EQUIVALENCE OF CHEMICAL REACTION NETWORKS IN A CRN-TO-DNA COMPILER FRAMEWORK

Stefan Badelt and Erik Winfree

DNA and Natural Algorithms (DNA) Group, Caltech

Oxford, July, 19th, 2018

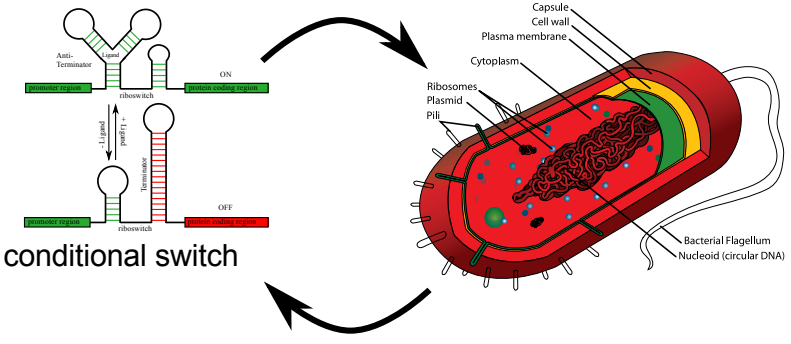
VEMDP 2018

MOLECULAR PROGRAMMING

(in terms of the nuskell compiler project)

nucleic acids are architecture to implement algorithms
chemical reaction networks are a programming language
formal/experimental verification of correct implementation

minimal/optimal components for biological systems

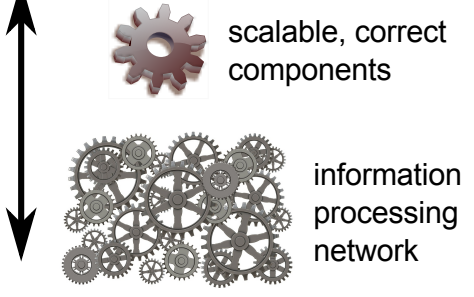


conditional switch

biological relevance is primary
→ if experiments fail, refine the method

verifyably correct artificial systems

arbitrary algorithm

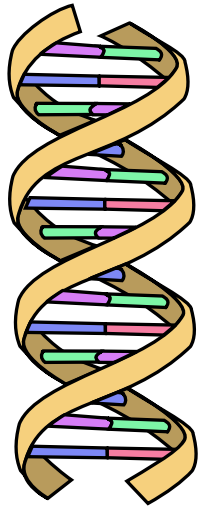


scalable, correct components

information processing network

formal description is primary,
biological relevance secondary

DNA STRAND DISPLACEMENT



DNA

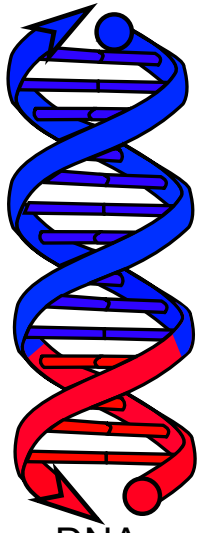
= Adenine

= Thymine

= Cytosine

= Guanine

= Phosphate backbone



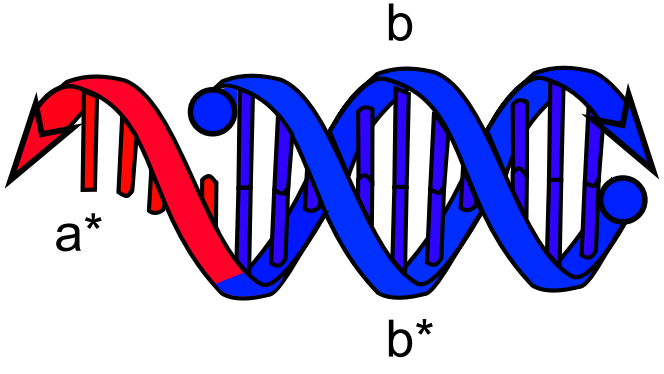
DNA

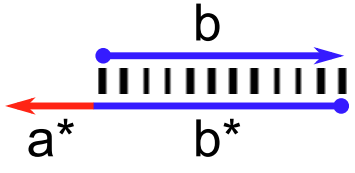
= long domain

= short domain

= 5' end

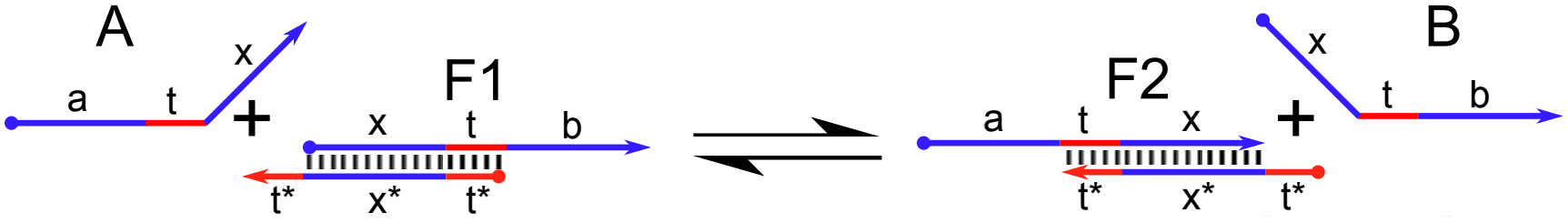
= 3' end



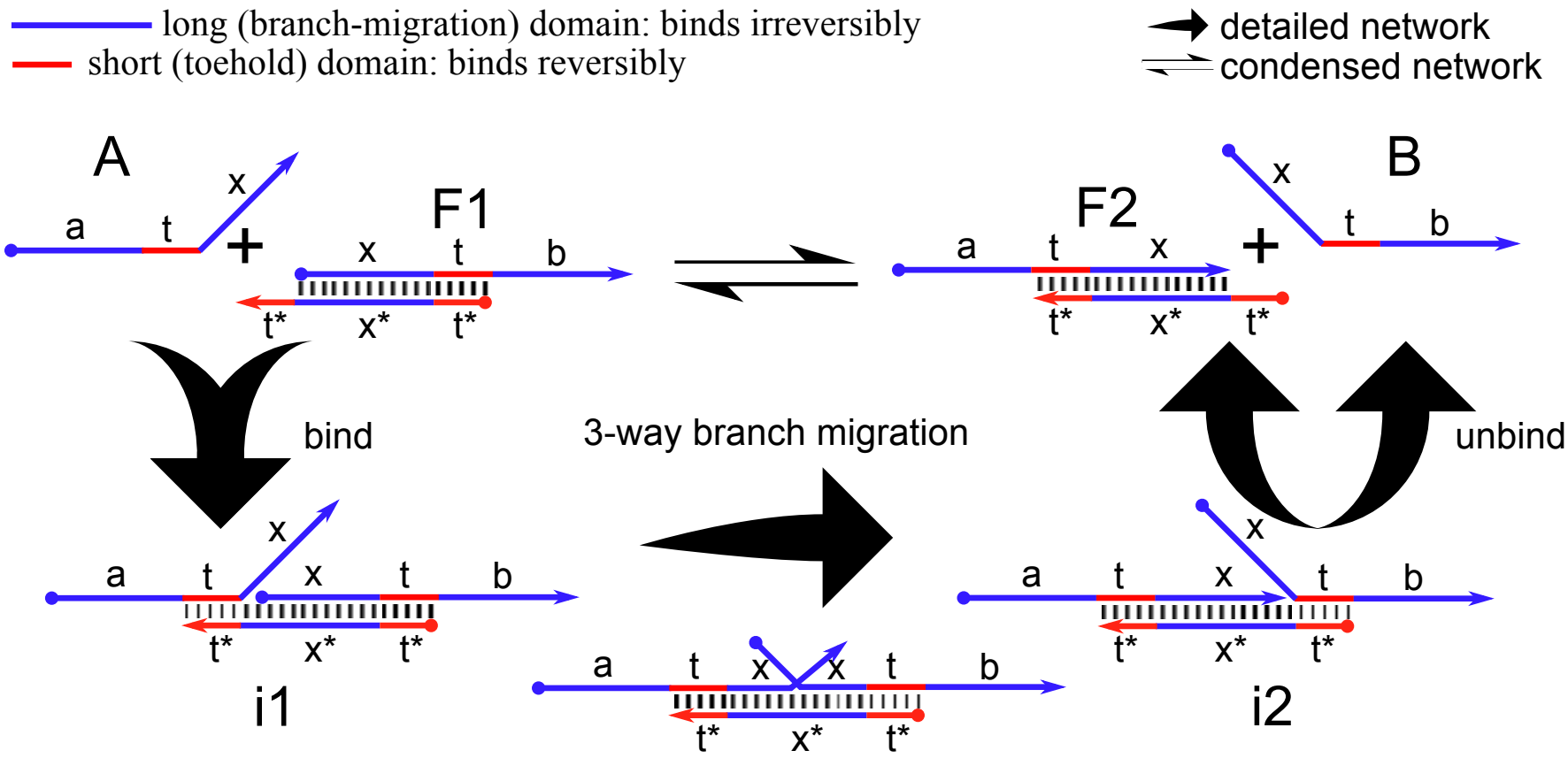


DOMAIN-LEVEL STRAND DISPLACEMENT

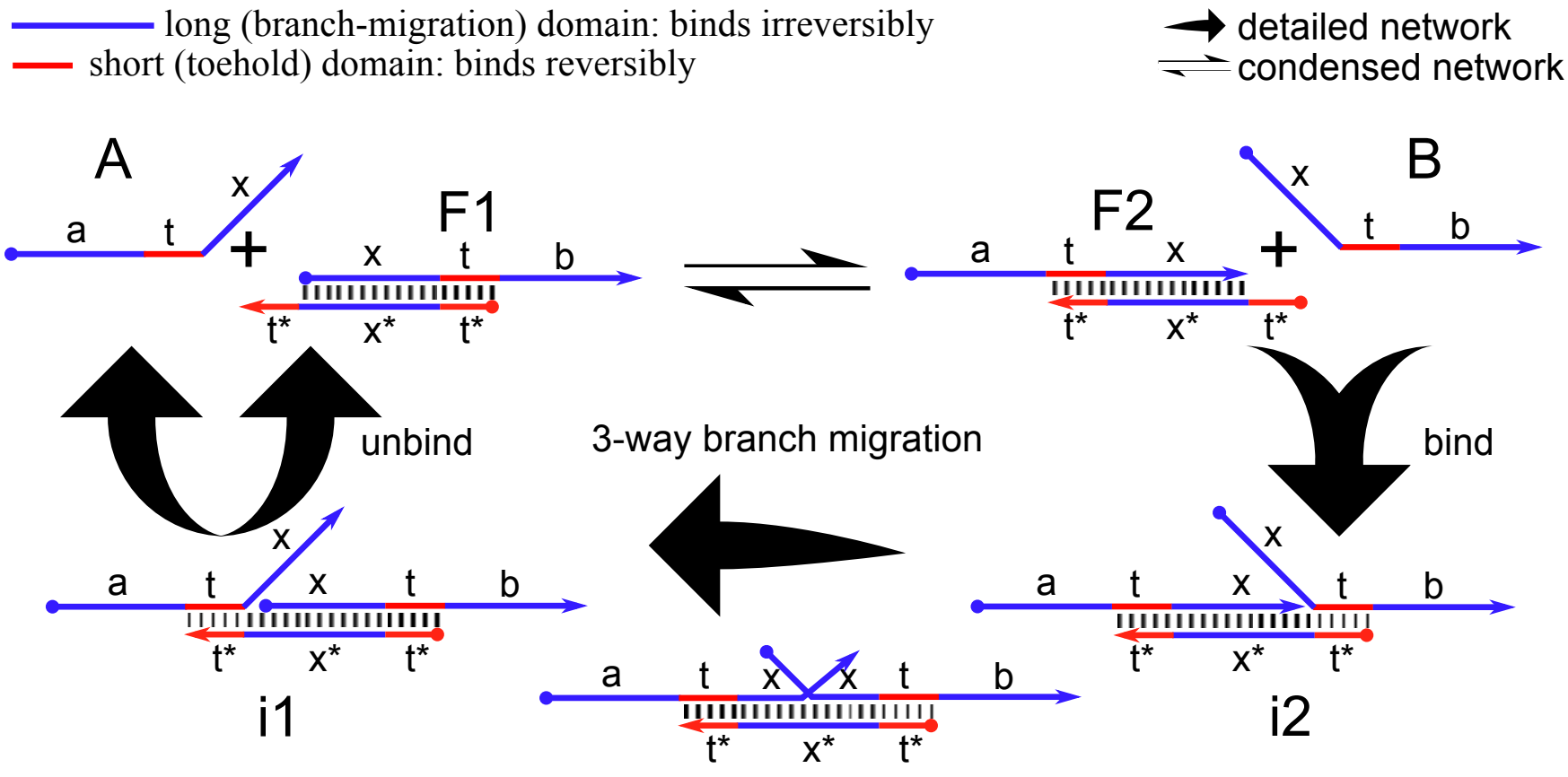
long (branch-migration) domain: binds irreversibly
short (toehold) domain: binds reversibly



DOMAIN-LEVEL STRAND DISPLACEMENT

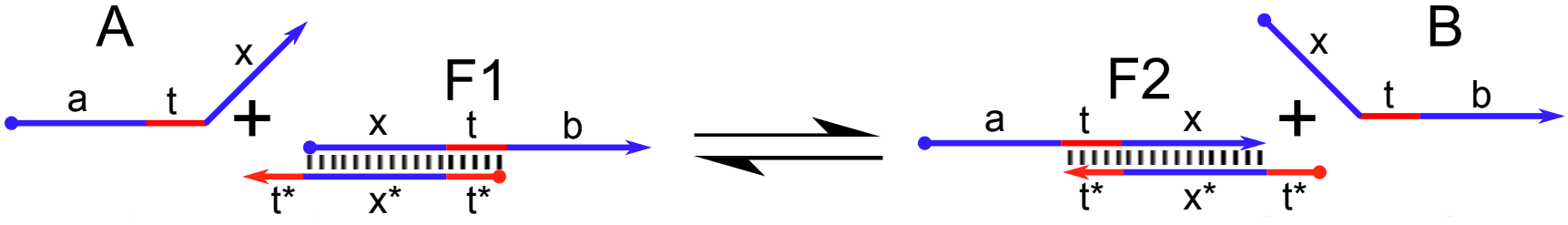


DOMAIN-LEVEL STRAND DISPLACEMENT



DOMAIN-LEVEL STRAND DISPLACEMENT

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



formal CRN

$$A \rightleftharpoons B$$

formal species: {A, B}

DSD sytem specification

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

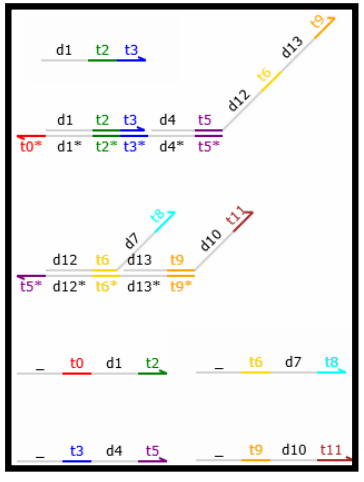
signal species (low concentation): {A, B}

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

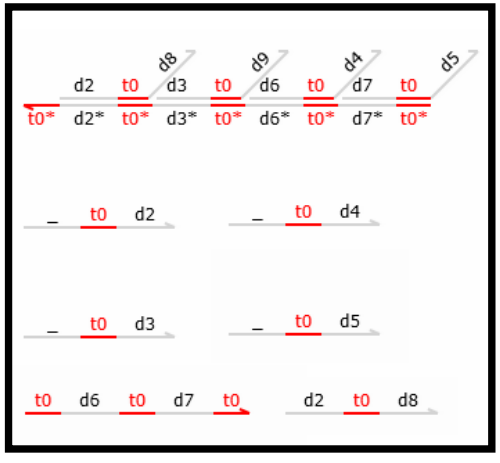
FROM CRN TO DSD SYSTEMS

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

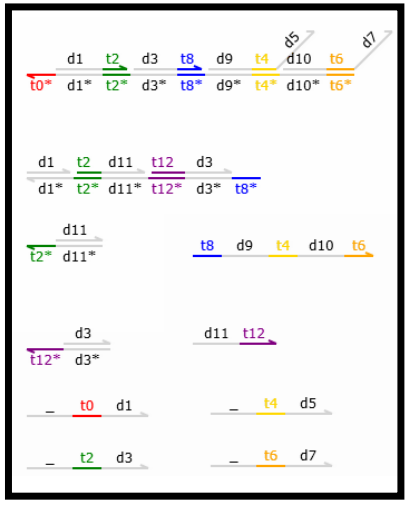
Soloveichik et al. (2010)



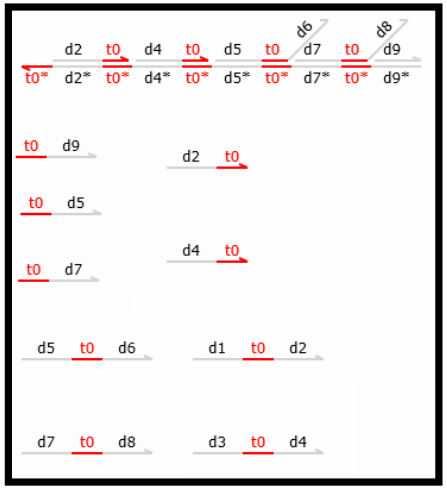
Lakin et al. (2012)



Cardelli (2011)



Qian et al. (2011)



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

Images drawn using VisualDSD, Lakin et al. (2012)

A CRN-TO-DSD COMPILER

try all
translation
schemes

automated
workflow

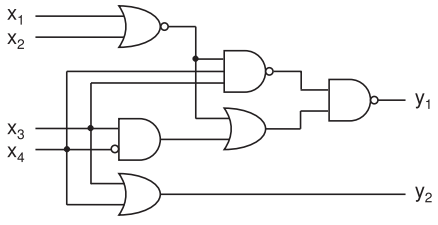
verify
correctness
- for all inputs
- experimental setup

choose
optimal scheme
- synthesis cost
- efficency / robustness
- experimental constraints

sequence design
experimental
testing

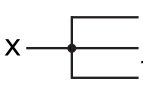
FROM A DIGITAL CIRCUIT TO DSD

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



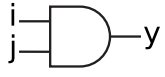
Qian et al. (2011)

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
```

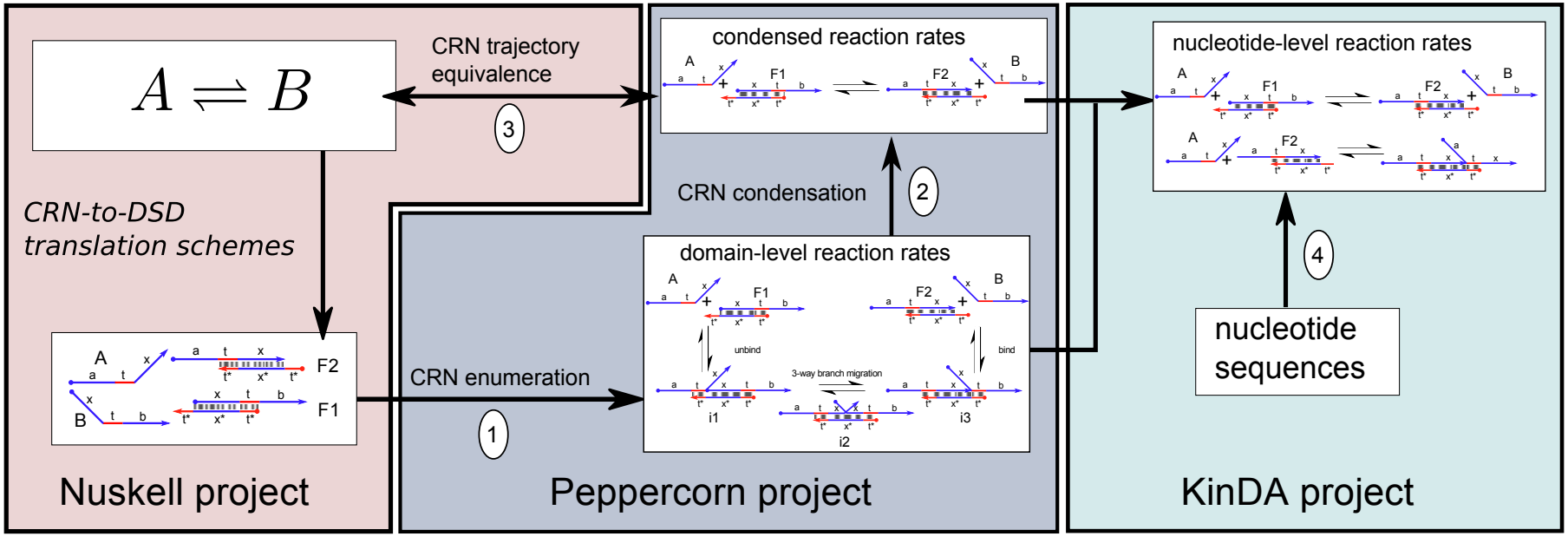
Input for the nuskell compiler: **32** formal reactions.

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

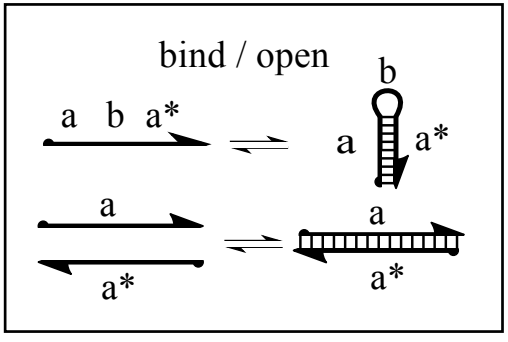
Badelt, Shin, Johnson, Dong, Thachuk and Winfree: A general-purpose CRN-to-DSD compiler with formal verification, optimization, and simulation capabilities. LNCS (2017)

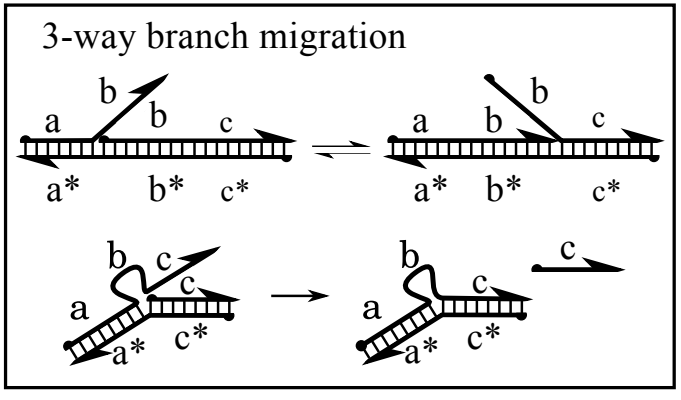
THE COMPILER FRAMEWORK

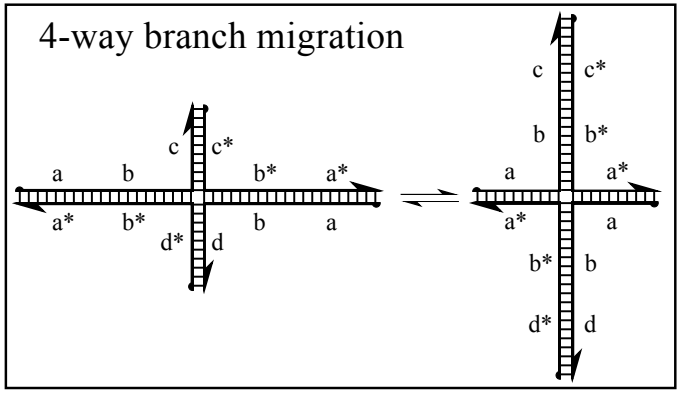


Badelt et al. (2017) - Nuskell
Grun et al. (2014) - Peppercorn
Shin et al. (2017) - CRN pathway decomposition equivalence
Johnson et al. (2018) - CRN bisimulation equivalence
Berleant et al. (submitted) - KinDA

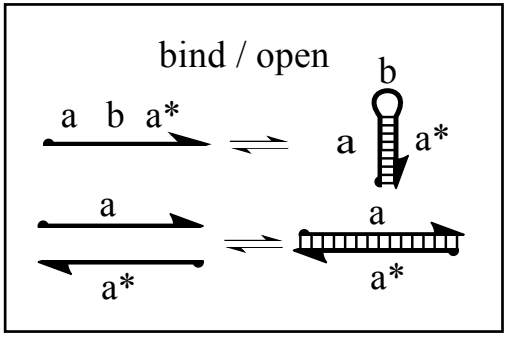
REACTION ENUMERATION







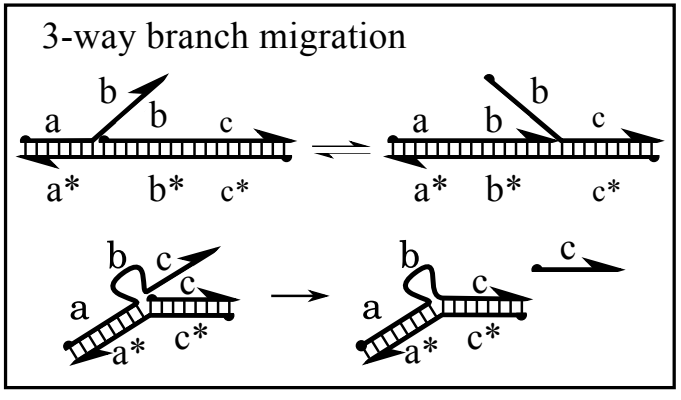
REACTION ENUMERATION

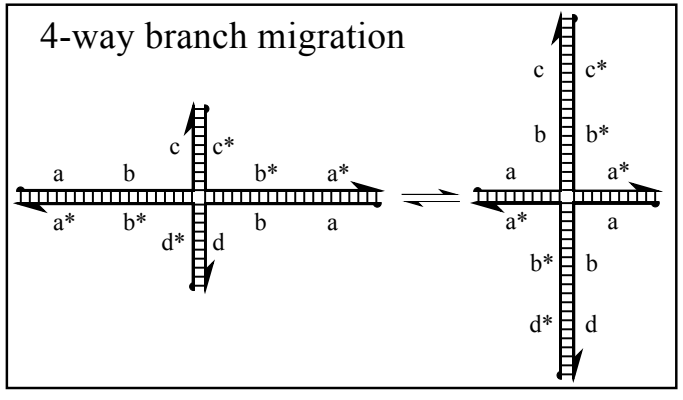


allows all secondary structures (pseudoknots excluded)
open reactions of domains with length > L are forbidden

open & branch migration reactions are always unimolecular, but may lead to dissociation.

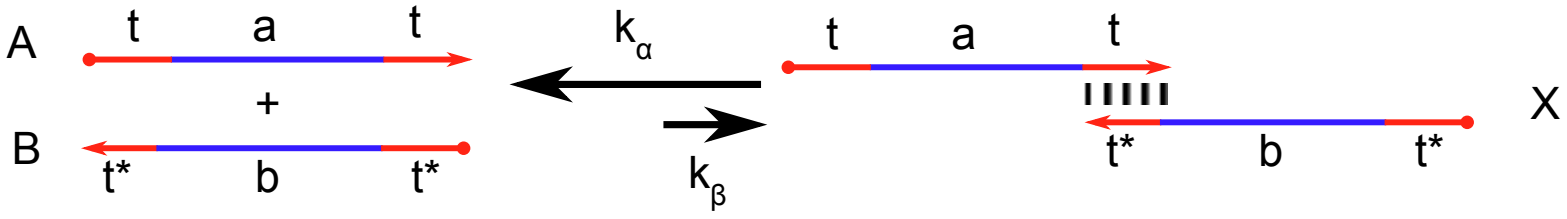
bind reactions are the only valid bimolecular reactions





SEPARATION OF TIMESCALES

unimolecular reactions are fast
bimolecular reactions are slow



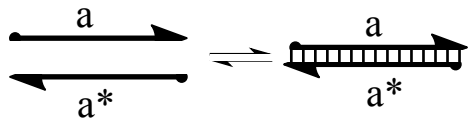
$$\{X \overset{k_\alpha}{\longrightarrow} A + B;\ A + B \overset{k_\beta}{\longrightarrow} X\}$$

at low concentrations:

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

APPROXIMATE REACTION RATE CONSTANTS

bimolecular binding

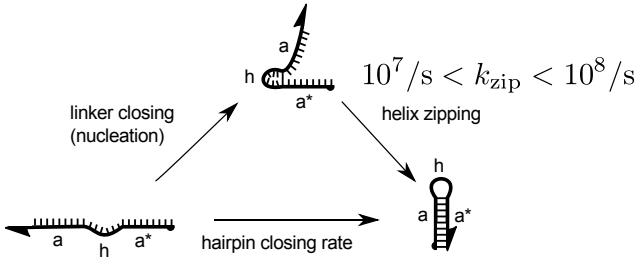


$$k_{\text{bind21}} = l * 3 * 10^5 \ \text{M}^{-1} \ \text{s}^{-1}$$

open rate

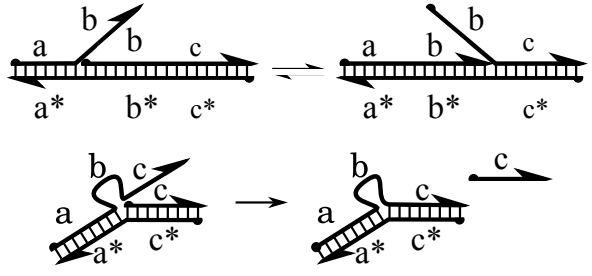
$$\frac{k_{\text{open}}}{k_{\text{bind21}}} = e^{\frac{\Delta G}{RT}}$$

unimolecular binding



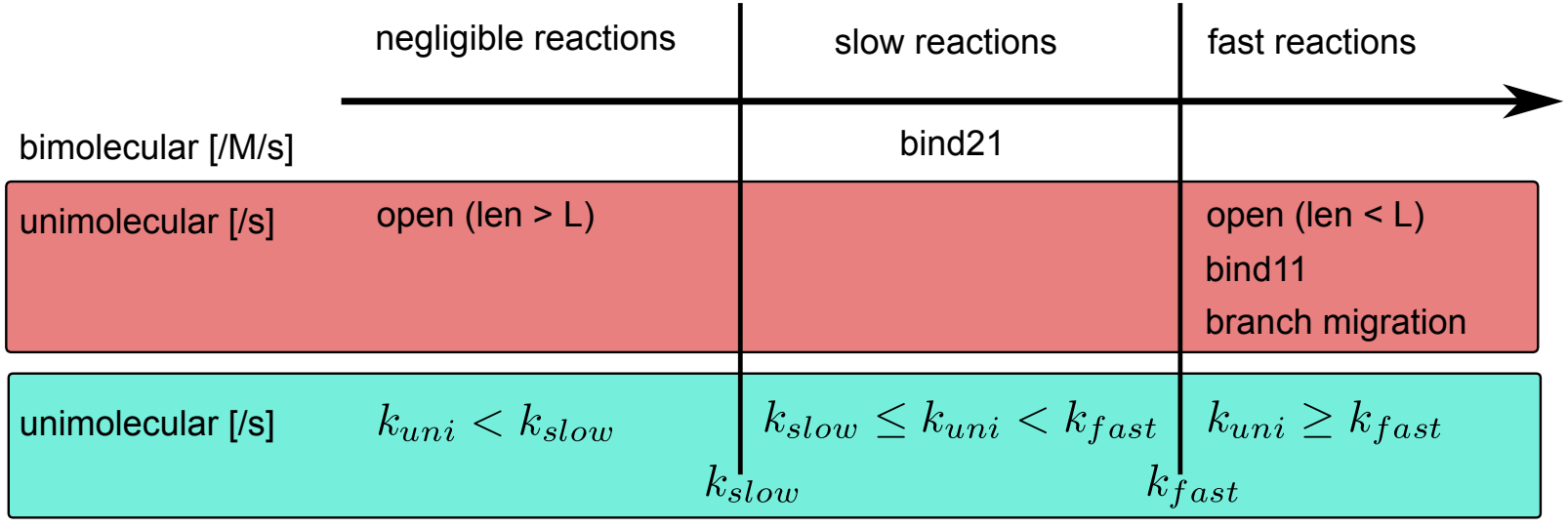
$$k_{\text{bind11}} = \mathcal{C}(l) * 10^6 \ \text{s}^{-1}$$

branch migration



$$k_{\text{bm}} = \frac{\mathcal{C}(l) * k_{\text{bm-init}}}{l} \text{s}^{-1}$$

MODEL PARAMETERS



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

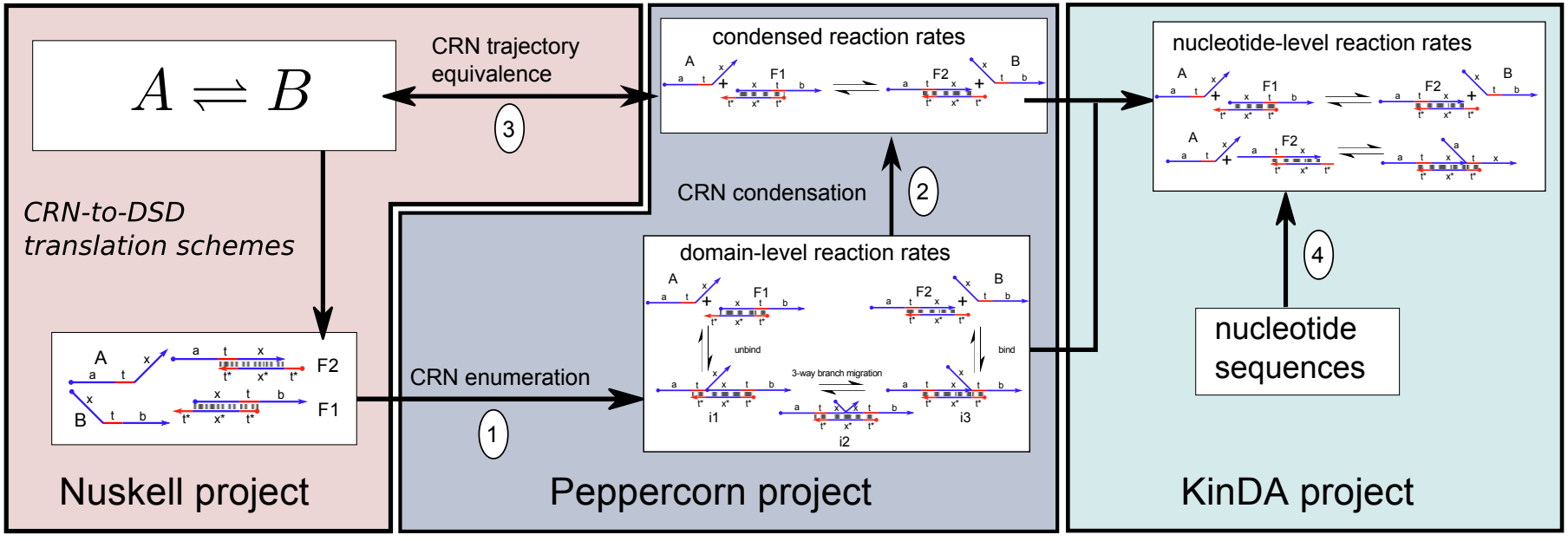
REACTION ENUMERATION

- 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

THE COMPILER FRAMEWORK



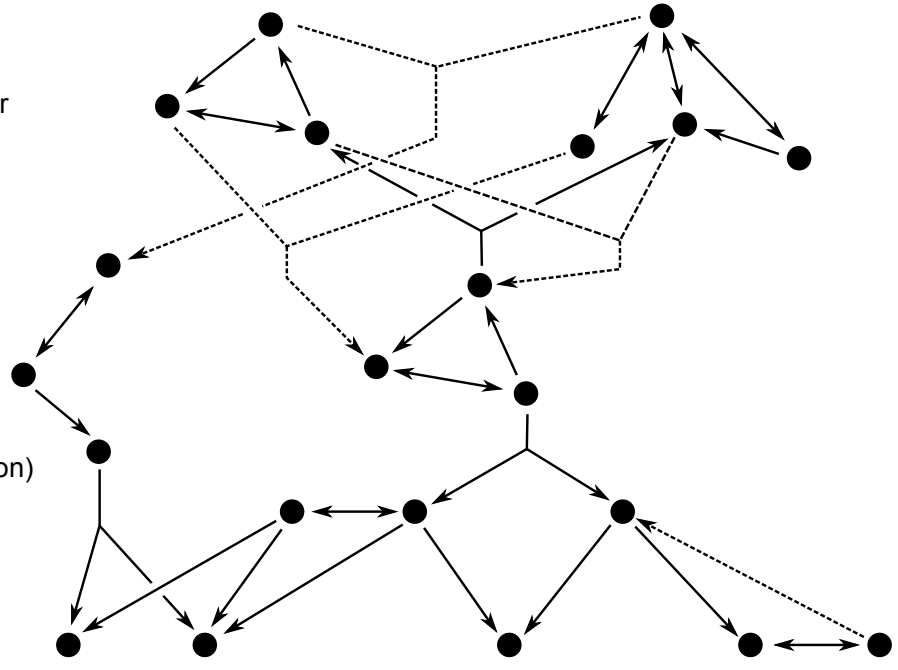
Badelt et al. (2017) - Nuskell
Grun et al. (2014) - Peppercorn
Shin et al. (2017) - CRN pathway decomposition equivalence
Johnson et al. (2018) - CRN bisimulation equivalence
Berleant et al. (submitted) - KinDA

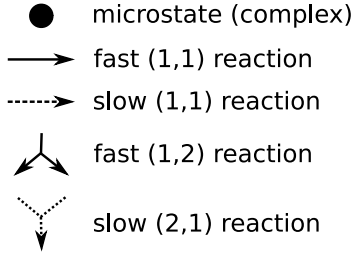
CRN CONDENSATION

Goal: represent CRN in terms of overall slow reactions

properties / requirements:

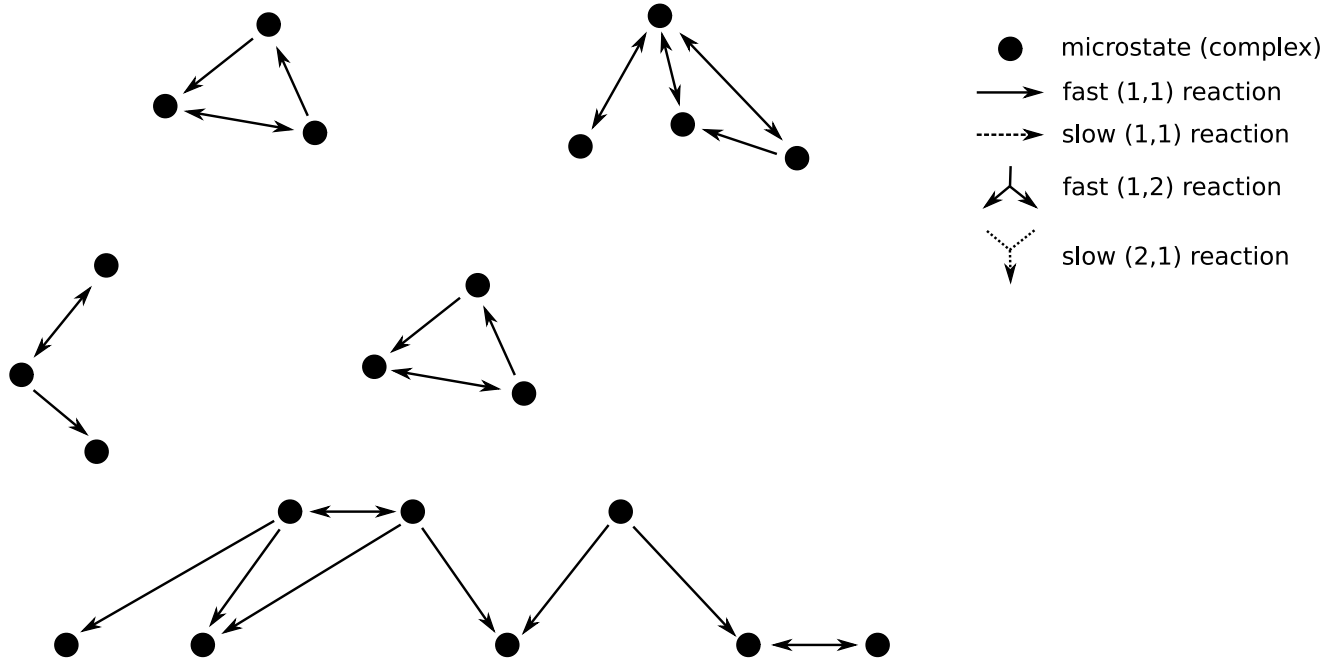
- all fast reactions are unimolecular
- reactions have arity (n,m) with n > 0 and m > 0
- reactants of slow reactions must be resting states
- reactants and products of fast (1-2) reactions are in different SCCs (mass conservation)





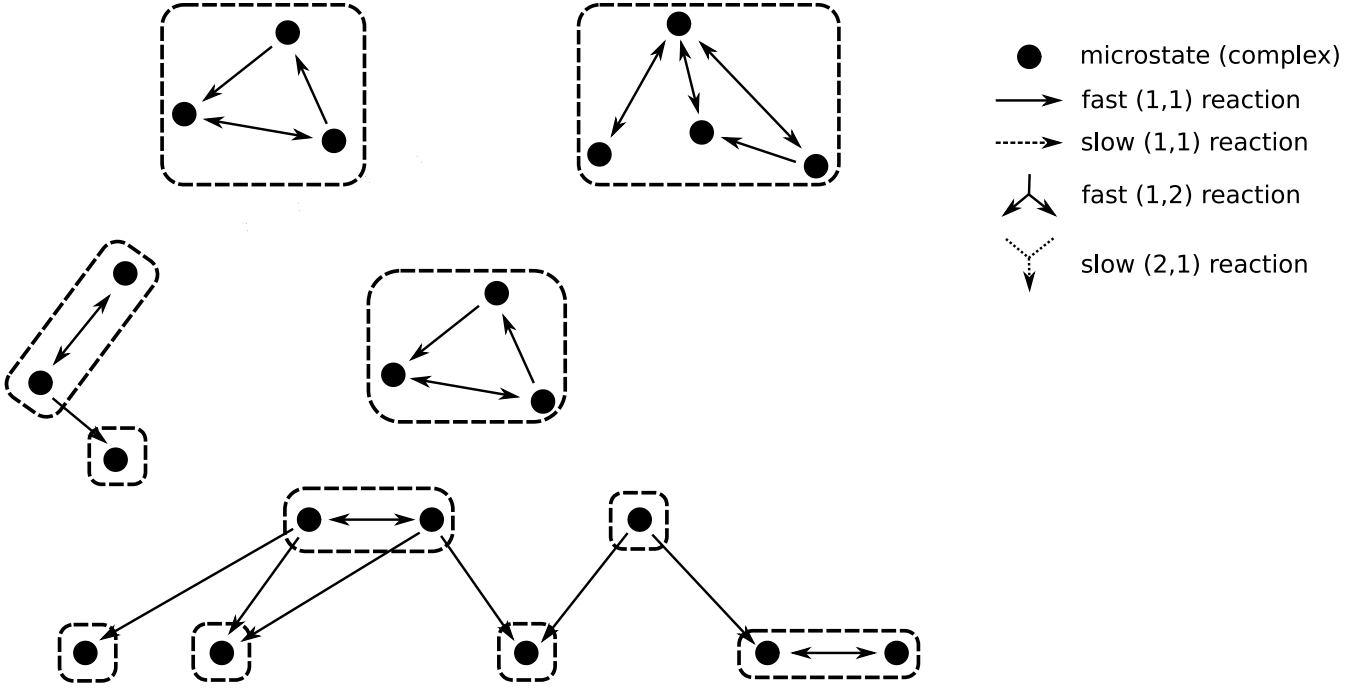
CRN CONDENSATION

Step 1: Make a graph that contains only fast (1,1) reactions



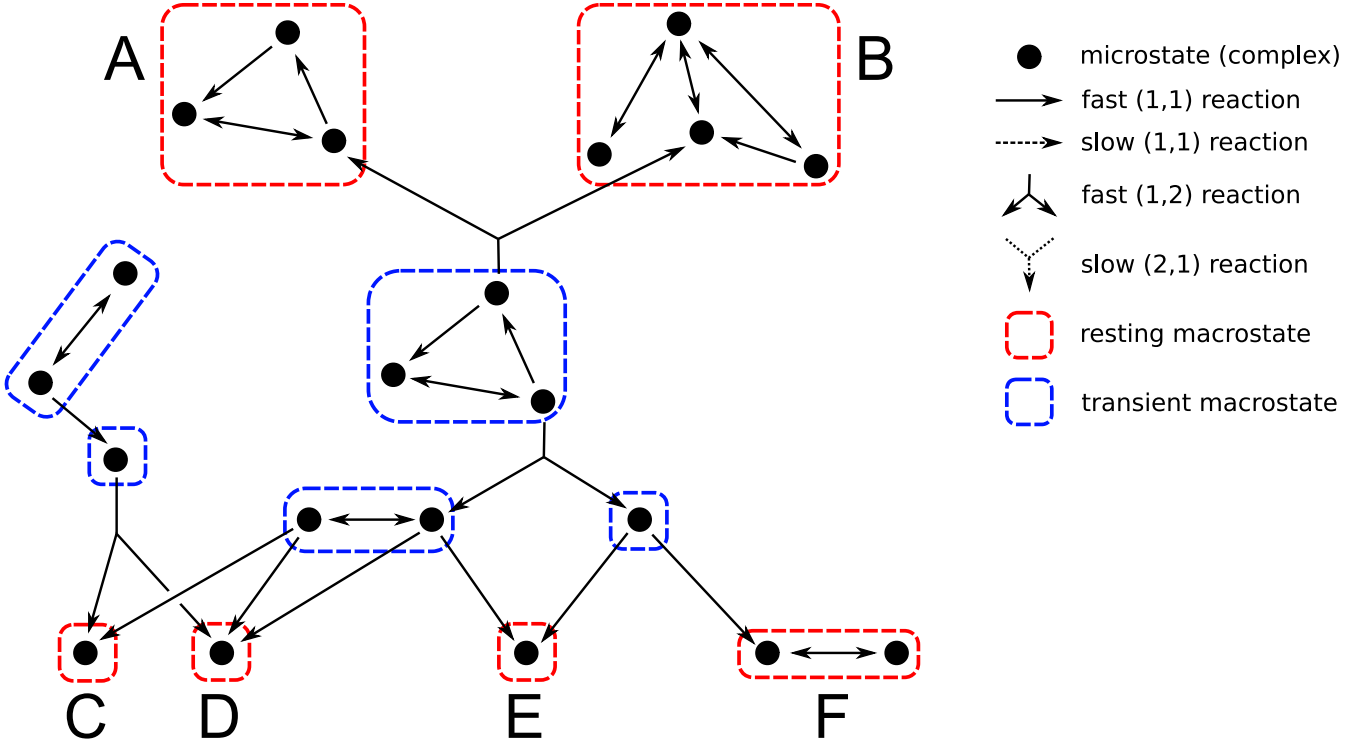
CRN CONDENSATION

Step 2: Identify strongly connected components (SCCs)



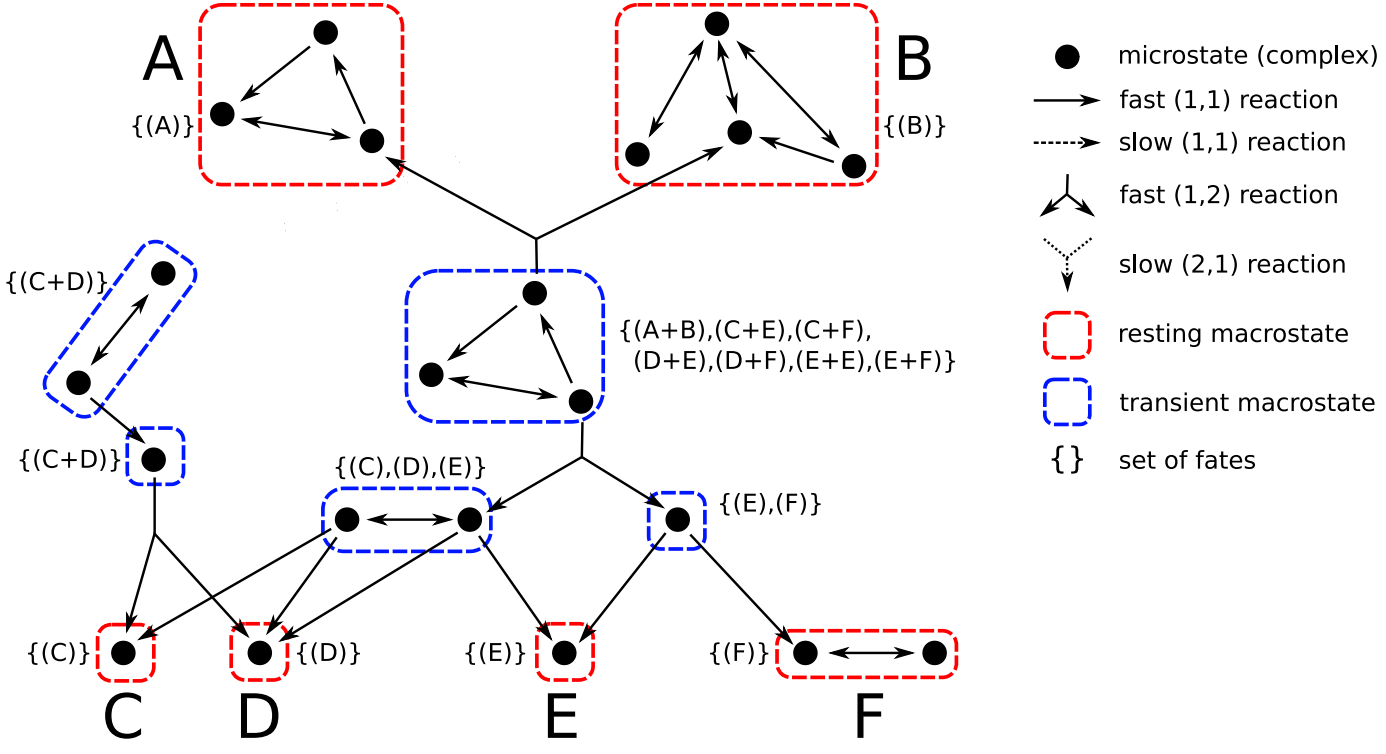
CRN CONDENSATION

Step 3: Define transient and resting macrostates



CRN CONDENSATION

Step 4: Assign fates to complexes (or macrostates)

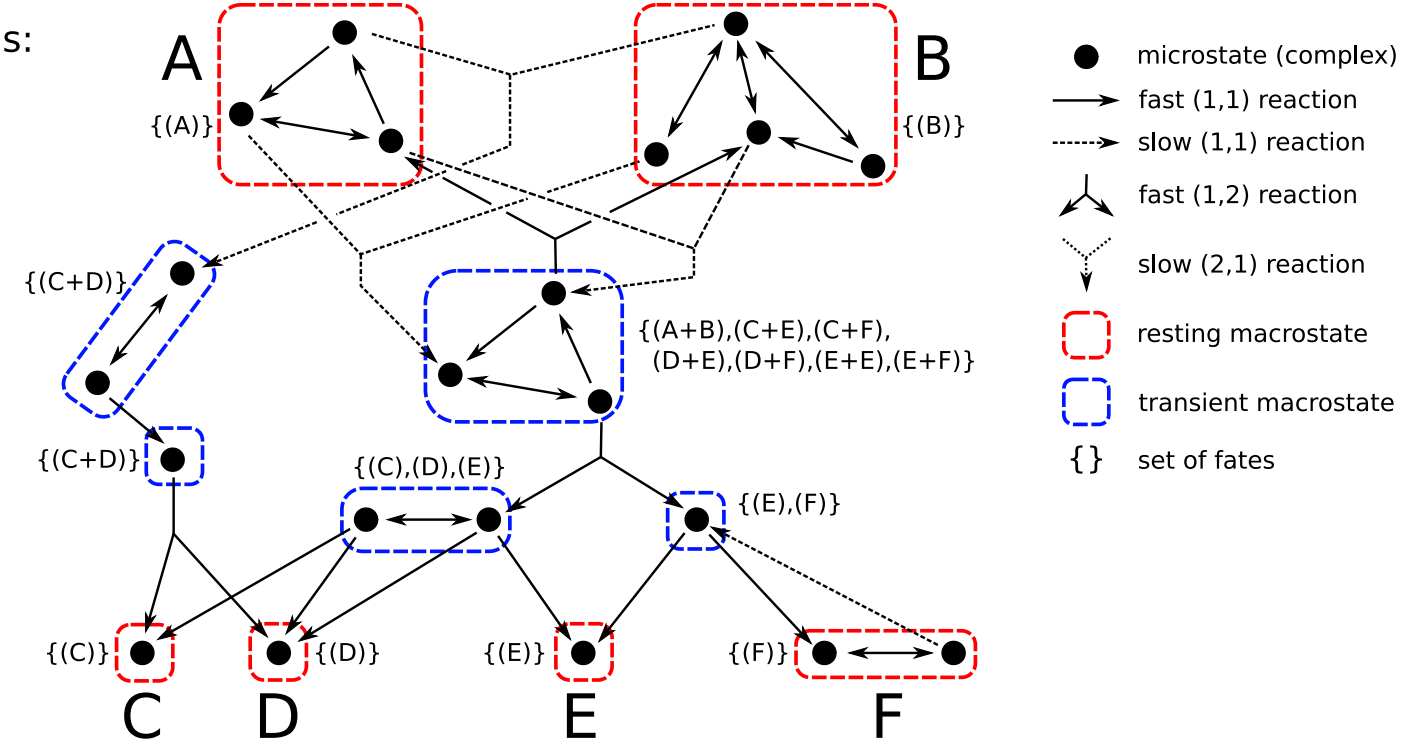


CRN CONDENSATION

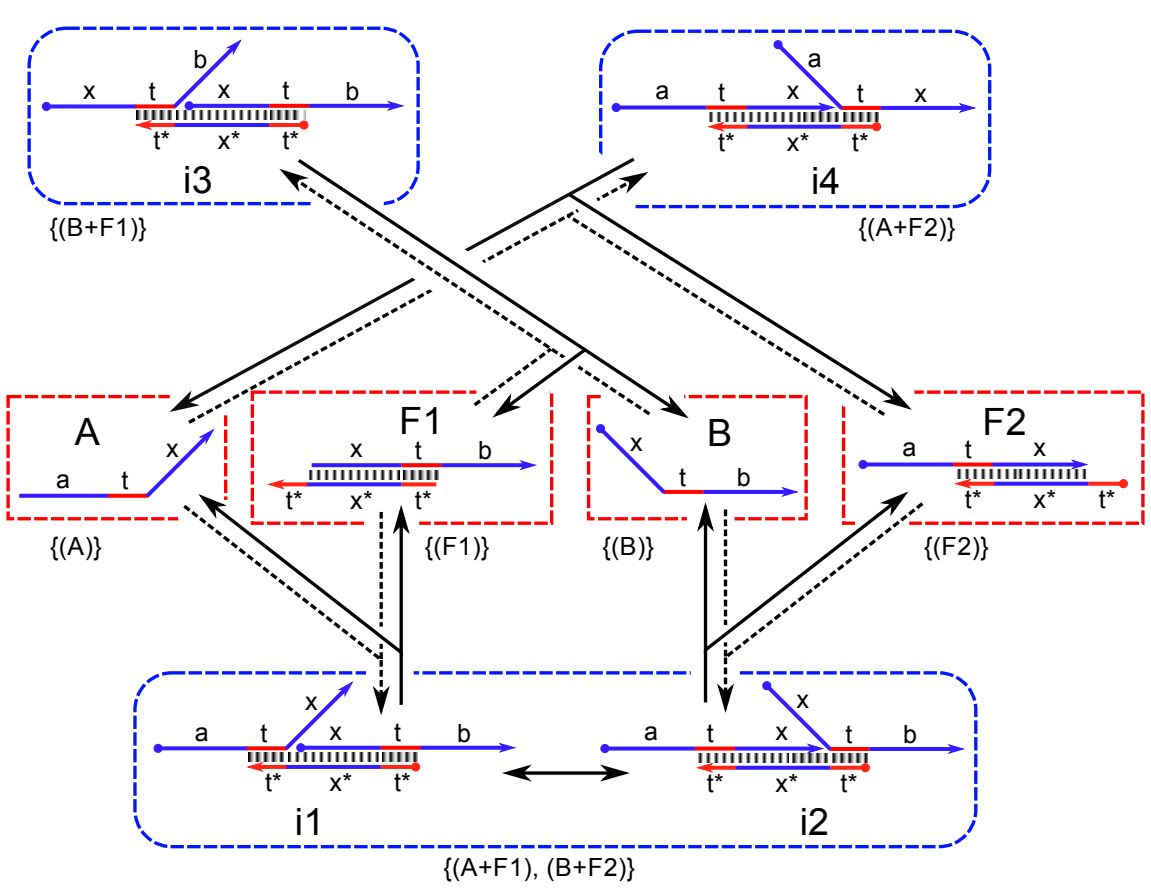
Step 5: Insert slow reactions & derive condensed reactions

condensed reactions:

```
A+B  ->  A+B
A+B  ->  C+D
A+B  ->  C+E
A+B  ->  C+F
A+B  ->  D+E
A+B  ->  D+F
A+B  ->  E+E
A+B  ->  E+F
F  ->  F
F  ->  E
```



DSD CONDENSATION



fast (1,1) reaction

fast (1,2) reaction

slow (2,1) reaction

resting macrostate

transient macrostate

{} set of fates

detailed reactions:

```
A + F1 -> i1
i1 -> i2
i2 -> B + F2
B + F2 -> i2
i2 -> i1
i1 -> A + F1
A + F2 -> i4
i4 -> A + F2
B + F1 -> i3
i3 -> B + F1
```

condensed reactions:

```
A + F1 -> B + F2
B + F2 -> A + F1
```

REACTION RATE CONDENSATION

Consider a condensed reaction:

$$P + Q \rightarrow K + L + M$$

It is composed of all detailed slow reactions:

$$p + q \rightarrow I$$

weighted by the decay probability over all pathways:

$$I \rightarrow \cdots \rightarrow k + l + m$$

where $p \in P, q \in Q, k \in K, l \in L, m \in M$
and I is a multiset of intermediate species

REACTION RATE CONDENSATION

Notation:

detailed reaction: $r = (A, B) \qquad A = \{|a_i|\}$

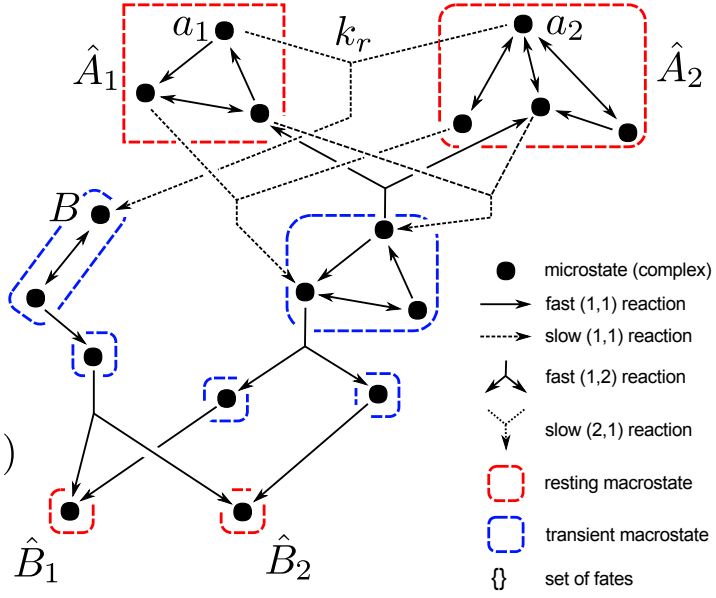
condensed reaction: $\hat{r} = (\hat{A}, \hat{B}) \qquad \hat{A} = \{|\hat{A}_i|\}$

given: $\hat{A} = (\hat{A}_1, \hat{A}_2) \quad \hat{B} = (\hat{B}_1, \hat{B}_2)$

define: $R_{\hat{A}} = \{r = ((a_1, a_2), B) : a_1 \in \hat{A}_1, a_2 \in \hat{A}_2\}$

then the condensed rate is:

$$k_{\hat{r}} = \sum_{r=((a_1,a_2),B)\in R_{\hat{A}}} P(a_1|\hat{A}_1) \cdot P(a_2|\hat{A}_2) \cdot k_r \cdot P(T_{B\to\hat{B}})$$



REACTION RATE CONDENSATION

general form:

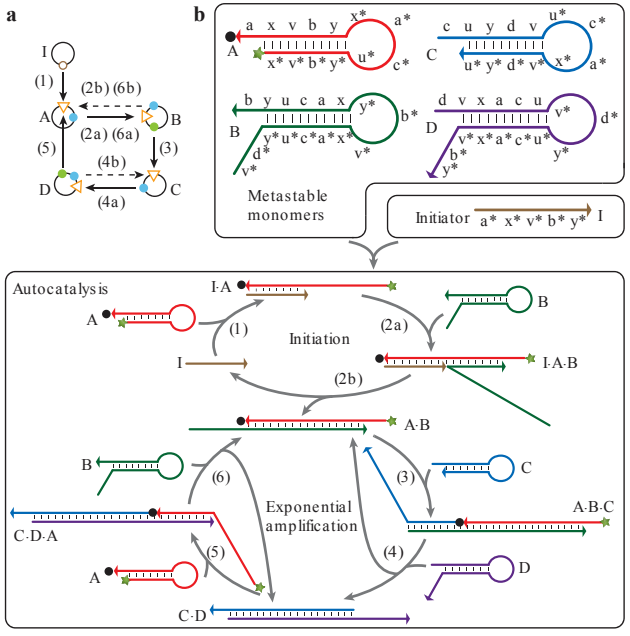
$$k_{\hat{r}} = \sum_{r=(A,B)\in R_{\hat{A}}} k_r \cdot \mathbb{P}[T_{B\to\hat{B}}] \cdot \prod_{a_i \in A} \mathbb{P}[a_i : \hat{A}_i]$$

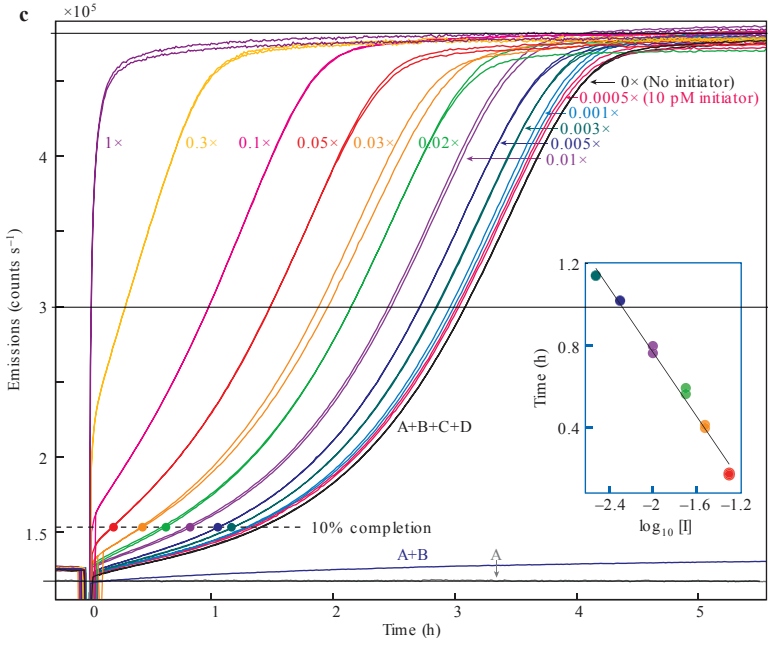
where

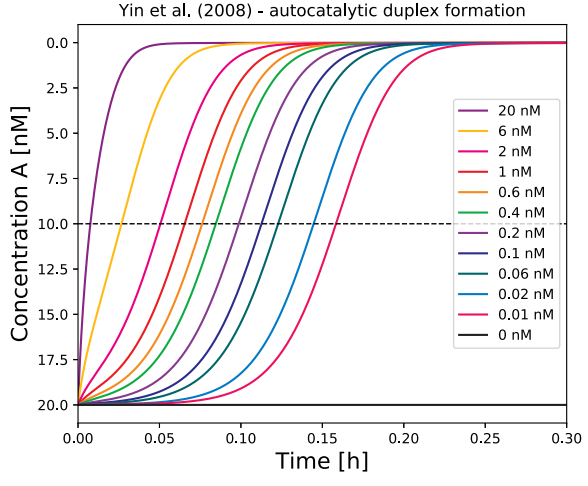
$\mathbb{P}[a_i : \hat{A}_i] =$ stationary distribution

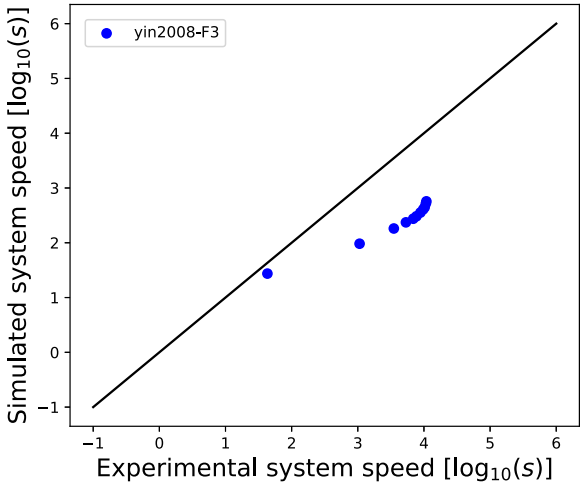
$\mathbb{P}[T_{B\to\hat{B}}] =$ reaction decay probability

EXPERIMENTAL DATA: YIN ET AL. (2008)

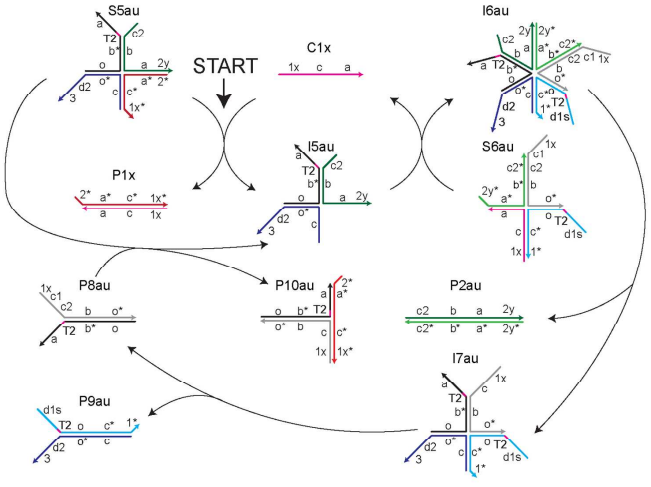


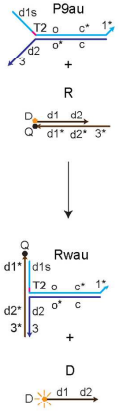


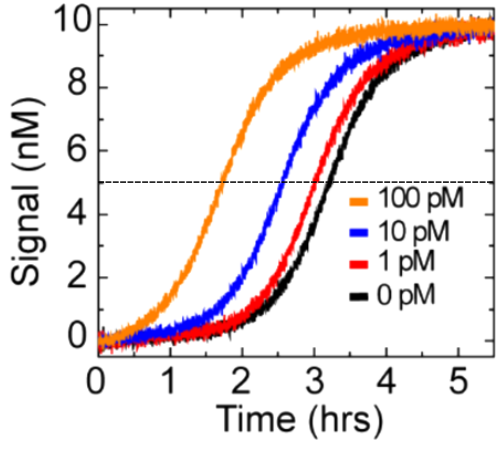


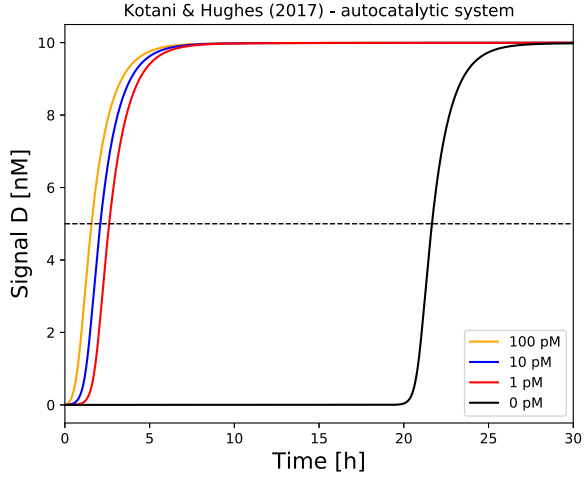


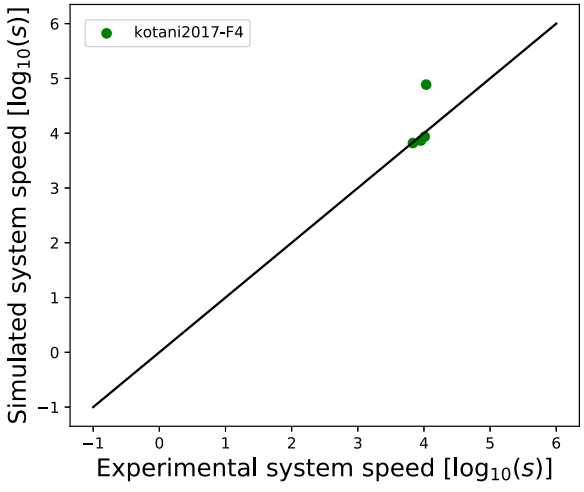
EXPERIMENTAL DATA: KOTANI & HUGHES (2017)



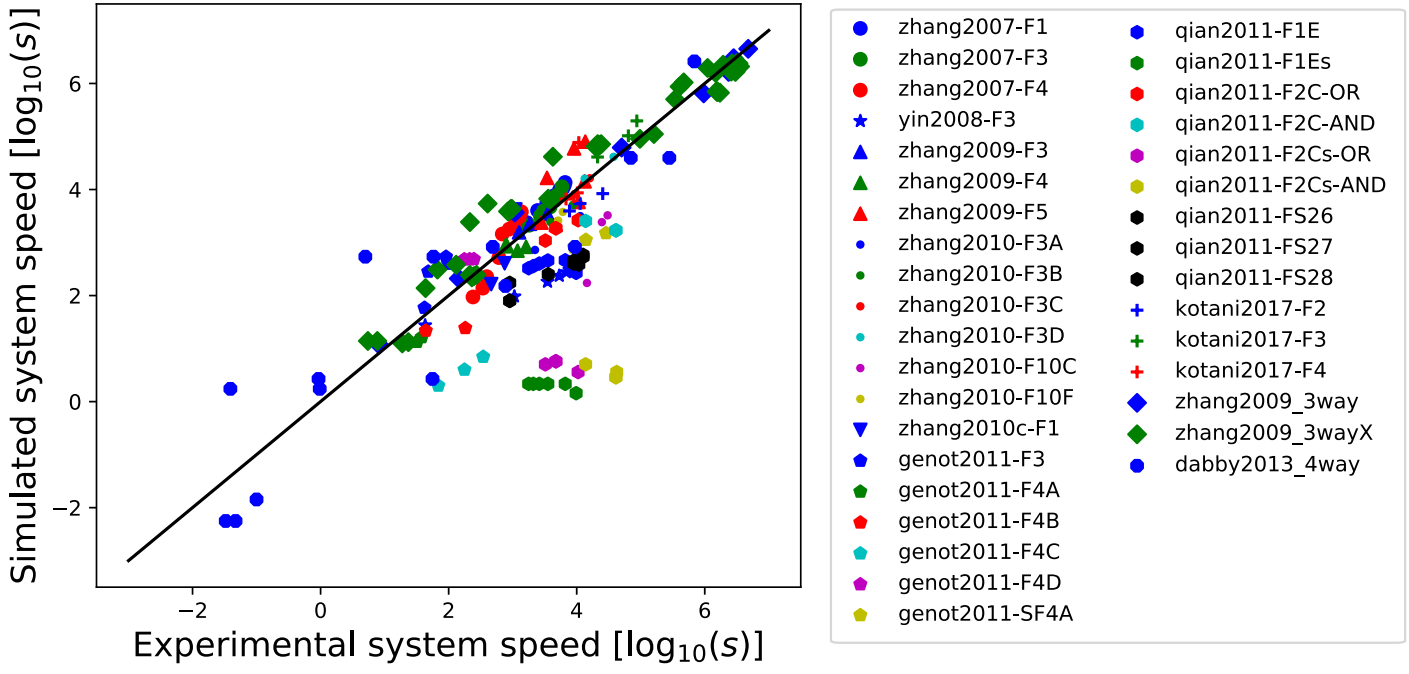




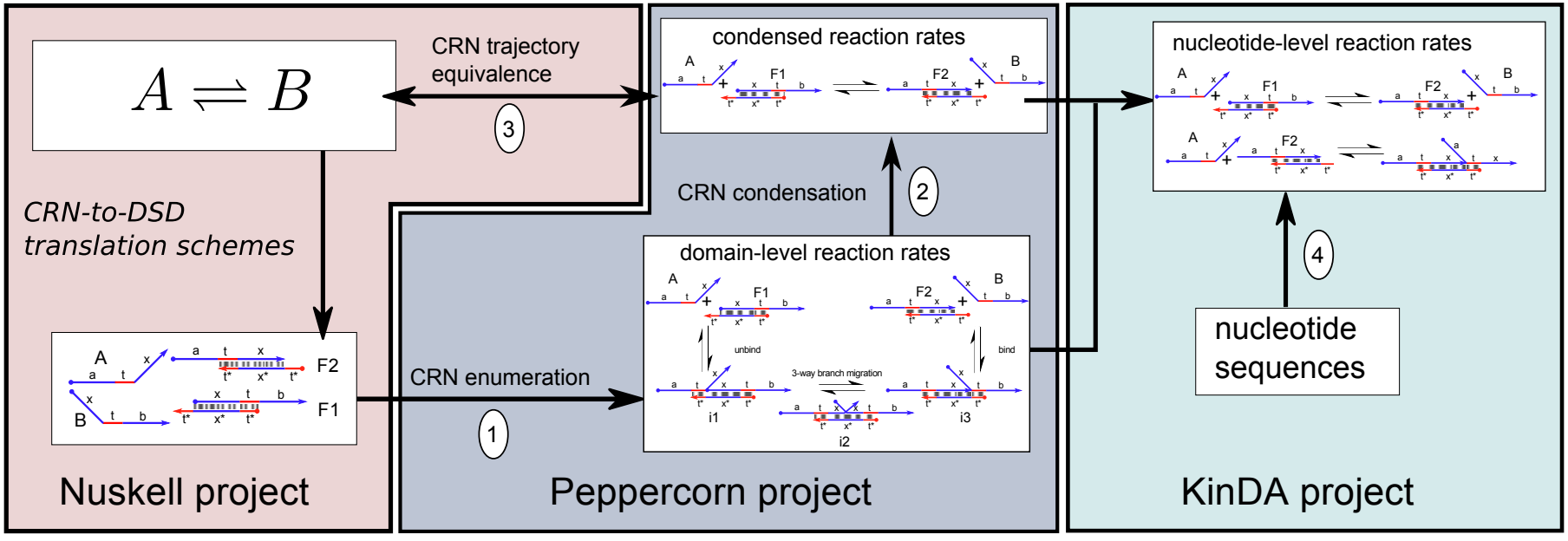




MORE EXPERIMENTAL DATA (2009 - 2017)



THE COMPILER FRAMEWORK



Badelt et al. (2017) - Nuskell
Grun et al. (2014) - Peppercorn
Shin et al. (2017) - CRN pathway decomposition equivalence
Johnson et al. (2018) - CRN bisimulation equivalence
Berleant et al. (submitted) - KinDA

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

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

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

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
```

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
```

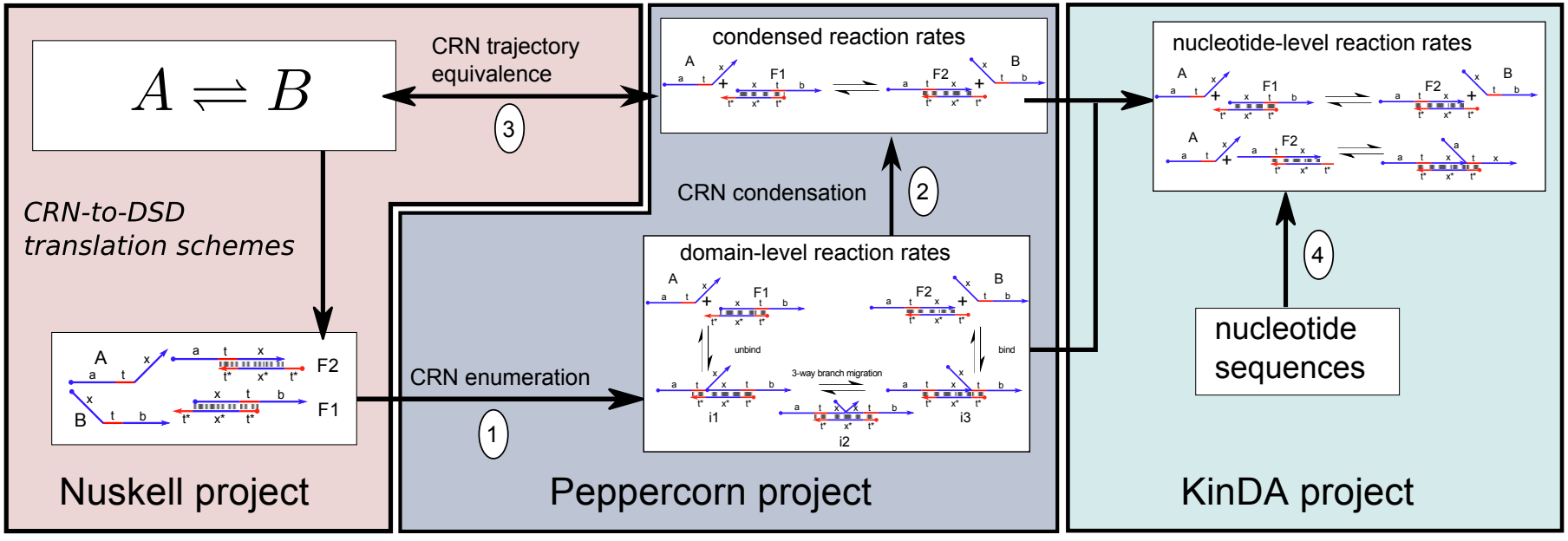
enumerated CRN
360 species
668 reactions

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
```

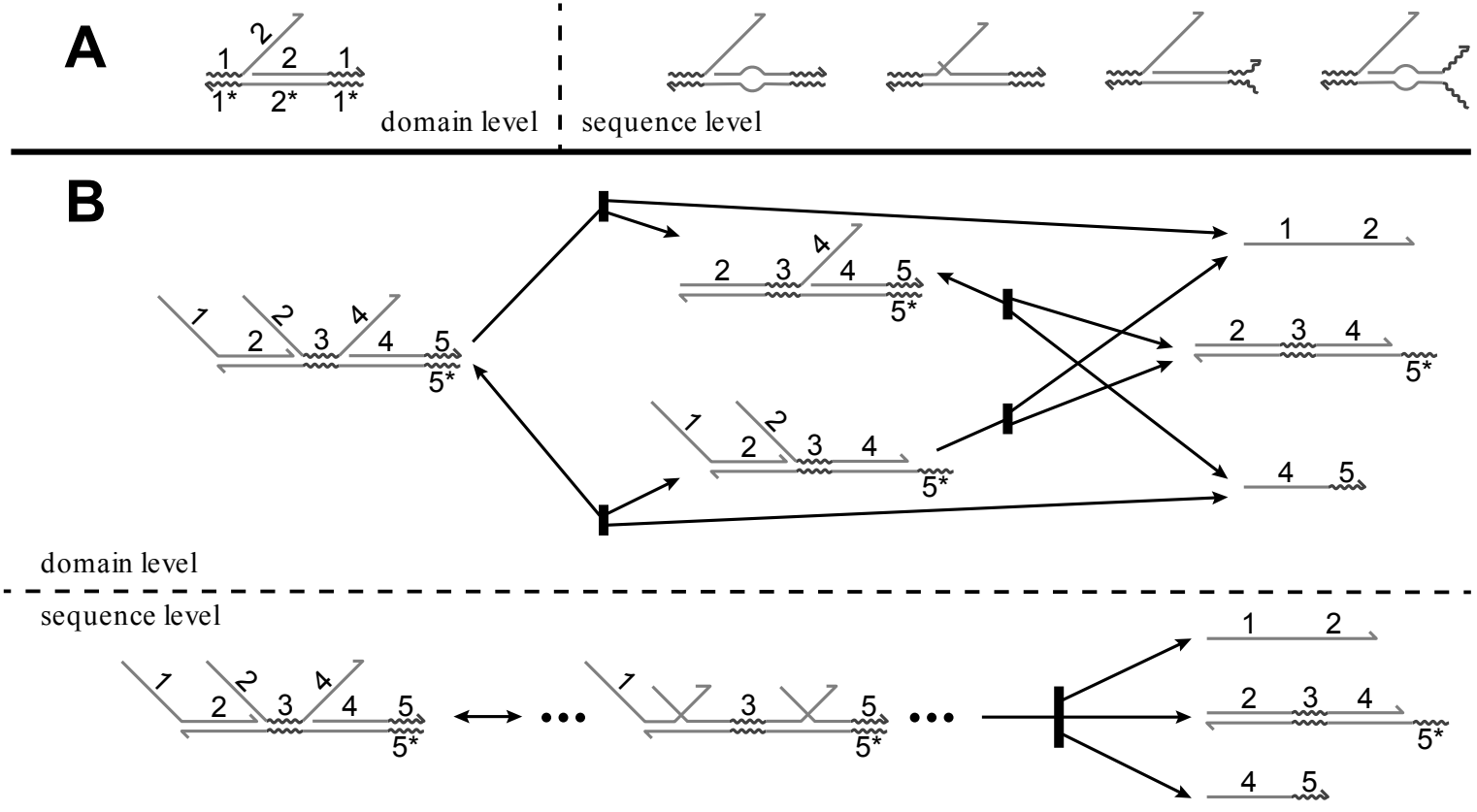
Johnson et al. (2016) - CRN bisimulation equivalence
translation scheme: qian2011_3D_var1.ts

THE COMPILER FRAMEWORK



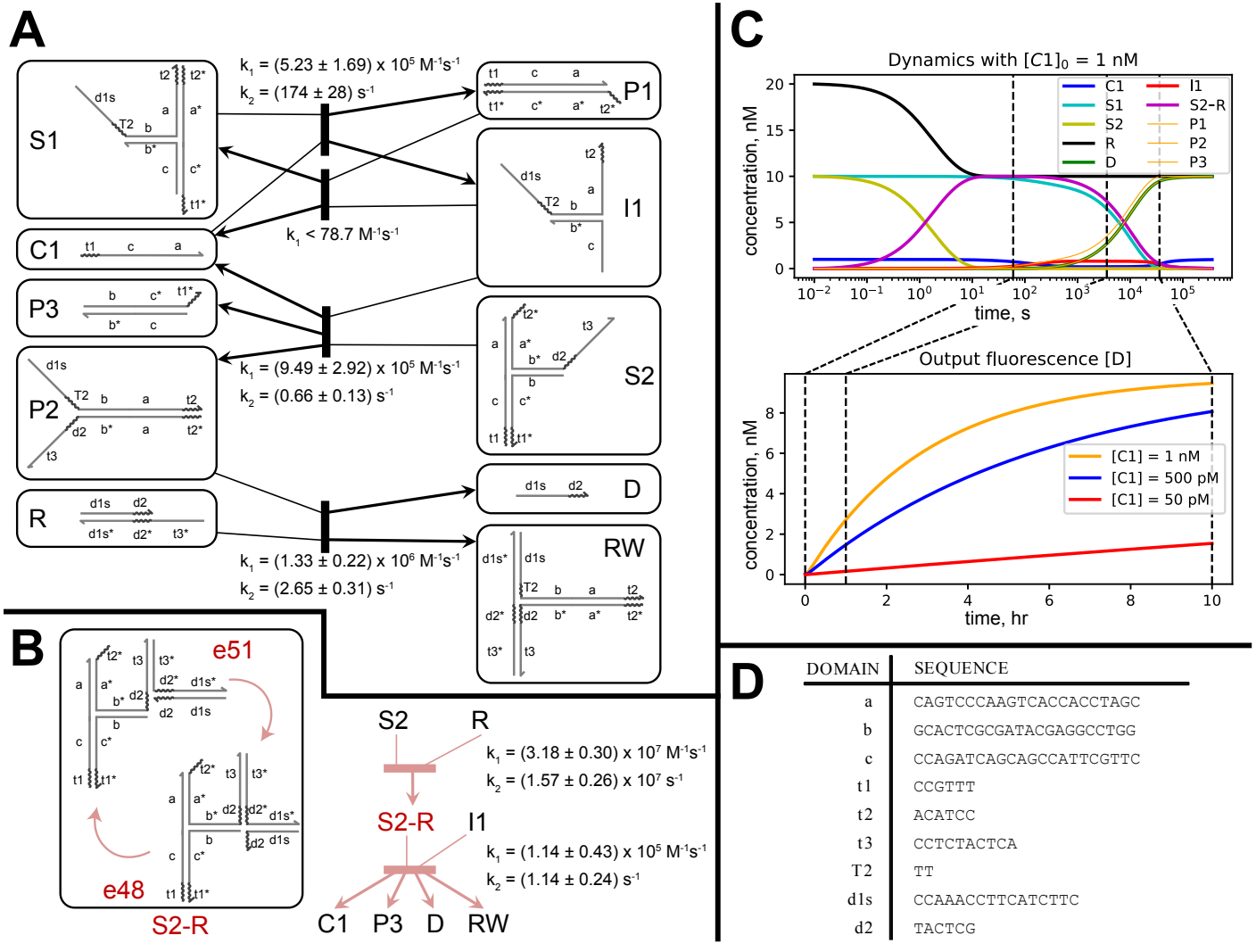
Badelt et al. (2017) - Nuskell
Grun et al. (2014) - Peppercorn
Shin et al. (2017) - CRN pathway decomposition equivalence
Johnson et al. (2018) - CRN bisimulation equivalence
Berleant et al. (submitted) - KinDA

KINETIC NUCLEOTIDE-LEVEL ANALYSIS



Berleant, Berlind, Badelt, Dannenberg, Schaeffer, Winfree (submitted) - Automated Sequence-Level Analysis of Kinetics and Thermodynamics for Domain-Level DNA Strand-Displacement Systems

KINETIC NUCLEOTIDE-LEVEL ANALYSIS

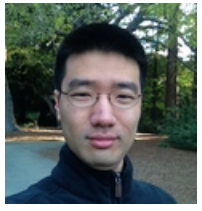


DOMAIN	SEQUENCE
a	CAGTCCCAAGTCACCACCTAGC
b	GCACTCGCGATACGAGGCCTGG
c	CCAGATCAGCAGCCATTCGTTC
t1	CCGTTT
t2	ACATCC
t3	CCTCTACTCA
T2	TT
d1s	CCAAACCTTCATCTTC
d2	TACTCG

Berleant, Berlind, Badelt, Dannenberg, Schaeffer, Winfree (submitted) - Automated Sequence-Level Analysis of Kinetics and Thermodynamics for Domain-Level DNA Strand-Displacement Systems

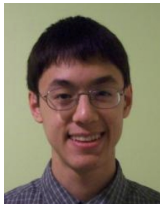
THANKS TO

Erik Winfree



Seung Woo Shin

Casey Grun



Joseph Berleant

Robert F. Jonhson

Chris Thachuk

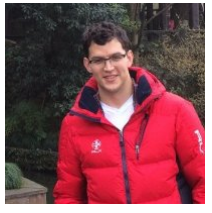
Frits Dannenberg

Chris Berlind

Karthik Sarma

Qing Dong

Joseph 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).