

RESEARCH ARTICLE

Sparse Regression Based Structure Learning of Stochastic Reaction Networks from Single Cell Snapshot Time Series

Anna Klimovskaia^{1,2,3}, Stefan Ganscha^{1,2,3}, Manfred Claassen^{1,2*}

1 Institute for Molecular Systems Biology, ETH Zurich, Zurich, Switzerland, **2** Swiss Institute of Bioinformatics, Zurich, Switzerland, **3** Life Science Zurich Graduate School, Zurich, Switzerland

* claassen@imsb.biol.ethz.ch



OPEN ACCESS

Citation: Klimovskaia A, Ganscha S, Claassen M (2016) Sparse Regression Based Structure Learning of Stochastic Reaction Networks from Single Cell Snapshot Time Series. PLoS Comput Biol 12(12): e1005234. doi:10.1371/journal.pcbi.1005234

Editor: Daniel A Beard, University of Michigan, UNITED STATES

Received: June 8, 2016

Accepted: November 2, 2016

Published: December 6, 2016

Copyright: © 2016 Klimovskaia et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Software implementing the reactionet lasso and datasets can be found at http://www.imsb.ethz.ch/research/claassen/Software/reactionet_lasso.html.

Funding: The authors were supported by internal funds from ETH Zurich for this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Abstract

Stochastic chemical reaction networks constitute a model class to quantitatively describe dynamics and cell-to-cell variability in biological systems. The topology of these networks typically is only partially characterized due to experimental limitations. Current approaches for refining network topology are based on the explicit enumeration of alternative topologies and are therefore restricted to small problem instances with almost complete knowledge. We propose the *reactionet lasso*, a computational procedure that derives a stepwise sparse regression approach on the basis of the Chemical Master Equation, enabling large-scale structure learning for reaction networks by implicitly accounting for billions of topology variants. We have assessed the structure learning capabilities of the reactionet lasso on synthetic data for the complete TRAIL induced apoptosis signaling cascade comprising 70 reactions. We find that the reactionet lasso is able to efficiently recover the structure of these reaction systems, ab initio, with high sensitivity and specificity. With only < 1% false discoveries, the reactionet lasso is able to recover 45% of all true reactions ab initio among > 6000 possible reactions and over 10^{2000} network topologies. In conjunction with information rich single cell technologies such as single cell RNA sequencing or mass cytometry, the reactionet lasso will enable large-scale structure learning, particularly in areas with partial network structure knowledge, such as cancer biology, and thereby enable the detection of pathological alterations of reaction networks. We provide software to allow for wide applicability of the reactionet lasso.

Author Summary

Virtually all biological processes are driven by biochemical reactions. However, their quantitative description in terms of stochastic chemical reaction networks is often precluded by the computational difficulty of structure learning, i.e. the identification of biologically active reaction networks among the combinatorially many possible topologies. This work describes the *reactionet lasso*, a structure learning approach that takes advantage of novel, information-rich single cell data and a tractable problem formulation to achieve structure learning for problem instances hundreds of orders of magnitude larger

than previously reported. This approach opens the prospect of obtaining quantitative and predictive reaction models in many areas of biology and medicine, and in particular areas such as cancer biology, which are characterized by significant system alterations and many unknown reactions.

This is a *PLOS Computational Biology* Methods paper.

Introduction

Cellular processes are essentially implemented by networks of biochemical reactions. The topology of such networks is typically only partially known, rendering the identification of the correct network from experimental data a key challenge. Despite the importance of this task, only little progress has been made in devising methods to systematically and comprehensively infer topologies of non-trivial chemical reaction networks. In this work, we propose a sparse regression approach tailored to the task of large-scale model selection for chemical reaction networks.

Different model classes have been developed to describe biochemical reaction systems. In order of increasing level of detail these comprise statistical time series models, such as autoregressive models and dynamic Bayesian networks, deterministic ordinary differential equation or stochastic differential equation based kinetic models [1]. The choice of model class depends on prior information for the system of interest and type of experimental data. Single cell technologies furnish further statistical information about component distributions, e.g. variances and covariances, aiding in systems identification [2] and are expected to become increasingly prevalent in routine biological research [3].

Two main computational tasks arise when learning any of these models from data: parameter inference, and structure learning. Parameter inference aims at finding model parameters (e.g. kinetic rate constants). Parameter inference has been performed by sampling from posterior parameter distributions, or global non-convex or convex optimization methods [4]. Structure learning aims at additionally identifying the reaction network topology governing the dynamics of the system components.

Parameter inference becomes increasingly computationally intensive for larger systems with numerous parameters [1]. Structure learning for these systems is an even more daunting task since parameter inference has to be performed for each of the possibly very many different system topologies. Therefore, structure learning is typically confined to comparison of a small, carefully selected set of candidate topologies by means of model selection criteria, such as information criteria (e.g. AIC, BIC) or Bayes Factors [5–7]. However, this approach requires substantial prior knowledge about the studied system in order to identify reasonable candidate models. Systematic approaches to enumerate a subset of sensible topologies have not been reported until recently. These approaches implement greedy strategies that either iteratively reduce the number of reactions of an overcomplete system of reactions or add reactions one at a time to a system with a minimal set of reactions [8]. However, such greedy approaches do not guarantee finding globally optimal topologies for non-convex fitting objectives. Furthermore, exploration of the multitude of local optima in the context of combinatorially many possible topologies becomes computationally prohibitive due to the requirement to explicitly

evaluate every considered candidate topology. No global approaches have been reported to perform structure learning by comprehensively evaluating model candidates for stochastic chemical reaction networks.

We propose the *reactionet lasso*, a convex relaxation of the structure learning task. This approach yields a single best sparse reaction set from all possible reactions by translating a recent sparse identification approach for nonlinear dynamic systems [9] to operate on and deal with non-trivial application specific parameter and noise structure for time series snapshot data acquired for stochastic chemical reaction networks.

Results

Sparse regression for structure learning of stochastic reaction networks

Structure learning by the reactionet lasso takes advantage of the formal link between the chemical reaction model and the observed data that is defined by the Chemical Master Equation. This differential equation system describes the temporal evolution of the abundance distributions of species governed by a stochastic chemical reaction network [10]. The moment generating functions of the Chemical Master Equation give rise to the moment equations, a system of ordinary differential equations for the temporal evolution of the central moments M_r of the abundance distributions (see [Methods](#)).

$$\dot{M}_r = \sum_l k_l F_{rl}(t; \mathbf{M}), \quad (1)$$

with rate constants k_b , time t and set of all central moments of individual species \mathbf{M} . For mass action kinetics the terms $F_{rl}(t; \mathbf{M})$ are polynomials over these moments such as abundance means and variances of individual species. $F_{rl}(t; \mathbf{M})$ will be referred to as stoichiometric moment functions herein (see also [S1 Text](#)).

The moment equations constitute the formal link between the time series snapshot data and the rate constants of the underlying chemical reactions. Rate constant estimation for stochastic mass action kinetics reaction networks in this context therefore reduces to parameter estimation for the ordinary differential equation system [Eq 1] with stoichiometric moment functions determined from the time series data.

Parameter estimation for a mass action kinetics network typically requires the costly integration of the moment equations for every considered parameter configuration. Imputation of the moment gradients by gradient matching procedures (see [Methods](#)) circumvents these type of evaluations and, in conjunction with the empirical moments, allows for parameter inference by means of a non-negative linear regression task with the least squares estimate $\hat{\mathbf{k}}$ for rate constants \mathbf{k} given by:

$$\hat{\mathbf{k}} = \arg \min_{\mathbf{k} \geq 0} \|\hat{\mathbf{b}} - \hat{\mathbf{A}}\mathbf{k}\|_2^2, \quad (2)$$

where the response vector elements b_j corresponds to the vector of empirical gradient estimates for $\dot{M}_j(t)$ from the gradient matching procedure (see [Methods](#)) and the design matrix

entries \hat{A}_{ji} correspond to the estimates of the stoichiometric moment functions $F_{ji}(t; \mathbf{M})$:

$$\mathbf{b} = \begin{bmatrix} \hat{M}_{r_1}(t_1) \\ \vdots \\ \hat{M}_{r_1}(t_T) \\ \hat{M}_{r_2}(t_1) \\ \vdots \\ \hat{M}_{r_2}(t_T) \\ \vdots \\ \hat{M}_{r_N}(t_1) \\ \vdots \\ \hat{M}_{r_N}(t_T) \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \hat{F}_{r_1 1}(t_1) & \hat{F}_{r_1 2}(t_1) & \cdots & \hat{F}_{r_1 L}(t_1) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{F}_{r_1 1}(t_T) & \hat{F}_{r_1 2}(t_T) & \cdots & \hat{F}_{r_1 L}(t_T) \\ \hat{F}_{r_2 1}(t_1) & \hat{F}_{r_2 2}(t_1) & \cdots & \hat{F}_{r_2 L}(t_1) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{F}_{r_2 1}(t_T) & \hat{F}_{r_2 2}(t_T) & \cdots & \hat{F}_{r_2 L}(t_T) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{F}_{r_N 1}(t_1) & \hat{F}_{r_N 2}(t_1) & \cdots & \hat{F}_{r_N L}(t_1) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{F}_{r_N 1}(t_T) & \hat{F}_{r_N 2}(t_T) & \cdots & \hat{F}_{r_N L}(t_T) \end{bmatrix}.$$

This linear regression formulation has been applied for parameter inference of deterministic chemical reaction models [6, 11].

Model selection across small sets of model variants has previously been performed with information criteria [11] or model averaging [6]. The Lasso constitutes another approach for efficient and comprehensive model selection in linear regression models [12]. It introduces an L1 norm ($\|\cdot\|_1$) regularization on the parameters \mathbf{k} to promote the identification of sparse solutions, i.e. solutions with many zero-valued parameter estimates.

$$\hat{\mathbf{k}} = \arg \min_{\mathbf{k} \geq 0} \|\hat{\mathbf{b}} - \hat{\mathbf{A}}\mathbf{k}\|_2^2 + \lambda \|\mathbf{k}\|_1, \quad (3)$$

Various extensions of the Lasso method were introduced in literature to improve its shrinkage properties in the presence or absence of heteroscedasticity [13].

While the Lasso has been used in recent reports to identify general nonlinear dynamical systems [9] or to select the mechanism types (mass action or Hill kinetics) of a fixed reaction set defined by the deterministic Repressilator comprising six components [14], it still remains to adapt the regression model and regularization concepts to enable more comprehensive model selection for realistic reaction systems that exhibit stochasticity and larger amount of components/reactions. The next sections will delineate in detail the challenges and solutions implemented in the reactionet lasso to achieve this goal.

Reactionet lasso

This section introduces the reactionet lasso (Fig 1), a computational method for learning the structure of chemical reaction networks. The overarching strategy of this procedure consists of (1) enumerating all (or at least a significant fraction of reasonable) conceivable unary/binary reactions between the components of a reaction system of interest and (2) applying an appropriate stepwise sparse regression approach to select the sparse subset of reactions underlying the observed dynamics in the snapshot time series data.

The following properties of such structure learning instances preclude the application of conventional least squares based approaches for parameter estimation and selection: (1) noise and heteroscedasticity of the observed response $\hat{\mathbf{b}}$ (empirical moment gradient estimates) as

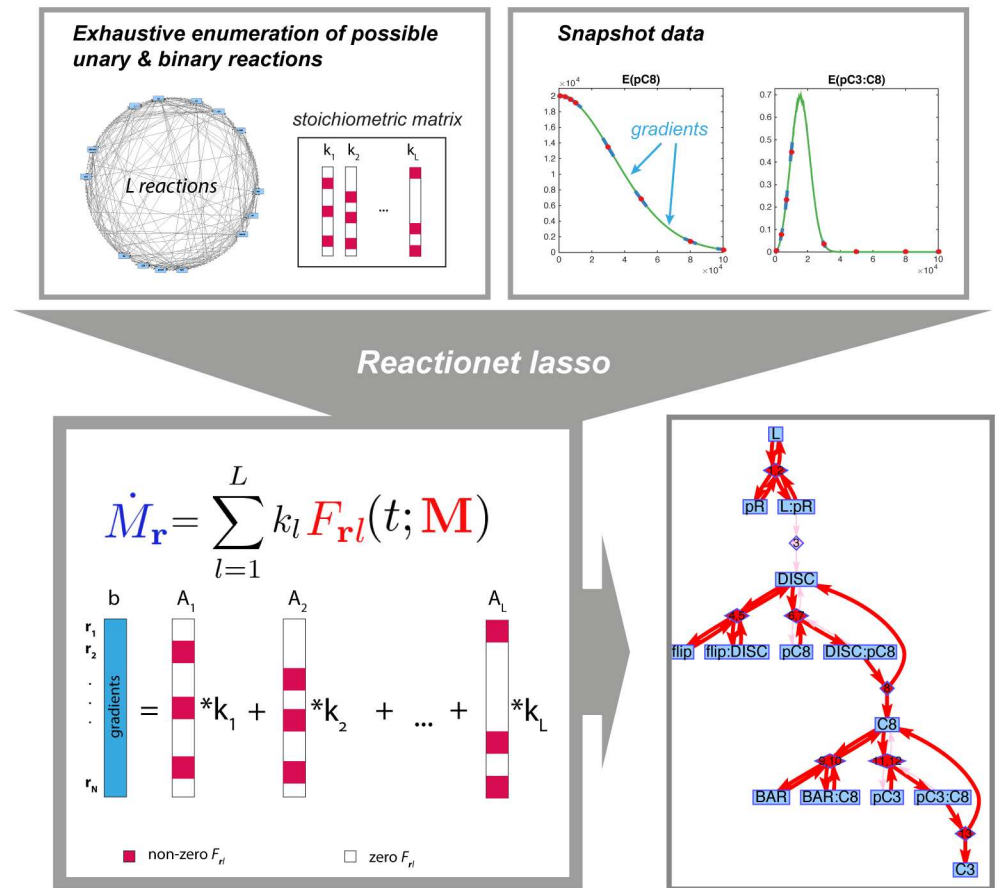


Fig 1. Schematic representation of *reactionnet lasso* procedure. For details see section [Methods](#).

doi:10.1371/journal.pcbi.1005234.g001

well as in the observed design matrix $\hat{\mathbf{A}}$ (stoichiometric moment function evaluations) and (2) different scales of individual parameters k_i (rate constants) resulting from the occurrence of large a spectrum of fast and slow reactions. The reactionnet lasso addresses each of these challenges in as delineated in the following.

The intrinsic variability of stochastic chemical kinetics result induces variability of the empirical estimates of moments and their gradients. Therefore the observed response vector as well as the stoichiometric moment functions in the design matrix are expected to deviate from the true latent correspondents. We capture this by defining $\mathbf{b} = \hat{\mathbf{b}} + \epsilon_b$ and $\mathbf{A} = \hat{\mathbf{A}} + \epsilon_A$ to be the true latent moment gradients and stoichiometric moment functions, and ϵ_A and ϵ_b to be their respective intrinsic variability induced deviations from the estimated/observed quantities. If we knew the true values of the latent variables, finding the rate constants \mathbf{k} would translate to solving the following equation:

$$\mathbf{b} = \mathbf{A}\mathbf{k}. \tag{4}$$

By substituting the variables in [eq 4](#) with the definitions for our empirical estimates of the latent variables we obtain:

$$\hat{\mathbf{b}} = \hat{\mathbf{A}}\mathbf{k} + \epsilon. \tag{5}$$

with $\epsilon := \epsilon_A \mathbf{k} - \epsilon_b$.

Eq 5 seems to motivate a straightforward optimization strategy to compute a maximum likelihood parameter estimate given the observations for moment gradients and stoichiometric moment functions (e.g. least squares for independent and normally distributed residuals ϵ). However, it becomes apparent that this strategy is not valid due to the residual ϵ being a function of the parameters \mathbf{k} (by virtue of the noise in the observed design matrix).

The reactionet lasso implements a stepwise strategy to address this dependency. The first step (Step 1) is a Feasible Generalized Least Squares (FG) estimate. It comprises the estimation of the variances of the residuals ϵ_b and ϵ_A via bootstrapping of the gradient estimates and stoichiometric moment functions on the basis the single-cell data. A preliminary least squares fit is then performed to achieve an estimate \mathbf{k}^{LS} for eq 5. This estimate is expected to approximate the order of magnitude of the individual rate constants. In conjunction with the estimates of the variances of the residuals ϵ_b and ϵ_A , we use \mathbf{k}^{LS} to achieve an estimate of the component-wise variance $\Sigma_\epsilon = \text{diag}\{\sigma_{\epsilon_1}^2, \dots, \sigma_{\epsilon_R}^2\}$ of the residuals ϵ . To achieve this estimate we use only first order moments (means), as they are less subjected to noise in the design matrix and provide a more robust estimate of the covariance matrix Σ_ϵ . This estimate will allow us to operate with the rescaled observed response vector $\hat{\mathbf{b}}^s = \Sigma_\epsilon^{-1/2} \hat{\mathbf{b}}$ and design matrix $\hat{\mathbf{A}}^s = \Sigma_\epsilon^{-1/2} \hat{\mathbf{A}}$ to adjust for heteroscedasticity and enable effective linear regression [15].

$$\hat{\mathbf{k}}^{FG} = \arg \min_{\mathbf{k} \geq 0} \|\hat{\mathbf{b}}^s - \hat{\mathbf{A}}^s \mathbf{k}\|_2^2, \tag{6}$$

The subsequent steps aim at addressing the second challenge introduced above, i.e. the different scales of individual parameters k_i , which render conventional sparse regression approaches (such as the Lasso) suboptimal due to the uniform penalization strength of the L1 norm $\|\cdot\|_1$ across all components k_i of the parameter vector k . The adaptive Lasso [16] constitutes an alternative to the conventional Lasso. It defines a regularization penalty that is scaled component-wise by the expected order of magnitude \hat{k}_i of the respective component i .

In Step 2 of the reactionet lasso, we apply a combination of the adaptive and relaxed Lasso, stability selection based prioritization of reactions and an additional stepwise backward regression to achieve the final set of reported reactions. We use the parameter estimates from Step 1 (obtained with Moore–Penrose pseudoinverse matrix), i.e. $\tilde{\mathbf{k}} = \hat{\mathbf{k}}^{FG}$, in order to adapt the regularization penalty.

To improve shrinkage, the adaptive Lasso is followed by a relaxed Lasso [17] that recomputes optimal parameter estimates with respect to the objective specified in eq 6, while only considering the set $\Phi = \{l : k_l^{FG} \neq 0\}$ of parameters that were not set to zero in Step 1, for which the optimal solution is

$$\hat{\mathbf{k}}^{ARL} = \arg \min_{\mathbf{k} \geq 0} \|\hat{\mathbf{b}}^s - \hat{\mathbf{A}}^{s,\Phi} \mathbf{k}\|_2^2 + \lambda \sum_i |k_i / \tilde{k}_i|, \tag{7}$$

where $\hat{\mathbf{A}}^{s,\Phi}$ contains only that columns of $\hat{\mathbf{A}}^s$, which are in a set Φ .

The adaptive relaxed Lasso solution has been computed by optimizing the respective Alternating Direction Method of Multipliers (ADMM) formulations [18]. The adaptive relaxed Lasso is performed with five fold cross validation. We used stability selection to prioritize reactions according to their frequency of being selected across all cross validation folds [19]. Bayesian information criterion (BIC) was used as selection criterion (S2 Text).

In summary, the reactionet lasso procedure constitutes a stepwise sparse regression approach that addresses the parameter-dependent noise and heteroscedasticity in the response and design matrix for structure learning of stochastic chemical reaction systems. See also Fig 1

for a schematic overview of its steps. Software implementing the reactionet lasso can be found at http://www.imsb.ethz.ch/research/claassen/Software/reactionet_lasso.html.

Ab initio structure learning of chemical reaction networks

We first consider an extreme and yet conceptually simple scenario where we aim at learning the structure of a reaction network without any prior knowledge about the underlying reactions. While this scenario rarely occurs in a real world application because typically some prior knowledge of relevant reactions is available, we first investigate this scenario to demonstrate the structure learning capabilities of the reactionet lasso.

We study two systems varying in number of components and reactions: (1) the enzymatic reaction system with four components and three reactions, (2) the receptor subunit of a recently reported kinetic model of TRAIL induced apoptosis with fourteen components and thirteen reactions, which can be combined in a total of 2275 possible unary or binary reactions, giving a total of more than 10^{600} possible reaction network candidates. For these systems we simulated 5 replicates each with either 10^3 , 10^4 or 10^5 single cell trajectories with the stochastic simulation algorithm [20]. We then generated snapshot time series datasets from the single cell trajectories by defining pools of cells at selected sets of 7, 13, or 28 time points. Moment gradients were estimated either with the smoothing procedure, cubic splines or the finite difference scheme (see [Methods](#)).

The reactionet lasso achieves structure learning of chemical reaction networks via a two step sparse regression formulation that (1) specifically accounts for heteroscedasticity in the response vector and the design matrix of the regression instances and (2) assumes a regularizer that encourages sparse reaction sets by suppressing compensatory reaction sets with small rate constants ([Fig 1](#)). The first step aims at accounting for heteroscedasticity and, most importantly at reducing the number of reaction candidates for the second step that both capture the empirical moment gradients and select for correct reactions ([Fig 2](#)). The following results are based on moment equations for all moments up to order two, i.e. means, variances and covariances. Following Step 1 of the reactionet lasso, we achieve a substantial reduction to less than 100 candidate reactions that, regardless of the moment gradient estimation technique, retains at least ten of the thirteen true reactions ([Fig 1A](#)). The vast majority of the empirical moment gradients are well fit by the set of candidate reactions. The few moment gradients that are sub-optimally captured correspond to higher order moments such as variances or covariances whose highly dynamic behavior precluded accurate gradient estimation by either the finite difference or spline fit. ([Fig 2B](#)). Step 2 of the reactionet lasso procedure uses a relaxed adaptive Lasso estimator to estimate the rate constants of a sparse set of candidate reactions following from Step 1. The method recovers ten out of thirteen reactions correctly with one false positive reaction when assuming no prior knowledge and selecting a suitable model with BIC ([S1A Fig](#)). Similar performance is achieved for the enzymatic reaction network ([S1B Fig](#)). These results demonstrate that the stepwise sparse regression strategy of Step 2 completes the structure learning task from the candidate reactions supplied by Step 1 with great sensitivity and specificity. In summary, the reactionet lasso is able to ab initio reconstruct the reaction network structure of typically-sized signaling cascades such as the fourteen component receptor subunit of TRAIL induced apoptosis [21].

We further evaluated the impact of different gradient estimation approaches on structure learning performance ([S2–S4 Figs](#)). For benchmarking purposes we used the smoothed empirical moment gradient estimate as a ground truth which is not available in a real time series snapshot setting. According to these considerations, the cubic spline estimator achieves almost optimal performance for thirteen or more time points, whereas FDS is consistently inferior.

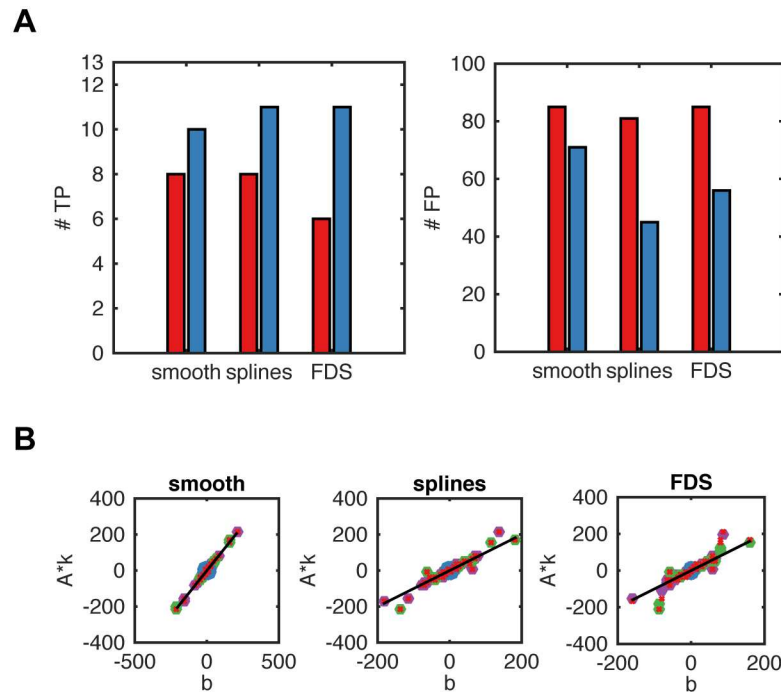


Fig 2. Performance assessment of the first reactionet lasso step for 10^5 single cell trajectories of the apoptotic receptor subunit. (A) Enrichment and depletion of true and respectively false positive reactions for the reaction rate estimates \hat{k}^{LS} (red) and \hat{k}^{FG} (blue). Results are reported for gradient estimation procedures smooth, FDS, splines (see main text for details). (B) Comparison of response (empirical moment gradients) and prediction with feasible generalized least square estimate for moments of different order: means (blue), variances (green), covariances (yellow) and prediction with true rate constants for all moments (red crosses).

doi:10.1371/journal.pcbi.1005234.g002

These results indicate that the cubic spline estimator provides the most favorable structure learning performance for empirical moment gradients.

We evaluated how measurement noise affects the ability of reactionet lasso to learn the reaction network structure. We assume a binomial measurement noise model that reflects the incomplete capture efficiency inherent to all single cell technologies (see [Methods](#), [S3 Text](#)). While structure learning performance is reduced with increasing levels of measurement noise, the reactionet lasso still recovers more than 50% of the reactions for the apoptotic receptor subunit at levels reported for single cell sequencing and mass cytometry approaches ([Fig 3](#), [S5 Fig](#)).

To assess the relative importance of the amount of available data, we varied the amount of time points and single cell recordings used at each time point. Interestingly, we found that the inclusion of additional measurement time points did not improve structure learning performance. However, the tradeoff between true and false positive reaction discoveries worsened considerably with fewer time points ([Fig 4A](#)). While we found that decreasing the amount of single cell measurements per time point did result in noticeable performance losses, this situation does not constitute a limitation for flow cytometry techniques, that are easily able to generate millions of single cell snapshots ([Fig 4B](#)). Cell count related performance losses can be associated with higher absolute variability and therefore reduced accuracy of empirical moment estimates ([S6 Fig](#)). We conclude that careful selection of amount of single cell measurements and number as well as position of time points ([S7 Fig](#)) translates to accurate

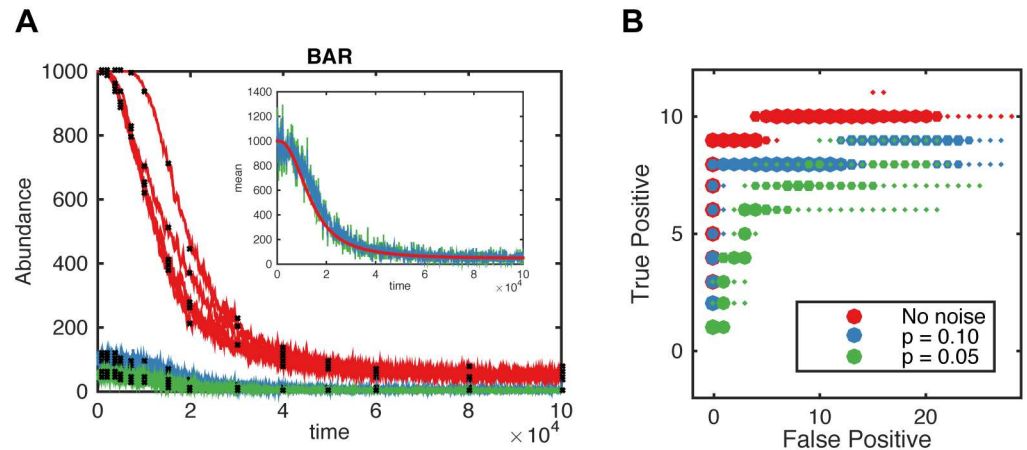


Fig 3. Influence of measurement noise on structure learning capacity. (A) Example of several single cell trajectories of one of the species (BAR) in apoptotic receptor subunit: without measurement noise (red), with measurement noise according to the binomial model with probability of success $p = 0.1$ (blue), 0.05 (green). Comparison of reconstructed means for known p between different noise levels shows how empirical moments are affected by measurement noise. Black dots represent snapshot measurements used for the inference procedure. (B) Overlay of five regularization paths in terms of true/false positive tradeoff for different measurement noise levels as indicated in the legend in terms of binomial capture efficiency. Structure learning performance for 10^5 single cell trajectories and thirteen time points of the apoptotic receptor subunit. Empirical moment gradients estimated with splines.

doi:10.1371/journal.pcbi.1005234.g003

interpolation and subsequent gradient fitting, thereby leading to good structure learning performance of the reactionet lasso.

We further investigated the impact of including different moment orders for structure learning. As expected, precisely estimated higher-order moments contain a substantial amount of information and therefore enhance the structure learning capability accordingly (Fig 5A). However, although this relationship still holds for medium levels of measurement noise (capture efficiency $p = 0.1$), (Fig 5B), the inclusion of second order moments becomes misleading

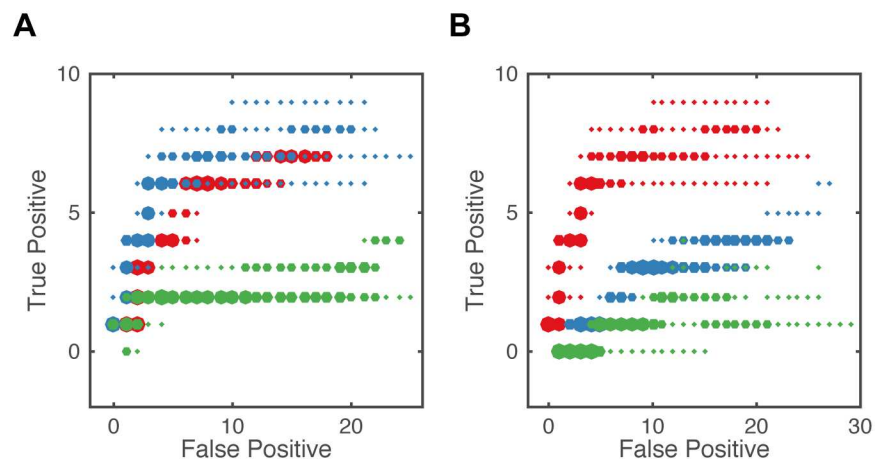


Fig 4. Overlay of five regularization paths in terms of true/false positive tradeoff over different data availability situations. Results for *reactionet lasso* application to apoptotic receptor subunit ($p = 0.05$). Empirical moment gradients estimated with cubic splines. (A) 10^5 single cell trajectories evaluated at different amount of time points: 28 (red), 13 (blue), 7 (green). (B) Different number of single cell trajectories: 10^5 (red), 10^4 (blue), 10^3 (green) evaluated at thirteen time points.

doi:10.1371/journal.pcbi.1005234.g004

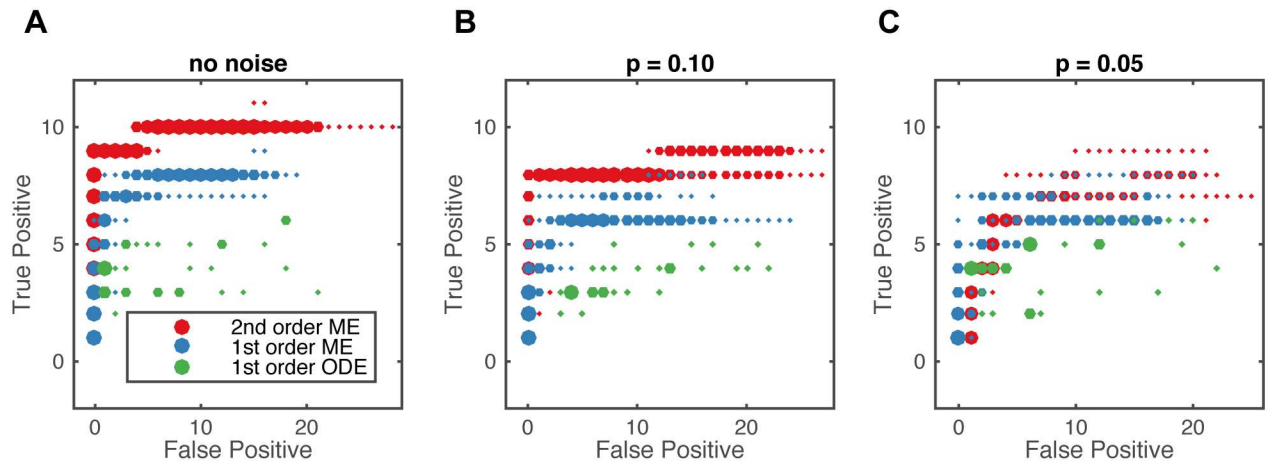


Fig 5. Overlay of five regularization paths in terms of true/false positive tradeoff for different moment orders of gradients considered for structure learning: up to 2nd order (red) Moment Equations, up to 1st order (blue) Moment Equations and only 1st order moments for deterministic mean ODE model (green). Structure learning performance for 10^5 single cell trajectories and thirteen time points of the apoptotic receptor subunit. Empirical moment gradients estimated with splines. Results represented for different levels of measurement noise: (A) no noise; (B) $p = 0.1$; (C) $p = 0.05$.

doi:10.1371/journal.pcbi.1005234.g005

for high levels of measurement noise (capture efficiency $p = 0.05$, Fig 5C). This observation is likely caused by the limited ability to accurately estimate higher order moments for high levels of measurement noise. However, the performance of the reactionnet lasso assuming stochastic kinetics modeled with moment equations (higher order ME) is consistently better than assuming a deterministic kinetics modeled with mean based ordinary equations (1st order ODE). This observation demonstrates that the incorporation of higher order moment information induced by the chemical kinetics and accessible by means of single cell measurements allows for significantly improved structure learning capacity.

In summary, the benchmarks above strongly advocate for the use of an experimental setup that allows for sufficiently dense sampling across time to ensure accurate empirical moment gradient estimates, as well as single cell technology, such as flow/mass cytometry, which provide 10^4 or more single cell measurements, for the accurate estimation of higher order moments. In these situations the reactionnet lasso is capable of ab initio recovery of almost the complete reaction network structure with more than a dozen components.

Structure learning of large chemical reaction networks with prior knowledge

We now consider a scenario where we aim at learning the structure of a large reaction network with partial knowledge about the underlying reactions. For this situation we demonstrate how reactionnet lasso is capable of recovering a sizable amount of the unknown reactions, for a reaction network as large as the 70 reaction TRAIL induced apoptosis cascade [21].

Structure learning tasks for chemical reaction networks typically aim at complementing already available partial knowledge on reaction sets. We assessed the ability of the reactionnet lasso to complement a set of known reactions for the 70 reaction TRAIL induced apoptosis cascade. Specifically, we defined six modules for this cascade following [21], and assumed a limited set of 22 reaction candidates connecting these modules (S8 Fig) and 33 uniformly distributed time points (S1 Dataset). For step 1 of the reactionnet lasso all possible unary and binary reactions between components within modules and the module connecting reactions serve as candidate reactions for structure learning, totaling 6828 reactions.

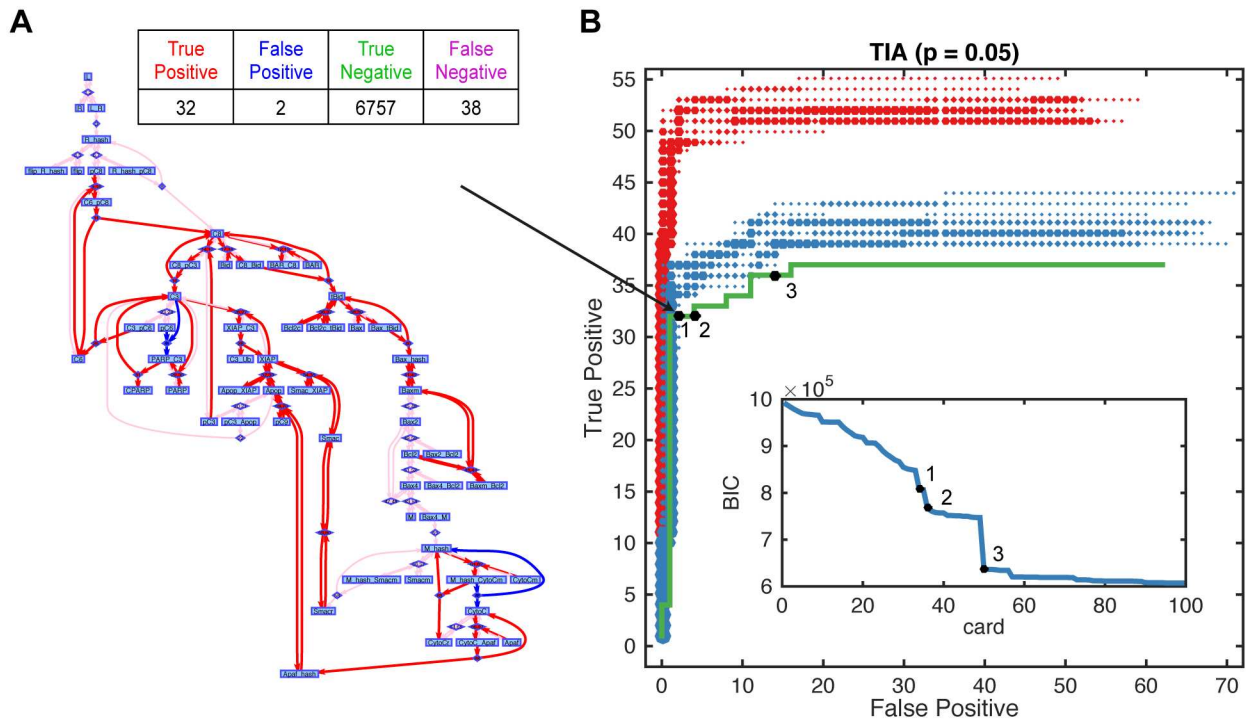


Fig 6. Structure learning with prior knowledge for the 70 component TRAIL induced apoptosis cascade. Structure learning performance for 10^5 single cell trajectories and 33 time points and capture efficiency $p = 0.05$. Empirical moment gradients estimated with splines. (A) Example of a recovered graph for the setting above and no prior knowledge. True positive reactions are red, false positive reactions are blue, false negative reactions are pink. (B) Regularization paths in terms of true/false positive tradeoff (including prior knowledge reactions, see Results for details) for different prior knowledge situations. Following prior knowledge situations are depicted: no prior knowledge (green). Additional prior knowledge comprise ten instances of 10%(blue)/ 50%(red) randomly chosen known reactions. Diameter of dots and color code indicate frequency of solutions with a specific true/false positive tradeoff. Black dots represent solutions coinciding with large improvements of BIC.

doi:10.1371/journal.pcbi.1005234.g006

In the absence of ground truth it is difficult to identify a regularization strength that achieves a desirable tradeoff between true and false positive reaction discoveries. We evaluated the BIC and report solutions that map to large initial improvements of BIC [22]. Structure learning without prior knowledge on the considered set of reactions achieves 32 true positive at the cost of 2 false positive reactions (10^5 single cell trajectories, 33 time points, capture efficiency 0.05, Fig 6A). Prior knowledge on a specific reaction was encoded by a positivity constraint on the corresponding reaction rate during all regression steps of the reactionet lasso. We considered different prior knowledge settings: (1) 10% or (2) 50% randomly chosen reactions considered to be known. Settings (1) and (2) were each evaluated using ten different subsets. For 10% known reactions almost 40 (including 7 known) out of 70 reactions are correctly recovered with five or less false positive discoveries (Fig 6B). For 50% known reactions the total number of true positive reactions is beyond 50 (including 35 known). The performance doesn't depend significantly on the choice of prior reactions. The reactionet lasso enables discovery up to dozens of novel reactions at the cost of few false positive reactions for a large signaling cascade comprising 70 reactions.

While published structure learning approaches are only available for problem instances of sizes hundreds of orders of magnitude smaller, we compared these results to more simple variants of the reactionet lasso procedure, either exhibiting inferior accuracy or exceedingly high computational complexity (S9 Fig).

The above results demonstrate how single cell snapshot time series data and the reactionet lasso can be used to complement prior mechanistic knowledge by a sizable set of candidate reactions that is highly enriched for true positive discoveries, and do so for systems and structure learning tasks of unprecedented size [23, 24].

Discussion

In this work we introduce the reactionet lasso for comprehensive structure learning of stochastic chemical reaction networks.

Chemical reaction networks constitute a highly detailed and mechanistic description for biological processes and are qualitatively different from other popular network models in biology. These comprise probabilistic graphical models seeking to discover statistical dependencies between measured system components. These approaches range from simple correlation [25] or regression analysis [26], to Bayesian networks [27] or more structured and robust module networks [28] and extensions thereof [29]. In contrast to chemical reaction networks, each of these model classes allows for detection of statistical dependencies without further elucidation of causality relationships and the possibly intricate dependency inducing biochemical mechanism. Physical interaction networks get closer to this goal and complement the information of reaction networks by summarizing measurements of static protein interactions [30].

By virtue of formulating the task of structure learning of chemical reaction networks as a sequence of convex optimization problems, this procedure is able to assess an unprecedented number of potential network topologies without need for explicit enumeration [23, 24]. We demonstrate the utility of the method for ab initio structure learning of whole signaling cascades such as the apoptotic receptor subunit. The reactionet lasso originally integrates a moment based description of stochastic reaction networks with sparse regression approaches via a gradient matching to achieve an efficient and scalable structure learning procedure, overcoming the limitations of available methods for structure learning which either explicitly enumerate a small set of models or greedily search for locally optimal topologies [8, 31]. Recent generic sparse regression approaches for identification of general nonlinear dynamical systems are in principle applicable for structure learning of biological reaction networks [9]. However, these approaches, in contrast to the reactionet lasso, do not take into account their (1) foundation in the Chemical Master Equation, (2) heteroscedastic and parameter dependent noise structure, as well as (3) parameter ranges varying across many scales, therefore failing to achieve competitive structure learning performance (S9 Fig).

The challenging structure learning task crucially depends on sensible experimental design yielding informative data. A central design choice concerns the selection of time points recording the relevant dynamical changes of the process of interest. These are typically chosen from prior knowledge or preliminary dense snapshot time series experiments with a cheap readout, such as population based instead of single cell measurements. Another important experimental parameter concerns the number of single cell snapshots. Our benchmarks advocate for having at least thousands of snapshots per time point. Flow cytometry experiments easily achieve snapshot counts in the order of 10^5 . For single cell transcriptomics experiments it seems advisable to resort to novel droplet based techniques achieving $> 10^4$ single cell snapshots per experiment [32, 33].

Structure learning performance of the reactionet lasso depends on the accuracy of the gradient estimates from the time series snapshots. We find that estimates based on gradients obtained by rather simple approaches such as finite difference approximations or spline curve fitting achieve competitive performance. Improvements are conceivable by resorting to other techniques specifically designed for gradient estimation in differential equation systems

[34, 35]. These approaches jointly fit parameters of the curve fitting procedure and the differential equation system. The success of this strategy relies on considering a problem instance where the differentiation equation systems strongly constrains the state space. However, the problem instances we consider for systematic structure learning with the reactionet lasso assume a differential equation system defined by the moment equation for all possible unary and binary reactions. Such a system will by definition impose little constraints on the state space. Gradient matching approaches would therefore have to be adapted to avoid expected parameter overfitting resulting from their application to problem instances with such an expressive differential equation system.

For this proof of principle study, we consider single time series experiments. Reactionet lasso analysis easily accommodates multiple replicates or perturbation experiments such as dose responses. Specifically, condition specific response vectors b_k and design matrices A_k for each condition k are utilized to construct a problem instance by concatenation. For this problem instance reactionet lasso can be applied as described (see also [S4 Text](#)). Additional experiments are expected to enhance structure learning performance. Indeed, we observe that this is the case for incorporating additional replicate time series experiments ([S10 Fig](#)).

The reactionet lasso is able to recover a significant proportion of missing reactions in various settings. However, integration of the moment equations for the component means assuming this set is not always able to recover the observed temporal dynamics of the system. This situation arises for instance when a single pivotal true reaction is missed and therefore precludes the correct reconstruction of downstream component dynamics. We frequently encounter this situation in the ab initio structure learning scenario ([S11 Fig](#)). This scenario though constitutes an artificial setting that we only report for a proof of concept of the reactionet lasso. Real world applications comprise prior knowledge about true reactions, typically comprising specifically those pivotal reactions. It turns out that we achieve good reconstruction of integrated trajectories for the structure learning settings assuming prior knowledge ([S11 Fig](#)).

Until now we consider reaction systems which obey mass action kinetics. Systems of this kind can be easily translated into a series of moment equations which depend linearly upon the reaction rates. However, systems with non-mass action kinetics, such as Michaelis-Menten, can still be addressed with the reactionet lasso. While appropriate moment closure approximations for certain rational rate law kinetics preserve convexity of the reactionet lasso objective [36], generally such kinetics might yield non-convex optimization problems that would have to be dealt with using appropriate optimization techniques.

The reactionet lasso generates a single point estimate for the optimal, sparse reaction network that might neglect other reasonable candidate network structures. Thus, it will be interesting to perform further in-depth analysis of the resulting network structures, for instance with Markov Chain Monte Carlo sampling techniques.

For our study we assume that all relevant molecular components can be measured. Many biological applications, however, do not allow monitoring all relevant components, as for instance antibodies might only be available for a subset of components of a signaling cascade. While the aim of our study was to demonstrate proof of concept for large scale structure learning of chemical reaction networks, it will be possible to account for missing measured components by either augmenting the model by introducing latent variables or ‘lumping’ them into more complex non-mass action reaction mechanisms [37].

The reactionet lasso can be applied in its current form to systems where a significant proportion of relevant components can be measured. Considering the steady advance of single-cell technologies, we expect an increasing number of cellular signaling and metabolic processes to be assayed at single-cell resolution. While mass cytometry approaches allow for

measurement of more than 30 protein components, sufficient e.g. to substantially map out the T cell receptor, epidermis growth factor and apoptosis signaling cascades, single-cell RNA sequencing opens the prospect of achieving genome-wide transcriptomic snapshots of single cells. Thus we anticipate a surge of relevant data in the near future for which the reactionet lasso can straightforwardly be applied for systematic and comprehensive structure learning of the underlying reaction networks, with direct implications for systems biology and health by providing quantitative and predictive models for scientific insight and rational intervention design.

Methods

Experimental setting: Single cell time series snapshot data

We assume time series data with single cell resolved population snapshots obtained at discrete time points. We denote by C the number of cells measured per experiment, T the number of time points at which measurements were performed and N the number of components (e.g. proteins) measured in each cell. For each measurement time point $t = 1, \dots, T$ for a cell $c = 1, \dots, C$ we denote a vector of measured N protein abundances $\mathbf{x}_t^c = \{x_{t,1}^c, \dots, x_{t,N}^c\}$. Therefore at each time point t vectors $x_{t,1}^c, \dots, x_{t,N}^c$ represent a sample from a high-dimensional distribution, which evolves according to the Chemical Master Equation.

Moment equations for the Chemical Master Equation

We assume a biochemical reaction network of N different chemical species with abundances X_1, \dots, X_N involved in L reactions. Each reaction l is characterized by stoichiometry vector s_l and propensity function $a_l(\mathbf{x}; k_l)$ with \mathbf{x} representing the collection of species abundances (system state) and k_l the reaction rate. In our work we consider systems described by mass action kinetics, resulting propensities $a_l(\mathbf{x}; k_l) = k_l^* g_l(\mathbf{x})$, where $g_l(\mathbf{x})$ is a known function of the system's state. The state of the system evolves probabilistically according to the possible reactions, with probability $P(\mathbf{x}, t)$ of occupying state \mathbf{x} at time t . The probabilistic evolution of the system's state is described by the Chemical Master Equation:

$$\frac{P(\mathbf{x}, t)}{dt} = \sum_{l=1}^L P(\mathbf{x} - \mathbf{s}_l, t) a_l(\mathbf{x} - \mathbf{s}_l) - P(\mathbf{x}, t) a_l(\mathbf{x}). \quad (8)$$

We denote by $M_{\mathbf{r}} = M_{r_1, \dots, r_N} = E(X_1 - EX_1)^{r_1} \dots (X_N - EX_N)^{r_N}$ the central moment of order $\mathbf{r} = (r_1, \dots, r_N)$. The moment generating function of the probability distribution $P(\mathbf{x}, t)$ can be used for the derivation of moment equations [38]. Assuming mass action kinetics, we obtain Eq 1 for the time evolution of a central moment (see also S1 Text).

Gradient matching for parameter estimation of ordinary differential equation systems

Gradient matching approaches avoid costly integration by instead interpolating the discrete snapshot time series data and estimating the empirical moment gradients $M_{\mathbf{r}}(t)$, rendering the initial ODE system an algebraic equation system with the parameters as unknowns. This formulation further eliminates the need for moment closure, in contrast to integration based techniques. Previously, gradients have been estimated with spline interpolators [34, 39, 40], Gaussian processes [35, 41] or finite difference approximations [6]. Parameter estimation has been performed by least squares minimization [34, 39, 40] or by approximation of the parameter posterior [6, 35, 41]. While deterministic chemical reaction networks frequently served as

application settings for gradient matching schemes, only little attention has been paid to networks with stochastic dynamics [6, 11].

We used and compared cubic spline interpolators (spline) and finite difference approximations (FDS) to estimate empirical moment gradients for the $M_r(t)$ of the moment equations. As a ground truth estimate for simulated data, we use a smoothed finite difference approximation of the single cell trajectories at the evaluation point of interest (smooth). Gradient estimates are obtained via a smoothing procedure that relies on a sliding window estimate of finite differences on the simulated trajectories using the smoothing function “smooth” in Matlab.

Moment estimation for noisy single cell data

Single cell data such as obtained from flow/mass cytometry and single cell sequencing exhibit measurement noise. These technologies each detect a random fraction of the total molecular content of every individual cell. This relationship between the measurement signal and cellular analyte abundance has been frequently modeled by a binomial distribution $Bi(X, p)$ whose success probability p corresponds to the capture efficiency for the analyte present at amount X [42, 43]. We have devised an estimator to subtract the misleading measurement noise component to provide the reactionet lasso with the appropriate noise-correct empirical moment estimates for structure learning.

We assume that measurement noise can be represented by the following binomial model. Let X represent the true abundance of one species at a given time point. Let X^{obs} be the corresponding measured signal, such that $X^{obs} \sim Bi(X, p)$, where p is the capture efficiency. The binomial noise model allows for specifying the following analytical relationships between the first and second order moments of X and X^{obs} :

$$E[X^{obs}] = pE[X], \tag{9}$$

$$E[(X^{obs})^2] = p(1 - p)E[X] + p^2E[X^2], \tag{10}$$

$$Var[X^{obs}] = p(1 - p)E[X] + p^2Var[X], \tag{11}$$

$$cov(X_1^{obs}, X_2^{obs}) = p^2cov(X_1, X_2), \tag{12}$$

For a derivation see S3 Text. We assume that the capture efficiency p of the single cell instrument is known [42, 43] and estimates the empirical moments of X on the basis of the empirical moments of X^{obs} by solving the above equations for the respective moment of X . The resulting moment estimates are then used in the regression procedure described above to perform structure learning.

Supporting Information

S1 Fig. Structure learning performance of the reactionet lasso. 10^5 single cell trajectories evaluated at 13 time points for (A) apoptotic receptor subunit (no measurement noise); (B) the enzymatic system. Empirical moment gradients estimated with cubic splines. Solution selected with Bayesian Information Criteria (BIC).

(TIF)

S2 Fig. Regularization paths in terms of true/false positive tradeoff over different data availability situations. Results for reactionet lasso application to apoptotic receptor subunit (no measurement noise). (A-B) Empirical moment gradients estimated with “smooth”

procedure: (A) 10^5 single cell trajectories evaluated at different amount of time points (tp) as indicated in the legend. (B) Different number of single cell trajectories: 10^3 , 10^4 , 10^5 evaluated at thirteen time points. (C-E) Results for different empirical moment gradient estimates: smooth (red), splines (blue), FDS (green) for different amount of time points: 28 (C), 13 (D), 7 (E). (TIF)

S3 Fig. Regularization paths in terms of true/false positive tradeoff over different data availability situations. Results for *reactionet lasso* application to enzymatic system (no measurement noise). (A-B) Empirical moment gradients estimated with “smooth” procedure: (A) 10^5 single cell trajectories evaluated at different amount of time points (tp) as indicated in the legend. (B) Different number of single cell trajectories: 10^3 , 10^4 , 10^5 evaluated at thirteen time points. (C-E) Results for different empirical moment gradient estimates: smooth (red), splines (blue), FDS (green) for different amount of time points: 28 (C), 13 (D), 7 (E). (TIF)

S4 Fig. Overlay of five regularization paths in terms of true/false positive tradeoff over different data availability situations. Results for *reactionet lasso* application to apoptotic receptor subunit ($p = 0.05$) with 10^5 trajectories. Results for different empirical moment gradient estimates: splines (red), FDS (blue) for different amount of time points: 28 (A), 13 (B), 7 (C). (TIF)

S5 Fig. Structure learning performance of the *reactionet lasso*. 10^5 single cell trajectories evaluated at 13 time points for apoptotic receptor subunit ($p = 0.05$). Empirical moment gradients estimated with cubic splines. Solution selected with Bayesian Information Criteria (BIC). (TIF)

S6 Fig. Analysis of standard deviation of moment and stoichiometric moment function estimates for high order moments for different sample sizes. Results for application to apoptotic receptor subunit ($p = 0.05$). (A) Absolute values of standard deviation of moment estimate estimated from bootstrap for the apoptotic receptor subunit with no noise with 10^5 (red), 10^4 (blue), 10^3 (green) trajectories, 13 time points. (B) Relative change of standard deviation of the moment estimates with decreasing number of trajectories compared to 10^5 . (C) Corresponding absolute and relative change of standard deviation of design matrix estimate (with stoichiometric moment functions as entries) with decreasing number of samples compared to 10^5 . (TIF)

S7 Fig. Overlay of five regularization paths in terms of true/false positive tradeoff over different data availability situations. Results for *reactionet lasso* application to apoptotic receptor subunit for uniform selection of time points. Results for different empirical moment gradient estimates: splines (red), FDS (blue) for different amount of time points and different levels of noise: 28 (A, D), 13 (B, E), 7 (C, F). (TIF)

S8 Fig. Original reaction network of TRAIL induced apoptosis. Different modules colored in different colors. Reactions connecting the models depicted in gray. (TIF)

S9 Fig. Comparison of the *reactionet lasso* with various baseline procedures. RL = *reactionet lasso*; STlsq = sequential thresholded regression, TF = Topological filtering. All methods applied to Moment Equations of 1st and 2nd order correspondingly. Results for: (A) the

apoptotic receptor subunit with noise ($p = 0.05$) with 10^5 trajectories, 13 time points; (B) TRAIL-induced apoptosis with noise ($p = 0.05$) with 10^5 trajectories, 33 time points. TF2 was interrupted after 2h hours and didn't produce any solution in the range of cardinality represented on the plot.

(TIF)

S10 Fig. Results for application of the matrix concatenation strategy for *reactionet lasso* for the case of multiple replicates. 5 replicates of the apoptotic receptor subunit ($p = 0.05$) were generated with 10^5 single cell trajectories each evaluated at 13 time points. Red dots correspond to different replicates. Size of the dot proportional to the frequency of the solution between the replicates. Blue line corresponds to the strategy of concatenating design and response matrices.

(TIF)

S11 Fig. Recovery of the dynamics of mean trajectories by the *reactionet lasso*. Red: observed data for 10^5 single cell trajectories evaluated at 13 time points for apoptotic receptor subunit without measurement noise. Solution selected with AIC for two distinct scenarios: *ab initio* learning (blue), *a priori* specified reaction identified false negative in *ab initio* learning setting (green).

(TIF)

S1 Text. Moment expansion.

(PDF)

S2 Text. Information criteria.

(PDF)

S3 Text. Inference of binomial noise correction for empirical moments.

(PDF)

S4 Text. Biological replicates.

(PDF)

S1 Dataset. Time points selection for TRAIL induced apoptosis signaling cascade.

(PDF)

Acknowledgments

We acknowledge Justin Feigelman, Will Macnair, Eirini Arvaniti and Dimitris Christodoulou for helpful discussions and feedback on the manuscript.

Author Contributions

Conceptualization: AK SG MC.

Formal analysis: AK MC.

Funding acquisition: MC.

Investigation: AK.

Methodology: AK MC.

Project administration: MC.

Software: AK.

Supervision: MC.

Validation: AK.

Visualization: AK MC.

Writing – original draft: AK MC.

Writing – review & editing: AK SG MC.

References

1. Raue A, Schilling M, Bachmann J, Matteson A, Schelker M, Schelke M, et al. Lessons learned from quantitative dynamical modeling in systems biology. *PLoS One*. 2013 30 Sep; 8(9):e74335. doi: [10.1371/journal.pone.0074335](https://doi.org/10.1371/journal.pone.0074335) PMID: [24098642](https://pubmed.ncbi.nlm.nih.gov/24098642/)
2. Zechner C, Ruess J, Krenn P, Pelet S, Peter M, Lygeros J, et al. Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences*. 2012 22 May; 109(21):8340–8345. doi: [10.1073/pnas.1200161109](https://doi.org/10.1073/pnas.1200161109) PMID: [22566653](https://pubmed.ncbi.nlm.nih.gov/22566653/)
3. de Vargas Roditi L, Claassen M. Computational and experimental single cell biology techniques for the definition of cell type heterogeneity, interplay and intracellular dynamics. *Curr Opin Biotechnol*. 2014 9 Nov; 34C:9–15. doi: [10.1016/j.copbio.2014.10.010](https://doi.org/10.1016/j.copbio.2014.10.010) PMID: [25461506](https://pubmed.ncbi.nlm.nih.gov/25461506/)
4. Moles CG, Mendes P, Banga JR. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res*. 2003 Nov; 13(11):2467–2474. doi: [10.1101/gr.1262503](https://doi.org/10.1101/gr.1262503) PMID: [14559783](https://pubmed.ncbi.nlm.nih.gov/14559783/)
5. Kuepfer L, Peter M, Sauer U, Stelling J. Ensemble modeling for analysis of cell signaling dynamics. *Nat Biotechnol*. 2007; 25(9):1001–1006. doi: [10.1038/nbt1330](https://doi.org/10.1038/nbt1330) PMID: [17846631](https://pubmed.ncbi.nlm.nih.gov/17846631/)
6. Oates CJ, Dondelinger F, Bayani N, Korkola J, Gray JW, Mukherjee S. Causal network inference using biochemical kinetics. *Bioinformatics*. 2014 1 Sep; 30(17):i468–74. doi: [10.1093/bioinformatics/btu452](https://doi.org/10.1093/bioinformatics/btu452) PMID: [25161235](https://pubmed.ncbi.nlm.nih.gov/25161235/)
7. Kirk P, Thorne T, Stumpf M. Model selection in systems and synthetic biology. *Curr Opin Biotechnol*. 2013;. doi: [10.1016/j.copbio.2013.03.012](https://doi.org/10.1016/j.copbio.2013.03.012) PMID: [23578462](https://pubmed.ncbi.nlm.nih.gov/23578462/)
8. Sunnåker M, Zamora-Sillero E, Dechant R, Ludwig C, Busetto AG, Wagner A, et al. Automatic Generation of Predictive Dynamic Models Reveals Nuclear Phosphorylation as the Key Msn2 Control Mechanism. *Sci Signal*. 2013; 6(277):ra41. doi: [10.1126/scisignal.2003621](https://doi.org/10.1126/scisignal.2003621) PMID: [23716718](https://pubmed.ncbi.nlm.nih.gov/23716718/)
9. Brunton SL, Proctor JL, Kutz JN. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc Natl Acad Sci U S A*. 2016 12 Apr; 113(15):3932–3937. doi: [10.1073/pnas.1517384113](https://doi.org/10.1073/pnas.1517384113) PMID: [27035946](https://pubmed.ncbi.nlm.nih.gov/27035946/)
10. Anderson DF, Kurtz TG. Continuous time Markov chain models for chemical reaction networks. *Design and Analysis of Biomolecular Circuits*. 2011;. doi: [10.1007/978-1-4419-6766-4_1](https://doi.org/10.1007/978-1-4419-6766-4_1)
11. Oates CJ, Mukherjee S. Network Inference and Biological Dynamics. *Ann Appl Stat*. 2012 Sep; 6(3):1209–1235. doi: [10.1214/11-AOAS532](https://doi.org/10.1214/11-AOAS532) PMID: [23284600](https://pubmed.ncbi.nlm.nih.gov/23284600/)
12. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc Series B Stat Methodol*. 1996 1 Jan; 58(1):267–288.
13. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc Series B Stat Methodol*. 2011 20 Jun; 73(3):273–282. doi: [10.1111/j.1467-9868.2011.00771.x](https://doi.org/10.1111/j.1467-9868.2011.00771.x)
14. Pan W, Yuan Y, Stan GB. Reconstruction of Arbitrary Biochemical Reaction Networks: A Compressive Sensing Approach. 2012 8 May;.
15. Wagener J, Dette H. Bridge estimators and the adaptive Lasso under heteroscedasticity. *Mathematical Methods of Statistics*. 2012; 21(2):109–126. doi: [10.3103/S1066530712020032](https://doi.org/10.3103/S1066530712020032)
16. Zou H. The Adaptive Lasso and Its Oracle Properties. *J Am Stat Assoc*. 2006; 101(476):1418–1429. doi: [10.1198/016214506000000735](https://doi.org/10.1198/016214506000000735)
17. Meinshausen N. Relaxed Lasso. *Comput Stat Data Anal*. 2007 Sep; 52(1):374–393. doi: [10.1016/j.csda.2006.12.019](https://doi.org/10.1016/j.csda.2006.12.019)
18. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*. 2011;p. 1–123. doi: [10.1561/22000000016](https://doi.org/10.1561/22000000016)
19. Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Series B Stat Methodol*. 2010 1 Sep; 72(4):417–473. doi: [10.1111/j.1467-9868.2010.00740.x](https://doi.org/10.1111/j.1467-9868.2010.00740.x)

20. Gillespie DT. Stochastic Simulation of Chemical Kinetics. *Annu Rev Phys Chem.* 2007; 58(1):35–55. doi: [10.1146/annurev.physchem.58.032806.104637](https://doi.org/10.1146/annurev.physchem.58.032806.104637) PMID: [17037977](https://pubmed.ncbi.nlm.nih.gov/17037977/)
21. Albeck JG, Burke JM, Spencer SL, Lauffenburger DA, Sorger PK. Modeling a snap-action, variable-delay switch controlling extrinsic cell death. *PLoS Biol.* 2008 2 Dec; 6(12):2831–2852. doi: [10.1371/journal.pbio.0060299](https://doi.org/10.1371/journal.pbio.0060299) PMID: [19053173](https://pubmed.ncbi.nlm.nih.gov/19053173/)
22. Chen Y, Shen K, Shan SO, Kou SC. Analyzing Single-Molecule Protein Transportation Experiments via Hierarchical Hidden Markov Models. *J Am Stat Assoc;* 0(ja):1–49. doi: [10.1080/01621459.2016.1140050](https://doi.org/10.1080/01621459.2016.1140050)
23. Meyer P, Cokelaer T, Chandran D, Kim KH, Loh PR, Tucker G, et al. Network topology and parameter estimation: from experimental design methods to gene regulatory network kinetics using a community based approach. *BMC Syst Biol.* 2014 7 Feb; 8:13. doi: [10.1186/1752-0509-8-13](https://doi.org/10.1186/1752-0509-8-13) PMID: [24507381](https://pubmed.ncbi.nlm.nih.gov/24507381/)
24. Villaverde AF, Henriques D, Smallbone K, Bongard S, Schmid J, Cicin-Sain D, et al. BioPreDyn-bench: a suite of benchmark problems for dynamic modelling in systems biology. *BMC Syst Biol.* 2015 20 Feb; 9:8. doi: [10.1186/s12918-015-0144-4](https://doi.org/10.1186/s12918-015-0144-4) PMID: [25880925](https://pubmed.ncbi.nlm.nih.gov/25880925/)
25. Rice JJ, Tu Y, Stolovitzky G. Reconstructing biological networks using conditional correlation analysis. *Bioinformatics.* 2005 Mar; 21(6):765–773. doi: [10.1093/bioinformatics/bti064](https://doi.org/10.1093/bioinformatics/bti064) PMID: [15486043](https://pubmed.ncbi.nlm.nih.gov/15486043/)
26. Rogers S, Girolami M. A Bayesian regression approach to the inference of regulatory networks from gene expression data. *Bioinformatics.* 2005 15 Jul; 21(14):3131–3137. doi: [10.1093/bioinformatics/bti487](https://doi.org/10.1093/bioinformatics/bti487) PMID: [15879452](https://pubmed.ncbi.nlm.nih.gov/15879452/)
27. Friedman N. Inferring Cellular Networks Using Probabilistic Graphical Models. *Science.* 2004; 303(5659):799–805. doi: [10.1126/science.1094068](https://doi.org/10.1126/science.1094068) PMID: [14764868](https://pubmed.ncbi.nlm.nih.gov/14764868/)
28. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet.* 2003; 34(2):166–176. doi: [10.1038/ng1165](https://doi.org/10.1038/ng1165) PMID: [12740579](https://pubmed.ncbi.nlm.nih.gov/12740579/)
29. Lee SI, Dudley AM, Drubin D, Silver PA, Krogan NJ, Pe'er D, et al. Learning a Prior on Regulatory Potential from eQTL Data. *PLoS Genet.* 2009; 5(1):e1000358. doi: [10.1371/journal.pgen.1000358](https://doi.org/10.1371/journal.pgen.1000358) PMID: [19180192](https://pubmed.ncbi.nlm.nih.gov/19180192/)
30. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 2011 Jan; 12(1):56–68. doi: [10.1038/nrg2918](https://doi.org/10.1038/nrg2918) PMID: [21164525](https://pubmed.ncbi.nlm.nih.gov/21164525/)
31. Sunnåker M, Zamora-Sillero E, López García de Lomana A, Rudroff F, Sauer U, Stelling J, et al. Topological augmentation to infer hidden processes in biological systems. *Bioinformatics.* 2014 15 Jan; 30(2):221–227. doi: [10.1093/bioinformatics/btt638](https://doi.org/10.1093/bioinformatics/btt638) PMID: [24297519](https://pubmed.ncbi.nlm.nih.gov/24297519/)
32. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell.* 2015 21 May; 161(5):1187–1201. doi: [10.1016/j.cell.2015.04.044](https://doi.org/10.1016/j.cell.2015.04.044) PMID: [26000487](https://pubmed.ncbi.nlm.nih.gov/26000487/)
33. Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell.* 2015 21 May; 161(5):1202–1214. doi: [10.1016/j.cell.2015.05.002](https://doi.org/10.1016/j.cell.2015.05.002) PMID: [26000488](https://pubmed.ncbi.nlm.nih.gov/26000488/)
34. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter Estimation for Differential Equations: A Generalized Smoothing Approach. *J R Stat Soc Series B Stat Methodol.* 2007 1 Jan; 69(5):741–796.
35. Dondelinger M, Filippone F, Rogers S, Husmeier D. ODE parameter inference using adaptive gradient matching with Gaussian processes. *AISTATS.* 2013; 31. doi: [10.1186/s12938-016-0186-x](https://doi.org/10.1186/s12938-016-0186-x)
36. Milner P, Gillespie CS, Wilkinson DJ. Moment closure approximations for stochastic kinetic models with rational rate laws. *Math Biosci.* 2011 Jun; 231(2):99–104. doi: [10.1016/j.mbs.2011.02.006](https://doi.org/10.1016/j.mbs.2011.02.006) PMID: [21338614](https://pubmed.ncbi.nlm.nih.gov/21338614/)
37. Nakakuki T, Birtwistle MR, Saeki Y, Yumoto N, Ide K, Nagashima T, et al. Ligand-specific c-Fos expression emerges from the spatiotemporal control of ErbB network dynamics. *Cell.* 2010 28 May; 141(5):884–896. doi: [10.1016/j.cell.2010.03.054](https://doi.org/10.1016/j.cell.2010.03.054) PMID: [20493519](https://pubmed.ncbi.nlm.nih.gov/20493519/)
38. Gillespie CS. Moment-closure approximations for mass-action models. *Systems Biology, IET.* 2009; 3(1):52–58. doi: [10.1049/iet-syb:20070031](https://doi.org/10.1049/iet-syb:20070031) PMID: [19154084](https://pubmed.ncbi.nlm.nih.gov/19154084/)
39. Varah JM. A Spline Least Squares Method for Numerical Parameter Estimation in Differential Equations. *SIAM Journal on Scientific and Statistical Computing.* 1982; 3(1):28–46. doi: [10.1137/0903003](https://doi.org/10.1137/0903003)
40. Poyton AA, Varziri MS, McAuley KB, McLellan PJ, Ramsay JO. Parameter estimation in continuous-time dynamic models using principal differential analysis. *Comput Chem Eng.* 2006 15 Feb; 30(4):698–708. doi: [10.1016/j.compchemeng.2005.11.008](https://doi.org/10.1016/j.compchemeng.2005.11.008)
41. Calderhead B, Girolami M, Lawrence ND. Accelerating Bayesian Inference over Nonlinear Differential Equations with Gaussian Processes. In: Koller D, Schuurmans D, Bengio Y, Bottou L, editors. *Advances in Neural Information Processing Systems 21.* Curran Associates, Inc.; 2009. p. 217–224.

42. Ulianov A, Müntener O, Schaltegger U. The ICPMS signal as a Poisson process: a review of basic concepts. *Journal of Analytical Atomic Spectrometry*. 2015; 30(6):1297–1321. doi: [10.1039/C4JA00319E](https://doi.org/10.1039/C4JA00319E)
43. Grün D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. *Nat Methods*. 2014; 11(6):637–640. doi: [10.1038/nmeth.2930](https://doi.org/10.1038/nmeth.2930) PMID: [24747814](https://pubmed.ncbi.nlm.nih.gov/24747814/)