Driving the Model to Its Limit: Profile Likelihood Based Model Reduction

Tim Maiwald[1☯], **Helge Hass**[1☯], **Bernhard Steiert**[1☯], **Joep Vanlier**[1], **Raphael Engesser**[1], **Andreas Raue**[2], **Friederike Kipkeew**[3], **Hans H. Bock**[3], **Daniel Kaschek**[1], **Clemens Kreutz**[1,4], **Jens Timmer**[1,4,5] *

1 Institute of Physics, University of Freiburg, Freiburg im Breisgau, Germany, **2** Merrimack Pharmaceuticals, Boston, MA, United States of America, **3** Department of Gastroenterology, Hepatology and Infectiology, University Hospital Duesseldorf, Duesseldorf, Germany, **4** Center for Biosystems Analysis (ZBSA), University of Freiburg, Freiburg im Breisgau, Germany, **5** BIOSS Centre for Biological Signalling Studies, University of Freiburg, Freiburg im Breisgau, Germany

☯ These authors contributed equally to this work.
* jeti@fdm.uni-freiburg.de

Abstract

In systems biology, one of the major tasks is to tailor model complexity to information content of the data. A useful model should describe the data and produce well-determined parameter estimates and predictions. Too small of a model will not be able to describe the data whereas a model which is too large tends to overfit measurement errors and does not provide precise predictions. Typically, the model is modified and tuned to fit the data, which often results in an oversized model. To restore the balance between model complexity and available measurements, either new data has to be gathered or the model has to be reduced. In this manuscript, we present a data-based method for reducing non-linear models. The profile likelihood is utilised to assess parameter identifiability and designate likely candidates for reduction. Parameter dependencies are analysed along profiles, providing context-dependent suggestions for the type of reduction. We discriminate four distinct scenarios, each associated with a specific model reduction strategy. Iterating the presented procedure eventually results in an identifiable model, which is capable of generating precise and testable predictions. Source code for all toy examples is provided within the freely available, open-source modelling environment Data2Dynamics based on MATLAB available at http://www.data2dynamics.org/, as well as the R packages dMod/cOde available at https://github.com/dkaschek/. Moreover, the concept is generally applicable and can readily be used with any software capable of calculating the profile likelihood.

Introduction

Mathematical models represent abstractions of reality that aid to understand a phenomenon of interest. Any model is a simplification, which manifests itself as some features being over-represented, while others are absent. It is important to balance model complexity with experimental observation, such that only those aspects of reality on which the observed data has

meaningful information are included in the model. Independent of whether modelling is performed by shrinking of an oversized model that was taken as starting point, or by adding complexity to a core model until the data is well-described, the result is usually a model that is larger than it would need to be. In turn, over-parameterisation in light of limited data is expressed through non-identifiability of parameters. On the one hand, this strongly restricts the applicability of statistical and numerical approaches to a small subset of computationally demanding modelling techniques. On the other hand, non-identifiable parameters often lead to imprecise model predictions. Therefore, including model reduction as an additional step in the modelling process ameliorates lack of predictability and restricts model complexity to an appropriate level.

A variety of methods have been published for reducing the complexity of models. In the following, we summarise relevant ideas and concepts available in the literature, as well as the most related methods to the one proposed here.

In the context of dynamic reaction networks, most methods intend elimination of irrelevant time scales. The *Methods of Invariant Manifolds (MIM)* [1, 2] originate from chaos theory and are based on the fact that after a transient phase, the dynamics usually reduce to low-dimensional manifolds in the phase space. However, the mechanistic interpretation of the dynamic states is lost if the state space is redefined in terms of such manifolds. Furthermore, identification of such manifolds requires detailed knowledge about the model parameters, which is usually not available in system biology. Another class of methods is based on the detection of *quasi steady-states (QSS)* [3–6] requiring again prior knowledge about parameter values, which is typically not available or unreliable. These methods reduce complexity by removing states that are approximately constant and/or reactions that are extremely fast. In [7], another method has been introduced which simplifies complex rational rate equations. The paper suggests further reduction steps by removing reaction rates with negligible fluxes as well as redundant terms in the ordinary differential equations (ODEs).

The aforementioned approaches are applied without direct use of experimental data. Therefore, these approaches rely on prior knowledge concerning parameter values, as well as ignoring both parameter uncertainties and the fact that some parameters can compensate the effect of others on observed model predictions. Alternative approaches have been developed which, like the method we suggest, reduce the complexity up to a level which is specified by available experimental data. These methods typically reduce a large model iteratively until a desired level of complexity is obtained. In statistics, this strategy is known as *backward elimination*. A well-established methodology for backward elimination is available in statistical literature for models where analytical mathematical solutions are available, e.g. for linear regression models as summarised e.g. in [8], or for machine learning applications as discussed in [9]. Methods for its application to non-linear models are briefly discussed in the following.

In the non-linear setting, the likelihood can be approximated by the Hessian when a sufficiently large amount of data is incorporated into the estimation process. This requirement however, is usually strongly violated for ODE models and it has been shown that the Hessian typically only provides an unreliable indication of parameter identifiability. Therefore, confidence intervals based on such an approach are typically inappropriate [10]. Nevertheless, methods have been suggested for removing parameters which appear non-identifiable based on evaluation of the Hessian [11]. An additional aspect emerging for non-linear models, namely the existence of local minima, was addressed in [11, 12]. In both publications, candidates for a backward elimination procedure were suggested. These were selected based on the range of parameter estimates over different local minima. However, both approaches require an arbitrary threshold for defining the local minima of interest and are only able to perform

single parameter eliminations, thereby missing model reduction steps that involve a combination of several coupled parameters.

In [13], the optimal reduction is obtained by fitting the dynamics of the reduced system for a pre-specified magnitude of reduction, e.g. for given number of remaining fluxes, to the dynamics of the full model. In this way, the local linearisation of the model, which corresponds to a quadratic approximation of the likelihood, is not required. However, the method again does not account for parameter uncertainties as well as for the availability of data.

The *Manifold Boundary Approximation Method (MBAM)* has been presented in [14] and exploits the fact that eigenvalues of the Hessian matrix range over several orders of magnitude. Starting from the direction with the smallest eigenvalue, a geodesic path is calculated until the Hessian becomes singular which corresponds to a boundary in the model prediction space. However, the initial step of this method approximates the likelihood by a quadratic function. The boundaries in the model prediction space obtained via geodesic paths typically involve arbitrary parameter combinations and therefore hampers mechanistic interpretation of a model reduction step and of the resulting reduced model.

In [15], backward elimination was discussed by defining *admissible regions*. Like in the standard backward elimination setting, the model is iteratively reduced by applying a heuristic strategy for testing whether parameters can be set to the bounds of the parameter space, e.g. to zero. This approach is conceptionally the most similar one to our method based on the profile likelihood. One advantage of our method is that it provides a clear strategy for determining which parameters have to be tested for reduction. In addition, inferring model reduction steps involving re-parameterisations of several coupled parameters are feasible.

In summary, the method introduced in the following does not rely on prior knowledge or likelihood approximation by the Hessian. Moreover, it takes into account parameter uncertainties given model and data and features a clear strategy for determining relationships of multiple parameters while preserving a mechanistic interpretation of the reduced model.

Methods

The underlying idea

Our method can be motivated by a prominent example from systems biology: consider the conversion

$$S + E \underset{k_-}{\overset{k_+}{\rightleftharpoons}} C \overset{k_2}{\rightarrow} P + E \tag{1}$$

of a substrate S into a product P by an enzyme E via formation of an intermediate complex C. The Michaelis-Menten kinetics

$$\dot{P} = k_2 E \frac{S}{K_m + S} \tag{2}$$

is a frequently used approximation which holds for $\dot{C} \approx 0$. This assumption is satisfied if, for instance, the rates k_+ and k_- are much larger than k_2. Indeed, in the joint limit $k_\pm \to \infty$, under the constraint $\frac{k_-}{k_+} = const.$, the kinetics of the full model converge to Michaelis-Menten kinetics with $K_m = \frac{k_- + k_2}{k_+} \to \frac{k_-}{k_+}$. In terms of balancing model complexity with the available data, such a simplification would be acceptable as long as no other evidence requires employing a more complex model.

This example summarises concisely the two fundamental principles underlying our model reduction approach. (1) Reduced models emerge in the limit of extreme parameter values—our

method constructs and investigates these limits systematically. (2) A reduced model is a valid approximation if it does not contradict the data at hand—our method systematically searches for reductions, which cannot be rejected by statistical testing.

Starting with the example above, the Michaelis-Menten approximation is obtained in the limit of rapid complex formation and decay compared to the conversion rate. This transition is rendered more precisely in Fig 1. The model response for different values of constant ratio $K_m = \frac{k_+}{k_-}$ is depicted in state space (A), parameter space (B) and the log-likelihood landscape (C). The different scenarios are parameterised by the complex formation rate k_+. For larger values, the dynamics of substrate and product concentrations approach a limiting behaviour indicated by a black line, Fig 1A, which is the prediction derived from Michaelis-Menten kinetics. Based on the hypothetical data points, models with slow complex formation/decay (orange curves in Fig 1A) deviate significantly from the data whereas those with fast complex formation/decay (blue curves) do not. This can be quantified by means of the least squares function or the log-likelihood, shown in Fig 1B and 1C. Every point in k_+-k_- space is associated with a
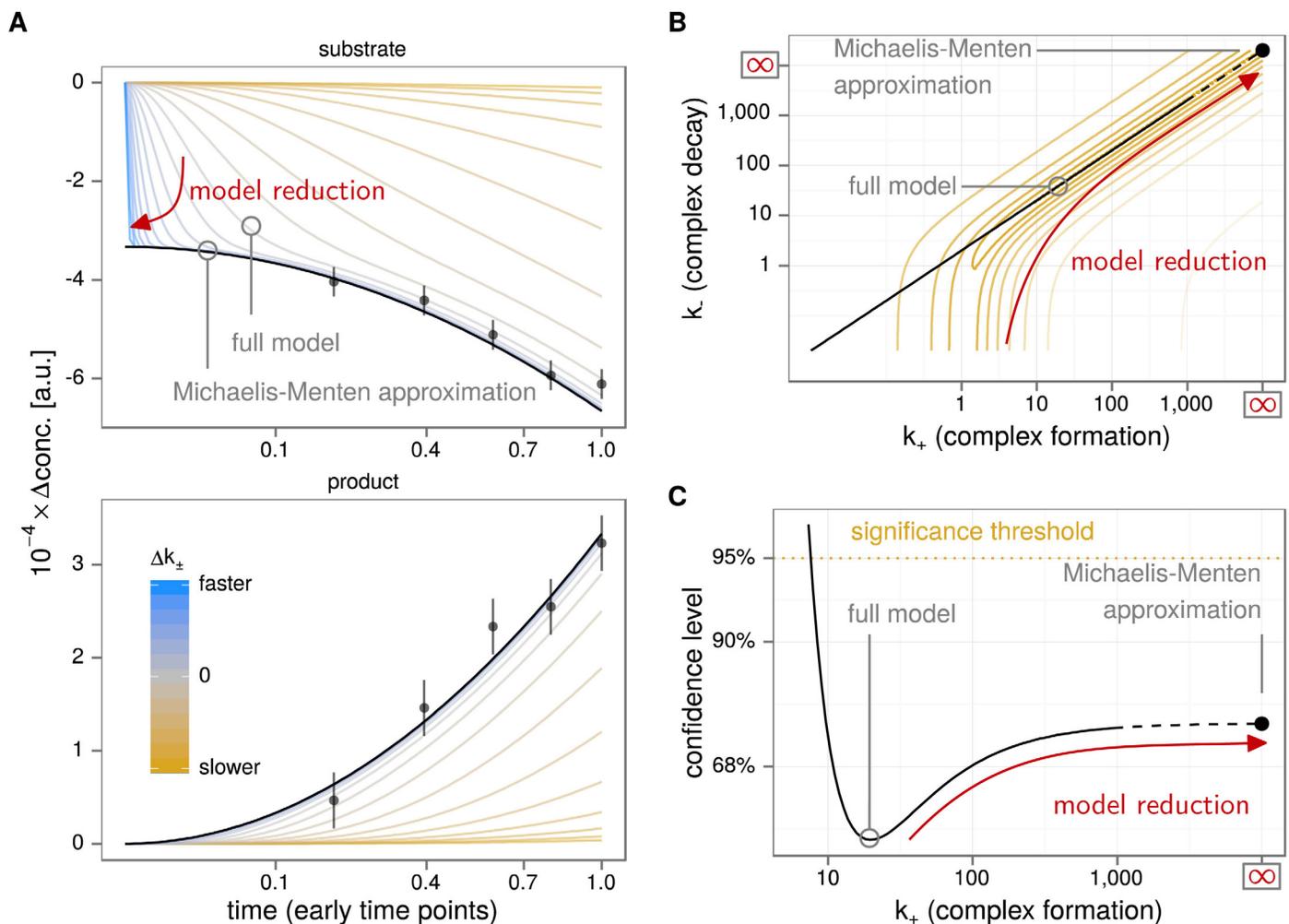
Fig 1. Emergence of the Michaelis-Menten approximation. A: Fast complex formation and decay (blue trajectories) result in the Michaelis-Menten approximation, slow formation and decay (orange trajectories) in a significant discrepancy to the data. **B:** The path in parameter space leading to the Michaelis-Menten approximation runs parallel to the contour lines of the log-likelihood function. **C:** The log-likelihood defines a significance threshold, which is not exceeded in the limit of fast formation/decay rates. Slower rates quickly lead to significant deviations from the data.

doi:10.1371/journal.pone.0162366.g001

unique model prediction and a corresponding log-likelihood value to measure the distance from the data points. The model reduction path in this space runs along a valley ending on a log-likelihood level that is not significantly increased in comparison to the optimal parameterisation of Eq (1) (full model). Since the end-point, the Michaelis-Menten approximation, was known, we could easily find this path. Conversely, the particular characteristics of our model reduction method is to construct these paths. This is achieved by computing the profile likelihood [16], which drives the model parameters systematically towards small and large values, respectively, while controlling and minimising the discrepancy between data and model prediction. The required mathematical tools for maximum-likelihood estimation in dynamic mathematical models and the profile likelihood method are introduced in the next sections.

Mathematical modelling

The rate-equation approach is a well-established methodology employing the mass-action law of chemical reactions for translating biochemical interactions into a mathematical model. Such a model is *mechanistic*, i.e. each component x and parameter of the network model has its counterpart within the biological process. Thus, it allows to infer knowledge about the underlying network and drives biological discoveries, e.g. in [17–19]. Furthermore, in [20], an overview about unravelling dynamical features from biological systems by ODE modelling is presented. The time evolution $x(t)$ of the concentrations of biochemical compounds is computed by solving the corresponding ODE system

$$\dot{x}(t) = f(x(t), \theta_x) \, , \tag{3}$$

with parameters θ_x, e.g. reaction rates or Hill coefficients. Initial concentrations x_0 are either fixed using prior assumptions or estimated from the data. The internal states $x(t)$ are mapped to observations $y(t)$ via an observation function g, i.e.

$$y(t) = g(x(t, \theta_x, x_0), \theta_y) + \epsilon(t) \, , \tag{4}$$

where independent additive Gaussian errors $\epsilon \sim N(0, \sigma^2)$ are assumed. In addition to the dependency on kinetic rates and initial concentrations, the observation function g may introduce observational parameters θ_y, which are subsumed in $\theta = \{\theta_x, x_0, \theta_y\}$. To quantify the discrepancy between model response and measurements, the scaled negative likelihood-function is calculated via

$$-2 \log(\mathcal{L}) := L(\theta) = \sum_{i=1}^{n} \left(\frac{y_i - g(x(t_i, \theta_x, x_0), \theta_y)}{\sigma_i} \right)^2 + const. \, , \tag{5}$$

for n measured data-points y_i with standard deviation σ_i. If the measurement noise is not normally distributed, transformations resulting in Gaussian errors can be applied in most cases. Often, a log-transformation is sufficient [21]. Eq 5 can be amended by penalisation terms representing prior knowledge. These can be e.g. quadratic terms for normally distributed parameters, with mean and standard deviation taken from literature, or for derived model components such as the ratio between protein concentrations in different cell compartments. To estimate model parameters, we apply maximum likelihood [22] by minimising the scaled negative log-likelihood

$$L(\hat{\theta}) = \min_{\theta} L(\theta) \, . \tag{6}$$

leading to an optimised parameter vector $\hat{\theta}$. To ensure positive values and improve numerical performance, all entries of θ are optimised on a logarithmic scale throughout the manuscript.

Finding the global optimum can be challenging due to the existence of multiple local minima. Therefore, we performed a deterministic multi-start optimisation strategy with widely dispersed initial guesses [23].

In general, no analytical solution of the ODE system (Eq (3)) exists or is available, hence numerical solvers are utilised to approximate the dynamics. Here, we used *CVODES* from the SUNDIALS suite [24] for ODE integration. The minimisation in Eq (6) is performed numerically using the trust-region-based large-scale non-linear optimisation algorithm *lsqnonlin* [25] as implemented in MATLAB. Gradient-based parameter estimation strategies depend on the sensitivities of the model function, i.e. inner derivatives of the likelihood. In order to ensure numerical accuracy, we computed and supplied forward sensitivities [26] by extending the ODE system with the appropriate sensitivity equations. All model analyses, optimisation and visualisation of model responses for the manuscript were performed using the freely available Data2Dynamics modelling environment [27] for MATLAB (http://www.data2dynamics.org/). Additionally, implementations of the toy models and analyses using the freely available dMod/cOde packages for R (https://github.com/dkaschek/) can be found in the Supplementary Section 1.

Parameter profile likelihood

When fitting models to data, the precision of parameter estimates is assessed by computing confidence intervals. In non-linear models, confidence intervals that are not confined to the lower and/or higher limit of a parameter may occur. In this case, a parameter is termed non-identifiable. In contrast, a parameter with a finite confidence interval is called identifiable.

The profile likelihood is an established concept to assess parameter identifiability in non-linear regression. It generalises Fisher Information based confidence intervals to the non-linear setting, resulting in appropriate confidence regions [10]. In short, a parameter of interest θ_i is profiled by scanning along its axis and re-optimising all other parameters $\theta_{j \neq i}$ for each value of θ_i. Thus, the profile likelihood is defined as

$$\text{PL}(\theta_i) = \min_{\theta_{j \neq i}} L(\theta) \ . \tag{7}$$

$\text{PL}(\theta_i)$ traces an optimal path through the parameter space for each θ_i, thereby providing global information in contrast to the curvature of Eq (5) evaluated at $\hat{\theta}$, as used for Fisher Information based confidence intervals. If $\chi^2_{\alpha,1}$ denotes the α quantile of the χ^2 distribution with one degree of freedom, the region for which the inequality

$$D := \text{PL}(\theta_i) - L(\hat{\theta}) \leq \chi^2_{\alpha,1} \tag{8}$$

is satisfied yields the confidence interval of the parameter θ_i to a given confidence level α. This means that under weak assumptions [28] $(1 - \alpha)$ specifies the probability that, for repeated experiments, the true value of θ_i lies within the boundaries of the confidence interval. $L(\hat{\theta}) + \chi^2_{\alpha,1}$ defines the threshold that $\text{PL}(\theta_i)$ may not exceed for an acceptable parameter. Thereby, the profile likelihood provides the range of parameter values supported by the available measurement data. The re-optimisation of other parameters during the profile likelihood calculation is crucial to probe the non-linear relationships between parameters, which are key for discovering suitable model reductions later in the manuscript.

Concerning parameter profiles, three scenarios can be distinguished: A parameter can be 1) identifiable, 2) structurally non-identifiable, or 3) practically non-identifiable [10]. Each case is associated with a recognisable shape of the profile likelihood.

An *identifiable* parameter has a confined confidence interval, i.e. the corresponding profile likelihood PL(θ_i) exceeds the threshold given by $L(\hat{\theta}) + \chi^2_{\alpha,1}$ for values above and below its maximum likelihood parameter estimate $\hat{\theta}_i$. For an identifiable parameter the shape of the profile likelihood is often close to quadratic. This setting is referred to as the *asymptotic* setting, because the profile likelihood based confidence intervals are similar to those calculated from Fisher Information. Thus, asymptotic confidence intervals calculated using only the local curvature of the likelihood at the optimum $\hat{\theta}$ are valid in this scenario.

Structurally non-identifiable parameters are characterised by the profile likelihood of θ_i remaining constant for arbitrary values of θ_i. Structural non-identifiabilities are a consequence of over-parameterisation in terms of too many parameters for describing the available data. Hence, the system harbours symmetries that can be detected by exploiting Lie group theory [29]. Once the symmetry transformations are found, they allow reconstruction of the full solution space from a reduced set of parameters. Therefore, one parameter of each detected transformation group is set to a fixed value. Thereby, the reduced set of parameters becomes structurally identifiable.

Finally, a non-identifiable parameter characterised by a profile with a (not necessarily unique) global minimum, but which does not exceed the statistical threshold in at least one direction is called *practically non-identifiable*. In such cases, the parameter does not have a finite confidence interval but could be restricted to one side, e.g. the parameter must not be larger than a certain value but could be arbitrarily small. Hence, asymptotic confidence intervals are inappropriate as they are finite by construction. This feature is inherited from the linearity assumption of asymptotic confidence intervals, making practical non-identifiability a property that only occurs in non-linear models. For practically non-identifiable parameters, repeating the same measurements can in principle lead to identifiability, given that the true solution is part of the parameter space. Experimental design based on parameter profiles can be applied to efficiently attain identifiability [30]. However, if new data acquisition is impossible or infeasible, practical non-identifiabilities can be resolved by applying the model reduction technique presented in the following, eventually resulting in a completely identifiable model.

Results

The reduction procedure described in this paper is based on the parameter profile likelihood introduced above. Identifiable parameters have a well-defined confidence interval and may not be reduced. Structurally non-identifiable parameters can be set to an user-defined value. This leaves practically non-identifiable parameters to serve as starting point for model reduction. Like in the introductory example (Fig 1), the possibility of practically non-identifiable rate constants being driven to infinity implies that the corresponding reactions may be arbitrarily fast and are subsequently *lumped* in a model reduction step.

More general, each point along a parameter profile PL(θ_i) can be interpreted as a likelihood-ratio test with the test-statistic D according to Eq (8). If a parameter θ_i is practically non-identifiable, the profile likelihood does not exceed the threshold and in most cases flattens out to a constant value in the unbound direction of the confidence interval. The parameter values along the profile likelihood define a sub-manifold in the parameter space, whose respective dynamics correspond to model behaviour which cannot be rejected by the likelihood-ratio test. Similar to the structurally non-identifiable case, the model exhibits a symmetry which can be exploited to reduce the model. Hence, for practically non-identifiable parameters, the profile likelihood method systematically drives the system towards limit behaviour. It provides direct information on the asymptotic coupling/uncoupling of parameters via the parameter paths and reveals the behaviour of the limit model by inspection of the model predictions along the sub-manifold. It

is this limit behaviour which ultimately corresponds to the idealised reduced model. In [10], the authors provide a more detailed description on parameter identifiability.

Four different scenarios leading to model reduction can be distinguished. On the one hand, a parameter's limit behaviour that cannot be rejected by the likelihood-ratio test can occur for parameter values towards plus (+) or minus (−) infinity. On the other hand, the paths of the remaining model parameters can be coupled (↕) to the parameter of interest or remain at the same value (|). This leads to the four scenarios (+|), (−|), (+↕), and (−↕). To make statements more concrete, we chose a representative *in-silico* example for each of these four scenarios. Details about their respective model equations, parameter values, and an extensive model reduction procedure for each example are provided in the Supplementary Section 1. Subsequently, reduction of a biological model of the Reelin signalling cascade is exerted using the presented model reduction strategy based on the profile likelihood.

Scenario 1: (+|)

The illustrative model used for the basic reductions is a cascade $X \xrightarrow{k_1} pX \xrightarrow{k_2} ppX$, where each arrow indicates an elementary reaction parameterised by k_1 and k_2, respectively ([Fig 2A]). The
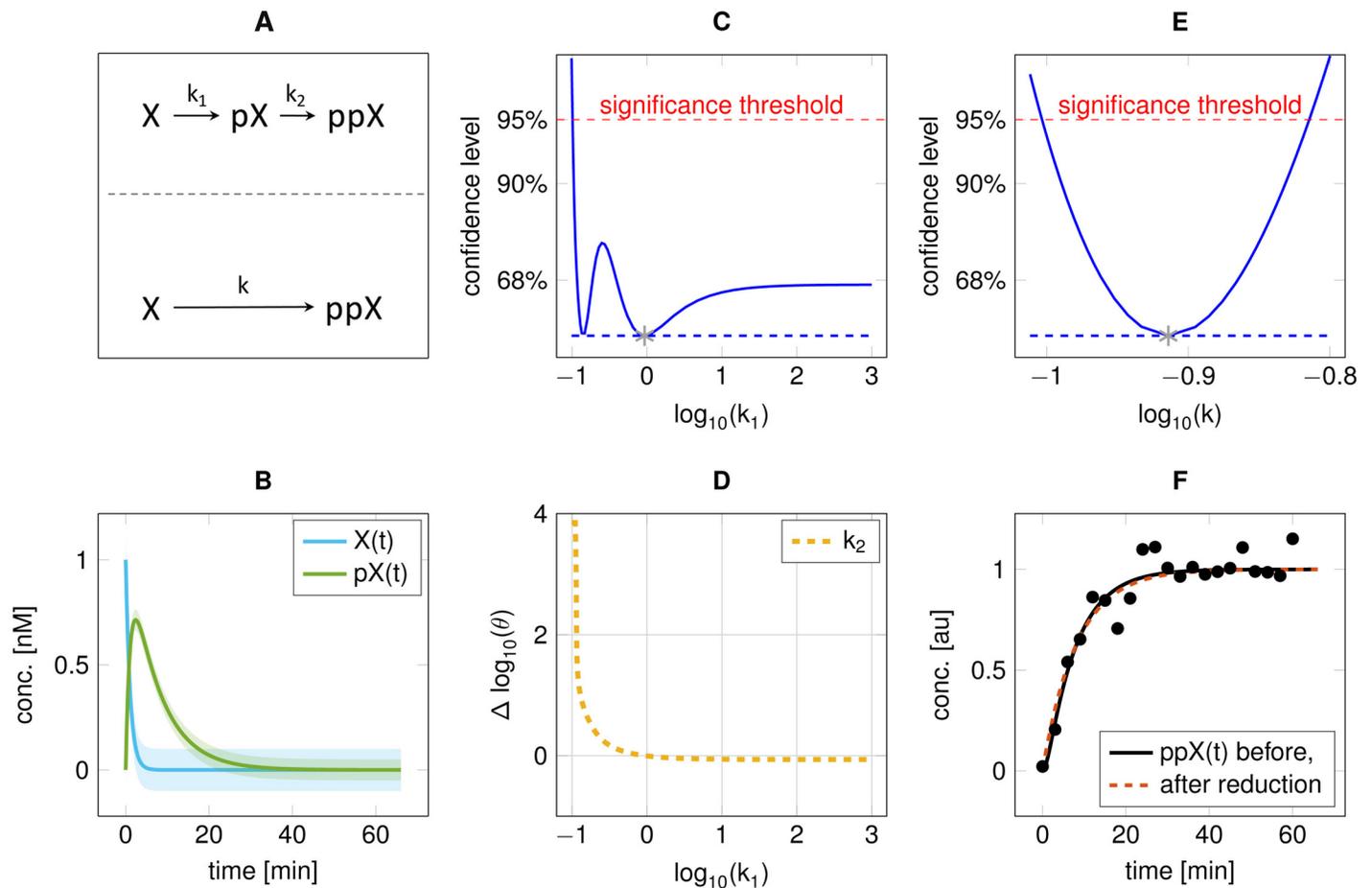
Fig 2. Reduction of the cascade toy model 1 (scenario 1). A: Model scheme before (upper) and after (lower) reduction with lumped states X and pX. **B:** $X(t)$ and $pX(t)$ for fitted parameters. **C:** Profile likelihood for parameter k_1. **D:** Relationship of k_1 to k_2 obtained by re-optimisation along the profile likelihood (panel C). **E:** Profile likelihood for the identifiable parameter k in the reduced model. **F:** $ppX(t)$ with simulated data after fitting, with comparison between a one and two step conversion.

doi:10.1371/journal.pone.0162366.g002

dynamics of $X(t)$ and $pX(t)$ are shown in Fig 2B. In this scenario, data is observed for $ppX(t)$ with $\sigma = 0.1$ (Fig 2F). By calculating the profile likelihood of the kinetic parameter k_1 towards $+\infty$ two observations can be made. (1) two local optima exist around -1 and 0, which reflect the interchangeability of the reaction rates of two consecutive steps in a conversion chain for unobserved intermediate pX. (2) the parameter k_1 is practically non-identifiable as the profile likelihood of k_1 is below the statistical threshold for large values (Fig 2C). Thus, high values of the parameter k1 do not decrease the ability of the model to describe the data. Furthermore, no other parameter is compensating the changes in parameter k_1, which manifests in k_2 flattening out for $k_1 \rightarrow \infty$ (Fig 2D). Since a large value for k_1 corresponds to fast conversion of X to pX, a reduction of the model by *lumping* the states X and pX is feasible. In the reduced model $X \xrightarrow{k} ppX$, the remaining rate constant is identifiable (Fig 2E), whereas the data is equally well described (Fig 2F).

Scenario 2: (−|)

Similar to the previous example, we use a model structure featuring a cascade X to pX and ppX, as shown in Fig 3A, with simulated data with $\sigma = 0.02$ for $X(t)$ and $pX(t)$ (Fig 3B). In addition,
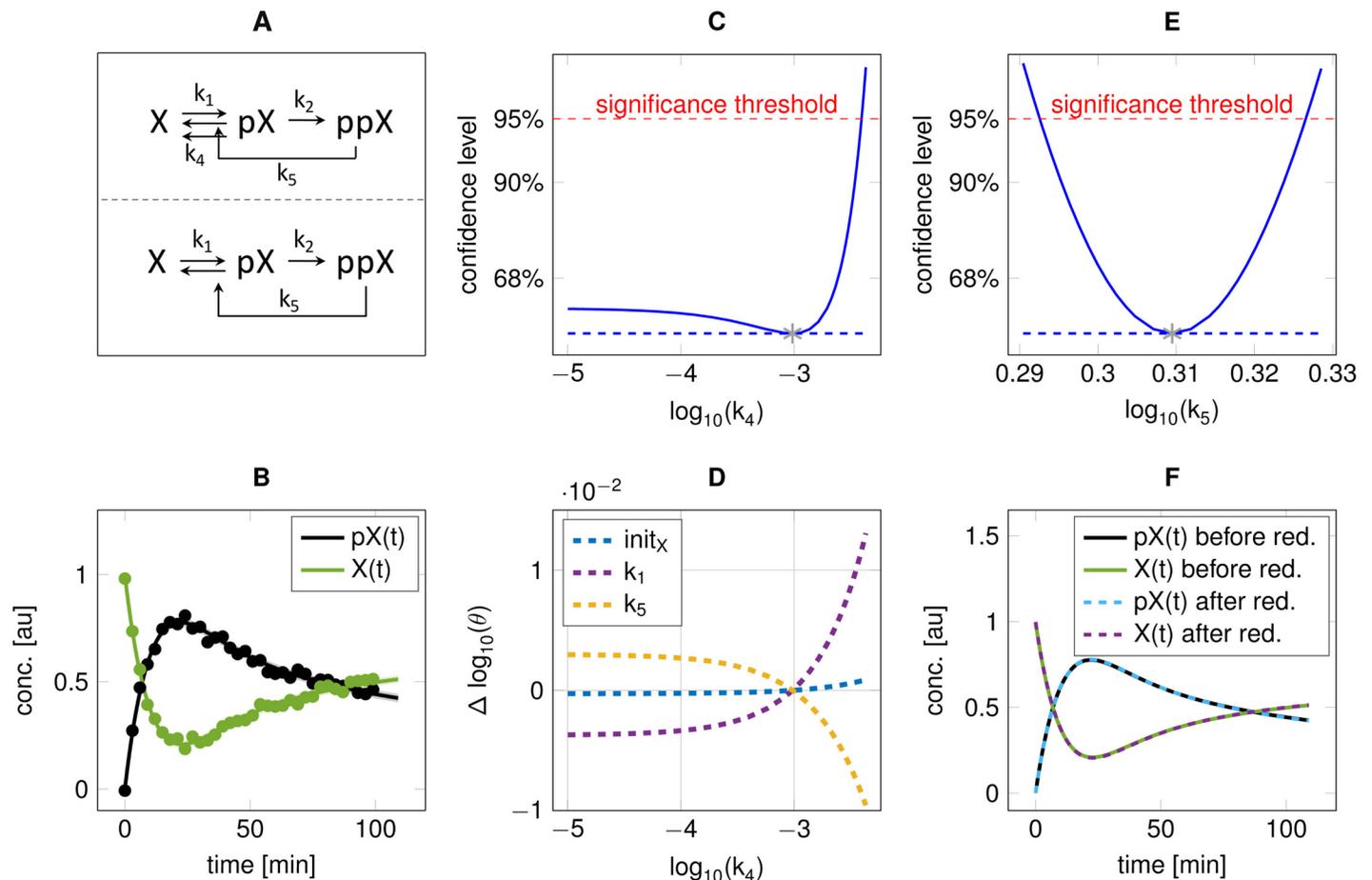
Fig 3. Reduction of the cascade toy model 2 (scenario 2). A: Model scheme before (upper) and after (lower) reduction. The difference is the basal deactivation from pX to X with rate constant k_4. **B:** $X(t)$ and $pX(t)$ with data for fitted parameters. **C:** Profile likelihood for parameter k_4. **D:** Relationship of the remaining parameters to the profile shown in C. **E:** The reaction associated with the parameter k_4 is removed by the reduction, while k_5 is identifiable. **F:** $pX(t)$ and $X(t)$ with comparison before and after reduction.

pX is dephosphorylated to X with a basal rate, and by a feedback through ppX. The profile likelihood for k_4 shows a practical non-identifiability of the basal dephosphorylation, as the lower boundary is not defined (Fig 3C) with constant values of the other parameters as k_4 approaches $-\infty$ (Fig 3D). On the other hand, the parameter k_5, reflecting the dephosphorylation through a feedback by ppX, is identifiable as shown by the corresponding profile likelihood (Fig 3E). Here, an appropriate reduction is given by the *removal* of the basal dephosphorylation of pX (Fig 3A), with similar dynamics of $pX(t)$ before and after reduction (Fig 3F).

Scenario 3: (+↕)

In the following, an example of a model is given where the ODE solution of one state can be replaced by an *algebraic equation* (Fig 4A). It features qualitatively similar transient dynamics on a different scale for two successive phosphorylations, $pY(t)$ and $pZ(t)$ (Fig 4B). Data is observed with $\sigma = 0.05$. Since there is no feedback, the activation of state Z cannot be faster than the one of state Y [31]. If the total amount of Z is not limiting, the dynamics $pY(t)$ and $pZ(t)$ have similar shape when the phosphorylation and dephosphorylation parameters tend to infinity. In this example, the kinetic parameter for the dephosphorylation of state Z, $k_{d,Z}$ is
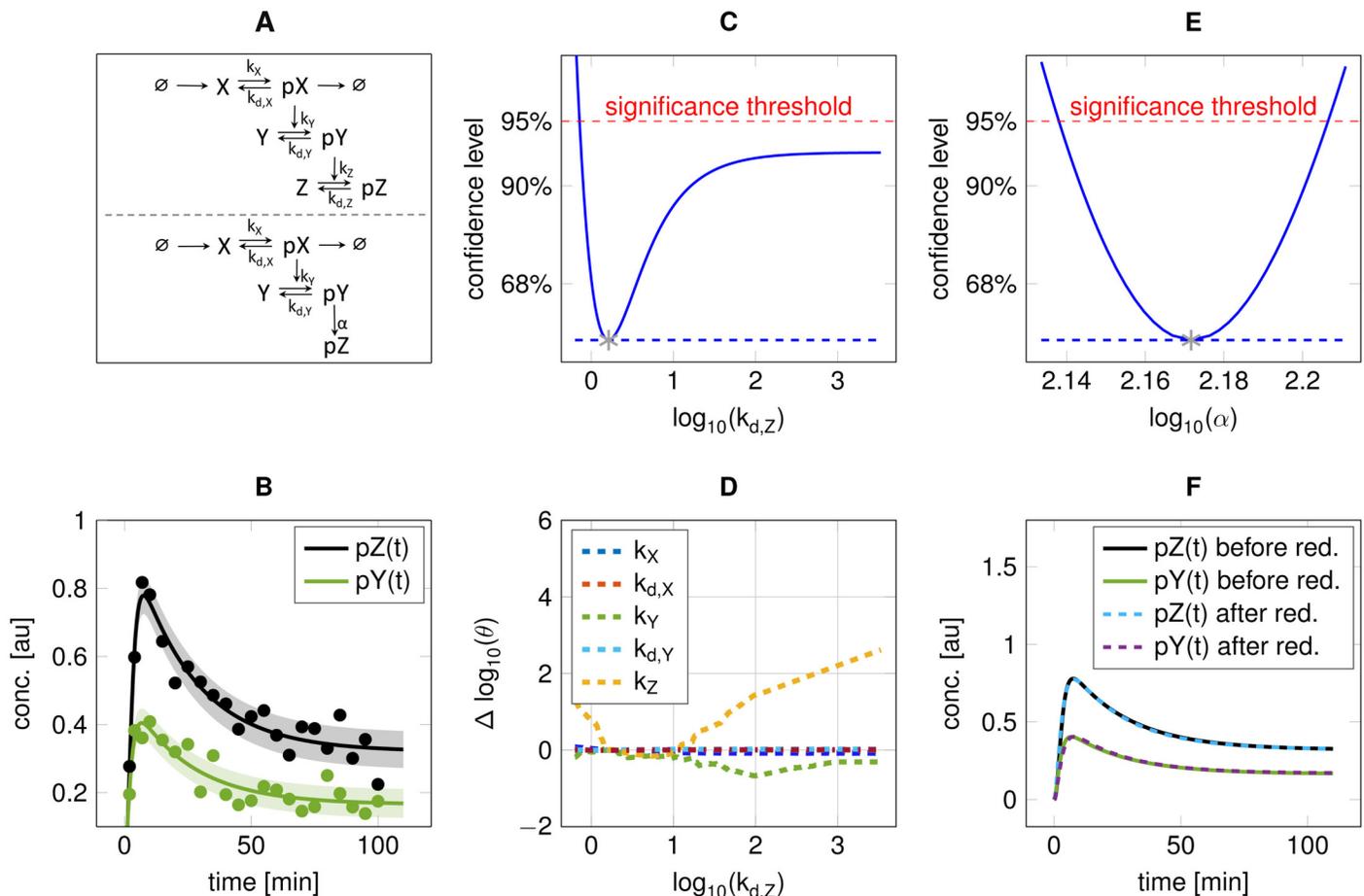
Fig 4. Reduction through functional relation (scenario 3). A: Model scheme before (upper) and after (lower) reduction. The difference is the removal of state Z, and the algebraic replacement $pZ(t) = \alpha \cdot pY(t)$. **B:** $pY(t)$ and $pZ(t)$ with simulated data after fitting. **C:** Profile likelihood for parameter $k_{d,Z}$. **D:** Re-optimisation of remaining parameters along the profile likelihood for $k_{d,Z}$. **E:** The profile likelihood for the identifiable parameter α in the reduced model. **F:** Comparison of trajectories for $pY(t)$ and $pZ(t)$.

doi:10.1371/journal.pone.0162366.g004

non-identifiable in its upper direction ([Fig 4C](#)). The phosphorylation rate k_Z will compensate for changes since $\frac{k_Z}{k_{d,Z}}$ determines the final concentration level of state Z ([Fig 4D](#)). To overcome this practical non-identifiability, the kinetic rate equation of state Z is replaced by a simple functional relation given by $Z = \alpha Y$, which makes α identifiable ([Fig 4E](#)). The reduced model is able to describe the measurements without a statistically significant decrease in the likelihood compared to the full model ([Fig 4F](#)).

For this scenario, a comparison with model reduction based on negligible fluxes is conducted in Supplementary Section 1.3.8, demonstrating the advances of profile likelihood based reduction for non-trivial relations between different model components. In addition, a parameter reduction not proposed by the presented reduction algorithm is illustrated in Supplementary Section 1.3.9, resulting in a reduced model which is rejected by the likelihood-ratio test.

Scenario 4: ($-\updownarrow$)

In the final toy example of a weakly activated signalling pathway, the reduction is based on the trajectories associated with the parameter profile likelihood. State X is simulated to be phosphorylated by an exponentially decaying input through k_{on}, whereby the phosphorylated state pX can be deactivated with rate constant k_{off} ([Fig 5A](#)). In addition, pX is mapped to the observations by a scaling factor, i.e. $pX_{obs} = scale \cdot pX$. Data is observed for $pX(t)$ with standard deviation $\sigma = 0.1$ ([Fig 5F](#)). The parameter profile likelihood reveals that k_{on} ([Fig 5C](#)) is practically non-identifiable and the concomitant change of the parameter *scale* along the profile likelihood reveals that only their product is identifiable due to the linear relationship of $\log(scale)$ to $\log(k_{on})$ in [Fig 5D](#). The reason that these are not structurally non-identifiable is revealed by plotting the trajectories associated with the parameter sets at each point of the profile likelihood ([Fig 5B](#)). As the initial level of X is fixed, the time course of its concentration level $X(t)$ tends towards zero for large k_{on}, i.e. X may become limiting (cyan trajectories in [Fig 5B](#)). If X is limiting, though, the shape of $pX(t)$ is different from the weakly activated case, where the time course of $X(t)$ converges to a nearly flat trajectory (blue trajectory in [Fig 5B](#)). These dynamics are associated with the flat direction of the parameter profile likelihood. Thus, the appropriate model reduction in this case is obtained by assuming that phosphorylation of X leads to a *negligible reduction* of the concentration of unphosphorylated X, which then serves as constant input for its phosphorylation. Thereby, the former practical non-identifiablility becomes a structural non-identifiability, which is depicted by the parameter profile likelihood of k_{on} after reduction ([Fig 5E](#)). Similar to [Fig 5D](#), the parameter *scale* is coupled to k_{on} but as their relationship is now structural the model can be reduced by fixing one of them, e.g. *scale*, to an arbitrary value. After this reduction, the data is equally well described ([Fig 5F](#)).

Flow-chart

A summary of the presented model reduction method is depicted in a flow-chart diagram in [Fig 6](#). The proposed reduction steps are context-specific and should not be applied without checking their validity. Especially for scenarios 3 and 4, where coupling of parameters is present ($\updownarrow$), careful investigation of the parameter dependencies and their influence on the model quantities is crucial. On the contrary, this context-specific step provides additional insight, since relationships between parameters are revealed by driving the model towards its limits through the profile likelihood. These dependencies yield sub-systems of the original model and show the variability present within this sub-system. This modularisation identifies well-informed parts, and allows for reduction of uninformed parts.
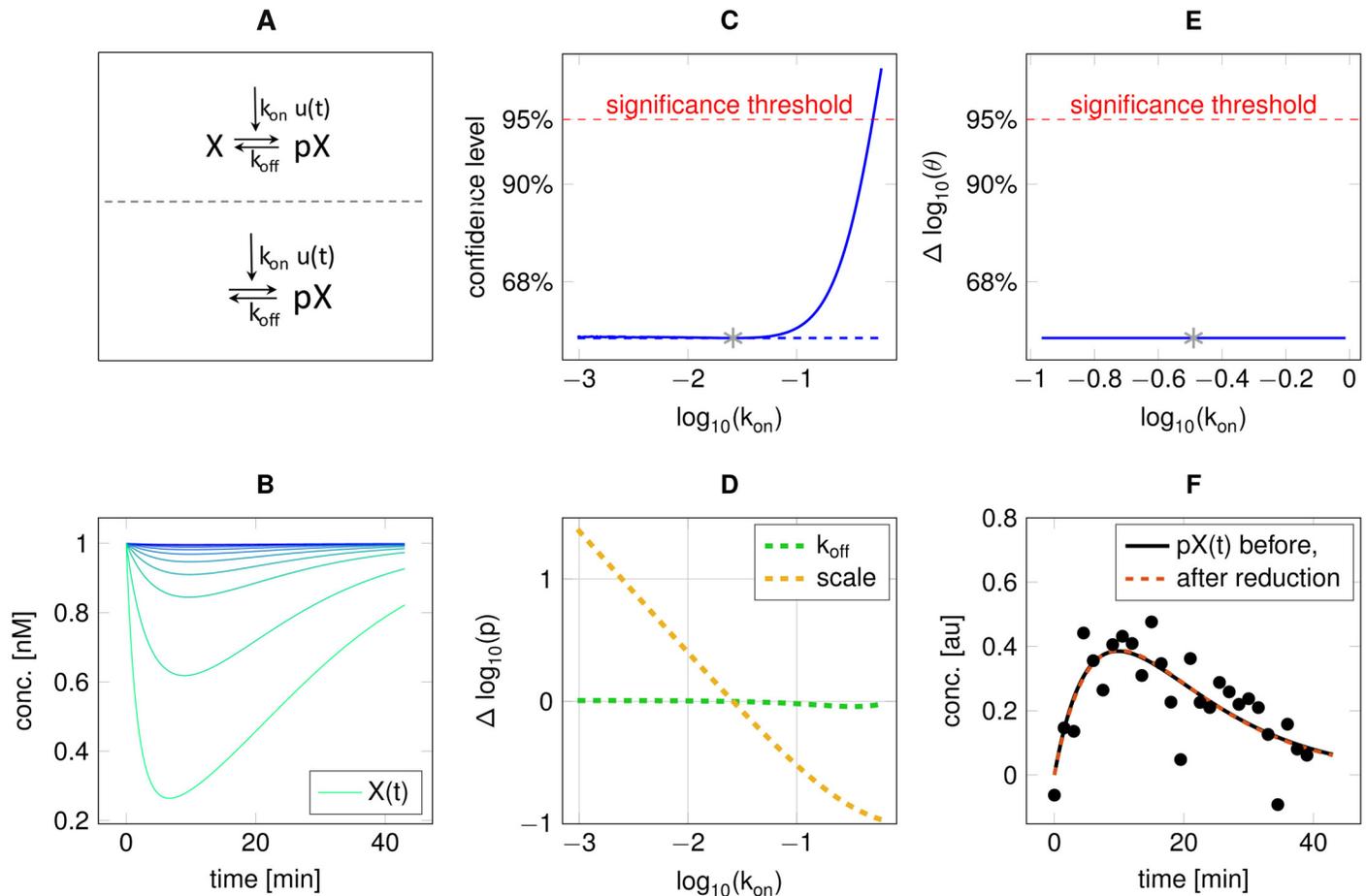
Fig 5. Model reduction of weakly activated signalling pathway (scenario 4). A: Model scheme before (upper) and after (lower) reduction. The difference is the omission of state X. **B:** $X(t)$ for parameter sets along the profile likelihood of k_{on} (high to low values of k_{on} from bottom to top, i.e. cyan to blue). **C:** Parameter profile likelihood of k_{on}, depicting a practical non-identifiability to small values. **D:** Relation of *scale* to k_{on}. **E:** After model reduction, the parameter k_{on} is structurally non-identifiable and can be set to an arbitrary value. **F:** Comparison of model fits before and after reduction. Both curves overlap and cannot be statistically distinguished based on the data.

doi:10.1371/journal.pone.0162366.g005

Model reduction of Reelin signalling cascade

In the following, the presented model reduction strategy will be carried out on a model of the early Reelin signalling cascade in order to demonstrate its applicability on real-world models of biochemical reactions. The Reelin signalling cascade is essential for the development of the mammalian brain by regulating the position of new-born neurons in the neocortex, hippocampus and cerebellum [32, 33]. Moreover, the pathway is involved in the modulation of synaptic plasticity, learning and memory in the adult brain, and defects in Reelin signalling are associated with neuropsychiatric diseases such as Alzheimer disease, schizophrenia, autism and epilepsy [34–36]. Although important aspects of Reelin signalling have been deciphered using classical biochemical approaches and mouse genetics, a superordinate view of the interaction of its components is needed [37].

Reelin exerts its function by binding to the lipoprotein receptors VLDLR and ApoER2, which induce tyrosine phosphorylation of the adaptor protein Dab1 through Src family kinases (SFKs) [38, 39]. As a feed-forward loop, phosphorylated Dab1 then trans-phosphorylates other Dab1 proteins bound to the receptor complex. In turn, Dab1 activates Akt that is, amongst
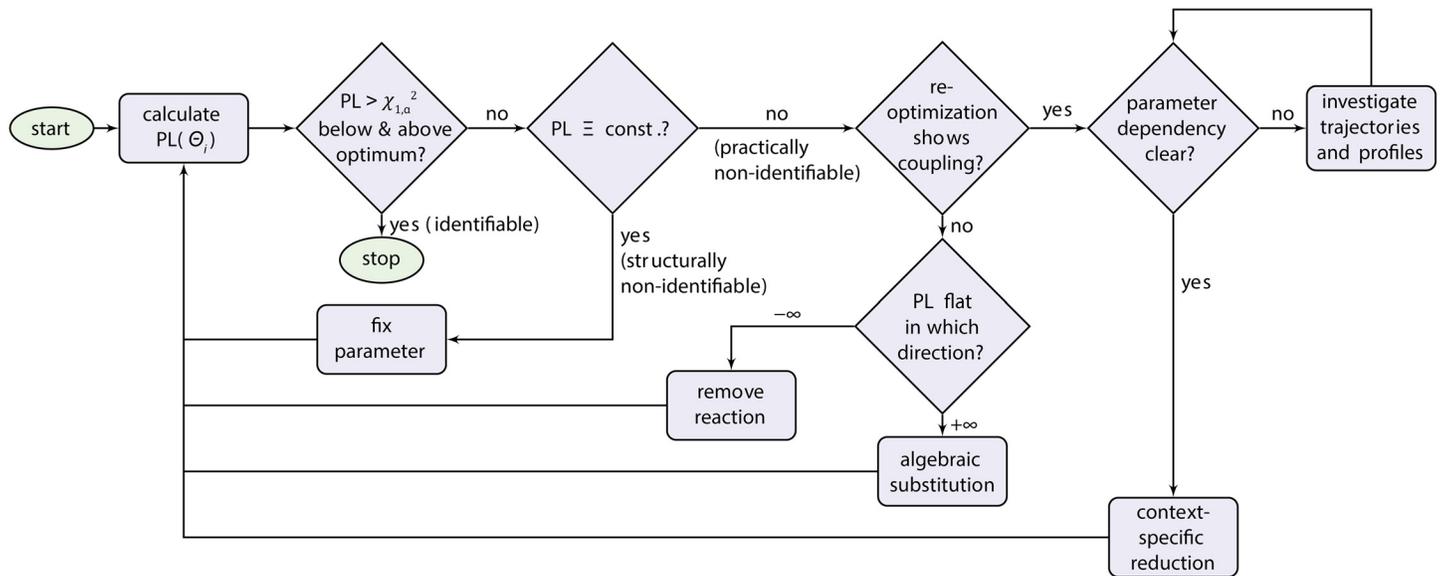
Fig 6. Typical flow-chart for model reduction based on the profile likelihood. The depicted steps are to be applied for each parameter individually, starting with the calculation of the respective profile likelihood. After detection, the procedure resolves non-identifiabilities by fixing parameters, removing reactions, performing algebraic substitutions, or context-specific reductions. The method terminates when all parameters of interest are identifiable.

others, involved in cell survival and migration (Fig 7A). The signalling is terminated by ubiquitination and degradation of Dab1. The proposed model reduction strategy is illustrated based on time-resolved data of total Dab1 and phosphorylated Dab1, SFKs and Akt, which were measured in cortical neurons after Reelin stimulation. In addition, an experiment applying a SFK inhibitor prior to Reelin stimulation was performed. Protein concentrations were measured in cell lysates of primary cortical neuron cultures by immunoblotting (see Supplementary Section 2). The measurements cover the first four hours after ligand stimulation, with dense measurements in the first 30 minutes (Fig 7B). Analytic steady state solutions, e.g. through basal phosphorylation, were determined for the initial concentration levels. The model consists of 12 states with initial concentrations implicitly given by steady state assumptions, 13 kinetic parameters, and 23 observational parameters. Taken together, the model therefore has 35 free parameters, which are fitted to 108 data-points. Computation of one parameter profile likelihood for this setting takes less than one minute on a standard laptop. Details about the data, model fits and differential equations before and after reduction are provided in Supplementary Section 2 and 3.

Considering the model parameters, two reductions can be conducted, which both correspond to scenario 3 ($+\updownarrow$). In Fig 7A, they are marked as red boxes named A and B. Regarding box A, the profile likelihood for the release of the SFK inhibitor, $SFK_{deInhib}$, does not exceed the 95% threshold for large values (Fig 8A) showing a positive relation to its binding to the SFKs, SFK_{Inhib} (Fig 8B). This indicates that a steady state is reached instantly at every time point, which results in a consistent partitioning between SFKs with and without bound inhibitor. The factor Inh_{part}, which determines the equilibrium concentration, is thereby identifiable (Fig 8E).

Further, the Akt deactivation Akt_{deact} shows a parameter profile that does not exceed the 95% threshold for large parameter values (Fig 8C), depicted by box B of Fig 7A. The Akt activation along the profile likelihood (Fig 8C) adapts to the Akt deactivation, as their ratio determines the steady state concentration of pAkt after Reelin stimulation (Fig 8D). A helpful fact in order to resolve this non-identifiability is that the time-course of pAkt will approach the time-
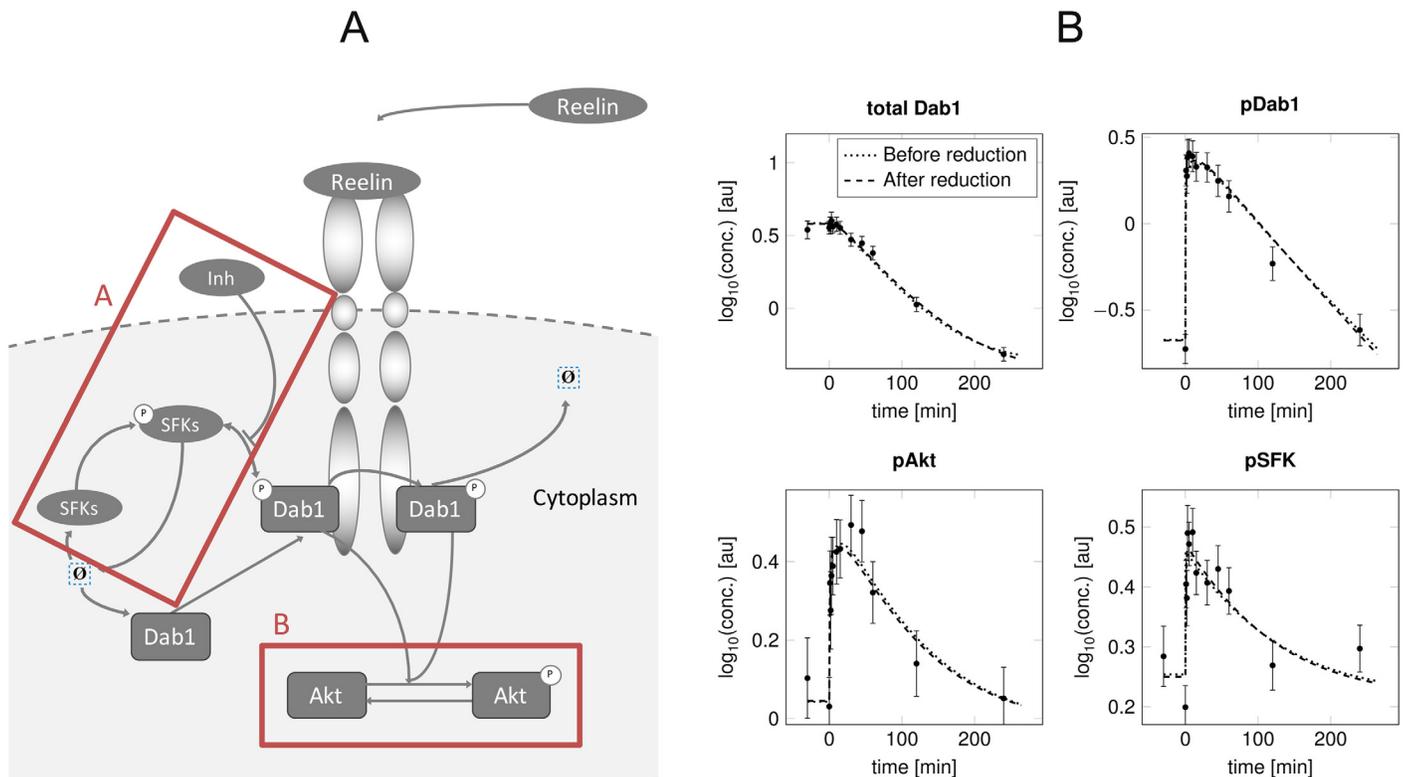
Fig 7. Reelin-induced signalling pathway. A: Scheme of the Reelin-induced signalling pathway. Sub modules that underwent model reduction are framed. **B**: Experimental data and model trajectories of the Reelin model. Data points and their measurement errors are shown on a log-scale for the measurements of total Dab1, phosphorylated Dab1, Akt and SFKs for time points between zero and 240 minutes. Dotted and dashed lines indicate the model response for the parameter set before and after model reduction, respectively.

doi:10.1371/journal.pone.0162366.g007

course of the upstream protein Dab1 for large values of both Akt activation and deactivation. This result is not trivial, since scaling and offset linking both states to their respective observations are different. Replacing the Akt phosphorylation by a functional relation between pAkt and pDab1 resolves this non-identifiability. Thus, the pAkt dynamics follow the pDab1 dynamics, and can be mapped with an identifiable scaling factor $scale_{pAkt}$ to the pAkt observations (Fig 8F). Fig 7B shows the insignificant impact of all model reduction steps on the dynamics. Statistical agreement of both the original and the reduced model is assessed by the likelihood-ratio test described in Eq (8), taking the difference in degrees of freedom between both models into account.

Discussion and Conclusion

Traditionally, mathematical models are small since they phenomenologically describe processes like biochemical interactions at a simplified level. In such a setting, one question is whether an initially chosen level of abstraction turns out to be too stringent demanding for to model enlargement by additional effects. In contrast to this traditional situation, mathematical models in systems biology are usually mechanistic, i.e. all relevant processes in living cells have their counterparts in the model. Therefore, the models typically start out large and the adequate level of complexity has to be checked in both directions, i.e. on the one hand it has to be tested whether all essential processes and interactions are included and on the other hand
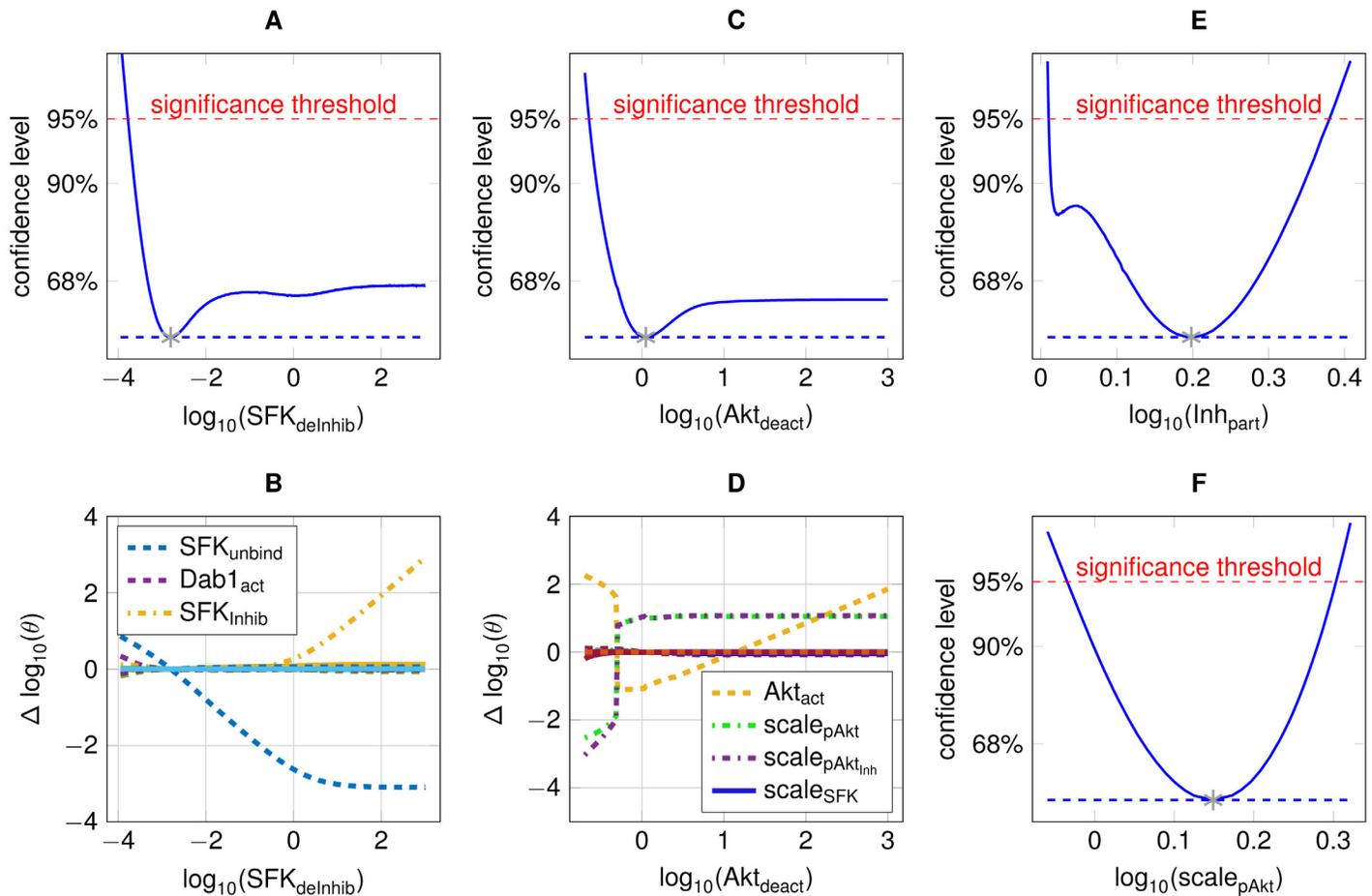
Fig 8. Summary of model reduction steps. **A**: The profile likelihood of the inhibitor release from SFKs. **B**: The re-optimised paths of the remaining parameters with respect to the profile of the inhibitor release. **C**: Parameter profile likelihood of the Akt deactivation. **D**: Coupling of the Akt activation, log (Akt_{act}), to the Akt deactivation shown in panel C. **E**: Identifiable partitioning of SFKs with and without bound inhibitor. **F**: The scaling factor $scale_{pAkt}$, which links the observations of pAkt to pDab1, is identifiable.

model reductions can be performed. Following such a combined strategy offers the possibility to obtain a comprehensive model, which is tailored to include only the required components.

The model reduction methodology presented in this manuscript is purely data-based, yet can take prior knowledge into account. We use the profile likelihood, which is a continuous representation of the likelihood ratio statistic, to suggest a model reduction strategy that iteratively eliminates practically non-identifiable parameters. It considers the distribution and magnitude of data noise in a statistically valid manner and is robust against observation noise. Investigating the model fit along the profile likelihood and considering parameter coupling provides educated guesses on where the model may be amenable to reduction. Increasing levels of data noise would affect the method in a way that more simulation trajectories would be able to describe the data within the statistical boundaries set by the χ^2 function.

Our suggested approach permits simplifications, which would not be rejected by the likelihood-ratio test. However, because only experimentally observed model components and incorporated prior knowledge enter the likelihood ratio, such model simplifications face the risk that unobserved but biologically relevant parts are removed which would manifest in biased predictions. Unfortunately, significance considerations are not applicable to this issue since

statistical tests only permit of model simplifications but not of generalisations. However, the risk of removing relevant parts can be diminished by restricting to biologically meaningful simplifications and by performing validation experiments. Except for this important aspect, the presented model reduction strategy can be performed iteratively until an identifiable model is obtained which typically enables precise predictions with small confidence intervals that facilitate experimental validation or falsification of the current model structure. Moreover, an identifiable model permits summarising the experimental outcomes to estimated parameters and confidence intervals of finite size which is an important requirement for transferring a model to new applications. Additionally, identifiable models require less demanding computations at several levels. Optimisation of the parameters for instance is less hampered by ill-conditioned step size control. Moreover, the performance of ODE integration deteriorates less frequently because rate constants are restricted to meaningful ranges and the ODEs tend to be non-stiff.

In this study, we do not consider the specific order in which the model reduction steps are performed. This issue occurs in any model reduction and model discrimination problem and is usually approached by either completely searching all admissible cases or by stepwise forward or backward selection or combinations thereof. As an example, [15] consider the order of model reduction steps by comparing it to the traveling salesman problem. Although not shown here, established approaches targeting this issue can be readily combined with the presented model reduction steps.

The profile likelihood has become a well-established method for the analysis of parameter uncertainty. Its calculation is often routinely performed in systems biology applications. In such cases, our model reduction analysis can be performed without additional effort. In addition, the computation of parameter profiles can be performed in parallel on a multi-core processor or by means of distributed computing.

To illustrate our profile likelihood-based model reduction strategy, we focused on examples with general impact covering a broad range of model redundancies. However, these examples are not exhaustive and additional reductions may occur in special cases. If such context-specific model reductions are required, investigation of model predictions, e.g. trajectories of the internal states $x(t)$, for parameter sets sampled along the profile likelihood is often helpful.

Thereby, the presented method reveals features which are crucial for model predictions while the simplifications can be traced back to their mechanistic origin. Thus, the approach iteratively builds up confidence in the model, and constitutes a helpful tool for improving understanding, interpretability, and usefulness of systems biology models.

Supporting Information

S1 File. Implementations and analyses of toy models and further documentation on the Reelin pathway model.
(PDF)

Author Contributions

Conceived and designed the experiments: TM HH BS JV RE FK HHB DK CK JT.

Performed the experiments: TM HH BS JV RE FK DK.

Analyzed the data: TM HH BS JV RE FK DK CK.

Contributed reagents/materials/analysis tools: AR HHB DK CK.

Wrote the paper: TM HH BS JV RE AR FK HHB DK CK JT.

References

1. Gorban AN, Karlin IV. Method of invariant manifold for chemical kinetics. Chemical Engineering Science. 2003; 58(21):4751–4768. doi: 10.1016/j.ces.2002.12.001

2. Roussel MR, Fraser SJ. Invariant manifold methods for metabolic model reduction. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2001; 11(1):196–206. doi: 10.1063/1.1349891

3. Segel IH, Segel AH. Biochemical Calculations: How to solve mathematical Problems in General Biochemistry. Wiley; 1968.

4. Segel LA, Slemrod M. The quasi-steady-state assumption: a case study in perturbation. SIAM Review. 1989; 31(3):446–477. doi: 10.1137/1031091

5. Boulier F, Lefranc M, Lemaire F, Morant PE. Model reduction of chemical reaction systems using elimination. Mathematics in Computer Science. 2011; 5(3):289–301. doi: 10.1007/s11786-011-0093-2

6. Sunnåker M, Cedersund G, Jirstrand M. A method for zooming of nonlinear models of biochemical systems. BMC Systems Biology. 2011; 5(1):140. doi: 10.1186/1752-0509-5-140 PMID: 21899762

7. Schmidt H, Madsen MF, Danø S, Cedersund G. Complexity reduction of biochemical rate expressions. Bioinformatics. 2008; 24(6):848–854. doi: 10.1093/bioinformatics/btn035 PMID: 18267948

8. Quinn GP, Keough MJ. Experimental Design and Data Analysis for Biologists. Cambridge University Press; 2002.

9. Radulescu O, Gorban AN, Zinovyev A, Noel V. Reduction of dynamical biochemical reactions networks in computational biology. Frontiers in Genetics. 2012; 3:131. doi: 10.3389/fgene.2012.00131 PMID: 22833754

10. 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; 25(15):1923–1929. doi: 10.1093/bioinformatics/btp358 PMID: 19505944

11. Quaiser T, Dittrich A, Schaper F, Mönnigmann M. A simple work flow for biologically inspired model reduction-application to early JAK-STAT signaling. BMC Systems Biology. 2011; 5(1):30. doi: 10.1186/1752-0509-5-30 PMID: 21338487

12. Maurya M, Bornheimer S, Venkatasubramanian V, Subramaniam S. Reduced-order modelling of biochemical networks: application to the GTPase-cycle signalling module. IEE Proceedings-Systems Biology. 2005; 152(4):229–242. doi: 10.1049/ip-syb:20050014 PMID: 16986265

13. Petzold L, Zhu W. Model reduction for chemical kinetics: An optimization approach. American Institute of Chemical Engineers AIChE Journal. 1999; 45(4):869. doi: 10.1002/aic.690450418

14. Transtrum MK, Qiu P. Model reduction by manifold boundaries. Physical Review Letters. 2014; 113(9):098701. doi: 10.1103/PhysRevLett.113.098701 PMID: 25216014

15. Apri M, de Gee M, Molenaar J. Complexity reduction preserving dynamical behavior of biochemical networks. Journal of Theoretical Biology. 2012; 304:16–26. doi: 10.1016/j.jtbi.2012.03.019 PMID: 22465110

16. Venzon D, Moolgavkar S. A method for computing profile-likelihood-based confidence intervals. Applied Statistics. 1988; 37(1):87–94. doi: 10.2307/2347496

17. Schoeberl B, Eichler-Jonsson C, Gilles ED, Müller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nature Biotechnology. 2002; 20(4):370–375. doi: 10.1038/nbt0402-370 PMID: 11923843

18. Swameye I, Müller T, Timmer J, Sandra O, Klingmüller U. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(3):1028–1033. doi: 10.1073/pnas.0237333100 PMID: 12552139

19. Becker V, Schilling M, Bachmann J, Baumann U, Raue A, Maiwald T, et al. Covering a broad dynamic range: information processing at the erythropoietin receptor. Science. 2010; 328(5984):1404–1408. doi: 10.1126/science.1184913 PMID: 20488988

20. Kholodenko BN. Cell-signalling dynamics in time and space. Nature Reviews Molecular Cell Biology. 2006; 7(3):165–176. doi: 10.1038/nrm1838 PMID: 16482094

21. Kreutz C, Rodriguez MB, Maiwald T, Seidl M, Blum H, Mohr L, et al. An error model for protein quantification. Bioinformatics. 2007; 23(20):2747–2753. doi: 10.1093/bioinformatics/btm397 PMID: 17768165

22. Fisher RA. On an absolute criterion for fitting frequency curves. Messenger of Mathematics. 1912; 41:155–160.

23. Raue A, Schilling M, Bachmann J, Matteson A, Schelker M, Kaschek D, et al. Lessons learned from quantitative dynamical modeling in systems biology. PloS ONE. 2013; 8(9):e74335. doi: 10.1371/journal.pone.0074335 PMID: 24098642

24. Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, Shumaker DE, et al. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. ACM Transactions on Mathematical Software (TOMS). 2005; 31(3):363–396. doi: 10.1145/1089014.1089020

25. Coleman TF, Li Y. An interior, trust region approach for nonlinear minimization subject to bounds. SIAM Journal on Optimization. 1996; 6:418–445. doi: 10.1137/0806023

26. Leis JR, Kramer MA. The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. ACM Transactions on Mathematical Software (TOMS). 1988; 14(1):45–60.

27. Raue A, Steiert B, Schelker M, Kreutz C, Maiwald T, Hass H, et al. Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. Bioinformatics. 2015; 31(21):3558–3560. doi: 10.1093/bioinformatics/btv405 PMID: 26142188

28. Feder PI. On the distribution of the log likelihood ratio test statistic when the true parameter is "near" the boundaries of the hypothesis regions. The Annals of Mathematical Statistics. 1968; 39(6):2044–2055. doi: 10.1214/aoms/1177698032

29. Merkt B, Timmer J, Kaschek D. Higher-order Lie-symmetries in identifiability and predictability analysis of dynamic models. Phys Rev E. 2015; 92(1):12–20. doi: 10.1103/PhysRevE.92.012920

30. Steiert B, Raue A, Timmer J, Kreutz C. Experimental Design for Parameter Estimation of Gene Regulatory Networks. PLoS ONE. 2012; 7(7):e40052. doi: 10.1371/journal.pone.0040052 PMID: 22815723

31. Heinrich R, Neel BG, Rapoport TA. Mathematical models of protein kinase signal transduction. Molecular Cell. 2002; 9(5):957–970. doi: 10.1016/S1097-2765(02)00528-2 PMID: 12049733

32. D'Arcangelo G. The reeler mouse: anatomy of a mutant. International review of neurobiology. 2005; 71:383–417. doi: 10.1016/S0074-7742(05)71016-3 PMID: 16512359

33. Tissir F, Goffinet AM. Reelin and brain development. Nature Reviews Neuroscience. 2003; 4(6):496–505. doi: 10.1038/nrn1113 PMID: 12778121

34. Herz J, Chen Y. Reelin, lipoprotein receptors and synaptic plasticity. Nature Reviews Neuroscience. 2006; 7(11):850–859. doi: 10.1038/nrn2009 PMID: 17053810

35. Drakew A, Frotscher M, Deller T, Ogawa M, Heimrich B. Developmental distribution of a reeler gene-related antigen in the rat hippocampal formation visualized by CR-50 immunocytochemistry. Neuroscience. 1997; 82(4):1079–1086. doi: 10.1016/S0306-4522(97)00326-6

36. Fatemi S. Reelin mutations in mouse and man: from reeler mouse to schizophrenia, mood disorders, autism and lissencephaly. Molecular psychiatry. 2001;. doi: 10.1038/sj.mp.4000129 PMID: 11317213

37. Caffrey JR, Hughes BD, Britto JM, Landman KA. An in silico agent-based model demonstrates Reelin function in directing lamination of neurons during cortical development. PloS one. 2014; 9(10): e110415. doi: 10.1371/journal.pone.0110415 PMID: 25334023

38. Bock HH, Herz J. Reelin activates SRC family tyrosine kinases in neurons. Current Biology. 2003; 13 (1):18–26. doi: 10.1016/S0960-9822(02)01403-3 PMID: 12526740

39. Feng L, Cooper JA. Dual functions of Dab1 during brain development. Molecular and Cellular Biology. 2009; 29(2):324–332. doi: 10.1128/MCB.00663-08 PMID: 18981215