Symposium

Journal of INTERNAL MEDICINE

doi: 10.1111/j.1365-2796.2011.02492.x

Predictive mathematical models of cancer signalling pathways

J. Bachmann[1,*], A. Raue[2,3,*], M. Schilling[1], V. Becker[1,†], J. Timmer[2,3] & U. Klingmüller[1]

From the [1]Systems Biology of Signal Transduction, DKFZ-ZMBH Alliance, German Cancer Research Center, Heidelberg; [2]Institute of Physics, University of Freiburg, BIOSS Centre for Biological Signalling Studies; and [3]Freiburg Institute for Advanced Studies (FRIAS) & Zentrum für Biosystemanalyse (ZBSA), University of Freiburg, Freiburg; Germany

Abstract. Bachmann J, Raue A, Schilling M, Becker V, Timmer J, Klingmüller U (German Cancer Research Center, Heidelberg; BIOSS Centre for Biological Signalling Studies, Freiburg; and University of Freiburg, Freiburg; Germany). Predictive mathematical models of cancer signalling pathways (Key Symposium). *J Intern Med* 2012; **271**: 155–165.

Complex intracellular signalling networks integrate extracellular signals and convert them into cellular responses. In cancer cells, the tightly regulated and fine-tuned dynamics of information processing in signalling networks is altered, leading to uncontrolled cell proliferation, survival and migration. Systems biology combines mathematical modelling with comprehensive, quantitative, time-resolved data and is most advanced in addressing dynamic properties of intracellular signalling networks. Here, we introduce different modelling approaches and their application to medical systems biology, focusing on the identifiability of parameters in ordinary differential equation models and their importance in network modelling to predict cellular decisions. Two related examples are given, which include processing of ligand-encoded information and dual feedback regulation in erythropoietin (Epo) receptor signalling. Finally, we review the current understanding of how systems biology could foster the development of new treatment strategies in the context of lung cancer and anaemia.

Keywords: cancer, cell biology, cytokines, hematology, lung cancer.

Introduction

Molecular and cell biologists frequently question why we need mathematical models to understand biological processes. However, owing to the complexity of biological systems, it is evident that a higher level of abstraction is required to decode the 'language' of cells and to gain insights into how signals from the environment are integrated and how decisions for life, death, proliferation or differentiation are regulated. In recent years, much progress has been made in the qualitative analyses of biological systems. Yet, to advance the understanding of complex diseases and the process of drug discovery (including efficacy, safety and consequently the outcome in patients), computational tools are essential that are able to integrate the plethora of experimentally observed information and facilitate the prediction of cellular responses. The reliability of these predictions critically depends on the availability of quantitative data to capture cellular events over time with sufficient quality to calibrate the mathematical models.

Cellular signal transduction pathways process extracellular signals that are received by cell surface receptors; these receptors are activated and translate this information via signalling networks to cellular responses. As the 'omics' technologies facilitated the identification of key components of signalling pathways in high-throughput systems, the current focus is the investigation of the connectivity, crosstalk and dynamics of these networks. From systems-based approaches, we have learned that temporal dynamics [1, 2], spatial distribution [3, 4] and cell-to-cell variability [5–7] are key systems properties that lead to context-specific cellular responses. These insights serve as inspiration to further investigate the emergent properties of signalling pathways and how they are quantitatively linked to decisions concerning cell fate.

In this review we provide an overview of modelling concepts describing biological systems, in particular cancer signalling pathways, to demonstrate the power of modelling approaches in addressing urgent biological questions. We do not intend to give a theoretical introduction to modelling strategies in general as this has been previously provided elsewhere [8–10].

*These authors contributed equally.
†Present address: Department of Systems Biology, Harvard Medical School, Boston, MA, USA.

Rather, we emphasize that when using mathematical models to predict cellular behaviour, certain requirements concerning the modelling strategy have to be accomplished to facilitate the prediction of cellular behaviour and a deeper understanding of the biological system. Therefore, a summary of modelling signalling pathways using ordinary differential equations (ODEs) is provided. Because the identifiability of parameters in these models is one prerequisite for achieving models with high predictive power, we introduce the theoretical background on identifiability in ODE-based models.

A signal transduction network that has been intensively studied by systems-based approaches is the erythropoietin receptor (EpoR) signalling system. Therefore, we highlight the most recent advances in this field and demonstrate how mathematical modelling enabled the identification of key system properties. Furthermore, we discuss how systems-based approaches can be employed to address complex questions in pharmacology. In particular, advances in systems biology approaches that investigate the risk of Epo treatment in patients with cancer are presented.

Two directions in systems biology

To unravel complex signalling networks and their underlying regulatory mechanisms, two different approaches have evolved in the field of systems biology (Fig. 1a). The 'top down' approach combines large-scale 'omics' data including qualitative and static information about cellular components with bioinformatics tools to analyse network topologies. The other approach, 'bottom-up', generates low to medium-throughput data with quantitative and dynamic information content relevant for calibrating medium-sized mathematical models with high predictive power. The current challenge in systems biology is to combine both approaches to achieve large quantitative network models that are detailed enough to decipher regulatory mechanisms and to accurately predict cellular behaviour as well as the potential impact of pharmaceutical intervention.

Mathematical models address different levels of complexity

In the last decade, different mathematical methods have been applied and refined to describe and analyse cellular processes at different levels of complexity (Fig. 1b). Starting with the lowest level, regression analysis is the simplest form of a statistical model that identifies correlations between biological entities. Boolean models are often applied if qualitative information and details of the temporal order of cellular events are available. In these logic models, reactions of signalling pathways or transcriptional networks are represented as logic gates and the reaction partners can have two states: on/off. Despite the fact that quantitative information about the individual components cannot be integrated, logic models provide a good starting point to analyse large-scale biological systems where details of the system are only partially defined [11–13]. The most common form to describe molecular interactions and transformations of signalling pathways in a mathematical form is a chemical reaction network described by ODEs. ODE networks represent the rates of the reactions (i.e. association, dissociation, production and degradation of the individual species) usually in terms of mass action kinetics. This approximation is derived from physicochemical theory and states that the rates are proportional to the concentrations of the reactants [8]. For example, the Michaelis–Menten approximation is derived from mass action ODEs

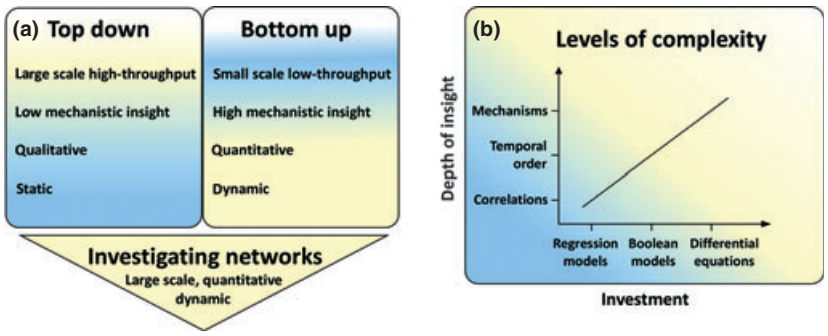

Fig. 1 *Different levels of complexity in systems biology approaches. For details see text.*

and applied in enzyme kinetics [9]. Because ODE models are able to integrate quantitative and time-resolved information, these models allow for interpretation and analysis of complex dynamic behaviour, which can lead to nonintuitive insights into biological systems [3, 14]. To integrate spatial information, however, partial differential equations are required that are more difficult to use especially with regard to parameter estimation in combination with high complexity of the interaction between reactants. Therefore, ODE modelling that considers reactions to occur under well-mixed conditions is currently the most widely applied trade-off between computational effort and feasibility. If the abundance of reactants is low, it can be important to consider the stochastic nature of chemical reactions. However, in most cases, abundances are larger than ~100; therefore, the ODE description is a good approximation [15].

Establishing ODE models of cancer signalling pathways

When establishing an ODE model for a specific biological system, such as a signalling cascade, a familiar schematic of a 'preliminary' pathway from the published literature or databases (e.g. the KEGG pathway [16]) can be translated to an ODE network as a starting point. After constructing the topology of the model and selecting the required level of abstraction, the parameters of the system (i.e. the rates of the molecular interactions and transformations and the abundance of the reactants for the initial values of the ODE system) have to be estimated or determined. These parameters determine the behaviour of the ODE model and are of critical importance for obtaining a reliable mathematical description of the biological system. However, for a given cell type and species, parameter values are often not available nor can they be measured directly. In particular, if models are representations of signalling pathways, parameters that were determined *in vitro* are often not applicable as they differ substantially from the *in vivo* situation. Moreover, parameters could be cell context-specific depending, for example, on the cell type. For these reasons, a reverse engineering approach is most suitable. Here, the dynamic behaviour of experimentally accessible components is determined using quantitative techniques such as quantitative immunoblotting [17, 18], protein arrays [19–21], quantitative mass spectrometry [22–25] or quantitative RT-PCR. Of importance, defining standardized conditions for experimental protocols [26, 27] and modelling tools are crucial [28]. After obtaining time-resolved and quantitative data, the parameters of the ODE model are estimated by calibrating the model with these measurements. The advantage of the reverse engineering approach is that the dynamic behaviour of the mathematical model is tailored to the specific cell type and species. There are two possible scenarios after obtaining the estimates for the parameters and the 'fitted' mathematical model (Fig. 2).

The ODE model is not able to describe the experimental data

The first scenario is that the ODE model cannot describe the data. A key advantage of a statistical model is that all available hypotheses regarding the biological system are explicitly considered in the mathematical description. If the current model fails to explain the data, it can be rejected, indicating that the actual biological system is different from expected. As an iterative process, several additional hypotheses can now be included and tested. For example, these can be different molecular interactions between the reactants, unconsidered reactants or different levels of abstraction concerning the reaction rate equations or assumptions about model parameters. This process is iterated until one or more candidate models are defined that are able to reproduce the experimental data. Subsequently, model selection strategies [29] can be used to rank the candidates. Furthermore, experimental design techniques [30] enable new

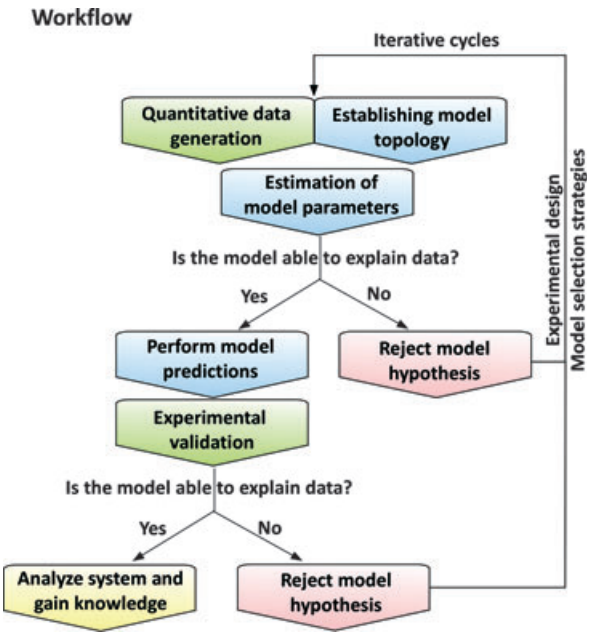

Fig. 2 *Workflow for establishing an ODE model. See text for details.*

experimental data to be created under suitable conditions that allow better differentiation between the models (Fig. 3a).

The ODE model is able to describe the experimental data

In the second scenario, the ODE model can describe the data. In this case, the model can be used for the prediction of cellular behaviour, such as the dynamics of components that are inaccessible by experiments or the dynamics of all network reactants under altered conditions (e.g. overexpression or knockdown of components or inhibitor treatments). Subsequently, the predictions can be used for validating or rejecting the current model. After gaining confidence in this particular mathematical representation of the biological system, its behaviour can be investigated. Here, it is important to consider that the model's behaviour is a statistical prediction. Therefore, uncertainties in the experimental data and in any prior assumptions used in the model have to be analysed to assess the accuracy of the model predictions.

Accurate predictions of cellular behaviour requires models with identifiable parameters

The scope of the ODE model should always corresponds to the availability of data and the biological question being addressed. ODE models are typically nonlinear with regard to the dependency of model trajectories on the parameters. In combination with limited availability of biochemical data, the problem of parameter identifiability may arise, which reduces the predictive power of the model [31]. Therefore, it is crucial to investigate the identifiability of the model parameters to facilitate accurate predictions. In the following section, we give a short theoretical overview of parameter estimation and methods to investigate the identifiability of model parameters.

Maximum likelihood estimation (MLE)

A maximum likelihood estimation (MLE) can be applied to calibrate the mathematical model by estimating the parameters of the ODE system from experimental data [32]. For mathematical definitions, refer to Box 1. If the measurement error of the experimental technique is normally distributed, the minimum of the residual sum of squares χ^2 yields the MLE of the model parameters. Optimization algorithms (e.g. see [33]) facilitate the numerical determination of the χ^2 minimum. For biochemical measurement techniques such as immunoblotting, the measurement noise is often log-normally distributed [34] and a log-transformation yields normally distributed noise. In MLE, parameter uncertainties are usually determined using confidence intervals. An interval indicates that the true value of the parameter is expected to be inside this interval with a certain probability [35]. It is critically important that uncertainties in the estimated parameter are considered and propagated to the model predictions if model predictions are the goal of the investigation. If the information contained in the experimental data is not sufficient, parameter uncertainties and

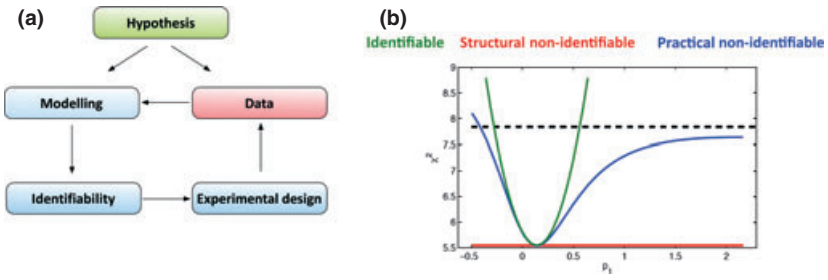

Fig. 3 *(a) To test a given hypothesis, experimental data are used for mathematical modelling. Frequently, model parameters cannot be uniquely determined based on the existing data. To resolve this issue, experimental design techniques allow for the planning of new experiments, resulting in additional data. (b) Investigating parameter identifiability using the profile likelihood approach. The coloured lines indicate profiles of the residual sum of squares χ^2 with respect to parameter p_1. Three typical cases can arise. The red profile is perfectly flat and indicates that the parameter is structurally nonidentifiable. The blue curve indicates a practically nonidentifiable parameter. Its lower confidence bound is finite, given by the crossing of the threshold indicated by the dashed line. Its upper confidence bound is infinite. The green curve indicates an identifiable parameter with finite confidence intervals.*

consequently uncertainties in the model predictions can be very large. Some parameters can even be undefined (i.e. nonidentifiable), leading to unconstrained model predictions. Identifiability and confidence intervals can be investigated by calculations of the profile of the likelihood [31].

Box 1 definitions

Ordinary differential equation (ODE) model

Cellular processes such as molecular interactions can be described by ODE models. The concentration dynamics of n compounds $\vec{x}$ such as proteins in different phosphorylation states is given by an ODE system

$$\dot{\vec{x}}(t,\theta) = \vec{f}(\vec{x}(t,\theta), \vec{u}(t), \theta). \tag{1}$$

The dynamical behavior may depend on a time dependent input function $\vec{u}(t)$ such as a treatment as well as model parameters $\theta = \{\theta_1 \ldots \theta_l\}$ such as rate constants and initial concentrations. The dynamical variables $\vec{x}$ are mapped to m model outputs $\vec{y}$, the quantities that are accessible by experiments at discrete times t_i, via a function

$$\vec{y}(t_i,\theta) = \vec{g}(\vec{x}(t_i,\theta),\theta) + \vec{\varepsilon}_i. \tag{2}$$

The outputs may depend on additional parameters such as scaling or offset parameters that are included in θ. Often, only a subset or combinations of compounds are accessible by experiments, meaning that $m < n$. The distribution of the measurement noise, e.g. $\varepsilon_{ki} \sim N(0, \sigma_{ki}^2)$, is assumed to be known.

Maximum likelihood estimation (MLE)

Often, many model parameters θ are unknown and have to be estimated from experimental data. The agreement of experimental $y_k^\dagger(t_i)$ with the predicted model out-put $y_k(t_i, \theta)$ for parameters θ is measured by an *objective function*, commonly the weighted sum of squared residuals

$$x^2(\theta) = \sum_{k=1}^{m} \sum_{i=1}^{d_k} \frac{1}{\sigma_{ki}^2} \left(y_k^\dagger(t_i) - y_k(t_i,\theta) \right)^2 \tag{3}$$

where d_k denotes the number of data points for each observable $k = 1 \ldots m$, measured at time points t_i with $i = 1 \ldots d_k$. The variances σ_{ki}^2 of the measurement noise are assumed to be known. The parameters can be estimated by finding the parameter values $\hat{\theta}$ that minimize $\chi^2(\theta)$, i.e. the best model fit. For normally distributed measurement noise, $\chi^2(\theta)$ is proportional to the log-likelihood and minimizing (3) corresponds to MLE.

Profile likelihood (PL)

The profile likelihood is defined by

$$\chi_{\mathrm{PL}}^2(\theta_i) = \min_{\theta_j \neq i} \left[\chi^2(\theta) \right]. \tag{4}$$

The idea of this approach is to detect flatness of the likelihood in a high dimensional space by exploring each parameter in the direction of least increase in $\chi^2(\theta)$. For each parameter θ_i a section is individually computed along the minimum of the objective function with respect to all the other parameters $\theta_j \neq i$. The profile likelihood enables to calculate *likelihood-based* confidence intervals. Here, a threshold Δ_α in the likelihood defines a confidence region

$$\left\{ \theta | \chi^2(\theta) - \chi^2(\hat{\theta}) < \Delta_\alpha \right\}. \tag{5}$$

Its borders represent confidence intervals. The threshold Δ_α is the α quantile of the χ_{df}^2-distribution, e.g. $\alpha = 0.95$ yields confidence intervals that contain the true value of the parameter with 95% probability.

The choice of *df* yields confidence intervals that hold jointly for *df* number of parameters. Often, $df = 1$ is applicable yielding confidence intervals that hold individually for each parameter.

Non-identifiability

Non-identifiability is characterized by flatness of the likelihood. The profile likelihood allows to investigate non-identifiability also in a high dimensional space, see Fig. 3b. A structural non-identifiability is characterized by a perfectly flat profile (red line). This indicates that the parameter cannot be determined at all. A practical non-identifiability is characterized by a profile that flattens out and stays below the threshold Δ_α for confidences intervals (blue line). This indicates that the lower and/or the upper confidence bound is infinite. Figure 3b shows a case where a lower bound around −0.5 can be determined. A profile of an identifiable parameter allows for determination of both lower and upper bound (green line).

Non-observability

The uncertainty of parameter estimates $\hat{\theta}$ translates to uncertainty of model trajectories. In particular non-identifiability can induce non-observability. For structurally non-identifiable parameters θ_{sub}, those components of $\vec{x}$ affected by θ_{sub} can be non-observable, whereas the model observables $\vec{y}$ are by definition invariant. In contrast, for practical non-identifiable parameters, the model observables $\vec{y}$ are affected but stay in agreement with the uncertainties of the experimental data because the likelihood stays below the threshold Δ_α. Nevertheless, some components of $\vec{x}$ might be affected strongly by a practical non-identifiability and hence might be non-observable. Also, confidence intervals of parameter estimates translate to confidence intervals of model trajectories.

(c.f. Raue *et al.* 2009)

Structural (non) identifiability

Given that only a subset of the molecular compounds involved in the model can be measured, some of the model parameters may be structurally nonidentifiable [36]. This means that several parameters can potentially cancel out their effects and produce exactly the same model output for the measurable compounds. Therefore, structural nonidentifiability indicates that these parameter values cannot be determined at all, given the current experimental conditions and measurements. Structural nonidentifiability can be detected by calculating the profile likelihood [37]. A perfectly flat profile indicates a structural nonidentifiable parameter (Fig. 3b, red line).

Practical (non) identifiability

A structural nonidentifiable parameter can still be practically nonidentifiable. This case can arise for nonlinear models and if the amount and quality of data is limited. It is characterized by having a unique MLE, but infinite confidence intervals. Likelihood-based confidence intervals can be calculated by setting a threshold for the residual sum of squares χ^2 [38]. Using the profile likelihood approach [31], a profile that increases but plateaus to a level below the threshold to one or both sides indicates practical non-identifiability (Fig. 3b, blue curve). Consequently, likelihood-based confidence intervals are infinite. If the amount or quality of experimental data is increased, a practical nonidentifiability will finally be resolved (Fig. 3b, green curve). Experimental design based on the profile likelihood can be used to plan suitable experiments [39]. Of note, it is often more informative to measure additional molecular compounds than to increase the quality of already available measurements.

After parameter uncertainties have been investigated, they can be translated to confidence intervals for the model predictions. Nonidentifiability of the model parameters often induces nonobservability of the model prediction that is affected by these parameters. This indicates that the predictions are undetermined given the current experimental setup and data.

Examples of predictive models revealing key system properties

Two examples of erythropoietin (Epo)-induced signalling in erythropoiesis are provided to demonstrate mathematical modelling of signalling pathways

including identifiability analysis. The development of red blood cells (erythrocytes) is a complex and fine-tuned process that results in a precisely controlled number of erythrocytes in the blood. This system ensures a continuous oxygen supply to tissues and at the same time avoids elevated haematocrit levels that potentially lead to thrombosis. Epo binds to the haematopoietic cytokine receptor, the EpoR, which is primarily present on erythroid progenitor cells, but has been recently identified on tumour cells. By eliciting signalling cascades such as the JAK2/STAT5, the PI3K/AKT and the MAPK pathways, the EpoR coordinates decisions for differentiation, proliferation and survival. Although most of the components that are involved in EpoR signalling have been identified, signal processing strategies and the underlying regulatory mechanisms that lead to quantitative and controlled cellular decisions in erythroid progenitor cells are not entirely understood. Therefore, the bottom-up systems biology approach is ideal for studying this system.

Example 1: strategies for processing ligand-encoded information

In physiological situations, the concentration of Epo in the blood can vary up to 1000-fold [40]. Becker *et al.* [41] showed by mathematical modelling of quantitative data and by experimental validation how the Epo–EpoR system on the membrane can process such a broad range of Epo concentrations to a linear signal response. With conventional biochemical techniques, however, this issue was difficult to address because receptor endocytosis, recycling and turnover are highly interlinked processes that are difficult to discern experimentally. To separate these nonlinear processes, a mathematical model of ligand–receptor interaction and trafficking kinetics was developed. This model was first calibrated with quantitative experimental data based on radiolabelled Epo and identifiability analysis was performed using the profile likelihood. After confirming structural and practical identifiability of the parameter estimates, the model could be applied to accurately predict key dynamic properties of the EpoR system. In this way, the turnover rate of EpoR was found to be most important for a linear signal transmission, which enables the cell to detect the broad range of physiological ligand concentrations (Fig. 4).

Example 2: linking the integral signal response of transcription factor activity with survival decisions

Extracellular Epo concentrations are linearly transmitted to the inside of the cell, but how is this information quantitatively forwarded to the nucleus to elicit cellular responses? Binding of Epo to its cognate receptor leads to rapid activation of JAK2 phosphorylation followed by phosphorylation of the latent transcription factor STAT5. Although STAT5 is known to be a crucial regulator of survival in erythroid progenitor cells [42–44], the quantitative link between survival and STAT5 responses has been difficult to address. This is the case, in particular, because phenotypic assays that determine survival responses are usually

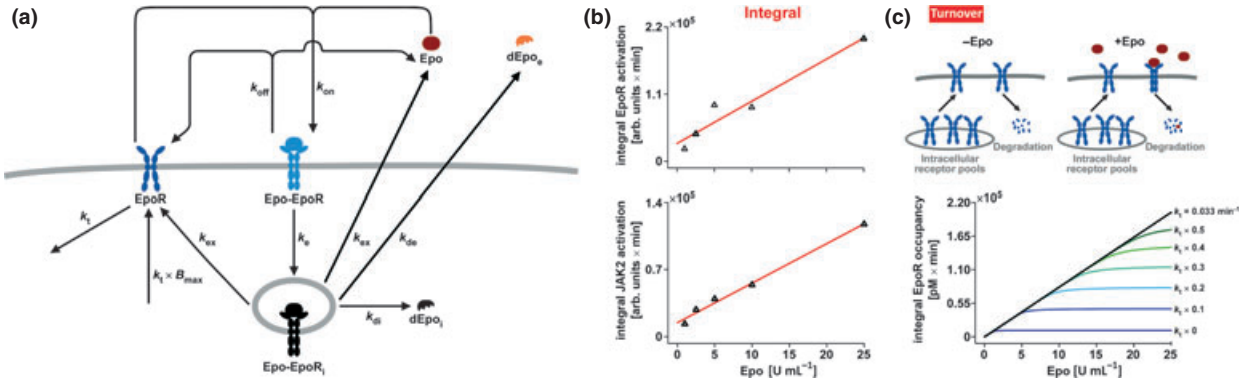

Fig. 4 *Information processing through the erythropoietin receptor (EpoR). (a) Graphical representation of a dynamic mathematical model for the EpoR system encompassing both ligand–receptor interactions and trafficking processes. (b) Time-course data for EpoR and JAK2 phosphorylation were acquired for different Epo concentrations by quantitative immunoblotting in BaF3-EpoR cells. A linear function was fitted to the data for the amount of activated EpoR and JAK2 integrated over time (integral activation, triangles). (c) Based on the calibrated and fully identifiable model, simulations for the amount of cell surface Epo–EpoR complexes (integral EpoR occupancy) were performed for parameter values lower than the estimated rate for EpoR turnover* k_t. Adapted from [41].

performed at Epo concentrations below the concentration threshold of biochemical experiments used to measure the activity of signalling proteins. To investigate the link between the integral STAT5 responses and survival, a dynamic pathway model of JAK2/STAT5 signalling was developed and calibrated with extensive quantitative and time-resolved data [45] (Fig. 5a,b). After assessing the identifiability of the parameters, the model was used to predict STAT5 responses over the entire range of Epo concentrations, including doses that are not accessible by experimental techniques. To evaluate the accuracy of the model predictions, the uncertainty of the parameter estimates were transferred to the model predictions by computing confidence bands of the model trajectories. It is interesting that the early signalling phase of STAT5 (1 h post stimulation) was most predictive of survival decisions (Fig. 5c). In line with this, it was demonstrated that early signalling events up to 90 min after receptor activation correlated best with apoptosis–survival decisions in HT-29 cells treated with tumour necrosis factor-alpha in combination with epidermal growth factor or insulin [46]. The next question to answer was: how is the STAT5 signal attenuated by the transcriptional feedback regulators CIS and SOCS3? Facilitated by the model,

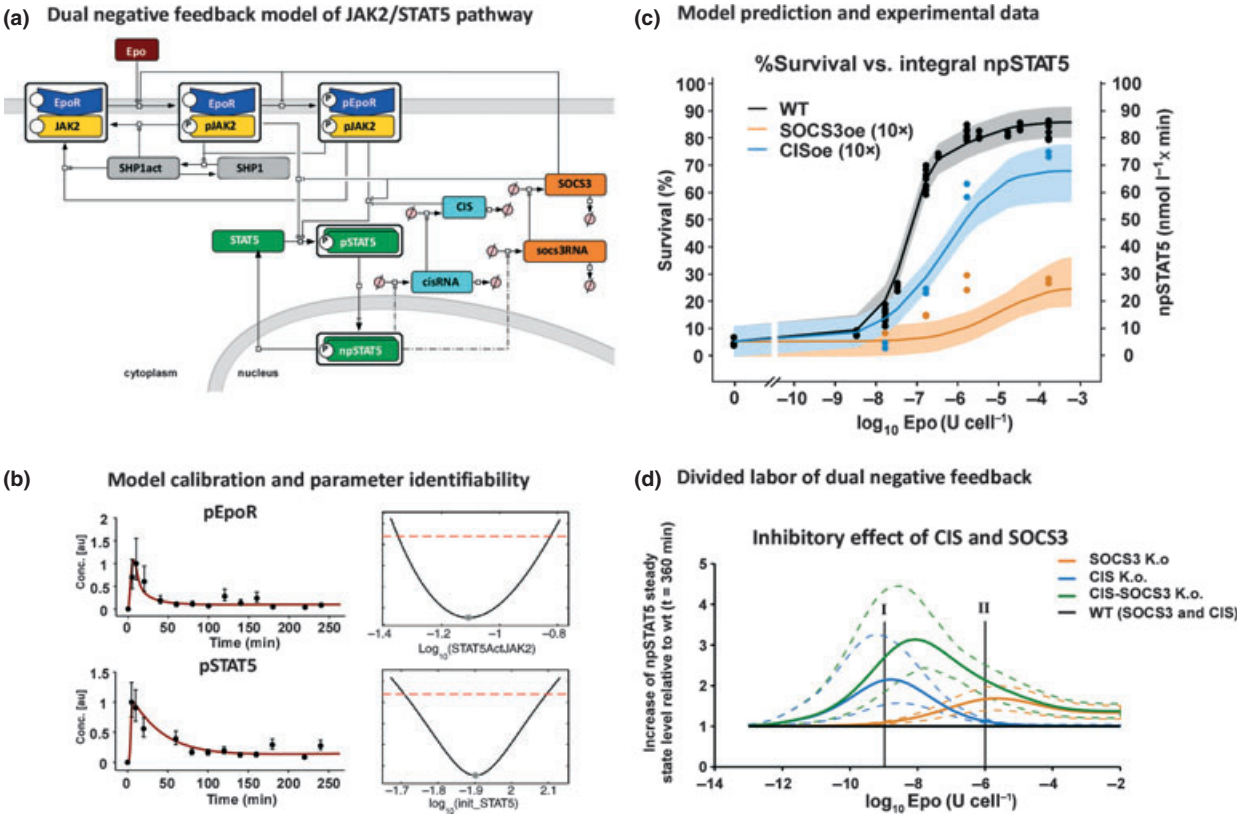

Fig. 5 *Information processing through the Epo-induced JAK2/STAT5 pathway. (a) Graphical representation of a dynamic mathematical model for the dual negative feedback of JAK2/STAT5 signalling. (b) Time-course data of JAK2/STAT5 signalling were acquired by quantitative immunoblotting in primary CFU-E cells. The model was fitted to the data and identifiability analysis was performed. Representative examples of the dynamics of phosphorylated EpoR and STAT5 are shown as well as parameter identifiability using the profile likelihood approach. (c) Based on the calibrated and identifiable parameters, simulations for the integral signal strength of phosphorylated STAT5 in the nucleus were performed including 95% confidence bands (shaded areas). Extent of survival for wild-type CFU-E cells, CIS- and SOCS3-overexpressing cells were experimentally determined at different Epo concentrations (circles). The overlay shows that the simulated integral of STAT5 (t = 60 min) correlates well with the survival rates. (d) The increase in steady-state pSTAT5 levels in the nucleus relative to wild-type cells (black line) was simulated in the presence of only one transcriptional negative regulator, CIS or SOCS3, and in the absence of both. For (I) Epo = 10^{-9} U per cell, the absence of CIS impacts STAT5 phosphorylation levels whereas the absence of SOCS3 has the major influence for (II) Epo = 10^{-6} U per cell. Dashed lines indicate upper and lower 95% confidence bands for the prediction. Adapted from [45].*

the individual inhibitory effects of these two transcriptional feedback regulators were analysed. The simulations revealed a division of labour by the two feedback proteins as the key property to control STAT5 responses. Model simulations of the STAT5 phosphorylation level in the nucleus demonstrated that the absence of CIS resulted in an increase in the amplitude and the steady-state level of STAT5 phosphorylation at low Epo concentrations (Fig. 5d, I) whereas the absence of SOCS3 caused an increase in the steady-state level at high Epo concentrations (Fig. 5d, II). The observation that dual feedback facilitates the tight regulation of transcription factor activity over a broad range of ligand concentrations suggests a new strategy of feedback control. In the future, studies that investigate STAT5 responses in single cells could clarify how the all-or-none response of survival–apoptosis is controlled by downstream effectors of STAT5. Here, a combination of theoretical approaches and single-cell observation techniques can provide new insights.

Perspective: from molecules to patients

Future goals in the combined approach of data-driven mathematical modelling are the integration of multiple cooperating and counteracting signalling pathways in large network models that have been comprehensively calibrated with quantitative dynamic data. Only then can predictive models with identifiable parameters be established that will be able to simulate cellular behaviour of experimentally unobservable conditions, to reveal regulatory mechanisms and to predict cellular responses to drugs. To achieve these aims, advances in mathematical methods, computational tools and experimental technologies are required. Quantitative experimental techniques have to be adapted to high-throughput as well as high-content measurement, requiring robotics and automated quantification methods. It will be essential to connect population-based and single-cell studies and therefore adapt methods such as mass spectrometric analysis and antibody-based techniques and combine them with single-cell analysis such as flow cytometry and imaging to quantitatively measure the dynamics of activated signalling proteins. Computationally, more efficient and faster methods for parameter estimations must be developed, as large models lead to parameter spaces with high dimensionality that require more elaborate methods to ensure that the global optimum is found. Ultimately, multiscale models need to be established that can describe cellular behaviour from individual cells to the multicellular level and to the organ and organism levels. This will lead to a deeper understanding of complex disease states, such as cancer, and facilitate a more target-oriented development of drugs with higher success rates and fewer adverse effects from the preclinical to the clinical phases.

Identification of the potential risk to patients with lung cancer who are treated with recombinant erythropoietin is a complex medical problem. The leading cause of cancer-related deaths worldwide is lung cancer, the most frequent form of which is non-small-cell lung cancer. Patients who are diagnosed at an advanced stage of disease are treated with surgery in combination with chemotherapy or angiogenesis inhibitors. As a common side effect, patients often develop cancer-related or chemotherapy-induced anaemia, which is frequently corrected with recombinant human Epo (rHuEPO). However, several clinical trials have shown that the outcome of patients with cancer receiving rHuEpo is impaired, despite successful correction of anaemia. In addition, the EpoR was recently identified on tumour cells [47, 48]. Consequently, the safety of Epo treatment in patients with cancer is now considered to be controversial (Fig. 6). To some extent, the higher mortality in patients with tumour treated with Epo may be attributed to an increase in thromboembolic events. However, potential tumour-promoting effects of Epo that affect angio-

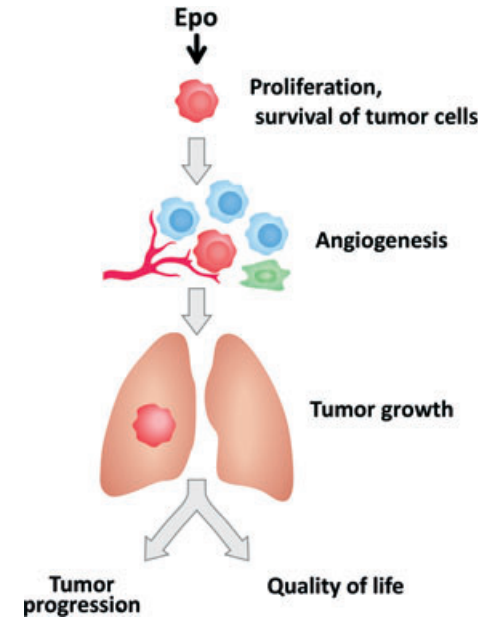

Fig. 6 *Potential risk of Epo treatment for patients with lung cancer.*

genesis [49] or counteract the efficacy of chemotherapy [50] have been reported.

As conventional experimental studies to investigate Epo-induced tumour-promoting effects resulted in contradictory results, multidisciplinary approaches that comprehensively elucidate the effects of Epo on emerging properties of intra- and intercellular communication in lung cancer (e.g. the German Ministry for Education and Research-funded MedSys Project LungSys) have been initiated. By employing comparative mathematical modelling of the dynamics of Epo-mediated receptor activation and intracellular signalling in tumour cells and endothelial cells, the Epo-specific effects on the tumour microenvironment are addressed. Furthermore, spheroid cultures and xenograft models are used to establish a cell-based multiscale model of the spatiotemporal organization of tumour growth and angiogenesis. Together with patient-derived data, this combined approach addresses the issue of whether the presence of EpoR is related to tumour-promoting functions. The established data-based mathematical models of Epo–EpoR dynamics and signalling pathways enable the prediction of new Epo treatment strategies. In summary, this multidisciplinary approach of addressing the complex pharmacological problem of Epo treatment will help to stratify the risk for patients with lung cancer and thereby contribute to improve their quality of life. As data-based mathematical models enable rapid testing of hypotheses, targeted design of experiments and the prediction of steps most suitable for intervention, this approach will contribute to a more rapid development of effective anticancer therapies.

Conflict of interest statement

No conflicts of interest to declare.

Acknowledgements

We acknowledge funding by the German Federal Ministry of Education and Research (BMBF)-funded MedSys-Network LungSys and the SBCancer network in the Helmholtz Alliance on Systems Biology.

References

1 Batchelor E, Loewer A, Mock C, Lahav G. Stimulus-dependent dynamics of p53 in single cells. *Mol Syst Biol* 2011; **7:** 488.
2 Marshall CJ. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* 1995; **80:** 179–85.
3 Kholodenko BN, Hancock JF, Kolch W. Signalling ballet in space and time. *Nat Rev Mol Cell Biol* 2010; **11:** 414–26.
4 Niethammer P, Bastiaens P, Karsenti E. Stathmin-tubulin interaction gradients in motile and mitotic cells. *Science* 2004; **303:** 1862–6.
5 Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 2009; **459:** 428–32.
6 Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 2008; **453:** 544–7.
7 Waks Z, Klein AM, Silver PA. Cell-to-cell variability of alternative RNA splicing. *Mol Syst Biol* 2011; **7:** 506.
8 Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol* 2006; **8:** 1195–203.
9 Chen WW, Niepel M, Sorger PK. Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev* 2010; **24:** 1861–75.
10 Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG. Systems biology: parameter estimation for biochemical models. *FEBS J* 2009; **276:** 886–902.
11 Ay A, Arnosti DN. Mathematical modeling of gene expression: a guide for the perplexed biologist. *Crit Rev Biochem Mol Biol* 2011; **46:** 137–51.
12 Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger DA. Logic-based models for the analysis of cell signaling networks. *Biochemistry* 2010; **49:** 3216–24.
13 Saez-Rodriguez J, Alexopoulos L, Zhang M, Morris MK, Lauffenburger DA, Sorger PK. Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. *Cancer Res* 2011; **71:** 5400–11.
14 Swameye I, Müller TG, Timmer J, Sandra O, Klingmüller U. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proc Natl Acad Sci U S A* 2003; **100:** 1028–33.
15 Taniguchi Y, Choi PJ, Li GW *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 2010; **329:** 533–8.
16 Kanehisa M, Araki M, Goto S *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 2008; **36:** D480–4.
17 Schilling M, Maiwald T, Bohl S *et al.* Quantitative data generation for systems biology: the impact of randomisation, calibrators and normalisers. *Syst Biol (Stevenage)* 2005; **152:** 193–200.
18 Aksamitiene E, Hoek JB, Kholodenko B, Kiyatkin A. Multistrip Western blotting to increase quantitative data output. *Electrophoresis* 2007; **28:** 3163–73.
19 Korf U, Derdak S, Tresch A *et al.* Quantitative protein microarrays for time-resolved measurements of protein phosphorylation. *Proteomics* 2008; **8:** 4603–12.
20 Sevecka M, Wolf-Yadlin A, MacBeath G. Lysate microarrays enable high-throughput, quantitative investigations of cellular signaling. *Mol Cell Proteomics* 2011; **10:** M110.005363.
21 Ciaccio MF, Wagner JP, Chuu CP, Lauffenburger DA, Jones RB. Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nat Methods* 2010; **7:** 148–55.
22 Hahn B, Böhm M, Raia V *et al.* One-source peptide/phosphopeptide standards for accurate phosphorylation degree determination. *Proteomics* 2011; **11:** 490–4.
23 Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* 2011; **80:** 273–99.

24 Eyrich B, Sickmann A, Zahedi RP. Catch me if you can: mass spectrometry-based phosphoproteomics and quantification strategies. *Proteomics* 2011; **11:** 554–70.
25 Bodenmiller B, Aebersold R. Quantitative analysis of protein phosphorylation on a system-wide scale by mass spectrometry-based proteomics. *Methods Enzymol* 2010; **470:** 317–34.
26 Schilling M, Pfeifer AC, Boh S, Klingmüller U. Standardizing experimental protocols. *Curr Opin Biotechnol* 2008; **19:** 354–9.
27 Klingmüller U, Bauer A, Bohl S *et al.* Primary mouse hepatocytes for systems biology approaches: a standardized in vitro system for modelling of signal transduction pathways. *Syst Biol (Stevenage)* 2006; **153:** 433–47.
28 Saez-Rodriguez J, Alexopoulos LG, Stolovitzky G. Setting the standards for signal transduction research. *Sci Signal* 2011; **4:** pe10.
29 Samanta T, Ghosh JK. Model selection - An overview. *Curr Sci* 2001; **80:** 1135–44.
30 Kreutz C, Timmer J. Systems biology: experimental design. *FEBS J* 2009; **276:** 923–42.
31 Raue A, Kreutz C, Maiwald T *et al.* Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 2009; **25:** 1923–9.
32 Fisher R. On the mathematical foundations of theoretical statistics. *Philos Trans R Soc Lond A* 1922; **222:** 309–68.
33 Conn AR, Gould NIM, Toint PL. *Trust-Region Methods*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2000.
34 Kreutz C, Bartolome Rodriguez MM, Maiwald T *et al.* An error model for protein quantification. *Bioinformatics* 2007; **23:** 2747–53.
35 Lehmann E, Leo E. *Theory of Point Estimation*. Weinheim, Germany: Wiley, 1983.
36 Walter E. *Identifiability of Parametric Models*. Elmsford, NY, USA: Pergamon Press, 1987.
37 Raue A, Kreutz C, Maiwald T, Klingmüller U, Timmer J. Addressing parameter identifiability by model-based experimentation. *IET Syst Biol* 2011; **5:** 120–30.
38 Meeker W, Escobar L. Teaching about approximate confidence regions based on maximum likelihood estimation. *Am Stat* 1995; **49:** 48–53.
39 Raue A, Becker V, Klingmüller U, Timmer J. Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos* 2010; **20:** 045105.
40 Jelkmann W. Molecular biology of erythropoietin. *Intern Med* 2004; **43:** 649–59.
41 Becker V, Schilling M, Bachmann J *et al.* Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science* 2010; **328:** 1404–8.
42 Socolovsky M, Nam H, Fleming MD, Haase VH, Brugnara C, Lodish HF. Ineffective erythropoiesis in Stat5a(-/-)5b(-/-) mice due to decreased survival of early erythroblasts. *Blood* 2001; **98:** 3261–73.
43 Cui Y, Riedlinger G, Miyoshi K *et al.* Inactivation of Stat5 in mouse mammary epithelium during pregnancy reveals distinct functions in cell proliferation, survival, and differentiation. *Mol Cell Biol* 2004; **24:** 8037–47.
44 Grebien F, Kerenyi MA, Kovacic B *et al.* Stat5 activation enables erythropoiesis in the absence of EpoR and Jak2. *Blood* 2008; **111:** 4511–22.
45 Bachmann J, Raue A, Schilling M *et al.* Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. *Mol Syst Biol* 2011; **7:** 516.
46 Gaudet S, Janes KA, Albeck JG, Pace EA, Lauffenburger DA, Sorger PK. A compendium of signals and responses triggered by prodeath and prosurvival cytokines. *Mol Cell Proteomics* 2005; **4:** 1569–90.
47 Hardee ME, Arcasoy MO, Blackwell KL, Kirkpatrick JP, Dewhirst MW. Erythropoietin biology in cancer. *Clin Cancer Res* 2006; **12:** 332–9.
48 Arcasoy MO, Jiang X, Haroon ZA. Expression of erythropoietin receptor splice variants in human cancer. *Biochem Biophys Res Commun* 2003; **307:** 999–1007.
49 Okazaki T, Ebihara S, Asada M, Yamanda S, Niu K, Arai H. Erythropoietin promotes the growth of tumors lacking its receptor and decreases survival of tumor-bearing mice by enhancing angiogenesis. *Neoplasia* 2008; **10:** 932–9.
50 Liang K, Esteva FJ, Albarracin C *et al.* Recombinant human erythropoietin antagonizes trastuzumab treatment of breast cancer cells via Jak2-mediated Src activation and PTEN inactivation. *Cancer Cell* 2010; **18:** 423–35.

Correspondence: Prof. Dr. Ursula Klingmüller, Systems Biology of Signal Transduction (A150), Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.
(fax: +49 6221 424488; e-mail: u.klingmueller@dkfz.de).