# A GENERAL-PURPOSE CRN-TO-DSD COMPILER FRAMEWORK WITH FORMAL VERIFICATION, OPTIMIZATION, AND SIMULATION CAPABILITIES

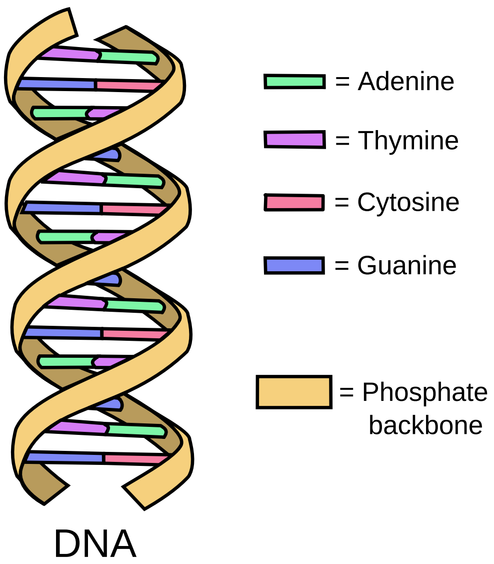
Stefan Badelt

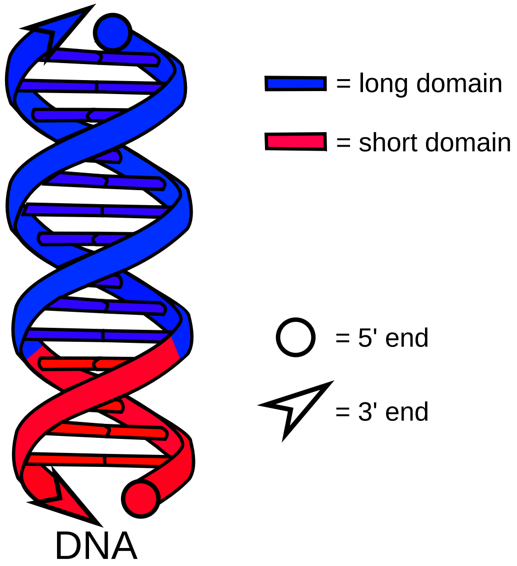
DNA and Natural Algorithms (DNA) Group, Caltech

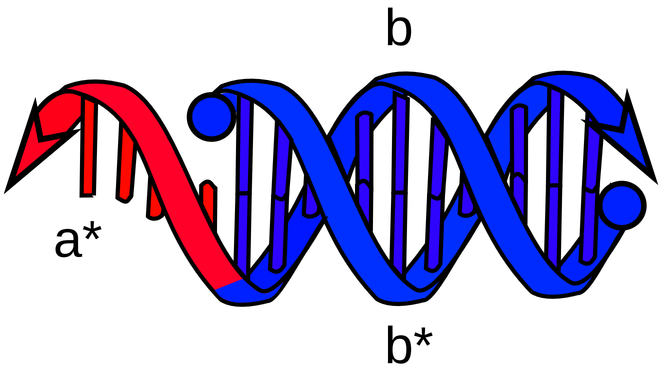
MPP2-Finale Workshop
Caltech, June $27^{th}$, 2019

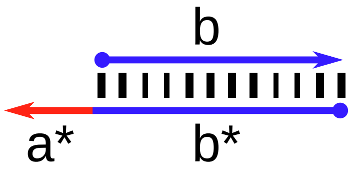
http://www.github.com/DNA-and-Natural-Algorithms-Group
**peppercornenumerator, nuskell, KinDA**

# DNA STRAND DISPLACEMENT



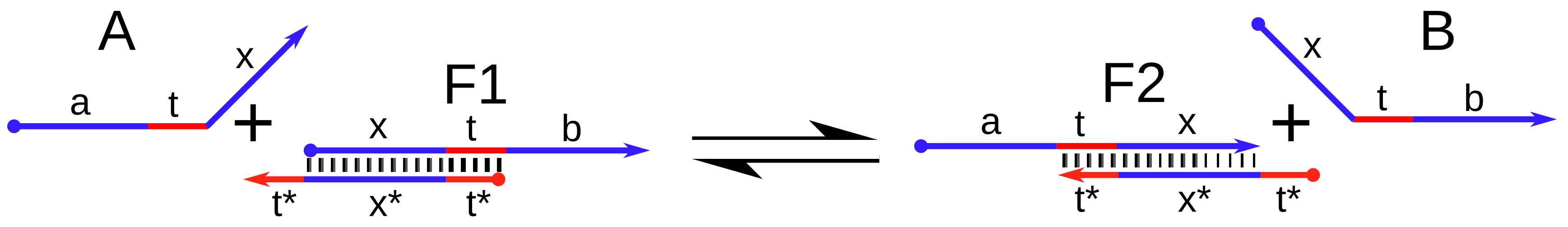



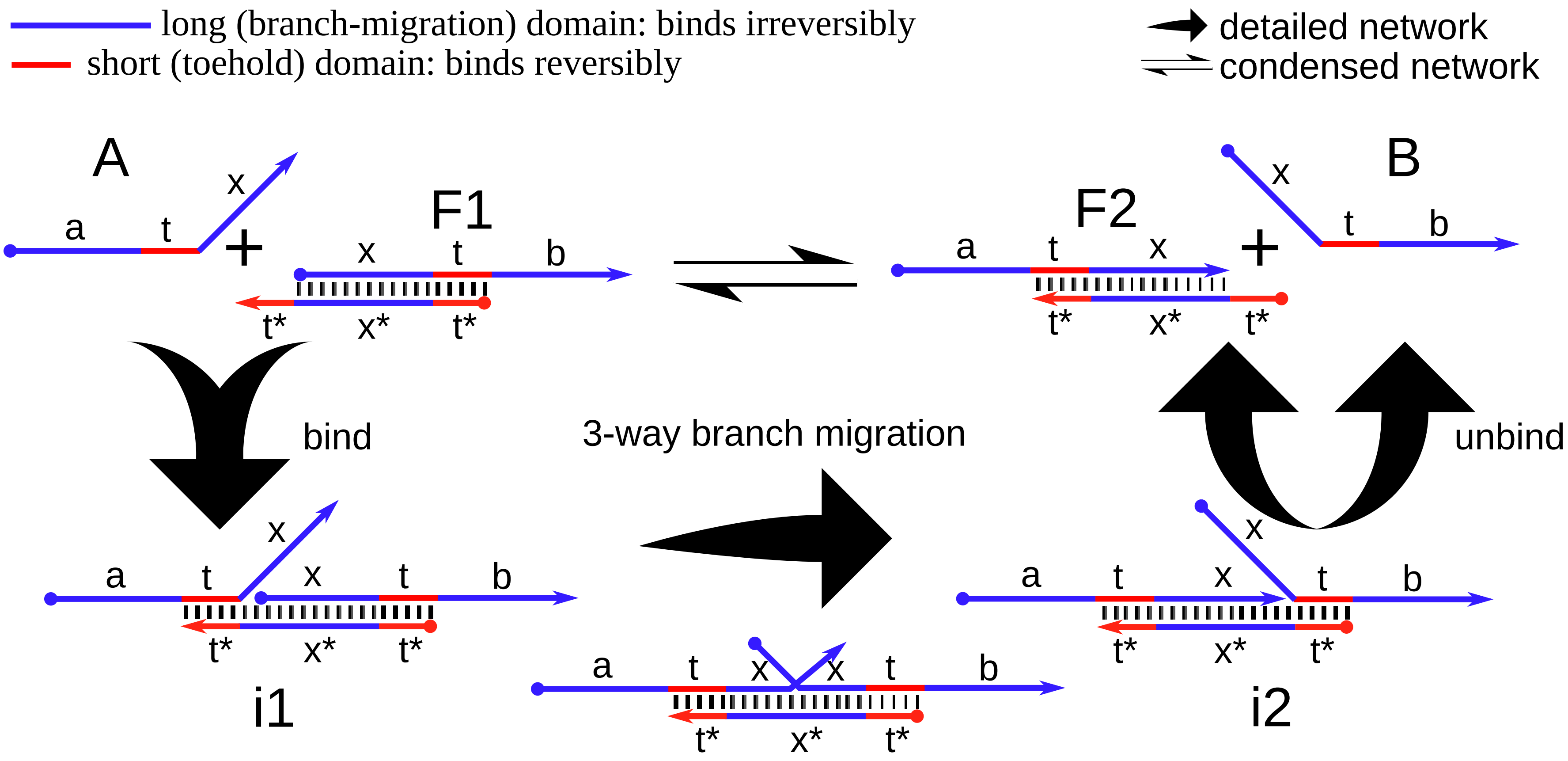



# DOMAIN-LEVEL STRAND DISPLACEMENT

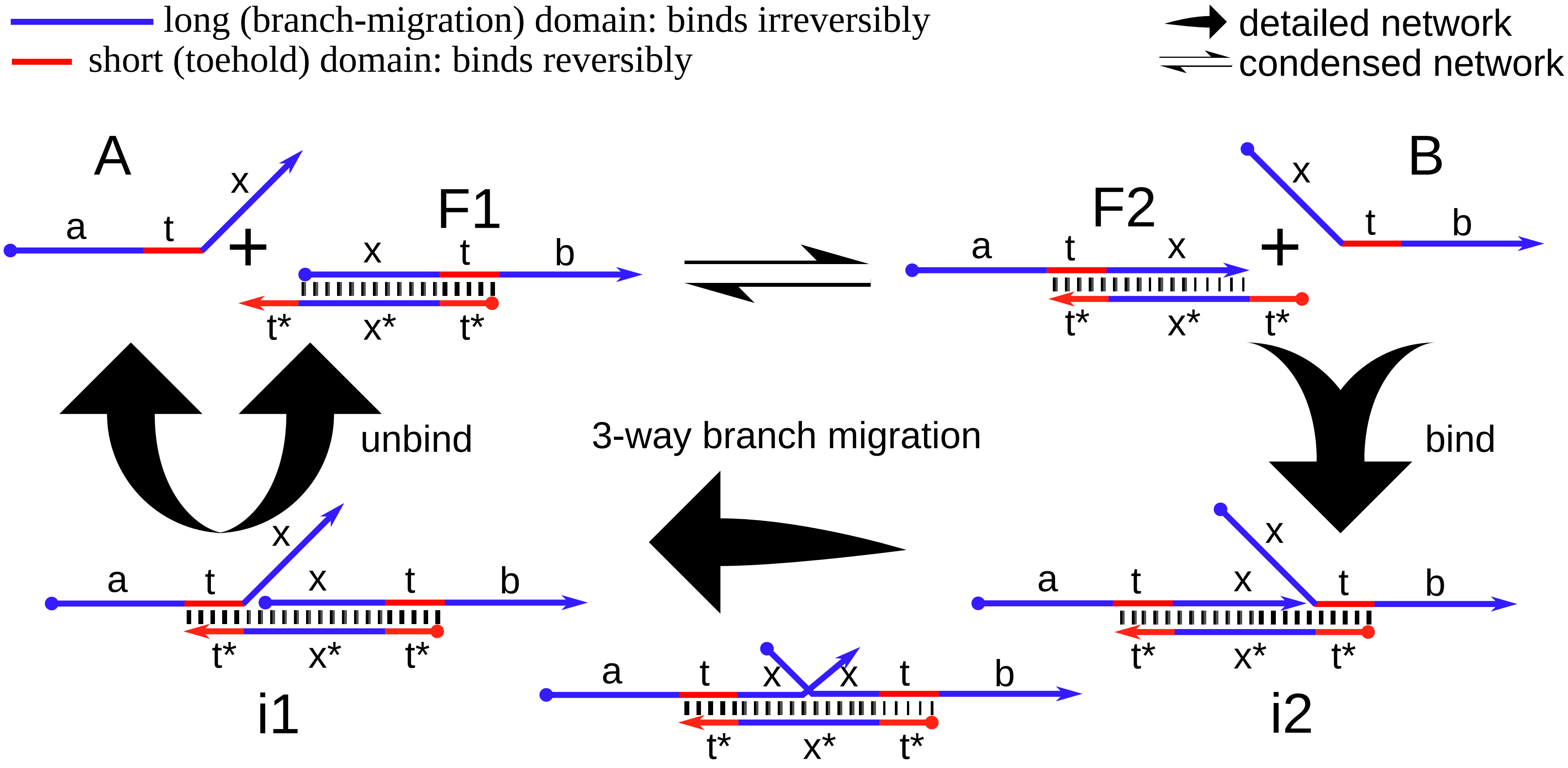
long (branch-migration) domain: binds irreversibly
short (toehold) domain: binds reversibly



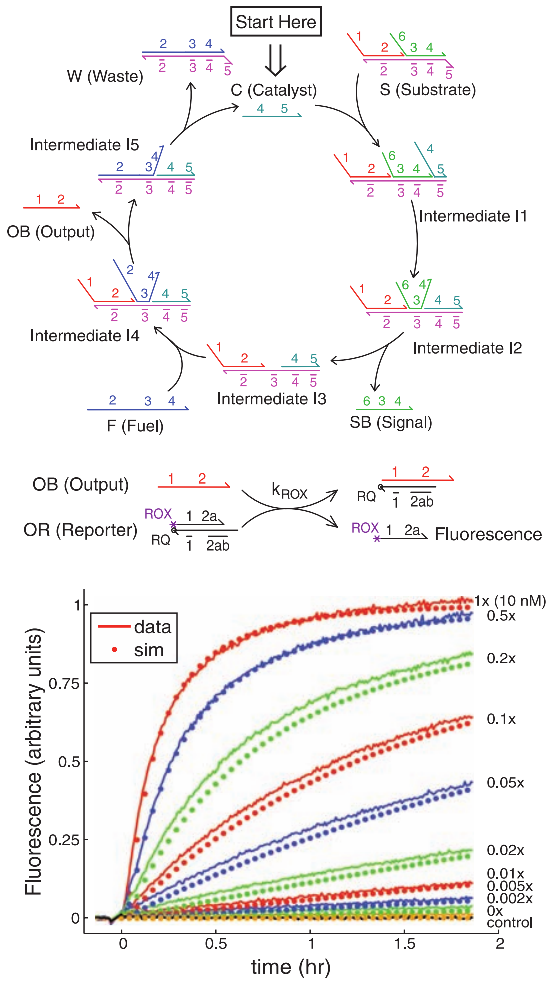
# DOMAIN-LEVEL STRAND DISPLACEMENT



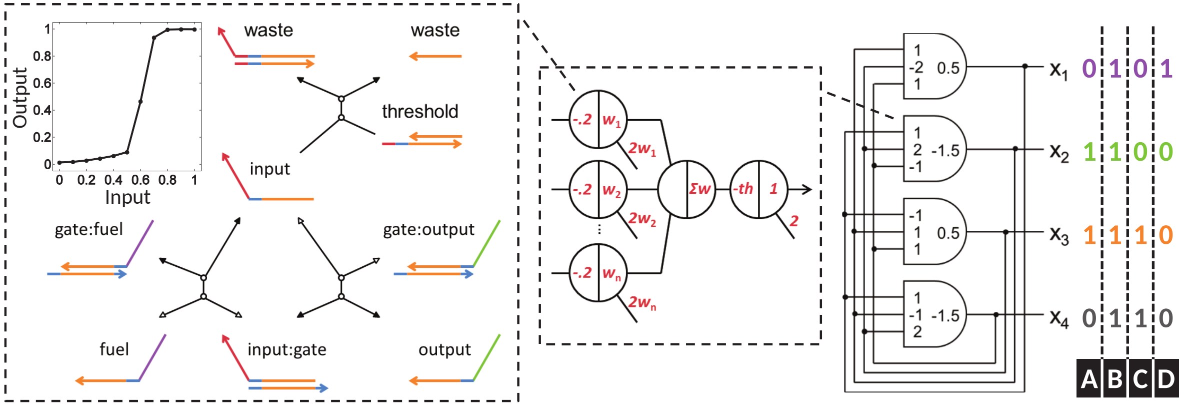
# DOMAIN-LEVEL STRAND DISPLACEMENT



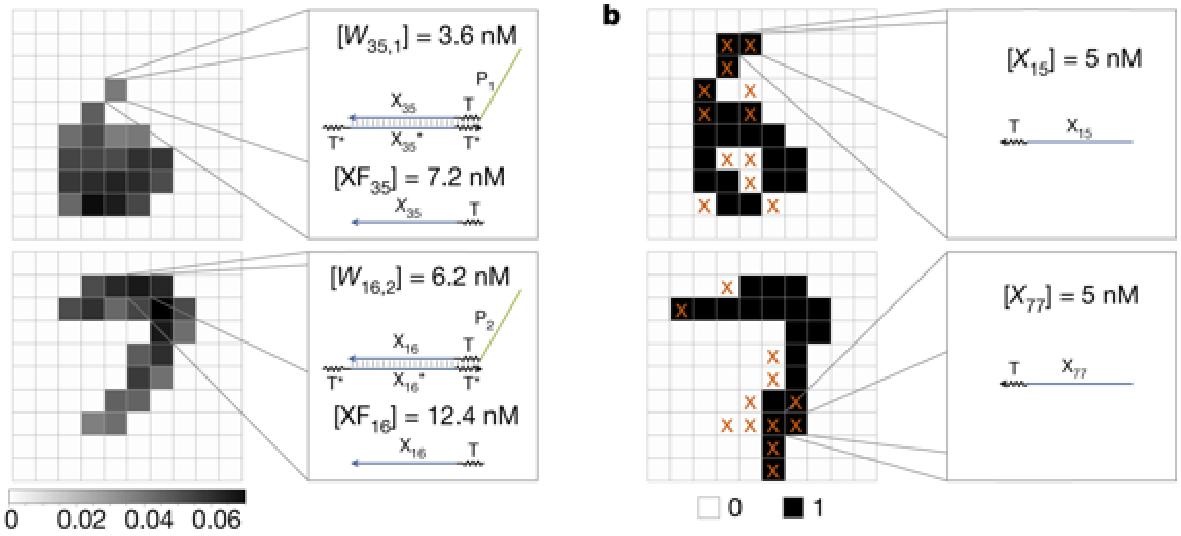
# MANY EXPERIMENTAL DEMONSTRATIONS …




Zhang et al. (2007)

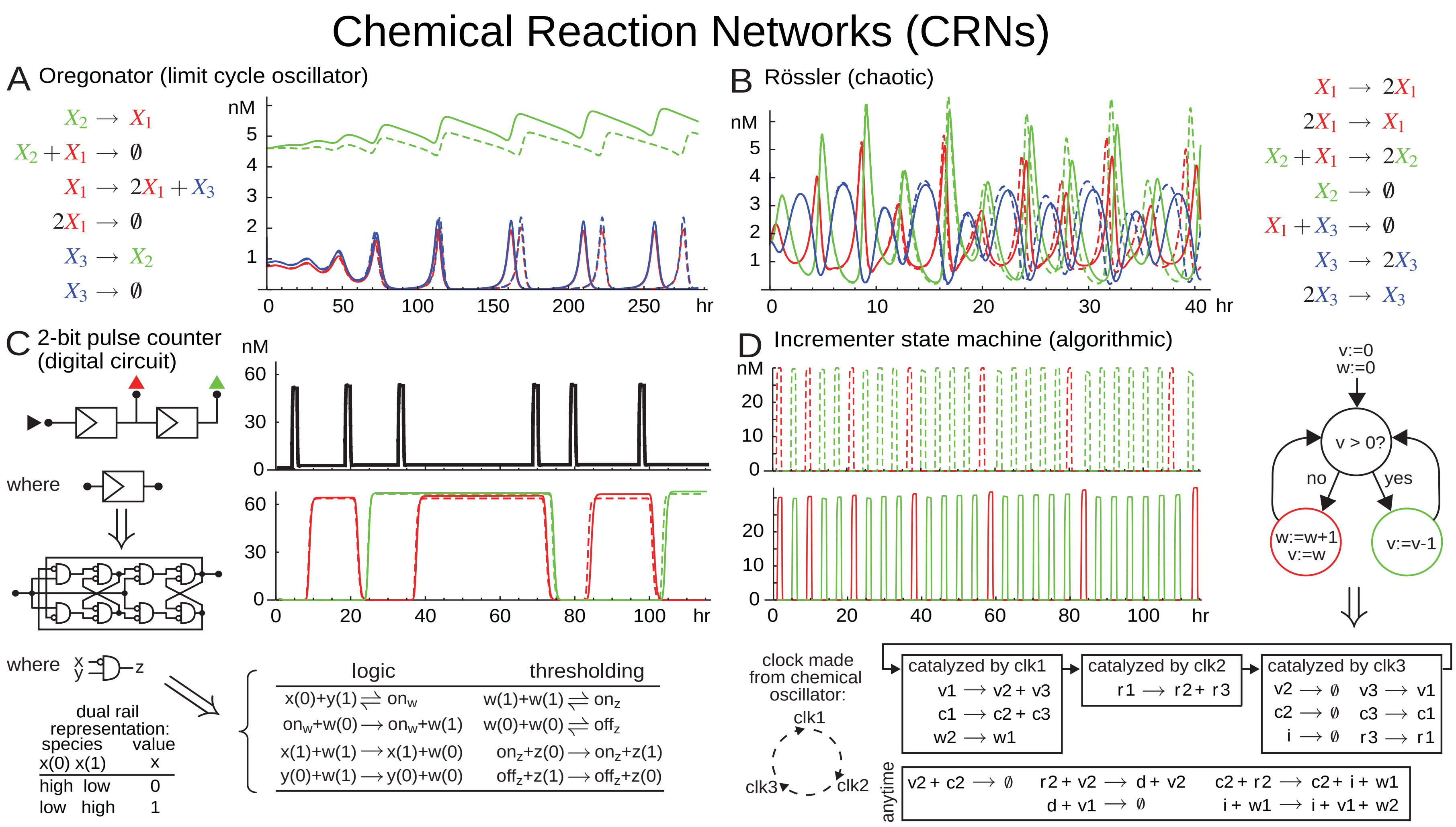



Qian, Winfree, Bruck (2011)



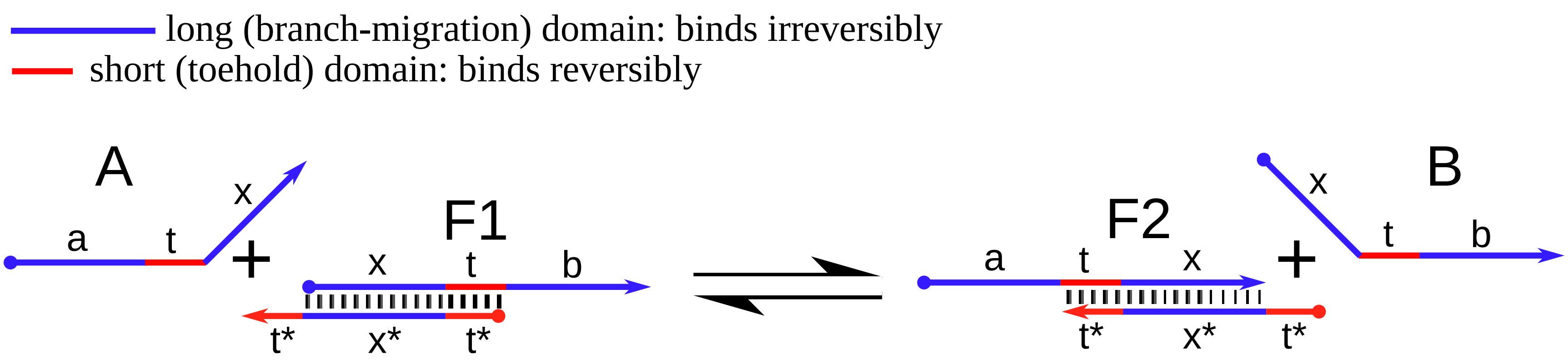

Cherry & Qian (2018)

# ... MANY MORE POTENTIAL APPLICATIONS.




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

# DOMAIN-LEVEL STRAND DISPLACEMENT




formal CRN

$$A \rightleftharpoons B$$

formal species: {A, B}

DSD sytem specification

$$A + F1 \rightleftharpoons F2 + B$$

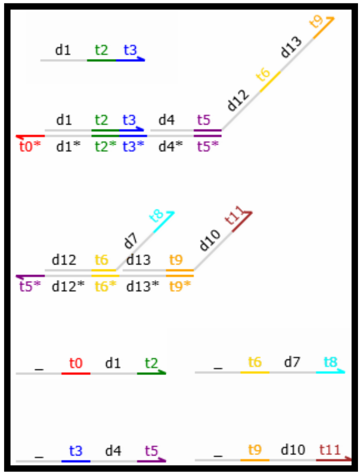
signal species (low concentation): {A, B}

fuel species (high concentration): {F1, F2}

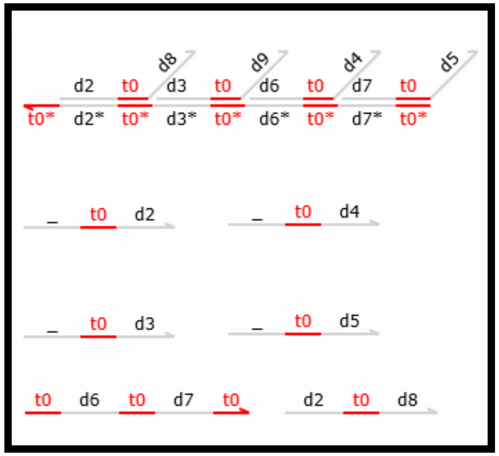
# FROM CRN TO DSD SYSTEMS

$$A + B \rightarrow C + D$$

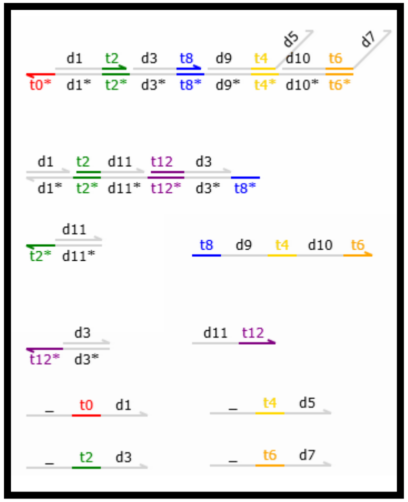
Soloveichik et al. (2010)



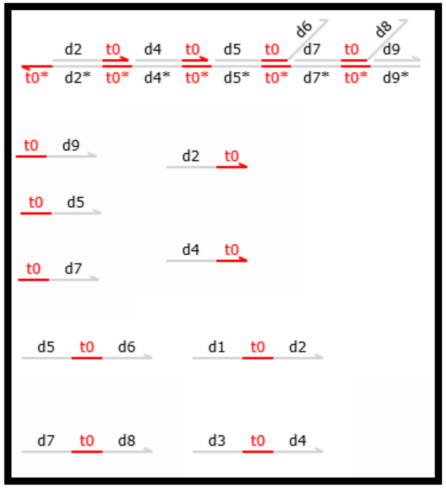
Lakin et al. (2012)



Cardelli (2011)



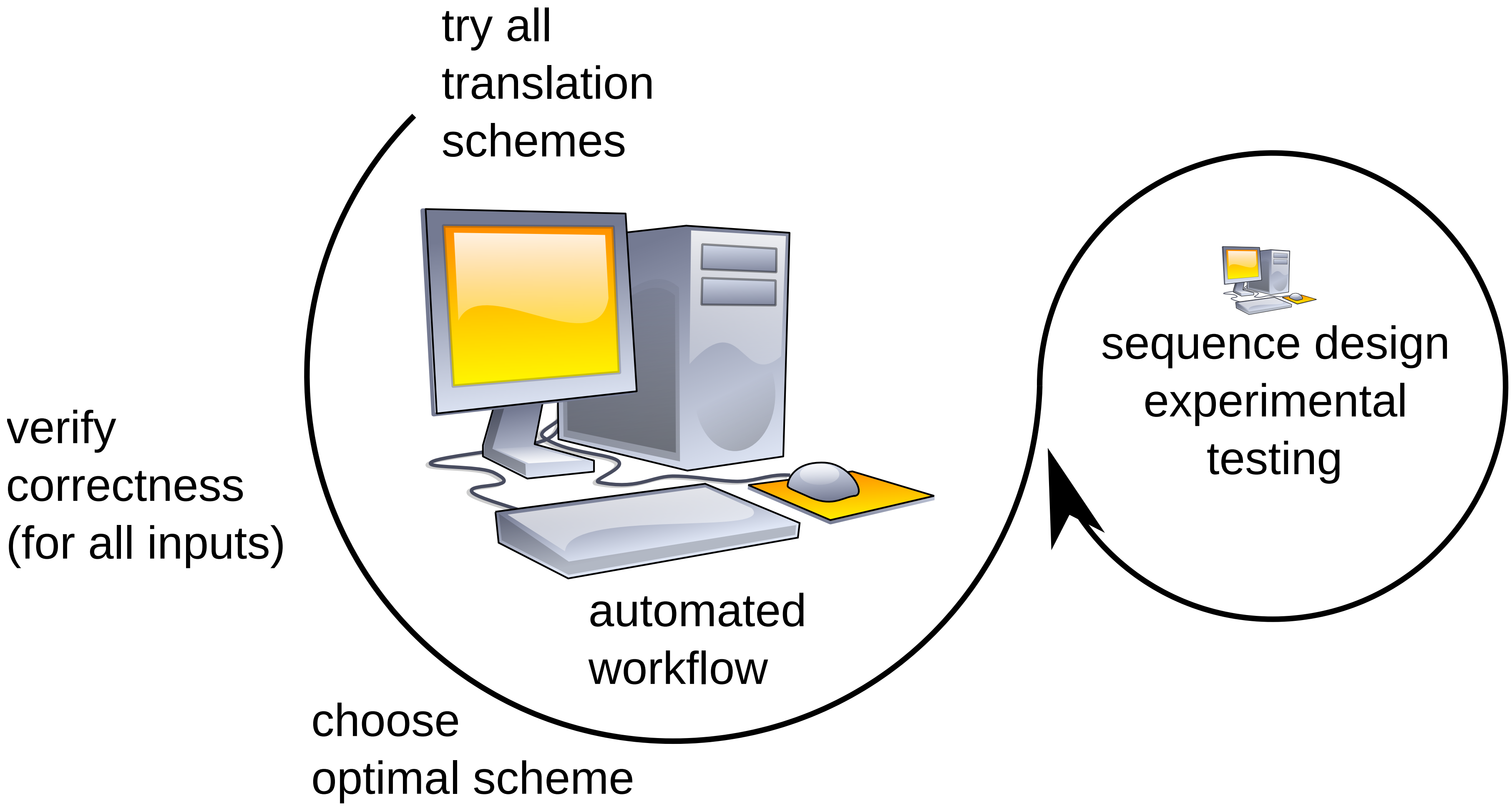
Qian et al. (2011)



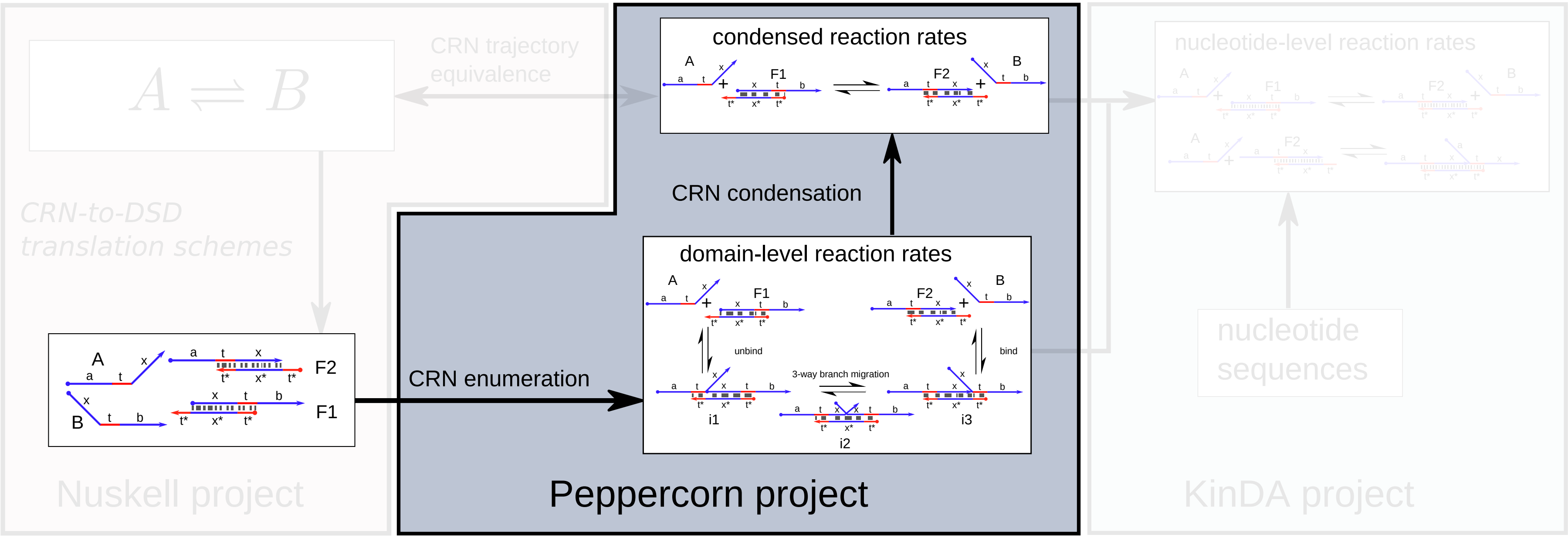
Chen et al. (2012), Cardelli (2013), Srinivas (2015), Lakin et al. (2016), ...

Images drawn using the DNA strand displacement analysis software VisualDSD:
Philipps & Cardelli (2009), ..., Spaccasassi et al. (2017)

# A CRN-TO-DSD COMPILER



# THE CENTRAL COMPONENT: PEPPERCORN



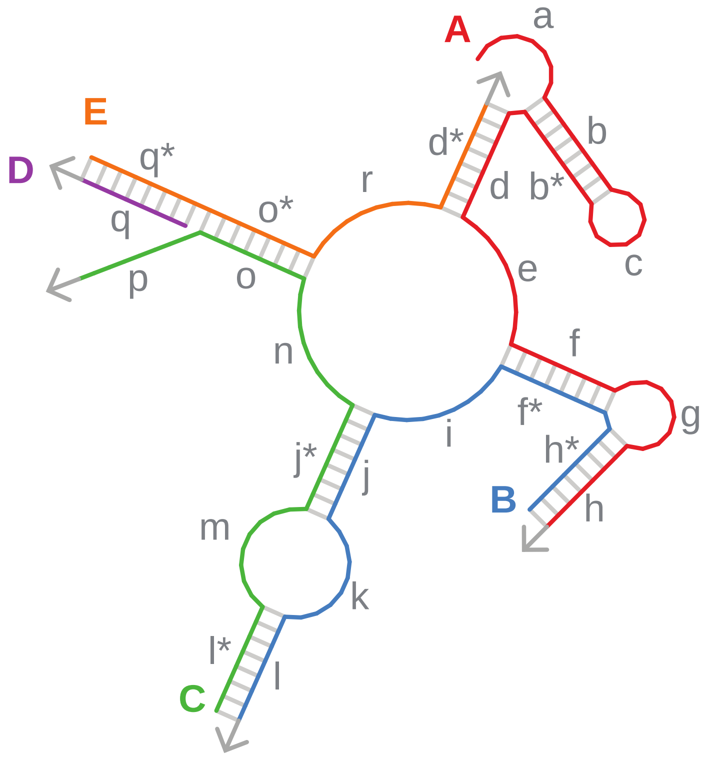

Grun et al. (2014) - arXiv
Badelt, Grun et al. (in perparation)

# 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*`

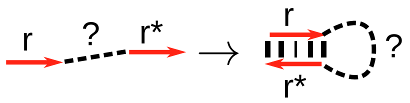
`. ( . ) ( . ( . ( + ) ) . ( . ( + ) . ) . ( . + ( + ) ) . )`

c) "kernel" notation

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

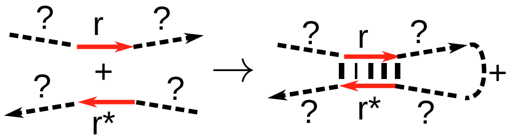
# REACTION TYPES & APPROXIMATE RATES 1/2

`bind11`:



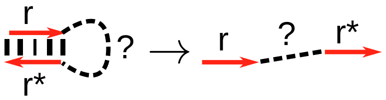

$$k_{\text{bind11}} = Pr(?) * 10^6\,\text{s}^{-1}$$

`bind21`:




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

`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

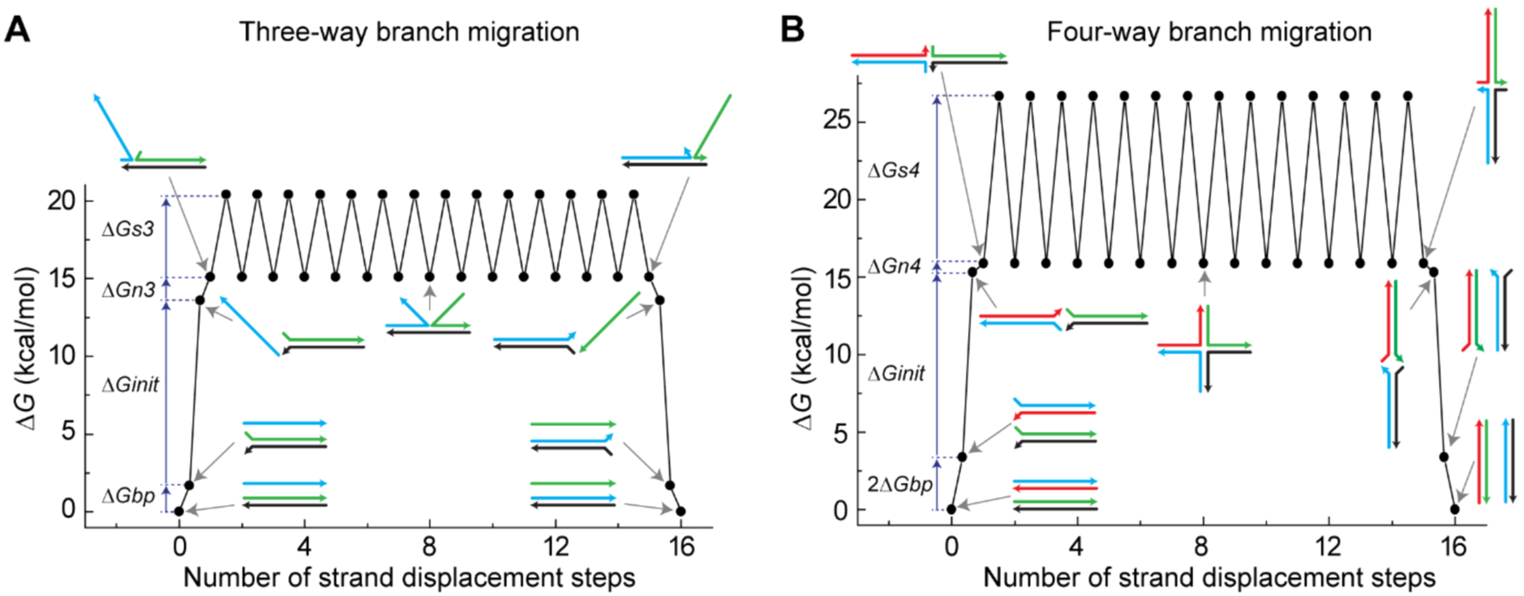
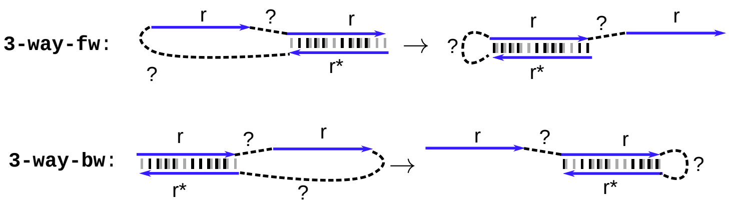
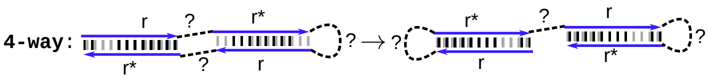



Figure from Kotani & Hughes (2017)

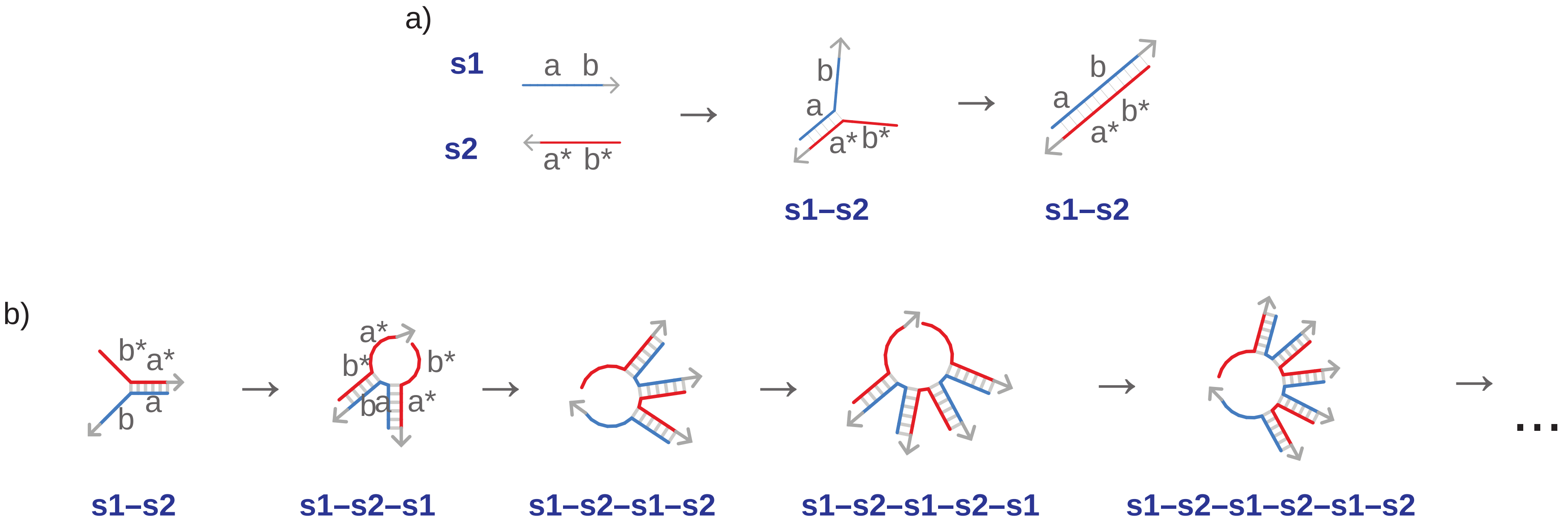


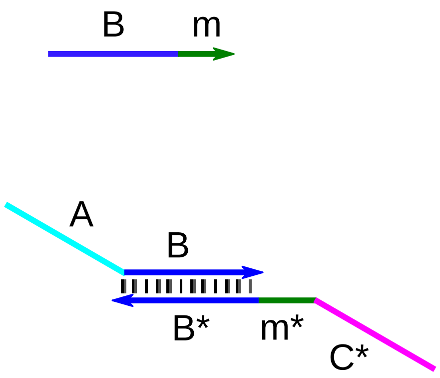

$$k_{\text{bm}} = \frac{Pr(\bigcirc ?) * k_{\text{bm-init}}}{|r|} \ \mathrm{s}^{-1}$$

unimolecular, but may lead to dissociation

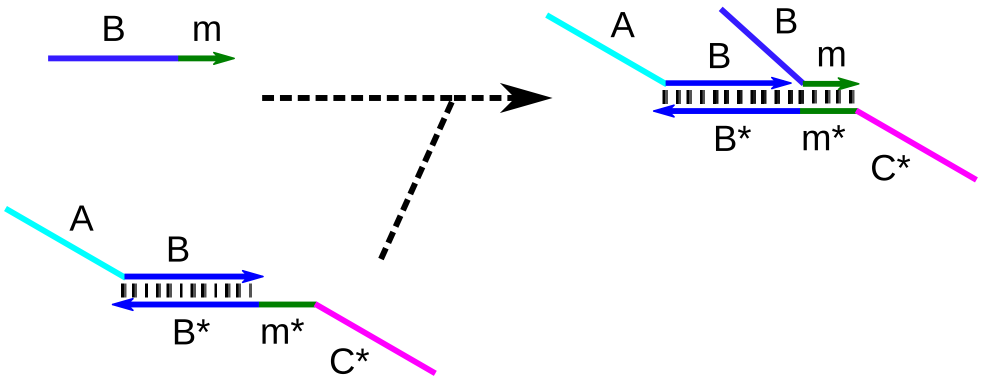
# POLYMERIZATION



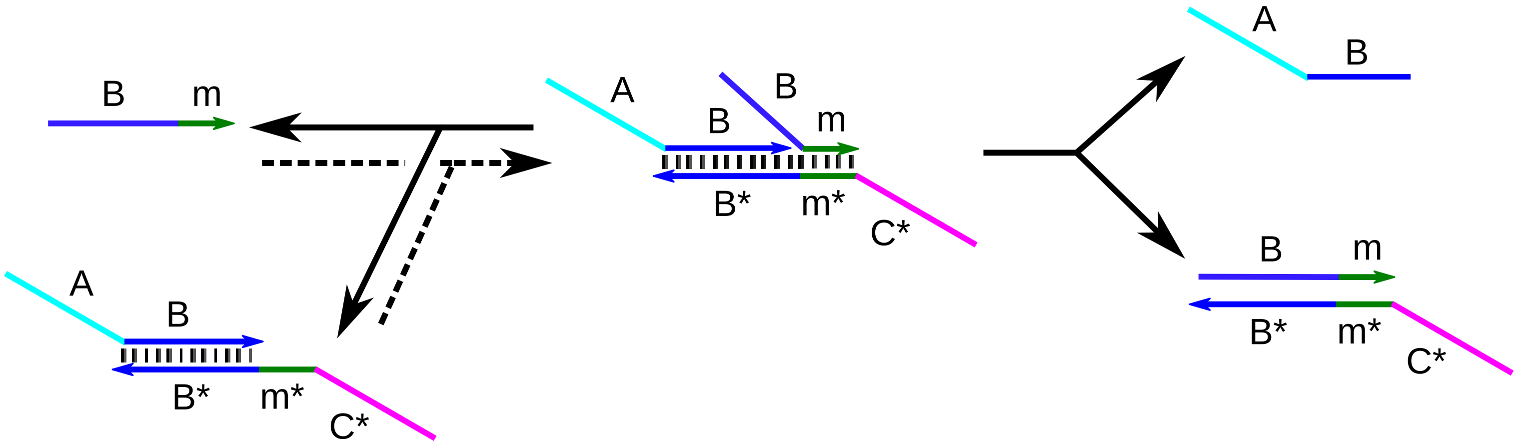
# ENUMERATION / CONDENSATION



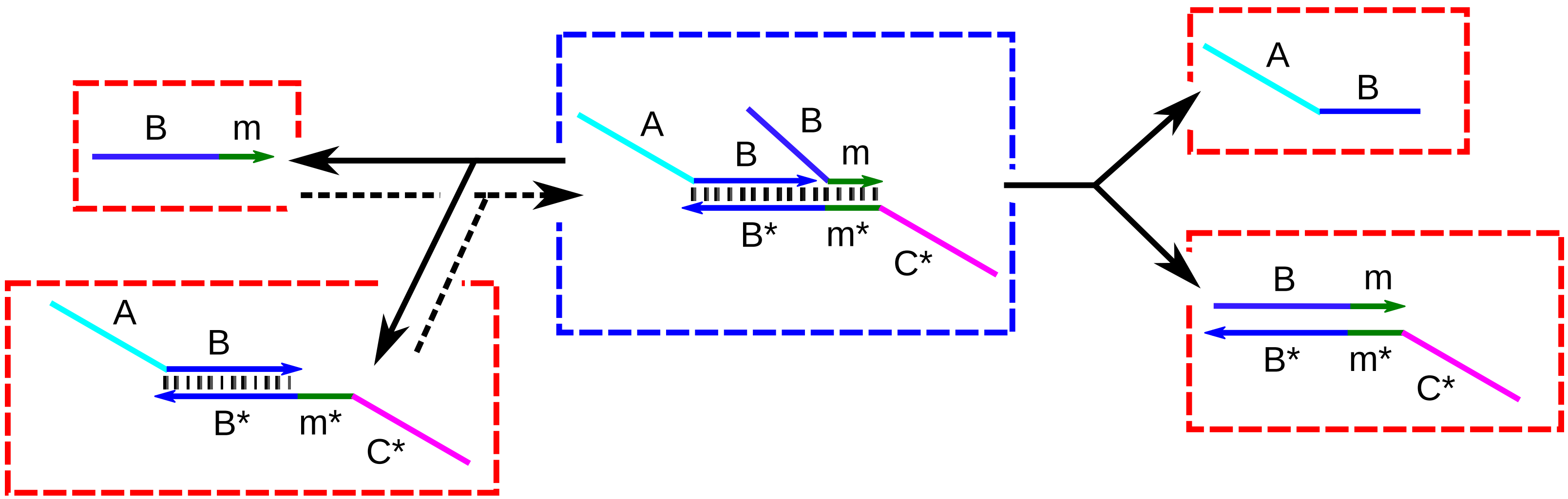
# ENUMERATION / CONDENSATION



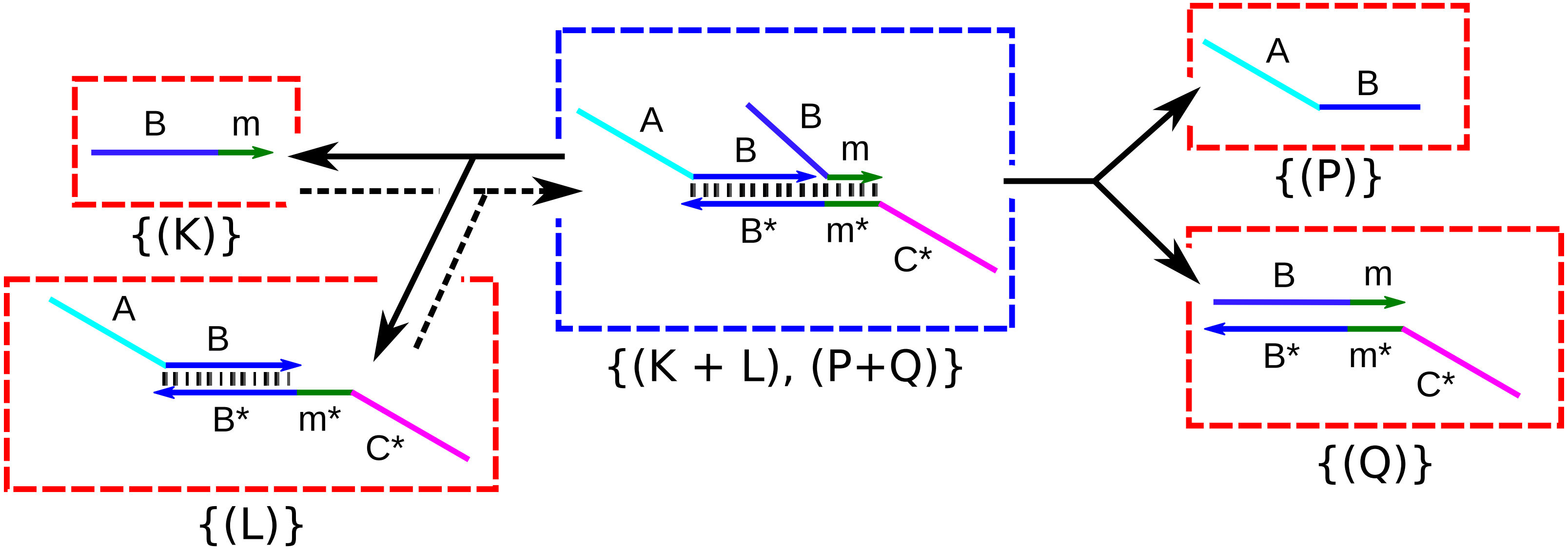
# ENUMERATION / CONDENSATION

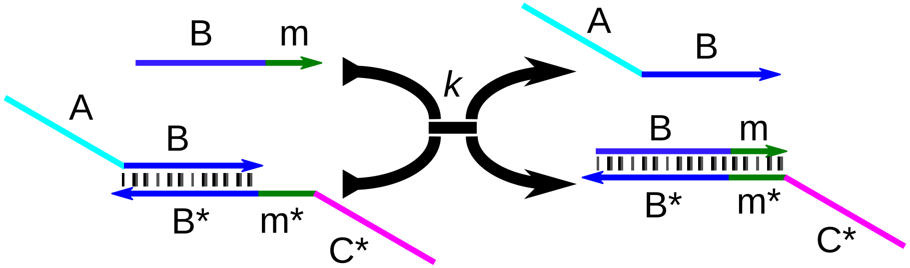


# ENUMERATION / CONDENSATION

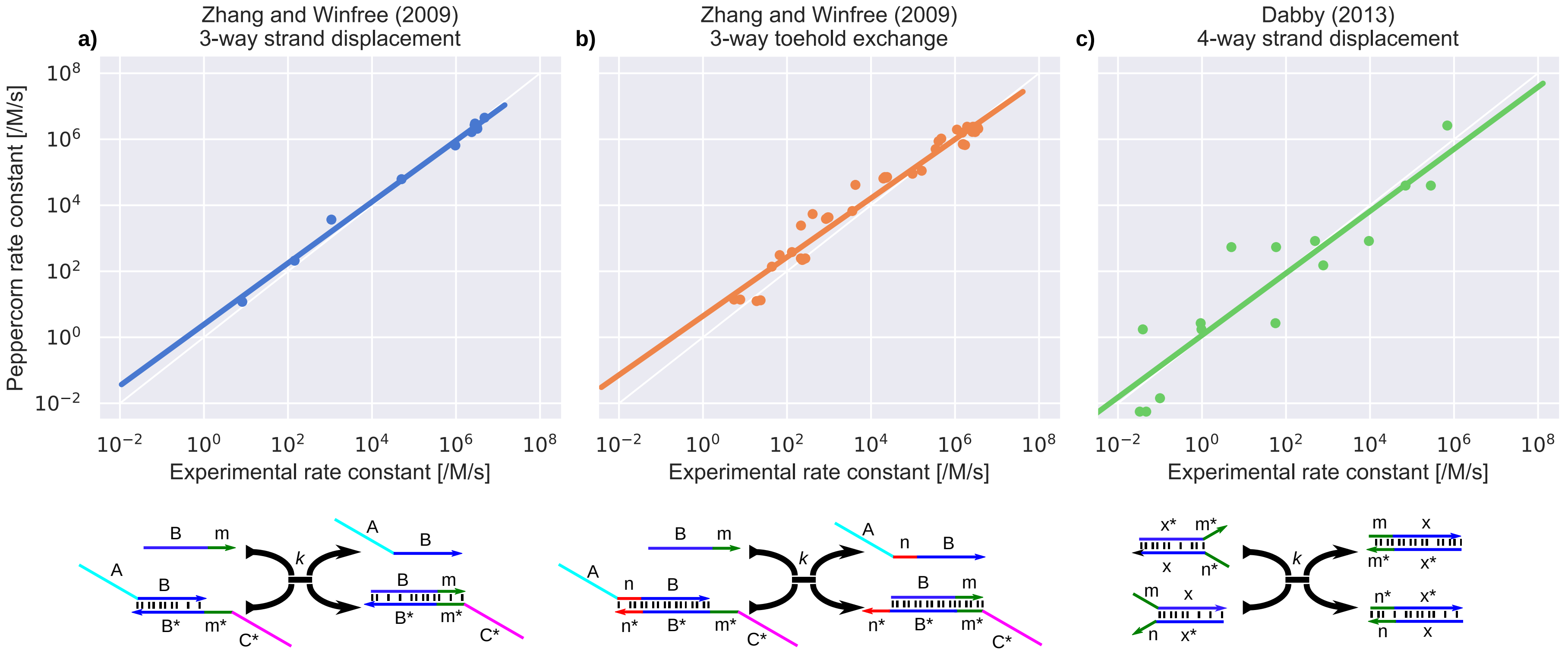


# ENUMERATION / CONDENSATION

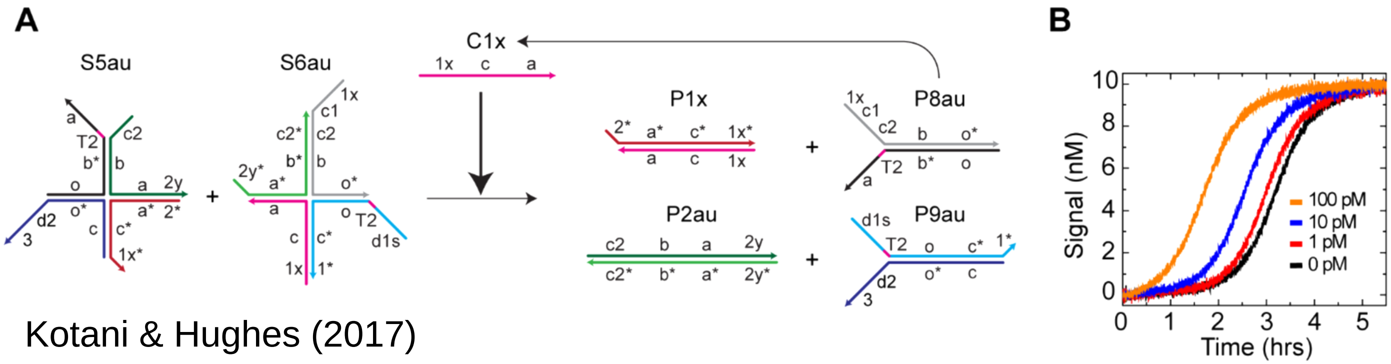


# CONDENSED REACTION RATES

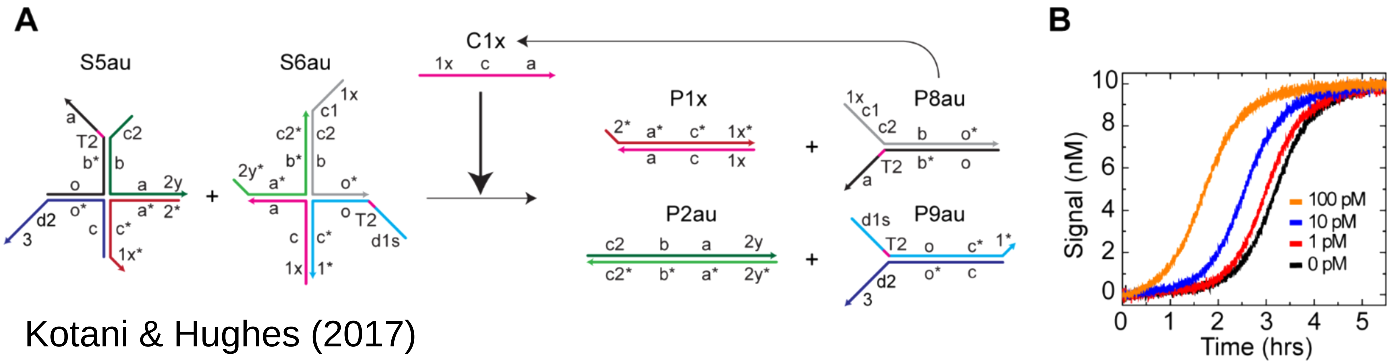


# AUTOCATALYTIC SYSTEM

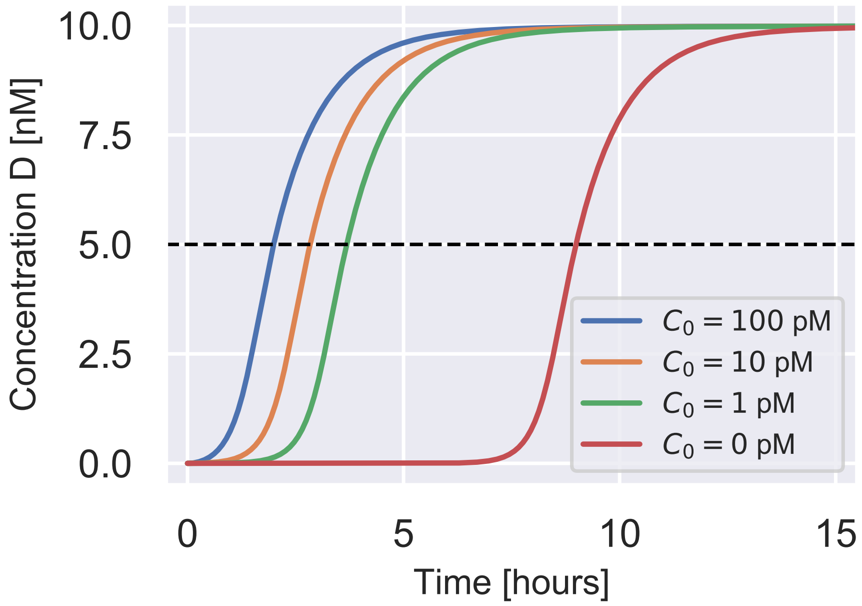


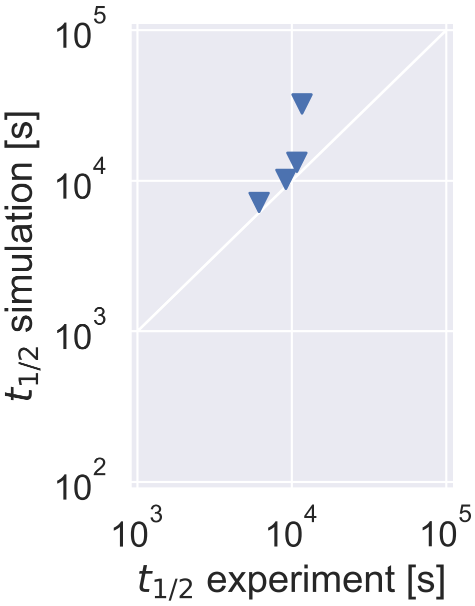

Kotani & Hughes (2017)

# AUTOCATALYTIC SYSTEM

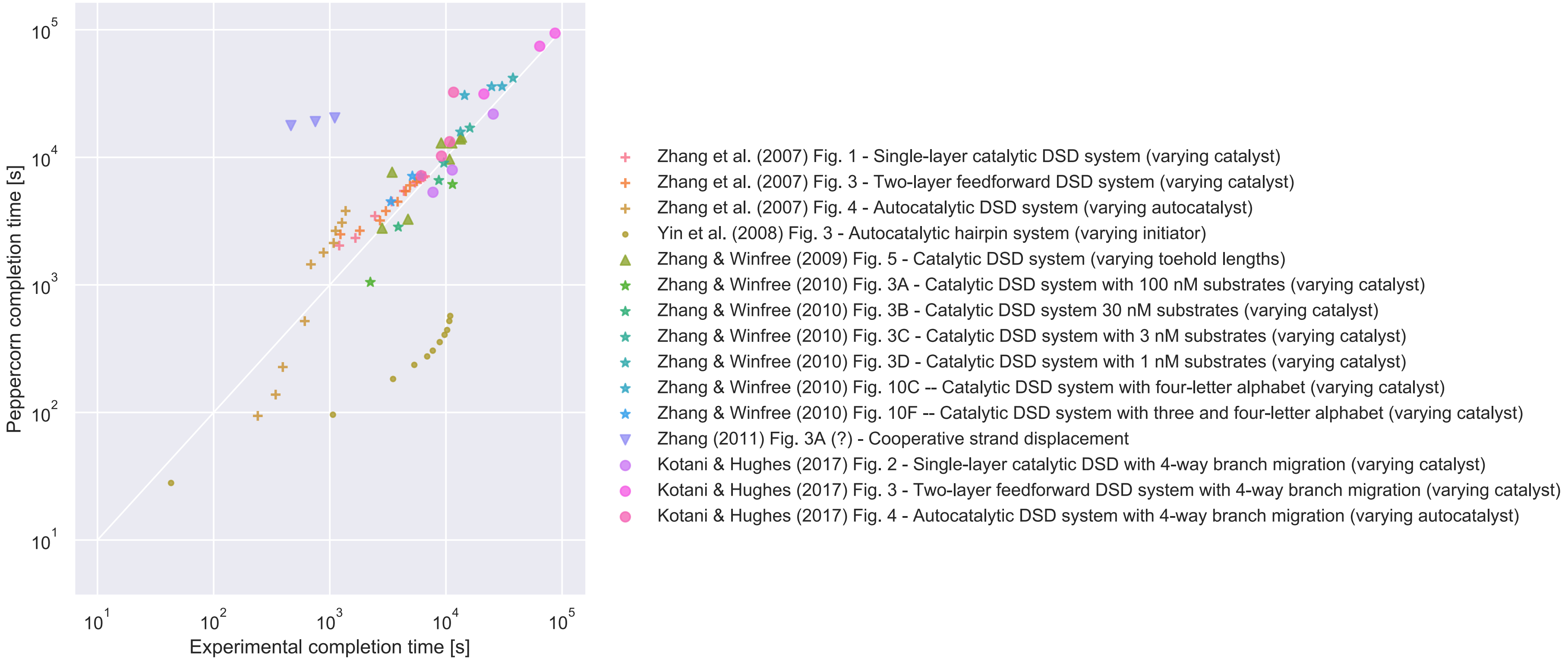



Kotani & Hughes (2017)

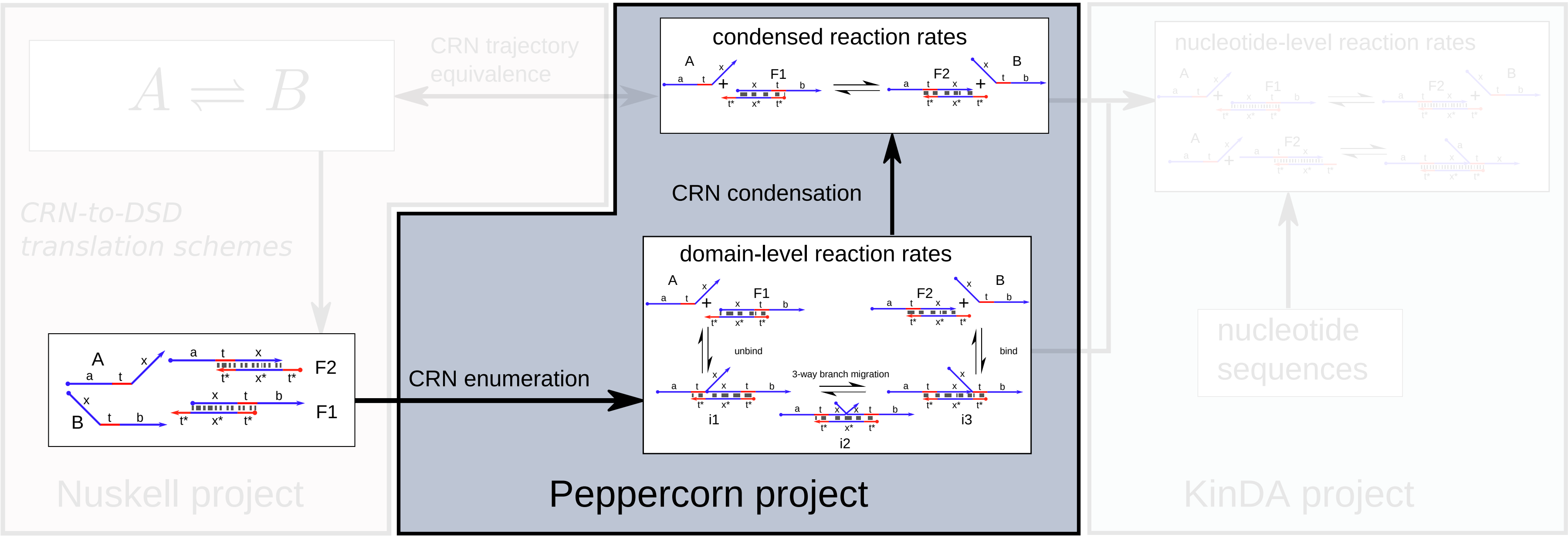


# MANY SYSTEMS



# THE COMPILER FRAMEWORK




**Peppercorn:** Badelt, Grun et al. (in perparation)

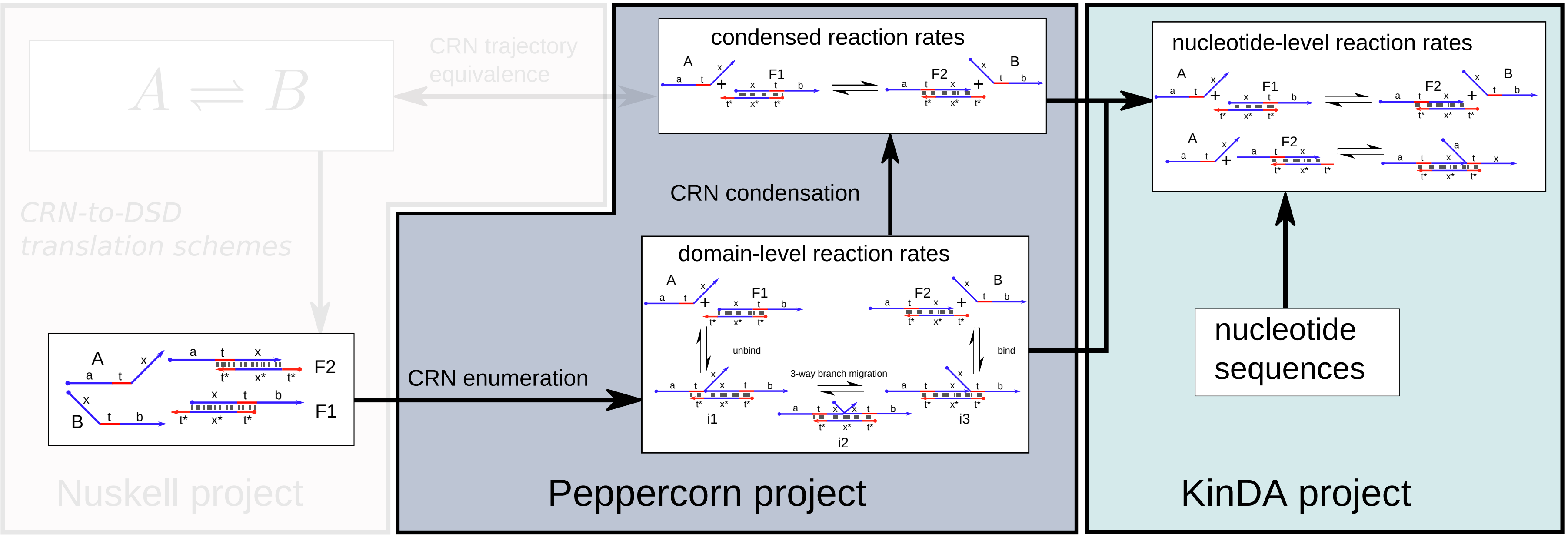
**KinDA:** Berleant et al. (2018)

- NUPACK: Dirks et al. (2007)
- Multistrand: Schaeffer et al. (2015)

**Nuskell:** Badelt et al. (2017)

- CRN pathway decomposition equivalence: Shin et al. (2017)
- CRN bisimulation equivalence: Johnson et al. (2018)

# THE COMPILER FRAMEWORK



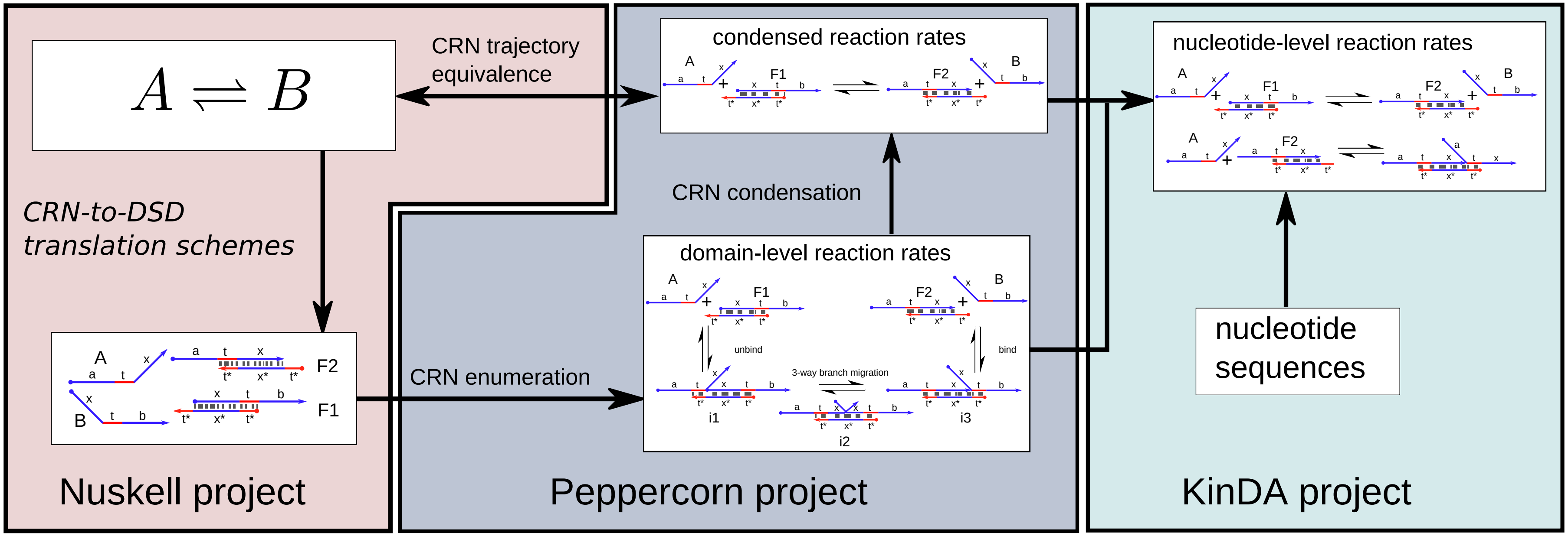

**Peppercorn:** Badelt, Grun et al. (in perparation)
**KinDA:** Berleant et al. (2018)

- NUPACK: Dirks et al. (2007)
- Multistrand: Schaeffer et al. (2015)

**Nuskell:** Badelt et al. (2017)

- CRN pathway decomposition equivalence: Shin et al. (2017)
- CRN bisimulation equivalence: Johnson et al. (2018)

# THE COMPILER FRAMEWORK




**Peppercorn:** Badelt, Grun et al. (in perparation)
**KinDA:** Berleant et al. (2018)

- NUPACK: Dirks et al. (2007)
- Multistrand: Schaeffer et al. (2015)

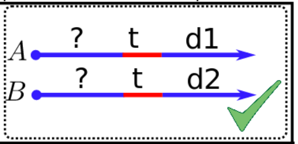
**Nuskell:** Badelt et al. (2017)

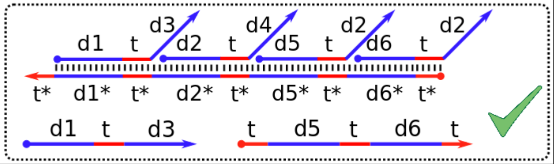
- CRN pathway decomposition equivalence: Shin et al. (2017)
- CRN bisimulation equivalence: Johnson et al. (2018)

# TRANSLATION SCHEMES

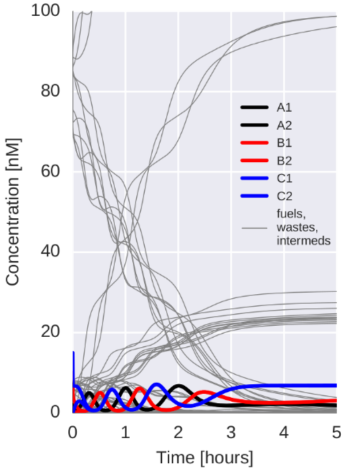
$A + B \to B + B$

```
# Translate formal reactions with two reactants and two products.
# Lakin et. al (2012) "Abstractions for DNA circuit design." [Figure 5]
# Define a global short toehold domain
global toehold = short();

# Define domains and structure of signal species
class formal(s) = "? t f" | ". . ."
  where { t = toehold; f = long() };

# Define fuel complexes for bimolecular reactions
class binary_fuels(r, p) =
  [ "a t i + b t k + ch t c + dh t d + t* dh* t* ch* t* b* t* a* t*"
  | "( ( . + ( ( . + (  ( . + (  ( . + )   )   )   )   )   )   )   )   . ",
    "a t i" | " . . . ", "t ch t dh t" | ". . . . ." ]
  where {
    a = r[0].f;
    b = r[1].f;
    c = p[0].f; ch = long();
    d = p[1].f; dh = long();
    i = long(); k = long();
    t = toehold };
```

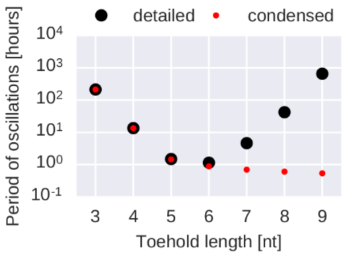



# DETAILED VS. CONDENSED ENUMERATION




$$A + B \to B + B$$
$$B + C \to C + C$$
$$C + A \to A + A$$

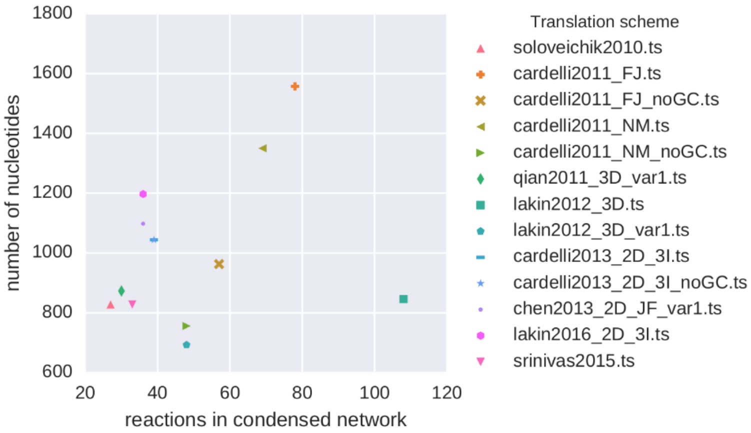



# OSCILLATOR ANALYSIS 1/2

| Translation scheme | pathway decompositon | CRN bisimulation |
|---|---|---|
| soloveichik2010.ts | True | True |
| cardelli2011_FJ.ts | False | False |
| cardelli2011_FJ_noGC.ts | False | True |
| cardelli2011_NM.ts | timeout | False |
| cardelli2011_NM_noGC.ts | False | True |
| qian2011_3D_var1.ts | True | True |
| lakin2012_3D.ts | False | False |
| lakin2012_3D_var1.ts | False | True |
| cardelli2013_2D_3I.ts | False | True |
| cardelli2013_2D_3I_noGC.ts | timeout | True |
| chen2013_2D_JF_var1.ts | timeout | True |
| lakin2016_2D_3I.ts | timeout | False |
| srinivas2015.ts | True | True |

CRN bisimulation equivalence: Johnson et al. (2016)
CRN pathway decomposition equivalence: Shin et al. (2014)

# OSCILLATOR ANALYSIS 2/2



# CRN EQUIVALENCE

**formal input CRN**
3 species
7 reactions

```
A     -> A + A
A + A -> A
A + B -> B + B
B     ->
A + C ->
C     -> C + C
C + C -> C
```

**enumerated CRN**
360 species
668 reactions

Johnson et al. (2018) - CRN bisimulation equivalence
translation scheme: qian2011_3D_var1.ts

# CRN EQUIVALENCE

**formal input CRN**
3 species
7 reactions

```
A     -> A + A
A + A -> A
A + B -> B + B
B     ->
A + C ->
C     -> C + C
C + C -> C
```

**condensed CRN**
42 species
32 reactions

```
f14 + C -> e1428 + f15
e853 + f12 -> C + f13
A + f4 -> f3 + e71
f2 + e25 -> A + f1
A + e25 -> f2 + e7
e996 + f3 -> A + f10
e1428 + f15 -> f14 + C
f3 + e71 -> A + f4
e465 + B -> e418 + f6
e614 + f9 -> e611 + e730
e996 + C -> e1040 + f12
e465 + f5 -> e514 + e368
e308 + f7 -> f8 + B
e418 + B -> e371 + f6
C + f13 -> e853 + f12
A + f1 -> f2 + e25
B + e71 -> e319 + f7
f2 + e7 -> A + e25
e1040 + f11 -> e1162 + e1163 + e1158
e319 + f7 -> B + e71
e308 + f9 -> e614 + e615 + e611
e371 + f6 -> e418 + B
e1040 + f12 -> e996 + C
f8 + B -> e308 + f7
e319 + f5 -> e372 + e371 + e368
f3 + e7 -> A + f0
e853 + f15 -> e1428 + C
e1428 + C -> e853 + f15
A + f10 -> e996 + f3
e1163 + f11 -> e1158 + e1246
e418 + f6 -> e465 + B
A + f0 -> f3 + e7
```

**enumerated CRN**
360 species
668 reactions

Johnson et al. (2018) - CRN bisimulation equivalence
translation scheme: qian2011_3D_var1.ts

# CRN EQUIVALENCE

**formal input CRN**
3 species
7 reactions

```
A     -> A + A
A + A -> A
A + B -> B + B
B     ->
A + C ->
C     -> C + C
C + C -> C
```

**verification CRN**
26 species (no fuel species)
32 reactions

```
C -> e1428
e853 -> C
A -> e71
e25 -> A
A + e25 -> e7
e996 -> A
e1428 -> C
e71 -> A
e465 + B -> e418
e614 -> e611 + e730
e996 + C -> e1040
e465 -> e514 + e368
e308 -> B
e418 + B -> e371
C -> e853
A -> e25
B + e71 -> e319
e7 -> A + e25
e1040 -> e1162 + e1163 + e1158
e319 -> B + e71
e308 -> e614 + e615 + e611
e371 -> e418 + B
e1040 -> e996 + C
B -> e308
e319 -> e372 + e371 + e368
e7 -> A
e853 -> e1428 + C
e1428 + C -> e853
A -> e996
e1163 -> e1158 + e1246
e418 -> e465 + B
A -> e7
```

**condensed CRN**
42 species
32 reactions

```
f14 + C -> e1428 + f15
e853 + f12 -> C + f13
A + f4 -> f3 + e71
f2 + e25 -> A + f1
A + e25 -> f2 + e7
e996 + f3 -> A + f10
e1428 + f15 -> f14 + C
f3 + e71 -> A + f4
e465 + B -> e418 + f6
e614 + f9 -> e611 + e730
e996 + C -> e1040 + f12
e465 + f5 -> e514 + e368
e308 + f7 -> f8 + B
e418 + B -> e371 + f6
C + f13 -> e853 + f12
A + f1 -> f2 + e25
B + e71 -> e319 + f7
f2 + e7 -> A + e25
e1040 + f11 -> e1162 + e1163 + e1158
e319 + f7 -> B + e71
e308 + f9 -> e614 + e615 + e611
e371 + f6 -> e418 + B
e1040 + f12 -> e996 + C
f8 + B -> e308 + f7
e319 + f5 -> e372 + e371 + e368
f3 + e7 -> A + f0
e853 + f15 -> e1428 + C
e1428 + C -> e853 + f15
A + f10 -> e996 + f3
e1163 + f11 -> e1158 + e1246
e418 + f6 -> e465 + B
A + f0 -> f3 + e7
```

**enumerated CRN**
360 species
668 reactions

Johnson et al. (2018) - CRN bisimulation equivalence
translation scheme: qian2011_3D_var1.ts

# CRN EQUIVALENCE

**formal input CRN**
3 species
7 reactions

```
A     -> A + A
A + A -> A
A + B -> B + B
B     ->
A + C ->
C     -> C + C
C + C -> C
```

Interpretation (CRN-bisimulation):

```
A => A
B => B
C => C
e1040 => A, C
e1158 =>
e1162 =>
e1163 =>
e1246 =>
e1428 => C
e25 => A
e308 => B
e319 => A, B
e368 =>
e371 => B, B
e372 =>
e418 => B
e465 =>
e514 =>
e611 =>
e614 =>
e615 =>
e7 => A, A
e71 => A
e730 =>
e853 => C, C
e996 => A
```

**interpreted CRN**
3 species
7 non-trivial reactions

```
C     -> C
C + C -> C
A     -> A
A     -> A
A + A -> A + A
A     -> A
C     -> C
A     -> A
B     -> B
      ->
A + C -> A + C
      ->
B     -> B
B + B -> B + B
C     -> C + C
A     -> A
B + A -> A + B
A + A -> A + A
A + C ->
A + B -> B + A
B     ->
B + B -> B + B
A + C -> A + C
B     -> B
A + B -> B + B
A + A -> A
C + C -> C + C
C + C -> C + C
A     -> A
      ->
B     -> B
A     -> A + A
```

**verification CRN**
26 species (no fuel species)
32 reactions

```
C -> e1428
e853 -> C
A -> e71
e25 -> A
A + e25 -> e7
e996 -> A
e1428 -> C
e71 -> A
e465 + B -> e418
e614 -> e611 + e730
e996 + C -> e1040
e465 -> e514 + e368
e308 -> B
e418 + B -> e371
C -> e853
A -> e25
B + e71 -> e319
e7 -> A + e25
e1040 -> e1162 + e1163 + e1158
e319 -> B + e71
e308 -> e614 + e615 + e611
e371 -> e418 + B
e1040 -> e996 + C
B -> e308
e319 -> e372 + e371 + e368
e7 -> A
e853 -> e1428 + C
e1428 + C -> e853
A -> e996
e1163 -> e1158 + e1246
e418 -> e465 + B
A -> e7
```

**condensed CRN**
42 species
32 reactions

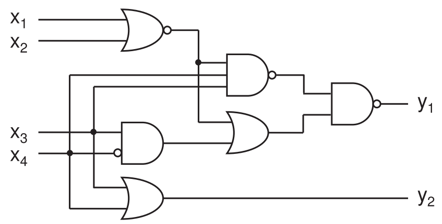
```
f14 + C -> e1428 + f15
e853 + f12 -> C + f13
A + f4 -> f3 + e71
f2 + e25 -> A + f1
A + e25 -> f2 + e7
e996 + f3 -> A + f10
e1428 + f15 -> f14 + C
f3 + e71 -> A + f4
e465 + B -> e418 + f6
e614 + f9 -> e611 + e730
e996 + C -> e1040 + f12
e465 + f5 -> e514 + e368
e308 + f7 -> f8 + B
e418 + B -> e371 + f6
C + f13 -> e853 + f12
A + f1 -> f2 + e25
B + e71 -> e319 + f7
f2 + e7 -> A + e25
e1040 + f11 -> e1162 + e1163 + e1158
e319 + f7 -> B + e71
e308 + f9 -> e614 + e615 + e611
e371 + f6 -> e418 + B
e1040 + f12 -> e996 + C
f8 + B -> e308 + f7
e319 + f5 -> e372 + e371 + e368
f3 + e7 -> A + f0
e853 + f15 -> e1428 + C
e1428 + C -> e853 + f15
A + f10 -> e996 + f3
e1163 + f11 -> e1158 + e1246
e418 + f6 -> e465 + B
A + f0 -> f3 + e7
```

**enumerated CRN**
360 species
668 reactions

Johnson et al. (2018) - CRN bisimulation equivalence
translation scheme: qian2011_3D_var1.ts

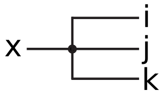
# FROM A DIGITAL CIRCUIT TO DSD

$$y_2 y_1 = \lfloor \sqrt{x_4 x_3 x_2 x_1} \rfloor$$



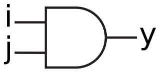

Qian et al. (2011)

fanout:




$$x_{on} \rightarrow i_{on} + j_{on} + k_{on}$$
$$x_{off} \rightarrow i_{off} + j_{off} + k_{off}$$

AND gate:




$$i_{off} + j_{off} \rightarrow y_{off}$$
$$i_{off} + j_{on} \rightarrow y_{off}$$
$$i_{on} + j_{off} \rightarrow y_{off}$$
$$i_{on} + j_{on} \rightarrow y_{on}$$

```
# The CRN computes the floor of the squareroot of a four-bit binary number.

# Fanout X3 -> F5 + F6 + F7
X3_OFF -> F5_OFF + F6_OFF + F7_OFF
X3_ON  -> F5_ON  + F6_ON  + F7_ON
# Fanout X4 -> F8 + F9 + F10
X4_OFF -> F8_OFF + F9_OFF + F10_OFF
X4_ON  -> F8_ON  + F9_ON  + F10_ON
# G11 = NOT(X1 OR X2) + Fanout G11 -> F14 + F15
X1_OFF + X2_OFF -> F14_ON  + F15_ON
X1_OFF + X2_ON  -> F14_OFF + F15_OFF
X1_ON  + X2_OFF -> F14_OFF + F15_OFF
X1_ON  + X2_ON  -> F14_OFF + F15_OFF
# G12 = F6 AND (NOT F9)
F6_OFF + F9_OFF -> G12_OFF
F6_OFF + F9_ON  -> G12_OFF
F6_ON  + F9_OFF -> G12_ON
F6_ON  + F9_ON  -> G12_OFF
# Y2 = F7 OR F10
F7_OFF + F10_OFF -> Y2_OFF
F7_OFF + F10_ON  -> Y2_ON
F7_ON  + F10_OFF -> Y2_ON
F7_ON  + F10_ON  -> Y2_ON
# G16b = F5 AND F8
F5_OFF + F8_OFF -> G16b_OFF
F5_OFF + F8_ON  -> G16b_OFF
F5_ON  + F8_OFF -> G16b_OFF
F5_ON  + F8_ON  -> G16b_ON
# G16 = NOT(F14 AND G16b)
F14_OFF + G16b_OFF -> G16_ON
F14_OFF + G16b_ON  -> G16_ON
F14_ON  + G16b_OFF -> G16_ON
F14_ON  + G16b_ON  -> G16_OFF
# G17 = F15 OR G12
F15_OFF + G12_OFF -> G17_OFF
F15_OFF + G12_ON  -> G17_ON
F15_ON  + G12_OFF -> G17_ON
F15_ON  + G12_ON  -> G17_ON
# Y1 = NOT(G16 AND G17)
G16_OFF + G17_OFF -> Y1_ON
G16_OFF + G17_ON  -> Y1_ON
G16_ON  + G17_OFF -> Y1_ON
G16_ON  + G17_ON  -> Y1_OFF
```

Input for the nuskell compiler: **32** formal reactions.

soloveichik2010.ts: **52** signal species, **92** fuel species, **172** intermediate species, **180** reactions.

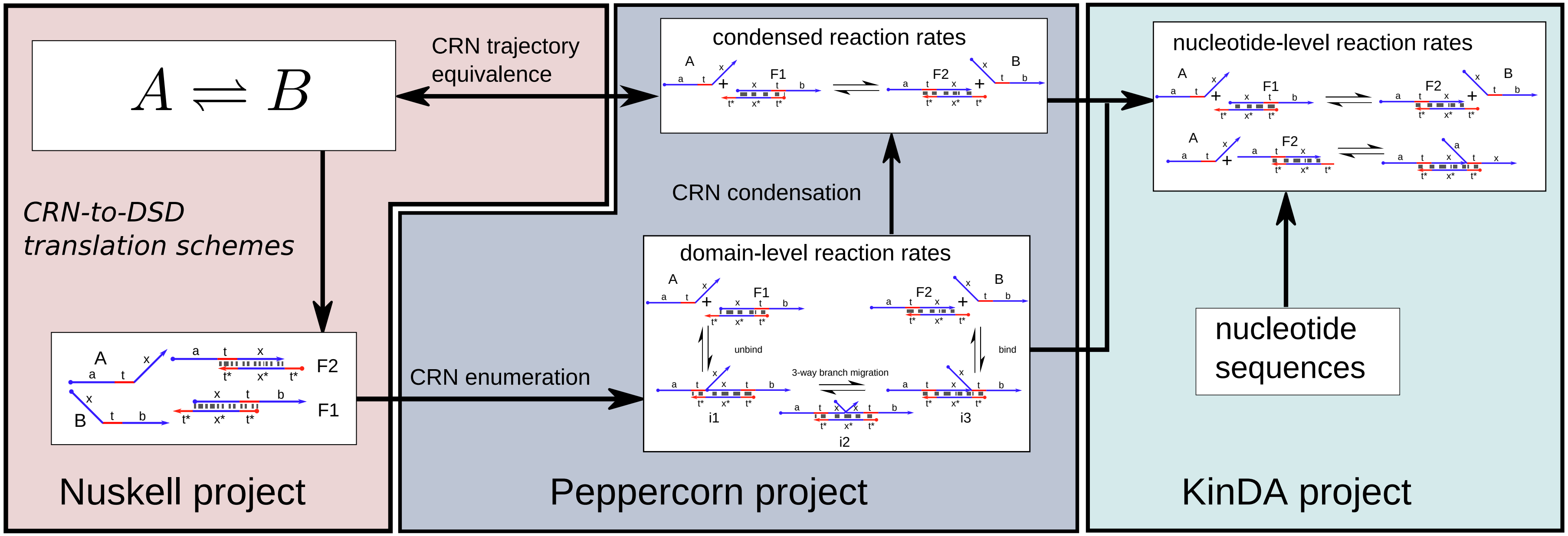
verifies as **correct** according to the **pathway decomposition** and **CRN bisimulation** equivalence

Badelt, Shin, Johnson, Dong, Thachuk and Winfree: A general-purpose CRN-to-DSD compiler with formal verification, optimization, and simulation capabilities. LNCS (2017)

# CONCLUSION

## STRAND DISPLACEMENT IS A KINETIC TOOLBOX ...

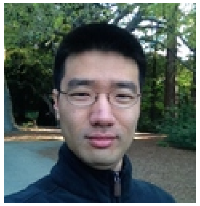
... that can be rigorously analyzed
within the domain of the thermodynamic energy model.




formal verification, choose optimal translation scheme,
simulation using approximate reaction rates

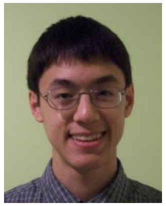
# THANKS TO

Erik Winfree



Seung Woo Shin

Casey Grun



Joseph Berleant

Robert F. Jonhson

Chris Thachuk

---

Frits Dannenberg

Chris Berlind

Karthik V. Sarma

Qing Dong

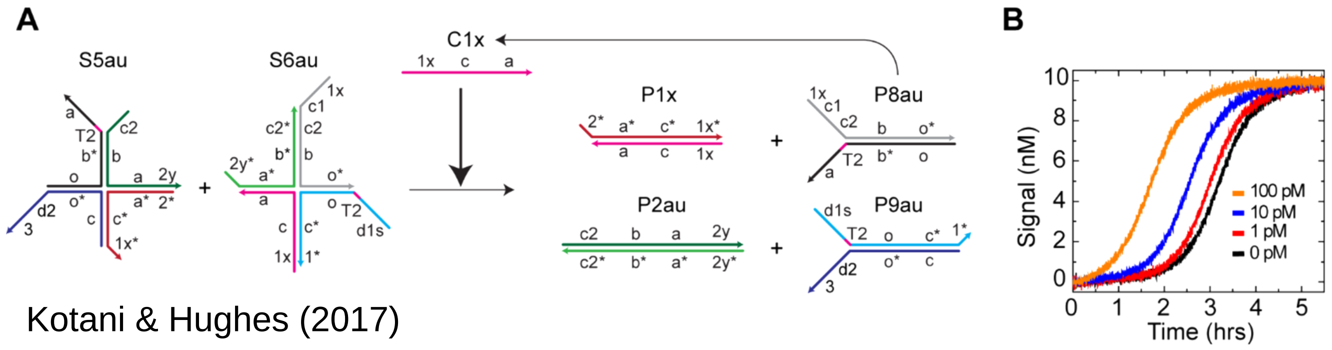
Joseph M. Schaeffer

Brian Wolfe

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

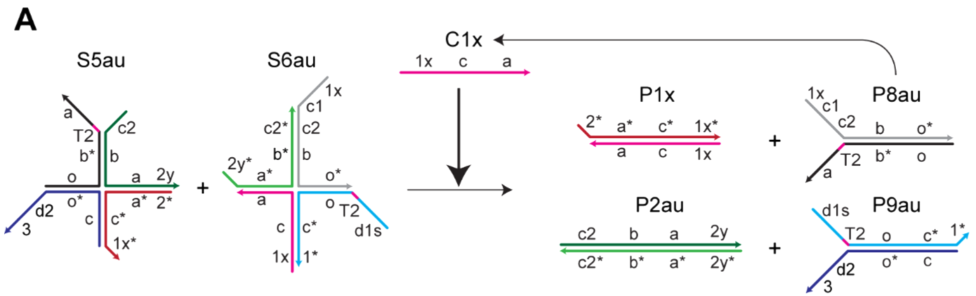
This research was funded in parts by:
The Caltech Biology and Biological Engineering Division Fellowship.
The U.S. National Science Foundation NSF Grant CCF-1213127 and NSF Grant CCF-1317694.
The Gordon and Betty Moore Foundation's Programmable Molecular Technology Initiative (PMTI).

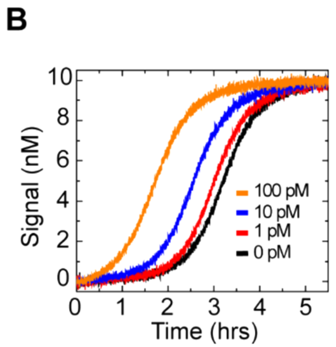
# CASE STUDIES: AUTOCATALYTIC SYSTEM



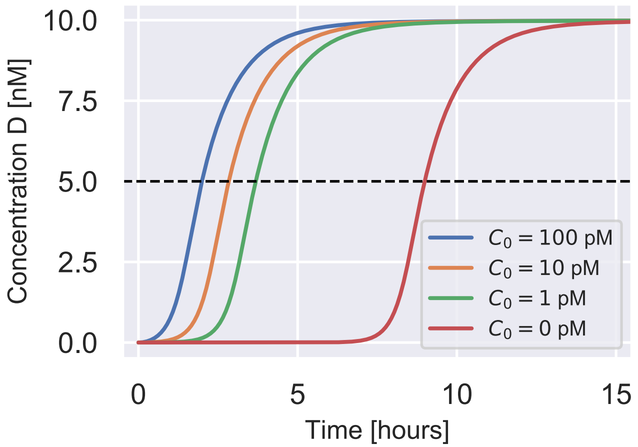

Kotani & Hughes (2017)

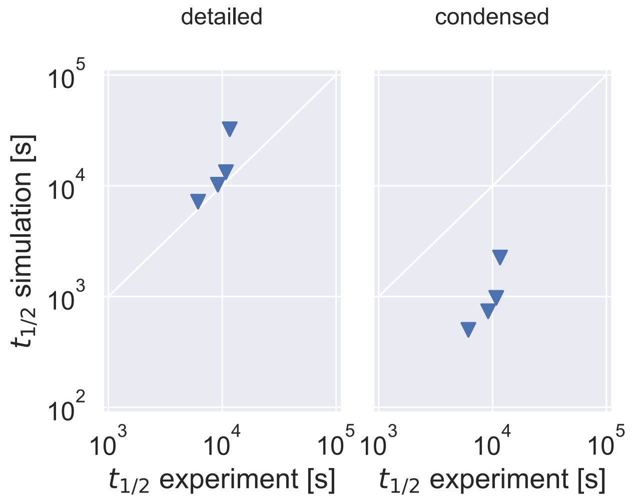
# CASE STUDIES: AUTOCATALYTIC SYSTEM




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 |