Mathematical models of cell cycle regulation

Hendrik Fuß, Werner Dubitzky, C. Stephen Downes and Mary Jo Kurth

Received in revised form: 7th April 2005

Abstract

The cell division cycle is a fundamental process of cell biology and a detailed understanding of its function, regulation and other underlying mechanisms is critical to many applications in biotechnology and medicine. Since a comprehensive analysis of the molecular mechanisms involved is too complex to be performed intuitively, mathematical and computational modelling techniques are essential. This paper is a review and analysis of recent approaches attempting to model cell cycle regulation by means of protein–protein interaction networks.

INTRODUCTION

Proliferating cells perform a series of coordinated actions collectively referred to as the cell cycle. The complex network of regulatory enzymes and cellular components that controls these processes enables cells to grow and divide, to control or prevent growth when appropriate, to carry out the different stages of growth and division in the correct order, and to respond to DNA damage by arresting progression through the cycle so as to allow time for repair to occur before more DNA is replicated or chromosomes are condensed.

Because the development of cancer is associated with loss of control over this regulatory system, the study of the mechanisms and functions of the eukaryotic cell cycle has gained increased attention in the past decades. A detailed understanding of the mechanisms underlying tumour growth, DNA damage repair, intercellular signalling and other cell cycle related processes is therefore of paramount importance to diagnosis, treatment and prognosis of cancer.

The underlying regulatory system is often described as complex. In fact, three different aspects contribute to the complexity of the cell cycle biochemistry:

- a large number of enzymes and proteins are involved in the cell cycle;
- a large number of interactions exist between the proteins and enzymes;
- the proteins and enzymes interact in different ways, performing stimulatory, inhibitory or other modulating functions.

Complexity makes the characterisation of this system difficult. A computational approach towards understanding cell cycle regulation therefore has two major objectives. The first objective is to provide a qualitative and quantitative explanatory model describing the inner workings of cell cycle regulation. Computer simulations can help linking observations to hypotheses by reproducing the expected behaviour from a theoretical model (see Figure 1A). The second objective is concerned with predictive models capable of extrapolating from a given state of the cell or its components to subsequent states (see Figure 1B). The prediction of anti-cancer drug response, for example, has applications in tumour therapy. Targeted tumour therapy uses computational model predictions to select the most effective medication and to reduce the overall costs of a therapy.

Expensive experimental techniques such as chip technologies are involved in both medical applications and in cell cycle research. Both fields can benefit from the...
Figure 1: Workflows of two mathematical modelling strategies. (A) Explanatory modelling using inductive reasoning. The experimental observations and the model’s predictions (possibly those of several alternative models) are compared in order to confirm or reject the hypotheses encoded in the mathematical model. The term ‘behaviour’ can refer to features of physiological parameters such as enzyme concentrations or macroscopic events, for example, the growth of a cell population. (B) Predictive modelling using deductive reasoning. A computational implementation of the mathematical model may be used to conduct artificial (in silico) experiments that can suggest novel hypotheses or support decisions on therapeutic strategies. Observations from the real system may be used to validate the new hypotheses or to provide parameters or initial conditions to the mathematical model.

Ordinary differential equations as mathematical representation of cellular processes

Predictive capabilities of a mathematical model.

In order to simulate the molecular processes inside a cell, a mathematical model captures the processes and its components by representing the system state $x(t)$ at time $t$ and by computing its subsequent state $x(t+1)$ based on the state $x(t)$. The variable $x$ denotes the concentrations of all molecular species captured by the model. In chemical kinetic theory\(^1\) the interactions between species are expressed using ordinary differential equations (ODEs).

ODEs are equations in which the unknown element is a function, rather than a number, and in which the known information relates that function to its ordinary (as opposed to partial) derivatives. Many phenomena in physics, chemistry and biology with a temporal dimension involve relationships between the values of variables at a given point in time and the changes in these values over time. Generally, an ODE takes the form:

$$F[t, x(t), \dot{x}(t), \ddot{x}(t), \ldots] = 0,$$

where $t$ is a scalar variable (normally interpreted as time), $F$ is a known function, $x$ is an unknown function, $\dot{x}(t)$ is the derivative of $x$ with respect to $t$, $\ddot{x}(t)$ is the second derivative of $x$ with respect to $t$, and so on.

The following simple example, adapted from supplementary material in Pomerening et al.,\(^4\) illustrates a system of two mutually activated enzymes A and B represented by two ODEs:

$$\dot{x} = -k_1 x + (a - x) G(y)$$

$$\dot{y} = -k_2 y + (b - y) F(x)$$

where $a$ and $b$ represent the total concentrations of the two proteins and $x$ and $y$ represent the concentrations of the active form of the proteins. The terms $F(x)$ and $G(x)$ describe the mutual interactions between the two proteins.

In the majority of cell cycle models these mathematical principles are applied to one of two biochemical domains: gene...
regulatory networks and protein–protein interaction networks. In gene regulatory networks interactions are modelled purely on a genetic level. Interactions between genes and proteins are implicit in these models. Cell cycle regulation depends heavily on interactions of proteins and enzymes and their modification through phosphorylation and dephosphorylation by kinases and phosphatases, respectively. These interactions and enzymatic reactions are represented in protein–protein interaction or regulatory networks. In contrast to gene regulatory networks such models explicitly represent known or hypothetical molecular interactions on the protein level.

Macroscopic effects, such as the migration of a cell within a tissue or the formation of buds in yeast cells, can be an explicit or implicit part of such a model. Patterns and shapes of higher-level structures sometimes arise purely from reaction dynamics. In many cases it is not clear how these macroscopic cellular processes connect to the cell’s physiology. Theoretical models that reproduce such effects can therefore make an important contribution to our understanding of the cell cycle.

For the successful application of mathematical modelling to cell cycle problems one has to answer a number of questions: What is the system in question? What is the environment? How can we model it and what information do we use as input? What do we expect to observe and what insights do we expect from the application of the model? What methods should we use for modelling, analysis and validation? How can we most effectively obtain experimental evidence to support new hypotheses?

Generally, the process of modelling the cell cycle and its regulatory mechanisms involves three distinct phases:

- **Model design phase.** This phase is concerned with the definition and formulation of a mathematical model of the biological process, system or problem under investigation.

- **Model application/analysis phase.** Here the computational implementation of the model is used to simulate and study the dynamic behaviour of the model under different conditions in relation to the studied biological system or process.

- **Model validation phase.** Here the behaviour and data generated by the computational simulation of the model are either compared against data obtained from analogous experiments on real biological systems or assessed epistemologically on the basis of existing knowledge.

Before we revisit each of the modelling phases outlined above in detail, we first review the relevant background of biology.

**THE CELL CYCLE – BIOLOGICAL BACKGROUND**

The eukaryotic cell cycle is divided into four main phases, the synthesis (S) phase, the mitosis (M) phase and two growth phases in between (G1 and G2, see Figure 2). A newborn cell resides either in G0, the quiescent non-dividing state, or in G1 until physiological parameters allow it to enter the S phase and to start replicating its genetic material. An exact copy of each DNA strand is created, and by the end of the subsequent G2 phase the chromosomes are aligned on the mitotic spindle. In the M phase the paired chromatids are separated and transported to the two opposite spindle ends, ready to form two new nuclei with identical genetic information. The cell cycle ends with cytokinesis, the physical division into two new cells. The period of this circular process varies between species and tissues from several minutes to several days. It should be noted that the distinction of these four phases is based on morphological observations and does not relate to the underlying biochemical mechanisms. Therefore not all of these
phases are sharply defined on a molecular basis. Surveillance systems called checkpoints prevent progression of the cell cycle in response to perturbations, such as DNA damage, nucleotide depletion or other defects (see Figure 2). Several classes of protein kinases and other enzymes perform essential functions in the implementation of the cell cycle and the checkpoint system. Protein kinases alter the activity of their enzyme substrates by attaching a phosphate group. In 2001, the Nobel Prize in medicine was awarded for the discovery of how cyclin-dependent kinases (CDKs), an essential class of cell cycle enzyme, control progression through the cell cycle. Different CDKs are activated in characteristic patterns in the distinct phases of the mammalian cell cycle (see Figure 3). Their activity is dependent on the availability of specific cyclin subunits, the phosphorylation state of the CDK itself and the presence of CDK inhibitors. These interactions form the biochemical manifestation of the cell cycle phases and checkpoints.

The term ‘checkpoint’ suggests that there exist certain entities in a cell whose purpose is to check whether a condition (e.g. complete DNA replication) is fulfilled and to emit signals to other active entities within the cell cycle. While this notion is part of the current consensus model of the cell cycle, such external monitoring entities have been subject to critique. The signal-emitting entity may well be an inherent part of the metabolism, such as the nucleotides used to replicate the DNA or the cytoskeleton that separates the chromatids rather than an external ‘checking’ instance. In this paper we will use the word checkpoint to describe any kind of regulatory mechanism that will halt the cell cycle in response to certain

Figure 2: Schematic representation of cell cycle phases and checkpoints. Proliferating cells proceed through a cycle of DNA synthesis (S phase) and mitosis (M phase) separated by two growth phases (G1 and G2). Quiescent cells remain in G0 state until conditions are favourable for cell division again. Checkpoints (represented by halt signs) prevent progression in response to delays or defects.

Cell cycle checkpoints respond to perturbations

Figure 3: Each cell cycle phase (bottom) is characterised by the activity of cyclin-dependent kinases (CDKs), which are activated by association with an appropriate cyclin subunit. While cyclins undergo specific patterns of synthesis and degradation, CDKs are present throughout the cell cycle (adapted from McBride).
in intracellular or extracellular conditions, whether it is evolved or inherent to the system.

**MODEL DESIGN**

Deriving a formal abstraction (i.e. a mathematical description) of the biological system or process under investigation is a crucial step in modelling. This step involves decisions on what precisely is understood by ‘the system’ and requires the definition of system boundaries and the components of the system. Furthermore, it is necessary to determine what types of interaction with the environment should be included and which simplifying assumptions should be made.

Detailed and comprehensive information about interactions between cell cycle-related proteins is available in the form of molecular interaction maps,9 pathway charts10 and public databases.11 However, mathematical models of biochemical and cellular processes require a precise and unambiguous formulation of the underlying reactions and their quantitative parameters in order to predict the real system’s behaviour. Resources such as the above-mentioned molecular interaction maps leave a large amount of uncertainty, as there are numerous ways to mathematically model interactions between two partners, especially when enzymes, inhibitors or other modulators are involved. This requirement forces us to systematically generate hypotheses and gather all the available information in order to mathematically describe the problem.12

Based on the mathematical formulation of the problem, the underlying quantitative information is finally encoded as an executable computational model, which can simulate the behaviour of the studied biological system or process by numerical integration. Thus, the computational model can be used to test hypotheses and to compare modelled system responses and behaviour to experimental observations (see Figure 1A).

**Scope and complexity**

**Cell cycle modules**

Mathematical modelling has been applied to numerous problems related to the cell cycle. Some modelling approaches focus on the basic function of the cell cycle ‘engine’ that is responsible for the different cell cycle phases and transitions between them. Modelling of protein interactions in well-studied species such as budding yeast has led to highly developed and complex models that reproduce and explain a major part of the physiological phenomena that are currently known to occur within the cell cycle.13

However, the complexity of the underlying biochemistry often makes it necessary to subdivide the problem into several parts, which can be studied and optimised individually.14 Complex mathematical models can be composed of smaller, well-studied modules. A systematic study of elementary protein-protein interaction modules with applications to cell cycle physiology was presented by Tyson et al.15 Biochemical switches, signal amplifiers, chemical oscillators and other modules resemble the components of an electronic circuit. They can provide a high-level structure to networks as well as inspire hypotheses on novel pathways to achieve certain signal-response properties.

Since many comprehensive and well-characterised cell cycle models are available, it is tempting to use these as a basis for larger models. Such an approach could be useful for determining the roles of novel regulatory proteins in the cell cycle, but also to extend the scope from molecular to cellular or multicellular level or to add a spatial component.

**Cell cycle phase transitions**

Modelling specific phase transitions often allows for a more detailed study of the features and roles of the key enzymes involved. Qu et al.16 discuss the importance of multisite phosphorylation for bistability and oscillation within the G1/S transition. Qualitatively different dynamics arise from the number of
phosphorylation sites on kinases and transcription factors. A larger number of sites increases the activation threshold and leads to ‘steeper’ response characteristics. Their model suggests that at least two phosphorylation sites are required for the regulation of this phase transition.

A recent study addressed the characterisation of the restriction point, an early cell cycle checkpoint first proposed by Pardee. Once the cell has passed this ‘point of no return’ it does not respond to certain cell cycle halting agents or deprivation of growth factors. The mathematical model quantitatively reproduces the effects of such interventions. The transient deprivation of growth factors may lead to an increased cell cycle length depending on the current position in the cycle. Restriction point control is believed to be an important oncological factor, although there are controversial views on the interpretation of Pardee’s experiments (see for example Cooper). In the light of this controversy mathematical models of these phenomena might serve as a basis for a profound discussion of the biological mechanisms underlying the observations, to eventually yield a theory that withstands critique.

Beyond the cell cycle

Developmental biology is the study of processes that give rise to tissues, organs and organisms in specific shapes and patterns. These processes, too, are controlled by biochemical networks, which are closely related to the cell cycle. Fertilised eggs of Xenopus laevis, for example, undergo a series of 12 rapid, synchronous cell divisions lacking growth phases in between. The embryo then enters the mid-blastula transition, after which its cells grow asynchronously and regulated by the full, somatic cell cycle system. In the model presented by Ciliberto et al., these features all emerge from a set of differential equations with appropriate parameters. The model thereby suggests two new biochemical mechanisms, introducing a hypothetical kinase required for oscillatory behaviour and a mechanism by which cyclins are removed or inactivated.

Models of social behaviour and intercellular messaging take whole populations of cells into account, each one proceeding through its own cell cycle, influenced by extracellular parameters such as growth factors and bonding with neighbouring cells. It seems reasonable to assume that models at this level do not necessarily require molecular details to provide interesting results. Walker et al. developed a simple rule-based computational model that is sufficient to reproduce cell migration phenomena under varying conditions. The model does not include any molecular interactions, but instead keeps track of the cell cycle phase and the cell size for each individual cell. Decisions are based on these two parameters and on the position of neighbouring cells. Such models offer interesting applications with regard to wound healing, tumour growth or apoptosis.

An illustrative simulation that integrates several of these levels is an agent-based system of cells simulating the growth dynamics of brain tumours by Athale et al. Each cell decides whether it should proliferate or migrate based on simulated molecular parameters. The cells can migrate within an extracellular matrix featuring a glucose gradient. The uptake of glucose then feeds back into the cell’s physiology. Using this approach, the expansion of tumours can be studied using a computational approach on both molecular and macroscopic levels.

Representing biochemical networks

There are many different possible mathematical representations that describe the same biochemical reactions in a cell. Some authors prefer to model protein interactions and enzymatic reactions using only basic mass-action kinetics, which leads to a high number of reactions (or variables in the model), while maintaining relatively simple individual equations. Using more abstract, higher-level kinetics,
such as Michaelis–Menten equations, Hill functions or threshold functions, it is possible to reduce the number of variables, but at the same time, complexity rises owing to the various types of reaction representations involved. This approach leads to fewer, but mathematically more complex, equations.

Few mathematical models work with actual physiological concentrations, since dimensionless terms greatly simplify the resulting equations by eliminating constants. Such a model’s variables represent fractions rather than absolute concentrations of proteins. Dimensionless terms are therefore useful, when quantitative predictions are not required.

**Graphical representations**

Graphical notations of biochemical networks are useful for two purposes: they are a compact form for displaying complex interaction networks and also allow scientists without a mathematical background to participate in the modelling process. However, many common forms of biochemical diagrams cannot be directly transformed into a mathematical model. Such diagrams do not contain the mechanistic details that are needed for simulating. Interactions such as the stimulatory influence denoted by the plus arrows in Figure 4A do not have a single, defined mathematical representation. The modeller must decide whether a Michaelis–Menten, basic mass-action or other kinetic law is appropriate for each interaction.

Kohn therefore distinguishes two types of diagram: explicit diagrams (as in Figure 4C), which contain all information required for simulation, and heuristic diagrams (as in Figure 4B), which are not sufficient, but possibly helpful for model creation. While heuristic diagrams provide a compact notation for depicting large networks of biochemical processes, explicit diagrams provide a graphical representation of a mathematical model. This means that an explicit diagram can be unambiguously converted into a set of ODEs, which, given appropriate parameters, is ready for numerical integration.

**Computer document formats**

To permit computational simulation it is necessary to transfer mathematical models to computational representations, i.e., programming languages, data structures and algorithms that efficiently manipulate those structures. Software for model building and simulation, as well as libraries and services, must agree on a common format or language to exchange the computational representations of biological models. In the context of cell cycle modelling, two XML-based document formats for this purpose exist: the *Systems Biology Markup Language* (SBML, see Figure 4E) and *CellML*. They both allow quantitative specification of reaction kinetics in multiple compartments.

Repositories containing cell cycle models in SBML and CellML format exist, but neither of these repositories nor the respective formats seem to have attracted great attention in the relevant modelling literature.

**MODEL ANALYSIS**

A central objective of mathematical cell cycle models is to facilitate understanding of cell cycle regulation. However, currently there does not seem to be a consensus on what ‘understanding’ precisely means. Merely reproducing the known biochemistry and the observed behaviour in a computer obviously does not further our knowledge about the cell cycle. Various analysis methods do actually provide insights into real biological systems that cannot be gained otherwise.

One important aspect of understanding is the derivation of principles and architectural or topological features. Scientists seek to understand how a system is constructed to be able to carry out its work and why it has been constructed that way. Owing to recent advances in graph theory, the notion of *scale-free networks* and the identification...
Figure 4: Examples of various representations of the same core cyclin/CDK subsystem (adapted from Qu et al.26). (A) A typical reaction scheme diagram. The double arrows denote the possible reactions between species, which in this diagram consist of phosphorylations, dephosphorylations and associations. Plus arrows indicate the stimulatory influence of an enzyme on another reaction. (B) and (C) Molecular interaction maps, a very compact display formalism that allows depiction of large interaction networks.9 The explicit diagram (C) contains nodes for enzyme substrate complexes to represent enzymatic activity, such as phosphorylation and dephosphorylation by Cdc25 and Wee1. The heuristic diagram (B) uses designated symbols (circle ended and T-shaped line) to represent these features. (D) Two equations from a system of ordinary differential equations (ODE) describing the change in concentration of each individual element over time. (E) Excerpt from an SBML model of this system.
Emergent properties correspond to biological features

So-called *emergent properties* provide insights gained from a systems-level view and analysis of a biological process or entity. These properties arise from the qualitative and quantitative structures of the system and its components and their interactions, similar to how pressure emerges as a global property from large numbers of interacting gas molecules. Even simple biochemical networks featuring a small number of interactions and feedback loops can exhibit *multistability*, oscillatory behaviour and can undergo different kinds of bifurcations (see below). These emergent properties directly correspond to biological features such as cell cycle phase transitions and checkpoint responses.

Furthermore, mathematical models and their computational implementations provide a framework for *in silico* or ‘artificial’ experiments. Experimentalists can execute and explore such models to decide whether an experiment is promising or not and thus reduce experimental time and costs. Cross et al. performed a series of experiments based on the prediction of a mathematical yeast cell cycle model. Since these studies often involved laborious genetic manipulations, computational modelling allowed them to specifically select the most insightful experiments.

**Bifurcation analysis**

**Parameter space and behaviour space**

Parameterisation and initialisation of a mathematical model is a prerequisite for model execution. Once all kinetic parameters and initial conditions have been specified, time-series data can be obtained from the simulation by numerical integration. Hakenberg et al. demonstrated how text mining can be used to determine kinetic model parameters from large volumes of scientific articles. However, often investigators are not interested in studying systems based on particular parameters but in how the system responds to perturbations of these parameters.

Perturbations lead to qualitatively different results that correspond to different behaviours or phenotypes of the biological system. In their comprehensive yeast cell cycle model, Chen et al. analysed simulation data for over 100 parameter sets. Each set corresponds to a certain genotype (e.g., deletion or over-expression mutants) and leads to an observed *in silico* behaviour (e.g., arrest in a particular cell cycle phase, production of abnormally large cells or wild-type behaviour), that can be compared to experimental observations.

The study of qualitative changes in dynamical behaviour of a system in response to parameter variation is called bifurcation analysis and several aspects of this theory have interesting implications for cell cycle modelling. Two of the two main implications, multistability and sensitivity analysis, are discussed below.

**Multistability**

Multistability refers to systems that have more than one stable state within a range of parameters. Such systems can alternate between two or more mutually exclusive states over time. It is widely believed that multistability is a critical property in the control of the cell cycle. Bistability of the G1/S transition has been demonstrated experimentally in yeast and mammalian cells. It involves a hysteretic switch that prevents the system from returning to G1 state once it has entered S phase.

The point at which the system undergoes a transition between two stable states usually depends on certain kinetic parameters. Novák and Tyson argued that this sensitivity to parameter variation plays a major role in the mechanisms underlying the cell cycle checkpoints.
Whenever conditions are unfavourable for progression into the next cell cycle phase, an allosteric modification (modification of a regulatory enzyme activity’s by the non-covalent binding of a particular metabolite at a site (the allosteric site) other than the active site) of an enzyme may change its kinetic behaviour, shifting its activation threshold to a virtually unreachable level.

Other mechanisms have been found to play a role in cell cycle transitions. For example, insoluble aggregates of cyclins can provide a kind of buffer for soluble cyclin in the cytoplasm. Since solvatation of cyclin is faster than cyclin synthesis, this allows for a rapid and sensitive response in the activation of the G1/M transition. The mathematical model proposed by Slepchenko and Terasaki simulates this phase transition. With cyclin aggregation the model displays bistable behaviour over a wider range of parameter values. In conclusion this mechanism increases the robustness of the system.

**Sensitivity analysis**

Characterisation of a mathematical model often involves parameter sensitivity analysis, which is used to determine the roles and importance of individual parameters. A kinetic parameter may, for example, directly affect major factors such as cell cycle period or mass at division, while others may raise or lower the threshold concentration level of a particular protein that is needed to activate an enzyme complex. The methods used for sensitivity analysis range from single parameter variation to complete randomisation tests. Clearly, changing only single parameters cannot provide a comprehensive overview over the robustness of the system. Cancerous cells normally have multiple defects that, taken together, cause the regulatory system to malfunction. Randomisation tests by von Dassow et al., on the other hand, have revealed a remarkable level of robustness in a developmental module in *Drosophila*. Out of their simulations with random parameters, 80 per cent reproduce the normal, healthy behaviour. This finding can also be taken as a measure of plausibility for the model, since in any biological system slight perturbations should not disrupt essential cellular functions. This is particularly important to systems exhibiting features that are well conserved over a wide range of species. While enzymes have changed in the course of evolution and exhibit altered kinetic constants, the basic function of the system remains intact.

Robustness can be used as a criterion to discriminate between several hypothetical models for a biological system. A quantitative analysis of robustness, comparing two published *Xenopus* cell cycle models, clearly demonstrates both the predictive and explanatory advantage of the more recent model, which differs from the earlier one only by the addition of two feedback regulation loops. Some of its basic properties, like the cell division frequency, are more robust to parameter variation, which makes the latter model more plausible. Furthermore, the earlier model required some kinetic parameters to assume physiologically unacceptable values. The latter model, which allows for a broader range of valid parameter values, eliminates this problem.

**MODEL VALIDATION**

In cell cycle simulations, experimental data are mainly used for three tasks:

- reverse-engineering of networks to construct a computer model from experimental data;
- estimating parameters of an existing mathematical model; or
- testing and discriminating between several alternative models.

Pure data-driven approaches are rare in protein–protein interaction modelling. Reverse-engineering networks from measured data using machine-learning, evolutionary or artificial intelligence...
methods are common in gene regulation studies, but have not been applied to protein-interaction networks. The main reason for this is that regulatory networks involve non-linear dynamics, which makes the reverse-engineering problem extremely complex. It is questionable whether a reverse-engineered regulatory network that was fitted to a set of training data has any explanatory capabilities, even when it has demonstrated predictive qualities. Another reason for the lack of data-driven approaches may be that available experimental data are too sparse to construct a network without prior knowledge. While large-scale time-series data can be obtained today using microarray techniques, protein chip technology is still in its infancy.

Most cell cycle simulations employ a hypothesis-driven approach. Experimental observations are then used as evidence to decide whether the hypotheses are plausible (see Figure 1A). It should also be noted that there are probably more insights to be gained from the finding that a model does not comply with a set of data. Ciliberto et al. found that it was necessary to add a hypothetical kinase to their model in order to reproduce the experimentally observed behaviour. The resulting model thereby suggests a novel pathway, involving a yet unknown protein, which remains to be discovered.

**Experimental methods**

Experimental data can be obtained on various levels of biological organisation. In cell cycle scenarios this involves data from the following groups of dynamic and static data. Dynamic data describe the temporal evolution of a system or its components:

- Concentration time-series data obtained using proteomic methods. This most direct and most common form of validation involves the detection of relevant proteins with specific antibodies or by using tagged or fluorescent labelled proteins. These techniques are often used in conjunction with polyacrylamide gel electrophoresis (PAGE) to separate the proteins in a sample. The obtained concentration data directly corresponds to the model variables. When combined with genetic engineering techniques, these methods allow for sophisticated experiments, which can give evidence for complex phenomena such as bistability, as demonstrated by Cross et al. However, the requirement of a specific detection method makes these procedures difficult to scale.

- Morphological and phenotypic observations. Much of what we know about the cell cycle was inferred from the behaviour of genetically altered strains. These genetic modifications are easy to reproduce in the computer and phenotypic features such as relative cell cycle times, cell size at division, resistance to perturbations such as DNA damage etc can be compared with experiments.

- Observations at population or developmental level. Automated cytometry methods such as flow-cytometry measure the distribution of a feature (cell size, DNA or protein concentration) across a population. Cross et al. have employed this technique to measure the change of protein concentrations during cell growth, which roughly correlates with progression through the cell cycle. Microscopic images, too, have been used as experimental evidence for cell cycle simulations to show that the model correctly reproduces cell migration or growth patterns.

Static data describe time-independent features of a system or its components:

- DNA and protein sequence data. Large amounts of data are available and accessible through public databases. For example, molecular dynamics...
Simulations with existing protein structural data have been used in other mathematical modelling fields to estimate rate constants for kinetic equations. Exploiting this type of data for use in cell cycle models is an open challenge.

- Protein interaction studies. Such experiments are mostly based on yeast two-hybrid systems or phage display libraries and contain qualitative information about cell physiology. While these types of experiments cannot reveal the function of protein interactions or even determine quantitative kinetic parameters, they show which proteins can associate due to their biochemical nature. The integration of such large-scale interaction data in actual cell cycle models has so far not been exploited, but numerous studies on topological system features rely on the availability of public interaction databases. It is often noted that experimental data for regulatory networks are difficult to obtain. It is true that most variables in a mathematical cell cycle model are not easily observable experimentally. On the other hand, indirect evidence for a model can be found on various levels. A key challenge for future research in cell cycle modelling lies in the correlation of these indirect measurements to the model.

Parameter estimation
It is often argued that as a mathematical model becomes more complex and more hypothetical model parameters are included in the model, the more the model’s ability to adjust to any arbitrary data set increases, and thus, the information content of the model decreases. Using the information encoded in experimental data to estimate model parameters can limit the model’s flexibility and thus makes it more likely to reflect the biological processes accurately. Common maximum likelihood methods, however, require a large number of data points to avoid over-fitting, but measurements are often time-consuming and expensive and may even disturb the system itself. In order to use laboratory resources most efficiently, experimental strategies need to be optimised, so that they can provide most useful information to the mathematical methods employed. Kutalik et al. have presented a mathematical method for optimising sampling times and the number of replicates required when collecting data for parameter estimation. This optimisation method can either achieve a given maximal tolerance level for the estimates or it can minimise the error for a given number of measurements. They illustrate their method using a simple mathematical model, which demonstrates that careful selection of sampling times can significantly increase estimation accuracy.

Parameter estimation is currently not widely used for cell cycle modelling. Models aiming at high predictive accuracy would certainly need this kind of data-driven support, but it may not be crucial to many cell cycle simulations that seek to qualitatively explain the observed behaviour.

CONCLUSIONS
A number of models for diverse cell cycle regulation problems have been proposed by the research community. They simulate physiological or developmental effects in a wide range of species. Most of the models considered in this paper pursue the all-important goal in the post-genomic era: characterisation of the roles and functions of genes and proteins. Put in the context of a particular system, the knowledge of these individual functions is expected to improve our understanding of the cell cycle as a whole and its constituent mechanisms.

Many of the molecular processes that govern the basic cell cycle ‘engine’ are known today, but this knowledge does not explain the highly developed organisation of the cell cycle. Events such as nuclear envelope breakdown,
chromosome condensation and alignment or cytokinesis have so far not been described mathematically. Also, while molecular targets involved in checkpoint responses have been identified, the concept of checkpoints in the cell cycle is not clearly defined and may need re-examination. It is currently not clear how checkpoints detect and respond to the various perturbations. For efficient use of the available knowledge, it is vital that model developers can both publish their work in a usable way and access existing models to apply analysis methods and experimental data or to build new models.

A consensus on a graphical formalism for protein interaction models is yet to emerge. Some promising approaches are underway. While differential equations allow detailed mathematical analyses and graphical displays, they also obscure certain features of a system, such as flow relationships between elements. The community should therefore agree on the use of standardised model formalisms as well as computer document formats that facilitate exchanging models between researchers and between software applications. Curated model repositories with standardised interfaces might comprise platforms as powerful as SwissProt for protein sequences.

Existing software tools, both commercial and open, support the systems biologist in some of the steps involved. The use and integration of experimental data for the purpose of parameter estimation or model discrimination are still open issues. SBML and CellML were not intended for storage of experimental data or resulting analyses. Emerging integration infrastructures like the Systems Biology Workbench are a vital contribution to efficient model development, because software solutions for common tasks such as numerical simulation and graphical visualisation can be reused to build upon. Still, the learning curve is very steep and building computer simulations typically requires advanced programming skills. This hinders scientists from various disciplines from ‘deep’ participation in the modelling process.

Current cell cycle models make little use of high-throughput methods such as mass spectrometry, hybridisation-based assays and 2D gel electrophoresis. While this is probably because of the difficulties associated with data-driven approaches in this field, overcoming these hurdles might dramatically change the ‘lack of data’ problem. Methods for obtaining dynamical data for mathematical models should be both specific to the system’s components and scalable to larger systems with more parameters.

References

10. Biocarta Pathway Charts (URL: http://www.biocarta.com/genes[cited 5th April, 2005]).
Fuß et al.

45. Kitano, H., ‘A Standard Graphical Notation for Biological Networks’ (URL: http://www.sbml.org/workshops/sixth/[cited 6th April, 2005]).