Kinetic modeling and parameter estimation of a prebiotic peptide reaction network

Authors

*Hayley Boigenzahn[1,2], Leonardo D. González[1], Jaron C. Thompson[1], Victor M. Zavala[1], John Yin[1,2]**

*Corresponding Author: john.yin@wisc.edu

Affiliations

[1]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706, USA

[2]Wisconsin Institute for Discovery, University of Wisconsin-Madison, 330 N. Orchard Street, Madison, WI 53715, USA

Abstract

Although our understanding of how life emerged on Earth from simple organic precursors is speculative, early precursors likely included amino acids. The polymerization of amino acids into peptides and interactions between peptides are of interest because peptides and proteins participate in complex interaction networks in extant biology. However, peptide reaction networks can be challenging to study because of the potential for multiple species and systems-level interactions between species. We developed and employed a computational network model to describe reactions between amino acids to form di-, tri-, and tetra-peptides. Our experiments were initiated with two of the simplest amino acids, glycine and alanine, mediated by trimetaphosphate-activation and drying to promote peptide bond formation. The parameter estimates for bond formation and hydrolysis reactions in the system were found to be poorly constrained due to a network property known as sloppiness. In a sloppy model, the behavior mostly depends on only a subset of parameter combinations, but there is no straightforward way to determine which parameters should be included or excluded. Despite our inability to determine the exact values of specific kinetic parameters, we could make reasonably accurate predictions of model behavior. In short, our modeling has highlighted challenges and opportunities toward understanding the behaviors of complex prebiotic chemical experiments.

Keywords

Peptides, chemical reaction networks, kinetics, parameter estimation, prebiotic chemistry

Introduction

The emergence of life on the early Earth is believed to have been preceded by the accumulation of an increasingly diverse and complex set of organic molecules (Orgel 2010). The reaction networks developed by these molecules laid the groundwork of functions critical for life, like energy and information processing. Understanding how the systems-level molecular interactions required for life-like behavior could emerge from simple precursors remains one of the key questions of prebiotic chemistry, but since this question is primarily about collective behaviors, complexity presents an ongoing challenge (Schwartz 2007; Johnson & Hung 2019). While studying a single type of molecule or a single reaction to establish its properties can be useful, it limits what conclusions can be drawn about potential broader community behavior. Experiments involving a greater variety of molecules and reactions can probe more interesting interactions, but have a large search space of variables, and the complexity of the systems make them inherently more difficult to analyze.

Models are useful for understanding complex systems because they can reveal the systematic dependence of various properties on each other and allow us to describe and make predictions about the system behavior. Computational models have been used to explore hypothetical prebiotic chemical networks for many years and have produced many interesting insights (Covney et al. 2012). However, our current interest is in models that are based on experimental data. Prior experimental works mainly used basic kinetic and thermodynamic governing equations to describe individual reactions or small networks involving fewer than five reactions. For example, Arrhenius expressions have been used to determine the free energies of activation for reactions in a small network (Sakata et al. 2010; Yu et al. 2016; Lee et al. 1996). More abstractly, parameters have been fit to empirical rate equations to describe specific elements of system behavior or distinguish between candidate models (von Kiedrowski 1986; Rout et al. 2022). These methods work well for small systems, but may not apply to larger systems with multiple reactions occurring simultaneously and potentially more intricate network interactions. Serov et al. (2020) approximated the parameters for multiple reactions simultaneously in a peptide reaction network, but the parameter fitting was performed manually, and the network was small. Manual approaches are less rigorous than using a computational strategy and can be difficult to implement for even moderately sized networks. On the other hand, results from more complex experiments have been analyzed using statistical methods, but these do not capture the system dynamics (Surman et al. 2019; Jain et al. 2022). There is a need for approaches to study the dynamical behavior of more complex experimental networks (Ruiz-Mirazo et al. 2014).

Complex network models are broadly applicable and have already been developed extensively for other fields (Newman 2003). One notable example is in systems biology, which has significant parallels to the origins of life. Both involve large interaction networks with potentially limited available data and may include community interactions that are critical to understanding system behavior. Bioinformatics models can be used to analyze experimental data and help understand the molecular interaction networks within living cells (Gauthier et al. 2019).

Similar approaches could be useful for furthering experimental chemical origins of life research, but aside from a few reviews and computational investigations, they have generally been overlooked (Johnson & Hung 2019; Ludlow & Otto 2008; Goldman et al. 2013).

Our goal in this study was to investigate how dynamical models, described by ordinary differential equations (ODEs), might be useful for studying origins of life chemistry. These models are theoretically generalizable, but as with all modeling approaches, there are limitations that make them more difficult to apply in some situations. Presenting the benefits and limitations of a model approach in a way that is accessible to experimentalists, which we aim to do, is an important step for linking theory and experiment. Differential equation models are not always suitable for large systems, since constructing them can become difficult, but in well-defined systems they can be used to study detailed mechanistic behavior (Maria 2004). Computational methods can be used to estimate all the parameters efficiently and simultaneously in a moderately complex dynamical network, but validating the physical meaning of the results can be more challenging since these problems may not have a unique and stable solution (Transtrum et al. 2015). However, parameter fitting has still been used to describe nonlinear networks in a variety of fields, including in systems biology for biochemical pathways (Raue et al. 2013; Rodriguez-Fernandez et al. 2006).

We focus specifically on fitting parameters to a set of nonlinear ODEs describing the kinetics of short peptide formation. Peptides are interesting candidates for emergent behavior because they can engage in a variety of intermolecular interactions and their development was likely an important step during the origin of life (Frenkel-Pinter et al. 2020). We studied a simplified network describing peptide formation in a system starting with only two amino acid species, glycine and alanine. By limiting ourselves to two amino acids, we were able to obtain quantitative data on the concentrations of most peptide species as they formed through a possible prebiotic reaction mechanism involving an inorganic phosphate activating agent, trimetaphosphate (TP) (Sibilska et al. 2018).

We found that our model exhibited “sloppiness”, a term originally used by the Sethna lab to describe models based on a set of highly imprecise parameters that still return reasonably accurate predictions (Gutenkunst et al. 2007a). Such models are significantly more sensitive to changes in certain parameter values while remaining largely unaffected by changes in others (Waterfall et al. 2006). We suspect sloppiness may be a common feature in networks relevant to the chemical origin of life. It is known to be extremely common in systems biology, and many of the features that contribute to it, like reversibility of reactions and limited experimental observations, are also common features of prebiotic chemistry networks (White et al. 2016).

Sloppiness occurs when parts of the parameter fitting problem are poorly constrained, resulting in highly imprecise parameter estimates. Our computational study reveals that the peptide network model is sloppy. Due to their high uncertainty, parameters fitted to a sloppy model cannot be treated as true kinetic reaction rates, limiting the hypotheses a sloppy model can

be used to evaluate (Gutenkunst et al. 2007a). However, the collective behavior predicted by fitting a sloppy model can be accurate even when fit to relatively sparse experimental data. This makes them useful for tasks such as exploring theoretical long-term behavior and model falsification (Brown et al. 2004; Gutenkunst et al. 2007b; Hettling & van Beek 2011). For these reasons, we concluded that this system was worth investigating and disseminating despite the high variability observed in the parameter estimates.

We attempted to reduce sloppiness using model reduction and statistical design of experiments, but without improvements. As such, it is important to recognize the inherent limitations of the model structure and of the experimental setup. We conclude that fitting accurate kinetic parameters using the approach we present might be difficult. However, ODE models can still be useful tools for characterizing the behavior and stability of prebiotic chemical reaction networks.

Methods

We studied the formation of peptides from amino acids using trimetaphosphate (TP) as an activating agent. For simplicity, our experiments only included two amino acids: glycine and alanine. To maximize peptide bond formation within 24 hours, samples at alkaline pH were allowed to dry completely (Sibilska et al. 2018). Various combinations of initial concentrations of glycine and alanine were used to increase the amount of relevant data for parameter fitting, and cover a larger range of potential conditions in the network, since concentrations of each species should not affect the values of the kinetic constants. The concentrations of each peptide product were determined using HPLC (see Experimental Methods for details). Each experimental data point is the average of three experimental replicates.

Parameters were fit to an ODE model describing peptide formation and decomposition in a mass-action style network, depicted in Fig. 1. The complete time-dependent ODEs for the model are provided in Supplementary Information 1. To keep the network a manageable size, we omitted many mechanistic details of peptide formation and only includes canonical peptides, not intermediates or possible side products. For example, no phosphate salts or intermediate products of TP activation were quantifiable in our analysis, so TP was not explicitly included anywhere in the network. To minimize any effect the concentration of TP might have on the kinetics studied, we used a constant ratio of TP to amino acids across all experiments. Isomers such as GGA, GAG, and AGG were grouped together to further reduce the number of parameters and avoid the need to resolve isomers, which tend to co-elute during HPLC analysis. A complete list of fitted parameters, organized by figure, are available on Github at https://github.com/haboigenzahn/OoL-KineticParameterEstimation.

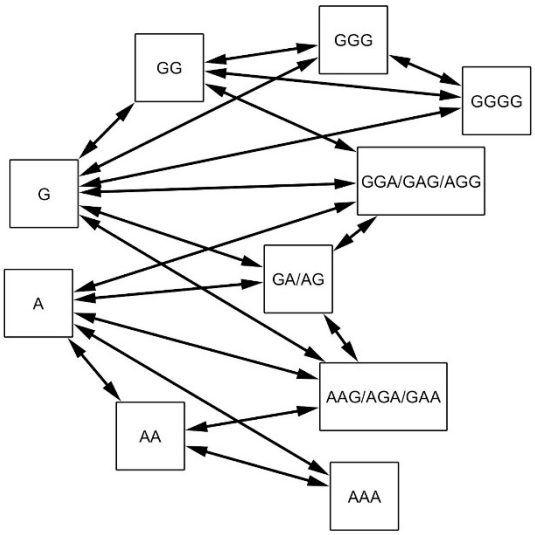


Figure 1: Peptide network. Double-headed arrows represent a reversible reaction connecting two species. Note that many edges share the same reaction parameter, such as the G → GGG and GG → GGG edges representing the reaction G+GG → GGG.

We expected that the network would provide a good baseline for understanding which reactions were occurring at higher rates. To improve the precision of the parameter estimates, we applied model reduction and statistical experimental design. Details about these approaches can be found in the Computational Methods section. Here we will describe the results of these tests and assess the feasibility of obtaining a predictive model and accurate parameter estimates from experimental data.

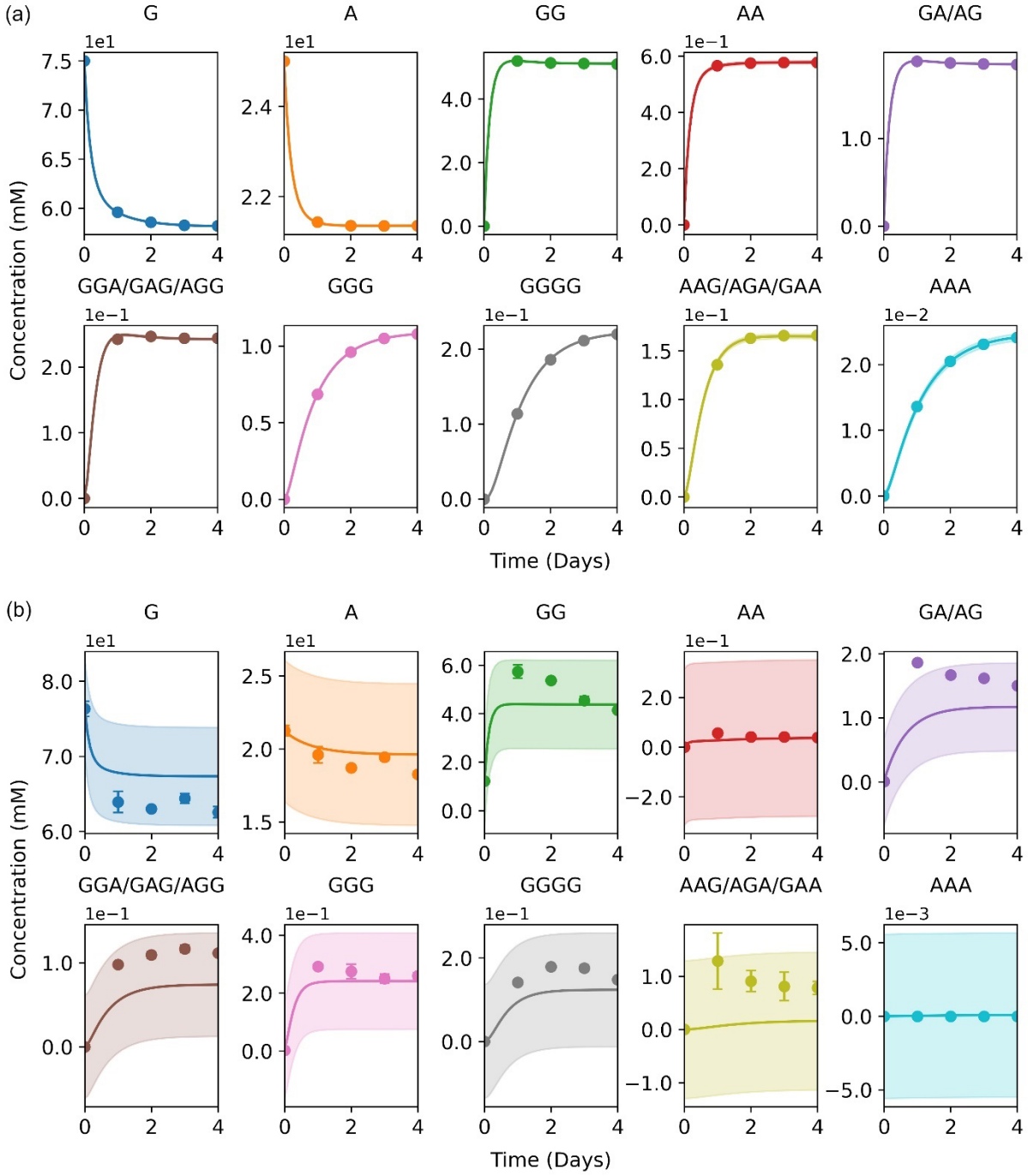
Results & Discussion

Parameter Estimation

Parameter fitting is performed by tuning the model parameters to minimize a *cost function* ($\mathcal{L}$) that calculates the difference between the model predictions and experimental data; $\mathcal{L}$is also called a loss function or a residual. We minimized $\mathcal{L}$ using the L-BFGS-B algorithm from Scipy's minimize function (Virtanen et al. 2020). We were also able to approximate the parameter uncertainties, which represent how well the parameters are constrained by experimental data using an asymptotic Gaussian approximation (Vanlier et al. 2013). Parameters determined using sparse or noisy experimental data are less precise than parameters fit with abundant, high precision data, but the structure of the model itself can also significantly contribute to the parameter uncertainty. Validating that the model can theoretically be solved can save time and experimental effort.

We first estimated the parameters for simulated data in the absence of noise, and we were able to accurately recover the parameters used to generate the data (Fig. 2a). When we applied the model to experimental data, it was able to capture general trends. However, the parameter uncertainties were undesirably high (Fig. 2b). For some species, the 95% confidence envelope

for the model prediction was larger than the peptide concentrations themselves. Since the optimization can find a local minimum, we repeated the parameter estimation for several different initial guesses. Although the number of initial guesses was limited by the fact that the parameter estimation method can take a full day to finish when all of the experimental data is included, we observed that none of the different initial guesses significantly improved the precision of the parameter estimates and that there did not appear to be any positive correlation between the MSE and the number of highly uncertain parameters (Supplementary Information 2). Trying many initial guesses to find the lowest possible value for the cost function may slightly improve the model predictions, but it does not seem to cause an improvement in the precision of the parameter estimates. Despite the extremely high parameter uncertainties, the accuracy of the model predictions initially seemed promising, so we began to explore the parameter fitting process in more detail to determine how to decrease the parameter uncertainty, starting with the identifiability of the network.



184 **Figure 2: Comparison of fitting data and model predictions.** Results are shown for (a)
185 simulated data and (b) experimental data, using initial conditions of 75 mM glycine and 25 mM
186 alanine. Both the simulated and experimental data sets included 65 data points and the simulated
187 data had no artificially added noise.

188

189 *Identifiability & Sloppiness*

Identifiability analysis determines the possibility of a unique and precise estimate of the unknown parameters in a network (Cobelli & DiStefano, 1980; Wieland et al. 2021). If a unique solution cannot be obtained, then the model is said to be *structurally unidentifiable.* A model is *practically unidentifiable* if its parameters cannot be estimated at an acceptable level of precision. The exact definition of what is considered an acceptable level of precision varies from case to case. Practical unidentifiability indicates that regions of the objective function are relatively flat, making it difficult to find a minimum and it typically results from overfitting (White et al. 2016). Finally, some models exhibit a property known as *sloppiness*, which occurs when its behavior is highly sensitive to changes in certain combinations of parameters and almost completely insensitive to changes in others (Gutenkunst et al. 2007a). Generally, sloppiness is a consequence of the model structure and its input range (White et al. 2016). Although sloppiness and practical unidentifiability are not synonymous, in practice they often coincide (Chis et al. 2014).

Sloppiness can be recognized by examining the spectrum eigenvalues of the Hessian matrix, sometimes called the sensitivity eigenvalues (see Computational Methods section for further detail) (Gutenkunst et al. 2007a). The sensitivity eigenvalues are an indirect estimate of the sensitivity of the cost function to changes in the parameter values and represents the confidence in the estimate of the parameter combination in the direction of the corresponding eigenvector. Small eigenvalues represent high uncertainties and large confidence intervals. Sloppy models have sensitivity eigenvalues that are roughly evenly spaced across three or more orders of magnitude. When the eigenvalue spectrum is this large, the smallest sensitivity eigenvalues tend to correspond to parameter combinations that have minimal effect on the model behavior – these combinations are ‘sloppy’ eigenvectors. The eigenvectors of the largest eigenvalues are referred to as ‘stiff’ and control most of the model behavior. In some models, there is a clear division between the large and small eigenvalues, usually corresponding to a clear separation in length or time scales that renders some of the physical details of the system irrelevant – for example, the kinetic models of many chemical reactions can be simplified when there is a known rate-limiting step (White et al. 2016). In sloppy models, no clear division exists, and the small eigenvalues are rarely united by a single physical phenomenon.

Since rigorously checking for structural identifiability in nonlinear models can be challenging, we tested the identifiability of our model by determining if it could recover the parameters used to generate a set of noiseless, simulated data. We found that all parameters could be recovered with acceptably high accuracy, suggesting that the model was identifiable. Here, we define acceptable accuracy to be when a parameter’s standard deviation is at least one order of magnitude smaller than value of the associated parameter. However, when we examined the effect of noise on model performance, we observed that the parameter standard deviations rise rapidly when even a small amount of noise is introduced (Fig. 3a). The error of the model predictions, on the other hand, rose relatively slowly as noise increased. This suggests that

despite the high parameter uncertainties, the general behavior predicted by the model can be accurate even when it is fit using noisy data (Fig. 3b).

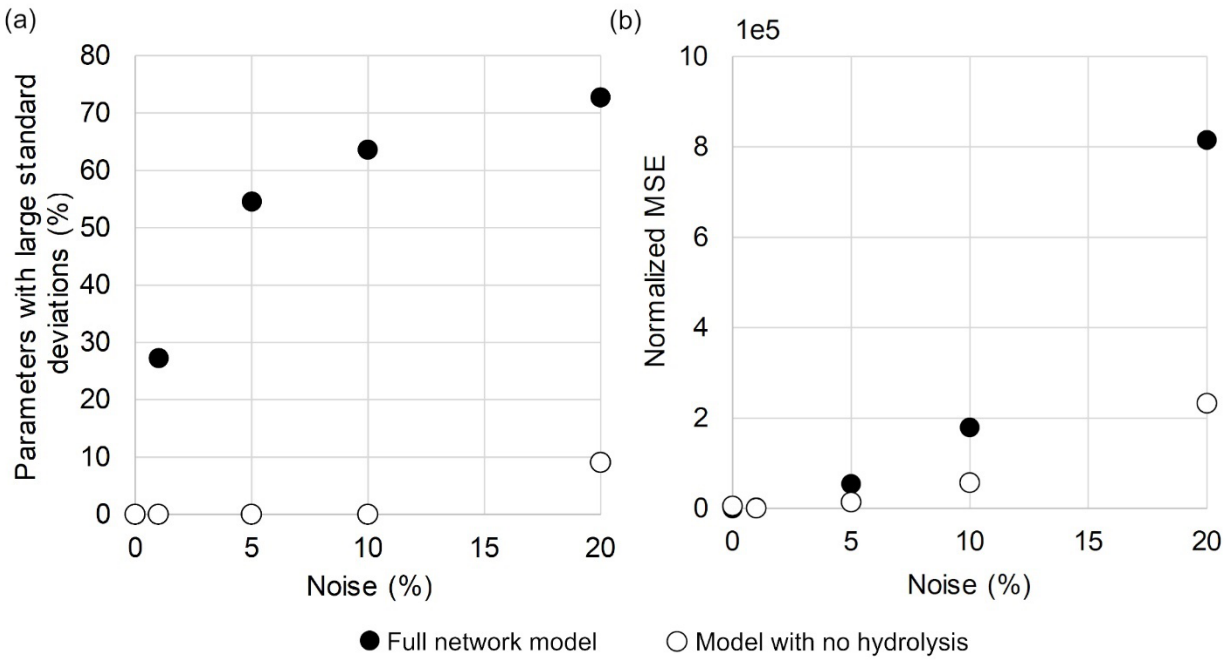


Figure 3: Comparison of parameter accuracy and mean squared error (MSE) for two different network structures at various noise levels. For the full reaction network, as the noise in the input data is increased, (a) the number of parameters with standard deviations within one order of magnitude of the parameter value rises rapidly compared to (b) the error of the model predictions. When the hydrolysis reactions are removed from the full network, the parameter estimates remain relatively precise as noise is introduced. The MSE of the model predictions are normalized to the MSE of the full network with no artificial noise (2.85e-11). All data sets used simulated experiments created from 25 different initial conditions and 125 data points. The added noise was normally distributed with a constant signal-to-noise ratio, and all negative values were set to zero to prevent negative concentrations.

Given that this behavior is typical in sloppy models, we checked the sensitivity eigenvalues for both our simulated data and experimental data (Fig. 4a, b). We found that the peptide reaction network is unambiguously sloppy, because the sensitivity eigenvalues of the simulated and experimental data span nearly nine and seven orders of magnitude respectively. To compare the behavior of the peptide reaction network with a similar model that was not sloppy, we modified the network to exclude all hydrolysis reactions (Supplementary Information 3a). Removing reversible pathways from the network eliminates many combinations of parameters that can compensate for one another, which significantly reduced sloppiness (Fig. 4c). To demonstrate that it was the modifications to the structure of the model, rather than its smaller size, that were responsible for the reduction in sloppiness, we also compared it to an even smaller network describing reversible homopolymer reactions (Supplementary Information 3b); this

model was determined to have a much larger eigenvalue span (Fig. 4d). To investigate whether the grouping of some species in the peptide network was responsible for the sloppiness of the model, we also checked the sensitivity eigenvalues for a network with the trimer species separated using simulated data (Supplementary Information 3c), and found it made the eigenvalue spread larger (Fig. 4e).

The parameter standard deviations were far more sensitive to noise in the full, sloppy network than in the network with no hydrolysis reactions (Fig. 3a). Despite the difference in the confidence of the parameter fits, the prediction accuracy was not significantly different between the two models until significant noise was added to the data (Fig. 3b). This demonstrates a previously mentioned key consequence of sloppy models – although they can make reasonably accurate predictions of system behavior, they should not be used to calculate the values of individual parameters, since the precision required for accurate parameter estimations cannot be experimentally realized.

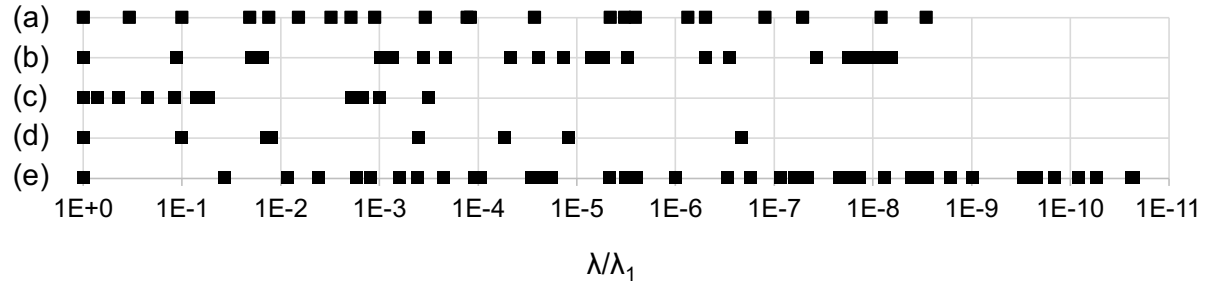


Figure 4: Sensitivity eigenvalues for different models: (a) simulated data for the full network (22 parameters, 35 data points), (b) experimental data generated from a mixture of glycine and alanine (22 parameters, 65 data points), (c) simulated data for a variation of the main network that excludes all hydrolysis reactions (11 parameters, 35 data points), (d) simulated data for network including only one amino acid forming peptides up to tetramer length with hydrolysis reactions included (8 parameters, 35 data points) and (e) simulated data for a network with separated trimers (40 parameters, 80 data points). Each system is normalized to its largest eigenvalue (λ_1). All simulated data has no additional noise included.

Sloppiness is a common property in systems biology models, and some of the characteristics that result in sloppiness are likely shared by prebiotic chemistry systems. Reversible reactions and cyclic behaviors can increase the likelihood of sloppiness because they create situations where a particular combination of parameters (for example, the ratio between forward and reverse rates defining an equilibrium constant) is more important for describing the system behavior than the individual parameters themselves. The parameters may become 'sloppy' because their individual values can essentially vary freely without affecting the overall model behavior, as long changes in other parameters can compensate to produce a similar overall prediction. Reaction networks that are mostly or entirely reversible, like the peptide reaction network, can therefore become significantly more difficult to fit with high precision than models

with comparable sizes, but fewer reversible reactions (Maity et al. 2020). The emergence of cycles and reversible reactions are expected to be important features in the emergence of life-like chemistry (Varfolomeev & Lushchekina 2014; Mamajanov et al. 2014). Therefore, we anticipate that sloppiness may be a common and potentially unavoidable feature of ODE models found in prebiotic chemistry, and its implications should be examined.

Consequences of Collective Fitting

Sloppy models can provide surprisingly accurate predictions despite having low confidence parameter estimates. The collective fit of all the parameters tends to be more accurate and require less data than the individual parameter uncertainties might suggest, since only the stiff parameter combinations must be constrained to achieve accurate predictions. One of the consequences of collective fitting is that the numerical values of parameters estimated for sloppy models cannot be treated as independent kinetic parameters whose quantitative values have physical meaning. Situations where a reaction occurs faster in the presence of one molecule than another are of interest to the chemical origins of life because of their semblance to catalysis. Unfortunately, in sloppy models, the numerical values of the parameters fit in each case are often not comparable. For example, even if the rate constant of one reaction in the peptide network was significantly higher than another, that is not necessarily good evidence that one reaction proceeds faster than the other. The parameters are only meaningful when the entire system is used to describe the specific environment to which they were fit. Fixing individual parameter values to reflect direct measurements or literature values can potentially break the collective fit and significantly increase the error of the prediction, often to the point that it is no longer useful. The lack of physical meaning of the individual parameter values is a significant drawback of sloppy models. However, such models can still be useful for certain tasks. For example, a sloppy model can still be used if the goal is to generate predictions about the behavior of a similar system with slightly different initial conditions, or to predict responses at longer time spans. Moreover, we highlight that sloppiness might simply be a fundamental property of the actual reaction network, that arises from inherent redundancies in the system.

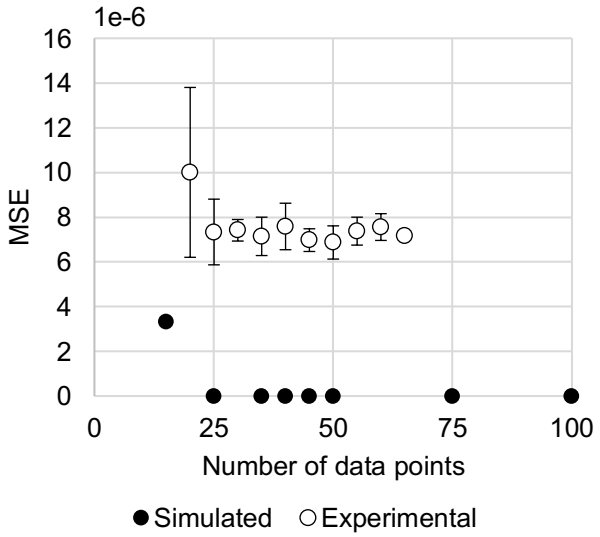


Figure 5: MSE of model predictions depend on quantity of experimental and simulated data. Except for the final points, which include all applicable data, parameters were estimated for three arbitrarily selected data subsets of varying sizes, then the average MSE of those models was determined. Noise was neglected. Error bars show the standard deviation of the three subsets, but are too small to be visible for the simulated data.

To estimate the minimal data required to get relatively accurate predictions, we created at least three different subsets of the data, trained the model individually with each subset, and compared their MSEs (Fig. 5). The simulated data was sampled at time intervals analogous to the experimental results, since those were the points that were physically relevant. When training the model using simulated data, increasing the amount of data used improved the model predictions up to about 40 data points, but with even 25 data points, the error was negligible compared to the experimental results. Similarly, when we repeated the process with experimental data, the average error did not decrease as more data was added beyond 25 data points.

We also investigated the effect of using more frequent measurements, as opposed to using a greater number of simulated experiments with different initial conditions. We compared the results of simulated data with a similar number of total data points, but double the usual sampling frequency to the simulated results in Fig. 5. Increasing the sampling frequency was comparable or slightly worse than including data from additional simulated initial conditions, except possibly when there is little data available overall (Supplementary Information 4). It did not improve the model's sensitivity to noise.

Different subsets of the data with the same number of data points could have fairly different MSEs, suggesting that some combinations of experiments may be better for parameter fitting than others. This subject will be discussed further in the section on the design of experiments (DoE). Overall, these results suggest that as few as 25 to 30 data points are required to fit the system as accurately as the model constraints allow; therefore reasonably accurate

predictive fits can be achieved with a realistically obtainable amount of data. The ability to extrapolate accurate model predictions from short-term experiments has some uses for studying prebiotic chemical reactions, since long time spans are potentially relevant. Models like the one we present here could be used to predict the expected equilibrium outcome of slow reactions based on data from a shorter time span and compare candidate model structures. They may also be a useful way to predict the outcomes of sequential or cyclic processes, provided that the parameters are fit in compatible experimental conditions. Sensitivity analysis can be used to validate the predictions from sloppy models independently from the parameter uncertainties (Gutenkunst et al. 2007a). Model selection, which involves comparing two or more different model structures to determine which one reflects the experimental data most accurately, can also still be performed with sloppy models (Brown & Sethna 2003). However, if finding physically meaningful terms for the parameter values is an important goal, then the aim should be to reduce the sloppiness of the model.

Model Reduction

To address high parameter uncertainty, one may seek to simplify the structure of the model, ideally without compromising the accuracy of the model predictions. This task is referred to as *model reduction* or *network reduction*, and it can be an effective way to improve overparameterized models (Apri et al. 2012; Transtrum et al. 2015). However, model reduction methods are generally based on statistical principles and not physical knowledge, and the results should be interpreted within an experimental context. The user must ensure that parameters that might be statistically problematic but are known to be physically significant are not removed from the model.

Since one of the main features of sloppy models is that they contain parameter combinations that are insensitive to changes, model reduction may initially appear to be a straightforward task for sloppy models. However, the fact that the sensitivity eigenvalues are evenly distributed over multiple orders of magnitude poses a challenge for accurate model reduction, as there is no clear cut-off between the parameter combinations that are important and those that are not. Additionally, in practice some parameters are so poorly constrained that they are randomly distributed throughout the sensitivity eigenvectors, so the components of the sensitivity eigenvectors are not entirely reliable indicators of what parameters are influencing them (Gutenkunst et al. 2007a).

We attempted model reduction with the peptide reaction network to determine if it was over-parameterized and if it might be possible to reduce the reactions considered. For example, we expected that some of the hydrolysis reactions could be ignored. Since we wanted to use a model reduction technique that is accessible and easily interpretable for experimentalists, we used sparse principal component analysis (SPCA). SPCA is an extension of principal component analysis (PCA), a popular dimensionality reduction method for linear models (Zou et al. 2006). Using SPCA, we can identify the inputs that capture most of the information in the data. It has

been used successfully in control theory and gene network analysis, and there are existing implementations of it in MATLAB and Python (Ma & Dai 2011).

When SPCA was applied to the peptide reaction network, the results were highly variable and unable to adequately represent the data. SPCA frequently suggested removing reactions known to be physically significant, such as the formation of dimers from monomers (Supplementary Information 5). Not only does this not make physical sense, but because these are the initial reactions that occur in the system, removing them severely limits the pathways for longer species to form. Other methods of network reduction may be more effective for sloppy models, but are less commonly used and may be more difficult to implement (Transtrum & Qiu 2014; Maiwald et al. 2016). If we choose to pursue additional model reduction efforts, one logical next step may be to inspect the inverse of the covariance matrix to identify which parameters are the most correlated and least constrained by the data (Wasserman, L. 2004). This information may be useful for determining which parameters are best to remove or to combine into a single term.

Design of Experiments

If the model structure cannot be altered, another method for reducing sloppiness is to determine if experimental data can be gathered strategically to explore the variable space more thoroughly (Apgar et al. 2010). However, to reduce parameter uncertainty, the selected experiments must provide new information not already captured in the model. *Design of experiments* (DoE), or experimental design, seeks to identify the experiments that would provide the most useful information for improving prediction accuracies. DoE methods such as factorial design (Fisher 1935), response surface methodology (Box & Wilson 1951), and screening (Shevlin 2017), have been widely adopted across various fields. However, there are several notable caveats in relation to sloppy models (Jagadeesan et al. 2022). First, the precision of parameter fitting for sloppy models is limited by the least accurately determined eigenvectors, so more data measured with the same uncertainty may not help. Second, there is some debate over whether DoE can be used with approximate models without risking the collective fit, as it can inadvertently place too much importance on details not included in the model (White et al. 2016).

In this work, we use a Bayesian experimental design (BED) method that selects experimental designs based on the expected reduction in parameter uncertainty as quantified by the determinant of the Fisher information matrix (FIM) (Transtrum et al. 2012; Thompson et al. 2022). To determine if there was any significant benefit obtained using DoE, we compared the reduction in parameter uncertainty from performing experiments suggested by the BED method to the reduction achieved from performing arbitrarily chosen experiments (Fig. 6). We evaluated the results using a couple of matrices – the percentage of parameters with standard deviations that were large (within an order of magnitude of the relevant parameter) to indicate the overall

precision of the parameter estimates, and by the MSE to indicate the accuracy of the model's predictions.

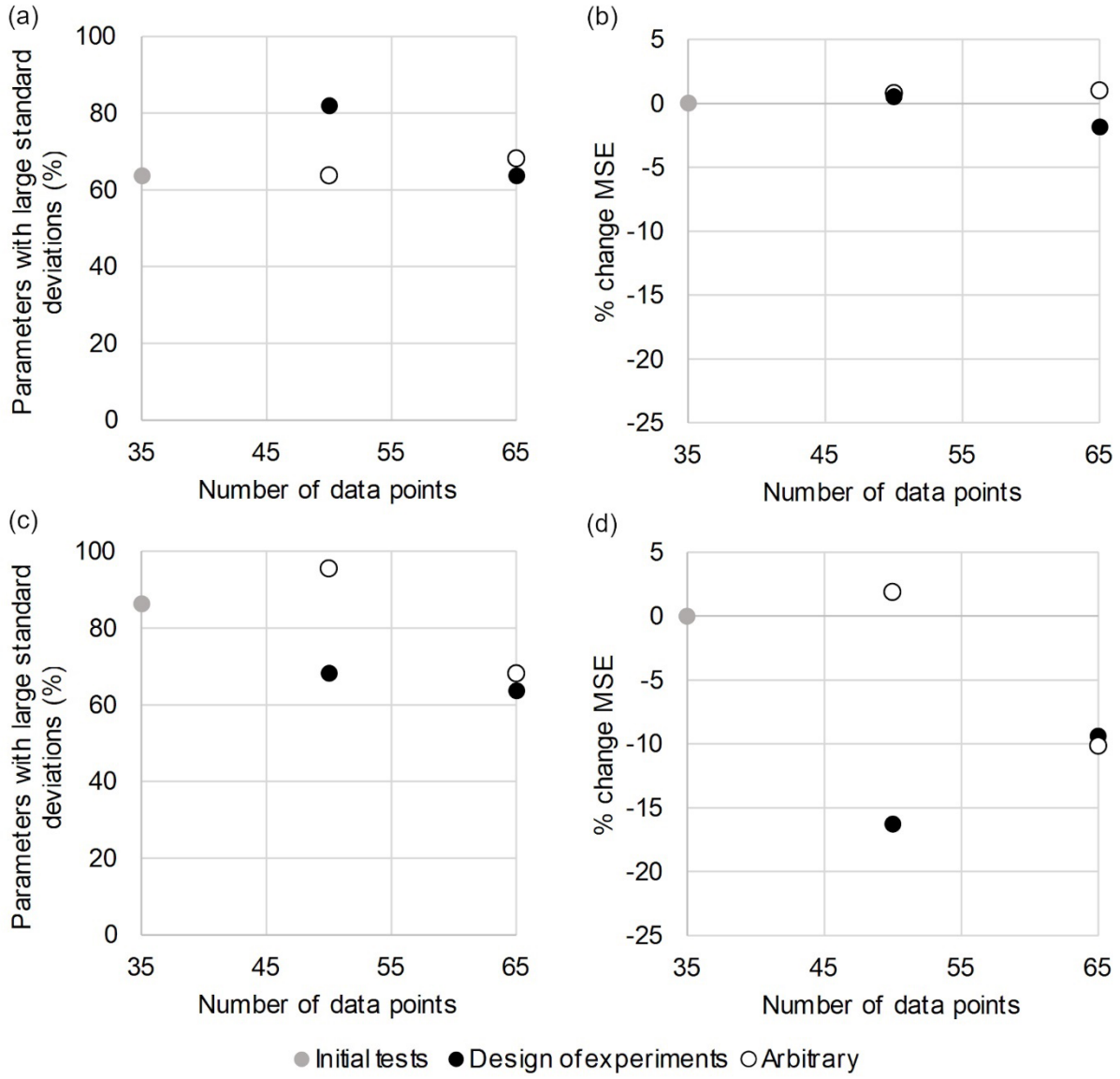


Figure 6: DoE slightly improved the precision of the parameter estimates and the model prediction accuracy. (a) Using simulated data with 15% noise, the percentage of large parameter uncertainties (standard deviation within one order of magnitude of the parameter value) remained consistent and (b) the MSE did not change significantly compared to the initial tests. (c) Using experimental data, the percentage of large parameter uncertainties decreased slightly and (d) the model predictions improved relative to the initial tests, but did not continue to improve as more data was added. Each round added 3 additional experiments, consisting of 5 time points measured for each experiment. For the DoE rounds, 3 experiments chosen from the top 20 experiments suggested by the DoE algorithm were added. For the control rounds, data

425 from 3 initial conditions not included in the DoE suggestions were added (50 mM Gly, 25 mM
426 Gly and 25 mM Ala, and 50 mM Ala).

427 In our preliminary tests using simulated data with artificial noise, adding results from
428 experiments suggested by the DoE method did not reduce the number of parameters with large
429 standard deviations or improve the accuracy of the model predictions. This suggests that the poor
430 precision of the parameter estimates may not be caused by poor data coverage, and is instead a
431 consequence of the model structure. When applied to our experimental data, the addition of
432 results suggested by the algorithm did decrease the number of parameters with large standard
433 deviations and improved the model predictions relative to the initial tests, however, there was
434 significantly less improvement from the second round of additional experiments than there was
435 in the first. The simulated results suggest a limit to how much additional data can improve the
436 parameter estimates and highlight that the model structure is responsible for sloppiness. Even
437 after nearly doubling the amount of data included in our original tests, neither the experimental
438 nor the simulated system ever had fewer than 60% of parameters with large standard deviations
439 and the model predictions were essentially unchanged. Overall, it seems unlikely that continued
440 cycles would significantly improve the parameter estimates to the extent that it would allow us to
441 attach any physical significance to their numerical values.

442 Data suggested by the DoE algorithm typically had similar or better performance than the
443 data that was added arbitrarily. However, we cannot conclude there is a significant improvement
444 from using the DoE algorithm, because during the second round of experiments using arbitrary
445 data produced very similar results in all cases. Concerning the experimental results, conclusively
446 determining whether the selections of the DoE algorithm are an improvement over randomly
447 selected conditions would require performing many additional experiments. Within the existing
448 results, we noted that model prediction errors occasionally increased when more data was added,
449 which can be a consequence of overfitting, however, there was no consistent trend of samples
450 outside of the training data set having significantly higher prediction errors, suggesting
451 overfitting is not likely (Supplementary Information 6). Because the increases in prediction error
452 are small, they are probably an incidental consequence of the noise in the data and the limited
453 sample size.

454 There are several possible reasons why DoE did not consistently improve the precision
455 of the parameter estimates this model. The precision of a sloppy model is limited by the most
456 variable parts, so experimental noise may be preventing key features from being determined
457 more precisely (Gutenkunst et al. 2007a). The prescribed range of initial conditions may have
458 also been too restrictive. We only included initial conditions with various concentrations of
459 monomers because amino acids and peptides can participate in different reaction mechanisms
460 with TP. Since these mechanisms were not being explicitly separated in the model, initial
461 conditions with large concentrations of peptides could have inadvertently led to measuring the
462 parameters for a different reaction mechanism. Rather than risk measuring the kinetics of a
463 different mechanism, which would undermine the assumption that each experiment had the same

kinetic parameter value, we chose to use a more limited system definition. However, this also may have limited our ability to constrain some parts of the network. Finally, as DoE methods are statistically based approaches that rely on existing results, they can be sensitive to noise in the data. As a result, it may be difficult to predict how parameter uncertainties will change as additional data is added. Therefore, because sloppy networks tend to be better at producing accurate predictions than accurate parameter estimates, approaches that aim to improve predictions rather than parameter uncertainties may be more useful.

Model Limitations

The mass-action style model used here is a significant simplification of the reactions occurring in the actual experimental system. TP-activated peptide bond formation involves not only multiple intermediates but likely multiple reaction mechanisms, which were not fully described in this model (Boigenzahn & Yin 2022). Certain products, like the cyclic dimers 2,5-diketopiperazine were not detectable or quantifiable in our analysis. Merging the isomeric peptide species also may have increased the experimental error slightly, since not all isomers have the same absorbance. However, on average, the species balances of glycine and alanine were about 90% accurate, suggesting that any products missed by our analysis were probably not dominant products in the system. While we acknowledge the simplifications and sources of noise in our experiments, it is important to note that the model generated high parameter standard deviations when extremely small amounts of noise were added to simulated data. It may not be possible to fit the current version of the peptide network with high precision from experimental data.

It might be possible to alleviate sloppiness by replacing the generic reversible reactions in this model with more detailed descriptions and measurements of intermediates. However, this would significantly increase the resources needed for experimental and statistical analysis. Additionally, this model does not account for increasing concentration of all species as the sample dries. The volume could be included as a dynamic term in the network model, but it complicates parameter estimation because of the infinite limits that occur as the volume approaches zero. There are also potential reactions that occur almost exclusively in the solid phase (Napier & Yin, 2006). We chose to neglect any concentration effects or details of the TP reaction mechanism and instead explored the feasibility of creating a model that predicted overall peptide production.

Conclusion

Although we were able to fit kinetic parameters to the peptide reaction network in our simulated tests, in practice the parameter estimations were poorly constrained due to sloppiness. Neither network reduction nor statistical design of experiments were particularly successful for reducing sloppiness or improving the precision of the parameter estimates for this example. Sloppiness precludes us from drawing any physical conclusions based on the individual values of the parameters estimated in these models, but this approach is still an effective way to make

model predictions based on relatively few time points. The predictive capacity of the model may be useful for forming hypotheses about the behavior of systems that pass through multiple conditions sequentially, or simply estimating equilibrium conditions based on short-term experiments.

Our goal was not only to explore the kinetics of these specific reactions, but to evaluate the potential challenges and opportunities of applying mathematical tools, which were originally developed for biological networks to prebiotic chemical systems. Sloppiness is a challenge when studying the kinetics of complex nonlinear system models but may be an interesting property in the broader context of the chemical origins of life; sloppiness has been suggested as a possible non-adaptive explanation for the robustness of many multiparameter biological systems (Daniels et al. 2008). This idea suggests that many complex networks, ranging from those found in biology to those that are randomly generated, have similar behavior across large areas of the parameter space. This implies that robustness, in this case a reaction network's ability to achieve similar outcomes despite variation in its parameter values, can emerge from complexity even when it is not specifically selected for. The feature of intrinsic robustness in sufficiently large multiparameter networks observed in deep neural networks, which can be dramatically complex but highly accurate, and is an open area of investigation in the machine learning community (Belkin, et al. 2019). As a result, there is a significant incentive to work towards studying more complex experimental origins of life systems.

Adapting systems biology tools to study complex origins of life experiments lends itself to an interdisciplinary approach, since many methods can be difficult to implement or even approach without expert assistance. Demonstrative studies like this one can improve experimentalists' understanding of what data analysis approaches are available, what their limitations are, and what results they can provide. We hope that using computational networks to analyze experiments will become more commonplace and enable the study of more complex origins of life reaction networks.

Computational Methods

The usefulness of a parametric model is limited by our ability to accurately determine the values of the corresponding parameters. A large body of work has detailed various parameter fitting or regression techniques that can be used to build these models (Bard 1974). The most popular parameter estimation method is maximum likelihood estimation (MLE). In MLE, the noise from experimental measurements (ϵ) is treated as a random variable that captures the error between the model predictions and the observed output values:

$$\boldsymbol{y} = m(X; \boldsymbol{\theta}) + \boldsymbol{\epsilon} \qquad (\mathbf{1})$$

Where $\epsilon \in \mathbb{R}^S$, S is the number of observations (measurements) available, m is the model and $\boldsymbol{\theta} \in \mathbb{R}^n$ are its n parameters. The set of output observations is stored in the vector $y \in \mathbb{R}^S$, and

$X \in \mathbb{R}^{S\times K}$, known as the design or feature matrix, is structured so that the s^{th} row corresponds to the s^{th} observation, $\mathbf{x}_s$, and the k^{th} column corresponds to the k^{th} input variable x_k. Combining MLE's assumption that $\boldsymbol{\theta}$ and X are deterministic variables with the most common noise model, the Gaussian or normal distribution ($\epsilon \sim \mathcal{N}(0,\Sigma)$), where Σ is the covariance of the noise) allows us to exploit the fact that the sum of normal distributions is also a normal distribution. We can use this to calculate the distribution for the observations vector, $y \sim \mathcal{N}(m(X,\boldsymbol{\theta}),\Sigma)$. The goal of MLE is then to find the values of $\boldsymbol{\theta}$ that best account for the experimental observations, or the values for $\boldsymbol{\theta}$ that best parameterize this output distribution. This is done by determining the values that maximize the log-likelihood function, $L(\boldsymbol{\theta})$:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, L(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, log\, f(\mathbf{y}|X,\boldsymbol{\theta},\Sigma) \tag{2}$$

where $f\left(\mathbf{y}\,\middle|\,X,\boldsymbol{\theta},\Sigma\right)$ is the likelihood (or conditional probability) that the outputs in $\mathbf{y}$ would be observed given values for X, $\boldsymbol{\theta}$, and Σ. For the given distribution of $\mathbf{y}$:

$$f(\mathbf{y}|X,\boldsymbol{\theta},\Sigma) = \left((2\pi)^{-\frac{S}{2}}|\Sigma|^{-\frac{1}{2}}\right)\exp\left(-\frac{1}{2}\big(\boldsymbol{y}-m(X;\boldsymbol{\theta})\big)^T\Sigma^{-1}\big(\boldsymbol{y}-m(X;\boldsymbol{\theta})\big)\right) \tag{3}$$

The well-known ordinary least squares regression problem is a special case of MLE where the model is linear and Σ is a diagonal matrix composed of identical values (σ^2).

A common issue with MLE is that (2) can have multiple solutions ($L(\boldsymbol{\theta})$ is nonconvex), as is often the case with nonlinear models. However, some of these solutions may contain parameter values that are not physically sensible, making the solution invalid. One way to overcome this limitation is to shift the goal of (2) from maximizing the probability of measuring the observed outputs given a set of parameters to maximizing the probability of a set of parameters being correct given a set of observations. Mathematically, this is done using Bayes' theorem, $f\left(\boldsymbol{\theta}\,\middle|\,\mathbf{y}\right) \propto f\left(\mathbf{y}\,\middle|\,\boldsymbol{\theta}\right)f(\boldsymbol{\theta})$, and changes the likelihood function to:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \log f(\mathbf{y}|X,\boldsymbol{\theta},\Sigma) + \log f(\boldsymbol{\theta}) \tag{4}$$

where now we no longer assume that $\boldsymbol{\theta}$ is deterministic but instead has some distribution (e.g., $\boldsymbol{\theta} \sim \mathcal{N}(\overline{\boldsymbol{\theta}},\Sigma_{\boldsymbol{\theta}})$) that is captured by the prior $f(\boldsymbol{\theta})$. This term can be used to input any prior knowledge or expectation one might have over the values of the model parameters (e.g., must have a certain sign, lay within a specified range, etc.) and thereby constrain the search to values of $\boldsymbol{\theta}$ that satisfy the desired criteria. If $\mathbf{y}$ and $\boldsymbol{\theta}$ are normally distributed, then (4) can be expressed as:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\frac{1}{2}\big(\mathbf{y}-m(X;\boldsymbol{\theta})\big)^T\Sigma^{-1}\big(\mathbf{y}-m(X;\boldsymbol{\theta})\big) + \frac{1}{2}(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}})^T\Sigma_{\theta}^{-1}(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}}) \tag{5}$$

Note that the first term will be minimized when the model predictions exactly match the output observations, while the second term will be minimized when $\boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$. To perform the optimization

of model parameters, we use the L-BFGS-B algorithm from SciPy's minimize function with a tolerance for termination of 1e-3. As a result, Bayes' estimation seeks to balance the fit of the model with the prior knowledge over the parameters that is available. We use an Expectation-Maximization (EM) algorithm to determine the covariance matrix of the measurement noise and the parameter prior that maximizes the model evidence (Thompson et al. 2022).

Due to the randomness in $\mathbf{y}$, the selected parameters $\boldsymbol{\theta}^*$ will exhibit an inherent uncertainty that is determined by how well the estimates are constrained by experimental data. The parameter uncertainty is largely controlled by the model structure as well as the quality and quantity of the available data. If a model is selected where certain inputs are not strong predictors of the outputs or are dependent on other inputs, or if the dataset is too small or contains redundant samples, then $\boldsymbol{\theta}^*$ will be imprecise. This is a major issue as it can lead to overfitting, where m is not able to make accurate predictions at values of x that are outside of the dataset.

An estimate of the parameter uncertainty can be obtained from the eigenvalues of the Hessian matrix, $\mathcal{H}(\mathbf{y};\boldsymbol{\theta})$, also known as the Fisher information matrix (FIM) in the context of parameter estimation, which is defined as:

$$\mathcal{H}_{i,j} = \frac{\partial^2 L}{\partial\theta_i \partial\theta_j} \tag{6}$$

The eigenvalues of the Hessian serve as an estimate of data sufficiency. From calculus we know that the second derivative of a function, f'', determines if a critical point $(f' = 0)$ is a maximum $(f'' < 0)$, a minimum $(f'' > 0)$, or an inflection point $(f'' = 0)$, which could be either a minimum, a maximum, or neither. Additionally, we can also estimate how sharp or defined an extremum is from the value of f''. As a result, we can use $\mathcal{H}(\mathbf{y};\boldsymbol{\theta})$ to gauge the quality of the obtained solution. For example, if all the eigenvalues of $\mathcal{H}(\mathbf{y};\boldsymbol{\theta})$ are large and positive $(\gg 0)$, this implies that $\boldsymbol{\theta}^*$ sits in a well-defined minimum and provides a precise estimate of the parameters. If all the eigenvalues are positive and one or more are small $(\ll 1)$, then the minimum is not sharp, and the parameter estimates will be ill-defined and exhibit high variability. Finally, if $\mathcal{H}(\mathbf{y};\boldsymbol{\theta})$ has any eigenvalues equal to zero, then $\boldsymbol{\theta}^*$ lays on a flat surface and cannot be uniquely estimated from the data; in other words, $\boldsymbol{\theta}^*$ has infinite variability.

If the precision of $\boldsymbol{\theta}^*$ is deemed to be too low, there are two methods that can be used to improve the quality of the estimates. The first, known as system identification, involves the structure of the model and the selection of the input variables. We can determine the relative importance of the input variables using a feature importance technique such as automatic relevance determination (ARD), or model class reliance (MCR), or as used in this paper, sparse principal component analysis (SPCA) (Zou et al. 2006). This information can then be used to restructure *m* to eliminate any redundant inputs.

If system identification is not able to reduce the uncertainty of the parameter estimates to a desired level, a second approach is to collect additional data. However, the data must provide

additional information beyond what is already contained in the current dataset to have any chance of improving the parameter estimates. One way to achieve this is by using a design of experiments (DoE) algorithm to select experiments that have a maximal value. Depending on the goal of the experiments (optimization, discovery, or both), their value can be measured by the information content they provide or by their predicted proximity to a desired set of properties. There is a rich variety of DoE algorithms to select from such as response surface methodology (RSM), screening, factorial design, etc. (Fisher 1937; Box & Wilson 1951; Shevlin 2017). A common metric to evaluate the optimality of candidate experimental designs is the determinant of the FIM. For any candidate experimental design, X, the FIM is computed as

$$\mathcal{H}_{i,j} = \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = \Sigma^{-1}_{\theta_{i,j}} + \frac{\partial m(X, \boldsymbol{\theta})}{\partial \theta_i} \Sigma^{-1} \frac{\partial m(X, \boldsymbol{\theta})}{\partial \theta_j} \tag{7}$$

where evaluations of the gradient with respect to model parameters is computed using the forward sensitivity equations (Ma et al. 2021).

While DoE can be very useful for improving parameter uncertainties, there are several challenges. Calculating the expected information gain (EIG) can be time consuming due to the number of operations that need to be performed for larger systems. As a result, obtaining a new batch of experiments can easily take on the order of hours depending on the size of the dataset and the number of parameters involved. Even for moderately sized models, the quantity or precision of an experimental system may not be sufficient for accurate predictions of the information generated by each experiment to be made in the first place, or the experiments that would provide the information may not be feasible in reality. Both cases seriously hinder the effectiveness of DoE methods.

Selection of experiments for the DoE method was performed as in Thompson et al. (2022). Experimental data was normalized using linear scaling to ensure that the concentration values for each species spanned $[0,1]$. Scaling the data ensures that low abundance species still affect the parameter fits, which was necessary since the experimental results span several orders of magnitude. Parameters values were limited to $[0,10]$ for simplicity, though we found that raising the upper bound had no effect if the initial guesses were single digit. Negative values had no physical meaning since both directions of the reversible reactions were already included. All computational methods were performed using Python 3.2.2. We used automatic differentiation in PyTorch to calculate the gradients of the loss function and SciPy to solve the initial value problems. Relevant code is available at https://github.com/haboigenzahn/OoL-KineticParameterEstimation.

Simulated data for testing was generated in Python 3.2.2 using SciPy 1.7.1 solve_ivp. The parameters for the simulated data were loosely based on the parameter fits of the experimental data, but were rounded to integers (Supplementary Information 7). Network figures were generated using Cytoscape 3.7.2 (Shannon et al. 1971).

Experimental Materials & Methods

All chemicals were of analytical grade purity and used without further purification. Materials were obtained from suppliers as follows: trisodium trimetaphosphate (TP) and trifloroacetic acid (TFA) from Sigma-Aldrich, sodium hydroxide from Fisher Scientific, acetone from Alfa Aesar, 9-fluorenylmethoxycarbonyl chloride (FMOC) from Creosalus, acetonitrile from VWR Chemicals, and sodium tetraborate anhydrous from Acros Organics. Reactions were carried out in 1.5 mL low-retention Eppendorf tubes. Peptide standards came from various sources: glycine, diglycine, triglycine, pentaglycine, dialanine and Ala-Gly from Sigma-Aldrich, tetraglycine from Bachem, Gly-Gly-Ala from Chem-Impex International, Ala-Gly-Gly from ChemCruz, Ala-Ala-Gly from Pepmic, and Gly-Ala-Gly, Gly-Ala-Ala, Ala-Gly-Ala and trialanine from Biomatik.

Samples were prepared with 0.15 M NaOH, various concentrations of glycine and alanine, and TP in equimolar concentration to the total amount of amino acid. Details of the initial conditions chosen are included in the supplemental information (Supplementary Information 8). Samples were placed on a heat block preheated to 90°C with the caps open and allowed to dry for 24 hours. At the end of each day of drying, samples were rehydrated with 1000 μL milliQ water preheated to about 65°C, capped and vortexed (Pulsing Vortex Mixer, Fisher Scientific) 3000 rpm until everything was dissolved, which took 1-3 minutes per sample.

To analyze the samples with UV-HPLC, they were first derivatized using FMOC, which increases the retention time and signal strength of peptide analytes. For the FMOC derivatization, 25 μL of sample was diluted with 75 μL milliQ water to put the large monomer peaks in a quantifiable range. Each sample was then mixed with 100 μL 0.1 M sodium tetraborate buffer for pH control. Finally, 800 μL 3.125 mM FMOC dissolved in acetone was added to each sample. For a sample of 0.1 M amino acid, this results in an equal concentration of FMOC and amino acid, and a slight excess of FMOC in any samples where peptide bond formation had occurred. Linear calibration curves were determined for all species using this approach (Supplementary Information 9), which were used to estimate peptide concentration based on the integrated absorbance values of the HPLC peaks of the samples.

Samples were analyzed with a Shimadzu Nexera HPLC with a C-18 column (Phenomenex Aeris XB-C18, 150 mm x 4.6 mm, 3.6 μL). Products were measured at 254 nm. UV-HPLC analysis was performed using Solvent A: milliQ water with 0.01% v/v trifluoroacetic acid (TFA) and Solvent B: acetonitrile with 0.01% v/v TFA. The following gradient was used: 0-4 min, 30% B, 4-12 min, 30-100% B, 14-15 min, 100-30% B, 15-17 min, 30% B. The solvent flow rate was 1 mL/min. Peak integration was performed using LabSolutions with the ‘Drift’ parameter set to 10000.

Declarations

Ethics Approval No approvals required.

Conflict of Interest The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Data Availability The data shown in this study is available from John Yin upon reasonable request.

Authors' Contributions All experimental data was collected and analyzed by HB. The main parameter estimation code was developed by JCT, with minor edits and applications performed by HB and LDG. The manuscript was drafted by HB and LDG and all authors contributed to subsequent editing.

Funding This research was funded by the Vilas Distinguished Achievement Professorship, the Office of the Vice Chancellor for Research and Graduate Education, the Wisconsin Institute for Discovery, all at the University of Wisconsin-Madison; and the Wisconsin Alumni Research Foundation (WARF); grant R01DK133605 from the National Institutes of Health; and grants MCB-2029281, CBET-2030750, and DMS-2151959 from the US National Science Foundation. We also acknowledge support from the GERS program at the University of Wisconsin-Madison.

Consent to Publish All authors read and approved the final manuscript.

Acknowledgements Thank you to Izabela Sibilska-Kaminski, who contributed to the experimental background for this project, and to other members of the Yin lab as well as Prof. David Baum for their thoughtful feedback. We further acknowledge financial support via the NSF-EFRI award 2132036 and the Advanced Opportunity Fellowship from the University of Wiscon- sin-Madison Graduate Engineering Research Scholars program.

714

715

716

717

718 **References**

719 Apgar, J. F., Witmer, D. K., White, F. M., & Tidor, B. (2010). Sloppy models, parameter uncertainty, and
720 the role of experimental design. *Molecular BioSystems*, *6*(10), 1890–1900.
721 https://doi.org/10.1039/b918098b

722 Apri, M., de Gee, M., & Molenaar, J. (2012). Complexity reduction preserving dynamical behavior of
723 biochemical networks. *Journal of Theoretical Biology*, *304*, 16–26.
724 https://doi.org/10.1016/j.jtbi.2012.03.019

725 Bard, Y. (1974). *Nonlinear parameter estimation* (No. 04; QA276. 8, B3.).

726 Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the
727 classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, *116*(32), 15849-
728 15854. https://doi.org/10.1073/pnas.1903070116

729 Box, G. E. P. and Wilson K. B. (1951) On the experimental attainment of optimum conditions. *Journal of*
730 *the Royal Statistical Society*, 13, 1-45. https://doi.org/10.1111/j.2517-6161.1951.tb00067.x

731 Brown, K. S., Hill, C. C., Calero, G. A., Myers, C. R., Lee, K. H., Sethna, J. P., & Cerione, R. A. (2004).
732 The statistical mechanics of complex signaling networks: Nerve growth factor signaling. *Physical*
733 *Biology*, *1*(3), 184–195. https://doi.org/10.1088/1478-3967/1/3/006

734 Brown, K. S., & Sethna, J. P. (2003). Statistical mechanical approaches to models with many poorly
735 known parameters. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related*
736 *Interdisciplinary Topics*, *68*(2), 9. https://doi.org/10.1103/PhysRevE.68.021904

737 Chaloner, K., & Verdinelli, I. (1995). Bayesian Experimental Design: A Review. *Statistical Science*,
738 *10*(3), 273–304.

739 Chis, O.-T., Banga, J. R., & Balsa-Canto, E. (2014). *Sloppy models can be identifiable*. 1–35.
740 http://arxiv.org/abs/1403.1417

741 Cobelli, C., & DiStefano, J. J. (1980). Parameter and structural identifiability concepts and ambiguities: a
742 critical review and analysis. *The American Journal of Physiology*, *239*(1), 7–24.
743 https://doi.org/10.1152/ajpregu.1980.239.1.R7

744 Coveney, P. V., Swadling, J. B., Wattis, J. A. D., & Greenwell, H. C. (2012). Theory, modelling and
745 simulation in origins of life studies. *Chemical Society Reviews*, *41*(16), 5430–5446.
746 https://doi.org/10.1039/c2cs35018a

747 Daniels, B. C., Chen, Y. J., Sethna, J. P., Gutenkunst, R. N., & Myers, C. R. (2008). Sloppiness,
748 robustness, and evolvability in systems biology. *Current Opinion in Biotechnology*, *19*(4), 389–395.
749 https://doi.org/10.1016/j.copbio.2008.06.008

750 Fisher, R. A. (1937). *Design of experiments.* Edinburgh: Oliver and Boyd, 1935.

751 Frenkel-Pinter, M., Samanta, M., Ashkenasy, G., & Leman, L. J. (2020). Prebiotic Peptides: Molecular
752 Hubs in the Origin of Life. *Chemical Reviews*, *120*(11), 4707–4765.
753 https://doi.org/10.1021/acs.chemrev.9b00664

754 Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019). A brief history of bioinformatics.
755 *Briefings in Bioinformatics*, *20*(6), 1981–1996. https://doi.org/10.1093/bib/bby063

756 Goldman, A. D., Bernhard, T. M., Dolzhenko, E., & Landweber, L. F. (2013). LUCApedia: A database
757 for the study of ancient life. *Nucleic Acids Research*, *41*(D1), 1079–1082.
758 https://doi.org/10.1093/nar/gks1217

759 Gutenkunst, R. N., Casey, F. P., Waterfall, J. J., Myers, C. R., & Sethna, J. P. (2007). Extracting
760 falsifiable predictions from sloppy models. *Annals of the New York Academy of Sciences*, *1115*,
761 203–211. https://doi.org/10.1196/annals.1407.003

762 Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. (2007a).
763 Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*,
764 *3*(10), 1871–1878. https://doi.org/10.1371/journal.pcbi.0030189

765 Hettling, H., & van Beek, J. H. G. M. (2011). Analyzing the functional properties of the creatine kinase
766 system with multiscale "sloppy" modeling. *PLoS Computational Biology*, *7*(8), 11–16.
767 https://doi.org/10.1371/journal.pcbi.1002130

768 Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K., & Ghadiri, M. R. (1996). A self-replicating peptide.
769 *Nature*, *382*, 525–528.

770 Ludlow, R. F., & Otto, S. (2008). Systems chemistry. *Chemical Society Reviews*, *37*(1), 101–108.
771 https://doi.org/10.1039/b611921m

772 Jagadeesan, P., Raman, K., & Tangirala, A. K. (2022). Bayesian Optimal Experiment Design for Sloppy
773 Systems. *IFAC-PapersOnLine*, *55*(23), 121-126. https://doi.org/10.1016/j.ifacol.2023.01.026

774 Jain, A., McPhee, S. A., Wang, T., Nair, M. N., Kroiss, D., Jia, T. Z., & Ulijn, R. V. (2022). Tractable
775 molecular adaptation patterns in a designed complex peptide system. *Chem*, *8*(7), 1894–1905.
776 https://doi.org/10.1016/j.chempr.2022.03.016

777 Johnson, E. O., & Hung, D. T. (2019). A Point of Inflection and Reflection on Systems Chemical
778 Biology. *ACS Chemical Biology*, *14*(12), 2497–2511. https://doi.org/10.1021/acschembio.9b00714

779 Ma, S., & Dai, Y. (2011). Principal component analysis based methods in bioinformatics studies.
780 *Briefings in Bioinformatics*, *12*(6), 714–722. https://doi.org/10.1093/bib/bbq090

781 Ma, Y., Dixit, V., Innes, M. J., Guo, X., & Rackauckas, C. (2021). A Comparison of Automatic
782 Differentiation and Continuous Sensitivity Analysis for Derivatives of Differential Equation
783 Solutions. *2021 IEEE High Performance Extreme Computing Conference, HPEC 2021*, *2*,
784 1–9. https://doi.org/10.1109/HPEC49654.2021.9622796

785 Maity, S., Ottelé, J., Santiago, G. M., Frederix, P. W. J. M., Kroon, P., Markovitch, O., … Roos,
786 W. H. (2020). Caught in the Act: Mechanistic Insight into Supramolecular Polymerization-
787 Driven Self-Replication from Real-Time Visualization. *Journal of the American Chemical*
788 *Society*, *142*(32), 13709–13717. https://doi.org/10.1021/jacs.0c02635

789 Maiwald, T., Hass, H., Steiert, B., Vanlier, J., Engesser, R., Raue, A., Kipkeew, F., Bock, H. H., Kaschek,
790 D., Kreutz, C., & Timmer, J. (2016). Driving the model to its limit: Profile likelihood based model
791 reduction. *PLoS ONE*, *11*(9), 1–18. https://doi.org/10.1371/journal.pone.0162366

792 Mamajanov, I., Macdonald, P. J., Ying, J., Duncanson, D. M., Dowdy, G. R., Walker, C. A., Engelhart,
793 A. E., Fernández, F. M., Grover, M. A., Hud, N. V., & Schork, F. J. (2014). Ester formation and
794 hydrolysis during wet-dry cycles: Generation of far-from-equilibrium polymers in a model prebiotic
795 reaction. *Macromolecules*, *47*(4), 1334–1343. https://doi.org/10.1021/ma402256d

796 Maria, G. (2004). A review of algorithms and trends in kinetic model identification for chemical and
797 biochemical systems. *Chemical and Biochemical Engineering Quarterly*, *18*(3), 195-222.

798 Monsalve-Bravo, G. M., Lawson, B. A. J., Drovandi, C., Burrage, K., Brown, K. S., Baker, C. M.,
799 Vollert, S. A., Mengersen, K., McDonald-Madden, E., & Adams, M. P. (2022). Analysis of
800 sloppiness in model simulations: Unveiling parameter uncertainty when mathematical models are
801 fitted to data. *Science Advances*, *8*(38). https://doi.org/10.1126/sciadv.abm5952

802 Napier, J., & Yin, J. (2006). Formation of peptides in the dry state. *Peptides*, *27*(4), 607–610.
803 https://doi.org/10.1016/j.peptides.2005.07.015

804 Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, *45*(2), 167-256.
805 https://doi.org/10.1137/S003614450342480

806 Nghe, P., Hordijk, W., Kauffman, S. A., Walker, S. I., Schmidt, F. J., Kemble, H., Yeates, J. A. M., &
807 Lehman, N. (2015). Prebiotic network evolution: Six key parameters. *Molecular BioSystems*,
808 *11*(12), 3206–3217. https://doi.org/10.1039/c5mb00593k

809 Orgel, L. E. (2010). The origin of life: A review of facts and speculation. *The Nature of Life: Classical*
810 *and Contemporary Perspectives from Philosophy and Science*, *0004*(December), 121–128.
811 https://doi.org/10.1017/CBO9780511730191.012

812 Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., Hug, S., Kreutz, C.,
813 Harms, B. D., Theis, F. J., Klingmüller, U., & Timmer, J. (2013). Lessons Learned from
814 Quantitative Dynamical Modeling in Systems Biology. *PLoS ONE*, *8*(9).
815 https://doi.org/10.1371/journal.pone.0074335

816 Rodriguez-Fernandez, M., Mendes, P., & Banga, J. R. (2006). A hybrid approach for efficient and robust
817 parameter estimation in biochemical pathways. *BioSystems*, *83*(2-3 SPEC. ISS.), 248–265.
818 https://doi.org/10.1016/j.biosystems.2005.06.016

819 Rout, S. K., Rhyner, D., Riek, R., & Greenwald, J. (2022). Prebiotically Plausible Autocatalytic Peptide
820 Amyloids. *Chemistry - A European Journal*, *28*(3). https://doi.org/10.1002/chem.202103841

821 Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv Preprint*
822 *arXiv:1609.04747*. http://arxiv.org/abs/1609.04747

823 Ruiz-Mirazo, K., Briones, C., & De La Escosura, A. (2014). Prebiotic systems chemistry: New
824 perspectives for the origins of life. *Chemical Reviews*, *114*(1), 285–366.
825 https://doi.org/10.1021/cr2004844

826 Sakata, K., Kitadai, N., & Yokoyama, T. (2010). Effects of pH and temperature on dimerization rate of
827 glycine: Evaluation of favorable environmental conditions for chemical evolution of life.
828 *Geochimica et Cosmochimica Acta*, *74*(23), 6841–6851. https://doi.org/10.1016/j.gca.2010.08.032

829 Schwartz, A. W. (2007). Intractable mixtures and the origin of life. *Chemistry and Biodiversity*, *4*(4),
830 656–664. https://doi.org/10.1002/cbdv.200790056

831 Serov, N. Y., Shtyrlin, V. G., & Khayarov, K. R. (2020). The kinetics and mechanisms of reactions in the
832 flow systems glycine–sodium trimetaphosphate–imidazoles: the crucial role of imidazoles in
833 prebiotic peptide syntheses. *Amino Acids*, *52*(5), 811–821. https://doi.org/10.1007/s00726-020-
834 02854-z

835 Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B.,
836 & Ideker, T. (1971). Cytoscape: A Software Environment for Integrated Models. *Genome Research*,
837 *13*(22), 426. https://doi.org/10.1101/gr.1239303.metabolite

838 Shevlin, M. (2017). Practical high-throughput experimentation for chemists. *ACS medicinal chemistry*
839 *letters*, *8*(6), 601-607. https://doi.org/10.1021/acsmedchemlett.7b00165

840 Sibilska, I., Feng, Y., Li, L., & Yin, J. (2018). Trimetaphosphate Activates Prebiotic Peptide Synthesis
841 across a Wide Range of Temperature and pH. *Origins of Life and Evolution of Biospheres*, *48*(3),
842 277–287. https://doi.org/10.1007/s11084-018-9564-7

843 Surman, A. J., Rodriguez-Garcia, M., Abul-Haija, Y. M., Cooper, G. J. T., Gromski, P. S., Turk-
844 MacLeod, R., Mullin, M., Mathis, C., Walker, S. I., & Cronin, L. (2019). Environmental control
845 programs the emergence of distinct functional ensembles from unconstrained chemical reactions.
846 *Proceedings of the National Academy of Sciences*, *116*(12), 5387–5392.
847 https://doi.org/10.1073/pnas.1813987116

848 Thompson, J. C., Zavala, V. M., & Venturelli, O. S. (2022). Integrating a tailored recurrent neural
849 network with Bayesian experimental design to optimize microbial community functions. *BioRxiv*, 1–
850 24.

851 Transtrum, M. K., Machta, B. B., Brown, K. S., Daniels, B. C., Myers, C. R., & Sethna, J. P. (2015).
852 Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *Journal of Chemical*
853 *Physics*, *143*(1). https://doi.org/10.1063/1.4923066

Transtrum, M. K., & Qiu, P. (2012). Optimal experiment selection for parameter estimation in biological differential equation models. *BMC bioinformatics*, *13*(1), 1-12. https://doi.org/10.1186/1471-2105-13-181

Vanlier, J., Tiemann, C. A., Hilbers, P. A. J., & van Riel, N. A. W. (2013). Parameter uncertainty in biochemical models described by ordinary differential equations. *Mathematical Biosciences*, *246*(2), 305–314. https://doi.org/10.1016/j.mbs.2013.03.006

Varfolomeev, S. D., & Lushchekina, S. V. (2014). Prebiotic synthesis and selection of macromolecules: Thermal cycling as a condition for synthesis and combinatorial selection. *Geochemistry International*, *52*(13), 1197–1206. https://doi.org/10.1134/S0016702914130102

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

von Kiedrowski, G. (1986). A Self-Replicating Hexdeoxynucleotide. *Angewandte Chemie - International Edition*, *25*(10), 932–935. https://doi.org/10.1002/anie.198609322

Wasserman, L. (2004). *All of statistics: a concise course in statistical inference* (Vol. 26, p. 86). New York: Springer.

Waterfall, J. J., Casey, F. P., Gutenkunst, R. N., Brown, K. S., Myers, C. R., Brouwer, P. W., Elser, V., & Sethna, J. P. (2006). Sloppy-model universality class and the vandermonde matrix. *Physical Review Letters*, *97*(15), 1–4. https://doi.org/10.1103/PhysRevLett.97.150601

White, A., Tolman, M., Thames, H. D., Withers, H. R., Mason, K. A., & Transtrum, M. K. (2016). The Limitations of Model-Based Experimental Design and Parameter Estimation in Sloppy Systems. *PLoS Computational Biology*, *12*(12), 1–26. https://doi.org/10.1371/journal.pcbi.1005227

Wieland, F. G., Hauber, A. L., Rosenblatt, M., Tönsing, C., & Timmer, J. (2021). On structural and practical identifiability. *Current Opinion in Systems Biology*, *25*, 60–69. https://doi.org/10.1016/j.coisb.2021.03.005

Yu, S. S., Krishnamurthy, R., Fernández, F. M., Hud, N. V., Schork, F. J., & Grover, M. A. (2016). Kinetics of prebiotic depsipeptide formation from the ester-amide exchange reaction. *Physical Chemistry Chemical Physics*, *18*(41), 28441–28450. https://doi.org/10.1039/c6cp05527c

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, *15*(2), 265–286. https://doi.org/10.1198/106186006X113430