to then undergo angiogenesis and invade the host cells. This occurs because in our model the tumour cells are the main source of VEGF. Hence when the normal cells are dominant the tumour recedes and VEGF secretion decreases (see Fig. 18.1(e)). Tumour vascularisation is quicker when the tumour consumes larger amounts of oxygen (i.e., as λ_2 increases; see Fig. 18.3(a)). Equally, increasing λ_2, and making the region hypoxic, in Fig. 18.2(a)–(b), increases the size of the region P thus making it more likely for the tumour to invade the host cells. This suggests that tumour invasion is stronger in hypoxic conditions and this is a new prediction of our model. Combining these results we predict that hypoxic conditions, brought about by large oxygen consumption by the tumour cells, render the tumour more invasive and able to vascularise more quickly.

Successful angiogenesis, in Fig. 18.4(c)–(d), is characterised by vascular profiles that propagate from the limbal vessels towards the tumour, with increasing speed and increasing maximum density (see Fig. 18.4(c)). In addition the capillary tip profiles precede the vessel profiles (compare Figs. 18.4(c) and 18.4(d)). These are features of what is called the brush-border effect associated with successful tumour vascularisation in the experiments by Muthukkaruppan [17]. By tracking the position of a point in the wave front over time, we estimate the speed of propagation of the tips to be
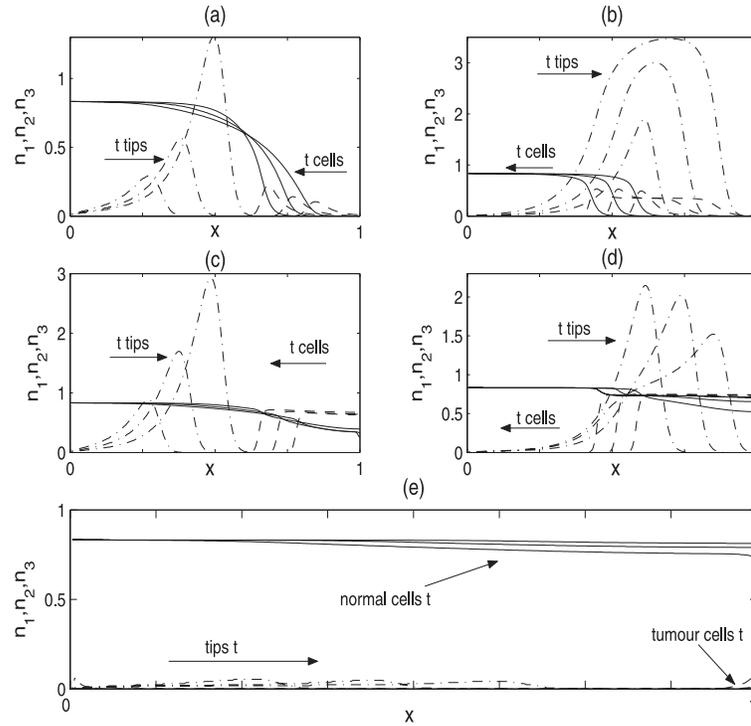
Fig. 18.1. Series of plots illustrating the types of behaviour that emerge from equations (18.1)–(18.10). The normal $n_1(-)$ and the tumour cell density $n_2(--)$ propagate as waves of normal cell regression and tumour invasion before and after successful vascularisation (a)–(d); or when the normal cells are better competitors the tumour density regresses (e) in which case angiogenesis is unsuccessful. Following successful vascularisation the tumour grows larger ((b) and (d)). During avascular tumour invasion in (a) and (c) the capillary tips $n_3(-.-)$ propagate from the limbus towards the tumour with increasing speed and increased maximum density. Post angiogenesis tip profiles propagate with constant speed and either increase to a maximum value within the tumour mass (b) or decrease towards the tumour centre (d). The results are shown at dimensionless $t = 5, 10, 15$ in (a), (c) and (e) and $t = 20, 25$ and 30 in (b) and (d). Parameter values: $r_1 = 4, \rho_1 = 8, R_1 = 1, r_2 = 10, \rho_2 = 15, R_2 = 2, d_{n_2} = 0.0007, h = 10, \lambda_1 = 0.1, r_3 = 0.1, r_4 = 10, p_1 = 10, d_c = 0.28, \gamma = 1, d_{n_3} = 0.0001, \psi = 0.8, \eta = 1.5, p_2 = 50, \beta_1 = 10, s_1 = 1, \delta = 0.25, k = 30, p_3 = 10, \epsilon_2 = 250, a_2 = 0.9, \epsilon_3 = \epsilon_4 = 250, a_3 = a_4 = 0$ and (a) $c_1 = 10, c_2 = 5, \lambda_2 = 50$; (b) $c_1 = 1, c_2 = 5, \lambda_2 = 0.5$; (c) $c_1 = 1, c_2 = 25, \lambda_2 = 50$.

approximately 0.03 in dimensionless units (or 0.11 mm day^{-1} in dimensional units) near the limbus and 0.13 in dimensionless units (or 0.4 mm day^{-1}) near the tumour in Fig. 18.4(c). This agrees with experimental measurements showing the vascular speed increasing from 0.1–0.2 mm day^{-1} near the limbus to 0.3–0.8 mm day^{-1} near the tumour [10]. Once the tumour is vascularised the speed of the vascular front becomes constant and approximately that near the limbus prior to angiogenesis. Therefore once the vessels penetrate the tumour their rate of propagation becomes constant. Further-
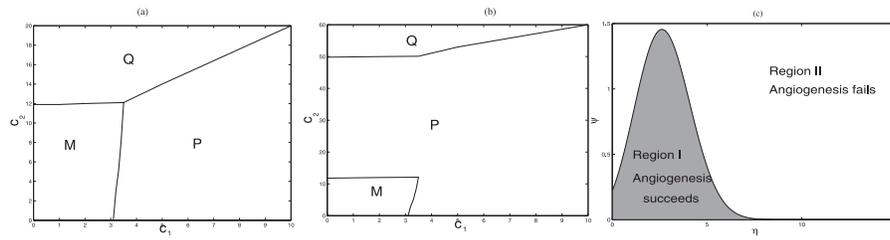
Fig. 18.2. (a)–(b) Diagram showing how the competition (c_1, c_2) parameter space can be decomposed into distinct regions depending on the long-time behaviour of the model solutions, for two different values of λ_2. In (a) $\lambda_2 = 5$ and in (b) $\lambda_2 = 50$. In regions M and P the tumour is a similar or better competitor than the normal cells, the numerical solutions evolve as travelling waves of tumour invasion of the host cells during the avascular phase followed by successful angiogenesis and vascular tumour growth. In region Q the normal cells dominate; the tumour regresses and fails to become vascularised. (c) Diagram showing the existence of a numerically calculated region where the tumour grows as avascular or vascular depending on the values of the chemotactic parameters η and ψ. This parameter space is determined with detailed numerical simulation. The rest of the parameter values are as per Fig. 18.1(a) with tumour being the better competitor and $(c_1, c_2) \in$ P from (a)–(b). Qualitatively the results are the same when $(c_1, c_2) \in$ M from (a)–(b).

more the maximum density of the capillary tips either reaches a maximum value within the tumour mass (see Fig. 18.1(b)) or decreases towards the tumour centre once the tips have penetrated the tumour (see Fig. 18.1(d)). The former case occurs when the tumour is a better competitor than the host cells, whereas the latter case occurs when the tumour cells coexist with the host cells. Therefore we predict that the outcome of
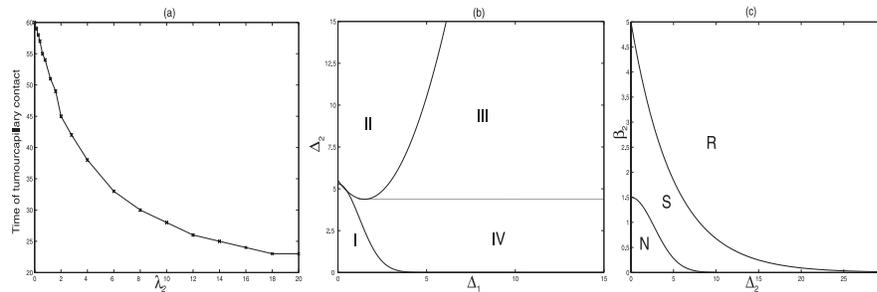
Fig. 18.3. (a) Numerically calculated decrease in the time when the tumour becomes vascularised as a function of the parameter λ_2 that controls oxygen consumption by the tumour cells. (b)–(c) Diagrams illustrating how the (Δ_1, Δ_2) and (Δ_2, β_2) parameter spaces can be decomposed into distinct regions depending on the effectiveness of an anti-proliferative therapy. In (b) in the long term, only tumour cells are killed (II), only normal cells are killed (IV), both cell types are killed (III) by the therapy, or it has no effect on tumour growth (I). In (c) the therapy is not effective in region N; the tumour regresses and is removed in region R and the tumour reaches saturated growth in region S.
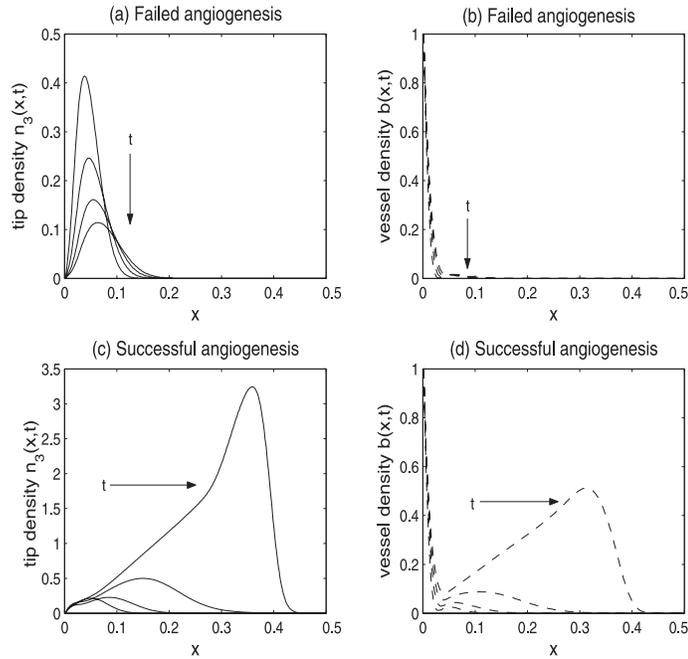
Fig. 18.4. Series of plots illustrating the profiles of the capillary tips and the vessel density during unsuccessful angiogenesis (a)–(b) and during successful angiogenesis with brush border (c)–(d). The parameters are as in Fig. 18.1(a) apart from $\gamma = 19, \eta = 50$ for (a)–(b) and $\gamma = 1$, $\eta = 1.5$ for (c)–(d). We plot the profiles at dimensionless $t = 5, 10, 15, 20$.

the tumour-host interaction affects the behaviour of the vasculature during vascular growth. A large vascular density (capillary tips in Fig. 18.1(b)) during vascular tumour growth is, in our simulations, present during successful invasion of the host tissue by the tumour cells. A small vascular density (capillary tip density in Fig. 18.1(d)) is associated with tumour-host coexistence during vascular growth.

The model also shows that angiogenesis enhances the ability of the tumour cells to invade the host tissue. For example, for the profiles depicted in Fig. 18.1(c) the tumour invasion speed increases from 0.09 (or 0.27 mm day^{-1}) to 0.105 (or 0.315 mm day^{-1}) following successful angiogenesis. In addition the tumour density is much larger following angiogenesis (compare Figs. 18.1(a) and 18.1(b)). These results suggest that in the later stages of tumour growth, following vascularisation, the tumour grows much larger and is more invasive than during the initial avascular stages of tumour growth. These observations agree with experimental observations of tumour growth *in vivo* [10].

18.4 Applications of the Model

The model (18.1)–(18.10) can be used to investigate the effect of different anti-cancer treatments. We now extend our model to include an equation for a blood-borne drug with different modes of action. For example, model simulations suggest that a tumour which in the absence of therapy invades the adjacent host tissue, when treated continuously with an anti-proliferative drug can saturate in growth or can regress. The effect of the drug in our simulations is thus similar to the effect of chemotherapeutic drugs such as doxorubicin [3] which target rapidly proliferating cells. In Fig. 18.3(b)–(c) we depict the bifurcation diagrams for the parameters that control the effectiveness of the anti-proliferative therapy: Δ_1 and Δ_2 associated with the potency of the drug on the healthy and the tumour cells and β_2 representing the rate of drug uptake by the tumour cells. We predict that the therapy is most effective when the potency on the tumour cells and the uptake of the drug by the tumour cells are large, whereas the potency on the normal cells is small.

Alternatively we study administration of a drug that destroys the vascular network. Such a drug may fall into two categories: one that targets the angiogenic stimuli (e.g., VEGF) via inhibitors such as endostatin [19] or angiostatin [20]; or a drug such as combretastatin (CAP4) [33] that directly targets the immature blood vessels. Within our model we are able to incorporate these different modes of action of the anti-vascular drug and compare the outcomes. Qualitatively the results are the same. Our simulations predict that upon administration of the anti-vascular drug tumour vascularisation can be prevented but tumour invasion of the host cells continues in this case. The tumour density resembles an avascular mass that invades with a constant speed. This is unrealistic, as we know that avascular tumours cannot grow indefinitely, and occurs because the model [11] does not properly include necrosis. Recently, an extension of the model in [11] has been proposed and shown to exhibit growth saturation [30]. A future extension of the present integrated model would be to include this new model.

When a combination of anti-proliferative and anti-vascular therapy is introduced into our model, the qualitative outcome is similar to that when only anti-proliferative therapy is applied. Tumour invasion into the host cells can either be halted and the tumour reaches a saturated growth, or tumour invasion is reversed and the tumour regresses. For more details of these and other therapeutic applications see [22] and [24].

18.5 Conclusions

We believe that the mathematical model presented here enables us to better understand the complex interactions that govern tumour growth. The continuum approach adopted in [22] allows us to make analytical predictions, for example, of wavespeed of invasion. Recently many cellular automata (CA) approaches have been developed to describe different aspects of tumour growth (see [16]). CA allows one to consider properties of individual cells but there is little in the way of mathematical theory developed for such models. Our model was the first deterministic model to study how tumour cells,

host cells and host blood vessels interact. With our results we can capture avascular followed by vascular tumour growth, as well as tumour elimination due to interactions with the host tissue. We can confirm that a vascular tumour is more aggressive and grows larger than its avascular counterpart. This agrees with *in vivo* observations of tumour growth and the modelling results presented in [4]. Furthermore we predict that it is not possible for a tumour that initially regresses to undergo angiogenesis and then invade the host cells. This may be a consequence of the fact that we do not allow for genetic mutations of the tumour cells. Our simulations also suggest that, during vascular growth, the maximum density of immature vessels within the tumour mass stays constant or decreases towards the tumour centre. In practice the situation which arises depends on the nature of the tumour-host interactions.

In terms of novel therapies, our numerical results suggest that targeting a vascular tumour with a highly potent anti-proliferative drug in combination with reducing the VEGF influence in the region (and thus preventing angiogenesis) is the most effective treatment. When the therapy only destroys the vascular network, we predict that angiogenesis can be prevented but tumour invasion will continue unaffected. We note that in our model there is continuous infusion of the drug. Questions remain as to whether such a therapy is feasible and, of course, our modelling framework does not account for the issue of side effects. A full critique of this modelling approach together with possibilities for further research are presented in [24].

Acknowledgments

References

1. Anderson, A.R.A., Chaplain, M.A.J.: Continuous and discrete mathematical models of tumour-induced angiogenesis. Bull. Math. Biol., **60**, 857–900 (1998).
2. Arakelyn, L., Merbi, Y., Agur, Z.: Vessel maturation effects on tumour growth: validation of a computer model in implanted human ovarian carcinoma spheroids. Eur. J. Cancer, **41**, 159–167 (2004).
3. Baxter, L.T., Yuan, F., Jain, R.K.: Pharmacokinetic analysis of the perivascular distribution of bifunctional antibodies and haptens: comparison with experimental data. Cancer Res., **52**, 5838 (1992).
4. Breward, C.J.W., Byrne, H.M., Lewis, C.E.: A multiphase model describing vascular tumour growth. J. Math. Biol., **65**, 609–640 (2003).
5. Breward, C.J.W., Byrne, H.M., Lewis, C.E.: Modelling the interaction between tumour cells and a blood vessel in micro-environment within a vascular tumour. Eur. J. Appl. Math., **12**, 529–556 (2001).

6. Byrne, H.M., Chaplain, M.A.J.: Mathematical models for tumour angiogenesis-numerical simulations and non-linear wave solutions. Bull. Math. Biol., **57**, 461–486 (1995).
7. Casciari, J.J., Sotirchos, S.V., Sutherland, R.M.: Variations in tumour growth rates and metabolism with oxygen concentration, glucose concentration and extracellular pH. J. Cell. Physiol., **151**, 386–394 (1992).
8. De Angelis, E., Preziosi, L.: Advection-diffusion models for solid tumour evolution in vivo and related free boundary problems. Math. Models Methods Appl. Sci., **10**, 379–407 (2000).
9. Edelstein, L.: The propagation of fungal colonies: a model for tissue growth. J. Theor. Biol., **98**, 679–701 (1982).
10. Folkman, J.: The vascularization of tumours In: Friedberg, E.C. (ed.) Cancer Biology: Readings from Scientific American. 115–124 (1986).
11. Gatenby, R.A., Gawlinski, E.T.: A reaction-diffusion model of cancer invasion. Cancer Res., **56**, 5745–5753 (1996).
12. Gimbrone, M.A., Cotran, R.S., Leapman, S.B., Folkman, J.: Tumour growth and neovascularisation: an experimental model using rabbit cornea. J. Natl. Cancer Inst., **52**, 413–427 (1974).
13. Griffiths, L., Daches, G.U.: The influence of oxygen tension and pH on the expression of platelet-derived endothelial cell growth factor thymidine phosphorylase in human breast tumour cells *in vitro* and *in vivo*. Cancer Res., **57**, 570–572 (1997).
14. Hahnfield, P., Panigraphy, D., Folkman, J., Hlatky, L.: Tumour development under angiogenic signalling: a dynamic theory of tumour growth, treatment response and post-vascular dormancy. Cancer Res., **59**, 4770–4775 (1999).
15. Jackson, T.L., Byrne, H.M.: A mathematical model to study the effects of drug resistance and vasculature on the response of solid tumours to chemotherapy. Math. Biosci., **164**, 17–38 (2000).
16. Moreira, J., Deutsch, A.: Cellular automaton models of tumour development: a critical review. Adv. Comp. Sys., **5**, 247–267 (2002).
17. Muthukkaruppan, V.R.M., Kubai. L., Auerbach, R.: Tumour-induced neovascularisation in the mouse eye. J.Natl.Cancer Inst., **69**, 699–705 (1982).
18. d'Onfrio, A., Gandolfi, A.: Tumour eradication by anti-angiogenic therapy: analysis and extension of the model by Hahnfeldt et al. (1999). Math. Biosc., **191**, 154–184 (2004).
19. O'Reilly, M.S., Boehm, T., Shing, Y., Fukai, N., Vasios, G., Lane, W.S., Flynn, E., Birkhead, J.R., Olsen, B.R., Folkman, J.: Endostatin: an endogenous inhibitor of angiogenesis and tumour growth. Cell, **88**, 277–285, (1997).
20. O'Reilly, M.S., Holmgren, L., Shing, Y., Chen, C., Rosenthal, R.A., Moses, M., Lane, W.S., Cao, Y., Sage, E.H., Folkman, J.: Angiostatin. A novel angiogenic inhibitor that mediates the suppression of metastasis by a Lewis lung carcinoma. Cell, **79**, 315–328, (1994).
21. Orme, M.E., Chaplain, M.A.J.: Two-dimensional models of tumour angiogenesis and anti-angiogenic strategies. IMA. J. Math. Appl. Med. Biol., **14**, 189–205 (1997).
22. Panovska, J.: Mathematical modelling of tumour growth and applications for therapy, D.Phil. Thesis, Oxford University, UK (2005).
23. Panovska, J., Byrne, H., Maini, P.: A mathematical model of vascular tumour growth, (in preparation).
24. Panovska, J., Byrne, H., Maini, P.: A theoretical study of the response of vascular tumours to different types of chemotherapy, (in preparation).
25. Pettet, G.J., Please, C.P., Tindall, M.J., McElwain, D.L.S.: The migration of cells in multicell tumor spheroids. Bull. Math. Biol., **63**, 231–257 (2001).

26. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes in Fortran: the art of scientific computing, 2nd edition. Cambridge University Press (1994).
27. Sherratt, J.A., Chaplain, M.A.J.: A new mathematical model for avascular tumour growth. J. Math. Biol., **43**, 291–312 (2001).
28. Sherratt, J.A., Nowak, M.A.: Oncogenes, anti-oncogenes and the immune response to cancer: a mathematical model. Proc. R. Soc. Lond., **248**, 261–271 (1992).
29. Sholley, M.M., Ferguson, G.P.: Mechanism of neovascularisation. Vascular sprouting can occur without proliferation of endothelial cells. Lab. Invest., **51**, 624–634 (1984).
30. Smallbone, K., Gavaghan, D.J., Gatenby, R.A., Maini, P.K.: The role of acidity in solid tumour growth and invasion. J. Theor. Biol., **235**, 476–484 (2005).
31. Sutherland, R.M.: Cell and environment interactions in tumour microregions: the multicell spheroid model. Science, **240**, 177–184 (1988).
32. Ward, J.P., King, J.R.: Mathematical modelling of avascular-tumour growth II: Modelling growth saturation. IMA J. Math. Appl. Med. Biol., **16**, 171–211 (1999).
33. West, C.M., Price, P.: Combretastatin A4 phosphate. Anticancer Drugs, **15**, 179–187 (2004).

A Stochastic Model of Glioblastoma Invasion

Andrew M. Stein,[1] David A. Vader,[2] Leonard M. Sander,[3] and David A. Weitz[2]

[1] Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109, USA; amstein@umich.edu

[2] Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA; vader@fas.harvard.edu, weitz@fas.harvard.edu

[3] Michigan Center for Theoretical Physics and Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA; lsander@umich.edu

Summary. Glioblastoma is the most malignant form of brain cancer. It is extremely invasive; the mechanisms that govern invasion are not well understood. To better understand the process of invasion, we conducted an *in vitro* experiment in which a 3D tumor spheroid is implanted into a collagen gel. The paths of individual invasive cells were tracked. These cells were modeled as radially biased, persistent random walkers. The radial velocity bias was found to be 19.6 µm/hr.

Key words: Glioblastoma, stochastic modeling, cell tracking.

19.1 Introduction

The outcome for patients with highly malignant brain tumors is extremely poor. Glioblastoma, the most malignant form of brain cancer, is responsible for 23% of primary brain tumors and has a 5-year survival rate below 2.1% [1]. One factor that makes glioblastoma multiforme (GBM) so difficult to treat is its high invasiveness [4]. It is known that the invasive cells are highly motile, but the mechanisms that govern their motility are not understood.

In this chapter, we present new results from experiments where fluorescently labeled tumor spheroids were grown in three-dimensional (3D) collagen gels. One day after implantation, the system was imaged once per minute for 12 hours using wide field microscopy. Postprocessing algorithms were used to derive individual cell paths from these images. These cells were modeled as radially biased random walkers using a 3D Ornstein–Uhlenbeck (OU) process. The model fits the data and also provides evidence for directed motility of the invasive cells away from the spheroid at a rate of 19.6 µm/hr.

19.2 The Experiments

19.2.1 Cell Culture

U87 glioblastoma cells were cultured in Dulbecco's modified essential medium (DMEM) with 10% fetal bovine serum (FBS), supplemented with 100 U/mL penicillin and 100 mg/mL streptomycin. The histone-GFP (H2B-GFP) fusion protein was stably expressed in U87 cells. The protein was transfected via the pBOS-H2BGFP vector, recently described by Kanda et al. [9] and commercially available (BD Pharmingen, San Diego, CA). Blasticidin (VWR, Westchester, PA) at 5 mg/mL was added to the cell culture medium to select for cells expressing the fusion protein.

19.2.2 Mullticellular Spheroids

Cells were grown to confluency in 10 cm diameter cell culture dishes, rinsed twice with phosphate-buffered saline (PBS), and trypsinized. After trituration, the cell suspension was centrifuged for 3 minutes at 2,000 RPM and the supernatant was replaced with new cell media. The cell suspension was then diluted to a cell number density of 2.5×10^4/mL. The hanging droplet method [10] was then used to produce spheroids. Briefly, 20 mL droplets of diluted cell suspension (∼500 cells) were pipetted on the inside of the cover of a cell culture dish; the cover was then turned right side up and placed on top of the culture dish; the droplets were allowed to hang for 3 days until cells accumulated at the bottom of the drop through gravity and adhered to one another to form a cell spheroid.

19.2.3 Tumor Model

The extracellular matrix was modeled, *in vitro*, by using type I bovine collagen (Angiotech Biomaterials, Palo Alto, CA) at a final concentration of 1.5 mg/mL. The collagen solution was prepared so as to contain 10% 10X DMEM, 10% FBS, 1% PS, and 0.025 M Na_2HCO_3 (Invitrogen, Carlsbad, CA). A neutral pH was achieved by adding NaOH 1 M to the collagen solution. The spheroid droplet was added after three days of hanging to 400 mL of collagen solution. The sample was then placed in an incubator (37°C, 5% CO_2, 100% humidity) for 1 hour to allow collagen to polymerize. Then, 100 mL of cell media was added on top of the polymerized sample to prevent drying.

19.2.4 Image Acquisition

We used a Leica DM IRB inverted microscope (Leica Microsystems, Wetzlar, Germany) with a 5X Leica objective to image our samples through a 640x480 Hamamatsu C7190 high sensitivity digital video camera (Hamamatsu Photonics, K. K., Hamamatsu City, Japan). The sample was placed on a heating stage (digital tempcontrol 37-2, Leica) and covered with a CO_2 control chamber (digital CTI controller 3700, Leica). We used a traditional FITC cube with a 100W mercury arc lamp (Ludl Electronic Products, Hawthorne, NY) to excite the GFP-labeled nuclei and acquire fluorescent
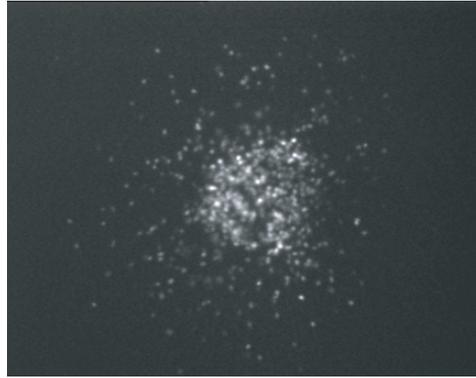
Fig. 19.1. Wide field confocal microscopy image of the invasive tumor spheroid.

images of our sample. A cell was visible if it was within approximately 25μm of the focal plane. The resulting image was a 2D projection of the 3D cell paths near the focal plane.

An example of an image taken from the setup is shown in Fig. 19.1. A total of 720 images were taken, one each minute for 12 hours. The paths taken by individual cells were traced. From the image, the spheroid was determined to be 212 μm in radius. It did not change significantly over the course of 12 hours. The coordinates for the center of the spheroid were obtained by eye.

19.2.5 Cell Tracking

Identifying the nuclei on the images was achieved by a particle tracking method previously described by Crocker and Grier [2] and implemented in the IDL (Research Systems, Inc, Boulder, CO) programming language. The output of the tracking routine is particle position, velocity, brightness, and radius for each time frame, as well as a particle ID assigned automatically by the program. All subsequent analysis of cell trajectories was done using in-house written Matlab (Mathworks, Inc. Natick, MA) code. Note that due to the high density of cells in the tumor core, the tracking algorithm produces errors there. Thus all tracks located at the center of the image in the region estimated to be the core were ignored. The paths are shown in Fig. 19.2.

19.3 Mathematical Model

A model that is frequently used to describe cell motility is the continuous, persistent random walk, described by the Ornstein–Uhlenbeck (OU) equation [6,8,11]

$$d\mathbf{V} = -\beta(\mathbf{V} - V_0\hat{\mathbf{r}})dt + \alpha d\mathbf{W}. \tag{19.1}$$

Here, $\mathbf{V}$ is a three-dimensional (3D) velocity vector, V_0 is the velocity bias, and multiplication of V_0 by $\hat{\mathbf{r}}$ indicates that this bias is in the radial direction, away from the

(a) Cell paths from experiment   (b) Cell paths simulated by (19.1)
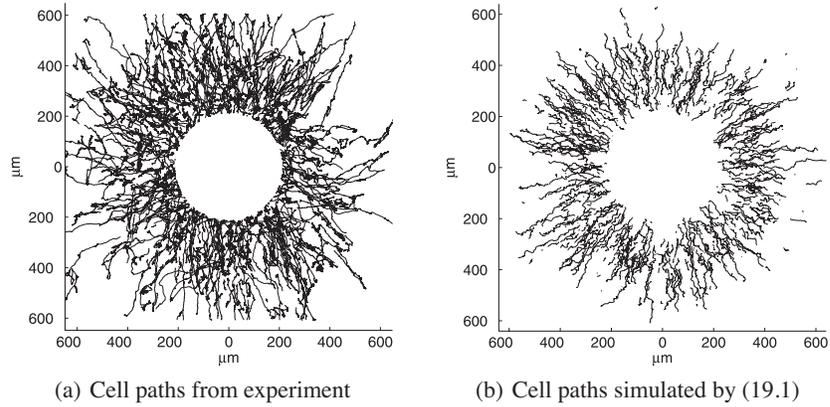
Fig. 19.2. Cell paths from experiment and from Ornstein–Uhlenbeck model.

tumor spheroid. $\mathbf{W}$ is the Wiener process that represents the unpredictable aspects of cell motion. The x, y, and z components of the velocity are modeled as independent processes. One can extract physical meaning from α and β by noting that the persistence time, P, is equal to $1/\beta$, and the diffusion constant, D, is equal to $n\alpha/\beta^2$, where n is the number of dimensions [11].

The tumor spheroid was modeled as a 3D reflecting sphere 157 μm in radius to match the size of the spheroid in the experiment. Each individual invasive cell was modeled as a random walker that obeyed (19.1). At time 0, random walkers were placed in a planar slice 50 μm thick which passed through the center of the spheroid. Their initial velocity was chosen to be radial and uniformly distributed in the interval $[0, 2V_0]$. When initially seeded, the cells were limited to be in a region within 200 μm from the spheroid surface to reflect the fact that, at time 0 in the experiment, all the invasive cells are in that region. The paths were recorded only when they stayed within the 50 μm planar slice. This reflected the fact that cells could leave and enter the field of view of the microscope.

19.4 Data Analysis

Using the above model, the paths from the experiment were analyzed so that the parameters for the model could be estimated. The track of an individual cell is denoted by $\mathbf{X_i}(t)$, where t indicates the time at which the measurement was taken and i is the index for a particular cell path. Paths may have missing data at some time points. The velocity of a cell is given by a forward difference, where h is one minute, the time between two successive images:

$$\mathbf{V_i}(t) = \frac{\mathbf{X_i}(t + h) - \mathbf{X_i}(t)}{h}. \tag{19.2}$$
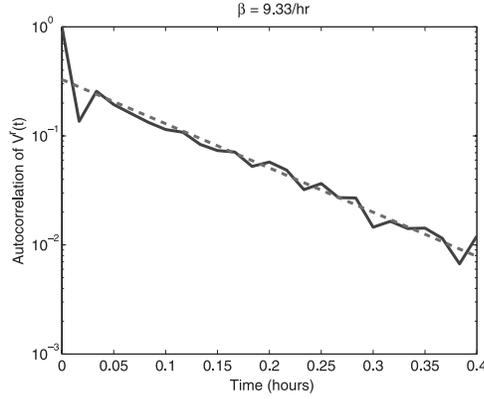
Fig. 19.3. The autocorrelation ensemble (*solid line*) and the least squares fit (*dashed line*).

The velocity vector, $\mathbf{V_i(t)}$, is expressed in polar coordinates oriented at the spheroid center. This vector is then split into its radial, $V_i^r(t)$, and angular, $V_i^\theta(t)$, components. The radial distance of the cell from the spheroid center is $R_i(t)$. It should be noted that while these cells moved in three dimensions, they were only imaged in two.

For the OU process, the autocorrelation function $r(\tau)$ decays at a rate $e^{-\beta\tau}$. Thus the autocorrelation function can be used to estimate β. We begin by finding the auto-covariance function of an individual cell, $c_i(\tau)$, using

$$c_i(\tau) = \frac{1}{N_i(\tau)} \sum_{t=t_1}^{t_m-\tau} (V_i^r(t) - \overline{V_i^r(t)})(V_i^r(t-\tau) - \overline{V_i^r(t)}). \qquad (19.3)$$

Here $N_i(\tau)$ is the number of pairs of points for a particular cell path that are spaced τ apart, t_1 is the first time point at which a cell path was measured, t_m is the last time point at which that cell path was measured, and $\overline{V_i^r(t)}$ is the average radial velocity of that cell, over its entire time course.

The autocorrelation for a particular cell is given by $r_i(\tau) = c_i(\tau)/c_i(0)$. The autocorrelation for the ensemble of cells, $r(\tau)$, is computed by averaging over all $r_i(\tau)$ at each time point. The result is shown in Fig. 19.3, along with a least squares fit. From the fit, we estimate that $\beta = 9.3$/hr. Note that the least squares fit does not go through the point $(0, 1)$. This is probably due to the inaccuracy of the velocity measurement at 1 minute time intervals caused by noise and pixelation. The diffusivity, D, was estimated to be 216 μm^2/hr, from [7] and the data from [5]. Using β, one finds $\alpha = D\beta^2/3 = 6300$ μm^2/hr^3.

To estimate V_0, we observe how the average radial velocity changes with the radial distance from the center of the spheroid. This quantity, $\langle V^r \rangle(d)$, is estimated, as shown in (19.4), by averaging the velocities of all the cells that fall in a ring of thickness $\Delta d = 20$ μm that is located d distance away from the center of the spheroid:

$$\langle V^r \rangle(d) = \frac{1}{N(d)} \sum_{i,t} \{V_i^r(t) \mid R_i(t) \in [d, d + \Delta d]\}. \qquad (19.4)$$
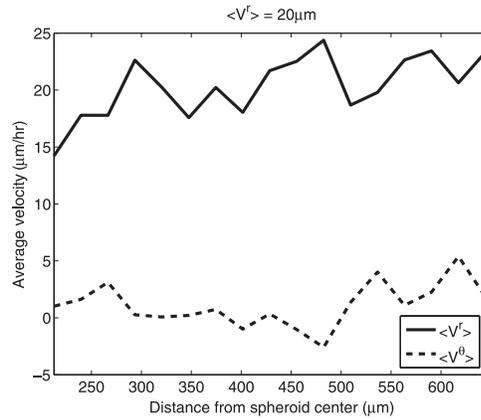
Fig. 19.4. The radial velocity (*solid line*) and angular velocity (*dashed line*) ensemble averaged over all cells and time as a function of distance from the center of the spheroid.

Here, $N(d)$ is the number of velocity measurements made in the ring of radius d. The average angular velocity, $\langle V^\theta \rangle(d)$ is found in the same way. The average radial and angular velocity are graphed in Fig. 19.4. From the figure, we see that $\langle V^r \rangle(d)$ and $\langle V^\theta \rangle(d)$ are approximately constant, with $\langle V^r \rangle(d) = 19.6$ µm/hr and $\langle V^\theta \rangle(d) = 0.8$ µm/hr.

Of course, cells without a radial bias that diffuse and reflect from a spherical boundary would also have a $\langle V^r \rangle(d)$ that is greater than zero because cells will diffuse away from the spheroid with time. To verify that this is not the cause of the bias, the model was run with the same parameter values for α and β above, but with $V_0 = 0$. The results are shown in Fig. 19.5.

(a) Simulated cell paths   (b) Radial and Angular Velocity
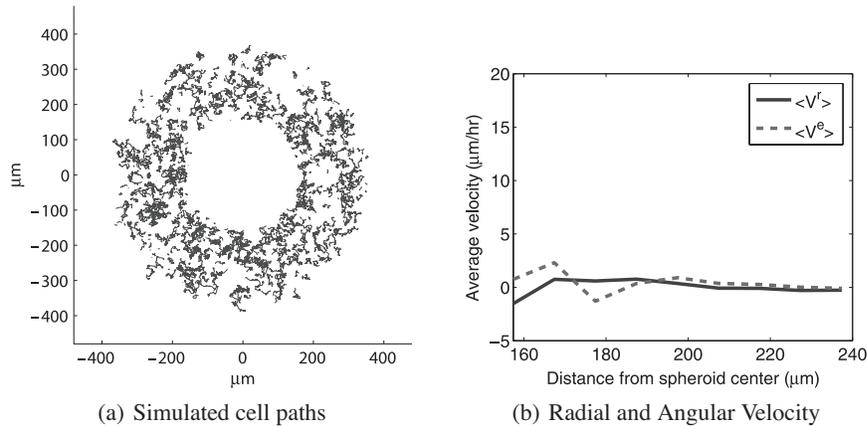
Fig. 19.5. Model simulation with the radial velocity bias set to zero. The radial bias that was present in the experiments is not seen here.

19.5 Discussion and Conclusions

The data provide clear evidence that when close to the spheroid, cells move away from it at a constant rate of 19.6 μm/hr. It is known that over longer times, this rate will decrease [3]. The cause of the radial bias is unknown. It may be due to chemicals in the gel that direct cell motion, or it may be due to the way in which the cells reorganize the collagen matrix as they move. Observing cell motion in a larger field of view and over longer times so that the decrease in directionality can be observed would help to answer this question. Since the images obtained were 2D projections of a 3D system, no direct information about velocity of the cells perpendicular to the focal plane, $\langle V^z \rangle$, was obtained. However, assuming spherical symmetry, $\langle V^z \rangle \approx \langle V^\theta \rangle \ll \langle V^r \rangle$, suggesting that the 2D projection alone provides sufficient information for estimating the cell velocity in 3D.

It should be noted that there are differences between the simulated and real cell paths. In the experiment, the path of an individual may have a short persistence time during one time interval and a long persistence time during a later time interval. In our model, persistence time is constant. It is not entirely clear what is causing the cell motility to change. The periods of short persistence time may be due to cell-cell adhesion when cells are in close proximity or it may be due to cells staying in place during mitosis. It may also be that, in isolation, a cell would naturally behave this way. It is not yet clear if it is necessary to consider this phenomenon when describing invasion over longer time scales.

Acknowledgments

References

1. CBTRUS: Primary brain tumors in the United States statistical report 1995–1999. Technical report (2002–2003).
2. Crocker, J.C., Grier, D.G.: Methods of digital viedo microscopy for colloidal studies. J. Coll. and Interf. Sci., **179**, 298–310 (1996).
3. Deisboeck, T.S., Berens, M.E., Kansal, A.R., Torquato, S., Stemmer-Rachamimov, A.O., Chiocca, E.A.: Pattern of self-organization in tumour systems: complex growth dynamics in a novel brain tumor spheroid model. Cell Proliferation, **34**, 115–134 (2001).
4. Demuth, T., Berens, M.E.: Molecular mechanisms of glioma cell migration and invasion. J. Neuro-Oncol., **70**, 217–228 (2004).
5. Demuth, T., Hopf, N.J., Kempski, O., Sauner, D., Herr, M., Giese, A., Perneczky, A.: Migratory activity of human glioma cell lines in vitro assessed by continuous single cell observation. Clin. and Exper. Metast., **18**, 589–597 (2001).

6. Dunn, G.A., Brown, A.F.: A unified approach to analysing cell motility. J. Cell Sci. Suppl., **8**, 81–102 (1987).

7. Hegedus, B., Zach, J., Czirok, A., Lpvey, J., Vicsek, T.: Irradiation and taxol treatment result in non-monotonous, dose-dependent changes in the motility of glioblastoma cells. J. of Neuro-Oncol., **67**, 147–57 (2004).

8. Ionides, E.L., Fang, K.S., Oster, G.F.: Stochastic models for cell motion and taxis. J. Math. Biol., **48**, 23–37 (2004).

9. Kanda, T., Sullivan, K.F., Wahl, G.M.: Histone-gfp fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. Curr. Biol., **8**, 377–385 (1998).

10. Kelm, J., Timmins, N., Brown, C., Fussenegger, M., Nielson, L.: Method for generation of homogeneous multicellular tumor spheroids applicable to a wide variety of cell types. Biotechn. and Bioeng., **83**, 173–80 (2003).

11. Stokes, C.L., Lauffenburger, D.A., Williams, S.K.: Migration of individual microvessel endothelial cells: stochastic model and parameter measurement. J. Cell Sci., **99**, 419–430 (1991).

Morphology of Tumor Vasculature:
A Theoretical Model

Katalin Bartha[1] and Heiko Rieger[2]

[1] Department of Medical Biochemistry, Semmelweis University, Budapest, Hungary;
katalin_bartha@puskin.sote.hu

[2] Theoretische Physik, Universität des Saarlandes, 66041 Saarbrücken, Germany;
h.rieger@mx.uni-saarland.de

Summary. A theoretical model based on the molecular interactions between a growing tumor and a dynamically evolving blood vessel network describes the transformation of the regular vasculature in normal tissues into a highly inhomogeneous tumor-specific capillary network. The emerging morphology, characterized by the compartmentalization of the tumor into several regions differing in vessel density, diameter and degree of tumor necrosis, is in accordance with experimental data for human melanoma. Vessel collapse, due to a combination of severely reduced blood flow and solid stress exerted by the tumor, leads to a correlated percolation process that is driven towards criticality by the mechanism of hydrodynamic vessel stabilization.

Key words: Tumors, angiogenesis, cooption, remodeling, blood flow, network morphology, fractals.

20.1 Introduction

Tumor vasculature, the network of blood vessels in and around a growing tumor, is in many respects different from the regular vasculature in normal tissues. Hypoxia, the lack of oxygen, that prevents a small tumor nucleus from further growth, induces the expression of various diffusible growth factors (GFs) by the tumor cells that trigger a coordinated response of angiogenesis—the formation of irregular blood vessels [1,2]. The expected increase in microvasular density (MVD) is usually observed in the periphery of the tumor, whereas the morphology of the vasculature in the central part of the tumor is characterized by a *decreased* MVD, dilated vessels and regions of necrotic tumor tissue [3,4]. The resulting tumor-specific capillary network is very heterogeneous, composed of dense and void regions and has a fractal dimension different from normal arterio-venous or normal capillary networks [5].

Although on the molecular level the main actors in the angiogenic game are rapidly identified, the physical principles that determine the global morphology of the vascular network in tumor tissues are not known. Since, for instance, MVD is used as a diagnostic tool in cancer therapy [6], a quantitative understanding of the mechanism that leads to the compartmentalization of the tumor vasculature into various regions

differing substantially in vessel density appears mandatory. Moreover, scale-invariant aspects like fractal dimension are used as hints towards the nature of the growth process underlying the formation of the tumor vasculature [7]. In this chapter we propose a theoretical model for the evolution of tumor vasculature that illuminates the physical principles leading to its global morphology [8]. The experimentally observed increase in MVD at the tumor perimeter and periphery and decrease in MVD and vessel dilation in the tumor center in human melanoma [4] appear as the general scenario in our theoretical model. Furthermore, we will argue that vessel collapses in the interior of the tumor lead to a percolation process which is driven towards criticality, the percolation threshold, via a mechanism of vessel stabilization by increased blood flow in the remaining vessels.

20.2 Model

There is a large amount of work on the mathematical modeling of tumor-induced angiogenesis (for reviews see, e.g., [9, 10]), which can be classified into two groups: Either they concentrate on blood vessel densities rather than vessel morphology (as in continuum partial differential equation [11, 12] or in locally coupled map lattice [13] approaches), or they represent vessels as interconnected lattice patterns, line segments or continuous curves [14–17] and assume a static tumor. A growing tumor in the vascular phase, however, remodels the blood vessel network via cooption, regression and growth—and the emerging morphology is determined by the interaction of the two dynamically evolving systems: the growing tumor and the remodeling vessel network.

20.2.1 Vessel Network

We describe the vessel network by a graph $G = (V, E)$, in which edges $e \in E$ represent tubular vessel segments of diameter $d(e)$ and nodes $v \in V$ represent vessel junctions, where two or more vessel segments join. For the moment we restrict to capillary networks and do not discriminate between arteries and veins, but a hierarchical structure of the original vessel network is easily incorporated into our model [18]. The network is fixed in an initial configuration representing features of the normal tissue vasculature like homogeneous microvascular density, typical vessel diameters, etc., but it can dynamically change over time: new vessels can be inserted, others can be removed, vessel diameter can change. For computational simplicity we allow only discrete locations of the nodes, i.e., they occupy certain sites on a square lattice of grid size 10 μm, by which each node gets a Cartesian coordinate $\mathbf{r}(v) = (x, y)$.

20.2.2 Blood Flow

Blood flow through this interconnected network of tubes is assumed to be an ideal pipe flow with flow conservation at all junctions v: $\sum_{e \in E(v)} q(e) = 0$, which is Kirchhoff's

law. $E(v)$ is here the set of all edges attached to the node v, and $q(e)$ the flow rate through vessel e. $q(e)$ and $f(e)$, which the shear force $f(e)$ acting upon the vessel wall, then follow Hagen–Poiseuille's law:

$$q(e) = (\pi/128)\eta_{d(e)}^{-1} \cdot d^4(e) \, \nabla P(e) \quad \text{and} \quad f(e) = (1/4) \cdot d(e) \nabla P(e), \quad (20.1)$$

where $\nabla P(e)$ is the pressure gradient in e, which is $\nabla P(e) = P(v_1(e)) - P(v_2(e))/l(e)$, with $P(v)$ the pressure at node v, $v_1(e)$ and $v_2(e)$ the start and end points of the edges e, and $l(e)$ the length of vessel e. In principle the viscosity $\eta_{d(e)}$ depends on the tube diameter $d(e)$, since blood is a non-Newtonian fluid, but for simplicity we set it to a constant, as it is correct initially, when all vessels have the same diameter.

Together with fixed boundary conditions for the blood pressure $P(v)$, the flow conservation equations establish an inhomogeneous system of linear equations for the blood pressures $P(e)$, which is solved numerically for the vessel network at hand. We choose boundary conditions that produce a homogeneous blood flow and shear stress in all capillaries in the initial network: $P(v)$ is fixed on all nodes v on the boundaries of the system such that it decreases linearly from $P_{\max}$ at the node at $\mathbf{r} = (x, y) = (L, L)$ along the boundary nodes at $\mathbf{r} = (L, y)$ and $\mathbf{r} = (x, L)$ to $(P_{\max} - P_{\min})/2$ at $\mathbf{r} = (L, 0)$ and $\mathbf{r} = (0, L)$; and from here further linearly along the boundary nodes at $\mathbf{r} = (x, 0)$ and $\mathbf{r} = (0, y)$ to $P_{\min}$ at $\mathbf{r} = (0, 0)$. One should note that these boundary conditions produce a pressure gradient and hence a global blood flow in the diagonal direction, which is somewhat unrealistic and will be repaired in initial network configurations that contain a hierarchy of arteries and veins.

20.2.3 Tumor Growth

The tumor in our model is defined on a square lattice, where each site represents an area of 10 μm $\times$ 10 μm. The tumor configuration is given by the set of lattice sites T that are occupied by tumor cells. Initially a nucleus of N_0 sites is occupied by tumor cells; proliferation can happen only at empty neighbor sites of already occupied sites [19] and removal (death) of tumor cells can happen everywhere. The restriction of tumor cell (TC) proliferation to the outer rim of the tumor is also observed in real tumors [20] and in theoretical models involving TC elasticity and increasing solid stress inside the tumor [21]. More sophisticated representations of the growing tumor are easily incorporated later into our model.

20.2.4 Oxygen Concentration

Proliferation and death of tumor cells depend on the supply of oxygen (or other nutrients), which is determined by the current vessel network: Oxygen is transported by blood flow through the vascular system and has to diffuse through the vessel wall to reach other cells in the extracellular matrix. In the case of a highly diffusible solute like oxygen, the transmural flux J_w is essentially driven by the difference between the oxygen levels inside and outside the vessel. With this boundary condition the distribution

of extracellular oxygen is described by a diffusion equation with sink terms, for which
the adiabatic approximation is completely sufficient [22,23], since the inter-vessel dif-
fusion time for oxygen is of the order of 1–10 seconds whereas the cell proliferation
time is several hours. The diffusion constant of oxygen in tissue is of the order of
$D \sim 10^{-5}$ cm^2/sec [24] and the mean square displacement of the underlying Brown-
ian motion of the O$_2$-molecules is $\langle x^2(t) = Dt \rangle$, hence the mean time for O$_2$ to pass
from one vessel to another which are $a = 100 \ \mu$m apart is ca. $\sqrt{a/D} = 10$ sec.

The Green's function method is an elegant and computationally tractable way to
solve the diffusion problem for the oxygen delivery to tissue by microvascular net-
works for large grid sizes [25,26]. The essential idea is to represent blood vessels as
a set of discrete O$_2$ sources and the O$_2$ field in the tissue as a superposition of fields
resulting from those sources. In the most general case the source and sink strengths are
unknown and have to be determined implicitly by solving a system of linear equations
for them.

Here we assume a uniform oxygen consumption rate M_0 of the normal tissue. Then
the oxygen distribution at site $\mathbf{r}$ resulting from a unit point source at $\mathbf{r}'$ is defined as the
Green's function $G(\mathbf{r}, \mathbf{r}')$ and given by the solution of

$$D \Delta_{\mathbf{r}} G - M_0 G = -\delta(\mathbf{r} - \mathbf{r}') \,, \tag{20.2}$$

where D is the oxygen diffusion constant and $\delta(\mathbf{r})$ is the delta function. The result-
ing $G(\mathbf{r}, \mathbf{r}')$ depends only on the distance from the point source $R = |\mathbf{r} - \mathbf{r}'|$ and
decays exponentially with R on a length scale $R_{\text{oxy}} = \sqrt{D/M_0})$. For computational
convenience we replace the exact Green's function by a piece-wise linear function that
decays to zero on the same length scale:

$$G(\mathbf{r}, \mathbf{r}') = \frac{3}{\pi R_{\text{oxy}}^2} \left(1 - \frac{|\mathbf{r} - \mathbf{r}'|}{R_{\text{oxy}}} \right) \cdot \theta(|\mathbf{r} - \mathbf{r}'| - R_{\text{oxy}}), \tag{20.3}$$

where $\theta(x)$ is the step function ($\theta(x) = 1$ for $x \geq 0, \theta(x) = 0$ for $x < 0$). Furthermore
here we assume for simplicity that the presence of TCs does not significantly alter the
oxygen consumption rate M_0. At least for melanoma this appears to be an acceptable
approximation, since the skin tissue MVD$_0 \approx 100$/mm^2 [4] indicates $R_{\text{oxy}} \approx 100 \ \mu$m
(see below), and data for pO$_2$ gradients in tumors also indicate $R_{\text{oxy}} \approx 100 \ \mu$m [1].
There are, however, various ways in which one could include the effect of an increased
cell density approximatively, for instance, by introducing a TC-density dependent O$_2$
diffusion range $R_{\text{oxy}}(\rho)$ that decreases monotonously with the local TC density ρ.

The total oxygen concentration $O_2(\mathbf{r})$ is then given by

$$O_2(\mathbf{r}) = \sum_{e \in E} \sum_{\mathbf{r}' \in e} J(\mathbf{r}') \cdot G(\mathbf{r}, \mathbf{r}'), \tag{20.4}$$

where $J(\mathbf{r}')$ is the source strength of a vessel segment with its center at $\mathbf{r}'$ [26]. This
depends on the difference between the blood O$_2$ partial pressure inside the vessel seg-
ment at $\mathbf{r}'$, $P_{\text{oxy}}(\mathbf{r}')$, and the tissue O$_2$ concentration at $\mathbf{r}'$: $J(\mathbf{r}') = P_{\text{oxy}}(\mathbf{r}') - O_2(\mathbf{r}')$.
Inserting this into (20.4) yields a system of linear equations for $O_2(\mathbf{r}')$ at all vessel

segments $\mathbf{r}'$, its solution determines $J(\mathbf{r}')$. We implemented this procedure for various (small) vessel network configurations and compared the resulting field $O_2(\mathbf{r})$ with one that we obtained by setting $J(\mathbf{r}') = 1$ for all vessel segments $\mathbf{r}'$. It turned out that both models lead to qualitatively the same O_2 fields for the range of MVDs that occur in our simulation runs (see below). Quantitatively the model with constant source strengths overestimates the O_2 concentration by ca. 40% for an inter-capillary distance of $0.5 \cdot R_{\mathrm{oxy}}$, corresponding to $2 \cdot \mathrm{MVD}_0$. Hence, for computational simplicity we assume constant source strengths $J(\mathbf{r}') = 1$ in our model.

20.2.5 Growth Factor Distribution

Tumor cells under hypoxia secrete increased amounts of GFs that can stimulate the formation of new blood vessels. In our model we assume that a TC at site $\mathbf{r}$ secretes GF if $O_2(\mathbf{r}) < c_{\mathrm{oxy}}$. The diffusion of the molecules into the extracellular matrix (ECM) can be described by a diffusion process with source terms, which can again be approximated to be adiabatic:

$$D_{\mathrm{GF}}\Delta GF(\mathbf{r}) - k_1 GF(\mathbf{r}) + s_{GF}(\mathbf{r}) = 0, \tag{20.5}$$

where $GF(\mathbf{r})$ is the GF concentration field, D_{GF} is the diffusion constant for GF in the ECM, k_1 is the degradation rate of the growth factors and $s_{GF}(\mathbf{r})$ the source strength at location $\mathbf{r}$. The latter we assume to be a delta function of unit weight at each TC under hypoxia. As a result this diffusion equation can again be solved by the Green function method and we replace the exact Green function by a piecewise-linear function as for the oxygen concentration field. The GF concentration therefore is

$$GF(\mathbf{r}) = \sum_{\mathbf{r}' \in T \text{ with } O_2(\mathbf{r}') < c_{\mathrm{oxy}}} \frac{3}{\pi R_{\mathrm{GF}}^2} \left(1 - \frac{|\mathbf{r} - \mathbf{r}'|}{R_{\mathrm{GF}}}\right) \cdot \theta(|\mathbf{r} - \mathbf{r}'| - R_{\mathrm{GF}}). \tag{20.6}$$

20.2.6 Dynamics

TCs proliferate/die when the local oxygen concentration is high/low. Vessels (edges) emerge when the local GF concentration is high enough, and they vanish (collapse) stochastically inside the tumor, when the hydrodynamic shear force acting on the vessel walls is too low. The biological and pathophysiological motivation for the details of the model definition is discussed in [8].

Starting with the initial configuration described above, the following updates are performed sequentially in each time step of duration $\Delta t = 1$ h. See Fig. 20.1 for an illustration.

(a) *TC proliferation*. TCs can proliferate at tumor surface sites if the local oxygen concentration is sufficient: If $\mathbf{r}$ is not occupied by a TC but has at least one neighboring TC and if $O_2(\mathbf{r}) > c_{\mathrm{oxy}}$: $T \to T \cup \{\mathbf{r}\}$ with probability $p_{\mathrm{TC}}^{\mathrm{new}} = \Delta t / t_{\mathrm{TC}}$.

(b) *TC death*. TCs that are extremely under-oxygenated for a long time are eliminated. We define the threshold for extreme under-oxygenation to be 10% of the threshold

(a)

(b)

(c)

(d)

(e)

(f)
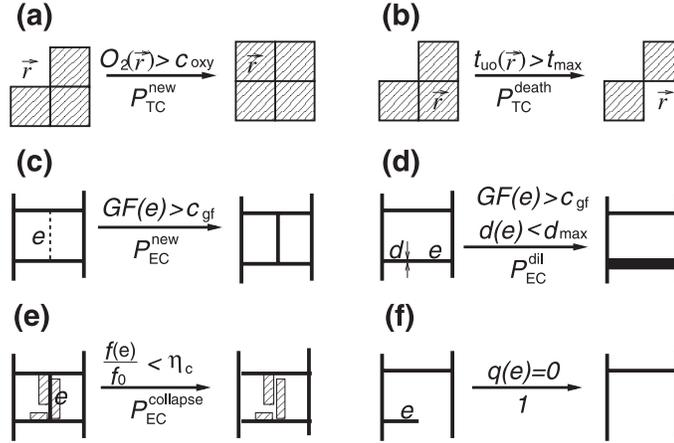
Fig. 20.1. Schematic illustration of the model: (a) TC proliferation, (b) TC death, (c) vessel growth, (d) vessel dilatation, (e) vessel collapse due to low shear force, and (f) collapse of non-circulated vessels.

beyond which they start to proliferate (i.e., $0.1 \cdot c_{\text{oxy}}$). If $O_2(\mathbf{r}) < 0.1 c_{\text{oxy}}$ for a TC at site $\mathbf{r}$ the counter for the time that a TC at site $\mathbf{r}$ spent in hypoxia is increased by one: $t_{\text{uo}}(\mathbf{r}) \rightarrow t_{\text{uo}}(\mathbf{r}) + 1$. When $t_{\text{uo}}(\mathbf{r}) > t_{\text{max}}$, the TC is eliminated: $T \rightarrow T - \{\mathbf{r}\}$ with probability $p_{\text{TC}}^{\text{death}} = 1/2$.

(c) *Vessel growth.* New straight vessel segments between two circulated vessels at site $\mathbf{r}$ and $\mathbf{r}'$ are introduced with probability $\Delta\tau/T_e$ (where T_e is the endothelial cell (EC) proliferation time) if: $GF(\mathbf{r}, t) > c_{\text{GF}}$, the neighbor of $\mathbf{r}$ on the migration path is not occupied by a TC, no site and no neighbor site of the migration path are occupied by ECs except $\mathbf{r}$ and $\mathbf{r}'$, and $|\mathbf{r} - \mathbf{r}'| < M_{\text{max}}$ (M_{max} being the maximum sprout migration distance). In such an event, $e(\mathbf{r}, t) = 1$ and $e_r(\mathbf{r}, t) = r_0$ along this path, and $O_2(\mathbf{r}, t)$ is updated.

(d) *Vessel dilatation.* In our model a vessel segment e at site $\mathbf{r}$ that is surrounded by TCs and has a GF concentration $GF(\mathbf{r}, t)$ larger than c_{GF} increases its radius $d(e)$ by an amount $r_0/2\pi$ with probability $\Delta\tau/T_e$ as long as $d(e) \leq d_{\text{max}}$. To mimic the smoothening effect caused by the surface tension of the vessel walls the location of the dilatation is shifted to a neighboring vessel segment if a radius difference larger than $r_0/2\pi$ would arise at the original location.

(e), (f) *Vessel regression and collapse.* Vessels can collapse due to solid stress exerted by the tumor and also long-term reduction of wall shear stress is associated with a dramatic reduction of the vessel diameter, up to complete vessel occlusion. We used both criteria to identify critical vessels: weakly perfused vessels e, which are surrounded by TCs, collapse with probability $p = \Delta\tau/T_{\text{collapse}}$ if the wall shear stress $f(e)$ is below a critical value f_{crit} [25]. After each collapse event the blood flow is re-computed and $O_2(\mathbf{r}, t)$ is updated. Vessels that are cut from the blood circulation ($q(e) = 0$) are instantaneously removed.

20.3 Results

In contrast to [8] we consider here a situation in which the MVD of the original vessel network is not sufficient for the tumor cells to proliferate and the secretion of GF is necessary to increase the MVD such that TCs can proliferate in certain regions of the tumor.

The original vasculature is modeled by a regular network of capillaries in a rectangular grid with vessel-to-vessel distance $\delta = 100 \ \mu$m. This implies that the original MVD is 10 vessel per mm (measured by counting the number of vessels in a vertical cut). If we extend the network into three dimensions with the same network parameters this implies an MVD of 100/mm^2, close to the value that is characteristic for skin tissue [4]. With $R_{\text{oxy}} = 100 \ \mu$m this yields an oxygen concentration that is nearly homogeneous and has a value of $\overline{O}_2 = 0.2$ (measured in the source strength of each vessel segment). Thus setting $c_{\text{oxy}} = 0.3$ implies that TCs can survive (since according to our definition they die only if their oxygen supply falls below $0.1c_{\text{oxy}} = 0.03$), but cannot proliferate in the original network.

We set R_{GF} to 150 μm. The value of c_{GF} is not crucial; if it is low it implies that new vessels can be generated within nearly the whole radius R_{GF} around a GF-secreting TC. We set it to $c_{\text{GF}} = 0.01$. Cells proliferate on the time scale of several hours, therefore we set the time step to $\Delta\tau = 1$ h. Other parameters are: collapse probability $\Delta\tau/T_{\text{collapse}} = 0.01$, critical shear force $f_{\text{crit}} = 0.5 f_0$ (where f_0 is the shear force in the original vasculature), sprout generation time $T_e = 40$ h, TC proliferation time $T_c = 10$ h, maximum sprout distance $M_{\text{max}} = 100 \ \mu$m, TC-survival time to $t_{\text{max}} = 100$ h, maximum vessel radius to $d_{\text{max}} = 35 \ \mu$m [4] and size of the initial tumor nucleus $N_0 = 1000$.

To obtain data for the stochastic time evolution of our model according to the dynamics defined above, we performed Monte Carlo simulations. The result of one representative run is shown in Fig. 20.2.

Tumor/vessel configurations at different times are shown in the middle column of Fig. 20.2: Starting from a regular vessel network the MVD in the peritumoral region is increased due to the supply of GFs from the tumor. Once the tumor grows over this highly vascularized region, vessels start to collapse, by which the MVD tumor center is continuously decreased until only a few thick vessels, surrounded by cuffs of TCs remain. TCs at a distance larger than R_{oxy} from these vessels die after time t_{max}, producing necrotic regions.

The left and right panels of Fig. 20.2 show the GF concentration $GF(x, y)$ and oxygen concentration $O_2(x, y)$, respectively, for the tumor/vessel configurations at time t. Both indicate roughly the spatial extent of the tumor and the region where new vessels can grow. Far away from the tumor it is $O_2(\mathbf{r}) = \overline{O}_2 = 0.2$ and only in the peritumoral region it is $O_2(\mathbf{r}) > c_{oxy} = 0.3$. Inside the tumor $O_2(\mathbf{r})$ is drastically reduced at later times, leading to under-oxygenation of TCs and thus to GF production. Consequently, inside the tumor $GF(\mathbf{r})$ is high, nearly one everywhere—except in the necrotic regions. Fig. 20.2 indicates a compartmentalization of the tumor into different shells characterized by MVD, vessel diameter and necrosis, as observed in real tumors [4]: A highly vascularized peritumoral regions, a well oxygenated tumor periphery and
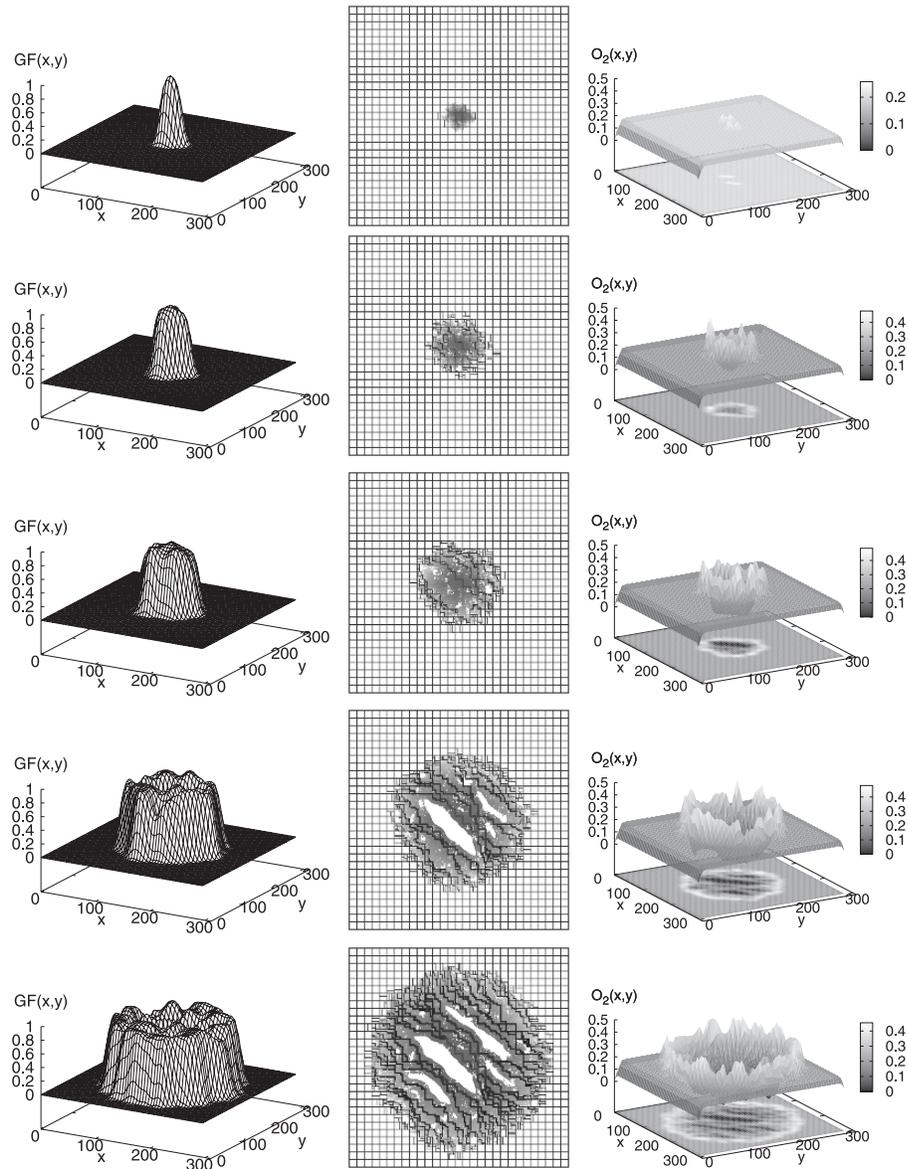
Fig. 20.2. Middle panel: Tumor and vessel network configuration at time $t = 1, 100, 200, 400$ and 600 (from top to bottom). The tumor is the grey area in the center—older TCs are darker than younger ones. The initial capillary network can be seen at $t = 1$ (with a few new vessels already there), vessel-to-vessel distance is $100 \, \mu$m. White areas are necrotic regions. **Left and right panels:** GF and oxygen concentration, respectively, for the configurations in the middle panel.
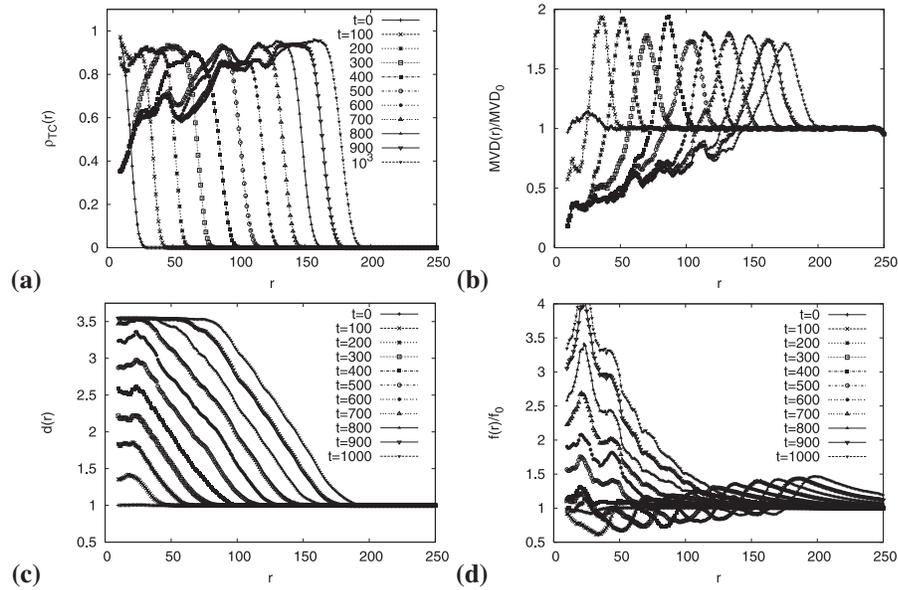
Fig. 20.3. (a) Tumor density as a function of the distance r from the tumor center for different equidistant times t. One sees clearly that the radius of the tumor grows linearly in time. **(b)** MVD/MVD$_0$ as a function of r for different equidistant times t (same symbols as in (a)). By comparison with the plots for tumor density in (a) one sees that the maximum MVD is localized at the tumor boundary, where it is up to 2 times larger than in the normal tissue. **(c)** Vessel diameter: It starts to increase linearly with decreasing r at the tumor boundary. For a fixed r below the actual tumor radius the average vessel radius increases linearly with time. **(d)** Shear force acting on the vessel walls, normalized to the shear force in the original capillaries. Note the pronounced dip: It is located at the maximum of the MVD in (b), i.e., at the tumor boundary: This is also the region where most vessels will collapse.

an hypoxic tumor center with decreased MVD, increased vessel diameter and large necrotic regions.

Fig. 20.3 presents a quantitative analysis of this dynamical evolution. Shown in Fig. 20.3(a) is the radial tumor density $\rho_{TC}(R)$. The tumor radius grows linearly with time t: $R_{TC}(t) - R_{TC}(0) \simeq 2t/t_{TC}$, where the factor 2 is typical for the Eden growth. The radial vessel density $MVD(R)$, shown in Fig. 20.3(b), is maximal at the tumor boundary at $R_{TC}(t)$. With increasing time both densities are substantially reduced inside the tumor, indicating the emergence of necrotic regions. The radial vessel diameter $d(R)$, shown in Fig. 20.3(c), increases linearly from 1 at $R \simeq R_{TC} + R_{GF}$ to d_{max} at the tumor center due to the continuous exposure of vessels to GF.

Such a characteristic vessel morphology is also in a quantitative agreement with experimental data presented in [4], where the morphometry of human malignant melanoma was analyzed and data for MVD and vessel perimeter were obtained in three different regions of the tumors: (I) the tumor center, (II) the tumor periphery—a 100 μm wide band of tumor immediately adjacent to the invasive edge and (III) the

peritumoral host tissue—a 200 μm wide band of host connective tissue immediately adjacent to the tumor periphery. It was found that for melanoma larger than 1.5 mm the MVD in (I) was less than 50% of the normal tissue MVD_0, in (II) it was ca. 50% more than MVD_0 and in (III) it was ca. two times MVD_0. Within the statistical error of the experimental data (up to 30%), this agrees reasonably well with our results.

We investigated extensively as to how far our conclusions depend on the parameters chosen here—a full account of this parameter dependence is reported in [8]. It turns out that the model behavior is robust and the parameters can be changed over a wide range without changing the qualitative results; in particular: the characteristic compartmentalization of the tumor, the vessel network morphology, the fractal dimension (to be studied below) and others. The question arises of which simplifications are crucial and what would change our conclusions, if they are abandoned in favor of a more realistic description. Obviously we cannot deal with all of them at once here, but we name a few:

(a) *Oxygen diffusion*. We modeled O_2 diffusion also by determining the source strength of each vessel implicitly as described in [26], also taking into account a TC density dependent O_2 consumption. Although this implies a major computational effort the quantitative change in the data (for fixed parameters) is only minimal—none of our conclusions is altered.

(b) *Growth factor diffusion*. Passive (as we assumed in our model) or active (via concentration gradients) diffusion changes the effective diffusion range (R_{GF}) slightly, but does not alter our conclusions.

(c) *Radius-dependent viscosity*. (to model the Fahraeus–Lindquist effect) does not change our results, as we checked.

(d) *Tumor growth*. More sophisticated tumor growth models can be incorporated than the simplified (Eden-growth inspired) model we use. We do not expect changes in our results as long as the tumor growth is restricted to a limited outer shell of TCs (as theoretically described in [21] and experimentally reported in [20]).

More serious assumptions are those of a regular network of original capillaries, including the boundary conditions for the pressure we have chosen. We are currently working on a version of the model that starts with a hierarchical arterio-venous network [18]; preliminary results indicate that the global picture that we present is maintained. Many other modifications are imaginable and will be topics of our future studies.

20.4 Fractal Dimension

The geometrical features of the emerging tumor vasculature in our model are obviously very different from the original, regular capillary network: the vasculature consists of a combination of dense and void regions that might possess fractal properties. We used the box-counting method to determine the fractal dimension D_f as $N_\epsilon \sim \epsilon^{-d_f}$ where N_ϵ is the number of boxes of volume ϵ^2 necessary to cover the tumor vessel network,
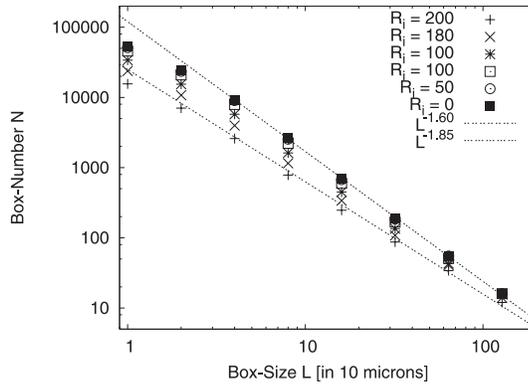
Fig. 20.4. Determination of the fractal dimension d_f of the vessel network at time $t = 1000$ via the box-counting method: The number of boxes of size L that is needed to cover completely the vasculature is plotted as a function of L in log-log scale. The slope d_f of the curve is the fractal dimension. We confined the measurement to concentric shell with fixed outer radius that is determined by the limit of the peritumoral plexus and with varying inner radius R_i. The slope of the curves decreases with increasing R_i: $d_f = 1.85 \pm 0.05$ for $R_i = 0$ (full squares, which corresponds to the complete tumor vasculature) and $d_f = 1.60 \pm 0.05$ for $R_i = 200$ (which corresponds to the peritumoral plexus exclusively), indicating that the fractal dimension is not a homogeneous measure over all regions of the tumor vasculature.

which is defined to lie within the outer limit of the peritumoral region. The plot of N_ϵ versus ϵ shown in Fig. 20.4 yields $d_f = 1.85 \pm 0.05$, which agrees with the value for the percolation cluster in conventional percolation in two dimensions [28]. We get the same value for a wide range of parameter values and also with other methods to estimate d_f. When we restrict the fractal analysis to concentric shells ($R_1 \leq R \leq R_2$) the estimates for d_f decrease systematically, decreasing the tumor center (see inset Fig. 20.4). Thus, the characteristic compartmentalization of the tumor vasculature is also reflected in the fractal properties.

Since d_f agrees with the value for the percolation cluster in two dimensions we conclude that the basic mechanism responsible for the fractal properties of the tumor vasculature in our model is the stochastic removal of vessels via vessel collapse and regression. In conventional percolation a critical cluster only emerges for an exactly tuned bond concentration. In our model the network is dynamically driven into this critical state without such a fine tuning since the removal of vessels is correlated with the blood flow: the collapse of weakly perfused vessels stabilizes the remaining ones due to an increase in blood flow. We propose that this mechanism is also at work in real tumors. Indeed the fractal analysis of two-dimensional photographs of vessel networks in human carcinoma yields a value of $d_f = 1.89 \pm 0.04$ [5], which agrees with d_f for the percolation cluster in 2D random percolation [28]. It has been suggested [5] that the origin of the fractal architecture of tumor vasculature might be based on an underlying invasion percolation process of the newly grown tumor vessels [29] due to inhomogeneities in the growth-supporting matrix. Our theoretical model does not involve any

such matrix inhomogeneities and we propose that it is rather the flow-correlated percolation process that determines the fractal properties of the tumor vasculature. Neovascularization mainly occurs at the tumor perimeter and a drastic reduction of vessel density is commonly observed in the interior of the tumor. Therefore, it appears unlikely that the fractal properties attained during growth in the periphery, independent of having characteristics of invasion percolation or not, survive the random dilution process in the tumor center. Thus, whereas in the proposal of [5] the fractal properties of the network are attained during *growth* of new vessels, we propose that it is mainly the *regression* of vessels that is responsible for the overall fractal morphology of the tumor vasculature.

20.5 Discussion

We have introduced a theoretical model for a dynamically evolving, two-dimensional vessel network interacting with a growing tumor, which is guided by experimental data for human melanoma. The emerging network morphology agrees well with those data and we find that the network is remodeled from a regular into a fractal structure with characteristics of conventional percolation. We have also implemented the model in three space dimensions and find similar results [30], where $d_f = 2.52$ turns out to correspond to 3D percolation. This also suggests for a large class of real solid tumor with decreased central MVD that the basic mechanism leading to the fractal features of the tumor vasculature is the random vessel collapse inside the tumor and not a stochastic vessel growth process.

An indirect experimental verification of this proposition is already given by the experimental data presented in [5, 7]: The fractal dimension of invasion percolation (growth-based mechanism), is 1.81 [29], whereas that of conventional percolation (collapse-based mechanism) is 1.891 [28]. Their data for the fractal dimension of the carcinoma vasculature is 1.88 ± 0.02, which is obviously closer to, if not identical with, conventional percolation, indicating at a collapse-based mechanism.

A direct proof of the proposition that vessel collapse is the relevant mechanism that leads to the fractal structure of the tumor vasculature is probably difficult but at least imaginable: We hypothesize that stabilization of tumor vessels without hampering the growth of new vessels could give evidence for one picture or the other. Antiangiogenic therapies with recurrent tumor growth may represent such a situation. Antiangiogenic blockade is known to lead to initial inhibition of vessel growth but also to vessel stabilization, and tumors may resume growth if the blockade continues for an extended period [31]. One of the possible mechanisms to such recurrence is the survival of some blood vessels which are stabilized by the adjacent smooth muscle cells and which survive vascular endothelial growth factor (VEGF) withdrawal, while other small vessels lacking vascular supportive cells collapse. Despite the upregulation of VEGF (possibility for new vessel growth) observed in all such tumors that resumed growth during prolonged antiangiogenesis, new capillaries were not detected in any. Instead large central vessels with significantly increased diameters and increased smooth muscle developed. If the structural characteristics of vessel network formation were growth

dependent (according to [5,7]) the newly synthetized VEGF and sprouting would result in growth dependent scale invariance, but this was not the case. Therefore the study of the fractal properties of tumor vasculature after antiangiogenic treatment could prove that in the long term vessel collapse is the relevant mechanism that leads to the fractal structure of the tumor vasculature.

Acknowledgments

References

1. Carmeliet, P., Jain, R.K.: Angiogenesis in cancer and other diseases. Nature, **407**, 249–257 (2000).
2. Acker, T., Plate, K.H.: A role for hypoxia and hypoxia-inducible transcription factors in tumor physiology. J. Mol. Med., **80**, 562–575 (2002).
3. Holash, J. Maisonpierre, P.C., Compton, D., Boland, P., Alexander, C.R., Zagzag, D., Yancopoulos, G.D., Wiegand, S.J.: Vessel cooption, regression, and growth in tumors mediated by angiopoietins and VEGF. Science, **284**, 1994–1998 (1999).
   Holash, J., Wiegand, S.J., Yancopoulos, G.D.: New model of tumor angiogenesis: Dynamic balance between vessel regression and growth mediated by angiopoietins and VEGF. Oncogene, **18**, 5356–5362 (1999).
4. Döme, B., Paku, S., Somlai, B., Tímár, J.: Vascularization of cutaneous melanoma involves vessel co-option and has clinical significance. J. Path., **197**, 355–362 (2002).
5. Gazit, Y., Berk, D.A., Leunig, M., Baxter, L.T., Jain, R.K.: Scale-Invariant Behavior and Vascular Network Formation in Normal and Tumor Tissue. Phys. Rev. Lett., **75**, 2428 (1995).
   Baish, J.W., Jain, R.K.: Cancer, angiogenesis and fractals [4]. Nature Med., **4**, 984 (1998).
6. Hlatky, L., Hahnfeld, P., Folkman, P.: Clinical application of antiangiogenic therapy: Microvessel density, what it does and doen't tell us. J. Nat. Canc. Inst., **94**, 883–893 (2002).
7. Baish, J.W., Jain, R.K.: Fractals and cancer. Canc. Res., **60**, 3683–3683 (2000).
8. Bartha, K., Rieger, H.: J. Theor. Biol., **241**, 903–918 (2006).
9. Mantzaris, N.V., Webb, S., Othmer, H.G.: Mathematical modeling of tumor-induced angiogenesis. J. Math. Biol., **49**, 111–187 (2004).
10. Preziosi, L. (ed.): Cancer Modelling and Simulation. Chapman & Hall/CRC, Boca Raton, FL, (2003).
11. Byrne, H.M., Chaplain, M.A.J.: Mathematical models for tumour angiogenesis: Numerical simulations and nonlinear wave solutions. B. Math. Biol., **57**, 461–486 (1995).
12. Levine, H.A., Sleeman, B.D., Nilsen-Hamiltion, M.: Mathematical modeling of the onset of capillary formation initiating angiogenesis. J. Math. Biol., **42**, 195–238 (2001).
13. Sansone, B.C., Scalerandi, M., Condat, C.A.: Emergence of taxis and synergy in angiogenesis. Phys. Rev. Lett., **87**, 128102 (2001).
14. Anderson, A.R.A., Chaplain, M.A.J.: Continuous and discrete mathematical models of tumor-induced angiogenesis. Bull. Math. Biol., **60**, 857–899 (1998).

15. McDougall, S.R., Anderson, A.R.A, Chaplain, M.A.J., Sherrat, J.A.: Mathematical modelling of flow through vascular networks: Implications for tumour-induced angiogenesis and chemotherapy strategies. B. Math. Biol., **64**, 673–702 (2002).
16. Tong, S., Yuan, F.: Numerical Simulations of Angiogenesis in the Cornea. Microvasc. Res., **61**, 14–27 (2002).
17. Sun, S.Y., Wheeler, M.F., Obeyesekere, M., Patrick, C.W.: A deterministic model of growth factor-induced angiogenesis. B. Math. Biol., **67**, 313–337 (2005).
18. Welter, M., Bartha, K., Rieger, H.: to be published.
19. Barabási, A.-L., Stanley, H.E.: Fractal Concepts in Surface Growth. Cambridge University Press, London, (1995).
20. Brú, A., Albertos, S., Subiza, J.L., García-Asenjo, J.L., Brú, I.: The universal dynamics of tumor growth. Biophys. J., **85**, 2948–1961 (2003).
21. Drasdo, D., Höme, S.: A single-cell-based model of tumor growth in vitro: Monolayers and spheroids. Phys. Biol., **2**, 133–147 (2005).
22. Patel, A.A., Gawlinsky, E.T., Lemieux, S.K., Gatenby, R.A.: A cellular automaton model of early tumor growth and invasion: The effects of native tissue vascularity and increased anaerobic tumor metabolism. J. Theor. Biol., **213**, 315–331 (2001).
23. Alarcon, T., Byrne, H.M., Maini, P.K.: A cellular automaton model for tumour growth in inhomogeneous environment. J. Theor. Biol., **225**, 257–274 (2003).
24. Popel, A.S.: Theory of oxygen transport to tissue. Crit. Rev. Biomed. Eng., **17**, 257–321 (1989).
25. Gödde, R., Kurz, H.: Structural and biophysical simulation of angiogenesis and vascular remodeling. Dev. Dyn., **220**, 387–401 (2001).
26. Hsu, R., Secomb, T.W.: A Green's function method for analysis of oxygen delivery to tissue by microvascular networks. Math. Biosci., **96**, 61–78 (1989).
27. Secomb, T.W., Hsu, R., Park, E.Y.H., Dewhirst, M.W.: Green's function methods for analysis of oxygen delivery to tissue by microvascular networks. Ann. Biomed. Eng., **32**, 1519–1529 (2004).
28. Stauffer, D., Aharony, A.: An Introduction to Percolation Theory, revised 2nd ed. Taylor and Francis, London (1994).
   Lorenz, C.D., Ziff, R.M.: Precise determination of the bond percolation thresholds and finite-size scaling corrections for the sc, fcc, and bcc lattices. Phys. Rev. E, **57**, 230–236 (1998).
29. Furuberg, L., Feder, J., Aharony, A., Jossang, T.: Dynamics of Invasion Percolation. Phys. Rev. Lett., **61**, 2117–2120 (1988).
   Sheppard, A.P., Knackstedt, M.A., Pinczewski, W.V., Sahimi, M.: Invasion percolation: New algorithms and universality classes. J. Phys. A, **32**, L521–L529 (1999).
30. S Lee, D., Rieger, H., Bartha, K.: Flow correlated percolation during vascular remodeling in growing tumors. Phys. Rev. Lett., **96**, 058104 (2006).
31. Huang, J.Z., Soffer, S.Z., Kim, E.S., McCrudden, K.W., New, T., Manley, C.A., Middlesworth, W., O'Toole, K., Yamashiro, D.J., Kandel, J.J.: Vascular Remodeling Marks Tumors That Recur During Chronic Suppression of Angiogenesis. Mol. Canc. Res., **2**, 36–42 (2004).

A Mathematical Model of the Cell Cycle and Its Circadian Control

Jean Clairambault,[1,4] Philippe Michel,[2,3] and Benoît Perthame[1,2]

[1] INRIA, projet BANG, Domaine de Voluceau, BP 105, F78153 Le Chesnay Cedex, France;
jean.clairambault@inria.fr

[2] Département de Mathématiques et Applications, UMR 8553, École Normale Supérieure, 45,
rue d'Ulm, F75230 Paris Cedex 05, France; Philippe.Michel@dma.ens.fr;
Benoit.Perthame@ens.fr

[3] CEREMADE, Université Paris 9 Dauphine, Place du Maréchal de Lattre de Tassigny,
F75775 Paris Cedex 16, France

[4] INSERM E 0354 "Rythmes biologiques et cancer", Hôpital Paul-Brousse, 14, Av.
Paul-Vaillant-Couturier, F94807 Villejuif Cedex, France

Summary. We address the following question: Can one sustain, on the basis of mathematical
models, that for cancer cells, the loss of control by circadian rhythm favours a faster popula-
tion growth? This question, which comes from the observation that tumour growth in mice is
enhanced by experimental disruption of the circadian rhythm, may be tackled by mathematical
modelling of the cell cycle. For this purpose we consider an age-structured population model
with control of death (apoptosis) rates and phase transitions, and two eigenvalues: one for pe-
riodic control coefficients (via a variant of Floquet theory in infinite dimension) and one for
constant coefficients (taken as the time average of the periodic case). We show by a direct proof
that, surprisingly enough considering the above-mentioned observation, the periodic eigenvalue
is always greater than the steady state eigenvalue when the sole apoptosis rate is concerned. We
also show by numerical simulations when transition rates between the phases of the cell cycle are
concerned, that, without further hypotheses, no natural hierarchy between the two eigenvalues
exists. This at least shows that, if such models are to take into account the above-mentioned ob-
servation, control of death rates inside phases is not sufficient, and that transition rates between
phases are a key target in proliferation control.

Key words: Cell cycle, age-structured population, population growth, circadian rhythm.

21.1 Cell Cycle and Circadian Rhythm

The goal of this chapter is to address by means of mathematical and numerical models
the following idea: circadian rhythms regulate cell proliferation, and their disruption
favours the growth of ill-controlled proliferative cell populations. In particular, tumour
growth has been shown to be favoured in mice by disruptions of the normal circa-
dian rhythm, as assessed, e.g., by central body temperature or rest-activity record-
ings [9, 10], both by surgical resection of suprachiasmatic nuclei and by jet-lag like

perturbations of the light-dark cycle. Also several epidemiological studies have shown that workers exposed to light-dark rhythm perturbations due to prolonged shift work are significantly more exposed to the risk of developing breast or colorectal cancer than others with regular work time schedules [8, 23]. It is thus suspected that a loss of circadian control on cell cycle dynamics may account for an acceleration in tumour progression. This is also supported by clinical observations according to which patients with cancer *and* disrupted circadian rhythms are less responsive to chemotherapy and have a poorer prognosis with shorter life expectancy than others with the same diseases but strong circadian rhythmicity [15, 20]. This idea is now sustained by a better understanding of the mechanisms underlying apoptosis and cell cycle phase transitions through proteins such as cyclins and p53. Indeed, the expression of cyclin-dependent kinases Cdk1 and Cdk2 and their dimerisation with Cyclins B and E, respectively, control the cell cycle phase transitions G2/M and G1/S, while protein p53 favours cell cycle arrest in G2/M or G1/S, by acting on these cyclins, and secondary apoptosis induction. And for instance, phosphorylation of the dimer CycB-Cdk1 by the kinase Wee1 is directly controlled by the circadian gene Bmal1, and p53 expression by the circadian gene Per1, see [3, 16, 21, 24, 26].

In this work (an abridged version of which has appeared in [6]), our approach relies on mathematical equations for the cell cycle which are well established nowadays. We introduce circadian control through periodic coefficients standing for phase transitions and apoptosis regulation by clock genes. We assess the hypothesis according to which periodicity diminishes the population growth as compared to constant coefficients (with the same average), i.e., we want to decide if a loss of circadian control theoretically favours tumour growth.

General references and experimental validations of the topic of age-structured population dynamics and cell cycle can be found in [1, 2, 4, 14, 17, 22]. For a more recent approach based on entropy properties, we refer to [18, 19]. Here and following earlier work [5], we model our population of cells by a partial differential equation for the density $n_i(t, x) \geq 0$ of cells with age x in the phase $i = 1, \ldots, I$, at time t.

$$
\begin{cases}
\dfrac{\partial}{\partial t} n_i(t, x) + \dfrac{\partial}{\partial x} n_i(t, x) + [d_i(t, x) + K_{i \to i+1}(t, x)] n_i(t, x) = 0, \\[2ex]
n_i(t, x = 0) = \displaystyle\int_{x' \geq 0} K_{i-1 \to i}(t, x') \, n_{i-1}(t, x') \, dx', \quad 2 \leq i \leq I, \\[2ex]
n_1(t, x = 0) = 2 \displaystyle\int_{x' \geq 0} K_{I \to 1}(t, x') \, n_I(t, x') \, dx'.
\end{cases}
\tag{21.1}
$$

Here and below we identify $I + 1$ to 1. We denote by $d_i(t, x) \geq 0$ the apoptosis rate, and by $K_{i \to i+1}$ the transition rates from one phase to the next; $(i = I)$ is mitosis where the two cells separate. These coefficients can be constant in time (no circadian control) or time T-periodic in order to take into account the circadian rhythm. Our assumptions are

$$
K_{i \to i+1}(t, x) \geq 0, \; d_i(t, x) \geq 0 \quad \text{are bounded,}
\tag{21.2}
$$

and, setting $\min_{0 \leq t \leq T} K_{i \to i+1}(t, x) := k_{i \to i+1}(x), \max_{0 \leq t \leq T}[d_i + K_{i \to i+1}] := \mu_i(x)$ (lower bound for the transition kernels and upper bound for the cell loss terms), $M_i(x) = \int_0^x \mu_i(y) dy$,

$$\prod_{i=1}^{I} \int_0^\infty k_{i \to i+1}(y) e^{-M_i(y)} dy > 1/2, \tag{21.3}$$

(thus ensuring positive population growth). Classically, one can introduce the growth rate of the system λ_{per} (Malthus parameter, first eigenvalue) such that there is a unique T-periodic **positive** solution to

$$\begin{cases} \dfrac{\partial}{\partial t} N_i(t, x) + \dfrac{\partial}{\partial x} N_i(t, x) + [d_i(t, x) + \lambda_{\text{per}} + K_{i \to i+1}(t, x)] N_i(t, x) = 0, \\[2mm] N_i(t, x = 0) = \displaystyle\int_{x' \geq 0} K_{i-1 \to i}(t, x') \, N_{i-1}(t, x') \, dx', \quad 2 \leq i \leq I, \\[2mm] N_1(t, x = 0) = 2 \displaystyle\int_{x' \geq 0} K_{I \to 1}(t, x') \, N_I(t, x') \, dx', \\[2mm] \displaystyle\sum_{i=1}^{I} \int N_i(t, x) \, dx = 1. \end{cases}$$

$$\tag{21.4}$$

Under our assumptions 21.2–21.3, the existence of a solution to 21.4, with $\lambda_{\text{per}} > 0$, follows from an infinite-dimensional version of Floquet theory and one has (see for instance [18])

$$\sum_i \int \left| n_i(t, x) e^{-\lambda_{\text{per}} t} - \rho N_i(t, x) \right| \varphi_i(t, x) dx \to 0 \qquad \text{as } t \to \infty,$$

where $\varphi_i(t, x)$ denotes the periodic positive solution to the adjoint problem to 21.4 normalised by $\sum_i \int N_i(t, x) \varphi_i(t, x) \, dx = 1$, and $\rho = \sum_{i=1}^{N} \int n_i(t = 0, x) \varphi_i(t = 0, x) \, dx$. In other words, the periodic solution is the observed stable state after renormalisation by the rate λ_{per}.

One can also introduce the coefficients averaged in time,

$$\langle K_{i \to i+1}(x) \rangle := \frac{1}{T} \int_0^T K_{i \to i+1}(t, x) \, dt, \qquad \langle d_i(t, x) \rangle := \frac{1}{T} \int_0^T d_i(t, x) \, dt,$$

and consider the associated steady state solution. This allows us to define another growth rate λ_s, and a steady state solution $\bar{N}_i$ to

$$
\begin{cases}
\dfrac{\partial}{\partial x}\bar{N}_i(x) + [\langle d_i(x)\rangle + \lambda_s + \langle K_{i\to i+1}(x)\rangle]\bar{N}_i(x) = 0, \\[2ex]
\bar{N}_i(x=0) = \displaystyle\int_{x'\geq 0} \langle K_{i-1\to i}(x')\rangle\, \bar{N}_{i-1}(x')\, dx', \quad 2 \leq i \leq I, \\[2ex]
\bar{N}_1(x=0) = 2\displaystyle\int_{x'\geq 0} \langle K_{I\to 1}(x')\rangle\, \bar{N}_I(x')\, dx', \\[2ex]
\displaystyle\sum_{i=1}^{I} \int \bar{N}_i(x)\, dx = 1.
\end{cases}
\tag{21.5}
$$

For these problems, we address the hypothesis that circadian control reduces the population growth, i.e., $\lambda_{\mathrm{per}} \leq \lambda_s$ (index *per* standing for "periodic" and *s* for "stationary"). In Section 21.2, we first study the effect of small variations, with respect to a circadian control, from constant transition and apoptosis rates on the resulting eigenvalue λ_{per}. Then in Section 21.3, we prove that, surprisingly enough, a result opposite to our experimental conjecture is true, i.e., $\lambda_{\mathrm{per}} \geq \lambda_s$, when the circadian control acts only on the apoptosis rate. In Section 21.4, we show by numerical experiments that no hierarchy exists between the two eigenvalues when the circadian control acts on the transition rate $K_{1\to 2}$ in a reduced two-phase model. These results give hints toward designing physiologically based models of the cell cycle for cancer therapeutics; that is discussed in Section 21.5. The results are summarised in Section 21.6.

21.2 Analysis of Local Variation by Small Oscillations

In this section, we study small variations, with respect to a circadian control, of the growth rate λ and we show that its effect is only of the second order.

To do so, we consider that the transition kernels and the death rates show small variation of order $\varepsilon > 0$ from their averages. Therefore we set

$$
\begin{aligned}
\tilde{K}^{\varepsilon}_{i-1\to i}(t, x) &:= \varepsilon \tilde{K}_{i-1\to i}(t, x) + \langle K_{i-1\to i}(x)\rangle, \\
\tilde{K}^{\varepsilon}_{I\to 1}(t, x) &:= \varepsilon \tilde{K}_{I\to 1}(t, x) + \langle K_{I\to 1}(x)\rangle, \\
d^{\varepsilon}_i(t, x) &:= \varepsilon \tilde{d}_i(t, x) + \langle d_i(x)\rangle,
\end{aligned}
$$

where the quantities $\tilde{d}_i$, $\tilde{K}_{i\to j}$ have vanishing averages:

$$
\langle \tilde{K}_{i-1\to i}(x)\rangle = \langle \tilde{K}_{I\to 1}(x)\rangle = \langle \tilde{d}_i(x)\rangle = 0.
$$

Then we define the solution $n^{\varepsilon}_i(t, y)$ to the cell cycle system

$$
\begin{cases}
\dfrac{\partial}{\partial t} n_i^\varepsilon(t, x) + \dfrac{\partial}{\partial x} n_i^\varepsilon(t, x) + [d_i^\varepsilon(t, x) + \tilde{K}_{i \to i+1}(t, x)] n_i^\varepsilon(t, x) = 0, \\[2mm]
n_i^\varepsilon(t, 0) = \displaystyle\int_{x' \geq 0} \tilde{K}_{i-1 \to i}^\varepsilon(t, x') n_{i-1}^\varepsilon(t, x') \, dx', \quad 2 \leq i \leq I, \\[2mm]
n_1^\varepsilon(t, x = 0) = 2 \displaystyle\int_{x' \geq 0} K_{I \to 1}^\varepsilon(t, x') \, n_I(t, x') \, dx'.
\end{cases}
\tag{21.6}
$$

Now, using the results recalled in Section 21.1, we know that, for all $\varepsilon \in [0, 1]$, there exist eigen elements associated with this problem, $(N_i^\varepsilon, \lambda_\varepsilon, \varphi_i^\varepsilon)$. They are solutions to

$$
\begin{cases}
\dfrac{\partial}{\partial t} N_i^\varepsilon(t, x) + \dfrac{\partial}{\partial x} N_i^\varepsilon(t, x) + [d_i^\varepsilon(t, x) + \lambda_\varepsilon + K_{i \to i+1}^\varepsilon(t, x)] N_i^\varepsilon(t, x) = 0, \\[2mm]
N_i^\varepsilon(t, x = 0) = \displaystyle\int_{x' \geq 0} K_{i-1 \to i}^\varepsilon(t, x') \, N_{i-1}^\varepsilon(t, x') \, dx', \quad 2 \leq i \leq I, \\[2mm]
N_1^\varepsilon(t, x = 0) = 2 \displaystyle\int_{x' \geq 0} K_{I \to 1}^\varepsilon(t, x') \, N_I^\varepsilon(t, x') \, dx' \qquad \displaystyle\sum_{i=1}^{I} \int N_i^\varepsilon(t, x) \, dx = 1,
\end{cases}
\tag{21.7}
$$

$$
\begin{cases}
-\dfrac{\partial}{\partial t} \varphi_i^\varepsilon(t, x) - \dfrac{\partial}{\partial x} \varphi_i^\varepsilon(t, x) + [d_i^\varepsilon(t, x) + \lambda_\varepsilon + K_{i \to i+1}^\varepsilon(t, x)] \varphi_i^\varepsilon(t, x) \\[2mm]
\qquad = \varphi_{i+1}^\varepsilon(t, 0) K_{i \to i+1}^\varepsilon(t, x), \quad 1 \leq i \leq I - 1, \\[2mm]
-\dfrac{\partial}{\partial t} \varphi_I^\varepsilon(t, x) - \dfrac{\partial}{\partial x} \varphi_I^\varepsilon(t, x) + [d_I^\varepsilon(t, x) + \lambda_\varepsilon + K_{I \to 1}^\varepsilon(t, x)] \varphi_I^\varepsilon(t, x) \\[2mm]
\qquad = 2 \varphi_1^\varepsilon(t, 0) K_{I \to 1}^\varepsilon(t, x),
\end{cases}
\tag{21.8}
$$

with

$$
\int_0^\infty \sum_{i=1}^{I} N_i^\varepsilon(t, x) \varphi_i^\varepsilon(t, x) \, dx = 1, \quad \forall t \geq 0.
\tag{21.9}
$$

With this notation, we clearly have $\lambda_0 = \lambda_s$. As a first step towards our main result, we gather some formulae that are used to prove the following theorem.

Theorem 1 *The function $\lambda \mapsto \lambda_\varepsilon$ is differentiable for all $\varepsilon \in\,]0, 1[$, and*

$$
\begin{aligned}
\frac{d\lambda_\varepsilon}{d\varepsilon} = {}& \frac{1}{T} \int_0^T \int_0^\infty N_I^\varepsilon(t, x) \Big[2 \tilde{K}_{I \to 1}(t, x) \varphi_1^\varepsilon(t, 0) \\
& \qquad\qquad - \big(\tilde{d}_I(t, x) + \tilde{K}_{I \to 1}(t, x) \big) \varphi_I^\varepsilon(t, x) \Big] dx \, dt \\
& + \frac{1}{T} \int_0^T \int_0^\infty \sum_{i=1}^{I-1} N_i^\varepsilon(t, x) \Big[\tilde{K}_{i \to i+1}(t, x) \varphi_{i+1}^\varepsilon(t, 0) \\
& \qquad\qquad - \big(\tilde{d}_i(t, x) + \tilde{K}_{i \to i+1}(t, x) \big) \varphi_i^\varepsilon(t, x) \Big] dx \, dt.
\end{aligned}
\tag{21.10}
$$

The proof of the theorem is somewhat lengthy and will be detailed in an appendix; see Section 21.7.

Corollary 1 *For small circadian effect ε, the variations of λ_ε are of order ε^2; in other words,*

$$\frac{d\lambda_\varepsilon}{d\varepsilon}\bigg|_{\varepsilon=0} = 0.$$

This corollary follows from the fact that for $\varepsilon = 0$ the functions $N_i^\varepsilon(t, x)$ and $\varphi_i^\varepsilon(t, x)$ are independent of time. Therefore, in (21.10), we are left only with the time averages of $\tilde{K}_{i \to i+1}(t, x)$ and $\tilde{d}_i(t, x)$, which vanish.

We can also deduce from Theorem 1 that, in the particular case when $K_{i \to i+1}$ is independent of time and d_i is independent of age, we cannot control locally the growth rate λ (see also Section 21.3 for a direct proof and a derivation of a global variation in a more general case). Indeed, we have the following.

Corollary 2 *Assume $\tilde{d}_i(t, x) = \rho_i(t)$, $\tilde{K}_{i \to i+1}(t, x) = 0$. Then*

$$\frac{d\lambda_\varepsilon}{d\varepsilon} = 0, \tag{21.11}$$

and $\lambda_s = \lambda_{\mathrm{per}}$.

Proof Using (21.10), we find

$$\frac{d\lambda_\varepsilon}{d\varepsilon} = -\frac{1}{T} \int_0^T \int_0^\infty \sum_{i=1}^I N_i^\varepsilon(t, x)\varphi_i^\varepsilon(t, x)dx\rho_i(t)\, dt,$$

but we have $\int_0^T \rho_i(t)\, dt = 0$ and (21.9); thus we find (21.11) and

$$\lambda_{\mathrm{per}} - \lambda_s = \int_0^1 \frac{d\lambda_\varepsilon}{d\varepsilon} = 0. \qquad \square$$

As a conclusion of this section, we see that a direct computation in the most general case, when $K_{i \to i+1}$ and d_i are time dependent, leads to hardly tractable formulae; the local variation of the first eigenvalue cannot be found directly because it is of second order in ε. For this reason it is natural to turn to numerical computations, as we do in Section 21.4.

21.3 Control Exerted on Apoptosis

In this section, we consider the case when the circadian control only acts on apoptosis, i.e., $K_{i \to i+1}$ depends only upon x.

Theorem 2 *Assume that $d_i(t, x) \geq 0$, $K_{i \to i+1}(x) \geq 0$ are bounded and that 21.3 holds. Then the eigenvalue problems 21.4, 21.5 have unique solutions $(\lambda_{\mathrm{per}}, N(t, x))$, $(\lambda_s, \bar{N}(x))$, and*

$$\lambda_{\mathrm{per}} \geq \lambda_s. \tag{21.12}$$

Proof The existence part for the two problems is standard and we do not prove it again (see [5, 18]). For the ordering of eigenvalues, consider the function $q_i(x) = \langle \log(N_i(t, x)/\bar{N}_i(x)) \rangle$. It satisfies

$$\frac{\partial}{\partial x} q_i + \lambda_{\text{per}} - \lambda_s = 0,$$

with

$$q_i(x = 0) = \left\langle \log\left[\int K_{i-1 \to i}(x) \frac{\bar{N}_{i-1}(x)}{\bar{N}_i(0)} \frac{N_{i-1}(t, x)}{\bar{N}_{i-1}(x)} dx \right] \right\rangle.$$

Since $d\mu_i(x) = K_{i-1 \to i}(x)[\bar{N}_{i-1}(x)/\bar{N}_i(0)] dx$ is a probability measure because of the condition $\bar{N}_i(0)$ (a factor 2 should be included for $i = 1$), we also have

$$q_i(x = 0) \geq \left\langle \int \log \frac{N_{i-1}(t, x)}{\bar{N}_{i-1}(x)} d\mu_i(x) \right\rangle \qquad \text{(by Jensen's inequality)}$$

$$= \int q_{i-1}(x) \, d\mu_i(x)$$

$$= \int [q_{i-1}(0) + (\lambda_s - \lambda_{\text{per}})x] \, d\mu_i(x).$$

Therefore, summing over i from 1 to I,

$$0 \geq (\lambda_s - \lambda_{\text{per}}) \sum_{i=1}^{I} \int_{x=0}^{\infty} x \, d\mu_i(x),$$

and the result follows. $\square$

Notice that in [7], the same question has been addressed for comparing the eigenvalues of matrices with positive coefficients, and with either constant or periodic diagonal terms.

21.4 Control Exerted on Phase Transitions

We have performed numerical tests for the cell cycle systems 21.4, 21.5 based on a classical upwind scheme with the CFL (Courant–Friedrichs–Levy) stability condition $CFL = 1$. This means in our context that we chose equal integration steps in age and time ($\Delta a = \Delta t$), which gives the exact transport solver (see [5] for details). We have taken a simplified version of the cell cycle with two phases ($I = 2$): G1-S-G2 and M. In other words, in the full cell cycle (G1, synthesis, G2, mitosis) we only retain as a major event the transition from G2 to M. The apoptosis rate has been taken as constant and the transition rates are

$$K_{1 \to 2}(t, x) = \mathbb{1}_{[x_*, +\infty[}(x), \qquad K_{2 \to 1}(t, x) = \mathbb{1}_{[x_{**}, +\infty[}(x),$$

Table 21.1. *The periodic and stationary eigenvalues for different duration G1-S-G2/M ratios and for two periodic phase transition functions: ψ_1 is a brief square wave (4 h/24 h), ψ_2 a longer one (12 h/12 h). For the reader's convenience, the greater of the two eigenvalues is underlined.*

Time ratio, ψ_1	λ_{per}	λ_s	Time ratio, ψ_2	λ_{per}	λ_s
1	<u>0.2385</u>	0.2350	1	0.2623	<u>0.2821</u>
2	0.2260	<u>0.2923</u>	2	0.3265	<u>0.3448</u>
3	0.2395	<u>0.3189</u>	3	. . .	. . .
4	0.2722	<u>0.3331</u>	4	. . .	. . .
5	0.3065	<u>0.3427</u>	5	. . .	. . .
6	0.3305	<u>0.3479</u>	6	. . .	. . .
7	0.3472	<u>0.3517</u>	7	0.4500	<u>0.4529</u>
8	<u>0.3622</u>	0.3546	8	<u>0.4588</u>	0.4575
10	<u>0.3808</u>	0.3588	10	<u>0.4713</u>	0.4641
20	<u>0.4125</u>	0.3675	20	<u>0.5006</u>	0.4818

where ψ is a periodic function of time t and $\mathbb{1}_I$ is the indicator function of interval I. We have in mind the following order of magnitudes for several animal tumour cells: total cycle duration is 21 h, 8 h for G1, 8 h for S, 4 h for G2, 1 h for M (therefore in this case $x_* = 20$ h and $x_{**} = 1$ h). But we will also consider different duration ratios x_*/x_{**} between the two phases G1-S-G2 and M, from 1 to 20. The reason for this is that although the G2/M transition is known to be a circadian control target with an identified mechanism (Bmal1→Wee1→Cdc2—the cyclin-dependent kinase Cdc2 being rather known as Cdk1 in mammals), another control target, with an as yet unidentified mechanism (though the genes *per* and *cMyc* have been shown to be involved [11, 12]), could take place at the G1/S transition, and the G1 phase may have a very variable duration. Thus, while in principle testing here the G2/M transition, we may also be testing the G1/S gate control by an unknown 24 h-rhythmic Cdc2 (Cdk1)-like factor. The function $\psi(t)$ has a 24 h period. We have tested several shapes for ψ (*cosine* and square wave functions), but eventually kept only two square waves, a brief one with 4 hours at value 1 and the remaining 20 hours at 0, the other one with 12 hours at 1 and 12 hours at 0. The first one mimics the shape of the Cdc2 kinase behaviour, with entrainment by 24 h-rhythmic Wee1, according to A. Goldbeter's model of the mitotic oscillator [13], the other a version of the same Cdc2 model, with no entrainment, but with fixed coefficients also yielding a 24 h period. In Table 21.1 we show a comparison between the two eigenvalues (periodic and stationary), for the two tested ψ periodic transition functions.

It is apparent from this table that no clear hierarchy can be seen between the two eigenvalues, even if some regularity may be suspected, and these simulations show cases favourable to our initial hypothesis in the interval $2 \leq$ G1-S-G2/M ≤ 7.

21.5 Discussion

Circadian control in general. Theorem 1 in Section 21.2 shows that in the general case one cannot drive any conclusion on the initial question: Is the growth of the population hampered by a periodic control on phase transitions and death rates? This negative result invited us to examine particular cases.

Circadian control of population growth by targeting apoptosis. The rather surprising result, given previous experimental observations, that periodic control of apoptosis in one phase resulted in enhanced proliferation, as compared to constant control (Theorem 2 in Section 21.3), may indicate that apoptosis is no physiological target for population growth control by the circadian clock. As regards p53, its circadian expression is thus likely to be linked more to its cell cycle arrest than to its apoptosis-inducing capacities.

Circadian control of population growth by targeting phase transitions. This case was simulated in a simplified way, with two different sorts of square waves in a reduced two-phase cell cycle model, and even in these simple settings, we obtained contrasting results, which did not enable us to answer the initial question. It is likely that two phases only in the model may not be sufficient to account for the physiopathological observation which guided us for this modelling work, and that, as it is, this model aggregates in an inaccurate way physiological effects of the G1/S and G2/M transition controls. Future work on the basis of this experimental observation should encompass three phases: G1, S-G2, and M, better knowledge of circadian control both at the G1/S and G2/M transitions, and synchronisation between these transitions.

Possible medical implications. Circadian rhythms are nonnegligible regulating factors of cell population growth, as shown by results from experimental and clinical observations. Cancer chronotherapy [15,20] for fifteen years has been taking advantage of circadian variations in cytotoxic drug therapeutic efficacy and unwanted toxicity. These variations are likely due to the simultaneous regulation by the circadian clock of cell cycle progression on the one hand, and of drug detoxification mechanisms on the other hand. Elucidating both these control processes in a theoretical way might give a better rationale to anticancer treatment optimisation. But other *adjuvant* therapies may be used, aiming at resynchronisation of individual proliferating cells in growing populations by strengthening circadian control, possibly using hormones such as melatonin and cortisol, or even feeding schedule [10,25]. For this prospect, the present work may offer new guidelines for designing such therapeutic control processes.

21.6 Conclusion

To summarise these results:

1. This model allows us to study the interactions in proliferating tissues between the cell cycle and physiological control systems such as the circadian clock. We have

shown with Theorem 1 and its corollaries that classical perturbation analysis is unlikely to provide an answer to the initial question of the effect of circadian control on population growth.

2. More than two phases and a better knowledge of other mechanisms (cortisol, Cyclin E on G1/S) might be necessary to account for the physiopathological facts reported from animal experimentation and human clinical observations which guided us in this investigation of the first eigenvalues of the periodic and stationary problems.

3. The unexpected result $\lambda_{\text{per}} \geq \lambda_s$ for apoptosis control shown in Theorem 2 suggests that the sole control of death rate inside cell cycle phases is unable to describe control of proliferation by cytotoxic drugs in cancer treatment. Transition rates should be considered in a therapeutic perspective.

21.7 Appendix: Proof of Theorem 1

First we introduce more condensed definitions: $(N^\varepsilon, \lambda^\varepsilon, \varphi^\varepsilon)$ by

$$\forall\, (t, y) \in [0, \infty[^2, \quad N^\varepsilon(t, y) \in [0, \infty[^I, \quad N^\varepsilon(t, y)|_i := N_i^\varepsilon(t, y),$$
$$\forall\, (t, y) \in [0, \infty[^2, \quad \varphi^\varepsilon(t, y) \in [0, \infty[^I, \quad \varphi^\varepsilon(t, y)|_i := \varphi_i^\varepsilon(t, y),$$
$$\lambda^\varepsilon := \lambda_\varepsilon.$$

Then, we define the operator $\mathcal{L}_\varepsilon^*$ such that $\mathcal{L}_\varepsilon^* \varphi^\varepsilon = \lambda^\varepsilon \varphi^\varepsilon$,

$$\mathcal{L}_\varepsilon^*(g)|_i := \frac{\partial}{\partial t} g_i(t, x) + \frac{\partial}{\partial x} g_i(t, x) - [d_i^\varepsilon(t, x) + K_{i \to i+1}^\varepsilon(t, x)] g_i(t, x)$$
$$+ g_{i+1}(t, 0) K_{i \to i+1}^\varepsilon(t, x), \quad 1 \leq i \leq I - 1,$$
$$\mathcal{L}_\varepsilon^*(g)|_I := \frac{\partial}{\partial t} g_I(t, x) + \frac{\partial}{\partial x} g_I(t, x) - [d_I^\varepsilon(t, x) + K_{I \to 1}^\varepsilon(t, x)] g_I(t, x)$$
$$+ 2 g_1(t, 0) K_{I \to 1}^\varepsilon(t, x),$$

and its dual satisfying $\mathcal{L}^\varepsilon N^\varepsilon = \lambda^\varepsilon N^\varepsilon$. Thus, for all ε and ε' such that ε and $\varepsilon - \varepsilon' \in\,]0, 1[$, we have

$$\lambda^\varepsilon = \int_0^\infty \mathcal{L}_\varepsilon^*(\varphi^\varepsilon)(y) N^\varepsilon(y)\, dy.$$

Therefore, we find

$$\lambda^\varepsilon - \lambda^{\varepsilon - \varepsilon'} = \int_0^\infty \mathcal{L}_\varepsilon^*(\varphi^\varepsilon)(y) N^\varepsilon(y) dy - \int_0^\infty \mathcal{L}_{\varepsilon - \varepsilon'}^*(\varphi^{\varepsilon - \varepsilon'})(y) N^{\varepsilon - \varepsilon'}(y)\, dy.$$

But, the normalisation gives

$$\int_0^\infty \varphi^\varepsilon(y) N^\varepsilon(y)\, dy = \int_0^\infty \varphi^{\varepsilon - \varepsilon'}(y) N^{\varepsilon - \varepsilon'}(y)\, dy = 1, \qquad (21.13)$$

and so, we can write

$$\lambda^\varepsilon - \lambda^{\varepsilon - \varepsilon'} = \int_0^\infty \Big(\mathcal{L}_\varepsilon^*(\varphi^\varepsilon)(y) - \mathcal{L}_{\varepsilon - \varepsilon'}^*(\varphi^\varepsilon)(y) \Big) N^{\varepsilon - \varepsilon'}(y) \, dy$$

$$+ \int_0^\infty \mathcal{L}_\varepsilon^*(\varphi^\varepsilon)(y) \Big(N^\varepsilon(y) - N^{\varepsilon - \varepsilon'}(y) \Big) \, dy$$

$$- \int_0^\infty \Big(\mathcal{L}_{\varepsilon - \varepsilon'}^*(\varphi^{\varepsilon - \varepsilon'})(y) - \mathcal{L}_{\varepsilon - \varepsilon'}^*(\varphi^\varepsilon)(y) \Big) N^{\varepsilon - \varepsilon'}(y) \, dy.$$

Using the definition of $\mathcal{L}^*$, $\mathcal{L}$, and their duality, we find

$$\lambda^\varepsilon - \lambda^{\varepsilon - \varepsilon'} = \int_0^\infty \Big(\mathcal{L}_\varepsilon^*(\varphi^\varepsilon)(y) - \mathcal{L}_{\varepsilon - \varepsilon'}^*(\varphi^\varepsilon)(y) \Big) N^{\varepsilon - \varepsilon'}(y) \, dy$$

$$+ \lambda^\varepsilon \int_0^\infty \varphi^\varepsilon(y) \Big(N^\varepsilon(y) - N^{\varepsilon - \varepsilon'}(y) \Big) \, dy$$

$$- \lambda^{\varepsilon - \varepsilon'} \int_0^\infty \Big(\varphi^{\varepsilon - \varepsilon'}(y) - \varphi^\varepsilon(y) \Big) N^{\varepsilon - \varepsilon'}(y) \, dy.$$

Thus, using the normalisation (21.13), we deduce from the above identity

$$\lambda^\varepsilon - \lambda^{\varepsilon - \varepsilon'} = \int_0^\infty \Big(\mathcal{L}_\varepsilon^*(\varphi^\varepsilon)(y) - \mathcal{L}_{\varepsilon - \varepsilon'}^*(\varphi^\varepsilon)(y) \Big) N^{\varepsilon - \varepsilon'}(y) \, dy$$

$$+ \lambda^\varepsilon \int_0^\infty \varphi^\varepsilon(y) \Big(N^\varepsilon(y) - N^{\varepsilon - \varepsilon'}(y) \Big) \, dy$$

$$- \lambda^{\varepsilon - \varepsilon'} \int_0^\infty \varphi^\varepsilon(y) \Big(N^\varepsilon(y) - N^{\varepsilon - \varepsilon'}(y) \Big) \, dy,$$

and

$$\lambda^\varepsilon - \lambda^{\varepsilon - \varepsilon'} = \int_0^\infty \Big(\mathcal{L}_\varepsilon^*(\varphi^\varepsilon)(y) - \mathcal{L}_{\varepsilon - \varepsilon'}^*(\varphi^\varepsilon)(y) \Big) N^{\varepsilon - \varepsilon'}(y) \, dy$$

$$+ \lambda^\varepsilon \int_0^\infty \Big(\varphi^{\varepsilon - \varepsilon'}(y) - \varphi^\varepsilon(y) \Big) N^{\varepsilon - \varepsilon'}(y)$$

$$- \lambda^{\varepsilon - \varepsilon'} \int_0^\infty \Big(\varphi^{\varepsilon - \varepsilon'}(y) - \varphi^\varepsilon(y) \Big) N^{\varepsilon - \varepsilon'}(y).$$

And so, we arrive at

$$(\lambda^\varepsilon - \lambda^{\varepsilon - \varepsilon'}) \left(1 - \int_0^\infty \Big(\varphi^{\varepsilon - \varepsilon'}(y) - \varphi^\varepsilon(y) \Big) N^{\varepsilon - \varepsilon'}(y) \, dy \right)$$

$$= \int_0^\infty \Big(\mathcal{L}_\varepsilon^*(\varphi^\varepsilon)(y) - \mathcal{L}_{\varepsilon - \varepsilon'}^*(\varphi^\varepsilon)(y) \Big) N^{\varepsilon - \varepsilon'}(y) \, dy.$$

Using the equality

$$(\mathcal{L}_*^\varepsilon - \mathcal{L}_*^{\varepsilon - \varepsilon'})(g) = \varepsilon' \big(- [\tilde{d}_i + \tilde{K}_{i \to i+1}] g_i + g_{i+1}(0) \tilde{K}_{i \to i+1} \big),$$

we deduce

$$(\lambda^\varepsilon - \lambda^{\varepsilon-\varepsilon'}) \left(1 - \int_0^\infty \left(\varphi^{\varepsilon-\varepsilon'}(y) - \varphi^\varepsilon(y)\right) N^{\varepsilon-\varepsilon'}(y) \, dy\right)$$

$$= \varepsilon' \int_0^\infty \left(-[\tilde{d}_i + \tilde{K}_{i\to i+1}]\varphi_i^\varepsilon + \varphi_{i+1}^\varepsilon(0)\tilde{K}_{i\to i+1}\right) N^{\varepsilon-\varepsilon'}(y) \, dy.$$

And finally, we obtain

$$\frac{\lambda^\varepsilon - \lambda^{\varepsilon-\varepsilon'}}{\varepsilon'} = \frac{\displaystyle\int_0^\infty \left(-[\tilde{d}_i + \tilde{K}_{i\to i+1}]\varphi_i^\varepsilon + \varphi_{i+1}^\varepsilon(0)\tilde{K}_{i\to i+1}\right) N^{\varepsilon-\varepsilon'}(y) \, dy}{1 - \displaystyle\int_0^\infty \left(\varphi^{\varepsilon-\varepsilon'}(y) - \varphi^\varepsilon(y)\right) N^{\varepsilon-\varepsilon'}(y) \, dy}. \quad (21.14)$$

Using Lebesgue's dominated convergence theorem, we can pass to the limit and find that the function $\varepsilon \mapsto \lambda^\varepsilon$ (i.e., λ_ε) is differentiable and (21.10) is satisfied. □

Acknowledgments

References

1. Arino, O.: A survey of structured cell population dynamics. Acta Biotheor., **43**, 3–25 (1995).
2. Basse, B., Baguley, B.C., Marshall, E.S., Joseph, W.R., van Brunt, B., Wake, G., Wall, D.J.N.: A mathematical model for analysis of the cell cycle in cell lines derived from human tumors. J. Math. Biol., **47**, 295–312 (2003).
3. Bjarnason, G., Jordan, R.C.K., Sothern, R.B.: Circadian variations in the expression of cell-cycle proteins in human oral epithelium. Am. J. Pathol., **154**, 613–622 (1999).
4. Chiorino, G., Metz, J.A.J., Tomasoni, D., Ubezio, P. Desynchronization rate in cell populations: mathematical modeling and experimental data. J. Theor. Biol., **208**, 185–199 (2001).
5. Clairambault, J., Laroche, B., Mischler, S., Perthame, B.: A mathematical model of the cell cycle and its control. INRIA research report # 4892 (2003).
6. Clairambault, J., Mischler, S., Perthame B.: Circadian rhythm and tumour growth. C. R. Acad. Sci. Paris, Ser. I, **342**, 17–22 (2006).
7. Cohen, J.E., Friedland, S., Kato, T., Kelly, F.P.: Eigenvalue inequalities for products of matrix exponentials. Lin. Alg. Appl., **45**, 55–95 (1982).
8. Davis, S., Mirick, D.K., Stevens, R.G.: Night shift work, light at night, and risk of breast cancer. J. Natl Cancer Inst., **93**, 1557–1562 (2001).
9. Filipski, E., King, V.M., Li, X.M., Granda, T. G., Mormont, M.C., Liu, X.H., Claustrat, B., Hastings, M.H., Lévi, F.: Host circadian clock as a control point in tumor progression. J. Natl. Cancer Inst., **94**, 690–697 (2002).
10. Filipski, E., Innominato, P.F., Wu, M.W., Li, X.M., Iacobelli, S., Xian, L.J., Lévi, F.: Effect of light and food schedules on liver and tumor molecular clocks in mice. J. Natl. Cancer Inst., **97**, 507–517 (2005).

11. Fu, L., Pelicano, H., Liu, J., Huang, P., Lee, C.C.: The circadian gene *Per2* plays an important role in tumor suppression and DNA damage response in vivo. Cell, **111**, 41–50 (2002).
12. Fu, L., Lee, C.C.: The circadian clock: pacemaker and tumor suppressor. Nature reviews/Cancer, 2003(3): 350–361.
13. Goldbeter, A.: Biochemical Oscillations and Cellular Rhythms. Cambridge University Press (1996).
14. Lebowitz, J.L., Rubinow, S.I.: A theory for the age and generation time distribution of a microbial population. J. Math. Biol., **1**, 17–36 (1977).
15. Lévi, F. (ed.): Cancer chronotherapeutics. Special issue of Chronobiology International Vol. 19 (1) (2002).
16. Matsuo, T., Yamaguchi, S., Mitsui, S., Emi, A., Shimoda, F., Okamura, H.: Control mechanism of the circadian clock for timing of cell division in vivo. Science, **302**, 255–259 (2003).
17. Metz, J.A.J., Diekmann, O.: The dynamics of physiologically structured populations. LN in Biomathematics 68, Springer-Verlag (1986).
18. Michel, P., Mischler, S., Perthame, B.: The entropy structure of models of structured population dynamics. General relative entropy inequality: an illustration on growth models. J. Math. Pures et Appl., **84**, 1235–1260 (2005).
19. Mischler, S., Perthame, B., Ryzhik L.: Stability in a nonlinear population maturation model. M3AS, **12**, 1751–1772 (2002).
20. Mormont, M.-C., Lévi, F.: Cancer chronotherapy: principles, applications and perspectives. Cancer, **97**, 155–169 (2003).
21. Nagoshi, E., Saini, C., Bauer, C., Laroche, T., Naef, F., Schibler, U. Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. Cell, **119**, 693–705 (2004).
22. Rotenberg, M.: Transport theory for growing cell populations. J. Theor. Biol., **103**, 181–199 (1983).
23. Schernhammer, E.S., Laden, F., Speizer, F.E., Willett, W.C., Hunter, D.J., Kawachi, I., Fuchs, C.S., Colditz, G.A.: Night-shift work and risk of colorectal cancer in the Nurses' Health Study. J. Natl. Cancer Inst., **95**, 825–828 (2003).
24. Schibler, U.: Liver regeneration clocks on. Science, **302**, 234–235 (2003).
25. Wu M.W., Li X.M., Xian L.J., Lévi F.: Effects of meal timing on tumor progression in mice. Life Sciences, **75**, 1181–1193 (2004).
26. You, S., Wood, P.A., Xiong, Y., Kobayashi, M., Du-Quiton, J., Hrushesky, W.J.M.: Daily coordination of cancer growth and circadian clock gene expression. Breast Canc. Res. Treat., **91**, 47–60 (2005).

Bone Turnover Cycle Model with a Torus-Like Steady State

Adam Moroz and David Ian Wimpenny

Computing Science and Engineering, De Montfort University, 49 Oxford Street, Leicester, LE1 5XY, UK; amoroz@dmu.ac.uk

Summary. A quantitative understanding of the bone remodeling process is of considerable biomedical and practical biotechnological interest to support the application of layer manufacturing techniques to produce scaffolds for surgical applications. Osteoclasts and osteoblasts play a principal role in different models of the bone multicellular unit operating in bone and display a rich spectrum of behaviour. The goal of this work is to show that it is possible to capture the cyclic dynamics of operating cells. The central idea of the mathematical model is that the regulatory nature of osteocytes is the basis of the cyclic-like behaviour associated with the system (remodeling process) as a whole. We developed this model taking due account of the apoptosis of osteocytes as a possible regulation loop in bone remodeling control. By applying the ordinary differential equations technique to the model, we show cyclic modes over a wide range of constants that have clear biological relevance. Simulations show that for a particular range of constants the model exhibits a torus-like quasi-steady state. Further investigation into these simulations indicates the existence of a surface in the osteoclasts-osteoblasts-osteocytes-bone space, that could be interpreted as a conservative value confirming the substrate-energy regenerative capability of the bone remodeling system. It is suggested that the nature of this recovering potential is directed against both mechanical and biochemical damage to the bone.

Key words: Mathematical model, bone, remodeling, torus.

22.1 Introduction

The potential to apply layer manufacturing methods to the production of complex scaffolds for tissue engineering has been recognised for several years [1–4]. However, despite some promising results [5,6], the full potential of this approach can only be fully realised when the bone remodeling process is well understood. Moreover, if this knowledge is incorporated into a mathematical model of the basic biological remodeling processes, this could then provide an important tool to enable the design of scaffolds to be optimised. In our study we investigate possible cyclic (periodic) steady-state dynamical options for known model approaches but with the addition of osteocyte activity based rather on the interaction of bone cells (osteoblasts (OBl), osteoclasts (OCl), osteocytes (OCt)) and their relationship with bone material density.
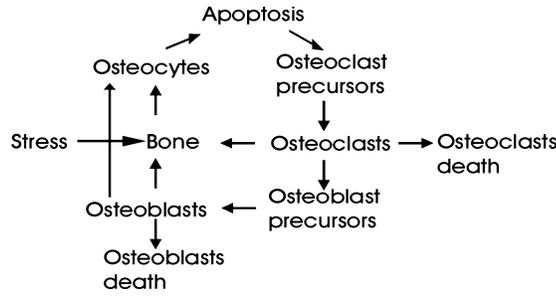
Fig. 22.1. Schematic representation of interaction between the cells in BMU. Arrows represent the control loops of regulation in BMU on the cellular level.

22.2 Model Development

Although the model developed by Komarova et al. [7] predicts different modes of dynamic behaviour of the basic multicellular unit (BMU) in bone remodeling control, a number of limitations to this model were identified by the authors of this paper, omission of autocrine and paracrine regulation loop parameters, for example. Furthermore, many publications indicate the importance of the level of osteocyte regulation, for example the role of the osteocyte apoptosis as a part of the mechanotransduction control mechanism [8,9] under the influence of mechanical stress [10]. These issues prompted us to take into account the osteocyte apoptosis in the bone remodeling regulation loop and the requirement to redevelop the model in such a way that the autocrine and paracrine control would be more biologically relevant from a generalised point of view. We refined the regulation loops, which control the activity of the BMU, and attempted to introduce a cybernetic approach in which the control would be minimised from both the energetic (catabolic) and metabolic points of view. Additionally, the authors believe that in order to produce a robust bone turnover process the regulation needs to be robust at all levels of control, and indeed at the cellular levels. Taking into account this point, we have attempted to develop and analyse the possible cellular model and robustness on the cellular level. The resulting cell-level control scheme based on the introduction of the osteocytes control loop could be presented as shown in Fig. 22.1, where osteocyte apoptosis initiates the osteoclast maturation from osteoclast precursors.

On the basis of this scheme we developed a set of various differential equation models using a particular range of functions in order to examine whether the model simulates the dynamic behaviour of the bone turnover cycle. Linear and quadratic dependence of functions, following [7] and modifying the model by introducing the equation for OCt dynamics (equation for x_3), gives the system

$$\frac{dx_1}{dt} = \alpha_{11}x_1 + \alpha_1 x_1(1 - x_3) - \beta_1 x_2 x_3 \tag{22.1}$$

$$\frac{dx_2}{dt} = \alpha_2(1 - x_4) - \beta_2 x_3 - \beta_{22}x_4 \tag{22.2}$$

$$\frac{dx_3}{dt} = \alpha_3 x_2 x_4 - s x_3 \tag{22.3}$$

$$\frac{dx_4}{dt} = -k_1(x_1 - 1) + k_2(x_2 - 1), \tag{22.4}$$

where x_1 is the relative concentration of osteoclasts; x_2 is the relative concentration of osteoblasts; x_3 is the relative concentration of osteocytes; x_4 is the relative bone mass, s is the level of mechanical stress and α_1, α_{11}, α_2, α_3, β_1, β_2, β_{22}, k_1 and k_2 are rate constants.

The first equation listed above describes autocrine and paracrine OCl regulation of OCt production and removal and follows the method of Komarova et al. The difference from the Komarova model [7] in terms of OCl dynamics is that we put in a paracrine term (which is the control feedback loop from the OCt) based on suggestions that OCt play a key role in this regulation [8–10]. The last negative term in this first equation describes OCl removal and reflects our proposition of the delayed paracrine control of OCl by OBl. In the second equation of the system (22.2), which concerns OBl, the first term describes the paracrine-like feedback control of bone material on OBl whilst the other terms describe OBl transformation into OCt and their apoptosis/death. The major difference from the Komarova model [7] is the introduction of the third equation (22.3) to the model, which describes OCt dynamics, where the first term is responsible for their transformation/differentiation from OBl and the second term describes the OCt apoptosis/death affected by the level of mechanical stress (s). The fourth equation of the system (22.4) follows the last equation from Komarova's model [7] but with the minor difference that the relative change was chosen as 1.0 instead of 100% as in the original model. The relative change was chosen as a variable to describe OCt population density, so all the variables describing OCl, OBl and OCt relative densities are dimensionless. This approach was adopted so that the cyclic modes of the model could be explored without requiring specific values for the cell variables. Indeed, the system (22.1–22.4) is partially scaled by using relative population densities. However, it is important to note that when the rate constants are specified in a particular unit, then the time scale of the system becomes this unit. Because we wished to explore the cyclic/periodic behaviour of the system in biologically relevant time scales (measured in hours, days and weeks) the smallest unit of time employed was 1 hour and this was used for our initial study. If Komarova's constants k_1, for example, are recalculated in cell$^{-1}*$ hour^{-1} this changes the constant from $k_1 = 0.24\% * $ cell$^{-1}*$ day^{-1} to $k_1 = 0.0001$ cell$^{-1}*$ hour^{-1}, which is actually the range of constants we have studied. The other rate constants α_1, α_2, β_1, β_2 have been considerably changed in sense and dimension compared to the values employed by Komarova et al. [7]. Rate constant α_1 is no longer the paracrine feedback parameter for OCl-OBl but is now OCl-OCt in our system. However, we retain its value in range of the original Komarova model. We have taken the value for parameter α_1 as $\alpha_1 = 0.002$ hour^{-1} or 3 cells $*$ day^{-1} (see Fig. 22.2 and Fig. 22.3), which is a biologically realistic rate for OCl generation at a single remodeling site. This is based on the value used by Komarova, which is independent of the type of paracrine feedback. A similar analysis and recalculation procedure was undertaken for all rate constants employed in the model.
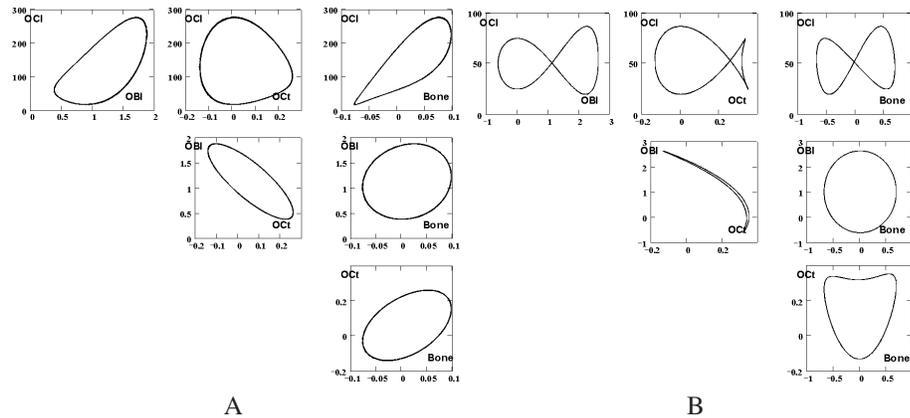
Fig. 22.2. Phase trajectories of 4D attractor of the system (22.1–22.4). The trajectories are shown for steady state. A, $\alpha_1 = 0.002$, $\alpha_{11} = 0.002$, $\alpha_2 = 0.01$, $\alpha_3 = 0.2$, $\beta_1 = 47.0$, $\beta_2 = 0.6$, $\beta_{22} = 0.045.0$, $s = 0.04$, $k_1 = 0.00001$, $k_2 = 0.01$. B, $\alpha_1 = 0.00002$, $\alpha_{11} = 0.000032$, $\alpha_2 = 0.000525$, $\alpha_3 = 0.007$, $\beta_1 = 5.0$, $\beta_2 = 0.05$, $\beta_{22} = 0.00001.0$, $s = 0.0004$, $k_1 = 0.000001$, $k_2 = 0.009$.

Thus, regarding the variables used, we believe that the relative population densities are dimensionless, but the time scale of our resulting data is in hours and all rate constants are in hour^{-1}. On the other hand, when the cycle period is scaled in certain time units, the cycles in all phase planes are actually in the same time units when the population densities are normalised to dimensionless units. Therefore, rate constants could change values that would render the model less biologically realistic. However, the underlying behaviour of the model (within the range of constants used) is unaffected. We also suggested that the new rate constant α_3 should be in the same range as the OBl first-order rate constant or even higher because the characteristic time of the bone regeneration could be shorter than the characteristic time of OBl production. In our study the α_3 rate constant varies between 0.0001 and 0.5 hour^{-1} (Fig. 22.3), which effectively covers the entire range of first-order kinetic processes between 1 hour and 1 second. This range of values is based on the osteocells production rate suggested by Komarova—effectively the OBl production constant in [7]. This value has been derived from the first-order kinetics data from experimental histomorphometric data [11,12]. However, this is a first-order rate constant when actually the OBl equation is second order and sometimes even fractal [7]. This illustrates that ultimately from first-order kinetic data it is possible only to conclude the range of rate constants. Moreover, the experiments to study kinetic parameters in bone are really semi-in vivo experiments because of the specific nature of hard tissue (bone) which prevents true in vivo results from being determined. Gathering experimental data on the kinetic parameters of bone enzymes, receptors and other complex molecular structures and then calculating and employing these rate constants in kinetic models is rather conjectural.
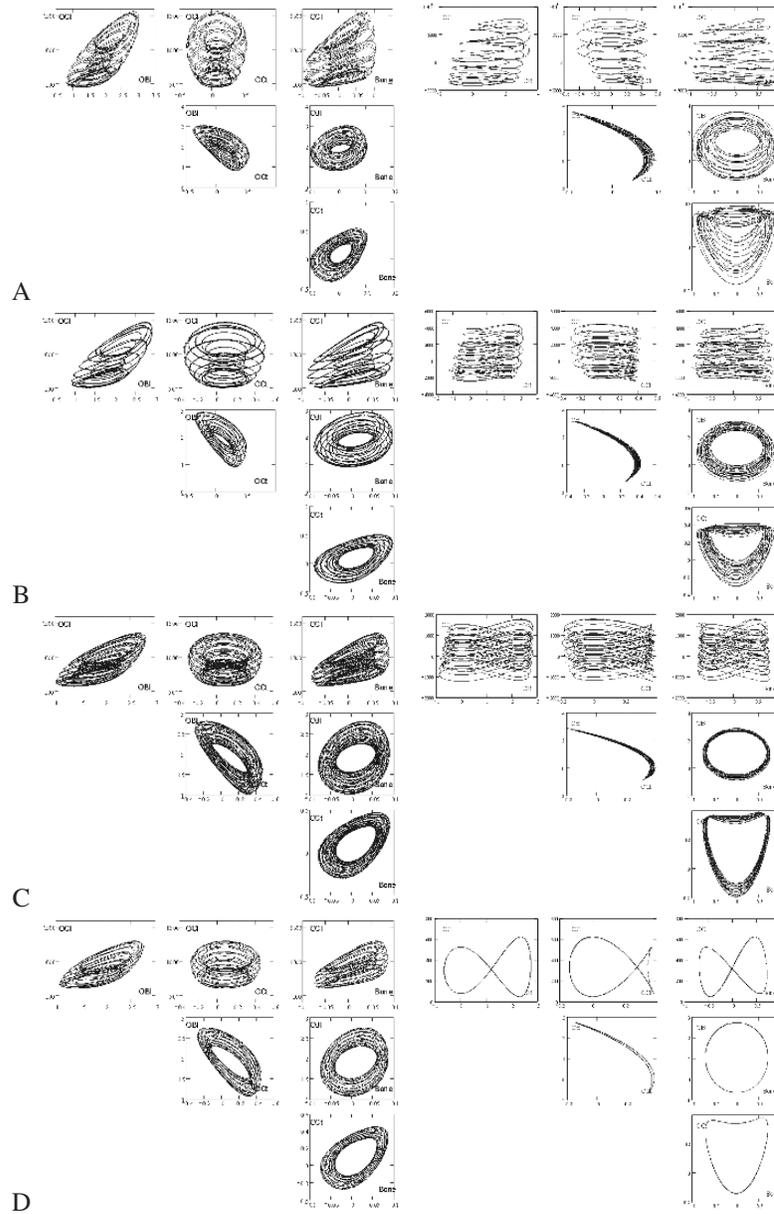
Fig. 22.3. Projections of the phase trajectories for the system (22.1–22.4). The graphs clearly show a collapse of 4D asymmetric torus to narrow figure 8-shaped cycle. The phase trajectories are shown in graphic matrix form at 4 intervals of time; start (row A), middle (rows B and C) and steady state (row D). Left column, $\alpha_1 = 0.002, \alpha_{11} = 0.002, \alpha_2 = 0.01, \alpha_3 = 0.2, \beta_1 = 25.0,$ $\beta_2 = 0.6, \beta_{22} = 0.045.0, s = 0.04, k_1 = 0.0001, k_2 = 0.01$. Right column, $\alpha_1 = 0.00002,$ $\alpha_{11} = 0.000032, \alpha_2 = 0.00055, \alpha_3 = 0.007, \beta_1 = 50.0, \beta_2 = 0.05, \beta_{22} = 0.00001.0,$ $s = 0.0004, k_1 = 0.000001, k_2 = 0.009$.

22.3 Results

For the wide range of constants discussed above we have numerically investigated the above mathematical model. We found that the system displays a characteristic cyclic behaviour with phase trajectories that show a steady-state asymmetric four-dimensional (4D) cycle (Fig. 22.2A), which is sometimes figure 8-shaped, see Fig. 22.2B.

We suggest that the cycle has a "toroidal" origin. For particular parameters, this torus is explicitly visible; see Fig. 22.3A. However, when other values are used, the torus clearly collapses with time into the cycle, when the curvilinear topological equivalent form is hidden; see Fig. 22.3, right column. Although for this particular case the trajectories of the system are difficult to interpret biologically, it is worth taking into account the existence of complex surfaces in 4D space.

The range of parameters employed in our study was selected to explore the complex nature of the trajectory and the relaxation of the system into the attracting cycle. For example, in Fig. 22.3, the left column shows that relaxation of the system occurs, commencing with a torus with a high amplitude oscillation, which partially collapses to form a distinctive attracting torus. However, in the right column of Fig. 22.3, the cycle commences from an indistinct torus with high amplitude that relatively quickly decays, which prevents a clear torus shape from being displayed (see Fig. 22.3 right column, second row). This collapses to form an 8-shaped attractor in the OCl-OBl phase plane and a "fish-shaped" attractor in the OCl-OCt phase plane (Fig. 22.3, right column, final row). Another important difference between the results shown in the left and right columns (Fig. 22.3) is that for the latter there is a certain area in the OBl-OCt phase plane when relations between OBl-OCt switch from a negative to a positive correlation (see the final three rows of the right-hand row). This result does not follow logically from the design of the system and raises important questions regarding the biological implications of this funding.

The steady cycle mode indicates periodic a type of behaviour with a time period of several days, taking into account the range of rate constants used (hour^{-1}). Over a certain range of parameters, the trajectories have a tendency to change shape, size and overlap, but their origin is invariably toroidal in nature; see, for example, Fig. 22.4A.

22.4 Discussion

In the introduction we stressed that the mathematical model developed in [7] predicts different modes of dynamic behaviour of the BMU while demonstrating the critical role for OCl autocrine regulation in the control of bone remodeling. In Komarova et al. [7] this role has been investigated in the framework of a dynamical system, where the OCt regulation at the cellular level was formally eliminated. In fact, these cells play a vital role in signaling mechanical damage. The introduction to the model of this regulation loop obviously increased the dimensions of the dynamical system. However, the system is still controlled with a minimal collection of cellular regulation loops. Over a certain range of constants that have realistic biological time scales for the bone

A   B
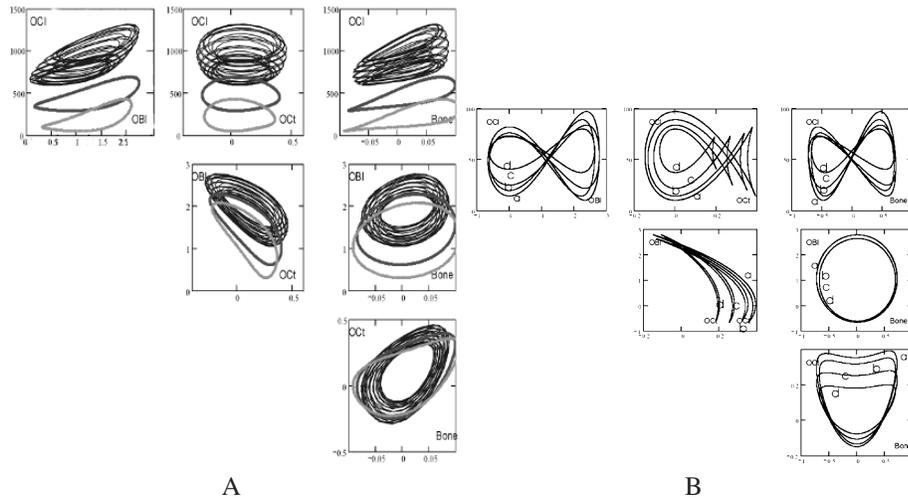
Fig. 22.4. Phase trajectories for the system (22.1–22.4), A, $\alpha_1 = 0.0002$, $\alpha_{11} = 0.002$, $\alpha_2 = 0.01$, $\alpha_3 = 0.02$, $\beta_2 = 0.6$, $\beta_{22} = 0.045$, $s = 0.04$, $k_1 = 0.0001$, $k_2 = 0.01$. spanning parameter a- $\beta_1 = 25.0$, b- $\beta_1 = 35.0$, c- $\beta_1 = 45.0$. B, $\alpha_1 = 0.00002$, $\alpha_2 = 0.000525$, $\alpha_{11} = 0.000035$, $\alpha_3 = 0.007$, $\beta_1 = 50.0$, $\beta_2 = 0.05$, $\beta_{22} = 0.00001$, $s = 0.0004$, $k_1 = 0.000001$, $k_2 = 0.009$. spanning parameters: a- $k_1 = 0.000001$, b- $k_1 = 0.00001$, c- $\alpha_3 = 0.005$, $k_1 = 0.000001$, d- $k_1 = 0.00001$, $\alpha_3 = 0.005$.

model, one can see the torus-like surface (Fig. 22.5) in the multidimensional phase space. In physics (mechanics), in the case of the classic 1D harmonic oscillator, such a surface is described by a circle in 2D phase space (coordinate, derivative). This circle represents a certain conservative value, i.e., the mechanical energy, which reversibly transforms from the potential form to kinetic form. In our multidimensional OCl-OBl-OCt-bone case, the shape of the toroidal surface reflects rather a certain interchange, i.e., the transformation from the "metabolic and kinetic" form of the BMU into the "bone material" form along the trajectory, which is supported in the steady state by the
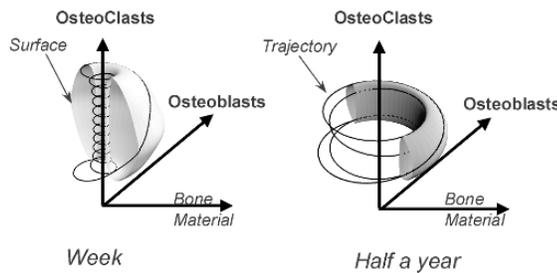
Fig. 22.5. 3D representation of the phase trajectories' evolution of the system (22.1–22.4).

blood substrate-energy resource. The asymmetry of this torus indicates the complexity of the phase relations in the BMU, OCt, bone mineral and organic components. In the final analysis a basic bone remodeling steady-state turnover could exist and all regulations in the direction of increasing or decreasing this level could be considered as the natural range of bone remodeling adaptability within the physiological activities of the body. This conclusion leads us to look for a cyclic process as the most optimised from the cellular regulation point of view. Another interesting result is that by adjusting the rate control of particular model parameters, for example β_1 (Fig. 22.4A), it is possible to control the shape and size of the attractor. This shows that the conservative quantity, along the steady-state trajectory, changes with the spanning of particular parameters and this in turn indicates that the substrate-energy metabolical cost of maintaining a steady state is also affected by changes in the parameters. If we consider how these findings can be mapped against realistic biological circumstances it can be seen that changes in amplitude and period could be relevant to some physiological situations. The increase in phase amplitude (Fig. 22.4) of oscillations and their frequency for oscillating systems in physics leads to an increase of the energy of the processes. Regarding the remodeling cycle, the biochemical nature of the existence of such a cyclic attractor indicates the need for biological resource (e.g., oxygen, ATP, other feeding substrates). Increasing the amplitude (Fig. 22.4B) of oscillations, and speeding or slowing down of the cycle are related to strengthening or weakening of the supply of this resource. When the speed of the cycle is too fast (it could be compared to Paget's disease, for example) this could result in "overfeeding" of the remodeling cycle. When the energy-like remodeling potential is undersupplied the risk of conditions such as osteoporosis may arise. To achieve optimal physiological control the metabolic resource required to support the cycle is balanced with the need to recover the mechanical function of the bone, in an appropriate physiological time period. In practice, the regulation of the cycle is probably maintained by a diverse range of molecular control loops and in a very precise way because the metabolic "cost" of correcting deviations from the ideal remodeling cycle places a significant burden on the system.

Additionally, the analysis of the phase shape of the cycle makes it possible to say that the relationships between OCl, OBl, OCt and bone mass are phase dependent. In practical terms this means that if the cycle period is longer than the sampling time, the resulting correlations established by these experiments could be different for different time periods of the cycle, e.g., phase dependent. For example, when the phase trajectory is figure 8-shaped, see Fig. 22.4B, OCl-OBl phase plane or OCl-bone phase plane, the correlation coefficients between OBl-OCl or OCl-bone could be positive after sampling on one time period and yet negative when sampled at another. Similar concerns also relate to the rate constants, mentioned previously (Section 22.2 Model Development). These findings reflect the complex nature of biochemical and cellular processes in the body and show the necessity of comparing the results of model-based and non-model-based approaches. Finally, within the framework of this rather phenomenological level of modeling, the role of the diverse molecular factors in bone regulation, such as receptors and mediators, the state of the membrane and the hormonal or genetic system, are difficult to discuss. The roles of these or any other molecular messenger

or substrate remain the subject of discussion in the biochemical literature, even for the generalised animal model. Thus, the development of a mathematical model, based on the molecular level of regulation in the bone, awaits more precise biochemical and biophysical data.

22.5 Conclusion

The developed model displays steady-quasi-cyclic behaviour when one employs a range of rate constants that is biologically realistic. We can conclude that the cyclic steady state of the system has a form which is topologically equivalent to a torus in multidimensional phase space. The existence of the surface in 4D osteoclast-osteoblast-osteocyte-bone space indicates that there could exist a conservative value for this dynamical system. In physics one usually interprets such a value as energy. In the case of the BMU it could be interpreted as substance (substrate) and energy recovery potential of the bone remodeling system. Metabolically and biochemically it could be associated with the continuously operating, genetically predetermined BMU. It provides a measure of the recovering ability of the BMU following mechanical and biochemical damage to bone.

References

1. Braddock, M., Houston, P., Campbell, C., Ashcroft, P.: Born again bone, tissue engineering for bone repair. News Physiol. Sci., 16, 208–213 (2001).
2. Xiong, Z., Yan, Y., Zhang, R., Sun, L.: Fabrication of porous poly(L-lactic acid) scaffolds for bone tissue engineering via precise extrusion. Scripta Materialia, **45**, 773–779 (2001).
3. Vozzi, G., Flaim C., Ahluwalia, A., Bhatia, S.: Fabrication of PLGA scaffolds using soft lithography and microsyringe deposition. Biomaterials, **24**, 2533–2540 (2003).
4. Vozzi, G., Flaim C.J, Bianchi, F., Ahluwalia, A., Bhatia, S.: Micro-fabricated PLGA scaffolds: a comparative study for application to tissue engineering, Mat. Sci. and Eng. C, **20**, 43–47 (2002).
5. Taboas, J.M., Maddox, R.D., Krebsbach, P.H., Hollister, S.J.: Indirect solid free form fabrication of local and global porous, biomimetric and composite 3D polymer-ceramic scaffolds. Biomaterials, **24**, 181–194 (2003).
6. Koegler, W.S., Griffith, L.G.: Osteoblast response to PLGA tissue engineering scaffolds with PEO modified surface chemistries and demonstration of patterned cell response. Biomaterials, **25**, 2819–2830 (2004).
7. Komarova, S.V., Smith, R.J., Dixon, S.J., Sims S.M., Wahl L.M.: Mathematical model predicts a critical role for osteoclast autocrine regulation in the control of bone modeling. Bone, **33**, 206–215 (2003).
8. Noble, B.: Bone microdamage and cell apoptosis, European Cells and Materials, **6**, 46–56 (2003).
9. Bonewald, L.F.: Osteocyte biology: Its implications for osteoporosis, J. Musculoskel. Neuron. Interact., **4**, 101–104 (2004).

10. Nakamura, H., Kumei, Ya., Morita, S., Shimokawa, H., Ohya, K., Shinomiya, K.: Suppression of osteoblastic phenotypes and modulation of pro- and anti-apoptic features in normal human osteoblastic cells under a vector-averaged gravity condition. J. Med. Dent. Sci., **50**, 167–176 (2003).
11. Parfitt, A.M.: Osteonal and hemi-osteonal remodelling: the spatial and temporal framework for signal traffic in adult human bone. J. Cell. Biochem., **55**, 273–286 (1994).
12. Vashishth, D., Gibson, G., Kimura, J., Schaffler, M.B., Fyhrie, D.P.: Determination of bone volume by osteocyte population. Anat. Rec., **276**, 292–295 (2002).

23

Modelling the Early Stages of Atherosclerosis

Michael J. Plank, Andrew Comerford, David J. N. Wall, and Tom David

University of Canterbury, Private Bag 4800, Christchurch, New Zealand;
m.plank@math.canterbury.ac.nz

Summary. Atherosclerotic lesions are predominantly localised to arterial bifurcations and bends, and are highly correlated with areas of low wall shear stress (WSS), but the underlying reason for this localisation is not fully understood. A key role is played by the endothelial cells, which regulate the transport of materials from the bloodstream to the artery wall and secrete vasoactive agents that modulate vascular tone. A mathematical model is presented, exploring the link between arterial geometry, WSS and factors related to atherogenesis. The model simulates the cellular response to the fluid shear stress on the cell membrane and the binding of ligands to cell surface receptors. This is used to calculate the rate of production of nitric oxide (NO), which is a potent vasodilator and anti-atherogenic factor. It is hypothesised that the section of endothelium adjacent to a region of recirculating flow is most at risk of developing atherosclerotic plaque, due to reduced bioavailability of NO.

Key words: Artery bend, calcium, endothelial cells, fluid shear stress, nitric oxide, recirculation zone.

23.1 Introduction

It is now well established that the formation of atherosclerotic lesions is not evenly distributed throughout the circulatory system, but occurs much more frequently near arterial bifurcations and in areas of high vessel curvature [5]. Atherosclerosis is strongly associated with disturbed haemodynamic characteristics, such as flow detachment from the artery wall, flow recirculation and oscillatory flow. In particular, the occurrence of atherosclerotic plaque is highly correlated with areas of the vasculature that experience low time-averaged wall shear stress (WSS) [8]. The causal link between low WSS and atherogenesis has long been the subject of investigation, but has yet to be fully elucidated; for a review of the role of fluid shear stress in endothelial physiology and pathology, see [11].

The intima (the portion of the artery wall most frequently affected by atherosclerosis) is separated from the bloodstream by a monolayer of endothelial cells (ECs). It has become increasingly recognised that this layer is not simply a passive barrier, but plays a crucial role in maintaining vascular homeostasis and regulating the passage of

materials between the blood and the vessel wall [16]. ECs are sensitive to mechanical stimuli such as fluid shear stress, as well as biochemical stimuli such as blood-borne agonists. ECs are also a source of numerous vasoactive factors, such as nitric oxide (NO), prostaglandin, endothelin and angiotensin [4].

One of the earliest clinical markers of atherosclerotic plaque formation is impaired EC-dependent vasodilation by NO. This state of endothelial dysfunction is frequently detectable before any structural changes in the arterial wall occur [3]. As well as being a potent vasodilator, NO inhibits proliferation of smooth muscle cells (SMCs) and aggregation and adhesion of platelets, and is thus a key atheroprotective factor [3,12]. The main enzyme governing synthesis and release of NO in ECs is endothelial NO synthase (eNOS). Endothelial production of NO is at least partially regulated by the inositol trisphosphate (IP_3)-dependent calcium (Ca^{2+}) signalling pathway, which may be activated by ligand-receptor binding and/or WSS. The elevated levels of free Ca^{2+} in the cytosol reversibly bind with calmodulin; the resulting Ca^{2+}-calmodulin complexes activate eNOS, thus causing an increase in the synthesis and release of NO. In addition, NO leads to increased levels of cyclic guanosine monophosphate (cGMP), which activates protein kinase G (PKG). This is thought to inhibit Ca^{2+} influx, thus providing a negative feedback mechanism and limiting the concentration of Ca^{2+} in the cytosol. For details of the IP_3-Ca^{2+} signalling pathway in ECs and associated NO production, see [7,13,15]. Reduced bioavailability of NO in atherosclerosis-prone regions is likely a consequence of disruption of this Ca^{2+}-dependent pathway.

In a previous model [10], the link between spatial haemodynamics and occurrence of atherosclerotic lesions was investigated. This model considered the effects of two types of stimuli: one biochemical (the binding of EC surface receptors by blood-borne ligand) and the other mechanical (the physical force exerted on the cell membrane by fluid shear stress). ECs are exposed to these stimuli in a spatially varying manner due to non-uniformities in the fluid flow field and consequent mass transport pattern. This results in differences in the behaviour of the intracellular Ca^{2+} pathway in different regions of the endothelium.

In this study, the effects of these differences on endothelial production of NO, which plays a crucial role in protecting the vasculature against atherogenesis, are examined. Reduced bioavailability of NO is a key factor in the onset of atherosclerotic plaque formation [4,12] and our model may thus be used to predict which areas of the artery wall, in a specific vascular geometry, are most susceptible to cardiovascular disease. The model thus directly addresses the key issue of the causative link, and its underlying biochemical mechanism, between the incidence of atherosclerosis and regions of low WSS.

A number of computational fluid dynamical (CFD) studies that model the concentrations of particular chemical species in specific two- or three-dimensional arterial structures may be found in the existing literature [6,14,17]. The model used in this study focuses on a simple two-dimensional flow geometry with spatially varying WSS. This is not intended to accurately replicate the complex situation *in vivo*, which is possible only in numerically intensive CFD studies, but rather to identify the most important features of the generic behaviour. The advantage of this approach is that it captures the important flow characteristics present in realistic arterial geometries (flow

separation, recirculation and subsequent reattachment), yet is sufficiently simple that the model equations are mathematically tractable and the cause-and-effect relationship between shear stress, advection, diffusion and reaction is not masked by complex three-dimensional flow. This enables the model analysis to identify the most important flow characteristics affecting the delivery of ligands to the endothelium and to elucidate the underlying relationship between mass transport in the bloodstream and chemical concentrations and reaction rates at the endothelial wall.

23.2 The Mathematical Model

23.2.1 Model for Wall Shear Stress and ATP Concentration

The mathematical model consists of two coupled components. The first concerns the transport of chemical species in the bloodstream in a region of spatially varying WSS. This model was developed in [1, 2, 9] and is used here to predict the concentration of ATP at the endothelial surface. This is an important factor in initiating the Ca^{2+} signalling pathway via activation of G-protein-coupled receptors; spatial variation in both ATP concentration and WSS will lead to variations in Ca^{2+} signalling and hence in NO production.

Under the assumption that ECs hydrolyse ATP via receptor binding to form ADP, a thin ATP concentration boundary layer will form in the bloodstream close to the endothelial surface. Denoting the WSS by $\tau_w(x)$, it has been shown [9] that the concentration ϕ of ATP at the endothelium is closely approximated by

$$\phi(x) = \beta(x)/(\beta(x) + IK),\tag{23.1}$$

where

$$\beta(x) = \tau_w(x)^{\frac{1}{2}}\left(9\int_0^x \tau_w(s)^{\frac{1}{2}}\,ds\right)^{-\frac{1}{3}},\tag{23.2}$$

$K = 1.7 \times 10^{-6}$ m/s is the rate of ATP hydrolysis by ECs and $I \approx 0.89$ is a constant resulting from the transformation to the similarity variable β (for details see [9]).

This gives a relationship between the WSS and the ATP concentration for any arterial geometry, provided that the Peclet number (i.e., the ratio of the rate of convective transport to diffusive transport) is high. This is valid for a wide range of blood-soluble factors, including adenosine nucleotides and lipoproteins, which have extremely low diffusion rates (of the order 10^{-10} m^2/s or less).

Attention is focused on the test geometry of a two-dimensional backward-facing step (see Fig. 23.1). In this geometry, the flow exhibits separation, recirculation and reattachment, thus comprising the fundamental haemodynamic characteristics of *in vivo* blood flow. Fig. 23.2 shows the WSS and ATP concentration profiles on the bottom surface of the backward-facing step ($y = 0$), under conditions of steady flow with Reynolds number Re $= 10$. The model predicts that the ATP concentration is maximal at or very near to the stagnation point. The concentration decreases gradually as
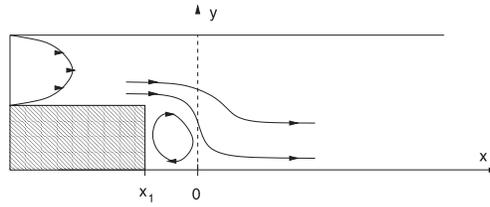
Fig. 23.1. Diagram of the model geometry of a backward-facing step. The flow exhibits detachment at $x = x_1$, with reattachment at the stagnation point, $x = 0$. There is a region of recirculating flow for $x_1 < x < 0$, whilst the flow becomes fully developed for $x > 0$.

one moves downstream and the flow becomes fully developed, but decreases rapidly in the recirculation zone between the points of flow separation and reattachment [9]. The profiles shown in Fig. 23.2 are considered to be the environmental stimuli to which the ECs on the bottom surface of the vessel are exposed and hence form the "inputs" to the model for intracellular Ca^{2+} signalling.

23.2.2 Model for Intracellular Signalling

The second component is the model for intracellular signalling. This consists of four coupled differential equations describing the rate of change of concentrations of IP_3 (i), cytosolic Ca^{2+} (C_c), Ca^{2+} in internal cellular stores (C_s) and activated eNOS (n):

$$\frac{di}{dt} = k_i \frac{\phi}{K_c + \phi} \frac{C_c}{K_1 + C_c} - \mu_1 i, \tag{23.3}$$

$$\frac{dC_c}{dt} = q_{rel} - q_{res} + q_{in} - q_{out}, \tag{23.4}$$

$$\frac{dC_s}{dt} = -V_r \left(q_{rel} - q_{res} \right), \tag{23.5}$$

$$\frac{dn}{dt} = \frac{k_{dis} C_c}{K_6 + C_c} - \mu_2 n. \tag{23.6}$$

(a)

(b)
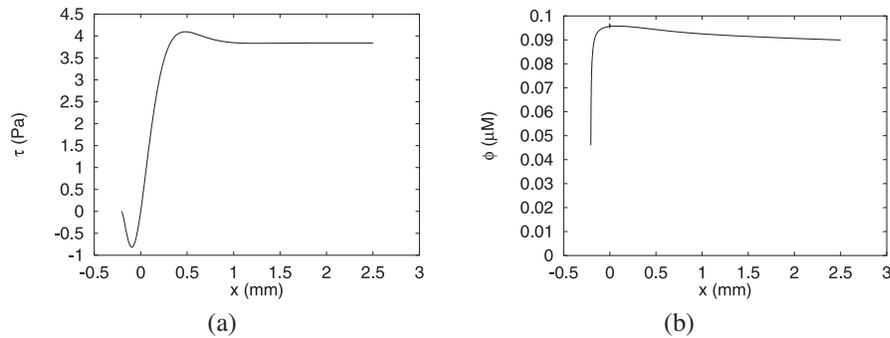
Fig. 23.2. Graphs of: **(a)** WSS τ_w; **(b)** surface ATP concentration ϕ against x under flow conditions with $Re = 10$.
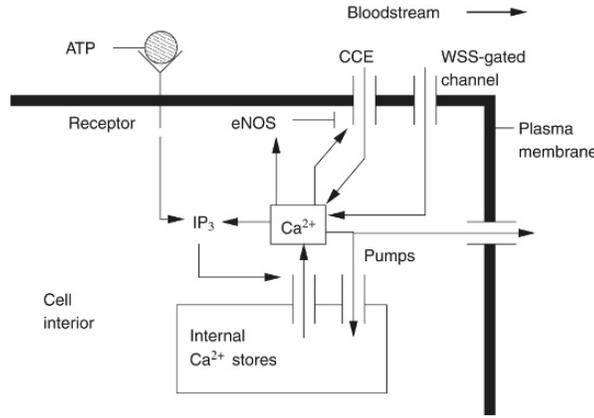
Fig. 23.3. Schematic illustration of the model of intracellular signalling. Receptor activation by external ATP stimulates IP$_3$ generation, which releases Ca^{2+} from internal stores into the cytosol. This can have a positive feedback effect by increasing the rate of IP$_3$ production. Depletion of internal stores stimulates capacitative Ca^{2+} entry (CCE). WSS can also increase Ca^{2+} influx via shear-gated plasma membrane Ca^{2+} channels. Free cytosolic Ca^{2+} activates eNOS, which has an inhibitory effect on the CCE channel. Cytosolic Ca^{2+} is resequestered back into the internal stores and is pumped across the plasma membrane out of the cell by Ca^{2+}-ATPases.

This is based on the models of [10, 18, 19]. The model is illustrated in Fig. 23.3 and a brief description of equations (23.3)–(23.6) is given here; for full details, see [10]. The meanings of the mathematical symbols and the values of the parameters are given in Table 23.1. The two terms in (23.3) represent IP$_3$ production (due to G-protein receptor binding by external ATP) and decay: k_i is the maximum rate of production, which is attained for large external ATP concentration ϕ; IP$_3$ production is accelerated by the presence of cytosolic Ca^{2+} as part of a positive feedback mechanism. The q_i in (23.4) and (23.5) are the rates of: Ca^{2+} release from the internal stores into the cytosol (q_{rel}); Ca^{2+} resequestration by the internal stores (q_{res}); Ca^{2+} influx (q_{in}); Ca^{2+} efflux (q_{out}). These are defined as follows:

$$q_{\text{rel}} = k_{\text{rel}} \left(\frac{i}{K_2 + i} \right)^3 C_s, \tag{23.7}$$

$$q_{\text{res}} = k_{\text{res}} \left(\frac{C_c}{K_3 + C_c} \right)^2, \tag{23.8}$$

$$q_{\text{in}} = \frac{k_{\text{CCE}}}{K_4 + n} \left(C_{s,0} - C_s \right) \left(C_{\text{ex}} - C_c \right) + F \left(\tau_w \right), \tag{23.9}$$

$$q_{\text{out}} = \frac{k_{\text{out}} C_c}{K_5 + C_c}. \tag{23.10}$$

q_{rel} is an increasing function of IP$_3$ concentration, with maximum rate k_{rel}. q_{in} consists of Ca^{2+} influx stimulated by depletion of the internal stores, termed capacitative Ca^{2+} entry (CCE), and WSS-induced influx, $F(\tau_w)$, where F is an increasing function of

Table 23.1. Table of parameter values.

Quantity	Symbol	Value and units
IP$_3$ production rate	k_i	5.46×10^{-3} μM s^{-1}
IP$_3$ decay rate	μ_1	0.2 s^{-1}
Ca^{2+} release rate	k_{rel}	6.64 s^{-1}
Ca^{2+} resequestration rate	k_{res}	5 μM s^{-1}
Ca^{2+} efflux rate	k_{out}	24.7 μM s^{-1}
eNOS-caveolin dissociation rate	k_{dis}	0.09 μM s^{-1}
eNOS-caveolin association rate	μ_2	0.0167 s^{-1}
Max. WSS-induced Ca^{2+} influx rate	q_m	17.6 μM s^{-1}
CCE rate	k_{CCE}	5.7×10^{-6} s^{-1}
Resting stored Ca^{2+} concentration	$C_{s,0}$	2830 μM
External Ca^{2+} concentration	C_{ex}	1500 μM
Michaelis–Menten constants	K_c	0.026 μM
	K_1	0 μM
	K_2	0.2 μM
	K_3	0.15 μM
	K_4	5 μM
	K_5	0.32 μM
	K_6	0.45 μM
Ratio of cytosol to ER volume	V_r	3.5

WSS (with maximum rate q_m). Note that the rate of CCE is attenuated in the presence of eNOS, reflecting the negative feedback of high eNOS levels on Ca^{2+} influx. q_{res} and q_{out} represent the actions of Ca^{2+}-ATPases on the cytosolic Ca^{2+}. Finally, the two terms in (23.6) represent activation and deactivation of eNOS: eNOS is activated by cytosolic Ca^{2+} at maximum rate k_{dis} and deactivated at constant rate μ_2.

23.2.3 Coupling the Model Components

A section of the endothelium on the bottom surface of a backward-facing step is now examined. The WSS and ATP concentration profiles along this surface (shown in Fig. 23.2) constitute the "inputs" to the intracellular signalling model. Specifically, the ATP concentration, at a given position on the endothelium, determines the rate of intracellular IP$_3$ production due to receptor binding by modulating the first term in (23.3). The WSS at a given position modulates the rate of Ca^{2+} influx into the cell via

the shear-stimulated influx term $F(\tau_w)$ in (23.9). These combined stimuli thus determine the behaviour of the IP_3-Ca^{2+} signalling pathway and its effect on eNOS activity, via the system of differential equations (23.3)–(23.6).

Hence at any given position on the endothelial surface, the system (23.3)–(23.6) can be solved numerically to give the local intracellular IP_3, Ca^{2+} and eNOS dynamics. By comparing the results in different locations along the endothelial surface, predictions can be made regarding the consequences for potential endothelial dysfunction and the onset of early atherosclerosis.

The intracellular dynamics do not impact the model for mass transport in the bloodstream, which is essentially decoupled. A more general model would allow for the blood transport of factors, such as NO, secreted by the ECs in a spatially varying manner. However, the extracellular concentration of Ca^{2+} is very high compared to intracellular levels [15], and intracellular Ca^{2+} will not significantly affect extracellular or fluid concentrations. Secretion of NO into the bloodstream may be significant, but the primary action of NO considered here is its vasodilatory effect on sub-endothelial SMCs. This action is paracrine in nature, so the local NO production rate, as indicated by the intracellular activated eNOS concentraion, is used as a predictor of arterial wall health.

23.3 Results

Fig. 23.4 shows the intracellular IP_3, Ca^{2+} and eNOS dynamics in response to a step increase in the extracellular ATP concentration from 0 μM to 0.06 μM (which is in the physiological range of ATP concentrations found in the bloodstream [13]). This stimulus initiates the Ca^{2+} mobilisation and influx pathways and the model produces a biphasic response. The initial phase consists of a transient increase in cytosolic Ca^{2+}, and a corresponding decrease in stored Ca^{2+}, representing a rapid release of Ca^{2+} from the internal stores. In the second phase, there is a gradual decay of cytosolic Ca^{2+} to a plateau level as the cell's Ca^{2+} extrusion mechanism takes effect. The activity of eNOS, which is stimulated by cytosolic Ca^{2+}, also exhibits a transient increase, followed by a gradual decay. The response to a step increase in WSS (i.e., sudden onset of flow) is very similar (data not shown). These results are in qualitative agreement with experimental evidence on the response of ECs to a physiological stimulus of this nature [20].

Notice that, after the initial transient, the signalling process settles down to a steady state. In particular, the cytosolic Ca^{2+} and eNOS concentrations tend towards plateau values, which are elevated above the levels in an unstimulated cell. The degree of elevation above resting conditions varies with the level of stimulation (by both WSS and ATP) and hence with position along the vessel. We therefore examine the relationship between spatial location and the steady-state concentrations of cytosolic Ca^{2+} and eNOS, disregarding the initial transient behaviour. Fig. 23.5 shows the steady-state concentrations of cytosolic Ca^{2+} and eNOS against distance downstream in the backward-facing step for $Re = 10$. Notice that eNOS is lower in the recirculation zone
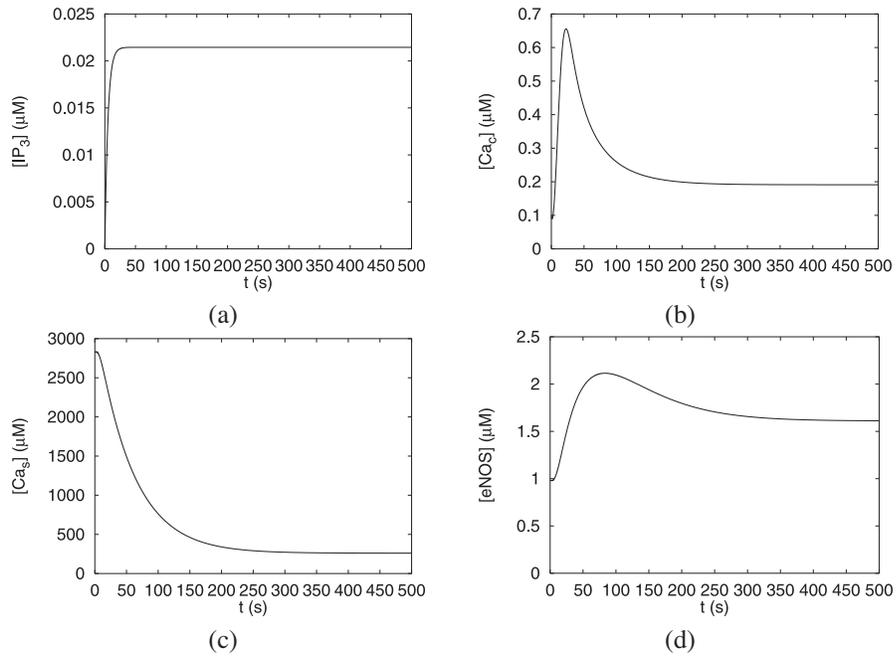
Fig. 23.4. Graphs of intracellular concentrations against time in response to a step increase in the extracellular ATP concentration from 0 μM to 0.06 μM: **(a)** IP$_3$; **(b)** cytosolic Ca^{2+}; **(c)** stored Ca^{2+}; **(d)** eNOS.

($x < 0$) than in the region of fully developed flow ($x > 1$ mm), suggesting that this area is more likely to develop atherosclerotic plaque.

To enable a comparison between model predictions for the backward-facing step and more realistic geometries, the model was extended to a three-dimensional artery bend. Simulations were run using the CFD package FLUENT to determine the WSS
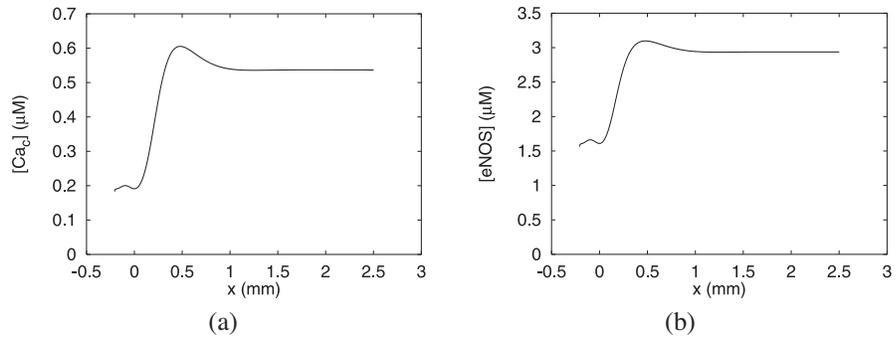
Fig. 23.5. Graphs of the steady-state intracellular concentration against x for Re $= 10$: **(a)** cytosolic Ca^{2+}; **(b)** eNOS.
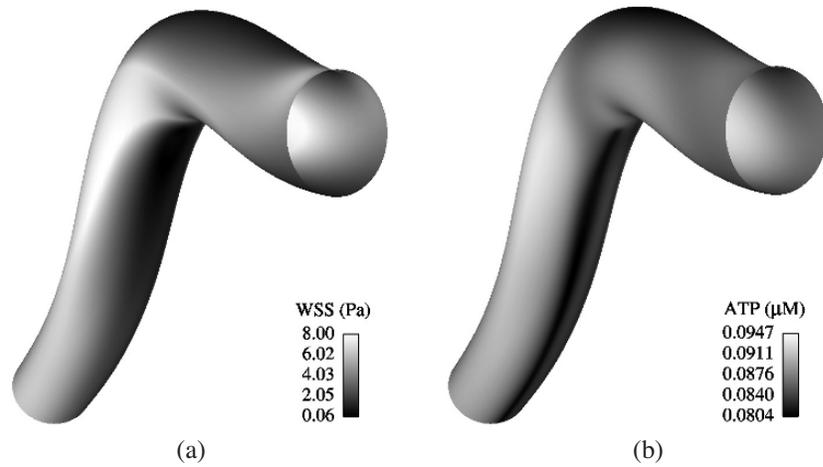
Fig. 23.6. Contour plots of: **(a)** WSS; **(b)** surface ATP concentration in an artery bend at Re = 300. Flow is from top-right to bottom-left, and there is a region of low WSS and low ATP on the inner wall downstream of the bend.

and ATP concentrations on the endothelium (see Fig. 23.6). A Reynolds number of 300 was used in these simulations: a higher value than that in the backward-facing step simulations was used in order to replicate flow in a larger artery with flow detachment, which occurs on the inner wall of the bend, resulting in a region of relatively low WSS on the inner wall (darker region in Fig. 23.6(a)). These data were again used as the inputs to the intracellular model, and the steady-state cytosolic concentrations of Ca^{2+} and activated eNOS were calculated on the artery wall (see Fig. 23.7). These are lower
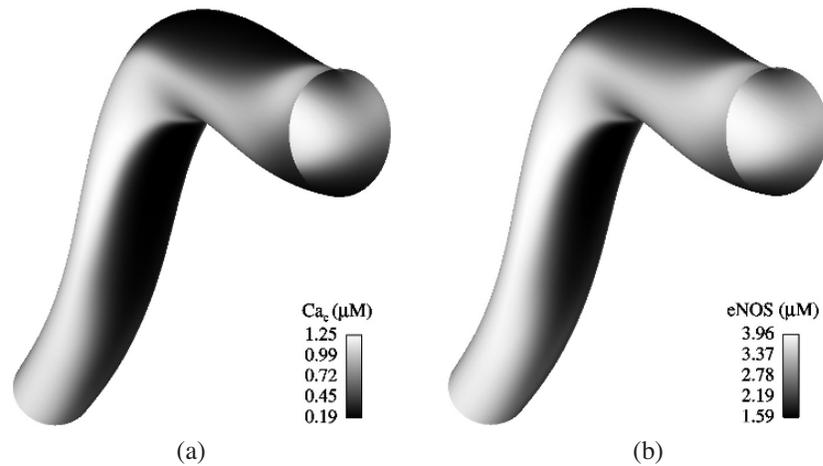
Fig. 23.7. Contour plots of steady-state intracellular concentrations in an artery bend for Re = 300: **(a)** cytosolic Ca^{2+}; **(b)** eNOS.

in the low WSS region on the inner wall than in regions of fully developed flow and high WSS.

23.4 Discussion

The behaviour of a section of vascular endothelium near a flow disturbance has been studied by coupling a model of fluid shear stress and ATP mass transport in the blood vessel (inputs) to a model of intracellular signalling and consequent eNOS activity (output). The results indicate that conditions favour atherogenesis (due to reduced eNOS activity) in the recirculation zone that forms immediately downstream of a backward-facing step. Clearly, the two-dimensional backward-facing step geometry used in this model is not an accurate representation of *in vivo* arteries. Nevertheless, the key fluid dynamical characteristics of flow separation, recirculation and reattachment are present in the simplified model geometry.

To verify the model predictions in a more realistic geometry, the model was applied to a three-dimensional artery bend. The results confirm the spatial relationship between disturbed flow (in particular low WSS) and reduced bioavailability of NO, with consequent adverse implications for atherogenesis. We therefore hypothesise that reduced NO production, due to impaired endothelial Ca^{2+} and eNOS signalling, is a key factor in the localisation of atherogenesis to regions of low WSS and recirculating flow. This hypothesis is consistent with empirical observations that atherosclerosis is strongly localised to arterial bifurcations and bends [5], regions which exhibit disturbed flow characteristics and low WSS [11].

Although atherogenesis occurs as a result of a complex interplay between a wide range of factors, the mathematical model provides insight into some of the mechanisms responsible. In regions of fully developed flow (i.e., unidirectional flow with a parabolic velocity profile), as occurs in relatively linear sections of the vasculature, the mechanical WSS on the EC is moderate or high (in the range 1–5 Pa), as is the convection of ATP into the thin mass transfer boundary layer near the endothelium. Thus the combined environmental stimulus for EC in such a region is in the physiological range, leading to activation of the IP_3-Ca^{2+} signalling pathway and consequent activation of eNOS and synthesis of NO. In contrast, the ECs in areas near a flow disturbance or recirculation zone, as occurs near arterial bifurcations and bends, usually experience lower time-averaged WSS (< 1 Pa). In addition, lower rates of convection and longer residence times in such regions mean that the ATP concentration in the mass transfer boundary layer can become substantially depleted. Hence the combined environmental stimulus is low, leading to reduced Ca^{2+} signalling and eNOS activity. As a consequence, this section of the endothelium is more prone to dysfunction and associated onset of atherosclerosis.

The hypothesis that the artery wall in a region of recirculating flow experiences impaired intracellular Ca^{2+} and eNOS signalling, and thus represents a "hot spot" for the initiation of atherosclerotic lesions, needs to be tested experimentally. This could be done by culturing vascular ECs on the bottom surface of a backward-facing step flow chamber (as shown in Fig. 23.1) and perfusing the chamber with medium

containing ATP. The cells will thus be exposed to spatially varying shear stress and ATP concentration and measuring these quantities will enable the theoretical results (e.g., those shown in Fig. 23.2) to be validated. The cytosolic free Ca^{2+} concentration at different points in the flow chamber can be measured by fluorescence and the results compared to the theoretical predictions of our model.

References

1. David, T.: Wall shear stress modulation of ATP/ADP concentration at the endothelium. Ann. Biomed. Eng., **31**, 1231–1237 (2003).
2. David, T., Thomas, S., Walker, P.G.: Platelet deposition in stagnation point flow: an analytical and computational solution. Med. Eng. and Phys., **23**, 299–312 (2001).
3. Davignon, J., Ganz, P.: Role of endothelial dysfunction in atherosclerosis. Circ., **109**, III23–32 (2004).
4. Galley, H.F., Webster, N.R.: Physiology of the endothelium. Brit. J. Anaesth., **93**, 105–113 (2004).
5. Gimbrone, M.A.: Vascular endothelium, haemodynamic forces and atherogenesis. Am. J. Path., **155**, 1–5 (1999).
6. John, K., Barakat, A.I.: Modulation of ATP/ADP concentration at the endothelial surface by shear stress: effect of flow-induced ATP release. Ann. Biomed. Eng., **29**, 740–751 (2001).
7. Lin, S., Fagan, K.A., Li, K.X., Shaul, P.W., Cooper, D. M.F., Rodman, D.M.: Sustained endothelial nitric-oxide synthase activation requires capacitative Ca^{2+} entry. J. Biol. Chem., **275**, 17979–17985 (2000).
8. Malek, A.M., Alper, S.L., Izumo, S.: Haemodynamic shear stress and its role in atherosclerosis. J. Am. Med. Assoc., **282**, 2035–2042 (1999).
9. Plank, M.J., Comerford, A., David, T., Wall, D. J.N.: Concentration of blood-borne agonists at the endothelium. Proc. Roy. Soc. A, **462**, 671–688 (2006a).
10. Plank, M.J., Wall, D. J. N., David, T.: Atherosclerosis and calcium signalling in endothelial cells. Prog. Biophys. Mol. Biol., **91**, 287–313 (2006b).
11. Resnick, N., Yahav, H., Shay-Salit, A., Shushy, M., Schubert, S., Zilberman, L. C.M., Wofovitz, E.: Fluid shear stress and the vascular endothelium: for better and for worse. Prog. Biophys. Mol. Biol., **81**, 177–199 (2003).
12. Shaul, P.W.: Endothelial nitric oxide synthase, caveolae and the development of atherosclerosis. J. Physiol., **547**, 21–33 (2003).
13. Shen, J., Luscinskas, F.W., Connolly, A., Dewey, C.F., Gimbrone, M.A.: Fluid shear stress modulates cytosolic free calcium in vascular endothelial cells. Am. J. Physiol., **262**, C384–390 (1992).
14. Stangeby, D.K., Ethier, C.R.: Computational analysis of coupled blood-wall arterial LDL transport. J. Biochem. Eng., **124**, 1–8 (2002).
15. Tran, Q.K., Ohashi, K., Watanabe, H.: Calcium signalling in endothelial cells. Cardiovasc. Res., **48**, 13–22 (2000).
16. Traub, O., Berk, B.C.: Laminar shear stress: mechanisms by which endothelial cells transduce an atheroprotective force. Arterioscler. Thromb. Vasc. Biol., **18**, 677–685 (1998).
17. Wada, S., Karino, T.: Theoretical prediction of low-density lipoproteins concentration at the luminal surface of an artery with a multiple bend. Ann. Biomed. Eng., **30**, 778–791 (2002).
18. Wiesner, T.F., Berk, B.C., Nerem, R.N.: A mathematical model of cytosolic calcium dynamics in human umbilical vein endothelial cells. Am. J. Physiol., **270**, C1556–1569 (1996).

19. Wiesner, T.F., Berk, B.C., Nerem, R.M.: A mathematical model of the cytosolic-free calcium response in endothelial cells to fluid shear stress. Proc. Natl. Acad. Sci. USA, **94**, 3726–3731 (1997).
20. Yamamoto, K., Korenaga, R., Kamiya, A., Ando, J.: Fluid shear stress activates Ca^{2+} influx into human endothelial cells via P2X4 purinoceptors. Circ. Res., **87**, 385–391 (2000).

Magnetic Nanoparticles for In Vivo Applications: A Numerical Modeling Study

Carlos Trenado[1] and Daniel J. Strauss[2]

[1] Leibniz Institute for New Materials, Saarbrücken, Germany; trenado@inm-gmbh.de
[2] Leibniz Institute for New Materials and Computational Diagnostics and Biocybernetics Unit, Saarbrücken/Homburg, Germany; strauss@inm-gmbh.de

Summary. In vivo applications of biocompatible magnetic nanoparticles in a carrier liquid controlled by an external magnetic field from outside the body have recently been proposed for specific drug delivery such as in locoregional cancer therapies or the occlusion aneurysms. They can also be used as guided contrast agents in myocardial imaging after myocardial infarction. However, the choice of the optimal clinical setting still remains a challenge for all of the mentioned applications. A numerical heterogeneous multiscale model can be used for an optimal a priori determination of the free parameters and might help to overcome this problem.

Key words: Navier–Stokes equations, Maxwell equations, multiscale, ferrofluids, biocompatible, nanoparticles.

24.1 Introduction

Current research on methods to target chemotherapy drugs in the human body includes the investigation of what are called biocompatible magnetic nanocarrier systems. For example, magnetic liquids such as ferrofluids can play an important role as drug carriers in the human body [1], while preventing damage to healthy human cells by allowing more strict control in the treatment dose and by weakening global exposure to the organism as in the case of ablation techniques.

Such is the case of modern locoregional cancer treatments using drug targeting techniques. However, a remaining challenge for this medical application is the choice of a safe and effective clinical setting, which must involve as a prerequisite the optimal adjustment of the external magnetic field as well as the choice of ferrofluid properties. In that direction, different theoretical mathematical models have already been proposed in order to address several aspects related to the validity and safety of such medical treatments [2].

In this chapter, we focus on the idea of a numerical heterogeneous multiscale model which can be used for an optimal a priori determination of the free parameters. In this study, we are going to present an approach to the implementation of such a multiscale model. We applied a hybrid scheme which is based on Maxwell and Navier–Stokes

equations to paramagnetic liquids. For a considered area of the body, the Maxwell equations for the static magnetic case were solved. The vector potential was then coupled to a liquid flow problem described by the Navier–Stokes equations by a volume force acting on the magnetic liquid. Time-dependent boundary conditions were used to describe the systolic blood flow regime. The system of coupled partial differential equations was solved by the finite element method on adaptive meshes. For a representative geometry, our hybrid model allowed the study of the hydrodynamics of the magnetic liquid. Moreover, our model is open for an embedding of a nanoscale model which represents the particle dynamics. It is concluded that the proposed model is a prerequisite for the optimal computational choice of the free parameters of magnetic liquids and the external magnetic field for in vivo applications.

24.2 Model Description

A simple and numerically effective geometric representation of a global vessel and an external magnetic field is achieved with a two-dimensional (2D) model. A blood vessel is fed with a liquid (blood) from the left in Fig. 24.1. The velocity and pressure field is calculated in the bloodstream. The magnetic field (magnetic vector potential) that is generated by the magnet is also calculated. This magnetic field generates a magnetic volume force that affects the flow field in the blood vessel.

24.2.1 Model Equations

We consider Maxwell's equations for the static case

$$\nabla \times \mathbf{H} = \mathbf{J}$$
$$\nabla \cdot \mathbf{B} = \mathbf{0}$$

and the constitutive relationship
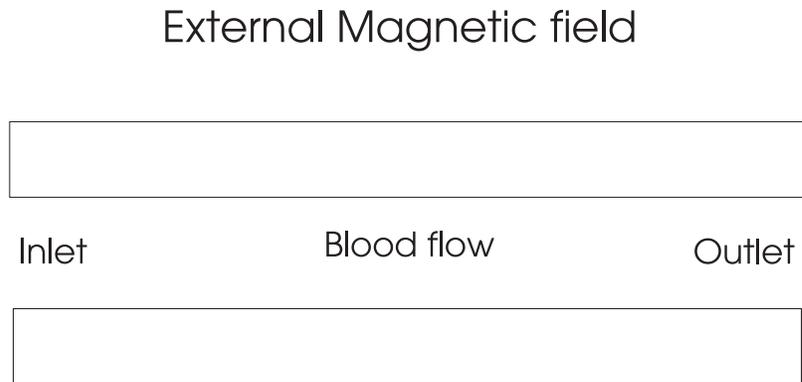
Fig. 24.1. The simplest 2D realization of a vessel with blood flow.

$$\mathbf{B} = \mu(\mathbf{H} + \mathbf{M}),$$

where μ is the permeability, $\mathbf{B}$ the magnetic flux density, $\mathbf{M}$ the magnetization vector, and $\mathbf{J}$ the current density. With the magnetic vector potential $\mathbf{A}$ ($\mathbf{B} = \text{div}\,\mathbf{A}$ and $\text{rot}\,\mathbf{A} = 0$) we obtain that

$$\nabla \times (\mu^{-1}\nabla \times \mathbf{A} - \mathbf{M}) = \mathbf{J}.$$

In the 2D case, this can be written as

$$-\nabla \cdot (\mu^{-1}\nabla A - \gamma) = J,$$

where γ is the magnetization of the ferrofluid given by

$$\gamma = \left(\alpha \arctan\left(\beta \frac{\partial A(x,\, y)}{\partial x} \right),\, \alpha \arctan\left(-\beta \frac{\partial A(x,\, y)}{\partial y} \right) \right)$$

as in [3]; also notice that $A = A_z$ and $J = J_z$.

It is important to point out that the vector potential was used to obtain the magnetic field $\mathbf{H}$, because of the numerical implementation. Note that $\mathbf{B} = \text{div}\,\mathbf{A}$ and $\mathbf{H} = \mu^{-1}\mathbf{B}$, which yields $\mathbf{H} = \mu^{-1}\text{div}\,\mathbf{A}$.

The time-dependent mass and momentum balances for an incompressible flow are described using the imcompressible Navier–Stokes equations

$$\rho\frac{\partial u}{\partial t} - \nabla \cdot \eta(\nabla u + (\nabla u)^T) + \rho(u \cdot \nabla)u + \nabla p = F,$$

where F involves gravity and magnetic forces $F_m + Fg$, the pressure term is represented by ∇p, and the viscous contribution is $\eta(\nabla u + (\nabla u)^T)$.

The components of the magnetic force are given as follows:

$$F_i^{\text{mag}}(x,\, y) = \mu_0(M_i(x,\, y) + M_j(x,\, y))^{\frac{1}{2}}\mu^{-1} \left(\frac{\frac{\partial A(x,y)}{\partial x}\frac{\partial^2 A(x,y)}{\partial x\partial x} + \frac{\partial A(x,y)}{\partial y}\frac{\partial^2 A(x,y)}{\partial x\partial y}}{\left(\left(\frac{\partial A(x,y)}{\partial y}\right)^2 + \left(\frac{\partial A(x,y)}{\partial x}\right)^2 \right)^{\frac{1}{2}}} \right)$$

$$F_j^{\text{mag}}(x,\, y) = \mu_0(M_i(x,\, y) + M_j(x,\, y))^{\frac{1}{2}}\mu^{-1} \left(\frac{\frac{\partial A(x,y)}{\partial x}\frac{\partial^2 A(x,y)}{\partial y\partial x} + \frac{\partial A(x,y)}{\partial y}\frac{\partial^2 A(x,y)}{\partial y\partial y}}{\left(\left(\frac{\partial A(x,y)}{\partial y}\right)^2 + \left(\frac{\partial A(x,y)}{\partial x}\right)^2 \right)^{\frac{1}{2}}} \right).$$

24.3 Materials and Methods

On the vessel a non-slip wall condition was applied, e.g., $u = 0$, $v = 0$. At the outlet we set up a pressure condition $p = 0$, and in the inlet boundary we applied a parabolic flow profile with x-velocity component $(v_0/4)S(1-S)(\gamma \sin(wt) + \sqrt{\sin^2(wt)})$, where

S is a segment length parameter that goes from 0 to 1 along the inlet boundary segment and v_0 is a limit velocity.

In order to avoid numerical instabilities that are inherent in the coupling of the two systems of differential equations as well as singularities that occur while getting approximating numerical solutions for each finite element, we made use of adaptive meshing techniques that allowed existence and smoothness of a global final numerical solution.

24.3.1 Numerical Results

Here we show the simulations of our modeling. Fig. 24.2 shows the contour lines of the external magnetic potential to be applied. Fig. 24.3 shows the streamline graph of the flow velocity field in the vessel at initial time. Fig. 24.4 shows the streamline graph of the flow velocity field in the vessel at time $t = 0.75$. Fig. 24.5 shows the effect of the magnetic field on the nanoparticles immersed in the blood flow. It is noticeable that above a time t, the magnetic nanoparticle carriers in the flow are disturbed in the vicinity of a magnetic field.

24.4 Conclusions

We have developed a numerical simulation for the behavior of magnetic nanoparticles in a carrier liquid. By means of our mathematical modeling we have given insight in the calculation of some of the parameters involved in drug targeting therapies.

The model that we have developed is a prerequisite for a more detailed and sophisticated model (multiscale model) which must incorporate the nanoscale effects and thus might allow a more realistic representation of the effects of magnetic fields and ferrofluids. Such an approach should also consider the regimes in which the magnetic fluid is likely to significantly influence the flow in a vessel.
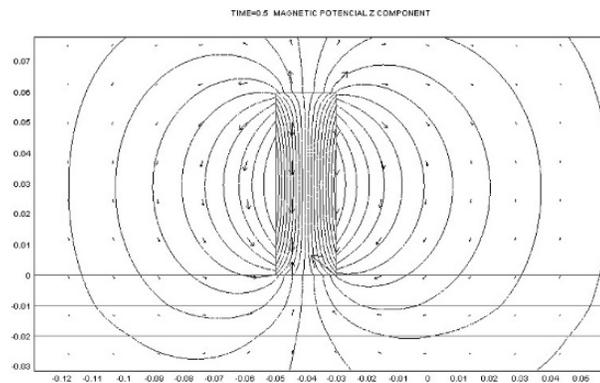
Fig. 24.2. Contour lines of the external magnetic potential.

TIME=0 FLUID VELOCITY FIELD
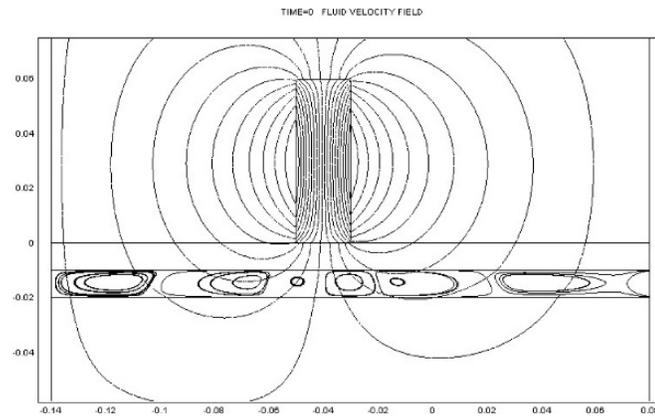
Fig. 24.3. Streamline of the vessel flow velocity field at time $t = 0$.

TIME=0.75 FLUID VELOCITY FIELD
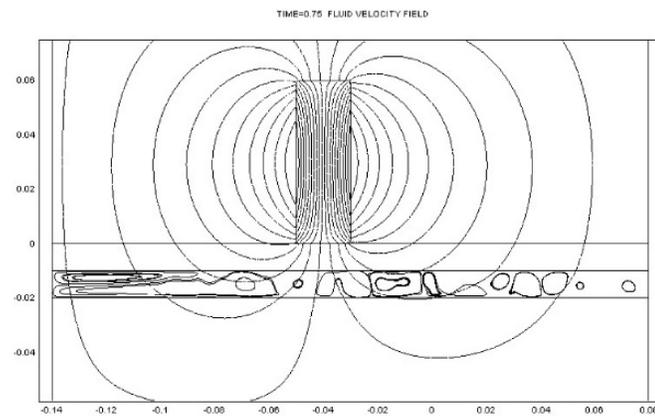
Fig. 24.4. Streamline of the vessel flow velocity field at time $t = 0.75$.

Time=0.75   Surface: Velocity field   Contour: Magnetic potential, z component
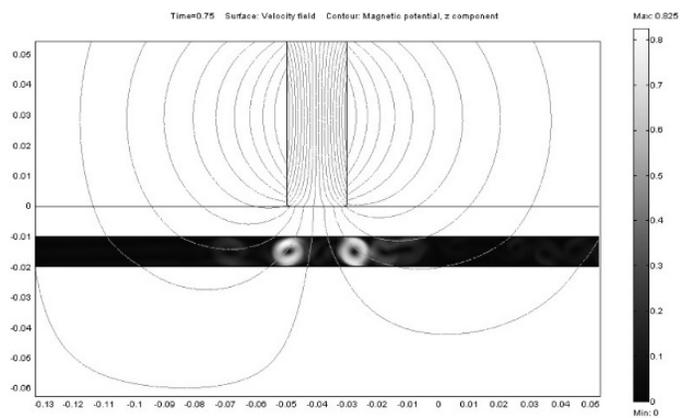
Fig. 24.5. Nanocarriers affected by the external magnetic field at time $t = 0.75$.

Acknowledgments

References

1. Voltairas, P.A., Fotiadis, D.I., Michalis, L.K.: Hydrodynamics of Magnetic Drug Targeting. J. of Biomechanics, **35**, 813–821 (2002).
2. Grief, A.D., Richardson, G.: Mathematical modelling of magnetically targeted drug delivery. Journal of Magnetism and Materials, **293**, 455–463 (2005).
3. Oldenburg, C.M., Borglin, S.E., Moridis, G.J.: Numerical simulation of ferrofluid flow for subsurface environmental engineering applications. Transport in Porous Media, **38**, 319–344 (2000).
4. Rosensweig, R.E.: Ferrohydrodynamics: Dover Publications, New York (1997).
5. Lübbe, S., Bergmann, C., Hunht, W., Frickmen, H., Riess, J., Brock, J.W. and Huhn, D.: Magnetically-controlled drug targeting. Cancer Research, **56** 4694 (1998).
6. Alexiou, Ch., Arnold, W., Klein, R.J., Klein, F.G., Parak, P., Hulin, C., Bergemman, W., Erhardt, W., Wagenpfeil, S. and Lübbe, A.S.: Locoregional cancer treatment with magnetic drug targeting. Cancer Research, **60**, 6641 (2000).

25

Fluid Transport in Peritoneal Dialysis:
A Mathematical Model and Numerical Solutions

Roman Cherniha,[1] and Vasyl' Dutka,[2] and Joanna Stachowska-Pietka,[3] and
Jacek Waniewski[3]

[1] Institute of Mathematics, NAS of Ukraine, Tereshchenkivs'ka Street 3, 01601 Kyiv,
   Ukraine; cherniha@imath.kiev.ua
[2] Institute of Superhard Materials, NAS of Ukraine, Avtozavods'ka Street 2, 02074 Kyiv,
   Ukraine
[3] Institute of Biocybernetics and Biomedical Engineering, PAS, Ks. Trojdena 4,
   02 796 Warszawa, Poland; jacekwan@ibib.waw.pl

Summary. A mathematical model of water flow between dialysis fluid in the peritoneal cavity
and blood through the capillary wall and homogeneous interstitium driven by high hydrostatic
and osmotic pressure of dialysis fluid is formulated. The model is based on nonlinear equations
of reaction-diffusion-convection type. Numerical simulations provide the distribution profiles
for hydrostatic pressure, glucose concentration, and water flux in the tissue for different times
from the infusion of dialysis fluid into the peritoneal cavity for different transport parameters
that represent clinical treatments of peritoneal dialysis.

Key words: Distributed model, numerical solution, water transport, glucose transport, peri-
toneal dialysis.

25.1 Introduction

Peritoneal dialysis is a life-saving treatment for chronic patients with end stage renal
disease [1]. Dialysis fluid is infused into the peritoneal cavity, and, during its dwell
there, small metabolites (urea, creatinine) and other uremic toxins diffuse from blood
to the fluid, to be removed after some time (usually a few hours) together with the
drained fluid. The treatment is repeated continuously. The peritoneal transport occurs
between dialysis fluid in the peritoneal cavity and blood passing down capillaries in
tissue surrounding the peritoneal cavity. The capillaries are distributed within the tis-
sue at different distances from the tissue surface that is in contact with the dialysis
fluid. Hydrostatic pressure of the dialysis fluid is higher than hydrostatic pressure of
the interstitial fluid within the tissue. Therefore, a flow of fluid from the peritoneal
cavity to the tissue occurs (called peritoneal absorption), which reduces the effective-
ness of peritoneal dialysis. The removal of excess water from the patient is achieved in
peritoneal dialysis by inducing high osmotic pressure in the dialysis fluid due to a high
concentration of glucose that creates high osmotic pressure of the fluid [1]. Glucose

(or another osmotic agent) diffuses from dialysis fluid dwelling in the peritoneal cavity into the tissue. Its increased concentration within the tissue induces osmotic flow of water across the wall of local capillaries from blood to the tissue [1]. How can this water pass from the tissue to the peritoneal cavity towards higher hydrostatic pressure? A mathematical description of osmotically induced fluid transport from blood to the peritoneal cavity has not been formulated fully yet, in spite of the well-known basic physical laws for these processes. A previous attempt to model water transport from a distributed system of capillaries was based on the assumption that a thin layer of cells (the mesothelium) on the surface of organs that surround the peritoneal cavity creates an osmotic barrier, similar to that formed by the capillary wall (the endothelium), and therefore a high osmotic pressure in the dialysis fluid overcomes its high hydrostatic pressure [2]. However, there is no evidence for such a function of the mesothelium [3]. Furthermore, that model predicted negative hydrostatic pressure in the tissue [2]. Experimental studies in rats demonstrated that hydrostatic pressure in the abdominal wall is positive during peritoneal dialysis with high glucose concentration in the dialysis fluid [4]. Thus, a model based on different assumptions should be formulated. In the present study we assume that the osmotic barrier is not created by the mesothelium layer on the surface of the tissue, but that this barrier is distributed within the tissue, i.e., an osmotic force may be created by a glucose concentration gradient at any point in the tissue. The consequences of such an assumption are studied by computer simulations. The model includes an additional assumption that, because of the high osmotic pressure induced by glucose, some thermodynamic forces, such as the oncotic pressure of proteins, may be neglected. A few exact solutions for the basic equations of the model were found using some analytical methods [5].

25.2 Mathematical Model

The mathematical description of transport processes within the tissue consists in the local balance of fluid volume and solute mass and is given by the following equations:

$$\frac{\partial v}{\partial t} = -\frac{\partial j_V}{\partial x} + q_V, \quad \frac{\partial (vC_G)}{\partial t} = -\frac{\partial j_G}{\partial x} + q_G, \tag{25.1}$$

where t is time, and x is the distance from the tissue surface in contact with dialysis fluid. Here the volumetric flux j_V and the solute flux j_G across the tissue are given by the formulas

$$j_V = -vK\frac{\partial P}{\partial x} + \sigma_{TG}vKRT\frac{\partial C_G}{\partial x}, \quad j_G = -vD_G\frac{\partial C_G}{\partial x} + S_{TG}C_G j_V, \tag{25.2}$$

while the density of fluid flux q_V and the density of glucose flux q_G from blood to tissue can be expressed as

$$q_V = -L_p a\sigma_{CG}RT(C_{GB} - C_G),$$
$$q_G = p_G a(C_{GB} - C_G) + S_{CG}q_V\left((1 - F_G)C_{GB} + F_G C_G\right). \tag{25.3}$$

We also assume that the fractional void volume, ν, is given by the function

$$\nu(P) = \nu_{\min} + \frac{\nu_{\max} - \nu_{\min}}{1 + \left(\frac{\nu_{\max} - \nu_{\min}}{\nu_0 - \nu_{\min}} - 1\right)\exp(-b(P - P_0))}, \tag{25.4}$$

which describes the experimental data on the increase of the volume filled by the interstitial fluid with the increase of hydrostatic pressure of this fluid [6].

Equations (25.1)–(25.4) can be united into two equations for finding the glucose concentration $C_G(t, x)$ and the hydrostatic pressure $P(t, x)$. The boundary conditions for a tissue layer of width L impermeable at $x = L$ and in contact with dialysis fluid at $x = 0$ are as follows:

$$x = 0 : P = P_D, \quad C_G = C_{GD},$$
$$x = L : \frac{\partial P}{\partial x} = 0, \quad \frac{\partial C_G}{\partial x} = 0. \tag{25.5}$$

The initial conditions describe equilibrium within the tissue without any contact with dialysis fluid at $x = 0$:

$$t = 0 : P = P_0, \quad C_G = C_{GB}. \tag{25.6}$$

The parameters in the above equations are selected according to the clinical and experimental data [4], [6]–[9]: $K = 5.14 \times 10^{-5}$ cm^2 $\times$ min^{-1} $\times$ mmHg^{-1}, hydraulic permeability of tissue; σ_{TG} is the Staverman reflection coefficient for glucose in tissue (varies from 0 to 0.01); $RT = 18 \times 10^3$ mmHg $\times$ mmol^{-1} $\times$ mL, the gas constant times temperature; $L_{Pa} = 7.3 \times 10^{-5}$ mL $\times$ min^{-1} $\times$ mmHg^{-1} $\times$ g^{-1}, the hydraulic permeability of the capillary wall; $\sigma_{CG} = 0.014$, the Staverman reflection coefficient for glucose in the capillary wall (as calculated using the two-pore model with small pore radius 60 Å and large pore radius 310 Å); $D_G = 12.11 \times 10^{-5}$ cm^2 $\times$ min^{-1}, the diffusivity of glucose in tissue divided by ν_0; $S_{TG} = 1 - \sigma_{TG}$, the sieving coefficient of glucose in tissue; $p_{Ga} = 3.4 \times 10^{-2}$ mL $\times$ min^{-1} $\times$ g^{-1}, the diffusive permeability of the capillary wall; $S_{CG} = 1 - \sigma_{CG}$, the sieving coefficient for glucose in the capillary wall; $F_G = 0.5$, a weighing factor; $C_{GB} = 6 \times 10^{-3}$ mmol $\times$ mL^{-1}, the glucose concentration in the blood; $C_{GD} = 180 \times 10^{-3}$ mmol $\times$ mL^{-1}, glucose concentration in the dialysate; $P_D = 3, 7.5$ and 12 mmHg, intraperitoneal hydrostatic pressure; $P_0 = 0$ mmHg, initial interstitial hydrostatic pressure; $L = 0.5$ cm; $\nu_{\min} = 0.17$; $\nu_{\max} = 0.35$; $\nu_0 = 0.175$; and $b = 1.7$ mmHg^{-1}. All pressures are scaled assuming atmospheric pressure is equal to zero. A detailed physical and physiological background for the model can be found in [5] (see also [1]–[4], [7,10]).

25.3 Numerical Results

A numerical technique described in [11] was used to solve the boundary-value problem (25.1)–(25.6). Numerical results for $P_D = 7.5$ mmHg and $\sigma_{TG} = 0.001$ (the values of other parameters are listed above and fixed) are presented in Fig. 25.1–25.3. Infu-
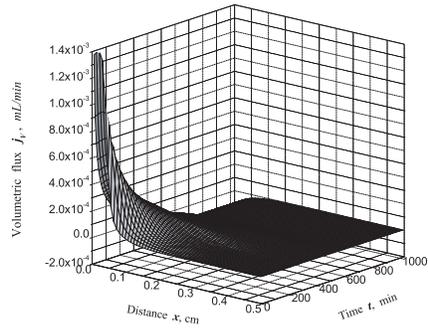
Fig. 25.1. Volumetric flux $j_V(t, x)$ for intraperitoneal pressure $P_D = 7.5$ mmHg and $\sigma_{TG} = 0.001$.
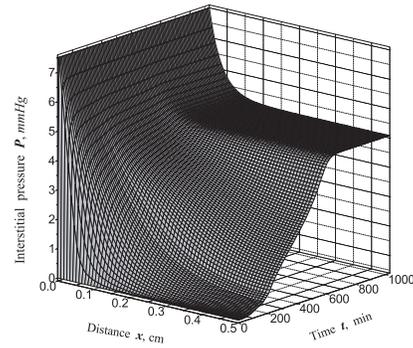
Fig. 25.2. Interstitial pressure $P(t, x)$ for intraperitoneal pressure $P_D = 7.5$ mmHg and $\sigma_{TG} = 0.001$.

sion of hypertonic solution into the peritoneal cavity increases intraperitoneal pressure and, consequently, causes rapid inflow of water into the surrounding tissue (Fig. 25.1). Inflowing water initially increases P only in the superficial layer, whereas P in the deeper tissue layers remains unchanged (Fig. 25.2). Later on, water penetrates deeper layers of the tissue, causing a local increase of P. Finally, a steady state is obtained at about 600 min after fluid infusion, with P decreasing with the distance from the cavity (Fig. 25.2). The inflow of water into the tissue decreases with time and finally, at the steady state, a reverse flow of water from the tissue into the cavity caused by the high concentration of glucose may be observed (Fig. 25.1). The glucose concentration in the tissue stabilizes quickly at a high level close to the cavity, but in the deeper part of the tissue it is close to the initial value (Fig. 25.3). A similar type of glucose distribution in the tissue was also obtained for $P_D = 3$ and 12 mmHg and for different values of σ_{TG} (data not shown). However, for higher values of P, deeper penetration of glucose may be observed.

Fig. 25.4–25.5 present results obtained for $\sigma_{TG} = 0.002$ and $\sigma_{TG} = 0.003$ with other parameters the same as in Fig. 25.2. It may be seen that the increase of osmotic pressure of glucose due to the increase of σ_{TG} has an effect on the quantitative behaviour of water flow and interstitial pressure (Fig. 25.4–25.5). An increase of σ_{TG} from 0.001 to 0.002 causes a reversed water flux, from the tissue into the cavity, at the beginning of dwell in the tissue layer close to the cavity. This outflow dehydrates the tissue and decreases P. Thus initially, negative P appears close to the peritoneal cavity, later on increasing due to the inflow of water into the tissue (Fig. 25.4). This effect may be even stronger for higher values of σ_{TG} (Fig. 25.5). Such a strong outflow of water, caused by the osmotic effect of glucose, permanently decreases P in the whole tissue layer below zero after about 100 min (Fig 25.5).

A similar behaviour of P may also be seen for $P_D = 3$ mmHg (not shown) and $P_D = 12$ mmHg (Fig. 25.6–25.8). Depending on the values of P_D and σ_{TG}, three
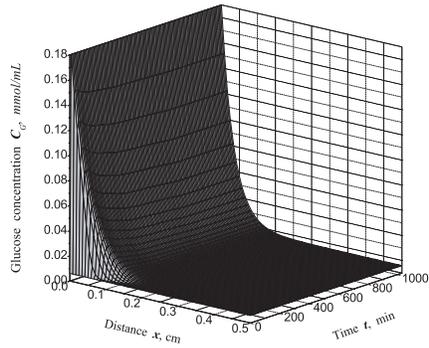
Fig. 25.3. Glucose concentration $C_G(t, x)$ for intraperitoneal pressure $P_D = 7.5$ mmHg and $\sigma_{TG} = 0.001$.
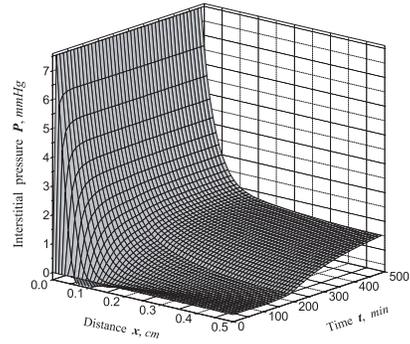
Fig. 25.4. $P(t, x)$ for $P_D = 7.5$ mmHg, $\sigma_{TG} = 0.002$.

types of behaviour of peritoneal pressure are possible. Pattern 1: P remains positive in the whole tissue layer (Fig. 25.6). Pattern 2: Negative pressure occurs at the beginning of dwell in the tissue layer close to the cavity (Fig. 25.7). If the reversed volumetric flux is not very strong and P_D is high enough, an increase of P follows with P positive at the steady state in the whole tissue (it occurs for $P_D = 3$ mmHg and $\sigma = 0.001$). Pattern 3: Negative pressure initially occurs only in the tissue layer close to the cavity, whereas in the deeper part of the tissue P is close to zero. However, the strong influence of glucose, which causes an outflow of water from the tissue, causes a further decrease
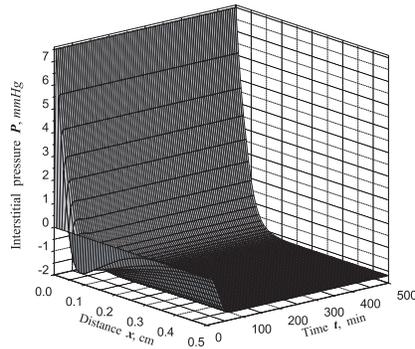
Fig. 25.5. $P(t, x)$ for $P_D = 7.5$ mmHg, $\sigma_{TG} = 0.003$.
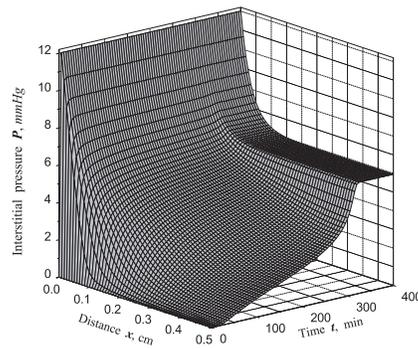
Fig. 25.6. $P(t, x)$ for $P_D = 12$ mmHg, $\sigma_{TG} = 0.002$.
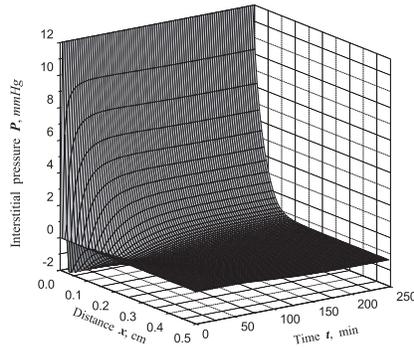
Fig. 25.7. $P(t, x)$ for $P_D = 12$ mmHg, $\sigma_{TG} = 0.004$.
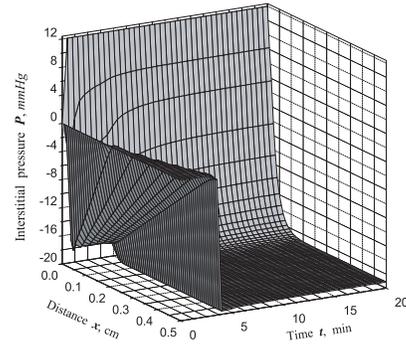
Fig. 25.8. $P(t, x)$ for $P_D = 12$ mmHg, $\sigma_{TG} = 0.010$.

of pressure in the tissue. Finally, at the steady state, P is negative in most of the tissue layer (Fig. 25.7 and 25.8).

The reverse water flow (i.e., out of the tissue) that may occur close to the cavity decreases with the distance from the peritoneal cavity to zero in the deeper tissue layers. The total fluid outflow from the tissue to the cavity, calculated assuming that the surface area of the contact between dialysis fluid and peritoneum is equal to 0.5 m^2, is similar at the steady state for all values P_D and σ_{TG} and varies from -0.24 mL/min to -0.5 mL/min. Furthermore, in simulations with high σ_{TG}, an outflow from the tissue was observed for a short initial period of the peritoneal dwell, with the maximal rate of -5 mL/min obtained for $\sigma_{TG} = 0.01$ and $P_D = 12$ mmHg.

25.4 Discussion

The model of fluid transport in peritoneal dialysis with hypertonic glucose solution presented in our study allows for analysis of the distribution of hydrostatic and osmotic pressures and water flux in the tissue. Our aim was to check if the assumption of the osmotic barrier distributed uniformly in the tissue allows for the high osmotic flux of water from blood to the peritoneal cavity and concomitant positive hydrostatic pressure in the tissue. The results show that the model is not able to describe these phenomena in agreement with clinical and experimental data. The rate of flow to the peritoneal cavity with hypertonic glucose solution is about 10–15 mL/min immediately after the infusion of dialysis fluid. In our simulations the maximal flow rate was about 5 mL/min. Although higher values of outflow can be easily obtained using higher values of σ_{TG}, the cost for this initial outflow from the tissue is a negative hydrostatic pressure (about -19.1 mmHg at $x = 0.5$ cm, Fig. 25.8) in the tissue. Moreover, this outflow is transient and results from dehydration of the tissue rather than a persistent outflow from the blood. Although such dehydration was discussed in the medical lit-

erature, it was not found in experimental studies [12]. The persistent outflow from the tissue can be observed in our simulations for low values of σ_{TG}, a weak osmotic barrier, but only after a long time (>9 hours, Fig. 25.1) and at a low rate (< 0.5 mL/min). This prediction is difficult to test, because in real dialysis the concentration of glucose quickly decreases. In any case, this steady state outflow from the tissue cannot explain the high initial values measured under clinical conditions. The presented model as well as the previous one [2] cannot describe the currently available clinical and experimental knowledge of the water transport in peritoneal dialysis. Although our model is simplified, there is little hope that its further development can yield better agreement with the data. Instead, a modelling based on the assumption of a heteroporous osmotic barrier (similar to that created by the capillary wall) distributed within the tissue may perhaps be able to describe water inflow to the tissue due to high intraperitoneal hydrostatic pressure and concomitant water outflow from the blood through the tissue to the peritoneal cavity due to high osmotic pressure of the dialysis fluid.

Acknowledgments

References

1. Gokal, R., Nolph, K.D.: The textbook of peritoneal dialysis. Kluwer, Dordrecht (1994).
2. Seames, E.L., Moncrief, J.W., Popovich, R.P.: A distributed model of fluid and mass transfer in peritoneal dialysis. Am. J. Physiol., **258**, R958–72 (1990).
3. Flessner, M.F.: Peritoneal transport physiology: insights from basic research. J. Am. Soc. Nephrol., **2**, 122–135 (1991).
4. Flessner, M.F.: Transport of protein in the abdominal wall during intraperitoneal therapy. I. Theoretical approach. Am. J. Physiol. Gastrointest Liver Physiol., **281**, G424–37 (2001).
5. Cherniha, R., Waniewski, J.: Exact solutions of a mathematical model for fluid transport in peritoneal dialysis. Ukrainian Math. J., **57**, 1112–1119 (2005).
6. Zakaria, E.R., Lofthouse, J., Flessner, M.F.: In vivo effects of hydrostatic pressure on interstitium of abdominal wall muscle. Am. J. Physiol., **276**, H517–529 (1999).
7. Waniewski, J.: Physiological interpretation of solute transport parameters for peritoneal dialysis. J. Theor. Med., **3**, 177–190 (2001).
8. Waniewski, J.: A mathematical model of local stimulation of perfusion by vasoactive agent diffusing from tissue surface. Int. J. Cardiovas. Eng., **4**, 115–123 (2004).
9. Imholz, A.L., Koomen, G.C., Voorn, W.J., Struijk, D.G., Arisz, L., Krediet, R.T.: Day-to-day variability of fluid and solute transport in upright and recumbent positions during CAPD. Nephrol. Dial. Transplant., **13**, 146–153 (1998).
10. Stachowska-Pietka, J., Waniewski, J., Flessner, M.F., Lindholm, B.: A mathematical model of peritoneal fluid absorption. Am. J. Physiol. Heart Circ. Physiol., **291**, 1862–1874 (2006).

11. Cherniha, R., Dutka, V.: Exact and numerical solutions of the generalized Fisher equation. Reports of Math.Phys., **47**, 393–411 (2001).
12. Flessner, M.F.: Osmotic barrier of the parietal peritoneum. Am. J. Physiol., **267**, F861–870 (1994).

Relevance of Intracellular Replication to the Evolution of Chagas' Disease

G.J. Sibona,[1] C.A. Condat,[2] and S. Cossy Isasi[3]

[1] Max-Planck-Institut für Physik Komplexer Systeme, Dresden, Germany;
    gsibona@mpipks-dresden.mpg.de
[2] CONICET and FaMAF, Universidad Nacional de Córdoba, Argentina
[3] Facultad de Cs. Médicas, Universidad Nacional de Córdoba, Argentina

Summary. We discuss a model for the interaction between the parasite *Trypanosoma cruzi* and the immune system in Chagas disease, by separately describing the intracellular and extracellular parasite stages. The solution of the case where two antibody species are active is worked out in detail, and a diagram showing the differents outcomes of the model is obtained. Our predictions accurately reproduce experimental data on the infection evolution during the acute phase of the disease and lead to an estimate of the damage generated by direct parasite action.

Key words: Chagas disease, parasite-antibody interaction, population dynamics.

26.1 Introduction

Chagas disease is an endemic parasitosis that affects nearly 20 million people in Central and South America [1]. This disease is caused by the protozoan parasite *Trypanosoma cruzi*, which is usually transmitted by a bloodsucking insect of the subfamily Triatominae. Parasite invasion is associated with strong antibody and cellular responses; the initial, acute, stage of the disease is characterized by a febrile illness; the disease then enters into a latent stage in which there are no clinical signals related to the infection. Many years after the infection, infected individuals often develop severe heart disorders including life-threatening electrical conduction abnormalities and dilated cardiomyopathy [2]. Although its social and economic effects are well known [1,3], there is, at present, no effective vaccine against Chagas.

   T. cruzi has a complex life cycle involving three main morphological stages: replicating epimastigotes are found in the insect vector, while an intracellular reproductive form (amastigote) and a blood circulating infective form (trypomastigote) are identified in the mammalian host. Trypomastigotes present in the insect feces invade the host through mucosae or skin microlesions. Inside the host they penetrate cells in a variety of tissues, where they differentiate into amastigotes. After several cycles of duplication by binary fission, they transform again into the nonreplicative trypomastigote stage, filling and bursting the cell, which releases them into the bloodstream.

We developed a model of the acute phase of the infection that predicts all possible outcomes of the disease: healing, death, and chronic infection, with stationary or quasicyclical populations [4,5]. We obtained very good agreement between the predictions of the model and available experimental results. Minor but definite discrepancies remained, however, especially when comparing the observed parasite numbers in the late infection stages with the model predictions. Very recently, and motivated by this limitation, we generalized the model by separately describing the parasitic intracellular and extracellular phases [6]. Much better fits were obtained with the improved model, as we show in the text. Since we need to consider two specific antibody species, it is advisable to undertake a more careful analysis of the model in this case. In addition to comparison with the experiments, we present here the explicit steady-state solutions of the two-antibody-species model and study the dynamics of the different populations numerically.

26.2 Materials and Methods

Our model is built upon the following assumptions:

1. A parasite batch invades or is inoculated into a mammalian host.
2. Trypomastigotes invade healthy cells, where they mutate into amastigotes and reproduce.
3. Newborn parasites break out of the cell and restart the invasive process.
4. Parasites activate antibody generation.
5. One or more antibody species mediate extracellular parasite removal.

We construct equations for the time dependence of the relevant cell and antibody populations. Defining the infectivity ζ as the rate at which a circulating parasite penetrates a host cell and the cytotoxicity η as the rate at which infected cells burst, the equation describing the number $r(t)$ of infected cells can be formulated as

$$\dot{r}(t) = \zeta n(t) - \eta r(t), \tag{26.1}$$

where $n(t)$ is the number of circulating trypomastigotes. The evolution equation for $n(t)$ is

$$\dot{n}(t) = \eta N_r r(t) - n(t) \sum_{i=1}^{N} \alpha_i(t) a_i(t) - \zeta n(t), \tag{26.2}$$

where N_r is the mean number of trypomastigotes emerging from a ruptured cell, $a_i(t)$ is the number of molecules of the ith antibody species, and $\alpha_i(t)$ is the removal efficiency, which is assumed to be a nondecreasing function of time: $\alpha_i(t) = \alpha_{A,i} + \alpha_{B,i}(1 - \exp[-t/T_i])$, where T_i is a "learning time."

Table 26.1 lists the main variables used in this work.

To construct the evolution equation for $a_i(t)$, we introduce N fixed values $a_{i,0}$ that represent a continuous source exactly compensating for the spontaneously inactivated

Table 26.1. Main variables used in this work.

Variable	Meaning
a_i	number of molecules of the ith antibody species
n	number of circulating trypomastigotes
r	number of infected cells
$a_{i,0}$	initial number of molecules of antibody species i
n_0	number of inoculated parasites
ζ	infectivity, rate at which a circulating parasite penetrates a host cell
η	cytotoxicity, rate at which infected cells burst
N_r	mean number of trypomastigotes emerging from a ruptured cell
α_i	parasite removal efficiency of antibody species i
γ_i	induced antibody production rate of species i
τ_i	intrinsic lifetime of antibodies belonging to species i
Ω	mean amastigote duplication time
κ_r	$\zeta(N_r - 1)$ effective reproduction rate
T_i	antibody learning time of species i

antibody molecules in the absence of parasites. If γ_i and τ_i are, respectively, the induced antibody production rate and the intrinsic lifetime of antibodies belonging to species i, antibody numbers evolve according to

$$\dot{a}_i(t) = (1/\tau_i)\left[a_{i,0} - a_i(t)\right] + \gamma_i n(t) - \alpha_i(t)a_i(t)n(t). \tag{26.3}$$

The parameters η and N_r can be related by considering a simple Malthusian model for the amastigote replication dynamics. If at time t_0 a parasite penetrates a healthy cell, the number $\tilde{n}(t)$ of amastigotes at a time $t \in [t_0, T_b]$, where T_b is the cell-burst time, is given by $\tilde{n}(t) = \exp[\beta(t - t_0)]$, where β is the amastigote duplication rate. Therefore, $N_r = \exp[\beta(T_b - t_0)]$. Since the mean amastigote duplication time is $\Omega = \beta^{-1}\ln 2$, we get

$$\Omega = (1/\eta)\left(\ln 2/\ln N_r\right). \tag{26.4}$$

Typically, $\Omega \sim 8$–21 hours [7]. Once the size n_0 of the initial inoculation and the form of the removal functions $\alpha_i(t)$ are known, the set of equations (26.1) to (26.3) can be solved to obtain the time dependence of the various populations.

26.3 Results and Discussion

26.3.1 Steady State: Two Antibody Species

The steady-state populations can be obtained by setting the time derivatives equal to zero and letting $\lim_{(t\to\infty)} \alpha_i(t) = \alpha_i$. From Eq. (26.3), the steady-state antibody numbers are given by

$$a_{is} = \frac{a_{i0} + \gamma_i \tau_i n_s}{1 + \alpha_i \tau_i n_s}. \tag{26.5}$$

Substituting Eqs. (26.5) and (26.1) into (26.2), it is easy to see that the steady-state parasite number, n_s, must satisfy an Nth-degree algebraic equation,

$$\sum_{i=1}^{N} \gamma_i - \kappa_r = \sum_{i=1}^{N} \frac{\gamma_i - \alpha_i a_{i0}}{1 + \alpha_i \tau_i n_s}, \tag{26.6}$$

where $\kappa_r = \zeta(N_r - 1)$. Of course, the trivial solution ($n_s = r_s = 0, a_{is} = a_{i0}$ for all i) is also possible. It would correspond to a healthy host.

The conditions for the existence and stability of the solutions to (26.5) and (26.6) can be worked out. In a previous work [6], we discussed the steady-state solutions of the one-antibody problem, finding the corresponding phase diagram. Here we extend our analysis to the case of two antibody species. If $N = 2$, (26.6) becomes a quadratic, whose solutions can be expressed as

$$n_{s\pm} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}, \tag{26.7}$$

where

$$A = \tau_1 \tau_2 \alpha_1 \alpha_2 (\gamma_1 + \gamma_2 - \kappa_r),$$
$$B = (\gamma_1 + \gamma_2 - \kappa_r)(\tau_1 \alpha_1 + \tau_2 \alpha_2) - \tau_1 \alpha_1 (\gamma_2 - \alpha_2 a_{20}) - \tau_2 \alpha_2 (\gamma_1 - \alpha_1 a_{10}),$$
$$C = \alpha_1 a_{10} + \alpha_2 a_{20} - \kappa_r. \tag{26.8}$$

We can now construct a three-dimensional phase diagram in the $\kappa_r - \gamma_1 - \gamma_2$ space. To build it, notice that the condition $C = 0$ defines a plane $\kappa_r = $ constant that divides the phase space into two domains. We consider them separately.

(i) $C < 0$. There are two cases. If $A < 0$, i.e., if $\gamma_1 < \kappa_r - \gamma_2$, no steady-state solution is possible: weak antibody formation combines with strong parasite reproduction to ensure indefinite growth of the parasite population, leading to host death (case III in Fig. 26.1). If $A > 0$, on the other hand, the Routh–Hurwitz criterion [8] can be used to show that the solution n_{s+} is stable: chronic disease develops (case I). By letting $A \to 0^+$, i.e., as we approach the host death case, we observe that n_{s+} diverges, as it should. The plane $\kappa_r - \gamma_1 - \gamma_2 = 0$ separates the death and chronic domains.

(ii) $C > 0$. If $A < 0$ (case IIIb), the outcome depends on the size of the initial inoculation: although n_{s+} is always possible, a small invasive parasite batch n_0 will be eliminated by the initial antibody population, while a large n_0 will lead to the chronic stage. If $A > 0$ the situation is somewhat more complex: If $B > 0$ or if $B < 0$ and $B^2 - 4AC < 0$, only the trivial solution is possible and the host always heals (case II). If $B < 0$ and $B^2 - 4AC > 0$ (case Ib), both solutions are possible, but the Routh–Hurwitz criterion indicates that n_{s-} is unstable. If $n_0 < n_{s-}$, the host heals, while if $n_0 > n_{s-}$, $n(t)$ tends to the upper steady-state solution n_{s+}. This situation is reminiscent of harvesting models [8,9]; here the role of "harvester" is played by the second antibody species. We remark that case Ib is not present when a single antibody
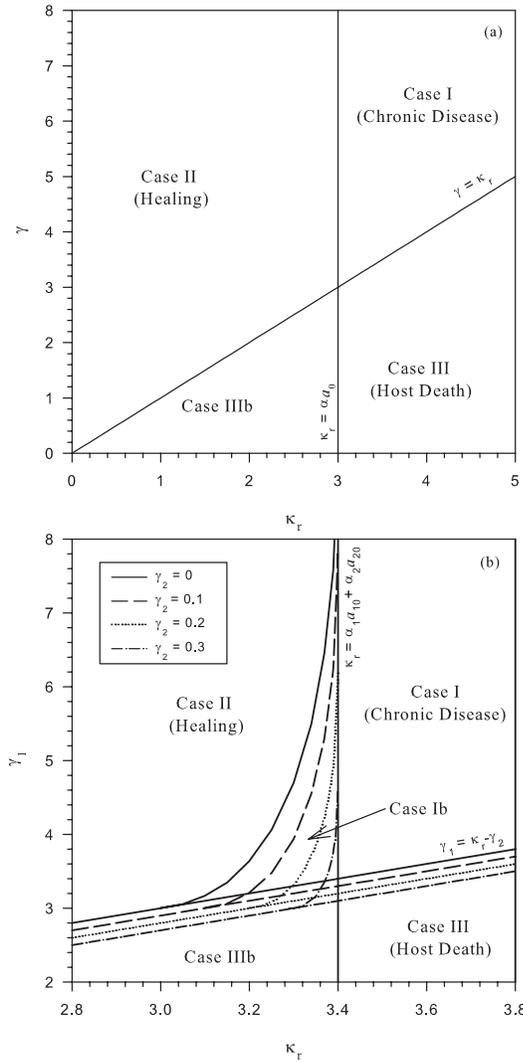
Fig. 26.1. (a) $\kappa_r - \gamma$ phase diagram corresponding to the one-antibody species problem. Here $\alpha = 1, \tau = 1$, and $a_0 = 3$. (b) Cross sections of the phase diagram describing the outcome of the acute phase of the parasite infection when two antibody species are active, for the γ_2 values specified in the figure. We chose $\alpha_1 = 1, \alpha_2 = 0.4, \tau_1 = 1, \tau_2 = 40, a_{10} = 3$, and $a_{20} = 1$. In cases Ib and IIIb the outcome depends on the initial conditions.

is active [6]. The domains corresponding to cases II and Ib are separated by the surface $B^2 = 4AC$.

Fig. 26.1(a) is the phase diagram corresponding to the single-antibody problem (see Ref. [6]), while Fig. 26.1(b) exhibits cross sections of the (three-dimensional) phase diagram taken for the indicated values of γ_2. Three differences with the single-

antibody solution stand out: the presence of region Ib, the rightward shift of the vertical phase boundary between cases I and II, and the downward shift of the slanted line corresponding to the separation between cases I and II on the upper side and case III on the lower side. By looking at Fig. 26.1(b), we notice that the size of region Ib decreases with increasing γ_2: healing is favored if the induced antibody production rate is high. We further note that the $\gamma_2 \to 0$ limit does not correspond to the single-antibody problem. Even if less plentiful, the second antibody species is still present.

26.3.2 Competition Dynamics

Time-dependent solutions to Eqs. (26.1) through (26.3) are easy to obtain numerically by using a standard Euler method. The parasite and antibody populations are shown in Fig. 26.2 as functions of time for parameter values corresponding to case Ib. Under these conditions, the coexistence of two steady states makes the model outcome dependent on the inoculation size. For the smaller n_0, the parasites are completely eliminated and the numbers of antibody molecules return to their initial values. For the larger inoculation size the parasites survive and the system reaches a steady state corresponding to chronic disease. In both cases we observe a fast initial reduction in the trypomastigote numbers. The initial absence of infected cells means that there are no parasite sources; therefore, those that are destroyed or that have infected new cells cannot be replaced. For intermediate times, the infected cells have already started to replenish the parasite numbers, which continue to decrease, albeit at slower rates, but the a_2 population also decreases due to the relative weakness ($\gamma_2 = 0.1$) of their pro-
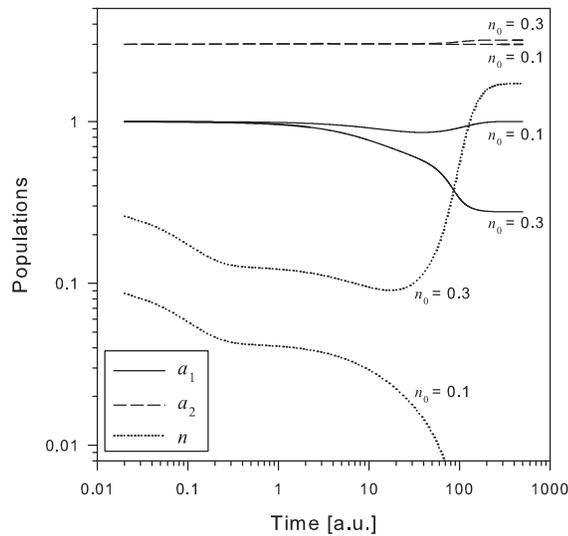
Fig. 26.2. Case Ib: evolution of the parasite and antibody populations for two inoculation sizes: $n_0 = 0.1$ and $n_0 = 0.3$. Parameter values are: $\kappa_r = \zeta = 3.3$, $\eta = 5$, $N_r = 2$, $r(0) = 0$, $\gamma_1 = 3.3$, $\alpha_1 = 1$, $\tau_1 = 1$, $a_{01} = 3$, $\gamma_2 = 0.1$, $\alpha_2 = 0.4$, $\tau_2 = 40$, and $a_{02} = 1$.

duction rate. At the end of this intermediate stage the parasite number may reach the threshold: if $n(t)$ decreases below the unstable steady-state value ($n_{s-} = 0.036$ for the parameter values chosen here), the production of new parasites cannot compensate for those destroyed by the immune system attack and the system converges towards the healing attractor. But, if this value is not reached, the parasite survives the immune reaction and the system enters the chronic stage.

26.3.3 Comparison with Experiments

We compared the predictions of the extended model (separate intracellular and extracellular populations) with those of the original version by fitting experimental data for BALB/c mice infected with *T. cruzi* (Tehuantepec strain) [10]. Three antibody species, IgM, IgG1, and IgG2a, were included in the model, although the influence of IgM turned out to be very weak. The observation that parasite control is almost completely effected by the two IgG species underlines the importance of achieving a clear understanding of two-antibody species processes. The parameters were chosen to attain the closest visual agreement between experimental data and model predictions. By keeping the antibody parameters as in the original model (see Table 26.2), we obtained: $N_r = 13.8$, $\zeta = 2.1$ [1/w], $\eta = 2.1$ [1/w], and $n_0 = 100$ [1/ml].

The results are presented in Fig. 26.3. It is evident that the modified version yields a better fit, especially towards the end of the acute stage. This has been confirmed by the application of the second-order (corrected) Akaike information criterion (AIC) [11], which shows a substantial reduction (from 32.5 to -7.5) in the value of the AIC estimator, defined as

$$AIC = N \ln\left(\frac{SS}{N}\right) + 2K + \frac{2K\,(K+1)}{N-K-1}, \tag{26.9}$$

where N is the number of data points, K is the number of parameters, and SS is the sum of squares of the differences between the model predictions and the experimental points.

We have also successfully fitted the data for various parasite strains (Tulahuen and CA-1) and mice (CBA/J and BALB/c), comparing the different model predictions of damage produced in the host through cell destruction with the in vivo observations. Remarkably, we found that the number of cells destroyed by the CA-1 strain, which

Table 26.2. Parameters for *T. cruzi* specific antibodies obtained fitting the experimental data of El Bouhdidi et al. [10]. Here w stands for weeks.

Isotype	a_0 [ml^{-1}]	α_A [ml/w]	α_B [ml/w]	γ [w^{-1}]	T [w]	τ [w]
IgM	6×10^{15}	4.6×10^{-20}	3.3×10^{-20}	2×10^{10}	2.4	8.3
IgG1	2.5×10^{17}	0	1.2×10^{-18}	8×10^{11}	6.1	7.1
IgG2a	1.2×10^{18}	0	4.1×10^{-18}	25×10^{11}	7.1	7.1
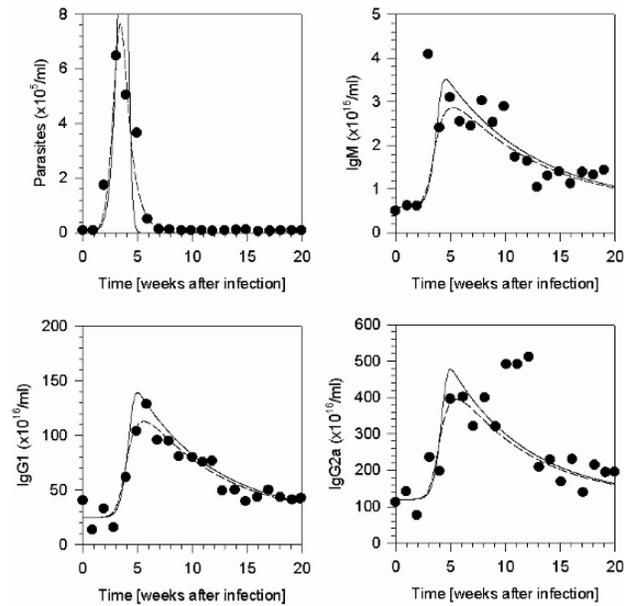
Fig. 26.3. Fits to the experimental data of Ref. [10] using the original [5] (smooth lines) and current (dashed lines) versions of the parasite-antibody competition model.

only causes a mild and late parasitemia, is much higher than the number corresponding to the "strong" Tulahuen strain, in agreement with the experimental observations [6].

The values of the amastigote doubling times obtained by fitting our model to different pairs of parasite/mouse strains are: Tehuantepec/BALBc: 21.2 hrs, Tulahuen/CBAJ: 18.7 hrs, Tulahuen/BALBc: 22.1 hrs, and CA-1: 82.1 hrs. These values of Ω are similar to those obtained experimentally [7].

26.4 Conclusions

We have reviewed a model for the *T. cruzi*-immune system competition in Chagas disease, showing that the separate modeling of the intracellular and extracellular parasite stages yields a very accurate agreement with observed parasite and immunoglobulin populations. This agreement validates the model assumptions. A new element strengthening our model formulation is the consistency found between the predicted and observed amastigote duplication times.

Since two antibody species are required to obtain accurate fits, we have analyzed the steady-state and time-dependent solutions of the dynamical system formed by intracellular and extracellular parasites and two antibody species. The steady states are conveniently presented using phase diagrams that reveal the various possible outcomes of the Chagas infection. The introduction of a second antibody species leads to the emergence of a new region in the phase diagram, where the infection outcome (healing

or chronicity) depends on the size of the parasite inoculation. If the induced antibody production rate of the second antibody is increased, this region shrinks, favoring host healing.

Acknowledgments

References

1. Report of a WHO Expert Committee: Control of Chagas Disease. WHO Technical Report Series, **811**, WHO, Geneva (1991).
2. Cunha-Neto, E., Moliterno, R., Coelho, V., Guilherme, L., Bocchi, E., et al.: Restricted heterogeneity of T cell receptor variable alpha chain transcripts in hearts of Chagas' disease cardiomyopathy patients. Parasite Immunol., **16**, 171–179 (1994).
3. Rodrigues-Coura, J., Junqueira, A.C.V., Fernandes, O., Valente, S.A.S., Miles, M.A.: Emerging Chagas disease in Amazonian Brazil. Trends in Parasitology, **18**, 171–176 (2002).
4. Cossy Isasi, S., Sibona, G.J., Condat, C.A.: A simple model for the interaction between *Tripanosoma cruzi* and its antibodies during Chagas infection. J. Theor. Biol., **208**, 1–13 (2001).
5. Condat, C.A., Cossy Isasi, S., Sibona, G.J.: Parasite-antibody competition in Chagas disease. Comments on Theoretical Biology, **8**, 587–607 (2003).
6. Sibona, G.J., Condat, C.A., Cossy Isasi, S.: Dynamics of the antibody-*T. cruzi* competition during Chagas infection: Prognostic relevance of intracellular replication. Phys. Rev. E, **71**, 020901(R), 1–4 (2005).
7. Engel, J.C., Doyle, P.S. Dvorak, J.A.: *Trypanosoma cruzi*: biological characterization of clones derived from chagasic patients. II. Quantitative analysis of the intracellular cycle. J. Protozool., **32**, 80–83 (1985).
8. Murray, J.D.: Mathematical Biology, 2nd Ed. Springer-Verlag, Berlin (1993).
9. Banchio, A.J., Condat, C.A.: Seasonality and harvesting, revisited. In: Delsanto, P.P. (ed.) Universality of Nonclassical Nonlinearity: Applications to Non-Destructive Evaluations and Ultrasonics. Springer, Berlin (2006).
10. El Bouhdidi, A., Truyens, C., Rivera, M.T., Bazin, H., Carlier, Y.: *Trypanosoma cruzi* infection in mice induces a polyisotypic hypergammaglobulinaemia and parasite-specific response involving high IgG2a concentrations and highly avid IgG1 antibodies. Parasite Immunol., **16**, 69–76 (1994).
11. Motulsky, H., Christopoulos, A.: Fitting Models to Biological Data Using Linear and Nonlinear Regression. A Practical Guide to Curve Fitting, 2nd Ed. GraphPad Software Inc., San Diego, CA (2003).

A Finite Volume Spatial Discretisation for Taxis-Diffusion-Reaction Systems with Axi-Symmetry: Application to Fracture Healing

Alf Gerisch[1] and Liesbet Geris[2]

[1] Martin-Luther-Universität Halle-Wittenberg, Institut für Numerische Mathematik, 06099 Halle (Saale), Germany; alf.gerisch@mathematik.uni-halle.de
[2] K.U. Leuven, Faculty of Engineering, Division of Biomechanics and Engineering Design, Celestijnenlaan 300C, B-3001 Leuven, Belgium; liesbet.geris@mech.kuleuven.be

Summary. We consider the numerical simulation of a time-dependent taxis-diffusion-reaction model of fracture healing in mice using the method of lines. The partial differential equation problem has an axi-symmetric structure and this is employed to properly reduce the model to an equivalent problem in two-dimensional (2D) space leading subsequently to an efficient spatial discretisation. Special care is given to respect conservation of mass and the non-negativity of the solution. The numerical simulation results are contrasted to those obtained from a simplistic reduction of the axi-symmetric model to 2D space (at the same computational cost). We observe quantitative and qualitative differences.

Key words: Taxis-diffusion-reaction model, axi-symmetry, numerical method, finite volume discretisation, fracture healing.

27.1 Introduction

In the theoretical study of bone fracture healing, the assumption of axi-symmetry of the bones is ubiquitous and taxis-diffusion-reaction (TDR) systems are often used. Numerical methods for the simulation of TDR systems in one, two and more space dimensions are investigated in [1] and results of their successful application to a diverse selection of TDR problems in two dimensions in the field of mathematical biology are presented in [2]. TDR systems are time-dependent partial differential equation (PDE) systems whose solution components describe the spatio-temporal evolution of chemical concentrations, and cell and other densities. For the numerical solution of these systems, the method of lines (MOL) is followed in [1]. The spatial discretisation of the PDE system follows a finite volume approach and leads to the MOL-ODE, a large system of ordinary differential equations (ODEs), which is then integrated in time to yield the final numerical solution of the TDR system. The spatial discretisation is designed such that two qualitative properties of the TDR system are mimicked by the exact solution of the MOL-ODE: first, the conservation of mass property, and

second, the property that concentrations/densities are non-negative quantities. Furthermore, the spatial discretisation is of second-order accuracy, in general. This higher than first-order property allows for suitably coarse grids while ensuring a sufficiently high accuracy (see [3] for an instructive example). A coarse grid translates into a MOL-ODE of more moderate dimension and hence a moderate computation time for the time integration.

In Sec. 27.4, we extend the finite volume spatial discretisation from [1,2] to axi-symmetric (or cylindrically symmetric) TDR problems in three space dimensions as defined in Sec. 27.3. This is achieved by reducing the axi-symmetric problem to an equivalent TDR system in 2D space and then applying the techniques described in [1,2] to obtain the MOL-ODE representing the axi-symmetric problem. That way we ensure the conservation of mass property and non-negativity of the exact solution of the MOL-ODE. Furthermore, that MOL-ODE has a dimension which scales only like h^{-2} with decreasing typical grid width h (in contrast to h^{-3} for general 3D problems) and the evaluation of its right-hand side has about the same computational cost as for the corresponding 2D problem. This renders our algorithm computationally efficient while capturing the 3D nature of the problem.

To further motivate our work, we give, in Sec. 27.2, an example with circular symmetry which demonstrates, by generalization, that the proper reduction of axi-symmetric TDR problems to equivalent TDR problems in two dimensions is important. In Sec. 27.5, we apply the discretisation developed in Sec. 27.4 to an axi-symmetric TDR model of bone fracture healing. Our conclusions are presented in Sec. 27.6.

27.2 An Instructive Example

In this section we present a simple, yet significant, example which demonstrates why a proper treatment of axi-symmetric TDR systems is necessary. For simplicity, we consider a problem with circular symmetry in 2D space instead of axi-symmetry in 3D space as considered in Sec. 27.3. However, the conclusions hold for both problem types.

Let $\Omega \subset \mathbb{R}^2$ be a ring-shaped domain with inner radius $r_0 = 0.01 \geq 0$ and outer radius $R_0 = 1.01 > r_0$. For $(x, y) \in \Omega$ and times $t > 0$ we consider the scalar diffusion-reaction equation for $u(t, x, y)$ given by

$$\partial_t u = \Delta u + g(u). \tag{27.1}$$

We consider two different nonlinear reaction functions,

$$g(u) = g_1(u) := \alpha u(1 - u) \quad \text{and} \quad g(u) = g_2(u) := -\alpha u(1 - u)(\beta - u).$$

Function g_1, with parameter $\alpha \geq 0$, is a logistic growth term and as such is present as part of many PDE models of biological processes. Function g_2, with parameters $\alpha \geq 0$ and $-1 \leq \beta < 1$, is a cubic nonlinearity; for $0 < \beta < 1$ the resulting diffusion-reaction equation is used as a model of heterozygote inferiority in genetics [4].
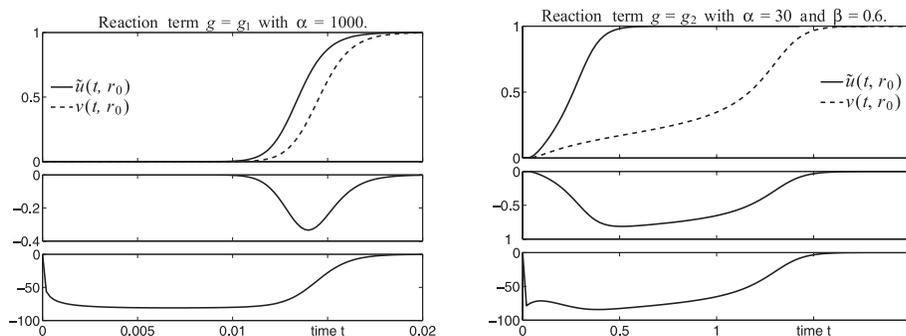
Fig. 27.1. Plots of $\tilde{u}(t, r_0)$ and $v(t, r_0)$ (top), and their absolute (middle) and relative (bottom, in %) differences as functions of time t. The reaction function is $g = g_1$ (left) and $g = g_2$ (right).

Let $r \equiv r(x, y) := \sqrt{x^2 + y^2}$ for $(x, y) \in \Omega$. The above PDE is supplemented with the initial condition $u(0, x, y) = u_{\text{init}}(r(x, y))$ on $\overline{\Omega}$, where $u_{\text{init}}(r) := (r_0 - r)^2 \exp(-50 \cdot (R_0 - r)^2)$ for $r \in [r_0, R_0]$, and with consistent boundary conditions: zero-flux for $r(x, y) = r_0$ and the Dirichlet boundary condition $u = 1$ for $r(x, y) = R_0$. So we consider the case of a concentration u which initially has a peak equal to one on the outer perimeter and rapidly decays to zero towards the inner perimeter of Ω; in particular, the initial and boundary data depend on $r = r(x, y)$ only. As time proceeds, the concentration rises in the inside of the spatial domain.

We consider solutions u of (27.1) with circular symmetry; that is, there exists a function $\tilde{u}(t, r)$ for $t \geq 0$ and $r \in [r_0, R_0]$, such that $u(t, x, y) = \tilde{u}(t, r(x, y))$. A change of coordinates $(x, y) \to (r(x, y), \theta(x, y)), \theta \in [0, 2\pi)$, and using the circular symmetry of u transforms PDE (27.1) and its side conditions to an equivalent PDE problem for $\tilde{u}$, namely

$$\partial_t \tilde{u} = \partial_{rr} \tilde{u} + \frac{1}{r} \partial_r \tilde{u} + g(\tilde{u}), \ \tilde{u}(0, r) = u_{\text{init}}(r), \ \partial_r \tilde{u}(t, r_0) = 0, \ \tilde{u}(t, R_0) = 1.$$

The above reduction of (spatial) dimension takes the geometry of the PDE problem for $u(t, x, y)$ properly into account. A more simplistic approach, see Sec. 27.5, to "reduce" the dimension of the PDE problem for $u(t, x, y)$ is to consider the whole problem in 1D space right away, i.e., to consider for a function $v(t, r)$ with $t \geq 0$ and $r \in [r_0, R_0]$ the problem

$$\partial_t v = \partial_{rr} v + g(v), \ v(0, r) = u_{\text{init}}(r), \ \partial_r v(t, r_0) = 0, \ v(t, R_0) = 1.$$

The simplistic approach does not take the geometry of the PDE problem for u appropriately into account and significant differences can arise between the reduced solutions $\tilde{u}$ and v, in particular for $r \approx r_0 \ll 1$. This is illustrated in Fig. 27.1, where we have plotted the solutions $\tilde{u}(t, r)$ and $v(t, r)$ at $r = r_0$ and also their absolute, $v(t, r_0) - \tilde{u}(t, r_0)$, and relative, $(v(t, r_0) - \tilde{u}(t, r_0))/\max\{|v(t, r_0)|, |\tilde{u}(t, r_0)|\}$, differences. We observe substantial relative differences of more than 50%. We also see, in accordance with theoretical expectation, a faster increase of the properly reduced solution $\tilde{u}(t, r_0)$ on the inner perimeter compared to $v(t, r_0)$. The aim of this work is

to show how the solution $\tilde{u}$ can accurately be computed numerically as efficiently as when using the potentially inaccurate model leading to v.

27.3 Axi-Symmetric TDR Systems

Let $\Omega \subset \mathbb{R}^3$ be a bounded domain with elements denoted by $\mathbf{x}$ and $t \in J := (0, T] \subset \mathbb{R}$ the time and a finite time interval, respectively. We consider for the unknown function $\mathbf{u} : \overline{J} \times \overline{\Omega} \to \mathbb{R}^n$, representing, e.g., concentrations of chemicals, or cell or matrix densities, the *TDR system*

$$\partial_t u_i = -\nabla \cdot \left(\sum_{j=1}^{n} f_{ij}(\mathbf{u}) \nabla u_j \right) + g_i(t, \mathbf{x}, \mathbf{u}) \quad \text{for all} \quad \begin{array}{l} (t, \mathbf{x}) \in J \times \Omega, \\ i = 1, 2, \ldots, n, \end{array} \quad (27.2)$$

with transport functions $f_{ij} : \mathbb{R}^n \to \mathbb{R}$ and reaction terms $g_i : \overline{J} \times \overline{\Omega} \times \mathbb{R}^n \to \mathbb{R}$. We assume that $f_{ii}(\mathbf{u}) = -p_{ii}(\mathbf{u})$ where either $p_{ii}(\mathbf{u}) \equiv 0$ or $p_{ii}(\mathbf{u}) > 0$ for all possible values $\mathbf{u}$. This implies that the ith term in the sum either vanishes (no diffusion for u_i) or represents the quasi-linear diffusive flux $-p_{ii}(\mathbf{u})\nabla u_i$ for u_i. For the functions f_{ij} with $i \neq j$ we assume the functional form $f_{ij}(\mathbf{u}) = p_{ij}(\mathbf{u})u_i$ without any sign restriction on $p_{ij}(\mathbf{u})$ such that the jth term in the sum represents the taxis flux $u_i p_{ij}(\mathbf{u})\nabla u_j$ of u_i with respect to u_j. The PDE (27.2) is to be supplemented with initial conditions $\mathbf{u}(0, \mathbf{x}) = \mathbf{u_{init}}(\mathbf{x})$ for all $\mathbf{x} \in \overline{\Omega}$ and suitable boundary conditions. We assume that all data is such that the existence and uniqueness of a solution $\mathbf{u}$ is ensured. Furthermore, we assume that this solution is non-negative on $\overline{J} \times \overline{\Omega}$. This is natural since in our applications $\mathbf{u}$ represents a vector of concentration-like functions.

The diffusion and taxis terms in (27.2) are in divergence form and hence the TDR system satisfies a conservation of mass principle: for all sub-domains $V \subset \Omega$ and all $i = 1, \ldots, n$, the temporal change of the mass of u_i in V equals the mass flow of u_i due to diffusion/taxis through the boundary ∂V of V plus the mass of u_i created/destroyed within V due to sources, sinks or reactions. This principle is embodied by the *integral form of the conservation law* (27.2),

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_V u_i \mathrm{d}\mathbf{x} = -\int_{\partial V} \sum_{j=1}^{n} f_{ij}(\mathbf{u}) \nabla u_j \cdot \mathfrak{n}(\mathbf{x}) \mathrm{d}\mathbf{x} + \int_V g_i(t, \mathbf{x}, \mathbf{u}) \mathrm{d}\mathbf{x}. \quad (27.3)$$

If (27.3) holds for all $t \in \overline{J}$, $V \subset \Omega$ and $i = 1, \ldots, n$, then $\mathbf{u}$ is a solution of the integral form of the TDR system. The integral form arises first when modelling mass-conservative systems, and under suitable assumptions on Ω as well as sufficient smoothness of $\mathbf{u}$ and the parameter functions of the model, it can be transformed to the differential form (27.2). The form (27.3) of a TDR system is therefore more fundamental than (27.2) and is the basis for the finite volume spatial discretisation.

We now specialise the above definition of TDR systems. A TDR system having the following two properties is called an *axi-symmetric TDR system*.

1. Ω is axi-symmetric along the x-axis, that is,

$$\Omega = \{\mathbf{x} = \varphi(\tilde{\mathbf{x}}, \theta) := (x, r \sin\theta, r \cos\theta)^{\mathsf{T}} : \tilde{\mathbf{x}} \equiv (x, r)^{\mathsf{T}} \in \tilde{\Omega}, \ \theta \in [0, 2\pi)\},$$

   for a subset $\tilde{\Omega} \subset \mathbb{R} \times [0, +\infty)$. We refer to $\tilde{\Omega}$ as the *cross section* of Ω.

2. The reaction function g_i as well as the initial and boundary conditions are axi-symmetric. This means that, for example,

$$g_i(t, \mathbf{x}, \mathbf{u}) \equiv g_i(t, \varphi(\tilde{\mathbf{x}}, \theta), \mathbf{u}) = g_i(t, \varphi(\tilde{\mathbf{x}}, 0), \mathbf{u}) \quad \text{for all } \mathbf{x} \in \overline{\Omega},$$

   i.e., g_i is independent of the angle θ. To emphasise this axi-symmetry, we introduce the notation $g_i(t, \mathbf{x}, \mathbf{u}) \equiv \tilde{g}_i(t, \tilde{\mathbf{x}}, \mathbf{u})$ and $\mathbf{u_{init}}(\mathbf{x}) \equiv \tilde{\mathbf{u}}_{\text{init}}(\tilde{\mathbf{x}})$.

Consider an axi-symmetric TDR system with axi-symmetric solution, i.e., $\mathbf{u}(t, \mathbf{x}) \equiv \tilde{\mathbf{u}}(t, \tilde{\mathbf{x}})$. Let V be an arbitrary axi-symmetric subdomain of Ω with cross section $A \subset \tilde{\Omega}$ (i.e., V is a cylindrical shell). By employing the axi-symmetric nature of the problem, we reduce the volume and surface integrals in (27.3) to area and line integrals and arrive at the (equivalent) integral form of the conservation law of an axi-symmetric TDR system given by

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_A r\tilde{u}_i \mathrm{d}\tilde{\mathbf{x}} = -\int_{\partial A} \sum_{j=1}^{n} r f_{ij}(\tilde{\mathbf{u}}) \widetilde{\nabla}\tilde{u}_j \cdot \mathfrak{n}(\tilde{\mathbf{x}})\mathrm{d}\tilde{\mathbf{x}} + \int_A r\tilde{g}_i(t, \tilde{\mathbf{x}}, \tilde{\mathbf{u}})\mathrm{d}\tilde{\mathbf{x}}. \tag{27.4}$$

Here $\widetilde{\nabla}$ denotes the gradient with respect to $\tilde{\mathbf{x}} \equiv (x, r)^{\mathsf{T}}$. With the help of the integral theorem of Gauss we obtain, under additional smoothness assumptions and for the area $|A|$ of A tending to zero,

$$\partial_t(r\tilde{u}_i) = -\widetilde{\nabla} \cdot \left(\sum_{j=1}^{n} r f_{ij}(\tilde{\mathbf{u}}) \widetilde{\nabla}\tilde{u}_j \right) + r\tilde{g}_i(t, \tilde{\mathbf{x}}, \tilde{\mathbf{u}}), \tag{27.5}$$

i.e., the differential form of the conservation law corresponding to (27.4).

We emphasise that (27.4) and (27.5) are conservation laws in two spatial dimensions which fully capture the properties of axi-symmetric TDR systems (27.3) and (27.2) in three spatial dimensions. Our derivation of a suitable and efficient spatial discretisation of axi-symmetric TDR systems will be based on the properly reduced 2D problem (27.4).

27.4 A Finite Volume Spatial Discretisation

In this section we derive, on a partition of $\tilde{\Omega}$, a finite volume spatial discretisation of the conservation law (27.4). This discretisation defines the MOL-ODE, which, in a second step of the MOL, is integrated in time to yield the fully discrete approximation to the solution of the TDR system. The latter is not discussed here but we refer to [1,5] and the references cited there for appropriate numerical schemes.

We continue to consider cross sections $A \subset \tilde{\Omega}$ and associated cylindrical shells $V \subset \Omega$. Both V and A are referred to as *finite volumes*. The volume $|V|$ and area $|A|$ are related by $|V| = 2\pi \bar{r} |A|$, where $\bar{r} := |A|^{-1} \int_A r \mathrm{d}\tilde{\mathbf{x}} > 0$ is the centre of mass of A with respect to r. For the axi-symmetric solution $\mathbf{u}(t, \mathbf{x})$ we define the average in V at time t, $\mathbf{U}(t; V)$, and for the corresponding function $\tilde{\mathbf{u}}(t, \tilde{\mathbf{x}})$ the radius-weighted average $\tilde{\mathbf{U}}(t; A)$ at time t in A, i.e.,

$$\mathbf{U}(t; V) := \frac{1}{|V|} \int_V \mathbf{u}(t, \mathbf{x}) \mathrm{d}\mathbf{x} \quad \text{and} \quad \tilde{\mathbf{U}}(t; A) := \frac{1}{|A|} \int_A r \, \tilde{\mathbf{u}}(t, \tilde{\mathbf{x}}) \mathrm{d}\tilde{\mathbf{x}},$$

respectively. Both quantities are related by $\bar{r} \, \mathbf{U}(t; V) = \tilde{\mathbf{U}}(t; A)$. From Eq. (27.4) we obtain the *exact* evolution equation for the time-dependent average $\tilde{U}_i(t; A)$

$$\frac{\mathrm{d}}{\mathrm{d}t} \tilde{U}_i(t; A) = -\frac{1}{|A|} \int_{\partial A} \sum_{j=1}^{n} r f_{ij}(\tilde{\mathbf{u}}) \widetilde{\nabla} \tilde{u}_j \cdot \mathfrak{n}(\tilde{\mathbf{x}}) \mathrm{d}\tilde{\mathbf{x}} + \frac{1}{|A|} \int_A r \tilde{g}_i(t, \tilde{\mathbf{x}}, \tilde{\mathbf{u}}) \mathrm{d}\tilde{\mathbf{x}}. \quad (27.6)$$

Having a partition $\{A_{k'l'}\}$ of $\tilde{\Omega}$, for instance as defined below, the purpose of the spatial discretisation is to approximate the right-hand side of (27.6) by using the set $\{\tilde{\mathbf{U}}(t; A_{k'l'})\}$ and so to arrive at the MOL-ODE system

$$\frac{\mathrm{d}}{\mathrm{d}t} \tilde{U}_i(t; A_{kl}) = \tilde{H}_{kl}(t, \{\tilde{\mathbf{U}}(t; A_{k'l'})\}). \quad (27.7)$$

From now on we consider a partition of $\tilde{\Omega}$ with rectangular finite volumes $A_{kl} := (\bar{x}_{kl} - \Delta x, \bar{x}_{kl} + \Delta x) \times (\bar{r}_{kl} - \Delta r, \bar{r}_{kl} + \Delta r) \subset \tilde{\Omega}$ with constant grid widths $2\Delta x, 2\Delta r > 0$. The left, right, lower and upper neighbours of A_{kl} are denoted by $A_{k-1,l} \; A_{k+1,l}$, $A_{k,l-1}$ and $A_{k,l+1}$, respectively. For such a partition, the line integral (excluding the factor $-1/|A|$) in (27.6) decomposes into four parts corresponding to the edges of A_{kl}. We define the *average total flux* through the right and the upper edges of A_{kl} (out of A_{kl}) by

$$F_{kl}^{[x,i]}(t, \tilde{\mathbf{u}}(t, \cdot)) := \sum_{j=1}^{n} F_{kl}^{[x,i,j]}(t, \tilde{\mathbf{u}}(t, \cdot)), \quad F_{kl}^{[r,i]}(t, \tilde{\mathbf{u}}(t, \cdot)) := \sum_{j=1}^{n} F_{kl}^{[r,i,j]}(t, \tilde{\mathbf{u}}(t, \cdot)),$$

respectively, where the individual flux contributions are

$$F_{kl}^{[x,i,j]}(t, \tilde{\mathbf{u}}(t, \cdot)) := \frac{1}{2\Delta r} \int_{\bar{r}_{kl} - \Delta r}^{\bar{r}_{kl} + \Delta r} \left[r f_{ij}(\tilde{\mathbf{u}}(t, \tilde{\mathbf{x}})) \partial_x \tilde{u}_j(t, \tilde{\mathbf{x}}) \right]_{x = \bar{x}_{kl} + \Delta x} \mathrm{d}r,$$

and for $F_{kl}^{[r,i,j]}(t, \tilde{\mathbf{u}}(t, \cdot))$ accordingly. Then, by conservation of mass, the average total fluxes through the left and the lower edges of A_{kl} (out of A_{kl}) are $-F_{k-1,l}^{[x,i]}(t, \tilde{\mathbf{u}}(t, \cdot))$ and $-F_{k,l-1}^{[r,i]}(t, \tilde{\mathbf{u}}(t, \cdot))$, respectively. Hence, we arrive at

$$-\frac{1}{|A_{kl}|} \int_{\partial A_{kl}} \sum_{j=1}^{n} r f_{ij}(\tilde{\mathbf{u}}) \widetilde{\nabla} \tilde{u}_j \cdot \mathfrak{n}(\tilde{\mathbf{x}}) \mathrm{d}\tilde{\mathbf{x}} = -\frac{F_{kl}^{[x,i]} - F_{k-1,l}^{[x,i]}}{2\Delta x} - \frac{F_{kl}^{[r,i]} - F_{k,l-1}^{[r,i]}}{2\Delta r}.$$

This kind of discretisation of the transport term is called *conservative* as it clearly respects the conservation of mass principle.

In order to define a spatial discretisation of the right-hand side of (27.6) in terms of the average values $\{\tilde{U}_i(t; A_{k'l'})\}$, it remains to define approximations $\mathcal{F}_{kl}^{[x,i,j]}$ and $\mathcal{F}_{kl}^{[r,i,j]}$ for $F_{kl}^{[x,i,j]}$ and $F_{kl}^{[r,i,j]}$, respectively, in terms of these average values alone. Accordingly, the integral over A_{kl} involving $\tilde{g}_i$ must be approximated by an expression $\mathcal{G}_{kl}^{[i]}$ also using only these averages. The definition of these approximations is the subject of the remainder of this section and is guided by two aims. First, $\tilde{H}_{kl}(t, \{\tilde{\mathbf{U}}(t; A_{k'l'})\})$ should approximate the right-hand side of (27.6) to second order, in general, and second, the exact solution of the MOL-ODE (27.7) should be non-negative whenever its initial data is such. See [5–7] for conditions on the MOL-ODE ensuring this property.

Numerically we can only operate with the values $\tilde{\mathbf{U}}(t; A_{kl})$ in all finite volumes A_{kl} of our computational grid. Let $\tilde{U}_{i;kl}(t)$ denote the exact value or a numerical approximation of $\tilde{U}_i(t; A_{kl})$, and $\tilde{\mathbf{U}}_{kl}(t) := (\tilde{U}_{1;kl}(t), \dots, \tilde{U}_{n;kl}(t))$. From this data we construct a time- and space-dependent *piecewise constant reconstruction of $\tilde{u}_i$* on $\bar{J} \times \tilde{\Omega}$ defined by

$$(\mathcal{P}_0 \tilde{u}_i)(t, \tilde{\mathbf{x}}) := \bar{r}_{kl}^{-1} \tilde{U}_{i;kl}(t) \quad \text{for all} \quad \tilde{\mathbf{x}} \in A_{kl}.$$

The radius-weighted average over A_{kl} of this reconstruction equals $\tilde{U}_{i;kl}(t)$. Higher-order reconstructions are possible but not considered here.

In the following, we use the function $(\mathcal{P}_0 \tilde{u}_i)$ as a replacement for $\tilde{u}_i$ in order to define approximations of the volume integral in (27.6) and of the diffusive/taxis fluxes $F_{kl}^{[x,i,j]}(t, \tilde{\mathbf{u}}(t, \cdot))$ and $F_{kl}^{[r,i,j]}(t, \tilde{\mathbf{u}}(t, \cdot))$. We approximate the integrands, e.g., by using finite differences to replace derivatives, such that, after substituting $(\mathcal{P}_0 \tilde{u}_i)$ for $\tilde{u}_i$, the integrals can be evaluated exactly.

Approximation of the volume integral in (27.6). We approximate for $\tilde{\mathbf{x}} \in A_{kl}$

$$r\tilde{g}_i(t, \tilde{\mathbf{x}}, \tilde{\mathbf{u}}(t, \tilde{\mathbf{x}})) \approx r\tilde{g}_i(t, (\bar{x}_{kl}, \bar{r}_{kl}), (\mathcal{P}_0\tilde{\mathbf{u}})(t, \tilde{\mathbf{x}})) = r\tilde{g}_i(t, (\bar{x}_{kl}, \bar{r}_{kl}), \bar{r}_{kl}^{-1}\tilde{\mathbf{U}}_{kl}(t)),$$

and obtain

$$\frac{1}{|A_{kl}|} \int_{A_{kl}} r\tilde{g}_i(t, \tilde{\mathbf{x}}, \tilde{\mathbf{u}}) d\tilde{\mathbf{x}} \approx \mathcal{G}_{kl}^{[i]}(t) := \bar{r}_{kl}\tilde{g}_i(t, (\bar{x}_{kl}, \bar{r}_{kl}), \bar{r}_{kl}^{-1}\tilde{\mathbf{U}}_{kl}(t)).$$

Approximation of the diffusive fluxes $F_{kl}^{[x,i,i]}$ and $F_{kl}^{[r,i,i]}$. We first consider the flux in the x-direction, $F_{kl}^{[x,i,i]}(t, \tilde{\mathbf{u}}(t, \cdot))$, and approximate for $r \in [\bar{r}_{kl} - \Delta r, \bar{r}_{kl} + \Delta r]$ the terms in the integrand by

$$\left[f_{ii}(\tilde{\mathbf{u}}(t, x, r)) \right]_{x=\bar{x}_{kl}+\Delta x} \approx f_{ii}\left(\frac{\tilde{\mathbf{u}}(t, \bar{x}_{kl} + 2\Delta x, \bar{r}_{kl}) + \tilde{\mathbf{u}}(t, \bar{x}_{kl}, \bar{r}_{kl})}{2} \right),$$

$$\left[\partial_x \tilde{u}_i(t, x, r) \right]_{x=\bar{x}_{kl}+\Delta x} \approx \frac{\tilde{u}_i(t, \bar{x}_{kl} + 2\Delta x, \bar{r}_{kl}) - \tilde{u}_i(t, \bar{x}_{kl}, \bar{r}_{kl})}{2\Delta x}.$$

Now, using the piecewise constant reconstruction of $\tilde{\mathbf{u}}$, this leads to

$$F_{kl}^{[x,i,i]}(t, \tilde{\mathbf{u}}(t, \cdot)) \approx \mathcal{D}_{kl}^{[x,i]}(t) := f_{ii}\left(\frac{\tilde{\mathbf{U}}_{k+1,l}(t) + \tilde{\mathbf{U}}_{kl}(t)}{2\bar{r}_{kl}}\right) \cdot \frac{\tilde{U}_{i;k+1,l}(t) - \tilde{U}_{i;kl}(t)}{2\Delta x}.$$

Along the same lines we deal with the diffusive flux in the radial direction and arrive at the approximation $F_{kl}^{[r,i,i]}(t, \tilde{\mathbf{u}}(t, \cdot)) \approx \mathcal{D}_{kl}^{[r,i]}(t)$, where

$$\mathcal{D}_{kl}^{[r,i]}(t) := \frac{\bar{r}_{kl} + \Delta r}{2\Delta r} f_{ii}\left(\frac{\tilde{\mathbf{U}}_{k,l+1}(t)}{2(\bar{r}_{kl} + 2\Delta r)} + \frac{\tilde{\mathbf{U}}_{k,l}(t)}{2\bar{r}_{kl}}\right) \cdot \left(\frac{\tilde{U}_{i;k,l+1}(t)}{\bar{r}_{kl} + 2\Delta r} - \frac{\tilde{U}_{i;kl}(t)}{\bar{r}_{kl}}\right).$$

Approximation of the taxis fluxes $F_{kl}^{[x,i,j]}$ and $F_{kl}^{[r,i,j]}$, $i \neq j$. Here we will only discuss the approximation of the radial fluxes $F_{kl}^{[r,i,j]}$; approximations for the axial fluxes $F_{kl}^{[x,i,j]}$ follow accordingly. Remember that we have $f_{ij}(\tilde{\mathbf{u}}) = p_{ij}(\tilde{\mathbf{u}})\tilde{u}_i$ for $i \neq j$. We exploit this structure in the following and consider, for each i, all fluxes $F_{kl}^{[r,i,j]}$ with $j \neq i$ simultaneously. That is, for a fixed value i, we consider the sum of fluxes

$$\sum_{j \neq i} F_{kl}^{[r,i,j]}(t, \tilde{\mathbf{u}}(t, \cdot)) = \frac{1}{2\Delta x} \int_{\bar{x}_{kl} - \Delta x}^{\bar{x}_{kl} + \Delta x} \left[r \, \tilde{u}_i(t, \tilde{\mathbf{x}}) \cdot \sum_{j \neq i} p_{ij}(\tilde{\mathbf{u}}) \, \partial_r \tilde{u}_j(t, \tilde{\mathbf{x}}) \right]_{r = \bar{r}_{kl} + \Delta r} dx.$$

We approximate, as in the diffusive case, the terms $p_{ij}(\tilde{\mathbf{u}}(t, x, r))$ and $\partial_r \tilde{u}_j(t, x, r)$ at $r = \bar{r}_{kl} + \Delta r$ for $x \in [\bar{x}_{kl} - \Delta x, \bar{x}_{kl} + \Delta x]$ and then plug in the piecewise constant reconstruction of $\tilde{\mathbf{u}}$. This leads to the following definition of the local transport velocity (in the radial direction) of component i due to taxis:

$$v_{kl}^{[r,i]}(t) := \frac{1}{2\Delta r} \sum_{j \neq i} p_{ij}\left(\frac{\tilde{\mathbf{U}}_{k,l+1}(t)}{2(\bar{r}_{kl} + 2\Delta r)} + \frac{\tilde{\mathbf{U}}_{kl}(t)}{2\bar{r}_{kl}}\right) \cdot \left(\frac{\tilde{U}_{j;k,l+1}(t)}{\bar{r}_{kl} + 2\Delta r} - \frac{\tilde{U}_{j;kl}(t)}{\bar{r}_{kl}}\right).$$

Depending on the sign of the velocity $v_{kl}^{[r,i]}$, we use either $\mathcal{S}_{kl}^{[r,i,+]}(t)$ or $\mathcal{S}_{kl}^{[r,i,-]}(t)$, defined below, as an approximation of $\tilde{u}_i(t, x, \bar{r}_{kl} + \Delta r)$ for $x \in [\bar{x}_{kl} - \Delta x, \bar{x}_{kl} + \Delta x]$. This *upwinding* technique ensures that information for the approximation is mainly taken from the upstream direction of the flow. Furthermore, in order to ensure the non-negativity of the resulting MOL-ODE solution, we must employ *limiting* techniques in the definition of these approximations (this is not necessary in the case of the diffusive fluxes discussed above). For a thorough discussion of flux limiting we refer to [5,8]. All approximated terms in the integrand are now independent of x and evaluation of the integral leads to $\sum_{j \neq i} F_{kl}^{[r,i,j]}(t, \tilde{\mathbf{u}}(t, \cdot)) \approx \mathcal{T}_{kl}^{[r,i]}(t)$, where

$$\mathcal{T}_{kl}^{[r,i]}(t) := (\bar{r}_{kl} + \Delta r) \cdot \left(\max\{0, v_{kl}^{[r,i]}\}\mathcal{S}_{kl}^{[r,i,+]}(t) + \min\{0, v_{kl}^{[r,i]}\}\mathcal{S}_{kl}^{[r,i,-]}(t)\right).$$

In order to define the approximations $\mathcal{S}_{kl}^{[r,i,+]}(t)$ and $\mathcal{S}_{kl}^{[r,i,-]}(t)$ we employ a *smoothness monitor function*, denoted by $\theta_{kl}^{[r,i]}(t)$, of the solution $\tilde{u}_i$, given by

$$\theta_{kl}^{[r,i]}(t) := \left(\frac{\tilde{U}_{i;k,l+1}(t)}{\bar{r}_{kl} + 2\Delta r} - \frac{\tilde{U}_{i;kl}(t)}{\bar{r}_{kl}} \right) \Big/ \left(\frac{\tilde{U}_{i;kl}(t)}{\bar{r}_{kl}} - \frac{\tilde{U}_{i;k,l-1}(t)}{\bar{r}_{kl} - 2\Delta r} \right)$$

and a (Lipschitz continuous) *limiter function* $\Phi : \mathbb{R} \to \mathbb{R}$ depending on $\theta_{kl}^{[r,i]}$. An example is the van Leer limiter $\Phi(\theta) = (\theta + |\theta|)/(1 + |\theta|)$. We then define

$$\mathcal{S}_{kl}^{[r,i,+]}(t) := \frac{\tilde{U}_{i;kl}(t)}{\bar{r}_{kl}} + \frac{1}{2}\Phi\left(\theta_{kl}^{[r,i]}(t)\right) \left(\frac{\tilde{U}_{i;kl}(t)}{\bar{r}_{kl}} - \frac{\tilde{U}_{i;k,l-1}(t)}{\bar{r}_{kl} - 2\Delta r} \right),$$

$$\mathcal{S}_{kl}^{[r,i,-]}(t) := \frac{\tilde{U}_{i;k,l+1}(t)}{\bar{r}_{kl} + 2\Delta r} + \frac{1}{2}\Phi\left((\theta_{k,l+1}^{[r,i]}(t))^{-1}\right) \left(\frac{\tilde{U}_{i;k,l+1}(t)}{\bar{r}_{kl} + 2\Delta r} - \frac{\tilde{U}_{i;k,l+2}(t)}{\bar{r}_{kl} + 4\Delta r} \right).$$

This completes the definition of $\tilde{H}_{kl}(t, \{\tilde{\mathbf{U}}(t; A_{k'l'})\})$ in (27.7) in finite volumes A_{kl} away from the boundary of $\tilde{\Omega}$. We do not discuss here the special approximations required in volumes close to $\partial\tilde{\Omega}$ (including the incorporation of boundary conditions) but instead refer to [1]. The MOL-ODE (27.7) can be simplified somewhat by dividing each equation by the appropriate value $\bar{r}_{kl}$. This results in a MOL-ODE for the averages $\{\mathbf{U}(t; V_{k'l'})\}$ instead of the radius-weighted averages $\{\tilde{\mathbf{U}}(t; A_{k'l'})\}$ and as such is more user friendly.

27.5 Application to a Model of Fracture Healing in Mice

Our interest in numerical methods for axi-symmetric TDR problems caused by their application to models of fracture healing of long bones. Considering the geometry of the problem we and, e.g., the authors of [9] regard an axi-symmetric model of the biological processes as a good starting point. In this study, the mathematical model of [9] was applied to our experimental set-up of a semi-stabilised tibial fracture in mice. Certain parameters of the mathematical model were adapted to obtain a good comparison between the numerically predicted and the experimentally observed results; for details we refer to [10].

Experimental model. A transverse fracture was made in the proximal tibia of 11-week-old mice (Fig. 27.2). The fracture was semi-stabilised with a thin-walled needle (0.4 mm) that spanned the bone longitudinally throughout the marrow cavity. Mice were sacrificed at different time points after fracture induction (post fracture day (PFD) 3, 8, 13, 21) and prepared for histological examination.

Mathematical model. The mathematical model describes the actions and interactions of mesenchymal stem cells, osteoblasts and chondrocytes, of which the differentiation and anabolic activity (matrix synthesis) are regulated by chondrogenic and osteogenic growth factors. "Ideal" mechanical and vascular conditions are assumed. Changes in cell, growth factor and extracellular matrix (ECM) densities/concentrations are captured in a system of seven highly coupled nonlinear differential equations, forming a taxis-diffusion-reaction (TDR) system:
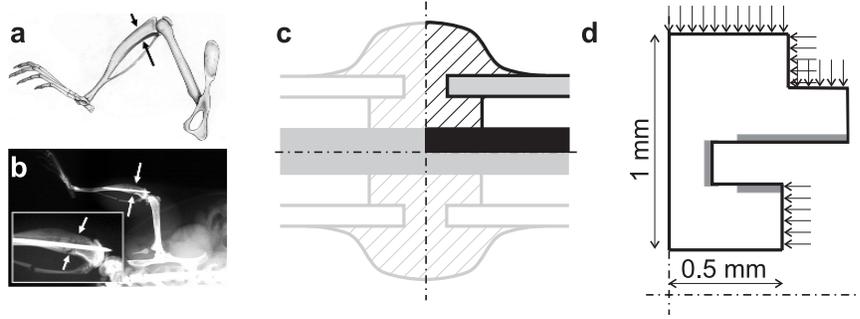
Fig. 27.2. Models of a semi-stabilised fracture in the tibia of 11-week-old mice. The arrows in (a) indicate the fracture site, (b) illustrates the stabilisation of the fracture by a needle. (c) is a generic model of the fracture site (black: needle, grey: cortex, hatched: callus), and (d) is the simplified generic model used in the simulations (arrows: mesenchymal cell source, shaded: growth factor source).

$$\frac{\partial c_m}{\partial t} = -\nabla \cdot \left[-D_{cm}(\mathbf{c})\nabla c_m + c_m \sum_{i=1}^{6} f_i(\mathbf{c})\nabla c_i \right] + f_0(c_m, \mathbf{c}),$$

$$\frac{\partial \mathbf{c}}{\partial t} = D\Delta\mathbf{c} + \mathbf{g}(c_m, \mathbf{c}).$$

(27.8)

This system is an example of the general TDR system (27.2) with $\mathbf{u} := (c_m, \mathbf{c})$. Here, $c_m(t, \mathbf{x})$ is the non-dimensional density of the mesenchymal stem cells and $\mathbf{c}(t, \mathbf{x})$ represents a vector of six non-dimensional concentrations or densities of chondrocytes, osteoblasts, ECM and growth factors, respectively. $D_{cm}(\mathbf{c})$ and D (non-negative diagonal matrix) represent the diffusion coefficients, $f_i(\mathbf{c})$ the taxis coefficients and $f_0(c_m, \mathbf{c})$ and $\mathbf{g}(c_m, \mathbf{c})$ are the reaction terms. The system (27.8) must be complemented by suitable initial and boundary data to ensure the existence, uniqueness and non-negativity of a solution $(c_m, \mathbf{c})$. This data as well as the domain Ω are chosen such that we obtain an axi-symmetric TDR system; see Fig. 27.2 for a sketch of the cross section $\tilde{\Omega}$ and the applied boundary conditions. The mathematical model considered in this work is precisely the same as that studied in [10].

Results. The amounts of cartilage and bone in the callus at different stages of the healing process in the animal experimental model were quantified from Safranin O and Von Kossa stained sections, respectively (Fig. 27.3). Cartilage levels peak around PFD 8 while bone levels keep rising throughout the healing process, until the entire callus is ossified at about PFD 21. These results correspond to the results obtained from the axi-symmetric mathematical model using the numerical procedure described in Sec. 27.4. At PFD 4, no mesenchymal cells remain in the callus, while at the periosteal surface intramembranous ossification starts. Chondrogenic differentiation proceeding from the peripheral callus is followed by cartilage formation visible from PFD 5 and building up to a maximum amount of cartilage in the callus at PFD 10, after which
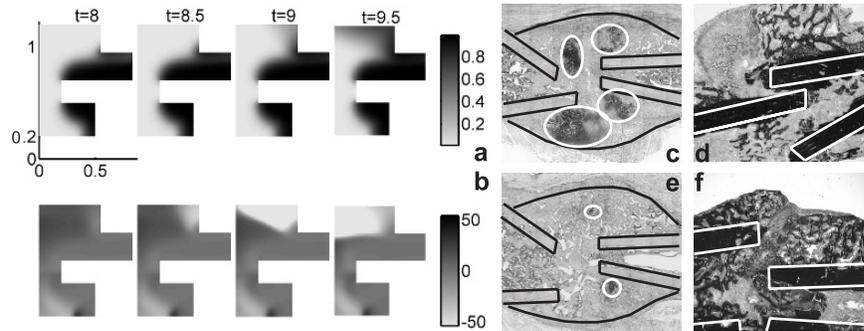
Fig. 27.3. Results showing the predicted density of bone [$\times 0.1$ g ml^{-1}] in the callus at different time points after fracture induction, using the properly reduced axi-symmetric TDR system (a). The relative difference [in %] from the solution using the simplistic "reduction" is represented in (b). The maximal absolute difference is of the order of magnitude of 10^{-3} g ml^{-1}. Plots (c) through (f) show the cartilage and bone content of the callus at PFD 8 (c,e) and 13 (d,f). The outer edges of the callus area and the old cortical bone are delineated, the cartilaginous areas are encircled and new bone is distinguishable as the spongy black tissue.

endochondral ossification starts. Complete ossification of the callus is reached three weeks after fracture induction.

Comparing the numerical solutions as obtained (i) from the properly reduced axi-symmetric TDR system (as described in this paper) and (ii) from a simplistic "reduction" of the system to 2D space as, e.g., used in [9, 10], we observe small transient differences in the time evolution of the solutions. An example of these differences is given in Fig. 27.3 (a,b), showing that the front of growing bone has different speeds in the two cases. The relative difference between both solutions is up to about 50%. No difference is visible between the result at final time, as the callus eventually completely ossifies.

27.6 Conclusion

In this work we have considered the MOL numerical solution of TDR systems with an axi-symmetric structure. We presented a finite volume spatial discretisation (first step of the MOL) of the problem exploiting that special structure. Its most important properties are:

1. Proper reduction of axi-symmetric TDR problems to equivalent 2D problems on the cross section of the axi-symmetric domain.
2. Second-order accuracy (in general) allowing for coarser grids than first-order discretisations would require and hence more favourable simulation times.
3. Ensuring of the conservation of mass property and the non-negativity of the exact solution in the numerical scheme.

This methodology is applied successfully in the simulation of an axi-symmetric TDR model of fracture healing in mice; see [10] for details on the model. The numerical simulation results obtained are in good agreement with available experimental data; see also [10].

Instead of, as above, using a proper reduction of an axi-symmetric TDR problem to an equivalent 2D problem, one could also model the processes simply as happening on the 2D cross section of the axi-symmetric domain only. The costs of the numerical simulation of both problems are about the same. However, the simplistic "reduction" neglects the 3D nature of the original problem and potentially leads to wrong simulation results. An instructive and biologically relevant example has been used to demonstrate this danger. We and others have used the simplistic "reduction" of axi-symmetric TDR systems in earlier work on fracture healing [9,10]. For the axi-symmetric model presented here and in [10], we have examined in this work the differences in the numerical solutions as obtained from (i) the properly reduced TDR model (see above) and (ii) the simplistic "reduction." We have observed subtle differences (e.g., slightly different wave speeds). However, these differences do not lead to conclusions different from those drawn in [10].

Based on the work presented, we conclude that the simplistic "reduction" of axi-symmetric TDR systems can lead to significant errors (despite its success for certain TDR models). Therefore, to avoid these potential errors, we advocate the use of the spatial discretisation described in this chapter, corresponding to a proper reduction of axi-symmetric TDR systems.

Acknowledgment

References

1. Gerisch, A.: Numerical Methods for the Simulation of Taxis-Diffusion-Reaction Systems. Ph.D. thesis, Martin-Luther Universität Halle-Wittenberg (2001).
2. Gerisch, A., Chaplain, M. A. J.: Robust numerical methods for taxis-diffusion-reaction systems: Applications to biomedical problems. Math. Comput. Modelling, **43**, 49–75 (2006).
3. Gerisch, A., Geris, L., Van Oosterwyck, H., Vander Sloten, J., Weiner, R.: Reliable and efficient numerical simulation of a model of tissue differentiation in a bone chamber. In: Bischoff, G., Hein, H.J. (eds) Micro- and Nanostructures of Biological Systems, 60–76. Shaker Verlag, 3rd edition, Aachen, Germany, (2005).
4. Kawahara, T., Tanaka, M.: Interactions of traveling fronts: An exact solution of a nonlinear diffusion equation. Phys. Lett. A, **97**, 311–314 (1983).
5. Hundsdorfer, W., Verwer, J. G.: Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations, volume 33 of Springer Series in Computational Mathematics. Springer-Verlag Berlin, Heidelberg, New York (2003).
6. Horváth, Z.: Positivity of Runge–Kutta and diagonally split Runge–Kutta methods. Appl. Numer. Math., **28**, 309–326 (1998).

 7. Hundsdorfer, W., Koren, B., van Loon, M., Verwer, J. G.: A positive finite-difference advection scheme. J. Comput. Phys., **117**, 35–46 (1995).
 8. LeVeque, R. J.: Finite volume methods for hyperbolic problems. Cambridge University Press, London, New York, (2002).
 9. Bailón-Plaza, A., van der Meulen, M. C. H.: A mathematical framework to study the effects of growth factor influences on fracture healing. J. Theor. Biol., **212**, 191–209 (2001).
10. Geris, L., Gerisch, A., Maes, Ch., Carmeliet, G., Weiner, R., Vander Sloten, J., Van Oosterwyck, H.: Mathematical modeling of fracture healing in mice: comparison between experimental data and numerical simulation results. Med. Biol. Eng. Comput., **44**, 280–289 (2006).

28

Information Content Toward a Neonatal Disease Severity Score System

Giulia Menconi,[1] Marco Franciosi,[2] Claudio Bonanno,[2] and Jacopo Bellazzini[2]

[1] Dipartimento di Matematica, Università di Bologna, Italy; Corresponding author: Giulia Menconi, menconi@mail.dm.unipi.it
[2] Dipartimento di Matematica Applicata, Università di Pisa, Italy

Summary. We introduce a score to classify the severity of patients by analysing the information content of clinical time series.

Key words: Information content, time series, neonatal illness severity.

28.1 Introduction

The assessment of neonatal illness severity is a crucial issue in the neonatal intensive care unit. Particular attention has to be paid to developing non-invasive methodologies for data analysis that should give an illness severity score in real time. Ideally, that score should reflect the current health status and suggest signs of incoming crises or aggravation. In the last decades of the twentieth century, several methods from nonlinear dynamics have been proposed to answer such a clinical query. One of the most significant is based on a measure of the information content in a time series. In an experimental setting, information content may be approximated by means of compression algorithms [1]. This chapter aims at showing an application of this kind of approach in order to analyse time series resulting from clinical data. To be more precise, starting from a set of time series related to a patient, the information content of each series is calculated and a score is defined for the patient. To evaluate the agreement between that score and the actual patient severity, we analysed a preliminary set of blind data collected by the Neonatal Intensive Care Unit, Siena University Hospital, Siena, Italy (Claudio De Felice, M.D.). Our methods produced good results, pertinent with the health status of the patients and in agreement with other totally different approaches of nonlinear time series analysis (for instance, see [5]). Our results are summarised as a *severity score* (see Fig. 28.3).

28.2 Materials and Methods

28.2.1 Information Content of Time Series

One of the most significant tools from the modern theory of nonlinear dynamics used to analyse time series of biological origin is related to the notion of *information content* of a finite word as introduced by Shannon (see [6]). The intuitive notion of information content of a finite word can be stated as "the length of the shortest message from which it is possible to reconstruct the original word." A formal mathematical definition of this notion has been introduced by Kolmogorov using the notion of the universal Turing machine (see again [6]). We will not enter into the details of the mathematical definition, but simply use the intuitive notion of information content that we stated.

The method we use to study the information content of a finite word is related to *compression algorithms*. These are a well-known tool present in every personal computer, used to store files in the most economic way from the point of view of space needed in the memory storage disks. The compression of a finite word reflects the intuitive meaning of the information content of the word.

Let $s = (s_1 s_2 \ldots s_n)$ be an n-long word written in the finite alphabet $\mathcal{A}$, that is, $s_i \in \mathcal{A}$ for all $i = 1, \ldots, n$. We will denote by $\mathcal{A}^n$ the set of all n-long words written using $\mathcal{A}$, and we denote $\mathcal{A}^* := \cup_n \mathcal{A}^n$. A *compression algorithm* can be defined as an injective function $Z : \mathcal{A}^* \rightarrow \{0, 1\}^*$, that is, a binary coding of the finite words written using $\mathcal{A}$. By using the algorithm Z we define the information content of a word s as the binary length of the compressed version of s, that is, $Z(s)$. Hence,

$$I(s) := \text{ Information Content of } s = |Z(s)|.$$

The notion of information content of a finite word can also be used for the problem of giving a notion of randomness of a word. Namely, we can think of a word as being more random the less efficient its compression achieved by a compression algorithm. This argument leads to the notion of *complexity $C(s)$* of a finite word, defined as the *compression ratio*, that is,

$$C(s) := \frac{I(s)}{|s|} = \frac{|Z(s)|}{|s|}.$$

The greater the complexity of a word, the higher the randomness of the word.

Let us consider now an application of these tools to time series. By a time series we mean a finite set $X = \{X_1, \ldots, X_N\}$ of data, where each data is an array of real numbers. The first step in the analysis is the reduction of a time series to a finite word. This is accomplished by a partition of the set of possible values of the data. Let $\mathcal{P}$ be such a partition into sets $\{I_1, \ldots, I_L\}$; then to the time series X we associate the word $s \in \mathcal{A}^N$, where $\mathcal{A} = \{1, \ldots, L\}$, with the rule that for all $j = 1, \ldots, N$, we choose $s_j = \ell \in \mathcal{A}$ if $X_j \in I_\ell$. We can then define the complexity of the time series X as

$$C(X, L) := C(s),$$

where the notation $C(X, L)$ points out the role of the number of symbols of the alphabet used to write the word s. In the following we will consider the effects of varying L.

When analysing a time series it is convenient to consider asymptotic properties, hence assuming an infinite time series. We can make this assumption to obtain some mathematical results on the complexity of a time series. For an infinite time series $X = \{X_i\}$ we can define the *asymptotic complexity*

$$K(X, L) := \lim_{N \to \infty} C((X_1, \dots, X_N), L)$$

with respect to an L-long alphabet.

If the time series X is the orbit of a dynamical system, then $K(X, L) \to h$ as $L \to \infty$, where h is the metric entropy of the system (see references in [1]). In fact, there is an integer L_0 such that $K(X, L) = h$ if $L \geq L_0$.

A particular case is obtained when the series is generated by a white noise on the unit interval $[0, 1]$. Let ξ be such a series. Then $K(\xi, L) \to \infty$ as $L \to \infty$, and the asymptotic behaviour is $K(\xi, L) \sim \log L$ (for two sequences (a_n) and (b_n) of real numbers, we say that $a_n \sim b_n$ asymptotically if $\lim_n (a_n/b_n) = 1$) (cf. [2]).

In [2] the case of a time series $(X + \xi)$ obtained as a random perturbation of a dynamical system was also studied. The results show that for big L the behaviour of $K(X + \xi, L)$ is analogous to that of $K(\xi, L)$, whereas for small enough values of L it is possible to achieve $K(X + \xi, L) = h$ where h is the metric entropy of the unperturbed dynamical system. In particular, it is shown that it is important to compare the size of the intervals of the partition used to obtain the symbolic word and the size of the noise.

The analysis we present in this chapter is performed with very short time series, and often the range of different values in a single series is very small. These features imply two drawbacks: (i) we are very far from obtaining an asymptotic behaviour, hence we are looking at the transients; (ii) we cannot use alphabets with many symbols, hence the limit in L can only be a rough approximation.

However, we can still analyse the series by comparing the results with those for white noise. It is reasonable to treat short time series as obtained by noisy systems. Moreover, we made partitions of the set of possible values of the data up to sets given by single values, that is using the best "resolution." Hence this analysis is comparable with the case of time series obtained as a random perturbation of a dynamical system with big values for L. For these reasons we decided to study the logarithmic behaviour of the values $K(X, L)$, and thus to make a linear interpolation of the set $\{(L, K(X, L))\}$ when plotted on a logarithmic linear scale. This gives a value $q(X)$ satisfying $K(X, L) \approx q(X) \log L$. For white noise we would have $q(X) = 1$, hence our results give values $q(X) \in (0, 1)$.

28.2.2 Experimental Data Sets

We applied our method to 23 triplets of time series related to 12 newborns admitted to the Neonatal Intensive Care Unit (NICU) of Azienda Ospedaliera Universitaria Senese, Policlinico "Le Scotte," Siena, Italy (Claudio De Felice, M.D.). The mean gestational

age was 32.5 ± 4.7 weeks (range: 23–38 weeks) while the mean birth weight was 1775 ± 766 g (range: 580–3380 g).

For each patient we considered a triplet concerning the following pulse oximetry-derived signals:

- perfusion index (PI)
- pulse rate (PR)
- oxygen saturation (SpO_2).

Pulse oximetry is a real-time, noninvasive technique to measure arterial oxygen saturation and is routinely used in the monitoring of critically ill patients and during surgical procedures. The principle of pulse oximetry is based on two light sources with different wavelengths (660 and 940 nm) emitted through the cutaneous vascular bed, usually of a finger or earlobe. As the deoxygenated hemoglobin (Hb) absorbs more light at 660 nm (red light section of the visible electromagnetic spectrum) while its oxygenated form (HbO_2) absorbs more light at 940 nm (infrared), the ratio between the red to infrared light absorbed gives the tissue oxygen saturation (SpO_2). A third light source at 800 nm provides an estimate of the overall hemoglobin concentration and can be used to determine the variations in arterial blood volume (pulsatile component) and calculate the pulse rate (PR). In addition, the ratio between the pulsatile component (originating from the arterial component) to the nonpulsatile component (originating from other tissues, including the connective tissue, venous blood, and bone) gives the perfusion index (PI). As an abnormality in peripheral perfusion is associated with a variation in the pulsatile component only, the value of this ratio reflects changes in peripheral perfusion (see [7]).

These data sets were collected with Radical SET pulse oximeter, Masimo Co., Irvine, CA, USA, probe placed at either foot, according to the methodology illustrated in [3] and [4]. They were registered every 4 seconds and the monitoring was carried on over a time from 1 to 1.5 hours. The perfusion index is a percentage, while pulse rate and saturation are integers. In order to reduce the time series into symbolic strings, we have considered uniform partitions $\mathcal{P}_L$ on the set of possible values $I = [\text{min val}, \text{max val}]$ made of L subintervals. For each time series, we have considered partitions where L ranges from 2 to 20 subintervals. We have used the compression algorithm called *CASToRe* developed by our group (see [1] for details). In Fig. 28.1 we give three different examples of triplets relative to different newborns.

When applying our analysis we were completely unaware of the health status of each patient.

28.3 Results

In Fig. 28.2 the behaviour of the complexity $K(X, L)$ with respect to the graining size L is shown for the whole data set #01. The plot is in a log-linear scale and the coefficient $q(X)$ is the slope of the growth (severity coefficient), for perfusion index, pulse rate, and saturation.
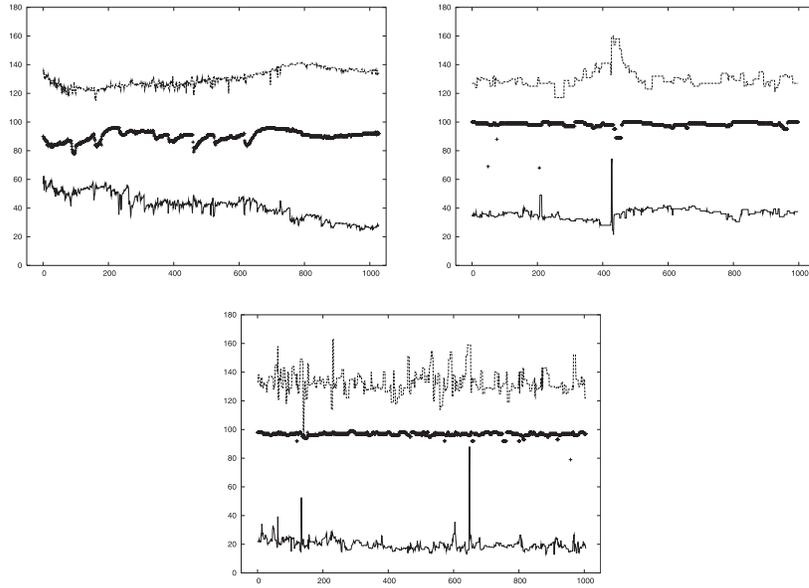
Fig. 28.1. Data sets for three different newborns: #01 (top left), #02 (top right), #08 (bottom). For each plot: PI series is at the bottom (rescaled, to be more readable), SpO$_2$ is the middle series, and PR is at the top.

28.3.1 Severity Score

For each patient, three coefficients q have been calculated. We denote by $\mathcal{S}(N) = (q(PI), q(PR), q(SpO_2))$ the severity array of data set $\#N$.

In order to classify the severity of each patient via their severity array, we have defined the severity score as the ℓ_1-norm of the array:
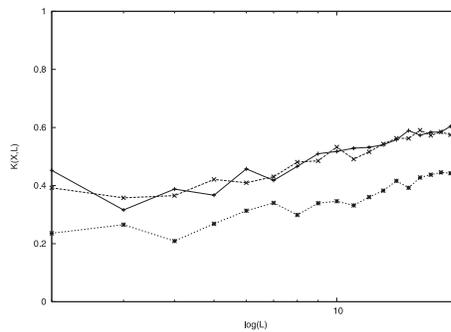
Fig. 28.2. Complexities $K(X, L)$ for data set #01. Perfusion index: top solid line; pulse rate: middle crossed-dashed line; saturation: bottom star-dotted line. The graining size L ranges from 2 to 20. The plot is log-linear.
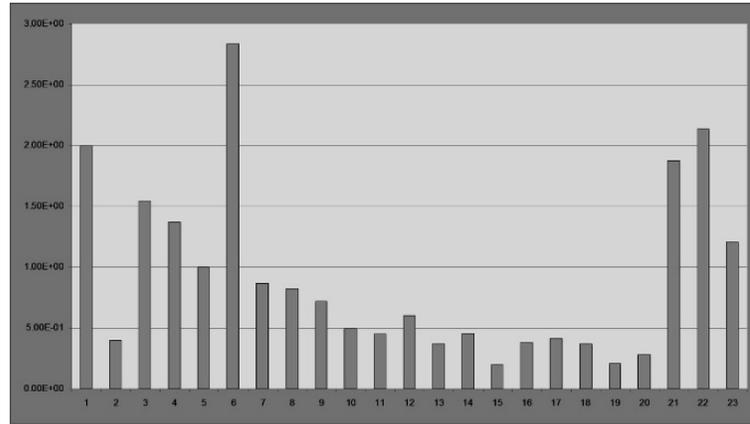
Fig. 28.3. Severity score for the 23 data sets.

$$||\mathcal{S}(N)|| = q(\text{PI}) + q(\text{PR}) + q(\text{SpO}_2). \qquad (28.1)$$

Fig. 28.3 shows how the score varies for each data set: as a matter of fact, the results show a high variability. The maximum is reached by data set #6, while the minimum is for data set #19. Two extremal groups of data sets are immediately identified: $\mathbf{A} = \{\#01, \#06, \#21, \#22\}$ have very high scores, while $\mathbf{B} = \{\#02, \#10, \#11, \#12, \ldots, \#20\}$ have a very poor score. The remaining data sets, with intermediate scores, are grouped in $\mathbf{C} = \{\#03, \#04, \#05, \#07, \#08, \#09, \#23\}$. It is a first hint for distinguishing the group of healthier patients from less healthy patients. It is well known that life parameters should show a high irregularity to be more adaptive. Therefore, it is reasonable to associate a higher severity to data sets in $\mathbf{B}$ whose severity score is lower and a lower severity to $\mathbf{A}$. A finer investigation is necessary to characterise the intermediate scores for the group of data sets $\mathbf{C}$.

A complementary useful tool is the *normalised severity array:*

$$\mathcal{S}^*(N) = \left(\frac{q(\text{PI}) - \min(\text{PI})}{\max(\text{PI}) - \min(\text{PI})}, \frac{q(\text{PR}) - \min(\text{PR})}{\max(\text{PR}) - \min(\text{PR})}, \frac{q(\text{SpO}_2) - \min(\text{SpO}_2)}{\max(\text{SpO}_2) - \min(\text{SpO}_2)} \right).$$

Normalised severity arrays belong to $[0, 1] \times [0, 1] \times [0, 1]$. For each $N = 1, \ldots, 23$ we have represented $\mathcal{S}^*(N)$ on a two-dimensional symplex, by drawing a ball with radius proportional to the severity score (Fig. 28.4).

28.4 Final Discussion

The *severity score method* (SSM) introduced here allows an efficient and user-friendly approach to early monitoring of a patient's health severity. Concerning the three groups identified in Section 28.3.1, their clinical status was the following, in agreement with the experimental classification:
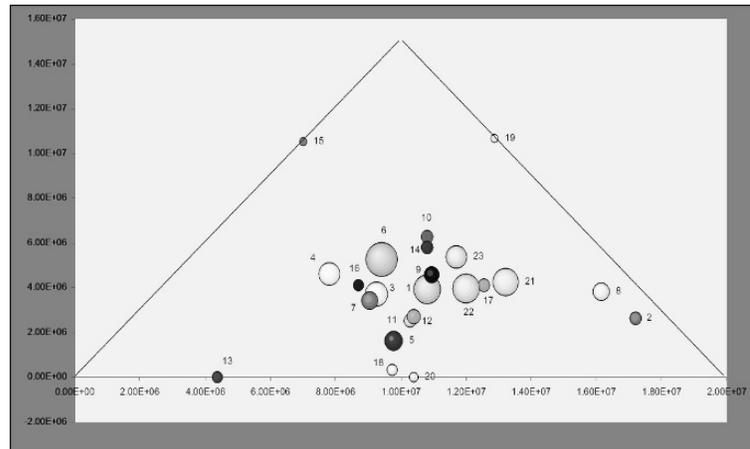
Fig. 28.4. Severity array: simplex representation.

A = {#01, #06, #21, #22}: very low clinical severity.

B = {#02, #10, #11, #12, . . . , #20}: high clinical severity.

C = {#03, #04, #05, #07, #08, #09, #23}: all but #09 were cases of histologic chorioamnionitis.

A few more details may be inferred from the severity array (Fig. 28.4). Patients #02 and #08 are isolated from the others: they had severe crises some hours after the end of the measurements. In particular, patient #08 shows an intermediate score, but an anomalous correlation among the three clinical parameters is highlighted by the simplex representation. Finally, the data sets from #09 through #20 are relative to the same patient at different times. They were clustered almost all together in group **B**. To sum up, this method gives interesting hints on a real-time assessment of disease severity in a high-risk newborn population and would allow a reliable prediction of possible disease aggravation episodes in order to institute a proper treatment. Also, it encourages interdisciplinary collaboration to develop some diagnostic protocols to help the medical team in deciding whether to take care of the patient in an intensive care unit. Finally, a more correct assessment of disease severity should ideally lead to a significant reduction in the total number of repeated invasive tests, including blood sampling. This would reduce both blood loss and experience of pain in this particular patient population. This has to be considered as one of the first experiments of the *ATTIS project*, started in 2004 with the aim of developing and consolidating different methods in time series analysis (ATTIS is the acronym for Approaches To TIme Series, see http://www.attis-project.org).

References

1. Benci, V., Bonanno, C., Galatolo, S., Menconi, G., Virgilio, M.: Dynamical systems and computable information. Disc. Cont. Dyn. Sys.-B, **4**, 935–960 (2004).

2. Bonanno, C.: The algorithmic information content for randomly perturbed systems. Dis. Cont. Dyn. Syst.-B, **4**, 921–934 (2004).

3. De Felice, C., Latini, G., Vacca, P., Kopotic, R.J.: The pulse oximeter perfusion index as a predictor for high illness severity in neonates. Eur. J. Pediatr., **161**, 561–562 (2002).

4. De Felice, C., Del Vecchio, A., Criscuolo, M., Lozupone, A., Parrini, S., Latini, G.: Early postnatal perfusion index changes in term newborns with subclinical chorioamnionitis. Arch. Dis. Child Fetal Neonatal Ed., **90**, 5 (2005).

5. De Felice, C., Bianciardi, G., Parrini, S., Verrotti, A., Goldstein, M., Latini, G.: Pulse oximetry signals: chaotic analysis in the evaluation of neonatal illness severity. Pediatric Res., **56**, 475 (2004).

6. Li, M., Vitanyi, P.: An introduction to Kolmogorov complexity and its applications. Springer-Verlag, Berlin (1993).

7. Lima, A., Bakker, J.: Noninvasive monitoring of peripheral perfusion. Intensive Care Med., **31**, 1316–1326 (2005).

Part V

Data Analysis and Model Validation

Statistical Analysis and Physical Modelling of Oligonucleotide Microarrays

Conrad J. Burden,[1,2] Yvonne E. Pittelkow,[1] and Susan R. Wilson[1]

[1] Centre for Bioinformation Science, Mathematical Sciences Institute, Australian National University, Canberra, A.C.T. 0200, Australia; conrad.burden@anu.edu.au

[2] John Curtin School of Medical Research, Australian National University, Canberra, A.C.T. 0200, Australia

Summary. The inference of regulatory networks from microarray data relies on expression measures to identify gene activity patterns. However, currently existing expression measures are not the direct measurements of mRNA concentration one would ideally need for an accurate determination of gene regulation. If the development of expression measures is to advance to the point where absolute target concentrations can be estimated, it is essential to have an understanding of physical processes leading to observed microarray data. We survey here the performance of existing expression measures for oligonucleotide microarrays and describe recent progress in developing physical dynamic adsorption models relating measured fluorescent dye intensities to underlying target mRNA concentration.

Key words: Gene expression, microarrays, adsorption models.

29.1 Introduction

Top-down inference of genetic regulatory networks frequently relies on large-scale gene expression data from microarray experiments [1]. These experiments are designed to enable the evaluation of simultaneous expression of large numbers of genes in prepared messenger RNA samples. Standard expression measures, which are not direct measurements of transcript or mRNA concentration, are typically used to group co-expressed genes in order to identify clusters and networks of causally connected genes. However, bottom-up studies of relatively simple regulatory networks (see, for example, [2] and references therein) show that gene regulation is often driven by subtle changes in mRNA concentrations of specific genes, and that these genes may not necessarily be highly expressed. Knowledge of absolute mRNA concentrations, rather than relative expression measures in arbitrary units, could therefore prove invaluable for the reconstruction of regulatory networks.

Data analysed from Affymetrix GeneChip arrays, the focus of this paper, usually consists of expression measures such as determined by Microarray Suite v.5 (MAS5) [3] or robust multi-chip average (RMA) [4]. Such indices are generally calculated by subtracting an estimate of background (estimated from mismatch readings in the case

of MAS5), and summarizing over readings from features within a probeset. The approach is largely empirical: little or no attempt is made to incorporate a model of the physical processes driving hybridisation, and consequently neither the effects of saturation at high target concentration nor the effects of probe sequence specificity are accounted for. As a consequence, expression indices are given in arbitrary units and not units of concentration.

Recently, studies have begun to address these issues by appealing to models based on well-established principles of physical chemistry [6–9]. Such models, known generically as chemical adsorption models, offer the possibility of predicting absolute target concentrations as opposed to relative expression measures, and hence have the potential to enable comparisons between expression levels of different genes.

In this chapter we review the performance of MAS5 and RMA expression measures in cases where the underlying mRNA concentration is known from spike-in experiments. We also review the current status of physical adsorption models and discuss an important open problem, the perfect match/mismatch differential at saturation concentrations. The eventual aim of such research is to provide a practical method of estimating the absolute concentration of mRNA in biological samples from microarray data.

29.2 Performance of Existing Expression Measures

Affymetrix GeneChip arrays consist of a substrate onto which short single strand DNA oligonucleotide probes have been synthesized using a photolithographic process. The microarray surface is divided into some hundreds of thousands of regions commonly 11 to 20 micrometers square, the probes within each region being synthesised to a specific nucleotide sequence. Throughout this chapter we use the word "probe" to refer to a single strand of synthesised DNA, and "feature" to refer to a region of identically synthesised strands. Depending on the array design, each gene is represented by a set of between 11 to 16 pairs of features. One element of each pair is synthesised as a perfect match (PM) sequence designed to be of length 25 bases, and the other a mismatch (MM) sequence identical with the PM sequence except that the middle (13th) base has been replaced by its complement. In practice, the stepwise photolithographic process for synthesising probes is not perfect, and each feature contains a population of probes of lengths up to 25 bases [10]. Each PM sequence is claimed to be a non-overlapping contiguous subsequence of the full gene sequence, chosen for its predicted hybridisation properties and specificity to the potential target gene. The target RNA sample is hybridized onto the chip to form probe-target duplexes, and the chip is scanned to obtain fluorescence intensity readings from dyes incorporated during the laboratory procedures.

The performance of the expression measures MAS5 and RMA can be evaluated using publicly available data from the Affymetrix Human HG-U95A Latin Square spike-in experiment (http://www.affymetrix.com/support/technical/sample_data/datasets. affx). In this experiment genes (or, more precisely, RNA transcripts) were spiked in at cyclic permutations of the set of known concentrations, together with a background
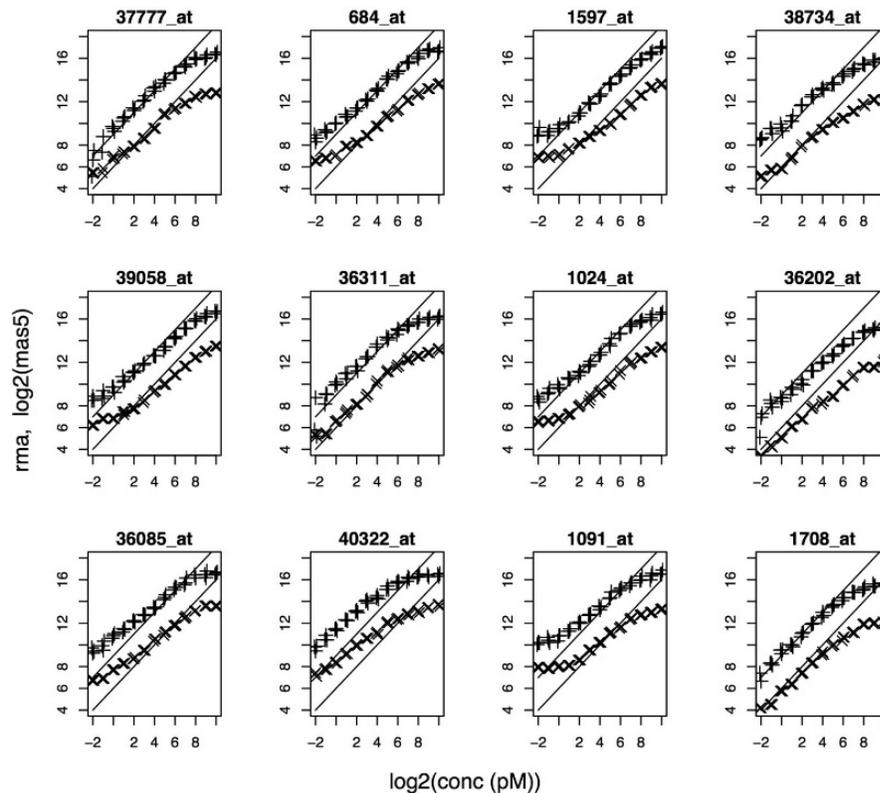
Fig. 29.1. log(MAS5) (upper data points) and RMA (lower data points) expression measures for the Affymetrix HG-U95A spike-in data set. The diagonal lines indicate the expected behaviour if expression measures correctly track target concentration fold changes. Units on the vertical axes are the arbitrary units in which these two expression measures are defined. The MAS5 and RMA units are not related.

of cRNA extracted from human pancreas. The data consists of fluorescence intensity values from a set of 14 probesets corresponding to 14 separate genes, each containing 16 pairs of features. For each probeset a set of fluorescence intensity values was obtained for the 14 spiked-in concentrations (0, 0.25, 0.5, 1, 2, 4, ... , 1024) pM. The experiment was replicated three times using microarray chips from different wafers. In common with previous analyses of this data set, we will concentrate here on data from 12 of the 14 genes, omitting data from two defective genes.

In Fig. 29.1 we plot MAS 5 and (PM-only) RMA expression measures as a function of spike-in cRNA concentration for each of 12 genes, together with the required behaviour of an expression measure if it were to correctly track fold changes. These measures were evaluated using the functions mas5() and rma() provided with the Bioconductor [5] package affy, Version 1.2. We see that with both measures fold changes are underestimated in all cases at high and low target concentrations. Shown in Fig. 29.2
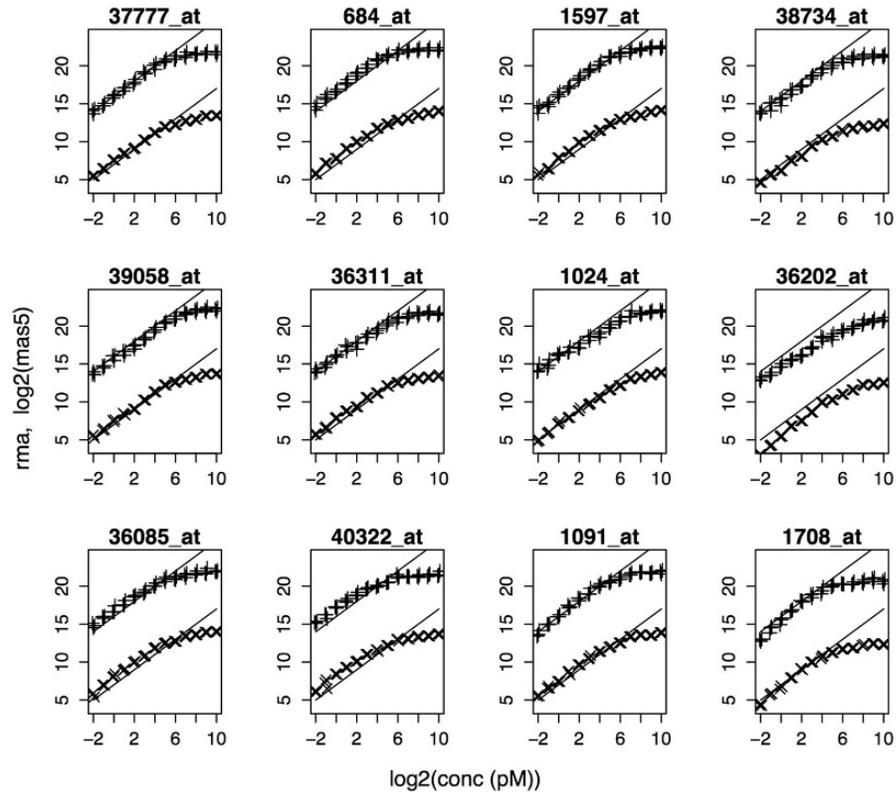
Fig. 29.2. The same as Fig. 29.1 for spike-in data without the human pancreas background cRNA.

are expression measures calculated from data provided by Affymetrix from a similar experiment in which no human pancreas cRNA background was present. For these data fold changes at low concentration are accurately reported by both MAS5 and RMA, indicating that the problem at these concentrations is caused by cross hybridisation from non-specific targets. Underreporting of fold changes at high concentrations, on the other hand, is known to be due to failure of expression measures to account adequately for saturation of features. In the following section we present a physical adsorption model which accounts for both of these phenomena.

29.3 The Langmuir Isotherm Model

The Langmuir model [6–9] is based on an assumption that there are two competing processes driving hybridisation: adsorption, i.e., the binding of target molecules to immobilized probes to form duplexes, and desorption, i.e., the reverse process of duplexes dissociating into separate probe and target molecules

$$\text{Probe} + \text{Target} \rightleftharpoons \text{Duplex}. \tag{29.1}$$
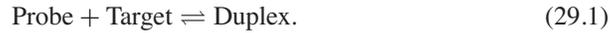
Herein we shall always use the word "probe" to indicate single strand DNA immobilised on the microarray, "target" to indicate RNA in solution and "duplex" to indicate a bound probe-target pair. Both the forward and reverse processes are determined by chemical rate constants which depend on a number of factors including activation energies and temperature.

We briefly survey here two approaches to the adsorption model in the presence of non-specific hybridisation, one by Hekstra et al. [6] based on balancing the chemical reaction Eq. (29.1), and one by Halperin et al. [7] based on equilibrium statistical mechanics. The approaches are shown to be equivalent and lead in both cases to a hyperbolic response function, or equilibrium Langmuir isotherm, relating RNA target concentration x to a measured equilibrium fluorescence intensity y, namely,

$$y(x) = y_0 + b \frac{x}{x + K}. \tag{29.2}$$

The isotherm is defined by three parameters: y_0 is the measured background intensity at zero specific target concentration, b is the saturation intensity above background at infinite specific target concentration and K is the specific target concentration required to reach half saturation.

29.3.1 Hekstra Approach

Hekstra et al. [6] determine the Langmuir parameters y_0, b and K for hybridisation in the combined presence of a specific cRNA target species and a single, non-specific target species. Here we extend their results to any number of non-specific species $i = 1, 2, \ldots$. We will define "specific" to mean PM specific. All other hybridisation will be referred to as "non-specific."

For a given feature on the microarray surface, whether PM or MM, let the concentration of target molecules specific to the PM feature of the neighbouring pair be x, and the concentration of the non-specific species i be z_i. Let the forward and backward rates of the chemical reaction Eq. (29.1) be k_f and k_b respectively for the formation of duplexes with specific target molecules, and k_{fi} and k_{bi} for the formation of duplexes with the ith species of non-specific target. Finally, let θ be the fraction of probe sites occupied by specific probe-target duplexes, and ϕ_i be the fraction of probe sites occupied by duplexes formed with the ith non-specific species of target.

The equilibrium equations are given by balancing forward and backward reactions:

$$0 = k_f x \left(1 - \theta - \sum_j \phi_j\right) - k_b \theta, \tag{29.3}$$

$$0 = k_{fi} z_i \left(1 - \theta - \sum_j \phi_j\right) - k_{bi} \phi_i. \tag{29.4}$$

Defining equilibrium constants $K_S = k_b/k_f$, $K_i = k_{bi}/k_{fi}$ and solving for θ and ϕ_i gives

$$\theta = \frac{x/K_S}{1 + x/K_S + \sum_i z_i/K_i}, \qquad \phi_i = \frac{z_i/K_i}{1 + x/K_S + \sum_j z_j/K_j}. \tag{29.5}$$

Introducing proportionality constants b_S and b_i for the specific and non-specific hybridisations and a physical optical background a, the measured fluorescence intensity is

$$y(x) = a + b_S\theta + \sum_i b_i\phi_i = y_0 + b\frac{x}{x + K}, \tag{29.6}$$

where

$$y_0 = a + A, \qquad b = b_S - A, \qquad K = K_S B, \tag{29.7}$$

and

$$A = \frac{1}{B}\sum_i \frac{b_i z_i}{K_i}, \qquad B = 1 + \sum_i \frac{z_i}{K_i}. \tag{29.8}$$

In spite of the presence of non-specific hybridisation the hyperbolic form of the Langmuir isotherm Eq. (29.2) is recovered. The purpose of Eqs. (29.7) and (29.8) is to relate the three estimated isotherm parameters y_0, b and K to the underlying physical parameters: a, b_S, b_i, K_S, K_i and a set of non-specific background target concentrations z_i.

29.3.2 Halperin Approach

The hyperbolic form of the equilibrium isotherm is equally derivable from statistical mechanics by considering a Gibbs distribution at constant chemical potential [11]. Halperin et al. [7] have used this approach to study adsorption in microarray chips in the presence of non-specific hybridisation. This approach has the benefit of relating equilibrium constants to duplex binding energies.

For the case of one specific and several non-specific target species, the free energy per mole of probe sites at the microarray surface is

$$\gamma = RT\left[\theta\ln\theta + \sum_i \phi_i\ln\phi_i + \left(1 - \theta - \sum_i \phi_i\right)\ln\left(1 - \theta - \sum_i \phi_i\right)\right]$$
$$+ \theta\mu_{pt}^0 + \sum_i \phi_i\mu_{pti}^0 + \left(1 - \theta - \sum_i \phi_i\right)\mu_p^0, \tag{29.9}$$

where the specific and non-specific coverage fractions θ and ϕ_i are defined in the previous section, μ_{pt}^0, μ_{pti}^0 and μ_p^0 are reference state chemical potentials per mole of specific and non-specific probe-target duplexes and unmatched probes respectively, R is the gas constant and T the absolute temperature. The exchange chemical potentials of the various species of probe-target duplexes are

$$\frac{\partial\gamma}{\partial\theta} = RT\left[\ln\theta - \ln\left(1 - \theta - \sum_i \phi\right)\right] + \mu_{pt}^0 - \mu_p^0,$$

$$\frac{\partial \gamma}{\partial \phi_i} = RT \left[\ln \phi_i - \ln \left(1 - \theta - \sum_i \phi \right) \right] + \mu_{\text{pt}i}^0 - \mu_{\text{p}}^0.$$

At equilibrium these exchange chemical potentials balance the chemical potentials of the corresponding target molecule species in solution. Assuming bulk concentrations of target molecules are not appreciably affected by hybridisation, these are given in terms of reference values μ_t^0 and μ_{ti}^0 at reference concentrations x_0 and z_{0i} of specific and non-specific target molecules by

$$\mu_t = \mu_t^0 + RT \ln \frac{x}{x_0}, \qquad \mu_{ti} = \mu_{ti}^0 + RT \ln \frac{z_i}{z_{0i}}. \qquad (29.10)$$

Matching exchange chemical potentials with target chemical potentials gives

$$RT \ln \frac{x}{x_0} = RT \left[\ln \theta - \ln \left(1 - \theta - \sum_i \phi \right) \right] + \Delta G,$$

$$RT \ln \frac{z_i}{z_{0i}} = RT \left[\ln \phi_i - \ln \left(1 - \theta - \sum_i \phi \right) \right] + \Delta G_i,$$

where we have defined the duplex binding free energies $\Delta G = \mu_{\text{pt}}^0 - \mu_{\text{p}}^0 - \mu_t^0$ and $\Delta G_i = \mu_{\text{pt}i}^0 - \mu_{\text{p}}^0 - \mu_{ti}^0$. Further defining $K_S = x_0 \exp(\Delta G/RT)$ and $K_i = z_{0i} \exp(\Delta G_i/RT)$ and rearranging gives the isotherms

$$\frac{\theta}{1 - \theta - \sum_i \phi_i} = \frac{x}{K_S}, \qquad \frac{\phi_i}{1 - \theta - \sum_i \phi_i} = \frac{z_i}{K_i}.$$

Solving for θ and ϕ_i, we recover Eq. (29.5) and hence Eq. (29.6). Note that this approach establishes a relationship between binding free energies and the chemical equilibrium constants introduced earlier.

29.4 Fits to the Spike-in Data

We have carried out an extensive statistical analysis of fits of the hyperbolic response function Eq. (29.2) to the PM probe data from the Affymetrix Human HG-U95A Latin Square spike-in experiment [9]. In Fig. 29.3 we show fits of Eq. (29.2) to fluorescence intensity data from the 16 PM and MM features corresponding to one of the 12 genes. These fits were estimated using a generalized linear model assuming the data at each spike-in concentration for each probe sequence to be drawn from a gamma distribution. The behaviour of this gene is typical of all 12 genes considered. Our findings [9] are summarised as follows:

1. Measured fluorescence values can be approximated by a gamma distribution with mean given by Eq. (29.2) and coefficient of variation of ≈ 0.17.
2. The equilibrium isotherm Eq. (29.2) tracks fold changes from both PM and MM probes over the range of spiked-in concentrations from < 1 pM to > 1000 pM.
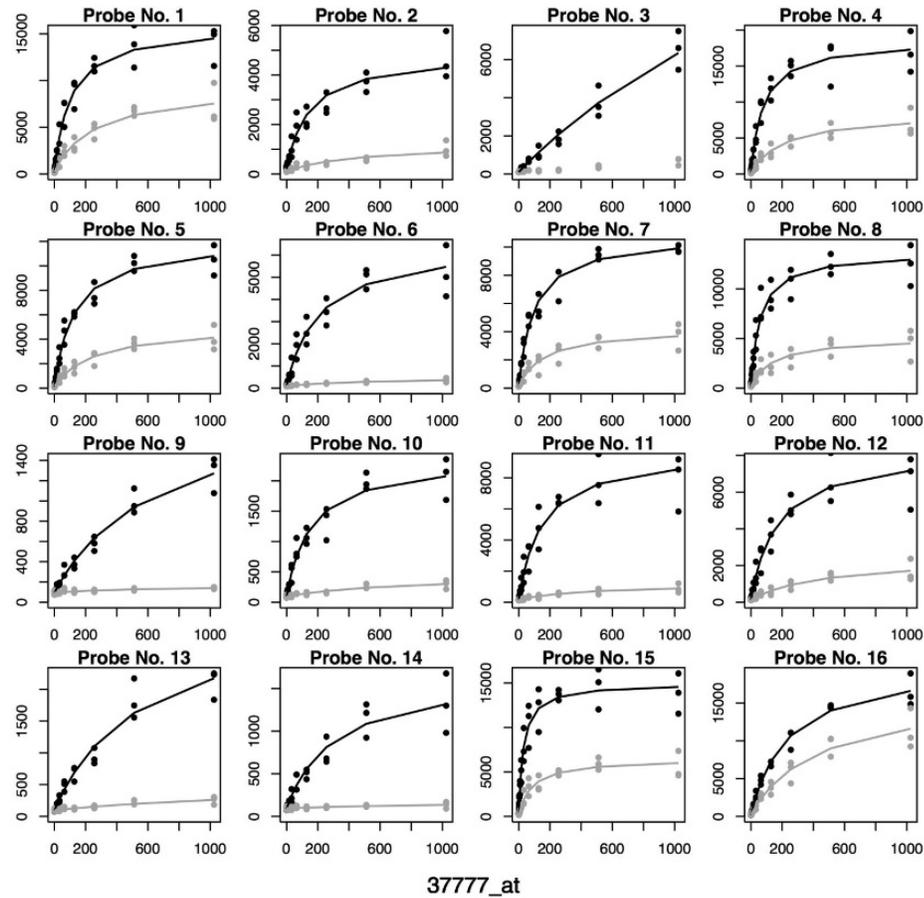
Fig. 29.3. Fits of Eq. (29.2) to fluorescence intensity data for the 16 PM (black) and 16 MM (grey) features of the gene 37777_at probeset of the Affymetrix spike-in experiment. Concentrations (horizontal axes) are in picomoles per liter and fluorescence intensities (vertical axes) are in the arbitrary units used in Affymetrix .cel files. The fit to MM probe No. 3 gave unphysical negative values to the parameters K and b and is not shown.

3. All three parameters y_0, b and K are probe sequence dependent (in contrast with the findings of Ref. [8]).
4. MM features almost invariably saturate at a lower asymptotic intensity $y_0 + b$ than their PM counterparts.

The final point has important consequences for the Hekstra/Halperin adsorption model described in the previous section. Below we argue that, if the Hekstra/Halperin model with non-specific hybridisation leading to the solution given by Eqs. (29.6) through (29.8) is assumed, we are led inescapably to a conclusion that the PM and MM intensity measurements for a given probe pair must saturate at the same asymp-

totic intensity value, in obvious contradiction to observation 4 above. This point has been inferred previously [10], but does not appear to be generally appreciated in the literature, with the exception of work by Peterson et al. [12].

Consider two neighbouring features on a microarray, one PM and one MM, their probe sequences differing only by the middle base. Note that here we define the word "specific" to mean those target RNA which are exact complements to the PM sequence, even when dealing with the MM feature. In what follows this definition will prove useful given that, for most probe pairs, the dominant part of the MM signal at high spike-in concentrations in the Affymetrix experiment appears to come from hybridisation of spiked-in target RNAs complementary to the PM sequence (see Fig. 29.3 for instance). Parameters relating to the PM and MM features will be indicated by superscripts PM and MM respectively.

Although the sums occurring in Eq. (29.8) are over the same set of non-specific targets for PM as for MM, one can expect $A^{MM} \neq A^{PM}$ since in general $K_i^{MM} \neq K_i^{PM}$. Considering the asymptotic intensities at high concentration, however, Eqs. (29.6) and (29.7) imply that, under the Hekstra/Halperin model, the non-specific hybridisation effects cancel out:

$$
\begin{aligned}
y^{MM}(\infty) &= y_0^{MM} + b^{MM} = a + b_S, \\
y^{PM}(\infty) &= y_0^{PM} + b^{PM} = a + b_S.
\end{aligned}
\tag{29.11}
$$

Note that a and b_S do not differ between intensity measurements from a neighbouring PM/MM pair of features. The physical background a depends only on physical properties of the chip in the absence of any hybridisation, such as reflectance, and such properties will not vary significantly over a distance of a few micrometers. The parameter b_S depends only on the amount of fluorescent light emitted per hybridized specific target RNA molecule, and as such is a function of the target sequence only, and not the probe sequence. By our current definition of "specific" this is the *same sequence* for PM and MM, and so b_S is common to both. The Hekstra/Halperin model formulated above then necessarily entails that $y_0^{MM} + b^{MM} = y_0^{PM} + b^{PM}$, in obvious contradiction to the values of y_0 and b obtained by fitting the spike-in data.

The source of the problem is that any model leading to the coverage fraction of Eq. (29.5) entails that, at sufficiently high specific target concentration, all probes form duplexes: as $x \to \infty, \theta \to 1$. That is, all probes in the feature are predicted to form duplexes if saturated with enough specific target, even in the case of the MM feature. This prediction was not borne out by the experiments of Forman et al. [10] who observed less than 10% coverage of PM features by probe-target duplexes at saturation target concentrations.

The problem of differential PM/MM saturation has also been recognized in the context of the Langmuir model without non-specific hybridisation by Peterson et al. [12], who explain their experimental data by invoking a Sips isotherm [13] to explain a lower MM response curve at high target concentrations. Peterson's experimental results are indeed a good fit to the Sips isotherm; however, the Peterson experiment was carried out at a hybridisation temperature of 20°C, while Affymetrix microarrays are hybridized at 45°C, which is much closer to the duplex melting temperature. Furthermore, Peterson et al. found that heating the hybridisation buffer to 37°C and then

cooling back to 20°C almost completely removed any difference in equilibrium saturation intensities between PM and MM probes at a temperature well below the melting temperature.

To determine whether the hyperbolic or Sips isotherm is more appropriate for the Affymetrix spike-in data we have carried out a statistical analysis comparing the fits of the MM data to both isotherms. Our results, which will be presented elsewhere [14], show that for the Affymetrix spike-in data the extra parameters involved in invoking the Sips isotherm are not significant, and that a hyperbolic response function adequately describes the data. We conclude that, at a hybridisation temperature of 45°C, the more appropriate empirical fit to the spike-in data is Eq. (29.2), with $y^{MM}(\infty) < y^{PM}(\infty)$.

29.4.1 The PM/MM Differential at Saturation Concentrations

To resolve the discrepancy between the observed PM/MM intensity differential at saturation concentrations and the theoretical predictions of the models set out in Sect. 29.3 we have recently studied a number of possible extensions to the Hekstra/Halperin model. Our analysis [14] dismisses a number of potential explanations for the discrepancy, including non-equilibrium models of hybridisation (including multistep models which take into account a slow initial step followed by a rapid zipping-up of probe-target duplexes) the effects of bulk target-target hybridisation, and partial zippering of duplexes. We have also considered the effect of competitive hybridisation from probe-probe duplexes at the microarray surface which may render a fraction of DNA probes unavailable to target molecules. While we are unable to dismiss entirely that this may be the cause of the discrepancy, we find that such a model is unlikely to lead to a hyperbolic response curve of the form of Eq. (29.2).

In the light of this analysis, we believe the most promising explanation for the PM/MM intensity differential to lie not with the hybridisation process, but with the ensuing washing step which is designed to remove unbound or loosely bound target molecules from the microarray surface before scanning. A similar idea has been proposed briefly by Zhang [15].

We return to the model prediction that, immediately prior to washing, the fraction θ of sites on a feature occupied by specific probe-target duplexes and the fraction ϕ_i covered by non-specific duplexes of species i are given by Eqs. (29.5). During the washing process some of the duplexes will be dissociated. Suppose that the probability that a given probe-target duplex has survived up to a washing time t_W is $s(t_W)$ for a specific duplex and $s_i(t_W)$ for the ith species of non-specific duplex. The survival functions s and s_i depend only on probe and target properties and not on the target concentrations x and z_i. They satisfy $s(0) = 1$ and are monotonically decreasing. The specific and non-specific duplex coverage fractions at time t_W are then

$$\theta(x, t_W) = \frac{s(t_W)x/K_S}{1 + x/K_S + \sum_i z_i/K_i}, \tag{29.12}$$

$$\phi_i(x, t_W) = \frac{s_i(t_W)z_i/K_i}{1 + x/K_S + \sum_j z_j/K_j}. \tag{29.13}$$

Repeating the assumption used in Sect. 29.3 that the measured fluorescence intensity is a linear function of the duplex coverage fractions, that is $y(x, t_W) = a + b_S \theta(x, t_W) + \sum_i b_i \phi(x, t_W)$, we find that the hyperbolic form

$$y(x, t_W) = y_0(t_W) + b(t_W) \frac{x}{x + K} \tag{29.14}$$

is maintained, and that the empirically observed parameters are now given by

$$y_0(t_W) = a + A(t_W), \quad b(t_W) = s(t_W)b_S - A(t_W), \quad K = K_S B, \tag{29.15}$$

where

$$A(t_W) = \frac{1}{B} \sum_i \frac{s_i(t_W)b_i z_i}{K_i}, \quad B = 1 + \sum_i \frac{z_i}{K_i}. \tag{29.16}$$

Note that the parameter K is unaffected by the length of the washing process, and depends only on duplex binding free energies via the hybridisation step. The asymptotic fluorescence intensity at high target concentration,

$$y(\infty, t_W) = y_0(t_W) + b(t_W) = a + s(t_W)b_S, \tag{29.17}$$

is depressed by the presence of the survival fraction $s(t_W)$.

The simplest reasonable assumption, that the rate at which duplexes are dissociated depends only on the current coverage fraction, leads to an exponential survival function

$$s(t_W) = e^{-\kappa t_W}, \tag{29.18}$$

where the constant κ is expected to increase with decreasing duplex binding affinity. Since the binding affinity of a PM-specific target to an MM probe is less than that to a PM probe, we expect in general that $\kappa^{MM} > \kappa^{PM}$, or equivalently, $s^{MM}(t_W) < s^{PM}(t_W)$ and hence $y^{MM}(\infty) < y^{PM}(\infty)$ as observed. We conclude that the washing hypothesis is promising in that it correctly predicts a hyperbolic isotherm with differential asymptotic PM and MM responses at high concentration.

29.5 Conclusions

The ultimate aim of this research is to provide a practical method of estimating absolute concentration of mRNA in biological samples taken in a typical laboratory situation. We are unaware of any existing downloadable software using adsorption models to infer absolute RNA target concentrations from measured fluorescence intensities and probe sequences. While the development of such a concentration estimator is still in its infancy, preliminary work has been done on testing the ability of adsorption models to recover known spike-in concentrations. Hekstra et al. [6] have used a simple linear model to infer probe sequence dependent parameters of the hyperbolic isotherm Eq. (29.2) from nucleotide abundances within each probe sequence. The inferred hyperbolic isotherm is then used to extract a concentrations estimate from each of the
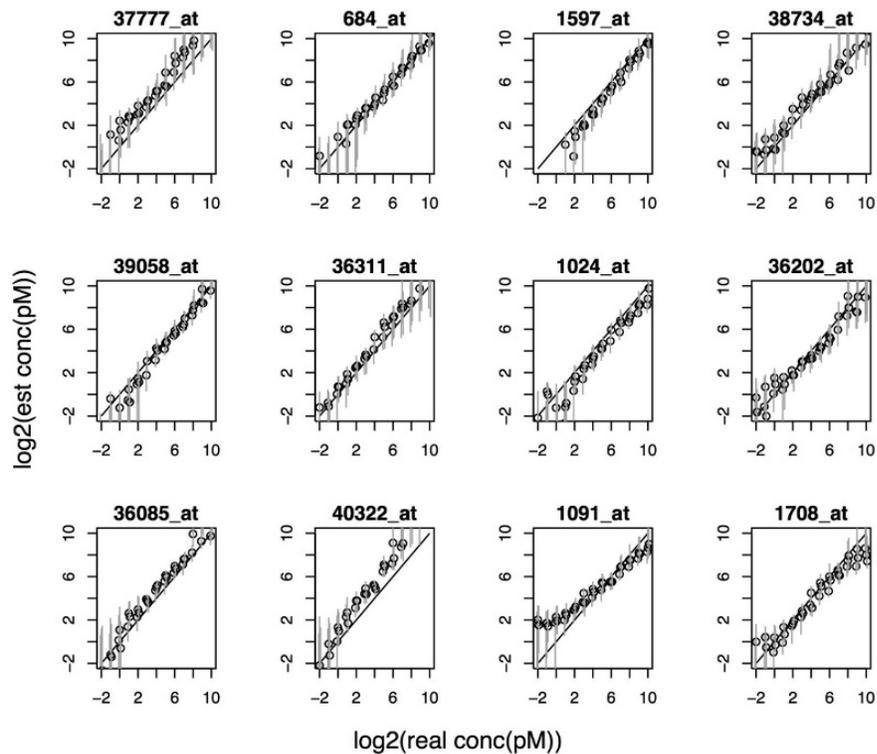
Fig. 29.4. Estimates of mRNA concentration in the Affymetrix HG-U95A spike-in experiment obtained by inverting the hyperbolic adsorption isotherm. For details of the calculaton, see Ref. [9]. Error bars are approximate 95% confidence intervals obtained by bootstrap resampling. The line indicates the perfect relationship between predicted and actual concentration.

16 PM features within a probeset, and an estimate of absolute mRNA concentration is then obtained by averaging over these 16 values. They applied this method to a cross-validated analysis of the Affymetrix HG -U95A spike-in experiment with reasonable success. In a refined version of this calculation [9], we have found that biases can be removed by replacing the average with the median, and have placed confidence intervals on our concentration estimates using bootstrapping. Our results are shown in Fig. 29.4. In most cases the actual spike-in concentration falls within the 95% confidence intervals, which generally cover less than an order of magnitude.

While this method performs at least as well as conventional expression measures in the case of the highly controlled Affymetrix spike-in experiment, it falls short of providing a universally applicable mRNA concentration estimator. In a clinical laboratory experiment, for instance, one would pay greater attention to normalisation effects caused by photomultiplier settings and variations in chip manufacture, and one could not necessarily assume the non-specific background to have the characteristics of the uniform background used in the spike-in experiment. Further analysis needs to

be done with other spike-in data sets, and we next intend to analyse the more extensive Affymetrix HG-U133A Latin Square data set. Also of interest are dilution data sets such as that in Ref. [16]. These data sets do not provide absolute spike-in concentrations, as they are more concerned with exploring microarray responses in terms of "fold changes." However, given the highly non-linear response of microarrays evident in Eq. (29.2) and Fig. 29.3, the concept of fold change can be very misleading. Nevertheless, these data sets could prove useful for testing any proposed concentration estimator based on adsorption models.

The above method performed particularly well at high, saturation concentrations (compare Fig. 29.4 with 29.1), but less well at low concentrations where non-specific hybridisation is important. Furthermore, no use was made of the MM intensities. Here we believe the model of Sect. 29.4.1 may be helpful. Assuming exponential survival functions, Eq. (29.14) is a sum of exponentials whose dominant contribution at large washing times t_W comes from the specific target. If fluorescence intensities could be continuously monitored in real time during the washing phase, using technology similar to the flow cell experiment described by Forman et al. [10], the potential exists to separate the specific signal from non-specific background as the linear contribution to a logarithmic plot at $t_W \to \infty$.

References

1. Segal, E., Friedman, N., Kaminski, N., Regev, A., Koller, D.: From signatures to models: Understanding cancer using microarrays. Nat. Genet., **37**, S38–45 (2005).
2. Gilman, A., Arkin, A.P.: Genetic "code": Representations and dynamical models of genetic components and networks. Annu. Rev. Genomics Hum. Genet., **3**, 341–369 (2002).
3. Affymetrix Inc., 2002, Statistical Algorithms Description Document, available at http://www.affymetrix.com/support/technical/whitepapers.affx.
4. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., et al.: Exploration, normalization and summaries of high density oligonucleotide array probe level data. Biostatistics, **4**, 249–264 (2003).
5. http://www.bioconductor.org/
6. Hekstra, D., Taussig, A.R., Magnasco, M., Naef, F.: Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays. Nucl. Acids Res., **31**, 1962–1968 (2003).
7. Halperin, A., Buhot, A., Zhulina, E.B.: Sensitivity, specificity and the hybridization isotherms of DNA Chips. Biophys. J., **86**, 718–730 (2004).
8. Held, G.A., Grinstein, G., Tu, Y.: Modeling of DNA microarray data by using physical properties of hybridisation. Proc. Natl. Acad. Sci., **100**, 7575–7580 (2003).
9. Burden, C.J., Pittelkow, Y., Wilson, S.R.: Statistical analysis of adsorption models for oligonucleotide microarrays. Stat. Appl. Gen. Mol. Biol., **3**, Article 35 (2004).
10. Forman, J.E., Walton, I.D., Stern, D., Rava R.P., Trulson, M.O.: Thermodynamics of duplex formation and mismatch discrimination on photolithographically synthesised oligonucleotide arrays. In: Leontis, N.B., SantaLucia, J. (eds.) Molecular Modeling of Nucleic Acids. American Chemical Society, Washington, D.C. (1998).
11. Hill, T.L.: An Introduction to Statistical Thermodynamics. Addison-Wesley, Reading, MA (1960).

12. Peterson, A.W., Wolf, L.K., Georgiadis, R.M.: Hybridization of mismatched or partially matched DNA at surfaces. J. Am. Chem. Soc., **124**, 14601–14607 (2002).
13. Sips, R.: On the structure of a catalyst surface. J. Chem. Phys., **16**, 490–495 (1948).
14. Burden, C.J., Pittelkow, Y., Wilson, S.R.: Adsorption models of hybridisation behaviour on oligonucleotide microarrays. J. Phys.: Condes. Matter, **18**, 5545–5565 (2006).
15. Zhang, L.: Modeling sequence dependence of probe signals on oligonucleotide microarrays. Talk at the Affymetrix Genechip Microarray low level workshop, Berkely, Ca., available at http://affymetrix.com/corporate/events/seminar/microarray.workshop.affx (2003).
16. Choe, S.E., Boutros, M., Michelson, A.M., Church, G.M., Haflon, M.S.: Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. Genome Biology, **6**, R16 (2005).

Validation of Human Alternative Splice Forms Using the EASED Platform and Multiple Splice Site Discriminating Features

Ralf Bortfeldt,[1,2] Alexander Herrmann,[2] Heike Pospisil,[3] and Stefan Schuster[1]

[1] Friedrich-Schiller University Jena, Dept. of Bioinformatics, Ernst-Abbe-Pl. 2, D-07743 Jena, Germany; {bortfeldt,schuster}@minet.uni-jena.de

[2] Dept. of Bioinformatics, Max-Delbrck Center for Molecular Medicine, D-13092 Berlin, Germany; alexander.herrmann@mdc-berlin.de

[3] University of Hamburg, Center for Bioinformatics, D-20146 Hamburg, Gemany; pospisil@zbh.uni-hamburg.de

Summary. We have shown for a dataset of computationally predicted alternative splice sites how inherent information can be utilized to validate the predictions by applying statistics on different features typical for splice sites. As a promising splice site feature we investigated the frequencies of binding motifs in the context of exonic and intronic splice site flanks and between the alternative and reference splice sites. We show that both partitions of splice sites can statistically be separated not only by their distance to the splice signal consensus but also via frequencies of splice regulatory protein (SRp) binding motifs in the splice site environment.

Key words: Splicing, alternative splicing, SR proteins, splicing enhancer.

30.1 Introduction

30.1.1 Alternative Splicing: From Sequence to Protein-Controlled Regulation

The excision of specific parts of mRNA strands before translation into proteins is called splicing. The increasing proteome complexity from protozoan eukaryotes via nematodes and arthropods to vertebrates is attributed to mechanisms of differential, that is alternative, splicing [1, 12, 21]. Furthermore, pathological complexity can be ascribed to misregulated splicing processes where cancer is one of the most prominent examples [29]. The mechanism of alternative splicing has attracted a wide range of scientific research addressing the problem with computational strategies and tools. Hence, up to 60% of the human genes have been predicted to produce alternative splice forms (*asf*) [18, 22], and a variety of databases have been designed to collect splice forms combined with experimental observations as tissue information or developmental and pathology information [10]. Experiments on the spliceosome—the transient megadalton complex that assembles the catalytic center to cut and splice pre-mRNA—showed

that splice regulatory proteins (SR proteins) are involved in initiating and maintaining this complex [3,13,25]. These proteins form a family of proteins rich in serine-arginine domains. Further, it was shown that SR proteins can additionally bind to specific sequence motifs (cis-elements) in order to promote alternative splice site usage and exon definition [12,19]. Considering the growing percentage of genes which are predicted to produce alternatively spliced transcripts, there must be a dense regulatory network, sensing and utilizing the information required to perform the surgical task of splicing and thus reacting appropriately to environmental conditions that are required by a specific splice form. Experimental studies in the last decade have provided examples of regulated alternative splice events [1,26], establishing working hypotheses and the basis for further computational studies. Accordingly, the following roughly sketched topics have been addressed with computational efforts:

(i) consensus splice signal prediction [30],
(ii) splice site plasticity [15],
(iii) exon/intron sequence composition [6],
(iv) splicing pattern statistics [31] and
(v) consensus enhancer/silencer signal prediction [5,9].

From these approaches we utilized categories (i), (iv) and (v) to channel their capabilities in the analysis of predicted *asf*. Furthermore, we tested whether repeats of splicing enhancer motifs in the neighborhood of splice sites (GT..AG consensus) may constitute a useful feature in discriminating alternative splice sites.

30.1.2 Prediction of Alternative Splice Forms

The EASED project[4] [23] constitutes one of the earliest efforts to combine an algorithm for *asf* prediction with a variety of additional medically relevant information such as reference to cancer, prevalence for specific tissue and developmental information [23]. Utilizing the EASED dataset of predicted *asf* this work combines splice site information content, transcript abundance and exonic splicing enhancer (ESE) occurrence in order to infer knowledge about the quality of a set of predicted *asf*. A central question was, whether multiple ESEs or enhancer repeats may contribute to the discrimination of predicted alternative splice sites. Since the prediction algorithm of the EASED pipeline relies on mRNA and expressed sequence tags (ESTs), the number of transcripts as well as the results of the tested hypotheses can be taken as a general validation criterion for the applicability of such comprehensive database systems and predicted *asf*. Up to now no perfect ab initio prediction algorithm for alternative splice events exists, and many characteristic features if taken isolated may not be strong enough to validate a splice site to certainty. The combination of several classes of information such as splice site score (S_{ss}), frequency of (splicing enhancing) SRp binding motifs (f_{ese}), transcript support (f_t), tissue type, disease and developmental state account for a strong ensemble that can shed more light on the processes of both constitutive and alternative splicing.

---

[4] http://eased.bioinf.mdc-berlin.de

30.1.3 SR Proteins and Their Binding Signals

SR proteins are members of a family of proteins possessing serine-arginine rich do-mains which can interact via these domains and thus provide bridging functions in spliceosome assembly [11]. SR proteins can recognize and bind a variety of ESEs and/or exonic splicing silencers (ESSs), which are motifs of approximately 6–12 nu-cleotides. Studies have shown that more than one copy of a high affinity SR protein (SRp) binding site can efficiently activate splicing, as was shown for three sequen-tial SRp binding motifs (ASF/SF2 and SRp40) [27, 28]. This finding was explained by the action of cooperative binding of a single factor, thus conveying specificity by outcompeting other trans-factors with a lower degree of RNA affinity in this region. Contrary to the single factor binding model, other studies have shown that combina-tions of SR proteins can promote splicing by binding to the same enhancer motif [24]. A variation of this model is shown by another example where human caldesmon exon 5 and 6 showed multiple purine rich repeats with an enhancer function of which a 32 nucleotide (nt) stretch enhanced an internal alternative 5′ss within exon 5 [16]. Since most of the reported ESEs show lengths between 5–19 nucleotides [11], longer mo-tifs may consist of several overlapping submotifs that increase the binding affinity of its recognizing trans-factors [8]. Even the possibility of a functional composite splice regulatory element has been demonstrated where motifs within CD44 exon 5v had to be direct neighbors to activate a functional splice complex, suggesting cooperating interactions of the binding proteins [17].

30.2 Materials and Methods

30.2.1 Source of Alternative and Reference Splice Sites

All analyses are based on the EASED version 19.34a from December 2003 that uses mRNA and gene sequence material of the underlying Ensembl database (same freeze) and ESTs from dbEST (isochronic version) [2]. The analysis was restricted to *Homo sapiens* as here the most abundant sequence information was available. Furthermore, SR proteins and their binding motifs have been investigated most extensively in this species. In general the prediction of the utilized *asf* is based on an mRNA-EST align-ment with a stringent set of parameters [23]. *Asf* arise through gaps in either one or the other alignment partner, be it mRNA or EST (Fig. 30.1). For each gene, the longest mRNA not classified as "alternatively spliced" was taken per definition as the reference transcript denoting a set of "constitutive" or reference donors and acceptors. Taking the longest mRNA as reference offers the chance to observe as many alignment gaps as possible against the available EST and, hence, to increase the number of consid-ered alternative splice junctions. According to these mRNA forms, all other differing splice junctions were considered alternative, whereas EST and, mRNA matching the reference splice sites was counted as evidence for the reference dataset of donors and acceptors. All gap coordinates in the transcripts were mapped back to genomic se-quence to extract the required splice site sequence context. Labeling any transcript as
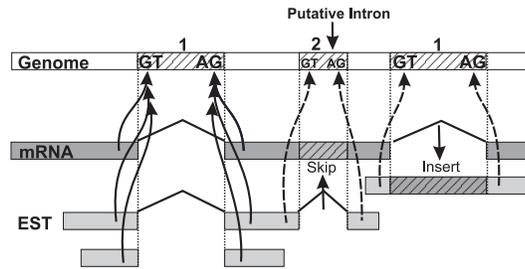
Fig. 30.1. Transcripts (EST + mRNA) supporting reference (1) and alternative (2) splice sites. Black arrows indicate the outcome of the EST-mRNA alignment; dashed arrows point to alternative splice sites predicted by the EST. Hatched boxes indicate introns.

the constitutively expressed form of its gene might be very debatable considering the highly dynamic proteome of a living cell where "constitutive" conditions may finally depend a on a number of external stimuli. Being aware of this uncertainty, we prefer the term "reference" instead of "constitutive."

30.2.2 Data Preparation and Refinement

For establishing the subset of EASED predictions used in the present work, the database was queried with a threshold of at least 5 *asf* (EST) per gene. Next, all genes whose supporting EST did not indicate alternative splice events utilizing the canonical splice site dinucleotides GT..AG were discarded. After refinement the dataset consisted of 2,624 genes with an average number of 8 alternative splice sites per gene and an average support of 3 ESTs per alternative splice site. In contrast $\approx$ 26 ESTs per reference splice sites were found to align in the same region of the transcript. The decrease of initially available ESTs and *asf* throughout the processing steps is documented in Table 30.1. As a further parameter, the genomic sequence window around each donor and acceptor was set to 100 nt in this study (Fig. 30.2). Finally, each splice site was stored as an alternative (*ass*) or reference (*rss*) splice site together with a number of supporting transcripts (mRNA and/or EST).

30.2.3 Splice Site Scoring

The reference and predicted alternative splice sites were evaluated (scored) by applying a maximum entropy model (MEM) implemented by Yeo and Burge [30]. This model estimates short sequence motifs of the donor and acceptor splice sites by approximating the maximum entropy distribution (MED). Minimizing the relative entropy

$$D(\hat{p}) = \sum \hat{p}(x) \log \frac{\hat{p}(x)}{q(x)}, \tag{30.1}$$

where p is the probability of a nucleotide at a specific DNA position and q a prior background distribution, yields the MED that has the lowest relative entropy of all distributions that satisfy a set of constraints relative to the background distribution. Those

Table 30.1. Dataset characteristics and refinement steps, satisfying *asf* prediction criteria (*) and GT..AG consensus (**).

Processing step	Number	Source
blasted ESTs	5,427,257	dbEST (freeze Dec 2003)
Ensembl Transcripts (ENST)	31,609	Ensembl 19_34a (freeze Dec 2003)
Ensembl Genes (ENSG)	23,531	Ensembl 19_34a
EST, matching $\geq$ 1 ENST	3,947,548	EASED
asf predicting EST*	428,474	EASED
predicted *as* sites	102,104	EASED
ENST matching $\geq$ 1 *asf*	21,044	EASED
ENSG matching $\geq$ 1 *asf*	15,426	EASED
ENSG matching $\geq$ 5 *asf*	2,624	EASED
filtered** *ass*Acceptors	3,862	inhouse software Jena
filtered *ass*Donors	3,705	inhouse software Jena
filtered *rss*Acceptors	25,526	inhouse software Jena
filtered *rss*Donors	25,103	inhouse software Jena

marginal constraints are imposed by neighborhood relations between nucleotide positions (estimated from experimental data) and nucleotide frequency counts. The MEM slightly outperforms maximum dependency decomposition (MDD) models and was shown to be clearly better than weight matrix models (WMM), both of which are generalized by the maximum entropy principle. The MED model was trained against a test set of 12,715 human introns utilizing a 9mer motif at the 5′ splice site (ss) and a 23mer at the 3′ss, excluding potential alternative splice junctions and those not showing the
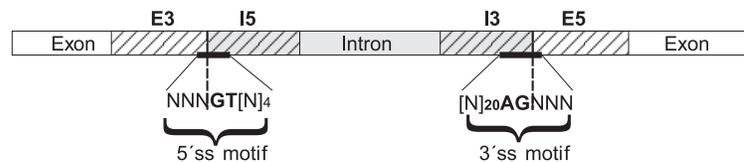
Fig. 30.2. Sequence windows used for determining the ESE frequencies around the scored splice sites. The four validation classes are defined as I5, I3, E5 and E3 with the following convention: E = exon, I = intron, 3 = downstream region (3′), 5 = upstream region (5′). Thus the intronic part (grey hatched) is separated from the exonic part of the donor and acceptor splice sites (white hatched); the position and size of the motifs used for scoring the 5′ss- and 3′ss are indicated with black horizontal bars.

GT..AG consensus [30]. The implementation is provided in Perl[5] and thus the splice site scoring was incorporated into the analysis pipeline.

30.2.4 ESE Scanning

To estimate the frequency of high scoring ESE motifs within the defined context of the *ass* and *rss*, we used scoring matrices implemented in the ESEfinder tool [5]. The matrices contain nucleotide frequencies of short motifs that were shown to bind the SR proteins SF2/ASF (7mer), SC35 (8mer), SRp40 (7mer) and SRp55 (6mer). Experiments have demonstrated the applicability of ESE consensus motifs in explaining splice events such as exon skipping and mis-splicing due to point mutations within those motifs [4]. Nevertheless, due to the degenerated nature of SRp binding motifs we did not consider the frequency of ESE motif hits alone but evaluated their occurrence in the context of additional characteristic information (splice site score, transcript support). We scanned for the four ESE motifs in a window of 100 nucleotides up- and downstream of the predicted *ass* and *rss*. Although scanning for exon splicing enhancer motifs would imply to search only in exonic sequence, given, i.e., two alternative donor splice sites, it depends which of the two donors is considered to clearly state whether the sequence between both donors is exonic or intronic. Thus, one could find a motif as an intronic enhancer in the flanking intron sequence of the upstream donor, which is viewed from the downstream donor an ESE. Besides covering up and downstream sequence flanks in our analysis, we restricted the search to 100 nt in each direction, as this splice site vicinity was previously shown to cover ESE locations [7, 14]. However, enhancer elements have also been reported at locations of more than 100 nt downstream of a donor site [20]. Additionally, restricting the window size may further result in missing regulatory motifs that are used in a structural context (farther distanced enhancer can bring their bound regulatory factors via RNA structure in splice site proximity).

30.2.5 Data Partitioning for Statistical Analysis

As shown in Table 30.1, four classes (data subsets or partitions) were created from the EASED data. These comprise the donor/acceptor sets of *ass* and *rss* splice sites. Due to the up- and downstream distinction of the splice site environment, we obtain data subsets that model either intron or exon specific features as, e.g., SRp binding motif frequency.

30.3 Results and Discussion

30.3.1 General Characterisation of the Validated Splice Site Attributes

Splice Site Score. Comparison of the main attributes *splice site score*, *transcript support* and *ESE frequency* was done first for the complete distribution of both splice

---

[5] http://genes.mit.edu/burgelab/maxent/

Table 30.2. Distributional characteristics of splice site score (S_{ss}) (a) and transcript support (f_t) (b) compared for donor and acceptor splice sites.

(a)				(b)		
	5′ss	**3′ss**			**5′ss**	**3′ss**
Minimum	−44,399	−46,658		Minimum	1	1
1st Quantile	6,640	6,483		1st Quantile	1	1
Median	8,456	8,450		Median	16	16
Mean	6,976	7,671		Mean	41	37
3rd Quantile	9,652	10,141		3rd Quantile	68	63
Maximum	11,807	15,589		Maximum	466	326

site types (donors/acceptors) independent from partitioning into reference (*rss*) and alternative (*ass*) splice sites. Our results indicate that the S_{ss} between the whole un-partitioned distributions of donor and acceptor sites vary significantly ($p < 10^{-3}$, Wilcoxon, t- and F-tests) in their means and variances. The mode of both distributions resides at S_{ss} of 5–10, whereas donor splice sites show lower scores in their quartiles and extrema than acceptor sites (Table 30.2).

Transcript Support. The number of transcripts that support donor and acceptor splice sites (f_t = frequency of transcripts that utilize a specific splice site) show different distributions in particular at higher transcript numbers per splice site (Fig. 30.3). Means and variances of the splice site type dependent f_t distributions vary significantly ($p < 10^{-3}$) with a tendency to more transcripts at donor splice sites. The mode of both donor and acceptor f_t distributions resides between 1 and 20 transcripts.

(a)   (b)   (c)
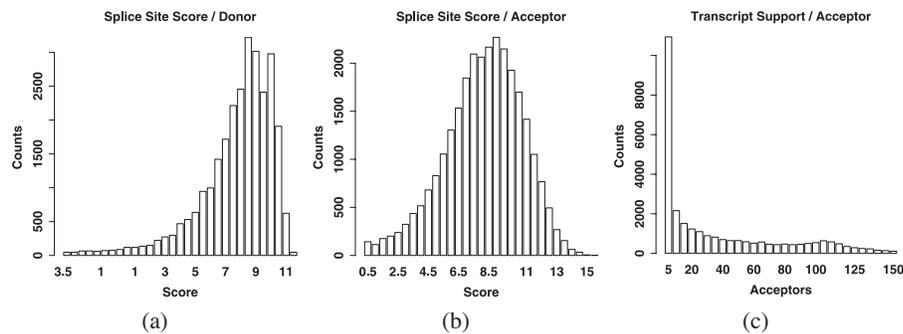
Fig. 30.3. Histograms of splice site scores (S_{ss}) of donor (a) and acceptor sites (b); histogram of transcript frequencies (f_t) supporting the acceptor sites, where bins join acceptors which are supported by a certain number (count) of observed transcripts (c).

Table 30.3. Summary of ESE motif frequency (f_{ese}) characteristics considering the whole set of splice sites while distinguishing the intronic and exonic parts of donor and acceptor splice sites; $p(\mu; \sigma^2)$ denotes p-values in comparing means and variances between the motif frequencies at the specified exonic/intronic region; p-values of significantly ($\alpha \leq 0.0001$) different motif frequencies between the compared splice site regions are marked with (*).

SRp	Splice site region	Median	Mean	Variance	E3 $p(\mu; \sigma^2)$	I3 $p(\mu; \sigma^2)$
ASF/SF2	E3	4	3,934	5,285	—	—
	E5	4	3,933	5,016	$< 1; < 10^{-4*}$	$< 10^{-4*}; < 10^{-4*}$
	I3	2	2,502	4,229	—	—
	I5	3	3,359	6,298	$< 10^{-4*}; < 10^{-4}$	$< 10^{-4*}; < 10^{-4*}$
SC35	E3	4	3,901	3,989	—	—
	E5	4	3,873	3,949	$< 10^{-1}; < 1$	$< 10^{-4*}; < 10^{-4*}$
	I3	3	2,928	3,467	—	—
	I5	3	3,622	4,560	$< 10^{-4*}; < 10^{-4}$	$< 10^{-4*}; < 10^{-4*}$
SRp40	E3	4	3,796	3,145	—	—
	E5	4	3,794	3,095	$< 1; < 1$	$< 10^{-4*}; < 10^{-4*}$
	I3	3	2,823	2,522	—	—
	I5	3	3,452	3,183	$< 10^{-4*}; < 1$	$< 10^{-4*}; < 10^{-4*}$
SRp55	E3	2	2,344	2,277	—	—
	E5	2	2,334	2,236	$< 1; < 1$	$< 10^{-4*}; < 10^{-4*}$
	I3	1	1,626	1,612	—	—
	I5	2	1,947	1,965	$< 10^{-4*}; < 10^{-4}$	$< 10^{-4*}; < 10^{-4*}$

Frequency of SRp Binding Motifs. In investigating the repeated occurrence of SRp binding motifs (f_{ese}) we considered first the total number of detected motifs. Comparison of this combined number of ESE frequencies showed no significant difference ($p > 0.1$, t-test) between the distributions in the exonic donor (E3) and acceptor (E5) regions but a significant difference between the intronic donor (I5) and acceptor (I3) regions ($p < 10^{-4}$). Also, the occurrence of ESE motifs between adjacent exonic (E5) and intronic (I3) splice site flanks differed significantly at acceptor sites ($p < 10^{-3}$). Looking at the motifs of the specific SR protein ASF/SF2, one observes similar f_{ese} distributions in the exonic part of 5' and 3' splice sites, but though they show a significant difference in their variances ($p < 10^{-4}$), their means are not different ($p > 0.1$, Wilcoxon test). In contrast, the number of ASF/SF2 motifs in the intron flanks of the splice sites vary significantly in their variances and means ($p < 10^{-4}$), and we found the same effect between adjacent exonic and intronic splice site environments. A summary of variances and means characterizing the f_{ese} distributions of all four tested SR proteins is listed in Table 30.3.

Occurrence of SRp Binding Motifs at Low and High Scoring Splice Sites. Addition-
ally, and in order to validate f_{ese} independently from our definition of *ass* and *rss*,
all splice sites were separated by their S_{ss} distributions (Fig. 30.3). Furthermore, only
the tails of the score distribution (< 0 and > 10) were selected. Besides addressing
the question of whether the f_{ese} distribution varies significantly between the low and
high scoring sets, this also served to measure the enrichment of predicted *asf* in both
score dependent data subsets (discussed in the section on *ass* and *rss* specific analyses).
The total donor/acceptor splice sites partition in 6%/4% with $S_{ss} < 0$ contrasted by
19%/25% with a score above 10 and thus almost consensus quality. Between these
two sets of extreme-scoring splice sites we find significant differences in the means of
their total (exonic) f_{ese} distributions. In particular, at both donor and acceptor sites the
motif distribution of the protein SC35 tends on average to at least one motif more at
low scoring than at high scoring splice sites. For two other SR proteins this effect was
found only either at donor splice sites (SRp55) or at acceptor splice sites (ASF/SF2).

30.3.2 Characterization Specific to Predicted Alternative and Reference Splice Site Attributes

Splice Site Score and Transcript Support. Comparing S_{ss} between reference and
alternative donors shows a significant difference in the mean values ($p < 10^{-4}$,
Wilcoxon test), although the mode of the distribution still resides at scores between
5 and 10. The same observation applies for the acceptor sites. Interestingly, the differ-
ence of the means between donor and acceptor score distributions is higher in *ass* than
in *rss* (underlined in Table 30.4).

Considering the number of transcripts that support the splice sites, the defined *rss*
are generally better covered with transcripts than the predicted *ass* (at both donor and
acceptor sites). The mode of both distributions can be found on a class level of 1–20
transcripts, though the overall number of transcripts is approximately six times higher
in the respective class level of the reference dataset. Between score and transcript sup-
port we found a strong dependency in both *ass* and *rss* ($p < 10^{-4}$, χ^2-test) whereupon
better scoring splice sites clearly show a better transcript support.

Table 30.4. Splice site score characteristics in the *ass* and *rss* datasets. Underlined values em-
phasize the significant difference between mean values of *ass* and *rss* score distributions.

Splice site type	# Total counts	Median	Mean	SD	Variance
all donors	28,808	8.456	6.976	5.475	29.971
ass donors	3,705	5.263	0.862	10.219	104.438
rss donors	25,103	8.626	7.878	3.557	12.652
all acceptors	29,388	8.450	7.671	4.446	19.765
ass acceptors	3,862	5.997	3.356	8.035	64.562
rss acceptors	25,526	8.678	8.324	3.122	9.747

Table 30.5. Analysis of variances (F-test) and means (Wilcoxon test) between ESE motif frequencies (f_{ese}) distributed over *ass* and *rss* splice sites in different pre-mRNA contexts (see Fig. 30.2); μ = sample mean, m = sample median, σ^2 = sample variance; p-values significant at $\alpha \leq 0.05$ are marked with (*).

SRp	Region	*ass* $\mu/m/\sigma^2$	*rss* $\mu/m/\sigma^2$	*ass*↔*rss* $p(\mu)$	*ass*↔*rss* $p(\sigma^2)$
ASF/SF2	E3	4,148/4/5,862	3,903/4/5,192	$< 10^{-4*}$	$< 10^{-4*}$
	E5	4,187/4/5,382	3,894/4/5,950	$< 10^{-4*}$	$< 10^{-3*}$
	I3	2,910/3/4,268	2,440/2/4,194	$< 10^{-4*}$	$> 10^{-1}$
	I5	3,864/4/6,580	3,285/3/6,213	$< 10^{-4*}$	$< 5 \cdot 10^{-2*}$
SC35	E3	4,026/4/4,068	3,883/4/3,975	$< 10^{-4*}$	$> 10^{-1}$
	E5	4,076/4/3,949	3,842/4/3,941	$< 10^{-4*}$	$> 10^{-1}$
	I3	3,254/3/3,434	2,878/3/3,454	$< 10^{-4*}$	$> 10^{-1}$
	I5	3,914/4/4,477	3,578/3/4,558	$< 10^{-4*}$	$> 10^{-1}$
SRp40	E3	3,769/4/3,299	3,800/4/3,122	$> 10^{-1}$	$< 5 \cdot 10^{-2*}$
	E5	3,893/4/3,191	3,779/4/3,079	$< 10^{-3*}$	$> 10^{-1}$
	I3	3,081/3/2,583	2,784/3/2,501	$< 10^{-4*}$	$> 10^{-1}$
	I5	3,629/3/3,232	3,426/3/3,170	$< 10^{-4*}$	$> 10^{-1}$
SRp55	E3	2,334/2/2,253	2,346/2/2,281	$> 10^{-1}$	$> 10^{-1}$
	E5	2,347/2/2,327	2,332/2/2,222	$> 10^{-1}$	$< 10^{-2*}$
	I3	1,695/2/1,719	1,616/1/1,595	$< 10^{-3*}$	$< 10^{-2*}$
	I5	2,098/2/2,101	1,924/2/1,941	$< 10^{-4*}$	$< 10^{-2*}$

SRp Binding Motifs in *ass* and *rss* Splice Site Environments. Considering f_{ese} at exonic and intronic flanks of splice sites, one observes only subtle differences between *ass* and *rss*. In the case of the ASF/SF2 motif, f_{ese} at the exonic 5' and 3' regions has its mode at a frequency of three motifs (except for the exonic 5' end of *ass*). Nevertheless, there is a significant difference in means and variances between the number of ASF/SF2 motifs at the exonic part of *ass* and *rss* (Table 30.5). In contrast the intronic regions of *ass* and *rss* show generally a higher ASF/SF2 motif abundance. This is unexpected since this motif should—as an *exonic* splicing enhancer—occur more frequently in exonic regions. There is no significant difference in the variances of the f_{ese} distributions between intronic *ass* and *rss* acceptor sites, but here also the means vary significantly. Considering the size of the dataset, the results clearly indicate a tendency to more ASF/SF2 motifs in the flanking regions of predicted *ass* than in the set of *rss*. Table 30.5 summarizes the observed distributions of SRp specific motif fre-

quencies f_{ese} for *ass* and *rss*, compared between an exonic and intronic splice site context. The ASF/SF2 and SC35 motif frequencies show in the exonic flanks of 5′ and 3′ splice sites a tendency to more motifs in the neighborhood of the *ass* than at the *rss* splice sites. The same result was found for SRp40 motif frequencies in the exon 3′ flanks. SRp55 motifs appeared at exonic flanks of 5′ splice sites less but at 3′ more frequently at *ass* than at *rss*. The means of the f_{ese} distribution of SRp SC35 show a significant difference between the E3 and E5 sites, although the variances do not convey this information. Based on these tests we concluded that f_{ese} is different between mRNA regions around splice sites of *ass* and *rss*. While these motif modules might be individually subtle (e.g., between two splice sites) they appear to be significant on the whole dataset.

30.4 Summary

We have shown for a dataset of computationally predicted alternative splice sites (*ass*) how inherent information can be utilized to validate the predictions by applying statistics on different features typical for splice sites. These features were compared between a set of predicted *ass* and splice sites arising from a set of mRNAs that did not predict or have (by current experimental knowledge) *ass*. We refrain from the term "constitutive" and use "reference splice site" (*rss*) instead, since our analysis suggests that in spite of not predicting *ass*, the reference transcripts and their splice sites share similar characteristics to the alternative ones as the overlapping regions in the splice site score (S_{ss}) distributions (+5 to -10) demonstrate. However, in the low scoring region both *ass* and *rss* separate clearly with the *ass* exhibiting scores below zero more frequently at both donor and acceptor sites. Thus, a significant part of the predicted *ass* possess motifs incongruent to the consensus motif found for human GT..AG consensus splice sites. In fact, this observation could still be due to pseudo splice sites but, as our test of S_{ss} against the transcript support (f_t) indicates, there exists a clear dependency between the number of transcripts that utilize these splice sites and the pertinent score in both the *ass* and *rss*. As a promising splice site feature we investigated the binding motif frequencies (f_{ese}) of splice-enhancing SR proteins (SR_p) in the context of exonic and intronic splice site flanks compared between *ass* and *rss*. First we analyzed the donor/acceptor site specific occurrence of SRp motifs independent from the classification into *ass* and *rss*. For both the exonic and intronic flanks we found a higher variance for f_{ese} of the SRp ASF/SF2 at donor and acceptor sites, but between the intronic splice site flanks only we found on average more motifs at the donor site. For the other SRp motifs a similar trend to more binding motifs at intronic donor compared to acceptor flanks can be observed. This finding further suggests a higher presence of these motifs at the intronic flank of donor sites, a surprising effect since the motifs are detected by consensus sequences made as *exonic* splicing enhancers (ESEs). Nevertheless, since the *rss* make up the major fraction in these donor/acceptor-intronic /exonic datasets, this finding needed to be further investigated to derive conclusions on an effect that is present also in predicted *ass*. Hence, we continued to analyze differences in f_{ese} of exonic donor and acceptor at *ass* and *rss*. The ASF/SF2 motif occurs in a signif-

icant fraction of predicted alternative splice sites (exonic flanks) more frequently than in the set of reference splice sites. While the medians of both frequency distributions suggest four motifs at both alternative and reference splice site flanks, the mean and variance indicate that at both 5′ and 3′ sites a motif pattern deviating from the *rss* exists, with a tendency to at least one additional motif. Comparing the intron flanks, however, showed no difference in the variances of the ASF/SF2 motif frequencies between *ass* and *rss*. Diverging characteristics (equal variances versus different means) in the *ass* and *rss* specific motif frequencies of the three other types of SR proteins suggest that both datasets exhibit little variation around significantly different ESE patterns. It will be interesting to follow up investigations on modules of SRp binding motifs between predicted *ass* and *rss* to get a computational grip on possible multimerized enhancer (or silencer) protein complexes as well as on their role in a prediction pipeline of *ass*. Summarizing, it can be concluded that the combination of statistical investigations on different splice site related attributes offers a suitable approach to refinement and evaluation of predicted alternative splice variants.

Acknowledgment

References

1. Black, D.L.: Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev. Biochem., **72**, 291–336 (2003).
2. Boguski, M.S., Lowe, T.M., Tolstoshev, C.M.: dbEST–database for 'expressed sequence tags'. Nat. Genet., **4**, 332–3 (1993).
3. Bourgeois, C.F., Lejeune, F., Stevenin, J.: Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. Prog. Nucleic. Acid. Res. Mol. Biol., **78**, 37–88 (2004).
4. Cartegni, L., Chew, S.L., Krainer, A.R.: Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat. Rev. Genet., **3**, 285–298 (2002).
5. Cartegni, L., Wang, J., Zhu, Z., *et al*.: ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic. Acids. Res., **31**, 3568–3571 (2003).
6. Clark, F., Thanaraj, T.A.: Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. Hum. Mol. Genet., **11**, 451–464 (2002).
7. D'Souza, I., Schellenberg, G.D.: Determinants of 4-repeat tau expression. Coordination between enhancing and inhibitory splicing sequences for exon 10 inclusion. J. Biol. Chem., **275**, 17700–17709 (2000).
8. Elrick, L.L., Humphrey, M.B., Cooper, T.A., Berget, S.M.: A short sequence within two purine-rich enhancers determines 5′ splice site specificity. Mol. Cell. Biol., **18**, 343–352 (1998).
9. Fairbrother, W.G., Yeh, R.F., Sharp, P.A., Burge, C.B.: Predictive identification of exonic splicing enhancers in human genes. Science, **297**, 1007–1013 (2002).

10. Galperin, M.Y.: The molecular biology database collection: 2005 update. Nucleic. Acids. Res., **33**, D5–D24 (2005).
11. Graveley, B.R.: Sorting out the complexity of SR protein functions. Rna, **6**, 1197–1211 (2000).
12. Graveley, B.R.: Alternative splicing: increasing diversity in the proteomic world. Trends. Genet., **17**, 100–107 (2001).
13. Hertel, K.J., Graveley, B.R.: RS domains contact the pre-mRNA throughout spliceosome assembly. Trends. Biochem. Sci., **30**, 115–118 (2005).
14. Hertel, K.J., Maniatis, T.: The function of multisite splicing enhancers. Mol. Cell., **1**, 449–455 (1998).
15. Hiller, M., Huse, K., Szafranski, K., et al.: Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. Nat. Genet., **36**, 1255–1257 (2004).
16. Humphrey, M.B., Bryan, J., Cooper, T.A., Berget, S.M.: A 32-nucleotide exon-splicing enhancer regulates usage of competing 5′ splice sites in a differential internal exon. Mol. Cell. Biol., **15**, 3979–3988 (1995).
17. Konig, H., Ponta, H., Herrlich, P.: Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. Embo. J., **17**, 2904–2913 (1998).
18. Lander, E.S., et al.: Initial sequencing and analysis of the human genome. Nature, **409**, 860–921 (2001).
19. Liu, H.X., Zhang, M., Krainer, A.R.: Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev., **12**, 1998–2012 (1998).
20. Lou, H., Neugebauer, K.M., Gagel, R.F., Berget, S.M.: Regulation of alternative polyadenylation by U1 snRNPs and SRp20. Mol. Cell. Biol., **18**, 4977–4985 (1998).
21. Maniatis, T., Tasic, B.: Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature, **418**, 236–243 (2002).
22. Modrek, B., Resch, A., Grasso, C., Lee, C.: Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic. Acids Res., **29**, 2850–2859 (2001).
23. Pospisil, H., Herrmann, A., Bortfeldt, R.H., Reich, J.G.: EASED: Extended Alternatively Spliced EST Database. Nucleic. Acids Res., **32**, D70–D74 (2004).
24. Ramchatesingh, J., Zahler, A.M., Neugebauer, K.M., et al.: A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. Mol. Cell. Biol., **15**, 4898–4907 (1995).
25. Shen, H., Green, M.R.: A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. Mol. Cell., **16**, 363–373 (2004).
26. Smith, C.W., Valcarcel, J.: Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem. Sci., **25**, 381–388 (2000).
27. Tacke, R., Chen, Y., Manley, J.L.: Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. Proc. Natl. Acad. Sci. USA, **94**, 1148–1153 (1997).
28. Tacke, R., Manley, J.L.: The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. Embo. J., **14**, 3540–3551 (1995).
29. Venables, J.P.: Aberrant and alternative splicing in cancer. Cancer Res., **64**, 7647–7654 (2004).
30. Yeo, G., Burge, C.B.: Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. RECOMB'03 April 10–13 Berlin, Germany (2003).
31. Zhang, M.Q.: Statistical features of human exons and their flanking regions. Hum. Mol. Genet., **7**, 919–932 (1998).

Gaussian Mixture Decomposition of Time-Course DNA Microarray Data

Joanna Polańska,[1,3] Piotr Widłak,[2] Joanna Rzeszowska-Wolny,[1,2] Marek Kimmel,[1,3] and Andrzej Polański[1]

[1] The Silesian University of Technology, Gliwice, Poland; Joanna.Polanska@polsl.pl
[2] Center of Oncology, Gliwice, Poland; widlak@io.gliwice.pl
[3] Rice University, Houston, TX, USA; kimmel@rice.edu

Summary. In this chapter we present the decomposition approach to the analysis of large gene expression profile data sets. We address the problem of analysis of transient time-course data of expression profiles. We accept the assumption that co-expression of genes can be related to their belonging to the same Gaussian component. We assume that parameters of Gaussian components, means and variances, can differ between time instants. However, the gene composition of components is unchanged between time instants. For such problem formulations we derive the appropriate version of expectation-maximization algorithm recursions for the estimation of model parameters. We apply the derived method to the data on gene expression profiles of human K562 erythroleukemic cells and we discuss the obtained gene clustering.

Key words: Gene expression profiles, maximum likelihood, EM method, Gaussian components.

31.1 Introduction

An approach that has the potential to support analyzing DNA microarray data is mixture modeling of the probability distribution of gene expression levels. This approach is based on the empirical observation that probability density functions of gene expressions can be represented by using mixtures of distributions and uses the hypothesis that parameters of approximating mixture distributions contain useful information. Several researches in this field have appeared in the literature [10], [4], [13], [12], [8], [7], [14]. As demonstrated in the referenced papers, approximating the probability density functions (pdfs) of logarithms of expression levels or fluorescence intensities of probes by mixtures of distributions can be used to solve several issues in the interpretation of DNA microarray data. Reference [10] surveys methods for obtaining parameters of mixtures of different distributions and proposes to use mixtures of factor analyzers for unsupervised classification of colon and leukemia gene expression data sets. In [4] the decomposition of expression level probability density functions into Gaussian components is used to set thresholds to classify expression levels as "change," "no change," "overexpressed," "underexpressed" etc. In the papers [13] and [12] different variants

of Bayesian mixture based clustering procedures were studied. Methods for estimating mixture parameters involved Gibbs sampling in conjunction with an infinite mixture technique. Arguments were given that the proposed approach has a superior ability to detect the number of components over classical, information criteria based methods. The elaborated methods were applied to yeast cell-cycle data obtained with the use of DNA microarrays. In [8] a hierarchical agglomerative clustering method was used for an initial guess of mixture parameters and the mixture model was combined with dimensionality reduction technique in the analysis of cutaneous melanoma data. In [7] a mixture model was used for determining differential expression of genes in the presence of mixed cell populations. In [14] mixture modeling was applied to the problem of missing measurements in DNA microarrays.

In this chapter we address the problem of analysis of transient time-course data of expression profiles. Like others, we accept the assumption that co-expression of genes can be related to their belonging to the same component. We assume that parameters of Gaussian components, means and variances, can differ between time instants. However, the gene composition of components is unchanged between time instants. For such problem formulations we derive the appropriate version of the expectation-maximization (EM) algorithm recursions for estimation of model parameters. We apply the derived method to the data on gene expression profiles of human K562 erythroleukemic cells and we discuss the obtained gene clustering.

31.2 Modeling Probability Distributions of Gene Expressions

We use the standard paradigm that information on the process behind the DNA microarray experiment is encoded in the ratios of RNA concentrations between different experiments. The researcher is interested in by how many fold the RNA concentration has increased or decreased from one measurement to the other. Consequently, we analyze DNA microarray fluorescence levels and we denote by x the base two logarithm of the normalized fluorescence signal. We assume that this signal is a realization of a random variable, which we denote by X.

31.2.1 One Microarray Chip

We assume that the probability distribution of the random variable X is described by a normal mixture model,

$$f(x) = \sum_{k=1}^{K} \alpha_k f_k(x). \tag{31.1}$$

In the above expression x denotes base two logarithm of expression level, $f(x)$ is its probability density function, α_k, $k = 1, 2, \ldots, K$ are weighting coefficients $\sum_{k=1}^{K} \alpha_k = 1$, and $f_k(x)$, $k = 1, 2, \ldots, K$ are pdfs of normal components of the mixture $f(x)$, i.e.,

$$f_k(x) = \frac{1}{\sqrt{2\pi}\,\sigma_k} \exp\left[-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right], \tag{31.2}$$

where μ_k is the expectation and σ_k^2 the variance of the kth normal component. The parameters to adjust are number of components, expectations and variances of each component and weighting coefficients.

We denote the number of gene probes at a DNA microarray chip by N. We treat genes as N independent realizations of random variable X with distribution (31.1). With N independent measurements, the parameters of (31.1) can be estimated by the maximum likelihood method, where the likelihood function is given by

$$L(x_1, \dots, x_N) = \prod_{n=1}^{N} \sum_{k=1}^{K} \alpha_k f_k(x_n), \tag{31.3}$$

where x_n stands for the logarithm of the expression level of the nth gene. For distributions with multiple components there are no analytical expressions for estimates of parameters, but efficient numerical approaches were elaborated based on the EM algorithm [5], [3], [11] or on the use of versions of the Metropolis–Hastings sampling algorithm [15]. Iterations of the EM algorithm for updating estimates $\alpha_1, \dots, \alpha_K$ and $\mu_k, \sigma_k, k = 1, 2, \dots, K$ assume the following form [11]:

$$\alpha_k^{\text{new}} = \frac{\sum_{n=1}^{N} p(k \mid x_n, p^{\text{old}})}{N}, \tag{31.4}$$

where

$$p(k \mid x_n, p^{\text{old}}) = \frac{\alpha_k^{\text{old}} f_k(x_n, p^{\text{old}})}{\sum_{\kappa=1}^{K} \alpha_\kappa^{\text{old}} f_\kappa(x_n, p^{\text{old}})}, \tag{31.5}$$

and p stands for the vector of all parameters, $p = [\mu_1, \dots, \mu_K, \sigma_1, \dots, \sigma_K]$. The expression (31.5) gives the conditional probability distribution of latent variable k given data x_n and given the estimate of parameters x_n. For location and scale parameters we have

$$\mu_k^{\text{new}} = \frac{\sum_{n=1}^{N} x_n p(k \mid x_n, p^{\text{old}})}{\sum_{n=1}^{N} p(k \mid x_n, p^{\text{old}})}, \qquad k = 1, 2, \dots, K, \tag{31.6}$$

and

$$(\sigma_k^{\text{new}})^2 = \frac{\sum_{n=1}^{N} (x_n - \mu_k^{\text{new}})^2 p(k \mid x_n, p^{\text{old}})}{\sum_{n=1}^{N} p(k \mid x_n, p^{\text{old}})}, \qquad k = 1, 2, \dots, K. \tag{31.7}$$

Superscripts "old" and "new" are added to denote two successive iterates. Expressions (31.6)–(31.7) follow from performing the M step of the EM algorithm involving maximization of expectation of the complete data likelihood function.

31.2.2 Time-Course Microarray Data

A probability density mixture model of the form (31.3) incorporates information about a single experiment with one DNA microarray chip. As stated in the introduction, our aim is the analysis of experiments with the use of DNA microarrays, where gene expression profiles are samples at time instants $t_1, t_2, \ldots, t_M$. We assume that at each time instant the gene expression profile can be represented as a mixture of components. The number of components as well as the composition (by composition we mean the assignment of genes to components) remain constant over time. However, Gaussian components are changing over time in the sense that their expectations and variances differ at different time instants.

The maximum likelihood approach can be modified to account for time-course microarray data. Let us denote by x_n^m the measurement of the logarithm of the expression of the nth gene, at time instant t_m. The likelihood function assumes the form

$$L(x_1^1, \ldots, x_1^M, \ldots, x_N^1, \ldots, x_N^M) = \prod_{n=1}^{N} \sum_{k=1}^{K} \alpha_k \prod_{m=1}^{M} f_{k,m}(x_n^m) \tag{31.8}$$

with $f_{k,m}(x_n^m)$ standing for the Gaussian probability density function of the kth component at the mth time instant,

$$f_{k,m}(x) = \frac{1}{\sqrt{2\pi}\sigma_{k,m}} \exp\left[-\frac{(x - \mu_{k,m})^2}{2\sigma_{k,m}^2}\right]. \tag{31.9}$$

The appropriate version of the EM algorithm for maximum likelihood function (31.8) can be obtained either by using (31.8) in the definitions of the E and M steps of the EM algorithm [11], or by replacing a univariate distribution $f_k(x)$ by a multidimensional (M-dimensional) normal probability density function in (31.4)–(31.7). The resulting EM iterations are as follows:

$$\alpha_k^{\text{new}} = \frac{\sum_{n=1}^{N} p(k \mid x_n^1, \ldots, x_n^M, p^{\text{old}})}{N}, \tag{31.10}$$

where

$$p(k \mid x_n^1, \ldots, x_n^M, p^{\text{old}}) = \frac{\alpha_k^{\text{old}} \prod_{m=1}^{M} f_{k,m}(x_n^m, p^{\text{old}})}{\sum_{\kappa=1}^{K} \alpha_\kappa^{\text{old}} \prod_{m=1}^{M} f_{\kappa,m}(x_n^m, p^{\text{old}})} \tag{31.11}$$

and p stands for the vector of all parameters, $p = [\mu_{1,1}, \ldots, \mu_{1,M}, \ldots, \mu_{K,1}, \ldots, \mu_{K,M}, \sigma_{1,1}, \ldots, \sigma_{1,M}, \ldots, \sigma_{K,1}, \ldots, \sigma_{K,M}]$. For location and scale parameters we then have

$$\mu_{k,m}^{\text{new}} = \frac{\sum_{n=1}^{N} x_n^m p(k \mid x_n^1, \ldots, x_n^M, p^{\text{old}})}{\sum_{n=1}^{N} p(k \mid x_n^1, \ldots, x_n^M, p^{\text{old}})}, \qquad k = 1, 2, \ldots, K, \tag{31.12}$$

and

$$(\sigma_{k,m}^{\text{new}})^2 = \frac{\sum_{n=1}^{N}(x_n - \mu_k^{\text{new}})^2 p(k \mid x_n^1, \dots, x_n^M, p^{\text{old}})}{\sum_{n=1}^{N} p(k \mid x_n^1, \dots, x_n^M, p^{\text{old}})}, \tag{31.13}$$

$$k = 1, 2, \dots, K, \ m = 1, \dots, M.$$

Again, superscripts "old" and "new" are added to denote two successive iterates.

31.2.3 Estimating Number of Components by Bayesian Information Criterion

An issue to solve in numerical computations is how many components K should appear in (31.1), (31.3), (31.8). If K is too large, the model becomes overparametrized and some components are unnecessary and non-informative. If K is too small, several different components may become merged into one. Detection of model overparametrization can be done by using information criteria. In order to estimate the number of parameters we used the Bayesian information index (BIC) [2]. Using the BIC can be understood as penalizing the number of parameters used in the model by subtracting the term weighted by the logarithm of the number of observations from the likelihood function. Since the number of parameters is $3K - 1$, then the BIC-corrected likelihood (31.8) becomes

$$L^{BIC}(x_1^1, \dots, x_1^M, \dots, x_N^1, \dots, x_N^M)$$

$$= L(x_1^1, \dots, x_1^M, \dots, x_N^1, \dots, x_N^M) - \frac{1}{2}(3K - 1)\ln(N * M). \tag{31.14}$$

31.3 Experimental Data

We applied a mixture decomposition model to the data [9] on cellular responses of human leukemic cells K562 to X-ray irradiation. Affymetrix U133A DNA microarrays with 22,000 human gene probes were used to record gene expression profiles at 5 time instants at and after irradiation, $t_1 = 0$ h, $t_2 = 12$ h, $t_3 = 24$ h and $t_4 = 36$ h. For this data we applied our method of decomposing the pdf into Gaussian components.

The EM method is known for sticking to local maxima. Therefore, we have repeatedly launched EM (31.10)–(31.13) starting from random initial guesses of parameters (10 repeats were applied for each K). We also used a method of bisection to compute the maximal BIC-corrected likelihood.

As the result we obtained 45 normal components. When using the rule "gene n belongs to component k^*" if

$$\arg \max_k p(k \mid x_n^1, \dots, x_n^M, p) = k^*, \tag{31.15}$$

then the numbers of genes that belong to components 1–45 range from 25 to 643.

31.3.1 Analysis of Mixture Decomposition with the Use of Gene Ontology (GO) Terms

The decomposition of a pdf into normal components is based on the hypothesis that genes with both similar values of expressions and similar patterns of expression changes should be classified to the same component. However, there is a lot of randomness in the data and the assumption that all 45 components can be treated as informative is rather naive. Therefore, components should be analyzed from the viewpoint of their possible relation to processes and signaling pathways.

Here we will analyze only the basic issue of non-randomness of the decomposition into components. We verify whether comparisons of gene ontology (GO) terms [6] of genes belonging to the obtained 45 components will be statistically significantly different from comparisons of genes belonging to components drawn randomly. We have decided to use a somewhat indirect approach involving pairwise comparisons of components. The reasons for that were (i) comparisons can be easily made with the use of internet GO browsers, and (ii) they provide enough statistical evidence for randomness or non-randomness of the gene composition of obtained Gaussian components. Consequently, in addition to the decomposition resulting from the maximum likelihood, we have assigned genes into 45 random components. We use the notation *EM components* for those resulting from maximum likelihood with the use of the EM algorithm and *random components* for those drawn randomly. The numbers of genes in random components were set equal to the numbers of genes in EM components. Then we drew randomly 80 pairs $(i_1, j_1), \ldots, (i_{80}, j_{80})$ and we performed parallel comparisons of EM component i_k with EM component j_k and of random component i_k with random component j_k, $k = 1, 2, \ldots, 80$. We used the GO browser FatiGO [1] to perform comparisons. Comparisons of gene contents of components are possible in three categories: biological process, cellular component and molecular process. For each of 80 pairs of comparisons we have also chosen randomly one of these three mentioned categories. By statistical comparison we specifically mean statistical test, where the null hypothesis is that two sets of genes do not differ in biological functions, molecular functions or cellular components. The results of comparisons can be quantitatively described in terms of the p-value of statistical tests using statistics based on a multinomial distribution. In Fig. 31.1 we show comparison histograms of the p-values. The upper plot presents the histograms resulting from comparing genes in EM components while the lower plot presents the histograms resulting from comparing genes in random components.

As seen from the figure, the p-values corresponding to comparisons of EM components are very significantly shifted towards the left. When comparing histograms corresponding to EM and random components using the non-parametric test of Smirnov and Kolmogorov, the null hypothesis of no shift between distributions can be rejected at a significance level lower than 10^{-10}.
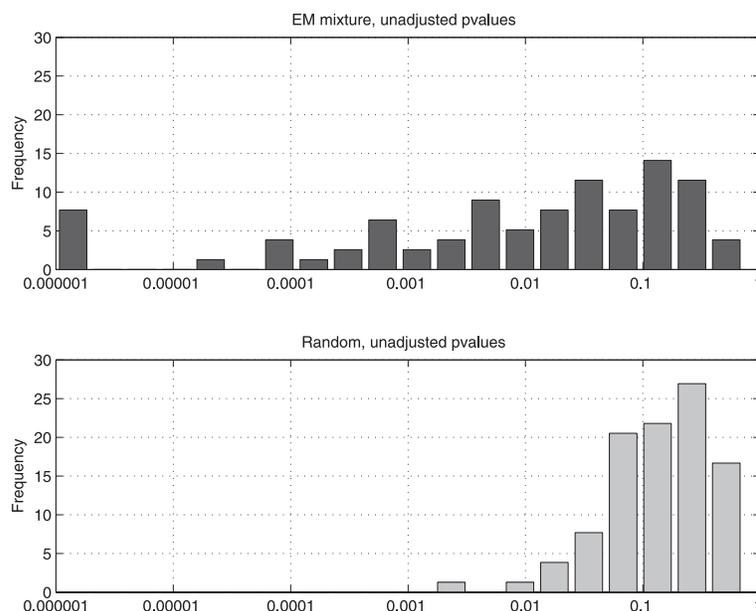
Fig. 31.1. Histograms of frequencies versus p-values for statistical tests of composition comparisons of two components of genes. Upper plot is for EM components and lower plot is for random components.

31.4 Conclusion

The analysis of DNA microarray data most often involves supervised or unsupervised classification or clustering. Due to the special structure of gene expression profiles, where the number of genes greatly exceeds the number of experiments or observations, classification and clustering are often combined with dimensionality reduction and principal component analysis techniques. Decomposition into probability distribution mixture components, discussed here, can also be viewed as a type of clustering technique since it leads to grouping of genes into classes defined by components composition. The advantages of the approach using mixture modeling are that the decomposition follows from maximization of the likelihood function and that there is a reasonable method of estimating the number of components.

In the case of time-course data, we can additionally analyze the time evolution of parameters of Gaussian components μ and σ, which can improve our understanding of the data. The analysis of experimental data indicates that the obtained decomposition into time-evolving Gaussian components is highly non-random. Therefore, by further analysis of the obtained components one can get some insight into the molecular processes behind the observed patterns of change in the gene expression profiles. We do not show detailed results due to space limitations, but some of the obtained components have nice biological interpretations resulting from their gene composition in conjunction with the pattern of time evolution of their expression levels.

The imposed restriction that gene composition of Gaussian components must remain unchanged between time instants is rather natural and allows interpretation of time evolution of clusters of genes in terms of biological or molecular processes. A Gaussian mixture decomposition of expression profiles of individual time points would include significantly lower numbers of components, ranging from 3 to 5. Each of these components would include a large number of genes and the structure of their GO composition would be less specific. Since both the numbers and gene composition of components change in time, they cannot be interpreted as being related to biological processes.

Acknowledgments

References

1. Al-Shahrour, F., Diaz-Uriarte, R., Dopazo, J.: FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes, Bioinformatics, **20**, 578–580 (2004).
2. Bernardo, J., Smith, A.: Bayesian Theory. Wiley, New York (1994).
3. Bilmes, J.: A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report, University of Berkeley, ICSI-TR-97-021, (1997). http://citeseer.ist.psu.edu/bilmes98gentle.html
4. Broet, P., Richardson, S., Radvanyi, F.: Bayesian hierarchical model for identifying changes in gene expression from microarray experiments. J. Comp. Biol. **9**, 671–683 (2002).
5. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statistics Soc., Ser. B., **39**, 1–38 (1977).
6. Gene Ontology Consortium, http://www.geneontology.org
7. Ghosh, D.: Mixture models for assessing differential expression in complex tissues using microarray data. Bioinformatics, **20**, 1663–1669 (2004).
8. Ghosh, D., Chinnaiyan. A.M.: Mixture modelling of gene expression data from microarray experiments. Bioinformatics, **18**, 275–286 (2002).
9. Konopacka, M., Rogolinski, J., Herok, R., Polanska, J., Fujarewicz, K., Hancock, R., Rzeszowska-Wolny, J.: Radiation induced bystander effects in human leukemia K562 cells in vitro. Acta Biochim. Pol., **50**(S1), 169–170 (2003).
10. McLachan, G.J., Bean, R.W., Peel, D.: A mixture model-based approach to the clustering of microarray expression data., Bioinformatics, **18**, 413–422 (2002).
11. McLachan, G.J., Peel, W.: Finite Mixture Distributions. Wiley, New York (2000).
12. Medvedovic, M., Sivaganesan, S.: Bayesian infinite mixture model based clustering of gene expression profiles. Bioinformatics, **18**, 1194–1206 (2002).
13. Medvedovic, M., Yeung, K.Y., Bumgarner, R.E.: Bayesian mixtures for clustering replicated microarray data. Bioinformatics, **20**, 1222–1232 (2004).

14. Ouyang, M., Welsh, W.J., Georgopoulos, P.: Gaussian mixture clustering and imputation of microarray data. Bioinformatics, **20**, 917–923 (2004).
15. Richardson, S., Green, P.J.: On Bayesian analysis of mixtures with an unknown number of components. J. Royal Statistics Soc., Ser. B., **59**, 731–792 (1997).

SVD Analysis of Gene Expression Data

Krzysztof Simek[1] and Michał Jarząb[2]

[1] Institute of Automatic Control, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland; krzysztof.simek@polsl.pl

[2] Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch, Wybrzeze Armii Krajowej 15, 44-100 Gliwice, Poland; mjarzab@io.gliwice.pl

Summary. The analysis of gene expression profiles of cells and tissues, performed by DNA microarray technology, strongly relies on proper bioinformatical methods of data analysis. Due to the large number of analyzed variables (genes) and the usually low number of cases (arrays) in one experiment, limited by the high cost of the technology, the biological reasoning is difficult without previous analysis, leading to a reduction of the problem dimensionality. A wide variety of methods have been developed; the most useful, from a biological point of view, are methods of supervised gene selection with estimation of false discovery rate. However, supervised gene selection is not always satisfying for the user of microarray technology, as the complexity of biological systems analyzed by microarrays rarely can be explained by one variable. Among unsupervised methods of analysis, hierarchical clustering and principal component analysis (PCA) have gained wide biological application. In our opinion, singular value decomposition (SVD) analysis, which is similar to PCA, has additional advantages that are very essential for the interpretation of the biological data. In this chapter we shall present how to apply SVD to unsupervised analysis of transcriptome data obtained by oligonucleotide microarrays. These results have been derived from several experiments, carried out at the DNA oligonucleotide microarray Laboratory at the Institute of Oncology, Gliwice, and are currently analyzed from a biological point of view.

Key words: Singular value decomposition, gene expression data, gene selection, hierarchical clustering.

32.1 Unsupervised Methods in Analysis of Gene Expression Data

Unsupervised analysis should be the first step in all microarray experiments. It reveals the intrinsic structure of the data, which helps to verify whether the assumptions of the microarray experiment are held and whether the major observed variability is related to the experimental variables or to confounding factors. Moreover, it allows for rapid detection of outliers, thus it is a valuable method of quality control. Last, but not least, unsupervised analysis is invaluable in genome-wide experiments, in which we aim to classify the samples based on their gene expression profiles and in this way to gain biological knowledge about subgroups of these samples. These three ap-

proaches are discussed later in this chapter. Various methods of unsupervised analysis have been developed. Initially, they were based on known statistical methods of clustering. At the moment, a widely accepted method is hierarchical clustering with various distance definitions and metrics. Other algorithms, which were used mainly to cluster genes and were not efficient in clustering of a limited number of microarray samples, are less widespread and are used only for specific occasions. Hierarchical clustering is a very useful method of analysis and visualization (by dendrograms) of differences/similarities between samples or genes. It allows one to calculate the distance between samples and divides the whole group into a chosen number of clusters. Moreover, being similar to methods used in taxonomy, dendrograms are easily understood by biologists. Dendrogram classification of genes and samples accompanied by a heatmap plot is probably one of the most informative methods of microarray data visualization.

However, when used in unsupervised analysis, hierarchical clustering has serious drawbacks. To be a fully unsupervised method, it should be used on all microarray genes or on genes filtered without the use of any variables in question. Clustering by all genes is a method which allows one to detect large differences in gene expression profile (e.g., it can separate expression profiles of different tissues), but when used on less variable experimental data, it will rather depend on technical parameters of arrays. This way it is more suitable for visualization of technical differences, but it has no place in routine use to answer biological questions. This is partly caused by the large number of genes on some arrays (more than 50K transcripts on routinely used oligonucleotide microarrays), from which only a small proportion is expected to be expressed. Filtering of genes by different variance measures may be useful in reducing the number of genes for clustering, but taking into account that some coordinated biological changes show a rather low amplitude, this strategy does not seem to be optimal. Moreover, in routine hierarchical clustering of normalized data every gene has equal weight, and thus some (e.g., not expressed) genes may bias the final result. At present there is no widely accepted method to select just the important transcripts in a gene expression profile. The use of singular value decomposition (SVD) for analysis of microarray data has been detailed in [1]. A profound description of SVD and a comprehensive survey of its applications in gene expression data analysis are given in [2] and references therein. SVD is a standard method of linear algebra and it may be easily performed on large matrices without significant computational cost. The most important feature of SVD which predisposes it to be used for the analysis of microarray data is that the characteristic modes obtained from decomposition of a gene expression matrix of various samples usually have a meaningful biological interpretation. In a homogeneous biological system, like in *in vitro* cell culture, the majority of genes are coordinately regulated by a limited number of signals, thus they exhibit similar expression profiles. Obviously, the complexity of this regulation is large and many genes are affected by numerous transcription factors, but it has been shown in yeast that common gene expression patterns in a cell cycle may be easily detected by SVD [3]. In experiments on cells from cultures the diversity of gene expression depends on the complexity of each cell transcriptome and the differences in the cellular cycle stage. However, microarray experiments are very often performed on even more complex biological systems, i.e.,

tissue samples. In tissues, the inherent feature is the heterogeneity of cells within the sample and until the microdissection techniques (allowing one to dissect the tissue into cells of various morphology) become widespread, the expression profile of the tissue depends on the transcriptional changes in cells and the differences in cellular content of the tissue. As we will show later, this second factor very often prevails over the whole gene expression profile and the analysis of characteristic modes then helps to interpret their biological meaning. However, the shape of the modes rarely provides us with direct conclusions, and thus we extend the method to select the genes correlated to each characteristic mode. Such an unsupervised gene selection method, when used together with hierarchical clustering, is a strong and powerful tool for biological analysis of microarray data. We briefly describe the principles of our method and show some examples of its application.

32.2 Singular Value Decomposition to Select Major Variability Genes in Transcriptome Data

32.2.1 Definition of SVD

The SVD of any $n \times m$ matrix A (gene expression matrix) has the form [4]

$$A = U S V^{\mathrm{T}}, \tag{32.1}$$

where U is an $n \times m$ orthonormal matrix, whose columns are called the left singular vectors of A (gene coefficient vectors), and V is an $m \times m$ orthonormal matrix, whose columns are called the right singular vectors of A (expression level vectors). S is a diagonal matrix $S = \mathrm{diag}(s_1, s_2, \dots, s_m)$. The diagonal elements of matrix S are, as a convention, listed in a descending order, $s_1 \geq s_2 \geq \cdots \geq s_m \geq 0$, and are called the singular values of A.

32.2.2 Characteristic Modes

Let us denote the rows of matrix SV^{T} by X_i, $i = 1, \dots, m$. The orthogonal vectors $X_i = s_i v_i^{\mathrm{T}}$ are called the *gene characteristic modes* associated with gene expression matrix A. In an analogous way, the rows of matrix SU^{T}, can be defined as *array characteristic modes*. The properties of both types of modes are similar so we present them only for the gene modes.

The profile of the jth gene, included in the row A_j of matrix A, can be obtained as linear combinations of the characteristic modes. The coefficients of the combination are the corresponding entries of matrix U,

$$A_j = \sum_{i=1}^{m} U_{ji} X_i. \tag{32.2}$$

The contribution of modes to the gene pattern decreases from the higher order to the lower order modes. Usually not all characteristic modes are needed to reconstruct gene expression patterns with a reasonable accuracy . We may use a truncated expression,

$$A_j = \sum_{i=1}^{l} U_{ji} X_i, \qquad l < m. \tag{32.3}$$

There are several heuristic methods to estimate the number l of the most significant characteristic modes [5]. One of the simplest is to retain just enough modes to capture a large percentage of the overall expression. Usually values of 70–90% are proposed. The other procedure is to exclude characteristic modes such that the fraction of expression p_i they capture is less than $(70/m)\%$. Another method is the examination of scree plots for s_i^2 or $\log s_i^2$. Using this method one can usually find a natural border between significant and insignificant singular values (called the elbow). The singular values which represent the magnitudes of the corresponding modes can be used as measures of the relative significance of each characteristic mode in terms of the fraction of overall expression that it captures:

$$p_i = \frac{s_i^2}{\sum_{j=1}^{m} s_j^2}, \qquad i = 1, \ldots, tgm. \tag{32.4}$$

A similar index, defining the contribution of the ith mode to the pattern of the kth gene, can be defined in the form

$$c_k^i = \frac{(U_{ki} s_i)^2}{\sum_{j=1}^{m} (U_{kj} s_j)^2}. \tag{32.5}$$

32.2.3 Gene Selection Using SVD

In the clustering literature, SVD is sometimes applied to extract the cluster structure in the data and reduce its dimensionality prior to clustering. Since characteristic modes are uncorrelated and ordered, the first few most significant ones, which contain most of the variations in the data, are usually used. Namely, characteristic mode coefficients (gene coefficient vectors), instead of original variables, are used for clustering. Our approach differs from that known from the literature. We apply SVD as a preprocessing step before cluster analysis of gene expression data. As a result, a small set of original genes is selected and then applied to cluster samples using one of the standard algorithms.

Algorithm of gene selection. The gene selection algorithm inspects gene coefficient vectors (columns of matrix U) corresponding to the set of the most significant characteristic modes. Each coefficient is compared to the threshold value [6], whose meaning is similar to a 3σ statistical significance cutoff, equal to $Wn^{-1/2}$, where n is the number of genes and W is a weight factor whose recommended value is greater than 3. If the magnitude of the element is greater than the threshold value, the corresponding

gene is selected to the clustering set. In practice we choose genes with sufficiently big coefficients for the most important characteristic modes, or in other words, genes for which values of index (32.5) for the most important modes are big enough. In the result we obtain a set of genes with patterns "similar" to the dominant modes.

32.3 Applications of Unsupervised Singular Value Decomposition Method to Microarray Data

We applied the SVD algorithm to four datasets of gene expression profiles related to cancer. The first data set was obtained from a cell culture experiment with murine melanoma cells exposed to hypoxic conditions (low oxygen tension) [8]. In this experiment oligonucleotide microarrays were used (Affymetrix MG-U74). Hypoxia was obtained by three different methods: chemical mimicry with cesium chloride, a Billups–Rothenberg chamber and an incubator with regulated O_2 tension, set to low O_2 level. Three other experiments were carried out with clinical material: tumor samples and corresponding normal tissues analyzed by gene expression profiling. In the first experiment we compared gene expression of papillary thyroid cancer to normal thyroid tissue from the same individual. The analyzed dataset consisted of 16 tumor and 16 normal samples hybridized to HG-U133A oligonucleotide microarrays and preprocessed by MAS5 algorithm [7]. In the second experiment we compared two different histological types of thyroid cancer. We performed SVD on a set of 57 thyroid neoplasms: 38 of papillary histology and 19 follicular adenomas or carcinomas [9]. We also compared gene expression profiles of laryngeal cancer and corresponding normal tissue and analyzed by HG-U133 Plus 2 microarray 17 samples of laryngeal cancer (9 microarrays) and normal tissue (8 samples) [10].

32.3.1 The Influence of Hypoxia on Gene Expression Profile

Hypoxia is an intrinsic feature of many malignant tumors and is highly related to their resistance to various treatment modalities (chemotherapy, radiation therapy). Thus, a sound understanding of the molecular mechanisms underlying hypoxia is crucial to the development of new molecular methods of therapy. In this experiment we obtained 13 microarrays from hypoxic cells and 8 from control samples. We performed SVD of the whole dataset and obtained 20 gene characteristic modes with the first mode accounting for 39.7% of variance and the first 5 modes describing 83% of the data variability (Fig. 32.1). This result was highly satisfying (more than 80% of the variability explained by only 5 expression patterns), but simultaneously showed that confounding factors influence the effect of hypoxia. We selected 154 genes correlated to the first mode profile. Hierarchical clustering of all samples based on these 154 genes (Fig. 32.2) ideally divided control and hypoxic samples which confirmed that the major source of variability in the analyzed dataset was the hypoxic-control difference. This was the proof that genes selected on the basis of this comparison are not strongly affected by technical factors. An important fact in understanding the SVD method was that the first mode profile itself was less clearly differentiating hypoxic and control
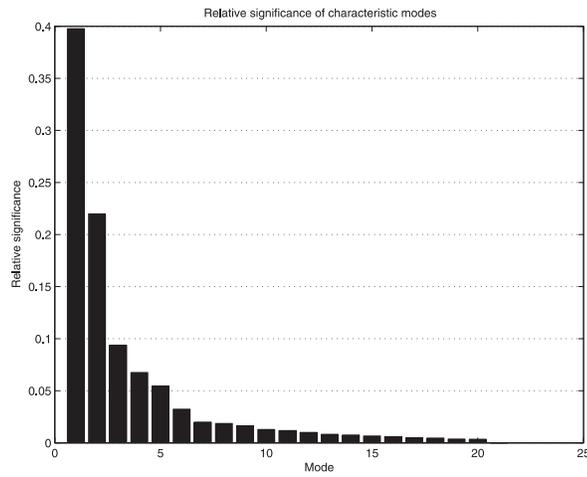
Fig. 32.1. Relative significance of gene characteristic modes in hypoxia experiment.
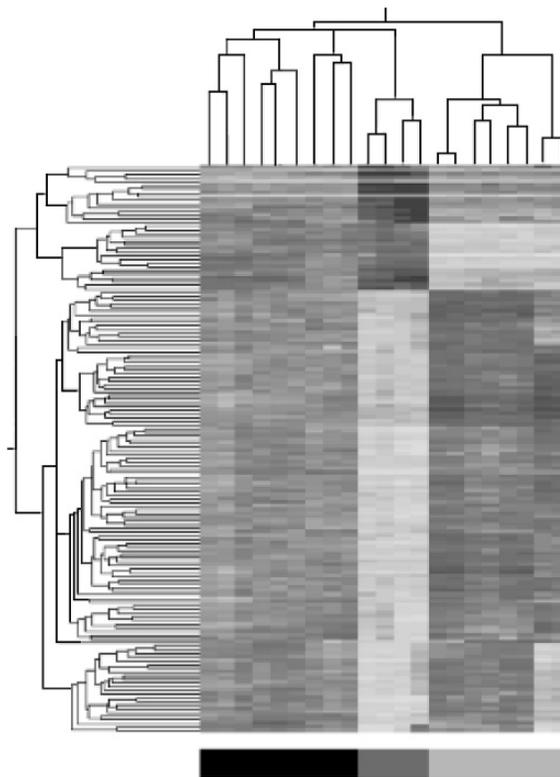
Fig. 32.2. Hierarchical clustering of first mode genes in hypoxia study. Real hypoxia (black) and hypoxia mimicry (dark grey) is distinctly different in expression profile from control samples (light grey).
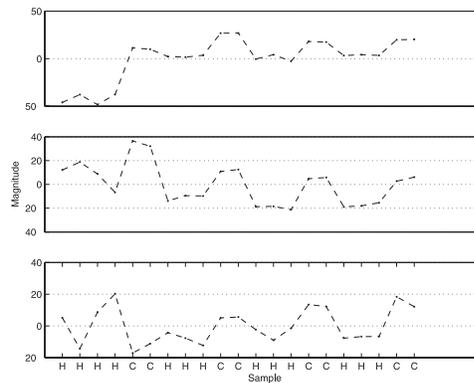
Fig. 32.3. First three gene characteristic modes in hypoxia study. Difference between hypoxic (H) and control (C) samples can be shown in all three modes, but it is influenced by other sources of variability (CoCl$_2$-treated specimens, first four from the left, are clearly different from other hypoxic samples).

samples. The shape of this pattern was influenced not only by the hypoxic-normal difference, but also by other factors (Fig. 32.3). Thus, direct analysis of mode patterns was not fully justified and we preferred to use gene selection. This approach is very powerful, not only because it allows one to properly classify samples, but because it also gives a number of genes related to each expression pattern. The gene content of this list can facilitate the biological reasoning, as shown later. In summary, SVD analysis allowed us to confirm that the control-treated difference was the major source of variability in the performed experiment, detected the important changes between two hypoxic conditions and thus influenced the further supervised analysis [8].

32.3.2 Gene Expression Profile of Papillary Thyroid Cancer

The next application of the SVD method is the analysis of the gene expression profile in papillary thyroid carcinoma. This malignant tumor is the most common cancer of the thyroid gland [7]. The first three characteristic modes in this dataset, which were considered significant, accounted for 40.4% of the variance in this dataset (Fig. 32.4). The much lower percentage of variability described by SVD analysis in comparison to the data described above is characteristic for the study of clinical specimens, where the samples differ not only by one variable (like hypoxia in the previous example), but by a whole set of features related to patients and the disease. Nevertheless, the most important factor of variability, revealed by the first mode, was the difference between normal thyroid and papillary thyroid cancer—all samples, clustered by 310 genes correlated to the first mode, were ideally separated into tumors and normals (Fig. 32.5). This was the proof of proper quality control in sample selection, and it confirmed that the tumor-normal difference in this type of cancer is large enough to be detected by unsupervised methods. Even more interesting were the results of clustering by the second and third modes: both these patterns were unexpected, and before the study we did
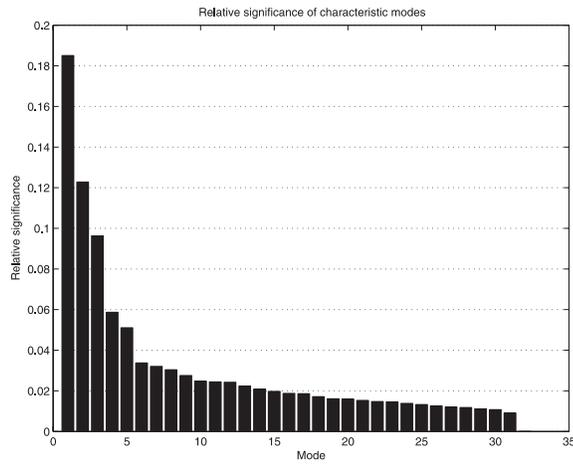
Fig. 32.4. Relative significance of characteristic modes in thyroid cancer study.

not have any knowledge as to what factors (except tumor-normal difference) might influence the expression profile. The clustering based on second mode expression profile revealed two groups of samples, both containing two subgroups of tumors and normals (Fig. 32.6). We could not attribute this subdivision to any clinical factor. The clustering based on the third mode genes did not show any attributable pattern, only some paired samples (tumor-normal) were co-clustering together. When we analyzed the content of the selected gene lists, we revealed that for the second and the third modes the significant proportion of transcripts was immune-related genes (40.3% in the second
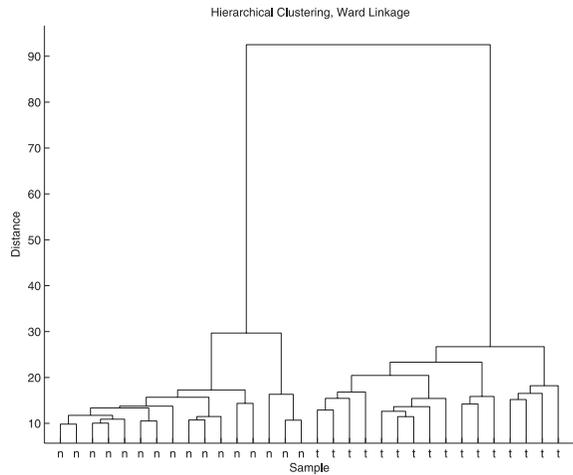
Fig. 32.5. Hierarchical clustering of thyroid tumors (t) and normal thyroid tissues (n) based on genes selected by the first mode.
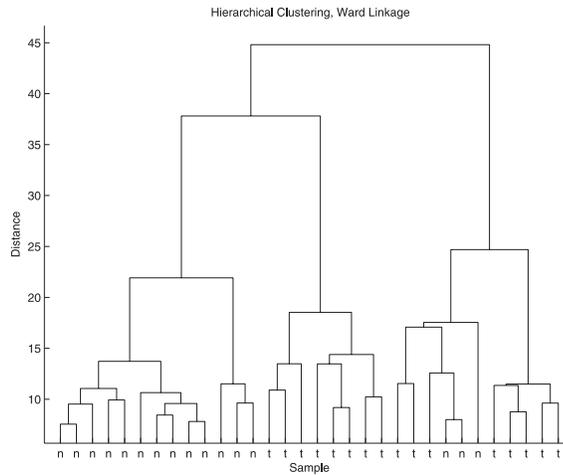
Fig. 32.6. Hierarchical clustering of thyroid tumors (t) and normal thyroid tissues (n) based on genes selected by the 2nd mode. The analysis reveals two distinct subgroups, each containing both tumor and normal tissues.

mode, 39.2% in the third mode). This suggested that the expression profile of papillary thyroid cancer is strongly influenced by expression of immune-related genes, the majority of them probably expressed in infiltrating leukocytes. However, the origin of these transcripts has still to be confirmed.

32.3.3 Distinguishing Between Histological Subtypes of Differentiated Thyroid Cancer

Papillary thyroid cancer is the most common malignant tumor of the thyroid, but there exist other histological types of thyroid neoplasms. The diagnosis of this tumor by microscopic analysis is related to numerous problems. The most difficult is the differentiation between two entities: follicular carcinoma and follicular adenoma, the latter one being a benign disease and not demanding the intensive treatment applied to both follicular and papillary cancers.

The situation is complicated further by the presence of a follicular variant of papillary thyroid cancer. The first four modes explained more than 45% of the variability (Fig. 32.7). The genes selected on the basis of mode correlation coefficients were very interesting from a biological point of view, and hierarchical clustering revealed distinct differences between tumors of both histological subtypes. However, the same conclusions were obtained by the analysis of array characteristic modes. Two-dimensional analysis of the first two modes (Fig. 32.8) revealed that tumors cluster not according to the difference between benign and malignant (which was expected from the difference in their clinical behavior), but according to the morphological distinction between papillary and follicular features. All follicular adenomas and carcinomas except one had positive values of the first mode coefficient, while all classic papillary tumors except
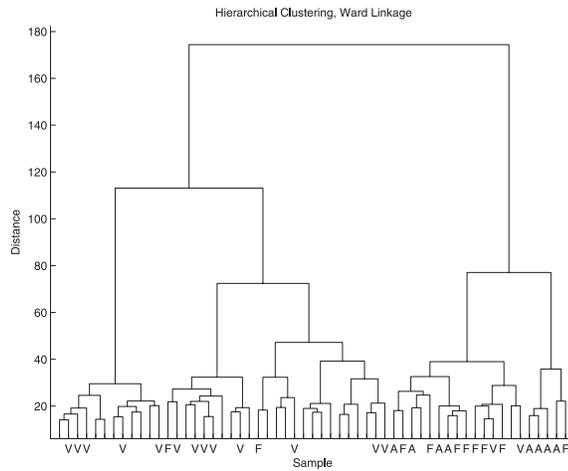
Fig. 32.7. Hierarchical clustering of thyroid cancer specimens of different histopathological subtype. Follicular cancers (F) and adenomas (A) are clustering together, while papillary cancers (not labelled) are within the other branch of dendrogram. Follicular variant of papillary cancer (V) samples are found within both subgroups.

two had negative coefficients. Benign and malignant follicular tumors could not be distinguished based on the gene expression profile. A very interesting category was the follicular variant papillary tumors: they were interdispersed between follicular tumors and papillary cancers, some of them with negative and some with positive coefficients. It is now a matter of debate and thorough analysis, whether this group is heterogeneous in biological nature or whether the histopathological criteria used to classify them are not adequate.
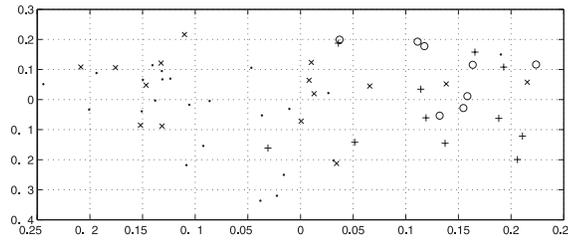
Fig. 32.8. Analysis of array characteristic modes shows distinction between follicular tumors (adenomas marked by circles, carcinomas by plus signs) and papillary tumors (papillary cancer marked by dots, follicular variant of papillary cancer denoted by x marks).
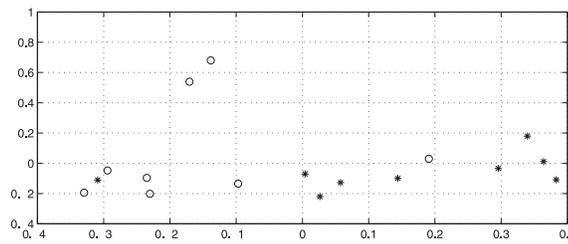
Fig. 32.9. Analysis of array characteristic modes in laryngeal cancer helps to detect outliers in gene expression profile: two mislabelled samples classifying in the inappropriate class of samples and two normal samples with distinctly different second mode coefficient values.

32.3.4 Detecting Outliers in Gene Expression Profile

SVD is a very powerful unsupervised method for detecting outliers in gene expression profiling experiments. A good example is our study of the gene expression profile in laryngeal cancer.

By analysis of array characteristic modes we found that two samples are probably mislabelled (one tumor clustered with normal tissues and one normal tissue with tumors), and two normal samples were distinctly different from all other samples within the second mode (Fig. 32.9). Using supervised methods of gene selection on the whole dataset we were unable to detect any genes significantly differentiating tumors and normal tissues. After exclusion of detected outliers, we determined a number of genes to have biological meaning, and we are further validating their significance.

32.4 Conclusions

Singular value decomposition is a reliable mathematical tool for revealing the main sources of variability in analyzed microarray datasets. When followed by a gene selection procedure based on *gene characteristic mode* coefficients, it is also a robust technique to provide biological interpretation of observed variability. Calculation and analysis of the *array characteristic modes* allow easy detection of outlier samples in microarray data.

Acknowledgments

References

1. Alter, O., Brown P.O., Botstein, D.: Singular value decomposition for genome-wide expression data processing and modeling. Proc. Natl. Acad. Sci. USA, **97**, 10101–10106 (2000).

2. Wall, M.E., Rechtsteiner, A., Rocha, L.M.: Singular value decomposition and principal component analysis. In: Berrar, D.P., Dubitzky, W., Granzow, M. (eds.) A Practical Approach to Microarray Data Analysis. Kluwer Academic Publishers, Boston Dordrecht London (2003).

3. Holter, N.S, Mitra, M., Maritan, A., Cieplak, M., Banavar, J.R., Fedoroff, N.F.: Fundamental patterns underlying gene expression profiles: simplicity from complexity. Proc. Natl. Acad. Sci. USA, **97**, 8409–8414 (2000).

4. Golub, G.H., van Loan, C.F.: Matrix Computations. Johns Hopkins University Press, Baltimore (1996).

5. Simek, K., Kimmel, M.: A note on estimation of dynamics of multiple gene expression based on singular value decomposition. Mathematical Biosciences, **182**, 183–199 (2003).

6. Simek, K., Fujarewicz, K., Swierniak, A., Kimmel, M., Jarzab, B., Wiench, M., Rzeszowska, J.: Using SVD and SVM methods for selection, classification, clustering and modeling of DNA microarray data, Artificial Intelligence. Engineering Application of AI, **17**, 417–427 (2004).

7. Jarzab, B., Wiench, M., Fujarewicz, K., Simek, K., Jarzab, M., Oczko-Wojciechowska, M., Wloch, J., Czarniecka A., Chmielik, E., Lange, D., Pawlaczek, A., Szpak, S., Gubala, E., Swierniak, A.: Gene expression profile of papillary thyroid cancer: sources of variability and diagnostic implications. Cancer Research, **65**, 1587–1597 (2005).

8. Olbryt, M., Jarzab, M., Jazowiecka-Rakus, J., Simek, K., Sochanik, A.: Gene expression profile in B16 (F10) murine melanoma cells in vitro under hypoxic conditions, paper submitted to Cancer Research.

9. Jarzab, B.: personal communication.

10. Markowski, J.: personal communication.

Index