ENUMERATION, CONDENSATION AND SIMULATION OF PSEUDOKNOT-FREE DOMAIN-LEVEL DNA STRAND DISPLACEMENT SYSTEMS

Stefan Badelt, Casey Grun, Karthik V. Sarma, Seung Woo Shin, Brian Wolfe, and Erik Winfree

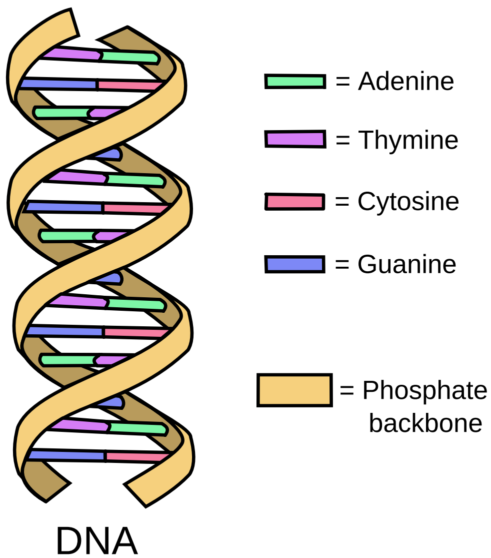
DNA and Natural Algorithms (DNA) Group, Caltech

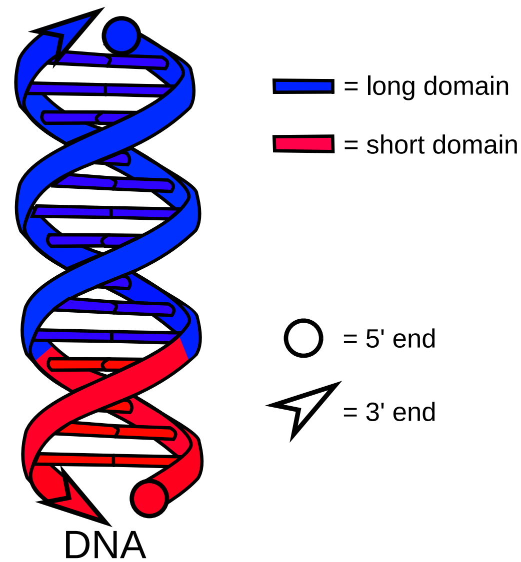
FNANO-19

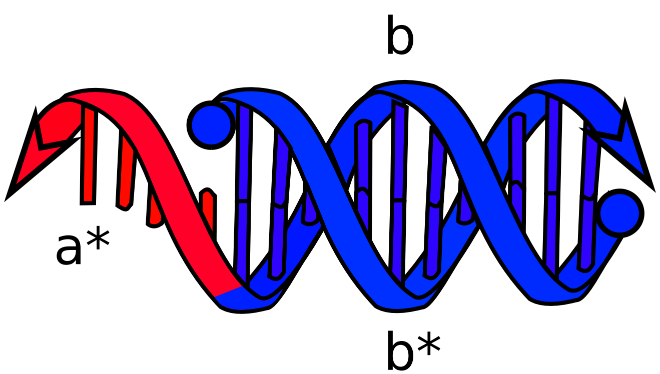
Snowbird, April 16^{th}, 2019

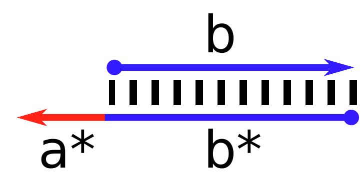
http://www.github.com/DNA-and-Natural-Algorithms-Group/peppercornenumerator

DNA STRAND DISPLACEMENT

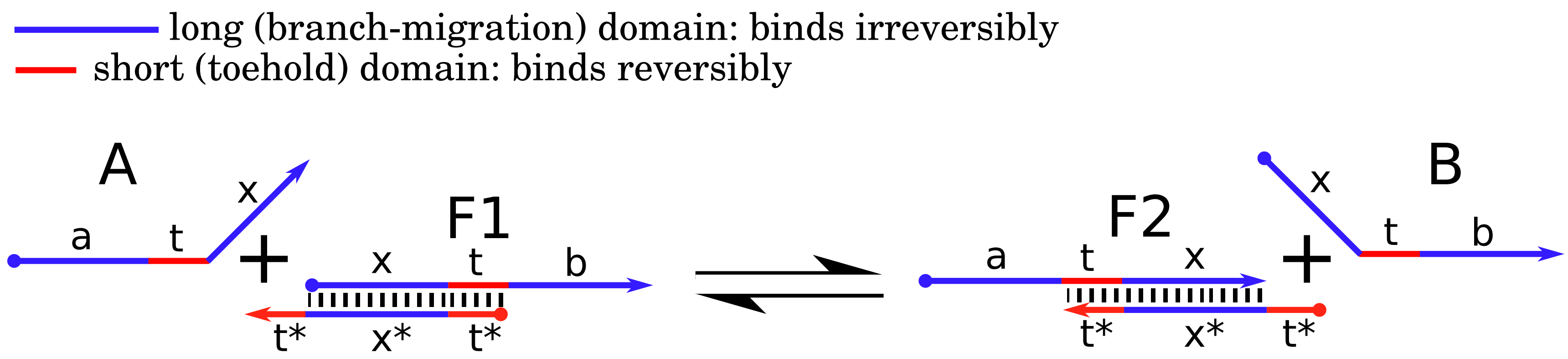




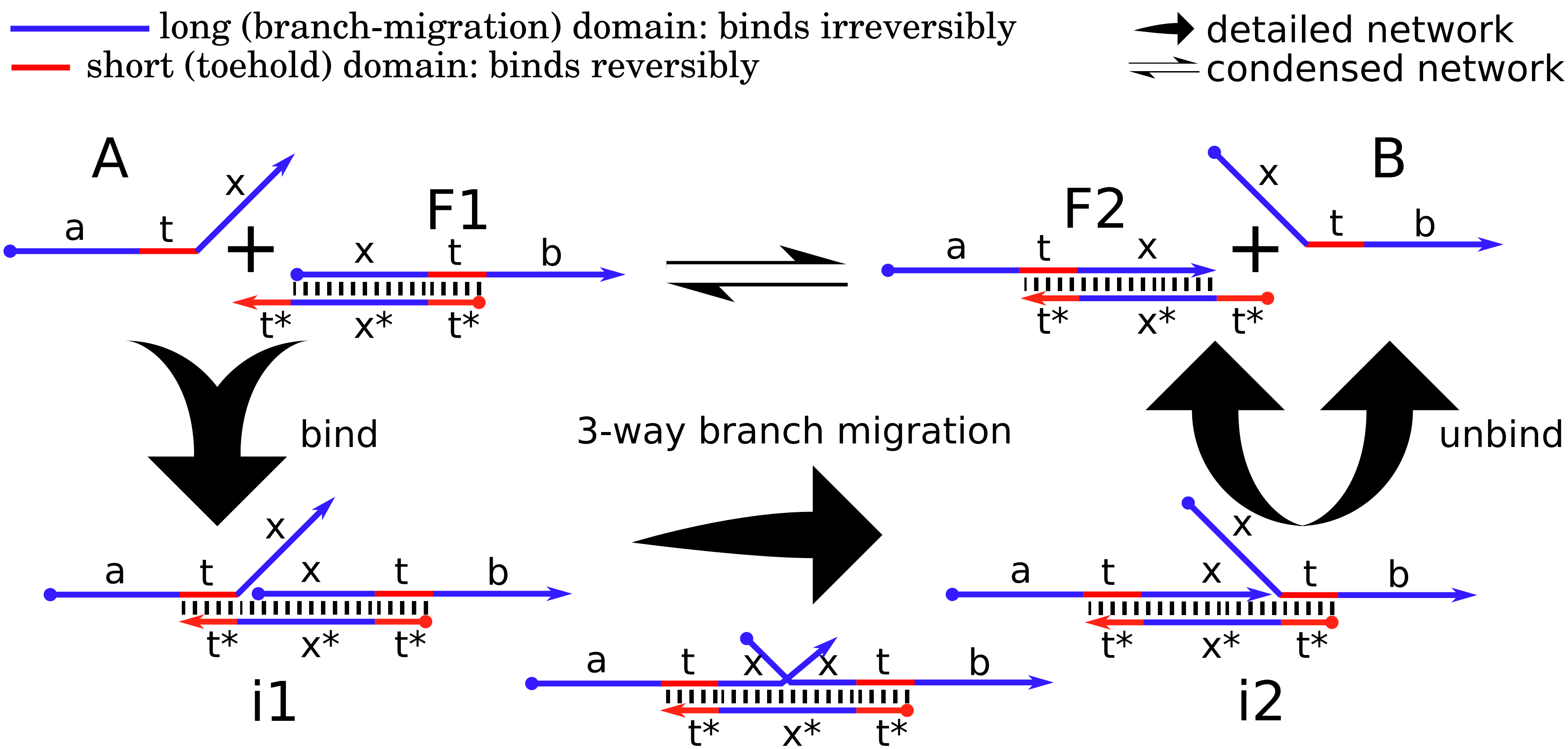




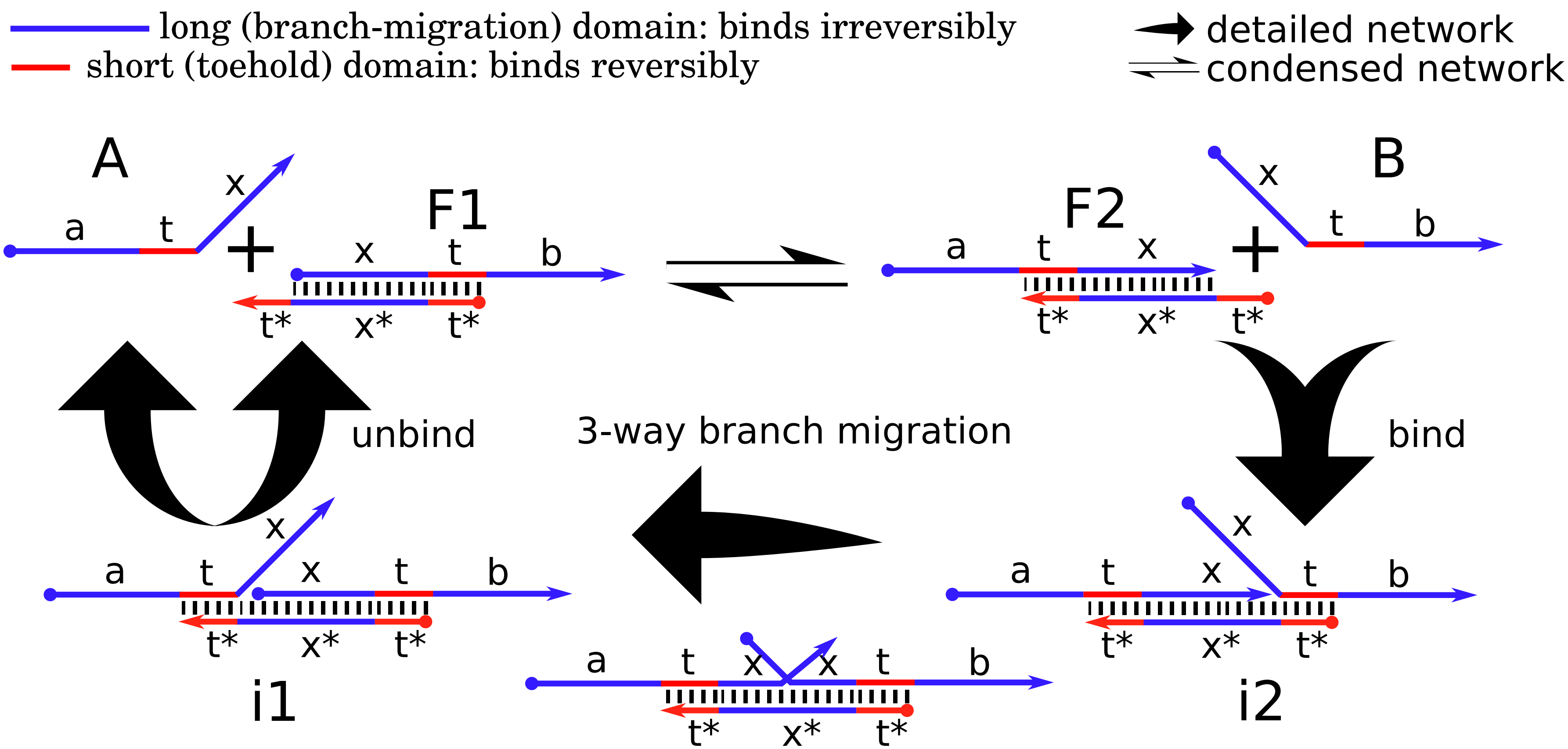
DOMAIN-LEVEL STRAND DISPLACEMENT



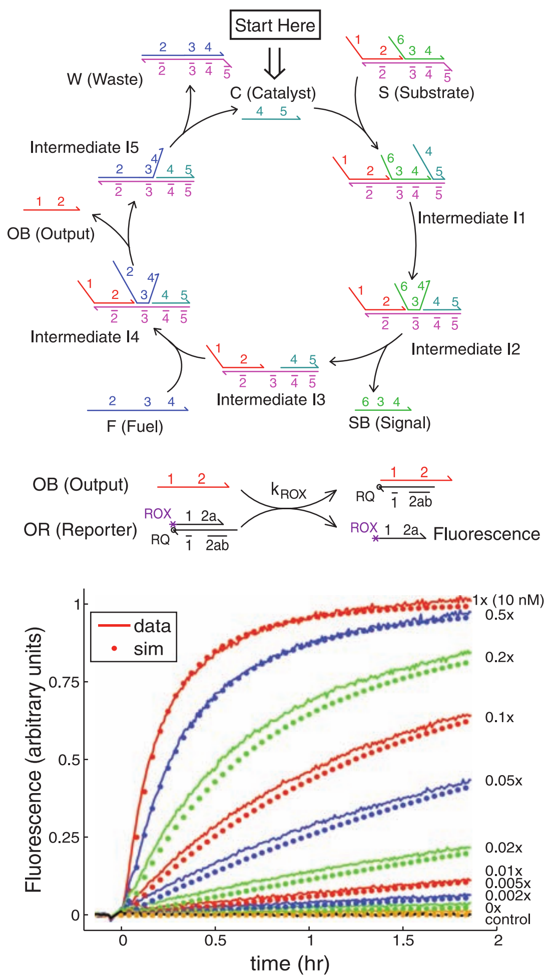
DOMAIN-LEVEL STRAND DISPLACEMENT



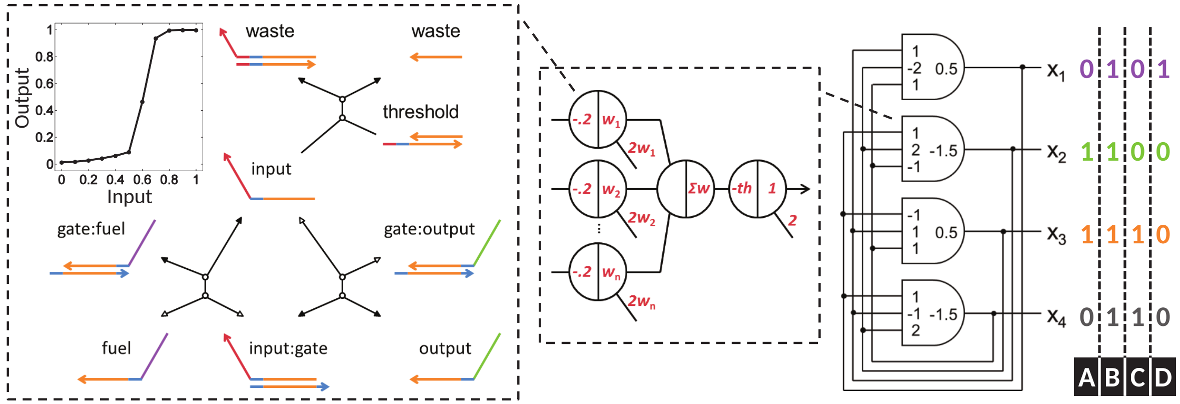
DOMAIN-LEVEL STRAND DISPLACEMENT



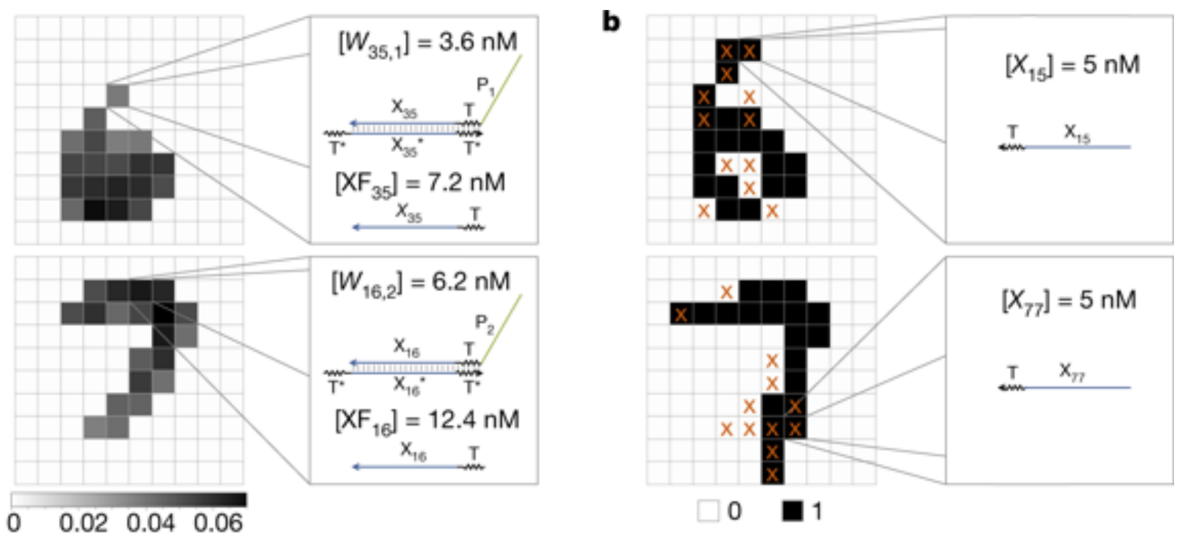
MANY EXPERIMENTAL DEMONSTRATIONS …



Zhang et al. (2007)



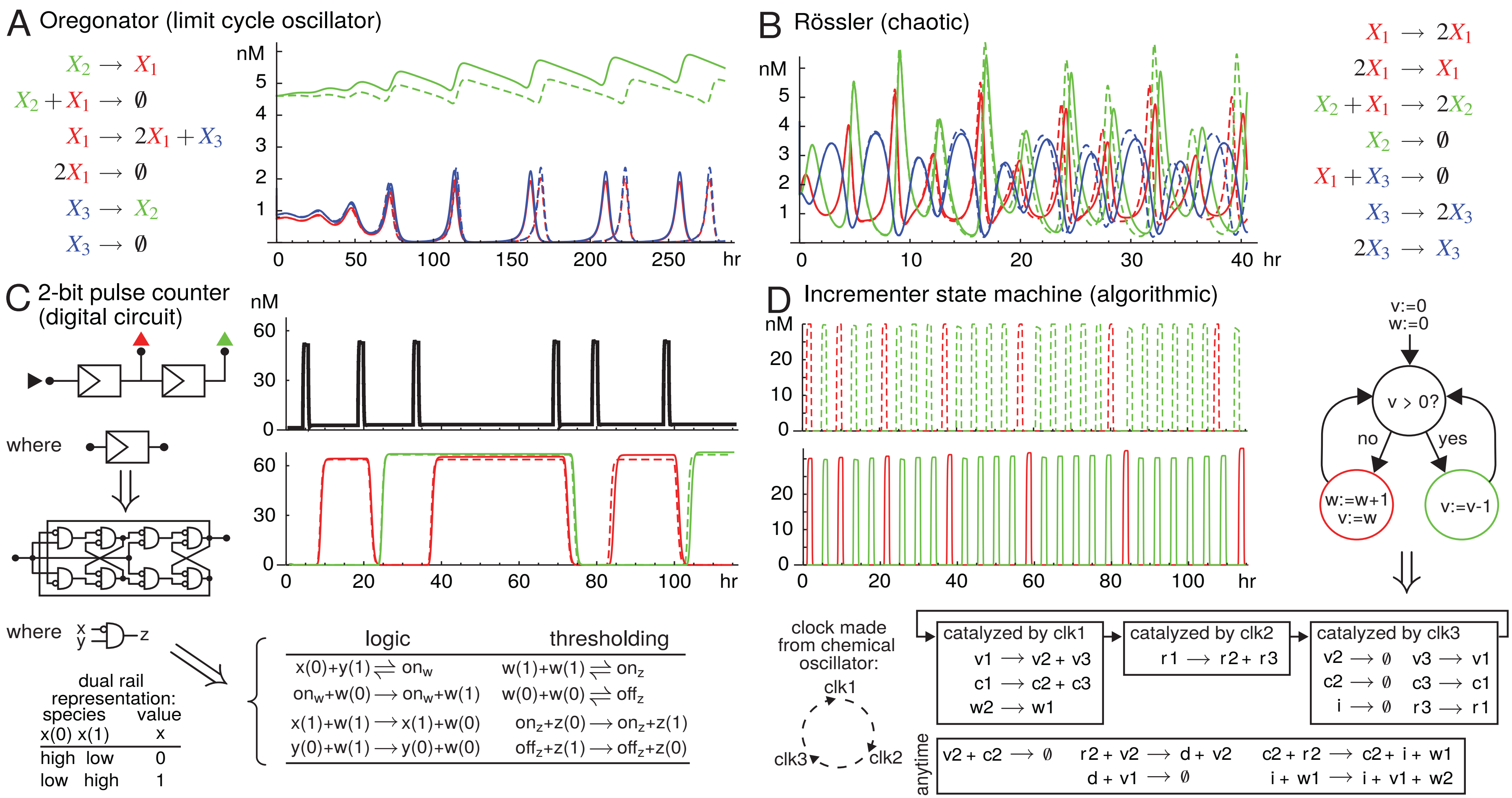
Qian, Winfree, Bruck (2011)



Cherry & Qian (2018)

... MANY MORE POTENTIAL APPLICATIONS.

Chemical Reaction Networks (CRNs)



Soloveichik et al. (2010) - DNA as a universal substrate for chemical kinetics

DSD IS A KINETIC TOOLBOX

... but how do you model your DSD system?

- per hand
- VisualDSD Phillips & Carelli (2009), ..., Sparcassi et al. (2018)
- other models Kawamata et al. (2012), Mokhtar et al. (2017), ... ?

You specify the reaction types. You specify the reaction rates. You may include all(?) types of pseudoknotted conformations, and even non-DSD reactions (e.g. enzyme cleavage reactions).

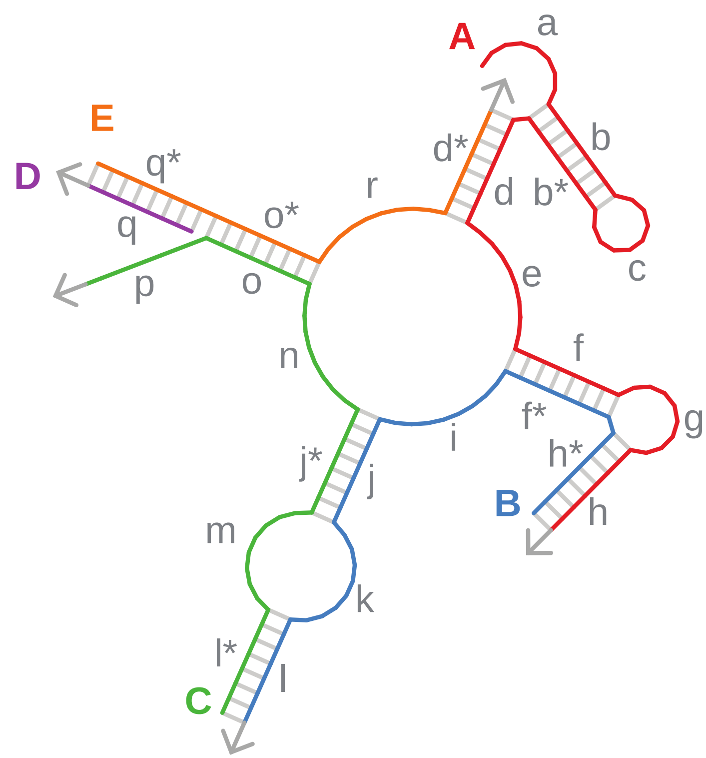
... so you better know what you are doing.

THERMODYNAMIC ENERGY MODEL

A secondary structure is a list of base pairs, where:

- A base may participate in at most one base pair
- Base pairs must not cross (no pseudoknots)
- Only specific base-pairs (GC, AT, GT) are allowed.

a)



b) "dot-bracket" or "dot-parens-plus" notation

```
a b c b* d e f g h + h* f* i j k l + l* m j* n o p + q + q* o* r d*
. ( . )  ( . ( . ( + )  )  . ( . ( + )  . )  . ( . + ( + )  )  . )
```

c) "kernel" notation

```
a b( c ) d( e f( g h( + ) ) i j( k l( + ) m ) n o( p + q( + ) ) r )
```

DSD IS A KINETIC TOOLBOX ...

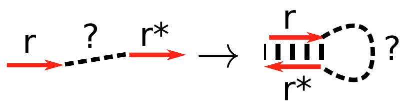
... that can be rigorously analyzed
within the domain of the thermodynamic energy model.

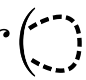
The **Peppercorn** software package:

- reaction enumeration
- reaction condensation
- approximate DNA reaction rate model

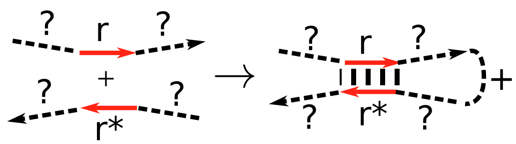
REACTION TYPES & APPROXIMATE RATES 1/2

`bind11:`



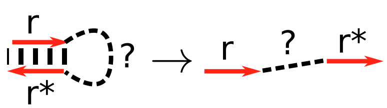
$k_{\text{bind11}} = Pr($$) * 10^{6}\,\mathrm{s}^{-1}$

`bind21:`



$$k_{\text{bind21}} = |r| * 3 * 10^{5}\,\mathrm{M}^{-1}\mathrm{s}^{-1}$$

`open:`



$$\frac{k_{\text{open}}}{k_{\text{bind21}}} = e^{\frac{\Delta G}{RT}}$$

Open reactions only for toeholds with parameter: L, k_{slow}

`bind21` is the only valid bimolecular reaction

REACTION TYPES & APPROXIMATE RATES 2/2

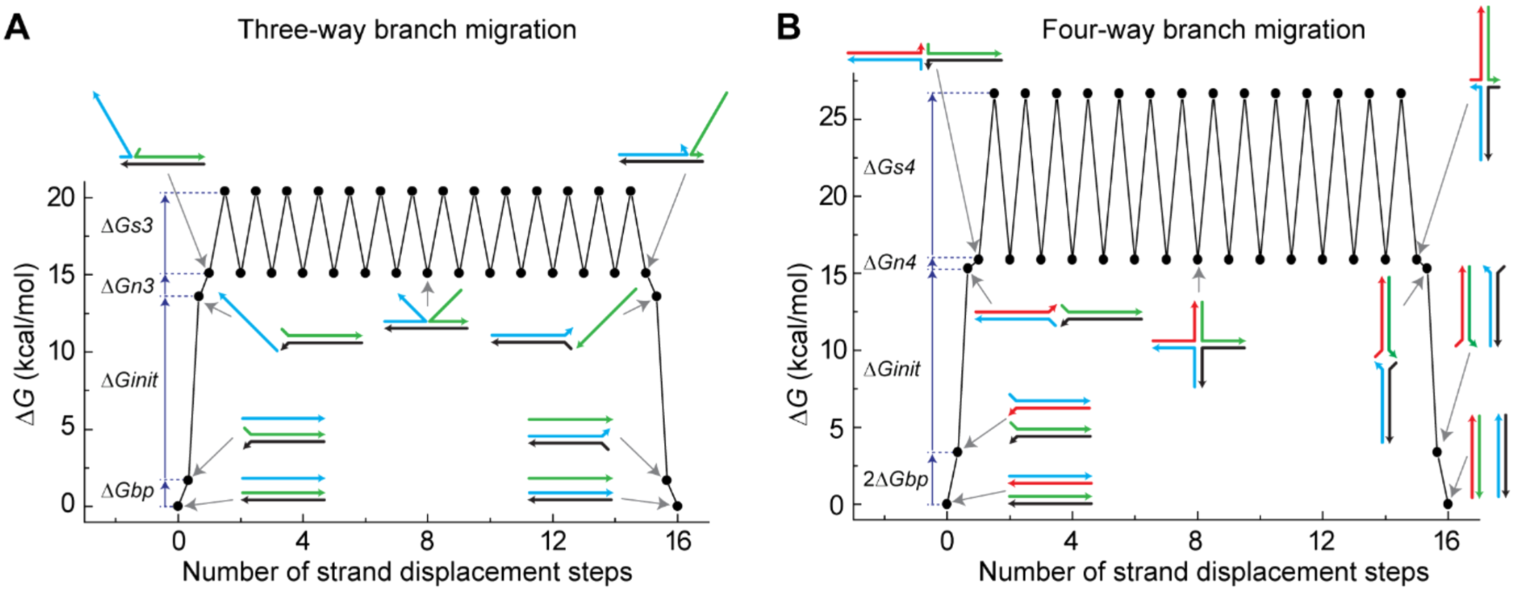
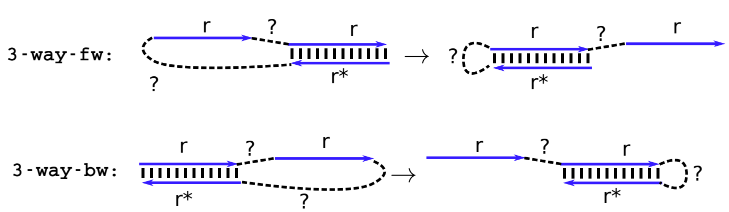
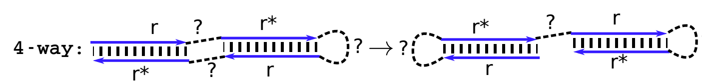


Figure from Kotani & Hughes (2017)





$$k_{\text{bm}} = \frac{Pr(\bigcirc ?) * k_{\text{bm-init}}}{|r|} \text{ s}^{-1}$$

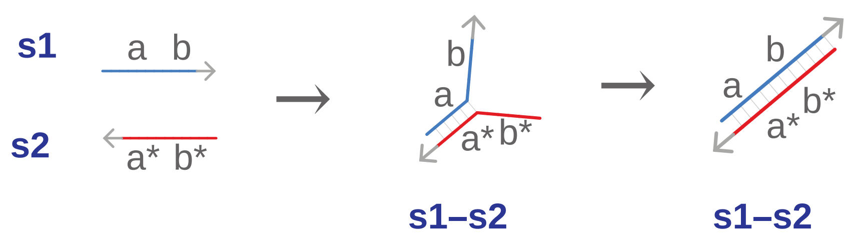
unimolecular, but may lead to dissociation

WHAT ARE THE CHALLENGES?

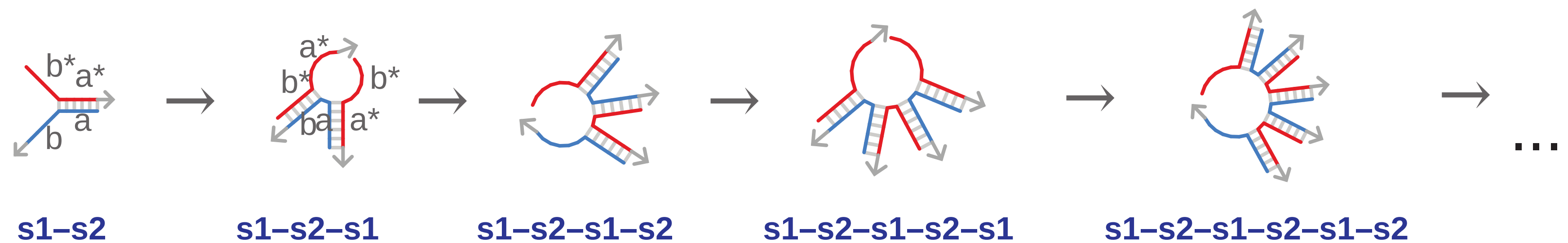
- polymerization
  => timescale separation
- size of the enumerated network
  => condensation

POLYMERIZATION

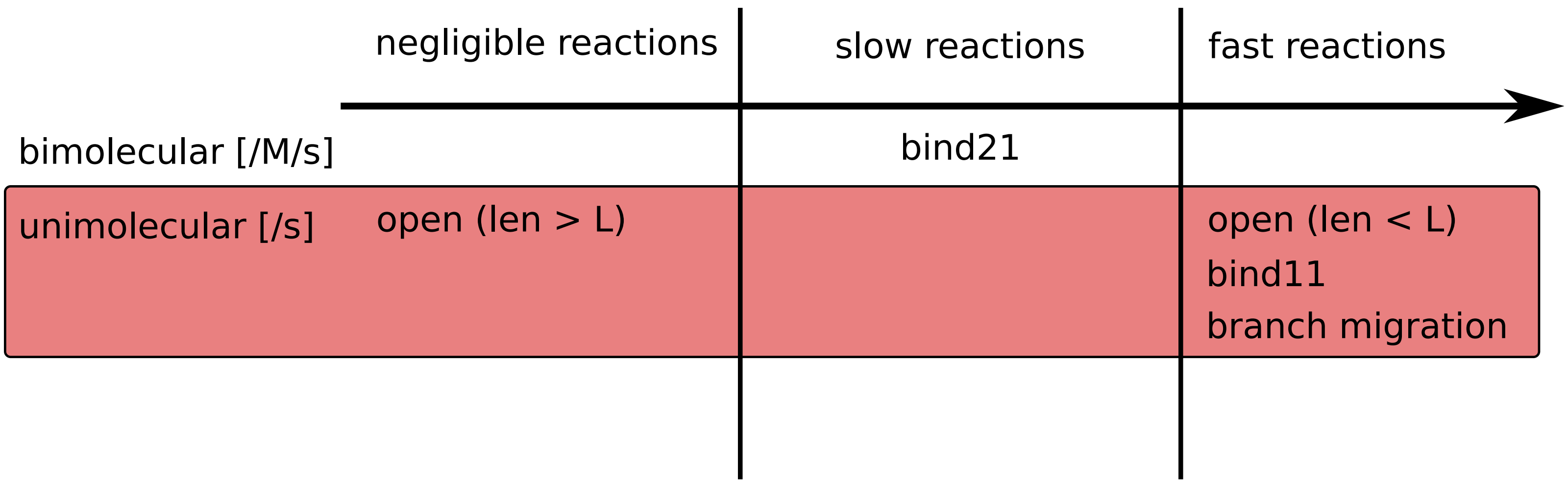
a)



b)

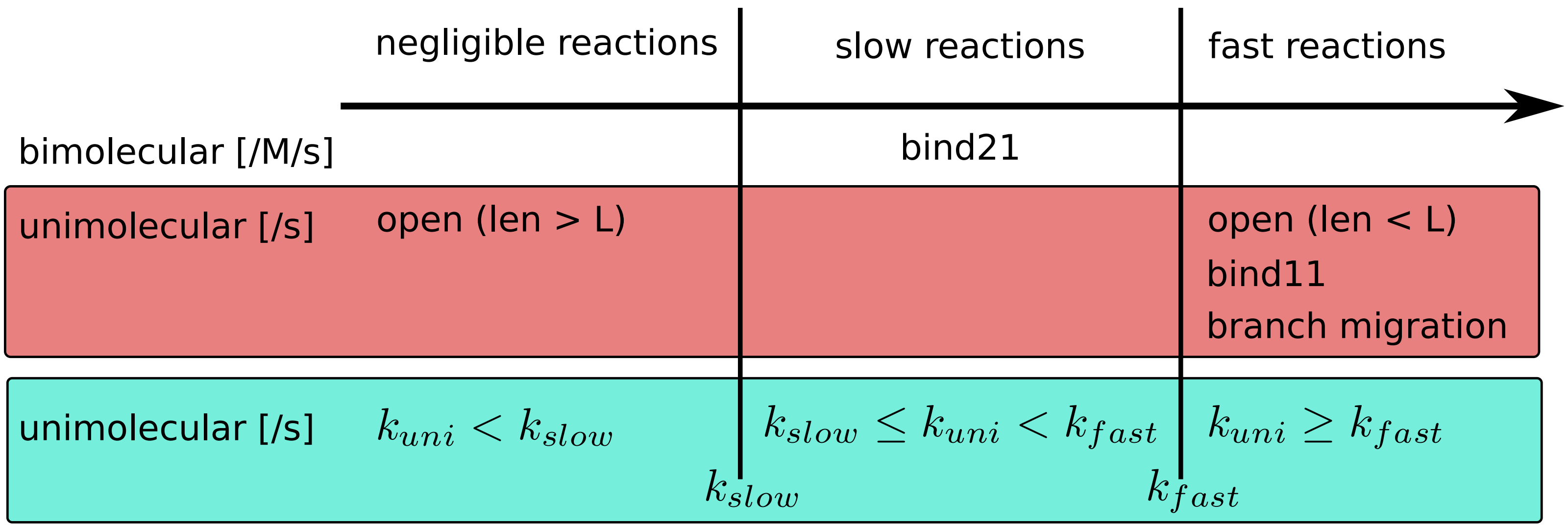


MODEL PARAMETERS



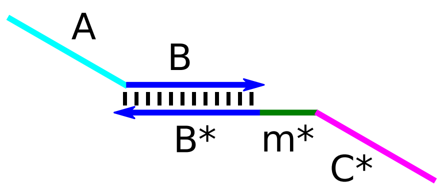
rate-independent model: simple, one parameter: L

MODEL PARAMETERS

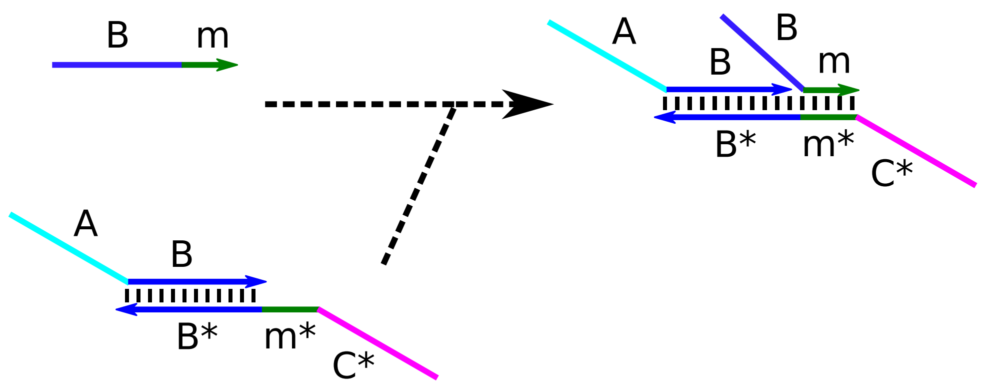


- rate-independent model: simple, one parameter: L
- rate-dependent model: flexible, two parameters: k-slow, k-fast

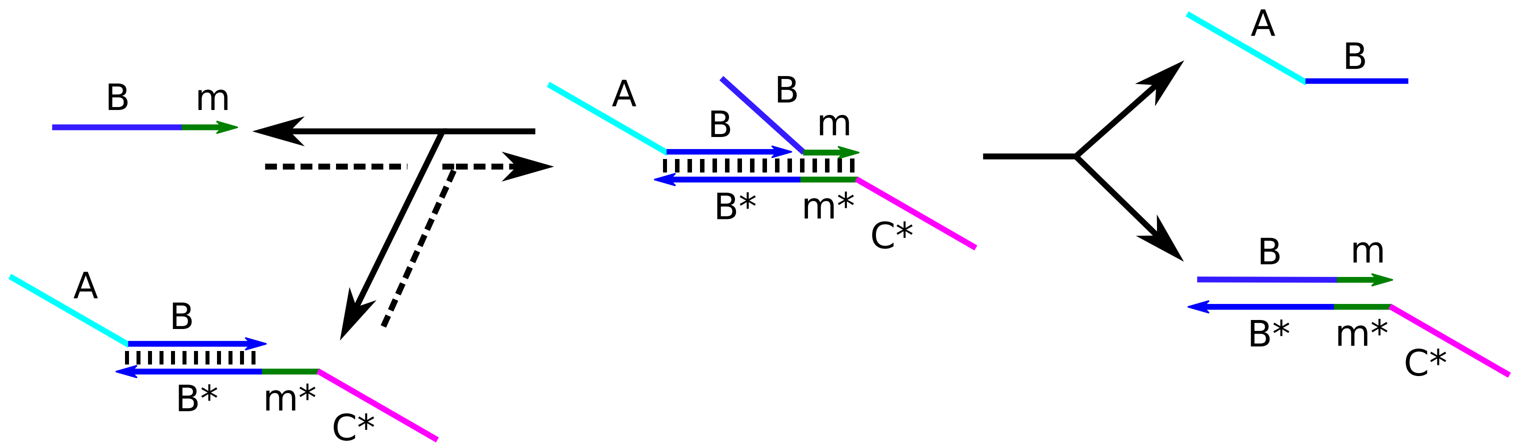
ENUMERATION / CONDENSATION



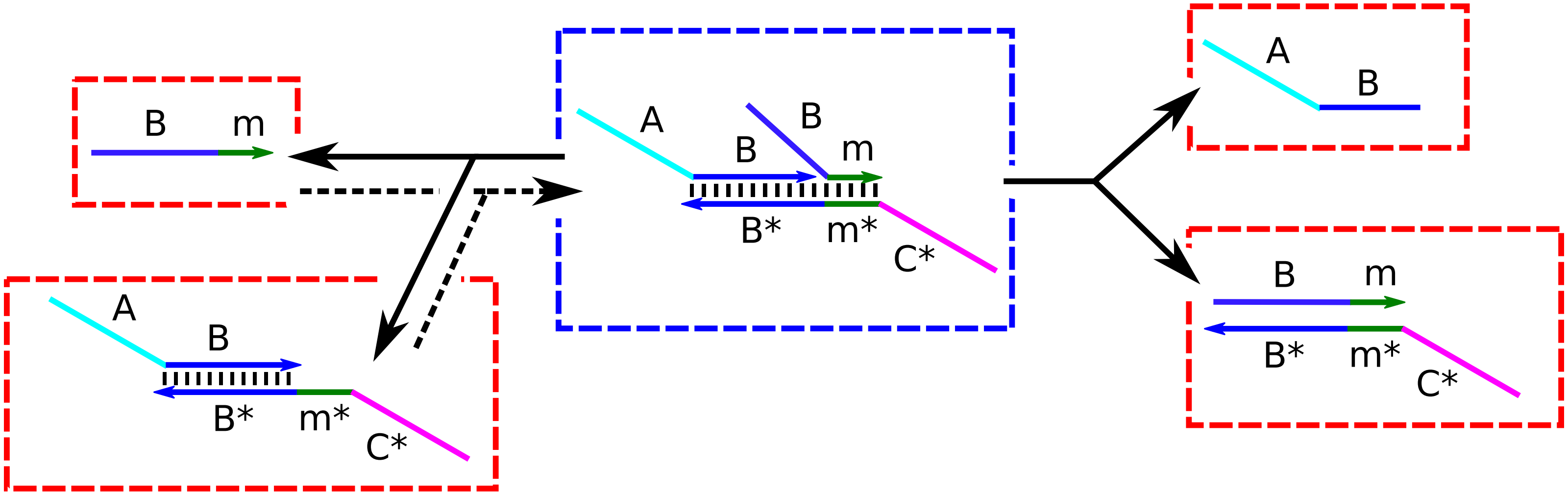
ENUMERATION / CONDENSATION



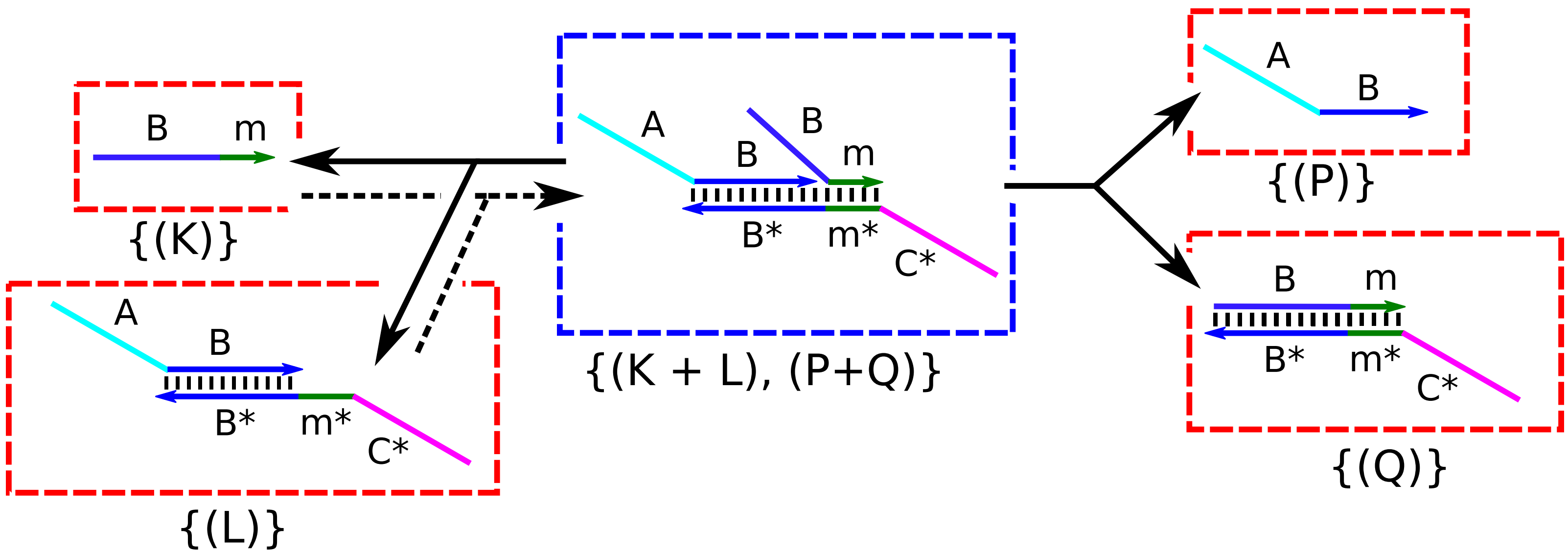
ENUMERATION / CONDENSATION

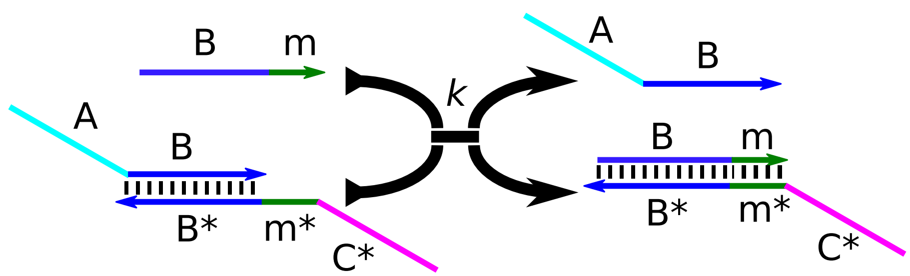


ENUMERATION / CONDENSATION

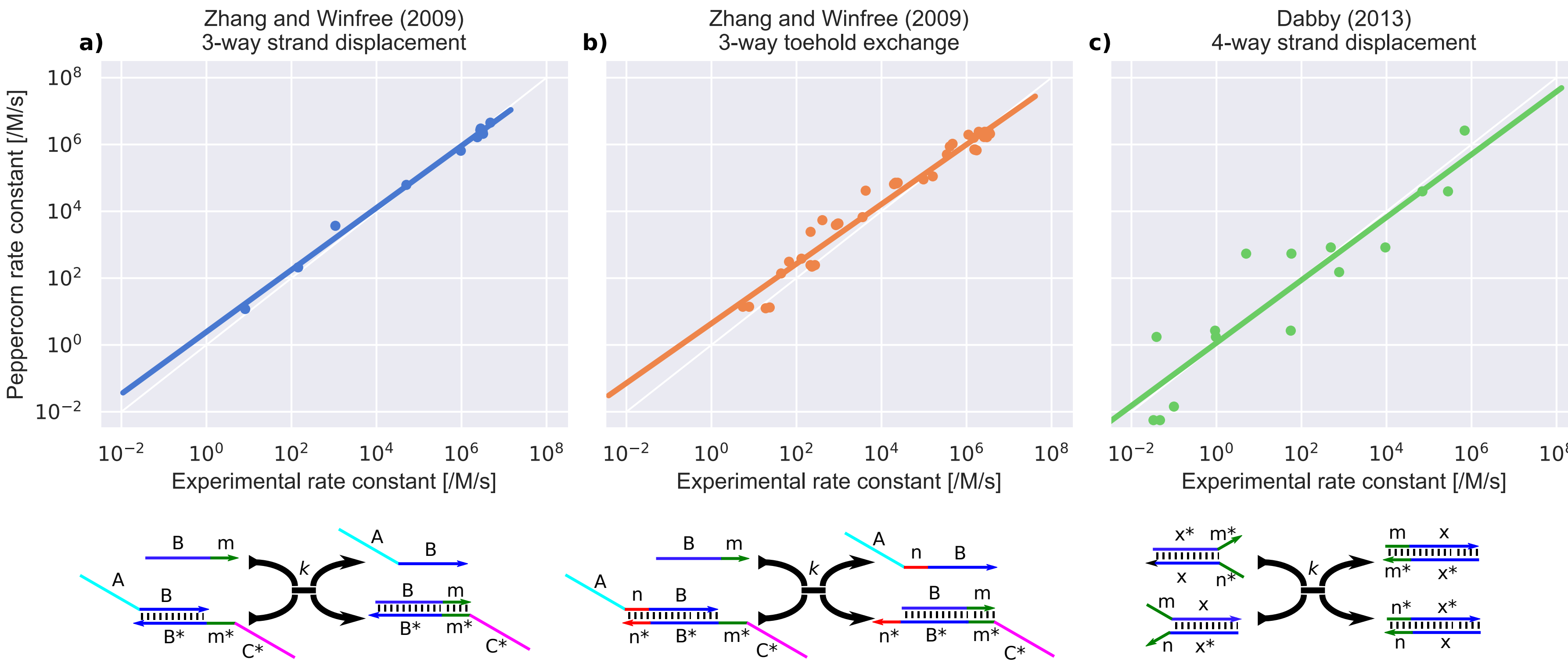


ENUMERATION / CONDENSATION

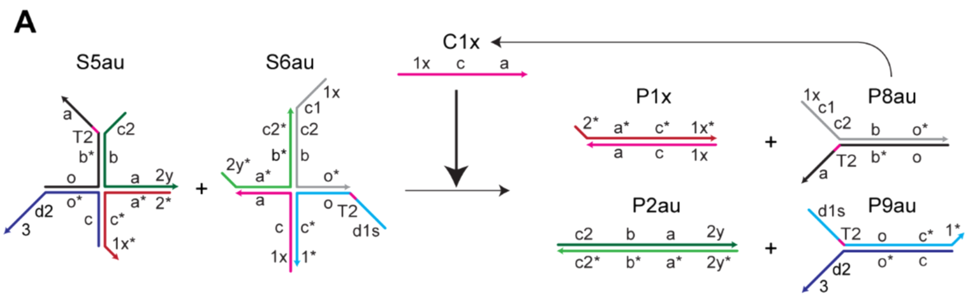




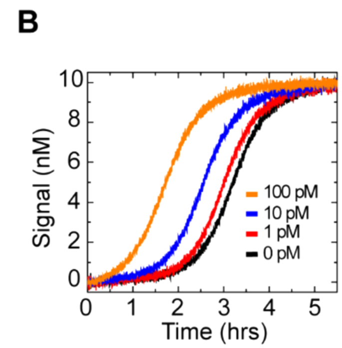
CASE STUDIES: CONDENSED REACTION RATES



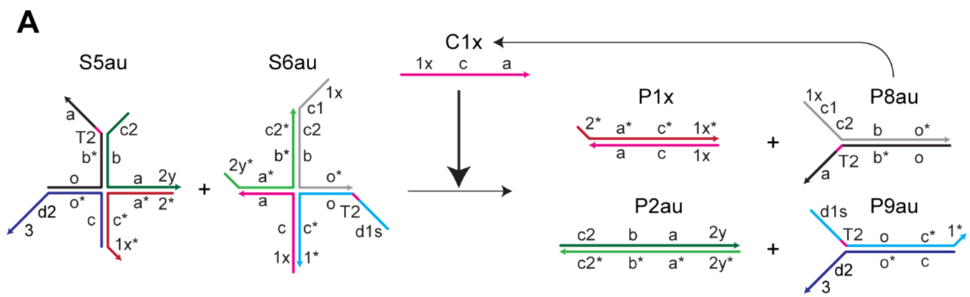
CASE STUDIES: AUTOCATALYTIC SYSTEM

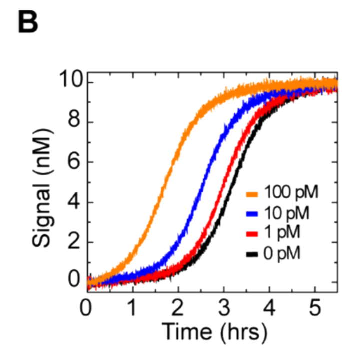


Kotani & Hughes (2017)

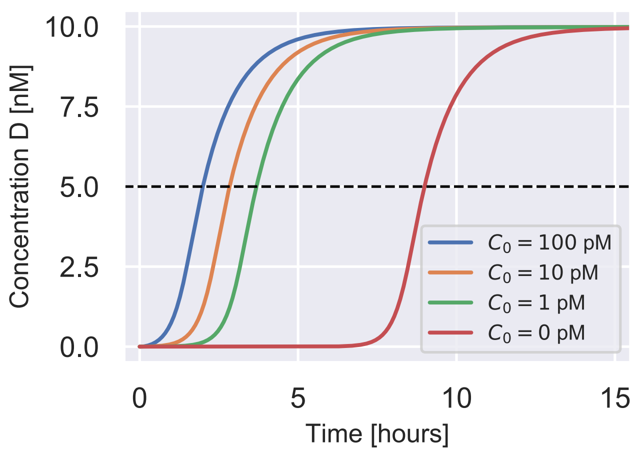


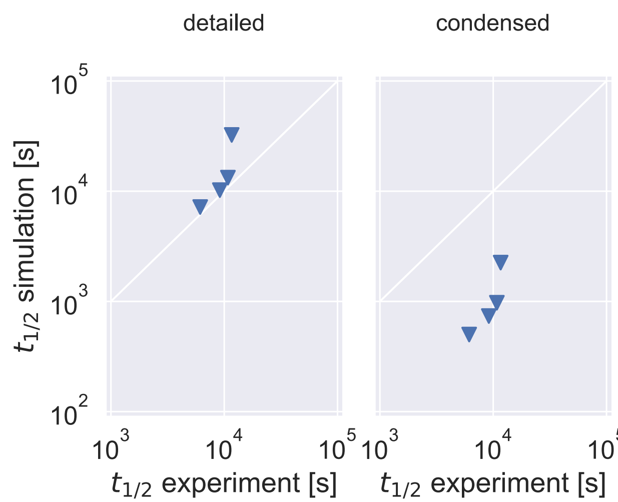
CASE STUDIES: AUTOCATALYTIC SYSTEM





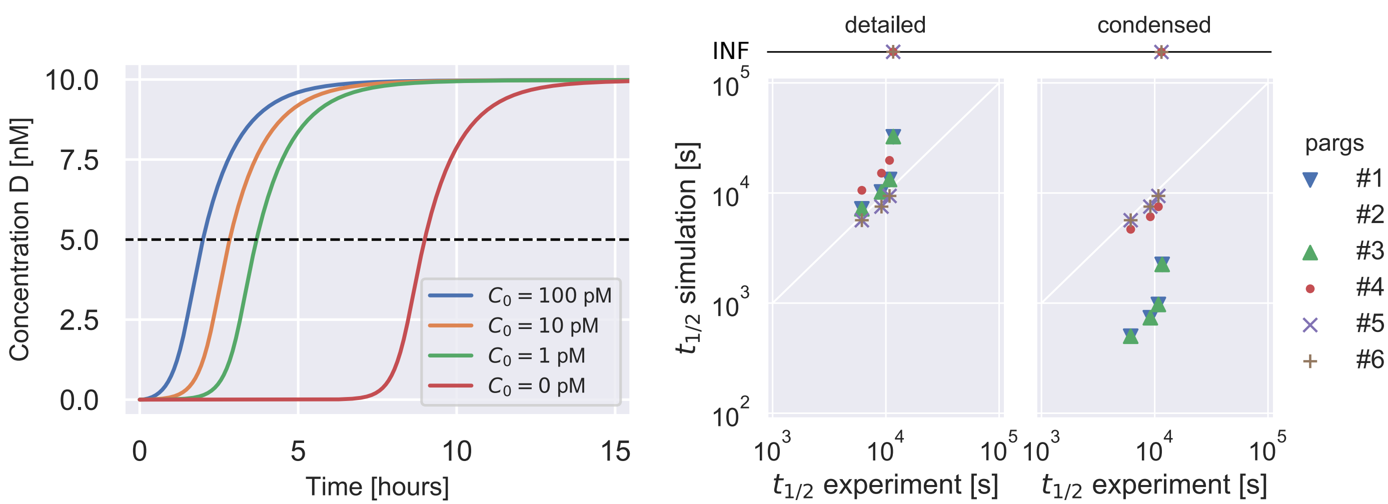
Kotani & Hughes (2017)



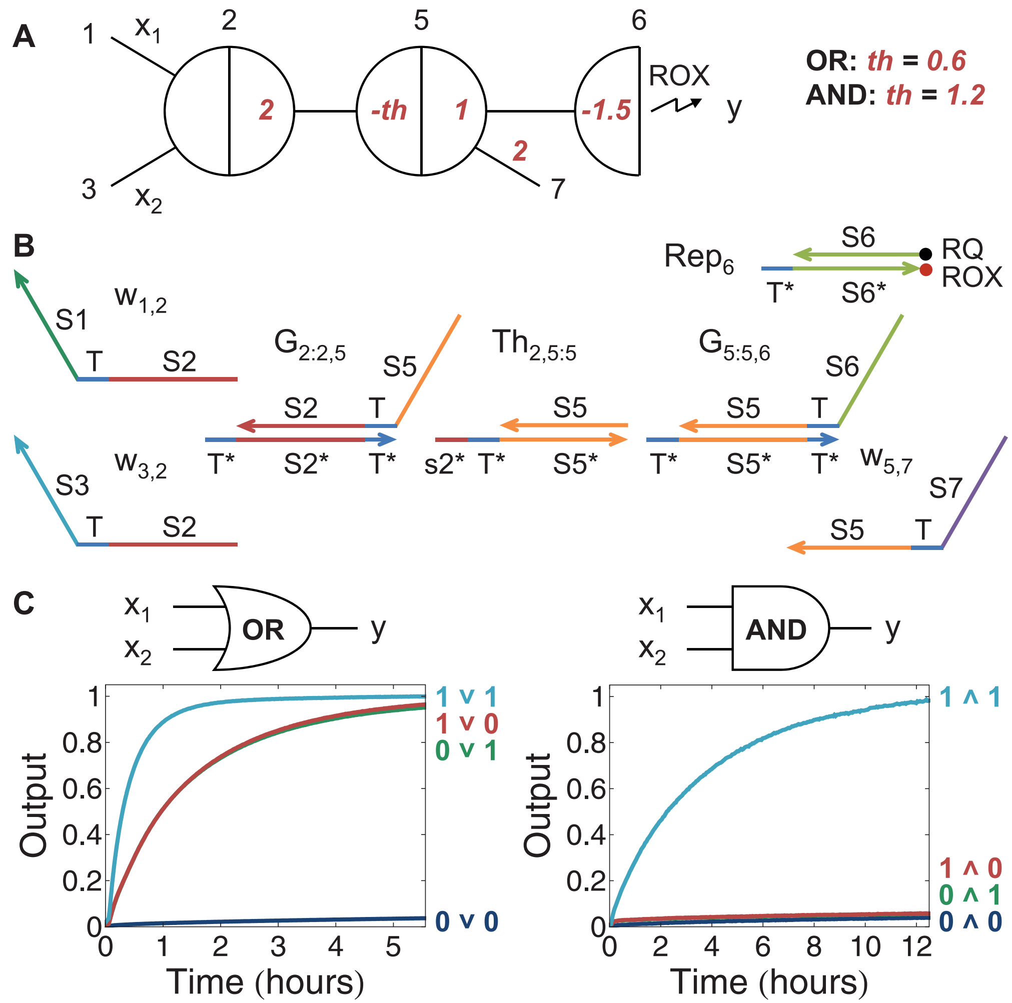


CASE STUDIES: AUTOCATALYTIC SYSTEM

#	parameters	MCS	RC + TC	DR	RM	CR
1	release-cutof = 8	10	16 + 159	556	14	15
2	$k_{slow} = k_{fast} = 10^{-3}$	16	13 + 82	265	10	11
3	$k_{slow} = k_{fast} = 10^{-4}$	10	16 + 164	599	14	15
4	$k_{slow} = 10^{-4}$, $k_{fast} = 10^{-3}$	16	20 + 164	488	17	22
5	$k_{slow} = 10^{-4}$, $k_{fast} = 10^{-2}$	24	55 + 1426	6628	28	62
6	$k_{slow} = 10^{-5}$, $k_{fast} = 10^{-2}$	24	55 + 1426	6652	28	75

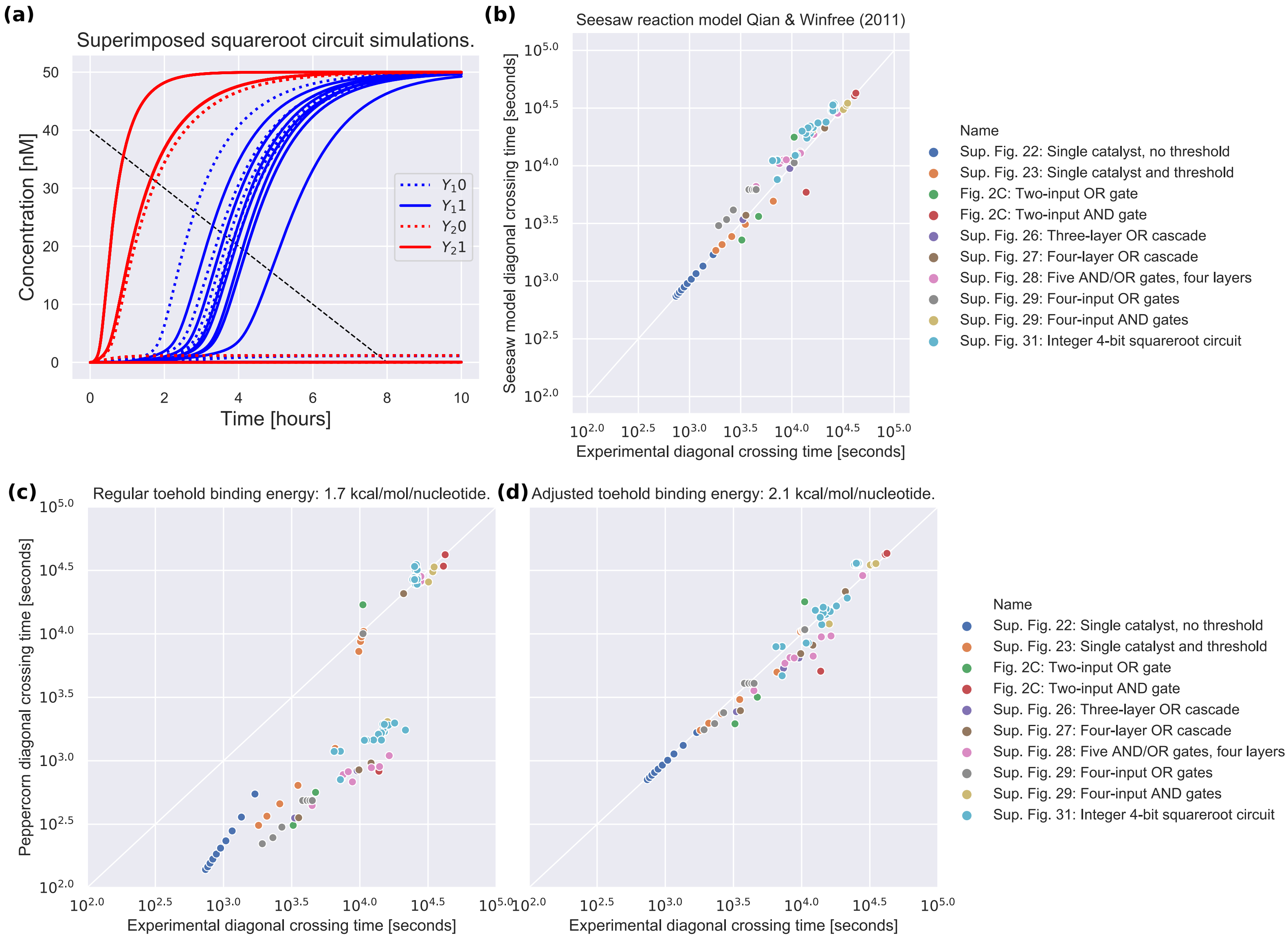


CASE STUDIES: SEESAW SYSTEMS

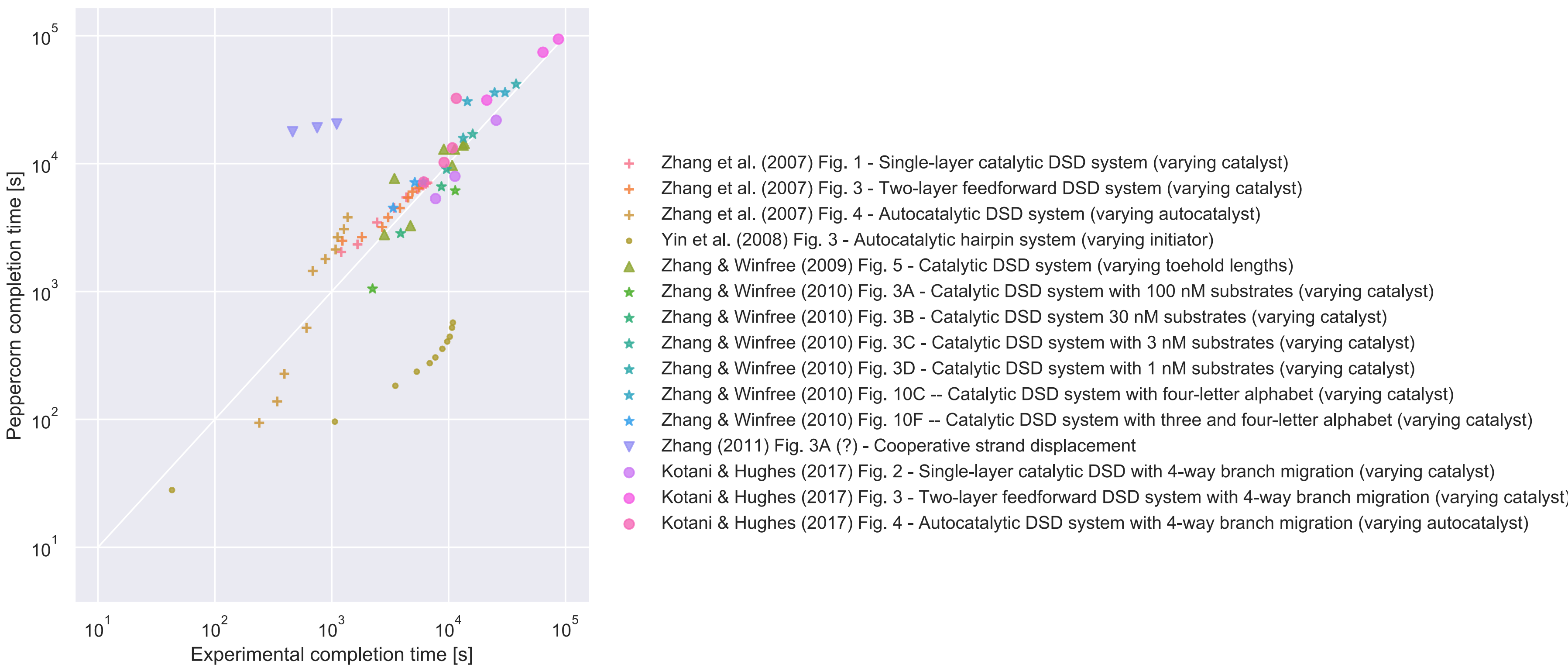


Qian & Winfree (2011)

CASE STUDIES: SEESAW SYSTEMS



CASE STUDIES: MANY SYSTEMS

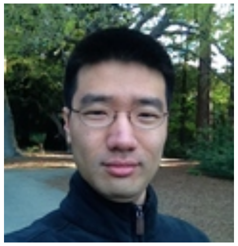


THANKS TO

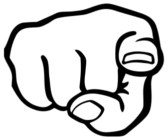
Erik Winfree

Casey Grun

Karthik Sarma



Seung Woo Shin



you

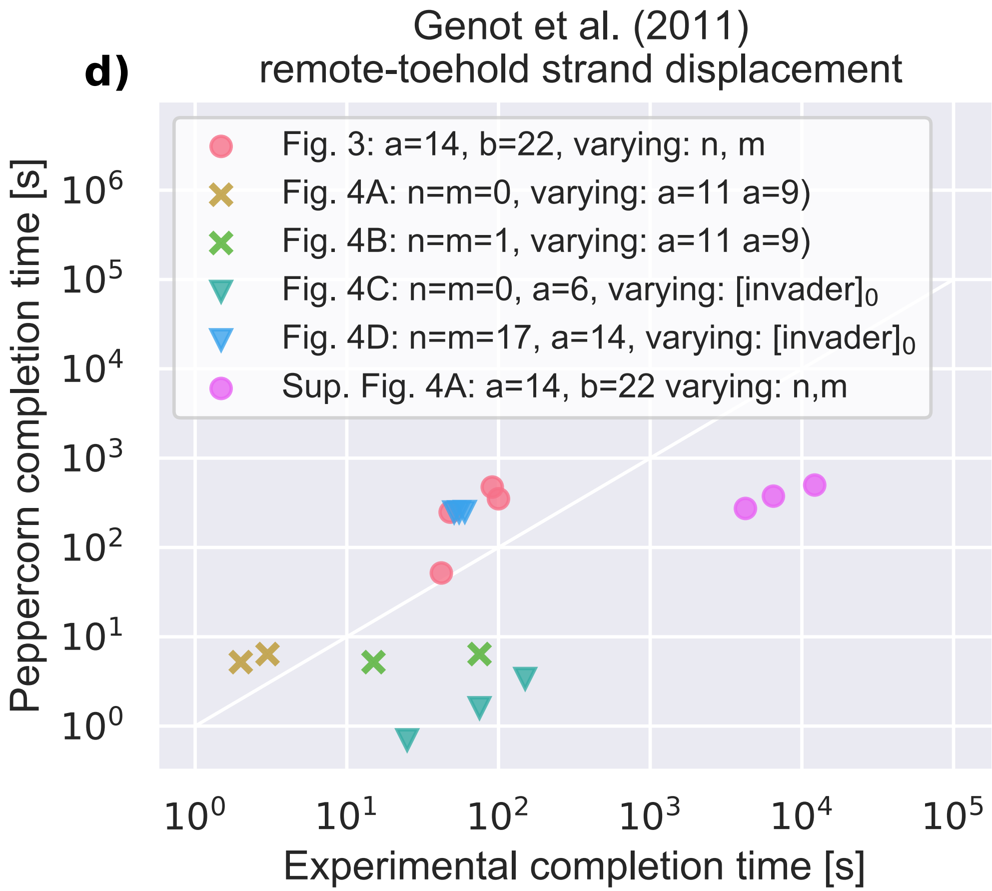
Brian Wolfe

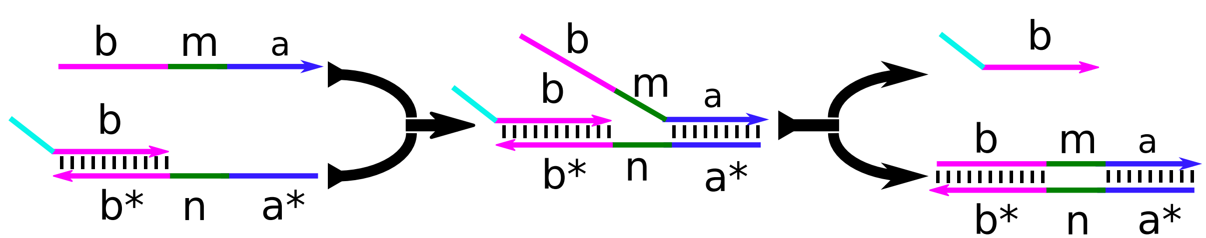
http://www.github.com/DNA-and-Natural-Algorithms-Group/peppercornenumerator

... don't forget to ask me about kernel notation.

This research was funded in parts by:
The Caltech Biology and Biological Engineering Division Fellowship.
The U.S. National Science Foundation NSF Grant CCF-1213127 and NSF Grant CCF-1317694.
The Gordon and Betty Moore Foundation's Programmable Molecular Technology Initiative (PMTI).

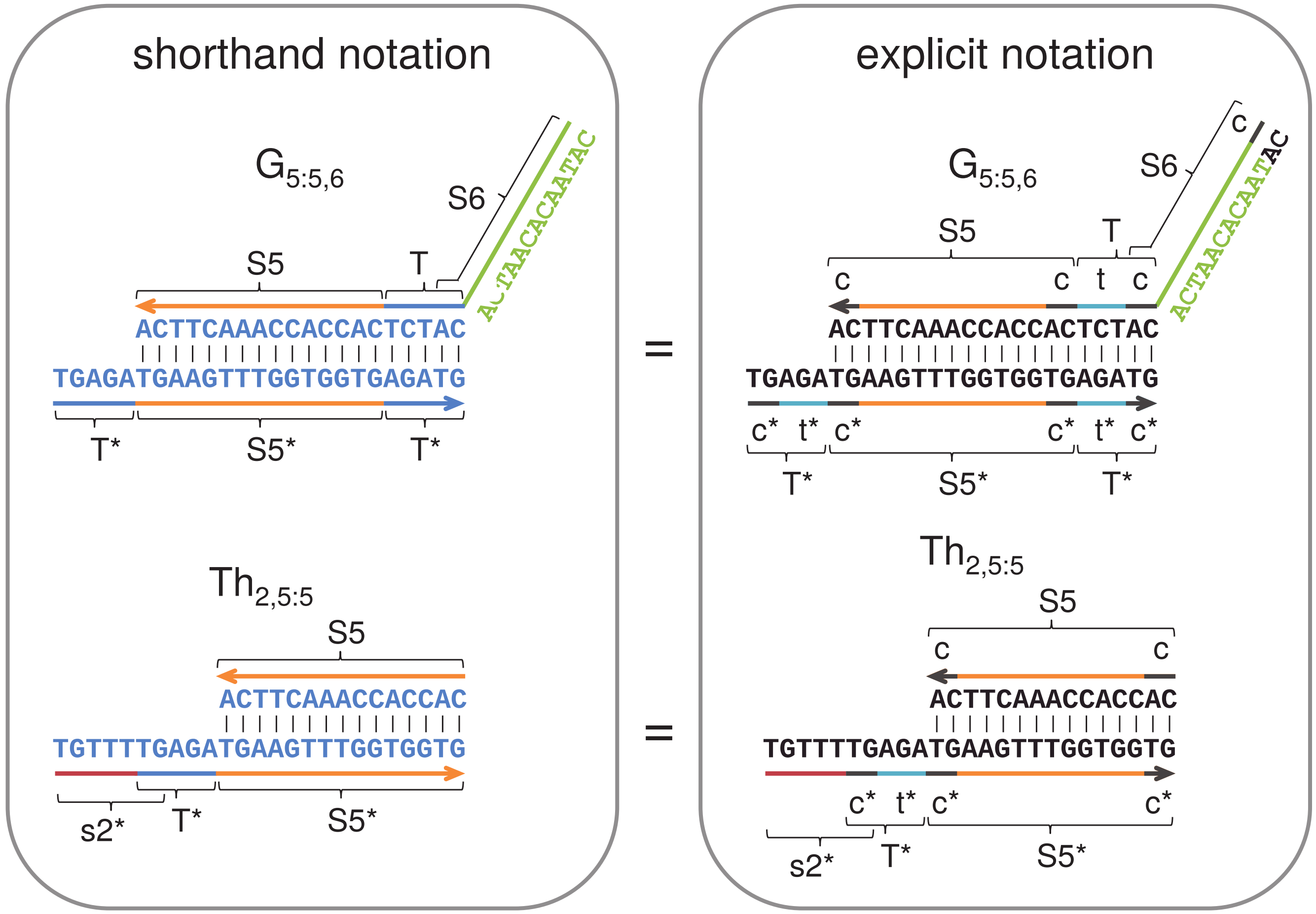
CASE STUDIES: REACTION COMPLETION



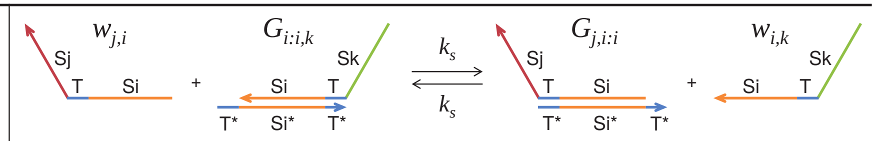
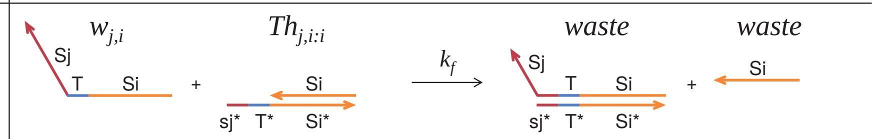
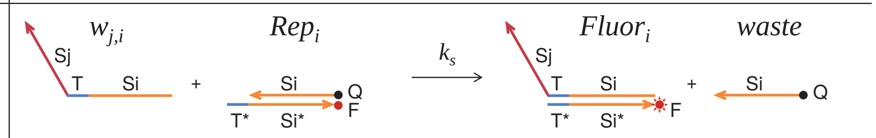
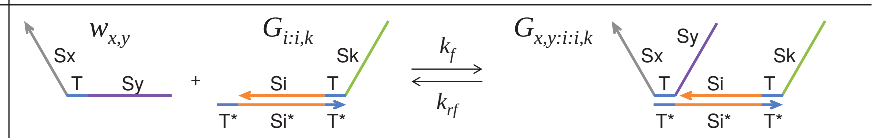
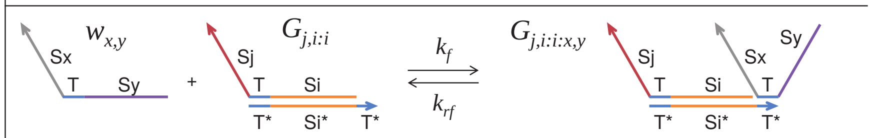
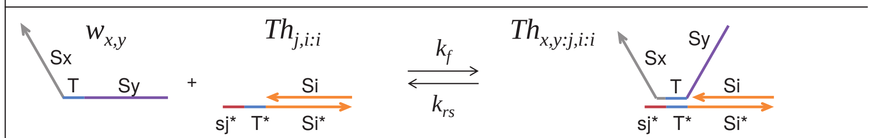
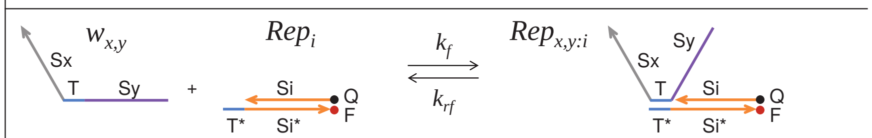
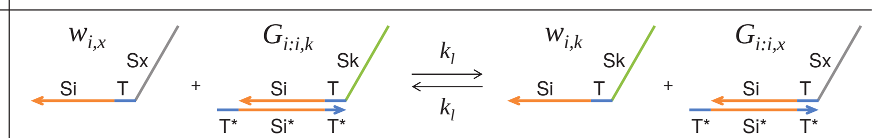
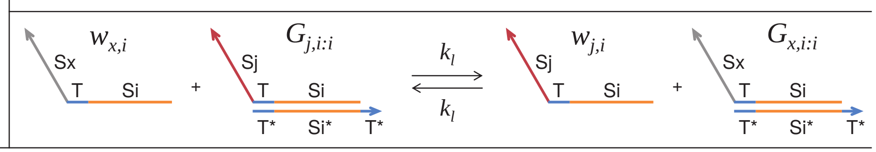


CASE STUDIES: SEESAW SYSTEMS

Default

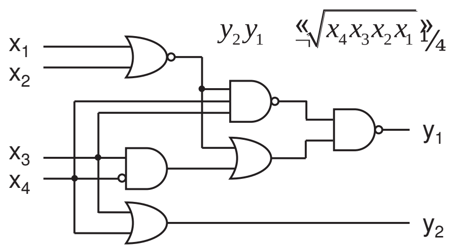


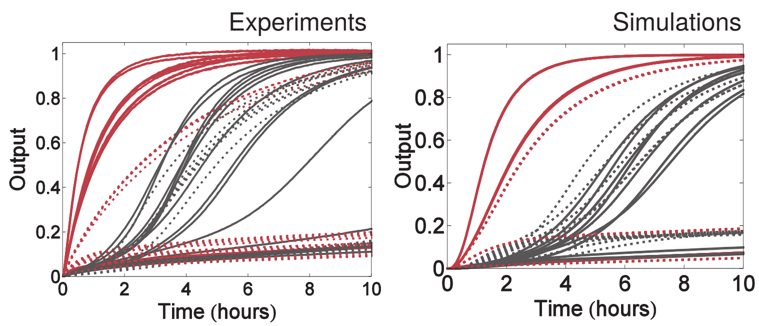
Qian & Winfree (2011) - Supporting Online Material

Designed reactions	Seesawing reactions	
	Thresholding reactions	
	Reporting reactions	
Side reactions	Universal toehold binding reactions	
		
		
		
	Leak reactions	
		

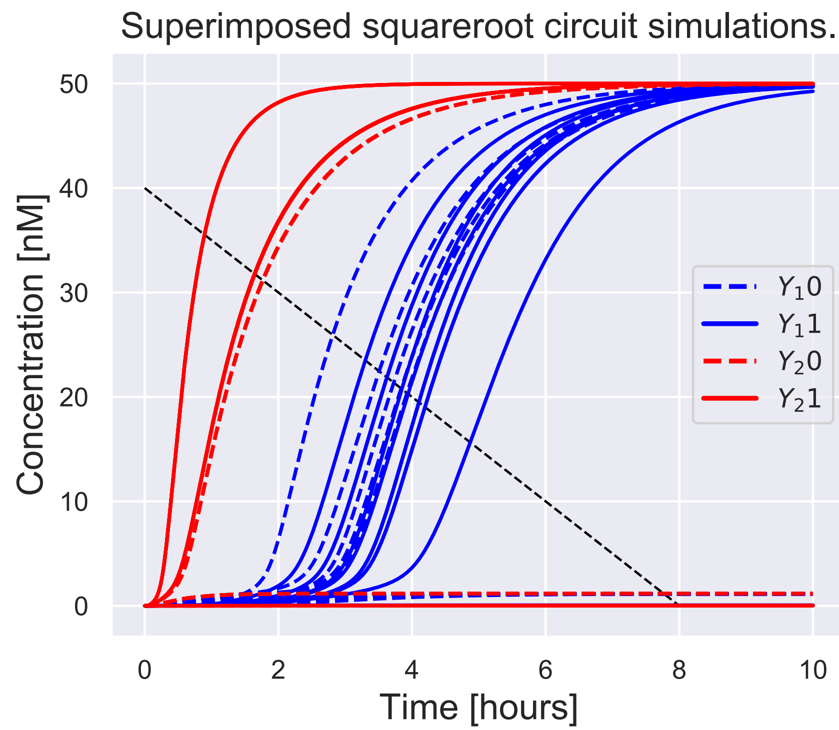
Qian & Winfree (2011) - Supporting Online Material

Temperature: 25 °C		
k_s	=	$5 \times 10^4\ \mathrm{M}^{-1}\mathrm{s}^{-1}$
k_f	=	$2 \times 10^6\ \mathrm{M}^{-1}\mathrm{s}^{-1}$
k_{rs}	=	$1.3\ \mathrm{s}^{-1}$
k_{rf}	=	$26\ \mathrm{s}^{-1}$
k_l	=	$10\ \mathrm{M}^{-1}\mathrm{s}^{-1}$

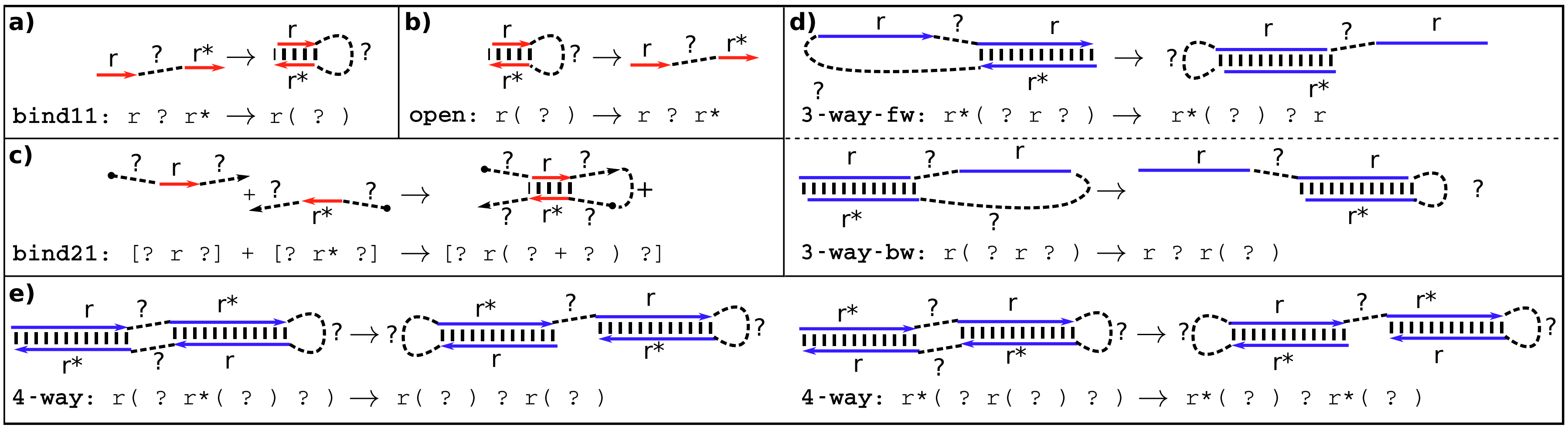




Qian & Winfree (2011) - Supporting Online Material

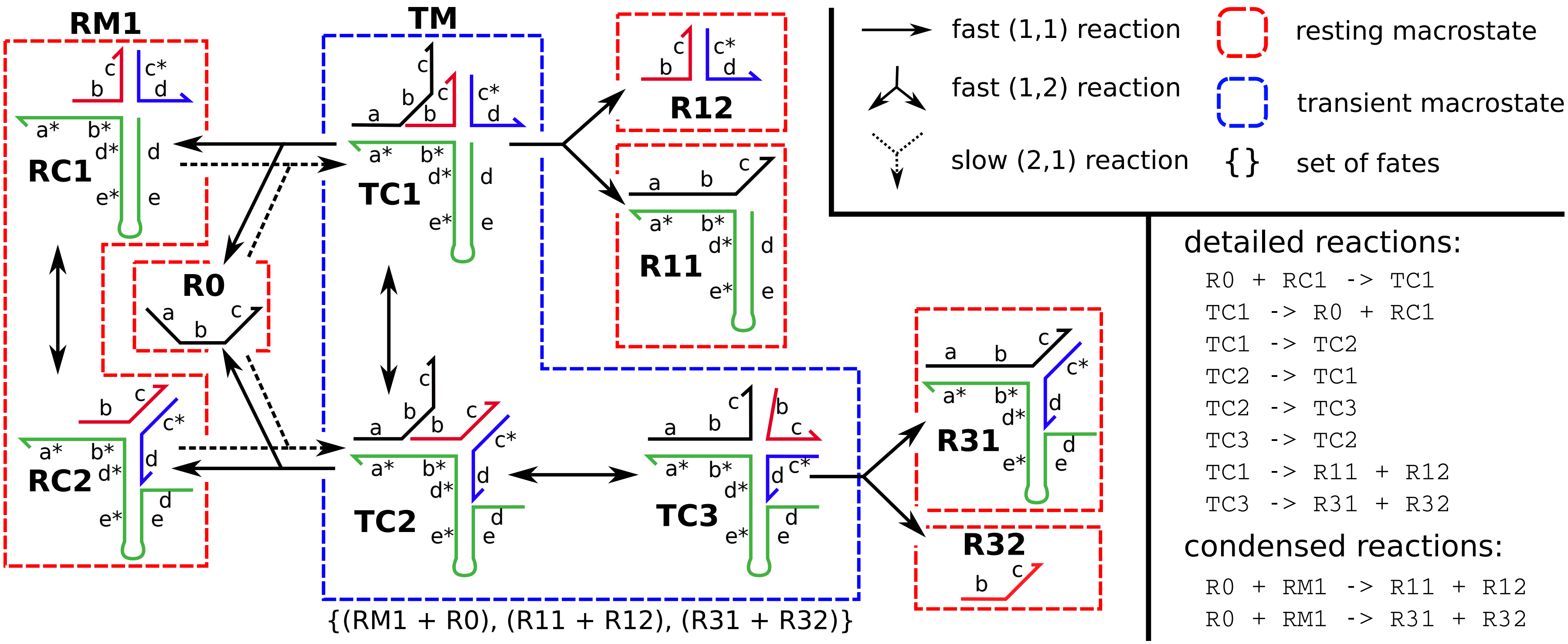


REACTION RULES



x(u(y)) t* = <x!j u!k y u!k x!j t*\>

CRN CONDENSATION



REACTION ENUMERATION

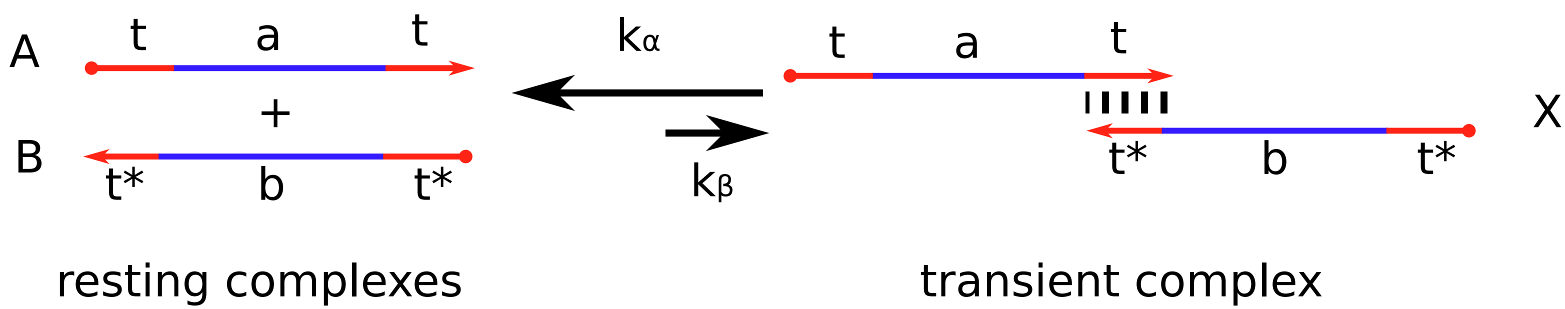
- all initial complexes are included
- every complex has all valid fast reactions enumerated
- transient complexes have no slow reactions enumerated
- resting complexes have all valid slow reactions enumerated

valid according to enumeration semantics:

- rate-dependent model
- rate-independent model
- max-helix semantics: reaction types are greedy
- reject-remote semantics: exclude remote-toehold branch migration

SEPARATION OF TIMESCALES

unimolecular reactions are fast
bimolecular reactions are slow



$$\{X \xrightarrow{k_\alpha} A + B;\ A + B \xrightarrow{k_\beta} X\}$$

at low concentrations:

$$k_\beta[A][B] << k_\alpha[X]$$