REACTION ENUMERATION & CONDENSATION
OF DOMAIN-LEVEL STRAND DISPLACEMENT SYSTEMS

Stefan Badelt
DNA and Natural Algorithms (DNA) Group, Caltech

Feb 14th, 2018
33rd TBI Winterseminar, Bled, Slovenia

Grun, Badelt, Sarma, Shin, Wolfe, and Winfree (manuscript in preparation)
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
DNA = Thymine
DNA = Cytosine
DNA = Guanine
DNA = Phosphate backbone

DNA = long domain
DNA = short domain

○ = 5' end
▼ = 3' end

b

a* b*

b

a* b*
DOMAIN-LEVEL STRAND DISPLACEMENT

- Long (branch-migration) domain: binds irreversibly
- Short (toehold) domain: binds reversibly

A

\[
\begin{array}{c}
\text{a} \quad \text{t} \\
\text{t}^* \quad \text{x}^* \quad \text{t}^*
\end{array}
\]

+ F1

\[
\begin{array}{c}
\text{a} \quad \text{t} \\
\text{x} \quad \text{t} \\
\text{b} \\
\text{t}^* \quad \text{x}^* \quad \text{t}^*
\end{array}
\]

\[
\begin{array}{c}
\text{a} \quad \text{t} \\
\text{b} \\
\text{t}^* \quad \text{x}^* \quad \text{t}^*
\end{array}
\]

\[
\begin{array}{c}
\text{a} \quad \text{t} \\
\text{b} \\
\text{t}^* \quad \text{x}^* \quad \text{t}^*
\end{array}
\]

B

\[
\begin{array}{c}
\text{x} \quad \text{t} \\
\text{t}^* \quad \text{x}^* \quad \text{t}^*
\end{array}
\]

\[
\begin{array}{c}
\text{a} \quad \text{t} \\
\text{b} \\
\text{t}^* \quad \text{x}^* \quad \text{t}^*
\end{array}
\]

\[
\begin{array}{c}
\text{a} \quad \text{t} \\
\text{b} \\
\text{t}^* \quad \text{x}^* \quad \text{t}^*
\end{array}
\]

bind

3-way branch migration

unbind
DOMAIN-LEVEL STRAND DISPLACEMENT

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

A

B

F1

F2

bind

3-way branch migration

unbind

i1

i2
DOMAIN-LEVEL STRAND DISPLACEMENT

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

A

F1

bind

3-way branch migration

i1

F2

unbind

i2

B

Detailed network

Condensed network
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 system specification**

\[ A + F_1 \rightleftharpoons F_2 + B \]

signal species (low concentration): \{A, B\}

fuel species (high concentration): \{F_1, F_2\}
FROM CRN TO DSD SYSTEMS

\[ A + B \rightarrow C + D \]


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

Images drawn using VisualDSD, Lakin et al. (2012)
FROM A DIGITAL CIRCUIT TO DSD

Qian et al. (2011)

\[ y_2y_1 = \sqrt{x_4x_3x_2x_1} \]

Input for the nuskell compiler: 32 formal reactions.


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

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

bind / open

3-way branch migration

4-way branch migration
REACTION TYPES

bind / open

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
\[ t \ a \ t + t^* \ b \ t^* + t \ a \ t + t^* \ b \ t^* \]
\[ \ldots (+ (\ldots +) \ldots (+) \ldots ) \]
multistranded pseudoknot

\[ t a t + t^* b t^* + t a t + t^* b t^* \]
\[ \ldots (+ (\ldots +) \ldots) \]
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:

\[k_\beta [A][B] << k_\alpha [X]\]
MODEL PARAMETERS

rate-independent model
open reactions where domain-length $> L$ are negligible
unimolecular reactions are fast
bimolecular reactions are slow

rate-dependent model
assume typical rate constant for every reaction:
\[ k = \text{rate(rtype, dlength)} \]
unimolecular reactions with $k < k_{\text{slow}}$ are negligible
unimolecular reactions with $k < k_{\text{fast}}$ are slow
unimolecular reactions with $k \geq k_{\text{fast}}$ are fast
bimolecular reactions are slow
REACTION ENUMERATION

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

*valid* according to enumeration semantics:

- all valid, except open > $L$
- max-helix semantics: reaction types are greedy
- probability threshold for reactants of bimolecular reactions.
- probability threshold for products of unimolecular reactions.
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 \rightarrow i1
i1 \rightarrow i2
i2 \rightarrow B + F2
B + F2 \rightarrow i2
i2 \rightarrow i1
i1 \rightarrow A + F1
A + F2 \rightarrow i4
i4 \rightarrow A + F2
B + F1 \rightarrow i3
i3 \rightarrow B + F1

condensed reactions:
A + F1 \rightarrow B + F2
B + F2 \rightarrow 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: \( r = (A, B) \quad A = \{|a_i|\} \)

condensed reaction: \( \hat{r} = (\hat{A}, \hat{B}) \quad \hat{A} = \{\hat{A}_i\} \)

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:

\[
k_{\hat{r}} = \sum_{r = ((a_1, a_2), B) \in R_{\hat{A}}} P(a_1|\hat{A}_1) \cdot P(a_2|\hat{A}_2) \cdot k_r \cdot P(T_{B \rightarrow \hat{B}})
\]
REACTION RATE CONDENSATION

general form:

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

where

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

\[ \mathbb{P}[T_{B \rightarrow \hat{B}}] = \text{reaction decay probability} \]
A DNA OSCILLATOR

Molecular program
i. B + A → 2B
ii. C + B → 2C
iii. A + C → 2A

(rock-paper-scissors oscillator)

Srinivas, Parkin, Seelig, Winfree, Soloveichik:
Enzyme-free nucleic acid dynamical systems. Science (2017)
A DNA OSCILLATOR

Molecular program
i. B + A → 2B
ii. C + B → 2C
iii. A + C → 2A

(rock-paper-scissors oscillator)

Srinivas, Parkin, Seelig, Winfree, Soloveichik:
Enzyme-free nucleic acid dynamical systems. Science (2017)
DETAILED VS. CONDENSED SIMULATION

\[ A + B \rightarrow B + B \]
\[ B + C \rightarrow C + C \]
\[ C + A \rightarrow A + A \]

translation scheme: srinivas2017.ts
REACTION ENUMERATOR

**model limitations**
- no multistranded pseudoknots
- assumption of low concentrations
  - assumption of "typical" reaction rate constants

**model parameters**
- multiple layers of reaction-semantics
  - reaction types
  - max-helix notion (representation-independent)
  - reaction rate dependent enumeration
What the domain level can do:
- enumerate intended reaction pathways
- detect unintended reaction pathways
- very fast assessment of overall dynamics
- define a CRN for sequence-level simulations

What the domain level cannot do:
- include sequence-level variations within the domains

What the domain level could do:
- detect and quantify particular leak reactions
- provide a coarse-graining for stochastic simulations
THANKS TO

Erik Winfree  |  Casey Grun  |  Karthik Sarma

Seung Woo Shin  |  you  |  Brian Wolfe


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).