Thermodynamic binding networks

slides © 2021, David Haley and David Doty
ECS 232: Theory of Molecular Computation, UC Davis
In an electronic circuit, voltage can represent Boolean input.

In a well-mixed solution, concentration can represent Boolean input.
Chemical Identity Gate: Idealized vs. Actual Behavior

Experimental Implementation of Chemical Logic

Wang et al. *DNA* 23, 2017

Qian et al. *Science* 332, 2011

(Qan’t worry about the details in the pictures above)


Input Present

No Input (Ideal)

Output

Input

Time

1

0

$X_1X_2X_3=000$

$y = 0$

$0 \leq t \leq 10$

$y(t) = \frac{K}{1 + e^{-\frac{t-T}{\tau}}}$

$\tau = 20$
Levels of Abstraction

DNA

Base Pairs

Strands
DNA strand displacement

Bind

Displace

Release
\[ X \rightarrow Y + Z \]
Leak in strand displacement experiments

\[ y_2 y_1 = \sqrt{x_4 x_3 x_2 x_1} \]

Reducing Leak

Intended:

[Boya Wang, Chris Thachuk, Andrew Ellington, David Soloveichik. The Design Space of Strand Displacement Cascades with Toehold-Size Clamps DNA Computing Conference, 2017]
What causes leak “kinetically”?

$X \rightarrow Y + Z$
What causes leak “thermodynamically”?

Before:

less favorable

After:

more favorable

What causes leak “thermodynamically”?
Need a kinetic binding network model

- Can we design pathways that maintain local stability?
Levels of Abstraction

DNA
↓
Base Pairs
↓
Strands
↓
Thermodynamic Binding Network
Thermodynamic Binding Networks

Monomer = collection of domains
Configuration = how monomers are bound

Geometry-Free Model:
The domains within a monomer are unordered
Energetic favorability: Bonds and complexes

all else equal, more bonds = more favorable

all else equal, more complexes = more favorable

\[ a^*b^* \quad a \quad b \quad a^*b^* \quad \frac{<}{=} \quad a^*b^* \quad a \quad b \quad a^*b^* \]

\[ a \quad b \quad a^*b^* \quad \frac{<}{=} \quad a \quad b \quad a^*b^* \quad a \quad b \]
Tradeoff between #bonds and #complexes

• in general, there’s some weight parameter $w$:
  
  \[
  \text{energy} = w \times \text{#bonds} + \text{#complexes}
  \]

  \text{(physics notation: } \Delta G = \Delta H - T \cdot \Delta S \text{)}

• We often consider a natural limiting case:
  • favoring #bonds \underline{infinitely} over #complexes
  • require maximal #bonds formed; use #complexes only as tiebreaker
  • Corresponds to bonds that are so strong they cannot spontaneously dissociate, but can exchange with each other to find configurations with more complexes
Thermodynamic Binding Networks

saturated = maximum #bonds formed
stable = saturated, AND maximum #complexes

If we’re careful to make starred binding sites limiting, then saturated = all starred sites are bound
Computing via Thermodynamic Equilibrium
AND gate

output monomer to be released only if both inputs present
Issues with Boolean logic

• How to compose?
  – We don’t know how to prove the previous gate is composable, and used a more complex design in the paper

• Want “entropy gap”:
  – Need not merely that unwanted configurations are unstable (i.e., if saturated, they have lower entropy), but more strongly that they have much lower entropy.
  – We can use $O(n)$ domain/monomer types to achieve an entropy gap of $n$.

• Output convention?
  – Obvious one: “there’s a unique stable configuration with the correct output”
  – It’s problematic, so we have a one-sided convention:
    • if correct output is 0, unique stable configuration with correct answer
    • if correct output is 1, then both the “output=1” and “output=0” configurations are stable
Composable AND gate with entropy gap 3

Rather than release a single output monomer, it suffices to gather all output domains on one complex.
Kinetic pathways and energy barriers
**Pathways**

**Thermodynamics:** Which configurations are energetically favorable

**Kinetics:** How a system moves between configurations over time
\( X \rightarrow Y + Z \)
$X \rightarrow Y + Z$
What causes leak “kinetically”? 

The reaction is represented as $\emptyset \rightarrow Y + Z$.

A diagram shows the energy and progress with points indicating the energetics of the reaction. The diagram illustrates the energetics of the reaction process:

- **Energy**
  - A red line indicating the energy changes with progress.

- **Progress**
  - The progress axis with marked energetics.

- **Energetics**
  - Merge (less favorable)
  - Split (more favorable)

The energetics of the reaction are shown with markers indicating the transition points.

The reaction process is summarized as:

- **Merge** (less favorable)
- **Split** (more favorable)
Kinetic Binding Networks

• Favorability is a combination of bond count and complex count

  \[
  \text{Weighted average:} \\
  \text{Energy} := -w_H(\# \text{ bonds}) - (\# \text{ complexes})
  \]

• Define pathways to consist of merges and splits

• But for \( w_H \geq 2 \), only saturated pathways need be considered

Since all saturated configurations have an equal number of bonds, we can focus solely on the number of complexes

Large Energy Barriers

Reaction Pathway

α

β

Configuration space

Energy

Merge (less favorable)

Split (more favorable)
A Network with a Programmable Energy Barrier

\[
\begin{align*}
X_{11} & \quad X_{12} & \quad X_{13} & \quad X_{14} \\
X_{21} & \quad X_{22} & \quad X_{23} & \quad X_{24} \\
X_{31} & \quad X_{32} & \quad X_{33} & \quad X_{34} \\
X_{41} & \quad X_{42} & \quad X_{43} & \quad X_{44}
\end{align*}
\]

Energy
\[
\begin{align*}
\text{Merge} & \quad \text{(less favorable)} \\
\text{Split} & \quad \text{(more favorable)}
\end{align*}
\]

Progress

\[
\begin{align*}
X_{11}^{*} & \quad X_{12}^{*} & \quad X_{13}^{*} & \quad X_{14}^{*} \\
X_{21}^{*} & \quad X_{22}^{*} & \quad X_{23}^{*} & \quad X_{24}^{*} \\
X_{31}^{*} & \quad X_{32}^{*} & \quad X_{33}^{*} & \quad X_{34}^{*} \\
X_{41}^{*} & \quad X_{42}^{*} & \quad X_{43}^{*} & \quad X_{44}^{*}
\end{align*}
\]

\[
\begin{align*}
X_{21} & \quad X_{22} & \quad X_{23} & \quad X_{24} \\
X_{31} & \quad X_{32} & \quad X_{33} & \quad X_{34} \\
X_{41} & \quad X_{42} & \quad X_{43} & \quad X_{44}
\end{align*}
\]
Catalysis

Energy

Merge
(less favorable)

Split
(more favorable)

progress

\[ x_{11} \quad x_{21} \quad x_{31} \quad x_{41} \]
\[ x_{12} \quad x_{22} \quad x_{32} \quad x_{42} \]
\[ x_{13} \quad x_{23} \quad x_{33} \quad x_{43} \]
\[ x_{14} \quad x_{24} \quad x_{34} \quad x_{44} \]
Autocatalysis
Multiple Stable Configurations

For a grid of prime size $n \times n$, there can be at most $n+1$ different stable configurations with barrier $n$ to pass between any of them.
Directed Catalysis

Along a catalyzed pathway, the barrier is 1
Otherwise the barrier is n/2
Social Golfer Problem

• Can $25 (n^2)$ golfers play in 5-somes ($n$-somes) for 6 ($n+1$) days, so that no two golfers play together more than once?

• First studied by Euler.

• True if $n$ is a prime power (2,3,4,5,7,8,9,11,13,...)

• False for smallest non-prime power $n=6$: can only play for 3 days!


• Unknown for next prime power $n=10$:
  • trivial upper bound is 11 days
  • best known lower bound is 3
(Feasible?) DNA implementation
Thermodynamic self-assembly

Grafting the TBN model onto self-assembly
A modest goal

• **Informal**: Design monomers that self-assemble arbitrarily large complexes.
  
  • *size of a complex* = # monomers in the complex

• **Formal**: Design a set of monomer types so that, for all \( S \in \mathbb{N} \), there is a stable complex of size at least \( S \).

• Easy to do in Abstract Tile Assembly Model:

  
  set of monomer types: \[ \{a, a^*\} \]

  size-8 complex (assembly) formed with 8 copies of monomer

  ![Diagram of monomer types and assembly](attachment:diagram.png)
Difficulty of self-assembling large complexes

- Not stable! (or even saturated)
  - attempt 2:
    - More complexes ⇒ higher entropy ⇒ more stable
      - These have more complexes, and each is self-saturating (every domain can be bound within the complex)

These diagrams illustrate the difficulty of assembling large complexes, where each domain can be bound within the complex to achieve stability.
An even more modest goal

Original goal: Design a set of monomer types so that, for all $S \in \mathbb{N}$, there is a stable complex of size at least $S$.

Revised goal: For all $S \in \mathbb{N}$, design a set of $M$ monomer types using $D$ domain types with a stable complex of size at least $S$.

How large can we make $S$ relative to $D$ and $M$?

$D, M = O(1), \ S = \text{arbitrarily large}$
How large can we make $S$ relative to $D$ and $M$?

$S \approx 2^D$?
Stable complexes have at most exponential size

**Theorem:** Any thermodynamic binding network with

- $D$ domain types,
- $M$ monomer types,
- $\leq A$ domains per monomer type  

Has stable complexes of size $\leq 2(M+D)(AD)^{2D+3} = \text{poly}(D^D)$ if $A = O(1)$
Easy proof if binding graph is acyclic (tree)

- Since monomers have $O(1)$ domains, binding graph is bounded degree
- # nodes of tree is at most exponential in depth (longest path length \leq 2 \cdot \text{depth})
- If some path has > $2D$ edges, it must repeat some ordered pair \((d_i, d_i^*)\) or \((d_i^*, d_i)\)
- Break into two saturated complexes as shown.
Monomers as vectors

• monomer \{a, \ b*, b*, \ d, d, d, d, d*, \ e, e*\} represented as (1, -2, 0, 3, 0)
• sum of many monomers gives the number of excess domains in a fully bound (saturated) complex with those monomers
  • i.e., 2 copies of above monomer \(2 \cdot (1, -2, 0, 3, 0) = (2, -4, 0, 6, 0)\) have an excess of 2 a’s, 4 b*’s, 0 c’s, 6 d’s, 0 e’s
Somewhat easy proof that unbounded size complexes cannot be assembled

**Original goal:** Design a set of monomer types so that, for all $S \in \mathbb{N}$, there is a stable complex $P$ of size $\geq S$.

**Theorem:** Original goal is impossible.

**Proof:**

1. Suppose otherwise, let $P_1, P_2, \ldots$ in $\mathbb{N}^m$ be an infinite sequence of stable complexes increasing in size. $m$ is number of monomer types, $P_i(j)$ = # monomers of type $j$ in complex $P_i$.
2. Represent each monomer type as a vector in $\mathbb{Z}^d$ as on previous slide.
   1. $P_i$ is composed of monomers $m_{1i}, m_{2i}, \ldots, m_{ki}$.
   2. Let $S_i = m_{1i} + m_{2i} + \ldots + m_{ki}$. Note that there is a $m \times d$ matrix $M$ such that $S_i = M \cdot P_i$.
3. Take several infinite subsequences:
   1. Since there are a finite number of domain types, some infinite subsequence of $P_i$’s agrees on which set of domain types are unbound.
   2. By Dickson’s Lemma we may assume $P_1 < P_2 < \ldots$ and $S_1 < S_2 < \ldots$ i.e., each has all the monomers of the previous, plus some more, and each has all the unbound domains of the previous, plus some more.
4. Let $d = P_2 - P_1$. Then $M \cdot d = M \cdot P_2 - M \cdot P_1 = S_2 - S_1 \geq 0$.
5. i.e., $S_2 = S_1 + M \cdot d$ and all three are nonnegative,
6. i.e., we can split $S_2$ into 2 disjoint nonempty nonnegative subsets, $S_1$ and $M \cdot d$. QED
A digression into computational complexity

- **INTEGER-PROGRAMMING** problem  
  Given: integer matrix $A$, integer vector $b$  
  Question: is there a nonnegative integer vector $x$ such that $Ax = b$?

- **0/1-INTEGER-PROGRAMMING** is **NP-complete** (Karp 1972).

- **Non-obvious fact**: INTEGER-PROGRAMMING is in **NP**. (independently due to [Borosh and Treybig 1976], [Gathen and Sieveking 1978], [Kannan and Monma 1978])  
  If $Ax = b$ has a solution, it has a “small” solution... $\max_i x_i \leq \exp(\max_{ij}(A_{ij}, b_j))$

- **Papadimitriou’s proof**: [On the complexity of integer programming. Papadimitriou, JACM 1981]  
  - If $x$ is a *large enough* solution, there is $0 < y < x$, $y \in \mathbb{N}^m$, such that $Ay = 0$.  
  - Defining $z = x - y$, $Az = A(x - y) = Ax - Ay = Ax - 0 = b$.  
  - So $z$ is a strictly smaller solution than $x$: $x$ cannot be the *smallest* solution.
Farkas’ Lemma

Given vectors $m_1, m_2, \ldots$, they obey one of two constraints:

a) are directions of balanced forces

$$\exists c \quad \text{(counts of monomers)} \quad (m_1 \, m_2 \, m_3)c = 0$$

b) lie on one side of some hyperplane

$$\exists h \quad \text{(hyperplane orthogonal vector)} \quad h \cdot m_i \geq 1$$
How to prove exponential complex size bound for complexes with cycles in binding graph?

• $A = d \times m$ matrix: $A_{ij} = \text{monomer } m_j$’s excess of domain $d_i$ over $d_i^*$

• If $Ac = b$, then $b_i = \text{total \# unbound } d_i$ in any saturated configuration of $c$

• If $|c| > \text{exponential in } D$, Papadimtriou’s proof gives us subcollection $y < c$ such that $Ay = 0$, (Farkas’ Lemma says that if this fails, then monomer vectors all lie on one side of a hyperplane, see next slide)

  • i.e., $\#d_i \text{ in } y = \#d_i^* \text{ in } y$, so $y$ is self-saturating.

• So whatever bonds were broken to separate $y$ can be re-bound within $y$.

• By symmetry, the same bonds in $z = c - y$ can be re-bound within $z$. 

monomer collection $c \in \mathbb{N}^M$

$Ac = (2,1,0)$

![Diagram showing the relationship between monomer vectors and unbound complexes]
If all monomer types lie on one side of hyperplane $h$...

- Consider “slack monomers” $\{d_1^*\}, \{d_2^*\}, \ldots$, adding just enough to bind to all the excess $d_i$ domains, so **saturated** (fully bound) == **all domains bound**
- If $c$ is count of all monomers including slack monomers ($c(i) =$ count of $m_i$), then $Ac = 0$, where each column of $A$ represents a monomer (counts of domains).
- dot-product $h$ on both sides: $h \cdot Ac = h \cdot 0 = 0$, distribute through: $\sum_i (h \cdot m_i)c(i) = 0$
- Let $S$ be set of monomers with “small” counts, move them to one side:
  $$\sum_{i \in S} (h \cdot m_i)c(i) = \sum_{i \notin S} (h \cdot m_i)c(i)$$
- Then “small” $c(i)$ (count of $i$’th monomer) is small by definition, and $h \cdot m_i = O(1)$ above since $h \cdot m_i \geq 1$
Applying thermodynamic model to tile assembly

• Let’s incorporate the thermodynamic binding network model into the abstract tile assembly model.

• How can we create a large assembly from a small number of tile types?
A thermodynamically unstable tile assembly counter
A thermodynamically **stable** tile assembly counter

Difference is that each row (corresponding to bits of the same significance) has glues labeled with the row number.
Conclusions

• Strong bonds (surprisingly) aren’t sufficient to self-assemble large thermodynamically stable structures. *Geometry helps!*

• Kinetically self-assembling a thermodynamically stable structure has very strong guarantees on errors:
  • target structure eventually results despite arbitrary kinetic errors.
  • If it’s the only stable structure, and free energy of other structures is much less, then it’s the only result you’ll see.

• Bad news: **NP-complete** to tell if a given configuration is unstable... even **NP-hard** to approximate entropy of stable configuration: