

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Kinetic Models in Life Science
Contributions to Methods and Applications

Joachim Almquist



Department of Biology and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2017

Kinetic Models in Life Science
Contributions to Methods and Applications
JOACHIM ALMQUIST
ISBN 978-91-7597-653-2

© JOACHIM ALMQUIST, 2017.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4334
ISSN 0346-718X

Department of Biology and Biological Engineering
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Phone (switchboard) +46 31 772 1000

Typeset by the author using L^AT_EX.

Printed by Chalmers Reproservice
Göteborg, Sweden 2017

Kinetic Models in Life Science
Contributions to Methods and Applications
JOACHIM ALMQUIST
Department of Biology and Biological Engineering
Chalmers University of Technology

ABSTRACT

Kinetic models in life science combine mathematics and biology to answer questions from areas such as cell biology, physiology, biotechnology, and drug development. The idea of kinetic models is to represent a biological system by a number of biochemical reactions together with mathematical expressions for the reaction kinetics, i.e., how fast the reactions occur. This defines a set of mass balance differential equations for the modeled biochemical variables, whose solution determines the variables' temporal dynamics. Good kinetic models describe, predict, and enable understanding of biological systems, and provide answers to questions which are otherwise technically challenging, unethical, or expensive to obtain directly from experiments.

This thesis investigates the workflow for building and using kinetic models. Briefly, the model question determines a suitable mathematical framework for the mass balance equations, prior knowledge informs selection of relevant reactions and kinetics, and unknown parameters are estimated from experimental data. A validated model is used for simulation and analysis, which is interpreted to gain biological insights.

Three kinetic models were created to illustrate the workflow. First, a model of the antiplatelet drug ticagrelor and the investigational antidote MEDI2452 was developed for the mouse. The model unraveled the biological mechanisms of the pharmacokinetic interaction and predicted free ticagrelor plasma concentration, thereby contributing to the pharmaceutical development of MEDI2452. Second, a model of the Kv1.5 potassium ion channel was integrated within an existing electrophysiological model of a canine atrial cell. The effect of Kv1.5 block on the action potential was simulated, which improved understanding of blocking mechanisms and enabled assessing pharmacological treatment of atrial fibrillation. Third, a nonlinear mixed effects (NLME) model, with population-level distributions of kinetic parameters, was successfully used to describe cell-to-cell variability of the yeast transcription factor Mig1. This model demonstrated the innovative idea of applying NLME modeling to single cell data.

Two studies of kinetic model-building methods are also presented. First, a novel parameter estimation algorithm for NLME models is explained. It computes exact gradients using sensitivity equations, and represents a substantial advancement over its predecessor. Second, a modeling framework is proposed that combines stochastic differential equations with NLME modeling. This promising framework extends the current scope of NLME models by considering uncertainty in the model dynamics.

Keywords: kinetic models; pharmacokinetics; ticagrelor; atrial fibrillation; Kv1.5; cell-to-cell variability; Mig1; NLME modeling; parameter estimation; FOCE

Acknowledgments

This thesis is the result of an industrial PhD project carried out at the Fraunhofer-Chalmers Centre (FCC), in collaboration with the group of Systems and Synthetic Biology (SysBio) at the Department of Biology and Biological Engineering, Chalmers University of Technology. I am very grateful for this opportunity, and I would like to thank my supervisors Associate Professor Mats Jirstrand at FCC and Professor Jens Nielsen in the SysBio group. The work presented in this thesis has to a large extent been supported by funding from the Swedish Foundation for Strategic Research, but also by the EU FP7 project SYSINBIO — Systems Biology as a Driver for Industrial Biotechnology (grant no 212766), the EU FP7 project UNICELLSYS (grant no 201142), the Gothenburg Mathematical Modelling Centre, and the EU FP6 through the BioSim Network of Excellence (grant no 005137). This support is gratefully acknowledged.

I would also like to thank all coauthors, coworkers, and friends at FCC, SysBio, AstraZeneca R&D Gothenburg, and elsewhere. A special recognition goes out to my coauthors and close collaborators Mikael Wallman, Jacob Leander, Tim Cardilin, Helga Kristín Ólafsdóttir, Robert Andersson, Monika Sundqvist, Peter Gennemark, Holger Becker, Loubna Bendrioua, and Johan Gabrielsson, from whom I have learnt a lot. Finally, I would like to thank my family for their wonderful support.

Joachim Almquist
Göteborg, November 2017

List of Publications

This thesis is based on the following appended papers:

- A Almquist J**, Cvijovic M, Hatzimanikatis V, Nielsen J, and Jirstrand M (2014). Kinetic models in industrial biotechnology — Improving cell factory performance. *Metab. Eng.* 24: 38–60.
- B Almquist J**, Penney M, Pehrsson S, Sandinge AS, Janefeldt A, Maqbool S, Madalli S, Goodman J, Nylander S, and Gennemark P (2016). Unraveling the pharmacokinetic interaction of ticagrelor and MEDI2452 (ticagrelor antidote) by mathematical modeling. *CPT Pharmacometrics Syst. Pharmacol.* 5 (6): 313–323.
- C Almquist J**, Wallman M, Jacobson I, and Jirstrand M (2010). Modeling the effect of Kv1.5 block on the canine action potential. *Biophys. J.* 99 (9): 2726–2736.
- D Almquist J**, Bendrioua L, Adiels CB, Goksör M, Hohmann S, and Jirstrand M (2015). A nonlinear mixed effects approach for modeling the cell-to-cell variability of Mig1 dynamics in yeast. *PLoS One* 10 (4): e0124050.
- E Almquist J**, Leander J, and Jirstrand M (2015). Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood. *J. Pharmacokinet. Pharmacodyn.* 42 (3): 191–209.
- F** Leander J, **Almquist J**, Ahlström C, Gabrielsson J, and Jirstrand M (2015). Mixed effects modeling using stochastic differential equations: Illustrated by pharmacokinetic data of nicotinic acid in obese Zucker rats. *AAPS J.* 17 (3): 586–596.

Other relevant publications coauthored by Joachim Almquist:

Bendrioua L, Smedh M, **Almquist J**, Cvijovic M, Jirstrand M, Goksör M, Adiels CB, and Hohmann S (2014). Yeast AMP-activated protein kinase monitors glucose concentration changes and absolute glucose levels. *J. Biol. Chem.* 289 (18): 12863–12875.

Tapani S, **Almquist J**, Leander J, Ahlström C, Peletier LA, Jirstrand M, and Gabrielsson J (2014). Joint feedback analysis modeling of nonesterified fatty acids in obese Zucker rats and normal Sprague-Dawley rats after different routes of administration of nicotinic acid. *J. Pharm. Sci.* 103 (8): 2571–2584.

Cvijovic M, **Almquist J**, Hagmar J, Hohmann S, Kaltenbach HM, Klipp E, Krantz M, Mendes P, Nelandér S, Nielsen J, Pagnani A, Przulj N, Raue A, Stelling J, Stoma S, Tobin F, Wodke JA, Zecchina R, and Jirstrand M (2014). Bridging the gaps in systems biology. *Mol. Genet. Genomics* 289 (5): 727–734.

Cardilin T, **Almquist J**, Jirstrand M, Sostelly A, Amendt C, El Bawab S, and Gabrielsson J (2017). Tumor static concentration curves in combination therapy. *AAPS J.* 19 (2): 456–467.

Andersson R, Kroon T, **Almquist J**, Jirstrand M, Oakes ND, Evans ND, Chappel MJ, and Gabrielsson J (2017). Modeling of free fatty acid dynamics: insulin and nicotinic acid resistance under acute and chronic treatments. *J. Pharmacokinet. Pharmacodyn.* 44 (3): 203–222.

Pehrsson S, Johansson KJ, Janefeldt A, Sandinge AS, Maqbool S, Goodman J, Sanchez J, **Almquist J**, Gennemark P, and Nylander S (2017). Hemostatic effects of the ticagrelor antidote MEDI2452 in pigs treated with ticagrelor on a background of aspirin. *J. Thromb. Haemost.* 15 (6): 1213–1222.

Cardilin T, **Almquist J**, Jirstrand M, Zimmermann A, El Bawab S, and Gabrielsson J (2017). Model-based evaluation of radiation and radiosensitizing agents in oncology. *CPT Pharmacometrics Syst. Pharmacol.* Accepted.

Abbreviations

AF	-	Atrial fibrillation
AP	-	Action potential
APD	-	Action potential duration
BFGS	-	Broyden-Fletcher-Goldfarb-Shanno
FD	-	Finite difference
EKF	-	Extended Kalman filter
FOCE	-	First-order conditional estimation
GEM	-	Genome-scale metabolic model
GFP	-	Green fluorescent protein
NLME	-	Nonlinear mixed effects
ODE	-	Ordinary differential equation
PD	-	Pharmacodynamic(s)
PK	-	Pharmacokinetic(s)
RNC	-	Ramirez-Nattel-Courtemanche
SDE	-	Stochastic differential equation
STS	-	Standard two-stage

Contents

Abstract	iii
Acknowledgments	v
List of Publications	vii
Abbreviations	ix
I Introductory Chapters	1
1 Introduction	3
1.1 Kinetic Models in Life Science	3
1.2 The Idea of Modeling	4
1.3 Similarities and Differences of Kinetic Models	6
1.3.1 Biological Systems	6
1.3.2 Mathematical Frameworks	7
1.4 The Aim of This Thesis	8
2 Mathematical Background	9
2.1 The Standard Kinetic Model	9
2.2 Nonlinear Mixed Effects Models for Populations	10
2.3 Models with Stochastic Dynamics	11
2.4 Nonlinear Mixed Effects Models with Stochastic Dynamics	12
3 Results and Discussion	13
3.1 Short Summary of the Appended Papers	13
3.2 The Workflow for Kinetic Modeling	13
3.3 Pharmacokinetics of Drug-Antidote Interaction	18
3.4 Effect of Kv1.5 Blockers on the Canine Atrial Action Potential	22

3.5 Cell-to-Cell Variability of Transient Glucose Sensing in Yeast	26
3.6 Exact Gradients for Nonlinear Mixed Effects Models	30
3.7 Nonlinear Mixed Effects Modeling with Stochastic Dynamics	33
4 Discussion and Conclusions	37
4.1 Kinetic Models Fit Many Life Science Questions	38
4.2 The Question Guides the Framework	38
4.3 The Model Structure Is the Heart of the Model	39
4.3.1 Be Aware of Uncertainty	39
4.3.2 Mechanistic Models Are Preferred	39
4.3.3 Avoid Unnecessary Complexity	40
4.4 Parameter Estimation Is a Critical Step	41
4.5 In Silico Experiments Are the Reward	42
4.6 Answering the Model Question	42
4.6.1 Pharmacokinetics of Drug-Antidote Interaction	43
4.6.2 Effect of Kv1.5 Blockers on the Canine Atrial Action Potential .	43
4.6.3 Cell-to-Cell Variability of Transient Glucose Sensing in Yeast .	43
4.7 A Brief Outlook	44
4.8 A Final Remark	45
Bibliography	47
II Appended Papers	59
A Kinetic Models in Industrial Biotechnology — Improving Cell Factory Performance	61
B Unraveling the Pharmacokinetic Interaction of Ticagrelor and MEDI2452 (Ticagrelor Antidote) by Mathematical Modeling	87
C Modeling the Effect of Kv1.5 Block on the Canine Action Potential	101
D A Nonlinear Mixed Effects Approach for Modeling the Cell-to-Cell Variability of Mig1 Dynamics in Yeast	115
E Using Sensitivity Equations for Computing Gradients of the FOCE and FOCEI Approximations to the Population Likelihood	149
F Mixed Effects Modeling Using Stochastic Differential Equations: Illustrated by Pharmacokinetic Data of Nicotinic Acid in Obese Zucker Rats	171

Part I

Introductory Chapters

CHAPTER 1

Introduction

This chapter gives a short introduction to what kinetic models are and how they are used in life science. It includes a brief description of the idea of modeling in general and of mathematical and kinetic models in particular, and points out some of the similarities and differences among kinetic models. These ideas lead up to the aims of the thesis, which are stated at the end of the chapter together with an outline of how they are addressed in the appended papers.

Following this introduction, a brief background on selected mathematical topics is given in Chapter 2. Then, results and discussions are presented in Chapter 3. Finally, additional discussions and conclusions are given in Chapter 4.

1.1 Kinetic Models in Life Science

Kinetic models are found across all the life sciences. They are used both in basic research, such as cell biology (Klipp et al. 2005; Ramsey et al. 2006) and physiology (Karaaslan et al. 2005; Silber et al. 2007), and in applied fields, like biotechnology (Stephanopoulos et al. 1998; Wiechert and Noack 2011) and drug development (Gabrielsson and Weiner 2016; Gennemark et al. 2017). Just like other types of mathematical models, kinetic models can be used to describe, understand, and predict how biological systems of interacting components function and behave (Kitano 2002; Riel 2006; Motta and Pappalardo 2013). Biological systems of interest in life science can be anything from a collection of molecules to a cell, a multicellular organ or organism, or even a population or ecosystem. Due to features such as a high degree of connectivity, strong feedbacks and regulations, and the presence of nonlinear interactions, the behaviors of these systems and their corresponding models tend to be highly complex (J Nielsen 2017), even for apparently simple systems consisting of just a couple of components (Tyson et al. 2003).

Kinetic models are primarily built from mass balance equations that involve different reactions and their kinetics, i.e., how fast the reactions proceed. This makes kinetic modeling especially suitable for the study of dynamic processes. The mass balances are usually defined by a set of time-dependent first-order ordinary differential equations (ODEs), which means that these models are essentially equivalent to what control engineering refers to as state-space representations. The reactions in kinetic models are often synonymous with well-defined biochemical reactions, e.g., enzymatic conversions of metabolites. They may however equally well describe processes on a higher level of organization, such as the rate of cellular proliferation, or changes of more abstract entities like transitions between healthy and diseased states of an individual. The notion of a reaction should therefore be interpreted in its broadest sense. The number of mass balance equations and reactions can differ a lot between models but are typically in the range of a few up to a hundred. These models are normally implemented and solved numerically using computers (Mendes and Kell 1998).

Given the high flexibility of what can be represented by kinetic mass balance equations and these equations' inherent ability of accounting for change, it is not surprising to learn that kinetic modeling is used for a wide range of topics involving dynamic processes such as adaptive bacterial metabolism (Kotte et al. 2010), aging of yeast (Erjavec et al. 2008), the development of diabetes (Topp et al. 2000), therapies for tumor growth inhibition (Cardilin et al. 2017a), and vaccination strategies for epidemics (Shulgin et al. 1998), just to name a few. The scope of kinetic models is as diverse as the field of life science itself.

1.2 The Idea of Modeling

Modeling is frequently driven by the desire to solve a certain problem or finding the answer to one or more specific questions. Typically, real world problems or questions involve complicated systems and the solutions and answers are not easily obtained. The universal idea of modeling is to create a simplification or abstraction of the system under study. By doing so, one can imagine the question being transferred from the real world into the model world.¹ Once posed in the model world the question will, hopefully, appear more comprehensible as the problem is stripped of irrelevant and confusing complexity, leaving only the essential core of the matter left. The answer can now be worked out more easily and subsequently be formulated in relevant terms of the real world context. This idea is depicted in Fig. 1.1 (which incidentally happens to be a model in itself).

Mathematical models are a special type of models. They express the abstractions of the model world quantitatively using different kinds of equations and mathematical statements (Wolkenhauer 2014; Torres and Santos 2015; Helmlinger et al. 2017; J Nielsen 2017). This is a powerful approach since (i) it allows a precise and

¹Sometimes the model world involves different kind of physical models that are part of the real world, e.g., the use of cultured cells or animal models. The behaviors of those models are not necessarily less complex as such, but they may nevertheless be considered simplifications for experimental and ethical reasons.

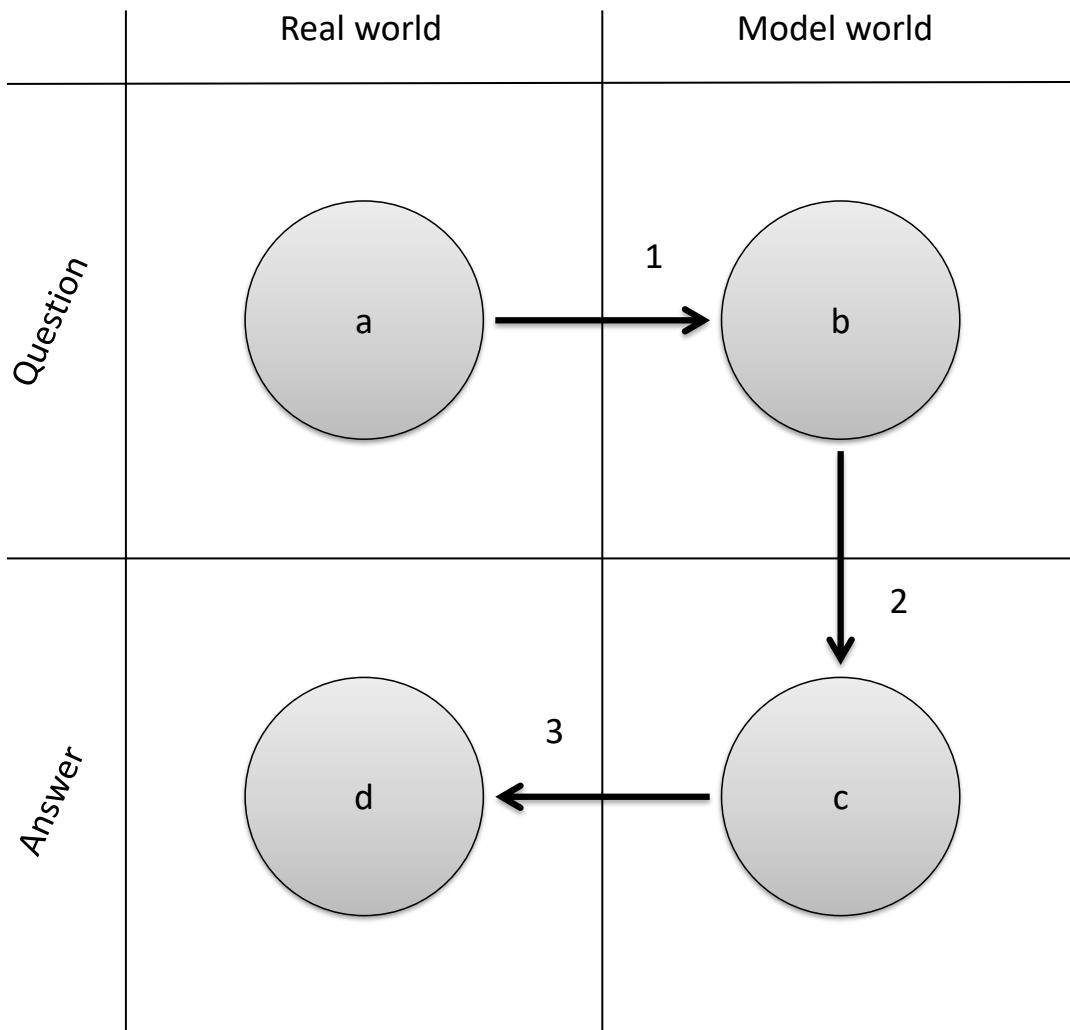


Figure 1.1: The idea of modeling. Starting with a question in the real world (a), the process of creating a model (1) moves the question into the model world (b), where it more easily can be answered (2). When the question has been answered in the model world (c), the answer must be translated back (3) to the real world (d) in a meaningful way.

unambiguous definition of the model; and (ii) the modeler can make use of a large volume of already existing results from the mathematical and statistical disciplines to support the modeling process. It is also a difficult approach because things that would have only vaguely been expressed in a verbal model must now instead be made much more concrete and quantitative. Mathematical models, properly defined, leave no wiggle room for alternative interpretations as far as their mathematical meaning is concerned. This stringency requires the modeler to carefully weigh different considerations when deciding how to define the model. Challenging as this may be, the fact that mathematical modeling forces this process of critical thinking to take place may be one of its strongest merits (Wolkenhauer 2014).

For kinetic models in life science, step 1 of Fig. 1.1 is usually the hardest. It involves most of the assumptions and decisions that eventually determines whether the modeling effort will be fruitful or not. For example, it involves understanding what the essential components of the system under study are, choosing an appropriate mathematical framework and defining the model equations, and deciding what type of data that should be collected in order to facilitate the modeling. Many of these activities requires the modeler to collaborate with non-modeler experts, and probably also requires the modeler to at least acquire some basic level of knowledge in the relevant field. Step 2 is typically quite straightforward for kinetic models, although it may occasionally involve extensive computational power. It does normally not rely on the interaction with non-modelers. Depending on the purpose of the modeling, it may involve tasks like optimization (Alvarez-Vasquez et al. 2000), derivation of mathematical relations (Cardilin et al. 2017a), explorative simulation of what-if-scenarios (Bajpai and Reuss 1980), or inspecting the estimate of one or more important parameter values (Pfeffer et al. 2011). These activities are sometimes referred to as *in silico* experiments. Step 3 again involves collaborations with the non-modeler experts to interpret the findings and to critically challenge the conclusions of the modeling. It may also trigger additional questions that lead to another round of modeling.

1.3 Similarities and Differences of Kinetic Models

Two ideas are central to all kinetic models. The first is to define the rate of change of one or more state variables due to some reactions, i.e., a set of mass balance equations. The second is to mathematically define the kinetic rate expressions (or rate laws) of these reactions. The rate expressions may depend on the state variables, on a set of kinetic parameters, and on external time-varying entities. The model state variables and the reactions thus have a mutual and dynamic dependence on one another; the state variables, i.e., the reactants, are changing according to the magnitude of the reaction rates, which in turn are determined by the magnitude of the reactants (and any potential effectors). Apart from this common foundation, kinetic models in life science are actually a quite heterogenous group of models. They differ with respect to the type and abstraction-level of the systems they describe, as well as with respect to the mathematical framework being used.

1.3.1 Biological Systems

The type of systems and processes described by kinetic models are very different, ranging for instance from the kinetics of enzymes and pathways at the subcellular level (Rizzi et al. 1997), to the whole-body pharmacokinetics (PK) of drugs (Boger et al. 2016), and the kinetics of animal populations (Lotka 1925; Volterra 1926). As a consequence of the multiplicity among kinetic models, the state variables may represent anything from specific metabolites, to aggregated pools of related molecular species, the cell membrane electrical potential, net body weight, or population sizes.

Occasionally, a subset of the state variables does not even have a clear interpretation — they have just been introduced for making the model better at fitting some observed phenomena. The same diversity is found in the spectrum of reaction kinetics. Some models are built around authentic and mechanistic representations of multi-step reactions by considering a detailed account of all elementary reactions or by mechanistically derived approximations like the Michaelis-Menten kinetics (Michaelis and Menten 1913; Chen et al. 2010), while the kinetics of other models is described by so-called empirical or phenomenological models (Monod 1942; Menezes et al. 1994; Gabrielsson and Weiner 2016), or by generic expressions with some mathematically or biophysically favorable properties (Heijnen 2005; Liebermeister et al. 2010).²

1.3.2 Mathematical Frameworks

There are several different types of mathematical frameworks that can be used to formulate the mass balance equations and the kinetic rate expressions. Kinetic models are almost always operating in continuous time, but examples of discrete-time models implemented by means of difference equations exist (Petersson et al. 2010). Many models are spatially dependent, most commonly by considering different discrete compartments (Hammarlund-Udenaes et al. 2008), but there are also many examples of continuous temporal-spatial descriptions using partial differential equations of the reaction-diffusion type (Kholodenko 2006), or even more complex methods like the Euler-Lagrange approach for single cells in turbulent flows of stirred bioreactors (Lapin et al. 2004). Another distinction can be made between models with deterministic kinetics and models with stochastic kinetics (Ullah and Wolkenhauer 2010). The stochastic models can be further separated into those that consider discrete stochastic events (Gonze et al. 2008) and those that consider continuous stochastic processes (Hasty et al. 2000). Kinetic models also differ with respect to what their parameters mean and how they are defined mathematically. The most common type of model is still based on a single fixed set of parameter values but different kind of more ambitious approaches that recognize the importance of parameter uncertainty are gaining ground (Chakrabarti et al. 2013). This includes the branch of formal Bayesian methods that aim at determining the whole posterior probability of the parameter values (Saa and LK Nielsen 2016). Yet other approaches look at the inter-individual parameter variability by defining a probability distribution of the parameter values at the population level, a modeling paradigm referred to as nonlinear mixed effects (NLME) modeling (Lindstrom and Bates 1990; Kuhn and Lavielle 2005) or hierarchical Bayesian modeling (Huang et al. 2006). Moreover, different mathematical frameworks are sometimes combined to form new ones, such as the cross-fertilization of NLME modeling with kinetics described by stochastic differential equations (SDEs) (Tornøe et al. 2004).

²Although Michaelis-Menten kinetics and Monod kinetics are equivalent from a mathematical point of view, the former is a mechanistically derived approximation from the underlying elementary kinetics while the latter is an empirically established rate law.

1.4 The Aim of This Thesis

This thesis addresses the use of kinetic models in life science. Although these models can appear to be very different on the surface, there is, at least to some degree, a common workflow and a shared set of mathematical and statistical modeling methods that unify the kinetic modeling approach. The aim is to

- A1** identify and describe the different steps in the workflow for building and using a kinetic model,
- A2** apply the kinetic modeling approach to address some relevant questions within life science, and
- A3** contribute to the development of new methods for building kinetic models.

The thesis contains six papers, which roughly speaking addresses the three aims in the following way:

Paper A is a review of kinetic modeling in biotechnology. In addition to providing many examples of applied modeling, it describes the workflow for kinetic modeling in detail, thereby addressing **A1**.

Paper B presents a kinetic model of the interaction between the drug ticagrelor and its antidote MEDI2452. This modeling application from pharmacokinetics addresses **A2**, and to some extent **A1**.

Paper C investigates the impact on the action potential of blocking the ion channel Kv1.5, using a kinetic model. This modeling application from electrophysiology addresses **A2**, and to some extent **A1**.

Paper D shows how a kinetic model can quantify cell-to-cell variability of transient glucose sensing in yeast. This modeling application from microbial signal transduction addresses **A2**, and to some extent **A1** and **A3**.

Paper E develops a parameter estimation algorithm for NLME models based on sensitivity equations. This method contribution addresses **A3**, and to some extent **A1**.

Paper F explores a modeling approach based on combining NLME modeling with SDEs. This method contribution addresses **A3**, and to some extent **A1** and **A2**.

CHAPTER 2

Mathematical Background

This chapter gives a brief mathematical background to the contributions of this thesis. First, the mathematical notation for a standard kinetic model is introduced. Second, it is shown how this model can be expanded to a NLME model in order to account for inter-individual parameter variability in populations of individuals. Third, the expansion of a standard kinetic model to a model based on SDEs is outlined. Finally, inter-individual parameter variability is combined with stochastic kinetics to form the SDE-NLME framework.

2.1 The Standard Kinetic Model

The dominating mathematical framework for kinetic models in life sciences is a set of first-order time-dependent ODEs. It is used to formulate a set of mass balance equations

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{S}\mathbf{r}(t), \quad (2.1)$$

where t is the independent time variable, $\mathbf{x}(t)$ is an m -dimensional vector of time-dependent state variables, \mathbf{S} is an $m \times n$ -dimensional stoichiometric matrix, and $\mathbf{r}(t)$ is an n -dimensional vector of time-dependent reaction rates. The reaction rates are further defined to be dependent on $\mathbf{x}(t)$, a set of parameters $\boldsymbol{\theta}$, and on a time-dependent input function $\mathbf{u}(t)$,

$$\mathbf{r}(t) = \mathbf{r}(\mathbf{x}(t), \boldsymbol{\theta}, \mathbf{u}(t)). \quad (2.2)$$

The input function is used to represent known and varying quantities such as an experimental protocol for the application of some stimuli to the system. The stoichiometric matrix and the reaction definitions are usually not written out explicitly.

It is instead more common to see the complete right hand side lumped together into a function \mathbf{f} , i.e., the mass balance equations are written as

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\theta}, \mathbf{u}(t)). \quad (2.3)$$

The differential equations always have a set of accompanying initial conditions

$$\mathbf{x}(t_0) = \mathbf{x}_0(\boldsymbol{\theta}), \quad (2.4)$$

which may depend on the parameters.

To model observations of the system under study, i.e., experimental measurements, a set of observation equations are also required. These equations combine a deterministic function \mathbf{h} of the model state variables, the parameters, and the input function, with a vector of stochastic observation errors \mathbf{e}_t , to describe a vector of discrete-time observations at time t ,

$$\mathbf{y}_t = \mathbf{h}(\mathbf{x}(t), \boldsymbol{\theta}, \mathbf{u}(t)) + \mathbf{e}_t. \quad (2.5)$$

It is standard practice to let the errors \mathbf{e}_t be normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\mathbf{x}(t), \boldsymbol{\theta}, \mathbf{u}(t))$. Observations made at different time points are furthermore assumed to be independent with respect to the stochastic error.

In addition to the differential equations and the observation equations, auxiliary variables are sometimes introduced to simplify the model formulation or to make it more intuitive. An example of this is the frequently applied approach of initially formulating mass balance equations in terms of time-dependent amounts, e.g., $a(t)$, but then replace the amount variable with the product of the corresponding concentration $c(t)$ and a volume parameter V , $a(t) = c(t)V$. The auxiliary concentration variable can then be used to more conveniently define some of the reaction rates, e.g., inter-compartment transport reactions which depend on concentration differences. It can also simplify the observation equations since most experimental methods measure concentrations rather than amounts. However, it is important to note here that balance equations in terms of concentrations can in general not be set up for multi-compartment models since concentrations are not conserved quantities (which explains the notion of “mass” in mass balance equations). If concentrations are desired as model state variables they have to be introduced afterwards by a change of variables, as in the example above.

2.2 Nonlinear Mixed Effects Models for Populations

The standard kinetic model considers the stochastic variable in the observation equation as the only source of variability in experimental data. As an additional source of data variability, NLME models introduce variability in the model parameters (Lindstrom and Bates 1990; Kuhn and Lavielle 2005). This is done by changing the meaning of model parameters from constants to stochastic variables such that

different instances of the model have different realizations of the parameter values. The point of having different instances with respect to the parameter values is to create unique models for all individuals in a population, but still use the same model equations for every individual. At the population level, the realization of these stochastic parameters can be assumed to follow a parametric probability distribution. The choice of this distribution and its parameter values is now also part of the model definition. This approach of expanding the model with an additional layer for the statistics of the individual parameters is sometimes referred to as hierarchical modeling. It is important to note that the inter-individual parameter variability is a one-off realization that affects all observations for an individual. Thus, in contrast to the observation errors \mathbf{e}_t which are independent both between and within individuals, variability of the parameter values contributes to correlated observations within individuals.

It is common to subdivide parameters of NLME models into constant *fixed effects* that are the same for all individuals, and stochastic *random effects* that are different between individuals, hence the term mixed effects models. By combining fixed and random effects, a set of composite parameters for the i th individual may be written

$$\varphi_i = \varphi(\boldsymbol{\theta}, \boldsymbol{\eta}_i), \quad (2.6)$$

where φ can be any nonlinear function, $\boldsymbol{\theta}$ are the fixed effect parameters similar to those of the standard kinetic model, and $\boldsymbol{\eta}_i$ are the random effect parameters for the i th individual. The individual parameters φ_i are used in the same way as $\boldsymbol{\theta}$ are used in the standard kinetic model. The random effects are furthermore defined to be normally distributed with zero mean and covariance matrix $\boldsymbol{\Omega}$. However, the function φ enables the transformation of the normal distribution into other distributions. Both fixed effect parameters and random effect parameters, once realized, can be viewed from either the frequentist's perspective, i.e., they are deterministic but unknown, or from the Bayesian perspective, as probability distributions. In the latter case, the model is often referred to as a hierarchical Bayesian model (Huang et al. 2006).

2.3 Models with Stochastic Dynamics

Another way of introducing additional variability to the standard kinetic model is to consider randomness in the dynamics of the state variables. This can be done by expanding the ODEs to SDEs by adding a stochastic term to the right-hand side (Jazwinski 1970). Written on differential form, the model becomes

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\theta}, \mathbf{u}(t))dt + \mathbf{G}d\mathbf{w}, \quad (2.7)$$

where the function \mathbf{f} defines the deterministic part of the dynamics like previously, \mathbf{G} is a weight matrix of dimension $m \times q$, and $d\mathbf{w}$ is a q -dimensional vector of Wiener increments (independent, zero mean, and variance dt). If \mathbf{G} is chosen to be the stoichiometric matrix \mathbf{S} , possibly scaled by some diagonal matrix, the randomness of the dynamics become completely associated with the reactions in an additive way.

It is then the kinetics of the individual reactions that are stochastic, rather than the net dynamics of the state variables. This furthermore ensures that the model equations still preserve the balances of, e.g., mass or charge.

The solution to the SDEs can be seen either as a particular realization of a stochastic process, or in a probabilistic sense in which the dynamics of a probability distribution for the state variables is described by a deterministic partial differential equation known as the Fokker-Planck or Kolmogorov forward equation (Jazwinski 1970). In either case, since particular realizations of the Wiener increments affects all future states of the model, SDE-based models lead to correlated observations within particular model realizations.

2.4 Nonlinear Mixed Effects Models with Stochastic Dynamics

The standard kinetic model can simultaneously be expanded with both inter-individual parameter variability and stochastic dynamics, as described separately above. This results in an SDE-NLME framework (Tornøe et al. 2004). Here, three sources of data variability are taken into account at the same time, namely (i) the *intra*-individual variability due to randomness in the observations; (ii) the *inter*-individual variability due to parameter variability; and (iii) the *intra*-individual variability due to uncertainty in the dynamics.

CHAPTER 3

Results and Discussion

This chapter summarizes and discusses the results of the appended papers. First, a short summary of all papers is given in the form of a table. Then, one section is devoted to each of the papers. The primary objective of this chapter is not to repeat the contents of the papers, but rather to lift some selected topics and to offer new perspectives based on the thesis aims.

3.1 Short Summary of the Appended Papers

A short summary of the appended papers is given in Table 1. This table provides an overview of the key aspects both within and across the papers of this thesis—all in one spread.

3.2 The Workflow for Kinetic Modeling

Paper A reviews the use of kinetic models in metabolic engineering and industrial biotechnology. Although the scope of the review is narrower than the scope of this thesis, the same modeling principles are generally applicable for most life science research areas involving kinetic models. As such, the review may be valuable to a larger audience.

The core of Paper A consists of a detailed account of the workflow for how kinetic models are set up. The steps of the workflow, and how they are interrelated, are depicted in Fig. 3.1. This is a similar illustration to that in Fig. 1 in Paper A, but provides more detail regarding the role of experimental data and the iterative aspects of the workflow. The process underlying the construction of most kinetic models in life science can be understood from this workflow, including the contributions in this thesis. A brief explanation of these steps now follows.

Table 3.1: Short summary of papers.

	Paper A	Paper B	Paper C
Title	Kinetic models in industrial biotechnology — Improving cell factory performance.	Unraveling the pharmacokinetic interaction of ticagrelor and MEDI2452 (ticagrelor antidote) by mathematical modeling.	Modeling the effect of Kv1.5 block on the canine action potential.
Background	Kinetic models guide genetic engineering and support the design of bioprocesses.	MEDI2452 is an antidote for the platelet aggregation inhibitor ticagrelor.	Prolonging the action potential duration may prevent atrial fibrillation.
Objective	Describe the workflow for kinetic modeling and review methods and applications.	Understand interaction between ticagrelor and MEDI2452.	Investigate impact of Kv1.5 potassium ion channel block on the atrial action potential.
Method	Review literature and perform survey in the modeling-community.	Derive a joint pharmacokinetic model of drug and antidote in the mouse.	Integrate mechanistic Markov-model of open-channel Kv1.5 block with existing model of the canine atrial action potential.
Results	Detailed account of how kinetic models are built and applied within biotechnology.	Model explains counter-intuitive experimental results and predicts free ticagrelor concentration following MEDI2452 treatment.	Model predicts action potential dynamics in the presence of Kv1.5-targeting drugs.
Limitations	Industry use of kinetic modeling is not always disclosed in scientific publications.	Model does not account for population variability.	Model studies electrophysiology at the cell level, but fibrillation is occurring at the organ level.
Impact	Fills literature gap by summarizing theory and applications in a single document.	Model has been used to design and interpret a new study in the pig.	First model to link the detailed kinetics of Kv1.5 block with cell-level variables like the action potential duration.

Table 3.1: Short summary of papers. (continued)

	Paper D	Paper E	Paper F
Title	A nonlinear mixed effects approach for modeling the cell-to-cell variability of Mig1 dynamics in yeast.	Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood.	Mixed effects modeling using stochastic differential equations: Illustrated by pharmacokinetic data of Nicotinic acid in obese Zucker rats.
Background	Populations of isogenic cells display cell-to-cell variability, but kinetic models seldom address this.	Parameter estimation in NLME models can be slow and unstable.	Models with uncertain kinetics may compensate for incomplete and/or incorrect model structures.
Objective	Characterize cell-to-cell variability of the transient re-localization of the yeast transcription factor Mig1.	Improve the speed and robustness of the FOCE and FOCEI parameter estimation methods.	Demonstrate a kinetic modeling framework that can handle both population variability and uncertain kinetics.
Method	Combine the nonlinear mixed effects modeling framework with single cell time series data.	Compute gradients of the FOCE(I) likelihood using sensitivity equations instead of finite differences.	Apply the SDE-NLME modeling framework to both synthetic data and pharmacokinetic data of nicotinic acid in obese Zucker rats.
Results	Model describes Mig1 dynamics and predicts population distributions of the response time, amplitude, and duration.	Algorithm computes numerically robust gradients and gives considerable speed-up compared to existing methods.	Framework allows the identification of parameter variability, uncertain kinetics, and measurement errors, and also reduce estimate bias.
Limitations	The phenomenological character of the model limits biological interpretations and conclusions.	Requires symbolic differentiation capability, and uses a likelihood approximation.	Synthetic data were generated with the SDE-NLME model used for re-estimation, and not with a different ODE-NLME model.
Impact	One of the first applications of nonlinear mixed effects models to single cell data.	Algorithm has been incorporated in industry standard commercial software.	SDE-NLME modeling is a promising approach that may become the next generation of population models.

Question Modeling starts with a question, as argued in Section 1.2.¹ The question defines the modeling task and will influence all subsequent steps of the modeling workflow. Occasionally, the question may already from the beginning be phrased as a specific hypothesis, otherwise this happens (indirectly) when the model is formed.

Model structure The model structure can be viewed as the combination of a so-called network structure and the kinetic rate expressions. The network structure or network topology is the “wiring diagram” of the model. It can sometimes be translated more or less directly from the pathway or interaction diagrams that biologists frequently use. The network structure defines which reactions that are taking place, including any reaction modulators, but does not give any further quantitative information about the reactions. This is instead determined by specification of the kinetic rate expressions which act as sub-models for each reaction. Together with the network structure they form a complete model by means of a set of mass balance equations. The model structure also involves a decision about what mathematical framework that should be used to represent the balance equations (see Chapter 1 and 2 for the most common examples).

Parameter values Kinetic models differ with respect to the philosophy of how values of model parameters are being determined. Parameter values are sometimes determined from a priori knowledge, obtained for instance from a literature review or from experiments designed to measure a specific parameter directly. A competing approach for parameter determination is based on estimation using experimental data from the system under study. This works by simultaneously trying to tune the model parameters such that the model behavior matches the experimental observations as closely as possible. It may be necessary to adjust the complexity of the model structure to ensure that parameter values can be estimated properly.

Experimental design and data The choice of model structure and the estimation of parameter values are coupled with the design and generation of new experimental data within the scope of the modeling project. Determination of the experimental protocol can be considered as an integrated part of modeling and may involve activities such as identifiability analysis and optimal design.

Validation When a model structure and parameter values have been proposed it is common to perform some kind of model validation to increase confidence in the model. This can be done in several ways, ranging from sanity checks and consistency checks with prior knowledge, to comparison with newly generated data from the system under study. A failure of this step should lead to a reconsideration of the model structure and the parameter values.

Model usage and answer The final part of the modeling consists of using the model to obtain an answer to the question that initiated the modeling. It may involve various types of analysis and predictive simulations—sometimes referred to as *in silico* experiments—as well as an educated interpretation of the results in the context of the model question.

¹This step was referred to as “model purpose” in Paper A, but had the same meaning.

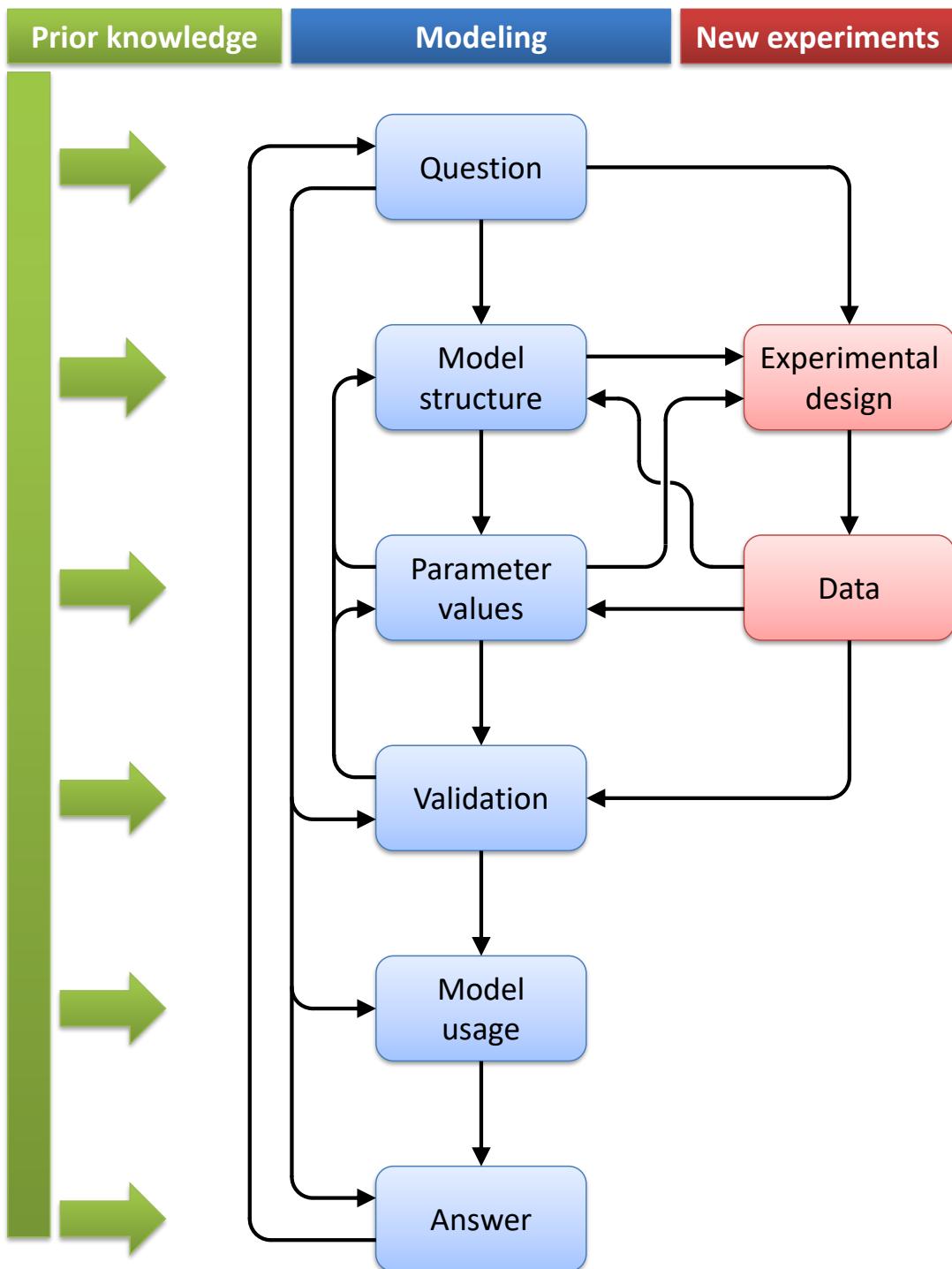


Figure 3.1: The workflow for kinetic modeling. Boxes represent different steps of the modeling workflow and arrows indicate the influence of one step on the other. Starting with a question, a model structure is defined and parameter values are determined. This is usually coupled with the design and generation of new experimental data. The quality of the model is often assured through some kind of validation before it is used to, for instance, perform some simulations that contribute towards answering the question. All steps are influenced by prior knowledge, and the order in which they are traversed is typically characterized by an iterative cycling.

The workflow in Fig. 3.1 is rarely performed in a linear manner from question to answer. It is rather a highly iterative process that involves repeated cycling through various steps (Riel 2006). The exact path will be unique for each modeling effort and cannot be predicted beforehand. The iterative signature of the workflow can even go beyond the original question, as the resulting answer may trigger entirely new questions. Also, what can be considered prior knowledge changes over time as science moves forward and new data become available. It is therefore natural to see models as temporary; some models will require slight alterations while others become obsolete or evolve into something different when new information is presented.

The kinetic models reviewed in Paper A span a wide range of biological questions and mathematical frameworks. Yet, the way that they have been set up is mostly well explained by the workflow in Fig. 3.1. The fact that many of the reviewed modeling projects involved generation of new data also clearly demonstrates a key point of Fig. 3.1, namely that mathematical modeling is closely related to the work of experimentalists (Klipp et al. 2005; Marucci et al. 2011; Bowden et al. 2014; Pehrsson et al. 2017).

3.3 Pharmacokinetics of Drug-Antidote Interaction

Ticagrelor is an antiplatelet drug (Van Giezen et al. 2009) approved for the treatment of acute coronary syndrome (Wallentin et al. 2009) and for long-term preventive use in patients with prior myocardial infarction (Bonaca et al. 2015). Antiplatelet therapies reduce the risk of blood clots, but at the same time also increase the risk of bleeding (Wallentin et al. 2009). Unlike several other antiplatelet medicines, ticagrelor binds reversibly to its target receptor (Van Giezen et al. 2009) which provides an opportunity for developing a specific antidote that reverses the effect of ticagrelor. Such an antidote would be a valuable treatment option for a patient on ticagrelor in the event of a major bleeding. The ticagrelor-neutralizing antibody fragment, MEDI2452, is currently in a preclinical development program (Buchanan et al. 2015; Pehrsson et al. 2017). If successful, MEDI2452 would be the first antidote for an antiplatelet drug.

Paper B presents a kinetic model of the drug-antidote interaction of ticagrelor and MEDI2452. This modeling work exemplifies most, if not all, aspects of the kinetic modeling workflow previously described and illustrated in Fig. 3.1. The question driving the modeling effort was to understand the PK interaction between ticagrelor and MEDI2452 in the mouse. In particular, a model was desired that could predict the time course of free ticagrelor resulting from different administration schemes for the drug and the antidote.

To form the model structure it was important to identify all biochemical species, reactions, and compartments that could be considered to be relevant for the drug-antidote interaction. The resulting model network structure is shown in Fig. 3.2. Based on this, kinetic mass balance equations were set up for all biochemical species.

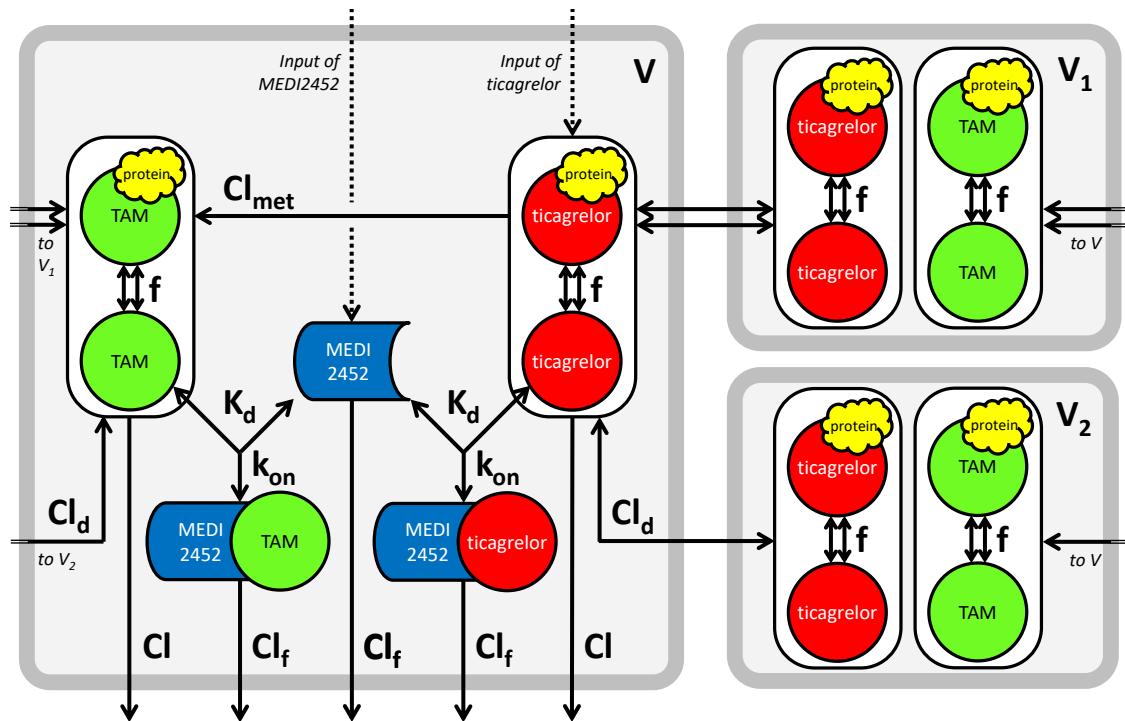


Figure 3.2: Ticagrelor-MEDI2452 pharmacokinetic model. Reactions assumed to equilibrate instantaneously are indicated by double arrows. Input to the system (ticagrelor and MEDI2452) are shown as dashed arrows. The rapid equilibriums of free and protein-bound ticagrelor and ticagrelor active metabolite (TAM) are depicted by encapsulated entities. The fractions of free ticagrelor and TAM within these entities are determined by the parameter f . The total contents of free and protein-bound ticagrelor and TAM in the plasma compartment (V) are cleared at the rate Cl , and ticagrelor is additionally being metabolized to TAM at the rate Cl_{met} . The total content of the encapsulated ticagrelor entity may furthermore distribute instantaneously to one peripheral compartment (V_1), and more slowly, with the intercompartmental clearance Cl_d , to another (V_2). Free ticagrelor and TAM in the plasma compartment can reversibly bind to free MEDI2452 with the rate k_{on} , forming complexes with dissociation constant K_d . Both the complexes and free MEDI2452 are cleared at the rate Cl_f .

As an example, the equation for ticagrelor in the plasma, $TicaV(t)$, was defined as

$$\begin{aligned}
 V \times TicaV'(t) = & \\
 & + TicaInput(t) \\
 & - Cl_{fast} \times (TicaV(t) - TicaV_1(t)) \\
 & - Cl_d \times (TicaV(t) - TicaV_2(t)) \\
 & - Cl_{met} \times TicaV(t) \\
 & - Cl \times TicaV(t) \\
 & - V \times k_{on}(f \times TicaV(t) \times FabV(t) - K_d \times FabTicaV(t))
 \end{aligned} \tag{3.1}$$

where $TicaInput(t)$ is the administration of ticagrelor, $TicaV_1(t)$ and $TicaV_2(t)$ are ticagrelor concentrations in the two other compartments, $FabV(t)$ and $FabTicaV(t)$ are the plasma concentrations of MEDI2452 and ticagrelor-bound MEDI2452, V is the plasma volume, Cl_{fast} , Cl_d , Cl_{met} , and Cl are kinetic clearance parameters, k_{on}

is a second-order association rate constant, f is the unbound fraction of ticagrelor, and K_d is the affinity of MEDI2452 for ticagrelor. The terms of the right-hand side of this equation describes, in order, external administration of ticagrelor, the inter-compartment clearance of ticagrelor to and from V_1 and V_2 , the specific metabolism of ticagrelor to its active metabolite, the clearance of ticagrelor, and the ticagrelor-MEDI2452 complex-formation. The mass balance equations in Paper B were initially formulated in terms of state variables for amounts, before concentrations were introduced as state variables according to the explanation in Section 2.1.

Before the model shown in Fig. 3.2 could be set up, two independent PK models of separately administered ticagrelor and MEDI2452 were created as an intermediate step. The combined PK model for ticagrelor and MEDI2452 was then formed by merging these models. Parameters of the combined model were determined using prior information from the literature and from parameter estimation using time series data from experiments of separately administrated ticagrelor or MEDI2452. This illustrates how the model structure was identified iteratively, which is very common as argued previously. Moreover, during the modeling project, new data were generated and used for validation and subsequent refinement of the model, adding further to the iterative aspect of the workflow.

The model was used to answer the initiating question in different ways. First of all, the model could explain the mechanism behind why total ticagrelor and free ticagrelor in plasma show opposite response after administration of MEDI2452. It was also shown how the predicted time-dependent concentration of free ticagrelor could drive the pharmacodynamic (PD) response of platelet aggregation. Finally, an interesting prediction was made about how free ticagrelor is being measured. According to the model, an *in vivo* blood sample may be far from equilibrium with respect to the complex-formation between ticagrelor and MEDI2452, but the equilibrium will eventually be reached *in vitro* before the bioanalysis is complete. This requires a special observation model for the equilibrium concentration of free ticagrelor. By comparing the model of *in vivo* free ticagrelor with the special observation model of the measured *in vitro* free ticagrelor, it was concluded the measurements may severely underestimate the actual free concentration of ticagrelor (see Fig. 3.3). The exact extent of this effect depends on the experimental protocol and varies over time, but underestimation of free concentrations by roughly an order of magnitude may occur for time periods of up to an hour.

During the modeling process it became apparent that potential recycling of ticagrelor from the cleared ticagrelor-MEDI2452 complex might be important for the behavior of the model. Thus, in line with the reasoning in Section 1.2, a new question arose as a consequence of the critical thinking that is forced to take place during modeling. The question of potential recycling was approached by modifying the model structure to account for various levels of recycling and then examining the feasibility of those alternative models. The result of this analysis favored a scenario where no significant recycling occurs, and where ticagrelor bound to MEDI2452 is eliminated as a complex via the urine.

When the model was complete and Paper B had been published, the mouse PK model was translated to the pig and used to support the evaluation of MEDI2452's

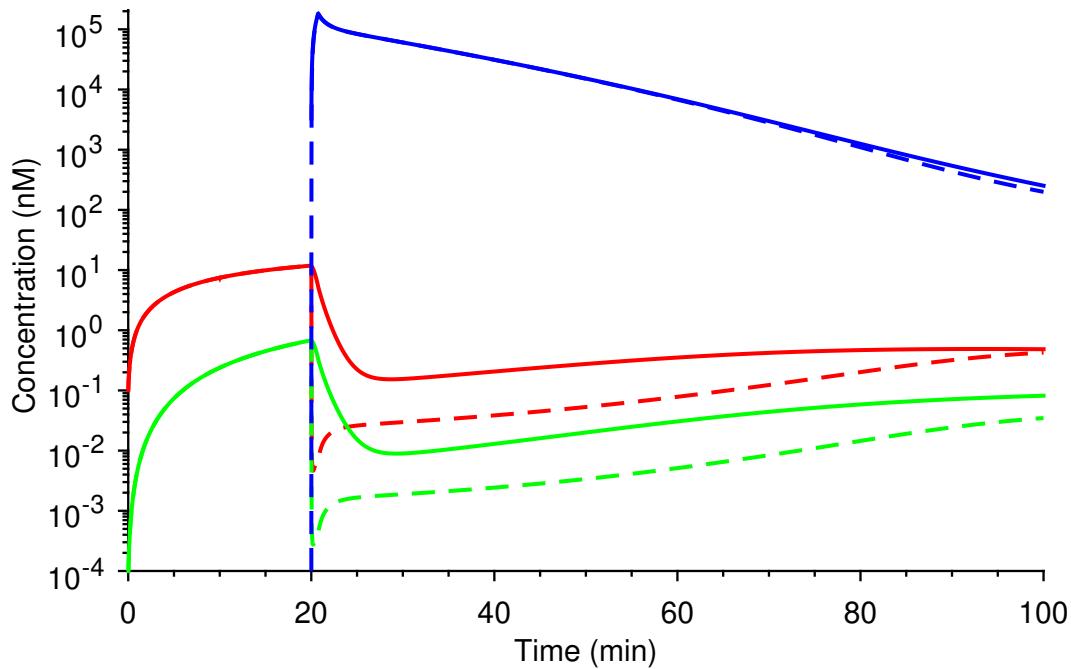


Figure 3.3: The observation model. Comparison of key model state variables and the corresponding observation model for Study design 3: free ticagrelor (red solid line) versus observed free ticagrelor (red dashed line) in plasma, free TAM (green solid line) versus observed free TAM (green dashed line) in plasma, and free MEDI2452 (blue solid line) versus observed free MEDI2452 (blue dashed line) in plasma. For further details, see Paper B.

ability to restore hemostasis in a pig animal-model of major bleeding (Pehrsson et al. 2017). This pig PK model was used to design the experiments and interpret the data presented in the main paper (Pehrsson et al. 2017). The details of the pig PK model are explained in the supplementary material (Pehrsson et al. 2017). Furthermore, the pig study measured the urine concentrations of ticagrelor, which can be used as a test of the mouse model prediction that ticagrelor is not recycled. After administration of only ticagrelor, the concentration of ticagrelor in the urine was just above the lower level of quantification. However, when both ticagrelor and MEDI2452 was administered, ticagrelor concentration in the urine was roughly a thousand times greater. It was intriguing to see that the pig PK model predicted these observations in a good way (Pehrsson et al. 2017). This provides further validation to the basic principles of both the mouse and pig PK models.

The data used for the drug-antidote interaction model of Paper B was collected from many different animals and displayed a clear inter-individual variability. Still, a standard kinetic model was used and the parameters were estimated according to the so-called naïve-pooled approach (Ette and Williams 2004). Expanding the PK interaction model from a standard kinetic model to a NLME model, as explained in Section 2.2, would therefore be a possible direction for future work. Accounting for inter-individual variability would most likely be essential to make the model clinically relevant, if translated to a human setting in the future.

3.4 Effect of Kv1.5 Blockers on the Canine Atrial Action Potential

Atrial fibrillation (AF) is a common form of heart arrhythmia that is characterized by a fast and unorganized beating of the upper chambers (Nattel 2002). It is associated with an increased risk of death and an increased risk of cardiovascular and renal disease (Odutayo et al. 2016). AF is a complex electrophysiological process that spans over multiple scales of space, time, and biological organization. At the tissue- or organ-level, AF manifests itself as self-sustained propagation of spiraling and circling electrical waves. It is a collective behavior of a large number of atrial cells that emerges from the single cell level where brief electrical impulses that change the cellular membrane potential are transmitted from one cell to another. These impulses are known as action potentials (APs) and in addition to their role as carriers of information between cells, they couple electrical activity with the mechanical contraction of cells that ultimately makes the heart beating. The AP, in turn, has a highly complex dependency on several ionic currents that flow across the cellular membrane. Many of these currents are mediated by voltage-gated ion channels that can open and close on a millisecond time scale depending on the current state of the membrane potential. Thus, one aspect of understanding AF is to understand the dynamic relation between the properties of the atrial AP and the underlying ion channel currents.

An important characteristic of the AP is the so-called refractory period, the period during which the cell cannot be stimulated to fire another AP. The duration of the refractory period puts a lower limit on the tissue length-scale of a possible re-entrant circuit. In this way, a short refractory period increases the risk for sustained re-entry propagation of APs, which is believed to be a main mechanism behind AF. Many pharmacological treatment strategies for AF therefore aim to increase the duration of the refractory period by using drugs that block ion channels responsible for the repolarizing potassium currents (Dobrev et al. 2012). This includes the repolarizing current carried by the voltage-gated potassium ion channel Kv1.5 (Wettwer and Terlau 2014). This ion channel is of particular interest since it is only expressed in the atrium and a specific Kv1.5 blocker could thereby possibly avoid the undesired prolongation of the refractory period in ventricular cells.

The study of the AP using mathematical modeling has a long history and has been made famous through the pioneering work of Hodgkin and Huxley, who already in 1952 used a kinetic model based on ODEs to explain the AP generation in the giant squid axon (Hodgkin and Huxley 1952). Since that, many AP models have been developed for different kind of excitable cells (Noble et al. 2012; Glynn et al. 2014; Heijman et al. 2016). Due to the use of animal models this includes mathematical models tailored for specific species, such as the Ramirez-Nattel-Courtemanche (RNC) model of the canine atrial AP (Ramirez et al. 2000).

Paper C uses a kinetic model to investigate the effect on the canine atrial AP when drugs with different properties are used to block the Kv1.5 channel. An example of this is shown in Fig. 3.4. In particular, the action potential duration (APD) is studied. The APD is closely related to the duration of the refractory period, and

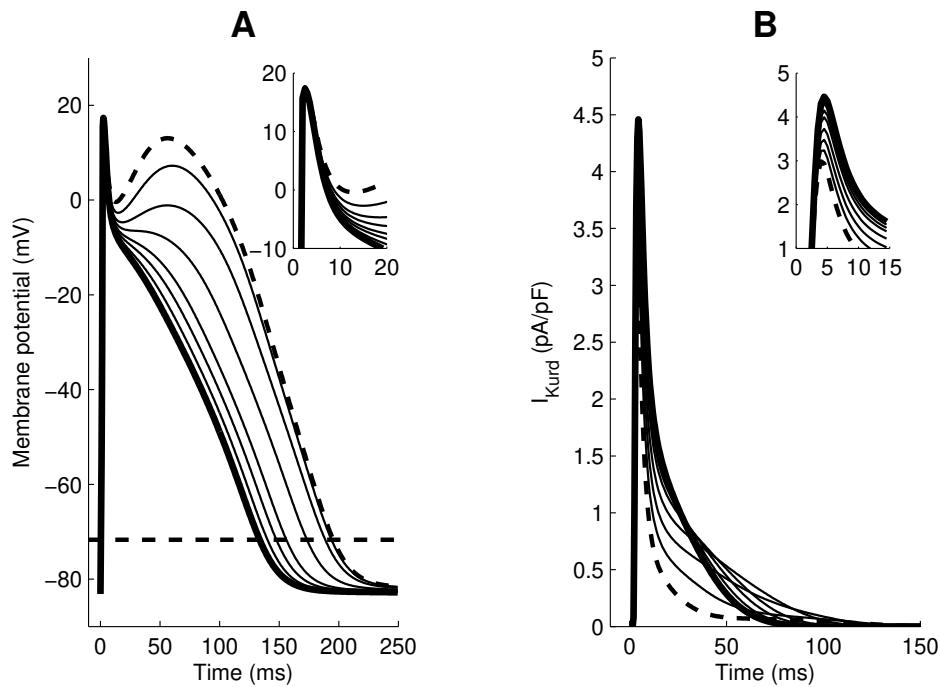


Figure 3.4: Model of the action potential in the presence of a hypothetical drug. The action potential (**A**) and the I_{Kurd} current carried by Kv1.5 (**B**) are shown for a remodeled PM-cell stimulated at 1 Hz. **Insets** show magnification of the traces during the spike. Bold traces correspond to no drug, solid traces to drug concentrations of 1, 2, 3, 5, 8, and 13 μM , respectively, and dashed traces to 21 μM . The APD₉₀ level is shown as a horizontal dashed line in (A). For further details, see Paper C.

it is used as a marker for assessing the impact of pharmacological treatment of AF. The modeling starts off from the well-established RNC model of the canine atrial AP (Ramirez et al. 2000). Then, the part of the model describing the I_{Kurd} current through the Kv1.5 ion channels is replaced by a mechanistic Markov-type model that allows the precise action of a blocking drug to be incorporated (Rudy and Silva 2006). Specifically, the new Kv1.5 model describes a selective open-channel block, the mechanism by which most Kv1.5 blockers are believed to operate (Dobrev et al. 2012). The RNC-integrated Markov model of Kv1.5 offers a possibility to explore the impact on the AP resulting from Kv1.5 blocking drugs with kinetic parameters and electrical properties of different types. Using in silico experiments like the one shown in Fig. 3.4, both actual compounds and a large range of hypothetical drugs with different properties could be evaluated. The corresponding systematic exploration of drug parameters in living cells in the wet lab would have required a huge amount of experimental work, if even at all possible.

Model simulations suggested that the APD increased with both the effective rate of receptor binding (product of drug concentration and rate of association to the open state) and with the rate of drug-receptor complex dissociation. Compared to a naïve model with state-independent block, differences in prolongation of the APD were large, suggesting that it is crucial to choose a model that matches the actual blocking mechanism. It was also found that open-channel block produce a

reverse use-dependence, i.e., that the increase in APD becomes smaller for higher frequencies of AP stimulation. Furthermore, the Kv1.5 model was modified beyond the straightforward mechanism of open-state block to account for uncharged drugs that display a voltage-dependent recovery from block. To the author's knowledge, this is the first Kv1.5 model capable of capturing this behavior. This extended version of the model was subsequently used to analyze two actual compounds that display such characteristics (Lagrutta et al. 2006). The results of that analysis were in line with previous experimental work (Lagrutta et al. 2006), thereby providing some confidence with respect to the validity of the model.

The Kv1.5 model in Paper C is a good example of how model structure complexity of an initial model can be reduced by imposing some assumptions. The Kv1.5 ion channel is composed of four identical subunits that can transition between a closed and an open conformation. If all four subunits of the channel are in the open conformation, a pore is formed between the intra- and extracellular spaces that selectively let potassium ions flow through. An initial model considers all $2^4 = 16$ possible configurations of the subunits within a channel and keeps track of the reactions that transforms one configuration into another, see Fig. 3.5A. If it now is assumed that subunit transition kinetics are independent of the state of other subunits within a channel (Fig. 3.5B), several configurations become equivalent and a much simpler model structure with only 6 states can be used instead (Fig. 3.5C). If a state-dependent block of the channel had not been considered, the model could have been reduced further to only describe the two possible states of each subunit. This would have been enough since the state of the subunits within a channel are always independent. Such a model is essentially a Hodgkin-Huxley model with a single gating variable and a gating variable exponent equal to 4 (Rudy and Silva 2006). However, state-dependent block introduces dependencies between the states of the subunits within a channel, and the model used in Paper C becomes necessary.

Model reduction of the type described in Fig. 3.5 is a common approach in modeling of ion channel kinetics and this part was therefore only mentioned in the passing of the results section in Paper C. It nevertheless deserves attention since it

Figure 3.5: Model reduction of potassium ion channel model. **(A)** Illustration of all possible ion channel configurations under the assumptions that each of the four subunits are either in a closed or open conformation and that there is a single drug-blocked state which is only reachable when all subunits are open, i.e., from the fully open configuration. Each channel configuration is shown within a box with black edges. Possible transitions between different configurations are indicated with black lines. Within the boxes, closed subunits are shown as red disks, open subunits as green disks, potassium ions as yellow disks, and drug molecules as blue diamonds. The positioning of the red disks is done to highlight that the closed conformation physically hinders the potassium ions from passing through the channel. Ions can only pass when all subunits are open and when the channel is not blocked by a drug. **(B)** An assumption is made that the transition rates between the closed and open conformations of a subunit are independent of the complete channel configuration. **(C)** Equivalent channel configurations are lumped together, forming a reduced model. The equivalent configurations in (A) have the same number of open/closed subunits and are vertically aligned with one another and with the states of the reduced model.

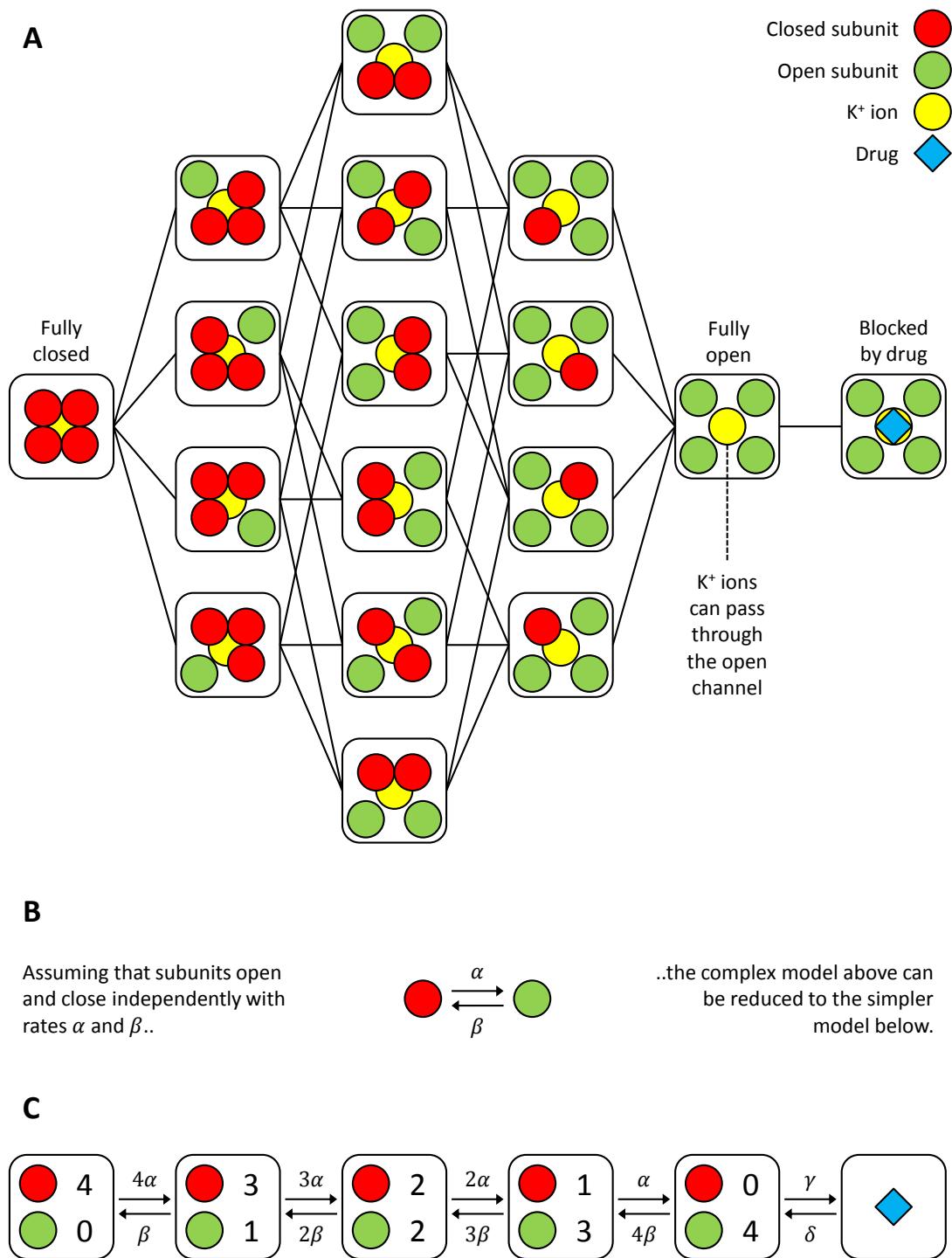


Figure 3.5: Model reduction of potassium ion channel model.

illustrates that model reduction is a central part of the kinetic modeling workflow and that it is something that modelers do all the time in order to limit the complexity of the model structure. It is perhaps especially important for modeling of ion channels where the combination of multiple subunits and multiple states of each subunit quickly can result in an explosion of possible ion channel configurations.

3.5 Cell-to-Cell Variability of Transient Glucose Sensing in Yeast

The yeast *Saccharomyces cerevisiae* is a well-studied microorganism. Already in 1996 it became the first eukaryotic genome to be completely sequenced (Goffeau et al. 1996). This yeast has been used in the production of food and beverages for a long time and it is nowadays also serving as a so-called cell factory for industrial fermentations (Hong and J Nielsen 2012). Last but not least, it is also an important model organism for eukaryotes (Petranovic et al. 2010; Botstein and Fink 2011). Lately, there has been an increasing awareness that yeasts and other microbes experience significant cell-to-cell variability, even among isogenic populations cultured under homogenous conditions. The variability concerns both static levels of mRNA, proteins, and other molecules, as well as their temporal profiles (Lidstrom and Konopka 2010; Gustavsson et al. 2012; Bendrioua et al. 2014; Durandau et al. 2015). These insights have largely been driven by the emergence of new experimental technology allowing different kinds of single cell measurements. As a result of the experimental development there is now a need for an accompanying advance of mathematical methods for modeling cell-to-cell variability.

Contributions to variability between cells can be separated into so-called intrinsic and extrinsic noise. Intrinsic noise refers to biochemical reactions that are inherently noisy or stochastic to their nature, for instance due to low copy-number effects. Not only do such reactions play out differently in different cells, but they can also be thought of as giving different outcome if realized repeatedly within the same cell. Extrinsic noise can be understood as cell-specific differences in for instance cell cycle stage or in enzyme concentrations. Extrinsic noise may affect a given reaction differently in different cells, but can be considered stable within a cell, at least on certain time scales. There is a lot of support to the consensus belief that extrinsic noise is the dominating source of variability in many biological processes, see (Elowitz et al. 2002; Raser and O’Shea 2004; Pedraza and Oudenaarden 2005; Colman-Lerner et al. 2005; Kollmann et al. 2005; Hilfinger and Paulsson 2011; Gaudet et al. 2012) and references within. Additional variability may furthermore arise from differences in the external environment of cells, such as the heterogeneity of industrial-scale bioreactors (Lapin et al. 2004). From a modeling perspective, intrinsic noise is synonymous with a model defined by stochastic kinetics. Extrinsic noise may on the other hand in many cases still be modeled by a standard deterministic kinetic model at the single cell level, but different instances of the model—each with its own set of parameter values—must be used for different cells. These single cell parameters can then be assumed to originate from a random event that follows some probability

distribution at the cell population level. From a model-identification perspective, there is however a risk that the information content in each single cell data set is too low compared to the complexity of a proposed model, a scenario which can make estimation of single cell parameters and of the corresponding population distributions problematic (Sheiner and Beal 1981; PM Wright 1998).

Under the premises described above, Paper D argues that NLME modeling can be a suitable framework for studies of cell-to-cell variability. The advantage of the NLME approach is that all single cell data are considered simultaneously rather than on an isolated per-cell basis. This increases the possibility of correctly estimating the population-level parameter distributions. Paper D then introduces cell-to-cell variability of the yeast transcription repressor Mig1 as a case study where NLME modeling may be applied. In short, the following dynamics of Mig1 is observed. When extracellular glucose levels are lowered from a high to an intermediate level, nuclear Mig1 temporarily re-localizes to the cytosol during a brief period. This pattern is qualitatively present in all cells, but the quantitative details differ. As for many signaling pathways the exact mechanisms controlling Mig1 dynamics are unfortunately not yet fully understood, and therefore a simple phenomenological kinetic model was set up to describe the observations. The population distribution of three parameters was estimated with the algorithm presented in Paper E using time series data from almost 200 single cells, collected from four different experiments (Experiment 1–4). These data had previously been obtained from a Mig1-GFP expressing strain of *S. cerevisiae* using an experimental setup that combined microfluidics, optical tweezers, fluorescence microscopy, and image processing (Bendrioua et al. 2014).

Paper D demonstrates reasonable precision of the parameter estimates and show a good fit of the model to the time series data at the single cell level. To complement this perspective, Fig. 3.6A shows model simulations together with data at the population level for Experiment 1. The simulated 10th, 50th, and 90th percentiles of the nuclear Mig1 concentration according to the model are plotted together with all data as well as the corresponding empirical percentiles. There is generally a good agreement between model and data except for some underprediction of the lowest percentile around 200 s after the extracellular glucose shift. Of the in total 741 data points of Experiment 1, 57, 398, and 653 data points fall below the three simulated percentiles, which correspond to percentages of 8, 54, and 88.

Estimation of the population distribution of parameter values was also done according to the simpler standard two-stage (STS) approach (Ette and Williams 2004) in Paper D. This approach works by estimating parameters for each cell independently, and then looking at the statistics of those estimates. Estimates derived with the STS approach showed a much larger variability, which is a clear indication of overestimation (Sheiner and Beal 1981). To further demonstrate the difference between the NLME and STS approaches, Fig. 3.6B compare model simulations together with data at the population level for Experiment 4. This time, the simulated 10th, 50th, and 90th percentiles are shown for the NLME model and for two variants of the STS model: one that is naïvely based on all single cell parameter values and one that excludes suspected outliers. The details of these approaches are stated in Paper D. The fit of the NLME model is acceptable, but

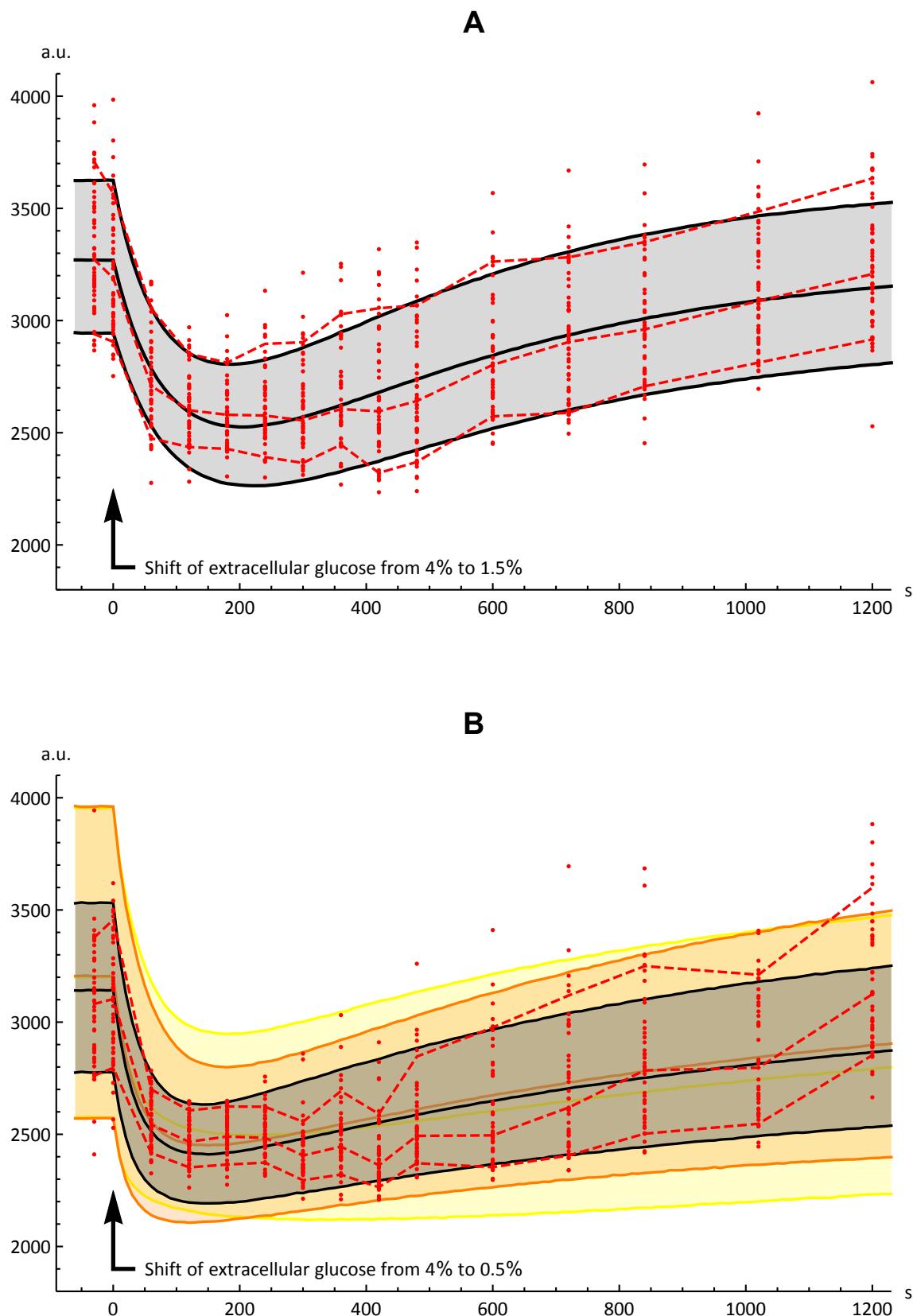


Figure 3.6: Visual predictive check.

Figure 3.6: Visual predictive check. **(A)** Model and data for Experiment 1, in which extracellular glucose is shifted from 4% to 1.5% at time zero. Simulated 10th, 50th, and 90th percentiles of the nuclear Mig1 concentration are plotted as black lines together with grey filling of the interval between the 10th and 90th percentiles. Data are plotted as red dots, and the 10th, 50th, and 90th percentiles of the data are plotted as red dashed lines. **(B)** Model and data for Experiment 4, in which extracellular glucose is shifted from 4% to 0.5% at time zero. The plot is constructed in the same way as (A), but with the addition of model simulations using the STS approach excluding outliers (orange) and the naïve STS approach (yellow).

still showing a bit of underestimation of the lowest percentile around 200 s, as well as an underestimation of all percentiles at the last time point. Percentages of data points falling below the three simulated percentiles were 6, 50, and 90. The STS approach excluding outliers clearly shows a general overestimation of the cell-to-cell variability of the population. This is particularly evident during the first 500 s of the experiment. The naïve STS approach display even more overestimation. Taken together, it can be concluded that the NLME approach should be preferred.

Paper D contributes new insights into the understanding of Mig1 signaling in *S. cerevisiae* by providing a cell-to-cell variability perspective. This kind of modeling is made possible by the access to single cell data (Bendrioua et al. 2014). However, the biological interpretations and conclusions of the current work unfortunately become limited by the use of a phenomenological model. A mechanistic model would have been preferred but a mechanistic hypothesis that accommodates the observed transient re-localization is still missing (García-Salcedo et al. 2014). Uncertainty of the wiring scheme in microbial signaling pathways is common, and it is a general problem for kinetic modeling (Schaber and Klipp 2011). Still, some biological interpretations of the modeling results could be made. It was for instance speculated that the transient pattern of Mig1 localization probably is already present in the dynamics of an upstream signaling component. The protein kinase Snf1 was pointed out as a particularly interesting candidate for carrying such a signal. With the availability of novel experimental methods like synthetic kinase activity relocation sensors (Durandau et al. 2015), the temporal phosphorylation pattern of Snf1 in live single cells may therefore be a possible focus for a future study.

Another merit of this work is the general idea of applying NLME modeling to single cell data. To the knowledge of the authors, Paper D represents the first journal article demonstrating this idea.² It will hopefully encourage other researchers to continue along this direction. Indeed, it appears to be a timely topic—two other articles investigating cell-to-cell variability using NLME modeling appeared shortly after the publication of Paper D (Karlsson et al. 2015; Llamosi et al. 2016).

The model of Paper D mainly serves to describe and quantify the degree of cell-to-cell variability in Mig1 signaling, but it does not provide an explanation for why this variability is present. Often, it is argued that cell-to-cell variability should be seen as a way of implementing multiple phenotypes from a single genotype, thereby achieving a risk-spreading strategy with evolutionary benefits (Veening et al.

²See Paper D for more details and references to previous contributions to single cell modeling.

2008; Lidstrom and Konopka 2010). It may also be interesting to speculate about whether cell-to-cell variability presents a limitation or a possibility when it comes to applied uses of, for instance, yeast, outside of its natural environment. As a model organism, yeast cell-to-cell variability may be potentially deceiving since in the cell being modeled, perhaps a human cell, cell-to-cell variability may play a different role. In the context of cell factories, cell-to-cell variability may have undesirable effects on the performance of a bioprocess (Delvigne et al. 2014), for instance if the total production of a certain product is due to a smaller fraction of high producing cells while most cells only consume resources (Lidstrom and Konopka 2010). On the other hand, cell-to-cell variability may also be exploited to enhance the performance of a cell factory (Xiao et al. 2016). Given the importance of yeasts and other microbes as both model organisms and as cell factories, kinetic models of cell-to-cell variability may therefore be expected to play a bigger role in the future.

3.6 Exact Gradients for Nonlinear Mixed Effects Models

Data collected from different individuals is encountered in both the preclinical and clinical phases of drug development. NLME modeling has emerged as a de facto standard of how these data are being analyzed and interpreted (Bonate 2011). NLME models have also been used successfully in a number of other scientific fields apart from pharmacology (Davidian and Giltinan 2003). Yet another example is the study of yeast cell-to-cell variability presented in Paper D of this thesis.

One of the most challenging aspects of NLME modeling is the estimation of model parameter values from experimental data, a task for which several computer programs exists (Beal et al. 2017; Certara 2011–2017; Lavielle 2014a). Within the maximum likelihood methods, which are dominating, two main approaches can be distinguished: the classical methods based on the Laplacian approximation (Bauer et al. 2007; Wang 2007), including the first-order conditional estimation (FOCE) method; and the new generation of methods based on various Monte Carlo techniques (Bauer et al. 2007; Leary et al. 2012; Lavielle 2014b). Although both approaches involve approximations of the population likelihood, the Monte Carlo methods have the advantage that the approximation can be made arbitrarily precise by increasing the number of samples (Leary et al. 2012), whereas methods like FOCE always are running the risk of producing more or less biased estimates. In practice, however, FOCE often yield “good enough” results, and can be faster than the Monte Carlo methods for some problems. Some Monte Carlo methods are also quite sensitive to different method-parameters, and there can sometimes be ambiguities as to whether the algorithm has converged or not. As a result of this, but also because of historical popularity, FOCE and the closely related FOCEI are still among the most commonly used methods for estimating parameter in NLME models.

Parameter estimation for NLME models with the FOCE approximation of the population likelihood is normally done using a standard gradient-based Quasi-Newton optimization method such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method (Nocedal and SJ Wright 1999). What makes this an extra challenging optimization problem is the complexity of the objective function, which arises from the way that the

FOCE method approximates the population likelihood. For every evaluation of the FOCE likelihood, one new optimization problem has to be solved for each individual that is part of the data set. In this way, the FOCE method leads to an optimization problem with two nested layers. Paper E proposes a new way of performing NLME parameter estimation based on the FOCE method. The main novelty lies in the way that the gradient of the FOCE likelihood is computed. As explained above, the gradient is critical for solving the maximum likelihood optimization problem. Although the central ideas of this work are applicable to any kind of NLME model, the advantage of the method is mainly expected to exert itself for models based on ODEs.

The FOCE approximation and how its gradient can be computed are illustrated in Fig. 3.7, and briefly explained in the following. The goal of all NLME maximum likelihood approaches is to find values of the fixed effect parameters $\boldsymbol{\theta}$ that maximize the population (log-)likelihood. As described in detail in Paper E, the joint individual likelihoods (Fig. 3.7A) have to be marginalized in the random effect dimension(s) to obtain their contributions to the population likelihood (Fig. 3.7B). During optimization of the population likelihood with respect to the fixed effect parameters $\boldsymbol{\theta}$, this marginalization is required for all individual likelihoods, and it will be repeated many times for different values of $\boldsymbol{\theta}$. For a given $\boldsymbol{\theta}$, the FOCE method approximates the population likelihood in the following way. First, the random effect parameters $\boldsymbol{\eta}_i^*$ that maximize the individual log-likelihoods l_i have to be determined for each individual (Fig. 3.7C). This constitutes the inner level nested optimization problem mentioned previously. Then, l_i is approximated with a second-order Taylor expansion around $\boldsymbol{\eta}_i^*$ (Fig. 3.7D), which allows a closed-form solution to the marginalization of the individual joint likelihood. The gradient-based optimization of the FOCE likelihood has traditionally relied on gradients computed by a finite difference (FD) approximation (Fig. 3.7E). The main result of Paper E is the derivation of how exact gradients can be computed (Fig. 3.7F). The details of this approach are mathematically involved, but a key step is the computation of the matrix $d\boldsymbol{\eta}_i^*(\boldsymbol{\theta})/d\boldsymbol{\theta}$, i.e., the sensitivity of the point for the Taylor expansion with respect to $\boldsymbol{\theta}$. This is achieved through an elaborate use of up to second-order state variable sensitivity equations in both the fixed and random effect parameters.

The new algorithm in Paper E was shown to have two advantages over existing versions: (i) it is faster; and (ii) it increases the precision and accuracy of the gradient. The exact speed-up of the parameter estimation depends on the type and complexity of the model, if exact gradients are considered for both levels of optimization, and whether it is compared to forward or central FDs. In the most favorable scenario, the time spent on estimation was reduced by a factor one hundred. As discussed in Paper E, a realistic estimate of the speed improvement of the new algorithm compared to previously existing FOCE implementations would perhaps rather be about three to five times depending on the model, but it is hard to say with certainty. The use of FDs for computing gradients may suffer from bias if the step size is too large, or suffer from severe noise if it is too small. Both of these effects are undesirable and may corrupt parameter estimation to various degrees. When using the exact method based on sensitivity equations, both accuracy and precision of the gradient

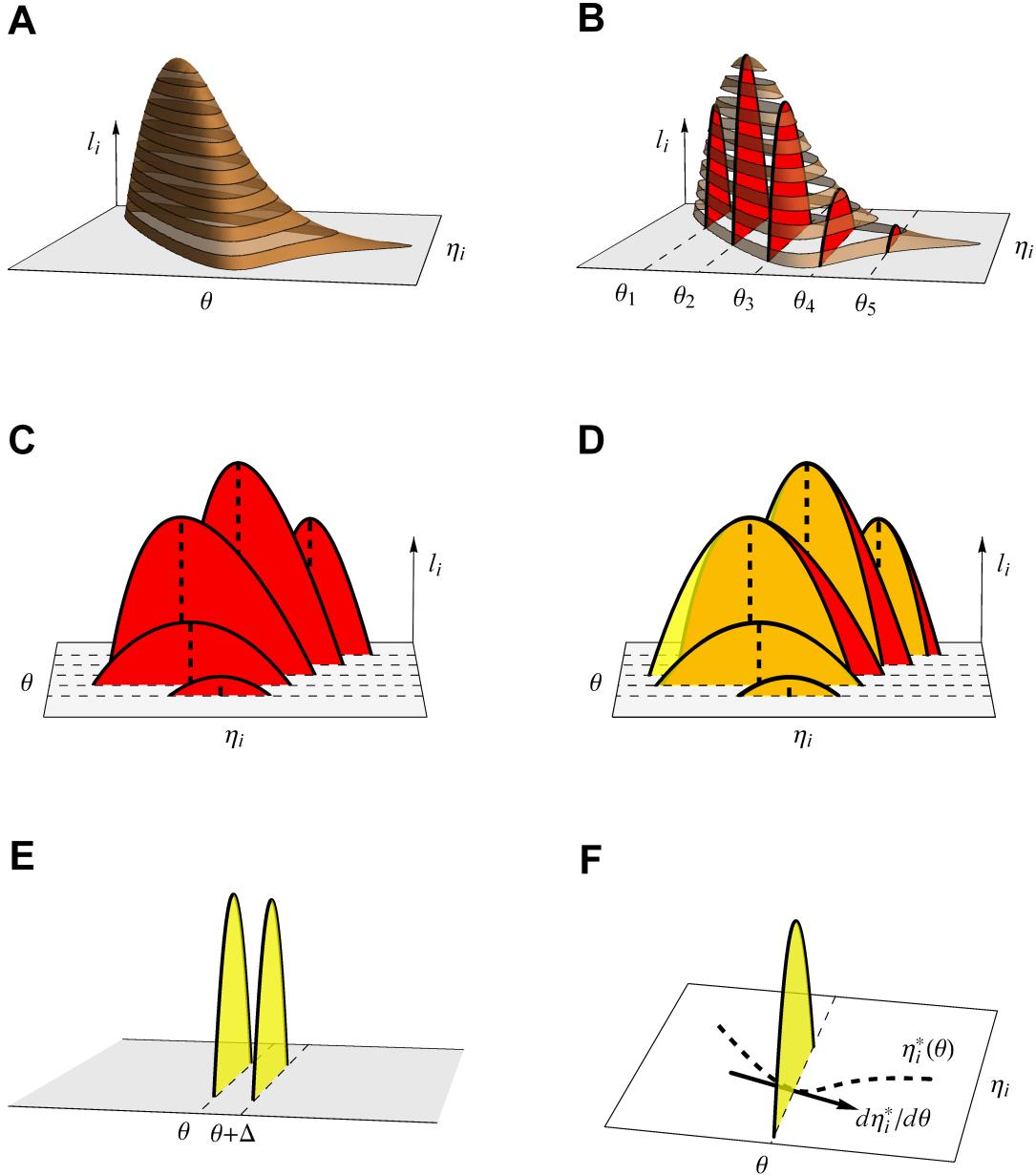


Figure 3.7: Approximating the population likelihood and its gradient.

was substantially improved.

The combination of challenging estimation scenarios that verge on practical non-identifiability with the nested optimizations of the FOCE method sometimes result in a lack of numerical robustness. This is manifested through convergence failure and problems with computing standard errors or confidence intervals for the parameter estimates (Bauer et al. 2007; Dartois et al. 2007; Chan et al. 2011; Bertrand et al. 2011; Aoki et al. 2016; Liu and Wang 2016), issues which are most apparent for complex models requiring numerical integration of ODEs. Part of the problem with lack of robustness may be attributed to the use of FDs for gradient computations. It therefore seems plausible that access to a gradient with higher

Figure 3.7: Approximating the population likelihood and its gradient. **(A)** Illustration of an individual joint log-likelihood, l_i , which is a function of both fixed effects, θ , and random effects, η_i . For simplicity, these two types of parameters are illustrated using only one dimension each. Different levels of the log-likelihood are shown with alternating solid and transparent brown bands. **(B)** The contribution of each l_i to the population (log)-likelihood is given by integrating out the random effect dimension (strictly speaking, the marginalization is performed on the individual likelihoods, not the log-likelihoods, but the cross-sections of the log-likelihood are shown for illustrative purposes). Such cross-sectional slices of the log-likelihood are shown in red for five different values of θ . These θ -values are shown as dashed lines across the η_i dimension. **(C)** Given a value of θ , the first step of the FOCE approximation of the population log-likelihood is to determine the η_i that maximizes l_i , denoted η_i^* in Paper E. The maxima for the different θ -values of (B) are shown as dashed black vertical lines. **(D)** The second step of the FOCE approximation is to perform a second-order Taylor expansion of l_i around η_i^* . The FOCE approximations of the l_i cross-sections are shown in yellow, on top of the actual cross-sections which still are shown in red. **(E)** The traditional way of computing gradients of the FOCE approximation of l_i is based on a FD approximation in which l_i is evaluated at θ and at $\theta + \Delta$. **(F)** The method presented in Paper E computes an exact gradient of the FOCE approximation of l_i . A key step for doing this is the computation of $d\eta_i^*(\theta)/d\theta$, i.e., the sensitivity of the point for the Taylor expansion with respect to θ .

accuracy and precision should overcome this issue. However, it remains to be proven that better accuracy and precision of gradients actually translate into increased robustness of the parameter estimation problem as a whole.

The majority of the algorithm proposed in Paper E was recently implemented in the 7.4 version of the computer program NONMEM. This program is probably the most widely used software for NLME modeling in population PKPD. According to the NONMEM developers, this has increased the speed of both parameter estimation and of the so-called covariance step by up to 3–4 fold for ODE-based models (Beal et al. 2017).

3.7 Nonlinear Mixed Effects Modeling with Stochastic Dynamics

In Section 1.2 it was explained how models are simplifications of the real world. These simplifications are a deliberate part of the modeling process. Since simplifications means that some aspects of reality have been omitted, all models could be considered to be more or less wrong. Modelers may furthermore base their models on incorrect prior knowledge, or make unrealistic assumptions. In contrast to the deliberate simplifications, this introduces undesirable discrepancies between model and reality. Modelers, of course, do their best to avoid this, but model errors of this sort are inevitable. Taken together, it is safe to conclude that models are never perfect descriptions of the corresponding (biological) systems. It can also be concluded that the degree to which a model undesirably deviates from a perfect description is typically uncertain (if it was certain the modeler would be able to correct for it).

The dynamics of the standard kinetic model is completely deterministic. The only

uncertain part lies in the equation for the observations. To reflect that not only the observations are uncertain, but also the kinetics encoded in the differential equations, additional uncertainty can be introduced. One approach is to formulate the kinetic model in terms of SDEs instead of ODEs. This does not necessarily mean that the kinetics are believed to be inherently stochastic, but rather that SDEs present a feasible approach to model the uncertainty about which deterministic model is the correct one.

Stochastic kinetics can replace deterministic kinetics in both the standard model and in the NLME model. A summary of the original and the resulting mathematical frameworks is shown in Fig. 3.8. Here, the frameworks are categorized according to whether parameter hierarchies are present or not, and according to the use of either deterministic or stochastic dynamics. This defines the four combinations that were previously introduced in Chapter 2: (i) the standard kinetic model; (ii) the NLME model; (iii) the SDE model; and (iv) the SDE-NLME model. The equations in Fig. 3.8 are arranged to emphasize the mathematical differences between the frameworks, making it clear which equations are added, removed, or altered when transitioning from one framework to another. For instance, the hierarchical parameterization of the NLME frameworks means that both state variables and parameters become individualized, as indicated by the index i . Fig. 3.8 also shows the distinction between the parts of the model kinetics that are deterministic and stochastic, respectively. As suggested in Fig. 3.8, the SDE-NLME framework can be seen either as expanding an NLME population model by adding stochastic dynamics, or as a model with stochastic dynamics that is expanded by adopting hierarchical population parameters. Regardless, the end result is a model that distinguishes three sources of variability or uncertainty for the observed data. These sources are variability in the parameters, variability or uncertainty in the model kinetics, and variability originating from making observations.

Paper F explores the use of SDE-NLME modeling for PK applications. It combines the FOCE method for approximation of the population likelihood (Wang 2007) with the extended Kalman filter (EKF) for state variable estimation (Jazwinski 1970). This approach was pioneered by Henrik Madsen and coworkers for similar applications, see for instance (Tornøe et al. 2004; Overgaard et al. 2005; Kristensen et al. 2005) and references within Paper F. Two case studies were presented in Paper F. First, an SDE-NLME one-compartment model with nonlinear elimination and an absorption compartment was used to simulate synthetic PK data following an oral bolus dose. The same model was subsequently used for re-estimation of its parameters. Based on a simulation-estimation study comprising 100 data sets, the three contributions to the total variability outlined above were shown to be identifiable. Moreover, parameter estimates in the SDE-NLME model had lower bias and higher precision compared to estimates from the corresponding ODE-NLME model. These results are encouraging, but it is a limitation that the synthetic data were generated with the exact same model that was used for re-estimation. It would also have been interesting to see whether similar results could have been obtained if the synthetic data were generated from a slightly different deterministic model. Second, an SDE-NLME PK model describing intravenous infusion of nicotinic acid

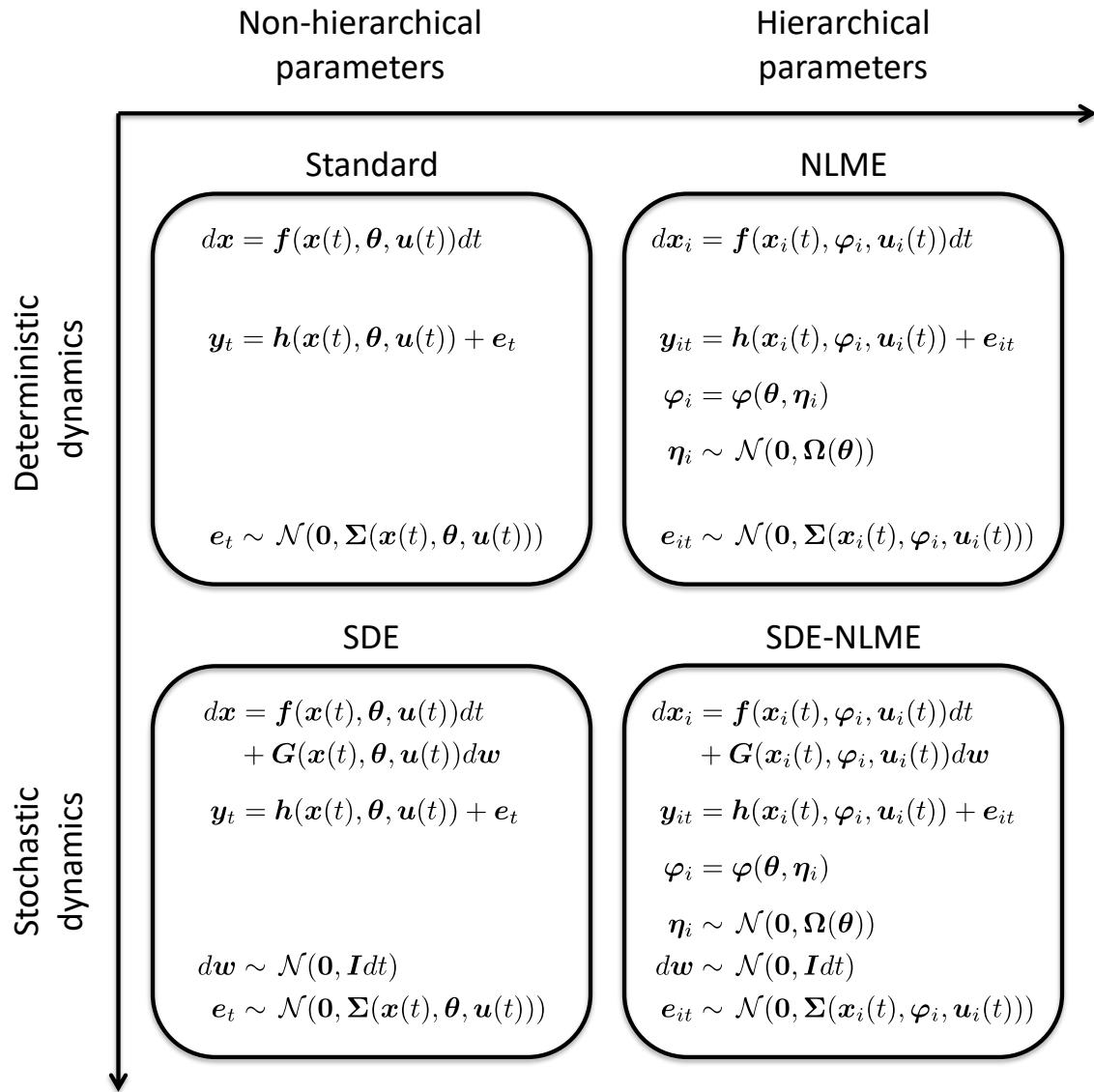


Figure 3.8: Mathematical modeling frameworks for kinetic models. The horizontal dimension of the illustration categorizes models based on whether parameters are hierarchical or not, and the vertical dimension categorizes model dynamics into deterministic or stochastic. Each of the four boxes contains model equations for the corresponding modeling framework. These equations are arranged to emphasize the similarities and differences between the frameworks. Initial conditions have been omitted for the sake of simplicity.

was applied to real data from obese Zucker rats. The SDE-NLME model gives different parameter estimates compared to the corresponding ODE-NLME model. For instance, a smaller value of the population variability in the upper capacity of the clearance, V_m , was estimated with the SDE-NLME model, suggesting that the ODE-NLME model may be overestimating the variability. This can however not be verified since the true model is not known. The second case study also involved the computation of a smoothed estimate of the observation distributions. This may be particularly useful for driving a PD model in a sequential PKPD analysis.

The SDE-NLME parameter estimation algorithm used in Paper F works in the same way as the algorithm for ODE-NLME described in Paper E, but also includes first- and second-order sensitivities of the EKF. The algorithm was implemented in Mathematica (Wolfram Research, Inc. 2012) and uses symbolic computation to derive the necessary sensitivity equations for each model. It was only later, in a master’s thesis supervised by the author of this thesis (Ólafsdóttir 2016), that a complete set of equations for the general case was derived and collected all in one place. Additional benchmarking of the SDE-NLME estimation algorithm has also recently been published (Ólafsdóttir et al. 2017).

The SDE-NLME modeling framework in Paper F may be used for other modeling applications where population variability, uncertain dynamics, and uncertain measurements need to be accommodated. One candidate area for the approach are tumor growth models (Cardilin et al. 2017a; Cardilin et al. 2017b). The ability to account for model imperfection may be extra useful here, since these models feature exponential or near-exponential growth dynamics. The reason is that exponential behavior tends to amplify any model errors, making tumor growth models particularly sensitive. Kinetic models based on SDE-NLME may also be suitable for modeling biological systems where reactions are best understood as truly stochastic. One such application could be the simultaneous modeling of intrinsic and extrinsic noise of the biochemistry at the single cell level, as discussed in Paper D. In this case, the stochastic dynamics would not be used as a “modeling trick” to accommodate uncertainty in how to formulate the reaction kinetics, but to actually represent (apparent) randomness due to low copy-numbers, etc. Yet another application involves using SDEs to regularize the likelihood function, so that local optimization methods have a better chance of finding the global optimum. This has been demonstrated for a non-hierarchical SDE model (Leander et al. 2014), but it is expected to work also in an SDE-NLME setting.

CHAPTER 4

Discussion and Conclusions

The aim of this thesis was to identify and describe the different steps in the workflow for building and using a kinetic model, apply the kinetic modeling approach to address some relevant questions within life science, and to contribute to the development of new methods for building kinetic models. Outlining the kinetic modeling process can be seen as the general overarching aim, whereas the two other aims should provide specific examples of methods and applications.

The workflow for kinetic modeling was identified and described in Paper A. The main steps of this process are summarized in Fig. 3.1. Although the review in Paper A was mainly underpinned by literature on kinetic modeling in biotechnology, the identified workflow is still relevant for kinetic modeling in general.

Three new kinetic models were developed as part of this thesis. In addition to serving as examples of the kinetic modeling workflow, these three models are valuable contributions in their own right. Paper B presented a PK model of the drug-antidote interaction between ticagrelor and MEDI2452. This model could explain counter-intuitive experimental results and make predictions that contributed to the drug development process. Paper C explored the impact of Kv1.5 potassium ion channel block on the canine atrial AP using a model-based approach. The modeling created a link between the kinetics of Kv1.5 ion channel block and the APD. Paper D proposed a descriptive model of cell-to-cell variability of transient glucose sensing in yeast. Even though biological conclusions of a mechanistic nature were limited, this work nonetheless demonstrated the potential of NLME modeling for single cell data.

Development of useful kinetic models requires that appropriate model-building methods are available. This thesis has contributed two studies on modeling methods. Paper E presented a novel method for parameter estimation in NLME models. The sensitivity-based exact gradients of this approach represent a substantial improvement to the already established FOCE method. Paper F explained and illustrated a kinetic modeling framework that combines SDEs with the NLME approach for population modeling. This is a promising framework that extends the current scope of the

population approach, but more examples of its application are required to fully assess its merits.

4.1 Kinetic Models Fit Many Life Science Questions

The combination of mass balance equations and reaction kinetics makes the kinetic modeling approach applicable to widely different questions throughout all of the life sciences. In addition to examples from Chapter 1 and the many examples in Paper A, this is also shown by the three models introduced in the thesis. These models span different scales of biological organization and size: from the different molecular states of ion channels in Paper C, to translocation of proteins at the organelle level in Paper D, to the simplified whole-body description of the mouse in Paper B. The models also cover very different time scales: from milliseconds in Paper C, to minutes in Paper D, and hours in Paper B. The diversity is furthermore reflected in the size of the models in terms of the number of equations and parameters, ranging from a fairly small model in Paper D, to a medium-sized model in Paper B, and a medium- to large-sized model in Paper C (here counting both the RNC model and the integrated Kv1.5 model).

4.2 The Question Guides the Framework

A successful modeling effort requires a mathematical framework that is suitable for the modeling question. As outlined in Chapter 1 and 2, and as shown in Fig. 3.8, there are many different mathematical frameworks that can be adopted for the kinetic modeling approach, and this is partly why it fits so many questions. For instance, the main driver behind Paper D was the desire to characterize cell-to-cell variability. Given the relatively sparse and noisy data, and the assumption of negligible intrinsic noise, an NLME modeling approach was advocated. The comparative analysis in Paper D, and the additional results presented in this thesis (Fig. 3.6), shows that the choice of this framework was indeed required to avoid overestimating the variability. Paper B, on the other hand, did not use a population approach even though data were collected from different animals; the construction of an NLME model is much more demanding compared to a naïve-pooled approach using a standard kinetic model, and the advantage of capturing between-individual variability was not considered sufficiently important given the modeling question. In Paper C, a kinetic model was used to describe transitions between ion channel states. Since the purpose of the model was to describe the whole pool of all Kv1.5 ion channels, the deterministic standard model was sufficient. However, if the goal would have been to model the transitions within a single ion channel, a stochastic model, such as a continuous-time Markov chain model, would have been more appropriate.

4.3 The Model Structure Is the Heart of the Model

The model structure should contain all relevant components of the modeled system, such as compartments, biochemical molecules, and reactions. It is the heart of the model, and both biological knowledge and modeling craftsmanship are required for its design. As new insights are gathered during the modeling process, it should be expected that the model structure has to be iteratively revised before taking on its final shape. To complement the mathematical definition of the model structure it is highly recommended to also provide a visual illustration, see for instance Fig. 3.2. This will quickly convey an overview of the model and will facilitate the communication of the model to others—both when forming the model and when interpreting the findings.

4.3.1 Be Aware of Uncertainty

It is not unusual for prior knowledge of the modeled system to be limited or contradictory. This leads to uncertainty about how to formulate the model structure (Brown et al. 2004; Kuepfer et al. 2007; Schaber and Klipp 2011), a topic that is underappreciated compared to uncertainty about parameter values. In some cases, structural uncertainty can be rephrased as a parameter estimation problem (Chou and Voit 2009; Srinath and Gunawan 2010). A similar scenario is exemplified in Paper B, where uncertainty regarding ticagrelor recycling was expressed with a parameter for the degree of recycling. This analysis was however limited to the investigation of a few discrete values of the recycling parameter, and contained no formal parameter estimation. In Paper F, a fundamentally different approach to uncertainty about the model structure was proposed. This approach accepts that errors in the model structure are inevitable, and instead focus on how to incorporate this as uncertainty in the mass balance equations using SDEs. Paper F showed that the total variability in data could be decomposed into inter-individual variability in the parameters, variability or uncertainty in the model kinetics, and variability originating from making observations. More applied examples of this modeling framework are however needed to fully assess its capabilities, ideally focusing further on its role in iteratively improving the deterministic part of models (Kristensen et al. 2005; Matzuka et al. 2016). Nevertheless, SDE-NLME modeling is a promising approach that could become the next generation of population modeling. If that were to happen, it remains to be seen whether the core mathematics will be based on conventional filters, like the EKF in Paper F, or on more general methods like particle filters (Doucet et al. 2001).

4.3.2 Mechanistic Models Are Preferred

Mechanistic models seek to incorporate specific biochemical, biophysical, or physiological knowledge, whereas phenomenological models settle for mathematical descriptions of the data that do not have to “mean” anything. Many models contain a mixture of both mechanistic and phenomenological components. The Kv1.5 model in Paper C

is, relatively speaking, a very mechanistic kinetic model. The model state variables have a clear connection to the molecular states of the ion channel, and the reactions describing transitions between states are modeled with both temperature and voltage dependencies in a biophysically rigorous way. In contrast, the model of Mig1 re-localization in Paper D is a phenomenological model that was set up due to the lack of a mechanistic hypothesis. This minimal model of perfect adaptation is only used to quantify the variability of the observed data, rather than trying to describe some molecular mechanism that is causing the observed phenomenon. The model of Paper B lies somewhere in between, but leans more toward the mechanistic side; there are molecular-level mechanistic interactions similar to the ideas of systems pharmacology models (Graaf and Benson 2011), but also the classical PK compartments, which can be seen as semi-mechanistic (Aarons 2005).¹ Good phenomenological models adequately describe data, and are predictive in an interpolating sense. This may be sufficient for some modeling questions, but can be a significant limitation for others. Phenomenological models do not usually provide any profound understanding of the real world and cannot be expected to give reliable predictions beyond the data that they are fitted to. Therefore, they are sometimes dismissed as simple “curve fitting”. Mechanistic models set higher demands on prior knowledge, but carries the potential for providing mechanism-based understanding, as well as a generally increased confidence in predictions of extrapolation-type. These are strong arguments in favor of mechanistic kinetic models. From the kinetic modeling examples in Paper B, C, and D, and from the many models reviewed in Paper A, it can be concluded that mechanistic models should in general be preferred over phenomenological models, but only if supported by sufficient prior knowledge and available data.

4.3.3 Avoid Unnecessary Complexity

The modeling approach to answering questions is to always look for opportunities to simplify and reduce unnecessary complexity. For instance, Paper B utilized a simplification from two- to one-compartment kinetics of MEDI2452, an approximation that was found to be valid for the time scale of interest. In Paper C, assumptions on the kinetics of ion channel subunits together with symmetry arguments resulted in lumping of molecular states when the state variables of the kinetic model were defined (see Fig. 3.5). Paper D involved a more technically driven model reduction, where a change of variables was performed to avoid a structurally unidentifiable model (Anguelova et al. 2012). What can be considered unnecessary complexity depends on the context, and will differ both between specific models and between different modeling paradigms. Kinetic models frequently have issues with determining both model structure and parameter values, and convincing arguments for model reduction are usually welcome. Often, “less is more” is applicable to kinetic models. This can be compared to so-called genome-scale metabolic models (GEMs), which are essentially stoichiometric models (i.e., steady state mass balance equations) that lack kinetic rate expressions and kinetic parameters (Förster et al. 2003). Due to their simpler structure, it is less complicated to build large models. In fact, for GEMs,

¹Paper B also featured a simple PD model that was purely phenomenological.

all known reactions of the cell are candidates to appear in the model, since more reactions mean a more complete—and therefore typically better—model. This comparison is not to say that large kinetic models cannot or should not be built, but that when it is done, it must be done carefully and only if the modeling question requires it.

4.4 Parameter Estimation Is a Critical Step

Determining values of the model parameters by collecting them from different external sources (e.g., literature, databases) has certain limitations, as discussed in Paper A. Parameter values should therefore ideally be estimated simultaneously from one or more data sets that belong together. Such collective estimation was the predominant approach in both the application- and method-oriented contributions to this thesis. This requires a balance between the model complexity and the information content of experimental data (Raue et al. 2011; Anguelova et al. 2012), a balance that tends to be hard to achieve for complex kinetic models with lots of parameters. To avoid the worst pitfalls, modelers should pay attention to identifiability analysis, model reduction, and experimental design, as reviewed in Paper A. Adding further to the problem of parameter estimation, it appears that many kinetic models are only sensitive to a rather small subset of parameter combinations, which will prevent precise estimates of many parameters when using collective fits (Brown et al. 2004; Gutenkunst et al. 2007; Daniels et al. 2008). This problem is unlikely to be resolved by direct parameter measurements, since the few sensitive combinations would require an unreasonable high precision in such measurements (Gutenkunst et al. 2007). This implies that modelers should critically think about whether high precision estimates are really needed to answer the model question. As shown in (Brown et al. 2004), there is not necessarily a conflict between precise predictions of model behavior and highly unprecise estimates of some model parameters. To some extent, this conclusion also applies to the parameter uncertainty analysis in Paper B. If high precision estimates are still essential, this may be solved by further improving the experimental design (Apgar et al. 2010; Nyberg et al. 2015).

Parameter estimation is often a difficult problem from a computational point of view (Mendes and Kell 1998; Chou and Voit 2009; Raue et al. 2013). A particularly challenging example is the maximum likelihood inference of parameters in NLME models based on differential equations. This is partly due to the high cost of evaluating the objective function, i.e., the population likelihood or an approximation thereof. If a gradient-based optimization method is used where gradients are computed from FDs, which has traditionally been common, this problem is amplified. The exact gradient algorithm proposed in Paper E makes FDs obsolete and thereby represents a significant advance of the FOCE method. The key step of this method may be attributed to the insight on how to compute the sensitivity of the solution to the inner level optimization problem with respect to the population-level fixed effect parameters (see Fig. 3.7). The benchmarking in Paper E and internal comparisons in the commercial software NONMEM (where the Paper E algorithm now is available through the FAST option (Beal et al. 2017)) both demonstrated a substantial speed-up

of the parameter estimation. As NLME modeling is a widely used approach with a secure position in the realm of kinetic modeling frameworks, this is an important achievement. The method of Paper E also yielded gradients with higher accuracy and precision. Although it appears plausible that this should improve estimate precision and algorithm robustness (Raue et al. 2013), further research remains to prove it. It is also worthwhile to note that this method could be generalized to also handle the estimation problem in SDE-NLME models (although this was not the focus of Paper F). Because of the added computational burden in the SDE-NLME framework, the speed-up from the exact-gradient method is perhaps even more valuable here. Without diminishing the results of Paper E, it should also be pointed out that the FOCE method as such is facing serious competition from the new generations of NLME estimation methods that build on various kind of Monte Carlo approaches (Bauer et al. 2007; Leary et al. 2012; Lavielle 2014b), and that these may become the standard tools of the future.

4.5 In Silico Experiments Are the Reward

A remarkable strength of kinetic modeling—or rather of mathematical modeling in general—is that in silico experiments can be performed once a model has been setup. In contrast to real experiments in the wet lab, in silico experiments can be performed more or less for free and without the fear of not being able to reproduce the results. This is one of the rewards for going through all the trouble of setting up a model. The three applied models introduced in this thesis all took advantage of in silico experiments in one way or another: Paper B simulated free levels of ticagrelor and investigated the impact of potential ticagrelor recycling during clearance of the ticagrelor-MEDI2452 complex, Paper C used in silico experiments exhaustively to characterize the dependence of the AP on kinetic parameters of real and hypothetical drugs, and Paper D performed Monte Carlo simulations that mapped the parameter variability of the yeast cell population to the behavior variability with respect to signaling characteristics such as response time, amplitude, and duration. It was at this stage of the kinetic modeling process, when the models were actually used for answering questions, that interesting findings could be made.

4.6 Answering the Model Question

Based on the different in silico experiments performed with three models introduced in this thesis, and on the identification of the models as such, some answers could be provided to the questions that initiated the modeling. Here, the key findings and conclusions are briefly stated.

4.6.1 Pharmacokinetics of Drug-Antidote Interaction

Paper B used a kinetic model to unravel the PK interaction between ticagrelor and MEDI2452 in the mouse. Of the three models, this was probably the most successful one in terms of answering both the original modeling question and additional ones that appeared during the modeling process. The model predicted the plasma concentration of free ticagrelor, it explained the behavior where free and total plasma ticagrelor showed opposite response, and it suggested that ticagrelor is not likely to be recycled when ticagrelor-bound MEDI2452 is eliminated in the kidneys. Furthermore, it was shown that a special model for the equilibration of *in vivo* samples is essential for distinguishing actual ticagrelor plasma concentration from measured ticagrelor plasma concentration (Fig. 3.3). Looking beyond the particular drug-antidote pair in this study, the modeling approach has great potential for similar scenarios.

4.6.2 Effect of Kv1.5 Blockers on the Canine Atrial Action Potential

Paper C studied the impact of Kv1.5 block on the canine atrial AP using kinetic modeling. Detailed kinetic models of open-channel block were found to give large differences in prolongation of the APD compared to a model with state-independent block. These simulation results recommend against the common but naïve modeling approach of simply decreasing the ion channel conductance, and instead advise modelers to choose an ion channel model that matches the actual blocking mechanism. It was also found that open-channel block produced a reverse use-dependence, i.e., that the increase in APD becomes smaller for higher frequencies of AP stimulation. Thorough understanding of Kv1.5 blocking kinetics is clearly important in order to accurately predict the AP following a pharmacological intervention (Fig. 3.4). However, to reach the even bigger goal of directly predicting the effects on AF, additional model layers are needed that couple the cellular level AP to the electrophysiology at the tissue or organ level.

4.6.3 Cell-to-Cell Variability of Transient Glucose Sensing in Yeast

Paper D applied NLME modeling to quantify the cell-to-cell variability of Mig1 signaling dynamics in yeast. Based on data from almost 200 cells, population distributions of parameters were estimated as well as parameter values for individual cells. Simulations were used to compute the population distribution of signaling characteristics like response time, amplitude, and duration. It was also suggested that the single cell dynamics of Snf1 phosphorylation might be crucial for understanding cell-to-cell variability in Mig1 localization patterns. Unfortunately, the phenomenological structure of the model severely limited conclusions regarding the exact mechanisms. Future work on single cell analysis using NLME modeling should therefore aim at using mechanistic kinetic models. Moreover, the comparison of modeling frameworks demonstrated that NLME modeling may be essential for single cell data. This is a valuable result in itself, since similar modeling scenarios are expected to become common.

4.7 A Brief Outlook

Parsimony has traditionally been a virtue of kinetic models in life science. This is now changing with the holistic view of the systems-paradigms, e.g., systems biology (Kitano 2002), systems medicine (Wolkenhauer et al. 2013), and systems pharmacology (Graaf and Benson 2011), and with large-scale or multi-scale modeling efforts in general (Hunter and Borg 2003; Holzhütter et al. 2012; Büchel et al. 2013). For instance, the k-ecoli457 kinetic model of *Escherichia coli* metabolism contains 457 reactions and 337 metabolites, and is perhaps one of the first models to deserve the epithet genome-scale kinetic model (Khodayari and Maranas 2016).² The sheer size of models like this pushes the limits of the technical aspects of model-building, making identifiability analysis, experimental design, and parameter estimation very challenging. But maybe even more important, many of these large-scale models come with a different modeling philosophy. Like the previously discussed GEMs, some of the models are constructed without a specific research question in mind. They are instead intended to serve as multi-purpose models for answering a whole range of questions that may arise in the future. Such kinetic platform models are of interest to, for instance, biotechnology, where the metabolism of a few selected microorganisms (platform cell factories) is repeatedly reengineered for the production of different fuels and chemicals, and in drug development, where disease-scale models are desired for evaluating different pharmacological targets. In terms of Fig. 1.1, this amounts to taking the first step (moving into the model world), possibly on somewhat vague premises, and then stopping there. Modeling effectively becomes a way of organizing information by constructing a knowledge database in the model world. Modeling communities are being built up around such projects and some modelers may eventually become permanently preoccupied with the model world they created.

When the visions and ambitions of kinetic models grow, it must not be forgotten that mathematical models are primarily tools for answering questions, and the path towards the answer is normally founded on the core idea of modeling: simplification and reduction of unnecessary complexity. While some questions may indeed require large models, it is also important to think about how to retain an appropriate level of model structure complexity (Helmlinger et al. 2017; Ataman et al. 2017; Ribba et al. 2017). Furthermore, the model size is just one of the model characteristics contributing to the total complexity of a model. The total complexity that a model can afford (given the prior knowledge and the access to experimental data) could in some cases also be allocated to the use of a more demanding mathematical framework. As shown in this thesis, the development of new frameworks and the tailoring of existing ones to fit new problems has potential in several life science disciplines, and may be just as important as merely increasing the model size.

²Both parameter estimation and evaluation of this model's predictive capabilities were performed using steady state metabolic flux data, and it is therefore unfair to compare its size directly with other *E. coli* models that aim to capture dynamic aspects of metabolism (Chassagnole et al. 2002; Kotte et al. 2010).

4.8 A Final Remark

The topic of kinetic models in life science is of course much too large to be fully covered within one thesis. It is the author's hope that this work can at least serve as a starting point for an organized way of thinking about the workflow for building and using kinetic models, and contribute some examples of kinetic models and modeling methods.

Bibliography

- Aarons L (2005). Physiologically based pharmacokinetic modelling: a sound mechanistic basis is needed. *Br. J. Clin. Pharmacol.* 60 (6): 581–583. DOI: [10.1111/j.1365-2125.2005.02560.x](https://doi.org/10.1111/j.1365-2125.2005.02560.x).
- Almquist J, Penney M, Pehrsson S, Sandinge AS, Janefeldt A, Maqbool S, Madalli S, Goodman J, Nylander S, and Gennemark P (2016). Unraveling the pharmacokinetic interaction of ticagrelor and MEDI2452 (ticagrelor antidote) by mathematical modeling. *CPT Pharmacometrics Syst. Pharmacol.* 5 (6): 313–323. DOI: [10.1002/psp4.12089](https://doi.org/10.1002/psp4.12089).
- Almquist J, Bendrioua L, Adiels CB, Goksör M, Hohmann S, and Jirstrand M (2015a). A nonlinear mixed effects approach for modeling the cell-to-cell variability of Mig1 dynamics in yeast. *PLoS One* 10 (4): e0124050. DOI: [10.1371/journal.pone.0124050](https://doi.org/10.1371/journal.pone.0124050).
- Almquist J, Cvijovic M, Hatzimanikatis V, Nielsen J, and Jirstrand M (2014). Kinetic models in industrial biotechnology — Improving cell factory performance. *Metab. Eng.* 24: 38–60. DOI: [10.1016/j.ymben.2014.03.007](https://doi.org/10.1016/j.ymben.2014.03.007).
- Almquist J, Leander J, and Jirstrand M (2015b). Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood. *J. Pharmacokinet. Pharmacodyn.* 42 (3): 191–209. DOI: [10.1007/s10928-015-9409-1](https://doi.org/10.1007/s10928-015-9409-1).
- Almquist J, Wallman M, Jacobson I, and Jirstrand M (2010). Modeling the effect of Kv1.5 block on the canine action potential. *Biophys. J.* 99 (9): 2726–2736. DOI: [10.1016/j.bpj.2010.08.062](https://doi.org/10.1016/j.bpj.2010.08.062).
- Alvarez-Vasquez F, González-Alcón C, and Torres NV (2000). Metabolism of citric acid production by *Aspergillus niger*: model definition, steady-state analysis and constrained optimization of citric acid production rate. *Biotechnol. Bioeng.* 70 (1): 82–108.
- Andersson R, Kroon T, Almquist J, Jirstrand M, Oakes ND, Evans ND, Chappel MJ, and Gabrielsson J (2017). Modeling of free fatty acid dynamics: insulin and nicotinic acid resistance under acute and chronic treatments. *J. Pharmacokinet. Pharmacodyn.* 44 (3): 203–222. DOI: [10.1007/s10928-017-9512-6](https://doi.org/10.1007/s10928-017-9512-6).

- Anguelova M, Karlsson J, and Jirstrand M (2012). Minimal output sets for identifiability. *Math. Biosci.* 239 (1): 139–153. DOI: [10.1016/j.mbs.2012.04.005](https://doi.org/10.1016/j.mbs.2012.04.005).
- Aoki Y, Nordgren R, and Hooker AC (2016). Preconditioning of nonlinear mixed effects models for stabilisation of variance-covariance matrix computations. *AAPS J.* 18 (2): 505–518. DOI: [10.1208/s12248-016-9866-5](https://doi.org/10.1208/s12248-016-9866-5).
- Apgar JF, Witmer DK, White FM, and Tidor B (2010). Sloppy models, parameter uncertainty, and the role of experimental design. *Mol. Biosyst.* 6 (10): 1890–1900. DOI: [10.1039/b918098b](https://doi.org/10.1039/b918098b).
- Ataman M, Hernandez Gardiol DF, Fengos G, and Hatzimanikatis V (2017). redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. *PLoS Comput. Biol.* 13 (7): e1005444. DOI: [10.1371/journal.pcbi.1005444](https://doi.org/10.1371/journal.pcbi.1005444).
- Bajpai R and Reuss M (1980). A mechanistic model for penicillin production. *J. Appl. Chem. Biotechnol.* 30: 332–344.
- Bauer RJ, Guzy S, and Ng C (2007). A survey of population analysis methods and software for complex pharmacokinetic and pharmacodynamic models with examples. *AAPS J.* 9 (1): E60–E83. DOI: [10.1208/aapsj0901007](https://doi.org/10.1208/aapsj0901007).
- Beal SL, Sheiner LB, Boeckmann AJ, and Bauer RJ (2017). NONMEM 7.4 users guides. (1989–2017). ICON plc. Gaithersburg, MD.
- Bendrioua L, Smedh M, Almquist J, Cvijovic M, Jirstrand M, Goksör M, Adiels CB, and Hohmann S (2014). Yeast AMP-activated protein kinase monitors glucose concentration changes and absolute glucose levels. *J. Biol. Chem.* 289 (18): 12863–12875. DOI: [10.1074/jbc.M114.547976](https://doi.org/10.1074/jbc.M114.547976).
- Bertrand J, Laffont CM, Mentré F, Chenel M, and Comets E (2011). Development of a complex parent-metabolite joint population pharmacokinetic model. *AAPS J.* 13 (3): 390–404. DOI: [10.1208/s12248-011-9282-9](https://doi.org/10.1208/s12248-011-9282-9).
- Boger E, Evans N, Chappell M, Lundqvist A, Ewing P, Wigenborg A, and Fridén M (2016). Systems pharmacology approach for prediction of pulmonary and systemic pharmacokinetics and receptor occupancy of inhaled drugs. *CPT Pharmacometrics Syst. Pharmacol.* 5 (4): 201–210. DOI: [10.1002/psp4.12074](https://doi.org/10.1002/psp4.12074).
- Bonaca MP, Bhatt DL, Cohen M, Steg PG, Storey RF, Jensen EC, Magnani G, Bansilal S, Fish MP, Im K, Bengtsson O, Oude Ophuis T, Budaj A, Theroux P, Ruda M, Hamm C, Goto S, Spinar J, Nicolau JC, Kiss RG, Murphy SA, Wiviott SD, Held P, Braunwald E, Sabatine MS, Committee PS, and Investigators (2015). Long-term use of ticagrelor in patients with prior myocardial infarction. *N. Engl. J. Med.* 372 (19): 1791–1800. DOI: [10.1056/NEJMoa1500857](https://doi.org/10.1056/NEJMoa1500857).
- Bonate PL (2011). Pharmacokinetic-pharmacodynamic modeling and simulation. Springer US.
- Botstein D and Fink GR (2011). Yeast: an experimental organism for 21st century biology. *Genetics* 189 (3): 695–704. DOI: [10.1534/genetics.111.130765](https://doi.org/10.1534/genetics.111.130765).
- Bowden LG, Maini PK, Moulton DE, Tang JB, Wang XT, Liu PY, and Byrne HM (2014). An ordinary differential equation model for full thickness wounds and the effects of diabetes. *J. Theor. Biol.* 361: 87–100. DOI: [10.1016/j.jtbi.2014.07.001](https://doi.org/10.1016/j.jtbi.2014.07.001).

- Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, and Cerione RA (2004). The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys. Biol.* 1 (3-4): 184–195. DOI: 10.1088/1478-3967/1/3/006.
- Buchanan A, Newton P, Pehrsson S, Inghardt T, Antonsson T, Svensson P, Sjögren T, Öster L, Janefeldt A, Sandinge AS, Keyes F, Austin M, Spooner J, Gennemark P, Penney M, Howells G, Vaughan T, and Nylander S (2015). Structural and functional characterization of a specific antidote for ticagrelor. *Blood* 125 (22): 3484–3490. DOI: 10.1182/blood-2015-01-622928.
- Büchel F, Rodriguez N, Swainston N, Wrzodek C, Czauderna T, Keller R, Mittag F, Schubert M, Glont M, Golebiewski M, Iersel M van, Keating S, Rall M, Wybrow M, Hermjakob H, Hucka M, Kell DB, Müller W, Mendes P, Zell A, Chaouiya C, Saez-Rodriguez J, Schreiber F, Laibe C, Dräger A, and Le Novère N (2013). Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.* 7: 116. DOI: 10.1186/1752-0509-7-116.
- Cardilin T, Almquist J, Jirstrand M, Sostelly A, Amendt C, El Bawab S, and Gabrielsson J (2017a). Tumor static concentration curves in combination therapy. *AAPS J.* 19 (2): 456–467. DOI: 10.1208/s12248-016-9991-1.
- Cardilin T, Almquist J, Jirstrand M, Zimmermann A, El Bawab S, and Gabrielsson J (2017b). Model-based evaluation of radiation and radiosensitizing agents in oncology. *CPT Pharmacometrics Syst. Pharmacol.* Accepted.
- Certara (2011–2017). Phoenix NLME.
- Chakrabarti A, Miskovic L, Soh KC, and Hatzimanikatis V (2013). Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnol. J.* 8 (9): 1043–1057. DOI: 10.1002/biot.201300091.
- Chan PLS, Jacqmin P, Lavielle M, McFadyen L, and Weatherley B (2011). The use of the SAEM algorithm in MONOLIX software for estimation of population pharmacokinetic-pharmacodynamic-viral dynamics parameters of maraviroc in asymptomatic HIV subjects. *J. Pharmacokinet. Pharmacodyn.* 38 (1): 41–61. DOI: 10.1007/s10928-010-9175-z.
- Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, and Reuss M (2002). Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol. Bioeng.* 79 (1): 53–73.
- Chen WW, Niepel M, and Sorger PK (2010). Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev.* 24 (17): 1861–1875. DOI: 10.1101/gad.1945410.
- Chou IC and Voit EO (2009). Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math. Biosci.* 219 (2): 57–83. DOI: 10.1016/j.mbs.2009.03.002.
- Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Pesce CG, and Brent R (2005). Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437 (7059): 699–706. DOI: 10.1038/nature03998.
- Cvijovic M, Almquist J, Hagmar J, Hohmann S, Kaltenbach HM, Klipp E, Krantz M, Mendes P, Nelander S, Nielsen J, Pagnani A, Przulj N, Raue A, Stelling J, Stoma S, Tobin F, Wodke JAH, Zecchina R, and Jirstrand M (2014). Bridging the gaps in

- systems biology. *Mol. Genet. Genomics* 289 (5): 727–734. DOI: [10.1007/s00438-014-0843-3](https://doi.org/10.1007/s00438-014-0843-3).
- Daniels BC, Chen YJ, Sethna JP, Gutenkunst RN, and Myers CR (2008). Sloppiness, robustness, and evolvability in systems biology. *Curr. Opin. Biotechnol.* 19 (4): 389–395. DOI: [10.1016/j.copbio.2008.06.008](https://doi.org/10.1016/j.copbio.2008.06.008).
- Darbois C, Lemenuel-Diot A, Laveille C, Tranchand B, Tod M, and Girard P (2007). Evaluation of uncertainty parameters estimated by different population PK software and methods. *J. Pharmacokinet. Pharmacodyn.* 34 (3): 289–311. DOI: [10.1007/s10928-006-9046-9](https://doi.org/10.1007/s10928-006-9046-9).
- Davidian M and Giltinan DM (2003). Nonlinear models for repeated measurement data: An overview and update. *J. Agric. Biol. Environ. Stat.* 8: 387–419.
- Delvigne F, Zune Q, Lara AR, Al-Soud W, and Sørensen SJ (2014). Metabolic variability in bioprocessing: implications of microbial phenotypic heterogeneity. *Trends Biotechnol.* 32 (12): 608–616. DOI: [10.1016/j.tibtech.2014.10.002](https://doi.org/10.1016/j.tibtech.2014.10.002).
- Dobrev D, Carlsson L, and Nattel S (2012). Novel molecular targets for atrial fibrillation therapy. *Nat. Rev. Drug Discov.* 11 (4): 275–291. DOI: [10.1038/nrd3682](https://doi.org/10.1038/nrd3682).
- Doucet A, Freitas N de, and Gordon N (2001). Sequential Monte Carlo methods in practice. Springer-Verlag.
- Durandau E, Aymoz D, and Pelet S (2015). Dynamic single cell measurements of kinase activity by synthetic kinase activity relocation sensors. *BMC Biol.* 13: 55. DOI: [10.1186/s12915-015-0163-z](https://doi.org/10.1186/s12915-015-0163-z).
- Elowitz MB, Levine AJ, Siggia ED, and Swain PS (2002). Stochastic gene expression in a single cell. *Science* 297 (5584): 1183–1186. DOI: [10.1126/science.1070919](https://doi.org/10.1126/science.1070919).
- Erjavec N, Cvijovic M, Klipp E, and Nyström T (2008). Selective benefits of damage partitioning in unicellular systems and its effects on aging. *Proc. Natl. Acad. Sci. U.S.A.* 105 (48): 18764–18769. DOI: [10.1073/pnas.0804550105](https://doi.org/10.1073/pnas.0804550105).
- Ette EI and Williams PJ (2004). Population pharmacokinetics II: estimation methods. *Ann. Pharmacother.* 38 (11): 1907–1915. DOI: [10.1345/aph.1E259](https://doi.org/10.1345/aph.1E259).
- Förster J, Famili I, Fu P, Palsson BØ, and Nielsen J (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 13 (2): 244–253. DOI: [10.1101/gr.234503](https://doi.org/10.1101/gr.234503).
- Gabrielsson J and Weiner D (2016). Pharmacokinetic and pharmacodynamic data analysis: Concepts and applications. Apotekarsocieteten.
- García-Salcedo R, Lubitz T, Beltran G, Elbing K, Tian Y, Frey S, Wolkenhauer O, Krantz M, Klipp E, and Hohmann S (2014). Glucose de-repression by yeast AMP-activated protein kinase Snf1 is controlled via at least two independent steps. *FEBS J.* 281 (7): 1901–1917. DOI: [10.1111/febs.12753](https://doi.org/10.1111/febs.12753).
- Gaudet S, Spencer SL, Chen WW, and Sorger PK (2012). Exploring the contextual sensitivity of factors that determine cell-to-cell variability in receptor-mediated apoptosis. *PLoS Comput. Biol.* 8 (4): e1002482. DOI: [10.1371/journal.pcbi.1002482](https://doi.org/10.1371/journal.pcbi.1002482).
- Gennemark P, Trägårdh M, Lindén D, Ploj K, Johansson A, Turnbull A, Carlsson B, and Antonsson M (2017). Translational modeling to guide study design and dose choice in obesity exemplified by AZD1979, a melanin-concentrating

- hormone receptor 1 antagonist. *CPT Pharmacometrics Syst. Pharmacol.* DOI: 10.1002/psp4.12199.
- Glynn P, Unudurthi SD, and Hund TJ (2014). Mathematical modeling of physiological systems: an essential tool for discovery. *Life Sci.* 111 (1-2): 1–5. DOI: 10.1016/j.lfs.2014.07.005.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, and Oliver SG (1996). Life with 6000 genes. *Science* 274 (5287): 546, 563–546, 567.
- Gonze D, Jacquet M, and Goldbeter A (2008). Stochastic modelling of nucleocytoplasmic oscillations of the transcription factor Msn2 in yeast. *J. R. Soc. Interface* 5 Suppl 1: S95–109. DOI: 10.1098/rsif.2008.0141.focus.
- Graaf PH van der and Benson N (2011). Systems pharmacology: bridging systems biology and pharmacokinetics-pharmacodynamics (PKPD) in drug discovery and development. *Pharm. Res.* 28 (7): 1460–1464. DOI: 10.1007/s11095-011-0467-9.
- Gustavsson AK, Niekerk DD van, Adiels CB, Preez FB du, Goksör M, and Snoep JL (2012). Sustained glycolytic oscillations in individual isolated yeast cells. *FEBS J.* 279 (16): 2837–2847. DOI: 10.1111/j.1742-4658.2012.08639.x.
- Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, and Sethna JP (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* 3 (10): 1871–1878. DOI: 10.1371/journal.pcbi.0030189.
- Hammarlund-Udenaes M, Fridén M, Syvänen S, and Gupta A (2008). On the rate and extent of drug delivery to the brain. *Pharm. Res.* 25 (8): 1737–1750. DOI: 10.1007/s11095-007-9502-2.
- Hasty J, Pradines J, Dolnik M, and Collins JJ (2000). Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 97 (5): 2075–2080. DOI: 10.1073/pnas.040411297.
- Heijman J, Erfanian Abdoust P, Voigt N, Nattel S, and Dobrev D (2016). Computational models of atrial cellular electrophysiology and calcium handling, and their role in atrial fibrillation. *J. Physiol.* 594 (3): 537–553. DOI: 10.1113/JP271404.
- Heijnen JJ (2005). Approximative kinetic formats used in metabolic network modeling. *Biotechnol. Bioeng.* 91 (5): 534–545. DOI: 10.1002/bit.20558.
- Helmlinger G, Al-Huniti N, Aksенov S, Peskov K, Hallow KM, Chu L, Boulton D, Eriksson U, Hamrén B, Lambert C, Masson E, Tomkinson H, and Stanski D (2017). Drug-disease modeling in the pharmaceutical industry - where mechanistic systems pharmacology and statistical pharmacometrics meet. *Eur. J. Pharm. Sci.* DOI: 10.1016/j.ejps.2017.05.028.
- Hilfinger A and Paulsson J (2011). Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc. Natl. Acad. Sci. U.S.A.* 108 (29): 12167–12172. DOI: 10.1073/pnas.1018832108.
- Hodgkin AL and Huxley AF (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117 (4): 500–544.

- Holzhütter HG, Drasdo D, Preusser T, Lippert J, and Henney AM (2012). The virtual liver: a multidisciplinary, multilevel challenge for systems biology. Wiley Interdiscip. Rev. Syst. Biol. Med. 4 (3): 221–235. DOI: [10.1002/wsbm.1158](https://doi.org/10.1002/wsbm.1158).
- Hong KK and Nielsen J (2012). Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. Cell. Mol. Life Sci. 69 (16): 2671–2690. DOI: [10.1007/s00018-012-0945-1](https://doi.org/10.1007/s00018-012-0945-1).
- Huang Y, Liu D, and Wu H (2006). Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. Biometrics 62 (2): 413–423. DOI: [10.1111/j.1541-0420.2005.00447.x](https://doi.org/10.1111/j.1541-0420.2005.00447.x).
- Hunter PJ and Borg TK (2003). Integration from proteins to organs: the Physiome Project. Nat. Rev. Mol. Cell Biol. 4 (3): 237–243. DOI: [10.1038/nrm1054](https://doi.org/10.1038/nrm1054).
- Jazwinski AH (1970). Stochastic processes and filtering theory. Academic Press, New York.
- Karaaslan F, Denizhan Y, Kayserilioglu A, and Gulcur HO (2005). Long-term mathematical model involving renal sympathetic nerve activity, arterial pressure, and sodium excretion. Ann. Biomed. Eng. 33 (11): 1607–1630. DOI: [10.1007/s10439-005-5976-4](https://doi.org/10.1007/s10439-005-5976-4).
- Karlsson M, Janzén DLI, Durrieu L, Colman-Lerner A, Kjellsson MC, and Cedersund G (2015). Nonlinear mixed-effects modelling for single cell estimation: when, why, and how to use it. BMC Syst. Biol. 9: 52. DOI: [10.1186/s12918-015-0203-x](https://doi.org/10.1186/s12918-015-0203-x).
- Khodayari A and Maranas CD (2016). A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. Nat. Commun. 7: 13806. DOI: [10.1038/ncomms13806](https://doi.org/10.1038/ncomms13806).
- Kholodenko BN (2006). Cell-signalling dynamics in time and space. Nat. Rev. Mol. Cell Biol. 7 (3): 165–176. DOI: [10.1038/nrm1838](https://doi.org/10.1038/nrm1838).
- Kitano H (2002). Systems biology: a brief overview. Science 295 (5560): 1662–1664. DOI: [10.1126/science.1069492](https://doi.org/10.1126/science.1069492).
- Klipp E, Nordlander B, Krüger R, Gennemark P, and Hohmann S (2005). Integrative model of the response of yeast to osmotic shock. Nat. Biotechnol. 23 (8): 975–982. DOI: [10.1038/nbt1114](https://doi.org/10.1038/nbt1114).
- Kollmann M, Løvdok L, Bartholomé K, Timmer J, and Sourjik V (2005). Design principles of a bacterial signalling network. Nature 438 (7067): 504–507. DOI: [10.1038/nature04228](https://doi.org/10.1038/nature04228).
- Kotte O, Zaugg JB, and Heinemann M (2010). Bacterial adaptation through distributed sensing of metabolic fluxes. Mol. Syst. Biol. 6: 355. DOI: [10.1038/msb.2010.10](https://doi.org/10.1038/msb.2010.10).
- Kristensen NR, Madsen H, and Ingwersen SH (2005). Using stochastic differential equations for PK/PD model development. J. Pharmacokinet. Pharmacodyn. 32 (1): 109–141. DOI: [10.1007/s10928-005-2105-9](https://doi.org/10.1007/s10928-005-2105-9).
- Kuepfer L, Peter M, Sauer U, and Stelling J (2007). Ensemble modeling for analysis of cell signaling dynamics. Nat. Biotechnol. 25 (9): 1001–1006. DOI: [10.1038/nbt1330](https://doi.org/10.1038/nbt1330).
- Kuhn E and Lavielle M (2005). Maximum likelihood estimation in nonlinear mixed effects models. Comput. Stat. Data Anal. 49: 1020–38.

- Lagrutta A, Wang J, Fermini B, and Salata JJ (2006). Novel, potent inhibitors of human Kv1.5 K⁺ channels and ultrarapidly activating delayed rectifier potassium current. *J. Pharmacol. Exp. Ther.* 317 (3): 1054–1063. DOI: 10.1124/jpet.106.101162.
- Lapin A, Müller D, and Reuss M (2004). Dynamic behavior of microbial populations in stirred bioreactors simulated with Euler-Lagrange methods: traveling along the lifelines of single cells. *Ind. Eng. Chem. Res.* 43: 4647–4656.
- Lavielle M (2014a). Monolix - A software for the analysis of nonlinear mixed effects models. The Monolix Group.
- Lavielle M (2014b). Mixed effects models for the population approach. models, tasks, methods and tools. Chapman & Hall/CRC Biostatistics Series.
- Leander J, Almquist J, Ahlström C, Gabrielsson J, and Jirstrand M (2015). Mixed effects modeling using stochastic differential equations: Illustrated by pharmacokinetic data of nicotinic acid in obese Zucker rats. *AAPS J.* 17 (3): 586–596. DOI: 10.1208/s12248-015-9718-8.
- Leander J, Lundh T, and Jirstrand M (2014). Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements. *Math. Biosci.* 251: 54–62. DOI: 10.1016/j.mbs.2014.03.001.
- Leary B, Dunlavay M, Chittenden J, Matzuka B, and Guzy S (2012). QRPEM - A new standard of accuracy, precision, and efficiency in NLME population PK/PD methods.
- Lidstrom ME and Konopka MC (2010). The role of physiological heterogeneity in microbial population behavior. *Nat. Chem. Biol.* 6 (10): 705–712. DOI: 10.1038/nchembio.436.
- Liebermeister W, Uhendorf J, and Klipp E (2010). Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics* 26 (12): 1528–1534. DOI: 10.1093/bioinformatics/btq141.
- Lindstrom ML and Bates DM (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* 46 (3): 673–687.
- Liu X and Wang Y (2016). Comparing the performance of FOCE and different expectation-maximization methods in handling complex population physiologically-based pharmacokinetic models. *J. Pharmacokinet. Pharmacodyn.* DOI: 10.1007/s10928-016-9476-y.
- Llamosi A, Gonzalez-Vargas AM, Versari C, Cinquemani E, Ferrari-Trecate G, Hersen P, and Batt G (2016). What population reveals about individual cell identity: Single-cell parameter estimation of models of gene expression in yeast. *PLoS Comput. Biol.* 12 (2): e1004706. DOI: 10.1371/journal.pcbi.1004706.
- Lotka AJ (1925). Elements of physical biology. Williams and Wilkins.
- Marucci L, Santini S, Bernardo M di, and Bernardo D di (2011). Derivation, identification and validation of a computational model of a novel synthetic regulatory network in yeast. *J. Math. Biol.* 62 (5): 685–706. DOI: 10.1007/s00285-010-0350-z.

- Matzuka B, Chittenden J, Monteleone J, and Tran H (2016). Stochastic nonlinear mixed effects: a metformin case study. *J. Pharmacokinet. Pharmacodyn.* 43 (1): 85–98. DOI: [10.1007/s10928-015-9456-7](https://doi.org/10.1007/s10928-015-9456-7).
- Mendes P and Kell D (1998). Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* 14 (10): 869–883.
- Menezes JC, Alves SS, Lemos JM, and Azevedo SF de (1994). Mathematical modelling of industrial pilot-plant penicillin-G fed-batch fermentations. *J. Chem. Tech. Biotechnol.* 61 (2): 123–138. DOI: [10.1002/jctb.280610207](https://doi.org/10.1002/jctb.280610207).
- Michaelis L and Menten ML (1913). Die kinetik der invertinwirkung. *Biochem. Z.* 49: 333–369.
- Monod J (1942). Recherches sur la croissance des cultures bactériennes. Hermann et Cie.
- Motta S and Pappalardo F (2013). Mathematical modeling of biological systems. *Briefings Bioinf.* 14 (4): 411–422. DOI: [10.1093/bib/bbs061](https://doi.org/10.1093/bib/bbs061).
- Nattel S (2002). New ideas about atrial fibrillation 50 years on. *Nature* 415 (6868): 219–226. DOI: [10.1038/415219a](https://doi.org/10.1038/415219a).
- Nielsen J (2017). Systems biology of metabolism. *Annu. Rev. Biochem.* 86: 245–275. DOI: [10.1146/annurev-biochem-061516-044757](https://doi.org/10.1146/annurev-biochem-061516-044757).
- Noble D, Garny A, and Noble PJ (2012). How the Hodgkin-Huxley equations inspired the Cardiac Physiome Project. *J. Physiol.* 590 (11): 2613–2628. DOI: [10.1113/jphysiol.2011.224238](https://doi.org/10.1113/jphysiol.2011.224238).
- Nocedal J and Wright SJ (1999). Numerical optimization. Springer-Verlag.
- Nyberg J, Bazzoli C, Ogungbenro K, Aliev A, Leonov S, Duffull S, Hooker AC, and Mentré F (2015). Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics studies. *Br. J. Clin. Pharmacol.* 79 (1): 6–17. DOI: [10.1111/bcp.12352](https://doi.org/10.1111/bcp.12352).
- Odutayo A, Wong CX, Hsiao AJ, Hopewell S, Altman DG, and Emdin CA (2016). Atrial fibrillation and risks of cardiovascular disease, renal disease, and death: systematic review and meta-analysis. *BMJ* 354: i4482. DOI: [10.1136/bmj.i4482](https://doi.org/10.1136/bmj.i4482).
- Ólafsdóttir HK (2016). Sensitivity-based gradients for parameter estimation in nonlinear mixed effects models with deterministic and stochastic dynamics. MSc. Chalmers University of Technology.
- Ólafsdóttir HK, Leander J, Almquist J, and Jirstrand M (2017). Exact gradients improve parameter estimation in nonlinear mixed effects models with stochastic dynamics. In proceedings of the 8th American Conference of Pharmacometrics.
- Overgaard RV, Jonsson N, Tornøe CW, and Madsen H (2005). Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm. *J. Pharmacokinet. Pharmacodyn.* 32 (1): 85–107. DOI: [10.1007/s10928-005-2104-x](https://doi.org/10.1007/s10928-005-2104-x).
- Pedraza JM and Oudenaarden A van (2005). Noise propagation in gene networks. *Science* 307 (5717): 1965–1969. DOI: [10.1126/science.1109090](https://doi.org/10.1126/science.1109090).
- Pehrsson S, Johansson KJ, Janefeldt A, Sandinge AS, Maqbool S, Goodman J, Sanchez J, Almquist J, Gennemark P, and Nylander S (2017). Hemostatic effects of

- the ticagrelor antidote MEDI2452 in pigs treated with ticagrelor on a background of aspirin. *J. Thromb. Haemost.* 15 (6): 1213–1222. DOI: 10.1111/jth.13680.
- Petersson KJF, Friberg LE, and Karlsson MO (2010). Transforming parts of a differential equations system to difference equations as a method for run-time savings in NONMEM. *J. Pharmacokinet. Pharmacodyn.* 37 (5): 493–506. DOI: 10.1007/s10928-010-9169-x.
- Petranovic D, Tyo K, Vemuri GN, and Nielsen J (2010). Prospects of yeast systems biology for human health: integrating lipid, protein and energy metabolism. *FEMS Yeast Res.* 10 (8): 1046–1059. DOI: 10.1111/j.1567-1364.2010.00689.x.
- Pfeffer M, Maurer M, Köllensperger G, Hann S, Graf AB, and Mattanovich D (2011). Modeling and measuring intracellular fluxes of secreted recombinant protein in *Pichia pastoris* with a novel 34S labeling procedure. *Microb. Cell Fact.* 10: 47. DOI: 10.1186/1475-2859-10-47.
- Ramirez RJ, Nattel S, and Courtemanche M (2000). Mathematical analysis of canine atrial action potentials: rate, regional factors, and electrical remodeling. *Am. J. Physiol.* 279 (4): H1767–H1785.
- Ramsey SA, Smith JJ, Orrell D, Marelli M, Petersen TW, Atauri P de, Bolouri H, and Aitchison JD (2006). Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. *Nat. Genet.* 38 (9): 1082–1087. DOI: 10.1038/ng1869.
- Raser JM and O’Shea EK (2004). Control of stochasticity in eukaryotic gene expression. *Science* 304 (5678): 1811–1814. DOI: 10.1126/science.1098641.
- Raue A, Kreutz C, Maiwald T, Klingmuller U, and Timmer J (2011). Addressing parameter identifiability by model-based experimentation. *IET Syst. Biol.* 5 (2): 120–130. DOI: 10.1049/iet-syb.2010.0061.
- Raue A, Schilling M, Bachmann J, Matteson A, Schelke M, Kaschek D, Hug S, Kreutz C, Harms BD, Theis FJ, Klingmüller U, and Timmer J (2013). Lessons learned from quantitative dynamical modeling in systems biology. *PLoS One* 8 (9): e74335. DOI: 10.1371/journal.pone.0074335.
- Ribba B, Grimm HP, Agoram B, Davies MR, Gadkar K, Niederer S, Riel N van, Timmis J, and Graaf PH van der (2017). Methodologies for quantitative systems pharmacology (QSP) models: Design and estimation. *CPT Pharmacometrics Syst. Pharmacol.* 6 (8): 496–498. DOI: 10.1002/psp4.12206.
- Riel NAW van (2006). Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Briefings Bioinf.* 7 (4): 364–374. DOI: 10.1093/bib/bb1040.
- Rizzi M, Baltes M, Theobald U, and Reuss M (1997). In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model. *Biotechnol. Bioeng.* 55 (4): 592–608.
- Rudy Y and Silva JR (2006). Computational biology in the study of cardiac ion channels and cell electrophysiology. *Q. Rev. Biophys.* 39 (1): 57–116. DOI: 10.1017/S0033583506004227.
- Saa PA and Nielsen LK (2016). Construction of feasible and accurate kinetic models of metabolism: A Bayesian approach. *Sci. Rep.* 6: 29635. DOI: 10.1038/srep29635.

- Schaber J and Klipp E (2011). Model-based inference of biochemical parameters and dynamic properties of microbial signal transduction networks. *Curr. Opin. Biotechnol.* 22 (1): 109–116. DOI: [10.1016/j.copbio.2010.09.014](https://doi.org/10.1016/j.copbio.2010.09.014).
- Sheiner LB and Beal SL (1981). Evaluation of methods for estimating population pharmacokinetic parameters. II. Biexponential model and experimental pharmacokinetic data. *J. Pharmacokinet. Biopharm.* 9 (5): 635–651.
- Shulgin B, Stone L, and Agur Z (1998). Pulse vaccination strategy in the SIR epidemic model. *Bull. Math. Biol.* 60 (6): 1123–1148. DOI: [10.1006/S0092-8240\(98\)90005-2](https://doi.org/10.1006/S0092-8240(98)90005-2).
- Silber HE, Jauslin PM, Frey N, Gieschke R, Simonsson USH, and Karlsson MO (2007). An integrated model for glucose and insulin regulation in healthy volunteers and type 2 diabetic patients following intravenous glucose provocations. *J. Clin. Pharmacol.* 47 (9): 1159–1171. DOI: [10.1177/0091270007304457](https://doi.org/10.1177/0091270007304457).
- Srinath S and Gunawan R (2010). Parameter identifiability of power-law biochemical system models. *J. Biotechnol.* 149 (3): 132–140. DOI: [10.1016/j.jbiotec.2010.02.019](https://doi.org/10.1016/j.jbiotec.2010.02.019).
- Stephanopoulos G, Aristidou AA, and Nielsen J (1998). Metabolic engineering: principles and methodologies. Academic press.
- Tapani S, Almquist J, Leander J, Ahlström C, Peletier LA, Jirstrand M, and Gabrielsson J (2014). Joint feedback analysis modeling of nonesterified fatty acids in obese Zucker rats and normal Sprague-Dawley rats after different routes of administration of nicotinic acid. *J. Pharm. Sci.* 103 (8): 2571–2584. DOI: [10.1002/jps.24077](https://doi.org/10.1002/jps.24077).
- Topp B, Promislow K, deVries G, Miura RM, and Finegood DT (2000). A model of beta-cell mass, insulin, and glucose kinetics: pathways to diabetes. *J. Theor. Biol.* 206 (4): 605–619. DOI: [10.1006/jtbi.2000.2150](https://doi.org/10.1006/jtbi.2000.2150).
- Tornøe CW, Jacobsen JL, Pedersen O, Hansen T, and Madsen H (2004). Grey-box modelling of pharmacokinetic/pharmacodynamic systems. *J. Pharmacokinet. Pharmacodyn.* 31 (5): 401–417.
- Torres NV and Santos G (2015). The (mathematical) modeling process in biosciences. *Front. Genet.* 6: 354. DOI: [10.3389/fgene.2015.00354](https://doi.org/10.3389/fgene.2015.00354).
- Tyson JJ, Chen KC, and Novak B (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* 15 (2): 221–231.
- Ullah M and Wolkenhauer O (2010). Stochastic approaches in systems biology. Wiley Interdiscip. Rev. Syst. Biol. Med. 2 (4): 385–397. DOI: [10.1002/wsbm.78](https://doi.org/10.1002/wsbm.78).
- Van Giezen JJJ, Nilsson L, Berntsson P, Wissing BM, Giordanetto F, Tomlinson W, and Greasley PJ (2009). Ticagrelor binds to human P2Y(12) independently from ADP but antagonizes ADP-induced receptor signaling and platelet aggregation. *J. Thromb. Haemost.* 7 (9): 1556–1565. DOI: [10.1111/j.1538-7836.2009.03527.x](https://doi.org/10.1111/j.1538-7836.2009.03527.x).
- Veening JW, Smits WK, and Kuipers OP (2008). Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.* 62: 193–210. DOI: [10.1146/annurev.micro.62.081307.163002](https://doi.org/10.1146/annurev.micro.62.081307.163002).
- Volterra V (1926). Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Acad. Lincei Roma* 2: 31–113.

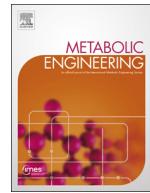
- Wallentin L, Becker RC, Budaj A, Cannon CP, Emanuelsson H, Held C, Horow J, Husted S, James S, Katus H, Mahaffey KW, Scirica BM, Skene A, Steg PG, Storey RF, Harrington RA, Investigators P, Freij A, and Thorsén M (2009). Ticagrelor versus clopidogrel in patients with acute coronary syndromes. *N. Engl. J. Med.* 361 (11): 1045–1057. DOI: 10.1056/NEJMoa0904327.
- Wang Y (2007). Derivation of various NONMEM estimation methods. *J. Pharmacokinet. Pharmacodyn.* 34 (5): 575–593. DOI: 10.1007/s10928-007-9060-6.
- Wettwer E and Terlau H (2014). Pharmacology of voltage-gated potassium channel Kv1.5—impact on cardiac excitability. *Curr. Opin. Pharmacol.* 15: 115–121. DOI: 10.1016/j.coph.2014.02.001.
- Wiechert W and Noack S (2011). Mechanistic pathway modeling for industrial biotechnology: challenging but worthwhile. *Curr. Opin. Biotechnol.* 22 (5): 604–610. DOI: 10.1016/j.copbio.2011.01.001.
- Wolfram Research, Inc. (2012). Mathematica. Version 9.0. Champaign, IL.
- Wolkenhauer O (2014). Why model? *Front. Physiol.* 5: 21. DOI: 10.3389/fphys.2014.00021.
- Wolkenhauer O, Auffray C, Jaster R, Steinhoff G, and Dammann O (2013). The road from systems biology to systems medicine. *Pediatr. Res.* 73 (4 Pt 2): 502–507. DOI: 10.1038/pr.2013.4.
- Wright PM (1998). Population based pharmacokinetic analysis: why do we need it; what is it; and what has it told us about anaesthetics? *Br. J. Anaesth.* 80 (4): 488–501.
- Xiao Y, Bowen CH, Liu D, and Zhang F (2016). Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis. *Nat. Chem. Biol.* 12 (5): 339–344. DOI: 10.1038/nchembio.2046.

Part II

Appended Papers

PAPER A

Kinetic Models in Industrial Biotechnology — Improving Cell Factory Performance



Minireview

Kinetic models in industrial biotechnology – Improving cell factory performance

Joachim Almquist ^{a,b,*}, Marija Cvijovic ^{c,d}, Vassily Hatzimanikatis ^e, Jens Nielsen ^b,
Mats Jirstrand ^a

^a Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Göteborg, Sweden

^b Systems and Synthetic Biology, Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

^c Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Göteborg, Sweden

^d Mathematical Sciences, University of Gothenburg, SE-412 96 Göteborg, Sweden

^e Laboratory of Computational Systems Biotechnology, Ecole Polytechnique Federale de Lausanne, CH 1015 Lausanne, Switzerland

ARTICLE INFO

Article history:

Received 11 September 2013

Received in revised form

7 March 2014

Accepted 9 March 2014

Available online 16 April 2014

Keywords:

Mathematical modeling

Kinetic models

Dynamic models

Parameter estimation

Cell factory

Industrial biotechnology

ABSTRACT

An increasing number of industrial bioprocesses capitalize on living cells by using them as cell factories that convert sugars into chemicals. These processes range from the production of bulk chemicals in yeasts and bacteria to the synthesis of therapeutic proteins in mammalian cell lines. One of the tools in the continuous search for improved performance of such production systems is the development and application of mathematical models. To be of value for industrial biotechnology, mathematical models should be able to assist in the rational design of cell factory properties or in the production processes in which they are utilized. Kinetic models are particularly suitable towards this end because they are capable of representing the complex biochemistry of cells in a more complete way compared to most other types of models. They can, at least in principle, be used to understand, predict, and evaluate the effects of adding, removing, or modifying molecular components of a cell factory and for supporting the design of the bioreactor or fermentation process. However, several challenges still remain before kinetic modeling will reach the degree of maturity required for routine application in industry. Here we review the current status of kinetic cell factory modeling. Emphasis is on modeling methodology concepts, including model network structure, kinetic rate expressions, parameter estimation, optimization methods, identifiability analysis, model reduction, and model validation, but several applications of kinetic models for the improvement of cell factories are also discussed.

© 2014 The Authors. Published by Elsevier Inc. On behalf of International Metabolic Engineering Society.

This is an open access article under the CC BY-NC-SA license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

1. Introduction

Throughout the World there is a desire to move towards sustainable production of energy, fuels, materials and chemicals, and biobased production of transportation fuels and chemicals is expected to contribute significantly towards reaching this objective. This has resulted in the advancement of industrial biotechnology, where microbial fermentation is used for the conversion of bio-based feedstocks to fuels and chemicals (Nielsen and Jewett, 2008; Tang and Zhao, 2009; Otero and Nielsen, 2010; Du et al., 2011; Sauer and Mattanovich, 2012). Not only has this resulted in a significant expansion of traditional processes such as bioethanol production, which has increased from 10 billion liters produced in

2010 to 75 billion liters produced in 2012, but it has also resulted in the introduction of novel processes for the production of chemicals that can be used for the production of polymers, e.g. lactic acid that goes into poly-lactate and 1,3 propanediol that goes into Sorona®. With these successes the chemical industry is looking into the development of other processes for the production of platform chemicals that can find application in the manufacturing of solvents and polymers. Traditionally the fermentation industry used naturally producing microorganisms, but today there is a focus on using a few microorganisms, often referred to as platform cell factories, and then engineering their metabolism such that they efficiently can produce the chemical of interest. This engineering process is referred to as metabolic engineering, and it involves the introduction of directed genetic modifications. Due to the complexity of microbial metabolism, both due to the large number of interacting reactions and the complex regulation, there has been an increasing focus on the use of mathematical models for the identification of metabolic

* Corresponding author at: Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Göteborg, Sweden.

E-mail address: joachim.almquist@fcc.chalmers.se (J. Almquist).

engineering targets (Patil et al., 2004; Cvijovic et al., 2011; Wiechert and Noack, 2011; Soh et al., 2012).

Industrial biotechnology can benefit from mathematical models by using them to understand, predict, and optimize the properties and behavior of cell factories (Työ et al., 2010). With valid models, improvement strategies can be discovered and evaluated in silico, saving both time and resources. Popular application of models thus includes using them to suggest targets for metabolic engineering leading to increases in yield, titer, and productivity of a desired product. Since these quantities not only depend on the genetic constitution of cells but to a large extent also on how the cells are utilized, models can additionally play a critical role in the optimization and control of the bioreactor and fermentation processes. Other possible model focus includes expanding the range of cell factory substrates, minimizing the formation of undesired by-products, increasing product quality, and guidance in the choice of cell factory when introducing a novel product.

Many biological processes or systems of importance to biotechnology, such as the metabolism of a cell culture during a fed-batch process, cellular stress responses, or the decision making during the cell cycle, are non-stationary in their nature. These systems are characterized by their dependence on time and the fact that the effect of inputs to the systems depends on the systems history. The most common way of modeling such dynamic systems is to set up mathematical expressions for the rates at which biochemical reactions of the systems are taking place. The reaction kinetics are then used to form mass balance equations which in turn describe the temporal behavior of all biochemical species present in the modeled system. Mathematical models of this type are usually referred to as kinetic models but the literature sometimes tends to use the terms dynamic and kinetic models interchangeably due to their largely overlapping concepts as far as biological models are concerned. Reaction kinetics being the fundamental building block of kinetic models, they are clearly distinguished from the large body of so-called genome-scale metabolic models (GEMs) which mainly focus on the stoichiometry of reactions (Thiele et al., 2009; Soh et al., 2010; Chung et al., 2010; Österlund et al., 2012). Although kinetic models are frequently being used to describe dynamic behaviors, they are equally important in the study of processes that may be stationary or close to stationary, such as cell metabolism

during exponential phase, since they can relate the properties of a (quasi) steady-state to the kinetic properties of the model components.

This review looks at the work-flow and methods for setting up, analyzing, and using kinetic models, focusing on models and modeling methodology with relevance for industrial biotechnology. The paper is divided into three main parts. The first part discusses and describes different aspects of the model building procedure, including defining the model focus, how to set up a model structure, determine parameter values and validate the model. The second part looks at how kinetic models have been used once they are set up. Applications of kinetic cell factory models for improving production, substrate utilization, product quality, and process design are reviewed. In the last part, a number of advantages and challenges of kinetic modeling are listed and some future perspectives of kinetic modeling in biotechnology are discussed. A complete overview of the organization can be found in Table 1. To increase the readability, especially for readers who are not experienced modelers, parts of the material which are of technical or mathematical nature are displayed in special boxes. The models and methods on which this review has been based have been supplied by the partners of SYSINBIO (Systems Biology as a Driver for Industrial Biotechnology, a coordination and support action funded by the European commission within the seventh framework programme) and through a thorough literature review.

2. Setting up kinetic models – Modeling framework

The kinetic modeling procedure can be divided into a number of steps which are illustrated in Fig. 1. Since the choices and decisions made at the different steps are dependent both on the objective of the modeling and on the previous steps, the exact details of how a model is set up will be different from case to case. Also, some steps will probably have to be iterated several times before a complete model can be presented (van Riel, 2006). For instance, the model structure will most certainly evolve during the model building process, having new elements added and other removed or changed. Parameter estimation may have to be performed again as new data sets are collected, and different types of analysis on the finished model may lead to new applications that was not initially foreseen. This type of iterative work-flow is not unique for kinetic models of cell factories, but apply for modeling efforts in general (Ljung, 1987). The steps of the kinetic modeling procedure are now described briefly, and then followed by elaboration and in-depth discussions on some of their aspects.

Purpose: The first step of modeling is to define the purpose of the model, an important step as it includes the very reason for setting up a model in the first place. Typical questions are: Why do we model? What do we want to use the model for? What type of behavior should the model be able to explain? The majority of the goals of modeling cell factories are related to understanding and predicting their behavior when perturbing them either internally through genetic modifications, or externally by changing various environmental factors. The model purpose defines the complexity of the modeling problem and will influence all subsequent steps of the modeling procedure.

Network structure: The model network structure is the wiring diagram of the model. It defines the network of interconnected elements that are assumed to be important for the modeling task in question. For instance, it will contain elements such as compartments, concentrations of metabolites, enzymes and transcripts, and reactions (including transport across membranes), including their effectors and stoichiometric coefficients. It also defines the interfaces of the model with the un-modeled exterior.

Table 1
Organization of this review.

Contents	
1	Introduction
2	Setting up kinetic models – Modeling framework
2.1	Purpose
2.2	Model structure
2.2.1	Representation of network structure
2.2.2	Kinetic rate expressions
2.2.3	Approximate kinetic rate expressions
2.2.4	Stochastic kinetics
2.3	Parameter determination
2.3.1	Computing the estimate
2.3.2	Identifiability analysis and experimental design
2.3.3	Model reduction
2.4	Validation
3	Using kinetic models
3.1	Improving production
3.1.1	Local parameter sensitivity analysis
3.1.2	Simulating larger changes
3.1.3	Optimization problems
3.2	Improving substrate utilization
3.3	Improving product quality
3.4	Improving process design
4	Advantages, challenges and perspectives
4.1	Advantages
4.2	Challenges
4.3	Perspectives

Box 1—Mass balance equations and model outputs.

Combining the stoichiometric information from the model network structure with the symbolic form of the kinetic rate expressions, mass balance equations with explicitly given kinetics can be set up for all dynamic components of the modeled system. In the deterministic, continuous case, these equations can be written as

$$\frac{dx(t)}{dt} = S \cdot v(x(t), u(t), \theta) \quad (1)$$

and their associated initial conditions are

$$x(0) = x_0(\theta). \quad (2)$$

Here, $x(t)$ denotes an m -dimensional vector of time-dependent state variables, S a stoichiometric matrix of dimension $m \times n$, and $v(x(t), u(t), \theta)$ an n -dimensional vector of reaction rates which are dependent on the state variables, a vector of input variables $u(t)$, and a set of parameters θ . Eq. (1) sometimes needs to be extended to take volume changes of the respective compartment into account, for example the dilution of intracellular species in growing cells. Additionally, it may be necessary to supplement the ordinary differential equations in Eq. (1) with a set of algebraic equations for certain models. Since the quantities measured in experiments are not necessarily the same as the model state variables, a function $h(x(t), \theta)$ is also needed to relate $x(t)$ to a vector of model outputs

$$y(t) = h(x(t), u(t), \theta). \quad (3)$$

Kinetic rate expressions: Having defined the model network structure, the next step in the modeling process is the determination of the mathematical expressions that define the interactions between the different components. The model network structure already delivers information about which elements should take part in the mathematical expressions. Kinetic rate expressions can be derived from actual reaction mechanisms, with different degrees of detail, or be represented by approximate expressions capturing the essential quantitative and qualitative features of a reaction. The complexity of a reaction's kinetics is defined by the scope of the reaction, the scope of the model and the biochemical knowledge about the reactions. Both deterministic and stochastic formulations of the reaction rates may be used.

Model structure: When the network structure and kinetic rate expressions have been determined, the structure of the kinetic model is complete. The model can now be written as a set of mass balance equations with explicitly given kinetic expressions, which determines the time trajectories of the modeled species, and a list of model outputs indicating which parts of the modeled system that are being observed in experiments, see Box 1.

Parameter determination: Next, the numerical values of the parameters appearing in the rate expressions, the initial conditions, and the outputs need to be determined. Parameter values are sometimes established one by one, either from targeted experiments measuring them directly or from other types of a priori information on individual parameter values. In contrast, parameter values can also be determined simultaneously in an inductive way by utilizing the implicit information in measurements of other quantities than the parameters themselves, using parameter estimation methods. If the parameter estimation problem does not have a unique solution, the space of admissible parameter values can be further constrained using physicochemical and thermodynamics laws. Subsequently, from such a reduced space parameter values can be determined by using Monte-Carlo sampling techniques.

Validation: With the parameter values determined, the quality of the model should be assessed. Such model validation can consist of both qualitative reasoning as well as formal statistical testing. In addition to explaining experimental data used for setting up the model, it is common to further validate the model's predictive power based on new sets of experimental data that was not used previously in the modeling process.

Usage: When a model has been established it can be used in a number of different ways to answer the questions for why it was

created. This involves various types of what-if analysis that explores different scenarios and investigates the impact of model assumptions. Examples of model usage include analysis of flux control in a pathway, *in silico* evaluation of metabolic engineering strategies, and design of optimal process conditions.

2.1. Purpose

Building models of biological systems is a way of collecting, organizing, and representing knowledge and hypotheses. The models can be thought of as formalized descriptions of what is known expressed by precise mathematical statements. They can be used for a variety of purposes including hypothesis testing, understanding how different components of a system work together to achieve some function or behavior, and learning about system components which are hard to access experimentally. Most importantly in the context of industrial biotechnology, they can be used for making predictions about the effects of genetic engineering, e.g. deleting or overexpressing a metabolic enzyme, and for optimizing the design and conditions of bioreactor or fermentation processes, e.g. determining the details of a fed-batch feeding strategy.

A common goal for many cell factory production processes, especially those for low-value products, is the desire to increase either yield, titer, or productivity, or combinations thereof. As a consequence, these quantities are ultimately what models should aim to describe and they are defined in Box 2. Which quantity is most relevant for a particular process is determined by a large number of factors such as the value and market size of the product, the substrate availability and cost, and the downstream processing. Although the models presented in this review do not always work directly with the above quantities, the models are usually describing aspects of cells and production processes that at least indirectly affect them and they should therefore always be kept in the back of the mind.

Essentially, any kinetic model whose purpose is to describe some aspects of the cellular machinery, or of the production process, that may impact the performance of a cell factory is of interest to biotechnology. Because there are many different types of cell factories and a plenitude of interesting products to be produced by them, the range of purposes and focus of potentially relevant kinetic models is wide. Depending on the problem they may address cellular processes such as metabolism, protein maturation and secretion, signaling, gene regulation, stress

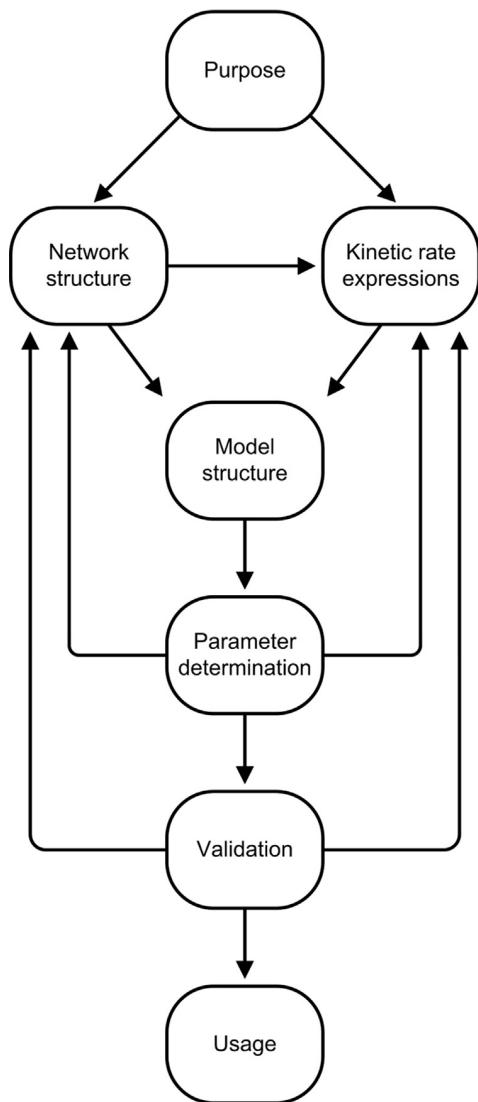


Fig. 1. Illustration of the main steps of the kinetic modeling procedure.

responses, cell cycle progression and apoptosis, as well as external or environmental factors like temperature, pH, osmolarity, product and by-product toxicity, and not least the type and operation mode of the bioreactor or fermentor. When describing the above features of cell factories a model may be specifically designed for a particular application, such as a specific pathway for the production of a special metabolite, or it can describe more general functions of the cell that may be exploited in different applications, such as primary metabolism or the protein synthesis machinery. The diversity in the purposes and scopes of kinetic models in biotechnology is reflected in the wide range of time-scales of commonly modeled processes. Fig. 2 shows how important processes such as signaling, the action of metabolic enzymes, gene expression, protein secretion, the cell cycle, and bioreactor processes have characteristic timescales that span and cover almost ten orders of magnitude. Also the size of kinetic models can be very different, ranging from single enzymes (Chauve et al., 2010; Hattersley et al., 2011), to entire pathways (Hynne et al., 2001), to larger models comprising several interacting modules or pathways (Klipp et al., 2005b; Kotte et al., 2010).

2.2. Model structure

Contemporary kinetic modeling is increasingly targeting cells at the molecular level, describing components like genes, enzymes, signaling proteins, and metabolites. From a metabolic engineering perspective this is in principle advantageous since it is at this level that genetic alterations eventually would take place. In a process referred to as a bottom-up or forward modeling, mechanistic descriptions of a system's components are integrated to form a description of the system as a whole (Bruggeman and Westerhoff, 2007). The central idea of this approach is that the behavior of a system emerges from the interaction of its components, and, importantly, that the behavior can be calculated if the properties of the components and their interactions have been characterized in sufficient detail. In principle the bottom-up concept can also be applied to merge already existing models of cellular sub-systems into larger models (Klipp et al., 2005b; Snoep et al., 2006). As indicated in Fig. 1 a kinetic model consists of a network structure, a corresponding set of rate expressions, and their associated parameter values. Knowledge of all three parts is needed to form a complete model.

Box 2—Production process quantities.

If we let the time dependent functions $x(t)$, $p(t)$, and $c(t)$ denote the biomass concentration, the specific productivity, and the specific substrate consumption, respectively, of a cell factory production process with a duration time T , the accumulative yield can be defined as

$$\frac{\int_0^T x(t)p(t) dt}{\int_0^T x(t)c(t) dt}, \quad (4)$$

the titer as

$$\int_0^T x(t)p(t) dt, \quad (5)$$

and the productivity as

$$\frac{1}{T} \int_0^T x(t)p(t) dt. \quad (6)$$

Note that the T in the expression of the productivity might itself be a parameter for optimization. For models that only consider situations where $p(t)$ and $c(t)$ are approximatively constant, such as for a continuous cultivation or perhaps for a population of cells growing in exponential phase, the yield can instead be quantified by p/c and the titer and productivity can both be replaced by looking at the specific productivity p if only a particular profile of $x(t)$ is considered.

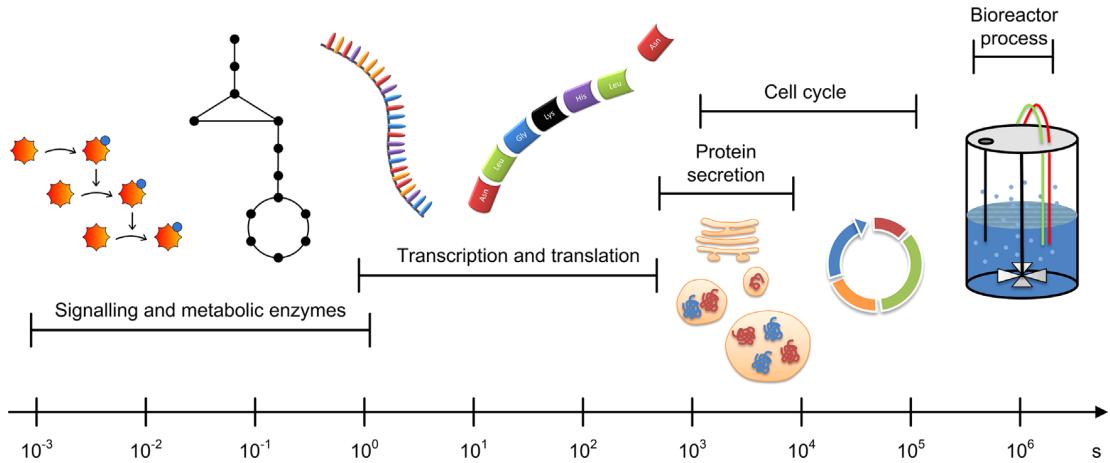


Fig. 2. Characteristic timescales for signaling, the action of metabolic enzymes, gene expression, protein secretion, the cell cycle, and bioreactor production processes.

Determination of the network structure and the symbolic structure of rate expressions in kinetic models are usually done according to the bottom-up approach (but exceptions exist, see for instance Mettetal and Muzey, 2008). It is dependent on experimental studies characterizing the properties of the individual components appearing in the model, information that is collected directly from the literature or compiled in databases. For some systems the components have been characterized in such detail that the bottom-up approach can be applied in its entirety (Bruggeman et al., 2005), also including the determination of all parameter values. However, it is common that some or all of the parameters are unknown and instead determined indirectly from system-level measurements of other quantities using parameter estimation methods, a strategy sometimes referred to as a top-down or inverse approach.

While the biochemistry and biophysics underlying the decision-making when setting up the model structure is in some cases well understood, this is generally far from true (Kaltenbach et al., 2009). Undoubtedly, network structures and rate expressions will be set up in incomplete or even incorrect ways. It may thus seem logical trying to infer the model structure from system-level data, in the same vein as the inverse problem of parameter estimation, but because of the countless possibilities of network structures and symbolic forms of rate expressions, a top-down approach is not feasible for this part of the model building process. One strategy for handling uncertainty in the model structure is to work with an ensemble of models with different structures. This approach has for instance been employed in a study of the TOR signaling pathway in *Saccharomyces cerevisiae* (Kuepfer et al., 2007). Other efforts have focused on the development of computational tools that support the handling of such model families (Haunschild et al., 2005). The problem can also in part be tackled by using different kinds of flexible standardized kinetic rate expressions that can display a large range of kinetic behaviors depending on their parameter values. In this way part of the structural identification problem can be turned into a parameter estimation problem (Chou and Voit, 2009; Srinath and Gunawan, 2010). In another variant of the bottom-up approach, addressing the issue of determining a suitable network structure, Hildebrandt et al. (2008) proposed a strategy where mechanistic modeling on the molecular level is combined with an incremental adding of model components in a systematic way. Starting from a basic backbone model, the effect of each added component can be evaluated to gain insight into its contribution to the overall behavior of the system. The authors of that study used the procedure to construct a model for optimizing the production of single-chain antibody fragment in *S. cerevisiae*, focusing on the chaperon binding protein and the foldase protein disulfide isomerase.

In contrast to the molecular level model structures, coarse-grained, lumped descriptions of biological systems and their parts are sometimes employed instead. Setting up models with less complex network structures can be a good way of capturing known higher-level mechanisms, such as the activity of a complete pathway, even though not all molecular mechanisms are understood. This is especially true for models of protein production and the protein secretion machinery where many details are still unidentified. For example, Wiseman et al. (2007) used a simplified treatment of the endoplasmic reticulum pathways for protein folding, degradation, and export to study their contributions to protein homeostasis and protein export efficiency. Similarly, the intricate details of the pathways of the unfolded protein response (Curtu and Diedrichs, 2010) were condensed into a minimal model featuring the basic mechanisms (Trusina et al., 2008; Trusina and Tang, 2010). Despite the simplified treatment the model could provide insight into the function of this homeostatic-restoring system, in particular in addressing the differences between yeast and mammalian cells and the role of translation attenuation. An even simpler, but nevertheless very useful, model of recombinant protein secretion in *Pichia pastoris* was presented by Pfeffer et al. (2011). This model is unique in that it was able to quantify the degree of intracellular protein degradation under production like conditions. A study addressing sustained oscillations in continuous yeast cultures is yet another example of successful modeling using a relatively simple model structure (Heinzle et al., 1982). At the extreme end of simple network structures there are of course also the so-called unstructured models which only use a single state variable to describe the cell biomass in addition to a few state variables accounting for extracellular substrates and products (Menezes et al., 1994; Portner and Schäfer, 1996; Carlsen et al., 1997; Ensari and Lim, 2003; Sarkar and Modak, 2003; Liu and Wu, 2008; Yüzeç et al., 2009). Such models are for the most part just phenomenological representations of what is empirically observed. An exception is a type of unstructured kinetic models that are derived using prior knowledge of intracellular reactions; based on a stoichiometric description of a metabolic network, a set of macroscopic reactions connecting the extracellular substrates and products are determined by decomposing the network into its elementary flux modes (Provost and Bastin, 2004; Haag et al., 2005; Provost et al., 2006; Teixeira et al., 2007; Dorka et al., 2009; Zamorano et al., 2013).

A drawback of all the less detailed network structure approaches mentioned above is the missing or complicated links between entities of the model and the actual molecular entities inside the cell. These links are particularly important if the model is to be used for identification of explicit targets for strain improvement by genetic engineering. However, depending on the purpose of modeling, a model with a simpler structure may still be useful. It can for instance foster a better general understanding of the system behavior or give

insights of the system that can be used as a starting point for further detailed modeling. A less detailed model can also be used for making predictions without explicit reference to the underlying, un-modeled reactions at the molecular level. For instance, to accurately simulate the concentration profiles of substrate, product, and biomass during a fermentation, which may be valuable for process design, a simple unstructured model may be sufficient. Thus, in situations where a simple model structure is believed to meet the requirements of the modeling purpose, nonessential details should be avoided since they will only make the modeling process unnecessary cumbersome.

2.2.1. Representation of network structure

The goal of the model network structure is the collection of all necessary and available biological information that will be converted into a mathematical representation. However, the network structure also serves as a basis for discussion between biologists and engineers, physicists or mathematicians. The graphical representation is therefore an important aspect of the model network structure. An accurate and standardized visual language facilitates the communication between researchers, especially for those with different backgrounds, and it rationalizes the interchange of models and biological knowledge, reducing the risk of misunderstandings and ambiguity. The Systems Biology Graphical Notation (SBGN) (Le Novère et al., 2009) was developed by members of the systems biology community to address these issues and is now emerging as a standard for graphical notation. The use of SBGN in biochemical modeling was recently reviewed by Jansson and Jirstrand (2010). Tools for visualization of model simulation results, arranged in the form of a network structure map, have also been developed (Oldiges et al., 2006; Noack et al., 2007).

2.2.2. Kinetic rate expressions

The kinetic rate expressions are the symbolic expressions that describe the reactions and interactions between the elements of the network structure. Determination of the numerical values of the parameters occurring in them are discussed later. A fundamental type of reaction kinetics is the so-called law of mass action. It states that the reaction rate is proportional to the concentrations of the reactants, or the reactant for a unimolecular reaction, and it is frequently used as a description for elementary reactions (reactions

with one step). Kinetics of multi-step reactions, such as those of enzymes and transporters, can be derived by combining the mass action kinetics of their elementary reactions (Goryanin and Demin, 2009). The resulting dynamical systems are usually simplified based on time-scale considerations (Klipp et al., 2005a; Almquist et al., 2010a), or on symmetries, such as the commonly used assumption of identical and independent behavior of ion channel subunits (Almquist et al., 2010b). Typically the simplification is done to the point where the internal dynamics of the reaction process is lost, and the description has reduced to an explicit function of the reactants and any effectors. The reduction also means that many of the parameters appearing in the final rate expressions are aggregates of elementary reaction parameters and therefore do not always have the same type of biochemical interpretability. An example of a well-known rate expression derived from elementary reactions is the Michaelis–Menten kinetics. It is obtained by separation of slow and fast dynamics and it is usually used to describe enzyme kinetics where the concentration of substrate is much higher than the concentration of the enzyme. A thorough treatment of the Michaelis–Menten approximation and its connection to the underlying dynamic system of elementary reactions was recently presented by Chen et al. (2010).

Determination of kinetic rate expressions is complicated by the fact that mechanisms of enzymes, transporters, and other complex biochemical reactions are often unknown (Costa et al., 2011). In those cases where reaction mechanisms have been derived through careful experimental studies, detailed modeling of the different reaction steps can produce rate expressions with complicated symbolic forms and large numbers of associated parameters (Goryanin and Demin, 2009), making subsequent model analysis and parameter determination difficult tasks. It must also not be forgotten that all kinetic rate expressions, no matter how comprehensive in their details, are just models. They have limitations in their applicability, they may be incomplete, or even incorrect. For instance, the experimental conditions under which a rate expression was established may differ from those of the living cell being modeled, making the kinetics inappropriate. In addition, reaction rates will to different degrees of extent depend on variables that were not considered in the derivation, such as pH, temperature, ionic strength, or the cooperative effect of enzyme effectors.

Box 3—Approximative kinetic formats.

Generalized mass action (GMA) describes reactions by power law kinetics with non-integer exponents (Savageau, 1976). GMA allows an analytical steady-state solution to be calculated for linear pathways.

S-systems also use power laws kinetics but here the individual reaction rates are aggregated into two reactions for every mass balanced biochemical species (Savageau, 1976). This approximation makes analytical solutions of steady-states possible also for branched pathways, but at the risk of introducing large errors and unrealistic results in certain situations (Heijnen, 2005).

Log-linear kinetics approximates reaction rates with a linear expression of logarithmic dependencies on reactants and effectors (Hatzimanikatis and Bailey, 1996, 1997). However, the enzyme concentration appears among the linear terms and the reaction rate is thus not proportional to the enzyme concentrations, something that is generally observed.

Lin-log kinetics (Visser and Heijnen, 2003; Heijnen, 2005) is also a linear expression of logarithms but with the difference that the enzyme concentration is a multiplicative factor to this linear sum, giving a reaction rate that is proportional to enzyme concentration. Like the power law approximations of GMA and S-systems, the log-linear and lin-log approaches enable analytic solutions of steady states. However, unlike the scale-free power laws, their concentration elasticities go towards zero for high concentrations, which is in agreement with the downward concave behavior of most enzymes' kinetics (Heijnen, 2005).

Convenience kinetics is a generalization of Michaelis–Menten kinetics that covers arbitrary reaction stoichiometries (Liebermeister and Klipp, 2006a). It can be derived from a non-ordered enzyme mechanism under the assumption of rapid equilibrium between the enzyme and its substrates and products. The convenience kinetics differs from the above rate laws in that it is saturable and can handle concentrations that are equal or close to zero, the latter situation being known to cause problems for kinetics containing logarithmic functions (Wang et al., 2007; del Rosario et al., 2008). It has also been described how to avoid violating the laws of thermodynamics by using thermodynamically independent system parameters (Liebermeister and Klipp, 2006a).

Modular rate laws is a family of different rate laws which were presented with an emphasis on thermodynamical correctness (Liebermeister et al., 2010).

Box 4—Stochastic kinetics.

Models with stochastic reaction kinetics can be based on either discrete or continuous state spaces. In a discrete stochastic model, the state of the system corresponds to the exact numbers of different types of molecules. Since it is impossible to predict the individual reactions changing the state of the system, the system must instead be described by the probability of being in each possible state. Knowing the transition probabilities between states, referred to as the reaction propensity, the time evolution of the probabilities for the different states can be described by a differential equation known as the master equation. Because of the large number of possible states even for the most simple biochemical systems it is not feasible to solve the master equation in most practical applications. What can be done, however, are (repeated) realizations of the stochastic process described by the master equation using the stochastic simulation algorithm (Gillespie's algorithm) (Gillespie, 1976), or extensions of it such as tau-leaping (Gillespie, 2001).

Another strategy to deal with the discrete stochastic process of the master equation is to approximate it by a continuous stochastic process. This is typically done by the use of stochastic differential equations known as Langevin equations, enabling simulations that are more efficient (Higham, 2001; Adalsteinsson et al., 2004). Although Langevin equations can be rigorously derived to approximate the discrete stochastic process described by the master equation (Gillespie, 2000; Lang et al., 2009), they can also be used to introduce randomness to an ordinary differential equation in an ad hoc manner (Hasty et al., 2000; Ghosh et al., 2012). The continuous process described by a Langevin equation can also be expressed by the corresponding deterministic partial differential equation for the dynamics of the probability distribution, the Fokker–Planck or Kolmogorov forward equation (Jazwinski, 1970; Gillespie, 2000; van Kampen, 2007).

2.2.3. Approximate kinetic rate expressions

Since most kinetic rate expressions are unknown, and because of the complexity and unreliability of those who are claimed to be known, a number of different approximative kinetic rate expressions have been suggested as alternatives. These rate expressions have in common that their symbolic structures are intended to be simple but yet flexible enough to describe many types of reaction kinetics. They aim for a small number of parameters to facilitate parameter determination, and some of them are designed to have good analytical properties or to guarantee correct parameterization from a thermodynamical point of view. Because of their standardized formats they simplify the modeling-building process, also encouraging automatic construction of kinetic models (Liebermeister and Klipp, 2006a; Borger et al., 2007; Adiamah et al., 2010; Liebermeister et al., 2010). Some of the approximative rate expressions used in kinetic modeling (generalized mass action, S-systems, log-linear, lin-log, convenience kinetics, and modular rate laws) are briefly described in Box 3.

The use of approximative rate expressions have been compared both to other approximative rate expressions as well as to traditional mechanistic formulations of reaction kinetics in a number of modeling studies. For example, a lin-log model (Visser et al., 2004) was derived based on a already established mechanistic model of the central carbon metabolism in *Escherichia coli* (Chassagnole et al., 2002), and was found to give similar simulation results despite its simpler structure and fewer parameters. In three parallel models of sphingolipid metabolism in yeast (Alvarez-Vasquez et al., 2004), the power law formats, GMA and S-systems, were compared to Michaelis–Menten kinetics. It was found that the models behaved similarly both with respect to steady states and dynamics responses. The performance of GMA, convenience kinetics, and Michaelis–Menten kinetics was compared in a number of model variants describing the biosynthesis of valine and leucine in *Corynebacterium glutamicum* (Dräger et al., 2009). Hybrid models consisting of both approximative kinetics and mechanistic kinetics have also been evaluated and concluded to be suitable approaches (Bulik et al., 2009; Dräger et al., 2009; Costa et al., 2010).

2.2.4. Stochastic kinetics

A deterministic formulation of reaction kinetics will gradually lose its validity as the number of reacting molecules becomes small. As a rule of thumb, there should be at least 10^2 – 10^3 molecules per reactant (Chen et al., 2010) when describing reactions with deterministic models. Metabolic reactions, the most commonly modeled aspects of cell factories, typically fulfill the requirements for deterministic modeling. However, low numbers of reacting molecules and stochastic

behavior can occur in for instance signaling (Wang et al., 2006), gene expression (Paulsson, 2004), and protein secretion (Love et al., 2010), processes potentially relevant in cell factory applications. Modeling of these and other processes is therefore in some cases best done using stochastic approaches that take the randomness of biochemical reactions into account (Ullah and Wolkenhauer, 2010). Such simulations have for instance been used for models of *S. cerevisiae* to study the GAL network (Ramsey et al., 2006), and the Ras/cAMP/PKA signaling pathway (Cazzaniga et al., 2008) including the nucleocytoplasmic oscillations of the downstream transcription factor Msn2 (Gonze et al., 2008). For more details on stochastic kinetics see Box 4.

2.3. Parameter determination

Parameters in kinetic models are essentially determined in two different ways; either one at a time, considering the different components and processes of the model individually, or by collectively calibrating the parameters to make the model fit measurements of the intact system. The two approaches are often combined by setting some parameters to previously known or measured values while simultaneously fitting the remaining ones (Zi et al., 2010).

Following the first approach, there are studies where the model building process has been complemented by experimental work aiming to measure parameter values directly (Teusink et al., 2000), but more commonly parameters are set to values already reported in the literature (Alvarez-Vasquez et al., 2004). These values can sometimes be found in databases compiling experimental information on kinetic parameters (Kanehisa and Goto, 2000; Rojas et al., 2007; Schomburg and Schomburg, 2010; Scheer et al., 2011). A serious problem with this approach is that it usually means that parameter values will have to be collected from different sources, involving different experimental conditions, different physiological states of the cells, different strains, or even different organisms (Costa et al., 2011). Notably, it is also common that such parameter values are derived from in vitro measurements, where conditions may differ drastically to those of in vivo systems (Minton, 2001, 2006), an approach which has been shown to have shortcomings even if great care is taken (Teusink et al., 2000). The above issues are being tackled by the development of standardized experimental systems imitating in vivo conditions for specific organisms or cell types (van Eunen et al., 2010). Sometimes model parameters are determined in even less accurate ways, for instance according to rule of thumb-like considerations such as using generic rate constants for protein–protein associations or by educated guessing of enzyme K_m values (Hoefnagel et al., 2002). Finally, there are many parameters whose values cannot be determined directly due to the limitations of experimental techniques.

The alternative to determining parameters one by one is to collectively calibrate the parameters to make the model reproduce experimental measurements of other quantities than the parameter themselves. This way of indirectly determining parameters is referred to as parameter estimation (but also as system identification, model fitting, or model calibration). The parameter estimation problem can be seen either as the geometrical problem of minimizing the distance between the model output and the corresponding experimental data, or it can be interpreted statistically as the problem of maximizing the likelihood of observing the data given a model that takes the experimental uncertainty into account. It can be shown that these views on parameter estimation are related. Specifically, when the geometrical approach uses a (weighted) sum of squares as the distance measure it is equivalent to when the statistical approach model

measurement errors as additive, independent, and normally distributed. For more details on how the parameter estimation problem is formulated see [Box 5](#). Some of the challenges of parameter estimation include large qualitative and quantitative uncertainties faced in biological systems, and parameter estimation for large-scale models. In these cases, it is common that multiple sets of parameter values can make the model reproduce the measurements. When the lack of sufficient information in experimental data results in a population rather than in a unique set of parameter values, an alternative to conventional parameter estimation methods might be more appropriate ([Miskovic and Hatzimanikatis, 2010](#); [Soh et al., 2012](#); [Chakrabarti et al., 2013](#)). In this approach, the space of admissible parameter values is first reduced by applying physicochemical and thermodynamic constraints integrated with available measurements.

Box 5—Formulating the parameter estimation problem.

The parameter estimation problem can be formulated as the following minimization problem. Consider N measured data points, $\mathcal{D}_N = d_1, \dots, d_N$, taken at time points t_1, \dots, t_N , which are described by a scalar-valued model output, $y(t)$ (at the expense of a little more notation the line of though easily extends to the case with vector-valued outputs, see for instance [Raue et al., 2009](#)). Now an objective function $V(\theta)$ can be defined for some distance measure of the vector of residuals, $[d_1 - y(t_1, \theta), \dots, d_N - y(t_N, \theta)]$. For instance, using a weighted sum of squares as a measure of the distance, the objective function, $V_{SS}(\theta)$, becomes

$$V_{SS}(\theta) = \sum_{i=1}^N \frac{(d_i - y(t_i, \theta))^2}{\sigma_i^2} \quad (7)$$

where σ_i^2 is the weight for the i th data point. The parameter estimate, $\hat{\theta}$, is then the set of parameters that minimizes $V_{SS}(\theta)$

$$\hat{\theta} = \arg \min_{\theta} V_{SS}(\theta). \quad (8)$$

The parameter estimation problem can also been seen from a statistical view point, treating experimental observations as realizations of random variables ([Ljung, 1987](#)). If the model is assumed to be a perfect description of the system, the deviation of each observed data point, d_i , from the model prediction, $y(t_i)$, must originate from a measurement error, ϵ_i , here assumed to be of additive nature

$$d_i = y(t_i) + \epsilon_i. \quad (9)$$

By changing the model of the outputs in Eq. (3) to

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t), \theta) + \epsilon, \quad (10)$$

the observed data can at any time point be seen as a deterministic part, as previously, plus the realization of the random numbers in the vector ϵ . If the measurement errors are assumed to be independent and normally distributed, with zero mean and variance σ_i^2 for the i th data point (again considering a scalar-valued model output), the likelihood of observing \mathcal{D}_N given θ , $\mathcal{L}(\theta)$, can be written as

$$\mathcal{L}(\theta) = c \prod_{i=1}^N \exp \left[-\frac{(d_i - y(t_i, \theta))^2}{2\sigma_i^2} \right] \quad (11)$$

where c is a constant not affecting the optimum of the likelihood function. The parameter vector $\hat{\theta}$ that maximizes $\mathcal{L}(\theta)$ is called the maximum likelihood estimate. Using the fact that the logarithm is a strictly monotonically increasing function, the problem of maximizing $\mathcal{L}(\theta)$ with respect to θ can be replaced with the problem of minimizing the negative logarithm of the likelihood function

$$-2 \ln \mathcal{L}(\theta) = -2 \ln c + \sum_{i=1}^N \frac{(d_i - y(t_i, \theta))^2}{\sigma_i^2}, \quad (12)$$

making the optimization problem equivalent to the sum of squares minimization described in Eq. (8). Therefore, the geometrical approach using a weighted sum of squares as discrepancy measure will coincide with the statistical approach if measurement errors are independent and normally distributed. More generally, any conceivable model of the measurement error like the one used here will correspond to some kind of distance measure of the vector of residuals.

The likelihood function above describes the probability of observing the data \mathcal{D}_N given the parameters θ . It is also possible to treat the parameters themselves as random variables ([Ljung, 1987](#); [Secrier et al., 2009](#)). Using Bayes' rule, the probability density function for the parameters given the data, $p(\theta|\mathcal{D}_N)$, or the posterior, can be written as

$$p(\theta|\mathcal{D}_N) = \frac{p(\mathcal{D}_N|\theta)p(\theta)}{p(\mathcal{D}_N)} \propto p(\mathcal{D}_N|\theta)p(\theta) \quad (13)$$

and the parameter set maximizing $p(\theta|\mathcal{D}_N)$ is called the maximum a posteriori estimate. The posterior distribution is a combination of the likelihood (of observing \mathcal{D}_N given the parameters) and any prior knowledge of the parameters. Prior knowledge could for instance come from typical distributions of similar parameters, or from previous estimates which did not include the data used for the likelihood. If there is no prior information about parameter values, i.e., the prior is a uniform distribution whose logarithm adds nothing but a constant to the objective function, the maximum a posteriori estimate is reduced to the maximum likelihood estimate. The Bayesian approach with maximum a posteriori estimation has for example been applied to a model of the threonine synthesis pathway ([Liebermeister and Klipp, 2006b](#)).

Then, the reduced solution space is sampled using Monte-Carlo techniques to extract a population of alternative sets of parameter values.

2.3.1. Computing the estimate

When an objective function describing a model's ability to reproduce the experimental data have been formulated – be it a likelihood function based on a probabilistic model of model prediction errors, or some other function – the parameter estimate is obtained by locating its optimum. This is accomplished by different ways of iteratively searching through the parameter space, usually taking constraints on admissible parameter values into account, and a large number of different optimization algorithms have been designed for this task, see Box 6. However, the problem is complicated by the fact that most models of biological systems contain nonlinearities and many of these models have large number of parameters to be estimated. A high dimensional parameter space in combination with strong nonlinearities can result in complexly shaped objective functions with many local optima. Such multimodality makes it hard to assess whether the global solution to the optimization problem has been located or if only a local optimum has been found. Adding further to the problem are the often vast and relatively flat parts of the parameter space, which only shows a weak response in the objective function (Transtrum et al., 2010) and consequently may delay the convergence of the search. As the objective function is not given as an explicit function of the model

parameters, its values for a certain parameter vector must be determined by solving the model equations. Every iterate of the optimization algorithm therefore requires one or more evaluations of the model equations and the majority of time spent on computing the estimate is typically used for integrating ODEs (Chou and Voit, 2009). The main challenges when optimizing the objective function are thus to locate the global optimum, and doing this in reasonable time.

2.3.2. Identifiability analysis and experimental design

An important but sometimes overlooked aspect of parameter estimation is the level of confidence in the obtained estimates and whether it is possible at all to uniquely assign values to the parameters (Cedersund, 2006; Gutenkunst et al., 2007b; Ashyraliyev et al., 2009; Roper et al., 2010; Raue et al., 2011; Erguler and Stumpf, 2011; Meshkat et al., 2011; Hattersley et al., 2011). To accurately estimate parameters requires a balance between the information content in the experimental data and the complexity level of the model. However, it is widely acknowledged that kinetic models often are overparameterized and too complex in their structures in relation to available quantitative data (Nikerel et al., 2006, 2009; Schmidt et al., 2008; Sunnåker et al., 2010; Schaber and Klipp, 2011). Some models have intrinsic symmetries that allow transformations of state variables and parameters in a way that does not change the model output. Such redundant parameterization leads to a likelihood function that instead of a unique minimum has a completely flat valley, meaning that there

Box 6—Optimization.

Two main categories of optimization methods can be distinguished, so-called local and global methods. Local methods require some kind of initialization of parameters, a position in the parameter space from where to start the optimization. This parameter set can come from in vitro measurements of reaction kinetics or other kinds of estimates, perhaps reported in the literature, but may also require guessing. The initial parameter set is then improved by repeated application of the optimization algorithm. Many local methods determine their direction of search in the parameter space based on the gradient and Hessian of the objective function at the present point in parameter space (Nocedal and Wright, 1999). The Newton method uses the exact Hessian, but quasi-Newton methods approximating the computationally costly Hessian using gradients, like the SR1 or BFGS algorithms, are more commonly used. For least squares problems, which are the most common in biochemical modeling, the Hessian approximation of the Gauss–Newton and Levenberg–Marquardt (Marquardt, 1963) methods are especially appropriate (Nocedal and Wright, 1999). The gradient of the objective function needed by these methods are typically computed by finite difference approximations. However, numerical solutions of the model equations using adaptive step length ODE solvers are known to introduce "quantification errors" to the objective function, making it non-smooth on small scales (Bohlin, 2006; Carlsson and Nordheim, 2011). The finite difference approximation may thus become an unreliable description of the gradient and gradient-based methods can as a consequence experience difficulties. To overcome such problems the gradient can instead be determined by integration of the so-called sensitivity equations (Ljung and Glad, 1994a; Skaar, 2008; Carlsson and Nordheim, 2011). Another strategy of handling issues with non-smooth objective functions is the use of non-gradient based methods like the Nelder–Mead method (Nelder and Mead, 1965), the Hooke–Jeeves method (Hooke and Jeeves, 1961), or the principal axis method (Brent, 1973). Although such methods are robust and easy to implement, they generally have much slower convergence in terms of the number of objective function evaluations.

Since the objective function typically has several local optima the choice of initial values is crucial for finding the global optimum using local methods. The inefficiency of local methods in finding the global optimum (Mendes and Kell, 1998; Moles et al., 2003) has spurred the development of global optimization methods that search the parameter space more comprehensively. A common drawback with these algorithms is a slower rate of convergence. Some of the popular global methods include simulated annealing (Kirkpatrick et al., 1983; Nikolaev, 2010), a large number of different genetic and evolutionary algorithms (Sarkar and Modak, 2003; Yüzgeç et al., 2009; Chou and Voit, 2009; Ashyraliyev et al., 2009), and particle swarms (Kennedy and Eberhart, 1995), and their performance has been compared in several studies (Moles et al., 2003; Dräger et al., 2009; Baker et al., 2010).

Most successful is the combination of local and global search methods. Such hybrid methods benefit both from the global methods' ability to explore the parameter space and from the faster convergence rate of the local methods once close to a (local) optimum. As an example, the results obtained by Moles et al. (2003) using the global SRES method (Runarsson and Yao, 2000, 2005) were substantially improved by different combinations with local methods (Rodríguez-Fernandez et al., 2006b), and further strengthened by a systematic strategy for when to switch from the global to the local method (Balsa-Canto et al., 2008). Even more promising results have been obtained with a hybrid approach based on a scatter search metaheuristic (Rodríguez-Fernandez et al., 2006a). An enhanced version of the scatter search (Egea et al., 2010) has also been shown to benefit from a cooperative parallelization (Balsa-Canto et al., 2012), as illustrated in a comparison with a non-cooperative parallelization of the algorithm on the parameter estimation problem of the 193 parameter *E. coli* model by Kotte et al. (2010).

Several of the local, global, and hybrid methods mentioned above are available through modeling software tools like SBML-PET (Zi and Klipp, 2006), the Systems Biology Toolbox (Schmidt and Jirstrand, 2006; Schmidt, 2007), COPASI (Hoops et al., 2006; Mendes et al., 2009), PottersWheel (Maiwald and Timmer, 2008), and AMIGO (Balsa-Canto and Banga, 2011).

are several parameter sets that are equally likely to have produced the measured data. Models of this type are said to be structurally unidentifiable (Bellman and Åstrom, 1970; Pohjanpalo, 1978). It should be emphasized that this property is only dependent on the model structure itself, including the set of measured model outputs and the known input variables, but not on the quality or quantity of data used for estimation. The analysis of structural identifiability can therefore be done a priori, meaning that neither experimental data, nor a certain parameterization, is required.

Structural identifiability is a necessary condition for an unambiguous estimation of parameters. It may however not be sufficient because it can happen that even though the likelihood function has a unique minimum for some parameter set, the surroundings of this minimum could be very flat. Consequently there may be other parameter sets with potentially very different values that are almost as likely. Such diverse parameter sets yielding very similar outputs have for example been observed in a model of the methionine cycle dynamics (Piazza et al., 2008) and in a model of monoclonal antibody production in Chinese hamster ovary (CHO) cells (McLeod et al., 2011). This situation is referred to as a lack of practical identifiability. Unlike structural identifiability, this property does depend on the amount, quality, and time points of experimental observations. Methods for determining practical identifiability also require that a parameter estimate has already been obtained, and can therefore not be

applied a priori. A review of methods for identifiability analysis is found in Box 7.

When estimating model parameters from experimental data, decisions have to be made about what kind of experiments to perform. It is rare that all state variables can be measured and typically there are several quantities appearing in the model for which experimental methods exist but come at a high cost in terms of time- or resource-consumption. In these situations, identifiability analysis can be a useful tool to guide the experimental design. For instance, the structural identifiability of a model depends on the set of model outputs but it is not only interesting to know whether a particular set of measured outputs renders the model identifiable but it is also of great interest to learn which potential sets of outputs that have to be measured in order to ensure structural identifiability. Addressing this question, an algorithm was developed in the group of Jirstrand and colleagues (Anguelova et al., 2012) that a priori finds so-called minimal output sets, which are sets of outputs that when measured results in an identifiable model. The algorithm has been implemented in Mathematica (Wolfram Research, Inc., Champaign, USA) and used successfully in the analysis of models with over 50 parameters (Anguelova et al., 2012). Since methods that only determine structural identifiability will not be able to detect practical identifiability, they can never be used to prove the feasibility of a certain experimental design. Rather, because approaches like the

Box 7–Identifiability analysis.

One algorithm for determining structural identifiability has been presented by Sedoglavic (2002) which is particularly interesting. Unlike previous efforts (Vajda et al., 1989; Audoly et al., 2001; Margaria et al., 2001) this method does not suffer from the limitation of only being applicable to smaller systems. In fact, a recent implementation of the algorithm, which was also extended to handle parameterized initial conditions, has been successfully applied to models with a size of about 100 state variables and 100 parameters using a standard desktop computer (Karlsson et al., 2012). The results obtained by Sedoglavic are so far unfortunately not disseminated in the biological modeling community, one of the reasons perhaps being the use of the related term observability instead of the, in the biological field, more common term identifiability. It should be noted that this method, and all other methods based on the so-called rank-test, are testing for so-called local structural identifiability. Thus, these methods will identify redundant parameterizations that correspond to completely flat and continuous regions in the likelihood function but there may still be an enumerable set of non-neighboring single points in the parameter space, also resulting in identical model output, which are not detected by this analysis. One situation, where multiple parameter sets are possible and where local structural identifiability analysis might be insufficient, is when measuring one or more components of a pathway containing an upstream reaction which is catalyzed by two or more isoenzymes whose concentrations and activities are not explicitly measured. If the different enzymes are described by the same type of model structure, permutations of concentrations and kinetic parameters for the set of isoenzymes results in models with identical output. The models themselves are however not identical because the different parameter sets have different implications when interpreting the properties and functions of the actual enzymes and their corresponding genes. Methods for the analysis of global structural identifiability exist (Ljung and Glad, 1994b; Bellu et al., 2007) but are typically only applicable to smaller systems with just a few state variables and parameters (Roper et al., 2010), or systems with a particular structure (Saccomani et al., 2010), and therefore so far of lesser interest in the analysis of most models addressed in this review. A notable exception is the successful application of the generating series approach to a medium-sized model of the NF κ B regulatory module (Chis et al., 2011). Though, potential issues with non-identifiability in the global sense could be eliminated if there is a priori knowledge about parameter values that can be used as a starting guess when computing the estimate or to discard an incorrect solution to the parameter identification problem.

A simple way of evaluating how accurately parameters can be identified in practice is to look at the standard parameter confidence intervals determined from a quadratic approximation of the log-likelihood function around its optimum. However, due to the frequent combination of limited amounts of experimental data and model outputs that depend non-linearly on the parameters, this type of confidence intervals can be unsuitable (Raue et al., 2011; Schaber and Klipp, 2011). Another way of assessing the accuracy of the parameter estimates is to use exact confidence intervals determined by a threshold level in the likelihood. A method to calculate such likelihood-based confidence intervals based on the profile likelihood was recently proposed (Raue et al., 2009, 2010). Here, all parameter directions of the likelihood function are explored by moving along the negative and positive directions of each parameter while minimizing the likelihood function with respect to the remaining parameters (which means that one studies the projection of the likelihood onto a specific likelihood-parameter axis plane). The confidence intervals are determined by the points where these likelihood profiles cross over a certain threshold, and the confidence levels are determined by the level of that threshold. If the profile likelihood for a parameter never reaches the threshold in either the negative or positive direction, or in neither, the confidence interval of this parameter extends infinitely in at least one direction. According to this approach, parameters with unbounded confidence intervals are defined as non-identifiable. This definition would make no sense for confidence intervals determined from the likelihood curvature at the point of the estimate, since these are always finite (with the exception of a completely flat likelihood resulting from a structural non-identifiable parameter). Profiling can also be applied to posterior distributions (Raue et al., 2013).

minimal output sets do not require any wet-lab efforts at all, the appropriate use of such structural identifiability analysis is to beforehand disprove any experimental design that is bound to fail in identifying the parameter values, and to give well-founded suggestions of which additional quantities that have to be measured to resolve the identifiability issues. Insights obtained in this way can potentially save a lot of valuable laboratory resources. The analysis of practical identifiability will on the other hand require an existing set of measurements, but it can not only determine which parameters that are impossible to estimate uniquely but also those that are too poorly constrained. This type of analysis is therefore able to confirm if a given set of measurements really is sufficient for parameter identification in practice. If this is not the case, and additional measurements are required, practical identifiability analysis can be used to improve the experimental design in more specific ways than methods like the minimal output sets, for instance by indicating certain time points at which the measurement of a particular quantity is most efficient in (further) constraining a parameter value (Raue et al., 2009, 2010). Thus, structural and practical identifiability analysis fulfills different needs and can be said to have complementary roles when used for experimental design.

2.3.3. Model reduction

It was shown above that identifiability analysis can guide the experimental design so that the correct type and amount of data required for system identification is collected. Another way of achieving the balance between model and data is to decrease the complexity of the model by different model reduction techniques. These techniques aim at simplifying models to reach an appropriate level of detail for experimental validation (Klipp et al., 2005a), and if done properly the reduced model retains the essential properties of the original model. Model reduction can also be performed on models where the parameters have already

been identified and whose applicability has been validated. In these cases the purpose of the reduction is to facilitate the understanding of essential structures and mechanisms of the model and to decrease the computational burden of simulation and analysis. Methods for model reduction are discussed in Box 8.

In addition to the more formal methods mentioned in Box 8, a lot of model reduction is often done by the modeler already when setting up the network structure and formulating the rate expressions. For instance, different post-translationally modified versions of a protein might be described by a single lumped state variable, concentrations of co-factors might be excluded as state variables and consequently not considered in the rate expressions of reactions in which they participate, reactions which are thought to be marginally relevant for the problem at hand might be left out from the model, known rate expression might be simplified and described by approximate kinetic formats as explained previously, and quantities that are changing slowly in the characteristic time scale of the model, such as the synthesis and degradation of enzymes during a much faster metabolic process, may be considered frozen and hence set constant. Decisions like these are usually dependent on a combination of the purpose of the model, the modelers experience and intuition, and prior knowledge of the modeled system.

2.4. Validation

Before a model is ready to be used its quality should be established. This is done not only by evaluating the model's ability to explain the experimental data used for parameter estimation but also by comparing some of its predictions to new data that was not used earlier in the model building process (Ljung, 1987). If a priori information is available on values of parameters with a biophysical interpretation, these should be compared to the estimated values as a feasibility check. Additionally, other aspects of the model, such as the predictions of unobserved state variables,

Box 8—Model reduction.

Two popular categories of model reduction methods are the ones based on time-scale separation and lumping. The time-scale separation approach is based on defining a time-scale of interest and neglecting changes in state variables that occur on slower time-scales and approximating state variables and processes associated with faster times-scales using the quasi-steady-state and the quasi-equilibrium approaches (Klipp et al., 2005a; Nikerel et al., 2009). Thus, the dynamics of some state variables will be replaced by either constants or algebraic relations. If the time-scales of the reactions in a system are not known, several reduced versions of a model may be considered (Almquist et al., 2010a) or further assumptions could be made (Almquist et al., 2010b). Lumping, on the other hand, transforms the original state variables to a set of new state variables in a lower dimensional state space (Okino and Mavrovouniotis, 1998). The choice of which state variables to lump together is frequently based on time-scale considerations, which results in groups of quickly equilibrating state variables being completely eliminated and replaced by a new state variable. One example of model reduction through lumping can be found in a study of secondary metabolism pathways in potato (Heinzle et al., 2007). Here, the steady-state assumptions which were used to motivate the lumping of different metabolites were derived from experimental work. Even though model reduction through lumping and time-scale separation often overlap, this is not always the case. Examples of time-scale separation not involving lumping include setting slowly varying variables to constant values, and examples of lumping not involving time-scale separation include mean concentration models of cellular compartments, i.e., reaction-diffusion equations represented without the spatial dimension. Other model reduction techniques include sensitivity analysis (Degenring et al., 2004; Danø et al., 2006; Schmidt et al., 2008) and balanced truncation (Liebermeister et al., 2005). The previously mentioned profile likelihood approach to practical and structural identifiability analysis can also be used for model reduction (Raue et al., 2009, 2010, 2011).

In most models with relevance for biotechnology the model components, such as state variables, their rates of change, and parameter values, have precise physical meaning. A successful model reduction should therefore not only preserve the input–output relations, which may be sufficient in other disciplines where models are used, but also preserve the interpretation of model components (Cedersund, 2006). These ideas are central in a recently developed method that reduce models by lumping (Sunnåker et al., 2010). Based on the approximation that state variables involved in fast reactions are in quasi-steady-state, interconnected groups of such quickly adjusting states are identified and lumped together. The distribution among the original states of a lump is determined analytically by so-called fraction parameters. These parameters can be used to retrieve the details of the original model, which is known as back-translation, thereby allowing better biochemical interpretation of analysis and simulations done with the reduced model. The method has also been extended to be able to handle nonlinear models and was successfully applied to a model of glucose transport in *S. cerevisiae* (Sunnåker et al., 2011).

may be interrogated with respect to their biological plausibility. Quality controls like the above are referred to as model validation. Strictly speaking, however, a model can never be validated. It may explain all experimental data generated so far but it can never be proven to correctly account for future experiments. What is meant by validation is rather that the model has withstood repeated attempts to falsify or invalidate it. The rationale here is that the more experiments that have been successfully explained by the model, and the more reasonable it is with respect to a priori information about the biological system, the more it can be trusted to correctly predict future experiments. If a model fails to pass the validation step, researchers need to revise their model by suitable iteration of the modeling steps outlined in Fig. 1.

The ability of a model to explain experimental data is frequently judged by visual inspection of the respective time-series (Heinzle et al., 2007; Li et al., 2011; Cintolesi et al., 2012) or by qualitative comparison of model characteristics (Gonzalez et al., 2001). A qualitative comparison may for instance involve an investigation of whether the model can produce certain observed behaviors such as oscillations, homeostasis, or switching. Such analysis is sometimes actually performed before parameters have been formally determined, typically using some initial estimate of the parameter values, which might result in models being discarded already at this point. While these less rigorous assessments may be a good first step of the validation procedure there are also formal statistical tests for determining the quality of a model, see Box 9. Regardless of the outcome of statistical tests and formal methods of validation, it should not be forgotten that these are best used as support for decisions made by the modeler (Cedersund and Roll, 2009) and that the ultimate validation is whether the model can fulfill the purpose for why it was created in the first place (Ljung, 1987).

Sometimes validation is done by qualitatively different types of data than what was used for model identification. For instance, the biological system can be measured under new external conditions (Shinto et al., 2007; Oshiro et al., 2009), resulting in a different operating point, new types of input schemes (such as steps, pulses, periodic pulses, or staircases) may be used (Klipp et al., 2005b; Zi et al., 2010), data can be collected on previously unmeasured molecular species, and validation experiments can be conducted on modified versions of the original system, i.e. mutants, where enzymes or other components are inactive, constitutively active, or have been underexpressed, overexpressed, or completely deleted

(Alvarez-Vasquez et al., 2005; Klipp et al., 2005b; Wang et al., 2006; Zi et al., 2010; Cintolesi et al., 2012). When models can successfully explain such new data, it is a strong indication that the mechanistic principles and assumptions behind the model are sound.

3. Using kinetic models

From a biotechnology perspective, a complete and validated model according to the steps outlined previously is usually not in itself the ultimate goal of modeling. The real value of a model lies instead in using it to predict, evaluate, and explore different scenarios or assumptions involving the modeled system and its surrounding environment. An established model should thus foremost be seen as a tool that can be used to answer questions about the cell factory and it should be used as a complement or alternative to performing actual experiments in the lab.

3.1. Improving production

A major question which has been attempted to be answered using kinetic models is how to rationally design directed metabolic engineering strategies that will improve a cell factory's ability to produce a desired product. This requires models that can predict the behavior of the cell in response to genetic alterations like gene deletion or overexpression. One way of using kinetic models to identify suitable targets is to perform a local parameter sensitivity analysis. A more thorough treatment of the problem involves simulating larger changes in the levels of enzymes and other components.

3.1.1. Local parameter sensitivity analysis

The aim of a local parameter sensitivity analysis is to determine the degree of change of some model property like a flux, a concentration, or a more complex quantity such as the area under the curve of some state variable, in response to a change in the model parameters. As the parameters may represent quantities that can be manipulated by genetic engineering, such as enzyme concentrations, the analysis provides predictive links between potential targets and their effect on the cell factory behavior. Since a local analysis only considers small or even infinitesimal perturbations around a point in parameter space, it is not intended to mimic any actual changes in, for example, an enzyme concentration. However, a

Box 9—Validation.

Model validation is typically done by analyzing the deviation between the measured data and the model outputs, $\epsilon_i = d_i - y(t_i)$. For a model to be good these residuals should be sufficiently small and uncorrelated. First of all, if the parameters have been collectively estimated, the model should be able to satisfactorily describe this 'training' data. For instance, the size of the residuals can be tested by a χ^2 test (Jaqaman and Danuser, 2006; Cedersund and Roll, 2009) and the correlation of residuals can be tested by a run test or a whiteness test (Cedersund and Roll, 2009). Secondly, the residual analysis should be performed also with new data that were not previously used. This is done to assure that a good fit is not just because a too complex model has been over-fitted to the particular data points of the estimation set. Validating a model with fresh data means that unless new data can be collected after parameter estimation, some data has to be saved. This can be a problem if there is not much data to begin with. A common approach to this situation is the use of resampling methods (Molinaro et al., 2005) where the model validation procedure is repeated and averaged over different partitions of the original data into training and validation sets. One such method is k -fold cross-validation, which has for instance been used in modeling of the TOR pathway (Kuepfer et al., 2007).

Model validation sometimes also involves comparison between competing models describing the same biological system, to see which one is 'most valid' (Schaber et al., 2012). Two common criteria used to find the most suitable model include the Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978), and two common tests that also address the statistical significance of model discrimination are the likelihood ratio test (Kreutz and Timmer, 2009; Cedersund and Roll, 2009) and the F-test (Jaqaman and Danuser, 2006; Cedersund and Roll, 2009). Other approaches to model discrimination, which included the dependence of model discrimination on experimental design, have been explored in studies on formate dehydrogenase production in *Candida boidinii* (Takors et al., 1997) and L-valine production in *C. glutamicum* (Brik Ternbach et al., 2005).

local parameter sensitivity analysis is easy to perform, gives a concise and transparent output, and despite its limitations it does have some predictive power allowing the results to be used as guidelines for identifying reasonable metabolic engineering targets.

A popular application of kinetic models is a special type of sensitivity analysis called metabolic control analysis (MCA) (Fell, 1992; Nielsen, 1998; Visser and Heijnen, 2002), the basis of which was already developed in the seventies (Kacser and Burns, 1973; Heinrich and Rapoport, 1974). It is concerned with the problem of quantifying how the control of steady-state flux is distributed among the enzyme-catalyzed reactions of a pathway. Two of the central quantities in MCA are the elasticity coefficients and the flux control coefficients (FCCs), defined in **Box 10**. They are both measures of sensitivity that have been scaled to obtain dimensionless numbers.

Normally there are many non-zero FCCs, meaning that there is no single rate-limiting enzyme but that the control of a flux instead is distributed over several reactions. However, it is likely that some reactions have larger values of their FCCs than others, indicating that these reactions are the ones primarily controlling the flux. The enzymes of those reactions may consequently be promising targets for successful metabolic engineering of the pathway. Given a kinetic model the FCCs can readily be calculated directly from its steady-state(s). The steady-state can be obtained either by simulations asymptotically approaching it, or by analytical or numerical solutions of the model equations. Alternatively, the FCCs may for linear pathways be determined indirectly from the summation and connectivity theorems using elasticity coefficients derived from the individual reaction rates of the model.

Sensitivity analysis in the form of MCA has been applied to a variety of kinetic models describing many different kinds of cell factories and types of products. It has for example been used to determine suitable genetic targets for improved production of lysine in *C. glutamicum* (Hua et al., 2000). This study found that lysine production was primarily controlled by the enzymes aspartokinase and lysine permease. The outcome of the analysis was verified experimentally by overexpression of aspartokinase, resulting in a significant increase in lysine production. However, the

lysine flux did not increase as much as would be expected from the sensitivity analysis, suggesting that model predictions of this type are best used as supporting guidelines and that they never should be taken as indisputable facts. Further verification of the model's predictive capability was obtained by overexpression of the low flux-control enzyme dihydridipicolinate, which only had a very limited effect on the production rate. Recently, Cintolesi et al. (2012) applied MCA to a model of ethanol production from glycerol in *E. coli*. Their analysis suggested that the control of the glycerol fermentation was almost exclusively shared between glycerol dehydrogenase and dihydroxyacetone kinase. The validity of this prediction was confirmed by the 2.4-fold increase in glycerol to ethanol flux observed when simultaneously overexpressing both enzymes. It was additionally seen that overexpression of other enzymes involved in glycerol metabolism, but whose flux control coefficients were close to zero, did not lead to increased rates of glycerol consumption and ethanol synthesis. The use of MCA is not limited to fluxes of metabolites but can be applied to the steady-state flux of any chemical entity. For instance, Gonzalez et al. (2001) used MCA to study monoclonal antibody synthesis in eukaryotic cells. They came to the conclusion that control of antibody production is shared between different steps of the synthesis pathway and that this division depends on the extracellular conditions and the physiological state of the cell. Their predictions were shown to compare qualitatively well with previously published experiments. Other examples of MCA applied to kinetic models include glycerol synthesis in *S. cerevisiae* (Cronwright et al., 2002), valine production in *C. glutamicum* (Magnus et al., 2009), the central carbon metabolism (Chassagnole et al., 2002) and production of threonine (Chassagnole et al., 2001), tryptophan (Schmid et al., 2004), and serine (Nikolaev, 2010) in *E. coli*, L-cysteine production in *Pseudomonas* sp. (Huai et al., 2009), production of lactic acid (Oh et al., 2011) and compounds of the acetolactate branch (Hoefnagel et al., 2002) in *Lactococcus lactis*, and the penicillin biosynthetic pathway in *Penicillium chrysogenum* (Theilgaard and Nielsen, 1999). Except for the work by Hoefnagel et al., which is further discussed in the next subsection, the model predictions of those studies were not tested by actually constructing the correspondingly modified

Box 10—Metabolic control analysis.

Consider a pathway, possibly containing branching points, consisting of metabolites x_i and reaction rates v_j which are catalyzed by enzymes with concentrations e_j . The elasticity coefficients (ECs) are then defined as

$$\epsilon_{x_i}^j = \frac{x_i}{v_j} \frac{\partial v_j}{\partial x_i}, \quad (14)$$

which means that for each reaction of the pathway there is a set of ECs measuring its sensitivity to the concentrations of the different metabolites. Each EC is a property of an individual enzyme and is therefore independent of the activity of the other enzymes in the pathway. For any steady-state flux J in the pathway the flux control coefficients (FCCs) are defined as

$$C_j^J = \frac{e_j}{J} \frac{\partial J}{\partial e_j}. \quad (15)$$

They quantify the degree of control exerted by the different enzymes on a steady-state flux of the pathway as a whole. This means that an FCC for one of the enzymes can depend on the properties of the other enzymes, and the FCCs are therefore system properties. The ECs and the FCCs are related by the summation theorem

$$\sum_j C_j^J = 1 \quad (16)$$

which states that the sum of all FCCs is 1, and by the connectivity theorem

$$\sum_j C_j^J \epsilon_{x_i}^j = 0 \quad (17)$$

which states that for each metabolite, the sum of the product of the FCCs and the ECs with respect to that metabolite is zero. The full details of MCA comprise additional sensitivity coefficients which are related through similar theorems. Thus, MCA is not just a sensitivity analysis but also a theoretical framework that formally describes the connection between properties of a system and its components.

microbes. There are also computational studies where MCA has been combined with parameter sampling approaches in order to examine the effect of parameter uncertainty (Pritchard and Kell, 2002; Wang et al., 2004; Wang and Hatzimanikatis, 2006a, 2006b; Miskovic and Hatzimanikatis, 2010).

Other types of local parameter sensitivity analysis are also abundantly represented in the literature. For example, Oshiro et al. (2009) determined the impact of parameter perturbations in a kinetic model describing the dynamics of lactic acid production in xylose fermenting *L. lactis*. Based on their results, the enzymes around the pyruvate node were proposed as targets for genetic manipulation. Some of these authors had also previously carried out a similarly designed sensitivity analysis of acetone–butanol–ethanol production in *Clostridium acetobutylicum* (strain N1-4, formerly known as *Clostridium saccharoperbutylacetonicum*) (Shinto et al., 2007). Here, several genetic engineering strategies for increased butanol production were suggested, including decreasing the activity of CoA transferase for butyrate and increasing the activity of the reverse pathway of butyrate production. This model (Shinto et al., 2007) was later improved by Li et al. (2011) who arrived at similar conclusions in their sensitivity analysis. In addition to looking at single parameters, their analysis also considered all combinations of parameter pairs. Though, as the size of parameter perturbations was small, the combined effects of simultaneously changing two parameters always equaled the sum of the separate parameter effects and no nonlinear crossover effects were thus found. Unfortunately did neither the *L. lactis* study nor the *Clostridium* studies genetically implement the proposed strategies. In recent work on CHO cells, McLeod et al. (2011) used sensitivity analysis to investigate which cellular process that controlled the production of a recombinant monoclonal antibody. The sensitivity analysis was repeated to specifically target different days of a two week fed-batch process. Unlike the modeling study by Gonzalez et al. (2001) it was found that control was divided almost exclusively between transcription, degradation, and translation of mRNA, and that this control structure did not change appreciably during the different phases of culturing. The authors consequently suggested that genetic engineering strategies for their system should focus on these processes, but the validity of their predictions was not tested experimentally.

3.1.2. Simulating larger changes

The theory behind MCA and other local sensitivity approaches is based on small perturbations of the parameter values and the resulting sensitivities are normally only valid in the vicinity of the nominal parameter values. Realistic cases of genetic manipulation will on the other hand likely involve larger changes in the levels of gene product concentrations. The extent to which the results of a local analysis of the model can be extrapolated to larger perturbations differs from case to case and cannot generally be determined (Visser et al., 2004; Schmid et al., 2004; Nikolaev, 2010). However, if a kinetic model has been formulated there is usually no reason for limiting the model analysis to local parameter sensitivities. Just as the control coefficients of MCA can be calculated directly from model simulations, the model can in principle be used for simulating any kind of perturbation of its components. By simulating more extensive changes to models, metabolic engineering scenarios can be explored in more realistic ways. In this respect, such approaches are more powerful compared to the traditionally used sensitivity analysis like MCA, and the predictions made have the potential to be much more accurate. Though, performing simulations that involve large changes in the model parameters, or even changes to the model structure, may require that the model has good predictive power not only for the specific

physiological setting for which it was developed but also for other operating points, something that cannot generally be assumed to be true. Therefore, the more extensive the perturbations to the model are, the more careful one should be when interpreting the results.

One example of model-based analysis of actual metabolic engineering strategies was provided by Hoefnagel et al. (2002). Based on MCA-derived candidate targets for increasing the production of acetoin and diacetyl in *L. lactis*, they proceeded with simulations of larger changes in the concentrations of two enzymes. First, a mutant with a lactate dehydrogenase deletion was simulated. This did indeed lead not only to a substantial flux towards the acetolactate synthase branch but also to a reduction in glycolytic flux, indicating potential problems with growth rates for such a strain. Then, a 40-fold overexpression of NADH-oxidase was simulated also resulting in some of the flux being diverted into the acetolactate branch. Finally, a simulation combining the two modifications was performed and it predicted that 92% of the flux through the pyruvate node would go into the desired direction, and that the glycolytic flux would be less affected. This fraction should be compared to a negligible 0.1% measured in the wild-type strain. The model prediction was tested, and at least to a certain degree confirmed, by an experiment which showed that 75% of the pyruvate ended up as acetoin in a strain where lactate dehydrogenase had been knocked out and NADH-oxidase was overexpressed. In another purely computational study, Chen et al. (2012) developed two separate kinetic models of glycolysis and the pentose phosphate pathway in *S. cerevisiae* and CHO cells. The authors then used the yeast model to analyze the impact of metabolic engineering targeting the production of dihydroxyacetone phosphate. Specifically, a deletion of the enzyme triose phosphate isomerase was simulated by setting its activity to zero. The rate of dihydroxyacetone phosphate production, and its yield on glucose, for this *in silico* deletion mutant was subsequently determined under different glucose uptake rates. Yet other studies have simulated the effects of realistically sized perturbations in the central metabolism of *E. coli* (Usuda et al., 2010; Kadir et al., 2010), comparing their results to experimental data.

3.1.3. Optimization problems

Even though the analysis of a specific metabolic engineering strategy is relatively easy to implement *in silico* given a kinetic model, there are still at least in theory infinitely many possible strategies to consider (assuming a continuum of expression levels) and it may be unclear which particular ones to try out in simulations. To overcome this difficulty, scenarios involving perturbations to model parameters are sometimes formulated as optimization problems (Hatzimanikatis et al., 1996a, 1996b; Mendes and Kell, 1998; Chang and Sahinidis, 2005; Pozo et al., 2011). A typical objective function to be optimized would be the rate of formation of the desired product and the optimization procedure may moreover be subject to constraints regarding the maximum changes in levels of enzymes and metabolites. All methods that can be used to compute the parameter estimates, described in Box 6, are typically applicable also for these problems. Using the output from kinetic models to set up optimization problems is perhaps the most rigorous and ambitious way of approaching the search for metabolic engineering targets.

Different optimization approaches to determining appropriate levels of metabolic enzymes have been used in a number of purely simulation-based studies for various aspects of microbial metabolism, including the production of ethanol in *S. cerevisiae* (Polisetty et al., 2008), citric acid in *Aspergillus niger* (Alvarez-Vasquez et al., 2000; Polisetty et al., 2008), and of serine (Visser et al., 2004; Vital-Lopez et al., 2006; Nikolaev, 2010), tryptophan (Marín-Sanguino and Torres, 2000; Schmid et al., 2004), and L-(–)-carnitine (Alvarez-

Vasquez et al., 2002) in *E. coli*. The simultaneous production of serine and tryptophan in *E. coli* has also been considered using a multi-objective optimization strategy (Lee et al., 2010).

3.2. Improving substrate utilization

An interesting prospect in the development of competitive cell factories is the expansion of their range of substrates. One example is the improved use of lignocellulosic biomass for ethanol production made possible by the introduction of genes for xylose utilization in *S. cerevisiae*. To assist in the evaluation of directed genetic engineering efforts towards improved efficiency in catabolism of this pentose sugar, Parachin et al. (2011) used a kinetic modeling approach for analyzing two different scenarios. One model was set up to represent a strain in which extracellular xylose reaches the pentose phosphate pathway through membrane transport followed by conversions by the enzymes xylose reductase, xylitol dehydrogenase, and xylulokinase. This model also featured a membrane transport reaction for the excretion of the intermediate metabolite xylitol. Additionally, another model was constructed for an alternative pathway comprising xylose isomerase and xylulokinase. In both these models the effects of a 10-fold overexpression as well as a severe knockdown (a 10-fold decrease in activity) of the different enzymes were examined. Simulating overexpression of xylose reductase in the first catabolic pathway slightly not only increased the ability to consume xylose but it also led to an increased excretion of xylitol. Conversely, not only the knockdown of xylose transport capacity decreased the xylitol formation but also the xylose consumption. Changing the activity of xylitol dehydrogenase, either by overexpression or knockdown, only had a marginal effect on model simulations. The best outcome was observed when xylulose kinase was overexpressed. This resulted in a slight increase in xylose consumption combined with a dramatic decrease in excretion of xylitol. Similarly, for the pathway using xylose isomerase, the model analysis suggested overexpression of xylulose kinase to be the best alternative for improving the utilization of xylose. The predictions of the models were in essence validated experimentally by aerobic and anaerobic cultivation of the correspondingly engineered yeast strains.

Other kinetic modeling contributions aiming for improved utilization of substrate includes enhancement of glucose uptake in *E. coli* (Visser et al., 2004; Nishio et al., 2008; Nikolaev, 2010) and application of the aforementioned MCA to the catabolism of L-arabinose (de Groot et al., 2005) and xylose (Prathumpai et al., 2003) in *Aspergillus nidulans* and *Aspergillus niger*. Of those studies, only Nishio et al. (2008) proceeded to validate their predictions in experimental follow-ups.

3.3. Improving product quality

For complex products such as glycoproteins, the quality of the product may be subject to improvement by genetic manipulations. The patterns of glycosylation have impact on in vivo activity, immunogenicity, and product half-life and their importance has encouraged the development of glycoengineered yeasts (Hamilton and Gengross, 2007; Ye et al., 2011; Nett et al., 2012), serving as an alternative to production of human-like glycoprotein in animal cells. Kinetic modeling has been employed to describe glycosylation in mammalian cells (Krambeck and Betenbaugh, 2005; Hossler et al., 2007). Expanding on an earlier model (Umaña and Bailey, 1997), Krambeck and Betenbaugh (2005) set up a model describing the non-linear kinetics of the enzymes involved in N-linked glycosylation in CHO cells. They were able to use their model to simulate how the glycosylation profile changes when the concentration of maturing protein in the Golgi increases. Specifically, they investigated a scenario where the concentration was increased 4-fold in order to represent a hypothetical cell line with

an increased specific productivity. The results of their simulations showed that the distribution of different glycoforms changes in response to the increased productivity, indicating a potential problem with reduced product quality. The authors then explored in silico the possibility of restoring the original glycosylation pattern in the high producer by means of changing the levels of glycosylation enzymes and the availability of uridine diphosphate *N*-acetylglucosamine. By just adjusting the level of a single enzyme, *N*-acetyllactosaminide α -2,3-sialyltransferase, more than half of the deviation could be reverted. Other solutions, involving changes in several targets, that almost completely restores the glycan distribution were also proposed based on the analysis. Models such as this are clearly interesting tools for making predictions of how to preserve correct glycosylation in high producing cells, but possibly also for how to engineer new glycan patterns.

3.4. Improving process design

In addition to predicting the effects of internal perturbations to a cell factory, kinetic models are also useful for predicting their behavior in response to various external conditions. Understanding the interplay between the cell and its environment is valuable since it can be used for improving the fermentation or bioreactor process. To describe the complete production process, a model which can reproduce cellular properties such as the rate of growth, substrate consumption, and product formation, is combined with a model of the bioreactor in which the cells are cultivated. Bioreactor models are usually set up as quite simple dynamical systems based on mass balances of substrates, products, biomass, and viable cells, normally assuming ideal mixing (however, highly complex models also exists Lapin et al., 2010). Considering the bioreactor as part of the modeled system is necessary for calculating the quantities discussed in the section on model purpose, like productivity and final titer. Not only are both the time trajectories for biomass concentration and the specific rates of consumption and production need for their determination but the dynamics of these variable are usually dependent on one another (Maurer et al., 2006; Douma et al., 2010) and their dynamics must be dealt with simultaneously.

To be useful for the design or optimization of fermentation or bioreactor processes, the models of the cell metabolism need only be predictive in a input–output sense. As long as this is the case it does not matter whether they are mechanistically correct representations of intracellular biochemistry or just empirical models. Because of the challenges of setting up mechanistic models on the molecular level, the production processes have traditionally been, and commonly still are, modeled with either unstructured kinetic models or by other simplified model designs (DiMasi and Swartz, 1995). Such kinetic models have been used to describe both continuous and fed-batch cultivations. For continuous cultures, modeling has for instance been used to study growth and metabolism of mammalian cells (DiMasi and Swartz, 1995) and the effect of oxygen uptake on L-lysine production in *Corynebacterium lactofermentum* (Ensari and Lim, 2003). Models have also been used for optimizing operating conditions such as the dilution rate in order to maximize production of protein in *S. cerevisiae* (Carlsen et al., 1997) and L-(–)-carnitine in *E. coli* (Alvarez-Vasquez et al., 2002). The model-based predictions in both of these studies turned out to agree very well with experiments. In industry many processes are run in fed-batch mode and kinetic models of fed-batch processes have for example been used to study penicillin fermentation in *P. chrysogenum* (Menezes et al., 1994) and production of proteins in mammalian cell lines like baby hamster kidney (Teixeira et al., 2007), murine hybridoma (Dorka et al., 2009), and

CHO (Xing et al., 2010). Several investigators have also used similar models for the optimization of fed-batch feeding profiles (Sarkar and Modak, 2003), for instance to maximize astaxanthin production in *Xanthophyllumyces dendrophous* (Liu and Wu, 2008) or protein production in *E. coli* (Levisauskas et al., 2003) and *P. pastoris* (Maurer et al., 2006), and for maximizing production of biomass while at the same time minimizing ethanol formation in a *S. cerevisiae* fermentation (Yüzgeç et al., 2009). The predicted optimal cultivation strategies in all of these studies were shown to compare well with validation experiments and did indeed lead to significantly improved fed batch processes. As well as using kinetic process models for optimizing operational strategies in advance, kinetic models are also potentially useful for online control (Chae et al., 2000; Teixeira et al., 2007; Yüzgeç et al., 2009).

Although models with simpler structures have proven useful in many cases, models of production processes and cell cultivation with increasing mechanistic details of intracellular reactions are now starting to appear (Bettenbrock et al., 2006; Shinto et al., 2007; Oshiro et al., 2009; Kadir et al., 2010; Li et al., 2011; Nolan and Lee, 2011). In fact, kinetic models of substantial complexity, which have been successful in describing how the metabolic state of cells varies with the external conditions, have recently been presented for both *E. coli* (Kotte et al., 2010; Usuda et al., 2010) and *S. cerevisiae* (Moisset et al., 2012). In addition to detailed representations of primary metabolic reaction networks, these models include genetic regulation of enzyme concentrations. One of the advantages of using more detailed models of the production process is that it allows the synergistic effects of metabolic engineering and process conditions to be evaluated, something which was recently explored in a mainly computational study addressing CHO cell metabolism (Nolan and Lee, 2012).

4. Advantages, challenges and perspectives

4.1. Advantages

The general strength of the kinetic modeling approach is that it quantitatively takes into account the factors that determine the rate of reactions. Compared to the modeling paradigm of the constraint-based stoichiometric models, which mainly is focusing on which reactions that can occur and the proportions of their reactants and products, kinetic models also define when and to what extent reactions take place. For an enzymatic reaction, for instance, not only can the effects of substrate and product concentrations be incorporated into the kinetic rate expression, but also the effects of co-factors, activators, inhibitors, and other modulators of enzyme activity. The ability of kinetic models to incorporate detailed information about reactions gives them a number of advantageous properties. Though, it must be emphasized that the advantages listed below partially reflect the potential capabilities of kinetics models, and not necessarily what is routinely achieved for all kinetic models.

The principles of kinetic modeling are applicable for all parts of the cell as well as the extracellular environment. Thus, a kinetic model can not only describe the rates of several interlinked enzyme-catalyzed reactions and the corresponding dynamics of the interconverted metabolites, but it may also include additional layers accounting for the rates of synthesis and degradation of transcripts and enzymes, as well as the rates of reactions involved in various sensing mechanisms and signal transduction. The many different levels of control, regulation and coordination of biochemistry are essential features of living cells (Heinemann and Sauer, 2010) and a modeling framework with a broad applicability is

clearly an advantage if one desires to study the integration of different cellular processes. A kinetic model of the cell factory is furthermore easily embedded in a dynamic model of the bioreactor process itself. GEMs, on the other hand, are less flexible and work best for modeling fluxes of metabolites.

Kinetic models can assist in understanding the complex behaviors of biological systems. Although the qualitative behavior may be intuitive, such as end-product inhibition in a linear pathway, understanding both the qualitative and quantitative aspects of how system behavior emerges from the properties of its components and their interactions is generally not trivial. In fact, even really small molecular circuits with just a few components are capable of producing non-intuitive dynamic behaviors such as adaptation, homeostasis, irreversible switching and oscillation (Tyson et al., 2003). Modeling behavior like these requires a kinetic approach and is beyond the scope of GEMs. One interesting example of how kinetic modeling has provided insight into the emergence of complex behaviors is the model of metabolic adaption in *E. coli* (Kotte et al., 2010). Here, a kinetic formulation of the reactions of the central metabolism, including their transcriptional and translational regulation, was shown to be capable of reproducing system-level metabolic adjustments through a mechanism termed distributed sensing of intracellular metabolic fluxes. This can be compared to the incorporation of Boolean rules for known gene regulation in GEMs (Herrgard et al., 2006). Since the regulatory information is explicitly hard-wired into the model, this strategy can never offer the same explanatory power in terms of actual molecular mechanisms.

Kinetic modeling can turn understanding of how cell factories work into predictions about how to improve them. When models have been set up linking relevant aspects on the system-level with the properties of the system components, they become valuable for predicting and optimizing the performance of cell factories. Ideally, the model components represent things that can be manipulated such as expression levels or process parameters, but also when model components are more abstract there may be general predictions achievable that still are useful. Kinetic models are unique in that predictions and optimizations are quantitative and can be very detailed, going beyond the regime of gene addition and deletion typically identified from flux balance analysis of GEMs. Thus, if such details are desired, kinetic models are conceptually superior to GEMs which instead are better suited for pathway-oriented problems involving prediction of the steady-state flux-capabilities of metabolic networks.

4.2. Challenges

A number of challenges must be addressed and overcome for biotechnology to capitalize from the advantages of kinetic modeling. The overall challenge lies in producing predictive models of high quality that really can make a difference for improving cell factory performance. Although this review has presented a number of studies where models have been used for predicting metabolic engineering targets, some of which have been experimentally verified, we are still far away from having kinetic models that are sufficiently good to be used for *in silico* design of industrially competitive cell factories. In this respect, kinetic models have not reached the same degree of maturity and industrial applicability as the much more successful GEMs.

The difficulty of producing high quality predictive models is that it requires a lot of detailed information about the system that one wishes to model. If too little information is available, the strengths and advantages of the kinetic modeling approach cannot be realized. Unfortunately kinetic modeling efforts frequently suffer from incomplete and uncertain knowledge of the underlying biochemistry with respect to both network structures, kinetic

rate expressions, and parameter values (Schaber et al., 2009; Kaltenbach et al., 2009; Soh et al., 2012). Thanks to the many reconstructions of genome-scale metabolic networks, the pathways and stoichiometry of metabolic reactions are often mapped out quite well, but regulatory mechanisms, both at the level of enzyme–metabolite interaction and at the transcriptional and translational level, are usually characterized to a lesser extent. For systems involving other types of reactions, such as protein secretion networks or signal transduction, knowledge of components and interaction is usually scarce, making the formulation of the network structure a challenging task (Schaber and Klipp, 2011). Little is also known about the exact mechanisms of the majority of reactions, meaning that the structure of kinetic rate expressions is also mostly unknown. Though, if a network structure as defined here (including qualitative knowledge about reaction modifiers) can be formulated, this issue can to some degree be circumvented using approximative kinetic rate expressions. The challenge of determining suitable structures for kinetic models should not be underestimated. While the literature often emphasizes the lack of quantitative information regarding parameter values, the lack of qualitative information based on which the model structure is set up may prove to be an even more difficult problem, at least when modeling certain parts of the cellular biochemistry. Nevertheless, the limited information about parameter values also deserves a lot of attention. Already in smaller models there are typically lots of parameters with unknown values and determining them is indeed a challenge. Even in those cases where parameter values are claimed to be known as the result of studying individual components, models that agree with experimental data of system properties do not automatically follow. A well-known example illustrating this point is the study of glycolysis in *S. cerevisiae* by Teusink et al. (2000). Here, kinetic parameters were determined experimentally under standard conditions *in vitro* for most of the glycolytic enzymes. When the individual enzyme kinetics were pieced together to form a model of the entire pathway, the model predictions deviated substantially from the *in vivo* behavior in some parts. This and other examples suggest that the accumulated uncertainty introduced by *in vitro* measurements, differences in experimental protocols, using data from different organisms, etc., make the resulting models questionable. If possible, modelers should try to make a transition from the bottom-up philosophy of determining parameter values, and instead collectively estimate them using *in vivo* data with the same scope as that of the system being modeled. As shown in this review, rigorous mathematical frameworks have been established for this task and there are several available methods for solving the resulting optimization problems. There are also established methodologies within identifiability analysis and model reduction which will help in achieving well-posed estimation problems.

Producing the right kind of data is critical for parameter estimation in kinetic models. Ideally, methods from identifiability analysis and experimental design should assist in laying down the directions for what data to collect, rather than uncritically basing these decisions common practice or on intuition. Performing relevant analysis and simulations before even a single experiment has been performed can potentially save both time and resources, and lead to better models. Of course, it may then turn out that the construction of a particular model is best done with the aid of data that is currently not routinely produced. To estimate parameter in larger scale models it is for instance expected that high-throughput time-series data will be a crucial factor. Thus, kinetic modeling can act as a driver for the development of new experimental techniques as well as a better use of existing ones.

The time it takes to set up kinetic models must be reduced. Since modeling projects can be very different in their scopes and

purposes it is hard to find a recipe that fits all scenarios and as a result the modeling procedure often becomes rather slow, typically involving a lot of manual work and case-to-case considerations by the modeler. One part of the solution towards a faster modeling cycle may be for the kinetic modeling community to continue to strive for a higher degree of standardization and automation. This is important not only for the representation and implementation of models, but foremost for the methods and workflows used to set them up. In this respect, valuable insights may come from looking at workflows for setting up GEMs (although the methods of course are different) where a substantial number of models have been produced in relatively short time.

If the routine generation of highly predictive kinetic models would become a reality, this will in turn pose new challenges for molecular biologists. As the predictions of metabolic engineering strategies derived from such models might be quantitatively very precise, an equally high precision in their implementation may potentially be needed to materialize the full potential of those predictions. This may require a precision in molecular biology methods that is currently not achievable, for instance such as very finely tuned expression levels or precise alteration of the catalytic properties of an engineered enzyme. In fact, the lack of such precision is already today preventing an exact implementation of the detailed results from the previously mentioned computational studies on optimal levels of metabolic enzymes.

4.3. Perspectives

It is not unrealistic to envision a future scenario where industrially relevant strategies for cell factory improvements based on classical methods gradually become exhausted or obsolete, and where the design is successively replaced by model-driven methods (Otero and Nielsen, 2010; Miskovic and Hatzimanikatis, 2010; Cvijovic et al., 2011). The most mature mathematical models of today, the GEMs, are the obvious candidates for this transition and they have already generated valuable results (Bro et al., 2006; Lee et al., 2006; Asadollahi et al., 2009; Becker et al., 2011; Neuner and Heinzel, 2011; Park et al., 2011). However, as the lower-hanging fruits of computational strategies are collected, the stoichiometric models will eventually also run into problems of predicting new targets. In this long-term perspective, kinetic models may well become a strong driving force for advancing the industrial application of cell factories. Two of the future aspects that are likely to be important for moving kinetic modeling forward are how their size and coverage can be increased and how they should deal with the previously mentioned limitations and uncertainty in the information needed to set them up.

In the future we will need to start producing large-scale kinetic models. The organization of the different biochemical reactions and pathways of the cell is characterized by a high degree of interconnectivity, for example through common precursor, energy, and redox metabolites. Because of this, changes in one part of the network of reactions may have unexpected consequences for other parts, rendering a global system perspective necessary. In the light of this complexity, one of the reasons for the successful application of stoichiometric models for predicting metabolic engineering targets is the fact that they can be set up on the genome-scale. Their aim for completeness means that they are re-usable for many kinds of problems and their popularity has even encouraged community consensus reconstructions of metabolic networks for *S. cerevisiae* (Herrgård et al., 2008). In addition to an extensive coverage of metabolism, recent work on GEMs is taking a genome-scale perspective also on the transcriptional and translational machinery (Thiele et al., 2009) as well as on protein secretion pathways (Feizi et al., 2013). The GEMs clearly have the advantage of being suitable for large scales, but they ultimately lack the

details required for a full characterization of the cell. Existing kinetic models, on the other hand, are usually set up on a small or at most medium scale. They are often built under rather specific assumptions which make them less adaptable for re-use in new situations and their lack of standard impedes the possibilities of merging smaller models into larger ones. Moving towards large-scale kinetic models, especially for the most important platform cell factories, will hopefully allow for better predictions and widen the possible model applications. Although the routine construction and use of genome-scale kinetic models definitely lies many years ahead, there are however already some emerging efforts towards the formulation of kinetic models with a more complete coverage. Various modeling methodologies have been proposed aiming to move large-scale modeling from stoichiometric constraint-based approaches to the kinetic domain (Famili et al., 2005; Smallbone et al., 2007; Jamshidi and Palsson, 2008; Ao et al., 2008; Adiamah et al., 2010; Smallbone et al., 2010), but their usefulness for cell factory improvements remains to be proven. Important results have been achieved for consistent reduction of metabolic networks (Soh, 2013), which may contribute towards genome-scale kinetic models through facilitating intermediate large-scale steppingstones (Chakrabarti et al., 2013). The idea of a community of modelers that together drives the development of large-scale kinetic models is potentially also interesting. Clearly, it is very difficult at this stage to predict which particular parts of the kinetic modeling procedure will be most crucial for eventually achieving kinetic genome-scale models. If it at all is possible given our current capabilities, it will likely involve a combination of the different topics covered within this review. In parallel with these developments, the constraint-based approaches have been modified to account for the dynamics of fluxes by sequential solutions of different steady states. These rather popular methods of so-called dynamic flux balance analysis (Mahadevan et al., 2002; Lee et al., 2008; Oddone et al., 2009) are however not addressing the kinetics of reactions, and are therefore still limited in this sense. Though, the concepts of dynamic flux balance might become useful in hybrid strategies where stoichiometric models and kinetic models are combined.

The awareness of the limited and uncertain information available for setting up kinetic models, especially large models, should lead to a critical examination of modeling strategies. One way in which modelers are facing up to the challenge posed by uncertainty is the inclusion of uncertainty itself as a part of the models. As shown in this review there is an increasing trend of publications addressing the uncertainty of both structure (Chou and Voit, 2009; Kaltenbach et al., 2009; Schaber et al., 2012), and of parameters, both directly (Pritchard and Kell, 2002; Liebermeister and Klipp, 2005; Piazza et al., 2008; Contador et al., 2009; Kotte and Heinemann, 2009; McLeod et al., 2011) and indirectly through sampling of enzyme state spaces (Wang et al., 2004; Wang and Hatzimanikatis, 2006a, 2006b; Miskovic and Hatzimanikatis, 2010, 2011). The presence of nested uncertainties of model structure and parameter values has also been emphasized (Schaber et al., 2009; Schaber and Klipp, 2011). Taking an even more fundamental approach to the uncertainty of molecular cell biology, researchers should also continue to investigate the prospect of accounting for uncertainty of parameters, reaction rates, and networks, using kinetic models based on stochastic differential equations, something which has been successful in other fields of biological modeling (Berglund et al., 2012). Related approaches can additionally be used to account for the uncertainty and variation between individual cells in a population (Almquist et al., 2008). Fueled by novel experimental techniques such as single cell metabolomics (Heinemann and Zenobi, 2011) and single cell level cultivation (Grünberger et al., 2012), cellular heterogeneity is a topic of growing interest (Lidstrom and Konopka, 2010) but its implications for cell factory design are largely uncharted territory. Based on these trends, further development of kinetic modeling strategies that can handle uncertainty will

likely be one essential ingredient for generating better cell factory models in the future. As part of such efforts it should be particularly important to elucidate the impact of uncertainty on the predictive power of models Gutenkunst et al. (2007a,b).

Acknowledgments

This work has been supported by the EU Framework VII project SYSINBIO – Systems Biology as a Driver for Industrial Biotechnology (call FP7-KBBE-2007-1).

References

- Adalsteinsson, D., McMillen, D., Elston, T.C., 2004. Biochemical network stochastic simulator (BioNetS): software for stochastic modeling of biochemical networks. *BMC Bioinf.* 5, 24, <http://dx.doi.org/10.1186/1471-2105-5-24>.
- Adiamah, D.A., Handl, J., Schwartz, J.M., 2010. Streamlining the construction of large-scale dynamic models using generic kinetic equations. *Bioinformatics* 26, 1324–1331, <http://dx.doi.org/10.1093/bioinformatics/btq136>.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control AC-19*, 716–723.
- Almquist, J., Lang, P., Prätzel-Wolters, D., Deitmer, J.W., Jirstrand, M., Becker, H.M., 2010a. A kinetic model of the monocarboxylate transporter MCT1 and its interaction with carbonic anhydrase II. *J. Comput. Sci. Syst. Biol.* 3, 107–116.
- Almquist, J., Sunnåker, M., Hagmar, J., Kvärnstrom, M., Jirstrand, M., 2008. System identification from spatiotemporal cell population data. In: The 9th International Conference on Systems Biology.
- Almquist, J., Wallman, M., Jacobson, I., Jirstrand, M., 2010b. Modeling the effect of Kv1.5 block on the canine action potential. *Biophys. J.* 99, 2726–2736, <http://dx.doi.org/10.1016/j.bpj.2010.08.062>.
- Alvarez-Vasquez, F., Cánovas, M., Iborra, J.L., Torres, N.V., 2002. Modeling, optimization and experimental assessment of continuous L-(−)-carnitine production by *Escherichia coli* cultures. *Biotechnol. Bioeng.* 80, 794–805, <http://dx.doi.org/10.1002/bit.10436>.
- Alvarez-Vasquez, F., González-Alcón, C., Torres, N.V., 2000. Metabolism of citric acid production by *Aspergillus niger*: model definition steady-state analysis and constrained optimization of citric acid production rate. *Biotechnol. Bioeng.* 70, 82–108.
- Alvarez-Vasquez, F., Sims, K.J., Cowart, L.A., Okamoto, Y., Voit, E.O., Hannun, Y.A., 2005. Simulation and validation of modelled sphingolipid metabolism in *Saccharomyces cerevisiae*. *Nature* 433, 425–430, <http://dx.doi.org/10.1038/nature03232>.
- Alvarez-Vasquez, F., Sims, K.J., Hannun, Y.A., Voit, E.O., 2004. Integration of kinetic information on yeast sphingolipid metabolism in dynamical pathway models. *J. Theor. Biol.* 226, 265–291.
- Anguelova, M., Karlsson, J., Jirstrand, M., 2012. Minimal output sets for identifiability. *Math. Biosci.* 239, 139–153, <http://dx.doi.org/10.1016/j.mbs.2012.04.005>.
- Ao, P., Lee, L.W., Lidstrom, M.E., Yin, L., Zhu, X., 2008. Towards kinetic modeling of global metabolic networks: *Methylobacterium extorquens* AM1 growth as validation. *Sheng Wu Gong Cheng Xue Bao* 24, 980–994.
- Asadollahi, M.A., Maury, J., Patil, K.R., Schalk, M., Clark, A., Nielsen, J., 2009. Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering. *Metab. Eng.* 11, 328–334, <http://dx.doi.org/10.1016/j.ymben.2009.07.001>.
- Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J.A., Blom, J.G., 2009. Systems biology: parameter estimation for biochemical models. *FEBS J.* 276, 886–902, <http://dx.doi.org/10.1111/j.1742-4658.2008.06844.x>.
- Audoly, S., Bellu, G., D'Angiò, L., Saccomani, M.P., Cobelli, C., 2001. Global identifiability of nonlinear models of biological systems. *IEEE Trans. Biomed. Eng.* 48, 55–65, <http://dx.doi.org/10.1109/10.900248>.
- Baker, S.M., Schallau, K., Junker, B.H., 2010. Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models. *J. Integr. Bioinform.* 7, <http://dx.doi.org/10.2390/biec09-jib-2010-133>.
- Balsa-Canto, E., Banga, J.R., 2011. AMIGO, a toolbox for advanced model identification in systems biology using global optimization. *Bioinformatics* 27, 2311–2313, <http://dx.doi.org/10.1093/bioinformatics/btr370>.
- Balsa-Canto, E., Banga, J.R., Egea, J.A., Fernández-Villaverde, A., de Hijas-Liste, G.M., 2012. Global optimization in systems biology: stochastic methods and their applications. *Adv. Exp. Med. Biol.* 736, 409–424, http://dx.doi.org/10.1007/978-1-4419-7210-1_24.
- Balsa-Canto, E., Peifer, M., Banga, J.R., Timmer, J., Fleck, C., 2008. Hybrid optimization method with general switching strategy for parameter estimation. *BMC Syst. Biol.* 2, 26, <http://dx.doi.org/10.1186/1752-0509-2-26>.
- Becker, J., Zelder, O., Häfner, S., Schroder, H., Wittmann, C., 2011. From zero to hero: design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. *Metab. Eng.* 13, 159–168, <http://dx.doi.org/10.1016/j.ymben.2011.01.003>.
- Bellman, R., Åström, K.J., 1970. On structural identifiability. *Math. Biosci.* 7, 329–339.
- Bellu, G., Saccomani, M.P., Audoly, S., D'Angiò, L., 2007. DAISY: a new software tool to test global identifiability of biological and physiological systems. *Comput.*

- Methods Progr. Biomed. 88, 52–61, <http://dx.doi.org/10.1016/j.cmpb.2007.07.002>.
- Berglund, M., Sunnåker, M., Adiels, M., Jirstrand, M., Wennberg, B., 2012. Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations. *Math. Med. Biol.* 29, 361–384, <http://dx.doi.org/10.1093/imammb/dqr021>.
- Bettenbrock, K., Fischer, S., Kremling, A., Jahreis, K., Sauter, T., Gilles, E.D., 2006. A quantitative approach to catabolite repression in *Escherichia coli*. *J. Biol. Chem.* 281, 2578–2584, <http://dx.doi.org/10.1074/jbc.M508090200>.
- Bohlin, T., 2006. Practical Grey-Box Process Identification: Theory and Applications. Springer-Verlag, London.
- Borger, S., Liebermeister, W., Uhlendorf, J., Klipp, E., 2007. Automatically generated model of a metabolic network. *Genome Inf.* 18, 215–224.
- Brent, R.P., 1973. Algorithms for Minimization without Derivatives. Prentice-Hall, Englewood Cliffs, NJ.
- Brik Ternbach, M., Bollman, C., Wandrey, C., Takors, R., 2005. Application of model discriminating experimental design for modeling and development of a fermentative fed-batch L-valine production process. *Biotechnol. Bioeng.* 91, 356–368, <http://dx.doi.org/10.1002/bit.20504>.
- Bro, C., Regenberg, B., Forster, J., Nielsen, J., 2006. In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab. Eng.* 8, 102–111, <http://dx.doi.org/10.1016/j.jymben.2005.09.007>.
- Bruggeman, F.J., Boogerd, F.C., Westerhoff, H.V., 2005. The multifarious short-term regulation of ammonium assimilation of *Escherichia coli*: dissection using an *in silico* replica. *FEBS J.* 272, 1965–1985, <http://dx.doi.org/10.1111/j.1742-4658.2005.04626.x>.
- Bruggeman, F.J., Westerhoff, H.V., 2007. The nature of systems biology. *Trends Microbiol.* 15, 45–50, <http://dx.doi.org/10.1016/j.tim.2006.11.003>.
- Bulik, S., Grimsb, S., Huthmacher, C., Selbig, J., Holzhütter, H.G., 2009. Kinetic hybrid models composed of mechanistic and simplified enzymatic rate laws—a promising method for speeding up the kinetic modelling of complex metabolic networks. *FEBS J.* 276, 410–424.
- Carlsen, M., Jochumsen, K.V., Emborg, C., Nielsen, J., 1997. Modeling the growth and proteinase A production in continuous cultures of recombinant *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 55, 447–454 <http://dx.doi.org/3.0.CO;2-C>.
- Carlsson, J., Nordheim, C., 2011. A Parameter Estimation Method for Continuous Time Dynamical Systems Based on the Unscented Kalman Filter and Maximum Likelihood (Master's thesis). Chalmers university of technology.
- Cazzaniga, P., Pescini, D., Besozzi, D., Mauri, G., Colombo, S., Martegani, E., 2008. Modeling and stochastic simulation of the Ras/cAMP/PKA pathway in the yeast *Saccharomyces cerevisiae* evidences a key regulatory function for intracellular guanine nucleotides pools. *J. Biotechnol.* 133, 377–385, <http://dx.doi.org/10.1016/j.biotec.2007.09.019>.
- Cedersund, G., 2006. Core-Box Modelling (Ph.D. thesis). Chalmers University of Technology.
- Cedersund, G., Roll, J., 2009. Systems biology: model based evaluation and comparison of potential explanations for given biological data. *FEBS J.* 276, 903–922, <http://dx.doi.org/10.1111/j.1742-4658.2008.06845.x>.
- Chae, H.J., Delisa, M.P., Cha, H.J., Weigand, W.A., Rao, G., Bentley, W.E., 2000. Framework for online optimization of recombinant protein expression in high-cell-density *Escherichia coli* cultures using GFP-fusion monitoring. *Biotechnol. Bioeng.* 69, 275–285.
- Chakrabarti, A., Miskovic, L., Soh, K.C., Hatzimanikatis, V., 2013. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnol. J.* 8, 1043–1057, <http://dx.doi.org/10.1002/biot.201300091>.
- Chang, Y., Sahinidis, N.V., 2005. Optimization of metabolic pathways under stability considerations. *Comput. Chem. Eng.* 29, 467–479.
- Chassagnole, C., Fell, D.A., Raïs, B., Kudla, B., Mazat, J.P., 2001. Control of the threonine-synthesis pathway in *Escherichia coli*: a theoretical and experimental approach. *Biochem. J.* 356, 433–444.
- Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M., 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol. Bioeng.* 79, 53–73.
- Chauve, M., Mathis, H., Huc, D., Casanave, D., Monot, F., Ferreira, N.L., 2010. Comparative kinetic analysis of two fungal beta-glucosidases. *Biotechnol. Biofuels* 3, 3, <http://dx.doi.org/10.1186/1754-6834-3-3>.
- Chen, N., Koumpouras, G.C., Polizzi, K.M., Kontoravdi, C., 2012. Genome-based kinetic modeling of cytosolic glucose metabolism in industrially relevant cell lines: *Saccharomyces cerevisiae* and Chinese hamster ovary cells. *Bioprocess. Biosyst. Eng.* 35, 1023–1033, <http://dx.doi.org/10.1007/s00449-012-0687-3>.
- Chen, W.W., Niepel, M., Sorger, P.K., 2010. Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev.* 24, 1861–1875, <http://dx.doi.org/10.1101/gad.1945410>.
- Chis, O.T., Banga, J.R., Balsa-Canto, E., 2011. Structural identifiability of systems biology models: a critical comparison of methods. *PLoS One* 6, e27755, <http://dx.doi.org/10.1371/journal.pone.0027755>.
- Chou, I.C., Voit, E.O., 2009. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math. Biosci.* 219, 57–83, <http://dx.doi.org/10.1016/j.mbs.2009.03.002>.
- Chung, B.K., Selvarasu, S., Andrea, C., Ryu, J., Lee, H., Ahn, J., Lee, H., Lee, D.Y., 2010. Genome-scale metabolic reconstruction and *in silico* analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. *Microb. Cell Fact.* 9, 50, <http://dx.doi.org/10.1186/1475-2859-9-50>.
- Cintolesi, A., Clomburg, J.M., Rigou, V., Zygourakis, K., Gonzalez, R., 2012. Quantitative analysis of the fermentative metabolism of glycerol in *Escherichia coli*. *Biotechnol. Bioeng.* 109, 187–198, <http://dx.doi.org/10.1002/bit.23309>.
- Contador, C.A., Rizk, M.L., Asenjo, J.A., Liao, J.C., 2009. Ensemble modeling for strain development of L-lysine-producing *Escherichia coli*. *Metab. Eng.* 11, 221–233, <http://dx.doi.org/10.1016/j.ymben.2009.04.002>.
- Costa, R.S., Machado, D., Rocha, I., Ferreira, E.C., 2010. Hybrid dynamic modeling of *Escherichia coli* central metabolic network combining Michaelis-Menten and approximate kinetic equations. *Biosystems* 100, 150–157, <http://dx.doi.org/10.1016/j.biosystems.2010.03.001>.
- Costa, R.S., Machado, D., Rocha, I., Ferreira, E.C., 2011. Critical perspective on the consequences of the limited availability of kinetic data in metabolic dynamic modelling. *IET Syst. Biol.* 5, 157–163, <http://dx.doi.org/10.1049/iet-syb.2009.0058>.
- Cronwright, G.R., Rohwer, J.M., Prior, B.A., 2002. Metabolic control analysis of glycerol synthesis in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 68, 4448–4456.
- Curtu, R., Diedrichs, D., 2010. Small-scale modeling approach and circuit wiring of the unfolded protein response in mammalian cells. *Adv. Exp. Med. Biol.* 680, 261–274, http://dx.doi.org/10.1007/978-1-4419-5913-3_30.
- Cvijovic, M., Bordel, S., Nielsen, J., 2011. Mathematical models of cell factories: moving towards the core of industrial biotechnology. *Microb. Biotechnol.* 4, 572–584, <http://dx.doi.org/10.1111/j.1751-7915.2010.00233.x>.
- Dano, S., Madsen, M.F., Schmidt, H., Cedersund, G., 2006. Reduction of a biochemical model with preservation of its basic dynamic properties. *FEBS J.* 273, 4862–4877, <http://dx.doi.org/10.1111/j.1742-4658.2006.05485.x>.
- de Groot, M.J.L., Prathumpai, W., Visser, J., Ruijter, G.J.G., 2005. Metabolic control analysis of *Aspergillus niger* L-arabinose catabolism. *Biotechnol. Prog.* 21, 1610–1616, <http://dx.doi.org/10.1021/bp0501890>.
- Degenring, D., Froemel, C., Dikta, G., Takors, R., 2004. Sensitivity analysis for the reduction of complex metabolism models. *J. Process Control* 14, 729–745.
- del Rosario, R.C.H., Mendoza, E., Voit, E.O., 2008. Challenges in lin-log modelling of glycolysis in *Lactococcus lactis*. *IET Syst. Biol.* 2, 136–149, <http://dx.doi.org/10.1049/iet-syb:20070030>.
- DiMasi, D., Swartz, R.W., 1995. An energetically structured model of mammalian cell metabolism. I. Model development and application to steady-state hybridoma cell growth in continuous culture. *Biotechnol. Prog.* 11, 664–676, <http://dx.doi.org/10.1021/bp00036a010>.
- Dorka, P., Fischer, C., Budman, H., Scharer, J.M., 2009. Metabolic flux-based modeling of mAb production during batch and fed-batch operations. *Bioprocess Biosyst. Eng.* 32, 183–196, <http://dx.doi.org/10.1007/s00449-008-0236-2>.
- Douma, R.D., Verheijen, P.J.T., de Laat, W.T.A.M., Heijnen, J.J., van Gulik, W.M., 2010. Dynamic gene expression regulation model for growth and penicillin production in *Penicillium chrysogenum*. *Biotechnol. Bioeng.* 106, 608–618, <http://dx.doi.org/10.1002/bit.22689>.
- Dräger, A., Kronfeld, M., Ziller, M.J., Supper, J., Planatscher, H., Magnus, J.B., Oldiges, M., Kohlbacher, O., Zell, A., 2009. Modeling metabolic networks in *C. glutamicum*: a comparison of rate laws in combination with various parameter optimization strategies. *BMC Syst. Biol.* 3, 5, <http://dx.doi.org/10.1186/1752-0509-3-5>.
- Du, J., Shao, Z., Zhao, H., 2011. Engineering microbial factories for synthesis of value-added products. *J. Ind. Microbiol. Biotechnol.* 38, 873–890, <http://dx.doi.org/10.1007/s10295-011-0970-3>.
- Egea, J.A., Martí, R., Banga, J.R., 2010. An evolutionary method for complex-process optimization. *Comput. Oper. Res.* 37, 315–324.
- Ensari, S., Lim, H.C., 2003. Kinetics of L-lysine fermentation: a continuous culture model incorporating oxygen uptake rate. *Appl. Microbiol. Biotechnol.* 62, 35–40, <http://dx.doi.org/10.1007/s00253-003-1266-y>.
- Erguler, K., Stumpf, M.P.H., 2011. Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Mol. Biosyst.* 7, 1593–1602, <http://dx.doi.org/10.1039/c0mb00107d>.
- van Eunen, K., Bouwman, J., Daran-Lapujade, P., Postmus, J., Canelas, A.B., Mensenides, F.I.C., Orij, R., 2010. Measuring enzyme activities under standardized in vivo-like conditions for systems biology. *FEBS J.* 277, 749–760, <http://dx.doi.org/10.1111/j.1742-4658.2009.07524.x>.
- Famili, I., Mahadevan, R., Palsson, B.O., 2005. k-Cone analysis: determining all candidate values for kinetic parameters on a network scale. *Biophys. J.* 88, 1616–1625, <http://dx.doi.org/10.1529/biophysj.104.050385>.
- Feizi, A., Österlund, T., Petranovic, D., Bordel, S., Nielsen, J., 2013. Genome-scale modeling of the protein secretory machinery in yeast. *PLoS One* 8, e63284, <http://dx.doi.org/10.1371/journal.pone.0063284>.
- Fell, D.A., 1992. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem. J.* 286 (Pt 2), 313–330.
- Ghosh, S., Banerjee, S., Bose, I., 2012. Emergent bistability: effects of additive and multiplicative noise. *Eur. Phys. J. E: Soft Matter* 35, 11, <http://dx.doi.org/10.1140/epje/i2012-12011-4>.
- Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical-reactions. *J. Comput. Phys.* 22, 403–434.
- Gillespie, D.T., 2000. The chemical Langevin equation. *J. Chem. Phys.* 113, 297–306.
- Gillespie, D.T., 2001. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* 115, 1716–1733.
- Gonzalez, R., Andrews, B.A., Asenjo, J.A., 2001. Metabolic control analysis of monoclonal antibody synthesis. *Biotechnol. Prog.* 17, 217–226, <http://dx.doi.org/10.1021/bp000165b>.

- Gonze, D., Jacquet, M., Goldbeter, A., 2008. Stochastic modelling of nucleocytoplasmic oscillations of the transcription factor msn2 in yeast. *J. R. Soc. Interface* 5 (Suppl. 1), S95–109, <http://dx.doi.org/10.1098/rsif.2008.0141.focus>.
- Goryanin, I., Demin, O., 2009. *Kinetic Modelling in Systems Biology*. Chapman and Hall/CRC, Boca Raton, FL.
- Grünberger, A., Paczia, N., Probst, C., Schendzielorz, G., Eggeling, L., Noack, S., Wiechert, W., Kohlheyer, D., 2012. A disposable picolitre bioreactor for cultivation and investigation of industrially relevant bacteria on the single cell level. *Lab Chip* 12, 2060–2068, <http://dx.doi.org/10.1039/c2lc40156h>.
- Gutenkunst, R.N., Casey, F.P., Waterfall, J.J., Myers, C.R., Sethna, J.P., 2007a. Extracting falsifiable predictions from sloppy models. *Ann. N.Y. Acad. Sci.* 1115, 203–211, <http://dx.doi.org/10.1196/annals.1407.003>.
- Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., Sethna, J.P., 2007b. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* 3, 1871–1878, <http://dx.doi.org/10.1371/journal.pcbi.0030189>.
- Haag, J.E., Vande Wouwer, A., Bogaerts, P., 2005. Dynamic modeling of complex biological systems: a link between metabolic and macroscopic description. *Math. Biosci.* 193, 25–49, <http://dx.doi.org/10.1016/j.mbs.2004.11.007>.
- Hamilton, S.R., Gerngross, T.U., 2007. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr. Opin. Biotechnol.* 18, 387–392, <http://dx.doi.org/10.1016/j.copbio.2007.09.001>.
- Hasty, J., Pradines, J., Dolnik, M., Collins, J.J., 2000. Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 97, 2075–2080, <http://dx.doi.org/10.1073/pnas.040411297>.
- Hattersley, J.G., Pérez-Velázquez, J., Chappell, M.J., Bearup, D., Roper, D., Dowson, C., Bugg, T., Evans, N.D., 2011. Indistinguishability and identifiability of kinetic models for the MurC reaction in peptidoglycan biosynthesis. *Comput. Methods Progr. Biomed.* 104, 70–80, <http://dx.doi.org/10.1016/j.cmpb.2010.07.009>.
- Hatzimanikatis, V., Bailey, J.E., 1996. MCA has more to say. *J. Theor. Biol.* 182, 233–242, <http://dx.doi.org/10.1006/jtbi.1996.0160>.
- Hatzimanikatis, V., Bailey, J.E., 1997. Effects of spatiotemporal variations on metabolic control: approximate analysis using (log)linear kinetic models. *Biotechnol. Bioeng.* 54, 91–104 <http://dx.doi.org/3.0.CO;2-Q>.
- Hatzimanikatis, V., Floudas, C.A., Bailey, J.E., 1996a. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. *AIChE J.* 42, 1277–1292.
- Hatzimanikatis, V., Floudas, C.A., Bailey, J.E., 1996b. Optimization of regulatory architectures in metabolic reaction networks. *Biotechnol. Bioeng.* 52, 485–500 <http://dx.doi.org/3.0.CO;2-L>.
- Haunschmid, M.D., Freisleben, B., Takors, R., Wiechert, W., 2005. Investigating the dynamic behavior of biochemical networks using model families. *Bioinformatics* 21, 1617–1625, <http://dx.doi.org/10.1093/bioinformatics/bti225>.
- Heijnen, J.J., 2005. Approximative kinetic formats used in metabolic network modeling. *Biotechnol. Bioeng.* 91, 534–545, <http://dx.doi.org/10.1002/bit.20558>.
- Heinemann, M., Sauer, U., 2010. Systems biology of microbial metabolism. *Curr. Opin. Microbiol.* 13, 337–343, <http://dx.doi.org/10.1016/j.mib.2010.02.005>.
- Heinemann, M., Zenobi, R., 2011. Single cell metabolomics. *Curr. Opin. Biotechnol.* 22, 26–31, <http://dx.doi.org/10.1016/j.copbio.2010.09.008>.
- Heinrich, R., Rapoport, T.A., 1974. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur. J. Biochem.* 42, 89–95.
- Heinzle, E., Dunn, I.J., Furukawa, K., Tanner, R.D., 1982. Modelling of sustained oscillations observed in continuous culture of *Saccharomyces cerevisiae*. In: *Modelling and Control of Biotechnical Processes*.
- Heinzle, E., Matsuda, F., Miyagawa, H., Wakasa, K., Nishioka, T., 2007. Estimation of metabolic fluxes, expression levels and metabolite dynamics of a secondary metabolic pathway in potato using label pulse-feeding experiments combined with kinetic network modelling and simulation. *Plant J.* 50, 176–187, <http://dx.doi.org/10.1111/j.1365-313X.2007.03037.x>.
- Herrgård, M.J., Lee, B.S., Portnoy, V., Palsson, B.Ø., 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.* 16, 627–635, <http://dx.doi.org/10.1101/gr.4083206>.
- Herrgård, M.J., Swainston, N., Dobson, P., Dunn, W.B., Arga, K.Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M.L., Oliveira, A.P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasić, I., Weichert, D., Brent, R., Broomhead, D.S., Westerhoff, H.V., Kirdar, B., Penttilä, M., Klipp, E., 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* 26, 1155–1160, <http://dx.doi.org/10.1038/nbt1492>.
- Higham, D.J., 2001. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.* 43, 525–546.
- Hildebrandt, S., Raden, D., Petzold, L., Robinson, A.S., Doyle, F.J., 2008. A top-down approach to mechanistic biological modeling: application to the single-chain antibody folding pathway. *Biophys. J.* 95, 3535–3558, <http://dx.doi.org/10.1529/biophysj.107.125039>.
- Hoefnagel, M.H.N., Starrenburg, M.J.C., Martens, D.E., Hugenholz, J., Kleerebezem, M., Van Swam, I.I., Bongers, R., Westerhoff, H.V., Snoep, J.L., 2002. Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. *Microbiology* 148, 1003–1013.
- Hooke, R., Jeeves, T.A., 1961. Direct search solution of numerical and statistical problems. *J. Assoc. Comput. Mach.* 8, 212–229.
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Palhe, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U., 2006. COPASI—a COmplex PAthway SImulator. *Bioinformatics* 22, 3067–3074, <http://dx.doi.org/10.1093/bioinformatics/btl485>.
- Hossler, P., Mulukutla, B.C., Hu, W.S., 2007. Systems analysis of N-glycan processing in mammalian cells. *PLoS One* 2, e713, <http://dx.doi.org/10.1371/journal.pone.0000713>.
- Hua, Q., Yang, C., Shimizu, K., 2000. Metabolic control analysis for lysine synthesis using *Corynebacterium glutamicum* and experimental verification. *J. Biosci. Bioeng.* 90, 184–192.
- Huai, L., Chen, N., Yang, W., Bai, G., 2009. Metabolic control analysis of L-cysteine producing strain TS1138 of *Pseudomonas* sp. *Biochemistry* 74, 288–292.
- Hynne, F., Danø, S., Sørensen, P.G., 2001. Full-scale model of glycolysis in *Saccharomyces cerevisiae*. *Biophys. Chem.* 94, 121–163.
- Jamshidi, N., Palsson, B.Ø., 2008. Formulating genome-scale kinetic models in the post-genome era. *Mol. Syst. Biol.* 4, 171, <http://dx.doi.org/10.1038/msb.2008.8>.
- Jansson, A., Jirstrand, M., 2010. Biochemical modeling with systems biology graphical notation. *Drug Discov. Today* 15, 365–370, <http://dx.doi.org/10.1016/j.drudis.2010.02.012>.
- Jaqaman, K., Danuser, G., 2006. Linking data to models: data regression. *Nat. Rev. Mol. Cell Biol.* 7, 813–819, <http://dx.doi.org/10.1038/nrm2030>.
- Jazwinski, A.H., 1970. *Stochastic Processes and Filtering Theory*. Academic Press, New York.
- Kacser, H., Burns, J.A., 1973. The control of flux. *Symp. Soc. Exp. Biol.* 27, 65–104.
- Kadir, T.A.A., Mannan, A.A., Kierzek, A.M., McFadden, J., Shimizu, K., 2010. Modeling and simulation of the main metabolism in *Escherichia coli* and its several single-gene knockout mutants with experimental verification. *Microb. Cell Fact.* 9, 88, <http://dx.doi.org/10.1186/1475-2859-9-88>.
- Kaltenbach, H.M., Dimopoulos, S., Stelling, J., 2009. Systems analysis of cellular networks under uncertainty. *FEBS Lett.* 583, 3923–3930, <http://dx.doi.org/10.1016/j.febslet.2009.10.074>.
- van Kampen, N.G., 2007. *Stochastic Processes in Physics and Chemistry*. North Holland, Amsterdam.
- Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Karlsson, J., Anguelova, M., Jirstrand, M., 2012. An efficient method for structural identifiability analysis of large dynamic systems. In: 16th IFAC Symposium on System Identification.
- Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Networks.
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. *Science* 220, 671–680, <http://dx.doi.org/10.1126/science.220.4598.671>.
- Klipp, E., Herwig, R., Kowald, A., Wierling, C., Lehrach, H., 2005a. *Systems Biology in Practice*. Wiley-VCH, Berlin.
- Klipp, E., Nordlander, B., Krüger, R., Gennemark, P., Hohmann, S., 2005b. Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.* 23, 975–982, <http://dx.doi.org/10.1038/nbt1114>.
- Kotte, O., Heinemann, M., 2009. A divide-and-conquer approach to analyze under-determined biochemical models. *Bioinformatics* 25, 519–525, <http://dx.doi.org/10.1093/bioinformatics/btp004>.
- Kotte, O., Zaugg, J.B., Heinemann, M., 2010. Bacterial adaptation through distributed sensing of metabolic fluxes. *Mol. Syst. Biol.* 6, 355, <http://dx.doi.org/10.1038/msb.2010.10>.
- Krambeck, F.J., Bettenbaug, M.J., 2005. A mathematical model of n-linked glycosylation. *Biotechnol. Bioeng.* 92, 711–728, <http://dx.doi.org/10.1002/bit.20645>.
- Kreutz, C., Timmer, J., 2009. Systems biology: experimental design. *FEBS J.* 276, 923–942, <http://dx.doi.org/10.1111/j.1742-4658.2008.06843.x>.
- Kuepfer, L., Peter, M., Sauer, U., Stelling, J., 2007. Ensemble modeling for analysis of cell signaling dynamics. *Nat. Biotechnol.* 25, 1001–1006, <http://dx.doi.org/10.1038/nbt1330>.
- Lang, M., Waldherr, S., Allgöwer, F., 2009. Amplitude distribution of stochastic oscillations in biochemical networks due to intrinsic noise. *PMC Biophys.* 2, 10, <http://dx.doi.org/10.1186/1757-5036-2-10>.
- Lapin, A., Klann, M., Reuss, M., 2010. Multi-scale spatio-temporal modeling: lifelines of microorganisms in bioreactors and tracking molecules in cells. *Adv. Biochem. Eng. Biotechnol.* 121, 23–43, http://dx.doi.org/10.1007/10_2009_53.
- Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M.I., Wimalaratne, S.M., Bergman, F.T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y., Villéger, A., Boyd, S.E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, T.C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D.B., Sander, C., Sauro, H., Snoep, J.L., Kohn, K., Kitano, H., 2009. The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741, <http://dx.doi.org/10.1038/nbt1558>.
- Lee, F.C., Pandu Ranganiah, G., Lee, D.Y., 2010. Modeling and optimization of a multi-product biosynthesis factory for multiple objectives. *Metab. Eng.* 12, 251–267, <http://dx.doi.org/10.1016/j.ymben.2009.12.003>.
- Lee, J.M., Lee, J.M., Gianchandani, E.P., Eddy, J.A., Papin, J.A., 2008. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput. Biol.* 4, e1000086, <http://dx.doi.org/10.1371/journal.pcbi.1000086>.
- Lee, S.J., Song, H., Lee, S.Y., 2006. Genome-based metabolic engineering of mannheimia succiniciproducens for succinic acid production. *Appl. Environ. Microbiol.* 72, 1939–1948, <http://dx.doi.org/10.1128/AEM.72.3.1939-1948.2006>.
- Levisauskas, D., Galvanauskas, V., Henrich, S., Wilhelm, K., Volk, N., Lübbert, A., 2003. Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*. *Bioprocess Biosyst. Eng.* 25, 255–262, <http://dx.doi.org/10.1007/s00449-002-0305-x>.
- Li, R.D., Li, Y.Y., Lu, L.Y., Ren, C., Li, Y.X., Liu, L., 2011. An improved kinetic model for the acetone–butanol–ethanol pathway of *Clostridium acetobutylicum* and

- model-based perturbation analysis. *BMC Syst. Biol.* 5 (Suppl. 1), S12, <http://dx.doi.org/10.1186/1752-0509-5-S1-S12>.
- Lidstrom, M.E., Konopka, M.C., 2010. The role of physiological heterogeneity in microbial population behavior. *Nat. Chem. Biol.* 6, 705–712, <http://dx.doi.org/10.1038/nchembio.436>.
- Liebermeister, W., Baur, U., Klipp, E., 2005. Biochemical network models simplified by balanced truncation. *FEBS J.* 272, 4034–4043, <http://dx.doi.org/10.1111/j.1742-4658.2005.04780.x>.
- Liebermeister, W., Klipp, E., 2005. Biochemical networks with uncertain parameters. *Syst. Biol. (Stevenage)* 152, 97–107.
- Liebermeister, W., Klipp, E., 2006a. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor. Biol. Med. Model.* 3, 41, <http://dx.doi.org/10.1186/1742-4682-3-41>.
- Liebermeister, W., Klipp, E., 2006b. Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data. *Theor. Biol. Med. Model.* 3, 42, <http://dx.doi.org/10.1186/1742-4682-3-42>.
- Liebermeister, W., Uhlenhoff, J., Klipp, E., 2010. Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics* 26, 1528–1534, <http://dx.doi.org/10.1093/bioinformatics/btq141>.
- Liu, Y.S., Wu, J.Y., 2008. Modeling of *Xanthophyllumyces dendrophorus* growth on glucose and overflow metabolism in batch and fed-batch cultures for astaxanthin production. *Biotechnol. Bioeng.* 101, 996–1004, <http://dx.doi.org/10.1002/bit.21978>.
- Ljung, L., 1987. *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.
- Ljung, L., Glad, T., 1994a. *Modeling of Dynamic Systems*. Prentice-Hall, Englewood Cliffs, NJ.
- Ljung, L., Glad, T., 1994b. On global identifiability for arbitrary model parametrizations. *Automatica* 30, 265–276.
- Love, K.R., Panagiotou, V., Jiang, B., Stadhheim, T.A., Love, J.C., 2010. Integrated single-cell analysis shows *Pichia pastoris* secretes protein stochastically. *Biotechnol. Bioeng.* 106, 319–325, <http://dx.doi.org/10.1002/bit.22688>.
- Magnus, J.B., Oldiges, M., Takors, R., 2009. The identification of enzyme targets for the optimization of a valine producing *Corynebacterium glutamicum* strain using a kinetic model. *Biotechnol. Prog.* 25, 754–762, <http://dx.doi.org/10.1002/btpr.184>.
- Mahadevan, R., Edwards, J.S., Doyle, F.J., 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* 83, 1331–1340, [http://dx.doi.org/10.1016/S0006-3495\(02\)73903-9](http://dx.doi.org/10.1016/S0006-3495(02)73903-9).
- Maiwald, T., Timmer, J., 2008. Dynamical modeling and multi-experiment fitting with PottersWheel. *Bioinformatics* 24, 2037–2043, <http://dx.doi.org/10.1093/bioinformatics/btn350>.
- Margaria, G., Riccomagno, E., Chappell, M.J., Wynn, H.P., 2001. Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences. *Math. Biosci.* 174, 1–26.
- Marín-Sanguino, A., Torres, N.V., 2000. Optimization of tryptophan production in bacteria. Design of a strategy for genetic manipulation of the tryptophan operon for tryptophan flux maximization. *Biotechnol. Prog.* 16, 133–145, <http://dx.doi.org/10.1021/bp9901441>.
- Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* 11, 431–441.
- Maurer, M., Kühlleitner, M., Gasser, B., Mattanovich, D., 2006. Versatile modeling and optimization of fed batch processes for the production of secreted heterologous proteins with *Pichia pastoris*. *Microb. Cell Fact.* 5, 37, <http://dx.doi.org/10.1186/1475-2859-5-37>.
- McLeod, J., O'Callaghan, P.M., Pybus, L.P., Wilkinson, S.J., Root, T., Racher, A.J., James, D.C., 2011. An empirical modeling platform to evaluate the relative control discrete CHO cell synthetic processes exert over recombinant monoclonal antibody production process titer. *Biotechnol. Bioeng.* 108, 2193–2204, <http://dx.doi.org/10.1002/bit.23146>.
- Mendes, P., Hoops, S., Sahle, S., Gauges, R., Dada, J., Kummer, U., 2009. Computational modeling of biochemical networks using COPASI. *Methods Mol. Biol.* 500, 17–59, http://dx.doi.org/10.1007/978-1-59745-525-1_2.
- Mendes, P., Kell, D., 1998. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* 14, 869–883.
- Menezes, J.C., Alves, S.S., Lemos, J.M., de Azevedo, S.F., 1994. Mathematical modelling of industrial pilot-plant penicillin-G fed-batch fermentations. *J. Chem. Technol. Biotechnol.* 61, 123–138, <http://dx.doi.org/10.1002/jctb.280610207>.
- Meshkat, N., Anderson, C., Distefano 3rd, J.J., 2011. Finding identifiable parameter combinations in nonlinear ODE models and the rational reparameterization of their input–output equations. *Math. Biosci.* 233, 19–31, <http://dx.doi.org/10.1016/j.mbs.2011.06.001>.
- Mettetal, J.T., Muzzey, D., Gómez-Uribe, C., van Oudenaarden, A., 2008. The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*. *Science* 319, 482–484, <http://dx.doi.org/10.1126/science.1151582>.
- Minton, A.P., 2001. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* 276, 10577–10580, <http://dx.doi.org/10.1074/jbc.R100005200>.
- Minton, A.P., 2006. How can biochemical reactions within cells differ from those in test tubes? *J. Cell Sci.* 119, 2863–2869, <http://dx.doi.org/10.1242/jcs.03063>.
- Miskovic, L., Hatzimanikatis, V., 2010. Production of biofuels and biochemicals: in need of an ORACLE. *Trends Biotechnol.* 28, 391–397, <http://dx.doi.org/10.1016/j.tibtech.2010.05.003>.
- Miskovic, L., Hatzimanikatis, V., 2011. Modeling of uncertainties in biochemical reactions. *Biotechnol. Bioeng.* 108, 413–423, <http://dx.doi.org/10.1002/bit.22932>.
- Moisset, P., Vaisman, D., Cintolesi, A., Urrutia, J., Rapaport, I., Andrews, B.A., Asenjo, J.A., 2012. Continuous modeling of metabolic networks with gene regulation in yeast and in vivo determination of rate parameters. *Biotechnol. Bioeng.* 109, 2325–2339, <http://dx.doi.org/10.1002/bit.24503>.
- Moles, C.G., Mendes, P., Banga, J.R., 2003. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* 13, 2467–2474, <http://dx.doi.org/10.1101/gr.1262503>.
- Molinaro, A.M., Simon, R., Pfeiffer, R.M., 2005. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 21, 3301–3307, <http://dx.doi.org/10.1093/bioinformatics/bti499>.
- Nelder, J., Mead, R., 1965. A simplex method for function minimization. *Comput. J.* 7, 308–313.
- Nett, J.H., Gomathinayagam, S., Hamilton, S.R., Gong, B., Davidson, R.C., Du, M., Hopkins, D., Mitchell, T., Mallem, M.R., Nylen, A., Shaikh, S.S., Sharkey, N., Barnard, G.C., Copeland, V., Liu, L., Evers, R., Li, Y., Gray, P.M., Lingham, R.B., Visco, D., Forrest, G., DeMartino, J., Linden, T., Potgieter, T.J., Wildt, S., Stadhheim, T.A., d'Anjou, M., Li, H., Sethuraman, N., 2012. Optimization of erythropoietin production with controlled glycosylation-PEGylated erythropoietin produced in glycoengineered *Pichia pastoris*. *J. Biotechnol.* 157, 198–206, <http://dx.doi.org/10.1016/j.jbiotec.2011.11.002>.
- Neuner, A., Heinzel, E., 2011. Mixed glucose and lactate uptake by *Corynebacterium glutamicum* through metabolic engineering. *Biotechnol. J.* 6, 318–329, <http://dx.doi.org/10.1002/biot.201000307>.
- Nielsen, J., 1998. Metabolic engineering: techniques for analysis of targets for genetic manipulations. *Biotechnol. Bioeng.* 58, 125–132.
- Nielsen, J., Jewett, M.C., 2008. Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 8, 122–131, <http://dx.doi.org/10.1111/j.1567-1364.2007.00302.x>.
- Nikerel, I.E., van Winden, W.A., van Gulik, W.M., Heijnen, J.J., 2006. A method for estimation of elasticities in metabolic networks using steady state and dynamic metabolomics data and linlog kinetics. *BMC Bioinf.* 7, 540, <http://dx.doi.org/10.1186/1471-2105-7-540>.
- Nikerel, I.E., van Winden, W.A., Verheijen, P.J.T., Heijnen, J.J., 2009. Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics. *Metab. Eng.* 11, 20–30, <http://dx.doi.org/10.1016/j.ymben.2008.07.004>.
- Nikolaev, E.V., 2010. The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems. *Metab. Eng.* 12, 26–38, <http://dx.doi.org/10.1016/j.ymben.2009.08.010>.
- Nishio, Y., Usuda, Y., Matsui, K., Kurata, H., 2008. Computer-aided rational design of the phosphotransferase system for enhanced glucose uptake in *Escherichia coli*. *Mol. Syst. Biol.* 4, 160, <http://dx.doi.org/10.1038/msb4100201>.
- Noack, S., Wahl, A., Qeli, E., Wiechert, W., 2007. Visualizing regulatory interactions in metabolic networks. *BMC Biol.* 5, 46, <http://dx.doi.org/10.1186/1741-7007-5-46>.
- Nocedal, J., Wright, S.J., 1999. *Numerical Optimization*. Springer-Verlag, New York, NY.
- Nolan, R.P., Lee, K., 2011. Dynamic model of CHO cell metabolism. *Metab. Eng.* 13, 108–124, <http://dx.doi.org/10.1016/j.ymben.2010.09.003>.
- Nolan, R.P., Lee, K., 2012. Dynamic model for CHO cell engineering. *J. Biotechnol.* 158, 24–33, <http://dx.doi.org/10.1016/j.jbiotec.2012.01.009>.
- Österlund, T., Nookaew, I., Nielsen, J., 2012. Fifteen years of large scale metabolic modeling of yeast: developments and impacts. *Biotechnol. Adv.* 30, 979–988, <http://dx.doi.org/10.1016/j.biotechadv.2011.07.021>.
- Oddone, G.M., Mills, D.A., Block, D.E., 2009. A dynamic, genome-scale flux model of *Lactococcus lactis* to increase specific recombinant protein expression. *Metab. Eng.* 11, 367–381, <http://dx.doi.org/10.1016/j.ymben.2009.07.007>.
- Oh, E., Lu, M., Park, C., Park, C., Oh, H.B., Lee, S.Y., Lee, J., 2011. Dynamic modeling of lactic acid fermentation metabolism with *Lactococcus lactis*. *J. Microbiol. Biotechnol.* 21, 162–169.
- Okino, M.S., Mavrovouniotis, M.L., 1998. Simplification of mathematical models of chemical reaction systems. *Chem. Rev.* 98, 391–408.
- Oldiges, M., Noack, S., Wahl, A., Qeli, E., Freisleben, B., Wiechert, W., 2006. From enzyme kinetics to metabolic network modeling—visualization tool for enhanced kinetic analysis of biochemical network models. *Eng. Life Sci.* 6, 155–162.
- Oshiro, M., Shinto, H., Tashiro, Y., Miwa, N., Sekiguchi, T., Okamoto, M., Ishizaki, A., Sonomoto, K., 2009. Kinetic modeling and sensitivity analysis of xylose metabolism in *Lactococcus lactis* IO-1. *J. Biosci. Bioeng.* 108, 376–384, <http://dx.doi.org/10.1016/j.jbiotec.2009.05.003>.
- Otero, J.M., Nielsen, J., 2010. Industrial systems biology. *Biotechnol. Bioeng.* 105, 439–460, <http://dx.doi.org/10.1002/bit.22592>.
- Parachin, N.S., Bergdahl, B., van Niel, E.W.J., Gorwa-Grauslund, M.F., 2011. Kinetic modelling reveals current limitations in the production of ethanol from xylose by recombinant *Saccharomyces cerevisiae*. *Metab. Eng.* 13, 508–517, <http://dx.doi.org/10.1016/j.ymben.2011.05.005>.
- Park, J.H., Kim, T.Y., Lee, K.H., Lee, S.Y., 2011. Fed-batch culture of *Escherichia coli* for L-valine production based on *in silico* flux response analysis. *Biotechnol. Bioeng.* 108, 934–946, <http://dx.doi.org/10.1002/bit.22995>.
- Patil, K.R., Akesson, M., Nielsen, J., 2004. Use of genome-scale microbial models for metabolic engineering. *Curr. Opin. Biotechnol.* 15, 64–69, <http://dx.doi.org/10.1016/j.copbio.2003.11.003>.
- Paulsson, J., 2004. Summing up the noise in gene networks. *Nature* 427, 415–418, <http://dx.doi.org/10.1038/nature02257>.

- Pfeffer, M., Maurer, M., Kollensperger, G., Hann, S., Graf, A.B., Mattanovich, D., 2011. Modeling and measuring intracellular fluxes of secreted recombinant protein in *Pichia pastoris* with a novel 34S labeling procedure. *Microb. Cell Fact.* 10, 47, <http://dx.doi.org/10.1186/1475-2859-10-47>.
- Piazza, M., Feng, X.J., Rabinowitz, J.D., Rabitz, H., 2008. Diverse metabolic model parameters generate similar methionine cycle dynamics. *J. Theor. Biol.* 251, 628–639, <http://dx.doi.org/10.1016/j.jtbi.2007.12.009>.
- Pohjanpalo, H., 1978. System identifiability based on the power series expansion of the solution. *Math. Biosci.* 41, 21–33.
- Polisetti, P.K., Gatzke, E.P., Voit, E.O., 2008. Yield optimization of regulated metabolic systems using deterministic branch-and-reduce methods. *Biotechnol. Bioeng.* 99, 1154–1169, <http://dx.doi.org/10.1002/bit.21679>.
- Pozo, C., Marín-Sanguino, A., Alves, R., Guillén-Gosálbez, G., Jiménez, L., Sorribas, A., 2011. Steady-state global optimization of metabolic non-linear dynamic models through recasting into power-law canonical models. *BMC Syst. Biol.* 5, 137, <http://dx.doi.org/10.1186/1752-0509-5-137>.
- Prathumpai, W., Gabelgaard, J.B., Wanchanthuek, P., van de Vondervoort, P.J.L., de Groot, M.J.L., McIntrye, M., Nielsen, J., 2003. Metabolic control analysis of xylose catabolism in *Aspergillus*. *Biotechnol. Prog.* 19, 1136–1141, <http://dx.doi.org/10.1021/bp034020r>.
- Pritchard, L., Kell, D.B., 2002. Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis. *Eur. J. Biochem.* 269, 3894–3904.
- Provost, A., Bastin, G., 2004. Dynamic metabolic modelling under the balanced growth condition. *J. Process. Control* 14, 717–728.
- Provost, A., Bastin, G., Agathos, S.N., Schneider, Y.J., 2006. Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells. *Bioprocess. Biosyst. Eng.* 29, 349–366, <http://dx.doi.org/10.1007/s00449-006-0083-y>.
- Portner, R., Schäfer, T., 1996. Modelling hybridoma cell growth and metabolism—a comparison of selected models and data. *J. Biotechnol.* 49, 119–135.
- Ramsey, S.A., Smith, J.J., Orrell, D., Marelli, M., Petersen, T.W., de Atauri, P., Bolouri, H., Aitchison, J.D., 2006. Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. *Nat. Genet.* 38, 1082–1087, <http://dx.doi.org/10.1038/ng1869>.
- Raue, A., Becker, V., Klingmüller, U., Timmer, J., 2010. Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos* 20, 045105, <http://dx.doi.org/10.1063/1.3528102>.
- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25, 1923–1929, <http://dx.doi.org/10.1093/bioinformatics/btp358>.
- Raue, A., Kreutz, C., Maiwald, T., Klingmüller, U., Timmer, J., 2011. Addressing parameter identifiability by model-based experimentation. *IET Syst. Biol.* 5, 120–130, <http://dx.doi.org/10.1049/iet-syb.2010.0061>.
- Raue, A., Kreutz, C., Theis, F.J., Timmer, J., 2013. Joining forces of bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. *Philos. Trans. A. Math. Phys. Eng. Sci.* 371, 20110544, <http://dx.doi.org/10.1098/rsta.2011.0544>.
- Rodriguez-Fernandez, M., Egea, J.A., Banga, J.R., 2006a. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinf.* 7, 483, <http://dx.doi.org/10.1186/1471-2105-7-483>.
- Rodriguez-Fernandez, M., Mendes, P., Banga, J.R., 2006b. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems* 83, 248–265, <http://dx.doi.org/10.1016/j.biosystems.2005.06.016>.
- Rojas, I., Golebiowski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Wittig, U., 2007. Storing and annotating of kinetic data. *In Silico Biol.* 7, S37–S44.
- Roper, R.T., Saccomani, M.P., Vicini, P., 2010. Cellular signaling identifiability analysis: a case study. *J. Theor. Biol.* 264, 528–537, <http://dx.doi.org/10.1016/j.jtbi.2010.02.029>.
- Runarsson, T.P., Yao, X., 2000. Stochastic ranking for constrained evolutionary optimization. *IEEE Trans. Evol. Comput.* 4, 284–294.
- Runarsson, T.P., Yao, X., 2005. Search biases in constrained evolutionary optimization. *IEEE Trans. Syst. Man Cybern.—Part C* 35, 233–243.
- Saccomani, M.P., Audoly, S., Bellu, G., D'Angiò, L., 2010. Examples of testing global identifiability of biological and biomedical models with the DAISY software. *Comput. Biol. Med.* 40, 402–407, <http://dx.doi.org/10.1016/j.combiomed.2010.02.004>.
- Sarkar, D., Modak, J.M., 2003. Optimisation of fed-batch bioreactors using genetic algorithms. *Chem. Eng. Sci.* 58, 2283–2296.
- Sauer, M., Mattanovich, D., 2012. Construction of microbial cell factories for industrial bioprocesses. *J. Chem. Technol. Biotechnol.* 87, 445–450.
- Savageau, M.A., 1976. Biochemical Systems Analysis. Addison-Wesley Pub. Co., Reading, MA.
- Schaber, J., Baltanas, R., Bush, A., Klipp, E., Colman-Lerner, A., 2012. Modelling reveals novel roles of two parallel signalling pathways and homeostatic feedbacks in yeast. *Mol. Syst. Biol.* 8, 622, <http://dx.doi.org/10.1038/msb.2012.53>.
- Schaber, J., Klipp, E., 2011. Model-based inference of biochemical parameters and dynamic properties of microbial signal transduction networks. *Curr. Opin. Biotechnol.* 22, 109–116, <http://dx.doi.org/10.1016/j.copbio.2010.09.014>.
- Schaber, J., Liebermeister, W., Klipp, E., 2009. Nested uncertainties in biochemical models. *IET Syst. Biol.* 3, 1–9, <http://dx.doi.org/10.1049/iet-syb:20070042>.
- Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Sohngen, C., Stelzer, M., Thiele, J., Schomburg, D., 2011. BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.* 39, D670–D676, <http://dx.doi.org/10.1093/nar/gkq1089>.
- Schmid, J.W., Mauch, K., Reuss, M., Gilles, E.D., Kremling, A., 2004. Metabolic design based on a coupled gene expression-metabolic network model of tryptophan production in *Escherichia coli*. *Metab. Eng.* 6, 364–377, <http://dx.doi.org/10.1016/jymben.2004.06.003>.
- Schmidt, H., 2007. Sbaddon: high performance simulation for the systems biology toolbox for matlab. *Bioinformatics* 23, 646–647, <http://dx.doi.org/10.1093/bioinformatics/btl668>.
- Schmidt, H., Jirstrand, M., 2006. Systems biology toolbox for matlab: a computational platform for research in systems biology. *Bioinformatics* 22, 514–515, <http://dx.doi.org/10.1093/bioinformatics/bti799>.
- Schmidt, H., Madsen, M.F., Danø, S., Cedersund, G., 2008. Complexity reduction of biochemical rate expressions. *Bioinformatics* 24, 848–854, <http://dx.doi.org/10.1093/bioinformatics/btn035>.
- Schomburg, D., Schomburg, I., 2010. Enzyme databases. *Methods Mol. Biol.* 609, 113–128, http://dx.doi.org/10.1007/978-1-60327-241-4_7.
- Schwarz, G., 1978. Estimating the dimension of model. *Ann. Stat.* 6, 461–464.
- Secrier, M., Toni, T., Stumpf, M.P.H., 2009. The abc of reverse engineering biological signalling systems. *Mol. Biosyst.* 5, 1925–1935, <http://dx.doi.org/10.1039/b908951a>.
- Sedoglavic, A., 2002. A probabilistic algorithm to test local algebraic observability in polynomial time. *J. Symb. Comput.*, 735–755.
- Shinto, H., Tashiro, Y., Yamashita, M., Kobayashi, G., Sekiguchi, T., Hanai, T., Kuriya, Y., Okamoto, M., Sonomoto, K., 2007. Kinetic modeling and sensitivity analysis of acetone–butanol–ethanol production. *J. Biotechnol.* 131, 45–56, <http://dx.doi.org/10.1016/j.jbiotec.2007.05.005>.
- Skaar, N., 2008. Parameter Estimation Methods for Continuous Time Dynamical Systems Given Discrete Time Measurements (Master's thesis). Chalmers University of Technology.
- Smallbone, K., Simeonidis, E., Broomhead, D.S., Kell, D.B., 2007. Something from nothing: bridging the gap between constraint-based and kinetic modelling. *FEBS J.* 274, 5576–5585, <http://dx.doi.org/10.1111/j.1742-4658.2007.06076.x>.
- Smallbone, K., Simeonidis, E., Swainston, N., Mendes, P., 2010. Towards a genome-scale kinetic model of cellular metabolism. *BMC Syst. Biol.* 4, 6, <http://dx.doi.org/10.1186/1752-0509-4-6>.
- Snoep, J.L., Bruggeman, F., Olivier, B.G., Westerhoff, H.V., 2006. Towards building the silicon cell: a modular approach. *Biosystems* 83, 207–216, <http://dx.doi.org/10.1016/j.biosystems.2005.07.006>.
- Soh, K.C., 2013. Computational Studies on Cellular Bioenergetics (Ph.D. thesis). Thèse École polytechnique fédérale de Lausanne EPFL.
- Soh, K.C., Miskovic, L., Hatzimanikatis, V., 2012. From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. *FEMS Yeast Res.* 12, 129–143, <http://doi.org/10.1111/j.1567-1364.2011.00771.x>.
- Sohn, S.B., Graf, A.B., Kim, T.Y., Gasser, B., Maurer, M., Ferrer, P., Mattanovich, D., Lee, S.Y., 2010. Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for *in silico* analysis of heterologous protein production. *Biotechnol. J.* 5, 705–715, <http://dx.doi.org/10.1002/biot.201000078>.
- Srinath, S., Gunawan, R., 2010. Parameter identifiability of power-law biochemical system models. *J. Biotechnol.* 149, 132–140, <http://dx.doi.org/10.1016/j.jbiotec.2010.02.019>.
- Sunnåker, M., Schmidt, H., Jirstrand, M., Cedersund, G., 2010. Zooming of states and parameters using a lumping approach including back-translation. *BMC Syst. Biol.* 4, 28, <http://dx.doi.org/10.1186/1752-0509-4-28>.
- Sunnåker, M., Cedersund, G., Jirstrand, M., 2011. A method for zooming of nonlinear models of biochemical systems. *BMC Syst. Biol.* 5, 140, <http://dx.doi.org/10.1186/1752-0509-5-140>.
- Takors, R., Wiechert, W., Weuster-Botz, D., 1997. Experimental design for the identification of macrokinetic models and model discrimination. *Biotechnol. Bioeng.* 56, 564–576 <http://dx.doi.org/3.0.CO;2-C>.
- Tang, W.L., Zhao, H., 2009. Industrial biotechnology: tools and applications. *Biotechnol. J.* 4, 1725–1739, <http://dx.doi.org/10.1002/biot.200900127>.
- Teixeira, A.P., Alves, C., Alves, P.M., Carrondo, M.J.T., Oliveira, R., 2007. Hybrid elementary flux analysis nonparametric modeling: application for bioprocess control. *BMC Bioinf.* 8, 30, <http://dx.doi.org/10.1186/1471-2105-8-30>.
- Teusink, B., Passarge, J., Reijenga, C.A., Esenalıoğlu, E., van der Weijden, C.C., Schepper, M., Walsh, M.C., Bakker, B.M., van Dam, K., Westerhoff, H.V., Snoep, J.L., 2000. Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochem.* 267, 5313–5329.
- Theilgaard, H.A., Nielsen, J., 1999. Metabolic control analysis of the penicillin biosynthetic pathway: the influence of the LLD-ACV:bisACV ratio on the flux control. *Antonie Van Leeuwenhoek* 75, 145–154.
- Thiele, I., Jamshidi, N., Fleming, R.M.T., Palsson, B.O., 2009. Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput. Biol.* 5, e1000312, <http://dx.doi.org/10.1371/journal.pcbi.1000312>.
- Transtrum, M.K., Machta, B.B., Sethna, J.P., 2010. Why are nonlinear fits to data so challenging? *Phys. Rev. Lett.* 104, 060201.
- Trusina, A., Papa, F.R., Tang, C., 2008. Rationalizing translation attenuation in the network architecture of the unfolded protein response. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20280–20285, <http://dx.doi.org/10.1073/pnas.0803476105>.
- Trusina, A., Tang, C., 2010. The unfolded protein response and translation attenuation: a modelling approach. *Diabetes Obes. Metab.* 12 (Suppl. 2), 27–31, <http://dx.doi.org/10.1111/j.1463-1326.2010.01274.x>.

- Tyo, K.E.J., Kocharin, K., Nielsen, J., 2010. Toward design-based engineering of industrial microbes. *Curr. Opin. Microbiol.* 13, 255–262, <http://dx.doi.org/10.1016/j.mib.2010.02.001>.
- Tyson, J.J., Chen, K.C., Novak, B., 2003. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* 15, 221–231.
- Ullah, M., Wolkenhauer, O., 2010. Stochastic approaches in systems biology. Wiley Interdiscip. Rev. Syst. Biol. Med. 2, 385–397, <http://dx.doi.org/10.1002/wsbm.78>.
- Umaña, P., Bailey, J.E., 1997. A mathematical model of N-linked glycoform biosynthesis. *Biotechnol. Bioeng.* 55, 890–908 <http://dx.doi.org/3.0.CO;2-B>.
- Usuda, Y., Nishio, Y., Iwatanai, S., Dien, S.J.V., Imaizumi, A., Shimbo, K., Kageyama, N., Iwahata, D., Miyano, H., Matsui, K., 2010. Dynamic modeling of *Escherichia coli* metabolic and regulatory systems for amino-acid production. *J. Biotechnol.* 147, 17–30, <http://dx.doi.org/10.1016/j.biote.2010.02.018>.
- Vajda, S., Godfrey, K.R., Rabitz, H., 1989. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. *Math. Biosci.* 93, 217–248.
- van Riel, N.A.W., 2006. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Br. Bioinf.* 7, 364–374, <http://dx.doi.org/10.1093/bib/bbl040>.
- Visser, D., Heijnen, J.J., 2002. The mathematics of metabolic control analysis revisited. *Metab. Eng.* 4, 114–123, <http://dx.doi.org/10.1006/mben.2001.0216>.
- Visser, D., Heijnen, J.J., 2003. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metab. Eng.* 5, 164–176.
- Visser, D., Schmid, J.W., Mauch, K., Reuss, M., Heijnen, J.J., 2004. Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics. *Metab. Eng.* 6, 378–390, <http://dx.doi.org/10.1016/j.ymben.2004.07.001>.
- Vital-Lopez, F.G., Armaou, A., Niklaev, E.V., Maranas, C.D., 2006. A computational procedure for optimal engineering interventions using kinetic models of metabolism. *Biotechnol. Prog.* 22, 1507–1517, <http://dx.doi.org/10.1021/bp0601560>.
- Wang, F.S., Ko, C.L., Voit, E.O., 2007. Kinetic modeling using S-systems and lin-log approaches. *Biochem. Eng. J.* 33, 238–247.
- Wang, L., Birol, I., Hatzimanikatis, V., 2004. Metabolic control analysis under uncertainty: framework development and case studies. *Biophys. J.* 87, 3750–3763, <http://dx.doi.org/10.1529/biophysj.104.048090>.
- Wang, L., Hatzimanikatis, V., 2006a. Metabolic engineering under uncertainty. II. Analysis of yeast metabolism. *Metab. Eng.* 8, 142–159, <http://dx.doi.org/10.1016/j.ymben.2005.11.002>.
- Wang, L., Hatzimanikatis, V., 2006b. Metabolic engineering under uncertainty. I. Framework development. *Metab. Eng.* 8, 133–141, <http://dx.doi.org/10.1016/j.ymben.2005.11.003>.
- Wang, X., Hao, N., Dohlman, H.G., Elston, T.C., 2006. Bistability, stochasticity, and oscillations in the mitogen-activated protein kinase cascade. *Biophys. J.* 90, 1961–1978, <http://dx.doi.org/10.1529/biophysj.105.073874>.
- Wiechert, W., Noack, S., 2011. Mechanistic pathway modeling for industrial biotechnology: challenging but worthwhile. *Curr. Opin. Biotechnol.* 22, 604–610, <http://dx.doi.org/10.1016/j.copbio.2011.01.001>.
- Wiseman, R.L., Powers, E.T., Buxbaum, J.N., Kelly, J.W., Balch, W.E., 2007. An adaptable standard for protein export from the endoplasmic reticulum. *Cell* 131, 809–821, <http://dx.doi.org/10.1016/j.cell.2007.10.025>.
- Xing, Z., Bishop, N., Leister, K., Li, Z.J., 2010. Modeling kinetics of a large-scale fed-batch CHO cell culture by Markov chain Monte Carlo method. *Biotechnol. Prog.* 26, 208–219, <http://dx.doi.org/10.1002/btpr.284>.
- Ye, J., Ly, J., Watts, K., Hsu, A., Walker, A., McLaughlin, K., Berdichevsky, M., Prinz, B., Sean Kersey, D., d'Anjou, M., Pollard, D., Potgieter, T., 2011. Optimization of a glycoengineered *Pichia pastoris* cultivation process for commercial antibody production. *Biotechnol. Prog.* 27, 1744–1750, <http://dx.doi.org/10.1002/btpr.695>.
- Yüzgeç, U., Türker, M., Hocalar, A., 2009. On-line evolutionary optimization of an industrial fed-batch yeast fermentation process. *ISA Trans.* 48, 79–92, <http://dx.doi.org/10.1016/j.isatra.2008.09.001>.
- Zamorano, F., Vande Wouwer, A., Jungers, R.M., Bastin, G., 2013. Dynamic metabolic models of CHO cell cultures through minimal sets of elementary flux modes. *J. Biotechnol.* 164, 409–422, <http://dx.doi.org/10.1016/j.jbbiotec.2012.05.005>.
- Zi, Z., Klipp, E., 2006. SBML-PET: a systems biology markup language-based parameter estimation tool. *Bioinformatics* 22, 2704–2705, <http://dx.doi.org/10.1093/bioinformatics/btl443>.
- Zi, Z., Liebermeister, W., Klipp, E., 2010. A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in *Saccharomyces cerevisiae*. *PLoS One* 5, e9522, <http://dx.doi.org/10.1371/journal.pone.0009522>.

PAPER B

Unraveling the Pharmacokinetic Interaction of Ticagrelor and MEDI2452 (Ticagrelor Antidote) by Mathematical Modeling

ORIGINAL ARTICLE

Unraveling the Pharmacokinetic Interaction of Ticagrelor and MEDI2452 (Ticagrelor Antidote) by Mathematical Modeling

J Almquist^{1,2,3}, M Penney⁴, S Pehrsson³, A-S Sandinge³, A Janefeldt³, S Maqbool⁴, S Madalli⁵, J Goodman⁴, S Nylander³ and P Gennemark^{3*}

The investigational ticagrelor-neutralizing antibody fragment, MEDI2452, is developed to rapidly and specifically reverse the antiplatelet effects of ticagrelor. However, the dynamic interaction of ticagrelor, the ticagrelor active metabolite (TAM), and MEDI2452, makes pharmacokinetic (PK) analysis nontrivial and mathematical modeling becomes essential to unravel the complex behavior of this system. We propose a mechanistic PK model, including a special observation model for post-sampling equilibration, which is validated and refined using mouse *in vivo* data from four studies of combined ticagrelor-MEDI2452 treatment. Model predictions of free ticagrelor and TAM plasma concentrations are subsequently used to drive a pharmacodynamic (PD) model that successfully describes platelet aggregation data. Furthermore, the model indicates that MEDI2452-bound ticagrelor is primarily eliminated together with MEDI2452 in the kidneys, and not recycled to the plasma, thereby providing a possible scenario for the extrapolation to humans. We anticipate the modeling work to improve PK and PD understanding, experimental design, and translational confidence.

CPT Pharmacometrics Syst. Pharmacol. (2016) 00, 00; doi:10.1002/psp4.12089; published online on 0 Month 2016.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC? Antiplatelet therapy for the prevention of atherothrombotic events in patients with acute coronary syndrome is known to increase the risk of bleeding complications. The unique reversible binding mode of action of the oral antiplatelet agent, ticagrelor, has enabled the development of MEDI2452, a ticagrelor-specific antidote for rare emergency situations. • WHAT QUESTION DOES THIS STUDY ADDRESS? This study seeks to unravel the PK interaction between ticagrelor and MEDI2452 through the use of mathematical modeling, aiming for both qualitative understanding as well as detailed quantitative predictions. • WHAT THIS STUDY ADDS TO OUR KNOWLEDGE A mathematical model describing the simultaneous PKs of ticagrelor and MEDI2452 in the mouse is presented. The model offers a mechanistic explanation for the complex kinetics and can predict the unobserved free ticagrelor plasma concentration that drives the platelet aggregation PDs. • HOW THIS MIGHT CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS The combined ticagrelor-MEDI2452 model can contribute to the development of MEDI2452 by assisting in interpretation of observed data, by prediction of free ticagrelor and TAM plasma concentrations, and by simulation of experimental designs.

Ticagrelor is a direct acting and reversibly binding P2Y₁₂ antagonist.¹ In the PLATO study, a positive benefit-risk profile for the prevention of atherothrombotic events in patients with acute coronary syndrome was established.² Based on the PLATO data, as well as data with the thienopyridines, clopidogrel in CURÉ,³ and prasugrel in TRITON,⁴ dual anti-platelet therapy, consisting of aspirin and a P2Y₁₂ antagonist, is critical for the treatment of acute coronary syndrome. In addition, in the recently completed PEGASUS TIMI-54, the benefit-risk profile for long-term treatment with ticagrelor in patients with a history of myocardial infarction and a high risk of developing an atherothrombotic event was documented supporting the use of ticagrelor for long-term treatment.⁵ However, all antiplatelet therapies are known to increase the risk of bleeding complications.^{2–5}

The unique reversibly binding mode of action of ticagrelor provides an opportunity for developing a specific reversal agent not possible for the thienopyridines, which are all irreversible P2Y₁₂ antagonists. The first data for the ticagrelor-specific neutralizing antibody fragment (Fab), MEDI2452, has recently been published.⁶

MEDI2452 specifically binds free (unbound to plasma proteins) ticagrelor and free AR-C124910XX—the ticagrelor active metabolite (TAM)—with a high affinity of about 20 pM.^{6,7} MEDI2452 thereby prevents ticagrelor's and TAM's (similar potency vs. P2Y₁₂ as ticagrelor) interaction with, and inhibition of, the platelet P2Y₁₂ receptor. MEDI2452 may prove valuable as an agent for patients on ticagrelor therapy who require urgent surgery or suffer an acute major bleed. If successful, MEDI2452 would be the

¹Fraunhofer-Chalmers Centre, Chalmers Science Park, Gothenburg, Sweden; ²Systems and Synthetic Biology, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden; ³Cardiovascular and Metabolic Diseases, Innovative Medicines, AstraZeneca R&D, Mölndal, Sweden; ⁴Clinical Pharmacology and DMPK, MedImmune, Cambridge, UK; ⁵Cardiovascular and Metabolic Diseases Research, MedImmune, Cambridge, UK. *Correspondence: P Gennemark (peter.gennemark@astrazeneca.com)

Received 8 February 2016; accepted 4 May 2016; published online on 0 Month 2016. doi:10.1002/psp4.12089

first antidote for an antiplatelet drug. Recently, specific antidotes for oral anticoagulants have been described and are undergoing clinical trials, including a specific Fab for the thrombin inhibitor dabigatran, idarucizumab,^{8–11} and a recombinant catalytically inactive Factor (F)Xa, andexanet alfa, which is targeted to reverse all the FXa inhibitors.¹²

The main objective of the present work is to better understand the pharmacokinetics (PKs) of ticagrelor, TAM, and MEDI2452, in general, and to be able to predict free plasma concentrations of ticagrelor and TAM in particular. Knowledge of free plasma concentrations is crucial as it is only the free fractions—and not the more than 99% protein-bound fraction or the dynamically changing MEDI2452-bound fraction—that can inhibit the P2Y₁₂ receptor and thereby drive the pharmacodynamic (PD) response. The PK analysis is complex because of the dynamic interaction that occurs between these compounds when administered to the same system. It is expected that mathematical modeling will be essential for unraveling their combined PK behavior. In addition, we investigate if a simple turnover model driven by the predicted free levels can describe PD data of platelet aggregation. The work was accomplished in three stages. First, a mathematical model of the combined ticagrelor-MEDI2452 PK in the mouse was set up based on data of separately administered ticagrelor and MEDI2452, and on assumptions supported by literature. Second, the model was validated and refined using several different combined ticagrelor-MEDI2452 PK datasets not used in the first stage. Finally, the model was used to understand how the complex PK emerges from the ticagrelor-MEDI2452 interaction. Specifically, we let the model predict free plasma levels of ticagrelor and TAM under different experimental designs, and, in turn, let these predictions drive the PD model.

MATERIALS AND METHODS

MEDI2452 prestudy: dosing only MEDI2452 to rats

MEDI2452 PK was observed following a 1,000 mg × kg⁻¹ i.v. bolus dose in conscious Sprague-Dawley rats. Venous serum samples were obtained predose and at 5, 15, and 30 minutes, and 1, 2, 6, 12, 24, and 48 hours postdose. The study was performed by Huntingdon Life Sciences (Huntingdon, UK) in compliance with the United Kingdom Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012.

Ticagrelor prestudies: dosing only ticagrelor to mice

Ticagrelor PK in nonfasted male C57Bl6 mice (Charles River, Sulzfeld, Germany; body weight in the range of 15–25 g) was observed in two studies, of which the first also included TAM observations. The studies were approved by the ethical committee for animal research at the University of Göteborg, Sweden. Mice were anesthetized with isoflurane vapor (Forene; Abbot Scandinavia AB, Sweden). A catheter was inserted in the left jugular vein for administration of the vehicle or drug. The body temperature was maintained at 38°C by external heating. In the first study, four mice were given ticagrelor as an i.v. infusion at a rate of 240 µg × min⁻¹ × kg⁻¹ for 5 minutes, followed by 30 µg × min⁻¹ × kg⁻¹ for 15 minutes. The ticagrelor and

TAM total concentrations (free and protein-bound) in plasma were observed at 20, 30, 40, 50, 65, and 80 minutes after the start of infusion (thus, the first sample was collected immediately poststop of infusion). In the second study, two mice were administered an i.v. bolus dose of 2,000 µg × kg⁻¹ ticagrelor. The total ticagrelor concentration in blood was measured in samples collected at 2, 5, 10, 30, 60, 90, 120, and 150 minutes after the dose.

Main studies: dosing MEDI2452 to ticagrelor-treated mice

Four different studies in nonfasted male C57Bl6 mice were used for model validation and refinement, labeled study 1 to 4. Their designs are illustrated in Figure 1, and the full details are given in Supplementary Text S1. Animal handling and experimental setup were as described for the ticagrelor prestudies. In common with these studies, the mice were first dosed with an i.v. infusion of ticagrelor, allowing the plasma concentration to reach steady-state. Then, a bolus of the antidote MEDI2452 was administered. The PD data from studies 1, 3, and 4 have previously been reported.⁶

Bioanalysis

Quantification of total (free and protein bound) as well as free (protein unbound) ticagrelor and TAM in plasma samples in the mouse was determined by protein precipitation and liquid chromatography mass spectrometry. The lower limits of quantification of free ticagrelor and TAM were 0.03 nmol L⁻¹ and 0.06 nmol L⁻¹. Quantification of MEDI2452 in mouse and rat plasma was performed with the Gyrolab nanoliter scale immunoassay assay platform. Mouse and rat assays had lower limits of quantification of 260 and 350 ng × mL⁻¹. Quantification of ADP-induced platelet aggregation was evaluated using the Multiplate impedance aggregometer (Dynabyte, Munich, Germany). Further details of the bioanalysis are given in Supplementary Text S1.

Data analysis

Parameter estimation was performed according to a maximum likelihood approach with a multiplicative lognormal error model for the PK model and a mixed additive and multiplicative error model for the PD model, using the naïve-pooled data approach. Uncertainty of parameter estimates was determined by bootstrapping, sampling single measurements randomly with replacement within each experiment ($N = 300$). Uncertainty of the final PK model used for predictions and for driving the PD model was generated based on Monte Carlo sampling from the parameter distributions obtained from bootstrapping, and from additional parameter uncertainties defined in Table 1. In addition to uncertainty with respect to parameter values, model predictions also incorporate the effect of uncertainty resulting from the model's residual variability (Supplementary Text S2).

Numeric analyses were performed in MATLAB (R2014a; The MathWorks, Natick, MA). Specifically, the Matlab function *fminsearch* was used for solving the optimization problems encountered during parameter estimation. The Matlab model code is provided in the Supplementary Model Code files. The analytical solution used in the observation model was derived in Mathematica 10 (Wolfram Research, Champaign, IL).

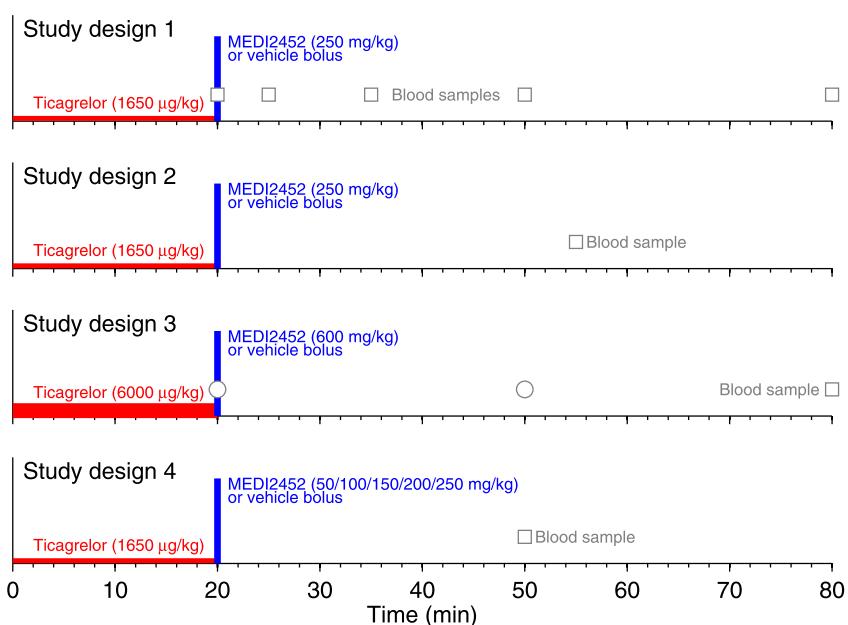


Figure 1 Designs of validation studies 1 to 4. Mice were given ticagrelor as an i.v. infusion during 20 minutes. Immediately after the ticagrelor infusion, an i.v. bolus dose of MEDI2452 or vehicle was given. Terminal blood samples for plasma exposure of ticagrelor and MEDI2452 were collected at various time-points. For study design 3, gray circles indicate blood samples from a separate pharmacokinetic experiment.

RESULTS

PK models for only ticagrelor or MEDI2452

Before the combined ticagrelor-MEDI2452 interaction model could be set up, two independent PK models were derived for the separate administration of ticagrelor in the mouse

and of MEDI2452 in the rat. The details of these models are described in **Supplementary Text S2**, **Supplementary Figure S1**, and **Supplementary Figure S2**. We found that ticagrelor (and TAM) follows two-compartment kinetics in the absence of MEDI2452, and that MEDI2452 kinetics in

Table 1 Model parameters

Name	Unit	Value	Estimated value	Uncertainty	Explanation
f	–	0.0020	0.0020 (not estimated)	0.0012, 0.0028 ^a	Fraction unbound ticagrelor (internal data, $n = 38$)
k_{on}	$\text{nM}^{-1} \times \text{min}^{-1}$	0.14	0.11	0.078, 0.122 ^b	Second-order rate constant. Initial point estimate obtained from different estimates of the rate constants of the similarly sized ligands methotrexate, dabigatran, and topotecan. ^{8,17,20,27}
K_d	nM	0.02	0.02 (not estimated)	0.013, 0.029 ^c	Affinity of MEDI2452 for ticagrelor and TAM. ⁶
Cl_{met}	$\text{L} \times \text{min}^{-1} \times \text{kg}^{-1}$	0.012	0.0080	0.0077, 0.0083 ^b	TAM-specific ticagrelor clearance
Cl	$\text{L} \times \text{min}^{-1} \times \text{kg}^{-1}$	0.058	0.022	0.019, 0.025 ^b	Remaining ticagrelor clearance
V_1	$\text{L} \times \text{kg}^{-1}$	1.63	1.12	0.95, 1.18 ^b	Initial estimate obtained by $V_{1,Tica} - V_{MEDI2452} = 1.68 - 0.05 \text{ L} \times \text{kg}^{-1}$
Cl_d	$\text{L} \times \text{min}^{-1} \times \text{kg}^{-1}$	0.042	0.041	0.050, 0.066 ^b	Ticagrelor intercompartmental clearance
V_2	$\text{L} \times \text{kg}^{-1}$	1.8	1.8	1.32, 1.88 ^b	Ticagrelor volume of second compartment
V	$\text{L} \times \text{kg}^{-1}$	0.05	0.05 (not estimated)	–	MED2452 (central compartment) volume of distribution. Standard plasma volume.
Cl_f	$\text{L} \times \text{min}^{-1} \times \text{kg}^{-1}$	0.0030	0.0025	0.0024, 0.0026 ^b	MED2452 Cl. Initial estimate scaled value from rat ($0.0048 \text{ L} \times \text{min}^{-1} \times \text{kg}^{-1}$) adjusted to 60% to match data.
Cl_{fast}	$\text{L} \times \text{min}^{-1} \times \text{kg}^{-1}$	10	10 (not estimated)	–	Rapid compared to other clearances in the system
σ^2_{Tica}	$\text{nM} \times \text{nM}$	0.076	0.064, 0.081 ^b		
σ^2_{TAM}	$\text{nM} \times \text{nM}$	0.080	0.076, 0.087 ^b		
σ^2_{MEDI}	$\text{nM} \times \text{nM}$	0.28	0.26, 0.31 ^b		
$\sigma^2_{freetica}$	$\text{nM} \times \text{nM}$	0.042	0.039, 0.045 ^b		
$\sigma^2_{freeTAM}$	$\text{nM} \times \text{nM}$	0.060	0.057, 0.069 ^b		

TAM, ticagrelor active metabolite; Tica, ticagrelor.

^aFifth and 95th percentiles obtained from internal data ($n = 38$). ^bFifth and 95th percentiles obtained from bootstrapping. ^c95% confidence interval.⁶

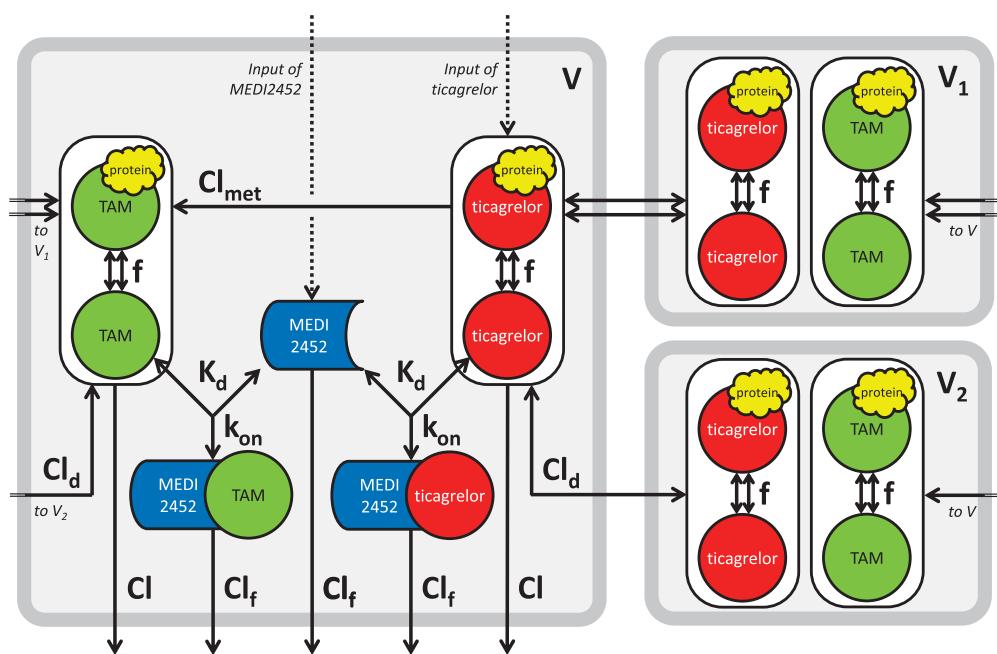


Figure 2 Schematic illustration of the combined ticagrelor-MEDI2452 pharmacokinetic-model. Reactions assumed to equilibrate instantaneously are indicated by double arrows. Input to the system (ticagrelor and MEDI2452) are shown as dashed arrows. The rapid equilibria of free and protein-bound ticagrelor and ticagrelor active metabolite (TAM) are depicted by encapsulated entities. The fractions of free ticagrelor and TAM within these entities are determined by the parameter f . The total contents of free and protein-bound ticagrelor and TAM in the plasma compartment (V) are cleared at the rate Cl , and ticagrelor is additionally being metabolized to TAM at the rate Cl_{met} . The total content of the encapsulated ticagrelor entity may furthermore distribute instantaneously to one peripheral compartment (V_1), and more slowly, with the intercompartmental clearance Cl_d , to another (V_2). Free ticagrelor and TAM in the plasma compartment can reversibly bind to free MEDI2452 with the rate k_{on} , forming complexes with dissociation constant K_d . Both the complexes and free MEDI2452 are cleared at the rate Cl_f .

the absence of ticagrelor was adequately described using a one-compartment model during the timescale of interest. This timescale was dictated by the experiments in which both compounds were administered and occurred over 80 minutes after drug dosing. The parameters of the rat MEDI2452 model were furthermore adjusted based on allometric scaling and *a priori* knowledge of mouse plasma volume to describe the presumed kinetics in the mouse (Supplementary Text S2).

A combined ticagrelor-MEDI2452 PK model

A combined ticagrelor-MEDI2452 PK model was hypothesized by integrating the independently derived models for the respective compounds. To this end, assumptions were required with respect to the reconciliation of compartment structures of the two independent models and with respect to certain kinetic matters brought to the fore by the model integration.

Because the central compartment of the independent MEDI2452 model equaled the mouse plasma volume, and because the larger central compartment of the independent ticagrelor and TAM model could be reasonably expected to contain the plasma volume, we assumed the central compartment of the independent ticagrelor and TAM model to be divided into two subcompartments for the hypothesized combined model. A plasma compartment that is identical to the central compartment of independent MEDI2452 model,

with the previously introduced volume V (Supplementary Text S2), and a compartment representing tissue rapidly exchanging with plasma, with the new volume V_1 , now representing the remainder of the previous ticagrelor and TAM central compartment. Thus, V accommodates all compounds, whereas V_1 is only for ticagrelor and TAM. The peripheral compartment with volume V_2 of the independent ticagrelor and TAM model (Supplementary Text S2) was kept unchanged.

We further assumed that: (I) binding and unbinding of ticagrelor and TAM to plasma protein is fast (i.e., the reaction is considered to reach equilibrium instantaneously); (II) distribution between V and V_1 also is fast in the above sense; (III) the ticagrelor-MEDI2452 complex is cleared at the same rate as MEDI2452; and (IV) the kinetics of TAM is governed by the same principles as ticagrelor (i.e., the equations are structurally equivalent, using the same parameters as for ticagrelor). Assumption (I) is a standard assumption within PK modeling because this equilibrium typically is reached on a timescale of seconds, and it is justified in our case in which the other dynamics occur on a timescale of minutes. Assumption (II) is reasonable because we were not able to improve the separate ticagrelor PK model by adding a third compartment. Assumption (III) is natural if no other information is available and it has been used in similar modeling situations.^{13–15} Regarding assumption (IV), all aspects of TAM kinetics have not been

investigated, but it is known that the affinities of both MEDI2452 and P2Y₁₂ for TAM,⁶ (data on file) are highly similar to the corresponding affinities for ticagrelor.^{6,16} The possibility of target-mediated drug disposition was excluded from the model according to the discussion in **Supplementary Text S2**.

The model is illustrated in **Figure 2**. It is defined by the following equations:

$$\begin{aligned} V \times TicaV(t)' &= -Cl_{fast} \times (TicaV(t) - TicaV_1(t)) \\ &\quad - Cl_d \times (TicaV(t) - TicaV_2(t)) \end{aligned} \quad (1)$$

$$\begin{aligned} - V \times k_{on} (f \times TicaV(t) \times FabV(t) - K_d \times FabTicaV(t)) \\ - Cl_{met} \times TicaV(t) - Cl \times TicaV(t) + TicalInput(t), \end{aligned}$$

$$V_1 \times TicaV_1(t)' = Cl_{fast} \times (TicaV(t) - TicaV_1(t)), \quad (2)$$

$$V_2 \times TicaV_2(t)' = Cl_d \times (TicaV(t) - TicaV_2(t)), \quad (3)$$

$$\begin{aligned} V \times TamV(t)' &= -Cl_{fast} \times (TamV(t) - TamV_1(t)) \\ &\quad - Cl_d \times (TamV(t) - TamV_2(t)) \end{aligned} \quad (4)$$

$$\begin{aligned} - V \times k_{on} (f \times TamV(t) \times FabV(t) - K_d \times FabTamV(t)) \\ + Cl_{met} \times TicaV(t) - Cl \times TamV(t), \end{aligned}$$

$$V_1 \times TamV_1(t)' = Cl_{fast} \times (TamV(t) - TamV_1(t)), \quad (5)$$

$$V_2 \times TamV_2(t)' = Cl_d \times (TamV(t) - TamV_2(t)), \quad (6)$$

$$\begin{aligned} V \times FabV(t)' &= - V \times k_{on} (f \times TicaV(t) \times FabV(t) - K_d \times FabTicaV(t)) \\ &\quad - V \times k_{on} (f \times TamV(t) \times FabV(t) - K_d \times FabTamV(t)) \\ &\quad - Cl_f \times FabV(t) + FabInput(t), \end{aligned} \quad (7)$$

$$\begin{aligned} V \times FabTicaV(t)' &= V \times k_{on} (f \times TicaV(t) \times FabV(t) - K_d \times FabTicaV(t)) \\ &\quad - Cl_f \times FabTicaV(t), \end{aligned} \quad (8)$$

$$\begin{aligned} V \times FabTamV(t)' &= V \times k_{on} (f \times TamV(t) \times FabV(t) - K_d \times FabTamV(t)) \\ &\quad - Cl_f \times FabTamV(t), \end{aligned} \quad (9)$$

where the state variables $TicaV(t)$, $TicaV_1(t)$, $TicaV_2(t)$, $TamV(t)$, $TamV_1(t)$, and $TamV_2(t)$ refer to the time-dependent concentrations of free and protein-bound (not including MEDI2452) ticagrelor and TAM in V , V_1 , and V_2 , respectively, where $FabV(t)$ refers to MEDI2452 concentration in V , and where $FabTicaV(t)$ and $FabTamV(t)$ refer to the ticagrelor-MEDI2452, and TAM-MEDI2452 complexes in V . Values of the parameters, and their justifications, are reported in **Table 1**.

Observation model

The reactions in which ticagrelor and TAM form complexes with MEDI2452 do, in general, never fully reach their equilibria

in the *in vivo* system. This is due to the continuously ongoing clearance and redistribution of the interacting species, which interferes with the processes of reaching the equilibria. When blood samples are collected, however, physiological reactions, like clearances and intercompartmental distribution, are interrupted, and there is sufficient time available for reaching the equilibria before the bioanalysis of the sample content is complete. Because of this, a model of the dynamic process taking place *in vitro* is required in addition to the original model of the *in vivo* system. Such an observation model can mathematically be constructed by setting the values of all clearance parameters to zero, reflecting the *in vitro* conditions of the bioanalysis, and then compute the closed-form analytical solution for the equilibria (**Supplementary Text S2**). The use of the observation model is critical for distinguishing between actual levels of free ticagrelor and TAM, and of measured values (**Supplementary Figure S3**). Specifically, the observed concentrations of free ticagrelor and free TAM are less than the concentrations of free ticagrelor and free TAM in the *in vivo* system. For free levels of MEDI2452, which is present in larger concentrations compared to free ticagrelor and TAM, the difference of using the observation model becomes marginal.

The hypothesized model explains observed data from four separate studies

The model was validated on experimental data from study 1 to 4, comprising different dosing and sampling schedules (**Figure 1**). By a reduction (40%) of the allometrically scaled MEDI2452 clearance (corresponding to scaling rat-mouse with exponent 1 and not 0.75), and without changing any other parameter value, we could reasonably well simulate all the qualitative characteristics of the observed data (**Figure 3**), as well as many of the quantitative characteristics. For studies 1 to 3, the predicted time profiles of plasma levels of total ticagrelor, total TAM, and free MEDI2452 were generally in accordance with data, although there was some underprediction of total ticagrelor and total TAM in the absence of MEDI2452, especially in study 3. For study 4, both model and data displayed increasing levels of total ticagrelor, total TAM, and free MEDI2452, as a function of dose, as well as decreasing levels of free ticagrelor and free TAM as a function of dose. Although the model performed reasonably well for total and free ticagrelor and TAM in study 4, there was a tendency to underpredict free ticagrelor and free TAM and at lower doses there was an overprediction of free MEDI2452.

The validation data were subsequently used to refine the model by reestimation of some of its parameters (**Table 1**). As shown in **Figure 3**, the refined model improved the fit to data in general, including – to us – the particularly important levels of free ticagrelor and free TAM. A residual plot for the refined model is presented in **Supplementary Figure S4**.

Ticagrelor is not likely to be fully recycled

When ticagrelor-bound or TAM-bound MEDI2452 is eliminated in the kidneys, it is currently not known if ticagrelor and TAM are eliminated, recycled, or a combination thereof. Fabs are, however, generally cleared through the kidneys and elimination is faster than for immunoglobulin G (IgG).¹⁷

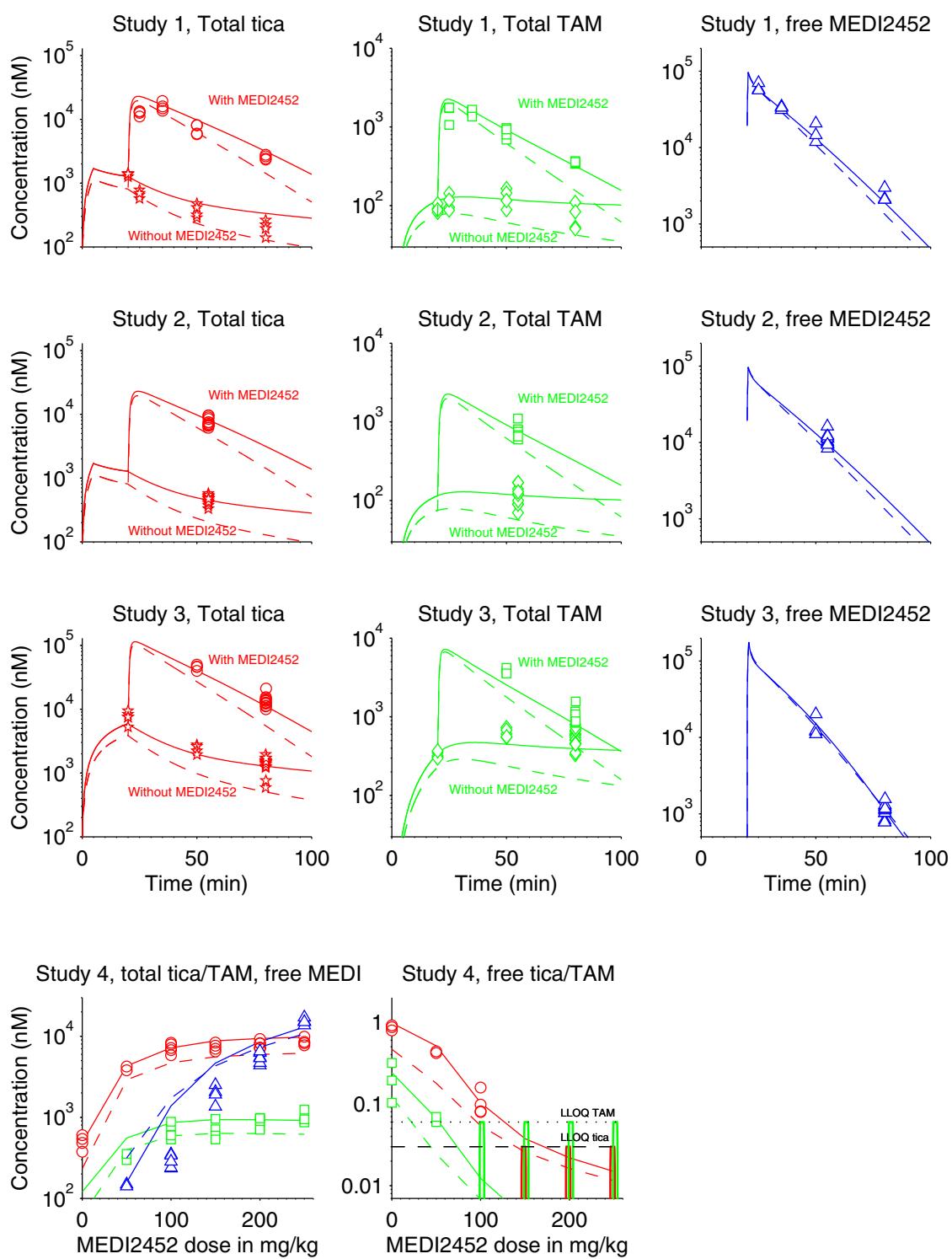


Figure 3 Model validation and refinement. Observed total and free ticagrelor (tica) in plasma are shown as red circles (when MEDI2452 has been coadministered) or red stars (without MEDI2452), observed total and free ticagrelor active metabolite (TAM) in plasma as green squares (with MEDI2452) or green diamonds (without MEDI2452), and observed free MEDI2452 in plasma as blue triangles. Model simulations of ticagrelor, TAM, and MEDI2452 are shown in red, green, and blue lines, respectively. Dashed lines indicate the initially hypothesized model, and solid lines the refined model. LLOQ, lower limit of quantification.

Our main hypothesis (assumption III) has therefore been that both compounds are eliminated intact via the urine as a complex together with MEDI2452, and the model based on this assumption fits data reasonably well. To challenge this assumption we instead assumed that a fraction r of cleared ticagrelor and TAM is recycled. This was implemented in the model by adding the terms $r \times Cl_f \times FabTicaV(t)$ and $r \times Cl_f \times FabTamV(t)$ to the right-hand sides of Eq. (1) and (4), respectively. For the values 0.25, 0.5, and 1 of the recycling fraction r , the previously refined model parameters were then reestimated. The resulting negative log-likelihood values (174, 287, and 613) show that the model becomes increasingly inferior for higher degrees of recycling. These values should also be compared to the model completely lacking recycling, whose negative log-likelihood value is 95 and thereby remains the best model. The same conclusion also follows from inspecting the model simulations in **Figure 4**; whereas the model of total ticagrelor and total TAM are only slightly worsened by the successive increase in the recycled fraction, free levels of ticagrelor and TAM are worse, and free MEDI2452 levels are drastically impaired.

Predicted total ticagrelor plus TAM and free ticagrelor plus TAM in plasma show opposite response after administration of MEDI2452

The refined model was subsequently used to predict the dynamics of both observed and unobserved variables in studies 1 to 4. For these predictions, the effects of both parameter uncertainties and residual variability were taken into account. This was done by Monte Carlo simulations in which parameter values were sampled according to the uncertainties derived from formal parameter estimation or from uncertainties reported in literature. **Figure 5** shows simulations of total ticagrelor plus TAM, free MEDI2452, and free ticagrelor plus TAM in V in response to coadministration of ticagrelor with either MEDI2452 or vehicle. For all designs, total ticagrelor plus TAM and free ticagrelor plus TAM in plasma show opposite response after administration of MEDI2452, with total levels rising while free levels drop. We also note that the levels of free ticagrelor plus TAM after administration of MEDI2452 always remain below the corresponding levels for the vehicle groups during the 100 minutes of the experiments. Importantly, the predictions in **Figure 5** were computed for the *in vivo* levels of free ticagrelor plus TAM (i.e., the observation model was not used). As previously explained, and illustrated in **Supplementary Figure S3**, true *in vivo* levels and measured levels in *ex vivo* samples of ticagrelor and TAM may differ significantly.

The PK model can drive the PD response using a standard turnover model

Inhibition of platelet aggregation stimulated by ADP is a commonly used PD marker for P2Y₁₂ receptor antagonists, both in animals and in the clinic. The predicted levels of free ticagrelor plus TAM were therefore used to drive a simple turnover model describing the dynamics of platelet aggregation, as described in detail in **Supplementary Text S2**. As seen in **Figure 6**, the combined ticagrelor-MEDI2452 PK model extended with a platelet aggregation PD model can describe

experimental data in a quite satisfactory way. A residual plot for the PD model is provided in **Supplementary Figure S5**.

DISCUSSION

Mathematical PK models accounting for the effects of an interfering antidote are scarce in the literature, indicating that this kind of modeling remains largely unexplored. A few existing examples include modeling of so-called inverse targeting strategies, in which antidrug antibodies are administered to reduce systemic drug toxicities.^{13–15,18} There are similarities between our model and the class of target-mediated drug disposition models^{19–21} in the sense that it mechanistically accounts for the buffering interaction of ticagrelor (and TAM) with another molecular species. In our case, the interacting entity is not the drug target, but instead an antidote. Like the target-mediated drug disposition models, which in addition to the interaction itself also includes other necessary aspects of the target dynamics, such as its turnover, our model also includes the parts of the antidote dynamics that are not directly affecting its interaction with ticagrelor (i.e., its distribution and clearance). However, in contrast to our model in which the dynamic aspect of ticagrelor-antidote binding is preserved, target-mediated drug disposition models are frequently simplified according to the rapid-binding assumption.²²

As a step toward validating the model, it was compared to completely fresh data that had not been used for setting up the model. Strictly speaking, models can never be validated but the confidence in a model gradually increases as it successfully predicts new experiments and repeatedly withstands attempts of falsification.²³ Given a minor change in the parameter value for MEDI2452 clearance, the model successfully described the qualitative characteristics of the experimental data of total and free ticagrelor and TAM, and free MEDI2452, in four different studies. Many of the quantitative characteristics were also acceptably well described. Considering the simplicity of allometric scaling it is fully reasonable that a parameter value not derived from mouse data would require a slight final adjustment. By additionally reestimating some of the model parameters, allowing the validation data to inform the inference, an even better description of the data with respect to the quantitative details was possible. Moreover, the successful application of free level predictions to platelet aggregation data further contributes to validating the combined ticagrelor-MEDI2452 PK model. The overall outcome of the model validation and refinement is encouraging and suggests that the principles and assumptions underlying the model are sound and that the model has true predictive capabilities. However, there are still possible improvements to be made, for instance, by considering a full nonlinear mixed effects approach.^{24,25}

A naïve analysis of the data may suggest that administration of MEDI2452 is a counterproductive strategy for neutralizing the effects of ticagrelor because it results in a rapid increase by more than an order of magnitude in the total plasma levels of both ticagrelor and TAM. Similar observations of rapidly increasing total levels after antidote administration have been made for colchicine,²⁶

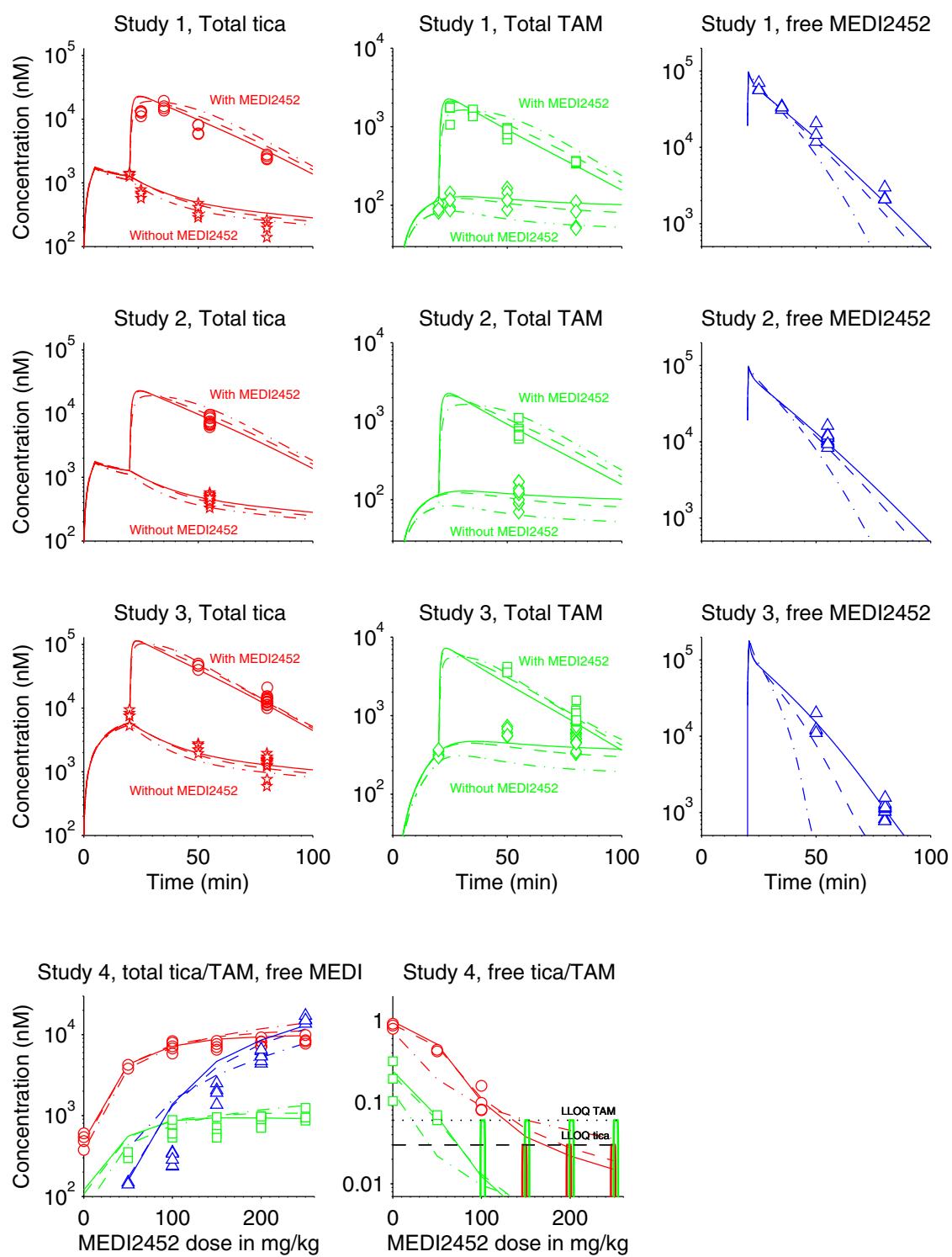


Figure 4 Comparison of alternative models with different degrees of recycling of MEDI2452-bound ticagrelor (tica) and ticagrelor active metabolite (TAM) upon MEDI2452 elimination. Model simulations of ticagrelor, TAM, and MEDI2452 are shown in red, green, and blue, respectively. Solid lines correspond to model simulations with 0% recycling, dashed lines to 25%, and dashed-dotted lines to 100%. Plotting of observed data follows the same organization as in **Figure 3**. The model's ability to describe the observations decreases as the fraction of recycling increases.

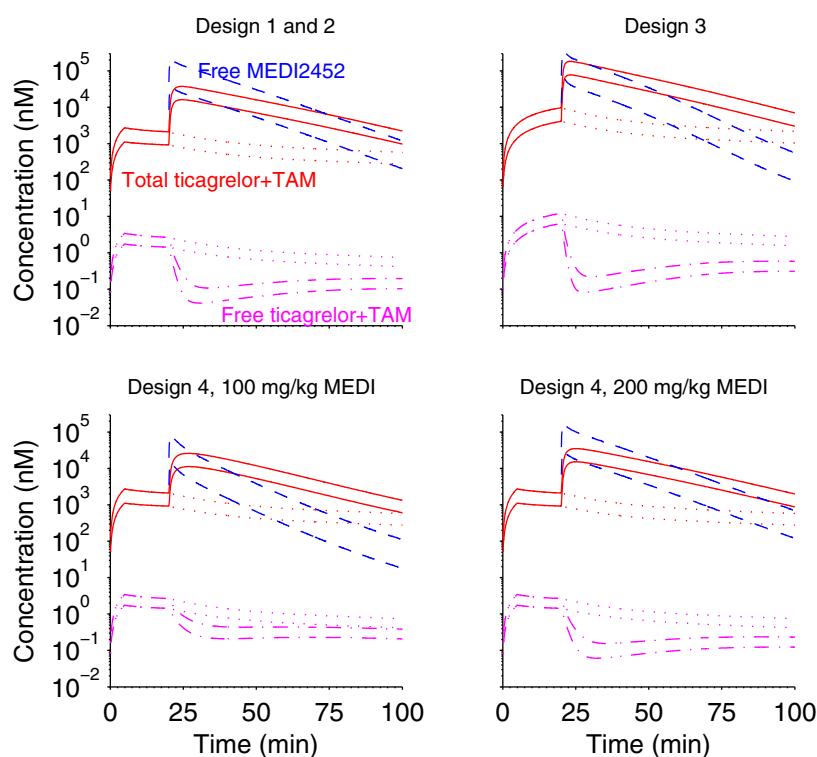


Figure 5 Prediction based on the validated model for studies 1 to 4. Curves represent the simulated 5th and 95th percentiles of the sum of free ticagrelor and free ticagrelor active metabolite (TAM) in plasma (purple dashed-dotted lines), sum of total ticagrelor and TAM in plasma (red solid lines), and free MEDI2452 (MEDI) in plasma (blue dashed line). The dotted lines indicate the scenario with ticagrelor infusion followed by vehicle instead of MEDI2452. The impact of both parameter uncertainty and residual variability is considered for the prediction (**Supplementary Text S2**).

dabigatran,²⁷ and rivaroxaban.¹² However, model simulations show that, despite the large increases in total plasma ticagrelor levels, the free levels in fact display an opposite response after administration of MEDI2452. Reasoning about the structure and parameter values of the model helps to explain the relationship between total and free plasma ticagrelor. Because the volume of distribution for ticagrelor is much larger compared to MEDI2452 (i.e., $V + V_1 + V_2 \gg V$), free ticagrelor that is bound by MEDI2452 is rapidly replenished from V_1 , and on a slower timescale also from V_2 . Thus, MEDI2452 is effectively acting like a buffer for ticagrelor in the plasma compartment, causing a substantial accumulation of total ticagrelor in the plasma. A similar explanation can be used for understanding the increase in total TAM. Moreover, if the molar dose of MEDI2452 is comparable to or larger than the total amount of ticagrelor and TAM in all three compartments at the time of dosing, MEDI2452 will bind sufficient ticagrelor and TAM to induce a significant reduction of their free levels in all compartments, including plasma. Although some of the qualitative behaviors of the combined ticagrelor-MEDI2452 PK can be understood from pure reasoning like above, the quantitative details are harder to grasp by intuition because of the nonlinear interaction of ticagrelor and MEDI2452, the delayed distribution of ticagrelor to V_2 , and the metabolism of ticagrelor to TAM, which, in turn, also competes for MEDI2452.

The ability to predict the plasma concentrations of free ticagrelor and free TAM are the key results from the present work. Knowledge about these levels is important because they drive the PD response and therefore also predict the efficacy of MEDI2452. Having access to a trustworthy model is advantageous because it can be used as a complement or an alternative to experimental observations.^{20,21} Although model simulations of the measured free ticagrelor and TAM have been validated, at least to some degree, all simulations of the actual free levels *in vivo* (**Figure 5**) are untested predictions. In fact, these predictions are not only untested but also untestable given the current experimental procedure of measuring free levels of ticagrelor and TAM. Mathematical modeling is therefore a necessity for unravelling the details of the combined ticagrelor-MED2452 PK.

The values of the model parameters—both the parameters estimated in the present work and those taken from literature—are uncertain to some degree, and, as a consequence, the model predictions are uncertain too. Providing a measure of uncertainty for the model state variables is often more interesting than the parameter uncertainties, but, unfortunately, also often overlooked. We have showed that, despite the parameter uncertainty, the behavior of the model predictions is relatively well defined. In addition to uncertainties in parameter values, there may be uncertainties with respect to the model structure. It is, for instance, not known what happens to ticagrelor and TAM when ticagrelor-bound

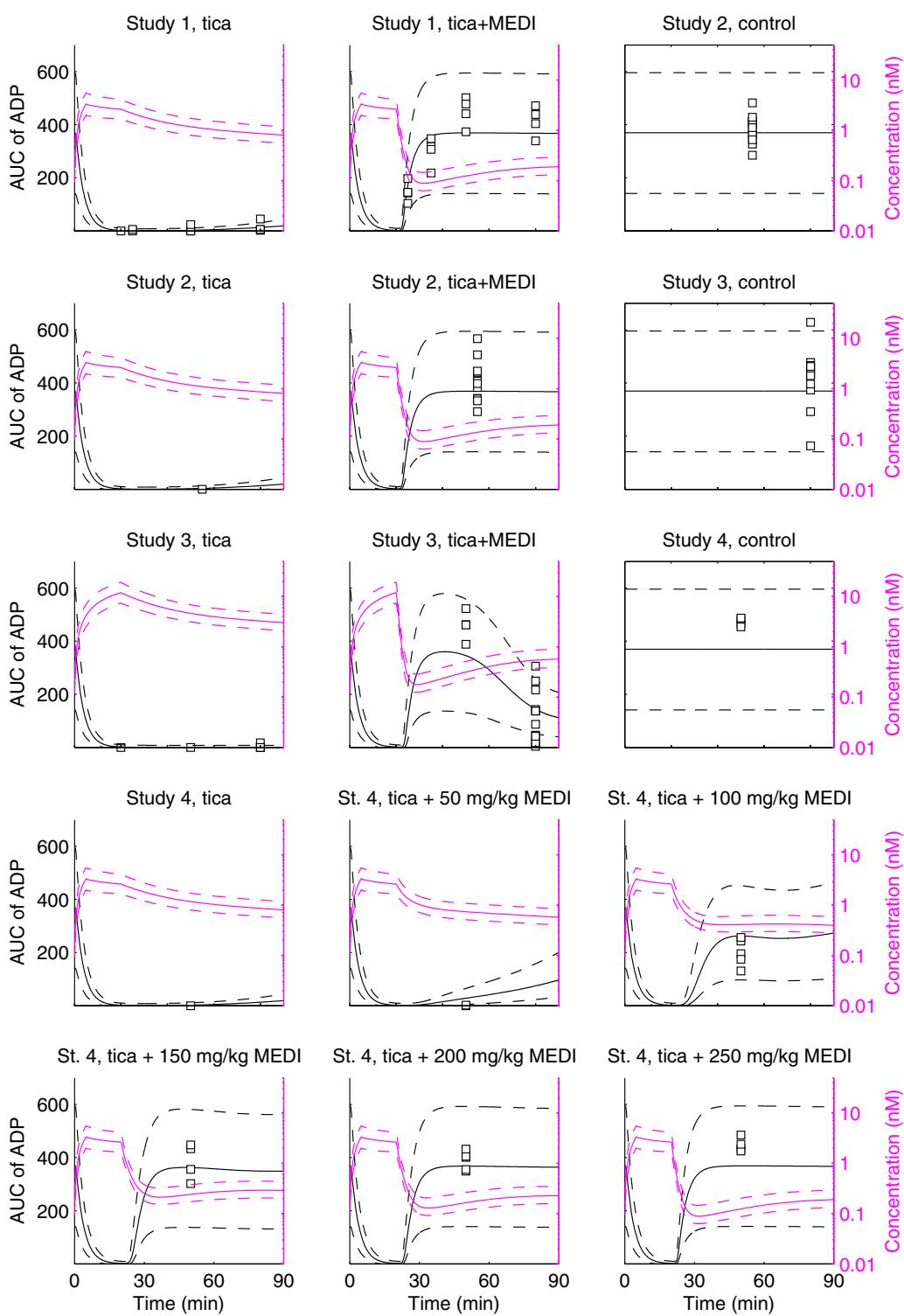


Figure 6 Pharmacokinetic-pharmacodynamic (PK-PD) model performance. The left axis shows simulated (solid black lines) and observed (squares) areas under the curve (AUCs) of ADP-induced platelet aggregation response in whole blood *ex vivo* (arbitrary units). The right axis displays (solid magenta lines) corresponding simulations of the sum of free ticagrelor (tica) and ticagrelor active metabolite (TAM) in plasma (V) which drive the pharmacodynamic effect. The impact of both parameter uncertainty and residual variability is considered in the prediction (Supplementary Text S2). Solid curves represent the median and dotted curves represent the simulated 5th and 95th percentiles. St. = Study, MEDI = MEDI2452.

or TAM-bound MEDI2452 is being eliminated, but because of the presumed renal elimination we assumed that no recycling occurs. There are examples of other modeling studies that have considered both full recycling of antibody ligands^{14,15} and partial (25%) recycling of Fab ligands.¹⁴ We showed that model behavior was highly sensitive to introducing a recycling mechanism, and that the corresponding alternative predictions successively became more and more difficult to reconcile with experimental data as the degree of recycling was increased from 0% to 100%. Thus, the modeling supports the view that, in the mouse, when MEDI2452 is eliminated, potential ticagrelor or TAM bound to MEDI2452 is not primarily recycled to the plasma (but eliminated together with MEDI2452), although we would not exclude that a smaller fraction may be. This is an important finding as it constitutes a possible scenario in the extrapolation to humans. The analysis of potential recycling rests on the notion that the clearance of MEDI2452 in the combined PK model represents elimination and not distribution. It was therefore crucial to ensure that the rapid phase of the separate MEDI2452 PK model was due to elimination, before reducing the initial two-compartment model to a one-compartment model (**Supplementary Text S2**).

The proposed model is a good starting point for scaling to model the PK of other species, including humans, and for expansion to a population model. Hence, we anticipate it to be valuable in the future clinical development of MEDI2452.

Conflict of Interest/Disclosure. The authors declared no conflict of interest.

Author Contributions: J.A., S.N., and P.G. wrote the manuscript. J.A., S.P., S.N., and P.G. designed the research. S.P., A-S.S., A.J., S.Maq., S.Mad., and J.G. performed the research. J.A., M.P., S.N., and P.G. analyzed the data.

1. Van Giezen, J.J. et al. Ticagrelor binds to human P2Y(12) independently from ADP but antagonizes ADP-induced receptor signaling and platelet aggregation. *J. Thromb. Haemost.* **7**, 1556–1565 (2009).
2. Wallentin, L. et al. Ticagrelor versus clopidogrel in patients with acute coronary syndromes. *N. Engl. J. Med.* **361**, 1045–1057 (2009).
3. Yusuf, S. et al. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N. Engl. J. Med.* **345**, 494–502 (2001).
4. Wiviott, S.D. et al. Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N. Engl. J. Med.* **357**, 2001–2015 (2007).
5. Bonaca, M.P. et al. Long-term use of ticagrelor in patients with prior myocardial infarction. *N. Engl. J. Med.* **372**, 1791–1800 (2015).
6. Buchanan, A. et al. Structural and functional characterization of a specific antidote for ticagrelor. *Blood* **125**, 3484–3490 (2015).
7. Dolgin, E. Antidotes progress for new oral clotbusters. *Nat. Biotechnol.* **33**, 117–118 (2015).
8. Schiele, F. et al. A specific antidote for dabigatran: functional and structural characterization. *Blood* **121**, 3554–3562 (2013).

9. Glund, S. et al. A randomised study in healthy volunteers to investigate the safety, tolerability and pharmacokinetics of idarucizumab, a specific antidote to dabigatran. *Thromb. Haemost.* **113**, 943–951 (2015).
10. Glund, S. et al. Safety, tolerability, and efficacy of idarucizumab for the reversal of the anticoagulant effect of dabigatran in healthy male volunteers: a randomised, placebo-controlled, double-blind phase 1 trial. *Lancet* **386**, 680–690 (2015).
11. Pollack, C.V. Jr. et al. Idarucizumab for dabigatran reversal. *N. Engl. J. Med.* **373**, 511–520 (2015).
12. Lu, G. et al. A specific antidote for reversal of anticoagulation by direct and indirect inhibitors of coagulation factor Xa. *Nat. Med.* **19**, 446–451 (2013).
13. Balthasar, J. & Fung, H.L. Utilization of antidiug antibody fragments for the optimization of intraperitoneal drug therapy: studies using digoxin as a model drug. *J. Pharmacol. Exp. Ther.* **268**, 734–739 (1994).
14. Lobo, E.D., Soda, D.M. & Balthasar, J.P. Application of pharmacokinetic-pharmacodynamic modeling to predict the kinetic and dynamic effects of anti-methotrexate antibodies in mice. *J. Pharm. Sci.* **92**, 1665–1676 (2003).
15. Shah, D.K. & Balthasar, J.P. PKTD modeling for prediction of the effects of 8C2, an anti-topotecan mAb, on topotecan-induced toxicity in mice. *Int. J. Pharm.* **465**, 228–238 (2014).
16. Springthorpe, B. et al. From ATP to AZD6140: the discovery of an orally active reversible P2Y12 receptor antagonist for the prevention of thrombosis. *Bioorg. Med. Chem. Lett.* **17**, 6013–6018 (2007).
17. King, D.J. *Applications and Engineering of Monoclonal Antibodies* (Taylor & Francis Ltd, London, 1988).
18. Balthasar, J.P. & Fung, H.L. Inverse targeting of peritoneal tumors: selective alteration of the disposition of methotrexate through the use of anti-methotrexate antibodies and antibody fragments. *J. Pharm. Sci.* **85**, 1035–1043 (1996).
19. Mager, D.E. & Jusko, W.J. General pharmacokinetic model for drugs exhibiting target-mediated drug disposition. *J. Pharmacokinet. Pharmacodyn.* **28**, 507–532 (2001).
20. Xiao, J.J. et al. Pharmacokinetics of anti-hepcidin monoclonal antibody Ab 12B9m and hepcidin in cynomolgus monkeys. *AAPS J.* **12**, 646–657 (2010).
21. Tang, C. & Prueksaritanont, T. Theoretical analysis of interplay of therapeutic protein drug and circulating soluble target: temporal profiles of 'free' and 'total' drug and target. *Pharm. Res.* **28**, 2447–2457 (2011).
22. Marathe, A., Krzyzanski, W. & Mager, D.E. Numerical validation and properties of a rapid binding approximation of a target-mediated drug disposition pharmacokinetic model. *J. Pharmacokinet. Pharmacodyn.* **36**, 199–219 (2009).
23. Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J. & Jirstrand, M. Kinetic models in industrial biotechnology – improving cell factory performance. *Metab. Eng.* **24**, 38–60 (2014).
24. Bonate, P. *Pharmacokinetic-Pharmacodynamic Modeling and Simulation* (Springer, New York, 2006).
25. Almquist, J., Leander, J. & Jirstrand, M. Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood. *J. Pharmacokinet. Pharmacodyn.* **42**, 191–209 (2015).
26. Sabouraud, A.E., Urtizberea, M., Cano, N.J., Grandjeorge, M., Rouzioux, J.M. & Schermann, J.M. Colchicine-specific Fab fragments alter colchicine disposition in rabbits. *J. Pharmacol. Exp. Ther.* **260**, 1214–1219 (1992).
27. van Ryn, J., Litzenburger, T., Gan, G., Coble, K. & Schurer, J. In vitro characterization, pharmacokinetics and reversal of supratherapeutic doses of dabigatran-induced bleeding in rats by a specific antibody fragment antidote to dabigatran. 54th ASH Annual Meeting and Exposition, Atlanta, GA, 8–11 December 2012. Abstract 3418.

© 2016 The Authors CPT: Pharmacometrics & Systems Pharmacology published by Wiley Periodicals, Inc. on behalf of American Society for Clinical Pharmacology and Therapeutics. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (<http://www.wileyonlinelibrary.com/psp4>)

PAPER C

Modeling the Effect of Kv1.5 Block on the Canine Action Potential

Modeling the Effect of Kv1.5 Block on the Canine Action Potential

Joachim Almquist,^{†*} Mikael Wallman,^{†‡} Ingemar Jacobson,[§] and Mats Jirstrand[†]

[†]Fraunhofer-Chalmers Centre, Chalmers Science Park, Göteborg, Sweden; [‡]Oxford University Computing Laboratory, Oxford, United Kingdom; and [§]AstraZeneca R & D, Mölndal, Sweden

ABSTRACT A wide range of ion channels have been considered as potential targets for pharmacological treatment of atrial fibrillation. The Kv1.5 channel, carrying the I_{Kur} current, has received special attention because it contributes to repolarization in the atria but is absent or weakly expressed in ventricular tissue. The dog serves as an important animal model for electrophysiological studies of the heart and mathematical models of the canine atrial action potential (CAAP) have been developed to study the interplay between ionic currents. To enable more-realistic studies on the effects of Kv1.5 blockers on the CAAP *in silico*, two continuous-time Markov models of the guarded receptor type were formulated for Kv1.5 and subsequently inserted into the Ramirez-Nattel-Courtemanche model of the CAAP. The main findings were: 1), time- and state-dependent Markov models of open-channel Kv1.5 block gave significantly different results compared to a time- and state-independent model with a down-scaled conductance; 2), the outcome of Kv1.5 block on the macroscopic system variable APD_{90} was dependent on the precise mechanism of block; and 3), open-channel block produced a reverse use-dependent prolongation of APD_{90} . This study suggests that more-complex ion-channel models are a prerequisite for quantitative modeling of drug effects.

INTRODUCTION

The ultrarapidly activating delayed rectifier K^+ current, commonly denoted I_{Kur} , conducted by the Kv1.5 channel, has emerged as a target for pharmacological treatment of atrial fibrillation (AF; see Tamargo et al. (1) and references within). I_{Kur} contributes to both early and late repolarization of the action potential (AP) of the human atrial myocyte and because a similar current largely seems to be lacking in ventricular tissue, a blocker of the Kv1.5 channel would have potential to selectively increase the duration of the atrial AP (APD) and, hence, the refractory period (AERP). I_{Kur} blockers have been shown to affect human atrial AP repolarization *in vitro*, and selectively increase atrial refractoriness and terminate atrial arrhythmias in several animal species *in vivo* (2–8). Human *in vitro* data as well as computer modeling of human atrial cells using Hodgkin-Huxley representations of I_{Kur} have also shown that the effect of a block of I_{Kur} on AP repolarization depend on the relative densities of all involved ion channels and may be more effective in diseased tissue (3,9). In both animal disease models and humans, ion channel densities changes during progression of AF (see Tamargo et al. (1) and references within), thus changing the relative contribution of each involved ion channel in the repolarization process. This remodeling process also facilitates the progression of the disease. In a complex biological system with interwoven dependencies like this, good knowledge of the behavior of individual components in isolation does not automatically provide good knowledge of the behavior of the system as a whole. Mathematical modeling may be especially helpful in elucidating important system properties by combining

mathematical formulations of the knowledge of single components. For example, complete models of the human atrial myocyte AP have been developed (10,11) and extensions of these models have been used to evaluate the effects of I_{Kur} block (3,9).

Animal models are used extensively throughout the drug development process from early discovery to late development, to improve understanding of biological mechanisms, and to allow for predictions of human responses to drug exposure. For this reason, mathematical models of animal models are of interest. Different animal models are used in different disease areas and for cardiovascular diseases, experimental dog models have been widely used to study, for example, atrial arrhythmia mechanisms *in vivo*. The Ramirez-Nattel-Courtemanche (RNC) model (12) is a mathematical model of the canine atrial AP. It is based on the Hodgkin-Huxley formalism, contains all major ionic currents, and has been validated using experimental measurements in canine atrial myocytes. In this model, the I_{Kur} current carried by Kv1.5 is known as I_{Kurd} .

However, most Kv1.5 blockers are selectively blocking only the activated or open state of Kv1.5 channels (2,13–18) and important features of the drug-ion channel interaction such as use- and voltage-dependence are not easily implemented in the Hodgkin-Huxley formalism. Rather than using the Hodgkin-Huxley formulation, state-dependent blocking mechanisms are usually modeled with Markov models. Therefore, a continuous-time Markov model of the guarded receptor type was set up for the canine Kv1.5 to enable studies on the effects of Kv1.5 open-channel blockers on the canine atrial AP *in silico*. Conceptually similar Markov models have previously been successfully used to describe the kinetics of block of human Kv1.5 by quinidine (13), quinine, clofibrate, and tetrapentylammonium (14),

Submitted March 29, 2010, and accepted for publication August 27, 2010.

*Correspondence: joachim.almquist@fcc.chalmers.se

Editor: Michael D. Stern.

© 2010 by the Biophysical Society
0006-3495/10/11/2726/11 \$2.00

doi: 10.1016/j.bpj.2010.08.062

loratadine (15), and bupi-, ropi-, and mepi-vacaine (16). The Markov model was subsequently inserted into the RNC model, replacing the original Hodgkin-Huxley expression. A subset of the parameters in our model reflects properties of a drug and drug-channel interactions such as the net charge of a drug, and the rates of binding to, and dissociation from, the receptor site on the channel. By performing simulations with the modified version of the RNC model, we have been able to examine the influence of these drug-defining parameters on the AP. The model has also been used to make predictions about the morphological changes of the AP for two particular I_{Kur} blockers, whose kinetic parameters have been determined experimentally elsewhere (2). In connection to this, our model of open-channel Kv1.5 block was extended to also account for drugs that, like the above-exemplified compounds, exhibit a voltage-dependent dissociation rate distinguished from the type of voltage dependence caused by drugs carrying a net charge. To our knowledge, this phenomenon has not been captured in any previous Kv1.5 Markov model.

METHODS

The RNC model

Canine atrial myocytes displays regional variations in ionic current density (12). In this study, values for cells from the pectinate muscle (PM) were used, and as described in Ramirez et al. (12), the maximum conductance of L-type Ca^{2+} channels, transient outward K^+ current, and Na^+ channels were set to 40%, 40%, and 50%, respectively, of their nominal values to account for the major changes in ion channel density that occurs during progression of AF in the dog (19). An often-encountered issue with models of electrophysiological systems is small drifts in ion concentrations. In our implementation, the ion concentrations of Na^+ , K^+ , and Cl^- were, for simplicity, fixed to their initial values as listed in Ramirez et al. (12). This did not alter the behavior of the model in any noticeable way except for the loss of dynamics for these ion concentrations.

Our Kv1.5 models

Transition rates between states in the models are assumed to have an exponential voltage dependence and are characterized by two parameters—the zero voltage rate and the equivalent charge movement up to the transition state. The parameters in our five-state Kv1.5 model were optimized to reproduce the behavior of the original RNC model. This optimization was performed using a multidimensional downhill simplex method (20) implemented in the Systems Biology Toolbox for MATLAB (The MathWorks, Natick, MA) (21,22). The model reduction used in connection with our 10-state model is based on an analytical equilibrium solution (23).

Simulations

Simulations were performed with the Systems Biology Toolbox for MATLAB (21,22), which uses the CVODE integrator (24). For all stimulation frequencies, a stimulation current of -2900 pA was applied during 2 ms. Simulation data used for analysis was always collected from the last of a series of 40 APs. This gave a maximum difference in membrane potential between the last two APs equal to 0.01 and 0.05 mV for AF remodeled cells stimulated at frequencies of 1 and 4 Hz, respectively. Visual inspections of APs and of the difference between consecutive APs was also performed to ensure that the AP waveforms really were converging toward

stable limit cycles. As a measure of APD, we used APD_{90} , which was defined as the time it takes, after a stimulation, to reach a voltage level of -72 mV . This corresponds to a 90% repolarization of a normal cell stimulated at a frequency of 1 Hz.

RESULTS

Modeling the Kv1.5 ion channel

Our strategy for implementing the continuous-time Markov model was in many respects similar to several modeling efforts of potassium channels found in the literature (13,15,16,23,25).

Kv1.5 lacks the so-called fast N-type inactivation but is affected by the slower C-type inactivation. From Fig. 3 in Ramirez et al. (12), it is apparent that the inactivation property becomes increasingly important for positive values of the membrane potential. However, for the range of voltages where the model will operate, inactivation is not very significant. At a membrane potential of 20 mV , the steady-state value of the inactivation variable, u_∞ , is ~ 0.95 , which means a reduction in I_{Kurd} by only 5% at steady state. For membrane potentials at or below zero, steady-state inactivations are clearly negligible. Also considering that inactivation is a rather slow process, the effective reduction in I_{Kurd} is even less for the typically short periods of more strongly depolarized membrane potentials during an action potential. As a consequence, the inactivation property was, for simplicity, omitted from our model. This approximation was validated by simulating the AP both with the inactivation property in place and, as a comparison, with the inactivation turned off. For the normal cell, the difference in membrane potential did not exceed 1.5 mV during the course of the AP and the maximum difference in I_{Kurd} was 0.1 pA/pF at a stimulation frequency of 4 Hz. In the AF setting, which was most frequently used in this study, differences were even less. Less difference were also observed when all simulations were repeated using a stimulation frequency of 1 Hz.

Under the assumption that transition rates between the open and closed conformations of a subunit are independent of the state of the other subunits and that all four subunits operate in an identical manner, the Kv1.5 channel can be represented by the five-state model in Fig. 1 A. It has four closed, nonconducting states, C_1-C_4 , and an open, conducting state, O . Here, α and β are the forward and reverse rate constants of the transition to the open conformation, respectively. A mathematical description of the scheme in Fig. 1 A is given by a system of ordinary differential equations

$$\frac{dC_1}{dt} = -4\alpha C_1 + \beta C_2, \quad (1a)$$

$$\frac{dC_2}{dt} = 4\alpha C_1 - \beta C_2 - 3\alpha C_2 + 2\beta C_3, \quad (1b)$$

$$\frac{dC_3}{dt} = 3\alpha C_2 - 2\beta C_3 - 2\alpha C_3 + 3\beta C_4, \quad (1c)$$

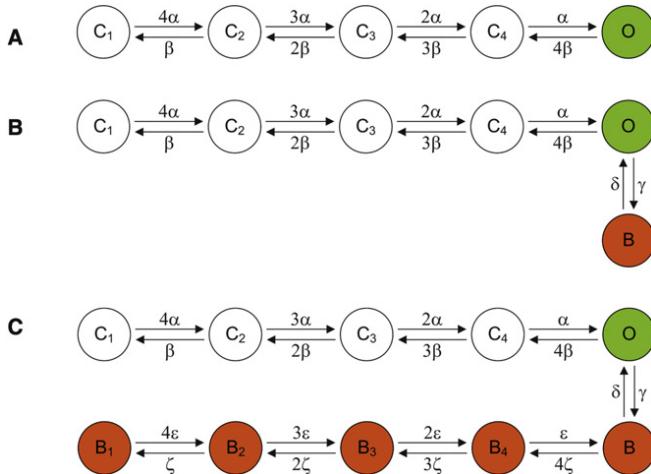


FIGURE 1 Five-state Kv1.5 model (A). Transitions between the states are determined by the rates α and β . Six-state Kv1.5 model with an open-channel block (B). Transitions between the open and the blocked state are determined by the rates γ and δ and by the drug concentration. Ten-state Kv1.5 model with several blocked states (C). Transitions between the blocked states are determined by the rates ζ and ε .

$$\frac{dC_4}{dt} = 2\alpha C_3 - 3\beta C_4 - \alpha C_4 + 4\beta O, \quad (1d)$$

$$\frac{dO}{dt} = \alpha C_4 - 4\beta O, \quad (1e)$$

where

$$\alpha = \alpha_0 \exp\left(\frac{Z_\alpha FV}{RT}\right), \quad (2a)$$

$$\beta = \beta_0 \exp\left(\frac{-Z_\beta FV}{RT}\right), \quad (2b)$$

and C_1 – C_4 and O represents the fraction of channels in the different states. The parameters α_0 and β_0 are the forward and reverse rate constants of the transition to the open conformation at zero membrane potential while Z_α and Z_β are the equivalent charge movements up to the transition state, defining the degree of voltage dependence of α and β . Together, the expressions in Eqs. 1 and 2 determine the time evolution of the fraction of open Kv1.5 ion channels. The conductance of the Kv1.5 channel was described by the voltage-dependent relation for the total conductance of Kv1.5 channel

$$g_{\text{Kurd}} = 0.00855 + \frac{0.0779}{1 + \exp(\frac{V+11}{-16})} \quad (3)$$

used in Ramirez et al. (12). This expression correctly relates experimental data of I_{Kurd} to the fraction of open channels (26). Multiplying the fraction of open channels with the conductance g_{Kurd} and with the deviation of the membrane potential from the Nernst potential of K^+ ions, gave the expression for I_{Kurd} ,

$$I_{\text{Kurd}} = g_{\text{Kurd}} O(V; \alpha_0, \beta_0, Z_\alpha, Z_\beta)(V - E_K), \quad (4)$$

to be compared with the Hodgkin-Huxley expression

$$I_{\text{Kurd}} = g_{\text{Kurd}} u_a^3 u_i (V - E_K) \quad (5)$$

used in Ramirez et al. (12).

Determining parameters

To maintain the properties of the RNC model in absence of a blocking agent, values of the parameters in the expressions in Eq. 2 were chosen accordingly. In practice, this was done by generating artificial data of the membrane potential of a remodeled cell during an 2 Hz AP using the RNC model with the original equations for I_{Kurd} . The I_{Kurd} part of the model was then replaced by the expressions in Eqs. 1 and 2, and the parameters were optimized to fit the artificially generated data. Optimizing the parameters during a complete AP, as opposed to using an *in silico* voltage-clamp protocol for I_{Kurd} alone, is appealing because the data set is naturally weighted over the desired operating range of the model. In this way, the parameter values were found to be

$$\alpha_0 = 0.6161 \text{ s}^{-1}, \beta_0 = 0.1001 \text{ s}^{-1}, Z_\alpha = 0.9470, \text{ and } Z_\beta = 0.8129.$$

The quality of the obtained parameter set is demonstrated in Fig. S1 in the Supporting Material. There, the membrane potential and I_{Kurd} of the Markov model is compared to the original RNC model, using both the remodeled and normal cell.

Including blocked state

A simple open-channel block of Kv1.5 was introduced by adding a new state connected to the open state in Fig. 1 A. The new state represents a nonconducting conformation where the drug only interacts with the open state of the ion-channel protein. This is the so-called foot-in-the-door mechanism, which has been described for several Kv1.5 blockers including AVE0118 (27) and the diphenyl phosphine oxides DPO-1 and DPO-2 (2,28). Returning to the open or closed states requires that the bound drug dissociates from its binding site. The extended model is shown in Fig. 1 B. In the mathematical description, Eq. 1 e was modified and a differential equation for the blocked state was added,

$$\frac{dO}{dt} = \alpha C_4 - 4\beta O - \gamma [Drug] O + \delta B, \quad (6a)$$

$$\frac{dB}{dt} = \gamma [Drug] O - \delta B, \quad (6b)$$

where $[Drug]$ denotes the concentration of the drug. The rates γ and δ are of the same form as α and β in the expressions in Eq. 2,

$$\gamma = \gamma_0 \exp\left(\frac{Z_\gamma FV}{RT}\right), \quad (7a)$$

$$\delta = \delta_0 \exp\left(\frac{-Z_\delta FV}{RT}\right). \quad (7b)$$

The potential voltage dependence of these rates (for nonzero Z_γ and/or Z_δ) then reflects the fact that the drug binding site may be located somewhere within the membrane electrical field. If the drug has a net charge, it will sense part of the electrical field and the binding and dissociation rates will be altered. The extended model of Kv1.5 current is then given by

$$I_{Kurd} = g_{Kurd} O(V; \alpha_0, \beta_0, \gamma_0, \delta_0, Z_\alpha, Z_\beta, Z_\gamma, Z_\delta, [Drug]) \times (V - E_K), \quad (8)$$

reflecting the introduced dependence of the drug.

Sensitivity analysis

The model was first used to illustrate how the AP of a remodeled PM cell changes in response to different concentrations of a typical, uncharged drug with $\gamma = 10 \mu\text{M}^{-1} \text{s}^{-1}$ and $\delta = 2 \text{s}^{-1}$, at a stimulation frequency of 1 Hz. The simulated membrane potential and I_{Kurd} current are shown in Fig. 2. In the absence of I_{Kurd} block, the AP of the remodeled cell has the characteristic triangular shape lacking a clearly defined plateau. At low drug concentrations, a plateau emerges and

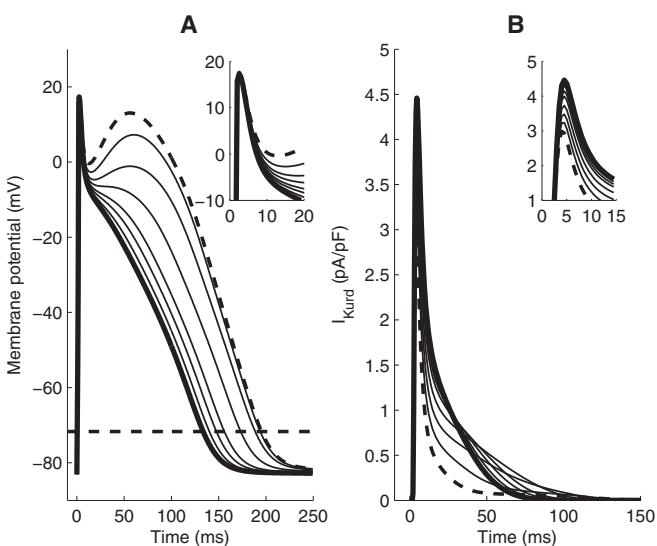


FIGURE 2 Action potential waveforms (*A*) and corresponding I_{Kurd} (*B*) generated by the remodeled PM-cell model stimulated at a frequency of 1 Hz. Bold traces corresponds to no drug, solid traces to drug concentrations of 1, 2, 3, 5, 8, and 13 μM , respectively, and dashed traces to 21 μM . The drug-blocking action was defined by the parameter values $\gamma = 10 \mu\text{M}^{-1} \text{s}^{-1}$, $\delta = 2 \text{s}^{-1}$, and $Z_\gamma = Z_\delta = 0$. The APD₉₀ is marked by a dashed line in the plot of the membrane potential. (*Inset*) Magnification of the traces during the spike.

the triangular shape is gradually lost. As the drug concentration increases further, the plateau phase becomes wider and more elevated before the AP finally culminates in a spike and dome morphology for the highest concentrations. In this way, the addition of the drug delays the repolarization and therefore increases the duration time of the AP. These changes in the AP morphology are accompanied by an attenuation of I_{Kurd} during the peak and the first part of the plateau phase. The AP peak, shown in the inset of Fig. 2 *A*, is not affected by the drug. However, it can be noted that the peak is lower in our simulation of the AF-remodeled AP, reaching only 17 mV, compared to the normal cell simulated in Ramirez et al. (12). The above effects on the AP qualitatively resembles the effect of changing the maximum I_{Kurd} conductance on the normal canine AP (12). A prolonged duration of the AP, and an emerging plateau, have also been reported from simulations of the AF remodeled human atrial myocyte but no dome-shaped plateau phase was observed even though conductance of I_{Kurd} was reduced to 20% (3) and 10% (9) of their nominal values, respectively.

A drug-induced increase of the APD and, as a consequence, in the AERP, has been considered as an effective antiarrhythmic mechanism. To quantify the effects of Kv1.5 blocking drugs on the AP we therefore used the APD₉₀-measure, as defined in the Methods. For all simulation results presented in this section, the similarly defined measure APD₇₀ was also calculated. As it gave qualitatively very similar results, these results are not shown. However, this suggests that the choice of APD measure does not have a critical influence on the conclusions drawn as long as it is chosen somewhere in mid- or end-repolarization phase. This can also be seen in Fig. 2 where the traces run in parallel during the later part of repolarization. Because the proposed antiarrhythmic mechanism is an increase in the refractory period, the different concentrations of the drug in Fig. 2 was used to investigate how the refractory period varies with APD₉₀. Both for the remodeled and normal cell, this relationship was very well described by a linear function (not shown). This suggests that APD₉₀ can be used as a relevant measure of the refractory period.

Having tested the model for one specific parameter setup, it was used, subsequently, to examine systematically how the APD₉₀ of a remodeled PM cell changes with respect to the drug concentration, the free parameters of the expressions in Eq. 7, and AP stimulation frequency. The results were also compared to a simple model featuring a time- and voltage-independent block obtained by just reducing the conductance by a constant factor.

First, the effect of an uncharged drug was investigated for two stimulation frequencies. Combining 21 values of the effective on-rate, the concentration of the drug times the binding rate, between 0 and 200 s^{-1} ,

$$[Drug]\gamma = 10n, n = 0, 1, \dots, 20$$

with 17 values of the dissociation rate between 0.25 and 64 s^{-1} ,

$$\delta = 2^n, n = -2, -1.5, \dots, 6$$

produced 375 different test situations. Results are shown as combined surface and contour plots in Fig. 3. At a stimulation frequency of 1 Hz, APD₉₀ increased for both decreasing values of the off-rate and for increasing values of the effective on-rate (see Fig. 3 A). At 4 Hz, the trend is comparable to 1 Hz, but saturation at high effective on-rates is not as evident and values of APD₉₀ are substantially lower overall (see Fig. 3 B). A very low frequency, 0.1 Hz, was also tested, and was found to be very similar to 1 Hz (results not shown). The difference in APD₉₀ between 1 Hz and 4 Hz, in absolute numbers, is shown in Fig. 3 C. It shows that APD₉₀ is 17 ms longer at 1 Hz compared to 4 Hz in absence of drug (points along the off-rate axis where the effective on-rate is zero). For almost all other points in the plane, the difference is larger, with a maximum difference of 41 ms.

Evidently, the presence of Kv1.5 targeting drugs in a remodeled PM cell tend to extend the duration of the AP more at lower frequencies than at higher. This reverse use-dependence also holds in terms of the relative increase in APD₉₀ (not shown). The frequency dependence was investigated further by looking at APD₉₀ and seeing how the fraction of open and blocked Kv1.5 developed over time in both the remodeled and the normal cell. Four different frequencies, 0.5, 1, 2, and 4 Hz, was used with and without $8 \mu\text{M}$ of the drug used in Fig. 2. The highest frequency, 4 Hz, could not be tested for the normal cell in presence of the drug because no stable limit cycle with a period of 250 ms was present. In Fig. 4 A the frequency dependence of

APD₉₀ is shown for the remodeled cell (*left*), and for the normal setting (*right*). Interestingly, in the normal setting, the reverse use-dependence seen for the remodeled cell is not present. Instead, the increase in APD₉₀ becomes larger at higher frequencies. Fig. 4 B shows the fraction of Kv1.5 in the open state during the first 100 ms of the AP in absence of the drug. The morphological changes in the AP due to rate adaption have the effect that lower fractions of Kv1.5 are in the conducting, drug-susceptible state at higher frequencies compared to lower. This observation holds both for the remodeled and normal cell. The same plot, but in the presence of the drug, is shown in Fig. 4 C. Because of the block, smaller fractions of conducting Kv1.5 are now seen. Finally, the fraction of blocked channels at different frequencies is shown in Fig. 4 D for the first 100 ms of the remodeled (*left*) and normal (*right*) AP. At the lowest frequency, there were virtually no channels blocked at the point of stimulation. At higher frequencies, significant fractions of channels were still blocked at the onset of stimulation.

The parameters Z_γ and Z_δ in Eq. 7 are the equivalent charge movements up to the transition state of a drug in the electrical field. By setting Z_γ and Z_δ to values other than zero, we examined the effects of various types of charged drugs on the APD. Assuming a drug with a net charge of +1, its binding site located at a point 20% into the electrical field, and a symmetrical barrier, we set $Z_\gamma = Z_\delta = 0$. These values are similar to experimentally determined values for two blockers of the human Kv1.5 (13,16). However, introducing a voltage-dependence of the interaction between the drug and its binding site had limited effect on APD₉₀. At 4 Hz, there was virtually no change in APD₉₀, except for very small

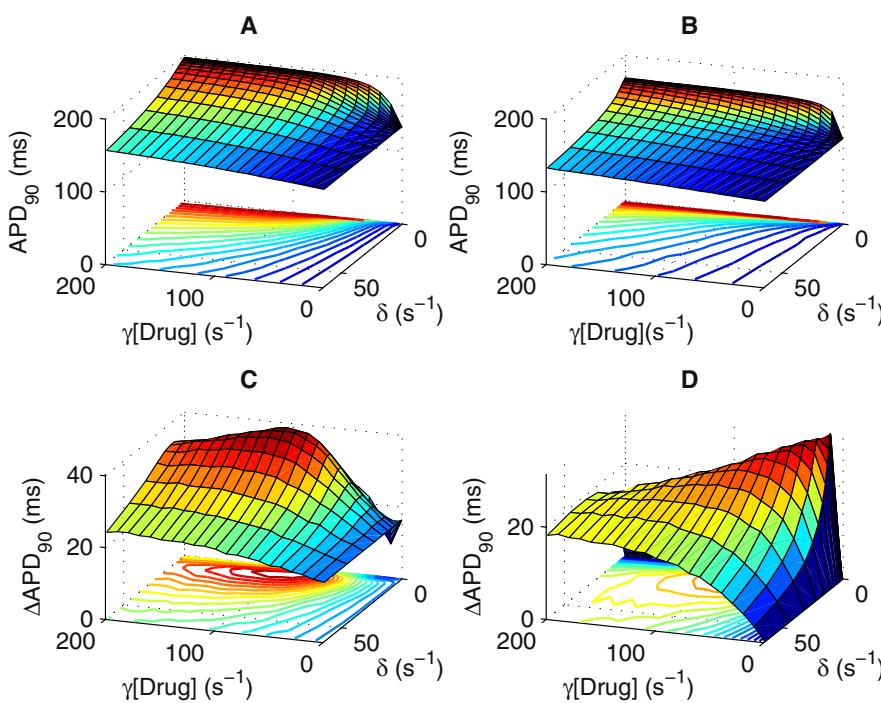


FIGURE 3 APD₉₀ as function of effective on-rate and off-rate of an uncharged drug at a stimulation frequency of 1 Hz (A). APD₉₀ as function of effective on-rate and off-rate of an uncharged drug at a stimulation frequency of 4 Hz (B). Absolute difference between APD₉₀ at stimulation frequencies of 1 and 4 Hz as function of effective on-rate and off-rate of an uncharged drug (C). Absolute difference between APD₉₀ calculated from the time- and voltage-independent model and from the six-state model, as function of effective on-rate and off-rate of an uncharged drug at a stimulation frequency of 4 Hz (D).

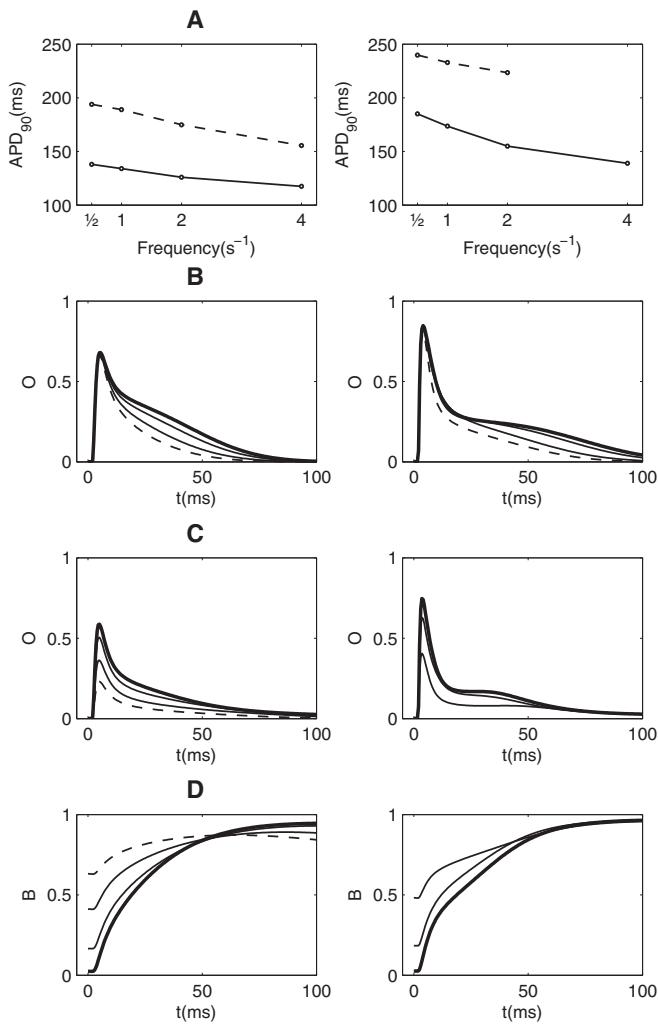


FIGURE 4 AP frequency dependence of 8 μM test drug defined by $\gamma = 10 \mu\text{M}^{-1} \text{s}^{-1}$, $\delta = 2 \text{s}^{-1}$, and $Z_\gamma = Z_\delta = 0$. Left column uses the AF setting, right column uses the normal cell setting. Row A shows APD_{90} as function of frequency in absence of the test drug (dashed) and with the test drug (solid). Row B shows the fraction open Kv1.5, O , during the first 100 ms of the AP while rows C and D show the fraction of open, O , and blocked, B , Kv1.5, respectively, in the presence of the test drug during the first 100 ms of the AP. Bold traces correspond to a stimulation frequency of 0.5 Hz, solid traces to 1 and 2 Hz, respectively, and dashed traces to 4 Hz.

off-rates, where the duration was decreased by $\sim 3\text{--}4$ ms. The effect was even more marginal at 1 Hz, producing only a small decrease in duration of 4 ms for combinations of low effective on-rate and low off-rate. These results suggest that Z_γ and Z_δ may only have a significant impact if the receptor site is located deep within the electrical field and/or the drug has several charged groups at physiological pH.

In several studies (3,11,29,30), the effect of an ion channel blocking drug has been modeled as a time- and voltage-independent decrease in channel conductance. If the degree of decrease in conductance in such a model is assumed to be the result of a simple state-independent bimolecular reaction in equilibrium (between the drug and its

binding site), a comparison of this approach to our dynamic modeling is possible. The constant degree of block in the simpler model, B_{const} , would then be a function of the effective on-rate k_{on} (the concentration of the drug times the binding rate) and the off-rate k_{off} (dissociation rate),

$$B_{\text{const}} = \frac{k_{\text{on}}}{k_{\text{off}} + k_{\text{on}}} \quad (9)$$

To see how the two models differ, the difference in their APD_{90} at a stimulation frequency of 4 Hz was calculated from simulations. The result is shown in Fig. 3 D. For large effective on-rates and small off-rates, as well as for the opposite situation with small effective on-rates and large off-rates, the two models gave similar results. In the other parts of the on-off plane the difference was bigger, peaking at 32 ms. We also looked at the difference between the models at a stimulation frequency of 1 Hz. There was a similarly located peak of the same height, but the differences in other parts were roughly halved (not shown).

Extending the model

It is important to emphasize that the drug dissociation rate in our open-channel block model only accounts for voltage dependence arising from the charge of the drug. For the human Kv1.5, voltage-dependent dissociation rates not related to the charge of the drug has been reported (2,15). The dissociation rates in Lagrutta et al. (2) were determined at two different membrane potentials and denoted k_{on} for the open state and k_{off} for the closed state. For one particular compound, the DPO-1, these rates differed by a factor of 60. Apparently, at least for the hKv1.5 channel, the state of the blocked channel can change with the membrane potential producing a strong voltage dependence of the dissociation rate in addition to any voltage dependence due to the charge of the drug.

To describe this, additional blocked states could be added in sequence to the blocked state of Fig. 1 B. This would introduce a state-dependent recovery from block, similar to the already existing state-dependent formation of block. The model outlined in Fig. 1 C is one out of many possible extensions to our model that could account for the state-dependent type of voltage dependence discussed above. A similar model structure has been used for the hERG channel (31). Four new blocked states were added, creating symmetry to the four closed states in the upper part. The reason for introducing precisely four new blocked states was consequently to reflect an assumption that the four subunits may fully or partially close independently and identically also when a drug is bound. Because these conformational changes would be expected to be different than in the absence of a drug, two new rates of the same type as earlier, ζ and ϵ , was introduced. Compared to the six-state model, the 10-state model can be seen as a more elaborate implementation of the foot-in-the-door mechanism.

For the rates ζ and ϵ to be determined, four parameter values are required. To reduce complexity of the extended model, it was assumed that the transition rates between the closed states were faster than drug binding and dissociation. Under this assumption, the blocked branch of the system can be reduced to a single lumped state. Let the blocked lumped state be defined as

$$L_B = B_1 + B_2 + B_3 + B_4 + B. \quad (10)$$

The occupancy of the original states are now determined from the equilibrium relations in each of the lumped states. In particular, the fraction of L_B in the blocked state B , f_B , is

$$f_B(V) = \frac{1}{1 + 4Q + 6Q^2 + 4Q^3 + Q^4}, \quad (11)$$

where Q is the ratio of ζ and ϵ . In the same way as O determines the fraction of unblocked channels accessible for drug binding, f_B determines the fraction of blocked channels that are subject to drug dissociation. Using this variable, the changes in O and in the lumped state L_B with respect to time can then be described by

$$\frac{dO}{dt} = \alpha C_4 - 4\beta O + \delta f_B L_B - \gamma [Drug], \quad (12a)$$

$$\frac{dL_B}{dt} = \gamma [Drug] - \delta f_B L_B, \quad (12b)$$

to be compared with Eq. 6.

Results from extended model

To explore further the impact of state-dependent recovery from block, literature values for voltage-dependent apparent association and dissociation rates (2) were implemented in the 10-state model. In particular, data from two compounds were chosen, DPO-1 and DPO-2, because of their similarity in all parameters but the apparent dissociation rate, where they displayed a large difference. These inhibitors have been characterized by measuring the apparent association rate at +40 mV, here denoted k_{on} , and the apparent dissociation rate at +40 mV and at -80 mV, here denoted k_{off}^{open} and k_{off}^{closed} , respectively. Although the kinetic parameters were determined for hKv1.5 channels, they will be assumed to be the same for the canine Kv1.5. In the 10-state model described above, the drug association rate is γ . Because basically all channels will open at +40 mV, the measured association rate, k_{on} , translates directly to our parameter γ . The apparent dissociation rate for the collection of blocked states in the 10-state model is δf_B , with f_B being a voltage-dependent function containing the variable Q . Written out, Q depends on the membrane potential, V , on Q_0 , the ratio of ζ_0 and ϵ_0 , and on Z_Q , the sum of Z_ϵ and Z_ζ ,

$$Q = \frac{\zeta}{\epsilon} = \frac{\zeta_0}{\epsilon_0} \exp\left(\frac{-(Z_\epsilon + Z_\zeta)FV}{RT}\right) = Q_0 \exp\left(\frac{-Z_Q FV}{RT}\right). \quad (13)$$

At the measured membrane potentials +40 mV and -80 mV, the apparent dissociation rate is known. The parameters Q_0 , Z_Q , and δ should therefore be set so that the two conditions

$$\delta f_B(40) = k_{off}^{open} \text{ and } \delta f_B(-80) = k_{off}^{closed}$$

are satisfied. With three parameters and only two relations, there is no unique solution. However, the solution can be parameterized in terms of, e.g., Z_Q . From the parameterized solution (not shown), it can be noted that the lower the value of Z_Q , the more linear the apparent dissociation rate will be as function of V between the known values at +40 mV and -80 mV. Conversely, the larger the value of Z_Q , the more nonlinear the apparent dissociation rate will be between the known values.

The different combinations of Q_0 and Z_Q that reproduce the kinetic parameters for the two compounds have different interpretations. The parameter Z_Q should be interpreted as the total equivalent charge moved during transition of one of the Kv1.5 subunits, when the drug is bound to the receptor site. In the absence of a bound drug molecule, the parameter optimization resulted in a total equivalent charge movement, $Z_\alpha + Z_\beta$, equal to 1.76. This sum constitutes an upper bound for Z_Q provided that binding of the drug, per se, does not induce a conformational change or that the interaction of the drug with the receptor site does not increase the distance translocated by a subunit during opening/closing. However, the binding of the drug may act as an obstacle, decreasing the distance moved in the electrical field, resulting in a value of Z_Q lower than $Z_\alpha + Z_\beta$. In any case, nothing is said about the location of the barrier, the exact values of Z_ϵ and Z_ζ , because of the model reduction. While Z_Q has the interpretation of equivalent charge movement, Q_0 should be interpreted as the relative affinity for the closed state of a subunit in the absence of an electrical force. One can hypothesize that the presence of a bound drug alters this relative affinity. Taken together, in the interpretation of the 10-state model a blocking drug leads to one or two things besides stopping the potassium current through the channel. The subunits may be hindered by the drug, making them move a shorter distance in the electrical field, and the bound drug may alter the subunits relative affinity for the open and closed states. Both events will affect the apparent dissociation rate through their impact on f_B .

Four versions of each compound were considered. For the values 1.75, 1.3, 0.85, and 0.4 of Z_Q , we calculated Q_0 and δ . We did not consider values of Z_Q lower than 0.4 because f_B at this point is already quite linear with respect to V , and little further change is anticipated. All parameter values, including the experimentally determined ones in Lagrutta et al. (2), are shown in Table S1 in the Supporting Material.

Having derived parameter sets for the four versions of each drug, their effect on the APD_{90} of a remodeled PM-cell was tested using a broad range of concentrations. The results from simulations are shown in Fig. 5. Upper and lower rows are results for a stimulation frequency of 1 Hz and 4 Hz, respectively. Column A shows APD_{90} for the four versions of compound 1, column B shows APD_{90} for the four versions of compound 2, and column C shows the differences in APD_{90} between columns A and B. The colors black, blue, green, and red corresponds to the different versions with Z_Q set to the values 1.75, 1.3, 0.85, and 0.4. The concentrations tested are $K_d 2^n$, where K_d is the dissociation constant, and where $n = -3, -2, \dots, 7$ is shown on the x axis. At 1 Hz, APD_{90} was 134 ms in the absence of a drug, increasing to values close to 200 ms at high concentrations for all parameter versions. For compound 1 (column A) the particular version was of considerable importance at intermediate concentrations. The fourth version (red curve), with $Z_Q = 0.4$, gave the strongest prolongation of the AP, and the first version (black curve) gave the weakest. This is reasonable, considering that the voltage dependence of f_B give rise to smaller apparent dissociation rates in the range between +40 mV and -80 mV for $Z_Q = 0.4$ than for $Z_Q = 1.75$. For compound 2 (column B), however, the dose-response curve was virtually independent of the parameter setup. The reason for this is that even though the behavior of f_B also differs for the different Z_Q values for compound 2, it has a ratio of k_{off}^{open} to k_{off}^{closed} equal to 2.27, while this number is 61.7 for compound 1. This means that the relative size of f_B at different Z_Q is less for compound 2 than for

compound 1, and hence that the variation of the apparent dissociation rate is less in its different versions. At 4 Hz, APD_{90} was 118 ms in absence of a drug, increasing to values close to 170 ms at high concentrations for all parameter setups. Also at this frequency, the versions of compound 2 showed practically no differences, in contrast to compound 1 where the parameter setup was again critical for the resulting APD_{90} . In column C, the differences in APD_{90} of all versions of the two compounds are displayed. At 1 Hz, the maximum difference ranged from 31 ms to 19 ms, and at 4 Hz from 26 ms to 10 ms. Based on these simulations it can be concluded, at least when looking at concentrations on a K_d scale, that DPO-1 is a more efficacious blocker than DPO-2, irrespective of the version implemented. This conclusion is consistent with the fact that DPO-1 is a more potent blocker of human Kv1.5 than DPO-2 (2). DPO-1 has also been shown to increase dog atrial refractoriness and to increase human atrial AP duration (2,8). It should also be noted from the simulations that both compounds display a reverse use-dependence, showing a larger increase in APD_{90} at 1 Hz than at 4 Hz.

DISCUSSION

To enable more-realistic studies on the effects of Kv1.5 blockers on the canine atrial AP in silico, two continuous-time Markov models of the guarded receptor type were set up for the Kv1.5 channel. Ion-channel model parameters were determined by using artificially generated data from the RNC model and available experimental data on the

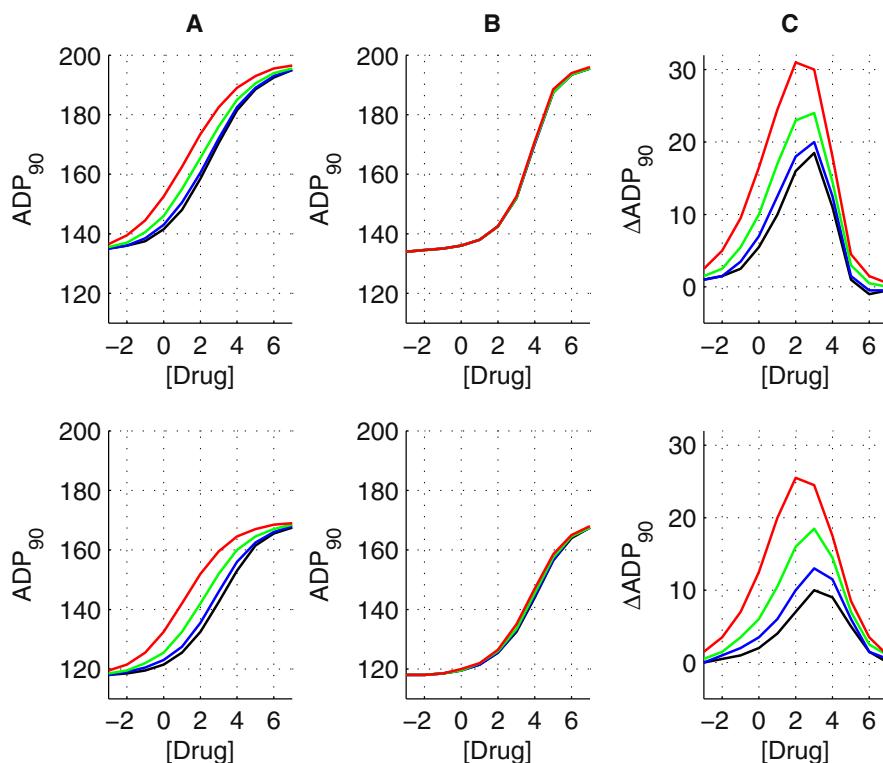


FIGURE 5 APD_{90} values and difference between APD_{90} values generated by the remodeled PM-cell model using the 10-state blocking model and four versions of compound 1 and compound 2, respectively. Versions 1–4 are encoded by black, blue, green, and red. Upper and lower row are for frequencies 1 Hz and 4 Hz, respectively. Columns A and B correspond to compound 1 and compound 2, respectively. Column C shows the difference between A and B. Drug concentration is given by the expression $K_d 2^n$, where K_d is the dissociation constant, and where $n = -3, -2, \dots, 7$.

kinetics of drug-ion channel interaction. The Kv1.5 Markov models were subsequently inserted to the simple AF-remodeled version of the RNC model.

The main findings from this study are:

1. Time- and state-dependent Markov models of open-channel Kv1.5 block gave significantly different results compared to a time- and state-independent model with a downscaled conductance.
2. The outcome of Kv1.5 block on the macroscopic system variable APD₉₀ was dependent on the precise mechanism of block.
3. Open-channel block produced a reverse use-dependent prolongation of the APD₉₀.

Our data clearly suggest that the choice of ion channel model has a large impact on the simulation results. When the dynamic six-state model was compared to a time- and state-independent model, differences in the simulated APD₉₀ as large as 32 ms were observed. This suggests that the simple down-scaling of I_{Kurd} conductance commonly used in many studies may be too unsophisticated for quantitative modeling of drug effects.

To be able to simulate drugs showing an apparent voltage dependence in their dissociation rates, a 10-state Markov model was formulated. To our knowledge, this is the first Markov model of Kv1.5 able to capture this phenomena. It was shown how the experimentally measured kinetic parameters of two clinically relevant drugs (2) could be implemented and the model was subsequently used to simulate the effect of these compounds. The simulations showed that the slower recovery of block at polarized potential of DPO-1 made it a more efficacious drug than DPO-2.

There are two frameworks commonly used for modeling state-dependent block—the modulated receptor theory and the guarded receptor theory. Sections reviewing these concepts and the means by which they have been applied in ion-channel modeling can be found in Brennan et al. (32). Briefly, in the modulated receptor theory, drugs interact with their receptor for all conformational states of a channel, each state differing in its kinetics of drug binding and dissociation. In the guarded receptor theory, drug interaction is limited to certain channel states only. The 10-state guarded receptor model of Kv1.5 developed in this study can be considered a special case of the more-general modulated receptor model. Hypothesizing nonzero affinities for all of the closed states in Fig. 1 C would transform the model into a modulated receptor. While adding vertical transitions between the closed and blocked states might allow for more complex and sophisticated behaviors, it would also make system identification from experimental data more demanding. As was seen for the two compounds, all parameters in our model could not even be uniquely determined. Because of the identifiability issue, and the fact that the 10-state model developed contained enough complexity to describe state-dependent recovery from

block, a general modulated receptor model was not considered in this study.

Tsujimae et al. (9) took another approach to incorporate the effects of voltage- and time-dependent block, including the voltage-dependent recovery from block. They extended a Hodgkin-Huxley formulation of human I_{Kur} by introducing a new multiplicative variable, describing the fraction of unblocked channels, in the expression for this current. Their modified formulation is given by

$$I_{Kur} = y_{Kur} g_{Kur} u_d^3 u_i (V - E_K), \quad (14)$$

where y_{Kur} is the fraction of I_{Kur} that is not blocked by a drug. The remaining variables have the same meaning as those of Eq. 5. Because the variables of Hodgkin-Huxley formulations are independent, open-state block was mimicked by designing the voltage-dependent steady-state profile for the fraction of unblocked channels to resemble, qualitatively, the opposite of the activation profile. Time constant profiles describing the kinetics of formation of and recovery from block were set up based on the same compounds investigated by us in this study. The authors connected known values of the time constant of block at two membrane potentials with a Boltzmann function with a half-value and slope of -40 mV and 5 mV, respectively. However, compared with our implementation of the two DPOs, it was not discussed in detail how to determine the voltage-dependent profiles for the relaxation time-constant and for the steady-state fraction of unblocked channels for some particular drug. Neither did their implementation address the effects from charged drugs. Furthermore, our model allows for any value of the drug concentration, whereas the Tsujimae model only considers one particular steady-state profile for the fraction of unblocked channels.

The outcome of the simulations with the 10-state model leads to important conclusions regarding the experimental protocols used to probe potential I_{Kurd} blockers. Because the precise choice of Q_0 and Z_Q had a major impact on the AP-prolonging effects of one of the compounds, additional measurements of the apparent dissociation rate at other membrane potentials may be crucial. Such additional measurements of the apparent dissociation rate would allow Q_0 and Z_Q to be determined. To increase the chances of good parameter identifiability besides just increasing the number of measurements, the preceding analysis of the 10-state model can be used for optimal experiment design with respect to the voltages at which measurements are to be performed, thereby increasing the predictive power of the model.

It was also found that open-channel block produced larger increases of the canine APD₉₀ at lower frequencies than at higher. This was observed for the six-state model as well as for two specific blockers encoded in the 10-state model, using the remodeled setting. On the other hand, when the six-state model was used with the normal cell setting,

a positive frequency dependence was observed. Because an open-state block can, in itself, never give rise to reverse use-dependence, it must be understood that it is an emergent property of the ensemble of all currents and the role played by Kv1.5 in this context. The environment of other ionic currents can be thought to interact with Kv1.5 and Kv1.5 block in two ways.

First, the voltage-dependent activation of Kv1.5 means that the fraction of Kv1.5 susceptible to block is dependent on how V develops during the AP. This, in turn, depends on the density and activity of all other currents. The rate adaptation to higher stimulation frequencies, described in Ramirez et al. (12), leads to a decreased APD and a lower fraction of open Kv1.5 channels susceptible to block during the AP. This was seen both in the AF setting and for the normal cell (see Fig. 4 B). However, a lower fraction of open channels did not lead to a lower fraction of blocked channels when the drug was added. This was seen in Fig. 4 D where the fraction of blocked channels is shown at different frequencies. Because the fraction of blocked channels during the initial phase of the AP in fact increases with frequency, it appears that the use-dependence of open-state block itself overcomes the decreasing fraction of susceptible channels at higher frequencies shown in Fig. 4 B.

Second, the impact of actually blocking Kv1.5 again depends on all other currents and the relative importance of I_{Kurd} compared to them. In Fig. 4 C, it was shown that in addition to the increase of the blocked fraction with higher frequencies in Fig. 4 D, the fraction of open channels is decreasing. Despite the decrease in the fraction of conducting Kv1.5 at higher frequencies observed for both the remodeled and normal cell, the increase in APD has a positive frequency dependence only in the normal cell. This illustrates the fact that there are several factors in addition to the drug-ion channel interaction itself that determine the outcome of ion channel block on a particular system's property, i.e., APD_{90} . It is most likely that such properties are different depending on species, tissue-type, the degree of electrical remodeling, etc. The finding that DPO-1 produced a positive frequency-dependent increase of APD in human atrial cells (2) does thus not necessarily imply a contradiction to our simulation data. DPO-1 has been shown to increase atrial AERP and thus most likely APD_{90} in a canine disease model, but no information is available regarding frequency dependency or effects in normal tissue (8). Interestingly, Wu et al. (33) showed that a Kv1.5 block produced a reverse use-dependent increase in pig atrial AERP *in vivo*.

The changes in conductances used in the AF setting were chosen as to mimic, in a simplistic way, the major ionic alterations observed during tachycardia-dependent remodeling in the dog according to references given. These ionic changes were experimentally observed and model AP morphology was then fitted to experimental data from cells where these ionic changes had been observed. It is clear

from our data and Ramirez et al. (12) that a block of I_{Kurd} in the native canine myocyte model, apart from increasing the duration of the plateau phase of the AP, most likely also prolongs late repolarization phases of the AP, suggesting an increased effective refractory period and, hence, less likelihood of AF initiating. This prolongation has also been observed experimentally in isolated canine myocytes (34). Although the simplistic remodeled atrial cell model used in this study does not fully capture all changes that occur during the remodeling process and the predictive value, thus, is uncertain, our data clearly suggest that more-complex ion-channel models are a prerequisite for quantitative modeling of drug effects.

SUPPORTING MATERIAL

Model validation and DPO parameter sets are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(10\)01058-1](http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)01058-1).

This study was funded by the Swedish Foundation for Strategic Research, both directly and via the Gothenburg Mathematical Modelling Centre, and by the European Commission through the BioSim Network of Excellence.

REFERENCES

1. Tamargo, J., R. Caballero, ..., E. Delpón. 2009. I_{Kur} /Kv1.5 channel blockers for the treatment of atrial fibrillation. *Expert Opin. Investig. Drugs.* 18:399–416, I_{Kur} .
2. Lagrutta, A., J. Wang, ..., J. J. Salata. 2006. Novel, potent inhibitors of human Kv1.5 K^+ channels and ultrarapidly activating delayed rectifier potassium current. *J. Pharmacol. Exp. Ther.* 317:1054–1063.
3. Wettwer, E., O. Hála, ..., U. Ravens. 2004. Role of I_{Kur} in controlling action potential shape and contractility in the human atrium: influence of chronic atrial fibrillation. *Circulation.* 110:2299–2306.
4. Blaauw, Y., H. Gögelein, ..., M. A. Allessie. 2004. “Early” class III drugs for the treatment of atrial fibrillation: efficacy and atrial selectivity of AVE0118 in remodeled atria of the goat. *Circulation.* 110:1717–1724.
5. Knobloch, K., J. Brendel, ..., K. J. Wirth. 2002. Electrophysiological and antiarrhythmic effects of the novel I_{Kur} channel blockers, S9947 and S20951, on left vs. right pig atrium *in vivo* in comparison with the I_{Kr} blockers dofetilide, azimilide, D,L-sotalol and ibutilide. *Naunyn Schmiedebergs Arch. Pharmacol.* 366:482–487.
6. Regan, C. P., A. A. Wallace, ..., J. J. Lynch, Jr. 2006. In vivo cardiac electrophysiologic effects of a novel diphenylphosphine oxide I_{Kur} blocker, (2-isopropyl-5-methylcyclohexyl) diphenylphosphine oxide, in rat and nonhuman primate. *J. Pharmacol. Exp. Ther.* 316:727–732.
7. Regan, C. P., L. Kiss, ..., J. J. Lynch, Jr. 2008. Atrial antifibrillatory effects of structurally distinct I_{Kur} blockers 3-[dimethylamino]methyl]-6-methoxy-2-methyl-4-phenylisoquinolin-1(2 H)-one and 2-phenyl-1, 1-dipyridin-3-yl-2-pyrrolidin-1-yl-ethanol in dogs with underlying heart failure. *J. Pharmacol. Exp. Ther.* 324:322–330.
8. Stump, G. L., A. A. Wallace, ..., J. J. Lynch, Jr. 2005. In vivo antiarrhythmic and cardiac electrophysiologic effects of a novel diphenylphosphine oxide I_{Kur} blocker (2-isopropyl-5-methylcyclohexyl) diphenylphosphine oxide. *J. Pharmacol. Exp. Ther.* 315:1362–1367.
9. Tsujimae, K., S. Murakami, and Y. Kurachi. 2008. In silico study on the effects of I_{Kur} block kinetics on prolongation of human action potential after atrial fibrillation-induced electrical remodeling. *Am. J. Physiol. Heart Circ. Physiol.* 294:H793–H800.
10. Courtemanche, M., R. J. Ramirez, and S. Nattel. 1998. Ionic mechanisms underlying human atrial action potential properties: insights from a mathematical model. *Am. J. Physiol.* 275:H301–H321.

11. Nygren, A., C. Fiset, ..., W. R. Giles. 1998. Mathematical model of an adult human atrial cell: the role of K^+ currents in repolarization. *Circ. Res.* 82:63–81.
12. Ramirez, R. J., S. Nattel, and M. Courtemanche. 2000. Mathematical analysis of canine atrial action potentials: rate, regional factors, and electrical remodeling. *Am. J. Physiol. Heart Circ. Physiol.* 279:H1767–H1785.
13. Snyders, J., K. M. Knoth, ..., M. M. Tamkun. 1992. Time-, voltage-, and state-dependent block by quinidine of a cloned human cardiac potassium channel. *Mol. Pharmacol.* 41:322–330.
14. Snyders, D. J., and S. W. Yeola. 1995. Determinants of antiarrhythmic drug action. Electrostatic and hydrophobic components of block of the human cardiac hKv1.5 channel. *Circ. Res.* 77:575–583.
15. Delpón, E., C. Valenzuela, ..., J. Tamargo. 1997. Block of human cardiac Kv1.5 channels by loratadine: voltage-, time- and use-dependent block at concentrations above therapeutic levels. *Cardiovasc. Res.* 35:341–350.
16. Longobardo, M., E. Delpón, ..., C. Valenzuela. 1998. Structural determinants of potency and stereoselective block of hKv1.5 channels induced by local anesthetics. *Mol. Pharmacol.* 54:162–169.
17. Choi, B. H., J.-S. Choi, ..., M. S. Kim. 2000. Direct block by bisindolylmaleimide of rat Kv1.5 expressed in Chinese hamster ovary cells. *J. Pharmacol. Exp. Ther.* 293:634–640.
18. Firth, A. L., and J. X.-J. Yuan. 2008. Antagonists of the Kv1.5 potassium channel. *Drugs Future.* 33:31–47.
19. Yue, L., P. Melnyk, ..., S. Nattel. 1999. Molecular mechanisms underlying ionic remodeling in a dog model of atrial fibrillation. *Circ. Res.* 84:776–784.
20. Press, W. H., S. A. Teukolsky, ..., B. P. Flannery. 1993. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, Cambridge, UK.
21. Schmidt, H., and M. Jirstrand. 2006. Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics.* 22:514–515.
22. Schmidt, H. 2007. SBADDON: high performance simulation for the Systems Biology Toolbox for MATLAB. *Bioinformatics.* 23:646–647.
23. Zagotta, W. N., T. Hoshi, and R. W. Aldrich. 1994. Shaker potassium channel gating. III: Evaluation of kinetic models for activation. *J. Gen. Physiol.* 103:321–362.
24. Hindmarsh, A. C., P. N. Brown, ..., C. S. Woodward. 2005. SUNDIALS: suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw.* 31:363–396.
25. Hille, B. 2001. Ion Channels of Excitable Membranes. Sinauer Associates, Sunderland, MA.
26. Yue, L., J. Feng, ..., S. Nattel. 1996. Characterization of an ultrarapid delayed rectifier potassium channel involved in canine atrial repolarization. *J. Physiol.* 496:647–662.
27. Decher, N., P. Kumar, ..., M. C. Sanguinetti. 2006. Binding site of a novel Kv1.5 blocker: a “foot in the door” against atrial fibrillation. *Mol. Pharmacol.* 70:1204–1211.
28. Du, Y.-M., X.-X. Zhang, ..., Y. H. Liao. 2010. Molecular determinants of Kv1.5 channel block by diphenyl phosphine oxide-1. *J. Mol. Cell. Cardiol.* 48:1111–1120.
29. Chay, T. R. 1996. Proarrhythmic and antiarrhythmic actions of ion channel blockers on arrhythmias in the heart: model study. *Am. J. Physiol.* 271:H329–H356.
30. Tveito, A., and G. T. Lines. 2009. A note on a method for determining advantageous properties of an anti-arrhythmic drug based on a mathematical model of cardiac cells. *Math. Biosci.* 217:167–173.
31. Brennan, T., M. Fink, ..., L. Tarassenko. 2007. Modeling effects of sotalol on action potential morphology using a novel Markov model of HERG channel. *MEDICON. IFMBE Proc.* 16:50–53.
32. Brennan, T., M. Fink, and B. Rodriguez. 2009. Multiscale modeling of drug-induced effects on cardiac electrophysiological activity. *Eur. J. Pharm. Sci.* 36:62–77.
33. Wu, S., A. Fluxe, ..., L. Djandjighian. 2006. Discovery and in vitro/in vivo studies of tetrazole derivatives as Kv1.5 blockers. *Bioorg. Med. Chem. Lett.* 16:6213–6218.
34. Fedida, D., J. Eldstrom, ..., D. R. Van Wagoner. 2003. Kv1.5 is an important component of repolarizing K^+ current in canine atrial myocytes. *Circ. Res.* 93:744–751.

PAPER D

A Nonlinear Mixed Effects Approach for Modeling the Cell-to-Cell Variability of Mig1 Dynamics in Yeast

RESEARCH ARTICLE

A Nonlinear Mixed Effects Approach for Modeling the Cell-To-Cell Variability of Mig1 Dynamics in Yeast

Joachim Almquist^{1,2*}, Loubna Bendrioua^{3,4}, Caroline Beck Adiels⁴, Mattias Goksör⁴, Stefan Hohmann³, Mats Jirstrand¹

1 Fraunhofer-Chalmers Centre, Chalmers Science Park, Göteborg, Sweden, **2** Systems and Synthetic Biology, Department of Chemical and Biological Engineering, Chalmers University of Technology, Göteborg, Sweden, **3** Department of Chemistry and Molecular Biology, University of Gothenburg, Göteborg, Sweden, **4** Department of Physics, University of Gothenburg, Göteborg, Sweden

* joachim.almquist@fcc.chalmers.se



OPEN ACCESS

Citation: Almquist J, Bendrioua L, Adiels CB, Goksör M, Hohmann S, Jirstrand M (2015) A Nonlinear Mixed Effects Approach for Modeling the Cell-To-Cell Variability of Mig1 Dynamics in Yeast. PLoS ONE 10(4): e0124050. doi:10.1371/journal.pone.0124050

Academic Editor: Jordi Garcia-Ojalvo, Universitat Pompeu Fabra, SPAIN

Received: October 23, 2014

Accepted: February 25, 2015

Published: April 20, 2015

Copyright: © 2015 Almquist et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work has been supported by funds from the Swedish Foundation for Strategic Research (SSF, to MJ), the Swedish Research Council (VR, to SH) as well as the European Commission FP7 through the UNICELLSYS project (grant 201142, to SH, MG and MJ) and the ISOLATE project (grant 289995, to SH and MG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The last decade has seen a rapid development of experimental techniques that allow data collection from individual cells. These techniques have enabled the discovery and characterization of variability within a population of genetically identical cells. Nonlinear mixed effects (NLME) modeling is an established framework for studying variability between individuals in a population, frequently used in pharmacokinetics and pharmacodynamics, but its potential for studies of cell-to-cell variability in molecular cell biology is yet to be exploited. Here we take advantage of this novel application of NLME modeling to study cell-to-cell variability in the dynamic behavior of the yeast transcription repressor Mig1. In particular, we investigate a recently discovered phenomenon where Mig1 during a short and transient period exits the nucleus when cells experience a shift from high to intermediate levels of extracellular glucose. A phenomenological model based on ordinary differential equations describing the transient dynamics of nuclear Mig1 is introduced, and according to the NLME methodology the parameters of this model are in turn modeled by a multivariate probability distribution. Using time-lapse microscopy data from nearly 200 cells, we estimate this parameter distribution according to the approach of maximizing the population likelihood. Based on the estimated distribution, parameter values for individual cells are furthermore characterized and the resulting Mig1 dynamics are compared to the single cell times-series data. The proposed NLME framework is also compared to the intuitive but limited standard two-stage (STS) approach. We demonstrate that the latter may overestimate variabilities by up to almost five fold. Finally, Monte Carlo simulations of the inferred population model are used to predict the distribution of key characteristics of the Mig1 transient response. We find that with decreasing levels of post-shift glucose, the transient response of Mig1 tend to be faster, more extended, and displays an increased cell-to-cell variability.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Cell biology data has traditionally been acquired by analyzing samples containing a large number of cells. However, data that has been produced by averaging the properties of individual cells may result in misleading interpretations of actual behaviors and underlying mechanisms [1–3]. Today, experimental methods are available that make it possible to measure certain quantities at the level of individual cells. These methods include techniques such as flow cytometry, fluorescence microscopy, and single cell transcriptomics, proteomics, and metabolomics. The development of experimental methods operating on single cells have enabled the study and characterization of cell-to-cell variability, adding a new dimension to the understanding of cell biology. For instance, flow cytometry has been used to study the population variability of the *GAL* regulatory network in yeast [4] and T cell activation [5]. This method produces snapshot data of the population at one or several time points. Each cell is only used for one single measurement, but the method can on the other hand be used to analyze a very large number of cells. For the generation of time-resolved data of the same particular cells, fluorescence microscopy of cells expressing proteins tagged with fluorescent proteins, e.g., GFP, has emerged as a powerful technique. Compared to the high-throughput capabilities of flow cytometry, time-lapse imaging using fluorescence microscopy is typically carried out on a low- or medium-throughput scale. However, this data is substantially richer in information than snapshot data due to the temporal tracking of the same individual cells. Time-resolved data from single cells generated by the combination of microscopy and fluorescent proteins have been used in a large number of studies, including for instance investigations of nuclear accumulation of transcription factor activator ERK2 [1], golgi maturation in yeast [6], and stress-induced nuclear translocation of yeast kinase Hog1 [7] and transcription factors Crz1 [8] and Msn2 [9]. Although various cell-to-cell variability aspects of such data are increasingly being quantified and classified, the development of appropriate mathematical models and modeling approaches is still in its infancy. The need for suitable modeling approaches to describe the variability in dynamic behavior of cell populations has previously been pointed out by the authors of the present work [10], and by others [11], and research activities within this field are expected to increase.

Cell-to-cell variability between genetically identical cells, cultured under the same conditions, originates from the inherently stochastic nature of biochemical reactions. The sources of contribution to variability in gene expression can be separated into the effect of intrinsic noise on the actual reactions themselves, and extrinsic noise in the concentration of components participating in gene expression [12–14]. The latter concentrations are in turn ultimately also determined under the influence of intrinsic noise. Similarly, cell-to-cell variability may additionally originate from the intrinsic and extrinsic fluctuations in other parts of the cellular machinery, such as signalling pathways, and may further be impacted by small local differences in the external environment of individual cells. To mathematically model aspects of variability that are dominated by intrinsic noise, thus displaying noisy dynamics, stochastic approaches are required [2, 15, 16]. These typically involve the chemical master equation, or more commonly, approximations thereof. However, in many cases noise will establish itself as different expression-levels of various proteins, such as metabolic and signalling enzymes [5, 11, 14] and it is in fact often argued that such extrinsic noise is the dominant source of variability [12, 14, 17–19]. Cell-to-cell variability caused by different levels of protein expression can be described by deterministic models, where the values of parameters describing protein concentrations, enzymatic rate constants, etc., are distributed across the population. This approach was taken in a computational study on the behavior of protein kinase cascades [20]. Here, the authors explored the variability in signalling activity through simulations where enzyme concentrations

were randomly sampled from log-normal distributions. In another study on the heterogeneous kinetics of ATK signalling [21], an ordinary differential equation model was fitted to average population data. The behavior of individual cells was then simulated by log-normal sampling of parameters representing enzyme concentrations. Still other examples can be found in modeling of the cell-to-cell variability of apoptosis signalling [18, 22, 23]. Importantly, in neither of these studies were the parameter distributions estimated using single-cell data.

Estimation of parameter distributions for models of heterogeneous cell populations has previously addressed the special case of single cell snapshot data. This has been done using Bayesian approaches, for models with either deterministic [24] or stochastic [19] dynamics, and using maximum likelihood approaches for deterministic models [25]. Recently, Bayesian estimation methods for models with stochastic dynamics have also been customized for the case of time series measurements of the same single cells [26, 27]. In this work we extend on the approaches of deterministic single-cell dynamic modeling by incorporating parameter variability by means of so called nonlinear mixed effects (NLME) modeling, and estimating parameters from time series data using a maximum likelihood approach. NLME is a well-established and wide-spread approach to describe inter-individual variability between subjects of a population. It has a long history with numerous successful applications within various scientific fields [28], in particular including dynamical models in population pharmacokinetics and pharmacodynamics, but is so far largely unexploited for addressing cell-to-cell variability in cell biology-oriented fields. An essential feature of the NLME framework is that all individuals of a population share the same model structure and that differences between subjects are due to different values of model parameters. Thus, the approach is suitable if it is reasonable to assume that the same mechanisms are controlling the behavior of different cells but quantitative details represented by parameter values may differ from one cell to another. This is implemented in the model by letting a subset of the parameters be described by a multivariate probability distribution, whose statistical properties are in turn parameterized by a set of additional parameters. Furthermore, as NLME facilitates the identification of parameters by considering the information from all individuals simultaneously, it is an especially appropriate modeling strategy when considering the often sparsely in time sampled data from single cells. We here apply NLME modeling in the novel context of single-cell data, using it to quantify the dynamic behavior of the yeast transcription factor Mig1.

Glucose and fructose are the most preferred carbon sources in *Saccharomyces cerevisiae* and the presence of any of these sugars activates the transcriptional repressor Mig1. This mechanism is referred to as glucose repression and involves genes required for the uptake and utilization of alternative carbon sources, gluconeogenic genes and the genes required for respiration [29]. A central role in glucose repression is played by the yeast AMP-activated protein kinase, Snf1 [30]. Snf1 is activated in response to glucose depletion by phosphorylation of the Thr210 residue within its activation loop [31]. This activation is promoted by any of the upstream activating kinases Sak1, Elm1 and Tos3 [32–34]. Snf1 phosphorylation is mainly antagonized by the activity of the Reg1-Glc7 protein phosphatase 1 (PP1) [35]. Active Snf1 phosphorylates the transcriptional repressor Mig1 promoting its dissociation from the co-repressor complex Ssn6 (Cyc8)-Tup1 and its nuclear export [36, 37]. Addition of glucose results in a rapid dephosphorylation of Snf1 and Mig1 and subsequently in nuclear accumulation of Mig1 [38, 39].

We recently published single-cell time-series data of Mig1 localization [39]. One of the interesting findings in that study was the behavior of Mig1 when glucose-grown cells experienced a shift in extracellular glucose from a high level (4%) to an intermediate level (1.5, 1.0, and 0.5%). In contrast to shifts to low concentrations of extracellular glucose, in response to which Mig1 persistently re-localized to the cytosol, shifts to intermediate levels of extracellular glucose caused Mig1 to first rapidly exit from the nucleus but then gradually return to its original

nucleocytoplasmic distribution. Thus, it appears that the Snf1-Mig1 system can respond to a change in glucose concentration but depending on the absolute concentration level the system may perform some kind of adaptation. Such a transient response was an unexpected finding and the mechanism behind the apparent adaptation is unknown. In fact, considering a recent study involving 24 different mechanistic mathematical model variants [40], all based on up-to-date understanding of the Snf1-Mig1 system on the molecular level, none of the investigated models would be able to account for the transiently cytosolic Mig1. This can be realized by recognizing that in response to a change in extracellular glucose concentration, the accumulation of activations and inhibitions of every possible path for going from extracellular glucose to Mig1 will drive the Mig1 localization equilibrium in the same direction. Hence, none of the pathway combinations which were implemented in the different model variants are sufficient to explain the non-monotonic nature of the re-entry response. Furthermore, our single-cell time-series data clearly indicated that the extent and timing of the transient re-localization differed between individual cells. Although previous mathematical modeling efforts of the Snf1-Mig1 system have had access to data at the single cell level [40, 41], cell-to-cell variability has not yet been addressed.

In the present work, we set out to describe and quantify the previously reported nuclear exit and re-entry observations, focusing especially on the population variability aspect. Due to the lack of a mechanistically based hypothesis, a simple phenomenological model is developed. Using the NLME approach we are able to show that this model successfully captures the main characteristics of the transient behavior as it varies between individual cells. Importantly, we provide a model-based quantification of the cell-to-cell variability. This variability is reported in terms of estimated distributions of the model parameters. We show that there is a strong correlation between the two parameters determining the time-scales of nuclear exit and re-entry, respectively. This is an interesting finding as it offers a clue to the actual mechanism behind the exit and re-entry behavior. The NLME approach is furthermore compared to the simpler two-stage-approach [42]. While the latter appears to provide reasonable estimates of the median parameter values, it severely overestimates the population variability of the parameters and thus clearly demonstrates why NLME should be preferred. Finally, once parameter estimates have been obtained, the parameter variability of the population can be translated into variability of any model-derived property through Monte Carlo simulations. This type of analysis is used to investigate three key characteristics of Mig1 behavior, namely the median and variability of 1) the response time of Mig1 to a glucose shift, 2) the maximal response of nuclear exit, and 3) the duration of Mig1 cytosolic re-localization. A comparison with a simple non-model-based analysis suggests that these characteristics may not be immediately accessible from data alone. Hence, from a data quantification point of view the model, although only of phenomenological character, is crucial for extracting quantitative information about the process generating the data.

Results

Data description

This study relies on single cell data that we recently published [39]. In brief, these data were acquired from a Mig1-GFP expressing yeast strain using an experimental setup that is combining microfluidics, optical tweezers, fluorescence microscopy, and image processing. We study the scenario where glucose-grown cells are experiencing an instantaneous shift in extracellular glucose, going from 4% glucose to an intermediate level. In total, data from nearly 200 yeast cells, divided over four different data sets, are being used. The experiments are listed in [Table 1](#).

Table 1. Experiments.

Exp Nr	Number of cells	From	To
1	56	4%	1.5%
2	46	4%	1.5%
3	46	4%	1%
4	46	4%	0.5%

List of experiments showing the experiment number, the number of cells used, and the levels of extracellular glucose.

doi:10.1371/journal.pone.0124050.t001

The data from experiments 1 to 4 is shown in Fig 1. The main feature of Mig1 behavior during these glucose exposure patterns is an initial rapid exit from the nucleus, followed by a slower re-entry, where Mig1 levels are readapting towards the baseline level prior to the glucose shift. Both the degree of Mig1 exiting the nucleus and the duration of the complete transient phase seem to increase with decreasing levels of extracellular glucose. All cells seem to share these characteristics but the baseline level of nuclear Mig1 and the timing and degree of exit and re-entry are varying between individual cells.

Setting up a model

Signalling pathways are notoriously challenging to model because of the limited and uncertain knowledge of their components and the interactions between them [43–45]. Since state-of-the-art mechanistic modeling of the Snf1-Mig1 system does not support the transient Mig1 behavior described here [40], we instead aim for a phenomenological model that is as simple as possible, yet flexible enough to describe the Mig1 data. The simplicity of such a model is particularly important in our cases since there is only one measured species from which to calibrate the model, and since we are looking to infer not only parameters values but parameter distributions.

A minimal model of perfect adaptation was considered for modeling the dynamics at the single cell level. This model structure captures the main characteristics of the observed Mig1 behavior, while still providing some degree of interpretability with respect to the components and interactions of the model. The model is illustrated in Fig 2. It consists of two state variables, one representing the time-dependent concentration of $\text{Mig1}(t)$ in the nucleus and one representing the time-dependent lumped effect, here denoted $X(t)$, of one or several unknown components involved in the adaptation. Since we do not know the scaling factor between the observed fluorescent light intensity and the underlying actual concentration of Mig1 molecules, we chose to formulate the model in terms of the observed light intensities. The rate of accumulation of both state variables respond linearly to the level of extracellular glucose, $\text{Glu}(t)$, which is treated as an experimentally controlled input to the system. Considering that the amounts of the involved components of the Snf1-Mig1 system are of the order 4 to 40 thousand molecules per cell [40], a deterministic model is assumed to be sufficient [46]. The mass balance equations for the state variables are defined by

$$\begin{aligned} \frac{d\text{Mig1}(t)}{dt} &= r_1 - r_2 \\ \frac{dX(t)}{dt} &= r_3 - r_4, \end{aligned}$$

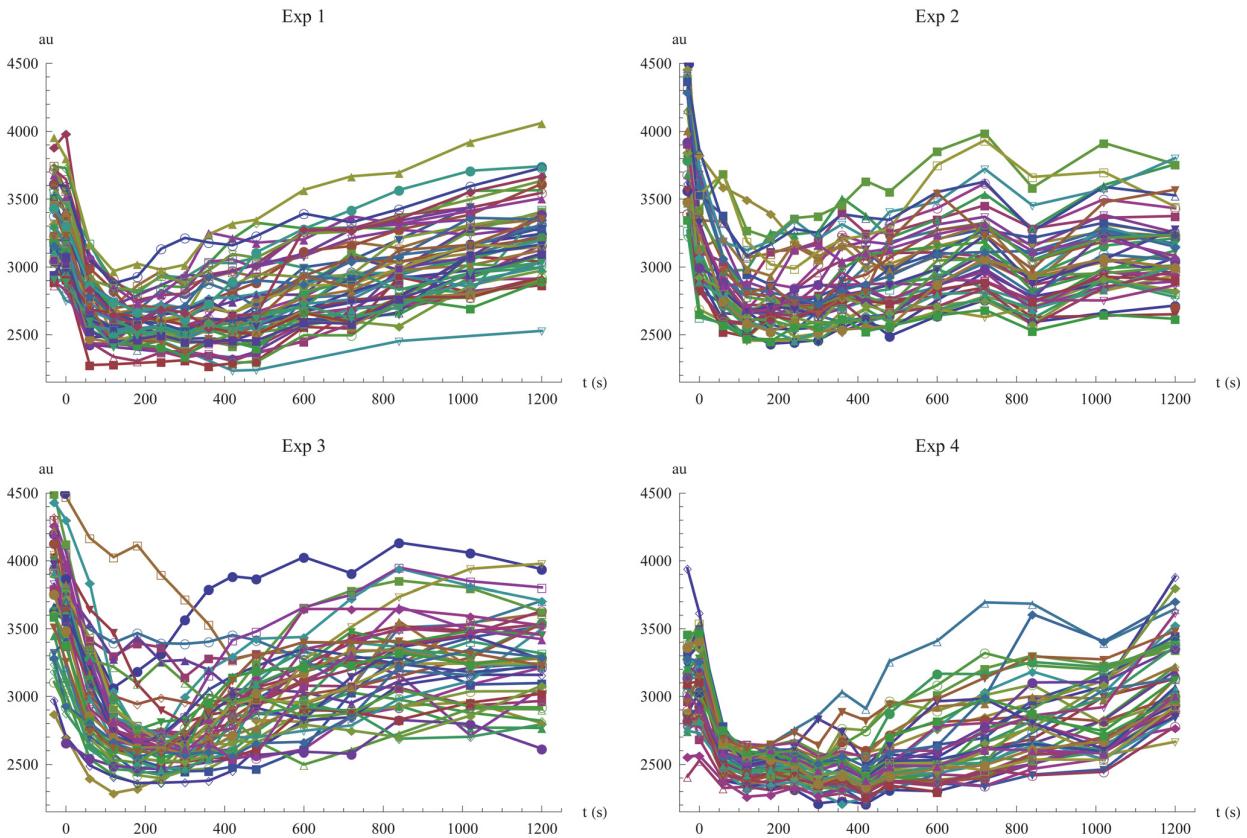


Fig 1. Visualization of all single cell data. Time-series data of fluorescent light intensity for nuclear Mig1 in single cells, shown for the four different experiments. At time zero, the extracellular glucose concentration is changed according to Table 1.

doi:10.1371/journal.pone.0124050.g001

where the rates are defined as

$$\begin{aligned} r_1 &= k_1 \cdot \text{Glu}(t) \\ r_2 &= k_2 \cdot X(t) \cdot \text{Mig1}(t) \\ r_3 &= k_3 \cdot \text{Glu}(t) \\ r_4 &= k_4 \cdot X(t). \end{aligned}$$

The initial conditions are

$$\begin{aligned} \text{Mig1}(-30) &= M_s \\ X(-30) &= X_s, \end{aligned}$$

where we have chosen the initial time to -30 s with the convention that the input to the system is changed at time 0. The input to the system, the extracellular level of glucose, is

$$\text{Glu}(t) = 4 - (4 - g) \cdot H(t),$$

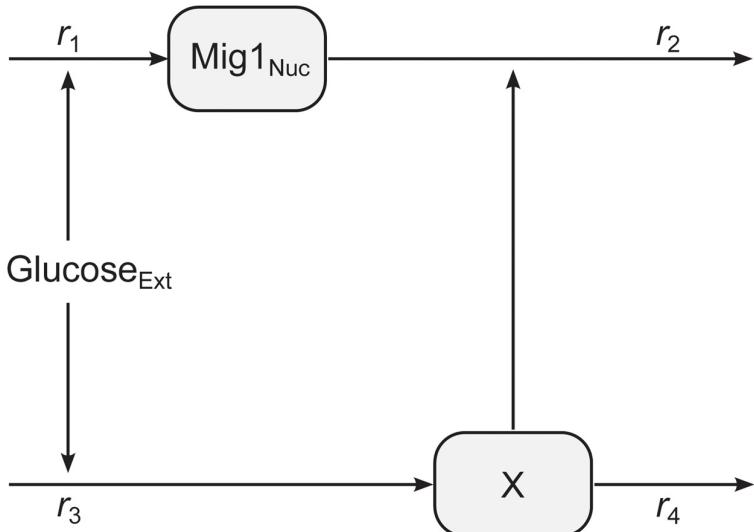


Fig 2. Illustration of the mathematical model. Extracellular glucose is controlling the rate of production of nuclear Mig1 and a hypothetical component X. The level of X in turn modulates the degradation of nuclear Mig1.

doi:10.1371/journal.pone.0124050.g002

where $H(t)$ is the Heaviside step function and g is equal to either 1.5, 1, or 0.5 depending on the experiment. An observation of nuclear Mig1 at time t , y_b , is modeled by introducing an additive error

$$y_t = \text{Mig1}(t) + e_t$$

where $e_t \sim \mathcal{N}(0, s)$, with s denoting the variance of the measurement error. In a previous study of GFP-Mig1 [41] a moderate bleaching effect was identified from averaged single cell data. However, these experiments involved a substantially larger number of measurements (80 per cell and experiment compared to our 15) and the samples were likely bleached to a higher degree. We did not include the effect of fluorophore bleaching in our model, as the majority of cells displayed intensity levels which eventually returned close to the starting levels. In fact, a comparison of the intensities before the glucose shift and at 20 minutes showed that there was an average recovery level of 96%, a number that was determined despite the fact that all cells might not fully have completed their re-entry during the course of the experiment.

It is straightforward to show that the steady-state value of the response variable in the model is independent of the input signal [47]. In the context of Mig1-observations, this means that the model is limited to the experiments where the re-entry phenomena with perfect adaptation is manifested. To be able to describe Mig1 localization in response to a general perturbation in the glucose level, it is clear that some other kind of model would be necessary.

An important question in modeling arises when a model structure has been proposed but parameter values need to be estimated from experimental observations; is there enough information in the data to uniquely determine the parameter values? If we in addition to $\text{Mig1}(t)$ had been able to measure $X(t)$, all parameters would have been *structurally identifiable* [48, 49]. However, when $X(t)$ is not measured it turns out that the model is not identifiable, irrespective of the amount and quality of the data being used. If we let $\tilde{X}(t) = \alpha X(t)$, $\tilde{k}_2 = k_2/\alpha$, and $\tilde{k}_3 = \alpha \cdot k_3$, and multiply the differential equation for $X(t)$ with α , the model equations can

be written

$$\begin{aligned}\frac{d\text{Mig1}(t)}{dt} &= k_1 \cdot \text{Glu}(t) - \tilde{k}_2 \cdot \tilde{X}(t) \cdot \text{Mig1}(t) \\ \frac{d\tilde{X}(t)}{dt} &= \tilde{k}_3 \cdot \text{Glu}(t) - k_4 \cdot \tilde{X}(t).\end{aligned}$$

This transformation leaves the measured state variable $\text{Mig1}(t)$ unchanged, and in this sense results in an equivalent model. Thus, there is a redundancy in the dependence between $X(t)$, k_2 , and k_3 which prevents us from uniquely identifying these parts of the model. The crucial point, however, is that by choosing α to contain either the factor k_2 or $1/k_3$, one of the parameters will cancel out and the transformed model will contain one parameter less. For instance, choosing $\alpha = k_4/(k_3 \cdot \text{Glu}(-30))$, the parameter k_3 will no longer appear in the equations and does not have to be estimated. This particular α also yields a very simple initial condition for $\tilde{X}(t)$. In this way we reduce the complexity of the original model but fully preserve its ability to describe the observed state variable $\text{Mig1}(t)$. The fact that $\tilde{X}(t)$, \tilde{k}_2 , and \tilde{k}_3 are different from the corresponding state variable and parameters of the original model is of no concern to us since they anyway represent aspects of a hypothesized process that is not defined on the molecular level, and hence there is no loss of interpretability. The model could have been reduced with respect to the parameter k_2 instead, but since k_2 will determine the turnover-timescale of $\text{Mig1}(t)$ reduction with respect to k_3 is more convenient.

For simplicity in notation, we will now drop the tildes and let the original names of variables and parameters refer to the reduced model. The equations defining the model in Fig 2 are

$$\begin{aligned}\frac{d\text{Mig1}(t)}{dt} &= k_1 \cdot \text{Glu}(t) - k_2 \cdot X(t) \cdot \text{Mig1}(t) \\ \frac{dX(t)}{dt} &= k_4 \frac{\text{Glu}(t)}{\text{Glu}(-30)} - k_4 \cdot X(t).\end{aligned}$$

Further model simplification can be achieved by acknowledging that the modeled system should be in steady-state at the beginning of each experiment. By assuming a steady-state at $t = -30$, we see that

$$\begin{aligned}0 &= k_1 \cdot \text{Glu}(-30) - k_2 \cdot X_s \cdot M_s \\ 0 &= k_4 - k_4 \cdot X_s,\end{aligned}$$

and thus that the values of the model parameters are constrained by the initial values. From the second equality, we require that $X_s = 1$. We furthermore let the parameter k_1 be a function of the other parameters and of the input according to

$$k_1 = \frac{k_2 M_s}{\text{Glu}(-30)}.$$

This particular choice of reparameterization is motivated by the fact that the parameter k_2 can be interpreted in terms of the turnover-timescale for $\text{Mig1}(t)$ and M_s as the basal level of Mig1 , making the resulting model most convenient.

We now turn to the population aspect of the mathematical model and how to account for the variability of the measured Mig1 dynamics in individual cells. In contrast to the non-random parameter values typically encountered in computational biology, variability between subjects is introduced by letting parameter values be described by probability distributions. Specifically, we chose to let the parameters of the dynamical model described above to be

defined as the product of a so called fixed effect parameter, which involves no randomness, and a so called random effect parameter according to

$$\begin{aligned} M_s &= \bar{M}_s e^{\eta_1} \\ k_2 &= \bar{k}_2 e^{\eta_2} \\ k_4 &= \bar{k}_4 e^{\eta_3}. \end{aligned}$$

Here, the vector of random effect parameters, $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)$, is normally distributed with zero mean and covariance matrix $\boldsymbol{\Omega}$. This means that the parameters M_s , k_2 , and k_4 are log-normally distributed. Their median values are determined by the parameters \bar{M}_s , \bar{k}_2 , and \bar{k}_4 , and their degree of variability is determined by $\boldsymbol{\Omega}$. The particular choice of a log-normal distribution is motivated by the universal appearance of this distribution in nature, ultimately originating from the fundamental laws of chemistry and physics [50, 51]. For instance, the concentrations of several mammalian signalling proteins have been shown to be log-normally distributed [5]. Since the proposed model is not a molecular-level mechanistic model, population variability of its parameters are meant to capture the aggregated effects of the underlying variability in all components relevant to Mig1 localization, ranging from proteins directly involved in Mig1 nucleocytoplasmic transport to proteins involved in sensing and signalling, etc.

Estimating parameters

The experimental data described previously was used to estimate the parameters of the dynamical population model. This was done by maximizing the so called FOCE approximation of the population likelihood, using a gradient-based optimization scheme [52]. Three types of parameters were included in the parameter estimation:

- The fixed effect parameters of the model, \bar{M}_s , \bar{k}_2 , and \bar{k}_4 .
- The variance of the measurement noise, s .
- The parameters used to define the random effect covariance matrix, ω_{11} , ω_{12} , ω_{13} , ω_{22} , ω_{23} , and ω_{33} . Details of the parameterization of the random effect covariance matrix $\boldsymbol{\Omega}$ are explained in the Methods section.

There are in a total 10 parameters to be estimated, collected in the vector

$$\boldsymbol{\theta} = (\bar{M}_s, \bar{k}_2, \bar{k}_4, s, \omega_{11}, \omega_{12}, \omega_{13}, \omega_{22}, \omega_{23}, \omega_{33}).$$

Each of the four experiments were considered separately, resulting in one set of estimates per experiment. The estimated values of the parameters for the different data sets are shown in Table 2. For each estimated parameter, its relative standard error (RSE) is shown within parenthesis. The estimate of the initial median level of nuclear Mig1, \bar{M}_s , is similar throughout the set of experiments. Experiments 1, 2 and 3 are similar with respect to the estimates of the parameters \bar{k}_2 and \bar{k}_4 , while experiments 4 shows a slightly larger \bar{k}_2 and a \bar{k}_4 that is roughly double in size. The estimates of the measurement error variance differ for the different experiments. Moreover, it is clear that the parameters of the dynamical model are determined with high certainty, especially \bar{M}_s which has a RSE of at most 2% in all of the four experiments. The values of the parameters used for constructing the covariance matrix for the random effect parameters are on the other hand somewhat more uncertain but the RSEs are in general still acceptable. One exception to this is RSE for ω_{13} in experiment 1. However, considering that RSE is a relative measure and that the estimate of this parameter value is close to zero, the absolute uncertainty is still low (standard error is 0.00128).

Table 2. Parameter estimates.

Parameter	Exp 1	Exp 2	Exp 3	Exp 4
\bar{M}_s	3.27×10^3 (1)	3.36×10^3 (1)	3.64×10^3 (2)	3.14×10^3 (1)
\bar{k}_2	0.00579 (4)	0.00473 (6)	0.00592 (9)	0.00815 (7)
\bar{k}_4	0.00846 (4)	0.00971 (9)	0.00999 (9)	0.0229 (8)
s	8.73×10^3 (6)	38.1×10^3 (6)	20.8×10^3 (6)	24.1×10^3 (6)
w_{11}	0.0653 (11)	0.0712 (20)	0.0624 (12)	0.0228 (34)
w_{12}	0.0391 (28)	0.0447 (53)	0.0568 (22)	0.0691 (13)
w_{13}	47.5×10^{-6} (25598)	-0.0377 (49)	-0.0649 (24)	-0.0322 (46)
w_{22}	0.231 (12)	0.144 (44)	0.313 (13)	0.252 (15)
w_{23}	0.0398 (108)	0.193 (35)	0.526 (16)	0.281 (25)
w_{33}	0.255 (12)	0.439 (17)	0.567 (12)	0.432 (16)

Estimated parameter values and their corresponding relative standard error (expressed in percentage in the parenthesis), considering each of the four experiments separately.

doi:10.1371/journal.pone.0124050.t002

Table 3 shows the covariance matrix for the random effect parameters, and the corresponding correlation matrix. For each matrix entry its RSE is shown within parenthesis. The correlation in population variability between η_1 and η_2 (associated with M_s and k_2 , respectively) is not very strong, and not showing a clear tendency across the experiments, but is on the other hand not very precisely estimated either. For experiments 2 to 4 there is a moderate negative correlation between η_1 and η_3 (associated with M_s and k_4 , respectively), and here correlation estimates are less uncertain. In these three experiments we also see that there is a substantial correlation

Table 3. Covariance and correlations matrices.

Exp Nr	Ω	Corr
1	$\begin{pmatrix} 0.0058(20) & 0.009(35) & 12. \times 10^{-6}(25684) \\ & 0.055(26) & 0.01(112) \\ & & 0.065(25) \end{pmatrix}$	$\begin{pmatrix} 1 & 0.51(24) & 620. \times 10^{-6}(24736) \\ & 1 & 0.17(102) \\ & & 1 \end{pmatrix}$
2	$\begin{pmatrix} 0.0085(27) & -840. \times 10^{-6}(735) & -0.017(58) \\ & 0.058(54) & 0.085(46) \\ & & 0.19(35) \end{pmatrix}$	$\begin{pmatrix} 1 & -0.038(652) & -0.41(41) \\ & 1 & 0.8(22) \\ & & 1 \end{pmatrix}$
3	$\begin{pmatrix} 0.011(22) & -0.016(66) & -0.037(31) \\ & 0.37(25) & 0.3(27) \\ & & 0.32(25) \end{pmatrix}$	$\begin{pmatrix} 1 & -0.25(60) & -0.61(17) \\ & 1 & 0.86(6) \\ & & 1 \end{pmatrix}$
4	$\begin{pmatrix} 0.0063(25) & 0.0083(75) & -0.014(55) \\ & 0.14(31) & 0.12(37) \\ & & 0.19(32) \end{pmatrix}$	$\begin{pmatrix} 1 & 0.28(69) & -0.41(39) \\ & 1 & 0.74(14) \\ & & 1 \end{pmatrix}$

Covariance and correlations matrices and their corresponding relative standard error (expressed in percentage in the parenthesis), considering each of the four experiments separately. The random effect parameters described by the first to the third row of these matrices, are associated with the fixed effect parameters \bar{M}_s , \bar{k}_2 , and \bar{k}_4 , respectively.

doi:10.1371/journal.pone.0124050.t003

between η_2 and η_3 (associated with k_2 and k_4 , respectively), with the precision in the estimates being quite good. Experiment 1 on the other hand only suggest a weak correlation between η_2 and η_3 , but may nevertheless be compatible with the other experiments since the estimated correlation is highly uncertain.

We additionally determined the maximum a posteriori estimates of the random effect parameters for each individual cell. These are the most likely values of η for an individual given the already estimated probability distribution for these parameters, and are also known as the empirical Bayes estimates (EBEs) [53]. To be able to trust further analysis involving the EBES we determined the so called η -shrinkage, defined as the relative decrease in standard deviation of the EBES compared to the standard deviation defined by the population estimate Ω . These values are shown in S1 Table. It is recommended that shrinkage should not be greater than 20 to 30% to avoid misleading conclusions in EBE-based diagnostics [53]. Although two of the percentages in experiments 2 are approaching such levels, the set of values as a whole should be considered feasible.

The EBES were used to further investigate the correlation between k_2 and k_4 . Fig 3 shows how the EBE values of the random effect parameters associated with \bar{k}_2 and \bar{k}_4 , namely η_1 and η_2 , are distributed in each of the four experiments. For each experiment, a normal distribution fitted to the EBE values is illustrated by two black ellipses, indicating the regions of one and

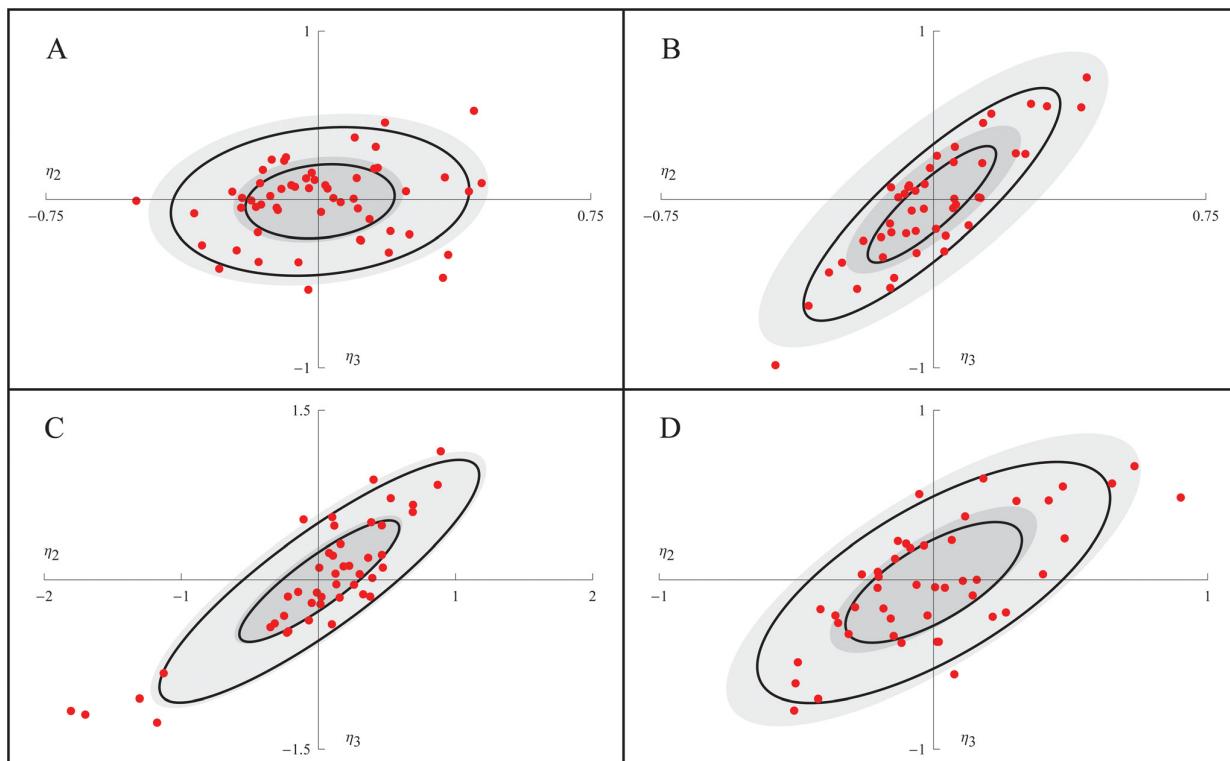


Fig 3. The distribution of maximum a posteriori η . For experiments 1 to 4 (A to D), the EBES of η_2 and η_3 are shown as red points. The regions of one and two standard deviations of a normal distribution fitted to the EBES, and the NLME population estimate of the distribution of η_2 and η_3 , are shown as black and filled gray ellipses, respectively.

doi:10.1371/journal.pone.0124050.g003

two standard deviations. The distribution of η_2 and η_3 defined by the population estimate Ω is similarly illustrated by filled grey ellipses. This analysis confirmed the results displayed in [Table 3](#). Again, there is only a slight correlation of the EBEs in experiments 1 (0.16), as shown in [Fig 3A](#), but a pronounced correlation for the other three experiments (0.86, 0.87, 0.70), as shown in [Fig 3B](#), [3C](#) and [3D](#). The somewhat worse shrinkages of experiment 2 are also seen in [Fig 3B](#) as a difference between the filled and non-filled ellipses, respectively (although yielding very similar variances, please note that the black ellipses are based on fitting the EBEs to a normal distribution while the η -shrinkage is just based on the variance of the EBEs). In experiment 3, [Fig 3C](#), five cells (with numbers #1, #2, #14, #26, and #29) stand out a bit from the others. Because of their comparatively more negative values of the random effect parameters, these cells have smaller effective values of k_2 and k_4 and should therefore display slower dynamics in response to the glucose shift. These cells may constitute a subgroup, but because of the relatively small sample size, and because of potential uncertainty in the EBEs of those cells, it is difficult to say with certainty. To make sure that these cells were viable and intact we went back to the raw images and inspected them manually. All cells looked normal although cell #29 appeared to be smaller and with a less developed nucleus.

Comparing the inferred model to data

The behavior of the model using the estimated parameter values was examined. We simulated the Mig1 dynamics of a typical cell by setting the random effect parameters to zero. For each of the four experiments, this simulation is shown together with the data from all cells in the first row of [Fig 4](#). Additionally, we used the derived EBEs to simulate the model for specific cells and compared the results to the experimental observations. This was done for four representative cells per experiment and the results are shown in rows two to five in [Fig 4](#). Plots of all individual cell data and model simulations for the four different experiments are shown in [S1](#), [S2](#), [S3](#), and [S4](#) Figs, respectively. Despite its simplicity the proposed model captures the different single cell Mig1 dynamics well, including cells with a “median response” (Exp 2 #21), high (Exp 2 #30) and low (Exp 3 #31) initial levels of Mig1, respectively, with fast (Exp 4 #9), and slow (Exp 4 #41) dynamics of the transient behavior, respectively, as well as cells with fewer data points (Exp 1, #30). We also note the unusually slow dynamics of cell #2 in experiment 3. This is one of the cells which we showed previously ([Fig 3C](#)) to have values of the EBEs that deviated from the others cells, and whose slower dynamics was already predicted at that point.

Accounting for background fluorescence

The model was built under the assumption that the observed fluorescent light intensities are proportional to the actual concentration of Mig1. This assumption does not account for the presence of background fluorescence. To test whether the simplification of disregarding any background fluorescence is critical for the outcome of the analysis, we repeated the parameter estimation using the modified observational model

$$y_t = b + \text{Mig1}(t) + e_t,$$

where b is a parameter to be estimated from data. The details of the parameter estimation are described in [S1 Text](#), the results of the parameter estimation is shown in [S2 Table](#) and the corresponding random effect covariance and correlation matrices are shown in [S3 Table](#), and plots of all individual cell data and model simulations for the four different experiments are shown in [S5](#), [S6](#), [S7](#), and [S8](#) Figs. In summary, changing the observation model to account for background fluorescence gave a marginally better fit to data but the parameter estimation

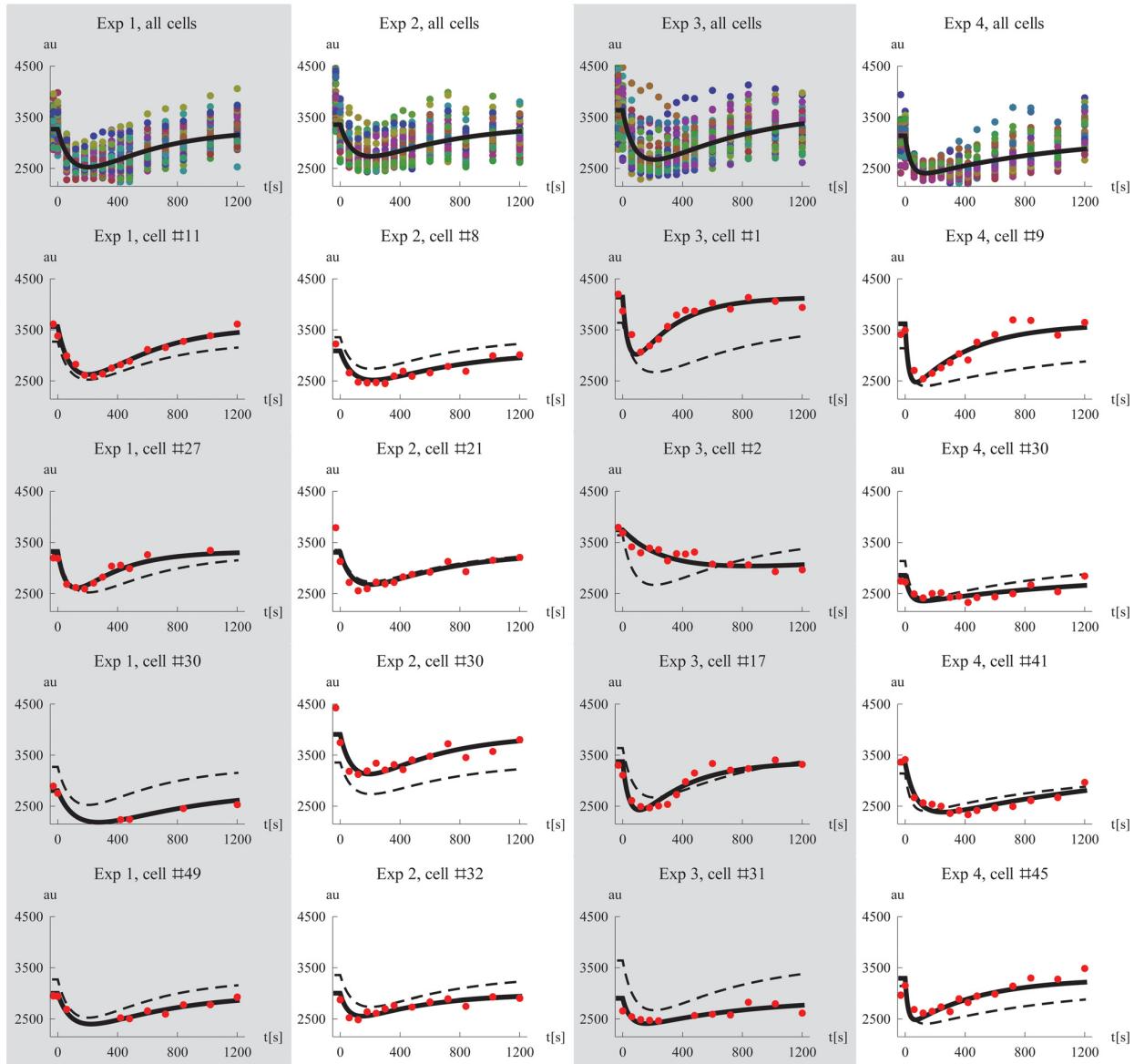


Fig 4. Model simulations and data. The first row show plots of all single cell data together with a simulation of a cell using the median parameters for each experiment, respectively. Rows two to five shows data and corresponding model simulations (derived using the EBEs) for a subset of all cells, exemplifying the fit on the individual level. The simulated median cell is shown in dashed for comparison. Columns one to four correspond to experiments 1 to 4.

doi:10.1371/journal.pone.0124050.g004

suffered from issues with *practical identifiability* [54] and this model variant was therefore not considered further.

Using all data sets simultaneously

Having estimated parameters successfully for each experiment separately, we decided to use all four data sets simultaneously for estimating the model parameters. The details of this analysis are described in [S2 Text](#), the results of the parameter estimation is shown in [S4 Table](#) and the

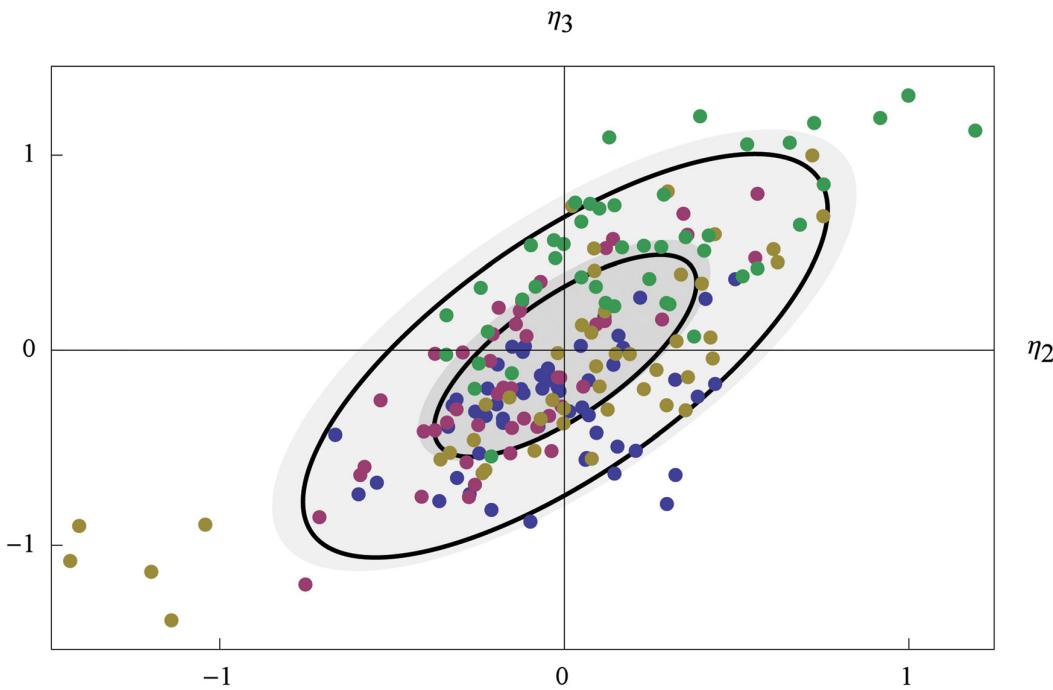


Fig 5. The distribution of EBEs of η for all cells in all experiments. The EBEs from individual cells are color-coded according to the experiments in which their data was produced using blue, pink, yellow, and green, for experiments 1 to 4, respectively.

doi:10.1371/journal.pone.0124050.g005

corresponding random effect covariance and correlation matrices are shown in [S5 Table](#), and plots of all individual cell data and model simulations for the four different experiments are shown in [S9](#), [S10](#), [S11](#), and [S12](#) Figs. We then reinvestigated the distribution of the EBEs of the random parameters associated with \bar{k}_2 and \bar{k}_4 , shown in [Fig 5](#). As in [Fig 3](#), a normal distribution fitted to the EBE values is illustrated by two black ellipses indicating the levels of one and two standard deviations, and the distribution of η_2 and η_3 defined by the population estimate Ω is similarly illustrated by filled grey ellipses. To separate the EBEs belonging to individual cells from the same experiment, we color-coded the dots for experiments 1 to 4 in blue, pink, yellow, and green, respectively. While the EBEs for experiments 1 to 3 display apparently similar distributions, though the five cells from experiment 3 still stand out, it is clear the cells from experiment 4 have consistently higher values of their random effect parameters, especially η_3 . Thus, even if the simulated Mig1 dynamics compare well with the single cell experimental observations, a model using the same parameter distributions $k_2 = \bar{k}_2 e^{\eta_2}$ and $k_4 = \bar{k}_4 e^{\eta_3}$ for all experiments is in some sense still misleading, and the results from the separate analysis should be considered more trustworthy.

Comparing population parameter estimates to the statistics of single subject estimation

If every cell contains sufficient information to precisely estimate the parameters of the dynamical model, the parameters describing the population variability could simply be derived by fitting a parameterized distribution to the collection of all individual estimates. This straightforward approach to population modeling is known as the standard two-stage (STS)

Table 4. Comparison of STS to NLME.

Parameter	Exp 1 All cells	Exp 1 3 cells removed	Exp 2 All cells	Exp 2 2 cells removed	Exp 3 All cells	Exp 4 All cells	Exp 4 2 cells removed
\bar{M}_s	101	100	100	101	100	102	103
\bar{k}_2	105	101	149	104	100	99	96
\bar{k}_4	94	98	140	99	100	97	88
s	30	67	31	58	70	27	64
Var[M_s]	101	97	123	107	109	412	394
Var[k_2]	217	132	8797	490	129	510	177
Var[k_4]	182	153	3329	206	142	839	217

The parameters estimates from the STS approach, either including all cells or removing cells with outlier estimates, expressed as percentages of the corresponding values derived from the NLME approach.

doi:10.1371/journal.pone.0124050.t004

approach [42]. However, even moderate issues with identifiability for the parameter estimation of single cells may lead to biased estimates of population median parameters and overestimation of parameter variability. Being a much easier method to implement, and requiring substantially shorter times for computing the estimates, we decided to test whether the STS approach would be a feasible alternative to NLME. For each of the four experiments, the values of all random effect parameters were set to zero and the values of M_s , k_2 , k_4 , and s were estimated for every cell. The resulting sets of parameter values were subsequently fitted to log-normal distributions. To avoid that extreme parameter estimates from uninformative single cell data sets had an unreasonably large impact on the estimated distributions, we repeated the analysis by removing single cell estimates that had at least one parameter value that differed more than 15 times from the median value of the set of individual estimates. This meant the exclusion of 2 cells from experiment 1, 3 cells from experiment 2, and 3 cells from experiment 4. No outliers were removed from experiment 3. The results of the comparison between STS and NLME is shown in Table 4, expressing the STS parameter estimates as percentages of the corresponding NLME estimates. The STS approach performed acceptably for estimating the median values of all experiments except for experiment 2 when all cells were used. When the outlier estimates had been removed it performs satisfactory for estimating median values in all experiments. The estimates for the measurement variance, s , were in all cases substantially lower. However, this parameter was not assigned to be distributed in the NLME approach, making the comparison more difficult. Importantly, with a few exceptions regarding the variance of M_s , there is a clear overestimation in the variance of the model parameters, and this bias is in some cases considerable. It is obvious that a naive application of the STS approach, i.e., without screening for deviating values first, will give highly questionable estimates of the variability. Additionally we observed that even with a more careful use of the STS approach, variances may still be severely overestimated. For instance, the variance of k_2 in experiment 2 is nearly five times larger when comparing the STS estimate to that of the NLME approach.

To illustrate why the STS approach gives different results than NLME three specific cells were examined more closely (Fig 6). Many cells contain an amount of information that is sufficient for the STS approach to produce similar estimates as the NLME at the single cell level. Fig 6A shows one such example where the simulations using the two different estimates practically look identical. In this example the NLME simulation used the value-triplet (3565, 0.00667, 0.00754) for the parameters (M_s , k_2 , k_4), and the STS simulation used the highly similar values (3578, 0.00668, 0.00738). When all cells are included in the analysis, a few rare time-series

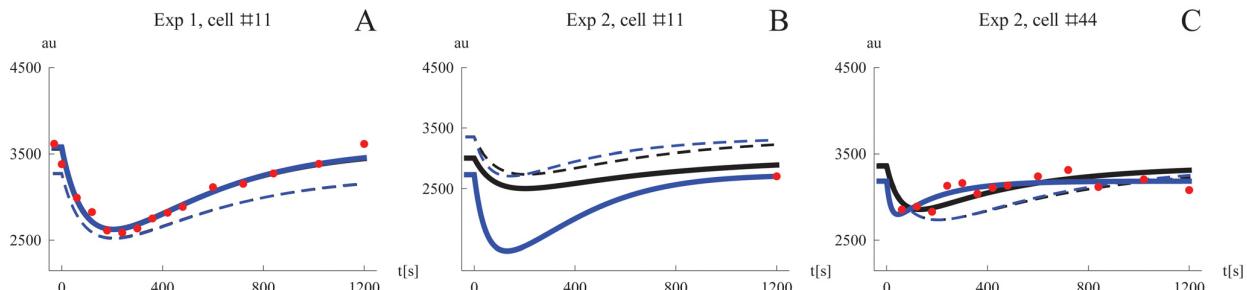


Fig 6. Comparing simulated Mig1 dynamics for individual cells using parameter from the STS and NLME approaches. Simulations with parameter values from the STS analysis are shown in blue, and in black for NLME. Simulations of typical cells are shown in dashed. A. An information-rich data set which by itself allows precise estimation of model parameters. The typical STS cell was simulated using the median parameters considering removal of outliers. B. The extreme case of an uninformative data set (only one data point). Here the STS approach may produce arbitrary parameter estimates which leads to questionable simulations as well as corrupting the population statistics of individual estimates. In this example the typical STS cell was simulated using the median parameters *without* considering removal of outliers, producing a different results compared to the typical NLME cell. C. A cell where the information content is too low for estimating all parameters with high precision. Model fits like this contribute to overestimation of parameter variability on the population level. The typical STS cell was simulated using the median parameters considering removal of outliers.

doi:10.1371/journal.pone.0124050.g006

containing only one or two data points will be used. Fitting all model parameters to such data will produce completely arbitrary estimates due to lack of identifiability. This kind of scenario is shown in Fig 6B. Because the NLME approach is “borrowing” information (in form of the empirical prior) from the other cells when computing the estimate for a single cell, this simulation still resembles the median cell, while the STS simulation on the other hand produces a much more extreme behavior. We can also see that the simulated median cell of the population differs when the median parameters has been determined from all individual estimates. In this example the NLME simulation used the values (3007, 0.00442, 0.0104), while the STS used the very different values (2732, 0.0201, 0.00509). The inclusion of cells like these in the analysis is the reason why the STS approach where no estimates were discarded performed so badly. Fig 6B also shows that the dynamics of a typical cell derived from the STS approach without discarding outliers may differ to the typical cell of the NLME approach. As shown in Table 4 the STS approach can be improved by removing obvious outliers from the set of individual cell parameters. Although it is straightforward to remove parameter estimates from obviously noninformative data sets, e.g., time-series containing only a single data point, such preprocessing will to some extent be arbitrary. Consider for instance the single cell data in Fig 6C where the NLME and STS simulations used the values (3360, 0.00619, 0.0171), and (3185, 0.0151, 0.0557), respectively. There are 13 data points for this cell, yet it lacks the good identifiability properties from the example in Fig 6A. In such cases the STS approach tend to produce exaggerated, but not extreme, estimates, which contributes to bias and variability overestimation on the population level.

Predicting the variability of the response activation time, amplitude, and duration

Having established an NLME model, it is possible to repeatedly simulate this model in order to determine the population-distribution of any property being described by the model. This was done to compute the population statistics of three quantitative measures of the transient Mig1 dynamics:

- **Response time.** The time it takes to reach the lowest concentration of nuclear Mig1 after a shift in extracellular glucose.
- **Amplitude.** The amplitude of the response measured in % below the baseline.
- **Duration.** The total time during which nuclear Mig1 remains below the level of half-maximum response.

These measures are illustrated in Fig 7A. According to the estimated population variability of the parameters M_s , k_2 , and k_4 , we randomly created 100 000 *in silico* cells per experiment and simulated their Mig1 dynamics. The distributions and the typical values (medians) of the response time, amplitude, and duration are shown in Fig 7B, 7C and 7D, respectively. The typical values are also shown in Table 5. We observe that the simulated median response time is similar for concentrations of 1.5% and 1% glucose, respectively, but decreases markedly at 0.5%. Additionally, there is an increased variability of the response time for the intermediate concentration. The simulated amplitude of the Mig1 response exhibited quite small differences between the three conditions, both with respect to the median and the variability. A clear increase in median duration of the simulated response was observed as glucose concentrations decreases. The variability of the duration also increased with decreasing glucose levels. A similar behavior of the response duration was observed also when this quantity was defined by other levels than 50% of the maximum response (not showed).

As a comparison to the model-based predictions of Mig1 dynamics, a simple non-model-based analysis was performed directly on the data and on dense data sets generated by smoothing and resampling the experimental data. The results from the simple analysis are shown together with the model-based predictions in Table 5. The simple analysis gave similar results for the amplitude, but did not identify an increased duration with decreasing extracellular glucose concentrations, and did furthermore suggest an opposite dependence of the typical response time on extracellular glucose concentrations when compared to the model-based predictions. Also, compared to the smooth distributions from the model-based analysis, the corresponding population histograms from the simple analysis were much less informative due to the limited number of cells and/or a binning based on the rather few discrete time points of the data, as shown for the simple analysis of the experimental data in S13 Fig.

Discussion

State-of-the-art experimental techniques such as fluorescence microscopy allow time-resolved data to be collected from individual living cells. This development has provided researchers with tools enabling them to investigate various aspects of cell-to-cell variability in cell populations. The progress of single cell experimental methods requires a parallel advancement in the development of mathematical models for describing cell population heterogeneity. We propose that so called nonlinear mixed effects (NLME) models, a class of models that for example is used for modeling variability between individuals in pharmacological studies, also may be adopted for modeling cell-to-cell variability in molecular biology. The usefulness of this framework was demonstrated by applying a model of this type to study the localization dynamics of the yeast transcription factor Mig1. This protein is a key component in the regulation of carbon metabolism, responsible for repressing a larger number of genes in the presence of glucose. Using an NLME model we were able to quantify and simulate the cell-to-cell variability of yeast cells with respect to their behavior of Mig1. Comparing the proposed modeling methodology to a second more intuitive approach, we showed that the former is crucial in order to not overestimate the variability. An additional comparison of the NLME model to a simple non-

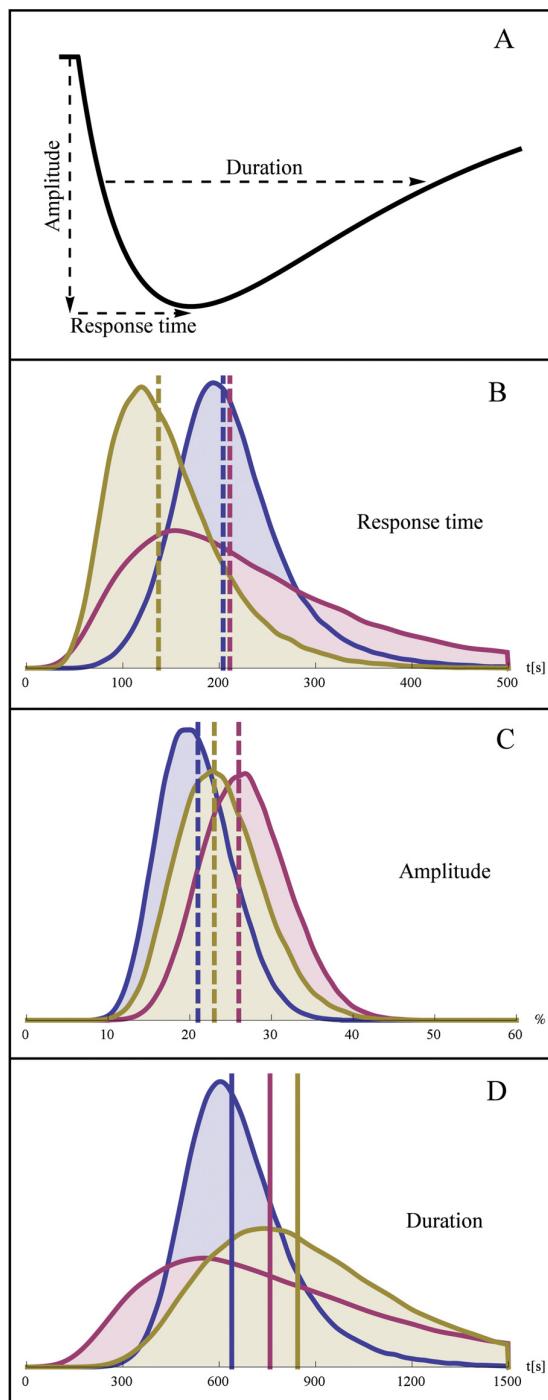


Fig 7. Distribution of the model-derived quantities response time, amplitude, and duration. A. Illustration of response time, (negative) response amplitude in % of baseline, and duration of half-maximal response. Distribution of activation (B), amplitude (C), and duration, for experiment 1 and 2 (blue), experiment 3 (pink), and experiment 4 (yellow). The typical cells (median response) are indicated by vertical dashed lines. Distributions of the model-derived quantities were determined from 100 000 Monte Carlo simulations per experiment.

doi:10.1371/journal.pone.0124050.g007

Table 5. Typical values of response time, amplitude, and duration.

Quantity	Method	Exp 1 and 2 (1.5%)	Exp 3 (1%)	Exp 4 (0.5%)
Response time [s]	Model	204	211	137
Response time [s]	Simple analysis, experimental data	240	270	420
Response time [s]	Simple analysis, smoothed data	229	265	413
Amplitude [%]	Model	21	26	23
Amplitude [%]	Simple analysis, experimental data	23	29	23
Amplitude [%]	Simple analysis, smoothed data	23	28	22
Duration [s]	Model	639	758	844
Duration [s]	Simple analysis, experimental data	780	630	780
Duration [s]	Simple analysis, smoothed data	768	597	757

Typical values of time to full response, the amplitude of the response, and the duration of the response, obtained from the NLME model and from a simple non-model-based analysis using either the original or smoothed experimental data. The typical values of derived using the model were determined from 100 000 Monte Carlo simulations per experiment.

doi:10.1371/journal.pone.0124050.t005

model-based analysis indicated that modeling may be required for reliable interpretation of population data.

Population model of Mig1 dynamics

We recently reported on a novel and unexpected aspect of Mig1 dynamics, namely the transient exit and subsequent nuclear re-entry of this protein in response to a shift from high to intermediate concentrations of extracellular glucose [39]. Similar transient responses followed by perfect adaption have been observed for other signalling proteins such as nuclear ERK2 in response to EGF levels [1], and the yeast kinase Hog1 in response to hyperosmotic shock [55]. Since the current understanding of the Snf1-Mig1 system does not provide a mechanistic basis for the apparent adaptation behavior, a simple phenomenological model of perfect adaptation was introduced to describe the observed Mig1 dynamics. The proposed model is a well known dynamical modeling motif and has previously been presented as one of the basic signal-response elements of regulatory networks [47]. Due to its simpler structure, the qualitative behavior of the model is limited to adaptation, with the parameter values controlling the quantitative details of this behavior, and it can therefore not be used as a general-purpose model of Mig1 localization in response to extracellular glucose. To account for the observed cell-to-cell variability of Mig1 dynamics so called random effect parameters were introduced to the model. In contrast to most dynamical models used in computational biology, a subset of the model parameter values are now stochastic variables characterized by a distribution rather than scalar values. Although the model was not based on known molecular mechanisms for Mig1 regulation, it was successful in describing the experimental observations of Mig1 dynamics. It is however clear that even though such a phenomenological model can fit the data it may not provide the same fundamental insights of a mechanistically based model. Though, given the circumstances of limited knowledge of the mechanistic details of the Snf1-Mig1 system, we believe that the proposed model has an appropriate level of complexity, especially considering the population variability aspect, and that it may be a stepping stone towards future mechanistic models.

Parameter estimates

The model performs well with similar median values of the time constants k_2 and k_4 both for 1.5% and 1% glucose, although with some variations in their variability. However, in experiment 4 the estimated time constant of the adaptation process, k_4 , was larger and a slightly larger value of k_2 was obtained as well. The fact that other parameter values are required for this particular experiment can be seen as an indication that this level, 0.5% glucose, is close to a threshold in the behavior of the Snf1-Mig1 system. Indeed, this was also observed in experiments, where the transient behavior disappears for extracellular glucose levels below 0.5% [39]. It also suggests that to model all four experiments simultaneously, the linear response to glucose, as defined by the adaptation model, may not be sufficient.

Although the estimates of \bar{k}_2 and \bar{k}_4 appeared to be determined with good precision in the separate analysis, we decided to fix these parameters, including their distribution within the population, and estimate them from all experiments simultaneously. The resulting estimates were close to the average of the separate estimates. However, from the distribution of the EBEs (the maximum a posteriori estimates of the random parameters η) it was obvious that the EBEs from the fourth experiment formed a separate cluster. This most likely violates the assumption that the random effect parameters from the different experiments are identically distributed and confirms what was already suspected based on the different and quite well-determined values of \bar{k}_2 , and \bar{k}_4 in the separate analysis of experiment 4. Thus, the simultaneous analysis of all four experiments again suggests that the characteristics of Mig1 regulation is changing at a glucose level around 0.5%.

To account for background fluorescence we set up an alternative model of the measurement process. This did only result in a marginal improvement in the ability to explain the data, and since parameter estimation for this model appeared to experience problems with practical identifiability, it was not considered further. We want to stress that this does not mean that there is no background fluorescence, only that with the alternative model and the available data it appears unfeasible to estimate it. Finally, even though the results from this altered model should be interpreted cautiously due to the issues with parameter identifiability, we note that the model behavior was highly similar to the original model and that the correlation in population variability between k_2 and k_4 remained.

Interpreting the model

Mig1 is continuously being transported in both directions across the nucleocytoplasmic interface and that its localization is dependent on the balance between these fluxes [39]. A change in Mig1 localization is thus due to a change in the balance between the rates of nuclear import and export. In light of this, the model can be interpreted as two counteracting mechanisms on Mig1 cellular localization: One quickly responding mechanism that promotes transport of Mig1 into the nucleus in response to an extracellular glucose signal (r_1), and another delayed mechanism that counteracts the first one by promoting nuclear exit in response to glucose (the modulation of r_2 by X). However, our present understanding of the signalling network controlling Mig1 activity does not include any mechanism that operates by favoring phosphorylation and cytosolic localization in response to the presence of glucose. Moreover, we observed a strong correlation in the cell-to-cell variability of k_2 and k_4 , the parameters which determine the time scales of the two counteracting mechanisms. This means that if a mechanism of the second kind existed, it must be highly coordinated with the first one. As this might require a very precise orchestration in the expression of the hypothetically involved signalling components, we consider the explanation of counteracting mechanisms for the transient behavior even less likely. Taken together, it appears more plausible that the transient pattern is already

present in the dynamics of an upstream pathway component that is controlling Mig1 localization according to the first type of mechanism. At least three candidate components may be considered for transmitting such a transient signal to Mig1:

- **Snf1.** This is a strong candidate since we know that on the averaged population level, Snf1 displays a temporal phosphorylation pattern that is similar to that of Mig1 localization [39]. However, the dynamics of Snf1 phosphorylation on the single cell level has not been investigated.
- **Glc7-Reg1.** Although mathematical modeling results and the lack of direct experimental evidence disfavor a direct regulation of Mig1 by Glc7-Reg1 [40], this scenario can not be ruled out. This phosphatase may alternatively transmit a transient signal indirectly via its effect on Snf1 phosphorylation.
- It has been observed that constitutively phosphorylated Snf1, as the result of overexpressing its upstream kinase Sak1, did not affect either Mig1 phosphorylation or its localization in the presence of glucose [40]. Based on this it was suggested that Snf1 activation is a necessary but not sufficient condition for mediating glucose de-repression, and that there must be a second glucose-regulated step directing Snf1 to Mig1. Such a mechanism may constitute the upstream source of the transient signal.

A combination of these scenarios would also be possible. Furthermore, the transient pattern need not emerge at the level of one of these components but could be present even further upstream, perhaps even in glycolysis itself which in a not fully understood manner generates the signal(s) for Snf1-Mig1 regulation. Further investigating the origin of the transient behavior, and the mechanisms behind its cell-to-cell variability, would be an interesting proposition for future single-cell studies.

A moderate negative correlation in the population variability of M_s and k_4 was also found. This suggests a negative correlation between the levels of Mig1 and the timescale of the hypothesized adaptation process. This may very well be reasonable considering that molecular processes of the Snf1-Mig1 system which directly involve the Mig1 protein, such as phosphorylation and inter-compartment transport, may be subject to saturation effects. Thus, in cells where Mig1 levels are higher than average, the adaptation tends to be slower since a higher number of molecules has to be regulated by a capacity-limited system.

Predicting the variability in response time, amplitude, and duration

Estimates of how parameters vary across the population can not only be analyzed as such, but they can also be used to derive the population variability of any system behavior described by the model. This can be achieved by Monte Carlo simulations using the inferred population model. Such model-based quantification is a powerful tool since it allows us to compute the cell-to-cell variability in aspects of Mig1 regulation which are not easily measurable directly from the time-series data. We used this approach to predict the population variability in three key determinants of the transient Mig1 response. From the results of this analysis (Fig 7) the following was concluded:

- The response time decreases as the level of the secondary glucose concentration decreases. Compared to the 1.5% level, the intermediate level (1%) additionally displays an increased cell-to-cell variability in the response time.

- The amplitude of the response, as determined relative to the baseline Mig1 level of each cell, appears to be largely independent of the glucose level, both with respect to its median value and with respect to its variability.
- The duration of the transient response is increased as the glucose level is decreased. Compared to the 1.5% level, there is also a clear increase in cell-to-cell variability of the duration.

To summarize, as the level of glucose after the shift is decreased, the transient Mig1 response tended to be faster and more extended, as well as showing an increased cell-to-cell variability in both of these two characteristics. Interestingly, we also note that all distributions of the investigated response characteristics appear to be log-normally shaped.

The model-based simulations of variability in response time, amplitude, and duration were also compared to a simple analysis based directly on the experimental data and on the corresponding smoothed and resampled data. Contrary to the model-based results, the simple data analysis did not identify an increasing duration of the response with decreasing extracellular glucose concentration, and did furthermore imply an increasing, rather than decreasing, response time with decreasing extracellular glucose concentration. Although such differences will depend on both the particular model used and on how the simple analysis is executed, the comparison suggests that a model-based approach may be more reliable for studying cell-to-cell variability in sparse or noisy data.

NLME should be preferred to STS

We compared the results from NLME modeling to the more naive STS approach, which consists of performing parameter estimation on single cell data separately and subsequently fitting parameterized distributions to the resulting set of point estimates. Since the estimation of parameters for individual subjects do not rely on information from the rest of the population, the STS approach may tend to over-fit the data, potentially leading to biased estimates but even more commonly to overestimation of parameter variability [42]. Although the two methods provided comparable estimates of the median parameter values, the STS approach severely overestimated parameter variability. The results were particularly bad when estimates from some of the most sparse data sets were included. On the other hand, the NLME modeling approach was fully capable of handling these sparse data sets. In fact, even individuals with just a single observation were feasible and added information to the estimation. For the present study this meant that we did not have to discard any data, allowing us to use the available measurements optimally. Our data included up to 15 data points per individual cell. It is however realistic to assume that some single cell studies may involve substantially sparser sampling of certain quantities, creating an even stronger motivation in favor of the NLME approach compared to STS.

It must be recognized that sparseness in data, determined from counting the number of observations as such, may be a poor indicator for determining if the STS approach will be appropriate. Our comparison of the individual fit of the STS approach to the EBE-based estimate resulting from the NLME population estimation in Fig 6C illustrates this point. If this data set had been very rich in information the NLME-derived population prior would have had a minor impact on the EBE parameter estimate, and the resulting dynamics, and the two approaches would have produced similar results for this cell. Since this was not the case, it is clear that data sets which are not obviously sparse in the sense of containing very few observations (this data set contained 13 data points) are not automatically suitable for the STS approach. The important question is rather whether the balance between information content in the data and the complexity of the model allows parameters to be estimated with high precision,

considering the individual data sets in isolation. Thus, the advantage of NLME over STS is ultimately determined not only by the data sets at hand but also by the particular model being used. Another way of looking at the NLME approach compared to the STS approach would therefore be that the complexity of the model can be allowed to increase, beyond the point of practical identifiability in single subjects, as long as there is enough data on the population level.

Parameter estimates of individual single cell data have previously been performed in a model of the NF- κ B signalling pathway [56]. Here, 6 parameters were estimated for 20 different cells using 15 data points per cell, and the authors noted that some of the parameters were estimated with a quite high uncertainty. Had parameter distributions been fitted to these single cell estimates, the risk of overestimating parameter variability would probably had been high. In another mathematical modeling study of cell-to-cell variability [23], parameter estimation at the single cell level was performed by complementing the single cell time lapse data with other types data, with the purpose of increasing parameter identifiability.

The need for population modeling frameworks

The idea of applying hierarchical modeling, such as NLME, to longitudinal population data acquired at the level of single cells has previously been acknowledged and outlined by the authors of this work [10]. Since then, initial efforts towards single cell modeling using the NLME approach have in fact been considered in a few cases [57, 58], but the full potential of the approach has yet to be realized. The present study is to our knowledge the first one to combine, and in detail cover, aspects of NLME modeling such as uncertainty of estimates, investigation of EBEs and comparison of simulations to single cell data, and using an estimated model for prediction. Also, this is the first study in which NLME has been applied not only with a focus on its technical aspects but also with an ambition to advance the understanding of cell biology.

In parallel with the developments within NLME modeling, single cell time series data have recently also been approached using hierarchical Bayesian methods [26, 27]. In addition to extrinsic variability these efforts also considered intrinsic noise. Although a deterministic approach seems to describe the single cell Mig1 data studied here quite well, an extension of the NLME approach to also cover uncertainty in the dynamics would be interesting. One way of achieving this would be to replace the ordinary differential equation by so called stochastic differential equations (SDEs). The combination of NLME and SDE has previously been considered in pharmacokinetics and pharmacodynamics [59–61]. Not only would this allow intrinsic noise to be addressed within the NLME framework, but the SDEs could also be used to account for miss-specification of the deterministic parts of a model. Applying dynamical modeling with SDEs towards this end has previously proven useful for guiding the process of model development [62]. This strategy may be especially rewarding for modeling of signalling transduction pathways, as these systems typically suffer from limitations and uncertainty in the information needed for setting up models.

The need for modeling frameworks that are able to address single cell data is perhaps most clearly demonstrated by a growing number of studies in which such data was collected but then averaged during the computational analysis [41, 63]. We predict that hierarchical modeling frameworks such as NLME modeling will become even more important as single cell experimental methods continue to develop, and as the biological questions will involve the single cell perspective to a larger extent. In the future, dynamical modeling of non-genetic cell-to-cell variability may not only be relevant for basic research but also become an important ingredient in various applied fields of life science such as quantitative pharmacology [11] and industrial biotechnology [64]. As previously pointed out [58], an intriguing future prospect of single cell

NLME modeling is the inclusion of so called covariates in the model. Covariates are known individual-specific variables which are used to account for predictable sources of the variability. In pharmacokinetic modeling, which frequently uses the NLME approach, covariates may for instance include weight, age, and sex. In the context of single cell modeling, the addition of covariates to the model could be used to incorporate cell-specific information such as size, shape, or age, in addition to the time-series data. Another important challenge for system identification from population data is the development of methods that can handle the combination of measurements at the single cell level with the traditional type of data produced from averaging over many cells.

Methods

The yeast strains, experimental setup, and imaging and image analysis, have been described previously [39].

Parameter estimation for NLME models

NLME models are often used in situations where sparse time-series data is collected from a population of individuals subject to inter-individual variability. These models contain both so called fixed effect parameters, being non-random, and so called random effect parameters, which are determined by some statistical model. Given a set of population data and a NLME model, the fixed effect parameters can be estimated according to the maximum likelihood approach. The likelihood subject to maximization is the so called population likelihood. This is a special kind of likelihood that has been marginalized with respect to all random effect parameters, and that is taking the observations from all individuals of the population into account. We now state the general form of a NLME model, the population likelihood, and its approximation by the so called FOCE method.

Consider a population of N subjects and let the i th individual be described by the dynamical system

$$\begin{aligned}\frac{d\mathbf{x}_i(t)}{dt} &= \mathbf{f}(\mathbf{x}_i(t), \mathbf{u}_i(t), \mathbf{Z}_i, \boldsymbol{\theta}, \boldsymbol{\eta}_i, t) \\ \mathbf{x}_i(t_0) &= \mathbf{x}_{0i}(\mathbf{u}_i(t_0), \mathbf{Z}_i, \boldsymbol{\theta}, \boldsymbol{\eta}_i),\end{aligned}$$

where $\mathbf{u}_i(t)$ is a time dependent input function, \mathbf{Z}_i a set of covariates, $\boldsymbol{\theta}$ a set of fixed effects parameters, and $\boldsymbol{\eta}_i$ a set of random effect parameters which are multivariate normally distributed with zero mean and covariance $\boldsymbol{\Omega}$. The covariance matrix $\boldsymbol{\Omega}$ is in general unknown and will therefore typically contain parameters subject to estimation. These parameters will for convenience of notation be included in the fixed effect parameter vector $\boldsymbol{\theta}$. A discrete-time observation model for the j th observation of the i th individual at time t_{ij} is defined by

$$\mathbf{y}_{ij} = \mathbf{h}(\mathbf{x}_{ij}, \mathbf{u}_{ij}, t_{ij}, \mathbf{Z}_i, \boldsymbol{\theta}, \boldsymbol{\eta}_i) + \mathbf{e}_{ij},$$

where

$$\mathbf{e}_{ij} \sim N(0, \mathbf{R}_{ij}(\mathbf{x}_{ij}, \mathbf{u}_{ij}, t_{ij}, \mathbf{Z}_i, \boldsymbol{\theta}, \boldsymbol{\eta}_i)),$$

and where the index notation ij is used as a short form for denoting the i th individual at the j th observation. Furthermore, we let the expected value of the discrete-time observation model be denoted by

$$\hat{\mathbf{y}}_{ij} = \mathbf{E}[\mathbf{y}_{ij}].$$

Given a set of experimental observations, \mathbf{d}_{ij} , for the individuals $i = 1, \dots, N$ at time points $j = 1, \dots, n_i$, we define the residuals

$$\boldsymbol{\epsilon}_{ij} = \mathbf{d}_{ij} - \hat{\mathbf{y}}_{ij},$$

and write the population likelihood

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N \int p_1(\mathbf{d}_i | \boldsymbol{\theta}, \boldsymbol{\eta}_i) p_2(\boldsymbol{\eta}_i | \boldsymbol{\theta}) d\boldsymbol{\eta}_i, \quad (1)$$

where

$$p_1(\mathbf{d}_i | \boldsymbol{\theta}, \boldsymbol{\eta}_i) = \prod_{j=1}^{n_i} \frac{\exp\left(-\frac{1}{2} \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \boldsymbol{\epsilon}_{ij}\right)}{\sqrt{\det(2\pi \mathbf{R}_{ij})}}$$

and

$$p_2(\boldsymbol{\eta}_i | \boldsymbol{\theta}) = \frac{\exp\left(-\frac{1}{2} \boldsymbol{\eta}_i^T \Omega^{-1} \boldsymbol{\eta}_i\right)}{\sqrt{\det(2\pi \Omega)}}.$$

The marginalization with respect to $\boldsymbol{\eta}_i$ in Eq 1 does not have a closed form solution. By writing Eq 1 on the form

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N \int \exp(l_i) d\boldsymbol{\eta}_i,$$

where the individual joint log-likelihoods are

$$\begin{aligned} l_i &= -\frac{1}{2} \sum_{j=1}^{n_i} \left(\boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \boldsymbol{\epsilon}_{ij} + \log \det(2\pi \mathbf{R}_{ij}) \right) \\ &\quad - \frac{1}{2} \boldsymbol{\eta}_i^T \Omega^{-1} \boldsymbol{\eta}_i - \frac{1}{2} \log \det(2\pi \Omega), \end{aligned}$$

a closed form solution can be obtained by approximating the function l_i with a second order Taylor expansion with respect to $\boldsymbol{\eta}_i$. This is the well-known Laplacian approximation. Furthermore, we let the point around which the Taylor expansion is done to be conditioned on the $\boldsymbol{\eta}_i$ maximizing l_i , here denoted by $\boldsymbol{\eta}_i^*$, and we approximate the Hessian used for the expansion with first order terms only. Thus, the approximate population likelihood L_a becomes

$$L(\boldsymbol{\theta}) \approx L_a(\boldsymbol{\theta}) = \prod_{i=1}^N \left(\exp(l_i(\boldsymbol{\eta}_i^*)) \det \left[\frac{-\Delta l_i(\boldsymbol{\eta}_i^*)}{2\pi} \right]^{-\frac{1}{2}} \right).$$

where

$$\Delta l_i(\boldsymbol{\eta}_i^*) \approx -\sum_{j=1}^{n_i} \nabla \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \nabla \boldsymbol{\epsilon}_{ij} - \Omega^{-1},$$

and

$$\nabla \boldsymbol{\epsilon}_{ij} = \frac{\partial \boldsymbol{\epsilon}_{ij}}{\partial \boldsymbol{\eta}_i} \Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*}.$$

This variant of the Laplacian approximation of the population likelihood is known as the first order conditional estimation (FOCE) method [65].

The maximum likelihood estimate of θ is obtained by maximizing the approximate population likelihood $L_a(\theta)$. The parameters being estimated are all parameter included in θ , namely the fixed effect parameters of the dynamical model, including the fixed effect parameters of the observational model, and any parameters appearing in the random effect covariance matrix Ω . The optimization problem resulting from the desire to maximize L_a with respect to θ was solved using the BFGS method [66]. Note that every evaluation of L_a requires the determination of η_i^* for all individuals due to the conditional nature of the FOCE approximation. Thus, the optimization of L_a with respect to θ involves a nested optimization of l_i with respect to η_i for every individual, making the parameter estimation a challenging problem. An exhaustive account of how the gradient-based optimization was performed for the FOCE approximation of the population likelihood can be found in [52].

Since the approximate population likelihood involves a marginalization over the random effect parameters η_i , these are not explicitly estimated. However, once the estimate of θ has been obtained, the maximum a posteriori estimates of the random effect parameters for each individual cell (referred to as empirical Bayes estimates in the [results](#) section) can be determined. These are in fact equivalent to η_i^* , meaning that they are already provided as an indirect effect of the final evaluation of L_a .

Parameterization of the random effect covariance matrix

The elements of the random effect covariance matrix Ω cannot be chosen independently from one another. To ensure that Ω will be positive semi-definite and symmetric, and thus a covariance matrix, it is decomposed into $\Omega = \mathbf{U} \mathbf{U}^T$, where \mathbf{U} is an upper triangular matrix which can be parameterized according to

$$\mathbf{U} = \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ & \omega_{22} & \omega_{23} \\ & & \omega_{33} \end{pmatrix}.$$

Such decomposition is only unique if Ω is strictly positive definite and if the diagonal elements of \mathbf{U} are positive. The sought-after covariance matrix can for practical purposes always be considered positive-definite, and since we are not interested in \mathbf{U} as such we do not care about the signs of its diagonal entries. With the parameterization above, Ω becomes

$$\Omega = \begin{pmatrix} \omega_{11}^2 + \omega_{12}^2 + \omega_{13}^2 & \omega_{12}\omega_{22} + \omega_{13}\omega_{23} & \omega_{13}\omega_{33} \\ \omega_{12}\omega_{22} + \omega_{13}\omega_{23} & \omega_{22}^2 + \omega_{23}^2 & \omega_{23}\omega_{33} \\ \omega_{13}\omega_{33} & \omega_{23}\omega_{33} & \omega_{33}^2 \end{pmatrix}.$$

Uncertainty of parameter estimates

The uncertainty of parameter estimates are reported as relative standard errors. The relative standard error is computed by taking the absolute value of the ratio between the standard error of the parameter estimate to the estimated value, expressed in percentage. Parameter standard errors are obtained by taking the square root of the diagonal elements of the inverse of the negative Hessian, calculated at the points of the estimated parameter values. Since the uncertainties

in the entries of Ω (and the corresponding correlation matrix) depend on various combination of parameter uncertainties, they were determined by computing the RSE from a large number of sampled covariance and correlation matrices.

We note that measures of confidence based on the exact likelihood, such as likelihood profiling, typically are superior to the results from asymptotic theory. Although preferable, such methods are too time-consuming for the large NLME problems considered here.

Starting values for the optimization algorithm

To reduce the time of computing the parameter estimates and to increase the chances of reaching a meaningful, and hopefully global, optimum, it is important to provide the optimization algorithm with starting values of the parameters that are as good as possible. By visually inspecting the data we were able to obtain educated guesses of some of the parameters. The same values were used to initiate the parameter estimation for all four data sets. The starting value of M_s was set to 3300. Noting that the parameter k_4 determines the relaxation time-scale of X and that the observed re-entry took place at roughly 200 s, we set $k_4 = 1/200 = 0.005$. If X is considered constant, the initial relaxation time-scale of Mig1 is given by k_2 . If we by a very crude visual assessment determine this time-scale to 50 s, we consequently set $k_2 = 1/50 = 0.02$. The measurement noise appear to be on the scale of a hundred to a few hundreds and its variance, s , was set to 40 000. Choosing staring values for the parameters of the random effect covariance matrix is more difficult. We have chosen $\omega_{11} = \omega_{22} = \omega_{33} = 0.1$ and $\omega_{12} = \omega_{13} = \omega_{23} = 0$. This roughly corresponds to parameter standard deviations of $\pm 10\%$ with no covariance between random effect parameters. When all experimental data was used simultaneously for estimation, the experiment-specific parameters inherited the starting values defined above.

Avoiding a constrained problem

The parameters M_s , k_2 , k_4 , and s are only meaningful for nonnegative values. To avoid a constrained optimization problem, any strictly positive parameter, θ , is transformed according to $\theta = e^{\tilde{\theta}}$, with the new starting value $\tilde{\theta}_s = \log \theta_s$. When the parameter estimates have been determined, the values of the transformed parameters must then be transformed back. However, when the parameter uncertainties are determined through the calculation of the Hessian, no parameter transformations are performed. In this case it is not needed since we only evaluate the likelihood function and its gradient for values of the parameters that are known to be positive. As a result, the Hessian and the coefficients of variations derived from it are valid for the original, untransformed parameterization of the model.

Simple analysis of Mig1 dynamics

A simple analysis was designed to extract the typical values (medians) of the response time, amplitude, and duration of Mig1 dynamics directly from data, without the use of a dynamical model. The amplitude for each cell was defined by maximal difference between the baseline and the subsequent data points, and the response time was defined as the time for the maximizing data point. The duration was defined as the difference in time between the first two data points to in each direction cross the level determined from 50% of the amplitude. The simple analysis was also applied to smoothed and densely resampled data. Smoothed data was generated for each cell by fitting a cubic B-spline to its experimental data, and from this smooth function sampling 1000 data points equidistantly in time.

Supporting Information

S1 Fig. Plots of all individual cell data and model simulations for experiment 1.
(PDF)

S2 Fig. Plots of all individual cell data and model simulations for experiment 2.
(PDF)

S3 Fig. Plots of all individual cell data and model simulations for experiment 3.
(PDF)

S4 Fig. Plots of all individual cell data and model simulations for experiment 4.
(PDF)

S5 Fig. Plots of all individual cell data and model simulations for experiment 1 when accounting for background fluorescence.
(PDF)

S6 Fig. Plots of all individual cell data and model simulations for experiment 2 when accounting for background fluorescence.
(PDF)

S7 Fig. Plots of all individual cell data and model simulations for experiment 3 when accounting for background fluorescence.
(PDF)

S8 Fig. Plots of all individual cell data and model simulations for experiment 4 when accounting for background fluorescence.
(PDF)

S9 Fig. Plots of all individual cell data and model simulations for experiment 1 when using all data sets simultaneously.
(PDF)

S10 Fig. Plots of all individual cell data and model simulations for experiment 2 when using all data sets simultaneously.
(PDF)

S11 Fig. Plots of all individual cell data and model simulations for experiment 3 when using all data sets simultaneously.
(PDF)

S12 Fig. Plots of all individual cell data and model simulations for experiment 4 when using all data sets simultaneously.
(PDF)

S13 Fig. Simple analysis histograms.
(PDF)

S1 Table. Shrinkage.
(PDF)

S2 Table. Parameter estimates when accounting for background fluorescence.
(PDF)

S3 Table. Estimates of random effect covariance and correlation matrices when accounting for background fluorescence.

(PDF)

S4 Table. Parameter estimates when using all data sets simultaneously.

(PDF)

S5 Table. Estimates of random effect covariance and correlation matrices when using all data sets simultaneously.

(PDF)

S1 Text. Parameter estimation when accounting for background fluorescence.

(PDF)

S2 Text. Parameter estimation when using all data sets simultaneously.

(PDF)

Acknowledgments

We thank Jacob Leander (FCC) for assisting in the setup of the parameter estimation using all experimental data simultaneously.

Author Contributions

Wrote the paper: JA LB. Conceived and performed the computational analysis: JA. Performed the interpretation of the results: JA LB MJ SH CBA MG. Edited the manuscript: JA LB MJ SH CBA MG.

References

1. Cohen-Saidon C, Cohen AA, Sigal A, Liron Y, Alon U. Dynamics and variability of ERK2 response to EGF in individual living cells. *Mol Cell*. 2009 Dec; 36(5):885–893. doi: [10.1016/j.molcel.2009.11.025](https://doi.org/10.1016/j.molcel.2009.11.025) PMID: [20005850](#)
2. Cheong R, Paliwal S, Levchenko A. Models at the single cell level. *Wiley Interdiscip Rev Syst Biol Med*. 2010; 2(1):34–48. doi: [10.1002/wsbm.49](https://doi.org/10.1002/wsbm.49) PMID: [20836009](#)
3. Love KR, Panagiotou V, Jiang B, Stadheim TA, Love JC. Integrated single-cell analysis shows *Pichia pastoris* secretes protein stochastically. *Biotechnol Bioeng*. 2010 Jun; 106(2):319–325. doi: [10.1002/bit.22688](https://doi.org/10.1002/bit.22688) PMID: [20148400](#)
4. Ramsey SA, Smith JJ, Orrell D, Marelli M, Petersen TW, de Atauri P, et al. Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. *Nat Genet*. 2006 Sep; 38(9):1082–1087. doi: [10.1038/ng1869](https://doi.org/10.1038/ng1869) PMID: [16936734](#)
5. Feinerman O, Veiga J, Dorfman JR, Germain RN, Altan-Bonnet G. Variability and robustness in T cell activation from regulated heterogeneity in protein levels. *Science*. 2008 Aug; 321(5892):1081–1084. doi: [10.1126/science.1158013](https://doi.org/10.1126/science.1158013) PMID: [18719282](#)
6. Losev E, Reinke CA, Jellen J, Strongin DE, Bevis BJ, Glick BS. Golgi maturation visualized in living yeast. *Nature*. 2006 Jun; 441(7096):1002–1006. doi: [10.1038/nature04717](https://doi.org/10.1038/nature04717) PMID: [16699524](#)
7. Babazadeh R, Adiels CB, Smedh M, Petelenz-Kurdziel E, Goksör M, Hohmann S. Osmostress-induced cell volume loss delays yeast Hog1 signaling by limiting diffusion processes and by Hog1-specific effects. *PLoS One*. 2013; 8(11):e80901. doi: [10.1371/journal.pone.0080901](https://doi.org/10.1371/journal.pone.0080901) PMID: [24278344](#)
8. Cai L, Dalal CK, Elowitz MB. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature*. 2008 Sep; 455(7212):485–490. doi: [10.1038/nature07292](https://doi.org/10.1038/nature07292) PMID: [18818649](#)
9. Bodvard K, Wrangborg D, Tapani S, Logg K, Sliwa P, Blomberg A, et al. Continuous light exposure causes cumulative stress that affects the localization oscillation dynamics of the transcription factor Msn2p. *Biochim Biophys Acta*. 2011 Feb; 1813(2):358–366. doi: [10.1016/j.bbamcr.2010.12.004](https://doi.org/10.1016/j.bbamcr.2010.12.004) PMID: [21167216](#)
10. Almquist J, Sunnåker M, Hagmar J, Kvarnström M, Jirstrand M. System identification from spatiotemporal cell population data. In: The 9th Int. Conf. on Systems Biology; 2008.

11. Niepel M, Spencer SL, Sorger PK. Non-genetic cell-to-cell variability and the consequences for pharmacology. *Curr Opin Chem Biol.* 2009 Dec; 13(5–6):556–561. doi: [10.1016/j.cbpa.2009.09.015](https://doi.org/10.1016/j.cbpa.2009.09.015) PMID: [19833543](#)
12. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science.* 2002 Aug; 297(5584):1183–1186. doi: [10.1126/science.1070919](https://doi.org/10.1126/science.1070919) PMID: [12183631](#)
13. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A.* 2002 Oct; 99(20):12795–12800. doi: [10.1073/pnas.162041399](https://doi.org/10.1073/pnas.162041399) PMID: [12237400](#)
14. Kollmann M, Løvdok L, Bartholomé K, Timmer J, Sourjik V. Design principles of a bacterial signalling network. *Nature.* 2005 Nov; 438(7067):504–507. doi: [10.1038/nature04228](https://doi.org/10.1038/nature04228) PMID: [16306993](#)
15. van Kampen NG. Stochastic processes in physics and chemistry. North Holland; 2007.
16. Wilkinson DJ. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat Rev Genet.* 2009 Feb; 10(2):122–133. doi: [10.1038/nrg2509](https://doi.org/10.1038/nrg2509) PMID: [19139763](#)
17. Koepl H, Zechner C, Ganguly A, Pelet S, Peter M. Accounting for extrinsic variability in the estimation of stochastic rate constants. *Int J of Robust Nonlinear Control.* 2012; 22:1103–1119. doi: [10.1002/rnc.2804](https://doi.org/10.1002/rnc.2804)
18. Gaudet S, Spencer SL, Chen WW, Sorger PK. Exploring the contextual sensitivity of factors that determine cell-to-cell variability in receptor-mediated apoptosis. *PLoS Comput Biol.* 2012; 8(4):e1002482. doi: [10.1371/journal.pcbi.1002482](https://doi.org/10.1371/journal.pcbi.1002482) PMID: [22570596](#)
19. Zechner C, Ruess J, Krenn P, Pelet S, Peter M, Lygeros J, et al. Moment-based inference predicts bimodality in transient gene expression. *Proc Natl Acad Sci U S A.* 2012 May; 109(21):8340–8345. doi: [10.1073/pnas.1200161109](https://doi.org/10.1073/pnas.1200161109) PMID: [22566653](#)
20. Jeschke M, Baumgärtner S, Legewie S. Determinants of cell-to-cell variability in protein kinase signaling. *PLoS Comput Biol.* 2013 Dec; 9(12):e1003357. doi: [10.1371/journal.pcbi.1003357](https://doi.org/10.1371/journal.pcbi.1003357) PMID: [24339758](#)
21. Meyer R, D'Alessandro LA, Kar S, Kramer B, She B, Kaschek D, et al. Heterogeneous kinetics of AKT signaling in individual cells are accounted for by variable protein concentration. *Front Physiol.* 2012; 3:451. doi: [10.3389/fphys.2012.00451](https://doi.org/10.3389/fphys.2012.00451) PMID: [23226133](#)
22. Schliemann M, Bullinger E, Borchers S, Allgöwer F, Findeisen R, Scheurich P. Heterogeneity reduces sensitivity of cell death for TNF-stimuli. *BMC Syst Biol.* 2011; 5:204. doi: [10.1186/1752-0509-5-204](https://doi.org/10.1186/1752-0509-5-204) PMID: [22204418](#)
23. Kallenberger SM, Beaudouin J, Claus J, Fischer C, Sorger PK, Legewie S, et al. Intra- and inter-dimeric caspase-8 self-cleavage controls strength and timing of CD95-induced apoptosis. *Sci Signal.* 2014 Mar; 7(316):ra23. doi: [10.1126/scisignal.2004738](https://doi.org/10.1126/scisignal.2004738) PMID: [24619646](#)
24. Hasenauer J, Waldherr S, Doszczak M, Radde N, Scheurich P, Allgöwer F. Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics.* 2011; 12:125. doi: [10.1186/1471-2105-12-125](https://doi.org/10.1186/1471-2105-12-125) PMID: [21527025](#)
25. Hasenauer J, Waldherr S, Doszczak NRM, Scheurich P, Allgöwer F. A maximum likelihood estimator for parameter distributions in heterogeneous cell populations. *Procedia Computer Science.* 2010; 1:1649–1657. doi: [10.1016/j.procs.2010.04.185](https://doi.org/10.1016/j.procs.2010.04.185)
26. Woodcock DJ, Vance KW, Komorowski M, Koentges G, Finkenstädt B, Rand DA. A hierarchical model of transcriptional dynamics allows robust estimation of transcription rates in populations of single cells with variable gene copy number. *Bioinformatics.* 2013 Jun; 29(12):1519–1525. doi: [10.1093/bioinformatics/btt201](https://doi.org/10.1093/bioinformatics/btt201) PMID: [23677939](#)
27. Zechner C, Unger M, Pelet S, Peter M, Koepl H. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat Methods.* 2014 Feb; 11(2):197–202. doi: [10.1038/nmeth.2794](https://doi.org/10.1038/nmeth.2794) PMID: [24412977](#)
28. Davidian M, Giltinan DM. Nonlinear models for repeated measurement data: An overview and update. *J Agric Biol Environ Stat.* 2003; 8:387–419. doi: [10.1198/1085711032697](https://doi.org/10.1198/1085711032697)
29. Gancedo JM. Yeast carbon catabolite repression. *Microbiol Mol Biol Rev.* 1998 Jun; 62(2):334–361. PMID: [9618445](#)
30. Carlson M, Osmond BC, Botstein D. Mutants of yeast defective in sucrose utilization. *Genetics.* 1981 May; 98(1):25–40. PMID: [7040163](#)
31. McCartney RR, Schmidt MC. Regulation of Snf1 kinase. Activation requires phosphorylation of threonine 210 by an upstream kinase as well as a distinct step mediated by the Snf4 subunit. *J Biol Chem.* 2001 Sep; 276(39):36460–36466. doi: [10.1074/jbc.M104418200](https://doi.org/10.1074/jbc.M104418200) PMID: [11486005](#)
32. Hong SP, Leiper FC, Woods A, Carling D, Carlson M. Activation of yeast Snf1 and mammalian AMP-activated protein kinase by upstream kinases. *Proc Natl Acad Sci U S A.* 2003 Jul; 100(15):8839–8843. doi: [10.1073/pnas.1533136100](https://doi.org/10.1073/pnas.1533136100) PMID: [12847291](#)

33. Nath N, McCartney RR, Schmidt MC. Yeast Pak1 kinase associates with and activates Snf1. *Mol Cell Biol*. 2003 Jun; 23(11):3909–3917. doi: [10.1128/MCB.23.11.3909-3917.2003](https://doi.org/10.1128/MCB.23.11.3909-3917.2003) PMID: [12748292](#)
34. Sutherland CM, Hawley SA, McCartney RR, Leech A, Stark MJR, Schmidt MC, et al. Elm1p is one of three upstream kinases for the *Saccharomyces cerevisiae* SNF1 complex. *Curr Biol*. 2003 Aug; 13(15):1299–1305. doi: [10.1016/S0960-9822\(03\)00459-7](https://doi.org/10.1016/S0960-9822(03)00459-7) PMID: [12906789](#)
35. Sanz P, Alms GR, Haystead TA, Carlson M. Regulatory interactions between the Reg1-Glc7 protein phosphatase and the Snf1 protein kinase. *Mol Cell Biol*. 2000 Feb; 20(4):1321–1328. doi: [10.1128/MCB.20.4.1321-1328.2000](https://doi.org/10.1128/MCB.20.4.1321-1328.2000) PMID: [10648618](#)
36. DeVit MJ, Johnston M. The nuclear exportin Msn5 is required for nuclear export of the Mig1 glucose repressor of *Saccharomyces cerevisiae*. *Curr Biol*. 1999 Nov; 9(21):1231–1241. doi: [10.1016/S0960-9822\(99\)80503-X](https://doi.org/10.1016/S0960-9822(99)80503-X) PMID: [10556086](#)
37. Papamichos-Chronakis M, Gligoris T, Tzamarias D. The Snf1 kinase controls glucose repression in yeast by modulating interactions between the Mig1 repressor and the Cyc8-Tup1 co-repressor. *EMBO Rep*. 2004 Apr; 5(4):368–372. doi: [10.1038/sj.embo.7400120](https://doi.org/10.1038/sj.embo.7400120) PMID: [15031717](#)
38. De Vit MJ, Waddle JA, Johnston M. Regulated nuclear translocation of the Mig1 glucose repressor. *Mol Biol Cell*. 1997 Aug; 8(8):1603–1618. doi: [10.1091/mbc.8.8.1603](https://doi.org/10.1091/mbc.8.8.1603) PMID: [9285828](#)
39. Bendrioua L, Smedh M, Almquist J, Cvijovic M, Jirstrand M, Goksör M, et al. Yeast AMP-activated protein kinase monitors glucose concentration changes and absolute glucose levels. *J Biol Chem*. 2014 May; 289(18):12863–12875. doi: [10.1074/jbc.M114.547976](https://doi.org/10.1074/jbc.M114.547976) PMID: [24627493](#)
40. García-Salcedo R, Lubitz T, Beltran G, Elbing K, Tian Y, Frey S, et al. Glucose de-repression by yeast AMP-activated protein kinase SNF1 is controlled via at least two independent steps. *FEBS J*. 2014 Apr; 281(7):1901–1917. doi: [10.1111/febs.12753](https://doi.org/10.1111/febs.12753) PMID: [24529170](#)
41. Frey S, Sott K, Smedh M, Millat T, Dahl P, Wolkenhauer O, et al. A mathematical analysis of nuclear intensity dynamics for Mig1-GFP under consideration of bleaching effects and background noise in *Saccharomyces cerevisiae*. *Mol Biosyst*. 2011 Jan; 7(1):215–223. doi: [10.1039/C005305H](https://doi.org/10.1039/C005305H) PMID: [20967382](#)
42. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoeponential model: routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm*. 1983 Jun; 11(3):303–319. doi: [10.1007/BF01061870](https://doi.org/10.1007/BF01061870) PMID: [6644555](#)
43. Schaber J, Liebermeister W, Klipp E. Nested uncertainties in biochemical models. *IET Syst Biol*. 2009 Jan; 3(1):1–9. doi: [10.1049/iet-syb:20070042](https://doi.org/10.1049/iet-syb:20070042) PMID: [19154080](#)
44. Kaltenbach HM, Dimopoulos S, Stelling J. Systems analysis of cellular networks under uncertainty. *FEBS Lett*. 2009 Dec; 583(24):3923–3930. doi: [10.1016/j.febslet.2009.10.074](https://doi.org/10.1016/j.febslet.2009.10.074) PMID: [19879267](#)
45. Schaber J, Klipp E. Model-based inference of biochemical parameters and dynamic properties of microbial signal transduction networks. *Curr Opin Biotechnol*. 2011 Feb; 22(1):109–116. doi: [10.1016/j.copbio.2010.09.014](https://doi.org/10.1016/j.copbio.2010.09.014) PMID: [20970318](#)
46. Chen WW, Niepel M, Sorger PK. Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev*. 2010 Sep; 24(17):1861–1875. doi: [10.1101/gad.1945410](https://doi.org/10.1101/gad.1945410) PMID: [20810646](#)
47. Tyson JJ, Chen KC, Novak B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol*. 2003 Apr; 15(2):221–231. doi: [10.1016/S0955-0674\(03\)00017-6](https://doi.org/10.1016/S0955-0674(03)00017-6) PMID: [12648679](#)
48. Karlsson J, Anguelova M, Jirstrand M. An efficient method for structural identifiability analysis of large dynamic systems. In: 16th IFAC Symposium on System Identification. vol. 16; 2012. p. 941–946.
49. Rae A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*. 2014 May; 30(10):1440–1448. doi: [10.1093/bioinformatics/btu006](https://doi.org/10.1093/bioinformatics/btu006) PMID: [24463185](#)
50. Limpert E, Stahel WA, Abbt M. Log-normal Distributions across the Sciences: Keys and Clues. *BioScience*. 2001; 51(5):341–352. doi: [10.1641/0006-3568\(2001\)051%5B0341:LNDATS%5D2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051%5B0341:LNDATS%5D2.0.CO;2)
51. Limpert E, Stahel WA. Problems with using the normal distribution—and ways to improve quality and efficiency of data analysis. *PLoS One*. 2011; 6(7):e21403. doi: [10.1371/journal.pone.0021403](https://doi.org/10.1371/journal.pone.0021403) PMID: [21779325](#)
52. Almquist J, Leander J, Jirstrand M. Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood. *J Pharmacokinet Pharmacodyn*. 2015; In press. doi: [10.1007/s10928-015-9409-1](https://doi.org/10.1007/s10928-015-9409-1) PMID: [25801663](#)
53. Savic RM, Karlsson MO. Importance of shrinkage in empirical bayes estimates for diagnostics: problems and solutions. *AAPS J*. 2009 Sep; 11(3):558–569. doi: [10.1208/s12248-009-9133-0](https://doi.org/10.1208/s12248-009-9133-0) PMID: [19649712](#)

54. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*. 2009 Aug; 25(15):1923–1929. doi: [10.1093/bioinformatics/btp358](https://doi.org/10.1093/bioinformatics/btp358) PMID: [19505944](#)
55. Muzzey D, Gómez-Uribe CA, Mettetal JT, van Oudenaarden A. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*. 2009 Jul; 138(1):160–171. doi: [10.1016/j.cell.2009.04.047](https://doi.org/10.1016/j.cell.2009.04.047) PMID: [19596242](#)
56. Kalita MK, Sargsyan K, Tian B, Paulucci-Holthauzen A, Najm HN, Debusschere BJ, et al. Sources of cell-to-cell variability in canonical nuclear factor- κ B (NF- κ B) signaling pathway inferred from single cell dynamic images. *J Biol Chem*. 2011 Oct; 286(43):37741–37757. doi: [10.1074/jbc.M111.280925](https://doi.org/10.1074/jbc.M111.280925) PMID: [21868381](#)
57. Janzén D. Standard two-stage and Nonlinear mixed effect modelling for determination of cell-to-cell variation of transport parameters in *Saccharomyces cerevisiae*. Linköpings universitet; 2012.
58. Gonzalez AM, Uhendorf J, Schaul J, Cinquemani E, Batta G, Ferrari-Trecate G. Identification of biological models from single-cell data: a comparison between mixed-effects and moment-based inference. In: European Control Conference 2013; 2013.
59. Klim S, Mortensen SB, Kristensen NR, Overgaard RV, Madsen H. Population stochastic modelling (PSM)—an R package for mixed-effects models based on stochastic differential equations. *Comput Methods Programs Biomed*. 2009 Jun; 94(3):279–289. doi: [10.1016/j.cmpb.2009.02.001](https://doi.org/10.1016/j.cmpb.2009.02.001) PMID: [19268387](#)
60. Berglund M, Sunnåker M, Adiels M, Jirstrand M, Wennberg B. Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations. *Math Med Biol*. 2012 Dec; 29(4):361–384. doi: [10.1093/imammb/dqr021](https://doi.org/10.1093/imammb/dqr021) PMID: [21965323](#)
61. Leander J, Almquist J, Ahlström C, Gabrielsson J, Jirstrand M. Mixed effects modeling using stochastic differential equations: illustrated by pharmacokinetic data of nicotinic acid in obese Zucker rats. *AAPS J*. 2015; In press. doi: [10.1208/s12248-015-9718-8](https://doi.org/10.1208/s12248-015-9718-8) PMID: [25693487](#)
62. Kristensen NR, Madsen H, Ingwersen SH. Using stochastic differential equations for PK/PD model development. *J Pharmacokinet Pharmacodyn*. 2005 Feb; 32(1):109–141. doi: [10.1007/s10928-005-2105-9](https://doi.org/10.1007/s10928-005-2105-9) PMID: [16215845](#)
63. Mettetal JT, Muzzey D, Gómez-Uribe C, van Oudenaarden A. The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*. *Science*. 2008 Jan; 319(5862):482–484. doi: [10.1126/science.1151582](https://doi.org/10.1126/science.1151582) PMID: [18218902](#)
64. Almquist J, Cvijovic M, Hatzimanikatis V, Nielsen J, Jirstrand M. Kinetic models in industrial biotechnology—Improving cell factory performance. *Metab Eng*. 2014 Jul; 24:38–60. doi: [10.1016/j.ymben.2014.03.007](https://doi.org/10.1016/j.ymben.2014.03.007) PMID: [24747045](#)
65. Wang Y. Derivation of various NONMEM estimation methods. *J Pharmacokinet Pharmacodyn*. 2007 Oct; 34(5):575–593. doi: [10.1007/s10928-007-9060-6](https://doi.org/10.1007/s10928-007-9060-6) PMID: [17620001](#)
66. Nocedal J, Wright SJ. Numerical optimization. Springer-Verlag; 1999.

PAPER E

Using Sensitivity Equations for Computing Gradients
of the FOCE and FOCEI Approximations to the
Population Likelihood

Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood

Joachim Almquist^{1,2} · Jacob Leander^{1,3} · Mats Jirstrand¹

Received: 11 September 2014/Accepted: 23 February 2015/Published online: 24 March 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract The first order conditional estimation (FOCE) method is still one of the parameter estimation workhorses for nonlinear mixed effects (NLME) modeling used in population pharmacokinetics and pharmacodynamics. However, because this method involves two nested levels of optimizations, with respect to the empirical Bayes estimates and the population parameters, FOCE may be numerically unstable and have long run times, issues which are most apparent for models requiring numerical integration of differential equations. We propose an alternative implementation of the FOCE method, and the related FOCEI, for parameter estimation in NLME models. Instead of obtaining the gradients needed for the two levels of quasi-Newton optimizations from the standard finite difference approximation, gradients are computed using so called sensitivity equations. The advantages of this approach were demonstrated using different versions of a pharmacokinetic model defined by nonlinear differential equations. We show that both the accuracy and precision of gradients can be improved extensively, which will increase the chances of a successfully converging parameter estimation. We also show that the proposed approach can lead to markedly reduced computational times. The

accumulated effect of the novel gradient computations ranged from a 10-fold decrease in run times for the least complex model when comparing to forward finite differences, to a substantial 100-fold decrease for the most complex model when comparing to central finite differences. Considering the use of finite differences in for instance NONMEM and Phoenix NLME, our results suggests that significant improvements in the execution of FOCE are possible and that the approach of sensitivity equations should be carefully considered for both levels of optimization.

Keywords Nonlinear mixed effects modeling · First order conditional estimation (FOCE) · Sensitivity equations

Introduction

Nonlinear mixed effects (NLME) models are suitable in situations where sparse time-series data is collected from a population of individuals exhibiting inter-individual variability [10]. This property has rendered NLME models popular in both pharmacokinetics and pharmacodynamics, and several public and commercial software packages have been developed for performing NLME modeling within these fields [13]. These modeling softwares include the well-known NONMEM [5], which was the first program to be developed and still is one of the most widely used, but also a number of other programs such as Phoenix NLME [21] and Monolix [15]. A core part of their functionality consist of various methods for addressing the problem of parameter estimation in NLME models, and several studies have been devoted to describing and comparing different aspects of these methods [4, 8, 9, 11, 12, 22].

✉ Joachim Almquist
joachim.almquist@fcc.chalmers.se

¹ Fraunhofer-Chalmers Centre, Chalmers Science Park,
41288 Gothenburg, Sweden

² Systems and Synthetic Biology, Department of Biology and
Biological Engineering, Chalmers University of Technology,
41296 Gothenburg, Sweden

³ Department of Mathematical Sciences, Chalmers University
of Technology and University of Gothenburg,
41296 Gothenburg, Sweden

The “mixed effects” in NLME refers to the fact that these models contain both fixed effect parameters, having the same value for all individuals, and random effect parameters, whose value differ from one individual to another and whose distribution in the population is determined by some statistical model. A common approach to the parameter estimation problem in NLME models is based on maximizing the so called population likelihood. The population likelihood is a function of the fixed effect parameters only, and it is obtained by marginalizing out the random effects from the joint distribution of data and random effects. However, the integral required for the marginalization lacks a closed-form solution for all realistic problems. Because of this, maximum likelihood parameter estimation for NLME models revolves around different numerical approximation methods for computing this integral. One of the main approaches for tackling the problem is a class of related methods based on the so called Laplacian approximation [25]. It includes the popular and widely used first order conditional estimation (FOCE) method, which is a special case of the closely related FOCE with interaction (FOCEI). With the FOCE and FOCEI methods, the approximation of the integral involves a Taylor expansion around the values of the random effect parameters that maximize the joint distribution. This means that one optimization problem per individual has to be solved for every evaluation of the approximated population likelihood. Since the aim is to maximize the (approximated) population likelihood, which constitutes the original optimization problem, conditional estimation methods such as FOCE produce a parameter estimation problem involving two nested layers of optimizations. For some NLME parameter estimation problems this results in long execution times, and in difficulties with numerical precision making the optimizations unstable and limiting the precision of estimates and the ability of obtaining confidence intervals. These issues are particularly pronounced for models that are formulated by systems of differential equations which are lacking analytical solutions [4, 7, 8].

The optimization problems resulting from the FOCE and FOCEI approximations, and other closely related approximations, are typically solved using gradient-based optimization methods such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton method [20]. For problems where analytical expressions for the function and its gradient are not available, it is common that gradients are computed by finite difference approximations. We instead propose another approach for determining the gradient of the FOCE and FOCEI approximations of the population likelihood. Our approach is based on formally differentiating the likelihoods used at the two levels of optimization, and computing the required derivatives of the model state variables using so called sensitivity equations. The proposed approach for computing gradients is readily applicable for the inner level of

the nested optimization problem. However, we also derive the necessary theory for computing gradients through the approach of sensitivity equations at the outer level optimization. This step is the more challenging, and requires that sensitivities up to second order of the state variables with respect to the parameters and random parameters are obtained. Being able to compute the gradient of the FOCE or FOCEI approximations of the population likelihood using the approach introduced in this paper is a great advantage as it circumvents the need for repeatedly having to solve the inner level optimization problem for obtaining the outer level gradients from a finite difference approximation.

This paper is organized in the following way. First, the mathematical theory is introduced. Here we recapitulate NLME models based on differential equations, including the formulation of the population likelihood and its approximations, as well as derive expressions for both the gradients of the individual joint log-likelihoods with respect to the random effect parameters, used for the inner level optimization problems, and the gradient of the approximate population likelihood with respect to the fixed effect parameters, used for the outer level optimization problem. Then, we apply the sensitivity approach for computing the gradients for different versions of a benchmark model. Compared to the finite difference approximation, the proposed approach leads to both higher precision and better accuracy of the gradient, as well as decreased computational times. Finally, the presented results are discussed and possible future extensions are outlined.

Theory

Various definitions and results from matrix calculus are used in the derivations of this section. These can be found in the “[Appendix 1](#)” section.

The nonlinear mixed effects model

Consider a population of N subjects and let the i th individual be described by the dynamical system

$$\begin{aligned} \frac{d\mathbf{x}_i(t)}{dt} &= \mathbf{f}(\mathbf{x}_i(t), t, \mathbf{Z}_i(t), \boldsymbol{\theta}, \boldsymbol{\eta}_i) \\ \mathbf{x}_i(t_0) &= \mathbf{x}_{0i}(\mathbf{Z}_i(t_0), \boldsymbol{\theta}, \boldsymbol{\eta}_i), \end{aligned} \tag{1}$$

where $\mathbf{x}_i(t)$ is a set of state variables, which for instance could be used to describe a drug concentration in one or more compartments, and where $\mathbf{Z}_i(t)$ is a set of possibly time dependent covariates, $\boldsymbol{\theta}$ a set of fixed effects parameters, and $\boldsymbol{\eta}_i$ a set of random effect parameters which are multivariate normally distributed with zero mean and covariance $\boldsymbol{\Omega}$. The covariance matrix $\boldsymbol{\Omega}$ is in general unknown and will therefore typically contain parameters

subject to estimation. These parameters will for convenience of notation be included in the fixed effect parameter vector $\boldsymbol{\theta}$. Fixed effects parameters will hence be used to refer to all parameters that are not random, not being limited for parameters appearing in the model differential equations. A model for the j th observation of the i th individual at time t_{ji} is defined by

$$\mathbf{y}_{ij} = \mathbf{h}(\mathbf{x}_{ij}, t_{ji}, \mathbf{Z}_i(t_{ji}), \boldsymbol{\theta}, \boldsymbol{\eta}_i) + \mathbf{e}_{ij}, \quad (2)$$

where

$$\mathbf{e}_{ij} \in N\left(\boldsymbol{\theta}, \mathbf{R}_{ij}(\mathbf{x}_{ij}, t_{ji}, \mathbf{Z}_i(t_{ji}), \boldsymbol{\theta}, \boldsymbol{\eta}_i)\right), \quad (3)$$

and where the index notation ij is used as a short form for denoting the i th individual at the j th observation. Note that any fixed effect parameters of the observational model are included in $\boldsymbol{\theta}$. Furthermore, we let the expected value of the discrete-time observation model be denoted by

$$\hat{\mathbf{y}}_{ij} = \mathbf{E}[\mathbf{y}_{ij}]. \quad (4)$$

The population likelihood

Given a set of experimental observations, \mathbf{d}_{ij} , for the individuals $i = 1, \dots, N$ at the time points t_{ji} , where $j = 1, \dots, n_i$, we define the residuals

$$\epsilon_{ij} = \mathbf{d}_{ij} - \hat{\mathbf{y}}_{ij}, \quad (5)$$

and write the population likelihood

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N \int p_1(\mathbf{d}_i | \boldsymbol{\theta}, \boldsymbol{\eta}_i) p_2(\boldsymbol{\eta}_i | \boldsymbol{\theta}) d\boldsymbol{\eta}_i, \quad (6)$$

where

$$p_1(\mathbf{d}_i | \boldsymbol{\theta}, \boldsymbol{\eta}_i) = \prod_{j=1}^{n_i} \frac{\exp\left(-\frac{1}{2} \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \boldsymbol{\epsilon}_{ij}\right)}{\sqrt{\det(2\pi \mathbf{R}_{ij})}} \quad (7)$$

and

$$p_2(\boldsymbol{\eta}_i | \boldsymbol{\theta}) = \frac{\exp\left(-\frac{1}{2} \boldsymbol{\eta}_i^T \boldsymbol{\Omega}^{-1} \boldsymbol{\eta}_i\right)}{\sqrt{\det(2\pi \boldsymbol{\Omega})}}, \quad (8)$$

and where \mathbf{d}_i is used to denote the collection of data from all time points for the i th individual.

The FOCE and FOCEI approximations

The marginalization with respect to $\boldsymbol{\eta}_i$ in Eq. 6 does not have a closed form solution. By writing Eq. 6 on the form

$$L(\boldsymbol{\theta}) = \prod_{i=1}^N \int \exp(l_i) d\boldsymbol{\eta}_i, \quad (9)$$

where the individual joint log-likelihoods are

$$\begin{aligned} l_i &= -\frac{1}{2} \sum_{j=1}^{n_i} \left(\boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \boldsymbol{\epsilon}_{ij} + \log \det(2\pi \mathbf{R}_{ij}) \right) \\ &\quad - \frac{1}{2} \boldsymbol{\eta}_i^T \boldsymbol{\Omega}^{-1} \boldsymbol{\eta}_i - \frac{1}{2} \log \det(2\pi \boldsymbol{\Omega}), \end{aligned} \quad (10)$$

a closed form solution can be obtained by approximating the function l_i with a second order Taylor expansion with respect to $\boldsymbol{\eta}_i$. This is the well-known Laplacian approximation. Furthermore, we let the point around which the Taylor expansion is done to be conditioned on the $\boldsymbol{\eta}_i$ maximizing l_i , here denoted by $\boldsymbol{\eta}_i^*$; I.e., the expansion is done at the mode of the posterior distribution. Thus, the approximate population likelihood, L_L , becomes

$$L(\boldsymbol{\theta}) \approx L_L(\boldsymbol{\theta}) = \prod_{i=1}^N \left(\exp(l_i(\boldsymbol{\eta}_i^*)) \det \left[\frac{-\Delta l_i(\boldsymbol{\eta}_i^*)}{2\pi} \right]^{-\frac{1}{2}} \right). \quad (11)$$

Here, the Hessian $\Delta l_i(\boldsymbol{\eta}_i^*)$ is obtained by first differentiating l_i twice with respect to $\boldsymbol{\eta}_i$, and evaluating at $\boldsymbol{\eta}_i^*$. If we let η_{ik} denote the k th component of $\boldsymbol{\eta}_i$, we have

$$\begin{aligned} \frac{dl_i}{d\eta_{ik}} &= -\frac{1}{2} \sum_{j=1}^{n_i} \left(2\boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\boldsymbol{\epsilon}_{ij}}{d\eta_{ik}} - \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \mathbf{R}_{ij}^{-1} \boldsymbol{\epsilon}_{ij} \right. \\ &\quad \left. + \text{tr} \left[\mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \right] \right) - \boldsymbol{\eta}_i^T \boldsymbol{\Omega}^{-1} \frac{d\boldsymbol{\eta}_i}{d\eta_{ik}}. \end{aligned} \quad (12)$$

Differentiating component-wise again, now with respect to the l th component of $\boldsymbol{\eta}_i$, we get the elements of the Hessian

$$\begin{aligned} \frac{d^2 l_i}{d\eta_{ik} d\eta_{il}} &= -\frac{1}{2} \sum_{j=1}^{n_i} \left(2 \frac{d\boldsymbol{\epsilon}_{ij}^T}{d\eta_{il}} \mathbf{R}_{ij}^{-1} \frac{d\boldsymbol{\epsilon}_{ij}}{d\eta_{ik}} - 2 \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{il}} \mathbf{R}_{ij}^{-1} \frac{d\boldsymbol{\epsilon}_{ij}}{d\eta_{ik}} \right. \\ &\quad \left. + 2 \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d^2 \boldsymbol{\epsilon}_{ij}}{d\eta_{ik} d\eta_{il}} - \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d^2 \mathbf{R}_{ij}}{d\eta_{ik} d\eta_{il}} \mathbf{R}_{ij}^{-1} \boldsymbol{\epsilon}_{ij} \right. \\ &\quad \left. + 2 \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{il}} \mathbf{R}_{ij}^{-1} \boldsymbol{\epsilon}_{ij} \right. \\ &\quad \left. - 2 \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \mathbf{R}_{ij}^{-1} \frac{d\boldsymbol{\epsilon}_{ij}}{d\eta_{il}} - \text{tr} \left[\mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{il}} \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \right] \right. \\ &\quad \left. + \text{tr} \left[\mathbf{R}_{ij}^{-1} \frac{d^2 \mathbf{R}_{ij}}{d\eta_{ik} d\eta_{il}} \right] \right) - \frac{d\boldsymbol{\eta}_i^T}{d\eta_{il}} \boldsymbol{\Omega}^{-1} \frac{d\boldsymbol{\eta}_i}{d\eta_{ik}}, \end{aligned} \quad (13)$$

where the last term is really just the kl th element of $\boldsymbol{\Omega}^{-1}$, Ω_{kl}^{-1} . The expression for the elements of the Hessian may be approximated in different ways, with the main purpose of avoiding the need for computing the costly second order derivatives. We apply a first order approximation, where terms containing second order derivatives are ignored, and write the elements of the approximate Hessian, \mathbf{H}_i , as

$$\mathbf{H}_{ikl} = -\frac{1}{2} \sum_{j=1}^{n_i} \left(\mathbf{a}_l \mathbf{B} \mathbf{a}_k^T + \text{tr}[-\mathbf{c}_l \mathbf{c}_k] \right) - \Omega_{kl}^{-1}, \quad (14)$$

where

$$\mathbf{a}_k = \left(\frac{d\epsilon_{ij}^T}{d\eta_{ik}} - \epsilon_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \right), \quad (15)$$

$$\mathbf{B} = 2\mathbf{R}_{ij}^{-1}, \quad (16)$$

and

$$\mathbf{c}_k = \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{ik}}. \quad (17)$$

This variant of the Laplacian approximation of the population likelihood is known as the first order conditional estimation with interaction (FOCEI) method. The closely related first order conditional estimation (FOCE) method is obtained by ignoring the dependence of the residual covariance matrix on the random effect parameters. The rationale for excluding the second order terms is that their expected values are zero for an appropriate model, as shown in the “Appendix 2” section. The Appendix also shows how the Hessian may be slightly further simplified, using similar arguments, to arrive at the variant of FOCE used in NONMEM. Those additional simplifications are however of relatively little importance from a computational point of view, since the components needed to evaluate these Hessian terms have to be provided for the remaining part of the Hessian anyway. We will therefore restrict the Hessian simplification by expectation to the second order terms only. Furthermore, we will from now on for convenience consider the logarithm of the FOCEI approximation to the population likelihood, L_F ,

$$\log L(\boldsymbol{\theta}) \approx \log L_F(\boldsymbol{\theta}) = \sum_{i=1}^N \left(l_i(\boldsymbol{\eta}_i^*) - \frac{1}{2} \log \det \left[\frac{-\mathbf{H}_i(\boldsymbol{\eta}_i^*)}{2\pi} \right] \right). \quad (18)$$

Gradient of the individual joint log-likelihood with respect to the random effect parameters

We now turn to the computation of the gradient of the individual joint log-likelihoods, $l_i(\boldsymbol{\eta}_i)$, with respect to the random effect parameters, $\boldsymbol{\eta}_i$, using the approach of sensitivity equations. Consider the differentiation done in Eq. 12. Given values of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}_i$, the quantities ϵ_{ij} , \mathbf{R}_{ij} , and $\boldsymbol{\Omega}$ can be obtained by solving the model equations. However, we additionally need to determine $d\epsilon_{ij}/d\eta_{ik}$ and $d\mathbf{R}_{ij}/d\eta_{ik}$. Expanding the total derivative of these quantities we see that

$$\frac{d\epsilon_{ij}}{d\eta_{ik}} = \frac{d(\mathbf{d}_{ij} - \hat{\mathbf{y}}_{ij})}{d\eta_{ik}} = - \left(\frac{\partial \mathbf{h}}{\partial \eta_{ik}} + \frac{\partial \mathbf{h}}{\partial \mathbf{x}_{ij}} \frac{d\mathbf{x}_{ij}}{d\eta_{ik}} \right), \quad (19)$$

and

$$\frac{d\mathbf{R}_{ij}}{d\eta_{ik}} = \frac{\partial \mathbf{R}_{ij}}{\partial \eta_{ik}} + \frac{\partial \mathbf{R}_{ij}}{\partial \mathbf{x}_{ij}} \frac{d\mathbf{x}_{ij}}{d\eta_{ik}}. \quad (20)$$

The derivatives of \mathbf{h} and \mathbf{R}_{ij} are readily obtained since these expressions are given explicitly by the model formulation. In contrast, the derivative of the state variables, \mathbf{x}_{ij} , are not directly available but can be computed from the so called sensitivity equations. The sensitivity equations are a set of differential equations which are derived by differentiating the original system of differential equations (and the corresponding initial conditions) with respect to each random effect parameter η_{ik} ,

$$\begin{aligned} \frac{d}{dt} \left(\frac{d\mathbf{x}_i}{d\eta_{ik}} \right) &= \frac{\partial \mathbf{f}}{\partial \eta_{ik}} + \frac{\partial \mathbf{f}}{\partial \mathbf{x}_i} \left(\frac{d\mathbf{x}_i}{d\eta_{ik}} \right) \\ \left(\frac{d\mathbf{x}_i}{d\eta_{ik}} \right)(t_0) &= \frac{\partial \mathbf{x}_{0i}}{\partial \eta_{ik}}. \end{aligned} \quad (21)$$

The solution to the sensitivity equations can be used to evaluate the derivatives in Eqs. 19 and 20, which in turn are needed for the gradient of the individual joint log-likelihoods. Importantly, these derivatives are also used for computing the approximate Hessian, Eq. 14, appearing in the approximate population log-likelihood.

In the unusual event that one or more of the random effect parameters only appear in the observational model, all sensitivities of the state variables with respect to those parameters are trivially zero. Note also that the sensitivity equations for all but trivial models involve the original state variables, which means that the original system of differential equations has to be solved simultaneously. Thus, if there are q non-trivial sensitivities and n state variables, the total number of differential equations that has to be solved in order to be able to compute l_i and $dl_i/d\eta_i$ for each individual is

$$n(1+q). \quad (22)$$

Gradient of the approximate population log-likelihood with respect to the fixed effect parameters

We now derive the expression for the gradient of the approximate population log-likelihood, $\log L_F(\boldsymbol{\theta})$, with respect to the parameter vector $\boldsymbol{\theta}$. Differentiating $\log L_F$ with respect to the m th element of $\boldsymbol{\theta}$ gives

$$\frac{\log L_F}{d\theta_m} = \sum_{i=1}^N \left(\frac{dl_i(\boldsymbol{\eta}_i^*)}{d\theta_m} - \frac{1}{2} \text{tr} \left[\mathbf{H}_i^{-1}(\boldsymbol{\eta}_i^*) \frac{d\mathbf{H}_i(\boldsymbol{\eta}_i^*)}{d\theta_m} \right] \right). \quad (23)$$

Here it must be emphasized that all derivatives with respect to components of the parameter vector $\boldsymbol{\theta}$ are taken *after* replacing $\boldsymbol{\eta}_i$ with $\boldsymbol{\eta}_i^*$. This is critical since $\boldsymbol{\eta}_i^*$ is an implicit

function of theta, $\eta_i^* = \eta_i^*(\theta)$. In other words, we have to account for the fact that the η_i maximizing the individual joint log-likelihood changes as θ changes.

To determine the total derivatives with respect to components of the parameter vector θ we will be needing the following result. Consider a function v which may depend directly on the parameters θ and η_i , and on the auxiliary function w representing any indirect dependencies of these parameters,

$$v = v(w(\theta, \eta_i), \theta, \eta_i). \quad (24)$$

We furthermore introduce the function z to denote the evaluation of v at $\eta_i = \eta_i^*(\theta)$,

$$z = z(w(\theta, \eta_i^*(\theta)), \theta, \eta_i^*(\theta)) = v|_{\eta_i=\eta_i^*(\theta)}. \quad (25)$$

Separating the complete dependence of z on θ into partial dependencies we get that

$$\begin{aligned} \frac{d}{d\theta} (v|_{\eta_i=\eta_i^*(\theta)}) &= \frac{dz}{d\theta} \\ &= \frac{\partial z}{\partial w} \frac{\partial w}{\partial \theta} + \frac{\partial z}{\partial \theta} + \frac{\partial z}{\partial \eta_i^*} \frac{d\eta_i^*}{d\theta} \\ &= \frac{\partial z}{\partial w} \frac{\partial w}{\partial \theta} + \frac{\partial z}{\partial w} \frac{\partial w}{\partial \eta_i^*} \frac{d\eta_i^*}{d\theta} + \frac{\partial z}{\partial \theta} + \frac{\partial z}{\partial \eta_i^*} \frac{d\eta_i^*}{d\theta} \\ &= \frac{\partial z}{\partial w} \frac{\partial w}{\partial \theta} + \frac{\partial z}{\partial \theta} + \frac{dz}{d\eta_i^*} \frac{d\eta_i^*}{d\theta} \\ &= \frac{\partial}{\partial w} (v|_{\eta_i=\eta_i^*(\theta)}) \frac{\partial w}{\partial \theta} + \frac{\partial}{\partial \theta} (v|_{\eta_i=\eta_i^*(\theta)}) \\ &\quad + \frac{d}{d\eta_i^*} (v|_{\eta_i=\eta_i^*(\theta)}) \frac{d\eta_i^*}{d\theta} \\ &= \left(\frac{\partial v}{\partial w} \frac{\partial w}{\partial \theta} \right) \Big|_{\eta_i=\eta_i^*(\theta)} + \left(\frac{\partial v}{\partial \theta} \right) \Big|_{\eta_i=\eta_i^*(\theta)} \\ &\quad + \left(\frac{dv}{d\eta_i^*} \right) \Big|_{\eta_i=\eta_i^*(\theta)} \frac{d\eta_i^*}{d\theta} \\ &= \frac{dv}{d\theta} \Big|_{\eta_i=\eta_i^*(\theta)} + \frac{dv}{d\eta_i^*} \Big|_{\eta_i=\eta_i^*(\theta)} \frac{d\eta_i^*}{d\theta}. \end{aligned} \quad (26)$$

Thus, the total derivative with respect to θ after insertion of η_i^* is equal to the sum of total derivatives with respect to θ and η_i before insertion of η_i^* , where the second derivative is multiplied with the sensitivity of the random effect optimum with respect to the parameters θ . It is straightforward to see that this result holds also when differentiating functions that only exhibit a subset of the possible direct and indirect dependencies of Eq. 24, for instance functions with just an indirect dependence on the two kind of parameters.

Applying the results from Eq. 26 to the first term within the summation of Eq. 23, we have that

$$\frac{dl_i(\eta_i^*)}{d\theta_m} = \frac{dl_i(\eta_i)}{d\theta_m} \Big|_{\eta_i=\eta_i^*(\theta)} + \frac{dl_i(\eta_i)}{d\eta_i} \Big|_{\eta_i=\eta_i^*(\theta)} \frac{d\eta_i^*}{d\theta_m}. \quad (27)$$

However, since $dl_i/d\eta_i$ evaluated at η_i^* is zero by definition, the second term of the right hand side of Eq. 27 disappears and

$$\begin{aligned} \frac{dl_i(\eta_i^*)}{d\theta_m} &= \frac{dl_i(\eta_i)}{d\theta_m} \Big|_{\eta_i=\eta_i^*(\theta)} \\ &= \left[-\frac{1}{2} \sum_{j=1}^{n_i} \left(2\epsilon_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\epsilon_{ij}}{d\theta_m} - \epsilon_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\theta_m} \mathbf{R}_{ij}^{-1} \epsilon_{ij} \right. \right. \\ &\quad \left. \left. + \text{tr} \left[\mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\theta_m} \right] \right) + \frac{1}{2} \eta_i^T \boldsymbol{\Omega}^{-1} \frac{d\boldsymbol{\Omega}}{d\theta_m} \boldsymbol{\Omega}^{-1} \eta_i \right. \\ &\quad \left. - \frac{1}{2} \text{tr} \left[\boldsymbol{\Omega}^{-1} \frac{d\boldsymbol{\Omega}}{d\theta_m} \right] \right] \Big|_{\eta_i=\eta_i^*(\theta)}. \end{aligned} \quad (28)$$

Using asterisks to denote that η_i has been replaced with η_i^* , we also get the following for the derivative of the second term within the summation of Eq. 23,

$$\begin{aligned} \frac{dH_{ikl}(\eta_i^*)}{d\theta_m} &= -\frac{1}{2} \sum_{j=1}^{n_i} \left(\frac{d\mathbf{a}_l^*}{d\theta_m} \mathbf{B}^* \mathbf{a}_k^{*T} + \mathbf{a}_l^* \frac{d\mathbf{B}^*}{d\theta_m} \mathbf{a}_k^{*T} + \mathbf{a}_l^* \mathbf{B}^* \frac{d\mathbf{a}_k^*}{d\theta_m} \right. \\ &\quad \left. + \text{tr} \left[-\frac{d\mathbf{c}_l^*}{d\theta_m} \mathbf{c}_k^* - \mathbf{c}_l^* \frac{d\mathbf{c}_k^*}{d\theta_m} \right] \right) - \frac{d\boldsymbol{\Omega}_{kl}^{-1}}{d\theta_m}, \end{aligned} \quad (29)$$

where

$$\begin{aligned} \frac{d\mathbf{a}_k^*}{d\theta_m} &= \frac{d}{d\theta_m} \left(\frac{d\epsilon_{ij}^T}{d\eta_{ik}} \right)^* - \frac{\epsilon_{ij}^{*T}}{d\theta_m} \mathbf{R}_{ij}^{*-1} \left(\frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \right)^* \\ &\quad + \epsilon_{ij}^{*T} \mathbf{R}_{ij}^{*-1} \frac{d\mathbf{R}_{ij}^*}{d\theta_m} \mathbf{R}_{ij}^{*-1} \left(\frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \right)^* \\ &\quad - \epsilon_{ij}^{*T} \mathbf{R}_{ij}^{*-1} \frac{d}{d\theta_m} \left(\frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \right)^*, \end{aligned} \quad (30)$$

$$\frac{d\mathbf{B}^*}{d\theta_m} = -2\mathbf{R}_{ij}^{*-1} \frac{d\mathbf{R}_{ij}^*}{d\theta_m} \mathbf{R}_{ij}^{*-1}, \quad (31)$$

and

$$\frac{d\mathbf{c}_k^*}{d\theta_m} = -\mathbf{R}_{ij}^{*-1} \frac{d\mathbf{R}_{ij}^*}{d\theta_m} \mathbf{R}_{ij}^{*-1} \left(\frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \right)^* + \mathbf{R}_{ij}^{*-1} \frac{d}{d\theta_m} \left(\frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \right)^*. \quad (32)$$

We now continue to expand the terms in Eqs. 28–32 containing derivatives with respect to θ_m . The terms $d\boldsymbol{\Omega}/d\theta_m$ and $d\boldsymbol{\Omega}_{kl}^{-1}/d\theta_m$ are obtainable by straightforward differentiation. Noting that the terms ϵ_{ij}^* , $(d\epsilon_{ij}/d\eta_{ik})^*$, \mathbf{R}_{ij}^* , and $(d\mathbf{R}_{ij}/d\eta_{ik})^*$, have indirect and/or direct dependence on θ

and $\boldsymbol{\eta}_i^*$, we apply the results from Eq. 26 and expand the remaining derivatives. First,

$$\frac{d\epsilon_{ij}^*}{d\theta_m} = \frac{d\epsilon_{ij}}{d\theta_m}\Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} + \frac{d\epsilon_{ij}}{d\boldsymbol{\eta}_i}\Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} \frac{d\boldsymbol{\eta}_i^*}{d\theta_m}. \quad (33)$$

Here, $d\epsilon_{ij}/d\boldsymbol{\eta}_i$ was determined previously in Eq. 19, and the derivative in the first term is given by

$$\frac{d\epsilon_{ij}}{d\theta_m} = \frac{d(\mathbf{d}_{ij} - \hat{\mathbf{y}}_{ij})}{d\theta_m} = -\left(\frac{\partial \mathbf{h}}{\partial \theta_m} + \frac{\partial \mathbf{h}}{\partial \mathbf{x}_{ij}} \frac{d\mathbf{x}_{ij}}{d\theta_m}\right). \quad (34)$$

The sensitivity of the random effect optimum with respect to the fixed effect parameters, $d\boldsymbol{\eta}_i^*/d\boldsymbol{\theta}$, must also be determined, which we will return to later. Then,

$$\frac{d\mathbf{R}_{ij}^*}{d\theta_m} = \frac{d\mathbf{R}_{ij}}{d\theta_m}\Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} + \frac{d\mathbf{R}_{ij}}{d\boldsymbol{\eta}_i}\Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} \frac{d\boldsymbol{\eta}_i^*}{d\theta_m}, \quad (35)$$

where $d\mathbf{R}_{ij}/d\boldsymbol{\eta}_i$ was determined in Eq. 20, and

$$\frac{d\mathbf{R}_{ij}}{d\theta_m} = \frac{\partial \mathbf{R}_{ij}}{\partial \theta_m} + \frac{\partial \mathbf{R}_{ij}}{\partial \mathbf{x}_{ij}} \frac{d\mathbf{x}_{ij}}{d\theta_m}. \quad (36)$$

Next,

$$\begin{aligned} & \frac{d}{d\theta_m} \left(\frac{d\epsilon_{ij}}{d\eta_{ik}} \Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} \right) \\ &= \left(\frac{d}{d\theta_m} \left(\frac{d\epsilon_{ij}}{d\eta_{ik}} \right) \right) \Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} + \left(\frac{d}{d\boldsymbol{\eta}} \left(\frac{d\epsilon_{ij}}{d\eta_{ik}} \right) \right) \Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} \frac{d\boldsymbol{\eta}_i^*}{d\theta_m} \\ &= \left(\frac{d}{d\theta_m} \left(\frac{d\epsilon_{ij}}{d\eta_{ik}} \right) \right) \Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} + \sum_l \left(\frac{d}{d\eta_{il}} \left(\frac{d\epsilon_{ij}}{d\eta_{ik}} \right) \right) \Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} \frac{d\eta_{il}^*}{d\theta_m} \\ &= -\left(\frac{\partial^2 \mathbf{h}}{\partial \eta_{ik} \partial \theta_m} + \frac{\partial^2 \mathbf{h}}{\partial \eta_{ik} \partial \mathbf{x}_{ij}} \frac{d\mathbf{x}_{ij}}{d\theta_m} + \left(\frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}_{ij} \partial \theta_m} + \frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}_{ij}^2} \frac{d\mathbf{x}_{ij}}{d\theta_m} \right) \frac{d\mathbf{x}_{ij}}{d\eta_{ik}} \right. \\ &\quad \left. + \frac{\partial \mathbf{h}}{\partial \mathbf{x}_{ij}} \frac{d^2 \mathbf{x}_{ij}}{d\eta_{ik} d\theta_m} \right) \Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} - \sum_l \left(\frac{\partial^2 \mathbf{h}}{\partial \eta_{ik} \partial \eta_{il}} + \frac{\partial^2 \mathbf{h}}{\partial \eta_{ik} \partial \mathbf{x}_{ij}} \frac{d\mathbf{x}_{ij}}{d\eta_{il}} \right. \\ &\quad \left. + \left(\frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}_{ij} \partial \eta_{il}} + \frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}_{ij}^2} \frac{d\mathbf{x}_{ij}}{d\eta_{il}} \right) \frac{d\mathbf{x}_{ij}}{d\eta_{ik}} + \frac{\partial \mathbf{h}}{\partial \mathbf{x}_{ij}} \frac{d^2 \mathbf{x}_{ij}}{d\eta_{ik} d\eta_{il}} \right) \Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} \frac{d\eta_{il}^*}{d\theta_m}, \end{aligned} \quad (37)$$

where we after the third equality have used the results from Eq. 19. The derivative of $(d\mathbf{R}_{ij}/d\eta_{ik})^*$ with respect to θ_m is done in a highly similar way and is left to the reader as an exercise.

In the above expressions, derivatives of \mathbf{h} and \mathbf{R}_{ij} are obtained by direct differentiation. The derivatives of the state variables are determined by the previously derived sensitivity equation in Eq. 21 and by the additional sensitivity equations

$$\begin{aligned} \frac{d}{dt} \left(\frac{d\mathbf{x}_i}{d\theta_m} \right) &= \frac{\partial \mathbf{f}}{\partial \theta_m} + \frac{\partial \mathbf{f}}{\partial \mathbf{x}_i} \left(\frac{d\mathbf{x}_i}{d\theta_m} \right) \\ \left(\frac{d\mathbf{x}_i}{d\theta_m} \right) (t_0) &= \frac{\partial \mathbf{x}_{0i}}{\partial \theta_m}, \end{aligned} \quad (38)$$

$$\begin{aligned} \frac{d}{dt} \left(\frac{d^2 \mathbf{x}_i}{d\eta_{ik} d\theta_m} \right) &= \frac{\partial^2 \mathbf{f}}{\partial \eta_{ik} \partial \theta_m} + \frac{\partial^2 \mathbf{f}}{\partial \eta_{ik} \partial \mathbf{x}_i} \frac{d\mathbf{x}_i}{d\theta_m} \\ &+ \left(\frac{\partial^2 \mathbf{f}}{\partial \mathbf{x}_i \partial \theta_m} + \frac{\partial^2 \mathbf{f}}{\partial^2 \mathbf{x}_i} \frac{d\mathbf{x}_i}{d\theta_m} \right) \left(\frac{d\mathbf{x}_i}{d\eta_{ik}} \right) \\ &+ \frac{\partial \mathbf{f}}{\partial \mathbf{x}_i} \left(\frac{d^2 \mathbf{x}_i}{d\eta_{ik} d\theta_m} \right) \end{aligned} \quad (39)$$

$$\left(\frac{d^2 \mathbf{x}_i}{d\eta_{ik} d\theta_m} \right) (t_0) = \frac{\partial^2 \mathbf{x}_{0i}}{\partial \eta_{ik} \partial \theta_m},$$

and

$$\begin{aligned} \frac{d}{dt} \left(\frac{d^2 \mathbf{x}_i}{d\eta_{ik} d\eta_{il}} \right) &= \frac{\partial^2 \mathbf{f}}{\partial \eta_{ik} \partial \eta_{il}} + \frac{\partial^2 \mathbf{f}}{\partial \eta_{ik} \partial \mathbf{x}_i} \frac{d\mathbf{x}_i}{d\eta_{il}} \\ &+ \left(\frac{\partial^2 \mathbf{f}}{\partial \mathbf{x}_i \partial \eta_{il}} + \frac{\partial^2 \mathbf{f}}{\partial^2 \mathbf{x}_i} \frac{d\mathbf{x}_i}{d\eta_{il}} \right) \left(\frac{d\mathbf{x}_i}{d\eta_{ik}} \right) \\ &+ \frac{\partial \mathbf{f}}{\partial \mathbf{x}_i} \left(\frac{d^2 \mathbf{x}_i}{d\eta_{ik} d\eta_{il}} \right) \\ \left(\frac{d^2 \mathbf{x}_i}{d\eta_{ik} d\eta_{il}} \right) (t_0) &= \frac{\partial^2 \mathbf{x}_{0i}}{\partial \eta_{ik} \partial \eta_{il}}. \end{aligned} \quad (40)$$

As noted previously, all sensitivity equations must be solved simultaneously with the original differential equations for all but trivial models. However, since one or more parameters in the vector $\boldsymbol{\theta}$ may not appear in the differential equation part of the model (such as parameters appearing only in Ω), there may be sensitivities which are trivially zero. If there are p non-trivial sensitivities among the parameters in $\boldsymbol{\theta}$, q non-trivial sensitivities among the parameters in $\boldsymbol{\eta}$, and n state variables, the total number of differential equations that has to be solved in order to be able to compute $\log L_F$ and $d \log L_F / d\boldsymbol{\theta}$ for each individual is

$$n(1+q)(1+p+q/2). \quad (41)$$

Finally, we need to determine $d\boldsymbol{\eta}_i^*/d\boldsymbol{\theta}$. At the optimum of each individual joint log-likelihood we have that

$$\frac{dl_i}{d\boldsymbol{\eta}_i} = \boldsymbol{\theta}, \quad (42)$$

or put differently,

$$\frac{dl_i}{d\boldsymbol{\eta}_i} \Big|_{\boldsymbol{\eta}_i=\boldsymbol{\eta}_i^*(\boldsymbol{\theta})} = \boldsymbol{\theta}. \quad (43)$$

This equality holds for any $\boldsymbol{\theta}$, and thus

$$\frac{d}{d\theta} \left(\frac{dl_i}{d\eta_i} \Big|_{\eta_i=\eta_i^*(\theta)} \right) = \mathbf{0}. \quad (44)$$

Recognizing that $dl_i/d\eta_i$ fulfills the requirements of applying the results from Eq. 26, we can write this as

$$\frac{d}{d\theta} \left(\frac{dl_i}{d\eta_i} \Big|_{\eta_i=\eta_i^*(\theta)} \right) = \frac{d^2 l_i}{d\eta_i d\theta} \Big|_{\eta_i=\eta_i^*(\theta)} + \frac{d^2 l_i}{d\eta_i^2} \Big|_{\eta_i=\eta_i^*(\theta)} \frac{d\eta_i^*}{d\theta} = \mathbf{0}. \quad (45)$$

By rearranging terms and inverting the matrix, we finally get that

$$\frac{d\eta_i^*}{d\theta} = - \left(\frac{d^2 l_i}{d\eta_i^2} \Big|_{\eta_i=\eta_i^*(\theta)} \right)^{-1} \frac{d^2 l_i}{d\eta_i d\theta} \Big|_{\eta_i=\eta_i^*(\theta)}. \quad (46)$$

The second order derivatives of the individual joint log-likelihoods with respect to the random effect parameters were previously derived in Eq. 13. In contrast to the first order approximation of the Hessian used in the approximate population log-likelihood, the second order derivatives of ϵ_{ij} and \mathbf{R}_{ij} are kept. These are obtained by differentiating Eqs. 19 and 20 once more with respect to η_i (not shown). This in turn requires the second order sensitivity equations of the state variables with respect to η_i , which were previously provided in Eq. 40. In addition to second order derivatives of the individual joint log-likelihoods with respect to the random effect parameters, Eq. 46 also requires the second order mixed derivatives, which are given by

$$\begin{aligned} \frac{d^2 l_i}{d\eta_{ik} d\theta_m} = & -\frac{1}{2} \sum_{j=1}^{n_i} \left(2 \frac{d\epsilon_{ij}^T \mathbf{R}_{ij}^{-1}}{d\theta_m} \frac{d\epsilon_{ij}}{d\eta_{ik}} - 2 \epsilon_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\theta_m} \mathbf{R}_{ij}^{-1} \frac{d\epsilon_{ij}}{d\eta_{ik}} \right. \\ & + 2 \epsilon_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d^2 \epsilon_{ij}}{d\eta_{ik} d\theta_m} - \epsilon_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d^2 \mathbf{R}_{ij}}{d\eta_{ik} d\theta_m} \mathbf{R}_{ij}^{-1} \epsilon_{ij} \\ & + 2 \epsilon_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\theta_m} \mathbf{R}_{ij}^{-1} \epsilon_{ij} \\ & - 2 \epsilon_{ij}^T \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{ik}} \mathbf{R}_{ij}^{-1} \frac{d\epsilon_{ij}}{d\theta_m} \\ & \left. + \text{tr} \left[\mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\theta_m} \mathbf{R}_{ij}^{-1} \frac{d\mathbf{R}_{ij}}{d\eta_{ik}} + \mathbf{R}_{ij}^{-1} \frac{d^2 \mathbf{R}_{ij}}{d\eta_{ik} d\theta_m} \right] \right) \\ & - \eta_i^T \boldsymbol{\Omega}^{-1} \frac{d\boldsymbol{\Omega}}{d\theta_m} \boldsymbol{\Omega}^{-1} \frac{d\eta_i}{d\eta_{ik}}. \end{aligned} \quad (47)$$

Here, all terms have previously been introduced except $d^2\epsilon_{ij}/d\eta_{ik}d\theta_m$ and $d^2\mathbf{R}_{ij}/d\eta_{ik}d\theta_m$, which are provided within the derivation of Eq. 37 and through a corresponding derivation involving \mathbf{R}_{ij} .

Better starting values for optimization of random effect parameters

Computing the approximate population log-likelihood and its gradient with respect to the parameters θ requires the determination of η_i^* for every individual. The first time $\log L_F$ and its gradient are evaluated it is reasonable to initiate the inner level optimizations for η_i^* with $\eta_i = \mathbf{0}$. However, in the subsequent steps of the optimization with respect to θ , better starting values for η_i can be provided. One way of choosing the starting values η_i^0 for the optimization of η_i is to set them equal to the optimized value from the last step of the outer optimization. If we for simplicity of notation from now on suppress the index of η_i denoting the individual, i , and instead let the the index s denote the step of the outer optimization with respect to θ , this can be expressed as $\eta_{s+1}^0 = \eta_s^*$. This will be particularly helpful as the optimization converges and the steps in θ become smaller. Using η^* from the evaluation of $\log L_F$ as starting value is also a good strategy when computing the gradient of $\log L_F$ by a finite difference approximation.

If the sensitivity approach is used for computing the gradient of $\log L_F$, even better starting values of η can be provided. This is accomplished by exploiting the fact that the sensitivity $d\eta^*/d\theta$ happens to be part of the gradient calculation. By making a first order Taylor expansion of the implicit function $\eta^*(\theta)$, we propose the following update of the starting values of the random effect parameters

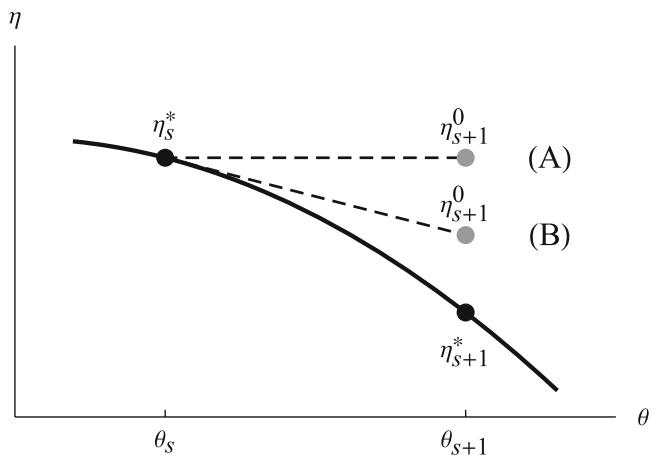


Fig. 1 Starting values for finding optimal random parameter values. The hypothetical relationship between a parameter θ and the optimal value of a random effect parameter η^* is depicted by the *solid curve*, and the optimal values of η for two consecutive θ of the optimization, θ_s and θ_{s+1} , are shown as *black points*. The two approaches for selecting starting values η_{s+1}^0 are shown as *dashed lines* and *gray points*, with the label (A) for using the previous value and (B) for using the gradient based update

$$\eta_{s+1}^0 = \eta_s^* + \frac{d\eta_s^*}{d\theta}(\theta_{s+1} - \theta_s). \quad (48)$$

The two approaches for choosing η_{s+1}^0 are illustrated in Fig. 1.

Results

Based on the theory presented in the previous section, we propose an alternative implementation of the FOCE and FOCEI methods for parameter estimation of NLME models based on differential equations. The steps of this novel approach are outlined in Algorithm 1. The crucial points are the computation of gradients using sensitivity equations, for both the inner and outer problem, and the way that starting values for the inner problem are determined.

covariance matrix was limited to a diagonal matrix. Observations were modeled using a normally distributed additive error. All parameters were estimated in model M2, including the full covariance matrix for the random effect parameters. In model M3, an additional random effect parameter was introduced and the full covariance matrix was extended accordingly. The observational model was also altered to include measurements from both compartments, and the error in the measurements from the first compartments was modeled with both an additive and proportional term. Model M4 is the same as M3 but for this model the parameter estimation was performed with FOCEI instead of FOCE.

Improving gradient precision and accuracy

We compared our proposed method of computing the gradient of the approximate population log-likelihood,

Algorithm 1 Parameter estimation algorithm

```

s := 0,  $\theta_s := \theta_{starting}$                                 ▷ Initialize algorithm
for all individuals do
    u := 0,  $\eta_s^u := 0$ 
end for
repeat
    for all individuals do
        u := 0
        repeat
            Solve for  $\mathbf{x}$  and the sensitivities  $d\mathbf{x}/d\eta$ 
            Compute  $l$  and  $dl/d\eta$ 
            Update  $\eta_s^{u+1}$  according to BFGS
            u := u + 1
        until  $\eta_s^*$  is obtained
    end for
    for all individuals do
        Set  $\eta := \eta_s^*$ 
        Solve for  $\mathbf{x}$  and the sensitivities  $d\mathbf{x}/d\eta$ ,  $d\mathbf{x}/d\theta$ ,  $d^2\mathbf{x}/d\eta^2$ , and  $d^2\mathbf{x}/d\eta d\theta$ 
    end for
    Compute  $\log L_F$  and  $d\log L_F/d\theta$ 
    Update  $\theta_{s+1}$  according to BFGS
    for all individuals do
         $\eta_{s+1}^0 = \eta_s^* + \frac{d\eta_s^*}{d\theta}(\theta_{s+1} - \theta_s)$                                 ▷ Set starting values for inner problem
    end for
    s := s + 1
until convergence of  $\theta$ 

```

The algorithm was evaluated using a two-compartment model with a capacity-limited elimination. This is a moderately complex pharmacokinetic model that requires the numerical solution of differential equations. All details regarding the model, including model equations, parameters used for simulating data, the starting values for the parameter estimation, and the parameter estimates, can be found in the “Appendix 3” section. A short summary of the model is shown in Table 1. Briefly, four versions of the model (M1–M4) were used. In model M1, some parameters were fixed to the true values, hence excluded from the estimation. Three random effect parameters were introduced but their

$\log L_F$, with respect to θ to the more straightforward approach of finite difference approximation. Two versions of the finite difference approximations were considered, a forward difference and a central difference. To investigate the precision and accuracy of these approximations, we first determined the estimate of θ for model M1. We then computed all 6 elements of the gradient at this point in parameter space using different values of the relative step size, 10^{-h} . The details of the comparison are explained in the methods section. In addition, we computed the gradient using the approach based on sensitivity equations. A comparison of the two approaches is shown in Fig. 2,

Table 1 Overview of benchmark models showing the method used, the numbers of different types of parameters, and the total number of ordinary differential equations (ODEs) per individual for the inner

and outer problem (including the number of sensitivity equations according to Eqs. 22 and 41)

Model	M1	M2	M3	M4
Method	FOCE	FOCE	FOCE	FOCEI
Total number of fixed effect parameters (θ)	6	12	18	18
Parameters in the ODE model	3	5	5	5
Parameters in the observational model	0	1	3	3
Parameters in the random effect covariance matrix	3	6	10	10
Number of random effect parameters (η)	3	3	4	4
ODEs per individual, inner problem	8	8	10	10
ODEs per individual, outer problem	44	60	80	80

where each row shows one element of the gradient at two levels of magnification.

The left column of Fig. 2 shows a pattern that appears to be consistent for all parameters; for large h , i.e. small step sizes, the result of the finite difference approach is dominated by numerical noise for both forward and central differences. Thus, for this particular model, and for this particular point in parameter space, the finite difference approximations have low precision as h increases beyond 3. For small h , i.e. large step sizes, there is a trend of severely decreased accuracy for the forward differences. Looking at the values of the gradient from the approach of sensitivity equations, it is clear that for h around 2 and smaller, forward differences produce values of elements of the gradient that are up to two orders of magnitude larger, and with a wrong sign in four of six cases. The behavior of the central difference approximation for small and intermediate h is best viewed in the right column, where the scales of the axis have been chosen differently. For the three first elements of the gradient, namely the derivatives of $\log L_F$ with respect to V_{max} , V_1 , and K_m , the central difference approximation appears to be accurate but, on the scale of the size of the gradient computed according to the sensitivity equation approach, the limits in precision are visible. For the derivatives with respect to the parameters of Ω , ω_{11} , ω_{22} , and ω_{33} , there are obvious issues with both accuracy and precision of the approximation, producing derivatives that are both of wrong size and sign. The fact that the approximation starts to deviate systematically for h less than 2 indicates that in these parameter directions, and on this scale, an expansion of the approximate log-likelihood function has a significant contribution of third order terms and higher, causing a bias in the approximation of the gradient using central differences.

The approach of determining the gradient using sensitivity equations is also subject to numerical errors. By repeated evaluation of the gradient using randomized values for the starting values of the inner optimization problem, we determined the relative standard error. For all 6

parameter directions of the gradients, the relative standard errors were between 0.1 and 1 %. Thus, these numerical errors are so small that they would not even be visible on the scales of Fig. 2.

Improving computational time

We investigated the improved computational times resulting from replacing finite difference approximations of the gradients in the inner and outer problem with gradients computed using sensitivity equations, and from using better starting values for the inner problems. The contribution from each of these three steps, as well as their accumulative effect, are shown in Fig. 3.

For the first step of improvement, using gradients based on sensitivity equations for the inner problem, computational times for models M1 and M2 (with 3 random effect parameters) decreased to almost a third compared to the approximation using forward differences, and to a fifth compared to central differences. The ratio of these two relative decreases is reasonable considering that the forward difference approximation requires 4 function evaluations and the central difference requires 7 evaluations. Model M3 and M4 contain one additional random effect parameter and the gains in speed were slightly larger compared to both variants of the finite difference approximation.

Replacing the finite difference approximation of the gradient in the outer problem with the approach based on sensitivity equations results in further improvement of computational times. As the number of parameters in the outer optimization problem increase from 6 to 18 for the models M1 to M3, the reduction in computational times improves from 29 to 14 % when compared to forward differences, and from 16 to 7 % compared to central differences. Although model M4 is identical to M3, the reduction in computational times are slightly less for this model. This is because M4 uses FOCEI for estimating parameters, which compared to FOCE requires more time

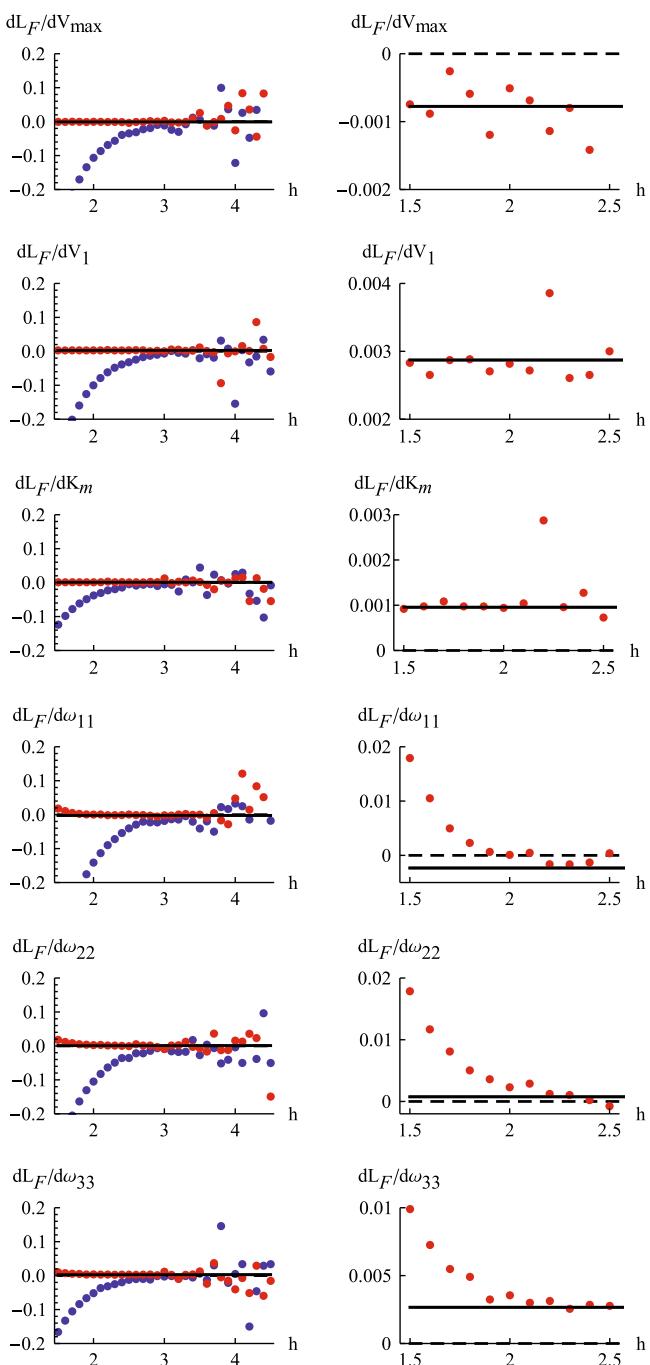


Fig. 2 Precision and accuracy of the approximate population log-likelihood gradient. Each row displays one element of the gradient, and the *left* and *right* columns show two different levels of magnification, respectively. Evaluations of the derivatives of $\log L_F$ using forward and central differences with different relative step sizes are shown as blue and red dots, respectively. A single evaluation of the derivatives using the approach based on sensitivity equations is indicated by a black line, and the value zero is shown as a dashed line for comparison

for putting together the more complex gradient expressions once the sensitivity equations have been solved. Again it is reasonable to expect a nearly doubled factor of decrease when comparing central and forward finite differences

since the former need almost twice as many function evaluations.

The final step of improvement is only applicable when gradients for both the inner and outer problem are computed using the approach based on sensitivity equations. Thus, the distinction between forward and central differences is no longer of importance. The decrease in computational times were around 70 % for models M1 to M3, and somewhat less for model M4, which again benefits less due to its larger overhead of having to compute all interaction terms.

The accumulated effect of all the steps range from a decrease in computational times to 7 % for the least complex model when comparing to forward differences, to the substantial decrease to 1 % for the most complex model when comparing to central differences.

Discussion

This article has demonstrated a novel approach to the computation of gradients needed for the FOCE and FOCEI approximation of the population likelihood encountered in NLME modeling. We have derived the analytic expressions for the gradients of both the individual and population log-likelihoods as well as the so called sensitivity equations, whose solution is a necessity for evaluating the gradient expressions.

Using sensitivity equations to compute the gradient for the inner problem is quite straightforward. As we understand it, approaches along these lines are in fact used for the inner problem, at least to some extent, in softwares such as NONMEM and Phoenix NLME. For the approximate population log-likelihood on the other hand, the sensitivity approach to gradient computation is complicated by the fact that this function depends on the nested optimization of the individual joint log-likelihoods. In this work we have, to the best of our knowledge, for the first time demonstrated how sensitivity equations can be used for computing the gradient of the FOCE and FOCEI approximations to the population log-likelihood. A key step to obtain this gradient involves the derivative of the optimal random effect parameters with respect to the fixed effect parameters. It was shown that this derivative could be determined given second order sensitivity equations.

Abandoning the finite difference approximation of gradients in favor of the approach of sensitivity equations were shown to have two advantages; gradients could be computed with a higher precision and computational times were substantially reduced. Though, implementation of the presented method is more challenging compared to finite difference FOCE/FOCEI, and the limitations of the Laplacian approximation are still present.

Increased precision and accuracy of gradients

The optimization of the approximate population log-likelihood $\log L_F$ with respect to θ would typically be performed with a quasi-Newton method. A straightforward approach to obtaining the gradient needed for such methods is to compute it from a finite difference approximation. However, the finite difference approach may result in issues with both precision and accuracy of the gradient. We demonstrated this for the computation of the gradient in the outer problem, evaluated close to the optimum of $\log L_F$. Although the use of central differences with an appropriate step length could avoid the worst problems, precision and accuracy were still inferior compared to the approach based on sensitivity equations. The potential limitations of combining NLME models based on differential equations with likelihood optimization using gradients computed by finite differences have previously been recognized [3]. The issues with the finite difference approximation depend both on numerical limitations and on the approximation itself. First of all, evaluation of $\log L_F$ can only be done to a certain precision. This is especially evident for models based on differential equations, whose solution involves adaptive schemes for numerical integration. In addition to the numerical precision of functions like \log , which is high, the precision of $\log L_F$ depends on the precision of the solutions to the differential equations, and the precision of computing derivatives with respect to η . The precision of $\log L_F$ also has a strong dependence on the precision of η^* , which in turn again depends on the solutions of differential equations and, if the inner level optimization problem is performed using a gradient-based method, depends on computing derivatives of the individual joint log-likelihoods with respect to η . Secondly, taking finite differences of $\log L_F$ will amplify numerical errors, resulting in increasingly poor precision of the gradient as the step size is decreased. On the other hand, taking too long steps will decrease the accuracy of the approximation due to the increasing impact of higher order terms in an expansion of $\log L_F$ (forward differences is only exact up to first order terms, and central differences is only exact up to second order terms). Even if it for a given model in some cases would be possible to customize the step length for the finite difference approximation (which typically would be different in each separate parameter direction) using an analysis like the one performed here, it would be infeasible in practice since such an investigation may take longer time than solving the parameter estimation problem itself. Adding further to the problem, the choice of a suitable step size will most certainly be different depending on the point in parameter space, thus constantly requiring a reevaluation of the step size.

There are several advantages of being able to compute gradients with an improved precision and accuracy

(i) Parameter estimates can be computed with higher precision, or alternatively, the same precision can be obtained but with shorter run times since we may afford to reduce the precision of the inner problem while still maintaining a similar precision in the outer problem [11]. (ii) Premature termination and convergence problems of the parameter estimation algorithm can be avoided or at least reduced [8, 24]. (iii) May enable the calculation of standard errors of the parameter estimates in cases where this was not possible due to the numerical issues of the finite difference approach [7]. However, we want to point out that for many points in the parameter space the limited precision and accuracy of the finite difference approach may not be crucial for the progression of the optimization as long as the approximation of the gradient results in a true ascent direction of the function being maximized.

Decreased computational times

The relative decrease in computational times were investigated for the successive application of three specific steps toward improvement, namely (i) Gradients based on sensitivity equations in the inner problem, (ii) Gradients based on sensitivity equations in the outer problem, and (iii) Better starting values for the inner problem. In all cases of applying the two first steps, we found that the decrease in computational times were substantially larger when comparing to central differences instead of forward differences. This was anticipated since central differences requires almost twice as many function evaluations as forward differences. Moreover, for both the inner and outer levels of optimization, the gains in computational times tended to be larger for models with higher number of parameters. For instance, the run time improvements of providing gradients from sensitivity equations in the outer problem were more than doubled for model M3 with 18 parameters compared to model M1 with 6 parameters. It was also observed that the improvement factor in the outer optimization was slightly lower for FOCEI compared to FOCE. Although the number of ODEs to be solved in both the inner and outer problem is the same, this was expected considering that the FOCEI method is based on more extensive expressions for both the likelihood and its gradient.

There are two main reasons why the approaches based on sensitivity equations should be faster. First of all, the right hand side of the sensitivity equations has lots of common subexpressions both with other sensitivity equations and with the original system of differential equations. Thus, the cost of evaluating the right hand side for the combined system of the original differential equations and the sensitivity equations can be surprisingly small. Furthermore, since the sensitivity equations are linear in the sensitivity state variables, there is typically little extra

effort needed in the adaptive time stepping of the differential equations solver for accommodating these additional equations. For the inner problem this means that it is faster to solve the combined system, yielding in total $n(1+q)$ differential equations, rather than having to solve the n original differential equations $1+q$ times, which would have been the case using forward finite differences. Secondly, the use of sensitivity equations in the outer level optimization avoids the repeated need of having to solve the inner problems for perturbed values of the outer parameters. The exact improvement made at this step depends on several factors of which perhaps the most important one is the desired precision (and hence the number of iterations required) of the inner optimizations needed for every parameter perturbation of a finite difference approximation (had this alternative been used instead).

We furthermore note that the computation of gradients based on sensitivity equations is highly amenable to parallelization, something which may be exploited to speed up computations considerably. The potential gains of doing this are expected to be similar to those of parallelizing the computation of the population log-likelihood itself [11].

In addition to the reduced computational times coming from the two steps of improved gradient computations, a third level of speed up was obtained by choosing more informed starting values for the inner problem. Although this improvement was not as substantial as the others, the gains from this step may be quite dependent on the starting values of the outer optimization problem. As the outer level optimization converges, the steps in θ become successively smaller, which in turn means that the linear approximation of $\eta^*(\theta)$ becomes better. Thus, the overall improvement in computational time will depend on how much of the optimization that was spent in these “later stages” of convergence. This means that it is likely that the relative improvement will be larger if the optimization had been started closer to the optimum.

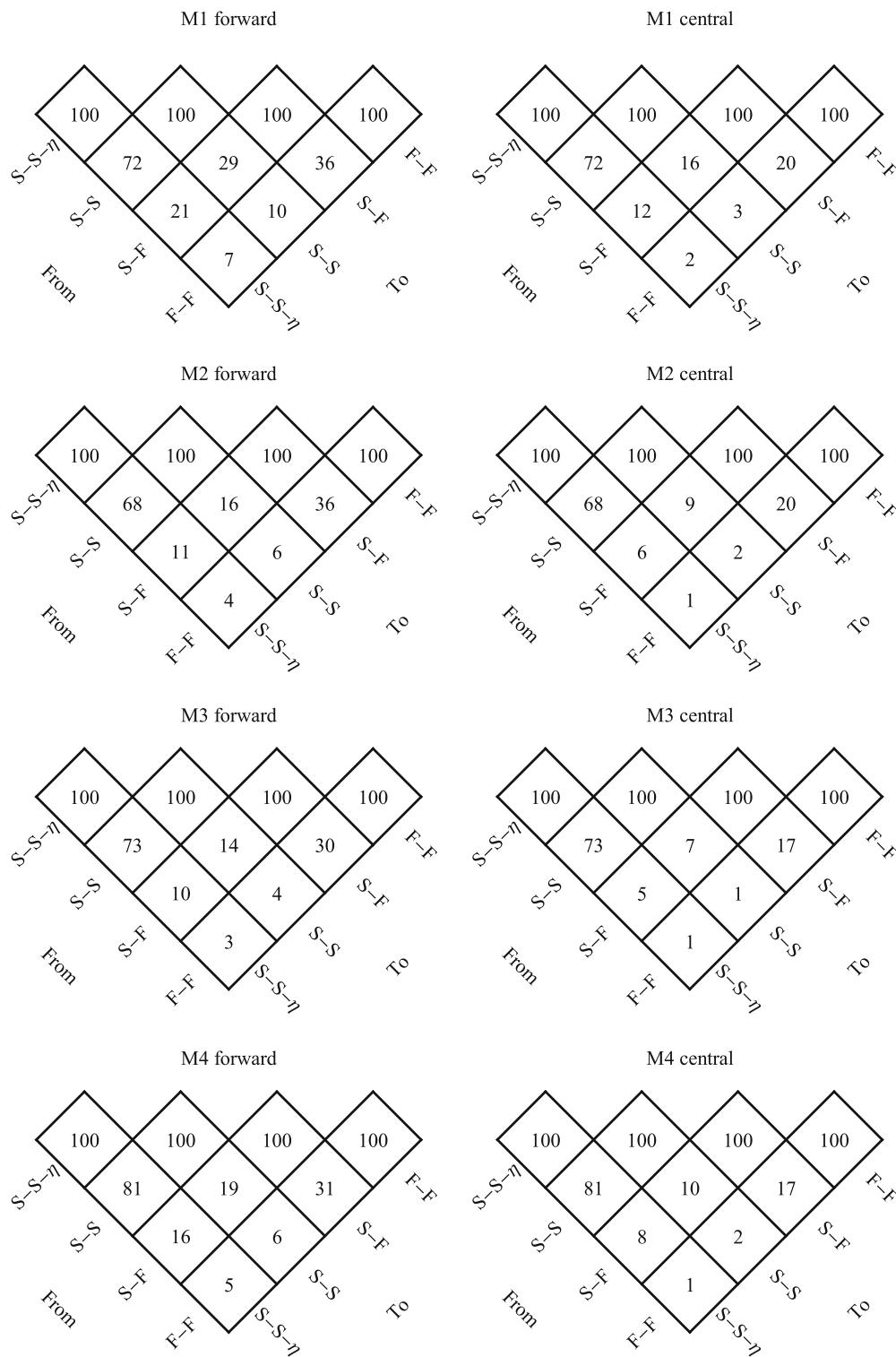
Setting the results of Fig. 3 in relation to commercial softwares for NLME parameter estimation, we would like to comment on a mixed analytical/finite difference approach to the differentiation of the FOCE likelihood with respect to the parameters of the random effect covariance matrix Ω , which is used as default by NONMEM (when the SLOW option is not selected). Since these parameters do not normally directly influence neither the residuals, nor the residual covariance matrix, their part of the likelihood gradient is less complicated compared to other parameters. As shown by the theory in this paper, their part of the gradient may be computed using only second order η sensitivities (Eq. 40), not requiring first order θ or second order mixed sensitivities (Eqs. 38 and 39, respectively). Although NONMEM FOCE does not use second order η

sensitivities, it still utilizes this technique by performing a central finite difference evaluation on the first order η sensitivities. While this is slower than performing completely analytical second derivatives, along with some erosion of precision, it is certainly faster than the SLOW FOCE method, which must perform the inner problem re-optimizations at each outer level perturbation of the Ω -parameters. The derivatives of the likelihood with respect to the remaining parameters are still obtained from finite differences.

The degree of improvement of speed for the S-S approach compared to an approach that is mixing finite differences and analytical methods at the outer level, i.e., an S-F/S approach, may therefore be less substantial than what can be achieved for going from S-F to S-S. Under the realistic assumption that all perturbed evaluations of $\log L_F$ are equally costly, and further assuming that the Ω -part of the gradient can be obtained at a computationally insignificant cost (ignoring the relatively few extra evaluations needed for the central finite difference of the first order η sensitivities), the reference time of 100 % for going from forward differences S-F to S-S in Fig. 3 would change to $((1 + P_\theta - P_\Omega)/(1 + P_\theta))100$ % if instead going from S-F/S to S-S, where P_θ is the total number of parameters and P_Ω is the number of Ω -parameters. The reference time for going from central differences S-F to S-S would for S-F/S to S-S similarly change to $((1 + 2P_\theta - 2P_\Omega)/(1 + 2P_\theta))100$ %. For model M1 this would mean that the improvements to 29 and 16, for forward and central differences, respectively, should be compared to the S-F/S references of 57 and 54, rather than to 100, and for model M3 the improvements to 14 and 7 should be compared to 47 and 46. In general, one would expect the advantage of the S-S approach to decrease as the fraction of Ω -parameters with respect to the total number of parameters increases, e.g., for problems with many random effect parameters when estimating the full random effect covariance matrix. It must however be emphasized that this is a mixed analytical/finite difference approach, and may as such have lower precision and accuracy compared to the S-S approach. Moreover, the remaining part of the gradient will still be completely derived from finite differences, and is expected to have the same comparable quality to the S-S approach as demonstrated in the results section.

Extending the line of thought, one could also consider a hybrid between the above S-F/S approach and the S-S approach, where the derivatives of $\log L_F$ with respect to the Ω -parameters are computed according to the exact approach presented in this work but where the derivatives for the remaining parameters of the outer level problem are obtained from a finite difference approach. This would indeed require the second order sensitivity

Fig. 3 Comparison of relative estimation times. The relative computation times expressed in percentage are shown for going from one scheme for obtaining gradients to another. Results are shown for the model variants M1–M4, using either a forward or central implementation of the finite difference approach. F-F denotes the use of finite differences for both the inner and outer problem, S-F the use of gradients based on sensitivity equations for the inner problem, S-S the use of gradients based on sensitivity equations for both inner and outer problems, and S-S- η denotes the additional implementation of the better starting values for the inner problem



equations with respect to η , but not the first order θ or the mixed second order sensitivity equations. The accuracy and precision would still be lower for the part of the gradient obtained from finite differences but the elements corresponding to the parameters of Ω would be of the same quality as the S-S approach, i.e., without approximations.

Challenges and limitations

Moving from a convenient proof-of-concept environment such as Mathematica, in which the proposed method currently is implemented, to a more stand-alone environment of a commercial software may present various challenges. One of the most obvious challenges is the integration of

functionality for performing symbolic differentiation. This is essential since the sensitivity equations, i.e., the differential equations in Eqs. 21, 38, 39, and 40, are model specific and have to be derived for every new model, in order to apply the results of this paper. It also applies to the derivatives of \mathbf{h} , \mathbf{R}_{ij} , and $\boldsymbol{\Omega}$, which too are model specific. Since differential equation models may be quite complex, and because second order derivatives are needed, it is not realistic to perform these derivations manually, and a tool that can perform symbolic differentiation will be required. To this end, one may consider to look at free symbolic packages such as SymPy [23]. The use of tools for symbolic analysis may furthermore be crucial to exploit the existence of common subexpressions, e.g., in the right hand sides of the sensitivity equations.

An alternative approach, which does not require symbolic differentiation, would be to use so called automatic differentiation (AD) [19]. The idea of AD is that every mathematical function that can be written as a computer program can be differentiated by applying the chain rule of differentiation, leading to the differentiation of every elementary operation of that computer program. Even though AD in principle could be applied directly to the approximate population likelihood, whose gradient we wish to compute, this would in practice be infeasible as this function is based on the execution of both optimization routines and adaptive numerical integration of differential equations. If used, AD would therefore not be applied to the population likelihood, but to the right hand sides of the model differential equations, and to the other model objects requiring differentiation. The parameter estimation would thus still proceed according to the steps laid out in Algorithm 1, but with symbolic differentiation replaced with AD. Following such an approach, the precision and accuracy of the gradients are not expected to differ, but it would have to be investigated how AD performs in terms of computational times. With a so called reverse mode AD it may actually be possible to improve run times even further compared to the current results.

Even if tools for differentiation can be provided for a stand-alone implementation, estimation methods which involve the direct differentiation of model state variables, etc., may experience limitations when considering other types of mathematical formalisms, such as models based on stochastic differential equations or hidden Markov models, since the required derivatives may be challenging to obtain. The method of computing gradients based on finite differences, on the other hand, do not care about the details of how a model is evaluated and has no limitations in this sense.

Finally, it should also be mentioned that although the approach for gradient computations presented here may improve the performance of FOCE and FOCEI, the

fundamental limitations of the Laplacian approximation as such still remains. Being only an approximation to the population likelihood, this class of methods do not guarantee the desirable statistical properties of a true maximum likelihood estimate. In this respect the new generation of estimation methods which are based on Monte Carlo expectation maximization methods, such as stochastic approximation expectation maximization and importance sampling, are superior to the classical ones since the parameter estimates and their confidence intervals, etc., are not biased by likelihood approximations. However, FOCE and FOCEI will likely be important complementary methods for a long time still, and improving their efficiency is therefore nonetheless relevant.

Possible extensions

The approach of computing gradients using sensitivity equations presented here could be modified for other variants of the population likelihood based on the Laplacian approximation. For instance, with some alterations it could be applied to the first order (FO) approximation of the population likelihood. Since the FO method does not rely on conditioning with respect to the optimal random effect parameters, the use of an approach based on sensitivity equations would be less complicated but at the same time also less rewarding. Gradients based on the approach of sensitivity equations could with some adjustments also be derived for the Laplace method. This would however require third order sensitivity equations but may be worthwhile since the potential gains should be at least as substantial as for FOCE and FOCEI. Because the theory presented in this article is derived for the FOCEI approximation, it accounts for the dependence of residual errors on the random effect parameters. This means that the gradient expressions stated here are suitable for prediction error-type NLME models, including models based on stochastic differential equations (see for instance [6, 14, 18]), since these typically display an interaction between residuals and random effects. The first step towards this end has in fact already been taken through the successful application of sensitivity equations for computing gradients in stochastic differential equation models on the single-subject level [16]. Furthermore, gradient computations based on sensitivity equations may be useful for the problem of optimal experimental design [1, 17].

Conclusions

The presented approach of computing gradients for both the individual- and population-level log-likelihoods of the FOCE and FOCEI approximations leads to more robust gradients and decreased computational times. We therefore

suggest that future implementations of these conditional estimation methods should include the approach based on sensitivity equations for computing the gradients. We eagerly await the further development of the proposed approach from the prototyped version used in the present study to its implementation in publicly or commercially available software packages.

Methods

The NLME parameter estimation algorithm investigated in this study was implemented in Mathematica 9. An executable version of the code, and the data sets used within this study, may be received from the authors upon request.

Comparison of performance

The performance of a computer program for parameter estimation in NLME models depends on several factors, such as the particular NLME model, the experimental data, how the estimation problem is formulated and possibly approximated, the choice and settings of the optimization method (including sub-methods such as line-searches, etc.), starting values of parameters, the differential equation solver used, the design of convergence criteria, etc. This paper is investigating the advantages of providing gradients by means of sensitivity equations for the FOCE or FOCEI approximation of the population likelihood. However, this paper is not claiming to address all the other factors that will impact on the parameter estimation. Comparing measures such as absolute run-times of our implementation with commercial software like NONMEM may therefore be misleading with respect to the advantages of gradient calculations. To avoid this the comparison is designed to look only at the improvements made by abandoning the finite difference approximation in our own implementation.

Comparison of precision and accuracy

The comparison of precision and accuracy was performed in the following way. At the optimal values of θ (found from the comparison of computational times), the elements of the gradient of the approximate log-likelihood function were approximated with finite differences, using a relative step size, according either to a forward difference

$$\frac{\log L_F(\theta_m(1 + 10^{-h})) - \log L_F(\theta_m)}{\theta_m 10^{-h}}, \quad (49)$$

or a central difference,

$$\frac{\log L_F(\theta_m(1 + 10^{-h})) - \log L_F(\theta_m(1 - 10^{-h}))}{2\theta_m 10^{-h}}. \quad (50)$$

For these function evaluations, the inner problem was solved to a precision of 4 digits (using the gradients from the approach of sensitivity equations). Furthermore, for forward differences the value of $\log L_F$ was recalculated for every h using randomized starting values for the inner problems. This was done to avoid correlations between differences with different step size that may otherwise have resulted from a single realization of the numerical error of $\log L_F$.

The approach of determining gradients using sensitivity equations does not involve any approximations, and is therefore expected to be correct on average. Its precision was assessed by computing the gradient 500 times using randomized starting values for the inner problems. For these gradient evaluations, the inner problem was solved to a precision of 4 digits.

Comparison of computational times

The comparison of computational times was done in the following way. Both the inner and outer problem were solved using gradients based on sensitivity equations, as outlined in the theory section. The inner problem was solved to a precision of 4 digits, and the outer to a precision of 3 digits. The comparison to finite differences was done by simultaneously clocking the time of computing gradients by a finite difference approximation but proceeding with the optimizations according to values of the gradient from the sensitivity approach. The reason for doing this is that the number of iterations, and the properties of every iteration (such as stiffness of the model equation with that certain set of parameters), for solving both the inner and outer problem might be affected by the choice of method for computing the gradients. Even small numerical differences in the results of the two methods may cause the paths taken in the parameter space to diverge substantially over the course of the optimizations, potentially making the comparison unfair. In this way we isolate the comparison to the actual computational times for the different methods of obtaining the gradients. Since the methods based on sensitivity equations were shown to have a higher precision in the evaluation of gradients, there may be additional gains in computational times to be made from traversing the parameter space based on more exact gradients. However, quantifying this type of contribution may require averaging over a large number of models and parameter starting values and was not considered. Thus, our implementation of the comparison focuses on the direct improvements in computational times and will therefore be a conservative measure of the gains in speed.

To make a fair implementation of timing the finite differences approach the following starting values of the random effect parameters for the inner problem were used.

When evaluating the approximate population log-likelihood at the unperturbed parameter values of the outer problem, the starting values for the parameters of the inner problem were set to the optimum from the previous outer evaluation, i.e., according to approach A in Fig. 1. For evaluating the approximate population log-likelihood at the perturbed parameter values of the outer problem, the starting values for the parameters of the inner problem were set to the optimum obtained for the unperturbed outer problem parameters. The relative size of each perturbation of the parameters in θ was 10^{-2} .

Compared to the finite difference approaches, using sensitivity equations had an overhead of evaluating the quite substantial mathematical expressions for the gradients once the differential equations are integrated, something which was carefully included in the comparison of computational times.

Optimization algorithm

Both the inner and outer optimization problems were solved using the BFGS method [20].

Derivation of sensitivity equations

Given an NLME differential equation model, the corresponding sensitivity equations were derived by symbolic differentiation in Mathematica.

Acknowledgments We thank the two anonymous reviewers for their constructive comments which significantly contributed to improving the quality of this publication. This project has in part been supported by the Swedish Foundation for Strategic Research, which is gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix 1: Matrix calculus

The default representation of a vector is a column vector,

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}. \quad (51)$$

The derivatives of vectors and matrices by scalars are defined as element-wise derivatives, according to

$$\frac{d\mathbf{y}}{dx} = \begin{pmatrix} \frac{dy_1}{dx} \\ \frac{dy_2}{dx} \\ \vdots \\ \frac{dy_m}{dx} \end{pmatrix}, \quad (52)$$

and

$$\frac{d\mathbf{A}}{dx} = \begin{pmatrix} \frac{da_{11}}{dx} & \frac{da_{12}}{dx} & \cdots & \frac{da_{1n}}{dx} \\ \frac{da_{21}}{dx} & \frac{da_{22}}{dx} & \cdots & \frac{da_{2n}}{dx} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{da_{m1}}{dx} & \frac{da_{m2}}{dx} & \cdots & \frac{da_{mn}}{dx} \end{pmatrix}, \quad (53)$$

respectively. The derivative of scalar by vector is given by

$$\frac{dy}{d\mathbf{x}} = \left(\frac{dy}{dx_1} \quad \frac{dy}{dx_2} \quad \cdots \quad \frac{dy}{dx_m} \right), \quad (54)$$

the derivative of vector by vector is given by

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \begin{pmatrix} \frac{dy_1}{dx_1} & \frac{dy_1}{dx_2} & \cdots & \frac{dy_1}{dx_n} \\ \frac{dy_2}{dx_1} & \frac{dy_2}{dx_2} & \cdots & \frac{dy_2}{dx_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dy_m}{dx_1} & \frac{dy_m}{dx_2} & \cdots & \frac{dy_m}{dx_n} \end{pmatrix}, \quad (55)$$

and the derivative of row-vector by vector is given by

$$\frac{d\mathbf{y}^T}{d\mathbf{x}} = \begin{pmatrix} \frac{dy_1}{dx_1} & \frac{dy_2}{dx_1} & \cdots & \frac{dy_m}{dx_1} \\ \frac{dy_1}{dx_2} & \frac{dy_2}{dx_2} & \cdots & \frac{dy_m}{dx_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dy_1}{dx_m} & \frac{dy_2}{dx_m} & \cdots & \frac{dy_m}{dx_m} \end{pmatrix}. \quad (56)$$

The derivative of a quadratic form is obtained in the following way. Let $y = \mathbf{b}^T \mathbf{A} \mathbf{b}$, where \mathbf{A} is a square matrix and \mathbf{b} a suitable vector. If \mathbf{A} is symmetric then

$$\begin{aligned} \frac{dy}{dx} &= \frac{d\mathbf{b}^T}{dx} \mathbf{A} \mathbf{b} + \mathbf{b}^T \frac{d\mathbf{A}}{dx} \mathbf{b} + \mathbf{b}^T \mathbf{A} \frac{d\mathbf{b}}{dx} \\ &= \mathbf{b}^T \mathbf{A}^T \frac{d\mathbf{b}}{dx} + \mathbf{b}^T \frac{d\mathbf{A}}{dx} \mathbf{b} + \mathbf{b}^T \mathbf{A} \frac{d\mathbf{b}}{dx} \\ &= 2\mathbf{b}^T \mathbf{A} \frac{d\mathbf{b}}{dx} + \mathbf{b}^T \frac{d\mathbf{A}}{dx} \mathbf{b}. \end{aligned} \quad (57)$$

The derivative of an inverse matrix is found by noting that

$$\begin{aligned} \frac{d\mathbf{A}^{-1}}{dx} &= \frac{d(\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1})}{dx} = \frac{d\mathbf{A}^{-1}}{dx}\mathbf{A}\mathbf{A}^{-1} + \mathbf{A}^{-1}\frac{d\mathbf{A}}{dx}\mathbf{A}^{-1} \\ &\quad + \mathbf{A}\mathbf{A}^{-1}\frac{d\mathbf{A}^{-1}}{dx}, \end{aligned} \quad (58)$$

and thus that

$$\frac{d\mathbf{A}^{-1}}{dx} = -\mathbf{A}^{-1}\frac{d\mathbf{A}}{dx}\mathbf{A}^{-1}. \quad (59)$$

The derivative of the logarithm of the determinant of a covariance matrix is given by the following expression. If \mathbf{A} is a real-valued, symmetric, positive-definite matrix, then

$$\frac{d}{dx}\log|\mathbf{A}| = \text{tr}\left[\mathbf{A}^{-1}\frac{d\mathbf{A}}{dx}\right]. \quad (60)$$

This can be seen by first writing \mathbf{A} as $\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^{-1}$, where Λ is a diagonal matrix. Now, the left-hand side of Eq. 60 becomes

$$\begin{aligned} \frac{d}{dx}\log|\mathbf{A}| &= \frac{d}{dx}\log(|\mathbf{Q}|\cdot|\Lambda|\cdot|\mathbf{Q}^{-1}|) = \frac{d}{dx}\log|\Lambda| \\ &= \frac{d}{dx}\sum_i \log\Lambda_{ii} = \sum_i \frac{1}{\Lambda_{ii}}\frac{d\Lambda_{ii}}{dx} = \text{tr}\left[\Lambda^{-1}\frac{d\Lambda}{dx}\right], \end{aligned} \quad (61)$$

which is equal to the right-hand side of Eq. 60 since

$$\begin{aligned} \text{tr}\left[\mathbf{A}^{-1}\frac{d\mathbf{A}}{dx}\right] &= \text{tr}\left[\mathbf{Q}\Lambda^{-1}\mathbf{Q}^{-1}\frac{d\mathbf{Q}}{dx}\mathbf{Q}\Lambda^{-1}\right] \\ &\quad + \text{tr}\left[\mathbf{Q}\Lambda^{-1}\frac{d\Lambda}{dx}\mathbf{Q}^{-1}\right] - \text{tr}\left[\frac{d\mathbf{Q}}{dx}\mathbf{Q}^{-1}\right] \\ &= \text{tr}\left[\frac{d\mathbf{Q}}{dx}\mathbf{Q}^{-1}\right] + \text{tr}\left[\Lambda^{-1}\frac{d\Lambda}{dx}\right] \\ &\quad - \text{tr}\left[\frac{d\mathbf{Q}}{dx}\mathbf{Q}^{-1}\right] = \text{tr}\left[\Lambda^{-1}\frac{d\Lambda}{dx}\right]. \end{aligned} \quad (62)$$

Appendix 2: Hessian approximation

For an appropriate model, it holds that

$$\mathbf{E}[\epsilon_{ij}] = \mathbf{0}, \quad (63)$$

and

$$\mathbf{E}\left[\epsilon_{ij}\epsilon_{ij}^T\right] = \mathbf{R}_{ij}, \quad (64)$$

where the expected values are taken with respect to data, which here are considered to be random variables whose values have not yet been realized. Based on these equations, the Hessian in Eq. 13 can be simplified to various degrees by approximating its different terms with their expected values. A minimal simplification for eliminating the second order derivative terms is achieved by noting that

$$\mathbf{E}\left[2\epsilon_{ij}^T\mathbf{R}_{ij}^{-1}\frac{d^2\epsilon_{ij}}{d\eta_{ik}d\eta_{il}}\right] = \mathbf{E}\left[2\epsilon_{ij}^T\right]\mathbf{R}_{ij}^{-1}\frac{d^2\epsilon_{ij}}{d\eta_{ik}d\eta_{il}} = 0, \quad (65)$$

and

$$\begin{aligned} &\mathbf{E}\left[-\epsilon_{ij}^T\mathbf{R}_{ij}^{-1}\frac{d^2\mathbf{R}_{ij}}{d\eta_{ik}d\eta_{il}}\mathbf{R}_{ij}^{-1}\epsilon_{ij} + \text{tr}\left[\mathbf{R}_{ij}^{-1}\frac{d^2\mathbf{R}_{ij}}{d\eta_{ik}d\eta_{il}}\right]\right] \\ &= \mathbf{E}\left[-\text{tr}\left[\epsilon_{ij}^T\mathbf{R}_{ij}^{-1}\frac{d^2\mathbf{R}_{ij}}{d\eta_{ik}d\eta_{il}}\mathbf{R}_{ij}^{-1}\epsilon_{ij}\right]\right] + \text{tr}\left[\mathbf{R}_{ij}^{-1}\frac{d^2\mathbf{R}_{ij}}{d\eta_{ik}d\eta_{il}}\right] \\ &= \mathbf{E}\left[-\text{tr}\left[\mathbf{R}_{ij}^{-1}\frac{d^2\mathbf{R}_{ij}}{d\eta_{ik}d\eta_{il}}\mathbf{R}_{ij}^{-1}\epsilon_{ij}\epsilon_{ij}^T\right]\right] + \text{tr}\left[\mathbf{R}_{ij}^{-1}\frac{d^2\mathbf{R}_{ij}}{d\eta_{ik}d\eta_{il}}\right] \\ &= -\text{tr}\left[\mathbf{R}_{ij}^{-1}\frac{d^2\mathbf{R}_{ij}}{d\eta_{ik}d\eta_{il}}\right] + \text{tr}\left[\mathbf{R}_{ij}^{-1}\frac{d^2\mathbf{R}_{ij}}{d\eta_{ik}d\eta_{il}}\right] = 0, \end{aligned} \quad (66)$$

where we are making use of the fact that the trace of a scalar is just the scalar, the order of the expectation and trace operators can be shifted, and the cyclic property of the trace operator. This simplification is used in the present study.

Further simplifications of Eq. 13 may be performed by noting that the expectation of additional terms vanishes,

$$\mathbf{E}\left[-2\epsilon_{ij}^T\mathbf{R}_{ij}^{-1}\frac{d\mathbf{R}_{ij}}{d\eta_{il}}\mathbf{R}_{ij}^{-1}\frac{d\epsilon_{ij}}{d\eta_{ik}}\right] = 0, \quad (67)$$

$$\mathbf{E}\left[-2\epsilon_{ij}^T\mathbf{R}_{ij}^{-1}\frac{d\mathbf{R}_{ij}}{d\eta_{ik}}\mathbf{R}_{ij}^{-1}\frac{d\epsilon_{ij}}{d\eta_{il}}\right] = 0, \quad (68)$$

and by taking the expected value and collecting terms,

$$\begin{aligned} &\mathbf{E}\left[2\epsilon_{ij}^T\mathbf{R}_{ij}^{-1}\frac{d\mathbf{R}_{ij}}{d\eta_{ik}}\mathbf{R}_{ij}^{-1}\frac{d\mathbf{R}_{ij}}{d\eta_{il}}\mathbf{R}_{ij}^{-1}\epsilon_{ij} - \text{tr}\left[\mathbf{R}_{ij}^{-1}\frac{d\mathbf{R}_{ij}}{d\eta_{il}}\mathbf{R}_{ij}^{-1}\frac{d\mathbf{R}_{ij}}{d\eta_{ik}}\right]\right] \\ &= \text{tr}\left[\mathbf{R}_{ij}^{-1}\frac{d\mathbf{R}_{ij}}{d\eta_{il}}\mathbf{R}_{ij}^{-1}\frac{d\mathbf{R}_{ij}}{d\eta_{ik}}\right]. \end{aligned} \quad (69)$$

Taken together, all simplifications yield the following Hessian

$$\tilde{\mathbf{H}}_{ikl} = -\sum_{j=1}^{n_i} \left(\frac{d\epsilon_{ij}^T}{d\eta_{il}}\mathbf{R}_{ij}^{-1}\frac{d\epsilon_{ij}}{d\eta_{ik}} + \frac{1}{2}\text{tr}\left[\mathbf{R}_{ij}^{-1}\frac{d\mathbf{R}_{ij}}{d\eta_{il}}\mathbf{R}_{ij}^{-1}\frac{d\mathbf{R}_{ij}}{d\eta_{ik}}\right] \right) - \Omega_{kl}^{-1}, \quad (70)$$

which is the variant used in NONMEM [2].

Appendix 3: Benchmark models and data

The equations for the two-compartment pharmacokinetic model are

Table 2 Parameter values used for simulating data (D), starting values for estimation (S), and parameter estimates (E) for the different models

Parameter	D	S, M1	S, M2	S, M3/M4	E, M1	E, M2	E, M3	E, M4
V_{max}	0.5	0.2	0.2	0.2	0.424	0.419	0.473	0.473
K_m	4	3	3	3	3.91	2.53	4.37	4.37
Cl_d	0.01	—	0.01	0.01	—	0.00976	0.00813	0.00813
V_1	0.3	0.1	0.1	0.1	0.288	0.285	0.321	0.321
V_2	0.1	—	0.1	0.1	—	0.0956	0.0959	0.0959
r_{a1}	$\sqrt{0.5} \approx 0.707$	—	$\sqrt{0.1} \approx 0.316$	$\sqrt{0.1} \approx 0.316$	—	0.414	0.644	0.644
r_{p1}	0*	—*	—*	$\sqrt{0.1} \approx 0.316$	—*	—*	0.00165	0.00163
r_{a2}	$\sqrt{0.5}^* \approx 0.707$	—*	—*	$\sqrt{0.1} \approx 0.316$	—*	—*	0.730	0.730
ω_{11}	$\sqrt{0.5} \approx 0.707$	1	1	1	0.616	0.553	0.559	0.560
ω_{12}	0	—	0	0	—	-0.0518	-0.123	-0.123
ω_{13}	0	—	0	0	—	0.439	-0.138	-0.138
ω_{14}^*	0*	—*	—*	0	—*	—*	0.0273	0.0275
ω_{22}	$\sqrt{0.5} \approx 0.707$	1	1	1	0.772	0.575	0.533	0.533
ω_{23}	0	—	0	0	—	-0.485	0.0174	0.0174
ω_{24}^*	0*	—*	—*	0	—*	—*	-0.0230	-0.0230
ω_{33}	$\sqrt{0.5} \approx 0.707$	1	1	1	0.994	1.39	0.776	0.776
ω_{34}^*	0*	—*	—*	0	—*	—*	-0.409	-0.409
ω_{44}^*	$\sqrt{0.5}^* \approx 0.707$	—*	—*	1	—*	—*	0.870	0.870

Parameters which were not estimated are indicated with a dash. The * indicate that a parameter is only used in models M3 and M4

$$\begin{aligned} V_1 \frac{dc_1(t)}{dt} &= u(t) + Cl_d(c_2(t) - c_1(t)) - \frac{V_{max} c_1(t)}{K_m c_1(t)} \\ V_2 \frac{dc_2(t)}{dt} &= Cl_d(c_1(t) - c_2(t)) \\ c_1(0) = c_2(0) &= 0, \end{aligned} \quad (71)$$

where $u(t)$ is an input function, which was used to model a constant infusion with the rate 0.67 per minute during the first 30 minutes followed by another 30 minutes of wash-out. For models M1 and M2, the scalar-valued observation model was defined by $\mathbf{y}_t = c_1(t) + \mathbf{e}_t$, where $\mathbf{e}_t \in N(0, \mathbf{R}_t)$ and

$$\mathbf{R}_t = (r_{a1}^2). \quad (72)$$

For models M3 and M4, the vector-valued observation model was defined by $\mathbf{y}_t = (c_1(t), c_2(t)) + \mathbf{e}_t$, where

$$\mathbf{R}_t = \begin{pmatrix} (r_{a1} + r_{p1} c_1(t))^2 & \\ & r_{a2}^2 \end{pmatrix}. \quad (73)$$

In models M1 and M2, the three parameters V_{max} , K_m , and V_1 , were defined to be log-normally distributed on the population level. This was accomplished by multiplying them with $\exp(\eta_1)$, $\exp(\eta_2)$, and $\exp(\eta_3)$, respectively, where $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)$ is normally distributed with zero mean. In the first variant of this model, M1, the covariance

matrix for the random effect parameters is defined by the diagonal matrix

$$\Omega = \begin{pmatrix} \omega_{11}^2 & & \\ & \omega_{22}^2 & \\ & & \omega_{33}^2 \end{pmatrix}, \quad (74)$$

and in the second variant, M2, the full matrix is estimated using the parameterization

$$\Omega = \begin{pmatrix} \omega_{11}^2 + \omega_{12}^2 + \omega_{13}^2 & \omega_{12}\omega_{22} + \omega_{13}\omega_{23} & \omega_{13}\omega_{33} \\ \omega_{12}\omega_{22} + \omega_{13}\omega_{23} & \omega_{22}^2 + \omega_{23}^2 & \omega_{23}\omega_{33} \\ \omega_{13}\omega_{33} & \omega_{23}\omega_{33} & \omega_{33}^2 \end{pmatrix} \quad (75)$$

to ensure positive definiteness. In models M3 and M4, an additional random effect parameter was in the same way introduced for the parameter Cl_d . A similarly defined full matrix for 4 random effect parameters was used for models M3 and M4.

The parameter values used for simulating data are shown in Table 2, together with information of which parameters are being estimated in the four model variants, and what the starting values of the estimation were. One data set consisting of 10 simulated individuals was used for models M1 and M2. Here, the values of c_1 were collected at the time points $t = 10, 15, 20, \dots, 60$. For models M3 and M4, another data set consisting of 20 simulated

individuals was used, where the values of c_1 and c_2 were collected at the time points $t = 10, 15, 20, \dots, 60$.

References

- Atkinson AC, Bogacka B (2002) Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. *Chemom Intell Lab Sys* 61:17–33
- Bauer R (2010) NONMEM7 Technical Guide
- Bauer R (2014) NONMEM User's Guides Introduction to NONMEM 7.3.0. Hanover, MD
- Bauer RJ, Guzy S, Ng C (2007) A survey of population analysis methods and software for complex pharmacokinetic and pharmacodynamic models with examples. *AAPS J* 9(1):E60–E83. doi:[10.1208/aapsj0901007](https://doi.org/10.1208/aapsj0901007)
- Beal S, Sheiner L, Boeckmann A, Bauer R (2009) NONMEM User's Guides. (1989–2009). Icon Development Solutions, Ellington City, MD
- Berglund M, Sunnåker M, Adiels M, Jirstrand M, Wennberg B (2012) Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations. *Math Med Biol* 29(4):361–384. doi:[10.1093/imammb/dqr021](https://doi.org/10.1093/imammb/dqr021)
- Bertrand J, Laffont CM, Mentré F, Chenel M, Comets E (2011) Development of a complex parent-metabolite joint population pharmacokinetic model. *AAPS J* 13(3):390–404. doi:[10.1208/s12248-011-9282-9](https://doi.org/10.1208/s12248-011-9282-9)
- Chan PLS, Jacqmin P, Lavielle M, McFadyen L, Weatherley B (2011) The use of the SAEM algorithm in MONOLIX software for estimation of population pharmacokinetic-pharmacodynamic-viral dynamics parameters of maraviroc in asymptomatic HIV subjects. *J Pharmacokinet Pharmacodyn* 38(1):41–61. doi:[10.1007/s10928-010-9175-z](https://doi.org/10.1007/s10928-010-9175-z)
- Dartois C, Lemenuel-Diot A, Laveille C, Tranchand B, Tod M, Girard P (2007) Evaluation of uncertainty parameters estimated by different population PK software and methods. *J Pharmacokinet Pharmacodyn* 34(3):289–311. doi:[10.1007/s10928-006-9046-9](https://doi.org/10.1007/s10928-006-9046-9)
- Davidian M, Giltinan DM (2003) Nonlinear models for repeated measurement data: an overview and update. *J Agric Biol Environ Stat* 8:387–419
- Gibiansky L, Gibiansky E, Bauer R (2012) Comparison of Nonmem 7.2 estimation methods and parallel processing efficiency on a target-mediated drug disposition model. *J Pharmacokinet Pharmacodyn* 39(1):17–35. doi:[10.1007/s10928-011-9228-y](https://doi.org/10.1007/s10928-011-9228-y)
- Johansson ÅM, Ueckert S, Plan EL, Hooker AC, Karlsson MO (2014) Evaluation of bias, precision, robustness and runtime for estimation methods in NONMEM 7. *J Pharmacokinet Pharmacodyn* 41(3):223–238. doi:[10.1007/s10928-014-9359-z](https://doi.org/10.1007/s10928-014-9359-z)
- Kiang TKL, Sherwin CMT, Spigarelli MG, Ensom MHH (2012) Fundamentals of population pharmacokinetic modelling: modelling and software. *Clin Pharmacokinet* 51(8):515–525. doi:[10.2165/11634080-00000000-00000](https://doi.org/10.2165/11634080-00000000-00000)
- Kristensen NR, Madsen H, Ingwersen SH (2005) Using stochastic differential equations for PK/PD model development. *J Pharmacokinet Pharmacodyn* 32(1):109–141. doi:[10.1007/s10928-005-2105-9](https://doi.org/10.1007/s10928-005-2105-9)
- Lavielle M (2014) Monolix: a software for the analysis of non-linear mixed effects models. The Monolix Group
- Leander J, Lundh T, Jirstrand M (2014) Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements. *Math Biosci* 251:54–62. doi:[10.1016/j.mbs.2014.03.001](https://doi.org/10.1016/j.mbs.2014.03.001)
- Mentré F, Mallet A, Baccar D (1997) Optimal design in random-effects regression models. *Biometrika* 84:429–442
- Møller JB, Overgaard RV, Madsen H, Hansen T, Pedersen O, Ingwersen SH (2010) Predictive performance for population models using stochastic differential equations applied on data from an oral glucose tolerance test. *J Pharmacokinet Pharmacodyn* 37(1):85–98. doi:[10.1007/s10928-009-9145-5](https://doi.org/10.1007/s10928-009-9145-5)
- Neidinger RD (2010) Introduction to automatic differentiation and MATLAB object-oriented programming. *SIAM Rev* 52:545–563
- Nocedal J, Wright SJ (1999) Numerical optimization. Springer, Berlin
- Pharsight, Cary, NC: Phoenix NLME
- Plan EL, Maloney A, Mentré F, Karlsson MO, Bertrand J (2012) Performance comparison of various maximum likelihood non-linear mixed-effects estimation methods for dose-response models. *AAPS J* 14(3):420–432. doi:[10.1208/s12248-012-9349-2](https://doi.org/10.1208/s12248-012-9349-2)
- SymPy Development Team: SymPy: Python library for symbolic mathematics (2014). <http://www.sympy.org>
- Tapani S, Almquist J, Leander J, Ahlström C, Peletier LA, Jirstrand M, Gabrielsson J (2014) Joint feedback analysis modeling of nonesterified fatty acids in obese Zucker rats and normal Sprague–Dawley rats after different routes of administration of nicotinic acid. *J Pharm Sci* 103(8):2571–2584. doi:[10.1002/jps.24077](https://doi.org/10.1002/jps.24077)
- Wang Y (2007) Derivation of various NONMEM estimation methods. *J Pharmacokinet Pharmacodyn* 34(5):575–593. doi:[10.1007/s10928-007-9060-6](https://doi.org/10.1007/s10928-007-9060-6)

Mixed Effects Modeling Using Stochastic Differential Equations: Illustrated by Pharmacokinetic Data of Nicotinic Acid in Obese Zucker Rats

Research Article

Mixed Effects Modeling Using Stochastic Differential Equations: Illustrated by Pharmacokinetic Data of Nicotinic Acid in Obese Zucker Rats

Jacob Leander,^{1,2} Joachim Almquist,^{1,3} Christine Ahlström,⁴ Johan Gabrielsson,⁵ and Mats Jirstrand^{1,6}

Received 22 August 2014; accepted 25 November 2014; published online 19 February 2015

Abstract. Inclusion of stochastic differential equations in mixed effects models provides means to quantify and distinguish three sources of variability in data. In addition to the two commonly encountered sources, measurement error and interindividual variability, we also consider uncertainty in the dynamical model itself. To this end, we extend the ordinary differential equation setting used in nonlinear mixed effects models to include stochastic differential equations. The approximate population likelihood is derived using the first-order conditional estimation with interaction method and extended Kalman filtering. To illustrate the application of the stochastic differential mixed effects model, two pharmacokinetic models are considered. First, we use a stochastic one-compartment model with first-order input and nonlinear elimination to generate synthetic data in a simulated study. We show that by using the proposed method, the three sources of variability can be successfully separated. If the stochastic part is neglected, the parameter estimates become biased, and the measurement error variance is significantly overestimated. Second, we consider an extension to a stochastic pharmacokinetic model in a preclinical study of nicotinic acid kinetics in obese Zucker rats. The parameter estimates are compared between a deterministic and a stochastic NiAc disposition model, respectively. Discrepancies between model predictions and observations, previously described as measurement noise only, are now separated into a comparatively lower level of measurement noise and a significant uncertainty in model dynamics. These examples demonstrate that stochastic differential mixed effects models are useful tools for identifying incomplete or inaccurate model dynamics and for reducing potential bias in parameter estimates due to such model deficiencies.

KEY WORDS: extended Kalman filter; model uncertainty; nonlinear kinetics; parameter estimation; state prediction.

INTRODUCTION

In pharmacokinetic and pharmacodynamic modeling, the physical system is often assumed to be described by a system of ordinary differential equations (ODEs). In pharmacokinetics, compartmental models are mostly used, whereas for pharmacodynamics, direct or turnover response models are common (1). The observed data are assumed to arise from a deterministic process under some measurement noise (additive, proportional, a combination of the two, or more general probabilistic models). However, the use of

deterministic modeling approaches for describing the dynamics often suffers from limited and uncertain knowledge regarding the details of such processes. Since it is up to the modeler to define a model describing the drug administration and its effect, there is also an uncertainty in the model itself. This uncertainty is not explicitly accounted for when considering a deterministic model, for example when using ODEs to describe the dynamics together with a measurement model to incorporate the error. This can lead to model deficiencies, such as correlated residuals, overestimated measurement noise, and incorrect inference (2).

Nonlinear mixed effects (NLME) models were introduced into the pharmaceutical field to analyze data from several individuals simultaneously (3–5). The individuals are assumed to be described by a common structural model with some of the model parameters varying within the population (so-called random effects parameters), while other parameters are invariant between subjects (so-called fixed effects parameters). The NLME approach can be of great benefit when the data is sparse and the information from a single subject is not sufficient to identify the model parameters. It is typically performed using a deterministic model describing the underlying system, for example by utilizing ODEs (6, 7).

¹Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-41288, Gothenburg, Sweden.

²Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden.

³Systems and Synthetic Biology, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden.

⁴CVMD iMed DMPK, AstraZeneca R&D, Mölndal, Sweden.

⁵Division of Pharmacology and Toxicology, Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, Uppsala, Sweden.

⁶To whom correspondence should be addressed. (e-mail: mats.jirstrand@fcc.chalmers.se)

We consider the extension of the NLME approach to allow for uncertainty in the model dynamics. This is done by considering stochastic differential equations (SDEs), which is an extension of ODEs to allow for a random part in the model dynamics. This approach has previously been advocated, see for example (8–10). We have previously also demonstrated the benefits of using SDEs when solving the inverse problem of parameter estimation (11). SDEs can furthermore be used to model the inherent randomness in pharmacokinetic and pharmacodynamic systems (12, 13), as an alternative to discrete models using a master equation approach and the Gillespie algorithm (14–16). Also, in systems without true randomness, SDEs can be used to model an incomplete or imperfect model structure. SDEs (see (17, 18) for references) have long been used in mathematical finance, for example to model the uncertainty in an asset (19). In contrast to the classical approach, where data variability arises from the measurements and the variability in parameters, a stochastic model also incorporates errors in the dynamics itself. Hence, this kind of modeling allows for three different sources of variability: population variability, measurement error, and system noise.

One often faces the inverse problem of estimating model parameters from observed noisy data. There are several approaches to the delicate problem of estimating parameters in stochastic differential mixed effects models. In general, there is no closed form solution of the likelihood function. Approaches to the parameter estimation problem on a population level include for example the first-order (FO) and the first-order conditional estimation (FOCE) method (20, 21) and stochastic approximation of expectation maximization (SAEM) method (22). State estimation on an individual level includes, for example, the Kalman filter (KF), the extended Kalman filter (EKF), and particle filters. For a combination of the FOCE approximation of the population likelihood and the EKF, see for example (10, 23–25). In (26), the authors propose a combination of the SAEM algorithm and EKF. For a review of parameter estimation methods in SDE population models, see (27).

In this paper, we consider approximation of the population likelihood by using the first-order conditional estimation with interaction (FOCEI) approximation of the population likelihood together with the EKF for state estimation on the individual level. In contrast to previous efforts (10, 23–25), we estimate the full covariance matrix describing population variability. That is, we allow for correlation between random parameters in the model. By utilizing the FOCEI method, we allow for interaction between output variance and random parameters. Furthermore, instead of adopting the commonly used finite difference approximation, we make use of sensitivity equations to evaluate the gradient of the objective function in the optimization procedure, previously mentioned in (28). Moreover, we produce illustrative plots describing state variable uncertainty and output uncertainty (which is a combination of state and measurement uncertainty). These plots serve as diagnostic tools of model appropriateness and illustration of the uncertainty in model output.

The extension of ODEs to SDEs is illustrated using two examples of pharmacokinetic data. First, a stochastic one-compartmental pharmacokinetic model with first-order input and nonlinear elimination is considered by using a simulated

data set consisting of 20 animals. Since we now account for three sources of variability in data, it is important to know if the three sources can be distinguished from each other. From the simulated data, the parameters of the model are estimated, including the system noise and the covariance matrices describing measurement error and parameter variability. Second, we consider a data set from a preclinical study of nicotinic acid (NiAc) turnover in obese rats, where the original NiAc disposition model and a NiAc disposition model extended to an SDE model are compared in terms of parameter estimates and model prediction.

MATERIALS AND METHODS

Mathematical Theory

In this section, we state the stochastic mixed effects model and derive the (approximate) maximum likelihood theory needed for parameter estimation. This section is recommended for readers not familiar with the concept of SDEs. We also introduce the concept of the EKF, which serves as a state estimator for the stochastic model (29).

The Stochastic Mixed Effects Model Framework

In population modeling, NLME models are used to describe data of the form

$$\mathbf{y}_{ij}, i = 1, \dots, N, j = 1, \dots, n_i, \quad (1)$$

where the vector \mathbf{y}_{ij} denotes the j :th observation for the i :th individual. The statistical model is the following

$$d\mathbf{x}_i = \mathbf{f}(\mathbf{x}_i, \mathbf{u}_i, t, \boldsymbol{\phi}_i)dt, \mathbf{x}_i(0) = \mathbf{x}_0(\boldsymbol{\phi}_i) \quad (2)$$

$$\mathbf{y}_{ij} = \mathbf{h}(\mathbf{x}_i, \mathbf{u}_i, t_{ij}, \boldsymbol{\phi}_i) + \mathbf{e}_{ij}, \quad (3)$$

where time is denoted by t . Note that the differential equation is written in differential form, which is the standard notation for SDEs that will be introduced later on. The vector-valued function $\mathbf{f}(\mathbf{x}_i, \mathbf{u}_i, t, \boldsymbol{\phi}_i)$ describes the dynamics of the system, and $\boldsymbol{\phi}_i$ denotes the individual parameters for individual i . The state variables of the system are denoted \mathbf{x}_i and may for example be the concentration of a drug or drug effect. The input to the system is denoted \mathbf{u}_i , which for example can be an infusion. Measurements are assumed to be taken at discrete points in time and characterized by the measurement function $\mathbf{h}(\mathbf{x}_i, \mathbf{u}_i, t_{ij}, \boldsymbol{\phi}_i)$ and the measurement error $\mathbf{e}_{ij} \sim N(\mathbf{0}, \mathbf{S}(\mathbf{x}_i, \mathbf{u}_i, t_{ij}, \boldsymbol{\phi}_i))$.

The individual parameters $\boldsymbol{\phi}_i$ are related to the population parameters $\boldsymbol{\theta}$ according to

$$\boldsymbol{\phi}_i = \mathbf{g}(\boldsymbol{\theta}, \mathbf{Z}_i, \boldsymbol{\eta}_i),$$

where $\boldsymbol{\theta}$ denote the fixed effect parameters, \mathbf{Z}_i denote the covariates for individual i and $\boldsymbol{\eta}_i \sim N(\mathbf{0}, \Omega)$ are the random effects for individual i , which are assumed to be multivariate normal distributed with mean zero and covariance Ω . The model described above is the commonly used NLME

model setup, and we refer the reader to (6, 7) for more information.

Instead of considering the deterministic model in the classic framework, we want to include some kind of uncertainty in the differential equations as well. This is achieved by expanding the ODEs to SDEs. The stochastic differential mixed effect model, abbreviated SDMEM (30), is defined as

$$dx_i = f(x_i, u_i, t, \phi_i)dt + \Sigma(x_i, u_i, t, \phi_i)dW_i, \quad x_i(0) = x_0(\phi_i) \quad (4)$$

$$y_{ij} = h(x_i, u_i, t_{ij}, \phi_i) + e_{ij}. \quad (5)$$

An SDE of the form (4) consists of two parts. First, we have the so-called drift function $f(x_i, u_i, t, \phi_i)$ corresponding to the deterministic part in the model, which is the same as in Eq. (2). Second, we have the random term $\Sigma(x_i, u_i, t, \phi_i)dW_i$, which corresponds to the uncertain part of the model. We will later refer to $\Sigma(x_i, u_i, t, \phi_i)dW_i$ as the system noise. The system noise is a continuous stochastic process, in contrast to the measurement noise, which is realized at discrete time points. In Eq. (4), dW_i corresponds to the increment of a q-dimensional Wiener process W_i . The elements of dW_i are independent and normally distributed with mean zero and variance dt . Moreover, the Wiener increments dW_i are considered independent across individuals and independent of the measurement error.

In contrast to the classic approach, where the only error arises in the measurement equation, an SDE setting provides a flexible framework to account for fluctuations in the underlying state variables. The system noise is a tool that accounts for all the unknown phenomena that are not captured by the deterministic model, for example approximations, modeling errors, and oversimplifications. In a mixed effects model setting, variability in response can now arise from three different sources, namely measurement noise, system noise, and parameter variability.

Parameter Estimation in the Stochastic Mixed Effects Population Framework

Given a collection of measurements of the form (2) and an underlying model of the form (4)–(5), the model parameters can be estimated using the maximum likelihood approach. This has previously been elaborated, see, e.g., (10, 25). For convenience of the reader, we here provide the necessary equations together with the extension to models with interaction between random effects and output covariance.

For a specific individual i , the optimal parameter values are found by maximizing the individual likelihood. Using the notation $Y_{ik} = [y_{i1}, y_{i2}, \dots, y_{ik}]$ to denote the measurements up to time point t_k for individual i , the combined likelihood becomes

$$L_i(\theta | Y_{in_i}) = \left(\prod_{j=2}^{n_i} p(y_{ij} | Y_{i(j-1)}, \theta) \right) p(y_{i1} | \theta)$$

where the probability for an observation given the previous observations and the parameters is $p(y_{ij} | Y_{i(j-1)}, \theta)$. Assuming

Gaussian densities, which are characterized by their first and second moments denoted by

$$\begin{aligned} \hat{y}_{ij} &= E(y_{ij} | Y_{i(j-1)}, \theta) \\ \mathbf{R}_{ij} &= Var(y_{ij} | Y_{i(j-1)}, \theta), \end{aligned}$$

we can write down the individual likelihood. Taking the logarithm, the individual log-likelihood is given by

$$\log L_i(\theta | Y_{in_i}) = -\frac{1}{2} \sum_{j=1}^{n_i} \left(\epsilon_{ij}^T \mathbf{R}_{ij}^{-1} \epsilon_{ij} + \log|2\pi\mathbf{R}_{ij}| \right),$$

where

$$\epsilon_{ij} = y_{ij} - \hat{y}_{ij},$$

is the prediction error, assumed to be normal distributed with mean $\mathbf{0}$ and variance \mathbf{R}_{ij} . We denote the collection of all individual measurements $Y = \{Y_{1n_1}, Y_{2n_2}, \dots, Y_{Nn_N}\}$. The population likelihood is simply a product of individual likelihoods,

$$L(\theta | Y) = \prod_{i=1}^N p(Y_{in_i} | \theta, \Omega)$$

Since the random effects are unobserved quantities, we marginalize over the random effects,

$$L(\theta | Y) = \prod_{i=1}^N \int p(Y_{in_i} | \theta, \eta_i) p(\eta_i | \Omega) d\eta_i = \prod_{i=1}^N \exp(l_i) d\eta_i, \quad (6)$$

where $l_i = l_i(\eta_i) = l_i(\eta_i; Y_{in_i}, \theta)$ is the *a posteriori* log-likelihood for the random effects of the i :th individual

$$l_i = -\frac{1}{2} \sum_{j=1}^{n_i} \left(\epsilon_{ij}^T \mathbf{R}_{ij}^{-1} \epsilon_{ij} + \log|2\pi\mathbf{R}_{ij}| \right) - \frac{1}{2} \eta_i^T \Omega^{-1} \eta_i - \frac{1}{2} \log|2\pi\Omega|, \quad (7)$$

In most cases, there is no closed form expression for the integral in Eq. (6). The integral can be approximated using the Laplace approximation, see (31–33). The Laplace approximation uses a second-order Taylor expansion of l_i around a point η_i^* . Here, the point is chosen to be the value of η_i which maximizes the individual log-likelihood (7),

$$\eta_i^* = \arg \max_{\eta_i} l_i(\eta_i).$$

Using this η_i^* , we end up with the approximate population likelihood function

$$L(\theta | Y) \approx \prod_{i=1}^N \exp(l_i(\eta_i^*)) \left| \frac{-\Delta l_i(\eta_i^*)}{2\pi} \right|,$$

where $\Delta l_i(\eta_i^*)$ denotes the Hessian of the individual log-likelihood (7) evaluated at the point η_i^* . Taking the logarithm, we have

$$\log L(\theta | Y) \approx \sum_{i=1}^N l_i(\eta_i^*) - \frac{1}{2} \log \left| \frac{-\Delta l_i(\eta_i^*)}{2\pi} \right|.$$

The expression for the element at index l,k of the Hessian matrix is

$$\begin{aligned} (\Delta l_i(\boldsymbol{\eta}))_{l,k} &= \frac{\partial^2 l_i(\boldsymbol{\eta})}{\partial \eta_l \partial \eta_k} \\ &\approx -\sum_{j=1}^{n_i} \frac{\partial \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1}}{\partial \eta_l} \frac{\partial \boldsymbol{\epsilon}_{ij}}{\partial \eta_k} + \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \frac{\partial \mathbf{R}_{ij}}{\partial \eta_l} \mathbf{R}_{ij}^{-1} \frac{\partial \boldsymbol{\epsilon}_{ij}^T}{\partial \eta_k} + \frac{\partial \boldsymbol{\epsilon}_{ij}^T}{\partial \eta_k} \mathbf{R}_{ij}^{-1} \boldsymbol{\epsilon}_{ij} \\ &\quad - \boldsymbol{\epsilon}_{ij}^T \mathbf{R}_{ij}^{-1} \frac{\partial \mathbf{R}_{ij}}{\partial \eta_l} \mathbf{R}_{ij}^{-1} \frac{\partial \mathbf{R}_{ij}}{\partial \eta_k} \boldsymbol{\epsilon}_{ij} - \frac{1}{2} \text{Tr} \left(-\mathbf{R}_{ij}^{-1} \frac{\partial \mathbf{R}_{ij}}{\partial \eta_l} \mathbf{R}_{ij}^{-1} \frac{\partial \mathbf{R}_{ij}}{\partial \eta_k} \right) - \Omega_{l,k}^{-1}, \end{aligned}$$

where only first-order partial derivatives are considered and higher-order contributions are assumed to be negligible in the calculation of the Hessian $\Delta l_i(\boldsymbol{\eta})$. This is called the FOCEI approximation. If we assume no interaction between the output covariance \mathbf{R}_{ij} and the random parameters, the approximate Hessian is given by

$$(\Delta l_i(\boldsymbol{\eta}))_{l,k} = \frac{\partial^2 l_i(\boldsymbol{\eta})}{\partial \eta_l \partial \eta_k} \approx -\sum_{j=1}^{n_i} \frac{\partial \boldsymbol{\epsilon}_{ij}^T}{\partial \eta_l} \mathbf{R}_{ij}^{-1} \frac{\partial \boldsymbol{\epsilon}_{ij}}{\partial \eta_k} - \Omega_{l,k}^{-1}, \quad (9)$$

referred to as the FOCE method (33). Finally, the maximum likelihood estimates are given by maximizing the approximate population likelihood as

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta} | Y). \quad (10)$$

Due to the stochastic part of the model in Eqs. (4)–(5), the state of the system is uncertain. There are several solutions to the state estimation problem in these situations, including for example particle filters and the Kalman filter (29). In this paper, we will utilize the so-called extended Kalman filter (EKF).

The Continuous Discrete Extended Kalman Filter

To calculate the individual log-likelihoods (7), we need the prediction errors $\boldsymbol{\epsilon}_{ij}$ and the output covariance matrices \mathbf{R}_{ij} . As noted in (10, 25), these identities can be recursively computed using the extended Kalman filter (EKF).

The continuous discrete EKF is a state estimator for continuous discrete state space models of the form (4)–(5) (29). From observations, the state variables of the system and their covariance are estimated in order to compute the residuals and the output covariance. From now on, we drop the individual notation i .

The EKF is an extension of the famous Kalman filter to nonlinear models (34). For linear dynamic models, the Kalman filter provides an optimal state estimator for a given parameter vector $\boldsymbol{\phi}$. For nonlinear models, the EKF uses a first-order linearization around the model trajectory. The EKF provides estimates of the conditional expectation of the state $\hat{\mathbf{x}}_{k|k} = E(\mathbf{x}_{t_k} | Y_k, \boldsymbol{\phi})$ and its covariance $\mathbf{P}_{k|k} = \text{Var}(\mathbf{x}_{t_k} | Y_k, \boldsymbol{\phi})$. Given initial conditions $\hat{\mathbf{x}}_{1|0} = \mathbf{x}_0$ and $\mathbf{P}_{1|0} = \mathbf{P}_0$ and linearizations

$$\begin{aligned} \mathbf{A}_t &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}_t} \Big|_{\mathbf{x}_t = \hat{\mathbf{x}}_{t|k}} \\ \mathbf{C}_k &= \frac{\partial \mathbf{h}}{\partial \mathbf{x}_t} \Big|_{\mathbf{x}_t = \hat{\mathbf{x}}_{k|k-1}}, \end{aligned}$$

the state variables and their covariance are predicted between two consecutive measurement time points according to

$$\begin{aligned} \frac{d\hat{\mathbf{x}}_{t|k}}{dt} &= \mathbf{f}(\hat{\mathbf{x}}_{t|k}, \mathbf{u}_t, t, \boldsymbol{\phi}), t \in [t_k, t_{k+1}] \\ \frac{d\mathbf{P}_{t|k}}{dt} &= \mathbf{A}_t \mathbf{P}_{t|k} + \mathbf{P}_{t|k} \mathbf{A}_t + \Sigma \Sigma^T, t \in [t_k, t_{k+1}]. \end{aligned}$$

From the predicted state variables and their covariance, we have the output prediction equations

$$\begin{aligned} \hat{\mathbf{y}}_{k|k-1} &= \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{u}_k, t_k, \boldsymbol{\phi}) \\ \mathbf{R}_{k|k-1} &= \mathbf{C}_k \mathbf{P}_{k|k-1} \mathbf{C}_k^T + \mathbf{S}. \end{aligned}$$

From the state covariance $\mathbf{P}_{k|k-1}$ and measurement covariance $\mathbf{R}_{k|k-1}$, the Kalman gain is given by

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{C}_k^T \mathbf{R}_{k|k-1}^{-1}.$$

Finally, the state and its covariance are updated according to

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \boldsymbol{\epsilon}_k \\ \mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{R}_{k|k-1} \mathbf{K}_k^T, \end{aligned}$$

where the residual $\boldsymbol{\epsilon}_k$ is given by

$$\boldsymbol{\epsilon}_k = \mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}.$$

Optimization of the Approximate Population Likelihood

To maximize the approximate population likelihood in Eq. (10), we have to solve a nested optimization problem. For every value of the population parameters $\boldsymbol{\theta}$ in the optimization of the approximate population likelihood, the individual likelihoods in Eq. (7) have to be maximized with respect to the random effects due to the Laplace approximation. We refer to the maximization of the individual likelihoods as the inner optimization problem and maximization of the approximate population likelihood as the outer optimization problem.

For the outer and inner optimization problems, we use a local gradient-based quasi-Newton optimization routine based on the Broyden-Fletcher-Goldfarb-Shannon (BFGS) updating formula (35). The BFGS updating formula is a popular optimization method because it performs well in many different problems.

Since the optimization methods are gradient-based, we need to calculate the gradient of the outer objective function (10) and the gradient of the inner objective function (7). We also need to calculate the approximate Hessians (8) or (9) of the inner objective function. As argued in (28), there are three approaches to this problem, namely (i) approximations based on finite differences, (ii) symbolic differentiation, or (iii) automatic differentiation tools. Instead of using finite difference approximation, as performed in (8, 10), we use

symbolic differentiation using the symbolic algebra capability in Mathematica. By using symbolic derivation, the system of ODEs is differentiated with respect to the model parameters to obtain the so-called sensitivity equations. These sensitivity equations are integrated together with the original system of ODEs to form the expression of the gradient.

Application 1: Simulation and estimation of a stochastic one-compartmental pharmacokinetic model

To illustrate the concept of SDEs explained in the previous section, we consider a one-compartment pharmacokinetic model with first-order input and nonlinear elimination. In this application, we validate the proposed stochastic modeling framework. A necessary condition is that the model parameters and the three sources of variability (measurement error, population variability, and model uncertainty) can be identified. Here, we consider data from a simulated population consisting of 20 animals. Moreover, we are interested in investigating the difference in parameter estimates with an assumption of a deterministic model. Consider the deterministic mixed effects pharmacokinetic model

$$\begin{aligned} \frac{dA_i}{dt} &= -k_{ai}A_i, & A_i(0) &= 20 \\ V\frac{dC_i}{dt} &= k_{ai}A_i - \frac{V_{mi}}{K_m + C_i}C_i, & C_i(0) &= 0, \end{aligned}$$

where A_i (mg) and C_i (mg L⁻¹) are the amount of drug in the GI tract and the concentration of drug in plasma for individual i , respectively. We assume that the k_{ai} and V_{mi} are multivariate log-normal distributed, that is

$$\begin{aligned} k_{ai} &= k_a \exp(\eta_{i1}), \\ V_{mi} &= V_m \exp(\eta_{i2}), \end{aligned}$$

where the random effects vector $\boldsymbol{\eta} = (\eta_{i1}, \eta_{i2})$ is assumed to follow a multivariate normal distribution with mean zero and covariance matrix Ω . The parameters of the model are the k_a (min⁻¹), V_m (mg min⁻¹), K_m (mg L⁻¹), and V (L). Moreover, we parameterize $\Omega = \mathbf{U}\mathbf{U}^T$, where \mathbf{U} is an upper triangular matrix of the form

$$\mathbf{U} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ 0 & \omega_{22} \end{pmatrix}.$$

The reason for the parameterization $\Omega = \mathbf{U}\mathbf{U}^T$ is to assure that Ω is a positive definite matrix, a necessary condition since it is a covariance matrix. We get

$$\Omega = \begin{pmatrix} \omega_{11}^2 + \omega_{12}^2 & \omega_{12}\omega_{22} \\ \omega_{12}\omega_{22} & \omega_{22}^2 \end{pmatrix}.$$

Using additive system noise, we end up with an SDE describing the concentration of drug.

$$VdC_i = \left(k_{ai}A_i - \frac{V_{mi}}{K_m + C_i}C_i \right) dt + \sigma dW_i, \quad C_i(0) = 0. \quad (11)$$

In the stochastic model, there exists a system noise σdW_i , where σ is the scaling factor and dW_i is the increment of a standard Wiener process. In Eq. (11), the system noise is independent of the drug concentration. This may not be a realistic assumption since it allows for a change in concentration even in the absence of drug. By defining the system noise dependent on the concentration level itself, such phenomena can be avoided. The final model, which we consider for simulation and estimation, is

$$\frac{dA_i}{dt} = -k_{ai}A_i, \quad A_i(0) = 20 \quad (12)$$

$$VdC_i = \left(k_{ai}A_i - \frac{V_{mi}}{K_m + C_i}C_i \right) dt + \sigma \hat{C}_i dW_i, \quad C_i(0) = 0, \quad (13)$$

where σ (L min⁻¹) is the system noise factor. In the SDE above, \hat{C}_i denotes the approximation of the conditional expectation of the concentration C_i computed by the EKF. The reason for this model is that the EKF does not allow for state-dependent system noise. However, by utilizing the approximation of the conditional expectation of the concentration, we can still use a proportional-like system noise. For SDEs with system noise dependent on the stochastic process itself, one can in rare cases use the Lamperti transform to obtain a state-independent system noise. The SDE (12)–(13) has a stochastic part in the equation describing the central compartment. Since dW_i is the increment of a Wiener process, this implies that dW_i is normally distributed with variance dt , which in turn implies that $\sigma \hat{C}_i dW_i$ is normally distributed with variance $\sigma^2 \hat{C}_i^2 dt$.

Mass-balance constrains for SDE models can be enforced by assign the structure of Σ according to the stoichiometry of the modeled system. For example, accounting for uncertainty in the drug uptake rate from the GI tract can be achieved by adding the same system noise term to the Eq. (12) with opposite sign. However, we have chosen to assume that the system noise describes an uncertainty in the elimination process from the central compartment, requiring no mass-balancing.

Furthermore, the measurement y_{ik} is the concentration for individual i measured at time t_{ik} under additive Gaussian noise according to

$$y_{ik} = C_i(t_{ik}) + e_{ik}, \quad (14)$$

where $e_{ik} \sim N(0, s^2)$. A population consisting of 20 animals is simulated according to the parameter values in the third column of Table I. A total of 100 data sets are simulated, and the model parameters are estimated using the FOCEI method. The response is measured at equidistant time points $t_k = 1, 9, 17, \dots, 97$ minutes for all animals.

Application 2: Stochastic NiAc disposition in obese Zucker rats

In this section, we extend a pharmacokinetic model of nicotinic acid (NiAc) in obese Zucker rats, previously used to drive a pharmacodynamic model describing nonesterified fatty acid (NEFA) turnover (36–41). The disposition of NiAc

Table I. Estimated parameter values for the one-compartmental model (12)–(13) using the ODE and the SDE model

Parameter	Definition	True value	Starting value	ODE model (RSE %)	SDE model (RSE %)
k_a	First-order absorption	0.1	0.2	0.078 (22.6)	0.103 (12.9)
V_m	Maximal velocity	0.5	1	0.637 (20.3)	0.508 (9.52)
K_m	Michaelis-Menten const.	3	1	5.43 (51.3)	3.04 (14.9)
V	Compartmental volume	1	2	0.793 (17.6)	1.00 (5.50)
s	Measurement error std.	0.1	0.5	0.609 (17.2)	0.099 (7.49)
ω_{11}	Interindividual variability k_a	0.5	0.1	0.403 (22.8)	0.456 (18.4)
ω_{12}	Interindividual correlation	0.1	0	0.098 (120.9)	0.092 (149.5)
ω_{22}	Interindividual variability V_m	0.3	0.1	0.332 (18.3)	0.279 (18.9)
σ	System noise factor	0.05	0.01	-	0.050 (6.56)

The relative standard errors (RSEs) in % are included in parenthesis

in obese rats was described by a one-compartment model with endogenous synthesis of NiAc and a capacity-limited elimination process. Instead of using a deterministic model for the pharmacokinetics, we consider the extension to a stochastic NiAc disposition model.

Original NiAc Disposition Model

The previously used pharmacokinetic model was a one-compartmental model with a synthesis rate $Synt$ ($\mu\text{mol min}^{-1} \text{kg}^{-1}$) of NiAc in the absence of drug with a nonlinear elimination with parameters V_{mi} ($\mu\text{mol min}^{-1} \text{kg}^{-1}$) and K_m ($\mu\text{mol L}^{-1}$) describing the maximal rate from the central compartment and the Michaelis-Menten constant, respectively. The mixed effects kinetics is described by the ODE

$$V_c \frac{dc_i}{dt} = u + Synt - \frac{V_{mi}c_i}{K_m + c_i}, \quad (15)$$

$$c_i(0) = \frac{Synt K_m}{V_{mi} - Synt}, \quad (16)$$

where u ($\mu\text{mol min}^{-1} \text{kg}^{-1}$) denotes the input. Two different infusion rates were used, $0.67 \mu\text{mol min}^{-1} \text{kg}^{-1}$ (corresponding to $20 \mu\text{mol kg}^{-1}$ over 30 min) and $0.17 \mu\text{mol min}^{-1} \text{kg}^{-1}$ (corresponding to $51 \mu\text{mol kg}^{-1}$ over 300 min).

In previous work, in addition to V_{mi} , $Synt$ was allowed to be distributed in the population. In this paper, we only consider V_{mi} to be distributed in the population. This choice was due to the fact that few data samples were taken at steady-state and that the estimation results in (39) showed a very high residual standard error on the interindividual variability parameters. We consider a log-normal distribution of the maximal rate V_{mi} . That is

$$V_{mi} = V_m \exp(\eta_i),$$

where $\eta_i \sim N(0, \omega^2)$. V_c (L kg^{-1}) denotes the central volume. The first group consisted of eight subjects and the second group of seven subjects. There was also one subject receiving placebo, giving a total of 16 subjects in the analysis.

Stochastic NiAc Disposition Model

We consider the extension to a stochastic NiAc model described by the SDE

$$V_c dc_i = \left(u + Synt - \frac{V_{mi}c_i}{K_m + c_i} \right) dt + \sigma \hat{c}_i dW_i \quad (17)$$

In the stochastic NiAc disposition model, the system noise $\sigma \hat{c}_i dW_i$ models the uncertainty in dynamics. Again, we assume a system noise proportional to the mode \hat{c}_i . We refer to σ ($\text{L min}^{-1} \text{kg}^{-1}$) as the system noise factor. In contrast to the original NiAc disposition model, the total error is now divided into measurement error and system noise. The choice of the stochastic model structure is to allow for an uncertainty in the elimination process dependent on the drug concentration.

The purpose of the extension is to identify the structural parameters together with the sources of variability. Most importantly, we are interested in identifying the system noise factor σ . Note that $\sigma=0$ corresponds to the original NiAc disposition model.

RESULTS

Application 1: Simulation and estimation of a stochastic one-compartmental pharmacokinetic model

In this example, the primary interest lies in how well the parameters in the model can be estimated from data. Since we have three sources of variability (parameter variability, measurement error, and system noise), it is important to know whether these sources of variability can be separated in the estimation.

The parameters in the model are estimated using the FOCEI method. The parameters in the model consist of the structural parameters k_a , V_m , K_m and V and the parameters s , ω_{11} , ω_{12} , ω_{22} and σ describing the three sources of variability. Hence, our vector of model parameters is $\theta=(k_a, V_m, K_m, V, s, \omega_{11}, \omega_{12}, \omega_{22}, \sigma)$ which gives a total number of 9 parameters to be estimated.

We estimate the model parameters using the SDE approach and compare this to the corresponding ODE model ($\sigma=0$). The estimated parameter values using the SDE model and the ODE model are shown in Table I. The relative

standard errors (RSEs) in % are included in parentheses. They are obtained by calculating the mean and standard deviations of the 100 estimates. The distribution of the estimated model parameters for the 100 simulated data sets are seen in Fig. 1.

Application 2: Stochastic NiAc disposition in obese Zucker rats

The measured NiAc concentrations for the two infusion groups are shown in Fig. 2.

Estimated Parameters

The parameters are estimated using the FOCE approximation of the individual Hessians. The reason for using the FOCE approximation is to guarantee that the individual Hessians are positive definite. The starting values for the structural parameters were adopted from Ahlström et al. (39) with values $V_m=1.8 \mu\text{mol min}^{-1} \text{kg}^{-1}$, $K_m=23 \mu\text{mol L}^{-1}$, $V_c=0.319 \text{ L kg}^{-1}$, and $Synt=0.00125 \mu\text{mol min}^{-1} \text{kg}^{-1}$. Moreover, the starting values for the variance components were $s=0.1$, $\omega=0.1$, and $\sigma=0.01 \text{ L min}^{-1}$ (for the SDE model).

Table II shows the estimated parameters for the two models of interest. As a reference, we also provide the estimates from Ahlström et al. (39). The relative standard errors (RSEs) in % are included in parentheses, calculated from the approximated Hessian at the optimum. Two comparisons are of interest. First, we used the original NiAc

disposition model (15) to compare the results from our estimation with the results from (39). Second, we are interested in the differences in parameter estimates using the original NiAc disposition model (15) and the stochastic NiAc disposition model (17).

Fitted Population and Individual Models

The fitted population models for the two approaches are illustrated in Fig. 2.

Given the estimated population parameters as priors, the individual likelihoods are maximized once again to obtain the *maximum a posteriori* estimates for the random effect parameters. These optimal parameter values are then inserted in the model equations to obtain the individual model fits. The original NiAc disposition model fit is obtained by simply solving the ODE describing NiAc concentration given an individual's parameter values. For the stochastic NiAc disposition model, the individual fit is slightly more complicated to obtain. Due to the stochastic component, the individual model fits for the stochastic model are obtained by a method called smoothing. This has previously been demonstrated by Kristensen et al. (2). When smoothing is used, the model is used to provide an optimal state estimator given all the measurements for a specific individual. The fitted individual models together with the output uncertainties are illustrated in Fig. 3 for three animals (rows 1–3) with the shorter infusion for the estimated ODE (a–c) and

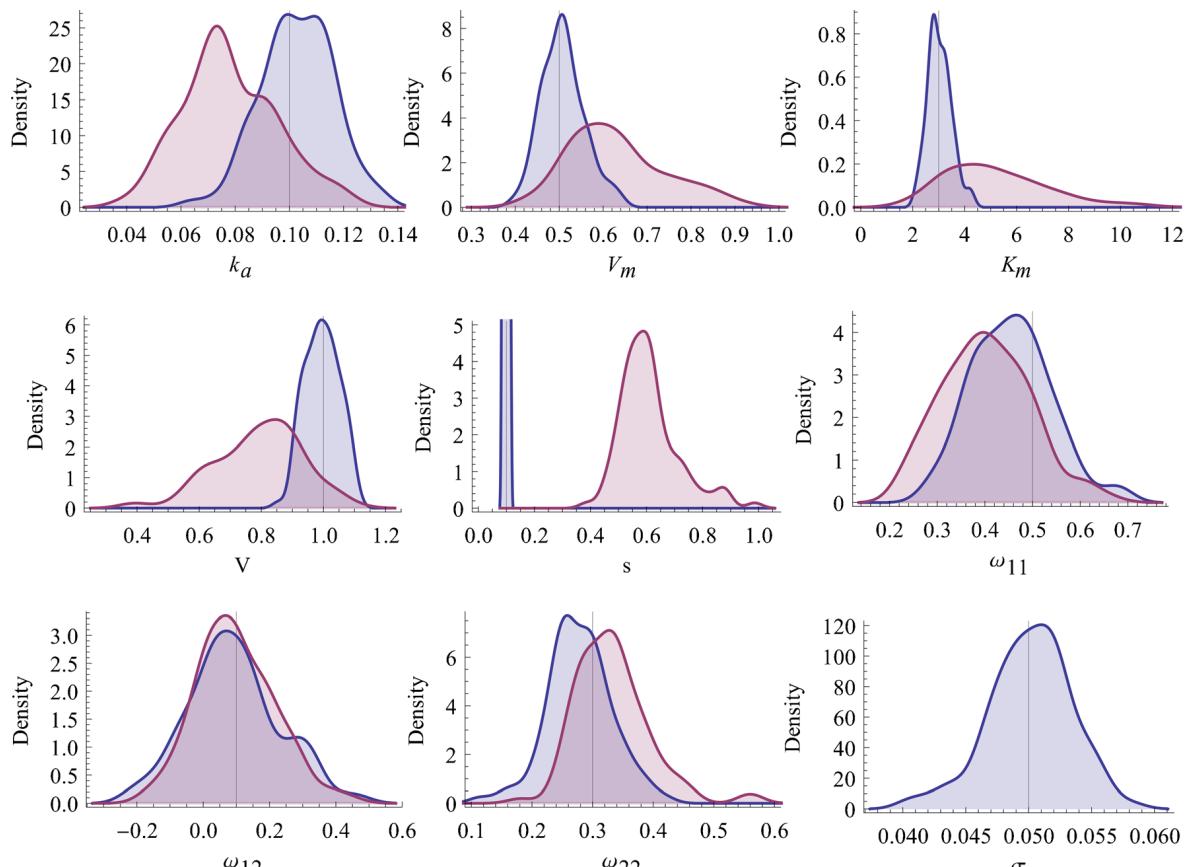


Fig. 1. Smoothed histograms over the estimated parameters from 100 simulated data sets. The estimated parameters using the SDE model is shown in blue and the estimates using the ODE model ($\sigma=0$) is shown in purple. The vertical lines show the parameter values used for simulation

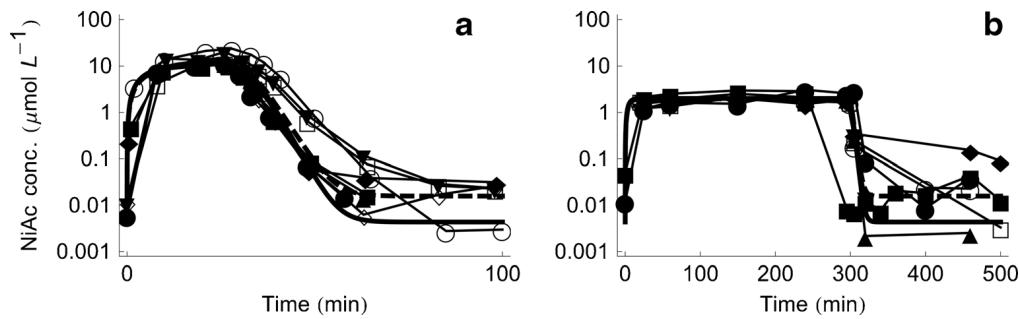


Fig. 2. Plots of the estimated ODE (solid) and SDE (dashed) NiAc model together with the observed concentration time courses of NiAc for the two infusion groups. **a** $20 \mu\text{mol kg}^{-1}$ over 30 min. **b** $51 \mu\text{mol kg}^{-1}$ over 300 min over 300 min. The concentration is shown on a log-linear scale

SDE (d-f) NiAc disposition model. The uncertainty bands represent one standard deviation of the output uncertainty (that is, the square root of the output covariance). Note that the concentration and its uncertainty are visualized on a linear scale, in contrast to the previous plots in Fig. 2, to emphasize the improved fit for large NiAc concentrations using SDEs.

DISCUSSION

The extension of NLME models to stochastic differential mixed effect models has been considered to provide a more general model to describe the error between model prediction and observed data. By utilizing the stochastic setting, the total error is divided into measurement noise and model uncertainty. Together with the population variability induced in a mixed effect model, we are able to account for a total of three sources of variability.

Maximum Likelihood Estimation

We have taken the maximum likelihood approach to parameter estimation by combining an approximation of the population likelihood together with an extended Kalman filter for state estimation. In contrast to the estimation method proposed in (10, 23–25), we have further developed the method by considering interaction between the output covariance and random effects, referred to as the FOCEI method (33). The interaction between output covariance and random effects occurs for example in ODE models with a proportional measurement noise and in stochastic models

where the output covariance depends on the Kalman update, which in turn depends on the individual response.

The parameters were estimated using the gradient-based method BFGS (35). In many applications, the gradient of the objective function is approximated using finite differences. However, numerical ODE solvers with an adaptive step length are known to introduce quantification errors to the objective function, making it nonsmooth on small scales (42). To overcome such problems, we utilized the sensitivity equations when calculating the gradient in the inner and outer optimization problem. The sensitivity equations were obtained by differentiating the system of differential equations and the extended Kalman filter equations with respect to the parameters in the model. This has been previously demonstrated by Leander *et al.* (11) in the single individual case, and we are preparing a manuscript that concerns the mixed effects case.

The extension to SDEs comes with an increased computational burden. Due to the fact that we now consider equations describing the time evolution of both mean and covariance of a stochastic process, a larger system of ODEs has to be solved. The implementation of the numerical machinery for parameter estimation also becomes more challenging and requires advanced numerical techniques such as the EKF. The stochastic mixed effects modeling framework has been implemented in Mathematica 9. An executable version of the code may be received from the authors upon request.

Application 1: Simulation and estimation in the simulated stochastic one-compartmental pharmacokinetic model

As a first application of stochastic mixed effects modeling, we used a stochastic one-compartmental pharmacokinetic

Table II. Estimated parameter values and interindividual variability (IIV) for the NiAc disposition model, with corresponding relative standard errors (RSE %)

Parameter	Definition	Ahlström <i>et al.</i>	Current investigation: ODE model	Current investigation: SDE model
V_m	Maximal velocity	1.59 (13.9)	1.46 (16.3)	1.35 (16.7)
K_m	Michaelis-Menten const.	18.9 (21.5)	15.2 (21.7)	13.6 (21.5)
V_c	Central volume	0.328 (12.4)	0.29 (4.3)	0.32 (5.5)
$Synt$	Endogenous synthesis rate	0.00280 (10.1)	0.0006 (29.5)	0.0018 (24.3)
s	Residual prop. error	0.400 (26.3)	0.460 (8.08)	0.241 (11.7)
ω	Variability V_m	0.214 (234)	0.174 (22.5)	0.133 (27.0)
σ	System noise factor	-	-	0.033 (15.7)

See Ahlström *et al.* (39) for reference

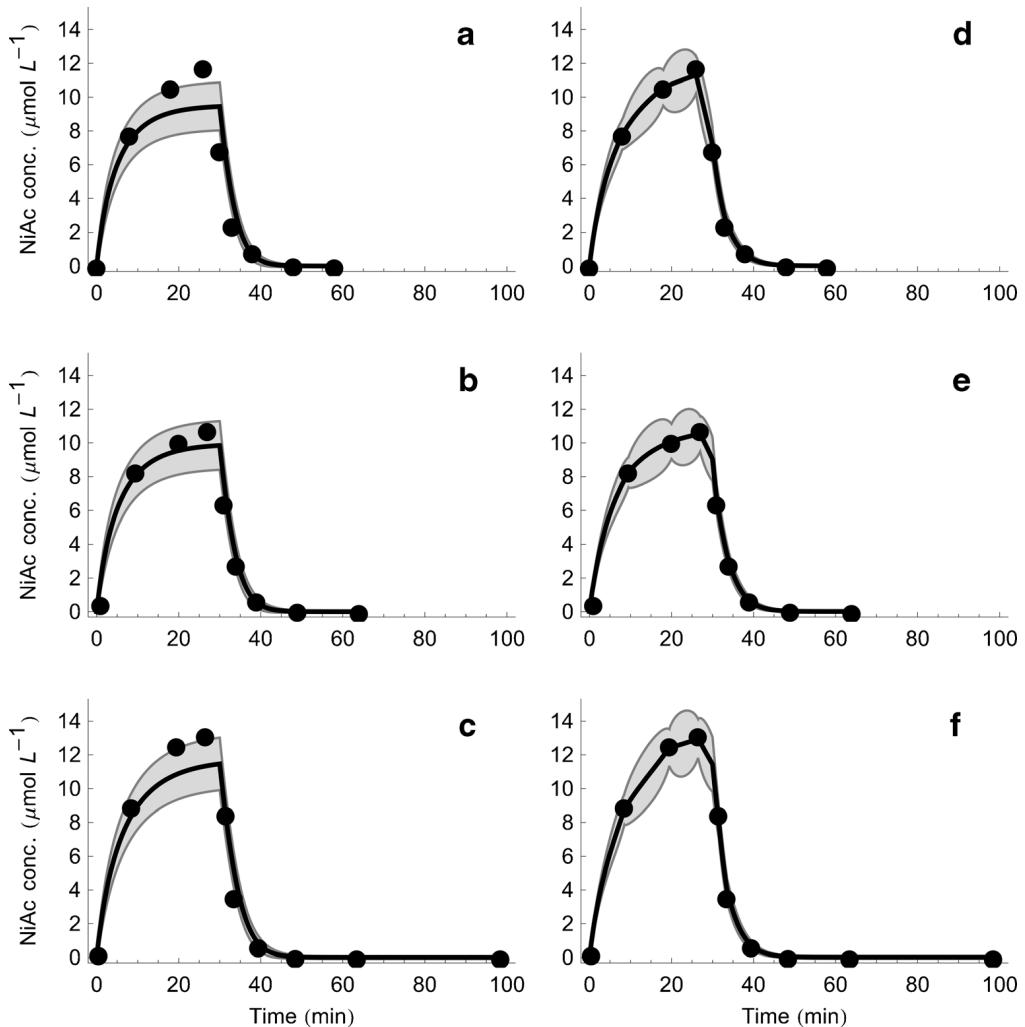


Fig. 3. Observed plasma NiAc concentration time profiles together with the estimated ODE (a–c) and SDE (d–f) NiAc disposition model for three animals (each row) from the first infusion group ($20\mu\text{mol kg}^{-1}$ over 30 min)

model for a simulation study. The aim was to investigate whether the three sources of variability (measurement error, population variability, and model uncertainty) could be separated using the proposed maximum likelihood estimation.

From the parameter estimates for the SDE model in Table I and the smoothed histograms in Fig. 1, we conclude that all estimated parameter values are close to their true values used for simulation. The measurement error standard deviation s and the system noise factor σ are identical to the true values using the SDE approach. Also, the correlation between the random effects is close to the true value for the SDE model. However, the RSE for the random effect correlation is very high, implying that the correlation is difficult to reliably estimate. This may depend on the fact that only 20 individuals were included in each data set.

In the ODE case, we conclude that all the structural parameters are biased and differ significantly from the true parameter values used for simulating the data. Comparing the results to the ODE case where the system noise is neglected ($\sigma=0$), we can see that the measurement error standard deviation s is increased six-fold. Hence, we conclude that when the system noise is set to zero (equivalent to the ODE case), the measurement error standard deviation is increased

to account for that variability, which is not unexpected since we neglect the variability in dynamics. We also conclude that the RSEs for the SDE model are generally lower than the RSEs for the ODE model, although the RSEs for ω_{12} and ω_{22} are slightly higher in the SDE model.

Most importantly, we conclude that we can successfully distinguish the three sources of variability, which is a necessary condition of the extended framework to be of practical value.

Application 2: Stochastic NiAc disposition in obese Zucker rats

Using the original NiAc disposition model, and comparing the results with previous work, we can conclude that most of the parameter values are similar. Our estimated values differ most in K_m and $Synt$. Note that the quotients V_m/K_m are similar in the work by Ahlström *et al.* (39) and in the current investigation (0.084 and 0.096, respectively). This may imply that there are problems estimating the parameters V_m and K_m , whereas the quotient as such is identifiable. One striking difference between our estimation and previous results is that the relative standard error for the parameter describing the interindividual variability (IIV) is significantly

lower in our investigation. The IIV for V_m was 234% in (39), while it is reduced to 22.5% in our investigation. This is most likely because instead of using finite difference approximation of the gradient in the optimization problem, we utilize sensitivity equations, yielding a more robust calculation of the gradient, which also seems to influence the precision of parameter estimates.

If we turn to our investigation and compare the ODE and the SDE NiAc disposition models, we can see that the endogenous synthesis rate $Synt$ is increased for the stochastic model, which is also seen in the population model fits in Fig. 2. Moreover, another difference in the parameter estimates is that the measurement error is much lower (two-fold) for the stochastic model whereas the system noise factor significantly differed from zero. This implies that the error that existed in the original model (measurement error) is now separated into two parts, namely the measurement error and the system noise. The interindividual variability for the maximal rate V_m is slightly decreased when the stochastic NiAc disposition model is used. Hence, some of the population variability in the maximal rate V_m seen in the original NiAc disposition model may instead be explained by a model uncertainty.

With respect to the individual fits in Fig. 3, there is a clear difference in terms of model fits. The original NiAc disposition model, proposed in (39), seems to underestimate the drug concentration during the infusion. The stochastic model that we propose is much closer to the measurements and seems to account for model deviation that the original model is not capable of. In the original NiAc disposition model, the output covariance is equal to the measurement covariance, whereas in the stochastic model, it is a combination of the state covariance and measurement covariance.

That means that the confidence band in the original model simply arises from the variance of measurements, which we assumed to be proportional to the concentration level. Using the stochastic model, we conclude that the uncertainty is highest between two consecutive time points. In contrast to the original model, we have a decreased uncertainty at the measurements, at which information is gained about the underlying system.

In previous work (39), the original NiAc disposition model was used to drive a pharmacodynamic model describing production of NEFA. By utilizing the stochastic NiAc disposition model, the fitted individual models seem to be able to capture the high concentrations during the infusion.

This is seen in Fig. 3 and may give a better input to the NEFA model. A better input to a pharmacodynamic model can be of broader interest in PKPD modeling, since a deterministic pharmacokinetic model often is used to drive a pharmacodynamic model. Using a stochastic pharmacokinetic model can better account for uncertainty in the drug kinetics.

CONCLUSIONS

We conclude that the stochastic modeling framework we proposed here leads to a more general framework for handling measurement error and model errors. This framework, together with an effective method for calculating the gradients in the nested optimization problem, provides us with a flexible, robust modeling framework for mixed effects models.

ACKNOWLEDGMENTS

This project has partly been supported by the Swedish Foundation for Strategic Research, which is gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

REFERENCES

1. Gabrielsson J, Weiner D. Pharmacokinetic and pharmacodynamic data analysis: concepts and applications. 4th ed. Swedish Pharmaceutical Press; 2006.
2. Kristensen NR, Madsen H, Ingwersen SH. Using stochastic differential equations for PK/PD model development. *J Pharmacokinet Pharmacodyn*. 2005;32(1):109–41.
3. Sheiner L, Beal S. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: Routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm*. 1980;8(6):553–71.
4. Sheiner L, Beal S. Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: Routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm*. 1983;11(3):303–19.
5. Sheiner L, Rosenberg B, Marathe V. Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokinet Biopharm*. 1977;5(5):445–479.
6. Davidian M, Giltinan DJ. Nonlinear models for repeated measurement data. 1st ed. Chapman & Hall; 1995.
7. Lindstrom MJ, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990;46(3):673–87.
8. Berglund M, Sunnåker M, Adiels M, Jirstrand M, Wennberg B. Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations. *Math Med Biol*. 2012;29(4):361–84.
9. Donnet S, Foulley JL, Samson A. Bayesian analysis of growth curves using mixed effects models defined by stochastic differential equations. *Biometrics*. 2010;66(3):733–41.
10. Overgaard R, Jonsson N, Tornøe C, Madsen H. Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm. *J Pharmacokinet Pharmacodyn*. 2005;32(1):85–107.
11. Leander J, Lundh T, Jirstrand M. Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements. *Math Biosci*. 2014;251:54–62.
12. Ramanathan M. An application of Ito's lemma in population pharmacokinetics and pharmacodynamics. *Pharm Res*. 1999;16(4):584–6.
13. Ramanathan M. A method for estimating pharmacokinetic risks of concentration-dependent drug interactions from preclinical data. *Drug Metab Dispos*. 1999;27(12):1479–87.
14. Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys*. 1976;22:403–34.
15. Qi X. A stochastic version of corticosteroid pharmacogenomic model. *AAPS J*. 2005;7(1):E134–40.
16. Siegel RA, Ramanathan M. Commentary: stochastic phenomena in pharmacokinetic, pharmacodynamic, and pharmacogenomic models. *AAPS J*. 2005;7(1):E141–2.
17. Klebaner FC. Introduction to stochastic calculus with applications. Imperial College Press; 2005.
18. Øksendal B. Stochastic differential equations. An Introduction with Applications. 6th ed. Springer; 2003.
19. Black F, Scholes M. The pricing of options and corporate liabilities. *J Polit Econ*. 1973;81(3):637–54.

20. Plan E, Maloney A, Mentré F, Karlsson MO, Bertrand J. Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for dose-response models. *AAPS J.* 2012;14(3):420–32.
21. Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS (Statistics and Computing). Springer; 2013.
22. Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of the EM algorithm. *Ann Stat.* 1999;27(1):94–128.
23. Klim S, Bousgaard Mortensen S, Kristensen NR, Overgaard RV, Madsen H. Population stochastic modelling (PSM)—an R package for mixed-effects models based on stochastic differential equations. *Comput Methods Prog Biomed.* 2009;94(3):279–89.
24. Mortensen S, Klim S, Dammann B, Kristensen N, Madsen H, Overgaard R. A matlab framework for estimation of NLME models using stochastic differential equations: applications for estimation of insulin secretion rates. *J Pharmacokinet Pharmacodyn.* 2007;34(5):623–42.
25. Tornøe C, Overgaard R, Agersø H, Nielsen H, Madsen H, Jonsson E. Stochastic differential equations in NONMEM: implementation, application, and comparison with ordinary differential equations. *Pharm Res.* 2005;22(8):1247–58.
26. Delattre M, Lavielle M. Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Stat Interface.* 2013;6:519–32.
27. Donnet S, Samson A. A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. *Adv Drug Deliv Rev.* 2013;65(7):929–39.
28. Picchini U, Ditlevsen S. Practical estimation of high dimensional stochastic differential mixed-effects models. *Comput Stat Data Anal.* 2011;55(3):1426–44.
29. Jazwinsky AH. Stochastic processes and filtering theory. Dover; 1970.
30. Picchini U, De Gaetano A, Ditlevsen S. Stochastic differential mixed-effects models. *Scand J Stat.* 2010;37(1):67–90.
31. Jonsson NE, Karlsson MO, Wade JR. Nonlinearity detection: advantages of nonlinear mixed-effects modeling. *AAPS Pharm Sci.* 2000;2(3):114–23.
32. Vonesh E. A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika.* 1996;83(2):447–52.
33. Wang Y. Derivation of various NONMEM estimation methods. *J Pharmacokinet Pharmacodyn.* 2007;34(5):575–93.
34. Kalman RE. A new approach to linear filtering and prediction problems. *Trans ASME J Basic Eng.* 1960;82(Series D):35–45.
35. Nocedal J, Wright S. Springer-Verlag New York: Numerical optimization; 1999.
36. Ahlström C. Modelling of tolerance and rebound in normal and diseased rats [PhD Thesis]. University of Gothenburg; 2011.
37. Ahlström C, Peletier LA, Gabrielsson J. Quantitative analysis of rate and extent of tolerance of biomarkers: application to nicotinic acid-induced changes in non-esterified fatty acids in rats. *Eur J Pharm Sci.* 2011;44(3):250–64.
38. Ahlström C, Peletier LA, Jansson-Löfmark R, Gabrielsson J. Feedback modeling of non-esterified fatty acids in rats after nicotinic acid infusions. *J Pharmacokinet Pharmacodyn.* 2011;38(1):1–24.
39. Ahlström C, Kroon T, Peletier LA, Gabrielsson J. Feedback modeling of non-esterified fatty acids in obese Zucker rats after nicotinic acid infusions. *J Pharmacokinet Pharmacodyn.* 2013;40(6):623–38.
40. Ahlström C, Peletier LA, Gabrielsson J. Challenges of a mechanistic feedback model describing nicotinic acid-induced changes in non-esterified fatty acids in rats. *J Pharmacokinet Pharmacodyn.* 2013;40(4):497–512.
41. Tapani S, Almquist J, Leander J, Ahlström C, Peletier LA, Jirstrand M, Gabrielsson J. Joint Feedback Analysis Modeling of Nonesterified Fatty Acids in Obese Zucker Rats and Normal Sprague–Dawley Rats after Different Routes of Administration of Nicotinic Acid. *J Pharm Sci.* 2014;103(8):2571–84.
42. Bohlin T. Practical Grey-box Process Identification: Theory and Applications. Springer-Verlag; 2006.