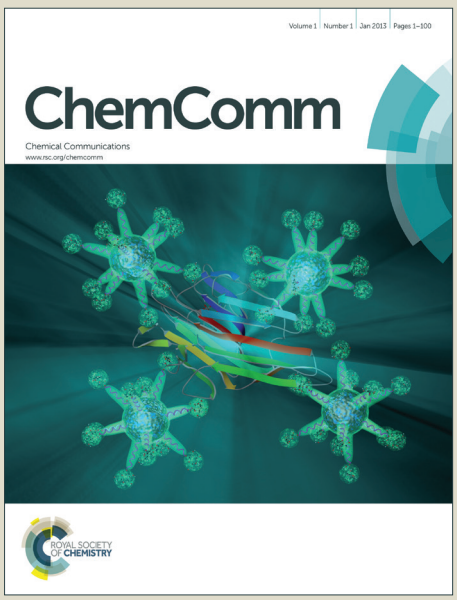
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

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

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

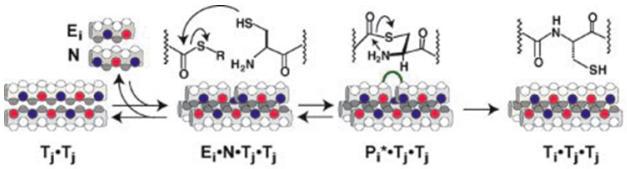
Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

Abstract: The production rate and selectivity of a cross-catalytic peptide network are optimized in a simulated continuous reaction process, under a peptide solubility constraint. The steady state of this open process is not the equilibrium state, and the optimal solution employs diverse cooperative components.

During the last 25 years, supramolecular chemistry has expanded the horizon of chemistry, using weak interactions to create functional macromolecular assemblies for molecular recognition,[1] catalysis,[2] and artificial photosynthesis.[3] However, control of self-assembly remains a hurdle to progress in the field.[4] Current ways of influencing the self-assembly include perturbation of the thermodynamic equilibrium via an engineered template,[5] creation of a kinetic trap by pH control,[6] and optimization of temperature rates and solvents to avoid undesired kinetic products.[7]

This communication presents the systematic modeling and optimization of a mass-action kinetic model for a chemical reaction network. Here, a template-directed peptide network based on the non-covalent α coiled-coil motif[9] is used to test how a chemical system developed in the systems chemistry community behaves under an open continuous process (different from closed batch processing), similar to an industrial process or a living being. This work also investigates how to drive the network to new steady states via the manipulation of the inlet stream in a continuous stirred tank reactor (CSTR).[10]

The data modeled here is from the paper by Ghadiri and coworkers[9] in which the authors utilized peptide fragments that were modified to undergo Kent ligation. The specific peptide sequences used in the Ghadiri system are shown in Figure 1. The nucleophile segment is referred to as N, while the two electrophile sequences are E_1 and E_4. The template T_1 is formed from the reaction of E_1 and N; similarly T_4 is formed from E_4 and N. The peptide sequences were designed to have an α coiled-coil folding motif that allows for non-covalent interactions between them. Thanks to this feature, T_i can be used as template-directed catalysts for the ligation reaction, since they provide an active surface for this reaction.[11] Previously, Severin et al.[12] elucidated that a duplex template T_iT_i is the corresponding catalyst structure. Given its design, the template-directed peptide network exhibits aspects of autocatalysis (T_i-T_iT_i), cross-catalysis (T_i-T_jT_j), and competition (for N). More recently, Wagner and Ashkenasy demonstrated complex system level dynamics[13] and function[14] using this peptide reaction network.



N: H_2N–CLELEVARLKKLVGE–$CONH_2$
E_1: Ar–RVARLEREVSELERKVA–COSR'
E_4: Ar–RVARLEKKVSALKKKVA–COSR'
T_1: Ar–RVARLEREVSELERKVA–CLELEVARLKKLVGE–$CONH_2$
T_4: Ar–RVARLEKKVSALKKKVA–CLELEVARLKKLVGE–$CONH_2$

Fig. 1 Graphical representation of the template-directed peptide ligation reactions that form the α coiled-coil network. The figure illustrates the mechanism for the Kent ligation process[15] to form the templates T_1 and T_4, as well as the corresponding electrophilic and nuclephilic peptide sequences incorporated in the network. Ar: 4-acetamidobenzoic acid; R': ethanesulfonic acid. Permission for reprint, Copyright (2004). National Academy of Sciences, U.S.A.

Mass-action kinetic models are used to describe replication networks;[16] here the peptide reaction network between T_1 and T_4 uses the reaction mechanism proposed by Ashkenasy et al.[9] for both auto-catalytic and cross-catalytic reaction pathways. A parameter estimation and sensitivity analysis[†,S1] is implemented to identify the parameters in the network, and to reduce the overall complexity of the network.[8] The result is an 18-reaction model for the auto- and cross-catalytic peptide network, which accurately represents the experimental data.[9] This model is the minimal model that can be identified from this particular data set, and thus may not capture all underly-

* School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, 311 Ferst Dr. NW, Atlanta GA. 30332, USA. Fax: +1 (404) 894-2866; Tel: +1 (404) 894-2878; E-mail: martha.grover@chbe.gatech.edu
† Electronic Supplementary Information (ESI) available: [Model identification of a template-directed peptide network and optimization in a continuous reactor]. See DOI: 10.1039/b000000x/

ing dynamics—with more data, a more detailed model might be uncovered. The reduced model optimized here is thus "inspired" by the α coiled-coil system. The model identification [†,S2–S4] indicates that the triplex peptide coiled-coil assemblies $T_1T_1T_1$ and $T_4T_4T_4$ serve as product inhibitors of new T_1 and T_4 molecules in the catalyzed pathways. Thus, the triplex assemblies are thermodynamic sinks, trapping the duplex assemblies and reducing the active catalyst concentration.

Figure 2 shows the reaction rates for the uncatalyzed, auto-catalytic and cross-catalytic pathways in the network. The peptide reaction network favors the production of T_1 over T_4, as suggested by the highest reaction rate being the cross-catalytic reaction T_1-T_4T_4. This cross-catalytic reaction is a more favorable route to produce T_1 than the uncatalyzed and auto-catalytic reactions for T_1. Also notice that both cross-catalytic reaction rates are higher than the corresponding auto-catalytic reaction rates, indicating a positive synergistic effect.

Using this reduced model, a system with constant inlet

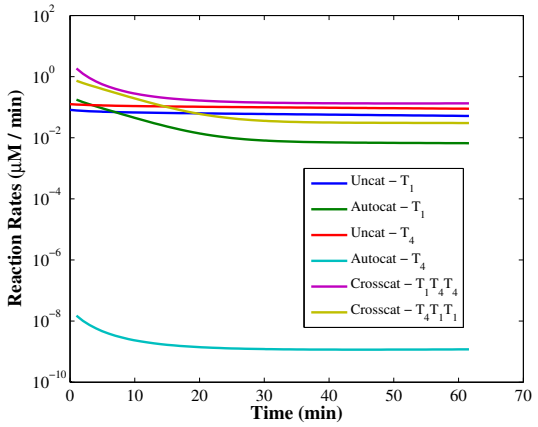


Fig. 2 Reaction rate profiles obtained from the parameter estimation of the 18-reaction reduced cross-catalytic mechanism between template T_1 and template T_4. The figure corresponds to the initial condition $E_{1,0}$ = 90 μM, $E_{4,0}$ = 90 μM, N_0 = 200 μM, $T_{1,0}$ = 20 μM, $T_{4,0}$ = 20 μM.

and outlet streams is simulated as a CSTR process.[†,S5] The inlet stream contains only electrophile (E_1, E_4) and nucleophile (N), but no template (T_1, T_4). According to the perfect mixing assumption, the output composition is identical to the composition inside the reactor.

The performance of the reaction network in the CSTR model is measured by the fraction of template T_1 relative to the total template concentration:

$$[T_1]^{tot}(\mu M) = [T_1] + 2[T_1T_1] + 3[T_1T_1T_1] + [T_1T_4T_4] + 2[E_4NT_1T_1] + 2[T_4T_1T_1]$$

$$[T_4]^{tot}(\mu M) = [T_4] + 2[T_4T_4] + 3[T_4T_4T_4] + [T_4T_1T_1] + 2[E_1NT_4T_4] + 2[T_1T_4T_4]$$

$$f = \frac{[T_1]^{tot}}{[T_1]^{tot} + [T_4]^{tot}} \quad (1)$$

and the overall production rate of T_1:

$$P\left(\frac{\text{mmol}}{\text{min}}\right) = F\,[T_1]^{tot} \quad (2)$$

where F is the inlet flowrate to the reactor $\left(\text{cm}^3/\text{min}\right)$.

The optimization seeks to maximize both the production P and T_1 fraction f as functions of $E_{1,in}$, $E_{4,in}$ and N_{in}, which are the inlet concentrations of the respective electrophilic and nucleophilic peptide sequences, at the steady state of the CSTR. While high production is the primary objective, high selectivity to T_1 is also important since separation of the catalyst T_4 after the reaction adds additional costs to the process. More broadly, selectivity is a desired feature in a DCN. The inlet electrophile concentrations are constrained in the optimization to account for solubility limitations. Based on experimental values,[9] the following constraints are applied: $E_{1,in}$ + $E_{4,in}$ = 200 μM and $N_{in} \leq$ 100 μM.

Figure 3 shows how the production rate P can be controlled by the inlet concentrations of the electrophilic and nucleophilic peptide species. The figure indicates that the maximum value of the production rate P is reached when some of the electrophile E_4 enters in the inlet stream. This finding indicates that the central role of the cross-catalytic pathway is to maximize production rates under constraints. The study also highlights a trade-off between maximizing the production rate P and the T_1 fraction f. The highest values of f are obtained when the only electrophile in the inlet stream is E_1. In those cases, the production rate P 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 further increase the production rate P in the CSTR is by adding E_4 to the inlet stream, at the cost of decreasing the T_1 fraction of the product. A similar behavior is observed in the cases of F = 0.05 cm^3/min and F = 0.1 cm^3/min.[†,S6] Although T_1 is also serving as a catalyst for T_4 replication, which competes for N, there is still a net advantage to adding T_4 for the production of T_1.

Conflicting operational objective functions are often found in industrial scenarios. Pareto front plots are frequently used to visualize the trade-offs in multi-objective optimization problems like the one presented in this study. Figure 4 shows the constructed Pareto fronts between the production rate P and the T_1 fraction f at different flowrates F. The Pareto front is constructed by converting the multi-objective optimization problem into an ε-constrained single-objective problem where the second objective function is imposed as a constraint in the optimization.[17] In the case of the peptide network, the T_1 fraction f is maximized, constrained by the production rate P.

Figure 4 shows the improvement in the production rate of

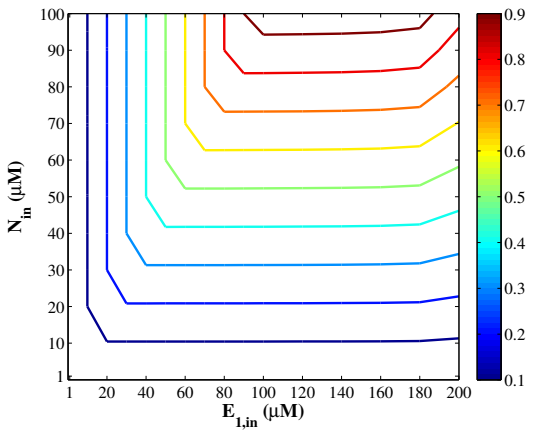


Fig. 3 Contour plot of the production rate P (mmol/min) at steady state as a function of the electrophile $E_{1,in}$ and nucleophile N_{in} inlet concentrations. The flow rate is $F = 0.01$ cm^3/min. The results in the figure satisfy the peptide solubility constraint $E_{1,in} + E_{4,in} = 200$ μM. Initial 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.

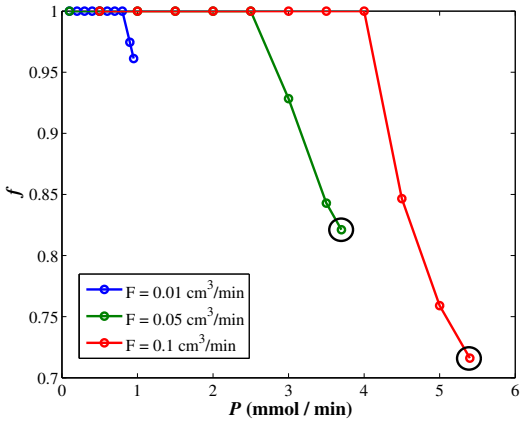


Fig. 4 Pareto fronts for the overall T_1 fraction f and the production rate P at steady state. The figure shows the trade-off between these two objective functions at different flow rates F in the CSTR.

the template T_1 due to the cross-catalytic pathway T_1-T_4T_4. For example, in the case when $F = 0.1$ cm^3/min, the production rate of T_1 increases from 4 mmol/min to 5.4 mmol/min, an overall improvement of 35% in the production. The optimization results at the points denoted by the black circles are $E_{1,in} = 84.6$ μM, $E_{4,in} = 115.4$ μM and $N_{in} = 100$ μM for $F = 0.05$ cm^3/min, and $E_{1,in} = 71.5$ μM, $E_{4,in} = 128.5$ μM and $N_{in} = 100$ μM for $F = 0.01$ cm^3/min. It is then counter-intuitive that to maximize T_1 production, the optimal operating condition requires an inlet stream richer in E_4 than E_1.

This study exploits the kinetics of the peptide network using a sustained driving force[18] (i.e. inlet and outlet streams) to reach a non-equilibrium steady state population. Rather than driving the system down a metastable pathway, this work shows how a continuous flow can hold the system at a nonequilibrium state. A continuous flow process does not exclude the use of other driving forces (like pH, temperature or engineered molecular templates), to be used simultaneously in the control of self-assembled systems. Finally, with the formulation of a kinetic model, it is possible to expand the optimization problem to include additional system-level metrics like the overall yield of the network, or other design variables such as the flow rate.

Acknowledgments

This work was jointly supported by a McDonnell Foundation 21st Century Science Initiative Grant on Studying Complex Systems No. 220020271, and a NSF and the NASA Astrobiology Program, under the NSF/NASA Center for Chemical Evolution, CHE-1004570.

References

1 J. Lewis, E. Gavey, S. Cameron and J. Crowley, *Chem. Sci.*, 2012, **3**, 778–784.
2 M. Yoshizawa, M. Tamura and M. Fujita, *Science*, 2006, **312**, 251–254.
3 R. Haycock, A. Yartsev, U. Michelsen, V. Sundstrom and C. Hunter, *Angew. Chem. Int. Ed.*, 2000, **39**, 3616–3619.
4 M. D. Ward and P. R. Raithby, *Chem. Soc. Rev.*, 2013, **42**, 1619–1636.
5 T. S. R. Lam, A. Belenguer, S. L. Roberts, C. Naumann, T. Jarrosson, S. Otto and J. K. M. Sanders, *Science*, 2005, **308**, 667–669.
6 B. Brisig, J. K. M. Sanders and S. Otto, *Angew. Chem. Int. Ed.*, 2003, **42**, 1270–1273.
7 P. A. Korevaar, C. Grenier, A. J. Markvoort, A. P. H. J. Schenning, T. F. A. de Greef and E. W. Meijer, *Proc. Nat. Acad. Sci. USA*, 2013, **110**, 17205–17210.
8 M. Rodriguez-Fernandez, P. Mendes and J. R. Banga, *BioSystems*, 2006, **83**, 248–265.
9 G. Ashkenasy, R. Jagasia, M. Yadav and M. R. Ghadiri, *Proc. Nat. Acad. Sci. USA*, 2004, **101**, 10872–10877.
10 H. S. Fogler, *Elements of Chemical Reaction Engineering*, Prentice Hall, 4th edn, 2005.
11 A. J. Kennan, V. Haridas, K. Severin, D. H. Less and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2001, **123**, 1797–1803.
12 K. Severin, D. H. Lee, J. A. Martinez and M. R. Ghadiri, *Chem. Eur. J.*, 1997, **3**, 1017–1024.
13 N. Wagner and G. Ashkenasy, *Journal of Chemical Physics*, 2009, **130**, 164907.
14 N. Wagner and G. Ashkenasy, *Chemistry: A European Journal*, 2009, **15**, 1765–1775.
15 P. Dawson, T. Muir, I. Clark-Lewis and S. Kent, *Science*, 1994, **266**, 776–779.
16 T. Achilles and G. von Kiedrowski, *Angewandte Chemie, Int. Ed. Engl.*, 1993, **32**, 1198–1201.
17 Y. Kawajiri and L. Biegler, *Ind. Eng. Chem. Res.*, 2006, **45**, 8503–8513.
18 J. S. Moore and M. L. Kraft, *Science*, 2008, **320**, 620–621.