INCOME

INtegrative COllaborative Modelling in Systems MEdicine

Integrative pathway modeling in systems biology and systems medicine

CONFERENCE & HACKATHON

NOVEMBER 25 - 29, 2019

BERLIN, GERMANY

—

INCOME is a collaboration project within systems medicine. INCOME aims to facilitate a networking community that brings models and datasets as well as their developers closer to each other.

Table of content

Table of content 1

Program 4

Monday, November 25 4

Tuesday, November 26 5

Wednesday, November 27 6

Thursday, November 28 7

Friday, November 29 7

Abstracts of keynotes 8

Monday, November 25 8

Rule-based modelling of protein regulation in the synapse 8

Mathematical models of infection and immune response 8

The role of quantitative systems pharmacology in translational medicine 8

Tuesday, November 26 8

Optimal control in systems medicine 8

Multi-scale modeling integrating cellular mechanistic and whole-body physiologically-based pharmacokinetic (PBPK) models 9

Computational modeling of liver function tests - Stratification and individualization using standardized models and data 9

Wednesday, November 27 9

Analytical approximations for spatial and non-spatial stochastic biochemical kinetics 9

Modelling signal transduction and gene regulation at the interface of single-cell and population data 10

Abstracts of selected talks 11

Monday, November 25 11

Model-based analysis of ligand- and drug-induced growth regulation in breast cancer 11

Predictive modeling, applied to genetically engineered mouse models of breast or lung cancer, provides insights into major oncogenic pathways 11

A modular, hierarchical model of melanoma cell signaling 11

Fetal lung functional module and COPD module are in close vicinity in the human interactome 12

Tuesday, November 26 12

From parameter and uncertainty estimation to optimal experimental design: Challenges in biological dynamical systems inference 12

Efficient parameter estimation methods for integration of qualitative data in quantitative models 12

Improving model understanding and reuse through standards and common knowledge representation 13

WikiPathways: Curation, visualization and analysis of biological pathway 13

Reusing provenance information captured in WebProv for automatic generation of experiment specifications 13

Wednesday, November 27 14

Sensitivity based input-response index to analyse and reduce large-scale signalling networks 14

An inactivation switch enables rhythms in a neurospora clock model 15

ODE-based modelling of binding kinetics from aptamer microarrays 15

Learning the topology of latent signaling networks from high dimensional transcriptional intervention effects 15

Software development for Bayesian parameter sampling of large intracellular models 16

Efficient sampling by marginalization of scaling parameters for mechanistic models with relative data 16

Abstracts of posters **17**

Modeling of signalling, metabolic and gene regulatory pathways 17

Modeling survival of long-lived plasma cells in the bone marrow 17

Modeling of diseases and therapies 17

Overexpression of CBL ubiquitin ligases causes paradoxical amplification of MAPK and PI3K signaling 17

Integrative analysis of peripheral N-acetylaspartate metabolism 17

The fetoplacental interface in the development of pre-eclampsia 18

The hallmarks of cancer metabolism, a genome-scale metabolic model approach 18

Phenotypic selection through cell death: stochastic modelling of O-6-methylguanine-DNA methyltransferase dynamics 18

Model and data standards 19

PK-DB: PharmacoKinetics DataBase for individualized and stratified computational modeling 19

Spatial cytokine gradients in T cell communication sensitively depend on cell-intrinsic properties 19

Dynamical modeling of complex diseases uncovers Critical Transitions in their phenotypes 19

Stochastic modeling of the dynamics of proliferating cell populations 19

Managing the INCOME pathway data on the FAIRDOMHub 20

Response-time modeling of T-helper cell differentiation 20

Model reuse, extension and integration 20

In silico mutagenesis of genes implicated in defects of neuronal migration 20

Model version merging 21

Parameter optimization, identifiability analysis and uncertainty analysis 21

Optimal paths between parameter estimates in nonlinear ODE systems using the nudged elastic band method 21

Noise up your ABC 21

Dirac mixture distributions for the approximation of mixed effects models 22

Integrating pyABC and Morpheus for the simulation and parameterization of computational models of multi-cellular processes 22

Model analysis 22

Modelling HFK20 methylation during embryogenesis in Xenopus laevis 22

Introducing an experimentally validated model for DNMT1 methylation dynamics comprising processive and distributive reaction mechanisms 23

Investigating heterogeneity in single cell data by applying statistical physics and random matrix approach to biological distribution 23

Construction of a kinetic model of the central carbon metabolism in E. coli describing the switch from aerobic to anaerobic growth 23

Detecting Oscillations in Neural Development 24

Distribution-free differential expression analysis for scRNA-seq data across patient groups 24

Automatic generation of priors for large-scale dynamic models 25

Venue and conference dinner 26

Venue 26

Conference dinner 27

Hackathon dinner 29

Program

Monday, November 25

12:00 *Registration opens*

13:30 - 13:50 Welcome by organizers

13:50 - 14:40 Keynote lecture by Melanie Stefan on **Rule-based modelling of protein regulation in the synapse**

14:40 - 15:30 Keynote Lecture by Lars Kaderali on **Mathematical Models of Infection and Immune Response**

Coffee break

16:15 - 17:05 Keynote Lecture by Birgit Schöberl on **The Role of Quantitative Systems Pharmacology in Translational Medicine**

17:05 - 17:25 Talk by **Svenja Kemmer** on **Model-based analysis of ligand- and drug-induced growth regulation in breast cancer**

17:25 - 17:45 Talk by **Bodo Lange** on **Predictive Modeling, Applied to Genetically Engineered Mouse Models of Breast or Lung Cancer, Provides Insights into Major Oncogenic Pathways**

17:45 - 18:05 Talk by **Charles Barker on A Modular, Hierarchical Model of Melanoma Cell Signaling**

18:05 - 18:25 Talk by **Annika Röhl** on **Fetal Lung Functional Module and COPD Module are in Close Vicinity in the Human Interactome**

Dinner and get-together

Bar is reserved for the evening

Tuesday, November 26

09:00 - 09:50 Keynote Lecture by Julio Banga on **Optimal control in systems medicine**

09:50 - 10:10 Talk by **Matthias Chung** on **From Parameter and Uncertainty Estimation to Optimal Experimental Design: Challenges in Biological Dynamical Systems Inference**

10:10 - 10:30 Talk by **Leonard Schmiester** on **Efficient Parameter Estimation Methods for Integration of Qualitative Data in Quantitative Models**

Coffee break & Poster Session 1

11:40 - 12:30 Keynote Lecture by Ursula Kummer on **Multi-scale modeling integrating cellular mechanistic and whole-body physiologically-based pharmacokinetic (PBPK) models**

Lunch

13:40 - 14:30 Keynote Lecture by Matthias König on **Computational modeling of liver function tests - Stratification and individualization using standardized models and data**

14:30 - 14:50 Talk by **Dagmar Waltemath** on **Improving Model Understanding and Reuse Through Standards and Common Knowledge Representation**

14:50 - 15:10 Talk by **Martina Kutmon** on **WikiPathways: Curation, Visualization and Analysis of Biological Pathway**

15:10 - 15:30 Talk by **Kai Budde** on **Reusing provenance information captured in WebProv for automatic generation of experiment specifications**

Coffee break & Poster Session 2

17:00 - 18:00 Plenary discussion

Conference Dinner at Berlin Fernsehturm

Wednesday, November 27

09:00 - 09:50 Keynote Lecture by Ramon Grima on **Analytical approximations for spatial and non-spatial stochastic biochemical kinetics**

09:50 - 10:10 Talk by **Wilhelm Huisinga** on **Sensitivity Based Input-Response Index to Analyse and Reduce Large-Scale Signalling Networks**

10:11 - 10:30 Talk by **Abhishek Upadhyay** on **An Inactivation Switch Enables Rhythms in a Neurospora Clock Model**

Coffee break

11:00 - 11:20 Talk by **Lukas Refisch** on **ODE-based Modelling of Binding Kinetics from Aptamer Microarrays**

11:20 - 11:40 Talk by **Zahra Sadat Hajseyed Nasrollah** on **Learning the Topology of Latent Signaling Networks from High Dimensional Transcriptional Intervention Effects**

11:40 - 12:00 Talk by **Andrei Kramer** on **Software Development for Bayesian Parameter Sampling of Large Intracellular Models**

12:00 - 12:20 Talk by **Elba Raimundez** on **Efficient sampling by marginalization of scaling parameters for mechanistic models with relative data**

Lunch

13:40 - 14:30 Keynote Lecture by Jana Wolf on **Modelling signal transduction and gene regulation at the interface of single-cell and population data**

14:30 - 14:40 Closing remarks

16:00 - 16:30 Presentation of hackathon topics

16:30 - 18:00 Formation of topic groups and tutorials

1. Tutorial on multi-cellular modelling
   - Jörn Starruß on Modelling and Simulation of Multi-Cellular Processes using Morpheus
   - Emad Alamoudi & Yannik Schälte on Parameter estimation using pyABC
2. Tutorial on model merging (by Tom Gebhardt)

Thursday, November 28

09:00 - 12:30 Hacking (with a coffee break at around 10:00)

12:30 - 13:30 Lunch

13:30 - 18:00 Hacking (with a coffee break at around 15:30)

18:00 - 20:00 Dinner

20:00 - Open end / Late night hacking

Friday, November 29

09:00 - 12:30 Hacking (with a coffee break at around 10:00)

12:30 - 13:30 Lunch

13:30 - 14:30 Hacking

14:30 - 15:00 Summary and closing remarks

Abstracts of keynotes

A. Monday, November 25

Rule-based modelling of protein regulation in the synapse

Melanie Stefan
University of Edinburgh, United Kingdom

Synapses are not static, but change in strength dependent on their activity. These processes form the cellular basis of learning and memory. Within the postsynaptic neuron, they rely on Calcium influx and the activation of Calcium signalling pathways that can induce either a strengthening or a weakening of the synapse. The balance between strengthening and weakening is finely tuned. Several of the proteins involved in synaptic Calcium signalling have complex regulatory mechanisms that involve several levels of regulation (e.g. conformational change, modulator binding, post-translational modifications, localisation).

These different regulation mechanisms co-exist and combine to fine-tune protein activity. Modelling this in detail raises the problem of combinatorial explosion, which can be addressed using rule-based modelling. Here, I present the general problem, discuss different methods for rule-based modelling, and show some results from rule-based models of synaptic protein regulation.

Mathematical models of infection and immune response

Lars Kaderali
University of Greifswald, Germany

Upon viral infection, antiviral innate immunity pathways induce an antiviral state of host cells to infer with viral replication and spread. The RIG-I like receptor (RLR) family plays a crucial role by sensing pathogen-associated molecular patterns (PAMPs) within the cytoplasm and triggering a signaling cascade leading to the expression of cytokines, most prominently type I and III interferons (IFNs). Upon secretion, IFNs trigger the expression of a large array of IFN stimulated genes (ISGs), which in concert establish a strongly antiviral state of the cell. In this study, we experimentally characterize the kinetic properties of RIG-I activation and the downstream signaling system and set up a mathematical model capable of accurately describing the dynamics from introduction of dsRNA to expression of ISGs upon IFN signaling. A previous study reported intriguing stochasticity in the activation of IRF7 upon virus infection of murine cells. Surprisingly, we found RIG-I signaling to be highly deterministic, suggesting that the previously observed stochasticity was largely due to staggered uptake of the stimulatory RNA during infection. Our time-resolved data was, hence, optimally suited to set up and calibrate a dynamic mathematical model of the core RIG-I pathway. This model allows the identification of sensitive steps in the regulation of the immune response by directly linking them to the expression of interferon-stimulated genes (ISGs). We validated this comprehensive pathway model by data from wildtype cells versus cells lacking the type I and III IFN system (IFNAR/IFNLR double knockout). Using additional activation dynamics of RIG-I pathway components in presence of viral antagonists, we identify the consequences of viral evasion strategies on the immune response and their most likely site-of-action.

The role of quantitative systems pharmacology in translational medicine

Birgit Schöberl
Novartis Institutes for BioMedical Research, Switzerland

The translation of preclinical science to anticipate clinical responses is still an unresolved challenge. As a field we made great progress in predicting human pharmacokinetics from preclinical species. However, the translation of drug response from *in vitro* and *in vivo* data collected in surrogate systems to the human disease remains challenging. In my presentation I will address how Quantitative Systems Pharmacology allows us to integrate seemingly disparate data and thus supports drug development via forward- and reverse-translation from preclinical to clinical data.

B. Tuesday, November 26

Optimal control in systems medicine

Julio Banga
IIM-CSIC, Spain

In this talk, I will review the links and opportunities between the concepts of optimality and regulation in biological systems. These concepts have been successfully applied by a number of researchers to explain dynamics at the cellular level, or the structures of physiological systems. After an overview of the main conceptual developments, I

will address how modern optimal control theory can be used to rationalize regulation mechanisms found in living systems, and how this can be exploited in systems medicine. In addition, I will also discuss some opportunities that optimal control theory offers to modify disease states.

Multi-scale modeling integrating cellular mechanistic and whole-body physiologically-based pharmacokinetic (PBPK) models

Ursula Kummer
COS Heidelberg, Germany

Many cellular kinetic models of metabolic pathways and signal transduction networks have been established in the past. These can be used to study the impact of chemicals, e.g. drugs on the cellular behaviour, if the mechanism of this interaction is known. At the same time, whole-body pharmacokinetic models have been used for decades that - often in a lot of detail - describe the distribution dynamics and concentration profiles of drugs in the body after administration. However, an integration of these two scales has been lacking until recently. In this talk, a methodology to combine whole-body physiologically-based pharmacokinetic (PBPK) models with mechanistic intracellular models of signal transduction is presented. As a use case IFN-α administration is studied. A multi-scale model infers the time-resolved concentration of IFN-α arriving at the liver after intravenous injection while simultaneously estimating the effect of the respective dose on the intracellular signalling behaviour in the liver.

Computational modeling of liver function tests - Stratification and individualization using standardized models and data

Matthias König
Humboldt Universität zu Berlin, Germany

Key requirements for the translation of computational models are: i) reproducibility of results; ii) reusability and extensibility of models; iii) availability of data; and iv) strategies for stratification and individualization of models. Here we will present a modeling workflow focused on these key aspects applied to liver function tests.

Assessment of liver function is a key task in hepatology but accurate quantification of hepatic function has remained a clinical challenge. Dynamic liver function tests are a promising tool for the non-invasive evaluation of liver function in vivo. These clinical tests evaluate the function of the liver via the clearance of a given test substance, thereby providing information on the metabolic capacity of the liver. We modeled these tests via whole-body physiological models of absorption, distribution, metabolization and elimination using multi-scale SBML models (core and comp package).

One class of such tests are breath tests based on the conversion of 13C-labeled substrates by the liver to 13CO2 subsequently measured in the breath. A commonly applied substrate is 13C-methacetin, converted to paracetamol and 13CO2 via cytochrome P450 1A2 (CYP1A2), used orally in the methacetin breath test (MBT) and intravenously in the LiMAx test. An important clinical question is which factors can affect MBT and LiMAx results. The aim of our study was to answer this question using computational modeling to derive basic information for a better understanding of the methacetin breath test and factors influencing its results.

A second example, the liver function test based on caffeine is known for long, but its clinical usability is hampered by large interindividual variability and dose-dependency. By applying a physiological based pharmacokinetics model (PBPK) for the evaluation of the caffeine clearance test, we were able to assess in silico the hepatic conversion of caffeine to paraxanthine via cytochrome P450 CYP1A2. The model is able to reproduce results from a wide range of reported studies under varying caffeine doses and application routes and accounts for interindividual differences based on distributions of CYP1A2 and modification of CYP1A2 activity via lifestyle factors, e.g. smoking, and pharmacological interactions, e.g. oral contraceptives. Validation was performed with an independent clinical trial (EudraCT 2011-002291-16, ClinicalTrials.gov NCT01788254) demonstrating an improved prediction using individualized models accounting for smoking status and contraceptive use. Hereby, we could reduce the large variability in the test results providing the basis for better sensitivity and specificity in diagnosing subjects with liver problems.

C. Wednesday, November 27

Analytical approximations for spatial and non-spatial stochastic biochemical kinetics

Ramon Grima
University of Edinburgh, United Kingdom

Stochastic effects in biochemical systems including gene regulatory networks are commonly studied using the stochastic simulation algorithm. An alternative means of exploring stochastic dynamics is by means of approximations of the chemical master equation or the reaction-diffusion master equation. In the first part of this

talk, I will present some of our recent work on deriving approximate solutions to a chemical master equation model of gene expression in mammalian cells which includes cell division, gene duplication, dosage compensation, growth-dependent transcription, mRNA maturation and translation. In the second part of the talk, I will discuss the construction of new stochastic spatial models that can approximately take into account excluded-volume interactions. The number distributions obtained from such an approach can be considerably different than those given by the conventional reaction-diffusion master equation (and the chemical master equation) thus suggesting that macromolecular crowding plays a significant role in controlling the dynamics of noisy intracellular reactions

Modelling signal transduction and gene regulation at the interface of single-cell and population data

Jana Wolf
Max Delbrück Center for Molecular Medicine, Germany

Signalling pathways and gene regulation are critically involved in cellular decisions how to respond to external signals and are often perturbed in tumour development. While the processes have classically been studied on the level of cell populations the focus has shifted to single cell investigations in the last years. This allowed a very detailed view revealing not only a wide range of dynamic phenomena but also the heterogeneity between cells.

In this talk I will discuss gene expression and protein synthesis at the interface of cell population and single cell data and introduce a method for advanced differential gene expression analysis that takes the role of the cell cycle in proliferating cell populations into account. Moreover, I will demonstrate how single cell data can be used to derive a quantitative signalling models and to decipher the cross-talk between signalling pathways involved in genotoxic stress.

Abstracts of selected talks

A. Monday, November 25

Model-based analysis of ligand- and drug-induced growth regulation in breast cancer

Svenja Kemmer
Institute of Physics University of Freiburg, Germany

Targeted therapies have shown striking success in the treatment of cancer over the last years. However, their specific effects on the tumor seems to be varying and difficult to predict. Using an integrative modeling approach combining mechanistic and regression modeling we could gain insights into the cellular response mechanism of breast cancer cells to different ligand-drug combinations. The multi-pathway model capturing ErbB receptor signaling, MAP kinase and PI3 kinase signaling pathways was trained on time-resolved data of the luminal cell lines MCF7 and T47D across a set of four ligands and five drugs applied in all possible ligand-drug combinations. Exploring the model system we can predict the proliferation response of cells to drug co-treatments experimentally validated with additional measurements.

Interestingly, the regulation of signaling pathways turned out to be conserved between these cell lines of the same breast cancer subtype. Analyzing the regulation of the system in other breast cancer subtypes we identified several differentially controlled, cell type specific mechanisms in the basal breast cancer cell line MDA MB 231 using L1 regularization.

The analysis of our model revealed the importance of different receptor homo- and heterodimers integrating the effect of applied ligands and drugs. Dependent on the receptors present on a cell and the ligands applied, cells show transient or more sustained signaling dynamics leading to different proliferation readouts. Drugs are able to alter and reduce this response in dependence of the receptor composition of the target cell and the ligand applied.

Predictive modeling, applied to genetically engineered mouse models of breast or lung cancer, provides insights into major oncogenic pathways

Bodo Lange
Alacris, Germany

The H2020 project CanPathPro is building and validating a predictive modelling platform applied to cancer. To this end, we develop and refine bioinformatic and experimental tools, utilized in generation and evaluation of systems biology modelling predictions. Key project components include: biological systems representing 3 levels of biological complexity (genetically engineered mouse models –GEMMs– of breast or lung cancer, organoids & cell lines derived thereof); NGS and SWATH-based phospho/proteomics; 2 large-scale computational mechanistic models.

The highly-defined biological systems are used (i) to activate selected oncogenic stimuli that modulate pathway components in a systematic manner; (ii) to characterize the signaling changes occurring during cancer development - thus generating temporally resolved datasets for model training; and (iii) to validate, in vitro and in vivo, the modelling predictions. The mechanistic models, based on ordinary differential equations, enable prediction of phenotypes and drug response in mouse or human. Model parameters are defined using project-derived experimental data, either via parameter estimation strategies or via selection of parameter distributions by a Monte Carlo approach. For simulations, the models are initialized with transcriptome data, either from GEMM-derived cell lines grown under variable conditions, or from GEMM-derived neoplastic and tumour tissue, representing lesion progression. The models also integrate the relevant mutations. Results, obtained by iterative rounds of in silico predictions and in vitro validation, include identification: (i) of the activation status of oncogenic pathways; (ii) of the drug response of the GEMMs; (iii) of the signaling changes induced by the in vitro growth conditions, or by the mutational profile, or by the lesion stage of each GEMM.

A modular, hierarchical model of melanoma cell signaling

Charles Barker
EMBL-EBI, United Kingdom

One of the central goals of biology is to explore the complex relationship between genotype and phenotype. In the cytoplasm, complex cell signaling circuits interact to process external information and determine cell fate. When genetic aberration causes dysregulation, a cell can proliferate in an uncontrollable fashion, manifesting itself in the human body as cancer. Studying the complexity of

these systems, and the nuances of how they interact can best be done through modeling in silico. Here, a modular and hierarchical approach to modeling cell signaling is illustrated, with melanoma as a case study.

Fetal lung functional module and COPD module are in close vicinity in the human interactome

Annika Röhl
Harvard University, United States

Many diseases are associated with genetic variants that may lead to its pathological features. However, the same variant can lead to different diseases and often the interplay between different genes are critical for the manifestation of a certain phenotype. To study the interactions of those genes, Human interactomes present a natural way and we use such a network to compute disease modules, which are connected subnetworks that can be linked to a particular disease phenotype.

In this study, we focus on the development of chronic obstructive pulmonary disease (COPD), which is known to be a disease that manifests especially in adults and often, the root can be found in other respiratory diseases. Specifically, we analyze the impact a disturbed lung growth makes on COPD and try to find biological mechanisms and common pathways, since they may present the molecular transition how altered lung development may influence and lead to COPD. We focus on the methylation of genes which are involved, as methylation plays a crucial role in gene regulation and organ development. We use differentially methylated genes from lung tissues from fetuses whose mothers smoked during pregnancy and from adult COPD patients.

To identify disease modules in the interactome we developed a method which maximizes the COnnectivity sigNiFicance of dIsease geNEs (CONFINE). For each data set we computed a module. The two modules do not overlap but they are connected, thus genes from both sets are linked to each other in the interactome. These connecting genes are candidates for answering the question which mechanisms lead from perturbed lung development to COPD. We studied these connecting genes further and identified enriched pathways that lie at the intersection of the two phenotypes.

B. Tuesday, November 26

From parameter and uncertainty estimation to optimal experimental design: Challenges in biological dynamical systems inference

Matthias Chung
Virginia Tech, United States

Inference through data and mathematical modeling is particularly challenging for biological systems with noisy data, model uncertainties, and unknown mechanisms. Here, parameter and uncertainty estimation problems are typically ill-posed, meaning solutions do not exist, are not unique, or do not depend continuously on the data. In this talk we will discuss challenges of parameter estimation problems and present new computational methods for parameter estimation and uncertainty quantification for dynamical systems. Moreover, we discuss novel techniques for optimal experimental design. Our techniques are illustrated by various biological applications such as infection disease models and medical tomography.

Efficient parameter estimation methods for integration of qualitative data in quantitative models

Leonard Schmiester
Helmholtz Zentrum München, Germany

In systems biology, mechanistic models are a valuable tool to understand biological processes such as signaling pathways and to derive new insights into the underlying processes. These models usually comprise unknown parameters, which have to be inferred from experimental data. This is commonly done using quantitative measurements by minimizing the difference between model output and measured data.

However, often biological experiments rather provide qualitative information or have a limited and often unknown linear detection range, e.g. for Western blots or microscopy images. Additionally the measurement uncertainties can be too large, to rely on the quantitative values. These type of measurements cannot be included in state-of-the-art parameter estimation methods and are therefore often neglected. Only recently, some methods have been developed to also handle qualitative data [Pargett et al. (2014), PLoS Comput Biol 10(3): e1003498; Mitra et al., Nature Communications 9, 3901 (2018)].

In this study, we evaluate and improve existing approaches to integrate qualitative data for parameter estimation. We derive a reduced formulation of the approach introduced in

Pargett et al., leading to substantially reduced computation times. We implemented the methods into a common framework in the open-source Python parameter estimation toolbox pyPESTO. This enabled the thorough performance comparison of different optimization algorithms. All methods were tested on several application examples, showing the benefit of including qualitative measurements.

Overall, we improve the efficiency and provide a unified, reusable implementation for the integration of qualitative data into the parameter estimation process.

Improving model understanding and reuse through standards and common knowledge representation

Dagmar Waltemath
University Medicine Greifswald, Germany

Research in computational biology depends on the availability of executable and reusable simulation models, including pathways. The complexity and diversity that biological systems have reached today require high-quality models to be available from public resources such as BioModels Database [1] or, specifically for pathway data, WikiPathways [2], the Reactome [3] or the KEGG pathway databases. In order to be reusable, models should be curated, well-annotated and FAIR [4]. Standardisation of models is coordinated in the COMBINE network [5]. Software tools for visualisation, exploration, comparison, retrieval and reuse of models are being provided by the broader community.

We will give brief impulses and highlight helpful tools, communities, ontologies and best practises for pathway representation, exploration, retrieval, visualisation, and distribution. In particular, we will introduce standards for pathway representation (BioPAX, SBGN, SBML), provide links to best practises and demonstrate tools for pathway visualisation. We will briefly point out pathway databases and their exploration and retrieval capabilities.

Finally, we will explain how to contribute to the community, and how to populate FAIR pathway models through repositories and data-oriented journals.

[1] Li, Chen, et al. "BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models." BMC systems biology 4.1 (2010): 92.
[2] Slenter, Denise N., et al. "WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research." Nucleic acids research 46.D1 (2017): D661-D667.
[3] Fabregat, Antonio, et al. "The reactome pathway knowledgebase." Nucleic acids research 46.D1 (2017): D649-D655.
[4] Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." Scientific data 3 (2016).
[5] Hucka, Michael, et al. "Promoting coordinated development of community-based information standards for modeling in biology: the COMBINE initiative." Frontiers in bioengineering and biotechnology 3 (2015): 19.

WikiPathways: Curation, visualization and analysis of biological pathway

Martina Kutmon
Maastricht University, Netherlands

AbstractWikiPathways (www.wikipathways.org) is a community curated pathway database that enables researchers to capture rich, intuitive models of pathways. Latest developments, newest features and ongoing projects of WikiPathways and the associated tools PathVisio and the WikiPathways app for Cytoscape will be highlighted. The database and the associated tools are developed as open source projects with a lot of community engagement.

Tools: The standalone pathway editor, analysis and visualization tool, PathVisio provides easy-to-use drawing and annotation tools to capture identities, relationships, comments and literature references for each pathway element and interaction. The WikiPathways app for Cytoscape can be used to import biological pathways in Cytoscape for data visualization and network analysis.

Data: The WikiPathways database is improved by continuous data curation and updates through an expanding community: close to 700 individual contributors and more than 45,000 edits on nearly 2,800 pathways since its beginning in 2008. In August 2019, we have reached a total number of 2,785 pathways for 25 different species. Recently, we have decided to adopt the Creative Commons CC0 waiver for our content on WikiPathways. Our data is available for download from our website, through our REST webservice or in RDF format from our SPARQL endpoint. We are also in the process of importing our content into WikiData (www.wikidata.org).

Reusing provenance information captured in WebProv for automatic generation of experiment specifications

Kai Budde
University of Rostock, Germany

The provenance of a simulation model provides important context about its generation in a structured and computer-accessible form which facilitates interpreting and thus reusing the simulation model. It includes information about (i) simulation experiments that have been executed, (ii) data that has been used as input for calibration or validation of the simulation model, or (iii) other simulation models it has been based upon or cross-validated with [1]. In [2] we have presented the web-based provenance tool WebProv to access, store, and display the provenance information of simulation studies using the PROV ontology.

In combination with WebProv, we will show how our template-based simulation experiment generation approach [3] can facilitate the automatic generation and execution of simulation experiments across simulation studies, thus supporting the development of valid simulation models. As an example, we have captured provenance information of different models of canonical Wnt signaling and discuss which information is needed for automatically generating various types of experiments such as sensitivity analyses or cross-validation experiments.

[1] Ruscheinski, A., Gjorgevikj, D., Dombrowsky, M., Budde, K., and Uhrmacher, A. M. (2018). Towards a PROV Ontology for Simulation Models. In: International Provenance and Annotation Workshop, pages 192-195. Springer.
[2] Budde, K., Smith, J., Ruscheinski, A., and Uhrmacher, A. M. (2019). WebProv: A Web-based Tool to Access, Store, and Display Provenance Information of Simulation Models. In: 10th Computational Modeling in Biology Network (COMBINE) Meeting, 15-19 Jul 2019, Heidelberg, Germany.
[3] Ruscheinski, A., Budde, K., Warnke, T., Wilsdorf, P., Hiller, B. C., Dombrowsky, M., and Uhrmacher, A. M. (2018). Generating Simulation Experiments Based on Model Documentations and Templates. In: Winter Simulation Conference (WSC 2018), 09-12 Dec 2018, Gothenburg, Sweden. Proceedings, published by IEEE, pp. 715-726.

C. Wednesday, November 27

Sensitivity based input-response index to analyse and reduce large-scale signalling networks

Wilhelm Huisinga
University of Potsdam, Germany

In systems biology and pharmacology, large-scale kinetic models are used to study the dynamic response of a system to a specific stimulus or perturbation (input). The quantitative analysis of the input-response behaviour, however, is often challenging due to the size of the model and the complexity of interactions. An approach that allows to identify the key dynamic constituents is therefore highly desirable.

We present a model order reduction approach based on a novel sensitivity based input-response index. It is linked to the product of two local sensitivity coefficients: The first coefficient quantifies the impact of the input on a given state variable at a given time; the second coefficient quantifies how a perturbation of a given state variable at a given time impacts the output in the future. We rank state variables according to their input-response index and reduce the complexity of a model in three steps by elimination of state variables of negligible impact, approximation of fast molecular species by their quasi-steady state; and exploitation of conservation laws. A key feature of the reduced model is its mechanistic interpretability in terms of quantities of the original system.

In application to a large-scale kinetic model of the humoral blood coagulation network, we identified a reduced model of substantially lower complexity for the time course of fibrinogen recovery after a snake bite (reduction of 62 to 8 state variables). The reduced model gives detailed insight into the coordinated action of specific coagulation factors. In application to the prothrombin-time (PT) test, we provide insight why the common test is insensitive to genetic variants of factors VIII and IX, while a modified test does allow to detect these variants.

The sensitivity based input-response indices demonstrate for the first time how local sensitivity coefficients can be effectively leveraged in the context of model order reduction.

An inactivation switch enables rhythms in a neurospora clock model

Abhishek Upadhyay[1,2]
[1]Institute for Theoretical Biology (ITB), Charite, Germany
[2]University of Medicine and Humboldt University of Berlin, Germany

Autonomous endogenous time-keeping is ubiquitous across many living organisms, known as the circadian clock when it has a period of about 24 h. Interestingly, the fundamental design principle with a network of interconnected negative and positive feedback loops is conserved through evolution, although the molecular components differ. Filamentous fungus Neurospora crassa is a well-established chrono-genetics model organism to

investigate the underlying mechanisms. The core negative feedback loop of the clock of Neurospora is composed of the transcription activator White Collar Complex (WCC) (heterodimer of WC1 and WC2) and the inhibitory element called FFC complex, which is made of FRQ (Frequency protein), FRH (Frequency interacting RNA Helicase) and CK1a (Casein kinase 1a). While exploring their temporal dynamics, we investigate how limit cycle oscillations arise and how molecular switches support self-sustained rhythms. We develop a mathematical model of 10 variables with 26 parameters to understand the interactions and feedback among WC1 and FFC elements in nuclear and cytoplasmic compartments. We performed control and bifurcation analysis to show that our novel model produces robust oscillations with a wild-type period of 22.5 h. Our model reveals a switch between WC1-induced transcription and FFC-assisted inactivation of WC1. Using the new model, we also study the possible mechanisms of glucose compensation. A fairly simple model with just three nonlinearities helps to elucidate clock dynamics, revealing a mechanism of rhythms' production. The model can further be utilized to study entrainment and temperature compensation.

ODE-based modelling of binding kinetics from aptamer microarrays

Lukas Refisch
Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany

Aptamer arrays constitute an alternative experimental technique for multiplexed detection and quantification of proteins. In order to circumvent limitations of antibody-based microarrays, DNA oligos are used to capture target proteins. In order to discover high-performance aptamer sequences, multiple oligos are tested on a microfluidic assay with a specific protein. Using fluorescent proteins attached to the investigated protein, a vast number of time-resolved binding curves is generated within a relatively short time span. On the other hand, it remains a major challenge to analyze this data and extract relevant features like association and dissociation constants in an automated manner with an appropriate handling of data variability including unknown time points at which experimental conditions change.

We applied a traditional ODE-based pathway modelling approach for analysis of the binding kinetics of aptamer microarrays. We first derived a proper model that describes commonly encountered data features, estimates delays caused by diffusion from the data and allows for systematic and reliable extraction of the binding kinetics. Furthermore, we can assess spatial microfluidic characteristics with a model-based approach. Uncertainty analysis of the model and its parameters yields insight about reliability of the results and technical limitations of the experimental technique. This approach allows to analyze the data comprehensively at a comparable rate as it is generated.

Learning the topology of latent signaling networks from high dimensional transcriptional intervention effects

Zahra Sadat Hajseyed Nasrollah
Institute of Medical Statistics and Computational Biology, Faculty of Medicine University of Cologne, Germany

Data based learning of the topology of molecular networks, e.g. via Dynamic Bayesian Networks (DBNs) has a long tradition in Bioinformatics. The majority of methods take gene expression as a proxy for protein expression in that context, which is principally problematic. Further, most methods rely on observational data, which complicates the aim of causal network reconstruction. Nested Effects Models (NEMs – Markowetz et al., 2005) have been proposed to overcome some of these issues by distinguishing between a latent (i.e. unobservable) signaling network structure and observable transcriptional downstream effects to model targeted interventions of the network.

In this study we developed a more principled and flexible approach for learning the topology of a dynamical system that is only observable through transcriptional responses to combinatorial perturbations applied to the system. More specifically, we focus on the situation in which the latent dynamical system (i.e. signaling network) can be described as a network of state variables with nonlinear activation functions. We show how candidate networks can be scored efficiently in this case and how topology learning can be performed via Markov Chain Monte Carlo (MCMC).

We extensively tested our approach by reconstruction of simple network motifs over a wide range of possible settings. Exertion of proposed nonlinear dynamics on the breast cancer proteomic dataset from the DREAM8 challenge highlighted most known interactions that has been reported in STRING. In other context, we aimed to expose interactions among 20 proteins that are involved in EGFR signaling and associated with drug sensitivity in Non Small Cell Lung Cancer (NSCLC). Results have been compared with reconstructed network from classical NEM model (Paurush et al., 2016).

Software development for Bayesian parameter sampling of large intracellular models

Andrei Kramer
KTH/SciLifeLab, Sweden

AbstractModel analysis via Bayesian parameter sampling is one of several useful techniques in systems biology alongside optimization and profile likelihood approaches. Ordinary differential equation models of intracellular mechanisms tend to grow to unwieldy sizes and have hundreds of parameters (such as reaction rate coefficients). Markov Chain Monte Carlo (MCMC) methods are often used for Bayesian parameter sampling [1] and many different MCMC-approaches have been developed to deal with specific problems, especially since simple methods are often slow or do not converge at all. Additionally, biological data itself poses new challenges as it is often unclear how to compare unnormalised data points from uncalibrated and very complex experimental setups to simulation results. The likelihood functions are sometimes complex and can require multiple simulations to replicate one data-point. Finally, the model and data have to be represented in a structured, simple and reusable format. We are developing a software [2], written in C, that deals with many of these challenges. We give an overview of this software and our approach to these problems, and also compare mcmc_clib's properties and organization to other software packages in systems biology today.

[1] Olivia Eriksson, Alexandra Jauhiainen, Sara Maad Sasane, Andrei Kramer, Anu G Nair, Carolina Sartorius, Jeanette Hellgren Kotaleski, Uncertainty quantification, propagation and characterization by Bayesian analysis combined with global sensitivity analysis applied to dynamical intracellular pathway models, Bioinformatics, Volume 35, Issue 2, 15 January 2019, Pages 284–292.
[2] Andrei Kramer, Vassilios Stathopoulos, Mark Girolami, Nicole Radde, MCMC_CLIB – an advanced MCMC sampling package for ODE models , Bioinformatics, Volume 30, Issue 20, 15 October 2014, Pages 2991–2992.

Efficient sampling by marginalization of scaling parameters for mechanistic models with relative data

Elba Raimundez
Helmholtz Zentrum München, 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 marginal sampling scheme for estimating the parameter uncertainties for ordinary differential equation models from relative data. We integrate 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 found that our approach outperforms standard sampling approaches by reducing auto-correlation and requiring a lower computation time.

Abstracts of posters

A. Modeling of signalling, metabolic and gene regulatory pathways

Modeling survival of long-lived plasma cells in the bone marrow

Gustav Gießler, Kevin Thurley
Deutsches Rheuma-Forschungszentrum, Germany

Plasma cells are a crucial part of the adaptive immune system and offer protection against a wide variety of antigens by secreting highly specific antibodies. Although plasma cells, in general, are short-lived effector immune cells, they can be kept alive in the bone marrow for decades under appropriate conditions, in order to provide a long-lasting immune response to residing or recurring threats. The complex signaling network enabling survival is comprised of dozens of components and some key elements have been identified. Based on recent experimental findings by our collaborators from the Radbruch group at DRFZ we developed a simplified model capturing signaling pathways, expression of pro- and anti-survival genes, and subsequent decision-making processes triggering plasma cell survival or apoptosis. Our model shows the switch-like behavior anticipated in such processes and lets us investigate qualitative differences in behavior depending on the included proteins and their properties. We believe this will lead to a better understanding of the underlying mechanisms of long-lived plasma cell survival and may contribute to advances in the treatment of chronic plasma cell-related diseases in the long term.

B. Modeling of diseases and therapies

Overexpression of CBL ubiquitin ligases causes paradoxical amplification of MAPK and PI3K signaling

Kai Hu[1], Chiara Di Ponzio[2], Nathalia Giese[1], Roland Eils[2], Oliver Strobel[1], Stefan Kallenberger[2]
[1]Department of General, Visceral and Transplantation Surgery, University of Heidelberg, Germany
[2]Center for Digital Health, Berlin Institute of Health and Charité Universitätsmedizin Berlin, Germany

Pancreatic ductal adenocarcinoma (PDAC) is one of the most malignant types of cancer. We investigated the role of casitas B-lineage lymphoma (CBL) ubiquitin ligases that were known as negative regulators of RTK signaling. In PDAC cell lines overexpressing CBL-c, we interestingly observed a paradoxical activation of MAPK and PI3K/Akt signaling after treatment with EGF and erlotinib. To understand this phenomenon, we established an ODE model describing epidermal growth factor receptor (EGFR) signaling, MAPK and PI3K/Akt pathways. Interactions among EGFR, CBL-c and adaptor proteins as GRB2 were further confirmed by immunoprecipitation experiments. Assuming that CBL-c increases the signal transduction from active EGFR complexes by acting as scaffold for mediators of downstream kinase signaling, the model could explain the observed increase in Erk and Akt activation in presence of erlotinib. This indicates that CBL-c transiently amplifies MAPK and PI3K signaling while acting as negative regulator on a longer timescale.

Integrative analysis of peripheral N-acetylaspartate metabolism

Polina Lakrisenko[1], Annette Bley[2], Andre Wegner[3], Daniel Weindl[1]
[1]Helmholtz Zentrum München, Germany
[2]Universitätsklinikum Hamburg Eppendorf, Germany
[3]Technische Universität Braunschweig, Germany

Canavan disease (CD) is a rare leukodystrophy caused by mutations in the ASPA gene, leading to severe neurodegeneration and short life expectancy. To date, the exact disease mechanism is poorly understood and therapeutic options are scarce. The ASPA gene encodes aspartoacylase, an enzyme catalyzing the degradation of N-acetylaspartate (NAA). Despite NAA being the second-most abundant metabolite in the mammalian brain, its functional role is poorly understood. Recently, several independent groups found NAA metabolism to also play important roles in non-nervous tissues, e.g. in adipocytes, immune cells, lung and prostate cancer cells, pointing towards a previously overlooked wide relevance of peripheral NAA metabolism.

Together will clinical and experimental partners, we are generating a computational model around NAA metabolism and infer model parameters from stable isotope-assisted metabolomics data, RNA-seq data, and other experimental data. We will use this model to iteratively generate and test various hypotheses around the NAA metabolism and its function. Thereby, we expect to obtain a more profound understanding of the roles of NAA in metabolism and

signaling, particularly in cells outside the brain, which is relevant for understanding and treatment of CD and other diseases.

The fetoplacental interface in the development of pre-eclampsia

Julia Scheel, Shailendra Gupta, Olaf Wolkenhauer

University of Rostock, Germany

Pre-eclampsia (PE) is a multi-systemic, complex disorder occurring during pregnancy and the postpartum. PE affects 2-8% of all pregnancies and often occurs with other complications. Its pathophysiology is largely unknown. To improve the diagnosis and treatment outcomes, a deeper understanding of the determinants of pathophysiology is urgently needed.

Any disturbance within the fetoplacental interface can lead to the development of pregnancy disorders. This is an overview of all essential steps for healthy placentation and highlighted steps that have been shown to be disrupted in pre-eclamptic patients. The pathophysiology of PE can be better understood by analyzing essential processes associated with healthy placentation and pinpointing molecules/processes that are disrupted in PE patients. Starting from openly accessible data we aim to create a disease map of PE, to study the fetoplacental interface, using a systems biology approach. The initial model will be improved, enriched and validated with data and resources provided by iPlacenta consortium partners.

The iPlacenta consortium is an innovative training network of 17 partners within the EU that aims to innovate modelling placenta for maternal and fetal health. Each partner exploits their expertise to study the topic. The fields of research span from engineering, over various experimental approaches (generation of in vitro models, organ-on-a-chip, placental vascular on-a-chip, genetics, epigenetics, metabolomics), collaboration with obstetrics and gynecology medical staff to systems biology and bioinformatics.

The hallmarks of cancer metabolism, a genome-scale metabolic model approach

Ali Salehzadeh-Yazdi[1], **Mahdi Jalili**[2], **Olaf Wolkenhauer**[1]

[1]Department of Systems Biology and Bioinformatics, University of Rostock, Germany

[2]Hematology, Oncology and SCT Research Center, Tehran University of Medical Sciences, Tehran, Iran

Ever since Otto Warburg discovered the characteristics of tumor cell metabolism, a large amount of experimental and computational works have been performed to explore the hallmarks of cancer metabolism. With the advent of genome-scale metabolic models (GEMs), understanding the genotype–phenotype relationship has a theoretical basis. One of the limitations of this approach is its disconnection from other

biological processes, such as genetic regulation. Therefore, for a better prediction of the metabolic activities, different approaches for integration of gene expression data into metabolic pathways have been developed during the past few years. To investigate the hallmarks of cancer metabolism, we integrated various omics datasets (transcriptome and proteome) from different sources (cell-line and patient) into Recon3 to examine functional and consistency performance of several existing genome-scale cancer modeling approaches. We found a metabolic pattern to distinguish cancers by metabolic models. The results show that, although the choice of integration methods has an impact on model functionalities, but we can distinguish cancers by their metabolic patterns.

Phenotypic selection through cell death: stochastic modelling of O-6-methylguanine-DNA methyltransferase dynamics

Ayoub Lasri, Viktorija Juric, Marc Sturrock
Royal College of Surgeons in Ireland

The high propensity of glioblastoma multiforme (GBM) towards genetic mutations often allows treatment-resistant GBM to arise. However, even without genetic mutations, fitness selection processes can expand resistant variants. Recent studies have shown that the development of drug resistance (which is the prime cause of failure in cancer therapy) can occur due to the dynamic non-genetic heterogeneity of cell populations continuously producing meta-stable phenotypic variants. In other words, stochastic expression of certain proteins can produce cells which are deemed “fit” in a stressful environment where “fitness” is usually defined in terms of cell growth rate (with faster growing cells assumed to be fitter). These fit cells which grow quickly are then selected for and this would explain the invariably rapid emergence of stem-like resistant cells. Though there has been some interest in this subject both mathematically and experimentally, there have been no studies to date with respect to GBM specifically. We are developing mathematical models to understand the mechanisms responsible for non-genetic GBM adaptation

towards the identification of novel targets for overcoming therapeutic resistance.

C. Model and data standards

PK-DB: PharmacoKinetics DataBase for individualized and stratified computational modeling

Jan Grzegorzewski, Matthias König
Humboldt University Berlin, Germany

A multitude of pharmacokinetics studies has been published. However, due to the lack of an open database, pharmacokinetics data, as well as the corresponding meta-information, have been difficult to access. We present PK-DB (https://pk-db.com), an open database for pharmacokinetics information from clinical trials as well as pre-clinical research. PK-DB provides curated information on (i) characteristics of studied patient cohorts and subjects (e.g. age, bodyweight, smoking status); (ii) applied interventions (e.g. dosing, substance, route of application); and (iii) measured pharmacokinetic time-courses; and (iv) pharmacokinetic parameters (e.g. clearance, half-life, area under the curve). Key features are the representation of experimental errors, the normalization of measurement units, annotation of information to biological ontologies, calculation of pharmacokinetic parameters from concentration-time profiles, a workflow for collaborative data curation, strong validation rules on the data, computational access via a REST API as well as human access via a web interface. PK-DB enables meta-analysis based on data from multiple studies and data integration with computational models. A special focus lies on meta-data relevant for individualized and stratified computational modeling with methods like physiologically based pharmacokinetic (PBPK), pharmacokinetic/pharmacodynamic (PK/DB), or population pharmacokinetic (pop PK) modeling.

Spatial cytokine gradients in T cell communication sensitively depend on cell-intrinsic properties

Lukas Kiwitz[1,2], Kevin Thurle[1,2]
[1]Institute for Theoretical Biology, Humboldt-Universität zu Berlin, Germany
[2]German Rheumatism Research Center Berlin, Germany

Immune responses are regulated by diffusible mediators, the cytokines, which act at sub-nanomolar concentrations. The spatial component of cytokine signaling is a crucial, yet poorly understood, functional property. Both containment of cytokine action in narrow junctions between immune cells and global signaling throughout entire lymph nodes have been proposed, but the conditions under which they might occur remain unclear. Here, we used analytical tools and three-dimensional finite elements analysis to model time-course and equilibrium distribution of cytokines in a multicellular environment. Using an established parameter set for the well-characterized cytokine interleukin-2, we observe a sensitive dependence of paracrine signaling intensity on receptor levels and receptor distribution among cells. With our explicit model of cytokine gradients in a heterogeneous, in silico T cell population, we seek to rationalize the spatial component of cell-fate decisions, which ultimately set the strength and type of an immune response.

Dynamical modeling of complex diseases uncovers Critical Transitions in their phenotypes

Daniele Proverbio, Jorge Goncalves
University of Luxembourg, Luxembourg

It has been identified that regime shifts in complex dynamical systems are not necessarily smooth but are often abrupt. In particular, evidence has been recently found suggesting that such phenomena exist in system biomedicine, e.g. during the progression of many complex diseases such as atrial fibrillation and epilepsy. The present work develops a modeling perspective, thus bridging nonlinear dynamical system theory and phenomenological system properties. This way we aim at uncovering bifurcations, noise effects or other characteristics of complex networks that lead to a qualitative transition between states occurring at the so-called “tipping points”. Of primary importance is the definition of phenomenological counterparts of abstract mathematical prototype equations. Thus, starting from a classification of said effects, and looking for universal dynamical patterns that are shared among different systems, we then pave the development of statistical tools for model selection and, ultimately, prediction of critical transitions. In addition to theoretical insights, case study applications will be presented and discussed in light of the discussed paradigm; among those, protein regulation and cancer recrudescence. This way, we show how mathematical modeling can be fruitfully coupled to experimental and computational efforts to better understand and intervene on complex biological systems.

Stochastic modeling of the dynamics of proliferating cell populations

Fan Bai, Jan Hasenauer

Hausdorff Center for Mathematics, University of Bonn, Germany

Cell proliferation and population dynamics play an essential role in the study of biological processes. In this work, we present a stochastic modeling and computational framework for proliferating cell populations which undergo symmetric cell division. A CME model is formulated for the processes of cell division and cell death and a SDE model is applied for label degradation. Means and co(variances) are calculated by applying the method of moments.

Managing the INCOME pathway data on the FAIRDOMHub

Dagmar Waltemath[1], Benjamin Winter[2], Xiaoming Hu[2], Yachee Gupta[2], Wolfgang Müller[2]

[1]University Medicine Greifswald, Germany
[2]Heidelberg Institute for Theoretical Studies HITS gGmbH, Germany

SEEK4SCIENCE [1] is a web-based system for the management of scientific research data sets, associated scientists and organisations. It offers flexible and detailed sharing permissions as well as publication pipelines. Furthermore it offers modelling capabilities as well as extraction of RDF descriptions from data sheets. The principal instance of SEEK4SCIENCE is the FAIRDOMHub [2,3]; it is tailored to the management of Systems Biology projects.

FAIRDOMHub manages pathway data and simulation descriptions generated during the INCOME project (https://www.integrative-pathway-models.de/). The system already supports the storage of pathways, datasets and metadata in standard formats (including SBML [4], RDF) and versioning of pathway models [5]. We will describe our ongoing INCOME work including lightweight compliance checking for data tables and improved tooling for storage and ranked retrieval of SBML models.

[1] Wolstencroft, Katherine, et al. "SEEK: a systems biology data and model management platform." BMC systems biology 9.1 (2015): 33. [2] Krebs, Olga, et al. "FAIRDOM approach for semantic interoperability of systems biology data and models." ICBO. 2015. [3] Wolstencroft, Katherine, et al. "FAIRDOMHub: a repository and collaboration environment for sharing systems biology research." Nucleic acids research 45.D1 (2016): D404-D407. [4] Hucka, Michael, et al. "The Systems Biology Markup Language (SBML): language specification for level 3 version 1 core." Journal of integrative bioinformatics 12.2 (2015): 382-549. [5] Scharm, Martin, Olaf Wolkenhauer, and Dagmar Waltemath. "An algorithm to detect and communicate the differences in computational models describing biological systems." Bioinformatics 32.4 (2015): 563-570.

Response-time modeling of T-helper cell differentiation

Philipp Burt, Kevin Thurley
Deutsches Rheuma-Forschungszentrum Berlin, ein Leibniz-Institut, Germany

CD4+ T cell differentiation is a key element of the adaptive immune system driving appropriate immune responses by selective recruitment and activation of effector immune cells. The decision on T cell differentiation into specific subtypes, such as Th1 or Th2 cells, is made by interacting immune cells in a collective process. Such immune cell communication is part of an extensive network involving multiple feedback mechanisms on the intracellular and intercellular level. The complexity of these cell-cell interaction networks confounds intuition and complicates quantitative, systems-level analysis. Here, employing response-time modeling, we develop and analyze data-driven models of T cell activation and decision-making. We find that response-time modeling of such differentiation circuits reveals qualitative and quantitative properties regarding robustness, reaction times and magnitude of the T cell response, which go beyond analysis with traditional rate equation models. We envision using response-time modeling to derive large-scale data-driven models of cell-cell communication circuits in autoimmune diseases, to elucidate decision-making processes and to derive testable predictions regarding therapeutic opportunities.

D. Model reuse, extension and integration

In silico mutagenesis of genes implicated in defects of neuronal migration

Salvatore Contino[1], Maurizio Elia[2] , Francesco Cali[2], Mirella Vinci[2], Valentino Romano[1]
[1]University of Palmero, Italy
[2]Oasi Institute (I.R.C.C.S), Italy

Migration of postmitotic neuroblasts in the fetal brain is an important step of neurodevelopment that brings cells into appropriate spatial relationships. Several cell surface adhesion molecules, extracellular matrix adhesion molecules, and associated signal transduction and cytoskeleton molecules mediate this process to ensure that the developing presynaptic and postsynaptic neurons be in the right place at the right time. Abnormal neuronal migration (NM) may result in cortical malformations that are

responsible for a significant proportion of cases of intellectual disability, psychiatric disorders (Schizophrenia and Autism Spectrum Disorders) and epilepsy.

We have used the Boolean model of 19 NM proteins developed by John et al. (2019) to assess - in silico – the effects of loss-of-function mutations implicated in the pathogenesis of some of these human disorders. Our results show that asynchronous simulation of Boolean networks performed with KO mutations for 15 proteins yielded an altered number and binary structure of attractors compared to the normal network. The most dramatic effects were seen with the KO DISC1 network. We have also undertaken an exome sequence screening of a cohort of epileptic patients with the aim of evaluating by this in silico mutagenesis approach the potential pathogenic role of mutations in NM genes detected in patients . Preliminary results of this work has already led to the identification of a missense mutation in the BBS44 gene predicted to have a deleterious effect on protein function according to the DANN, Mutation Taster, LRT algorithms. Simulation performed with the BBS4 KO Boolean network has uncovered an altered pattern of the attractors' landscape. In light of these interesting findings we propose the use of this Boolean model to gain a more comprehensive and systemic understanding of the cellular and molecular effects of mutations in NM genes in patients affected by inherited disorders of neuronal migration.

Model version merging

Tom Gebhardt, Martin Scharm, Olaf Wolkenhauer

University of Rostock, Germany

Development is often a non-linear process no matter if you plan buildings, program tools, or create biochemical models. Different approaches and improvements typically result in a large number of versions. Especially modellers of biochemical networks tend to keep several versions of their models, but how can two approaches be combined to one? Using a distributed version control system is not sufficient for biochemical models. Instead, the differences between two versions can be detected with the BiVeS algorithm by considering, for example, the model topology, ontologies and similarity scores. Based on BiVeS we present our approach for semi-automatic version merging to support the reuse of models and the collaboration between researchers.

E. Parameter optimization, identifiability analysis and uncertainty analysis

Optimal paths between parameter estimates in nonlinear ODE systems using the nudged elastic band method

Christian Tönsing[1], Jens Timmer[1], Clemens Kreutz[2]

[1]Institute of Physics, University of Freiburg, Germany
[2]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Germany

Mathematical modelling using ordinary differential equation (ODE) models is frequently used to understand the dynamic behavior of cellular components e.g. in biochemical reaction networks. Using experimental data, parameters of the model can be estimated by minimizing the discrepancy between data and the model trajectories using numeric optimization algorithms. Because of the non-linear solutions of the ODE models and the relative, sparse and noisy data in realistic applications, the objective function reveals several local optima in the high-dimensional model parameter space. Thus, parameter estimation in applications remains a challenging optimization problem and is considered as a major bottleneck for mathematical modelling in the context of systems biology.

For statistically valid conclusions about the local optima structure of the objective function and to be able to assess the optimizer performance in multistart optimization sequences, optimal paths in the parameter search space are of general interest. In this work, we adapt the nudged elastic band method and utilize it to calculate optimal paths between parameter estimates. The resulting path profiles can be used to assess the connectivity of multistart optimization results and may be utilized to improve a suboptimal optimization outcome by merging fits by connecting paths. By this, we check if suboptimal optimization outcomes are correctly interpreted as local optima originating from the non-convexity of the objective function as typically presumed, or if they are only a result of e.g. incompletely converged optimization runs and require adaptations in the choice or appropriate configuration of the optimization algorithm.

Noise up your ABC

Yannik Schälte[1], Jan Hasenauer[1,2]

[1]Helmholtz Zentrum München, Germany
[2]Bonn University, Germany

Approximate Bayesian Computation (ABC) is an increasingly popular method for likelihood-free parameter inference in systems biology and other fields of research, since it is broadly applicable to complex stochastic models.

However, the approximation error made by this method is often not clear, in particular in the presence of measurement noise, which is generally present in biological applications.

In this contribution, we thus first illustrate how neglecting noise in ABC, which is easy to do, yields erroneous parameter estimates. Then, we discuss practical ways of correcting for it. In particular, we present a novel sequential algorithm that allows to efficiently perform exact likelihood-free inference for general noise models in a self-tuned manner. The proposed algorithm could improve the accuracy of parameter estimates for a broad spectrum of applications.

Dirac mixture distributions for the approximation of mixed effects models

Dantong Wang[1], Paul Stapor[1], Jan Hasenauer[1,2]

[1]Helmholtz Zentrum München, Germany
[2]Bonn University, Germany

In systems biology, mechanistic models are widely used to describe the mechanism in biological reactions. However, for all these models, parameters need to be estimated for each cell or patient because of cell-to-cell and patient-to-patient variability. Therefore, mixed effect modeling methods are widely used to incorporate the variability in the distribution of parameters. Dirac mixture distributions obtained by Monte Carlo method, quasi-Monte Carlo methods or sigma point methods are developed to approximate the statistics of mixed effects models. Here, we propose an approximation method by optimizing the modified Cramér-von Misés distance (CMD). We assess the accuracy of the different methods using four problems and provide the first scalability study for the CMD method. Our results show that compared to Monte Carlo and quasi-Monte Carlo methods, the CMD method achieves better approximation accuracy with smaller number of points, and smaller uncertainty as well. In contrast to sigma point methods, the CMD method allows for a flexible number of points and better accuracy for nonlinear problems.

Integrating pyABC and Morpheus for the simulation and parameterization of computational models of multi-cellular processes

Emad Alamoudi, Jan Hasenauer
University of Bonn, Germany

Biological tissues tend to be dynamic and highly organized. Multi-cellular models are getting more attention as a way to explain and understand this organization. The parameterization of these models is essential to understand the multi-cellular systems, to predict perturbation experiments and to compare competing hypotheses. However, multi-cellular and multiscale models have been proven to be highly difficult to parameterize. A method that has been proven to be applicable to multi-cellular models is Approximate Bayesian Computation (ABC). Unfortunately, ABC is a computationally expensive approach, as it requires a large number of simulations. Thus, there is an increased need for a fast and general-purpose pipeline for modeling and simulating multi-cellular systems that can exploit HPC systems for faster computations. To this end, we started the development of a user-friendly, open-source, and scalable platform, called FitMultiCell, that can handle modeling, simulating, and parameterizing multicellular systems. To achieve the goal of FitMultiCell, we combine the modeling and simulation tool Morpheus with the advanced statistical inference tool pyABC. In this contribution, we present the current status of the project and demonstrate an application example. An HCV model that was described by Kumberger et al. (Viruses, 10(4), 2018) was used to illustrate the flow of the platform.

F. Model analysis

Modelling HFK20 methylation during embryogenesis in Xenopus laevis

Lea Schuh[1], Carolin Loos[1], Daniil Pokrovsky[2], Axel Imhof[2], Ralph Rupp[2], Carsten Marr[1]
[1]Institute of Computational Biology, Helmholtz Zentrum München, Germany
[2]Department of Molecular Biology, Ludwig-Maximilians-Universität München, Germany

Methylation of histone tails is an essential epigenetic control mechanism regulating gene expression. Particularly during embryogenesis, histone methylation coordinates global differentiation and development. Histone 4 lysine 20 (H4K20) methylation has been identified as a repressive mark important for full differentiation. H4K20 methylation dynamics comprise not only changes during the differentiation process but also methylation restoration upon cell division. However, the precise role of cell cycle on the global dynamics of H4K20 methylation during embryogenesis in Xenopus laevis is unknown.

We analyze snapshot mass spectrometry data of H4K20 methylation in two Xenopus populations - one population developing normally, the other being arrested in G1/S phase - and quantify the global distribution of un, mono- di- and trimethylated sites during the first 42 hours after fertilization. We deterministically model and infer the rate parameters of H4K20 methylation in both populations using a multi-start

maximum likelihood optimization approach. Model selection on over 200 competing model hypotheses identifies essential rate differences between the two populations. We find that the HUA population is best described with only one demethylation rate for all three demethylation events. Additionally, all three methylation rates are necessary for capturing the H4K20 dynamics.

Together, we present a general framework to investigate the dynamics of methylation events and identify the global effect of cell cycle on H4K20 methylation dynamics during embryogenesis.

Introducing an experimentally validated model for DNMT1 methylation dynamics comprising processive and distributive reaction mechanisms

Vincent Wagner, Nicole Radde
University of Stuttgart, Institute for Systems Theory and Automatic Control, Germany

DNA methylation is an important chemical process that plays a role in DNA activation. As an epigenetic reaction, methylation changes the functionality of DNA without changing its sequence [1]. In our project we consider time course data of the DNA methyltransferase 1 (DNMT1) mediated DNA methylation. A DNA strand comprises in total 44 methylation sites, and experimental data consist of frequencies of the numbers of methylated sites at four time points and three different initial DNMT1/DNA ratios. To gain further insights into the governing dynamics, we have developed modeling approaches based on stochastic chemical reaction kinetics to reproduce these data. Our models were calibrated by using moment matching methods, in which the data enter the objective function via their empirical moments [2], which scale with increasing number of data points and work well in this setting. Our main findings are: First, it is important to incorporate information on likelihoods of different methylation sites to become methylated into the models in order to capture the methylation dynamics. Second, by constructing two simplistic models we deduce that neither a solely distributive nor a purely processive reaction scheme can qualitatively reproduce the data, thus showing that DNMT1 is able to operate in both reaction modes. Third, we obtain a satisfying fit even with a very simple reduced model based on quasi-steady state assumptions which comprises only six effective rate constants as free parameters, indicating that we have captured the main mechanisms of the system with our modeling approach.

Investigating heterogeneity in single cell data by applying statistical physics and random matrix approach to biological distribution

Silvia Martina[1], Alexander Skupin[2]
[1]University of Luxembourg, Luxembourg
[2]Luxembourg Centre for Systems Biomedicine, Luxembourg

The aim of my PhD project is to develop a widely applicable theoretical framework tailored to the molecular characterization of single cells. Thanks to the development in the last decade of single-cell RNA-sequencing techniques it is currently possible to analyze transcriptome and genome at single-cell resolution. While these recent technical achievements are now used to characterize cellular heterogeneity, a systematic approach interpreting and exploiting the resulting high-dimensional data for identifying biological principles of development and pathogenesis is still largely lacking. We are addressing these challenges by further developing our statistical analysis and complementing it with a rigorous mathematical framework, starting from concepts based on distribution dynamics. Our goal is to better identify hidden cell states and the sudden critical transition between them during cellular development and differentiation.

In order to do that we are testing different statistical methods and computational standard approaches, the results of which are interpreted in the light of biological insights. We are going to compare the results of these well-known computational techniques with the alternative approaches which we are developing. One of them consists in exploiting the statistical properties of the random matrices, a mathematical framework largely investigated for its particular features in the large size limit. By combining biological constraints and mathematical tools we are investigating for instance the effects of mutations related to Parkinson's disease on the development of dopaminergic neurons. The ultimate goal is to approach heterogeneity by considering probability distributions covering the variability at different biological scales and levels and consequently to set up a corresponding theoretical description of the heterogeneity of living systems.

Construction of a kinetic model of the central carbon metabolism in E. coli describing the switch from aerobic to anaerobic growth

Giulia Slaviero, Steffen Klamt
Max Planck Institute for Dynamics of Complex Technical Systems, Germany

Escherichia coli is one of the most important model organisms and workhorses for biotechnological applications. In recent years, the use of mathematical models, especially of constraint-based metabolic models, helped getting a better understanding of its metabolism and also provided a basis for the computational design of production strains. However, inherent quantitative relationships and the transient behavior of E. coli's metabolism can only be understood with dynamic models. However, despite the rapidly growing amount of experimental data, still, only a very limited number of kinetic models of the central metabolism of E. coli can be found in the literature. This work seeks to improve an existing model [1] with the goal to simulate E. coli's metabolism under changing oxygen concentrations and to predict changes when intervening in its energy and central metabolism. We needed to introduce several modifications in the model including (1) a growth rate law that considers stoichiometric requirements of the precursors, (2) introduction of co-factors, such as ATP, ADP, NADP(H) as explicit metabolites, and (3) proton translocation and ATP synthesis in the electron transport chain. The derived model is validated using experimental data and shows a more robust behavior and improved predictive capabilities compared to the original model. We discuss different application examples including, for example, (1) the predicted behavior of strains with a targeted increase of cellular ATP turnover, and (2) the reorganization of metabolic fluxes when switching from aerobic to anaerobic production.

Detecting Oscillations in Neural Development

Adrian L. Hauber, Christian Sigloch, Samuel Böhm, Wolfgang Driever, Jens Timmer
University of Freiburg, Germany

HER/HES family transcriptional regulators are known to be dynamically expressed in progenitor cell cultures from neural stem cell niches. However, the temporal expression characteristics in the complex in vivo environments in brain stem cell niches are unknown. We use live imaging of fluorescent HER fusion proteins in the zebrafish larval brain to obtain concentration time series of multiple stem and progenitor cells and classify them depending on their characteristics of dynamic expression behavior. Therefore, we apply techniques of spectral analysis, fitting of autoregressive processes and Gaussian process regression to establish a robust classifier that includes cross-checking among different methods. We conclude that specific HER positive progenitor cells have oscillating HER expression.

Stem cells are known to feature intercellular communication via Notch signaling which is influencing HER protein expression. Therefore, coherence is expected between the HER time series of adjacent cells. We use cross-spectral and synchronization analysis to infer relative positions of individual cells from the information in the corresponding time series only.

We construct a model of ordinary differential equations including a Hill-type negative feedback as well as a time delay to mechanistically describe the occurrence of oscillations. By fitting time series data of HER oscillations, we are able to infer parameter values and corresponding confidence intervals. By sampling data from biologically relevant parameter values, we can identify regions in parameter space that determine certain features of oscillations such as amplitude or frequency.

The predictions by the model will be used in experimental designs to better understand the complex spatial and temporal control of stem and progenitor cells in neural stem cell niches.

Distribution-free differential expression analysis for scRNA-seq data across patient groups

Erika Dudkin[1], Jan Hasenauer[1,2]
[1]Faculty of Mathematics and Natural Sciences, University of Bonn, Germany
[2]Helmholtz Zentrum München, Germany

Single cell RNA-sequencing (scRNA-seq) data provide insights into gene expression profiles of individual cells on a large scale. This contributed in recent years substantially to the understanding and identification of cell types and differences between them. To unravel differences between cell populations, a multitude of differential expression (DE) methods has been introduced to compare clusters of cells. However, these methods are not suited for the identification of differences between patient groups for which scRNA-seq data are available. Typically, DE-analysis was performed on a single sample or across multiple samples, leading then to cross-condition analysis. The emergence of scRNA-seq datasets with replicated multi-conditions, for example multiple patients of one condition versus multiple patients of a second condition, demands the development of new particular methods which cover this issue.

In this work, we present a method for the statistical comparison of replicated multi-conditions. The method uses Wilcoxon rank sum test for the pairwise comparison of samples. Differences between patient combinations are evaluated while taking all single cell read counts into account. After calculating the test statistic, its significance is determined with a permutation test.

The method is applied on a scRNA-seq dataset with multiple controls and chronic obstructive pulmonary disease (COPD) patients. Differentially expressed genes were identified and underlying cellular mechanistic hypothesis of COPD could be confirmed by performing gene set enrichment analysis on the resulting DE-gene list.

Automatic generation of priors for large-scale dynamic models

Jakob Vanhoefer[1], Jan Hasenauer[1,2]
[1]Faculty of Mathematics and Natural Sciences, University of Bonn, Germany
[2]Helmholtz Zentrum München, Germany

Background: Mechanistic ordinary differential equation models are a powerful tool to understand biological networks. Recent advances in computational techniques facilitate the study of models with thousands of state variables and parameters, allowing for a more detailed and realistic description of the underlying process. However, when parametrizing a large-scale model, even large data sets do often not provide enough information to provide reliable estimates for all parameters. One way to tackle this problem is to use a Bayesian setting and include prior knowledge obtained from the literature or previous investigations. This facilitates the implicit integration of additional data into the own data analysis. However, the construction of priors for models with a large number of parameters can be tedious.

Method: Here, we present a tool for the automatic construction of priors for SBML models annotated by terms from the Systems Biology Ontology. The tool identified for each parameter the reaction type and kinetic law of the reactions using the parameter. Thereby we infer the biological meaning of a parameter (e.g. protein decay rate) and assign a corresponding prior distribution. The prior information is extracted from data bases like BRENDA.

Results: The resulting prior distributions are written in an PEtab parameter file. Thereby the priors can be used by any toolbox that can import problems specified in the PEtab format. In the future we plan to move from coarse grained to more fine grained and specific priors for the individual parameters.

Venue and conference dinner

Venue

We look forward to welcoming you to the conference location Schankhalle Pfefferberg, Schönhauser Allee 176, Berlin. The venue is centrally located ca. 5 minutes by train (S and U bahn) from Alexanderplatz.

Find information here about how to get there.

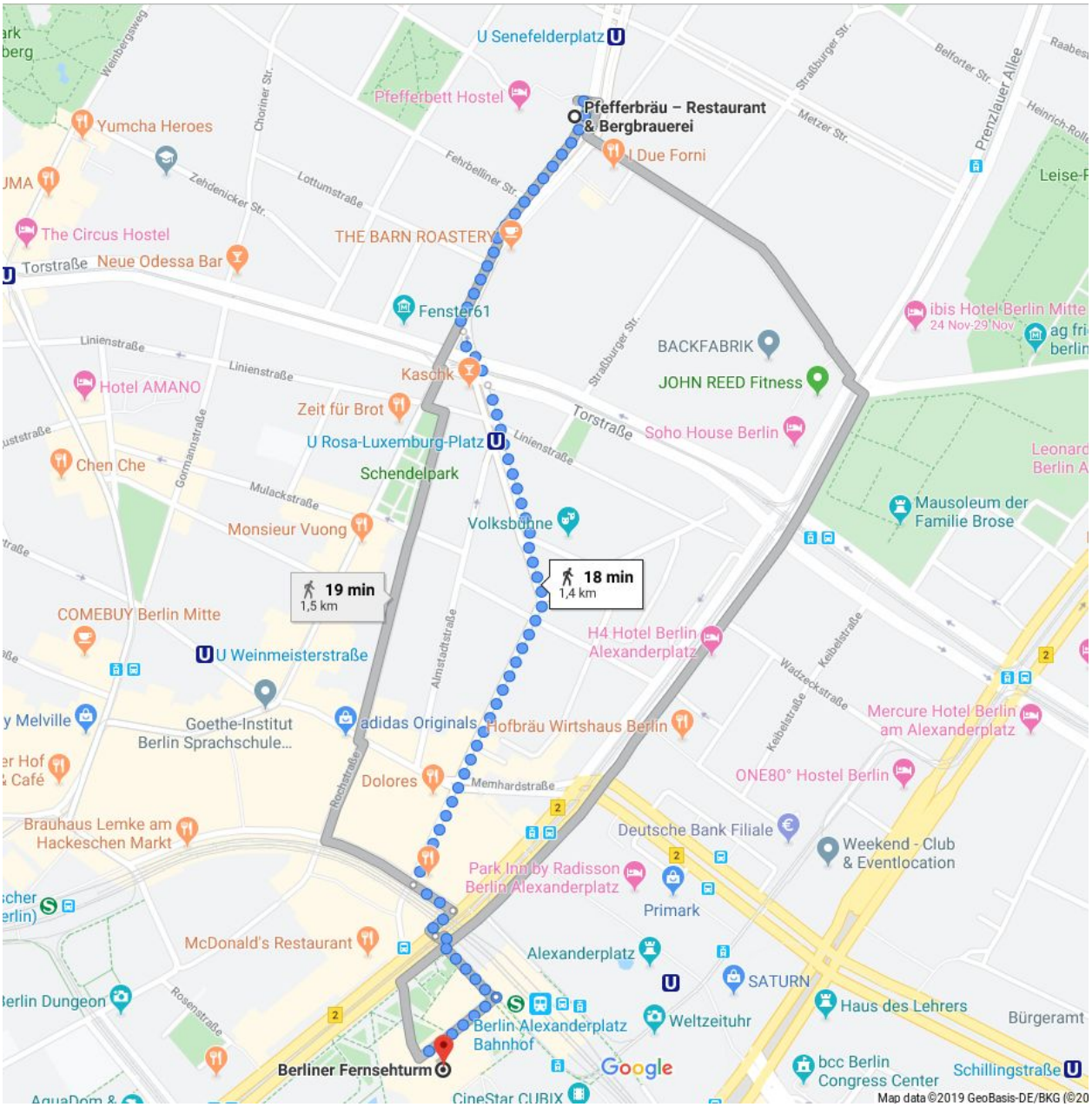
Conference dinner

Copyright @ Torsten Fritsche Die Lichtisten

Welcome to our night in the tallest building in Berlin - conference dinner - starting from 19:00 on the second day of the conference (Tuesday, November 26th)!

It will take place at the Berliner Fernsehturm, located in the heart of the city and near to the conference location.

Path from the conference venue to the restaurant:



Hackathon dinner

Copyright @ Clärchens Ballhaus

After the Hackathon session on Thursday (November 28th) a wonderful evening at Clärchens Ballhaus awaits us. We are looking forward to meeting you at 19:00 Clärchens Ballhaus, Auguststraße 24, 10117 Berlin, https://www.ballhaus.de/en/.