INCOME

INtegrative COllaborative Modelling in Systems MEdicine

ABSTRACTS OF INCOME2018

CONFERENCE & HACKATHON

OCTOBER 15-19, 2018

BERNRIED, LAKE STARNBERG, GERMANY

INTEGRATIVE PATHWAY MODELLING IN SYSTEMS BIOLOGY AND SYSTEMS MEDICINE

ABSTRACTS OF TALKS

MONDAY, OCTOBER 15

TUESDAY, OCTOBER 16

WEDNESDAY, OCTOBER 17

ABSTRACTS OF POSTERS

ABSTRACTS OF TALKS

KEYNOTE: SYSTEMATIC INTEGRATION OF MODELS AND DATA FOR YEAST GROWTH, DIVISION AND STRESS RESPONSE

EDDA KLIPP

Humboldt University of Berlin, Berlin, Germany

With the progress of genome-wide experimental approaches we witness the establishment of more and more libraries of genome-wide data for proteins or RNA or metabolites, especially for well-studied model organisms such as bakers' yeast. However, the separated consideration of metabolic networks or gene regulation networks does not tell us how these networks are integrated to allow a cell to grow, divide and respond to changing environments.

We use the yeast *Saccharomyces cerevisiae* as *the* model organism for eukaryotic cells allowing to comprehensively analyzing regulatory networks and their integration with cellular physiology. Here, we focus on processes during the cell division cycle and study the changes of signaling, metabolism, or ion transport during the growth of a single cell.

We use a modular and iterative approach that allows for a systematic integration of cellular functions into a comprehensive model allowing to connect processes that are strongly interlinked in cellular life, but measured separately. The modular concept also to zoom in and out if different aspects of regulation or dynamics become important.

ENSEMBLE MODELLING FOR DRUG RESISTANCE AND SENSITIVITY ON COLON CANCER CELL LINES

MIHALY KOLTAI, ANDREI ZINOVYEV AND EMMANUEL BARILLOT

Institut Curie, Paris, France

In the framework of a multi-group European project we use ensemble modeling to understand drug resistance and combinatorial drug effects in KRAS/BRAF mutant colon cancer cell lines. We use background knowledge (mutations, prior knowledge of signaling networks) and data from perturbation experiments to infer a logical model of drug resistance. With ensemble modeling we identify the required topological structures to explain combinatorial effects in KRAS/BRAF cell lines and to infer causality relations between cell cycle arrest and DNA repair components. One of the focus of our project is to use different data types to validate network models. Besides phosphoproteomics from perturbation experiments and flow cytometry of "cell fate" events, a novel aspect of our approach is using the results of a targeted CRISPRi screen for model selection, in the framework of stochastic logical models. A selected colon cancer cell line sensitive to both the single and double inhibition undergoes targeted deletions in a screening experiment. The deletions that have the most pronounced positive or negative growth effect will be used to distinguish between the different resistance mechanisms. Our aim is two-fold. On the one hand, by an analysis of network motifs, we want to explore the mechanisms leading to drug resistance in KRAS mutant cell lines. On the other hand, we are developing a larger logical model to predict drug effects and ultimately suggest potent drug combinations. In terms of biological relevance our study analyzes in detail the understudied interaction of two crucial cellular functions, namely MAPK signaling and DNA-repair and related cell cycle arrest. This is a question with high clinical relevance due to the importance of KRAS mutations in colon cancer patients who develop drug resistance.

INTEGRATIVE WORKFLOW FOR IDENTIFICATION OF DIAGNOSTIC AND THERAPEUTIC MARKERS IN TUMOR INVASION

FAIZ KHAN

University of Rostock, Rostock, Germany

In living cells, molecules interact with other molecules in a network to realize cellular functionality. Mutations in molecular factors

perturb the regulation of networks leading to dysregulation of cellular functionality, which associates with complex diseases such as cancer. Unraveling mechanisms underlying diseases has motivated the development of various systems biology approaches. Key challenges in systems biology approaches for mechanistic understanding of diseases are: (i) the large number of interacting components in molecular networks, and (ii) the nonlinear nature of spatio-temporal interactions constituting complex network structures including feedback/feedforward loops.

To address these challenges, I developed an integrative workflow (Khan el al. Nature Comm. 2017; doi:10.1038/s41467-017-00268-2) by combining techniques from bioinformatics and systems biology. The workflow combines network structure, omics and biomedical data, and dynamic modeling (logic-based) to understand the mechanism underlying complex diseases. Using the proposed workflow, I analyzed large-scale molecular interaction map of E2F1, a transcription factor involved in tumor invasion. It identified core-regulatory networks by ranking network sub-structures using multi-objective optimization function for epithelial-mesenchymal transition (EMT) in bladder and breast cancers, which are amenable for dynamical modeling. Using logic-based modeling formalism, the in silico stimulus-response analysis of the core networks detect molecular signatures for each cancer type. Further, I performed in silico perturbation experiments to identify therapeutic targets. The predicted molecular signatures and therapeutic targets were validated experimentally and through patient data. The computational analysis of biochemical networks can improve our understanding of disease processes in a mechanistic way. Ultimately, this shall provide the ability to manipulate and optimize processes towards treatment.

A BOOLEAN NETWORK MODEL TO PREDICT THE BEHAVIOUR OF COFILIN-1 IN PANCREATIC CANCER

SILKE KÜHLWEIN[1], JULIAN SCHWAB[1], MALTE BUCHHOLZ[2] AND HANS A. KESTLER[1]

[1]Ulm University, Institute of Medical Systems Biology, Ulm, Germany
[2]University of Marburg, Department of Gastroenterology, Endocrinology and Metabolism, Marburg, Germany

Pancreatic cancer is one of the most lethal cancer in developed countries with a five-year survival rate of less than 5% [1]. The actin severing protein cofilin-1 (CFL1) is overexpressed in pancreatic cancer which is associated with high invasiveness and poor prognosis [2]. In contrast, CFL1 knockdown experiments lead to cell-cycle arrest but no induction of apoptosis. We created a Boolean network model to predict reasons for the dysregulation of CFL1 and to evaluate its influence on the disease and its worse outcome. Boolean network models are simple mathematical models which can be used to study dynamic behaviour. The simplicity of these models arises through the assumptions that genes are considered as expressed or not expressed. For the setup of Boolean network models, we use qualitative knowledge about regulatory interactions taken from literature statements. Simulations reveal sequences of states, called attractors, determining the long-term behavior of the system. All initial states leading to the same attractor are part of its basin of attraction. In a biological context, these attractors can be associated to phenotypes [3]. In our simulation we obtain two attractors representing cancer (active CFL1) and apoptosis (inactive CFL1) while knockout of CFL1 yields attractors representing cell-cycle arrest without apoptosis. To validate our model, we performed in vitro laboratory experiments. Here, first predictions based on our established Boolean network model could be confirmed.

[1] Blum et al. Metabolism addiction in pancreatic cancer. Cell Death and Disease 5, e1065 (2014)

[2] Satoh et al. Immune-complex level of cofilin-1 in sera is associated with cancer progression and poor

prognosis in pancreatic cancer. Cancer Science. (2017)

[3] Kauffman. The origins of order. Self-organization and selection of evolution. Journal of Evolutionary Biology 7, 4 (1994), 518-519.

SIGNAL INTEGRATION IN ALTERNATIVE SPLICING NETWORKS

STEFAN LEGEWIE

Institute of Molecular Biology, Mainz, Germany

Alternative splicing increases protein diversity in eukaryotic cells, and plays an important role in development and tissue identity, but also in diseases such as cancer. Splicing reactions are catalyzed by the spliceosome and modulated by auxiliary RNA-binding proteins (RBPs) which recognize nearby RNA sequences and guide spliceosome activity. The rules how multiple protein complexes dynamically interact on pre-mRNA sequence to control splicing remain poorly understood. Here, we comprehensively characterize the cis-regulatory landscape controlling a cancer-relevant alternative splicing decision in the RON proto-oncogene using high-throughput random mutagenesis. Mathematical modelling of splicing kinetics reveals a dense regulatory landscape and enables us to identify more than 1,000 mutations affecting RON exon 11 skipping, which corresponds to the pathological isoform RONΔ165. Importantly, the effect correlate with RON alternative splicing in cancer patients bearing the same mutations. To characterize how RBPs interpret the pre-mRNA sequence and establish context-specific splicing patterns, we performed a comprehensive genetic interaction screen, in which the effect of RBP. knockdowns is assessed on the complete minigene library. Using a model-based synergy analysis, we pinpoint the functionally most relevant RBP binding sites, and follow up with independent validation experiments. In particular, we show how cooperative binding of the RNA-binding factor HNRNPH facilitates a splicing switch of RON exon 11. Our approach offers new insights into splicing regulation and the impact of mutations on alternative splicing in human patients

KEYNOTE: LOGIC MODELS OF LARGE SIGNALING NETWORKS FROM OMICS DATA

JULIO SAEZ-RODRIGUEZ

Heidelberg University, Heidelberg, Germany

Dynamic mechanistic models are a powerful tool to understand biological networks. Kinetic modeling can provide deep insights, but it requires detailed knowledge and abundant data, and does not scale up well. Logic modeling is an alternative formalism that for its simplicity can be used to model much larger networks with fewer knowledge and data. Over the years, we have developed methods and tools to apply this logic formalism to build context-specific models, with a focus on signaling networks and the use of data obtained upon perturbation. Our general pipeline involves obtaining existing prior knowledge on pathways from available public resources using our tool OmniPath (www.omnipathdb.org), building a logic model from this prior knowledge, and training it to data with our tools CellNOpt and Phonemes (www.cellnopt.org).

Formalism variants allow us to handle variables as either Boolean (binary) or continuous and even to be casted as differential equations. I will describe recent methodological developments, including the use of gene expression to build such models, and extensions to model metabolic regulation. I will illustrate their utility in cases of biomedical relevance, in particular to improve our understanding of cancer and develop novel therapeutic opportunities.

KEYNOTE: AUTOMATED MODEL ASSEMBLY FROM NATURAL LANGUAGE AND SCIENTIFIC LITERATURE USING INDRA

BENJAMIN M. GYORI

Harvard Medical School, Boston, USA

Building computational models of molecular pathways is a laborious process involving manual information gathering and model implementation. This makes it difficult to build models based on known mechanisms, creating a gap between the

scope of typical models, and the omics data they are meant to help interpret. We developed INDRA (the Integrated Network and Dynamical Reasoning Assembler), a system for automatically assembling models of biochemical mechanisms directly from English language (including from the scientific literature) and from pathway databases. INDRA interfaces with natural language processing systems to extract mechanisms from phrases such as "GRB2 binds EGFR that is phosphorylated on a tyrosine residue" and uses knowledge assembly algorithms to fix certain errors, resolve redundant knowledge, infer missing information, and assess belief. This intermediate knowledge can then be used to produce models in different formalisms, including rule-based mechanistic models, and causal graphs. We present applications of INDRA including (i) a dynamical model automatically built from an English language description which can reproduce resistance to targeted inhibition in BRAF-V600E-driven melanoma cells (ii) a model of growth-factor signaling built by machine-reading ~95,000 scientific publications, which can provide mechanistic explanations to drug perturbations in a melanoma cell line, and (iii) a human-machine dialogue system in which a user can gather information and build a mechanistic hypothesis for an observed phenomenon by talking with a computer partner.

KEYNOTE: MODELLING OF PATHWAYS: SOME RECENT METHODOLOGICAL INSIGHTS

CLEMENS KREUTZ

University of Freiburg, Freiburg, Germany

ODE models are frequently applied for describing cellular processes. Statistical and numerical approaches are challenged by the nonlinearity and the large number of parameters of these models. Since there are many technical pitfalls, well-established procedures and guidelines are required.

In this talk, ongoing methodological research and conceptual insights for modelling pathways with ODE models are presented. This covers topics like optimization for parameter estimation and approaches for uncertainty analysis as well as choosing proper observation functions and adequate noise models. Moreover, benchmarking results and strategies are presented for deriving general guidelines for data-based modelling.

OPEN PROBLEMS IN OBSERVABILITY AND STRUCTURAL IDENTIFIABILITY

ALEJANDRO F. VILLAVERDE AND JULIO BANGA

University of Vigo, Institute for Marine Research, Vigo, Spain

Observability is the ability to infer the internal state of a system from observations of its output. A related property, structural identifiability (SI), informs about the theoretical possibility of determining the values of the model parameters from the output. By considering the parameters as constant state variables, SI becomes a particular case of observability. It is thus possible to analyze the observability and SI of a model simultaneously, using the tools provided by differential geometry. The core element of this approach is the assessment of observability/SI by calculating the rank of an observability-identifiability matrix, which is constructed using Lie derivatives. This idea can be applied to systems of nonlinear ordinary differential equations, which are well suited for many biological systems such as pathway models.

This talk presents recent developments and discusses open problems in this area, such as: The role of initial conditions: some initial conditions may cause loss of identifiability; how can they be found (efficiently)?

- The role of inputs: extended Lie derivatives enable the analysis of systems with time-varying inputs. But...
  - ... is an experiment with time-varying input equivalent to several experiments with constant inputs?
  - ... how to deal with unknown time-varying inputs (dynamic perturbations)?
- Model reformulation: how to find symmetries that lead to identifiable parameter-state combinations?
- From local to global: the differential geometry determines local observability/SI; is it possible to obtain globally valid results?

- Predicting the time course of unmeasured states: is SI a requirement? Is observability sufficient?

Since the above issues are not fully answered, this talk seeks to engage the audience in an open discussion with the goal of coming up with possible directions.

BAYESIAN ANALYSIS COMBINED WITH GLOBAL SENSITIVITY ANALYSIS APPLIED TO DYNAMICAL INTRACELLULAR PATHWAY MODELS

OLIVIA ERIKSSON[1], ANDREI KRAMER[1] AND ALEXANDRA JAUHIAINEN[2]

[1]KTH Royal Institute of Technology, Stockholm, Sweden
[2]AstraZeneca, Sweden

Intracellular dynamical models increase in size and complexity as our knowledge about biological mechanisms improves, along with better algorithms and increased computing power. However, especially in the early stages of modeling the number of parameters far outweighs the available information from experimental data resulting in uncertainty in the parameter estimates as well as in the predictions made from the model. Finding good parameterizations is a difficult, ill-posed inverse problem, and we tackle it through Bayesian methods, global sensitivity analysis and iterative procedures via copula approximations [1]. We identify parameters which contribute to the uncertainty of the model predictions the most, as well as model parts (such as individual parameters, correlated parameter groups, and uncorrelated pairs) which have a large influence on qualitatively different behaviors of the model. We used approximate Bayesian computation (ABC) to fit the model and quantify the parameter uncertainty described by the posterior distribution. We then performed a global sensitivity analysis based on the posterior distribution while considering model behaviors for which we had no data. As a test case, this approach was applied on a simplistic model consisting of CaMKII and calcineurin, the balance of which is considered relevant for predicting synaptic plasticity.

[1] Eriksson, O. and Jauhiainen, A. et al.; Uncertainty quantification, propagation and characterization by Bayesian analysis combined with global sensitivity analysis applied to dynamical intracellular pathway models, Bioinformatics, bty607, https://doi.org/10.1093/bioinformatics/bty607

MODEL REDUCTION OF SMALL METABOLIC-GENETIC NETWORK

NEVEEN ALI SALEM ESHTEWY[1], LENA SCHOLZ[2] AND ALEXANDER BOCKMAYR[1]

[1]Freie Universität Berlin, Berlin, Germany
[2]TU Berlin, Berlin, Germany

Integrated modeling of metabolism and gene regulation continues to be a challenging problem in computational systems biology. In 2001, Covert et al. introduced a method called regulatory Flux Balance Analysis (rFBA), which combines a stoichiometric model of metabolism with Boolean rules for transcriptional regulation. rFBA models are solved in an iterative way, alternating between flux balance analysis and applying the Boolean rules. Due to the steady-state assumption of flux balance analysis, rFBA can handle only external metabolite concentrations. To overcome this limitation, we study the transformation of a rFBA model into a kinetic model, which includes concentrations for both internal and external metabolites. In order to reduce the dimension of the resulting model, we use proper orthogonal decomposition (POD), which is a model reduction method based on singular value decomposition of a snapshot matrix and Galerkin projection. We illustrate the approach on a small rFBA model of core carbon metabolism.

20 BENCHMARK PROBLEMS FOR DYNAMIC MODELING

CAROLIN LOOS[1], HELGE HASS[2], ELBA RAIMUNDEZ ALVAREZ[1], JAN HASENAUER[1] AND CLEMENS KREUTZ[2]

[1]Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany
[2]University of Freiburg, Freiburg, Germany

Ordinary differential equation (ODE) models which describe the temporal evolution of the abundance

of biochemical species are widely used to study dynamic cellular processes. These models generally have parameters, such as initial conditions or rate constants, which cannot directly be measured. Thus, methods to estimate these parameters from experimental data are continuously developed. However, a rigorous assessment of the methods in application settings is hindered by the deficiency of realistic data provided with biochemical signaling networks and lack of suitable benchmark models.

We compiled a collection of 20 benchmark models that can be used to evaluate and compare newly developed and existing methods for, e.g., parameter estimation and uncertainty analysis. We provide the models as ODEs and in the widely used machine-readable SBML format, together with experimental measurement data. The models cover a broad range of model properties, varying in problem size, complexity and numerical demands. Performing example analyses, in which we studied scaling properties of optimization algorithms and the influence of log-transformation of parameters on optimizers, we illustrate the possibilities enabled by our provided benchmark collection.

KEYNOTE: INTEGRATING TUMOUR DATA AND BIOLOGICAL KNOWLEDGE THROUGH DYNAMIC MODELLING

DIRK FEY[1,2]

[1]Systems Biology Ireland, University College Dublin, Belfield, Dublin 4, Ireland
[2]School of Medicine, University College Dublin, Belfield, Dublin 4, Ireland

Signaling networks are genetically encoded systems within cells that respond to environmental changes such as drug treatments. The behavior of these networks connects genotype to phenotype in a time-dependent manor, thereby dictating cell behavior and - to a great extent - clinical outcome. But how does this computation take place, and how can it be modelled on the level of individual patients? Here, I will address this questions focusing on the on the dynamic signaling networks that are activated in response to chemotherapeutical drugs, and in particular the interconnected JNK/p53/BRCA1 stress-response, DNA-damage and -repair networks in neuroblastoma, breast- and ovarian cancer. Firstly, I will present a generally applicable method for integrating static cancer-cell and tumor-profiling data into patient-specific dynamic models of cancer signaling [3]. Secondly, I will show how dynamic modelling of the JNK/p53/BRCA1 response network can explain drug-sensitivity data in a large panel of cancer cell-lines from the GDSC database. Thirdly, I will show how these patient-specific models can be used to stratify neuroblastoma patients (n=688 patients, 7 matched primary and relapsed tumors), with a particular focus on the robustness and reliability of these predictions. Going beyond mere biomarkers, these patient-specific simulations also revealed deep insight into disease mechanisms. Last but not least, I will present our progress on modelling the cell-to-cell variability during the development of drug resistance. Our theoretical and experimental results indicate that so-called cell-ensemble modelling can be used model the selective pressure of chemotherapy on cancer-cell populations.

[1] Kolch, W. and D. Fey, *Personalized Computational Models as Biomarkers.* J Pers Med, 2017. **7**(3).

[2] Fey, D., et al., *Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients.* Sci Signal, 2015. **8**(408): p. ra130.

[3] Fey, D., A. Kuehn, and B.N. Kholodenko, *On the personalised modelling of cancer signalling.* IFAC-PapersOnLine, 2016. **49**(26): p. 312-317.

KEYNOTE: MECHANISTIC MODELING OF PAN-CANCER SIGNALING NETWORKS

MARC R. BIRTWISTLE

Clemson University, Clemson, USA

Most cancer cells harbor multiple drivers whose epistasis and interactions with expression context clouds drug and drug combination sensitivity prediction. We constructed a mechanistic computational model that is context-tailored by omics data to capture regulation of stochastic proliferation and death by pan-cancer driver pathways. Simulations and experiments explore how the coordinated dynamics of RAF/MEK/ERK and PI-3K/AKT kinase activities in response to

synergistic mitogen or drug combinations control cell fate in a specific cellular context.

COMPARATIVE NETWORK RECONSTRUCTION TO IDENTIFY SELECTIVE ANTI-CANCER DRUG COMBINATIONS.

EVERT BOSDRIESZ[1], JOAO MANUEL NETO[1], ANJA SIEBER[2], RENE BERNARDS[1], NILS BLÜTHGEN[2] AND LODEWYK WESSELS[1]

[1]Netherlands Cancer Institute, Amsterdam, Netherlands
[2]Charité – Universitätsmedizin Berlin, Berlin, Germany

Targeted inhibition of proteins involved in signaling is an important treatment strategy in cancer. However, due to network effects, a single drug is often ineffective and combinations of multiple drugs are required. Because of the enormous number of possible combinations, a systematic method to identify drug combinations that potently and selectively inhibit cancer cells is required.

Here, we present a computational pipeline to predict selective drug combinations based on perturbation experiments, and its application to a normal breast cell line (MFC10A) and an clone with an oncogenic PI3K mutation [1]. Based on the response to MAPK and PI3K-pathway inhibitors, we first reconstruct the signaling networks of both cells lines using “Comparative Network Reconstruction” [2]. In the mutant, phospho-AKT is less responsive to changes in IGF1R, PI3K and EGFR. The network model was able to accurate predict the changes in phosphorylation of the network nodes in response to drug-combinations not used in training the model (Pearson-correlation ~ 0.84). Next, we trained a model relating changes in phospho-AKT and phospho-ERK to changes in cell viability. The good model performance (Cross-validated Pearson-correlation ~ 0.7) shows that cell viability can be predicted from changes in signaling output. Importantly, when we combined these two models (trained on single drug data only) to predict the cell-viability of drug-combinations, there was a good Pearson-correlation of 0.55 between simulated and measured viability.

These results show the feasibility of this approach to prioritize promising multi-drug combinations. We are currently combing the single and two-drug combination data to predict the effect of multi-drug combinations and use this to find combinations that are maximally selectivity for the PI3K mutant.

[1] F. Di Nicolantonio et al., PNAS 105, 20864–20869 (2008).

[2] E. Bosdriesz et al., Bioinformatics, (In press, Biorxiv dio: 10.1101/243709)

COMPUTING BIOLOGICALLY RELEVANT CRN STRUCTURES USING TIME-SERIES DATA

ZOLTAN TUZA AND GUY-BART STAN

Imperial College London, London United Kingdom

High-throughput data acquisition in Systems Biology leads to an abundance of data that need to be processed and aggregated into biologically relevant dynamical models. The dynamics of signaling, metabolic and gene regulatory networks are usually captured by ODE models, where the right hand side consists of polynomials or rational functions. Automatically building such ODE models, based on this wealth of data, is of paramount importance to gain insight into an existing biological process.

This type of model building approach is gaining traction in the literature where a dictionary of possible right-hand-side functions---assembled using domain expert knowledge---is presented to an algorithm along with the time-series data. Then, the algorithm builds the most suitable (e.g. parsimonious) combination of dictionary functions as an ODE model.

In this work, we combine this model building technique with Chemical Reaction Network Theory (CRNT) to automatically generate an ODE model, as well as all possible reaction graphs. For the model building part, we use Sparse Bayesian Learning, which automatically builds a nonlinear ODE model. It should be emphasized that compared to common `simulate-and-compare' parameter estimation approaches this algorithm does not require solving ODEs.

To compute the CRN structures from an ODE model, we exploit the fact that multiple network structures may exist for a given ODE; even if the parameter are known *a priori*. The existence of multiple network structures – without noise – for a given dynamic has been investigated extensively in the literature. We use these results to compute the network structures, from an automatically built model.

Using this workflow, the wealth of data can be translated into a biologically relevant ODE model as well as reaction graphs. The set of possible graphs give an overview of all possible network connections, and allows us to identify targets for further perturbation experiments.

DMOD - A DEVELOPMENT LIBRARY FOR DYNAMIC MODELING IN R

DANIEL LILL, DANIEL KASCHEK AND JENS TIMMER

University of Freiburg, Freiburg, Germany

The complexity of dynamic modeling projects in systems biology has led to the development of several toolboxes. We present the modeling environment dMod implemented in the popular programming language R. Key challenges of dynamic modeling such as parameter estimation, uncertainty assessment or model selection are met with state of the art methods such as integration with sensitivity equations, trust-region optimizers, profile likelihood and L1 regularization. The underlying design philosophy of dMod is to split the mathematical model into its three main components: Parameter transformation, dynamic model and observation function. This atomic approach greatly enhances the flexibility of the user since it is possible to incorporate for example new experimental conditions just by adding another parameter transformation function without having to recompile the dynamic model itself.

dMod has successfully been used in academic research as well as found its way into pharmaceutic research in industry. The implementation in R allows access to the new developments within the R language itself. dMod integrates seamlessly with popular R packages such as dplyr and many packages from Bioconductor for data pre-processing. Models can interactively be visualized in a shiny app, which enhances communication to biologists. An import for SBML model allows to take advantage of the sophisticated features of dMod for already existing models.

Furthermore, a newly introduced interface, the dMod.frame, allows to easily manage, test and compare multiple model hypotheses in parallel. As a single R object containing the whole modeling process, dMod.frames are readily interchanged between users and computers, e.g. for simple parallelization of time consuming computations or to communicate results among collaborators and enhance reproducibility of results.

ANNOTATION OF MODELS AND HOW TO BENEFIT FROM THEM

TOM GEBHARDT AND FAIZ KHAN

University of Rostock, Rostock, Germany

Model annotation is essential for efficient sharing, usage and combination of encoded knowledge in the biomedical research. MIRIAM (Minimal Information Required In the Annotation of Models) is an initiative to produce a set of guidelines for consistent annotation and curation of biological models. It established the essential, minimal set of information that is sufficient to annotate a model to enable its reuse.

In this talk we will present how to annotate models in CellDesinger using MIRIAM. We will show how to annotate nodes (which can be proteins, genes or complexes and their multiple instances) and reactions with a small example. Furthermore, will show how consistent annotations can enable collaborations in model development and knowledge combination.

FAIRDOMHUB AND ITS WRITE API

WOLFGANG MÜLLER

Heidelberg Institute for Technology (HITS) gGmbH, Heidelberg, Germany

Within this talk we will shortly present the FAIRDOMHub, a platform for FAIR collection, storage, annotation, as well as distributed cataloguing and data publicaton of life sciences

data, SOPs and models. It is built towards keeping track of assets in big projects, involving multiple sources and sinks of data, SOPs, and models. We give a short overview about the FAIRDOMHub's functionality, including a small glimpse on its use of JWS Online, and on the JSON API that allows automated read/write of assets, which is of strong interest when storing models along with rich sets of training data.

Throughout the hackathon WM will be available for questions and hacking sessions.

FAIRDOMHub website:

https://www.fairdomhub.org

Example of use of possibilities of FAIRDOMHub:

https://fairdomhub.org/investigations/51

ABSTRACTS OF POSTERS

A MODELING OF SIGNALING, METABOLIC AND GENE REGULATORY PATHWAYS

A1 TRACKING CARBONS IN GLYCOLYSIS AND THE TCA CYCLE IN NEURON CELLS

ROMAN RAINER AND EDDA KLIPP

Humboldt University of Berlin, Berlin, Germany

Immature (d0) and mature (d6) LUHMES cells are differently susceptible to neurotoxins. This difference in toxin resistance is not well studied in terms of metabolism changes from immature to mature neurons. A previous study concluded that LUHMES cells lose their glutamine dependency by differentiating from d0 to d6 cells. We built a kinetic ODE model based on 13 C glucose and glutamine experiments (pulsed stable isotype-resolved metabolomics, pSIRM) to track the carbons thereby finding differences in metabolic behavior. The experiment offers time-resolved 13 C incorporation for different mass shifts of metabolites. Additionally, metabolomics, proteomics and secretion/uptake rates are available. To track all the carbon, every combination of 12C and 13C for every metabolite has to be tracked (two to the power of #carbons), but through analysis of the pathway and cross-checking with the data we could reduce the number of needed ODE equations for all possible patterns of 12C and 13C significantly. The advantage of the tracking of carbon is that it allows estimating kinetic parameters in an otherwise steady system. To have estimates on the incoming/outgoing fluxes for the model we used TFA (Thermodynamics-based Flux Analysis) on a human GEM (Genome-Scale Metabolic Reconstructions). This ensured that we do not investigate an isolated system. Regarding the finding of differences between d0 and d6 cells, we are using L1 regularization to minimize the change of parameters from immature to mature cells. We are setting the hyperparameter of the regulation, due to lack of alternatives, via the scanning of possible values.

A2 MODEL REDUCTION OF A SMALL METABOLIC-GENETIC NETWORK

NEVEEN ALI SALEM ESHTEWY[1], LENA SCHOLZ[2] AND ALEXANDER BOCKMAYR[1]

[1]Freie Universität Berlin, Berlin, Germany
[2]TU Berlin, Berlin, Germany

Integrated modeling of metabolism and gene regulation continues to be a challenging problem in computational systems biology. In 2001, Covert et al. introduced a method called regulatory Flux Balance Analysis (rFBA), which combines a stoichiometric model of metabolism with Boolean rules for transcriptional regulation. rFBA models are solved in an iterative way, alternating between flux balance analysis and applying the Boolean rules. Due to the steady-state assumption of flux balance analysis, rFBA can handle only external metabolite concentrations. To overcome this limitation, we study the transformation of a rFBA model into a kinetic model, which includes concentrations for both internal and external metabolites. In order to reduce the dimension of the resulting model, we use proper orthogonal decomposition (POD), which is a model reduction method based on singular value decomposition of a snapshot matrix and Galerkin projection. We illustrate the approach on a small rFBA model of core carbon metabolism.

A3 MINIBATCH OPTIMIZATION: A METHOD FOR TRAINGING MECHANISTIC MODELS USING LARGE DATA SETS

PAUL STAPOR AND JAN HASENAUER

Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany

In recent years, a number of rich public data bases, such as the Cancer Cell Line Encyclopedia (CCLE), have emerged. They constitute a powerful tool for computational studies of complex diseases such as cancer and make it possible to develop comprehensive mechanistic models, e.g., based on ordinary differential equations (ODEs). Such models are necessary for a detailed understanding of the underlying cellular processes.

However, when dealing with ODE-models and big data sets, parameter estimation becomes the limiting factor in model development, as computation time increases linearly with the number of experimental data sets used for model training. This prohibits the use of ODE-models to either small models or limits amount of data to train the model on.

Mini batch optimization methods can be employed to circumvent the linear scaling of computation time with the number of the data sets. Although being widely used for model training of deep neural nets, these methods have never been applied for the training of ODE-models, to the best of the authors knowledge. One reason for this may be that this transfer is not straight forward. Typical optimization problems have different properties in both fields, e.g., in ODE-modeling, the underlying ODEs can be numerically not solvable for certain regions of the parameter space. This makes it necessary to adapt mini batch optimization methods for the application to ODE-model training.

We show that it is possible to transfer some of the common mini batch optimization algorithms to the field of ODE-model training. Using adequate adaptations, we can substantially reduce computation time for model training of large-scale ODE-models, while keeping good convergence properties. We moreover present additional approaches, which may help to further improve the performance of the presented methods.

A4 EFFICIENT PARAMETERIZATION OF LARGE-SCALE DYNAMIC MODELS USING RELATIVE PROTEIN, PHOSPHO-PROTEIN AND PROLIFERATION MEASUREMENTS

LEONARD SCHMIESTER[1], YANNIK SCHÄLTE[1], FABIAN FRÖHLICH[2], JAN HASENAUER[1] AND DANIEL WEINDL[1]

[1]Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany
[2]Laboratory of Systems Pharmacology, Harvard Medical School, 200 Longwood Ave, Boston, MA, 02115, USA

Mechanistic models of signaling pathways provide the means to integrate heterogeneous data and to gain a quantitative understanding of cellular physiology. However, many models only focus on individual pathways, neglecting many others and ignoring any cross talk. Larger models on the other hand come with big challenges for parameter estimation, not only in the form of high computational demand. In particular, we found that relative measurements - which is what most large-scale datasets are - drastically impair optimizer performance.

In this study, we consider a large-scale ordinary differential equation model of cancer-related signaling (>4000 kinetic parameters, >1000 state variables) and parameterize it using relative measurements of protein, phospho-protein and proliferation data from cancer cell lines. We demonstrate how a novel hierarchical optimization approach in combination with adjoint sensitivity analysis can be applied to recover optimizer performance and parameterize large models. We show how our approach allows computing proportionality factors, offset parameters and error model parameters analytically, thereby simplifying the numerical optimization problem considerably and rendering it solvable by previously failing optimizers. This hierarchical optimization approach only needs a fraction of the computation time of the standard optimization method to reach good

objective function values. Furthermore, we show that this approach allows us to estimate error model parameters with negligible computational overhead when no experimental estimates are available. The estimated distribution of measurement errors provides unbiased means to weight heterogeneous datasets.

Overall, our hierarchical optimization approach allows for the efficient parameterization of large-scale dynamic models based on heterogeneous relative measurements and can easily be adopted by other researchers.

B MODELING OF DISEASES AND THERAPIES

B2 COMPARATIVE NETWORK RECONSTRUCTION TO IDENTIFY SELECTIVE ANTI-CANCER DRUG COMBINATIONS.

EVERT BOSDRIESZ[1], JOAO MANUEL NETO[1], ANJA SIEBER[2], RENE BERNARDS[1], NILS BLÜTHGEN[2] AND LODEWYK WESSELS[1]

[1]Netherlands Cancer Institute, Amsterdam, Netherlands
[2]Charité – Universitätsmedizin Berlin, Berlin, Germany

Targeted inhibition of proteins involved in signaling is an important treatment strategy in cancer. However, due to network effects, a single drug is often ineffective and combinations of multiple drugs are required. Because of the enormous number of possible combinations, a systematic method to identify drug combinations that potently and selectively inhibit cancer cells is required.

Here, we present a computational pipeline to predict selective drug combinations based on perturbation experiments, and its application to a normal breast cell line (MFC10A) and a clone with an oncogenic PI3K mutation [1]. Based on the response to MAPK and PI3K-pathway inhibitors, we first reconstruct the signaling networks of both cells lines using “Comparative Network Reconstruction” [2]. In the mutant, phospho-AKT is less responsive to changes in IGF1R, PI3K and EGFR. The network model was able to accurate predict the changes in phosphorylation of the network nodes in response to drug-combinations not used in training the model (Pearson-correlation ~ 0.84). Next, we trained a model relating changes in phospho-AKT and phospho-ERK to changes in cell viability. The good model performance (Cross-validated Pearson-correlation ~ 0.7) shows that cell viability can be predicted from changes in signaling output. Importantly, when we combined these two models (trained on single drug data only) to predict the cell-viability of drug-combinations, there was a good Pearson-correlation of 0.55 between simulated and measured viability.

These results show the feasibility of this approach to prioritize promising multi-drug combinations. We are currently combing the single and two-drug combination data to predict the effect of multi-drug combinations and use this to find combinations that are maximally selectivity for the PI3K mutant.

[1] F. Di Nicolantonio et al., PNAS 105, 20864–20869 (2008).

[2] E. Bosdriesz et al., Bioinformatics, (In press, Biorxiv dio: 10.1101/243709)

B3 A BOOLEAN NETWORK MODEL TO PREDICT THE BEHAVIOUR OF COFILIN-1 IN PANCREATIC CANCER

SILKE KÜHLWEIN[1], JULIAN SCHWAB[1], MALTE BUCHHOLZ[2] AND HANS A. KESTLER[1]

[1]Ulm University, Institute of Medical Systems Biology, Ulm, Germany
[2]University of Marburg, Department of Gastroenterology, Endocrinology and Metabolism, Marburg, Germany

Pancreatic cancer is one of the most lethal cancer in developed countries with a five year survival rate of less than 5% [1]. The actin severing protein cofilin-1 (CFL1) is overexpressed in pancreatic cancer which is associated with high invasiveness and poor prognosis [2]. In contrast, CFL1 knockdown experiments lead to cell-cycle arrest but no induction of apoptosis. We created a Boolean network model to predict reasons for the dysregulation of CFL1 and to evaluate its influence on the disease and its worse outcome. Boolean network models are simple mathematical models which can be used to study dynamic behavior. The simplicity of these models arises through the assumptions that genes are considered as expressed or not expressed. For the setup of Boolean network models, we use qualitative knowledge about regulatory interactions taken from literature statements. Simulations reveal sequences of states, called attractors, determining the long-term behavior of the system. All initial states leading to the same attractor are part of its basin of attraction. In a biological context, these attractors can be associated to phenotypes [3]. In our simulation we obtain two attractors representing cancer (active CFL1) and apoptosis (inactive CFL1) while knockout of CFL1 yields attractors representing cell-cycle arrest without apoptosis. To validate our model, we performed in vitro laboratory experiments. Here, first predictions based on our established Boolean network model could be confirmed.

[1] Blum et al. Metabolism addiction in pancreatic cancer. Cell Death and Disease 5, e1065 (2014)

[2] Satoh et al. Immune-complex level of cofilin-1 in sera is associated with cancer progression and poor prognosis in pancreatic cancer. Cancer Science. (2017)

[3] Kauffman. The origins of order. Self-organization and selection of evolution. Journal of Evolutionary Biology 7, 4 (1994), 518-519.

B7 CANPATHPRO – DEVELOPTMENT OF A PLATFORM FOR PREDICTIVE CANCER PATHWAY MODELING USING GENETICALLY ENGINEERED MOUSE MODELS

CHRISTOPH WIERLING[1], DANIEL WEINDL[2], YANN HERAULT[3], JOS JONKERS[4], ASPASIA PLOUBIDOU[5], LUCIEN FRAPPART[5], JAN HASENAUER[2], JULIO BANGA[6], OLIVER RINNER[7], GLENN TERJE LINES[8], DAVID KOUBI[9] AND BODO LANGE[1]

[1]Alacris Theranostics GmbH, Berlin, Germany
[2]Institute of Computational Biology, Helmholtz Zentrum München, Germany
[3]PHENOMIN Mouse Clinic Institute, France
[4]Netherlands Cancer Institute, Netherlands
[5]Leibniz Institute on Aging, Germany
[6]Spanish National Research Council, Spain
[7]Biognysys AG, Switzerland
[8]Simula Research Laboratory AS, Norway
[9]Finovatis SAS, France

Omics technologies are generating complex molecular datasets that are exponentially increasing the cancer knowledge base. However, the challenge remains to optimally exploit this wealth of data for basic research, better treatment and stratification of patients and more efficient targeted drug development.

The Horizon 2020 project CanPathPro (www.canpathpro.eu), addresses this challenge and is building and validating a combined experimental and systems biology platform, to be utilized in testing cancer signaling hypotheses. It will combine, in a single platform, omics and quantitative immunohistopathological data of cancer mouse models with analytical, modelling, predictive and visualization computational tools. The platform will perform data integration and predictive modelling

(i.e. in silico predictions based on computational and mathematical modelling using large-scale datasets) of the relevant signaling networks, leading to an output of testable hypotheses. Components comprise highly defined mouse and organotypic experimental systems, next generation sequencing, SWATH-based proteomics and a systems biology computational model for data integration, visualization and predictive modelling.

The generated platform will enable in silico identification of cancer signaling networks critical for tumor development and predictions regarding the activation status of individual pathways, following integration of user (or public) datasets in the pathway models. The innovative approach taken by CanPathPro is set to have broad and significant impact on diverse areas, from cancer research and personalized medicine to drug discovery and development, and ultimately improving outcomes for cancer patients.

B8 COMPUTATIONAL MODELING OF HEMATOPOIESIS IN MYELODYSPLASTIC SYNDROMES AND AGE-MATCHED HEALTHY CONTROLS

LISA BAST[1], MICHELE KYNCL[2], LYNETTE HENKEL[3], FABIAN J. THEIS[1], ROBERT OOSTENDORP[2], CARSTEN MARR[1] AND KATHARINA GÖTZE[2]

[1]Institute of Computational Biology, Helmholtz Zentrum München, Germany
[2]Department of Medicine III, Technical University of Munich, Klinikum rechts der Isar, Germany
[3]2Department of Microbiology, Technical University of Munich, Klinikum rechts der Isar, Germany

Myelodysplastic syndromes (MDS) are a group of diverse bone marrow disorders in which abnormal growth and disrupted differentiation of hematopoietic stem and progenitor cells (HSPC) prevents the production of healthy blood cells, leading to cytopenia. Recently, HSPC in MDS were shown to carry acquired somatic mutations, most frequently in epigenetic regulators or splicing machinery factors likely responsible for the dysplastic features observed in this disease. Although clonal hierarchies have been inferred on the basis of mutational data, these only constitute a snapshot of hematopoiesis in the patient at a given time and may not accurately reflect the disease kinetics.

To understand how clonal dominance is achieved, we seek to identify the differences in proliferation kinetics and lineage fate decisions of MDS HSPCs in comparison to healthy hematopoiesis. HSPCs are isolated from the bone marrow of patients or age-matched healthy controls, respectively, and cultured in vitro for defined time periods in serum-free medium supplemented with specific growth factors. Proliferation, cell death and cell lineage are determined by multiparameter FACS analyses. Additional integration of genomic sequencing analyses of bulk bone marrow and HSPC differentiated in vitro will enable the assessment of the contribution of the malignant clones to the population of proliferating cells. The inferred cell counts are used to fit a mathematical model that describes the dynamics of lineage specification with ordinary differential equations (ODEs). Multistart optimization leads to parameter estimates, which uncover potential differences of dysfunction in MDS in comparison to healthy hematopoiesis. Integrating the estimated rates into a comprehensive mathematical model of human hematopoiesis will eventually help to establish new insights into the progression of MDS towards AML.

B9 CELL DENSITY INFLUENCES CLUSTERING AND MIGRATORY BEHAVIOUR OF TRIPLE-NEGATIVE BREAST CANCER CELLS

GERHARD BURGER

Division of Drug Discovery & Safety - Leiden Academic Centre for Drug Research, Leiden University, Netherlands

The ability of cancer cells to invade neighbouring tissue from primary tumours is an important determinant of metastatic behaviour. Quantification of cell migration characteristics such as migration speed and persistence helps to understand the requirements for such invasiveness. One factor that may play a role here is how local tumour cell density shapes the migration characteristics, which we here investigate with a

combined experimental and modelling approach. As a first step, we analysed time-lapse imaging data on three aggressive Triple-Negative Breast Cancer (TNBC) cell lines (MDA-MB-231, Hs578t, and HCC38) during 2D migration assays at different cell densities. HCC38 cells exhibited a counter-intuitive increase in speed and persistence with increasing density, whereas the other cell lines did not exhibit such an increase. Moreover, HCC38 cells could be distinguished from the other cell lines because they exhibited cluster formation, and the clusters were more stable at low than at high cell densities. This suggests that the cluster formation is related to the changes in migration characteristics.

In order to integrate these experimental data and obtain a mechanistic understanding of the density-dependent cell migration characteristics and cluster formation, as a second step we developed realistic spatial simulations using the Cellular Potts Model (CPM). Because all three cell lines exhibit high pseudopod activity, we included a description of pseudopod dynamics in our model. Analysis of the model suggests that pseudopod-driven motion is the source of directional persistence in these cells. We are currently investigating whether inclusion of different adhesion properties in the *in silico* cells can explain the experimentally observed density-dependent migration characteristics and presence or absence of cluster formation. Furthermore, experimental work is ongoing to investigate whether inhibition of cellular adhesion in HCC38 cells will prevent cluster formation and at the same time alter the density dependence of migratory behaviour. Our project shows how combined computational and experimental approaches can accelerate scientific progress in a synergistic fashion, hopefully leading to an increased discovery rate of novel treatment options.

C MODEL AND DATA STANDARDS

C2 DEVELOPMENT AND EXTENSION OF MODEL OF INTRINSIC APOPTOSIS

STEPHANIE MC KENNA, DIRK FEY, LUCIA GARCIA AND DAVID MATALLANAS

University College Dublin, Ireland

Apoptosis is a major form of programmed cell death which is essential in development and homeostasis. A Systems Biology approach has been employed in an effort to understand the tight signaling decisions involved in apoptosis. Caspases are the key drivers of intrinsic and extrinsic apoptotic cell death. Tight regulation of these cleaving proteins by pro- and anti- apoptotic proteins is important for normal physiological function. The aim of this work was to investigate how the bistable properties of apoptosis change with differential cell line or patient expression profiles. Bistability is a feature described in many cellular processes and has previously been described in apoptosis, with the occurrence of bifurcation into two possible states- cell death or cell survival (Bagci et al. 2006). Through integration of previously published information with experimental results, an ordinary differential equation model of intrinsic apoptosis was produced. Previous models published by Eissing et al. (Eissing et al. 2004), Lindner et al. (Lindner et al. 2013), and Legewie et al. (Legewie, Blüthgen, and Herzel 2006) were explored as possible base models. The model by Legewie, et al. was extended using parameters defined by literature and high throughput experimental results. The extended model was simulated and interrogated by altered pro- and anti- apoptotic protein concentrations to predict the impact on bistability of the system. For a group of 478 Melanoma patients, patient specific simulations were carried out and linked to survival outcomes. Pro-apoptotic proteins Bax and Diablo emerged as interesting signaling nodes within the intrinsic apoptotic pathway. Future work with the model will incorporate mass spectrometry

proteomic data from a range of resistant and sensitive melanoma cell lines in an effort to unveil the important signaling nodes involved in the resistance mechanism of the cell lines.

D MODEL REUSE, EXTENSION AND INTEGRATION

D2 DMD SIGMA POINT METHOD AND COMPARISON WITH UNSCENTED KALMAN FILTERS

DANTONG WANG AND JAN HASENAUER

Institute of Computational Biology, Helmholtz Zentrum München, Germany

Due to cell-to-cell variability, parameters in biological models are usually not constants, but distributions. Therefore, methods to propagate these parameters through the model are crucial. Monte Carlo sampling is an intuitive solution, but very time consuming. Sigma point method (Wan and Merwe, 2000) can capture the first two moments of the distribution with 2N+1 points, where N represents number of dimensions. But the accuracy becomes lower when model is nonlinear. In this paper, we implemented a DMD (Dirac Mixture Distribution) sigma point method to approximate Gaussian distributions, which computes the locations of sigma points by minimizing the distance between DMD and Gaussian distribution (Gilitschenski and Hanebeck, 2013). Several models including ODE models are implemented to assess the accuracy of this new method. We first compare it with Monte Carlo and quasi Monte Carlo method and show that it outperforms both these two methods. Furthermore, we also compare the DMD sigma point method with other Unscented Kalman Filters (including sigma point method) with different numbers of sigma points (Menegaz et. al., 2015). Results show that the DMD sigma point method is more accurate than the other methods especially in nonlinear models as well as in ODE models.

E PARAMETER OPTIMIZATION

E1 QUANTIFYING THE CURSE OF DIMENSIONALITY

LUKAS REFISCH[1], JENS TIMMER[2] AND CLEMENS KREUTZ[1]

[1]Center for Biological Systems Analysis, University of Freiburg,Germany
[2]Institute of Physics University of Freiburg,Germany

The dynamics of complex biochemical reactions as they occur in living cells can be modeled by ordinary differential equations (ODE). One major task is model calibration, i.e. to estimate parameters like rate constants and initial concentrations based on experimental data. Optimization-based estimation like maximum likelihood is often challenging due to the existence of local minima, the highly nonlinear model responses and the limited precision of numerical ODE solutions.

It is well known that both the number of estimated parameters, and the bounds of investigated parameter values put restrictions on the efficiency of model calibration. Due to the inflation of the size of parameter search space, computation time increases and reliability of convergence to the global optimum decreases. In addition, it is yet unclear whether kinetic rates, initial concentrations and observation parameters have similar impact on

the performance. To investigate performance differences dependent on these effects systematically, we utilize a set of established benchmark models. Results are quantified using multivariate statistical analyses.

E2 CUSTOMIZED STEADY-STATE CONSTRAINTS FOR PARAMETER ESTIMATION IN NON-LINEAR ORDINARY DIFFERENTIAL EQUATION MODELS

MARCUS ROSENBLATT

Institute of Physics, Univesity of Freiburg, Germany

Signaling pathways and chemical reaction networks in systems biology are frequently modeled by ordinary differential equations. Model parameters are often unknown and have to be estimated from experimental data, e.g. by Maximum-likelihood estimation. To reduce the dimensionality of the parameter space, steady-state information is incorporated in the parameter estimation process where possible.

For non-linear models, analytical steady-state calculation typically leads to higher-order polynomial equations for which no closed-form solutions can be obtained. This can be circumvented by solving the steady-state equations for kinetic parameters which likewise avoids multiple solutions that are problematic for optimization. When solved for kinetic parameters, however, steady-state constraints tend to become negative for particular model specifications, thus, inducing optimization aborts due to divergent model trajectories.

Here, we present an algorithm based on graph theory that derives non-negative, analytical steady-state expressions by stepwise removal of cyclic dependencies between dynamical variables. Thereby, multiple steady-state solutions are avoided by construction. Our method is applicable to most common classes of biochemical reaction networks comprising mass-action, Hill-type kinetics, inhibition and catalyzation. We show that our approach is especially well-tailored to guarantee a high success rate for optimization.

F IDENTIFIABILITY AND UNCERTAINTY ANALYSIS

F1 SYNCHRONIZATION EFFECTS ON MODEL PARAMETRIZATION FOR THE YEAST CELL CYCLE

JULIA KATHARINA SCHLICHTING, GABRIELE SCHREIBER, LISA MALLIS AND EDDA KLIPP

Humboldt University of Berlin, Berlin, Germany

Saccharomyces cerevisiae is a famous model organism to study the mitotic cell cycle in eukaryotic cells. Cln2 and Clb5 (cyclins) as well as Sic1 (CKI) are key players in the regulation of the G1-S transition, called START. We measured the absolute number of mRNA molecules of unsynchronized single cells for SIC1, CLN2 and CLB5 by smFISH. Each cell was assigned to a specific cell cycle phase by using morphological markers, so that we get mRNA distributions per cell cycle phase. We quantified the relative number of protein molecules in a synchronized cell population for Sic1, Cln2 and Clb5 by Western blotting. The number of proteins is given as time course over the cell cycle.

In this study, we analyze differences between synchronization methods, as well as the difference between synchronized and unsynchronized cell populations. Chemical synchronizations are stronger than physical ones. We used α-factor as the most common chemical synchronization. Cells start with G1 after release. Further, we applied hydroxyurea and nocodazol, which release cells in S and G2, respectively. Elutriation is used for physical

synchronization. Selected new born cells start with G1 after release. The goal is to gain insight, to what extent we can use synchronized cell populations to draw conclusions for unsynchronized cells. We combine both data types to parameterize a stochastic model focusing on the key regulators for the G1-S transcripts by using a maximum likelihood approach. Our 2-step-optimization method differentiates between mRNA and protein levels. In our algorithms, we used a Poisson error model for mRNAs and a Gaussian error model for proteins. We used deterministic optimizers and multistart conditions to identify potential global optima. By means of calculating profile likelihoods we analyzed parameter identifiabilities. Including regularization, we figured out where our data don't cover the model.

F3 OPTIMAL PATH BETWEEN PARAMETER ESTIMATES IN NONLINEAR ODE SYSTEMS

CHRISTIAN TÖNSING, JENS TIMMER AND CLEMENS KREUTZ

University of Freiburg, Freiburg, Germany

Frequently, ordinary differential equation (ODE) models are used to mathematically represent the dynamic behavior of cellular components, e.g. for signal transduction or gene regulatory networks. For parameter estimation in ODE models, the discrepancy between data and model output needs to be minimized by finding a parameter vector in the high-dimensional search space which is optimal in terms of a minimal value of an objective function. The usage of local deterministic optimizers from multiple initial guesses within the parameter search space typically reveals the existence of several local optima.

For statistically valid conclusions, it is of general interest, whether the identification of local optima based on the objective function's value of the fits is reasonable and if the appearance of local optima is only a result of a not completely converged local optimizer and may be solved by fine-tuning of the numeric algorithms or if this is a consequence of the local non-convexity of the objective function and hence an intrinsic property of the model and data. In order to clarify this question in application settings, we present a method for finding an optimal path between local optima on the objective function in the parameter space. By analyzing the path's profile, i.e., the value of the objective function along the path and its dependency on the parameters, it can be investigated if two optimization results are de facto separated or connected.

F4 PYABC: SCALABLE LIKELIHOOD-FREE INFERENCE

YANNIK SCHÄLTE[1], EMMANUEL KLINGER[2], ELBA RAIMUNDEZ[1] AND JAN HASENAUER[1]

[1]Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany
[2]Max Planck Institute for Brain Research Frankfurth, Germany

Parameter estimation for mathematical models of biological systems is an important part of computational biology. However, as models get more complex and stochastic, it can become infeasible to compute the likelihood function quantifying the support data lend to parameters. Examples of such models include stochastic differential equations, Markov processes, and agent-based models. Of particular importance is the emerging field of multi-scale models, which give a more holistic understanding of biological systems, but are usually computationally very expensive. For these models, it is often still possible to simulate data. Approximate Bayesian Computation (ABC) methods have been developed to perform reliable inference in this situation. In a nutshell, the likelihood evaluation is circumvented by assessing a distance measure between simulated and observed data. Accepted parameters then realize a sample from an approximation to the posterior distribution.

We developed pyABC, a distributed ABC framework which implements various state-of-the-art and unique features: To scale likelihood-free inference to computationally demanding models, we developed a runtime-minimizing dynamic strategy for distributed environments, and a scheme for the automatic adaptation of population sizes. Also, we implemented a scheme to automatically weight different data types, further adaptive acceptance

threshold schedules, transition kernels, distance functions, early rejection, and a novel assessment of measurement noise. The toolbox is easily accessible for non-expert users and offers multiple extension possibilities to advanced users. It has already been applied in diverse projects, e.g. to a multi-scale hybrid model of tumor spheroid growth. The toolbox is available at https://github.com/icb-dcm/pyabc.

This contribution will provide a comprehensive overview over the capabilities of pyABC and outline how many research projects can benefit from it.

F5 DATA-ENTRAINED MODELING OF CHEMICAL-INDUCED MITOCHONDRIAL TOXICITY: FROM SIGNALING TO MORPHOLOGY

HUAN YANG AND JOOST BELTMAN

Leiden University, Netherlands

Cellular stress pathways protect cells from potential stressors, e.g. chemical compounds. When these pathways become malfunctioning due to drug treatments, adverse outcomes like drug-induced liver injury can occur. The molecular mechanisms are still poorly understood and a paradigm shift towards a more quantitative understanding has therefore been advocated. One important organelle is mitochondrion, which plays an important role in oxidative phosphorylation to produce the ATP. Here we aim to understand how mitochondrial signaling network works to adapt against to compound exposure. First, to characterize oxidative phosphorylation, we adapted a differential-equation-based model consisting of essential components. We further optimized parameter values to fit in-house data for 24-hour dynamics of mitochondrial membrane potential from HepG2 cells exposed to various compounds. Second, our in-house microscopy-based data suggest dynamic changes in mitochondrial morphology, which involve fusion and fission processes. Towards further quantification, we applied the random forest algorithm with local features on a pixel level to segment the mitochondria from the gray-level images. Then we utilized mitochondrial morphology features (e.g. form factor and area size) to classify the mitochondria as either fragmented or fused types. Currently, we are adapting a published dynamic model to mimic amounts of these classified objects. Furthermore, our recent western blot data (OPA1 protein) suggests chemical-induced perturbation to the fusion process for morphological changes. Also by performing knockdown to regulators involved in fission and fusion, our analysis reported changes in the amounts of fragmented and fused objects. We are integrating all submodels above into one holistic model. Hopefully, this computational work could provide a more mechanistic understanding of how mitochondrial signaling network and morphology adapt against to compound exposure.

F8 MAXIMUM LIKELIHOOD ESTIMATION FOR DYNAMICAL SYSTEMS FROM DATA WITH CORRELATED ERRORS

RAPHAEL ENGESSER AND JENS TIMMER

University of Freiburg, Germany

Parameter estimation in nonlinear dynamical systems is a major challenge in mathematical modeling of biological systems. In recent years, several methods to this aim were developed, many of them based on maximizing the likelihood function [1]. These methods typically assume uncorrelated normally or log-normally distributed measurements. In order to model outlier corrupted data, further noise distributions with heavier tails were applied [2].

In this work, we extend these methods for experiments which incorporate correlations between the measured data points. First, we analyze the effects of such correlations between the errors on the existing maximum likelihood methods. To this end, we consider the realistic setting of positive correlations in a simulation study. Evaluating the coverage ratios shows that the parameter confidence intervals obtained by the profile likelihood [3] are highly underestimated when the data contains positive correlations. To obtain suitable parameter uncertainties, we present a novel method which introduces a parameterized correlation model, whose parameters are estimated simultaneously with the

remaining error parameters and dynamical parameters by applying a maximum likelihood approach. By this, one still can utilize the profile likelihood to obtain parameter confidence intervals. Analyzing the coverage ratio in toy models as well as in realistic models from literature, we show that the confidence intervals are more appropriate with the novel method and therefore results in a valid uncertainty analysis when estimating parameters form data with correlated residuals.

[1] Raue et al. PLoS ONE 8, 2013, e74335

[2] Maier et al. Bioinformatics 33, 2017, 718-725.

[3] Raue et al. Bioinformatics 25, 2009, 1923-1929

F10 HIERARCHICAL SAMPLING FOR THE EFFICIENT ASSESSMENT OF PARAMETER UNCERTAINTY OF ODE MODELS

ELBA RAIMUNDEZ ALVAREZ AND JAN HASENAUER

Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany

Mathematical models have become standard tools for understanding and unraveling the underlying mechanism of biological signaling pathways. In general, the parameters of these models are unknown a priori and they need to be inferred from experimental data using statistical methods. Many of the measurement techniques most commonly used, e.g. fluorescence or flow cytometry, only provide relative information about the absolute molecular state. In this context, introducing scaling parameters and corresponding noise parameters in the observables is necessary. These parameters substantially increase the dimensionality of the estimation problem as they also need to be estimated along with the kinetic rate constants. Sampling methods are widely used in systems biology to parameterize mathematical models, and to facilitate the assessment of parameter and prediction uncertainties. The evaluation of sampling methods is usually demanding, leaving these on the border of computational feasibility. To facilitate the often required rigorous statistical assessment of parameter probability distributions in systems biology applications, efficient sampling algorithms are required.

We propose a hierarchical sampling scheme for estimating the parameter uncertainties for ordinary differential equation (ODE) models from relative data. We marginalized out the scaling and noise parameters from the original problem, leading to a dimension reduction of the parameter space. Herewith, only kinetic rate constants have to be sampled. We evaluated accuracy, robustness, and computational efficiency. We found that our approach outperforms standard sampling approaches by decreasing the chain auto-correlation and requiring a lower computation time.

F11 COMPUTING BIOLOGICALLY RELEVANT CRN STRUCTURES USING TIME-SERIES DATA

ZOLTAN TUZA AND GUY-BART STAN

Imperial College London, London United Kingdom

High-throughput data acquisition in Systems Biology leads to an abundance of data that need to be processed and aggregated into biologically relevant dynamical models. The dynamics of signaling, metabolic and gene regulatory networks are usually captured by ODE models, where the right hand side consists of polynomials or rational functions. Automatically building such ODE models, based on this wealth of data, is of paramount importance to gain insight into an existing biological process.

This type of model building approach is gaining traction in the literature where a dictionary of possible right-hand-side functions---assembled using domain expert knowledge---is presented to an algorithm along with the time-series data. Then, the algorithm builds the most suitable (e.g. parsimonious) combination of dictionary functions as an ODE model.

In this work, we combine this model building technique with Chemical Reaction Network Theory (CRNT) to automatically generate an ODE model, as well as all possible reaction graphs. For the model building part, we use Sparse Bayesian Learning, which automatically builds a nonlinear ODE model.

It should be emphasized that compared to common `simulate-and-compare' parameter estimation approaches this algorithm does not require solving ODEs.

To compute the CRN structures from an ODE model, we exploit the fact that multiple network structures may exist for a given ODE; even if the parameter are known *a priori*. The existence of multiple network structures – without noise – for a given dynamic has been investigated extensively in the literature. We use these results to compute the network structures, from an automatically built model.

Using this workflow, the wealth of data can be translated into a biologically relevant ODE model as well as reaction graphs. The set of possible graphs give an overview of all possible network connections, and allows us to identify targets for further perturbation experiments.

F12 DRUG TARGET DETECTION IN METABOLIC NETWORKS AS AN OPTIMALITY PROBLEM

OLUFEMI BOLAJI AND EDDA KLIPP

Humboldt University of Berlin, Berlin, Germany

Optimality Principles have played a major role in biological systems, from describing mechanisms to being able to predict from first principles to the design of organisms.

In this work, we present a dynamic optimization strategy to determine drug targets of pathological dynamic metabolic networks. This methodology involves testing the influence of inhibitors, i.e. the control profiles, via different modes-of-action to the enzymes in the network, and driving the network to a desired healthy state through the maximization or minimization of one or multi-objectives set a priori.

The proposed solution of the optimization problem involves using a combination of ε - constraint and control vector parametrization (CVP) to obtain non-linear programming problem (NLP) and initial value problem (IVP), which are solved by enhanced scatter search (eSS) optimization metaheuristic implemented in the AMIGO2 toolbox [1].

Using a glycolysis dynamic-model of Trypanosoma brucei, we show a scan for vulnerable enzymes in the model that are probably good drug targets.

This project has received funding from the EU Horizon 2020 Research & Innovation Program under the Marie Skłodowska-Curie Grant Agreement #675585.

[1] Balsa-Canto E., et al., (2016), Bioinformatics, 32, 3357

F14 AN EFFICIENT PROCEDURE TO REPARAMETERISE STRUCTURALLY UNIDENTIFIABLE MODELS

DOMINIQUE JOUBERT, HANS STIGTER AND JAAP MOLENAAR

Wageningen University, Netherlands

An efficient method for reparametrizing structurally unidentifiable models is introduced. It significantly reduces computational demand by combining both numerical and symbolic identifiability calculations. This hybrid approach facilitates the parameterization of: 1) large unidentifiable ordinary differential equation models and 2) models where some of the initial values of states are unidentifiable, in which case state transformations are required. In our approach, a model is first assessed numerically, to establish which parameters might be unidentifiable and determine the nature of correlation between these redundant parameters. These numerical results are then used in symbolic calculations to compute viable parameterizations that will ensure a model's structural identifiability. This use of the preceding numerical results reduces the computational demand of these symbolic calculations. We illustrate our procedure and the parameterization process in detail using the well-known JAK/STAT model with 14 states and 23 parameters as example [1].

[1] A. Raue, J. Karlsson, M. P. Saccomani, M. Jirstrand, J. Timmer, Comparison of approaches for parameter identifiability analysis of biological systems, Bioinformatics 30 (10) (2014) 1440{1448.doi:doi.org/10.1093/bioinformatics/btu006.

G MODEL ANALYSIS

G1 MULTISCALE MODELING OF DRUG TRANSPORT IN SYSTEMS PHARMACOLOGY

JOSEPH LEEDALE[1], STEVEN WEBB[2] AND RACHEL BEARON[1]

[1]University of Liverpool, United Kingdom
[2]Liverpool John Moores University, United Kingdom

New drugs are tested for toxic side effects in the laboratory using isolated cells. These toxicity tests traditionally involve cells cultured in a flat, 2D environment. However, emerging experiments where cells are cultured in 3D have been shown to more closely resemble the functionality of cells within the body. While the increasing usage of 3D experiments represent more realistic biology, the underlying physical processes of what happens to the drug in these environments is not fully understood. Our research shows how mathematical models can be used to simulate the activity and transport of drugs in 3D, informing experimentalists on how best to use these systems to test for toxicity. A multiscale mathematical modelling framework to describe the temporal and spatial dynamics of drugs in multicellular environments will be presented. The model combines information relating to the diffusion, transport and metabolism of chemical species (drugs) in 3D environments. A simplified 3D microscale single-cell model was analyzed to study different transport mechanisms by varying boundary conditions on the cell membrane. A more complex multicellular model has been developed to study the effects of cellular arrangement and density on the transport and penetration of drugs to simulate the problem for in vitro microtissue environments. Following the preliminary theoretical work, integration of experimental data is incorporated to develop realistic geometries and parameterize the model for a range of pharmacologically realistic scenarios.

G2 20 BENCHMARK PROBLEMS FOR DYNAMIC MODELING

CAROLIN LOOS[1], HELGE HASS[2], ELBA RAIMUNDEZ ALVAREZ[1], JAN HASENAUER[1] AND CLEMENS KREUTZ[2]

[1]Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany
[2]University of Freiburg, Germany

Ordinary differential equation (ODE) models which describe the temporal evolution of the abundance of biochemical species are widely used to study dynamic cellular processes. These models generally have parameters, such as initial conditions or rate constants, which cannot directly be measured. Thus, methods to estimate these parameters from experimental data are continuously developed. However, a rigorous assessment of the methods in application settings is hindered by the deficiency of realistic data provided with biochemical signaling networks and lack of suitable benchmark models.

We compiled a collection of 20 benchmark models that can be used to evaluate and compare newly developed and existing methods for, e.g., parameter estimation and uncertainty analysis. We provide the models as ODEs and in the widely used machine-readable SBML format, together with experimental measurement data. The models cover a broad range of model properties, varying in problem size, complexity and numerical demands. Performing example analyses, in which we studied scaling properties of optimization algorithms and the influence of log-transformation of parameters on optimizers, we illustrate the possibilities enabled by our provided benchmark collection.

G3 REALISTIC DESIGN FOR DYNAMIC MODELS IN SYSTEMS BIOLOGY

JANINE EGERT[1], JENS TIMMER[2] AND CLEMENS KREUTZ[1]

[1]Centre for Biological Systems Analysis (ZBSA), University of Freiburg, Germany
[2]Center for Data Analysis and Modelling (FDM), University of Freiburg, Germany

Motivation: By analyzing complex non-linear systems in Systems Biology many methodological and computational questions raise up. Due to the limited amount of publicly available data for biological experiments, simulation studies represent a reasonable alternative for the methodological analysis. Here, an important task is to find a realistic reflection of biological applications.

Results: In this paper we present an implementation for setting up a realistic simulation study. On the basis of ten biological applications we analyze the distribution and measurement conditions of observables and define realistic observation time points. For any desired dynamic model, e.g. downloaded from an online database, the presented implementation analyzes the model dynamics and suggests a realistic experimental design comprising observables, measurement times and simulates data with typical measurement errors.

G4 ESTIMATING TIME DELAYS USING PROFILE LIKELIHOOD

ADRIAN LUKAS HAUBER, RAPHAEL ENGESSER, JOEP VANLIER AND JENS TIMMER

University of Freiburg, Germany

Time delays in biochemical reaction networks can arise from unobserved processes. Instead of using a delay differential equation, a widely used technique for modeling such delays is the linear chain trick, as it represents the actual reasons for the delay. Introducing intermediate states in the form of a linear chain delays and low-passes the output signal. However, the number of states in the linear chain seems arbitrary. As a consequence, a parameter non-identifiability outside of the linear chain occurs for too long chains, while the delay parameter governing the rate within the chain stays identifiable, which can be detected via a profile likelihood analysis.

In this work, we show how this parameter non-identifiability can be used as a criterion to decide whether the used linear chain is too long. An iterative algorithm can be applied to find the optimal chain length. In contrast to the analysis of goodness-of-fit values, in many cases this method allows determining whether the used chain is too long without having to compute fits for all feasible chain lengths.

The method has proved to be working for various toy models by reproducing the true chain length from simulated data. It has been applied to biological systems and yields reasonable results.

G5 INTEGRATIVE WORKFLOW FOR IDENTIFICATION OF DIAGNOSTIC AND THERAPEUTIC MARKERS IN TUMOR INVATION

FAIZ KHAN

University of Rostock, Germany

In living cells, molecules interact with other molecules in a network to realize cellular functionality. Mutations in molecular factors perturb the regulation of networks leading to dysregulation of cellular functionality, which associates with complex diseases such as cancer. Unraveling mechanisms underlying diseases has motivated the development of various systems biology approaches. Key challenges in systems biology approaches for mechanistic understanding of diseases are: (i) the large number of interacting components in molecular networks, and (ii) the nonlinear nature of spatio-temporal interactions constituting complex network structures including feedback/feedforward loops.

To address these challenges, I developed an integrative workflow (Khan el al. Nature Comm. 2017; doi:10.1038/s41467-017-00268-2) by combining techniques from bioinformatics and systems biology. The workflow combines network structure, omics and biomedical data, and dynamic modeling (logic-based) to understand the

mechanism underlying complex diseases. Using the proposed workflow, I analyzed large-scale molecular interaction map of E2F1, a transcription factor involved in tumor invasion. It identified core-regulatory networks by ranking network sub-structures using multi-objective optimization function for epithelial-mesenchymal transition (EMT) in bladder and breast cancers, which are amenable for dynamical modeling. Using logic-based modeling formalism, the in silico stimulus-response analysis of the core networks detect molecular signatures for each cancer type. Further, I performed in silico perturbation experiments to identify therapeutic targets. The predicted molecular signatures and therapeutic targets were validated experimentally and through patient data. The computational analysis of biochemical networks can improve our understanding of disease processes in a mechanistic way. Ultimately, this shall provide the ability to manipulate and optimize processes towards treatment.