Supporting Material

Model identification of a template-directed peptide network for optimization in a continuous reactor

Andres F. Hernandez, Michael J. Wagner and Martha A. Grover*

Keywords: peptide reaction networks, systems chemistry, model identification, optimization, auto-catalysis, cross-catalysis.

The chemical system used in this communication is part of a complex molecular network described by Ashkenasy et al. [1] as an example of rational design in systems chemistry. The molecular network is engineered using nine different peptide sequences that interact via a template-directed peptide fragment condensation reaction at neutral pH. The template-directed peptide network exhibits aspects of auto-catalysis, cross-catalysis and competition for limited resources. This supplemental material contains the documentation and analysis carried out for this chemical model including mechanisms for auto- and cross-catalytic activity, kinetic parameter estimation, and simulation of this peptide network in the presence of an influent stream of reactants and an effluent stream of products (i.e. open mass system). These results shows how one can take advantage of a rational systems chemistry design in a continuous flow process.

S1 Mathematical modeling of an α coiled-coil peptide network

The template-directed peptide network is created from a mixture of a common nucleophile peptide sequence N with nine different electrophile peptide sequences, E_{1-9}. The peptide fragments are modified to undergo coupling by Kent ligation [2], due to the C-terminal thiolester in the peptide fragments E_i and the N-terminal cysteine residue in the peptide fragments N. The resulting longer peptides are identified as T_{1-9}. This means that each electrophile E_i is competing for the common nucleophile N.

The sequences of the peptide fragments are designed to have an α coiled-coil folding motif that allows them to have non-covalent interactions. Thanks to these non-covalent interactions, it is possible to use the templates T_i as template-directed catalysts for the ligation

*Corresponding Author. E-mail: martha.grover@chbe.gatech.edu, School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0100. Tel.: +1 (404)894-2878. Fax: +1 (404)894-2866.

reaction, since they provide an active surface for this reaction [4]. Previously, Severin et al. [11] elucidated that a duplex template T_iT_i is the corresponding catalyst structure for the ligation reaction. From a broader perspective, the peptide fragments and their templates have the capability to create auto-catalytic and cross-catalytic cycles for the formation of new peptide molecules T_i, therefore increasing the complexity of the molecular reaction network. The chemical system in the manuscript uses two of the nine peptide sequences from Ashkenasy et al. [1]. The peptide fragments are

$$\begin{aligned} E_1 &= \text{RVARLEREVSELERKVA} \\ E_4 &= \text{RVARLEKKVSALKKKVA} \\ N &= \text{CLELEVARLKKLVGE} \end{aligned}$$

where the thiolester at the C-terminal of the electrophiles is ethanesulfonic acid. The reason for choosing these two templates is because they are the only ones in the original paper [1] for which kinetic data is available for both auto- and cross-catalysis. The characterization of the network reactions is made using chromatography (RP-HPLC), by extracting aliquots at different time points. Since the peptide assemblies are disassembled in the RP-HPLC measurement, only the total amount of each peptide can be quantified via RP-HPLC.

S1.1 Parameter estimation and identifiability analysis

An identification procedure is used to estimate the corresponding kinetic parameters of the reactions in the network. First, a system of ordinary differential equations is written from the proposed mechanism for the auto-catalytic and cross-catalytic cycles as described by Ashkenasy et al. [1] using mass-action kinetic rates. The mathematical model is written in the form of

$$\frac{dx_a}{dt} = \sum_{b=1}^{B} \nu_{a,b}\, r_b \quad i = 1, \ldots, A \tag{1}$$

where x_a corresponds to the concentration of species a [μM], r_b is the reaction rate expression for the reaction b, and $\nu_{a,b}$ is the stoichiometric coefficient of species a in reaction b. The expressions for the different reaction rates are specified in each of the following tables (Tables S1–S6) as a function of the corresponding species concentrations. The initial conditions for these differential equations are determined by the specific experiments that were performed [1].

The numerical solution of the system of equations is computed using the ordinary differential equation (ODE) solver `ode15s` in MATLAB R2013a [6]. The experimental data points from Figures 8a and 11 in Reference [1] are used to formulate a least-squares optimization problem to estimate the kinetic parameters in the model. The optimization procedure involves a Latin Hypercube sampling [8] over the kinetic parameter space to select a suitable initial guess for the derivative-free optimization routine `patternsearch` in MATLAB R2013a [6].

Due to the limited experimental information obtained from the original Ashkenasy paper, a model identifiability and sensitivity analysis was performed. This type of analysis, common in systems biology, is helpful to avoid over-fitting the data. The method eliminates parameters

and mechanisms from the model if they cannot be identified from the dataset [10]. The main concept in the model identification is the definition of the sensitivity matrix $S_i \in \mathbb{R}^{N_i \times J}$.

$$S_{i,n,j} = \left[\frac{\partial y_i\left(t_n, \boldsymbol{\theta}\right)}{\partial \theta_j}\right]\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{2}$$

The sensitivity matrix consists of the derivatives of each measured variable y_i at time point t_n with respect to the kinetic parameter θ_j, evaluated at the estimated parameter set $\hat{\boldsymbol{\theta}}$ from the optimization. In total there will be I sensitivity matrices ($i = 1, \ldots, I$), one for each measured species in the peptide network. The size of each of these S_i matrices is defined by the number of experimental time points N_i of the measured species i ($n = 1, \ldots, N_i$) and the number of kinetic parameters J ($j = 1, \ldots, J$), one for each reaction in the peptide network. In this study, the elements in the sensitivity matrices were not calculated analytically, but approximated by a centered finite difference formula.

Once an optimal solution is found for the full kinetic parameter set $\hat{\boldsymbol{\theta}} \in \mathbb{R}^J$, the local identifiability of this parameter set is evaluated by calculating the Fisher information matrix $FIM \in \mathbb{R}^{J \times J}$ [5, 7]

$$FIM = \sum_{i=1}^{I} \sum_{n=1}^{N_i} \frac{1}{\sigma_i^2}\, \mathbf{S}_{i,n}^T \mathbf{S}_{i,n} \tag{3}$$

where $\mathbf{S}_{i,n} \in \mathbb{R}^J$ is row n in the sensitivity matrix S_i. The values of σ_i^2 represent the measurement error covariance at each of the data points of the measured variable i, and it can be estimated from the sum-of-squares error in the optimization procedure as

$$\sigma_i^2 = \frac{\sum_{n=1}^{N_i} \left(y_i^{\exp}\left(t_n\right) - y_i\left(t_n, \hat{\boldsymbol{\theta}}\right)\right)^2}{N_i} \tag{4}$$

where $y_i^{\exp}\left(t_n\right)$ and $y_i\left(t_n, \hat{\boldsymbol{\theta}}\right)$ denote the measurement and model-predicted values at time point t_n.

If the FIM is singular, it indicates the presence of unidentifiable parameters, and/or correlations between parameters [9]. In this work, the LAPACK reciprocal condition estimator available in MATLAB R2013a, `rcond`, was used to determine the singularity of the FIM matrix [12]. If $\texttt{rcond}(FIM) < 10\epsilon$, where ϵ is the floating point relative accuracy (2.2×10^{-16}), the FIM was considered to be singular. When the FIM is singular, a sensitivity analysis is performed to identify insensitive parameters in the model. For the comparison of the different kinetic parameters, the coefficients in the sensitivity matrices S_i are normalized by the values of the parameter θ_j and the experimental data point $y_i^{\exp}\left(t_n\right)$. The normalized sensitivity coefficients are calculated as

$$nS_{i,n,j} = \frac{\theta_j}{y_i^{\exp}\left(t_n\right)} \left[\frac{\partial y_i\left(t_n, \boldsymbol{\theta}\right)}{\partial \theta_j}\right]\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{5}$$

Normalized parameter sensitivities of the model are calculated by summing up the normalized sensitivity coefficients over all time points n and all measured variables i by the equation

$$nS_j = \sum_{i=1}^{I} \sum_{n=1}^{N_i} \left|nS_{i,n,j}\right| \tag{6}$$

Low values of the normalized parameter sensitivities will indicate parameters that are insensitive with respect to other parameters in the model. The reaction with the least sensitive kinetic parameter is removed from the mechanism. The differential equations in Equation (1) are written again according to this change and a new optimization procedure is performed. The model reduction procedure continues until the estimated FIM matrix is no longer singular.

S2 Auto-catalytic pathway for template-directed peptide network. Template T_1

Using the auto-catalytic mechanism described by Ashkenasy et al. [1], the corresponding reactions for the formation of template T_1 from electrophile E_1 and nucleophile N are

$$E_1 + N \xrightarrow{k_1} T_1 + RSH$$
$$2\,T_1 \underset{k_3}{\overset{k_2}{\rightleftharpoons}} T_1T_1$$
$$E_1 + N + T_1T_1 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} E_1NT_1T_1$$
$$E_1NT_1T_1 \xrightarrow{k_6} T_1T_1T_1 + RSH$$
$$T_1T_1T_1 \underset{k_8}{\overset{k_7}{\rightleftharpoons}} T_1 + T_1T_1$$

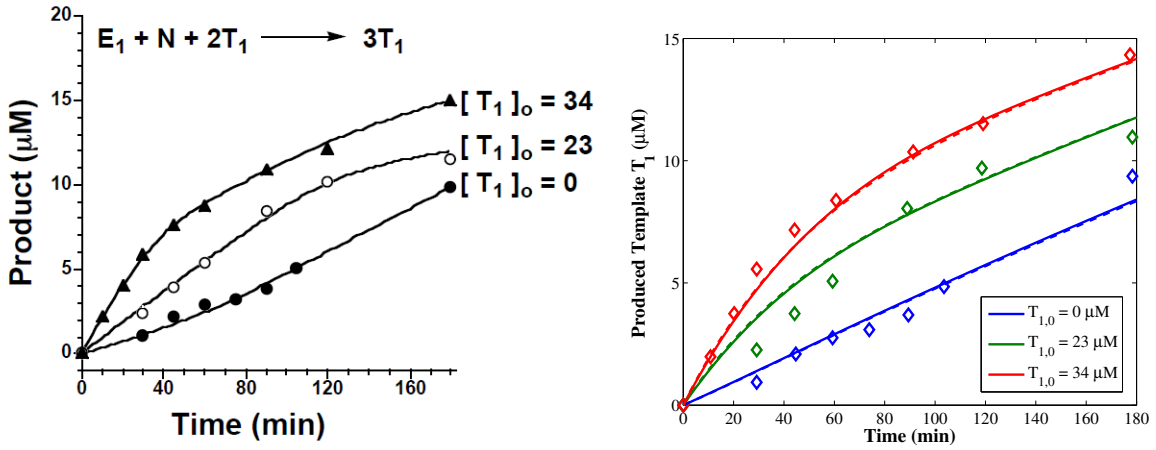
where T_1 and T_1T_1 are the single and duplex template structures that are present in the auto-catalytic pathway. $T_1T_1T_1$ is an intermediate triplex template structure in the auto-catalytic pathway and RSH is a secondary product formed after the Kent ligation, where R is the corresponding thiolester in the ligation. One minor difference between the above mechanism and the one presented by Ashkenasy et al. [1] is the omission of the reactions related with the rearrangement after the initial trans-thiolesterification between the peptide fragments. The assumption in the modified mechanism is that under neutral aqueous conditions, this intermediate spontaneously rearranges directly into the native peptide bond [4]. Therefore, the reaction is written from the quadruplex intermediate $E_1NT_1T_1$ directly to the triplex $T_1T_1T_1$.
Using the parameter sensitivity and identifiability analysis described in Section S1.1 the full eight-reaction auto-catalytic mechanism based on Ashkenasy's work is reduced to a four-reaction mechanism described as follows:

$$E_1 + N \xrightarrow{\hat{k}_1} T_1 + RSH$$
$$2\,T_1 \xrightarrow{\hat{k}_2} T_1T_1$$
$$E_1 + N + T_1T_1 \xrightarrow{\hat{k}_3} T_1T_1T_1 + RSH$$
$$T_1 + T_1T_1 \xrightarrow{\hat{k}_4} T_1T_1T_1$$

The model identification suggests that the reversible reactions in the eight-reaction mecha-

nism are not at equilibrium, and that there is an overall net reaction rate in each of these reversible reactions that can be represented in the model by an irreversible reaction in the appropriate direction. In addition, the parameter sensitivity analysis identifies high positive correlation between the parameters k_4 and k_6 in the auto-catalytic mechanism. This was addressed by combining these reactions and eliminating the intermediate $E_1NT_1T_1$ from the model. Bear in mind that the results of the model identification do not imply that there are not reversible reactions in the mechanism, nor that the intermediate species $E_1NT_1T_1$ does not exist. It only suggests that these effects cannot be identified by the available data. With a much larger set of data from these experiments, the remaining kinetic parameters from the model might be identified.

Figure S1 shows a comparison between the original experimental data and the results obtained in the parameter estimation procedure for both full and reduced auto-catalytic mechanisms, showing good agreement in both cases. A summary of the estimated parameters for both models is included in Tables S1 and S2. Because the full model is not identifiable, the parameters in Table S1 are not unique best-fit values, but rather are only one of the many parameter sets that minimizes the fitting error. Thus, no mechanistic interpretation should be made based on the values in Table S1. In contrast, the parameter values in Table S2 are the unique best fit values for the reduced model. It is also important to remember that, given more data, the set of reactions in the reduced model might be larger. The reduced model presented here is the minimal model associated with this particular data set.



(a) (b)

Figure S1: Model fitting for auto-catalytic rates of template T_1. (a) Original experimental data points from Reference [1]. The template reactions were performed by using 100 μM E_1 and 100 μM N, in the presence or absence of various initial concentrations of $T_{1,0}$ as indicated. Permission for reprint, Copyright (2004). National Academy of Sciences, U.S.A. (b) Results of the model identification procedure to estimate the values of the kinetic rate constants. The solid lines represent the parameter estimation results using the eight-reaction full auto-catalytic T_1 mechanism, and the dashed lines represent the parameter estimation results using the four-reaction reduced auto-catalytic T_1 mechanism.

Reaction	**Reaction Rate Law**	**Parameters**	**Parameter Units**
r_1	$k_1 [E_1] [N]$	$k_1 = 4.5739 \times 10^{-6}$	$\frac{1}{\mu M \text{ min}}$
r_2	$k_2 [T_1]^2$	$k_2 = 1.2303 \times 10^{3}$	$\frac{1}{\mu M \text{ min}}$
r_3	$k_3 [T_1T_1]$	$k_3 = 9.6248 \times 10^{-14}$	$\frac{1}{\text{min}}$
r_4	$k_4 [E_1] [N] [T_1T_1]$	$k_4 = 1.9355 \times 10^{-2}$	$\frac{1}{\mu M^2 \text{ min}}$
r_5	$k_5 [E_1NT_1T_1]$	$k_5 = 4.5179 \times 10^{0}$	$\frac{1}{\text{min}}$
r_6	$k_6 [E_1NT_1T_1]$	$k_6 = 1.4419 \times 10^{-2}$	$\frac{1}{\text{min}}$
r_7	$k_7 [T_1T_1T_1]$	$k_7 = 2.2262 \times 10^{-24}$	$\frac{1}{\text{min}}$
r_8	$k_8 [T_1] [T_1T_1]$	$k_8 = 2.7587 \times 10^{2}$	$\frac{1}{\mu M \text{ min}}$

Table S1: Estimated kinetic parameters for auto-catalytic mechanism of template T_1 using the full eight-reaction model. All concentrations in the reaction rate laws are measured in μM.

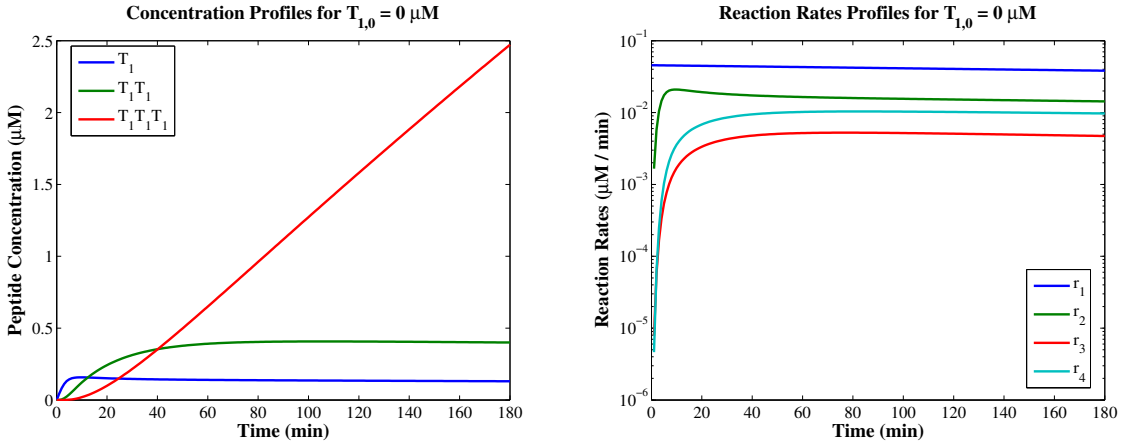
Reaction	**Reaction Rate Law**	**Parameters**	**Parameter Units**
r_1	$\hat{k}_1 [E_1] [N]$	$\hat{k}_1 = 4.5558 \times 10^{-6}$	$\frac{1}{\mu M \text{ min}}$
r_2	$\hat{k}_2 [T_1]^2$	$\hat{k}_2 = 8.4053 \times 10^{-1}$	$\frac{1}{\mu M \text{ min}}$
r_3	$\hat{k}_3 [E_1] [N] [T_1T_1]$	$\hat{k}_3 = 1.4041 \times 10^{-6}$	$\frac{1}{\mu M^2 \text{ min}}$
r_4	$\hat{k}_4 [T_1] [T_1T_1]$	$\hat{k}_4 = 1.8619 \times 10^{-1}$	$\frac{1}{\mu M \text{ min}}$

Table S2: Estimated kinetic parameters for auto-catalytic mechanism of template T_1 using the reduced four-reaction model. All concentrations in the reaction rate laws are measured in μM.

Figure S2 shows the concentration profiles of the species that contribute to the total T_1 concentration, as well as the reaction rate values as a function of time, for the reduced four-reaction auto-catalytic mechanism. The first conclusion from these studies is that the triplex $T_1T_1T_1$ is a very stable species, as indicated in the model identification in which the reaction $T_1T_1T_1 \rightarrow T_1 + T_1T_1$ is removed. Another conclusion from this study is that T_1 is mostly formed via the uncatalyzed reaction r_1 instead of the catalyzed reaction r_3. The catalyzed reaction r_3 shows some level of activity in the presence of an initial quantity of free T_1 (see Figures S2(d) and S2(f)).

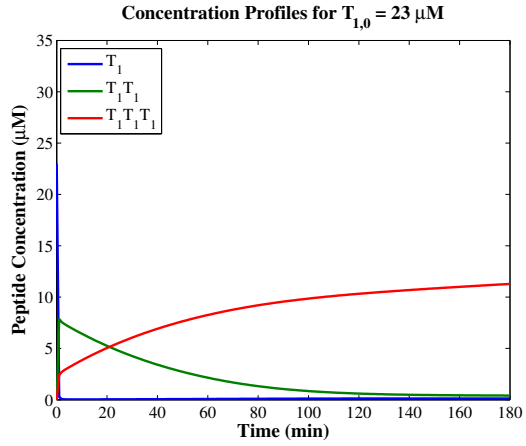
S3 Auto-catalytic pathway for template-directed peptide network. Template T_4

Section S3 covers all important aspects in the analysis of the auto-catalytic pathway for template T_4. The mechanism used to represent the auto-catalytic behavior of template T_4 is similar to the one used for template T_1, only changing the electrophile concentration from E_1 to E_4. Moreover, the model identification procedure suggests that both templates can be represented by similar reduced models. Therefore, the proposed eight-reaction auto-catalytic

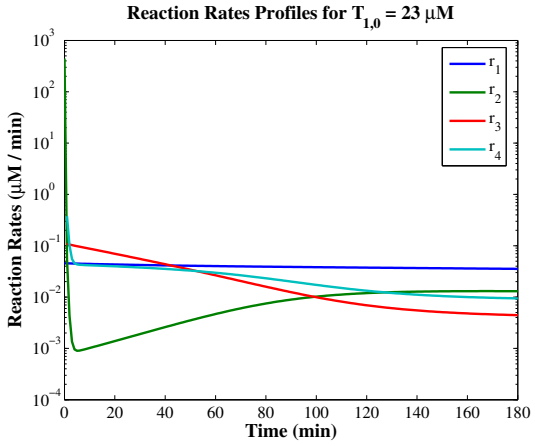


(a)

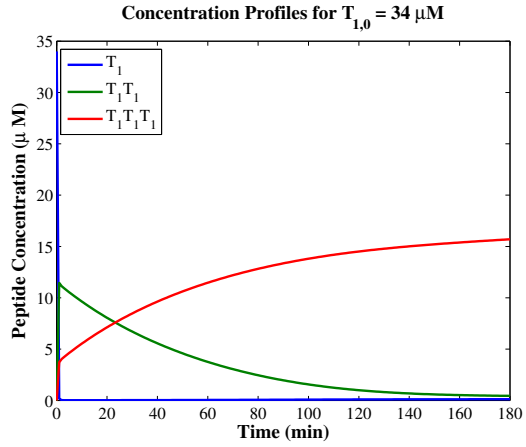
(b)



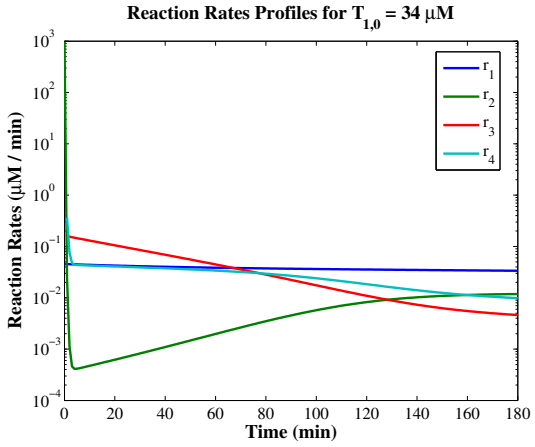
(c)



(d)



(e)



(f)

Figure S2: Concentration and reaction rate profiles obtained from the parameter estimation of the four-reaction reduced auto-catalytic mechanism of template T_1. Each figure represents the results at the different initial conditions used in the model identification.

mechanisms for template T_4 can be written as

$$E_4 + N \xrightarrow{k_1} T_4 + RSH$$
$$2\,T_4 \underset{k_3}{\overset{k_2}{\rightleftharpoons}} T_4T_4$$
$$E_4 + N + T_4T_4 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} E_4NT_4T_4$$
$$E_4NT_4T_4 \xrightarrow{k_6} T_4T_4T_4 + RSH$$
$$T_4T_4T_4 \underset{k_8}{\overset{k_7}{\rightleftharpoons}} T_4 + T_4T_4$$

and the proposed four-reaction auto-catalytic mechanism for the same template is written as:

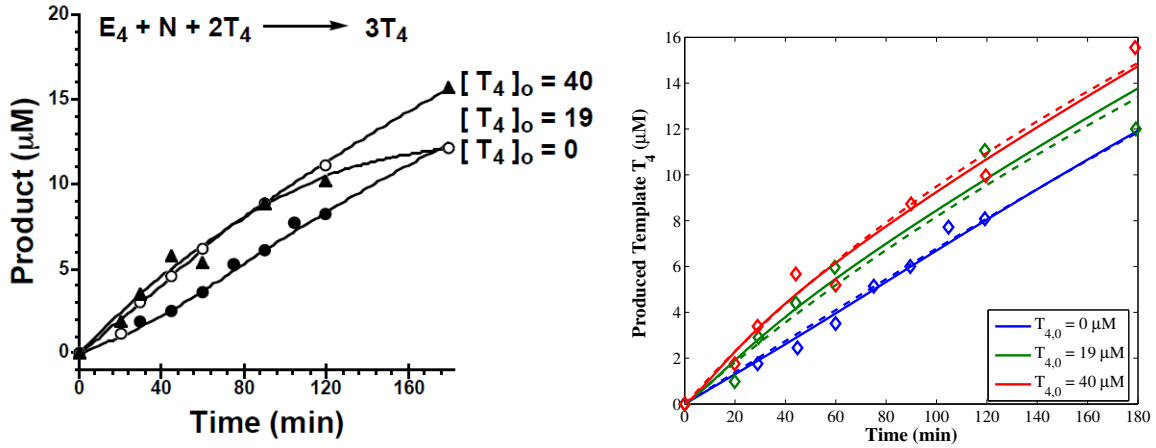
$$E_4 + N \xrightarrow{\hat{k}_1} T_4 + RSH$$
$$2\,T_4 \xrightarrow{\hat{k}_2} T_4T_4$$
$$E_4 + N + T_4T_4 \xrightarrow{\hat{k}_3} T_4T_4T_4 + RSH$$
$$T_4 + T_4T_4 \xrightarrow{\hat{k}_4} T_4T_4T_4$$

Figure S3 shows the results of the parameter estimation for both full and reduced auto-catalytic models, where the proposed auto-catalytic mechanisms are in good agreement with the experimental data. The estimated kinetic rate constants for both mechanisms are summarized in Tables S3 and S4. The values in Table S3 are not unique because the full model is not identifiable.

Reaction	**Reaction Rate Law**	**Parameters**	**Parameter Units**
r_1	$k_1\,[E_4]\,[N]$	$k_1 = 6.5564 \times 10^{-6}$	$\frac{1}{\mu M\ min}$
r_2	$k_2\,[T_4]^2$	$k_2 = 3.0919 \times 10^{-3}$	$\frac{1}{\mu M\ min}$
r_3	$k_3\,[T_4T_4]$	$k_3 = 7.6549 \times 10^{-19}$	$\frac{1}{min}$
r_4	$k_4\,[E_4]\,[N]\,[T_4T_4]$	$k_4 = 1.0771 \times 10^{-3}$	$\frac{1}{\mu M^2\ min}$
r_5	$k_5\,[E_4NT_4T_4]$	$k_5 = 1.4497 \times 10^{-2}$	$\frac{1}{min}$
r_6	$k_6\,[E_4NT_4T_4]$	$k_6 = 4.8134 \times 10^{-2}$	$\frac{1}{min}$
r_7	$k_7\,[T_4T_4T_4]$	$k_7 = 8.8541 \times 10^{-4}$	$\frac{1}{min}$
r_8	$k_8\,[T_4]\,[T_4T_4]$	$k_8 = 2.5825 \times 10^{0}$	$\frac{1}{\mu M\ min}$

Table S3: Estimated kinetic parameters for auto-catalytic mechanism of template T_4 using the full eight-reaction model. All concentrations in the reaction rate laws are measured in μM.

The analysis of the reaction rates for the reduced auto-catalytic mechanism in Figure S4 shows comparable results between the templates T_1 and T_4. As in the case of the template T_1, the formation of the duplex T_4T_4 and triplex $T_4T_4T_4$ template species is promoted in the model, and it is unlikely that these species dissociate back to single T_4 molecules. Also, single template T_4 molecules are mostly formed by the uncatalyzed reaction r_1, not



(a) (b)

Figure S3: Model fitting for auto-catalytic rates of template T_4. (a) Original experimental data points from Reference [1]. The template reactions were performed using 100 μM E_4 and 100 μM N, in the presence or absence of various initial concentrations of $T_{4,0}$ as indicated. Permission for reprint, Copyright (2004). National Academy of Sciences, U.S.A. (b) Results of the model identification procedure to estimate the values of the kinetic rate constants. The solid lines represent the parameter estimation results using the eight-reaction full auto-catalytic T_4 mechanism, and the dashed lines represent the parameter estimation results using the four-reaction reduced auto-catalytic T_4 mechanism.

the catalyzed reaction r_3, similar to the case for template T_1. The results suggest that the initial template concentrations $T_{4,0}$ used in the seeded experiments are not sufficiently high to make the catalyzed reaction rate r_3 to be at any point more active than the uncatalyzed reaction rate r_1.

S4 Cross-catalytic pathways for template-directed peptide network

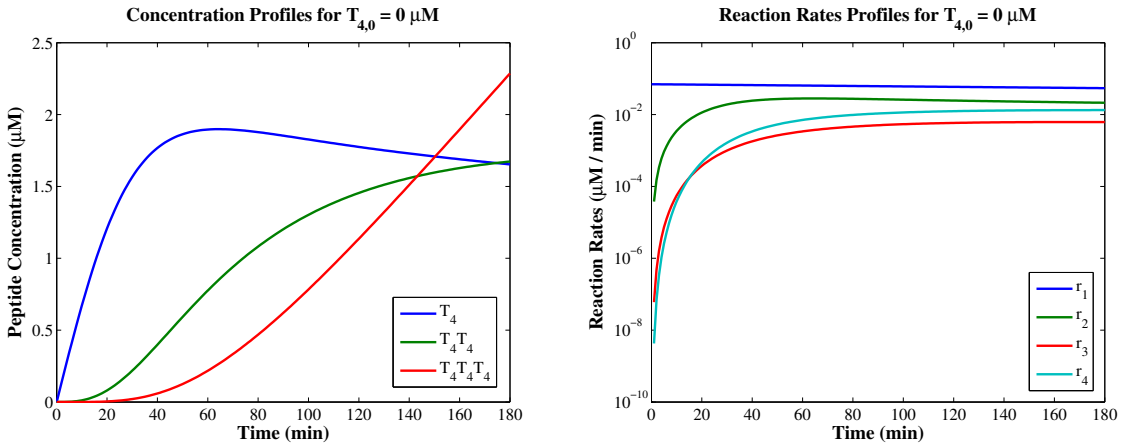
Based on the same reactions used in Sections S2 and S3 to describe the auto-catalytic mechanism, it is possible to write a cross-catalytic mechanism with 10 different kinetic parameters. The corresponding cross-catalytic reactions are:

$$E_1 + N + T_4T_4 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E_1NT_4T_4$$

$$E_1NT_4T_4 \xrightarrow{k_3} T_1T_4T_4 + RSH$$

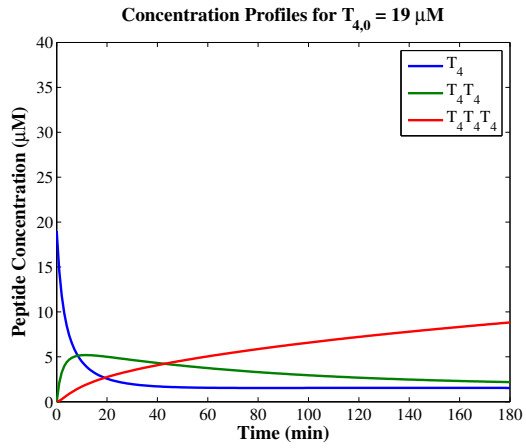
$$T_1T_4T_4 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} T_1 + T_4T_4$$

$$E_4 + N + T_1T_1 \underset{k_7}{\overset{k_6}{\rightleftharpoons}} E_4NT_1T_1$$

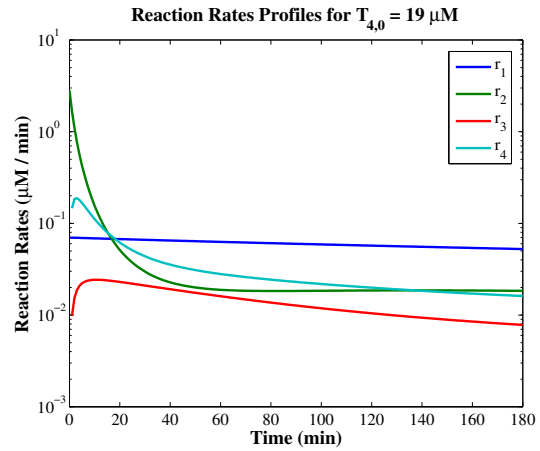


(a)

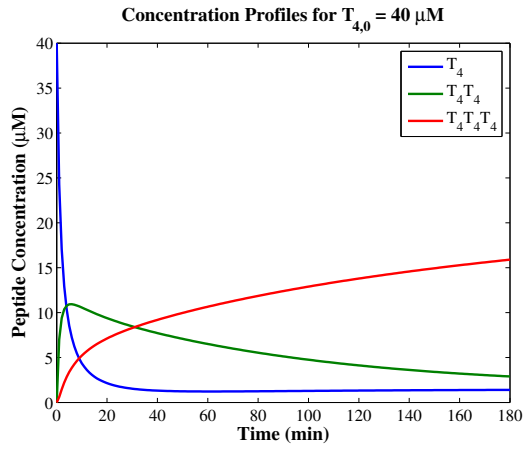
(b)



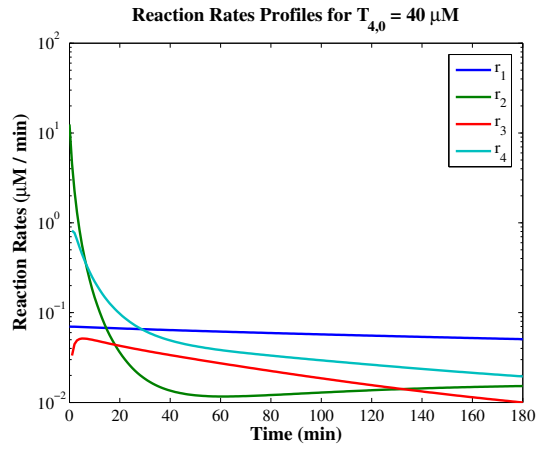
(c)



(d)



(e)



(f)

Figure S4: Concentration and reaction rate profiles obtained from the parameter estimation of the four-reaction reduced auto-catalytic mechanism of template T_4. Each figure represents the results at the different initial conditions used in the model identification.

Reaction	**Reaction Rate Law**	**Parameters**	**Parameter Units**
r_1	$\hat{k}_1 [\mathrm{E_4}] [\mathrm{N}]$	$\hat{k}_1 = 6.9912 \times 10^{-6}$	$\frac{1}{\mu\mathrm{M\ min}}$
r_2	$\hat{k}_2 [\mathrm{T_4}]^2$	$\hat{k}_2 = 7.7896 \times 10^{-3}$	$\frac{1}{\mu\mathrm{M\ min}}$
r_3	$\hat{k}_3 [\mathrm{E_4}] [\mathrm{N}] [\mathrm{T_4T_4}]$	$\hat{k}_3 = 4.7666 \times 10^{-7}$	$\frac{1}{\mu\mathrm{M^2\ min}}$
r_4	$\hat{k}_4 [\mathrm{T_4}] [\mathrm{T_4T_4}]$	$\hat{k}_4 = 4.8138 \times 10^{-3}$	$\frac{1}{\mu\mathrm{M\ min}}$

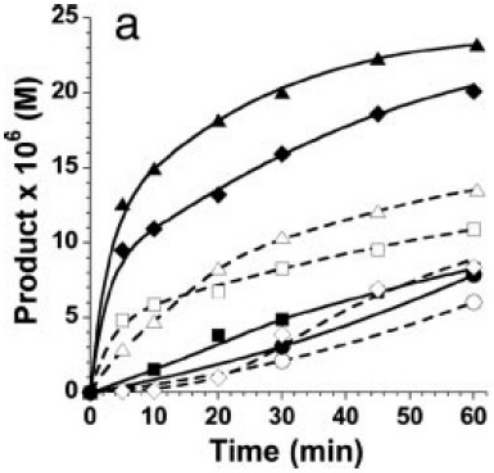
Table S4: Estimated kinetic parameters for auto-catalytic mechanism of template $\mathrm{T_4}$ using the reduced four-reaction model. All concentrations in the reaction rate laws are measured in μM.

$\mathrm{E_4NT_1T_1} \xrightarrow{k_8} \mathrm{T_4T_1T_1} + \mathrm{RSH}$
$\mathrm{T_4T_1T_1} \underset{k_{10}}{\overset{k_9}{\rightleftharpoons}} \mathrm{T_4} + \mathrm{T_1T_1}$

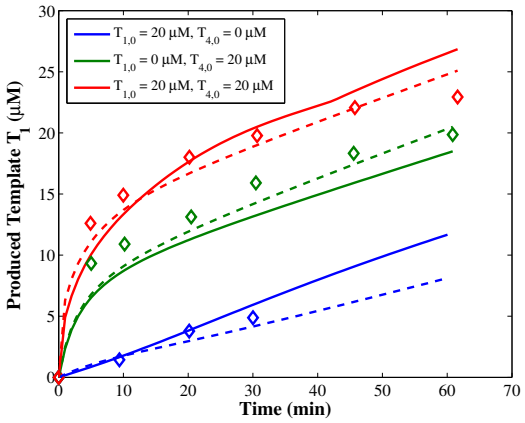
For the parameter estimation of the cross-catalytic mechanism, the auto-catalytic parameters obtained previously are held fixed during the optimization while the remaining ten parameters are estimated. The complete cross-catalytic ODE model in the parameter estimation could have either a full 26-reaction mechanism (which include the two full eight-reaction auto-catalytic mechanisms), or a reduced 18-reaction mechanism (based on the two reduced four-reaction auto-catalytic models). Figure S5 shows the results of the parameter estimation for the cross-catalytic mechanism in the peptide reaction network for both the full 26-reaction mechanism and the reduced 18-reaction mechanism. The main challenge in this parameter estimation procedure is the significant difference in the orders of magnitude between the auto-catalytic and cross-catalytic kinetic parameters, whose differences generate stiffness problems during the ODE simulations. Despite the difficulties, a reasonable agreement is obtained for both cross-catalytic models, with the advantage for the reduced 18-reaction cross-catalytic model since it has a lower computational cost (i.e. shorter simulation time). Tables S5 and S6 summarize the estimated kinetic parameters and specific reactions rate expressions for both cross-catalytic models.

Figure S6 shows the peptide concentration profiles and specific reaction rates of the reduced cross-catalytic mechanism under the initial conditions $\mathrm{E_{1,0}} = 90\ \mu$M, $\mathrm{E_{4,0}} = 90\ \mu$M, $\mathrm{N_0} = 200\ \mu$M, $\mathrm{T_{1,0}} = 20\ \mu$M, $\mathrm{T_{4,0}} = 20\ \mu$M. These initial conditions are important since the initial template concentrations $\mathrm{T_{1,0}}$ and $\mathrm{T_{4,0}}$ activate the auto- and cross-catalytic reactions in the network. Figure S6(a) shows that template $\mathrm{T_1}$ is mainly present in the form of the different triplex peptide species ($\mathrm{T_1T_1T_1}$, $\mathrm{T_1T_4T_4}$ and $\mathrm{T_4T_1T_1}$). This figure also indicates that single $\mathrm{T_1}$ molecules are rapidly incorporated into the different reaction pathways of the network, as evidenced by the low $\mathrm{T_1}$ concentration.

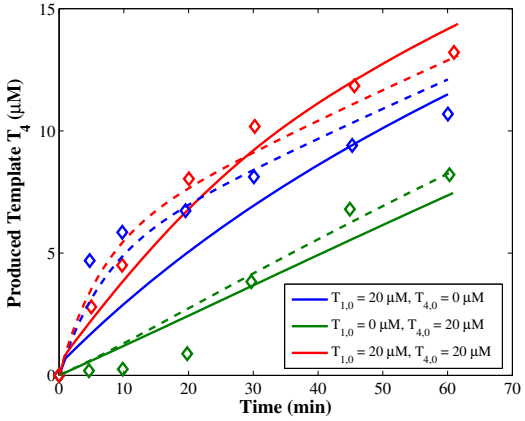
Figure S6(b) reveals a different story for the template $\mathrm{T_4}$. The main contributors to the production of this species are the triplex peptide species associated with template $\mathrm{T_1}$ ($\mathrm{T_1T_4T_4}$ and $\mathrm{T_4T_1T_1}$), and the single $\mathrm{T_4}$ molecule. The low concentration of the triplex $\mathrm{T_4T_4T_4}$ indicates that the activity of the auto-catalytic reaction is negligible, and the low concentration of the duplex $\mathrm{T_4T_4}$ suggests that this molecule is rapidly consumed in the cross-catalytic pathway for $\mathrm{T_1}$-$\mathrm{T_4T_4}$. Last, Figure S6(c) shows the reaction rates for the uncatalyzed, auto-



(a)

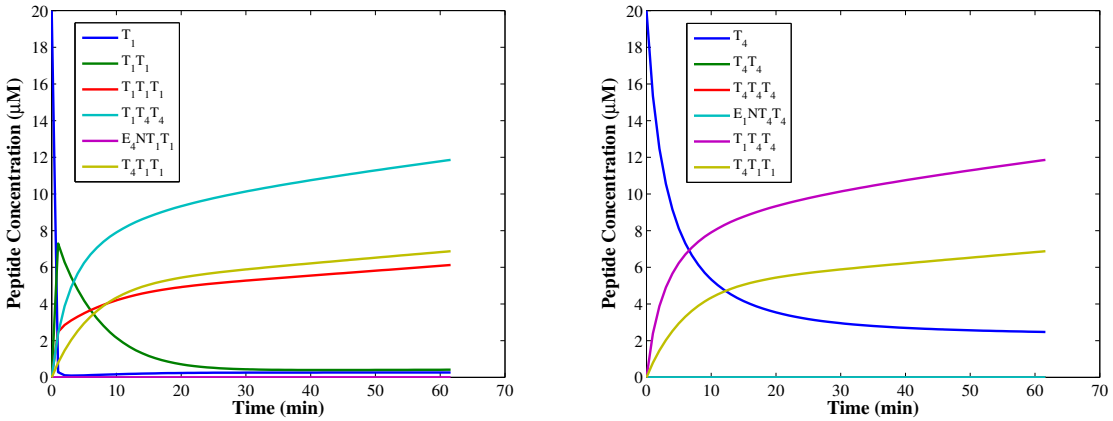


(b)



(c)

Figure S5: Model fitting results for the cross-catalytic pathways of Template T_1 and Template T_4. (a) Original experimental data points from Reference [1]. The template reactions were performed using 90 μM E_1, 90 μM E_4 and 200 μM N, in the presence or absence of various initial concentrations of $T_{1,0}$, $T_{4,0}$ as indicated in the legends of the figures. Permission for reprint, Copyright (2004). National Academy of Sciences, U.S.A. (b) and (c) Results of the model identification procedure to estimate the values of the kinetic rate constants based on the total template concentrations. The solid lines represent the parameter estimation results using the full auto-catalytic mechanisms for T_1 and T_4 (overall 26-reaction cross-catalytic mechanism), and the dashed lines represent the parameter estimation results using reduced auto-catalytic mechanisms for T_1 and T_4 (overall 18-reaction cross-catalytic mechanism).



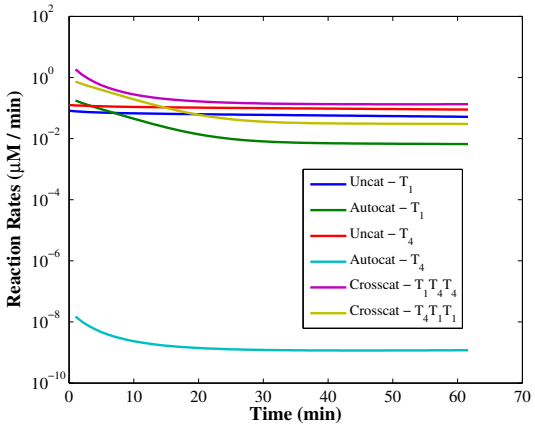


Figure S6: Concentration and reaction rate profiles obtained from the 18-reaction reduced cross-catalytic mechanism between template T_1 and template T_4. The figures correspond to the initial condition $E_{1,0} = 90$ μM, $E_{4,0} = 90$ μM, $N_0 = 200$ μM, $T_{1,0} = 20$ μM, $T_{4,0} = 20$ μM.

Reaction	Reaction Rate Law	Parameters	Parameter Units
r_1	$k_1\,[\mathrm{E_1}]\,[\mathrm{N}]\,[\mathrm{T_4T_4}]$	$k_1 = 6.7505 \times 10^8$	$\frac{1}{\mu\mathrm{M}^2\ \mathrm{min}}$
r_2	$k_2\,[\mathrm{E_1NT_4T_4}]$	$k_2 = 2.2221 \times 10^6$	$\frac{1}{\mathrm{min}}$
r_3	$k_3\,[\mathrm{E_1NT_4T_4}]$	$k_3 = 1.2466 \times 10^8$	$\frac{1}{\mathrm{min}}$
r_4	$k_4\,[\mathrm{T_1T_4T_4}]$	$k_4 = 3.1248 \times 10^{-3}$	$\frac{1}{\mathrm{min}}$
r_5	$k_5\,[\mathrm{T_1}]\,[\mathrm{T_4T_4}]$	$k_5 = 2.0586 \times 10^{-6}$	$\frac{1}{\mu\mathrm{M}\ \mathrm{min}}$
r_6	$k_6\,[\mathrm{E_4}]\,[\mathrm{N}]\,[\mathrm{T_1T_1}]$	$k_6 = 9.7300 \times 10^{-5}$	$\frac{1}{\mu\mathrm{M}^2\ \mathrm{min}}$
r_7	$k_7\,[\mathrm{E_4NT_1T_1}]$	$k_7 = 9.0271 \times 10^4$	$\frac{1}{\mathrm{min}}$
r_8	$k_8\,[\mathrm{E_4NT_1T_1}]$	$k_8 = 1.5442 \times 10^6$	$\frac{1}{\mathrm{min}}$
r_9	$k_9\,[\mathrm{T_4T_1T_1}]$	$k_9 = 2.5996 \times 10^8$	$\frac{1}{\mathrm{min}}$
r_{10}	$k_{10}\,[\mathrm{T_4}]\,[\mathrm{T_1T_1}]$	$k_{10} = 1.4517 \times 10^8$	$\frac{1}{\mu\mathrm{M}\ \mathrm{min}}$

Table S5: Estimated kinetic parameters for the cross-catalytic mechanism of the template-directed peptide network using the full 26-reaction model. All concentrations in the reaction rate laws are measured in μM.

catalytic and cross-catalytic pathways in the network, for the reduced 18-reaction model. It is important to recognize that the peptide reaction network between $\mathrm{T_1}$ and $\mathrm{T_4}$ will favor the production of the first over the second one, as is suggested by the highest reaction rate of the cross-catalytic reaction $\mathrm{T_1}$-$\mathrm{T_4T_4}$. This cross-catalytic reaction is a more favorable route to produce $\mathrm{T_1}$ than the corresponding uncatalyzed and auto-catalytic reactions. Also notice that both cross-catalytic reaction rates are higher than the corresponding auto-catalytic reaction rates, indicating a positive synergistic effect between the templates.

S5 From closed to open mass systems: the CSTR model

In a manufacturing / large-scale scenario, a continuous open process with a constant feed of reactants and removal of products is more economical compared to a batch closed process. For these purposes, this section of the supplemental material documents how the peptide network responds to constant inlet and outlet streams of material, such as in the case of an industrial continuous reactor. The studies involve the solution of a mass balance around a continuous stirred tank reactor (CSTR) [3] using the reduced 18-reaction kinetics model estimated previously. Figure S7 shows a simple scheme for the CSTR model. The inlet stream has a constant inlet flowrate F of reactants to the reactor. For the purposes of this study, the inlet stream contains the electrophilic $(\mathrm{E_1}, \mathrm{E_4})$ and the nucleophilic N species. No additional template molecules are added to the open system. The outlet stream contains all the chemical species involved in the peptide network, reactants and products, and according to the perfect mixing assumption, the output composition is identical to composition of the material inside the reactor.

Mathematically, the CSTR model is written as a system of differential equations relating the mass balances for each of the chemical species in the peptide network. The CSTR model

Reaction	**Reaction Rate Law**	**Parameters**	**Parameter Units**
r_1	$\hat{k}_1\,[E_1]\,[N]\,[T_4T_4]$	$\hat{k}_1 = 6.0502 \times 10^{1}$	$\frac{1}{\mu M^2\ \text{min}}$
r_2	$\hat{k}_2\,[E_1NT_4T_4]$	$\hat{k}_2 = 6.0358 \times 10^{4}$	$\frac{1}{\text{min}}$
r_3	$\hat{k}_3\,[E_1NT_4T_4]$	$\hat{k}_3 = 6.7330 \times 10^{6}$	$\frac{1}{\text{min}}$
r_4	$\hat{k}_4\,[T_1T_4T_4]$	$\hat{k}_4 = 7.2170 \times 10^{-3}$	$\frac{1}{\text{min}}$
r_5	$\hat{k}_5\,[T_1]\,[T_4T_4]$	$\hat{k}_5 = 1.2689 \times 10^{-5}$	$\frac{1}{\mu M\ \text{min}}$
r_6	$\hat{k}_6\,[E_4]\,[N]\,[T_1T_1]$	$\hat{k}_6 = 1.0530 \times 10^{1}$	$\frac{1}{\mu M^2\ \text{min}}$
r_7	$\hat{k}_7\,[E_4NT_1T_1]$	$\hat{k}_7 = 1.1394 \times 10^{9}$	$\frac{1}{\text{min}}$
r_8	$\hat{k}_8\,[E_4NT_1T_1]$	$\hat{k}_8 = 6.1971 \times 10^{2}$	$\frac{1}{\text{min}}$
r_9	$\hat{k}_9\,[T_4T_1T_1]$	$\hat{k}_9 = 1.0636 \times 10^{-6}$	$\frac{1}{\text{min}}$
r_{10}	$\hat{k}_{10}\,[T_4]\,[T_1T_1]$	$\hat{k}_{10} = 1.0401 \times 10^{-6}$	$\frac{1}{\mu M\ \text{min}}$

Table S6: Estimated kinetic parameters for the cross-catalytic mechanism of the template-directed peptide network using the reduced 18-reaction model. All concentrations in the reaction rate laws are measured in μM.

can be represented as

$$\frac{dx_a}{dt} = \frac{Fx_{a,in}}{V} - \frac{Fx_a}{V} + \sum_{b=1}^{B} \nu_{a,b} r_b \quad i = 1, \ldots, A \tag{7}$$

where x_a corresponds to the concentration of species a in the reduced cross-catalytic mechanism [μM], r_b is the reaction rate expression for reaction b, $\nu_{a,b}$ is the stoichiometric coefficient of species a in reaction b, $x_{a,in}$ is the concentration of species a in the inlet stream [μM], F is the constant flowrate (cm^3/min), and V is the reactor volume. Consistent with the work in Ashkenasy et al. [1], the reactor volume is $V = 85$ cm^3, and the initial concentrations inside the reactor in all the simulations are $E_{1,0} = 90$ μM, $E_{4,0} = 90$ μM, $N_0 = 200$ μM, $T_{1,0} = 20$ μM, $T_{4,0} = 20$ μM, which corresponds to one of the experiments reported in the article.
The performance of the CSTR system is evaluated using the fraction f of the template T_1 in the outlet stream:

$$[T_1]^{tot} = [T_1] + 2[T_1T_1] + 3[T_1T_1T_1] + [T_1T_4T_4] + 2[E_4NT_1T_1] + 2[T_4T_1T_1]$$
$$[T_4]^{tot} = [T_4] + 2[T_4T_4] + 3[T_4T_4T_4] + 2[E_1NT_4T_4] + 2[T_1T_4T_4] + [T_4T_1T_1]$$
$$f = \frac{[T_1]^{tot}}{[T_1]^{tot} + [T_4]^{tot}}$$

The T_1 fraction f is a system-metric to evaluate the efficiency of the CSTR process in manipulating the selectivity of the peptide network towards a desired template product.
Another variable that could be used to evaluate the CSTR performance is the production rate of template T_1. The overall production rate of template T_1 is defined by

$$P = F\ [T_1]^{tot} \tag{8}$$

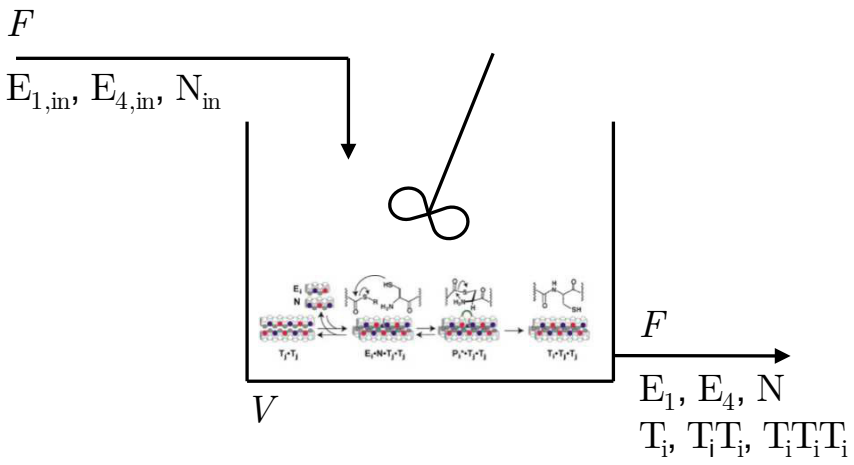


Figure S7: Continuous stirred tank reactor (CSTR) model for the auto and cross-catalytic peptide network. The model assumes a constant flowrate F for the inlet and outlet streams, in a constant reactor volume V. The figure describes the reaction mechanism of the Kent ligation in the template-directed peptide network. Permission for reprint, Copyright (2004). National Academy of Sciences, U.S.A.

where P is in units of mmol / min. Figure S8 compares the results between a closed batch system and the implemented CSTR open model using these two operational metrics. The figure shows how the implementation of the continuous flow process allows the system to reach a steady state which is different from the thermodynamic equilibrium.

S6 Optimization of selectivity and production rate of template T_1

With the definition of T_1 fraction f and the overall production rate P, the study explores the optimization of these two metrics as functions of $E_{1,in}$, $E_{4,in}$ and N_{in}, which are the inlet concentrations of the electrophiles and nucleophile, respectively. The optimization considers the operational conditions of the CSTR at steady-state. The steady state conditions of the CSTR model are obtained by setting the left-hand-side of Equation (7) equal to zero, and solving the nonlinear system of equations on the right-hand-side of Equation (7) simultaneously. In this work, the function `lsqnonlin` under the algorithm option `trust-region-reflective` in the program MATLAB [6] is used to compute the steady-state solutions of the CSTR model. In addition to the mentioned optimization procedure, the process is constrained in the inlet electrophile concentrations to account for potential solubility issues. Based on the concentrations used by Ashkenasy et al. [1] in the original paper, the optimization is constrained such that $E_{1,in} + E_{4,in} = 200$ μM and the maximum nucleophile inlet concentration N_{in} is 100 μM. Figure S9 shows the contour plots for both operational metrics, as a function of the inlet concentrations $E_{1,in}$, N_{in}, and the flowrate F in the reactor.

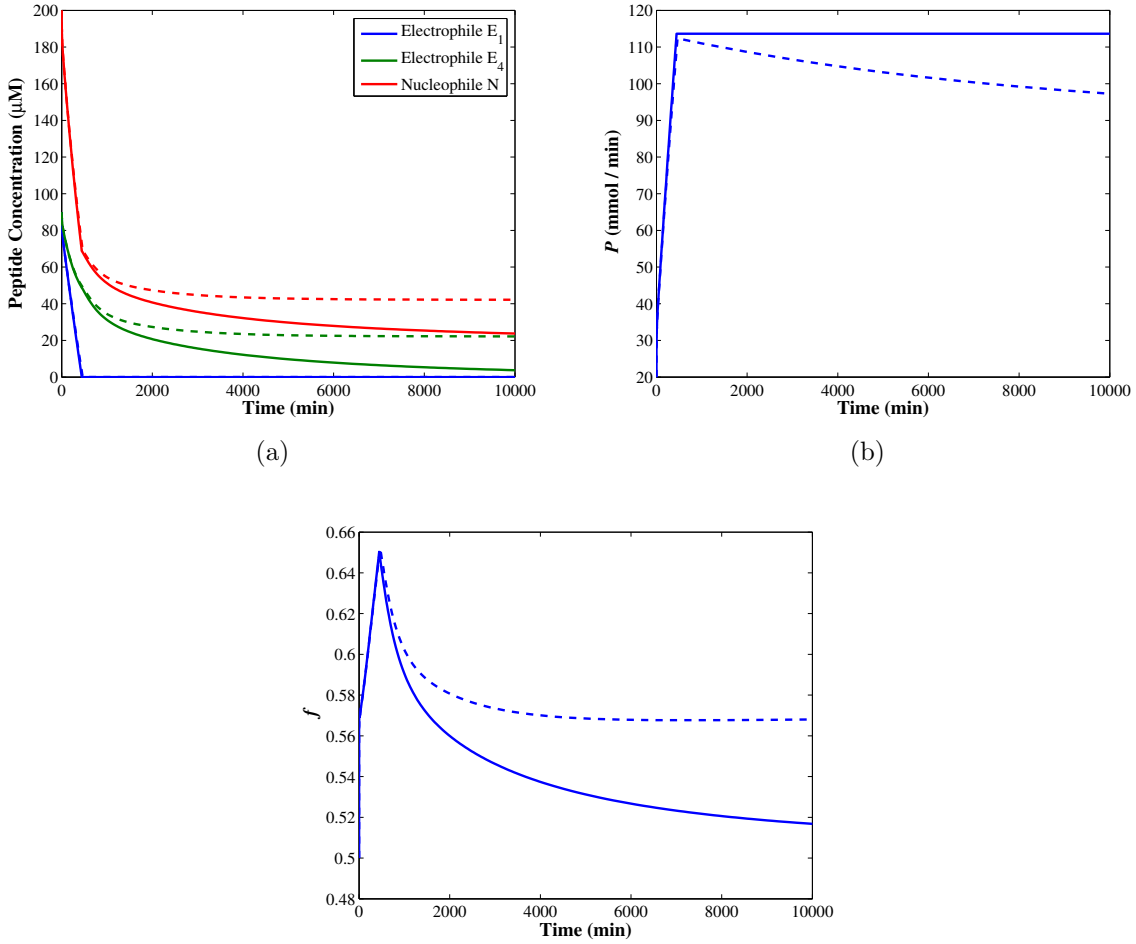


Figure S8: Comparison between closed and open systems for the template-directed peptide network between template T_1 and T_4. Initial conditions in the CSTR: $E_{1,0} = 90$ μM, $E_{4,0} = 90$ μM, $N_0 = 200$ μM, $T_{1,0} = 20$ μM, $T_{4,0} = 20$ μM. Solid lines represent the response of the closed system ($F = 0$ cm^3/min), dashed lines represent the response of the open system ($F = 0.01$ cm^3/min, $E_{1,in} = 90$ μM, $E_{4,in} = 90$ μM, $N_{in} = 200$ μM).

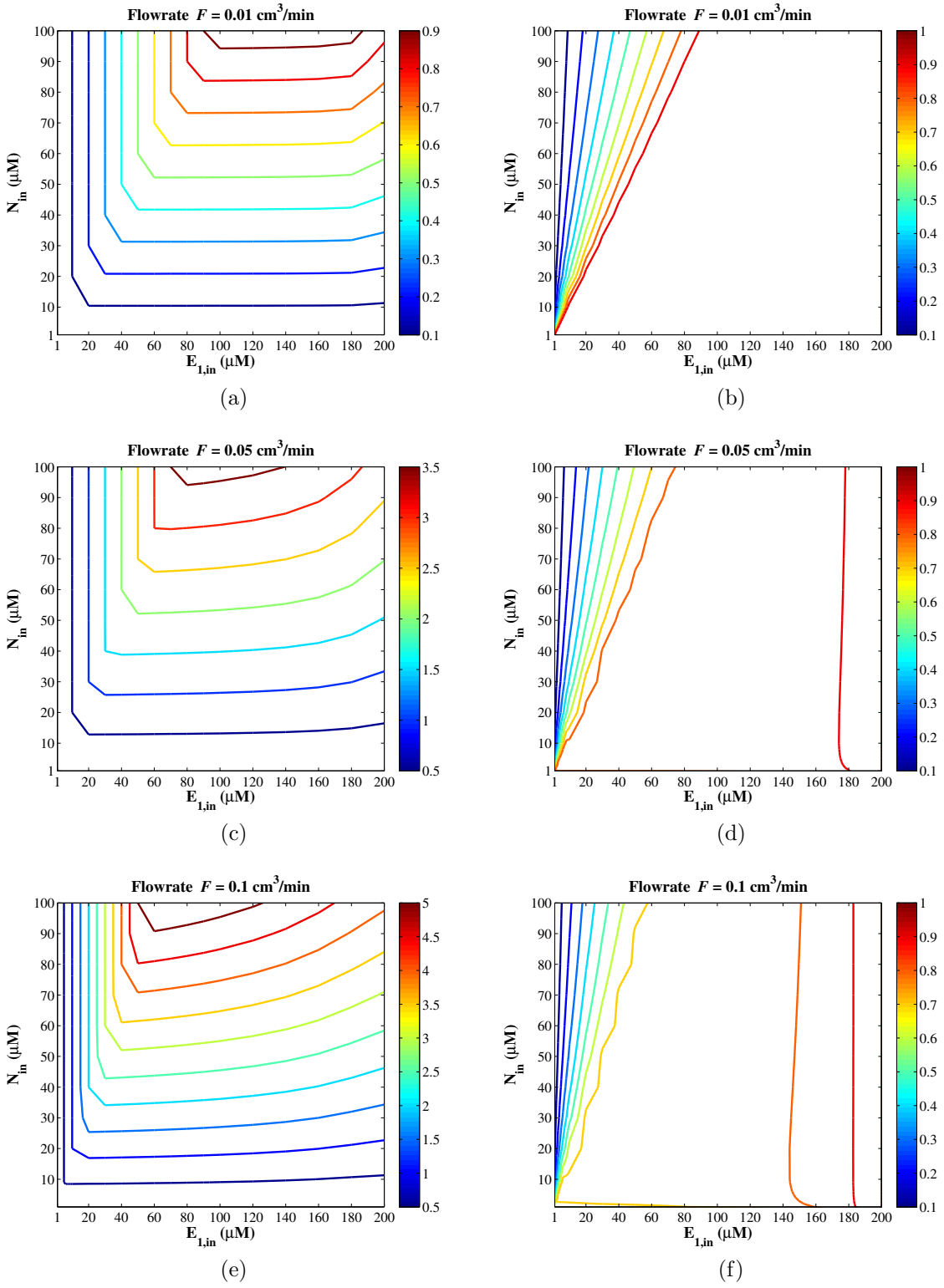


Figure S9: Effects of electrophile $E_{1,in}$, nucleophile N_{in} inlet concentrations, and flowrate F in the production of T_1 at steady-state. Figures (a), (c) and (e) indicate production rate P (mmol / min), and Figures (b), (d) and (f) indicate T_1 fraction f. Initial conditions in the reactor: $E_{1,0} = 90$ μM, $E_{4,0} = 90$ μM, $N_0 = 200$ μM, $T_{1,0} = 20$ μM, $T_{4,0} = 20$ μM.

Figures S9(a), S9(c), and S9(e) show the overall production rate of template T_1 for the CSTR model, and it is important to notice that the maximum value of the production rate occurs when some of the electrophile E_4 comes in the inlet stream. This finding indicates the important role of the cross-catalytic pathway to achieve greater production rates. Figures S9(b), S9(d) and S9(f) show that there is a trade-off between maximizing the production rate of template T_1 and maximizing the T_1 fraction f. It is evident from these figures that higher f values are obtained when the only electrophile in the inlet stream is E_1. In this case the production rate is limited by the amount of nucleophile N in the inlet stream. Once the nucleophile inlet concentration limit of 100 μM is reached, the only way to increase the production rate in the CSTR is by adding E_4 in the inlet stream at the cost of decreasing the selectivity.

References

[1] Gonen Ashkenasy, Reshma Jagasia, Maneesh Yadav, and M. Reza Ghadiri. Design of a directed molecular network. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):10872–10877, 2004.

[2] P.E. Dawson, T.W. Muir, I. Clark-Lewis, and S.B. Kent. Synthesis of proteins by native chemical ligation. *Science*, 266(5186):776–779, 1994.

[3] H. Scott Fogler. *Elements of Chemical Reaction Engineering*. Prentice Hall, 4th edition, 2005.

[4] Alan J. Kennan, V. Haridas, Kay Severin, David H. Less, and M. Reza Ghadiri. A *de novo* designed peptide ligase: A mechanistic investigation. *Journal of the American Chemical Society*, 123(9):1797–1803, 2001.

[5] Lennart Ljung. *System Identification: Theory for the User*. Prentice-Hall Inc., 2nd edition, 1999.

[6] MATLAB. *version 8.1.0.604 (R2013a)*. The MathWorks Inc., Natick, Massachusetts, 2013.

[7] H. B. Matthews, Stephen M. Miller, and James B. Rawlings. Model identification for crystallization: Theory and experimental verification. *Powder Technology*, 88:227–235, 1996.

[8] M.D. McKay, R.J. Beckman, and W.J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, 1979.

[9] Maria Rodriguez-Fernandez, Pedro Mendes, and Julio R. Banga. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *BioSystems*, 83:248–265, 2006.

[10] Maria Rodriguez-Fernandez, Markus Rehberg, Andreas Kremling, and Julio R. Banga. Simultaneous model discrimination and parameter estimation in dynamic models of cellular systems. *BMC Systems Biology*, 7:76, 2013.

[11] Kay Severin, David H. Lee, Jose A. Martinez, and M. Reza Ghadiri. Peptide self-replication via template-directed ligation. *Chemistry - A European Journal*, 3(7):1017–1024, 1997.

[12] Daniel E. Zak, Gregory E. Gonye, James S. Schwaber, and Francis J. Doyle III. Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: Insights from an identifiability analysis of an in silico network. *Genome Research*, 13:2396–2405, 2003.