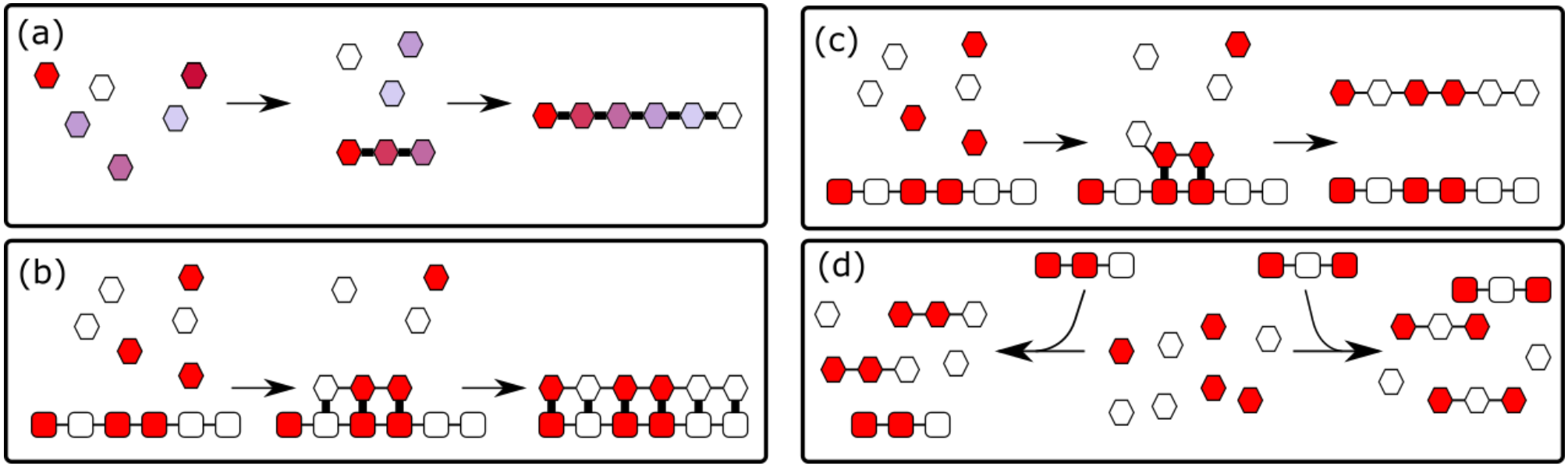
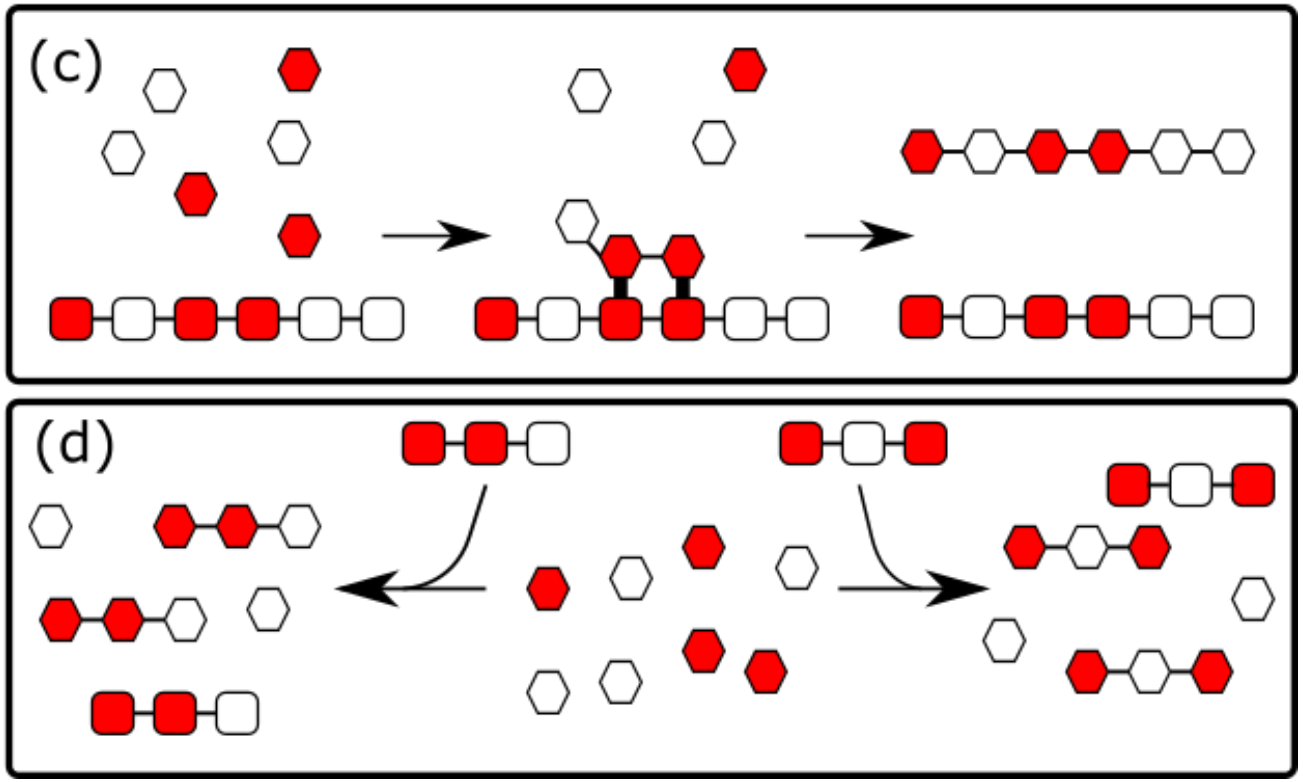
Non-equilibrium thermodynamics of catalytic information processing

Thomas Ouldridge
Imperial College London

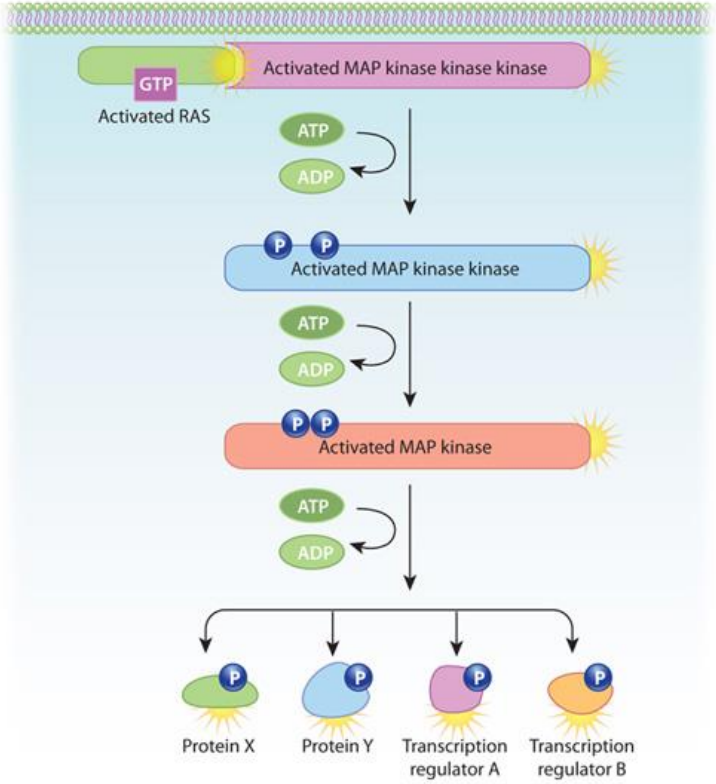


Mathematical Models in Biology: from Information Theory to Thermodynamics
July 2020



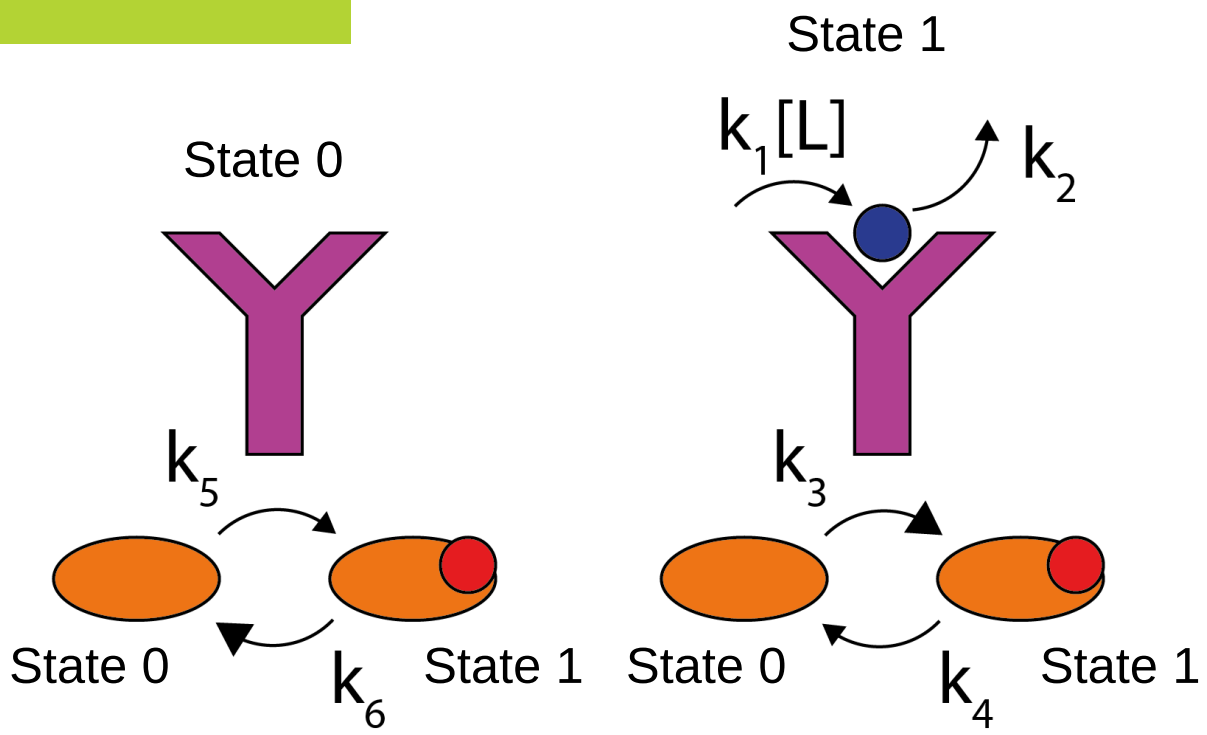
- What do I mean by catalytic information processing, and why is it important?
- Why are catalytic systems so much harder to design and build?
- How are we rising to that challenge?

What is catalytic information processing?



As introduced by Udo, catalytic readouts of cell surface receptors are a canonical example.

What is catalytic information processing?



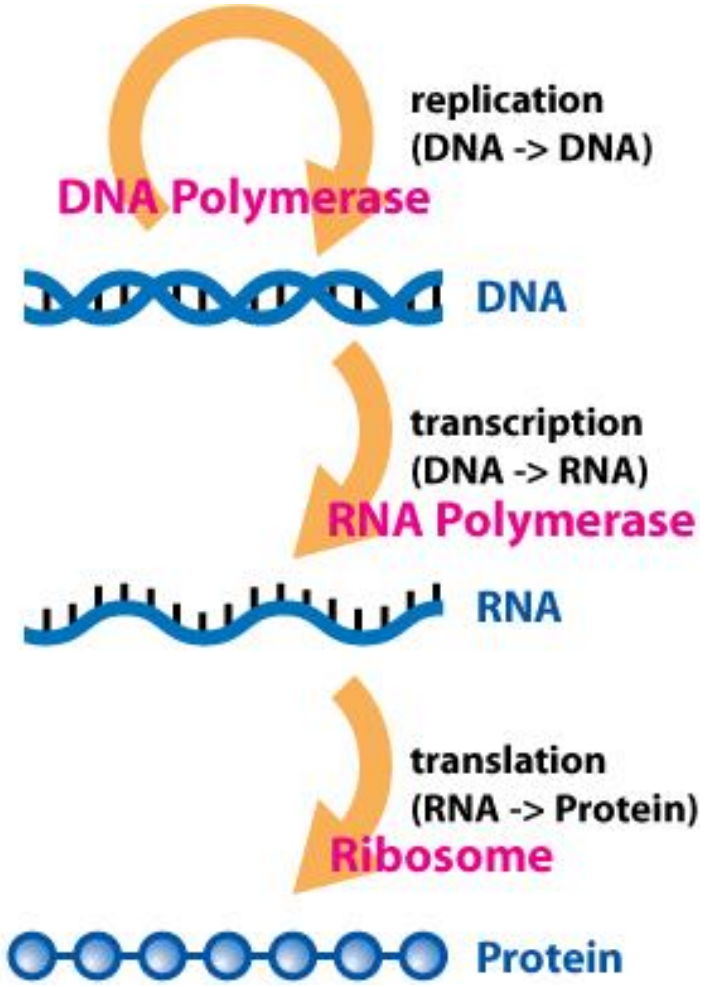
As introduced by Udo, catalytic readouts of cell surface receptors are a canonical example.

The state of the receptor is effectively “copied” into that of the readout molecule, without changing the receptor.

What is catalytic information processing?

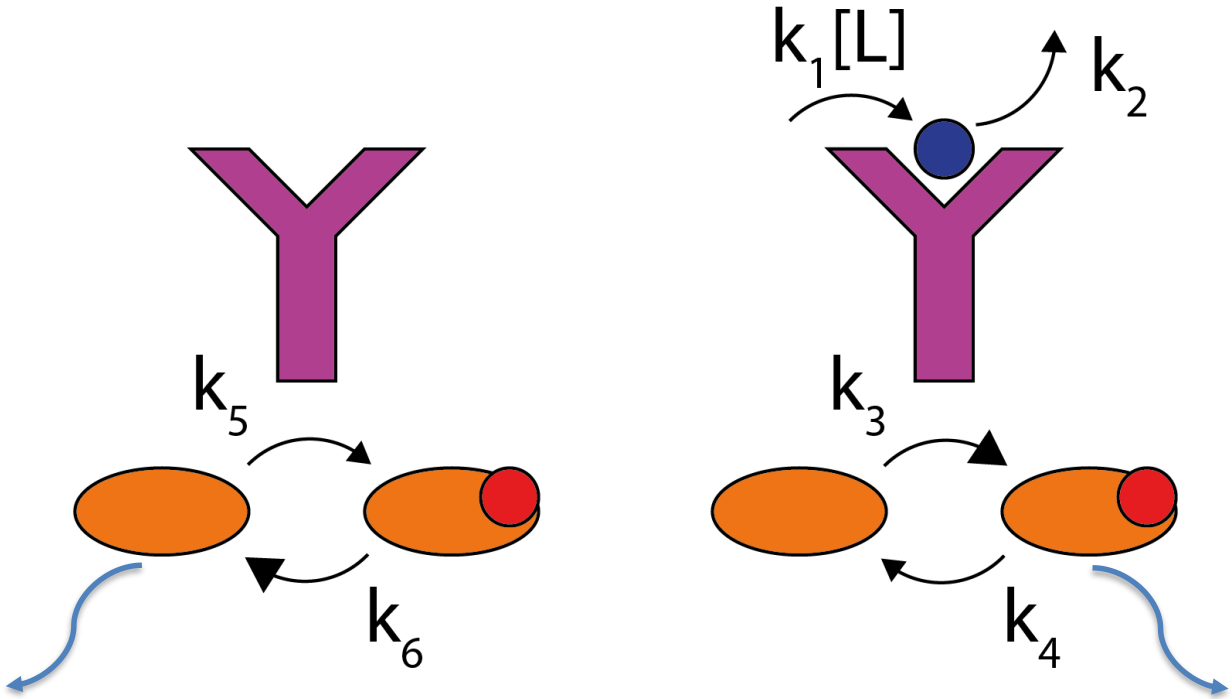
The polymer-copying reactions of the central dogma also qualify.

The templates themselves act as catalysts for the production of new polymers.



Why do these systems need to be catalytic?

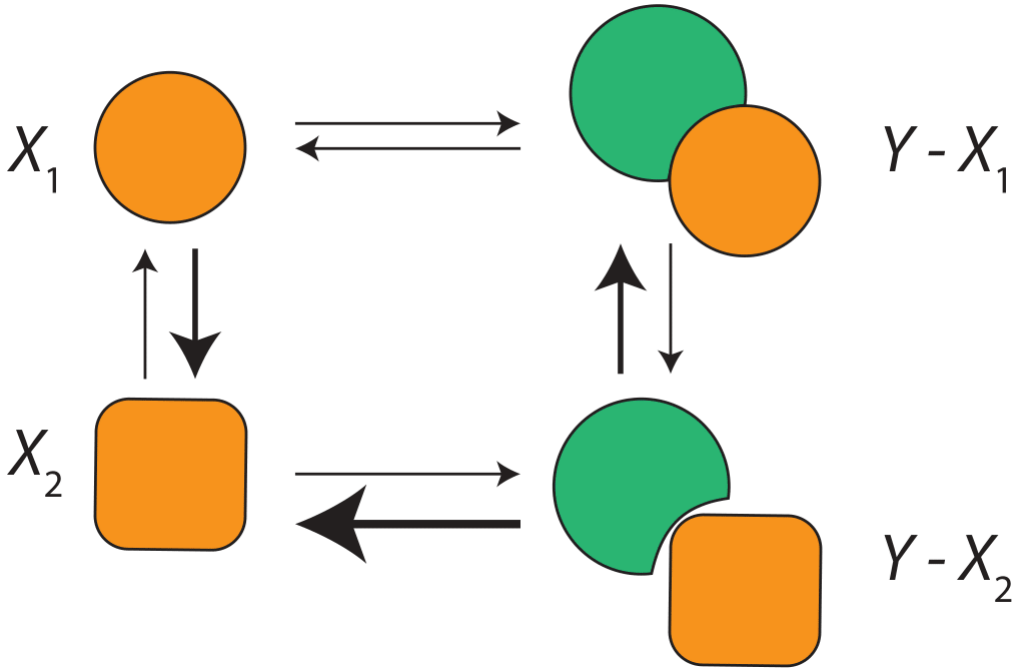
The influence of the catalyst on the product ***persists*** for a long period of time after the two have physically decoupled.



Why do these systems need to be catalytic?

The influence of the catalyst on the product ***persists*** for a long period of time after the two have physically decoupled.

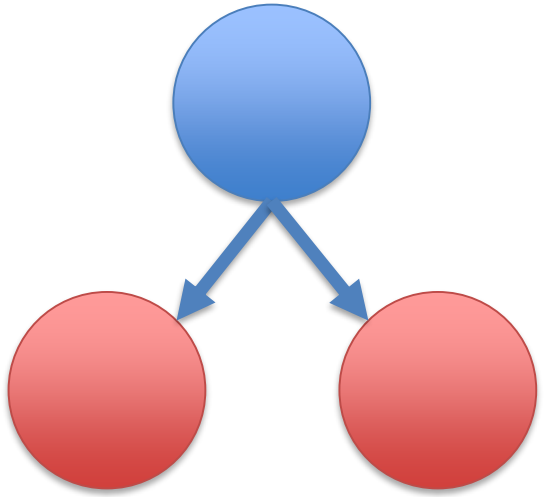
This is fundamentally ***not*** true for non-catalytic mechanisms



Why do these systems need to be catalytic in signal transduction?

In the case of catalytic activation, this persistence of effects allows:

- Signal amplification.

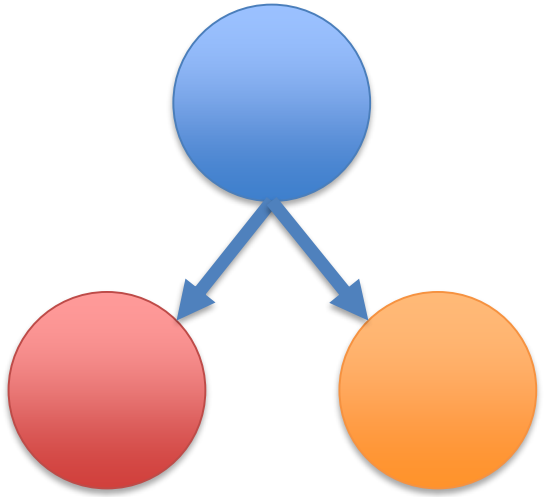


[1] P Mehta, AH Lang, DJ Schwab, J. Stat. Phys., 16:1153–1166, 2016.
[2] CC Govern, PR ten Wolde. PRL, 113:258102, December 2014.
[3] JP Barton and ED Sontag. Biophys. J., 104:2013.
[4]A Deshpande, TEO, Eng. Biol., 1:86–99, 2017.

Why do these systems need to be catalytic in signal transduction?

In the case of catalytic activation, this persistence of effects allows:

- Signal amplification.
- Signal splitting.

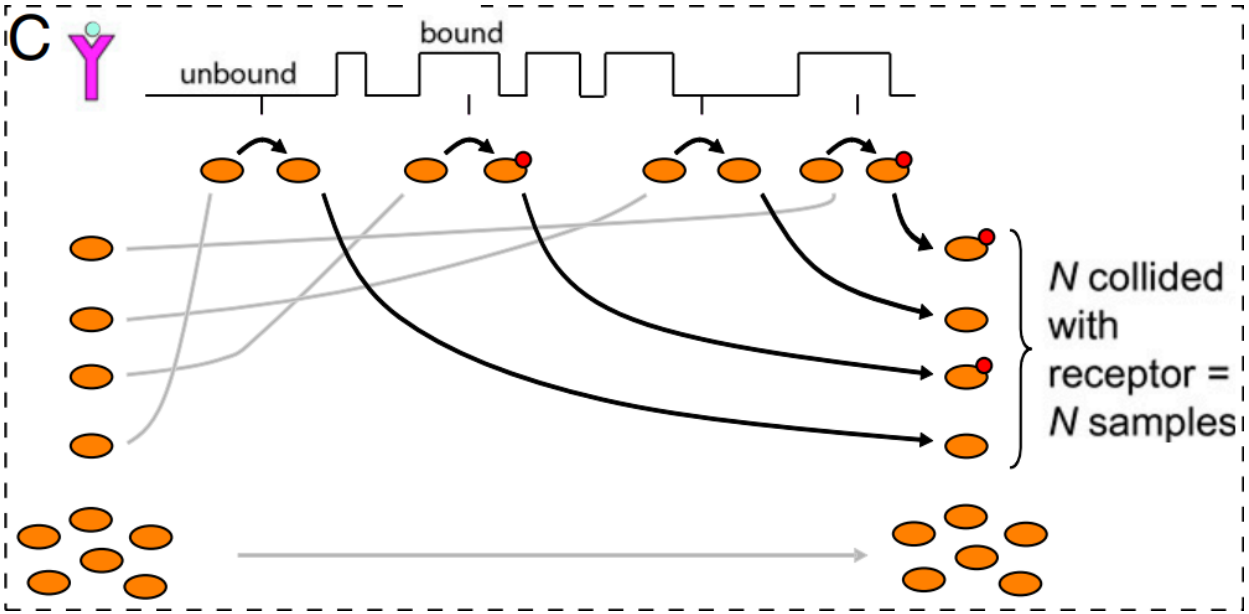


[1] P Mehta, AH Lang, DJ Schwab, J. Stat. Phys., 16:1153–1166, 2016.
[2] CC Govern, PR ten Wolde. PRL, 113:258102, December 2014.
[3] JP Barton and ED Sontag. Biophys. J., 104:2013.
[4]A Deshpande, TEO, Eng. Biol., 1:86–99, 2017.

Why do these systems need to be catalytic in signal transduction?

In the case of catalytic activation, this persistence of effects allows:

- Signal amplification.
- Signal splitting.
- Time integration.

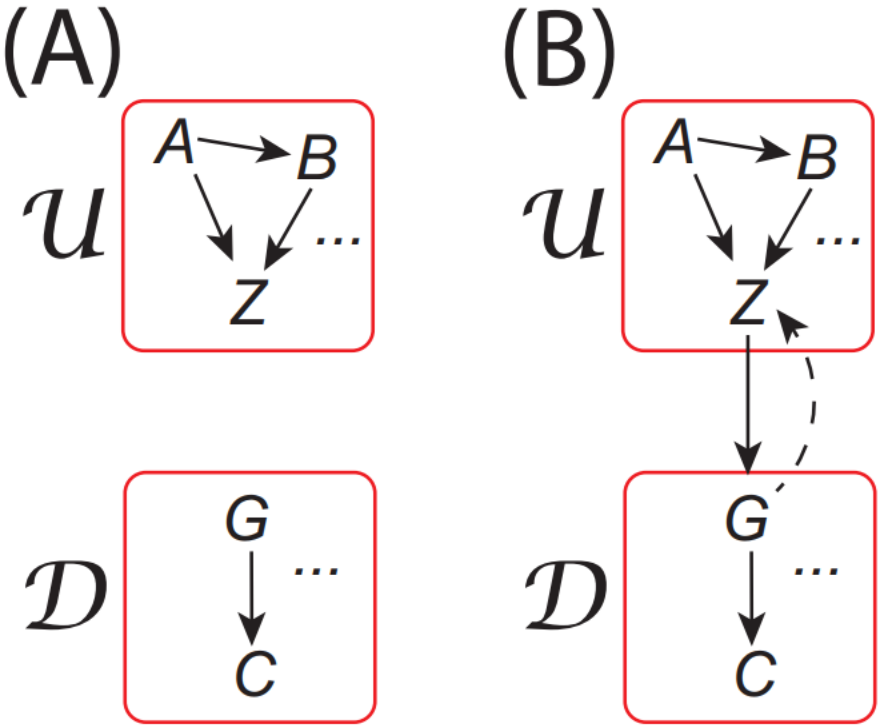


[1] P Mehta, AH Lang, DJ Schwab, J. Stat. Phys., 16:1153–1166, 2016.
[2] CC Govern, PR ten Wolde. PRL, 113:258102, December 2014.
[3] JP Barton and ED Sontag. Biophys. J., 104:2013.
[4]A Deshpande, TEO, Eng. Biol., 1:86–99, 2017.

Why do these systems need to be catalytic in signal transduction?

In the case of catalytic activation, this persistence of effects allows:

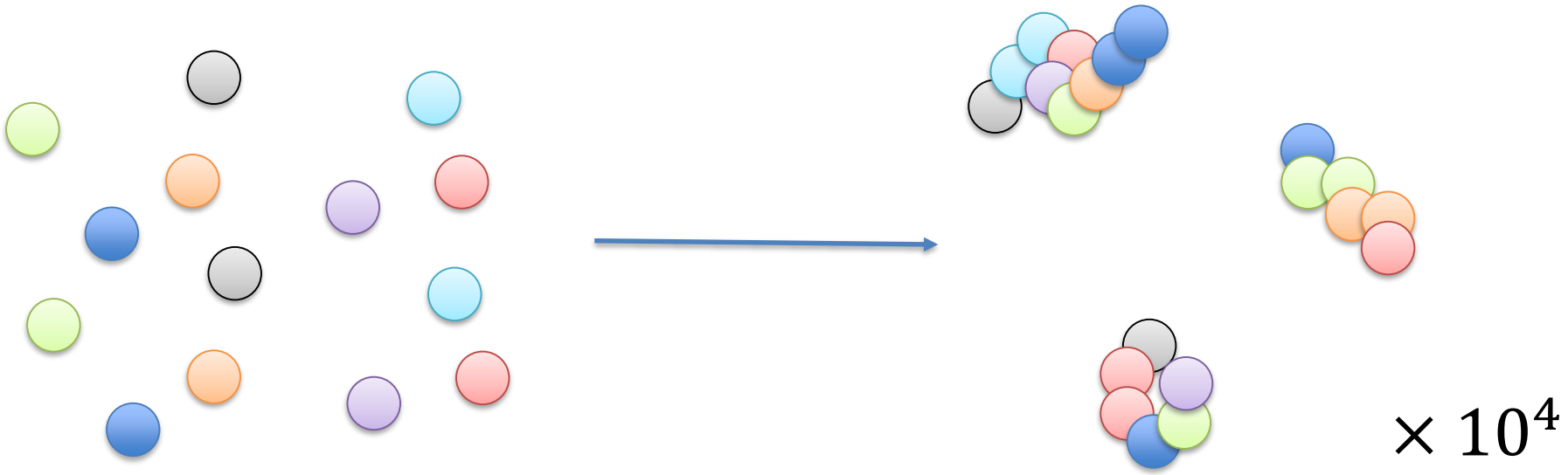
- Signal amplification.
- Signal splitting.
- Time integration.
- Modularity (suppression of retroactivty).



[1] P Mehta, AH Lang, DJ Schwab, J. Stat. Phys., 16:1153–1166, 2016.
[2] CC Govern, PR ten Wolde. PRL, 113:258102, December 2014.
[3] JP Barton and ED Sontag. Biophys. J., 104:2013.
[4]A Deshpande, TEO, Eng. Biol., 1:86–99, 2017.

Why do these systems need to be catalytic when templating polymer assembly?

Catalytic copying of polymer templates is necessary for the chemical complexity of life.



Why do these systems need to be catalytic when templating polymer assembly?

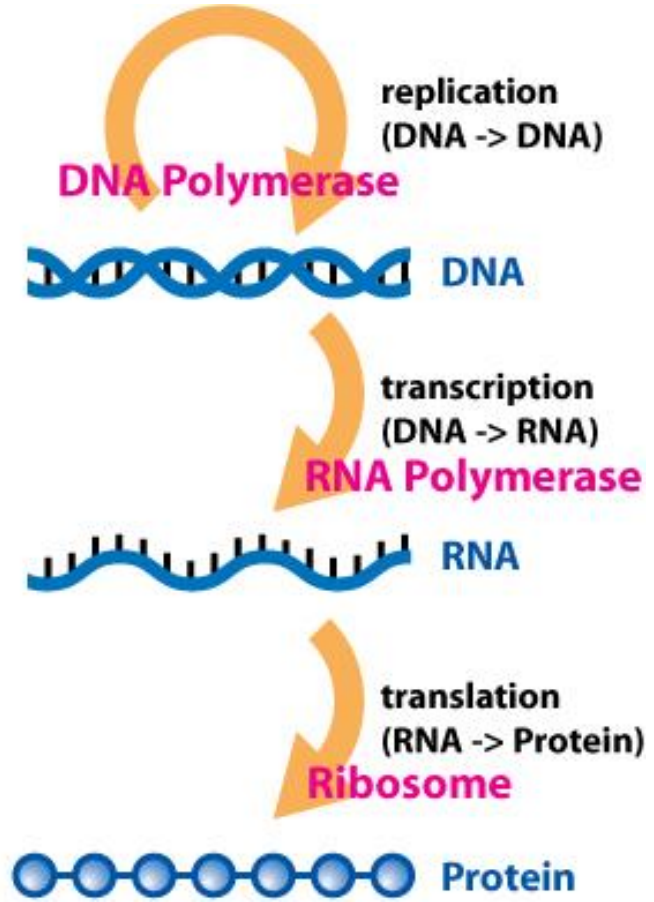
Amino acid interactions themselves do not have enough information to drive the formation of functional proteins with the desired yields.

[1] P Sartori, S Leibler, PNAS 117:114-120, 2020.

Why do these systems need to be catalytic when templating polymer assembly?

Nature gets round this problem by templating. But the templates *must* act catalytically, and be recovered.

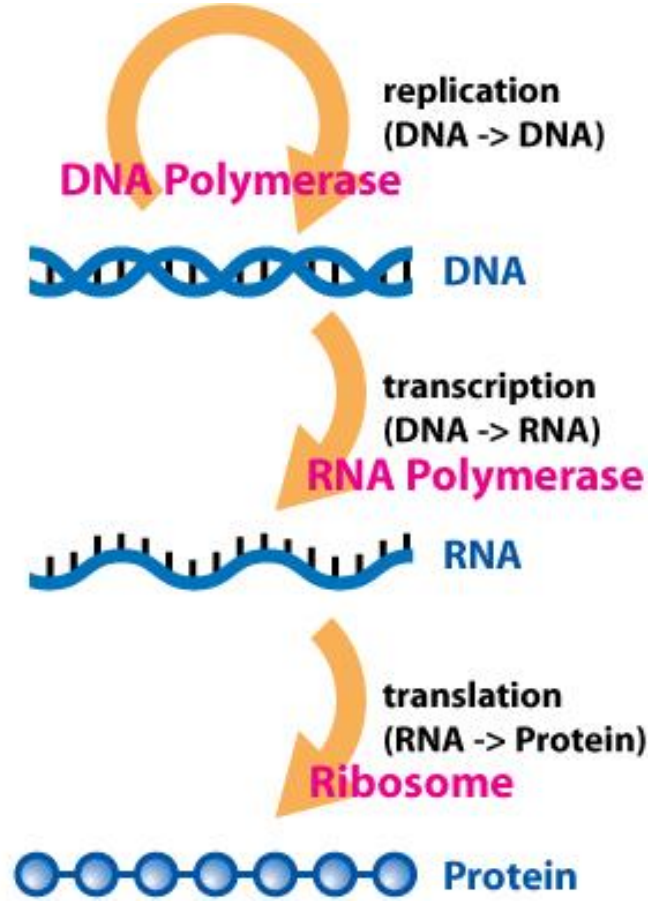
- Products must separate to be functional.

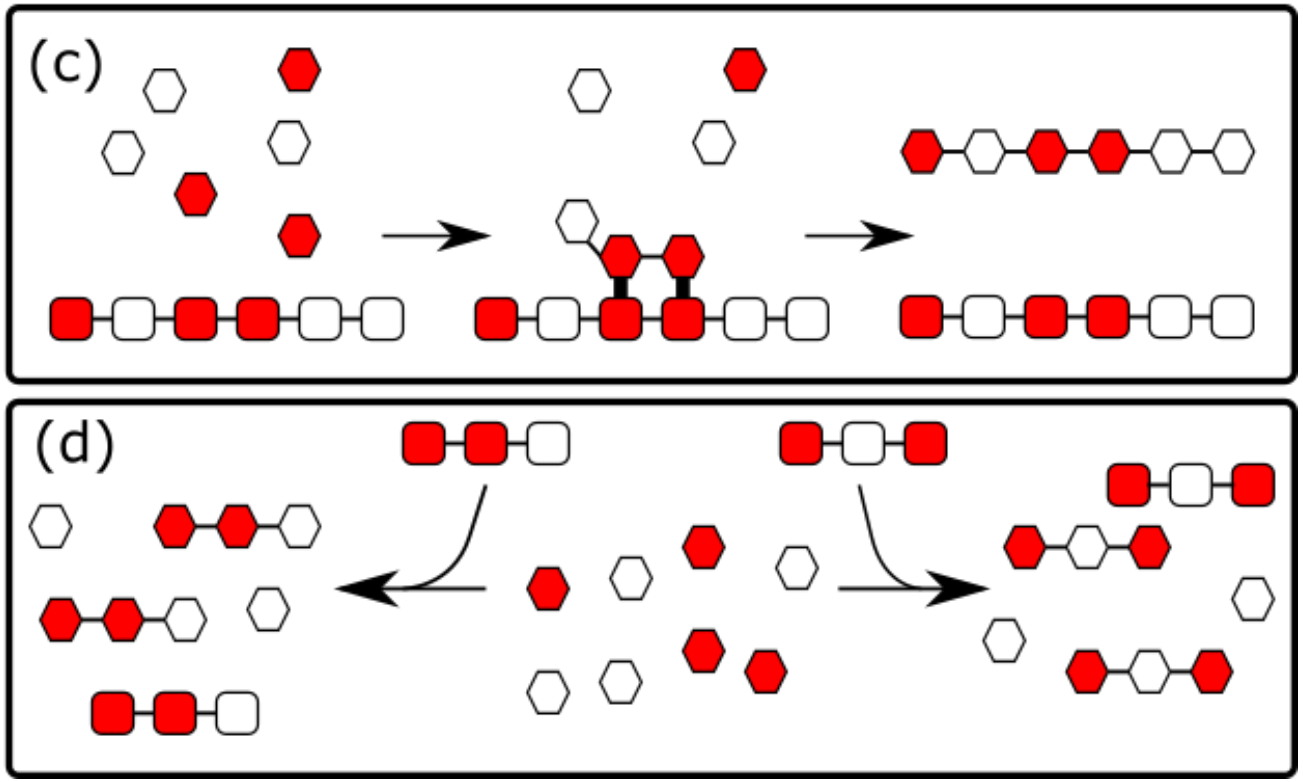


Why do these systems need to be catalytic when templating polymer assembly?

Nature gets round this problem by templating. But the templates *must* act catalytically, and be recovered.

- Products must separate to be functional.
- If you can't re-use the template, you haven't solved the problem.

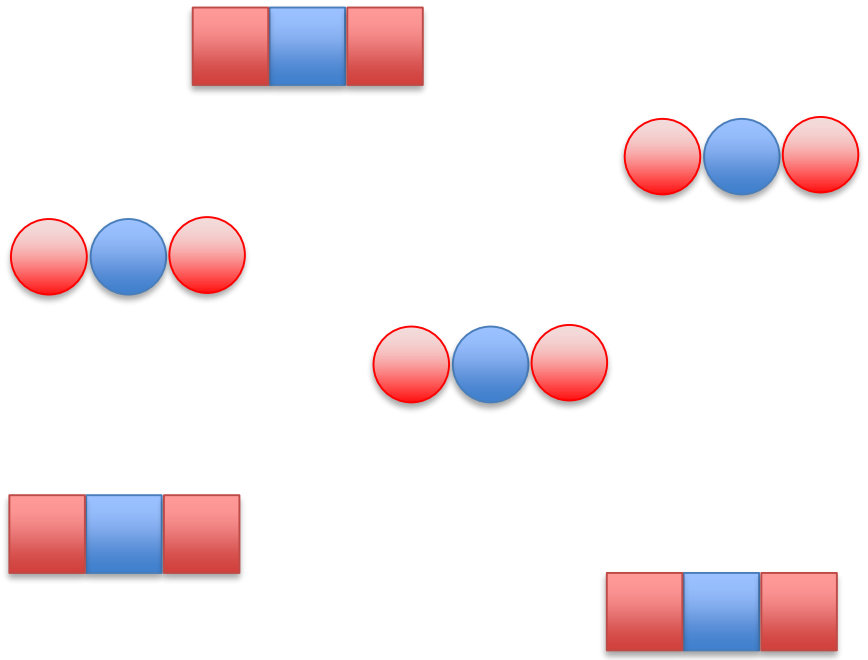




- What do I mean by catalytic information processing, and why is it important?
- Why are catalytic systems so much harder to design and build?
- How are we rising to that challenge?

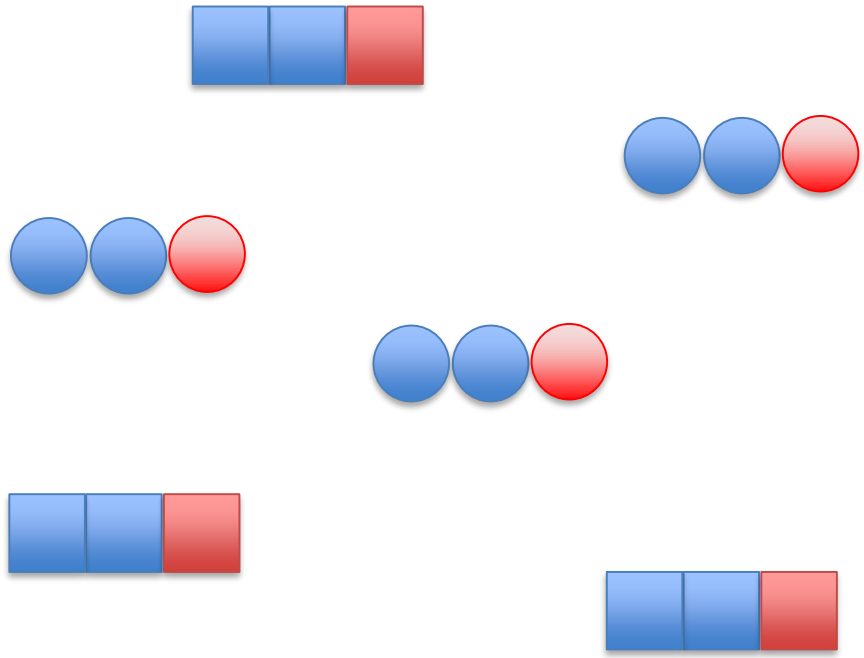
Catalytic information processing produces far-from-equilibrium states

The point of catalytic information processing is to correlate the state of one set of molecules with another set.



Catalytic information processing produces far-from-equilibrium states

The point of catalytic information processing is to correlate the state of one set of molecules with another set.



What does this mean thermodynamically? We need some more formalism…

Catalytic information processing produces far-from-equilibrium states

Two distinct macrostates (labelled by *m*) containing many microstates *y*.

vs

Catalytic information processing produces far-from-equilibrium states

Two distinct macrostates (labelled by *m*) containing many microstates *y*.

vs

$$p(y) \neq p_{\mathrm{eq}}\,(y) = \frac{\mathrm{e}^{-u(y)/kT}}{\sum_y \mathrm{e}^{-u(y)/kT}}$$

$$p(y|m) = p_{\mathrm{eq}}(y|m) = \frac{\mathrm{e}^{-u(y)/kT}}{\sum_{y\in m} \mathrm{e}^{-u(y)/kT}}$$

So $p(m)$ is the key quantity, determining how far system is from equilibrium.

[1] U Seifert, Eur. Phys. J. E 34:26, 2011.
[2] GE Crooks, Phys. Rev. E 60:2721, 1999.
[3] C Jarzynski, Annu. Rev. Cond. Matt. Phys. 2:329, 2011.

Catalytic information processing produces far-from-equilibrium states

Two distinct macrostates (labelled by *m*) containing many microstates *y*.

vs

$$p(m) \neq p_{\mathrm{eq}}(m) = \frac{\mathrm{e}^{-g(m)/kT}}{\sum_m \mathrm{e}^{-g(m)/kT}}$$

$$g(m) = \sum_{y \in m} u(y)\, p_{\mathrm{eq}}(y|m)$$

Average chemical energy of *m*

$$+ kT \sum_{y \in m} p_{\mathrm{eq}}(y|m) \ln p_{\mathrm{eq}}(y|m) = u(m) - Ts(m)$$

Entropy within macrostate *m*

[1] U Seifert, Eur. Phys. J. E 34:26, 2011.
[2] GE Crooks, Phys. Rev. E 60:2721, 1999.
[3] C Jarzynski, Annu. Rev. Cond. Matt. Phys. 2:329, 2011.

Catalytic information processing produces far-from-equilibrium states

$$G[p(m)] = U[p(m)] - TS[p(m)]$$

Generalised free energy of the whole probability distribution

$$g(m) = u(m) - Ts(m)$$

[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states

Generalised free energy of the whole probability distribution

$$G[p(m)] = U[p(m)] - TS[p(m)]$$

$$= \sum_m p(m)(u(m))$$

Average energy

$$g(m) = u(m) - Ts(m)$$

[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states

Generalised free energy of the whole probability distribution

$$G[p(m)] = U[p(m)] - TS[p(m)]$$

$$= \sum_m p(m)(u(m) - Ts(m)) + kT \sum_m p(m) \ln p(m)$$

$$g(m) = u(m) - Ts(m)$$

Entropy arising from average entropy of microstates

Entropy arising from uncertainty over microstates

[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states

Generalised free energy of the whole probability distribution

$$G[p(m)] = U[p(m)] - TS[p(m)]$$

$$= \sum_m p(m)(u(m) - Ts(m)) + kT \sum_m p(m) \ln p(m)$$

$$= \sum_m p(m) g(m) + kT \sum_m p(m) \ln p(m)$$

$$g(m) = u(m) - Ts(m)$$

[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states

Generalised free energy of the whole probability distribution

$$G[p(m)] = U[p(m)] - TS[p(m)]$$

$$= \sum_m p(m)(u(m) - Ts(m)) + kT \sum_m p(m) \ln p(m)$$

$$g(m) = u(m) - Ts(m)$$

$$= \sum_m p(m) g(m) + kT \sum_m p(m) \ln p(m)$$

$$= G_{\text{chem}}[p(m)] - TH[p(m)]$$

Average free energy of occupied macrostates

[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states

Generalised free energy of the whole probability distribution

$$G[p(m)] = U[p(m)] - TS[p(m)]$$

$$= \sum_m p(m)(u(m) - Ts(m)) + kT \sum_m p(m) \ln p(m)$$

$$g(m) = u(m) - Ts(m)$$

$$= \sum_m p(m) g(m) + kT \sum_m p(m) \ln p(m)$$

$$= G_{\text{chem}}[p(m)] - TH[p(m)]$$

Entropy due to uncertainty over macrostates.

[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

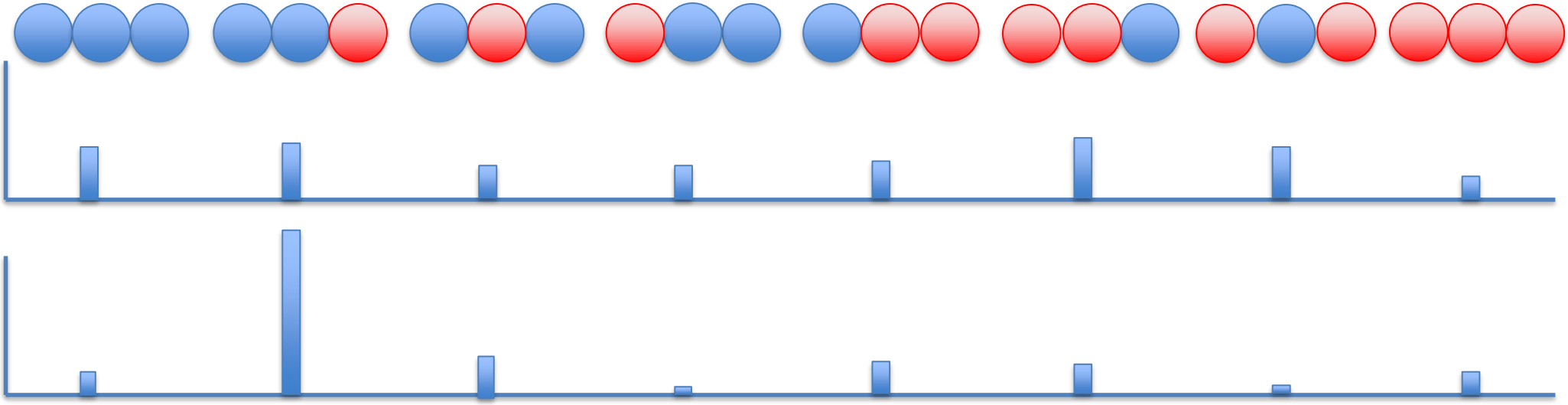
Catalytic information processing produces far-from-equilibrium states

- In a system which is only in contact with a thermal reservoir, $G[p(m)]$ can only decrease with time (2nd law of thermodynamics).

[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states

- In a system which is only in contact with a thermal reservoir, $G[p(m)]$ can only decrease with time (2nd law of thermodynamics).
- Thus $G[p(m)]$ is minimized by the equilibrium distribution $p(m) = p_{\mathrm{eq}}(m)$.

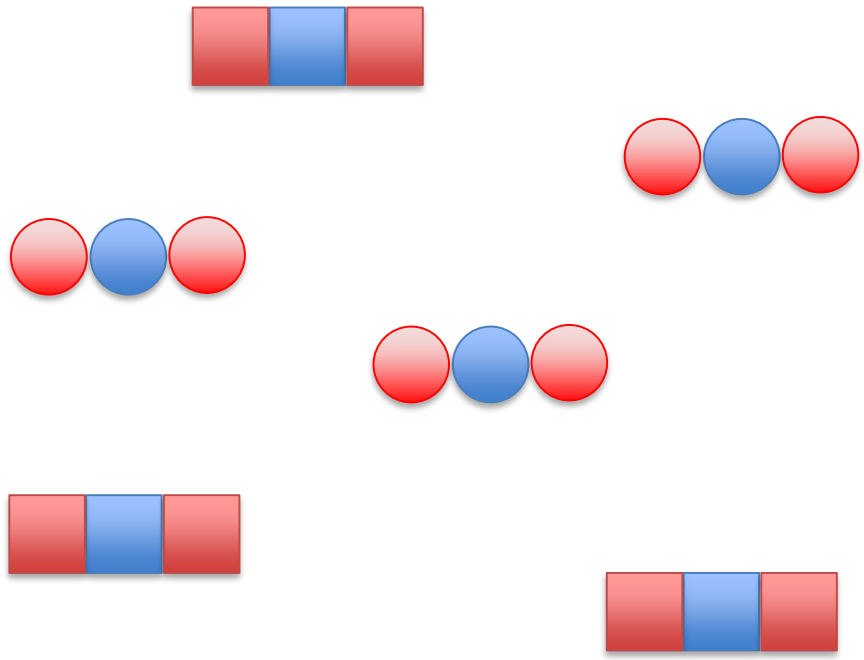


The equilibrium distribution balances $G[p(m)] = G_{\mathrm{chem}}[p(m)] - kTH[p(m)]$

[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states

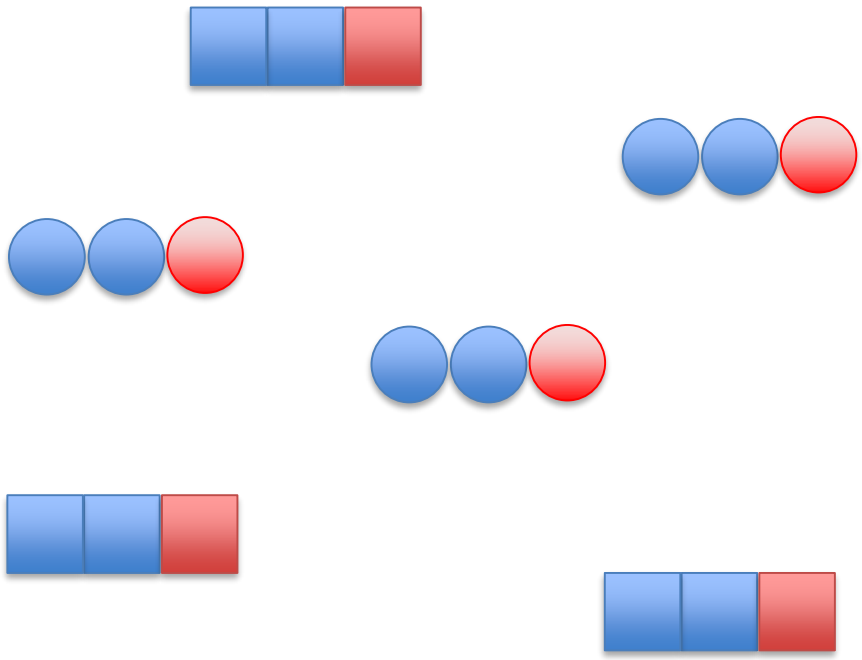
- Correlations imply low $H[p(m)]$.



[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states

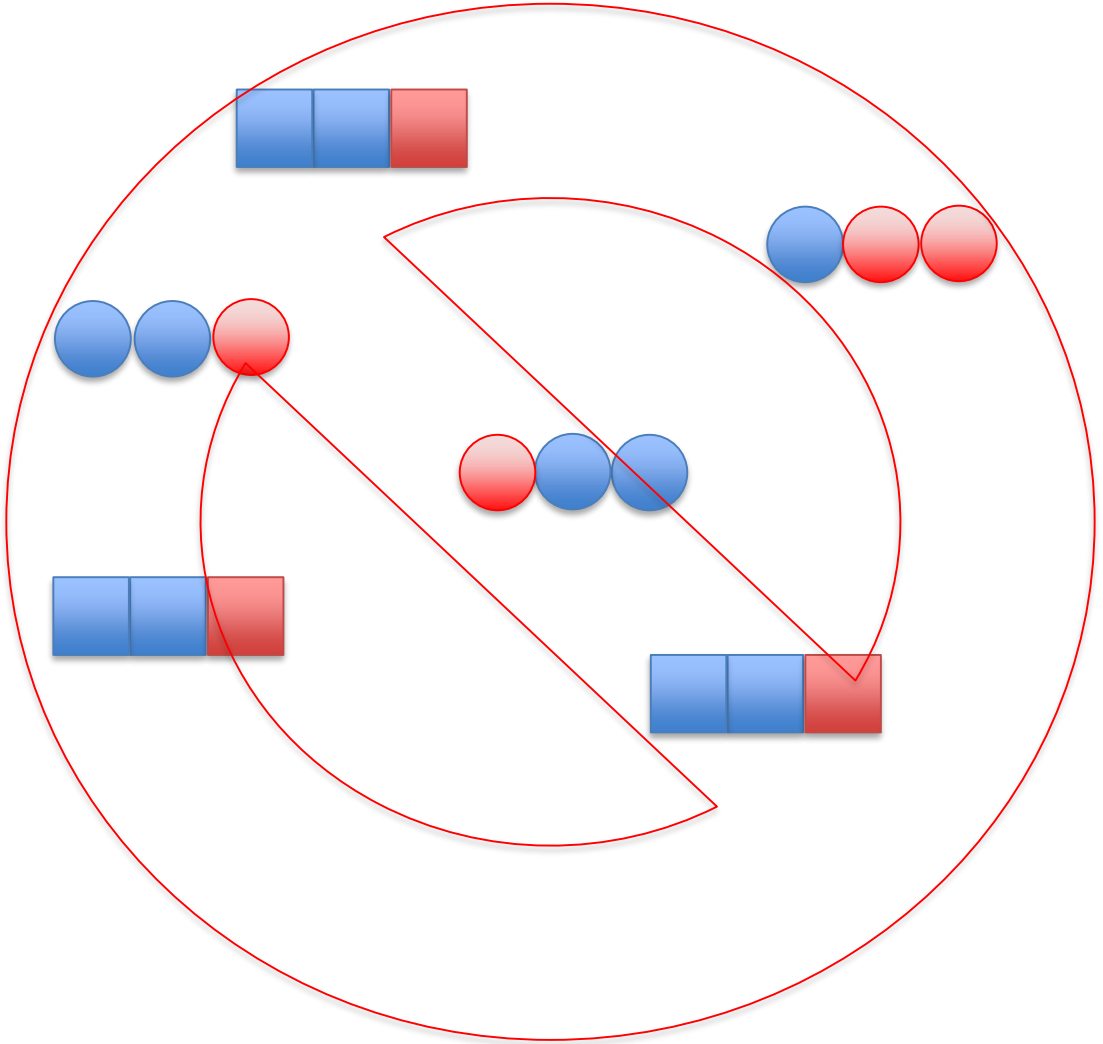
- Correlations imply low $H[p(m)]$.



[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states

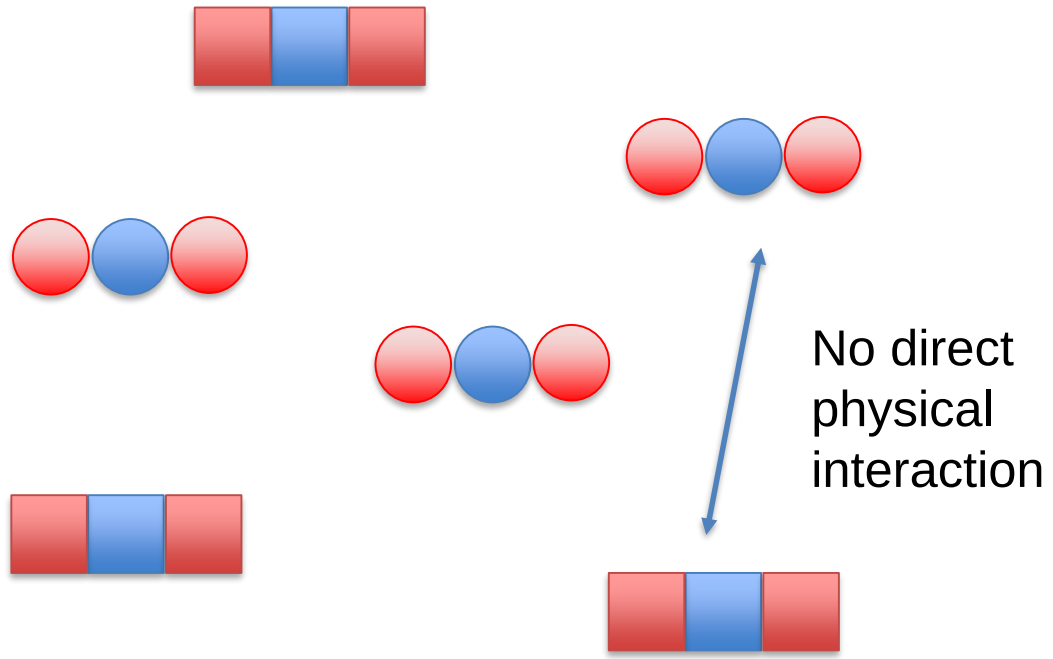
- Correlations imply low $H[p(m)]$.



[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

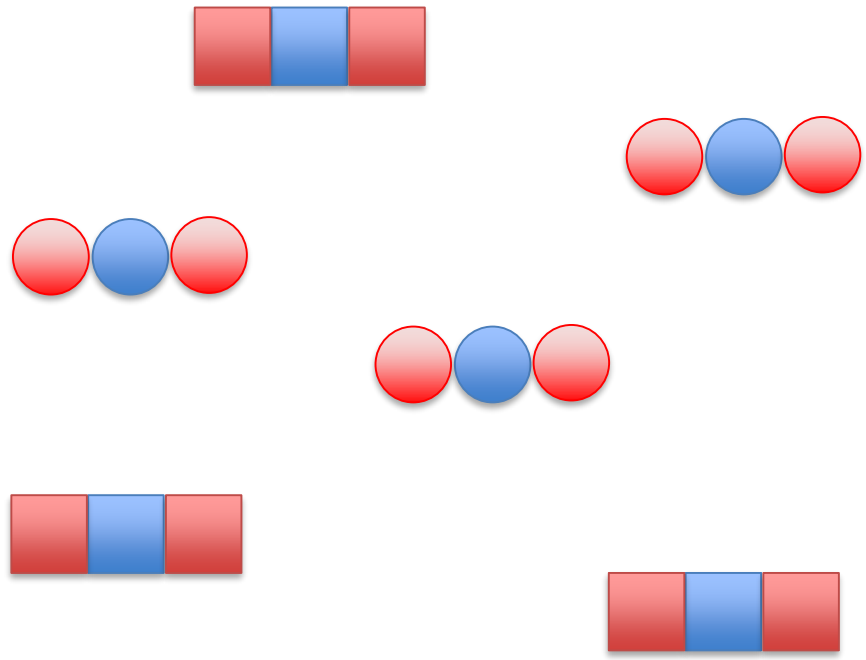
Catalytic information processing produces far-from-equilibrium states

- Correlations imply low $H[p(m)]$.
- $G_{\text{chem}}[p(m)]$ doesn't favour these low entropy states.



[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states



- Correlations imply low $H[p(m)]$.
- $G_{\text{chem}}[p(m)]$ doesn't favour these low entropy states.
- System is therefore *really far* from equilibrium: $p(m) \neq p_{\text{eq}}(m)$.

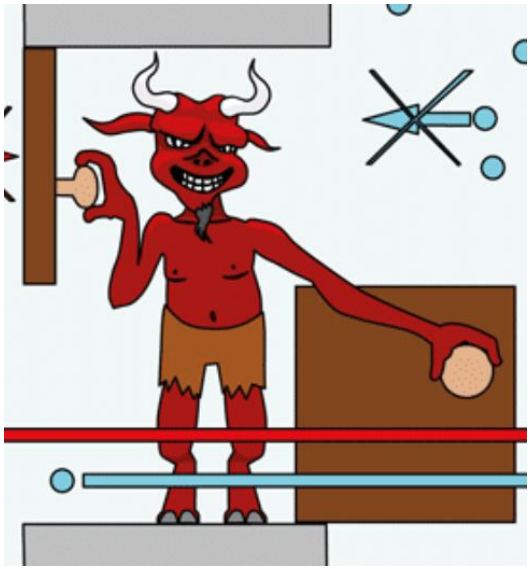
$$G[p(m)] = G_{\text{chem}}[p(m)] - kTH[p(m)]$$

[1] JM Parrondo J.M Horrowitz, T Sagawa, Nat. Phys. 11:131, 2015.
[2] M Esposito, C Van den Broeck, Europhys. Lett. 95: 40004, 2011.

Catalytic information processing produces far-from-equilibrium states

A deeper dive into the thermodynamics of non-equilibrium correlations in molecular systems, and their role in the Maxwell's demon paradox:

- Thermodynamics of Computational Copying in Biochemical Systems TEO, CC Govern, PR ten Wolde, PRX 7, 021004, 2017.
- Fundamental costs in the production and destruction of persistent polymer copies, TEO, PR ten Wolde, PRL 118, 158103, 2017.
- The power of being explicit: demystifying work, heat, and free energy in the physics of computation, TEO, RA Brittain, PR Wolde in "The Energetics of Computing in Life and Machines", SFI press, arXiv:1812.09572.
- Biochemical Szilard engines for memory-limited inference, RA Brittain, NS Jones, TEO, N. J. Phys. 21, 063022, 2018.
- Nonequilibrium correlations in minimal dynamical models of polymer copying, JM Poulton, PR ten Wolde, TEO, PNAS 116, 1946-1951, 2019.
- Edge-effects dominate copying thermodynamics for finite-length molecular oligomers, JM Poulton, TEO, arXiv:2005.11255.



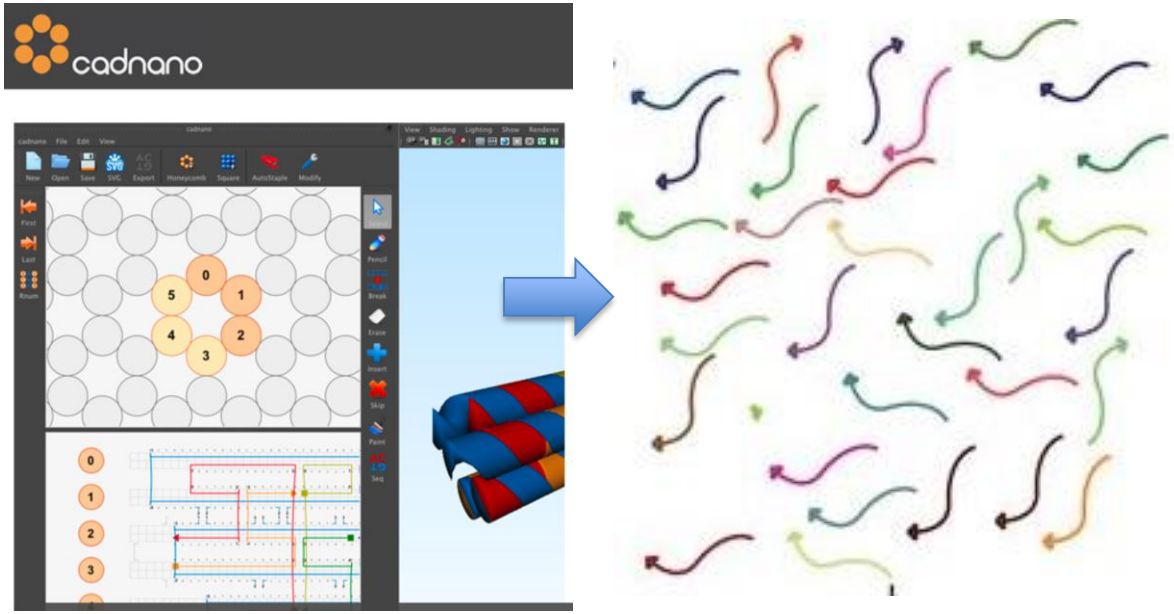
Producing specific non-equilibrium states is *really hard*

In recent years, we've got *really good* at engineering self-assembling molecular systems.

[1] SM Douglas *et al., Nucl. Acids Res.* 37:5001-5006, 2009.

Producing specific non-equilibrium states is *really hard*

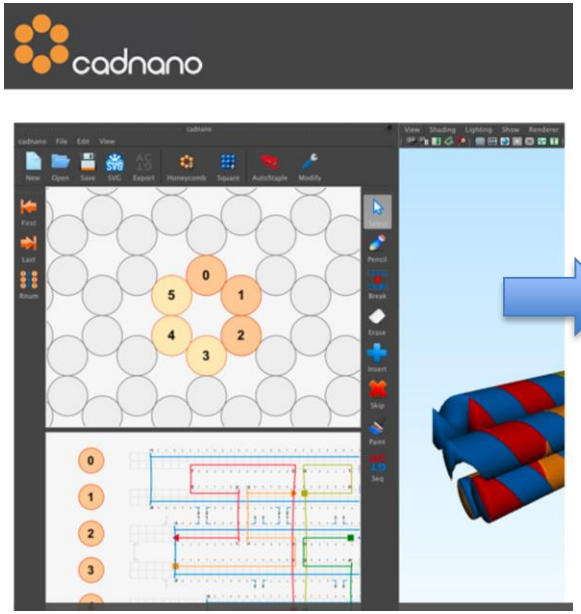
In recent years, we've got *really good* at engineering self-assembling molecular systems.

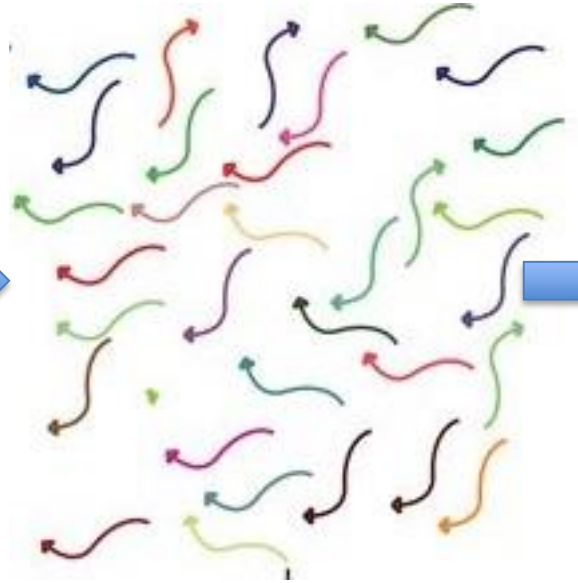


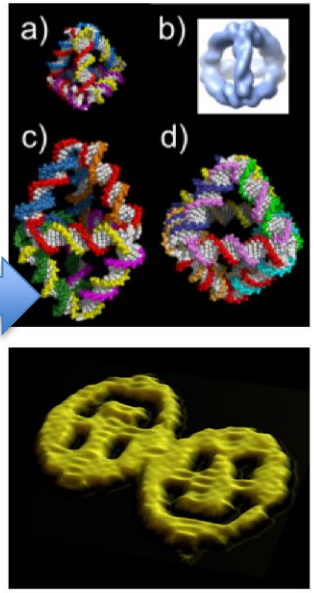
[1] SM Douglas *et al., Nucl. Acids Res.* 37:5001-5006, 2009.

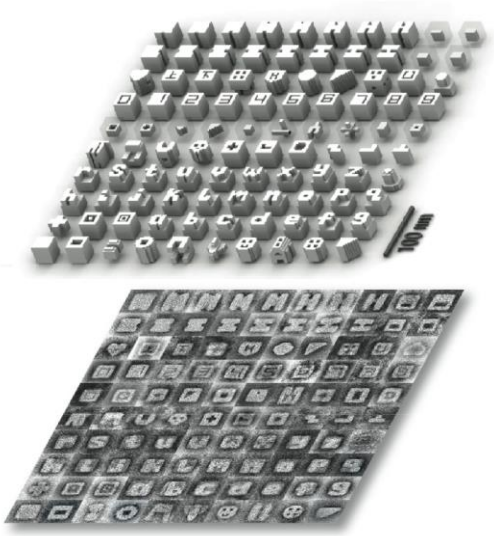
Producing specific non-equilibrium states is *really hard*

In recent years, we've got *really good* at engineering self-assembling molecular systems.





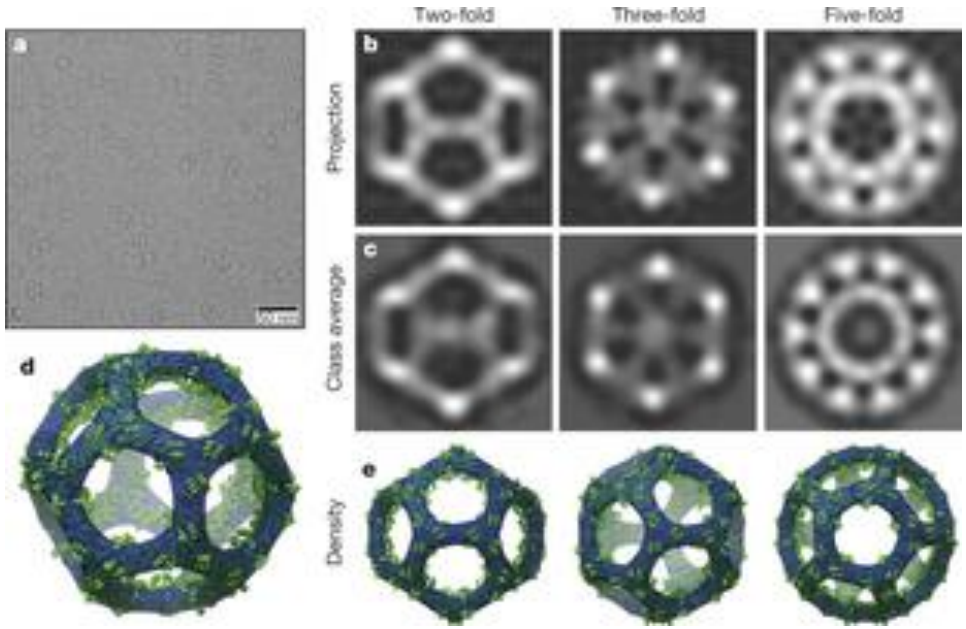




[1] SM Douglas *et al., Nucl. Acids Res.* 37:5001-5006, 2009.
[2] RP Goodman *et al., Science,* 310:1661-1665, 2005.
[3] PWK Rothemund, *Nature,* 440:297-302, 2006.
[4] Y Ke, LL Ong, WM Shih, P Yin, Science 338:177, 2012.

Producing specific non-equilibrium states is *really hard*

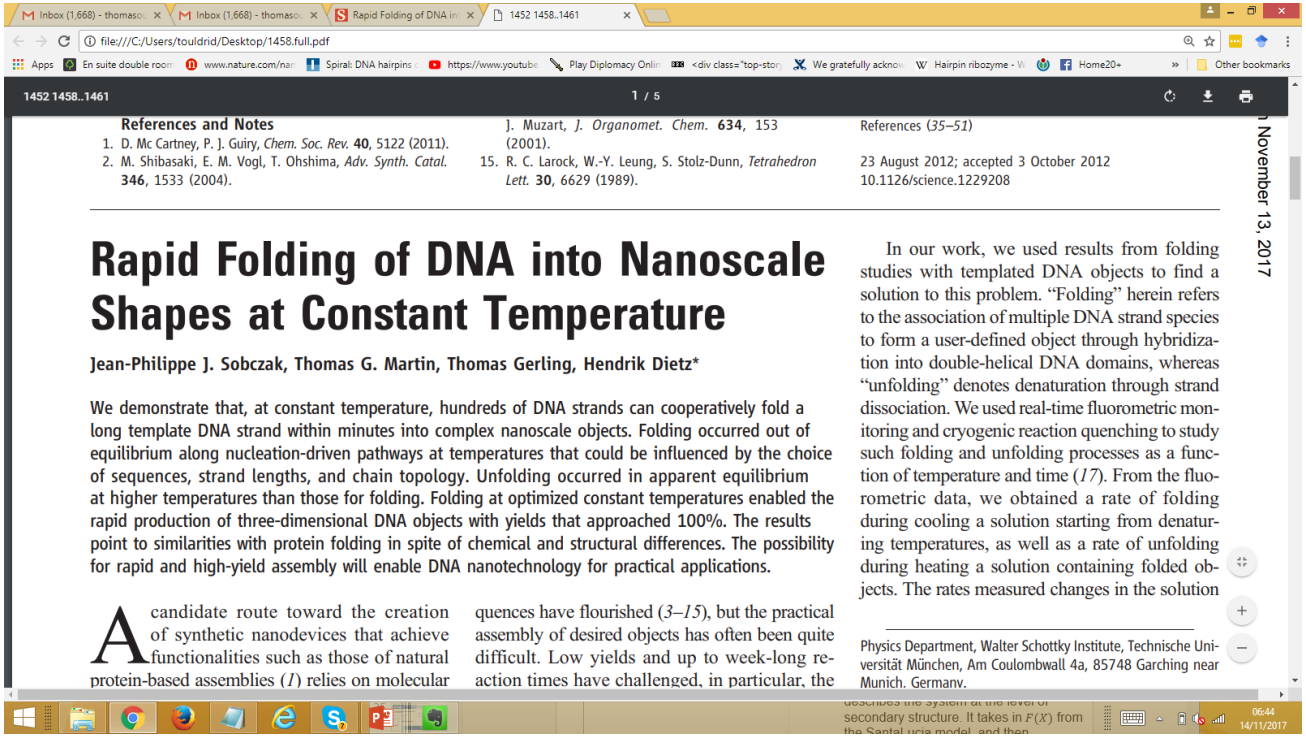
In recent years, we've got *really good* at engineering self-assembling molecular systems.



[4] Y Hsia *et al. Nature* 535:136-139, 2016.

Producing specific non-equilibrium states is *really hard*

In recent years, we've got *really good* at engineering self-assembling molecular systems.



References and Notes

1. D. Mc Cartney, P. J. Guiry, *Chem. Soc. Rev.* **40**, 5122 (2011).
2. M. Shibasaki, E. M. Vogl, T. Ohshima, *Adv. Synth. Catal.* **346**, 1533 (2004).

J. Muzart, *J. Organomet. Chem.* **634**, 153 (2001).
15. R. C. Larock, W.-Y. Leung, S. Stolz-Dunn, *Tetrahedron Lett.* **30**, 6629 (1989).

References (*35–51*)

23 August 2012; accepted 3 October 2012
10.1126/science.1229208

Rapid Folding of DNA into Nanoscale Shapes at Constant Temperature

Jean-Philippe J. Sobczak, Thomas G. Martin, Thomas Gerling, Hendrik Dietz*

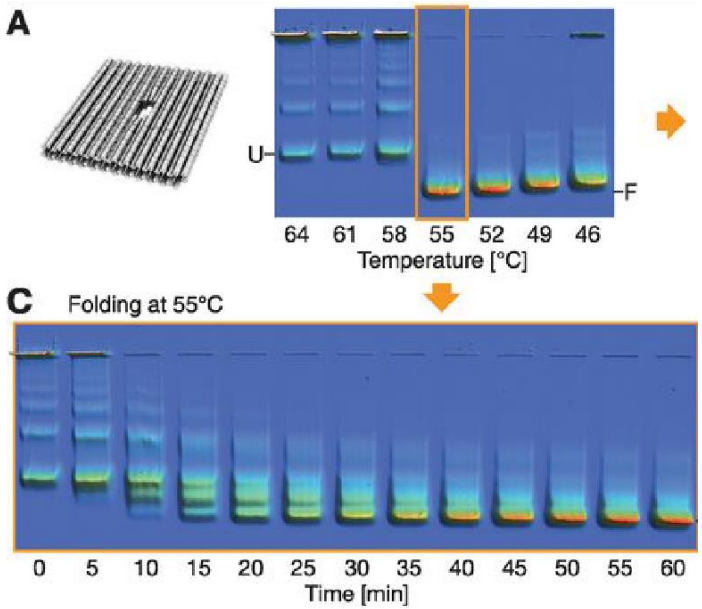
We demonstrate that, at constant temperature, hundreds of DNA strands can cooperatively fold a long template DNA strand within minutes into complex nanoscale objects. Folding occurred out of equilibrium along nucleation-driven pathways at temperatures that could be influenced by the choice of sequences, strand lengths, and chain topology. Unfolding occurred in apparent equilibrium at higher temperatures than those for folding. Folding at optimized constant temperatures enabled the rapid production of three-dimensional DNA objects with yields that approached 100%. The results point to similarities with protein folding in spite of chemical and structural differences. The possibility for rapid and high-yield assembly will enable DNA nanotechnology for practical applications.

A candidate route toward the creation of synthetic nanodevices that achieve functionalities such as those of natural protein-based assemblies (*1*) relies on molecular

quences have flourished (*3–15*), but the practical assembly of desired objects has often been quite difficult. Low yields and up to week-long reaction times have challenged, in particular, the

In our work, we used results from folding studies with templated DNA objects to find a solution to this problem. "Folding" herein refers to the association of multiple DNA strand species to form a user-defined object through hybridization into double-helical DNA domains, whereas "unfolding" denotes denaturation through strand dissociation. We used real-time fluorometric monitoring and cryogenic reaction quenching to study such folding and unfolding processes as a function of temperature and time (*17*). From the fluorometric data, we obtained a rate of folding during cooling a solution starting from denaturing temperatures, as well as a rate of unfolding during heating a solution containing folded objects. The rates measured changes in the solution

Physics Department, Walter Schottky Institute, Technische Universität München, Am Coulombwall 4a, 85748 Garching near Munich, Germany.



[1] J-P Sobczak *et al. Science* 338:1458-1461, 2012.

Producing specific non-equilibrium states is *really hard*

Q: Why are there very few synthetic catalytic information processing systems, given our success with self-assembly?

Producing specific non-equilibrium states is *really hard*

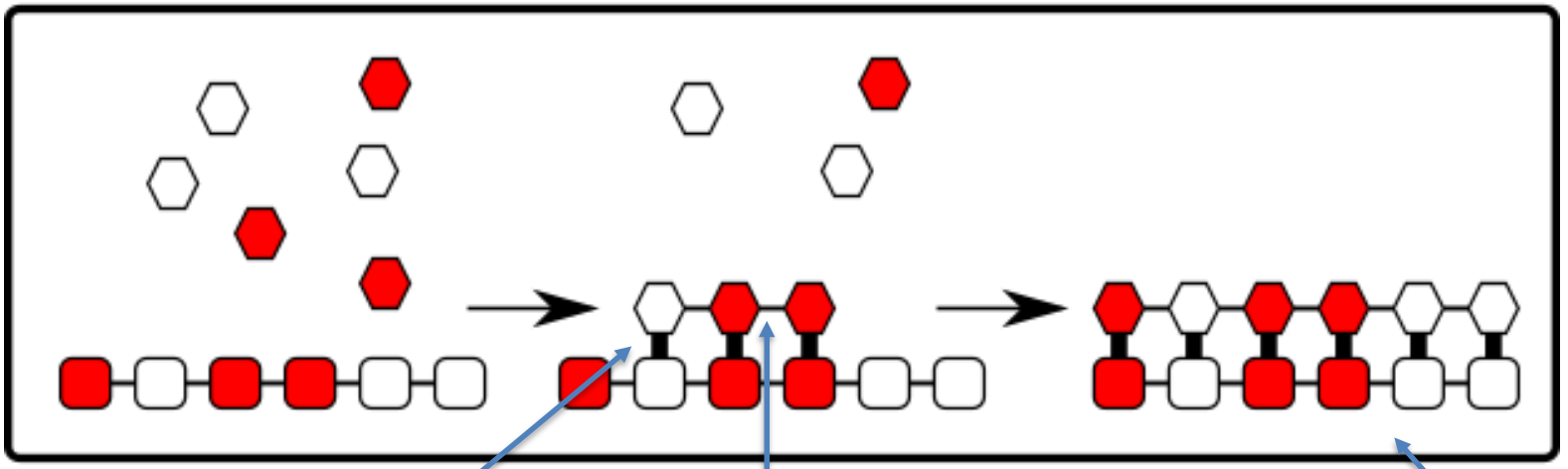
Q: Why are there very few synthetic catalytic information processing systems, given our success with self-assembly?

A: They're hard to engineer because you need strong, selectively-attractive interactions that are then disrupted later on (the practical consequence of being out of equilibrium).

Producing specific non-equilibrium states is *really hard*

The problem is at its worst for polymerisation

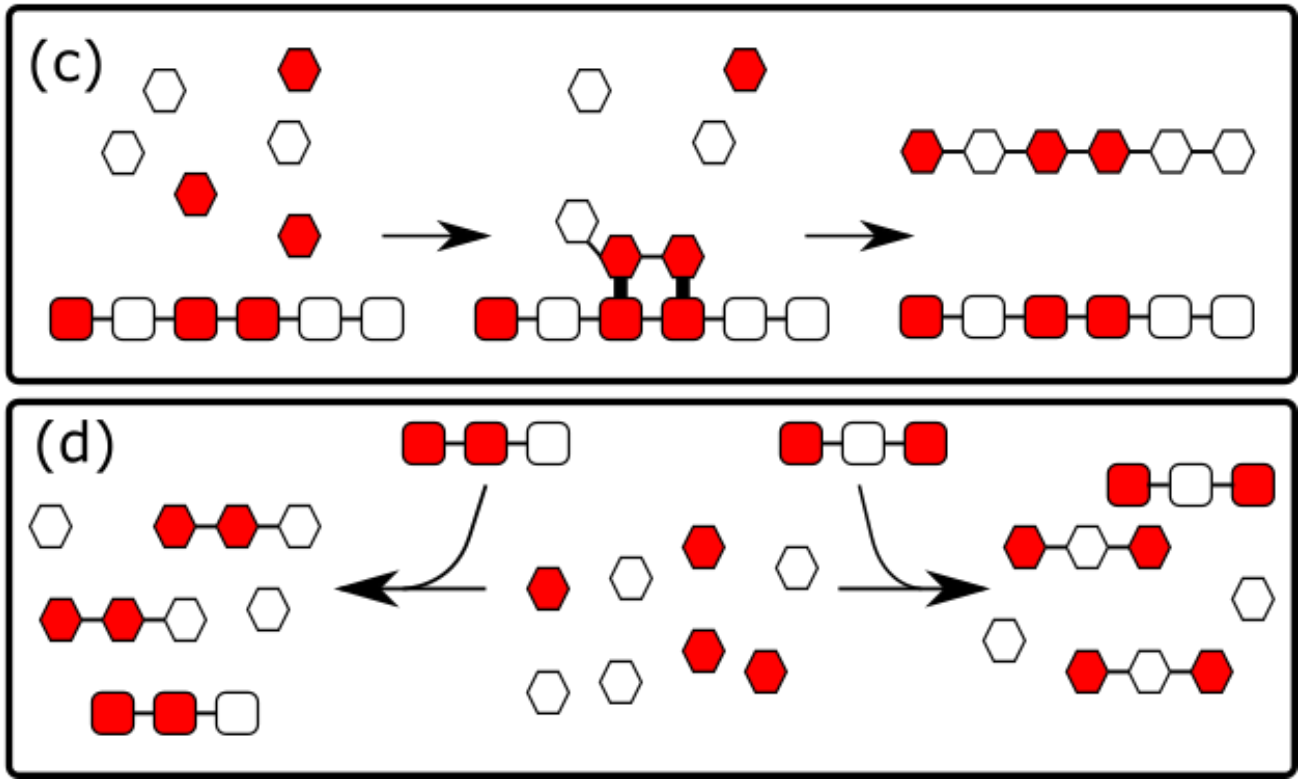
"Product Inhibition"



Selectively-attractive interaction

Polymerisation encouraged by proximity.

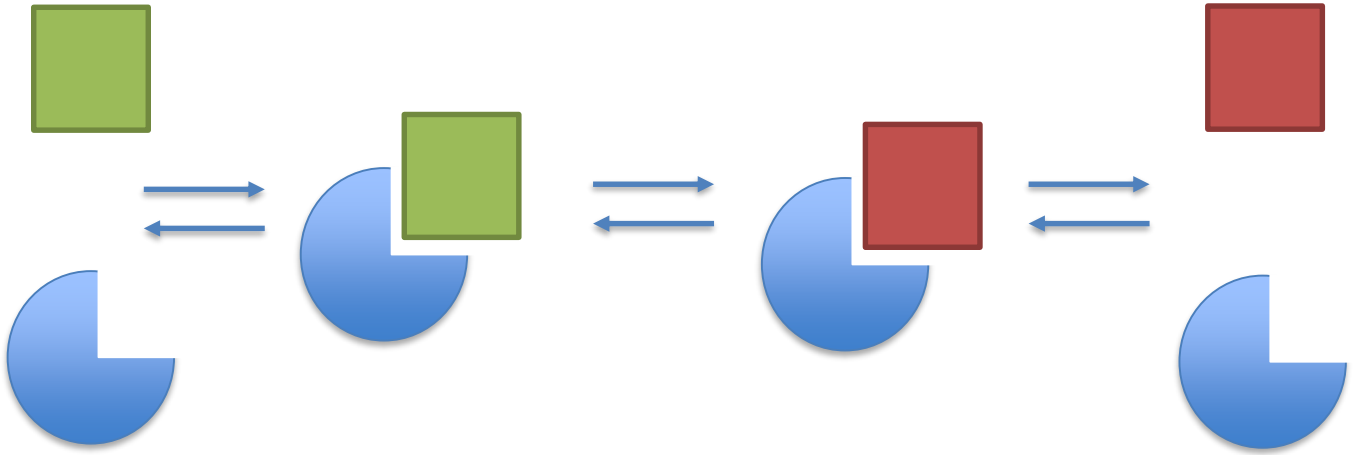
Daughter polymer cooperatively anchored to template.



- What do I mean by catalytic information processing, and why is it important?
- Why are catalytic systems so much harder to design and build?
- How are we rising to that challenge?

How should a catalyst be designed to minimize sequestration?

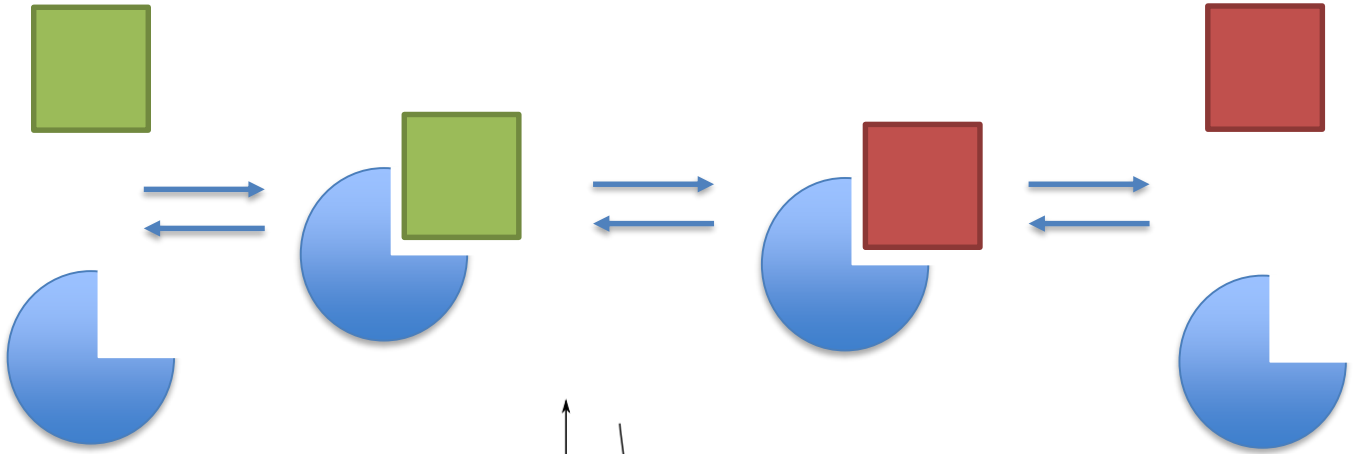
Abhishek Deshpande
(Wisconsin-Madison)



Optimizing enzymatic catalysts for rapid turnover of substrates with low enzyme sequestration,
A Deshpande, TEO, arXiv:1905.00555.

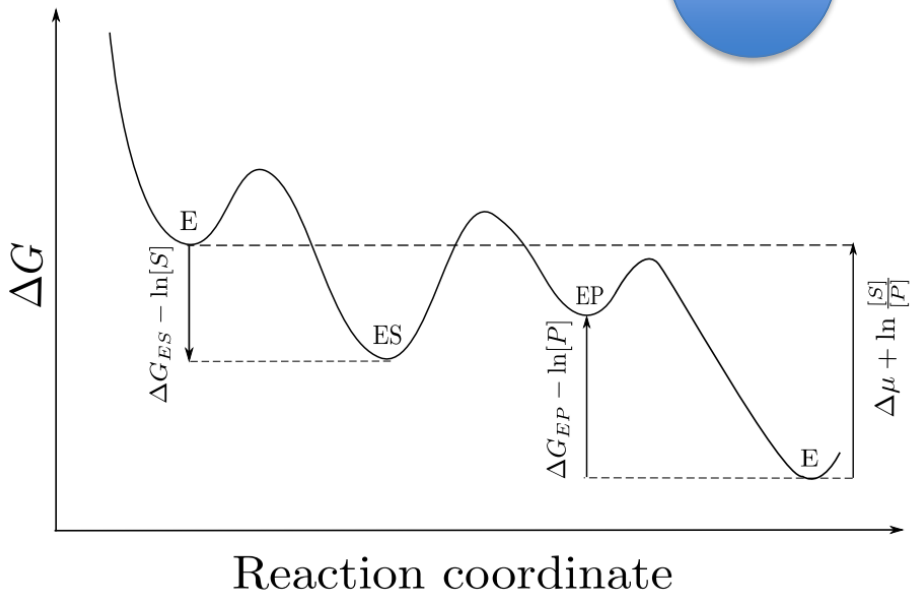
How should a catalyst be designed to minimize sequestration?

Given a fixed overall free-energy budget, diffusion-controlled binding reactions and a limited catalytic rate, how should the free energy of intermediate states be chosen to meet a target flux whilst minimizing sequestration?



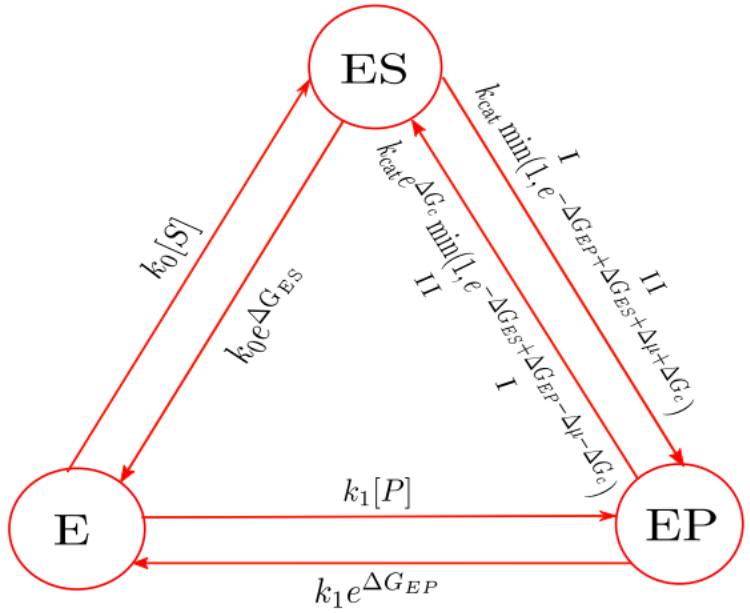
Forwards and backwards rates constrained by detailed balance:

$$\frac{k_+}{k_-} = \mathrm{e}^{-\Delta g/kT}$$



How should a catalyst be designed to minimize sequestration?

Given a fixed overall free-energy budget, diffusion-controlled binding reactions and a limited catalytic rate, how should the free energy of intermediate states be chosen to meet a target flux whilst minimizing sequestration?

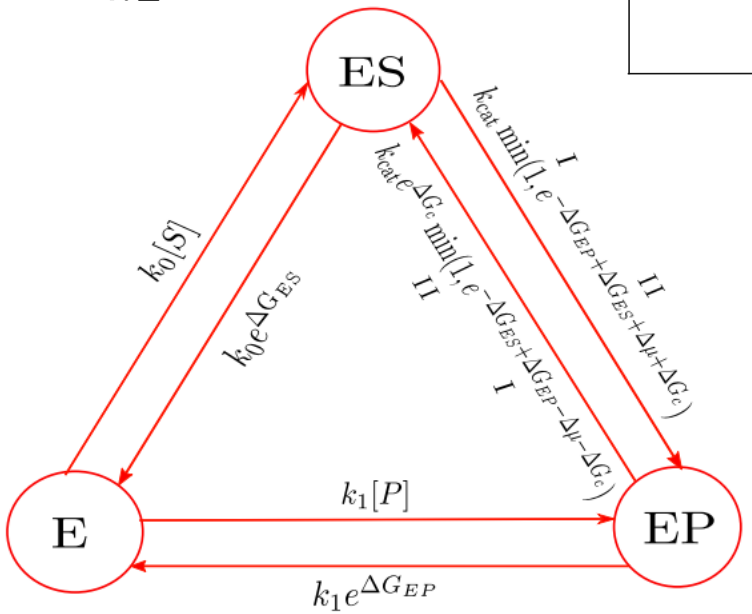


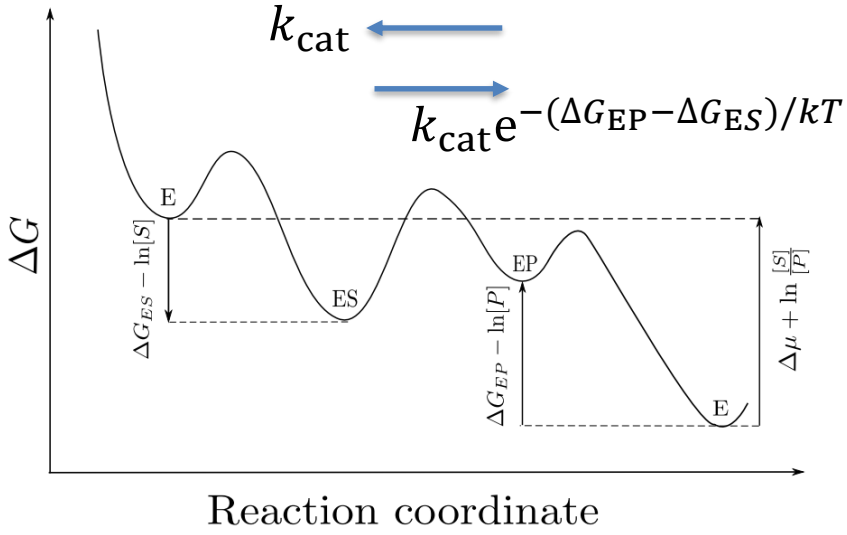
How should a catalyst be designed to minimize sequestration?

Given a fixed overall free-energy budget, diffusion-controlled binding reactions and a limited catalytic rate, how should the free energy of intermediate states be chosen to meet a target flux whilst minimizing sequestration?

Forwards and backwards rates constrained by detailed balance:

$$\frac{k_+}{k_-} = \mathrm{e}^{-\Delta g/kT}$$



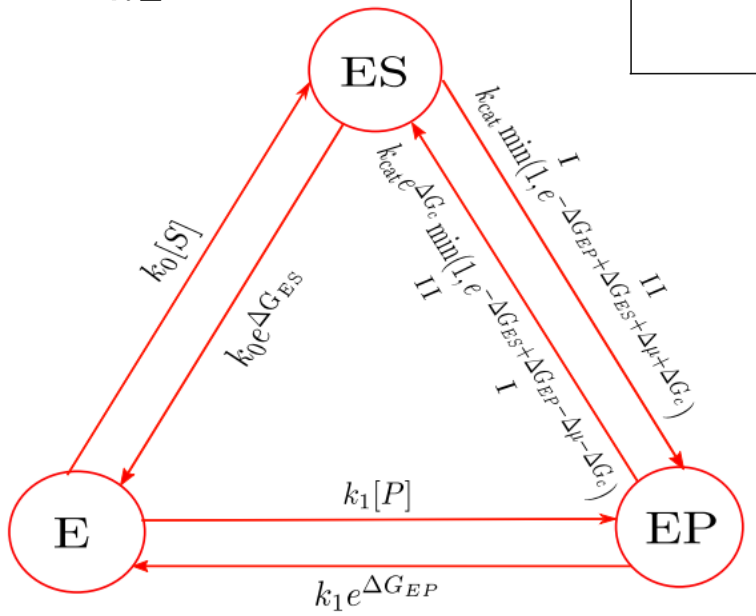


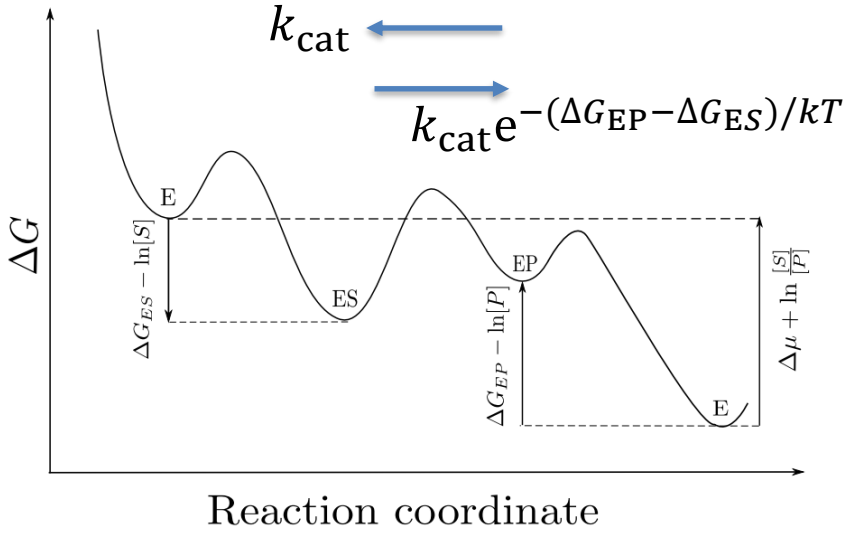
How should a catalyst be designed to minimize sequestration?

Given a fixed overall free-energy budget, diffusion-controlled binding reactions and a limited catalytic rate, how should the free energy of intermediate states be chosen to meet a target flux whilst minimizing sequestration?

Forwards and backwards rates constrained by detailed balance:

$$\frac{k_+}{k_-} = \mathrm{e}^{-\Delta g/kT}$$

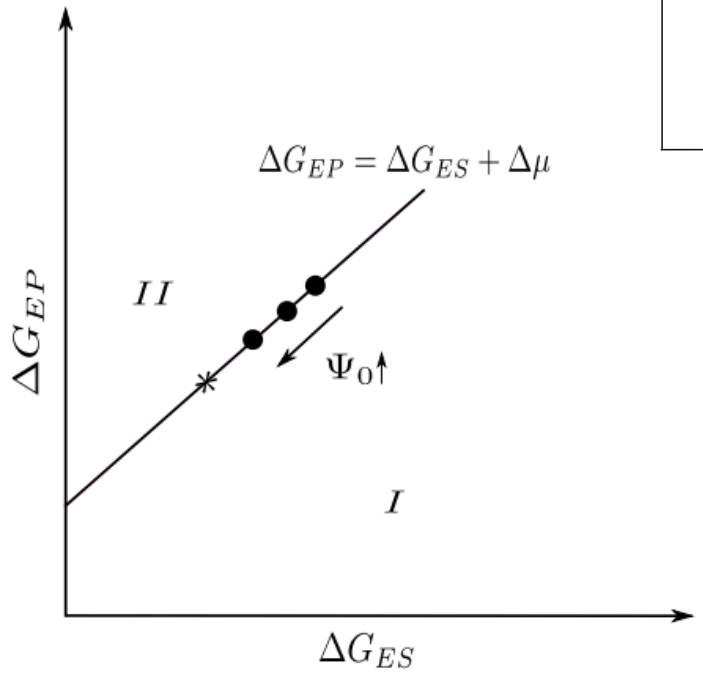


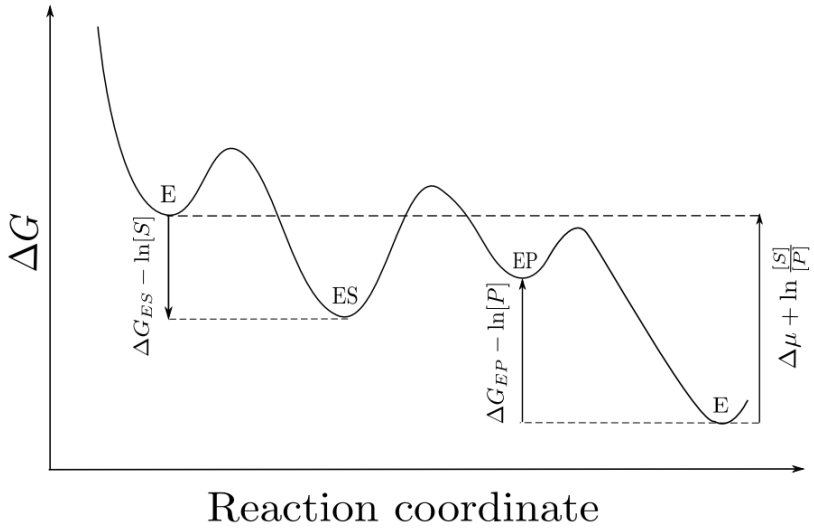


Flux and sequestration can be analysed using first-passage-time techniques.

How should a catalyst be designed to minimize sequestration?

Given a fixed overall free-energy budget, diffusion-controlled binding reactions and a limited catalytic rate, how should the free energy of intermediate states be chosen to meet a target flux whilst minimizing sequestration?

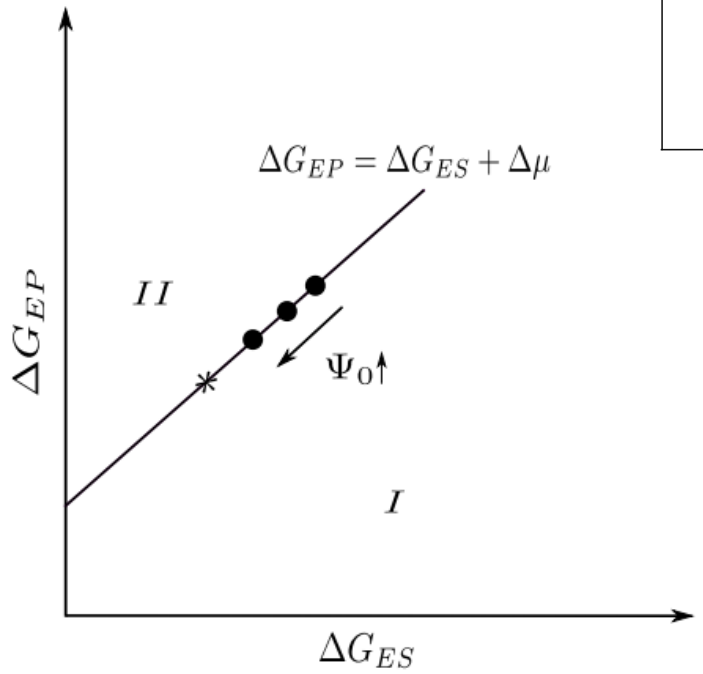


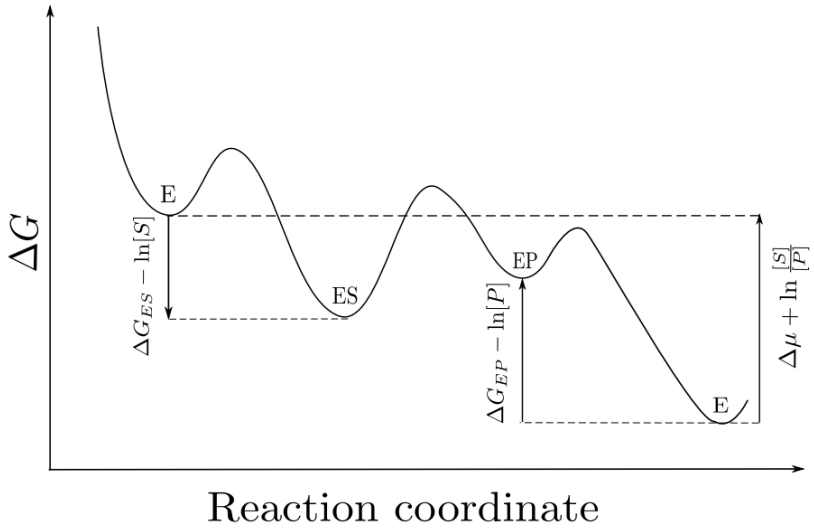


Free energies of intermediate states strongly constrained.

How should a catalyst be designed to minimize sequestration?

Given a fixed overall free-energy budget, diffusion-controlled binding reactions and a limited catalytic rate, how should the free energy of intermediate states be chosen to meet a target flux whilst minimizing sequestration?



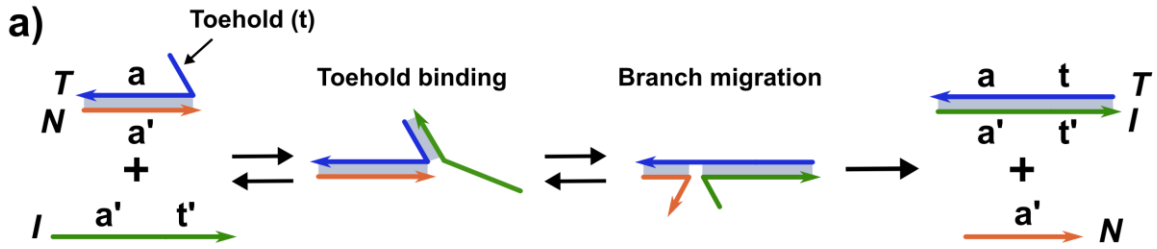


Optimal strategy is to use the intrinsic free-energy change $\Delta\mu$ to "pay" for an increase in G_{EP} relative to G_{ES}.

Can we use this insight to design synthetic systems for catalytic information propagation?

Javier Cabello Garcia

Handhold-mediated strand displacement: a nucleic acid-based mechanism for generating far-from-equilibrium assemblies through templated reactions
J Cabello-Garcia, W Bae, GBV Stan, TEO.

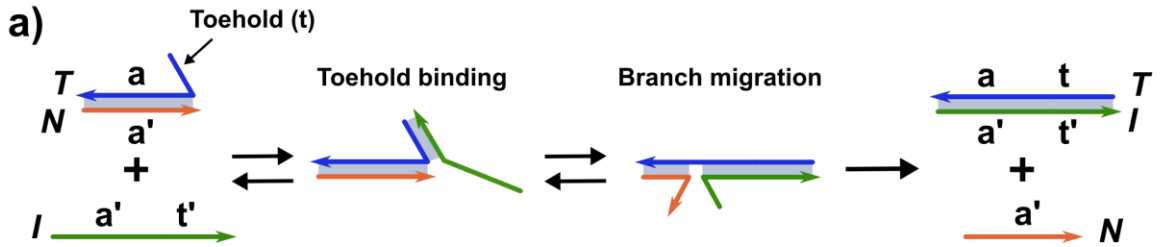


In DNA nanotech, toehold mediated strand displacement is a way to engineer dynamical systems based on recognition of molecular toeholds. But it tends to produce equilibrium products with sequestered recognition regions.

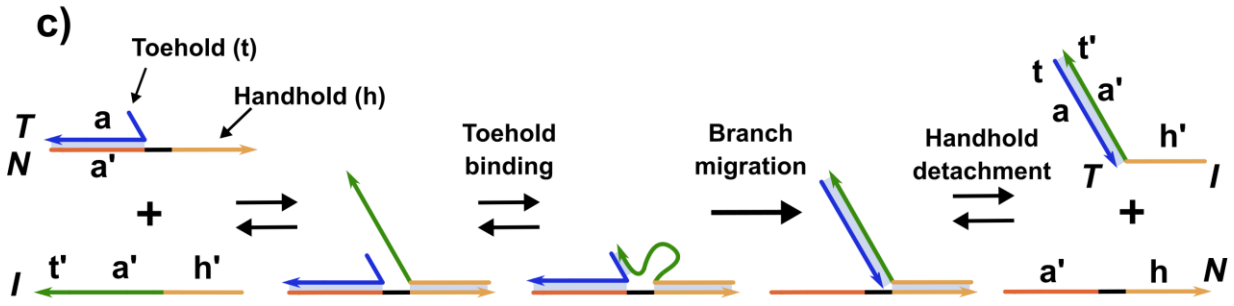
Can we use this insight to design synthetic systems for catalytic information propagation?

Javier Cabello Garcia

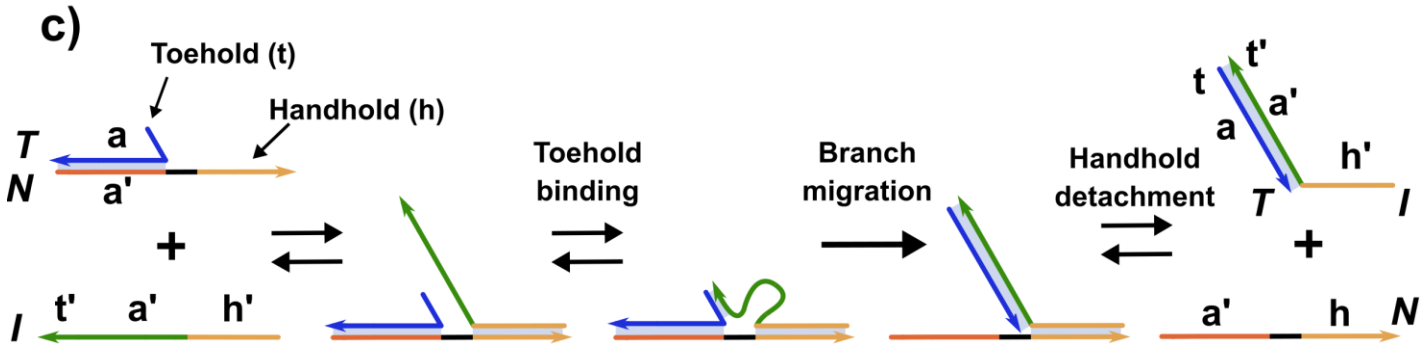
Handhold-mediated strand displacement: a nucleic acid-based mechanism for generating far-from-equilibrium assemblies through templated reactions
J Cabello-Garcia, W Bae, GBV Stan, TEO.



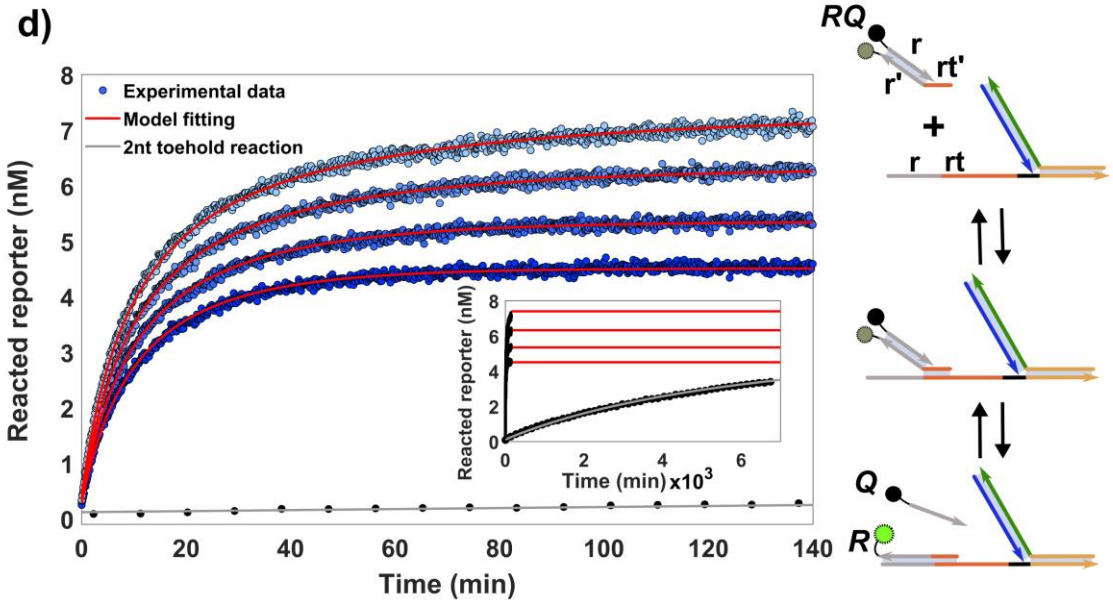
We propose handhold-mediated strand displacement as an alternative that allows for the templating of non-equilibrium assemblies.



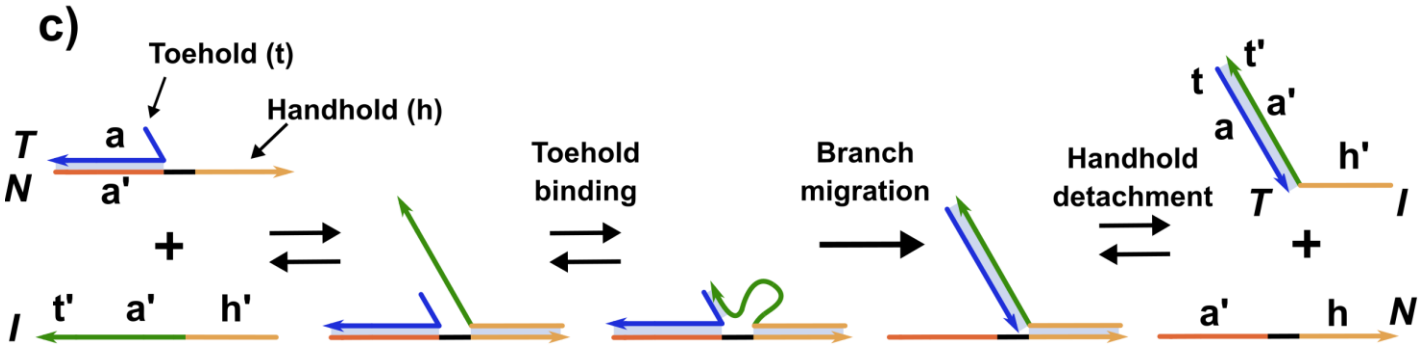
Can we use this insight to design synthetic systems for catalytic information propagation?



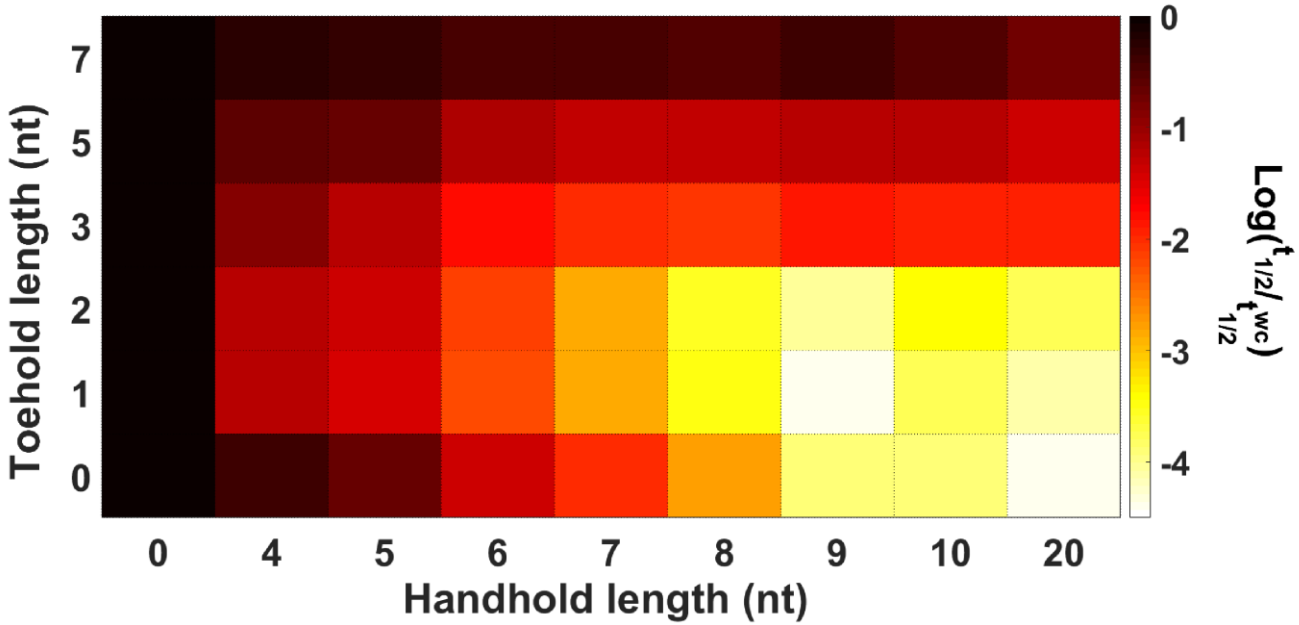
Progress of binding between Invader (I) and Target (T) can be followed by a fluorescent reporter.



Can we use this insight to design synthetic systems for catalytic information propagation?

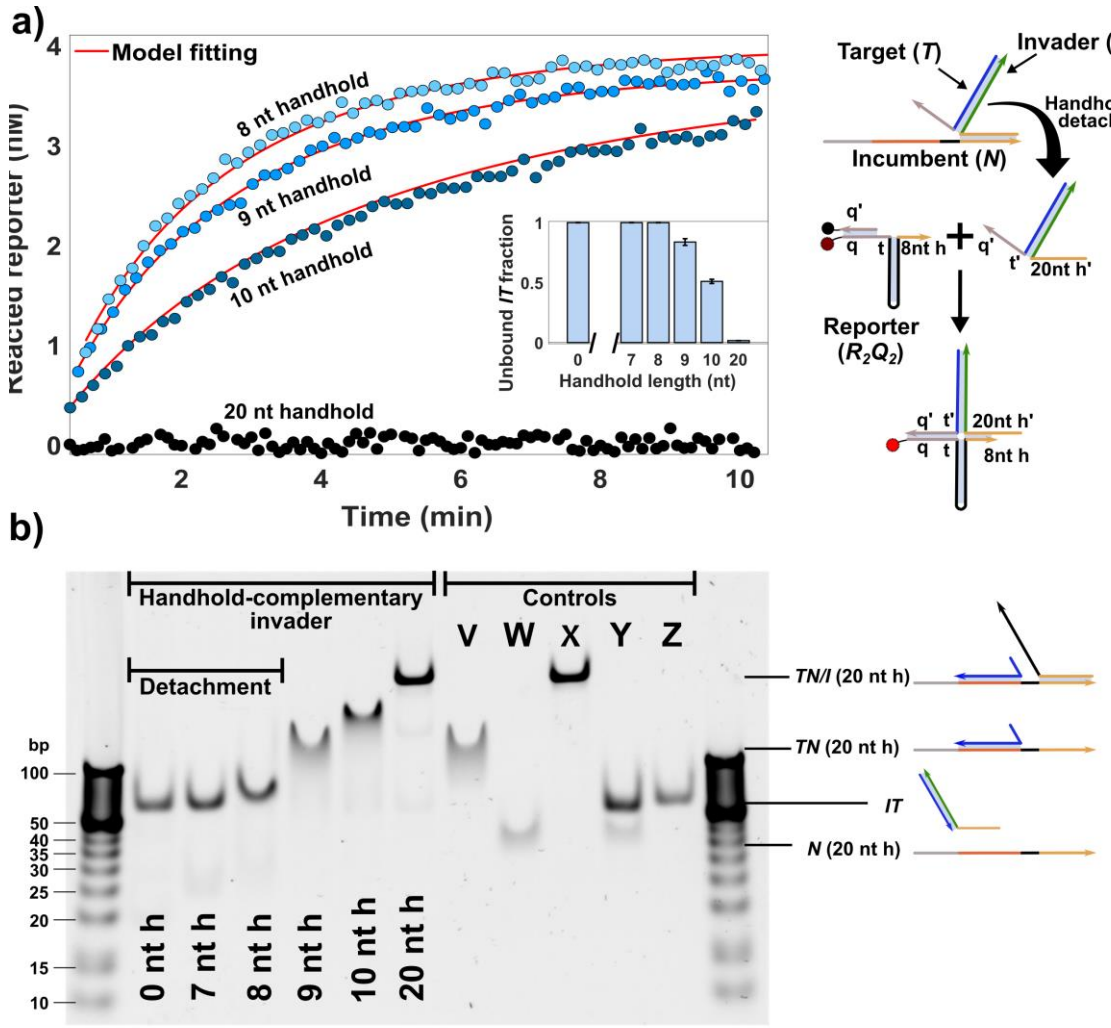


Handhold drives association of Invader (I) and Target (T).



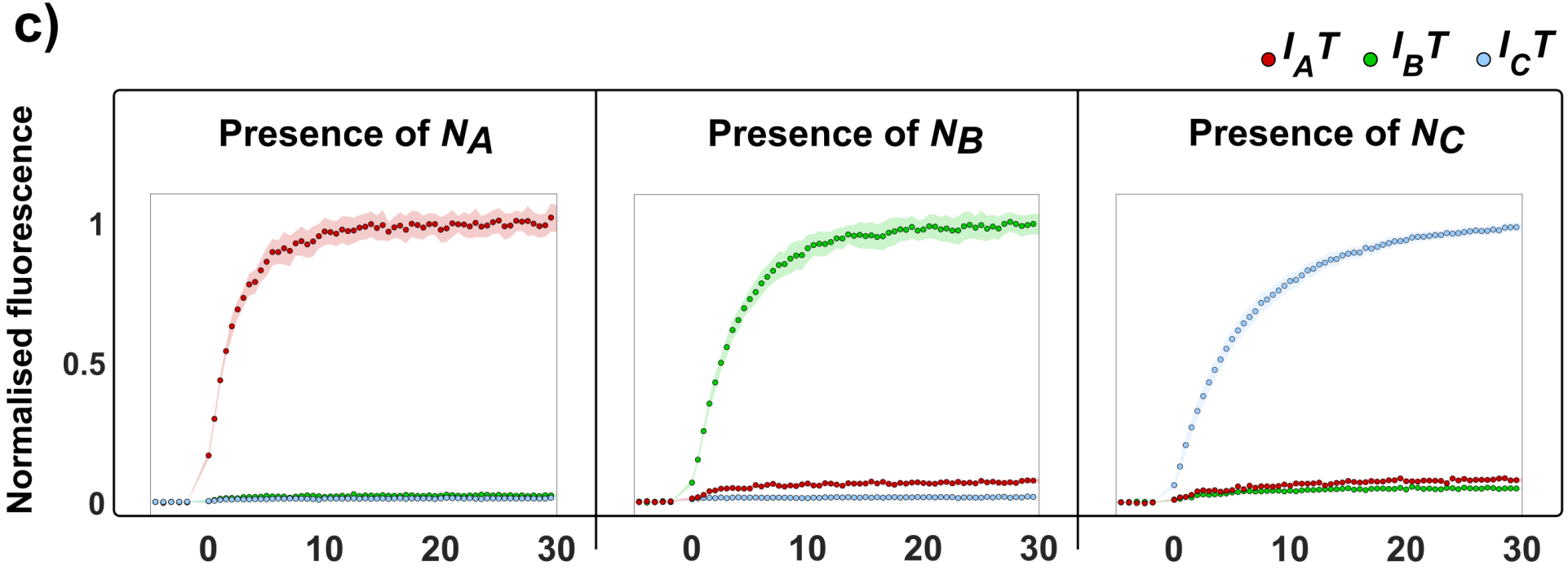
Can we use this insight to design synthetic systems for catalytic information propagation?

Moderate handholds also allow for detachment.



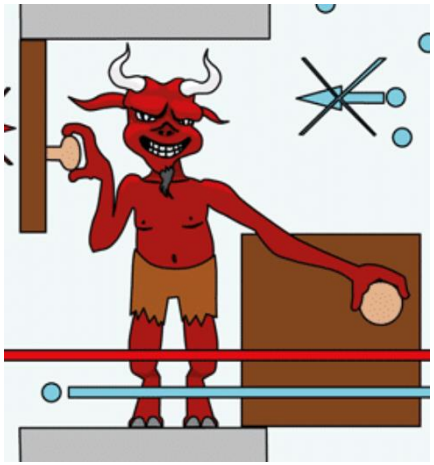
Can we use this insight to design synthetic systems for catalytic information propagation?

We are thus able to use HMSD to selectively template the assembly of non-equilibrium complexes.



3 invaders and 3 reporters in all experiments; product selected by template handhold.

Where next?



- Use HMSD to build a genuinely catalytic system for dimerisation.
- Use HMSD to template the assembly of longer polymers (lots of interesting theoretical and experimental questions here).
- Incorporate kinetic proofreading?
- Build a replicator, rather than just a copier?
- Build a multi-layer signal propagation system?
- Get cells to assemble the far-from equilibrium ingredients of these nucleic acid systems?

Thanks!

This work has been supported by:

THE ROYAL SOCIETY

erc
European Research Council
Established by the European Commission

Imperial College London

- Everyone in the “Principles of Biomolecular Systems” group at Imperial.
- Pieter Rein ten Wolde from FOM Institute AMOLF.
- Nick Jones and Guy-Bart Stan at Imperial.
- Andrew Turberfield and his group from Oxford.

ICL Department of Bioengineering

Imperial College Centre for Synthetic Biology