Minimum Free Energy, Partition Function & Kinetics Simulation Algorithms for a Multistranded Scaffolded DNA Computer

Ahmed Shalaby, Chris Thachuk, Damien Woods

Hamilton Institute, Computer Science

Maynooth University, Ireland

Hamilton Institute

Maynooth University
National University of Ireland Maynooth

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Science Foundation Ireland sfi For what's next

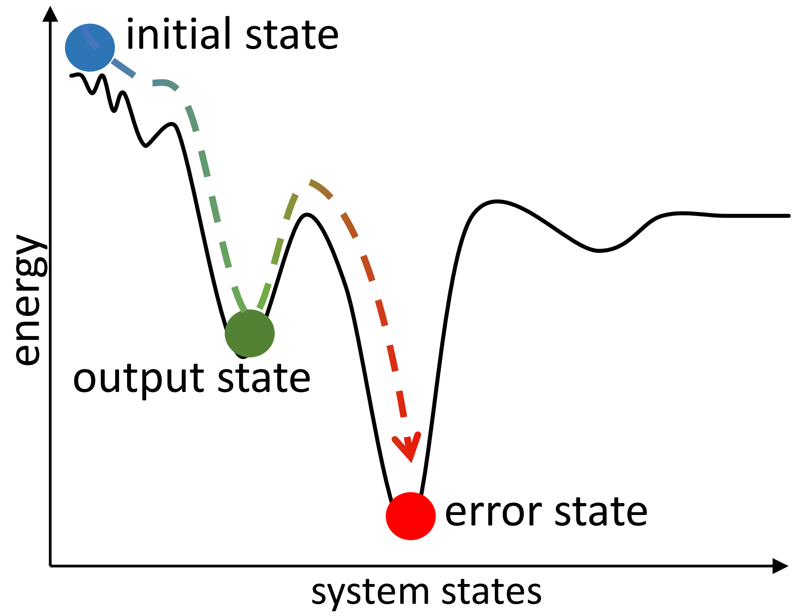
European Innovation Council

Funded by the European Union

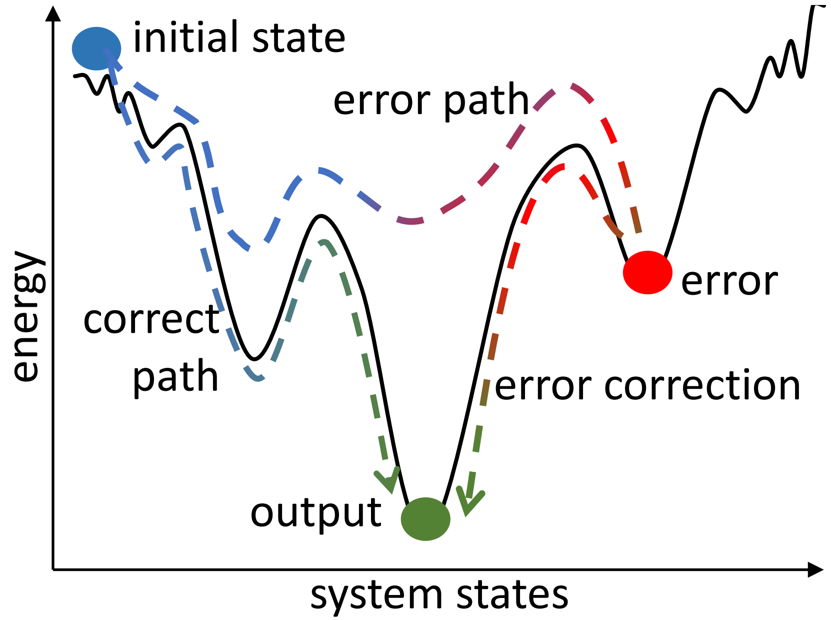
erc

Thermodynamically favourable computation

Typical molecular computers are thermodynamically unfavourable (leak, errors, spurious nucleation, etc.)



Today: Thermodynamically favoured computation on our **Scaffolded DNA Computer**

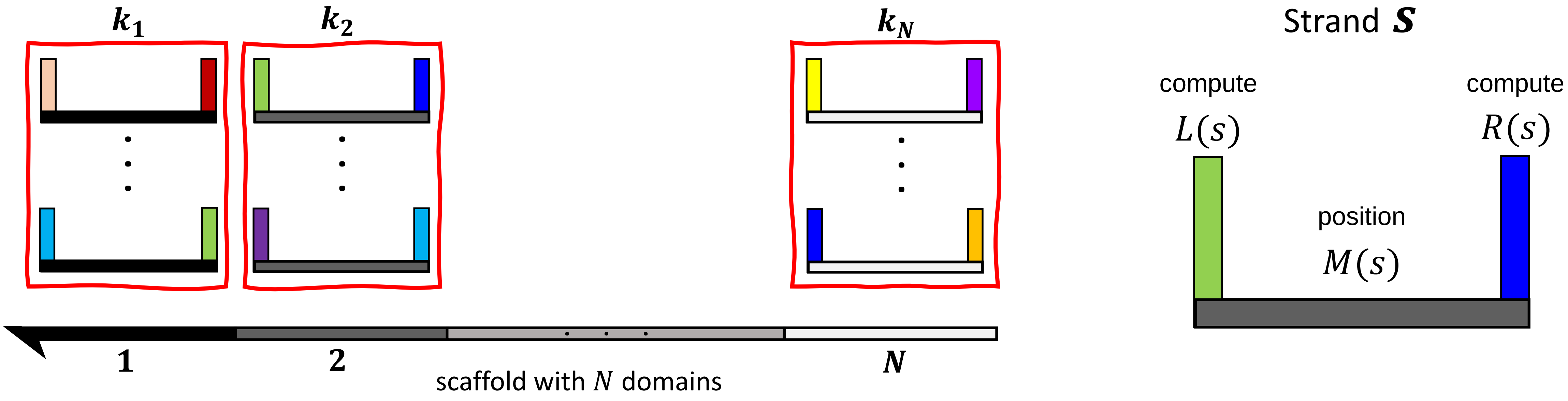


What the system is designed to reach! 🤝 What the system thermodynamically tends towards

Stérin, Eshra, Woods (DNA28 Track B, In prep)
Abeer Eshra Poster 21, DNA29
Doty, Soloveichik et al. 2017
Wang, Thachuk et al. 2017

Scaffolded DNA Computer: strand-based model

A. Scaffolded DNA Computer



scaffold with N domains

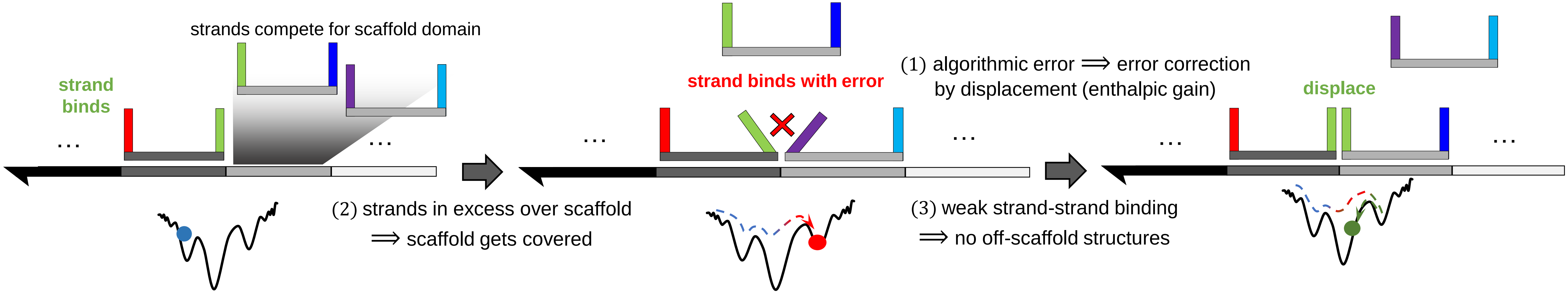
Strand S

compute $L(s)$

compute $R(s)$

position $M(s)$

B. Key design concepts for thermodynamically favoured computation

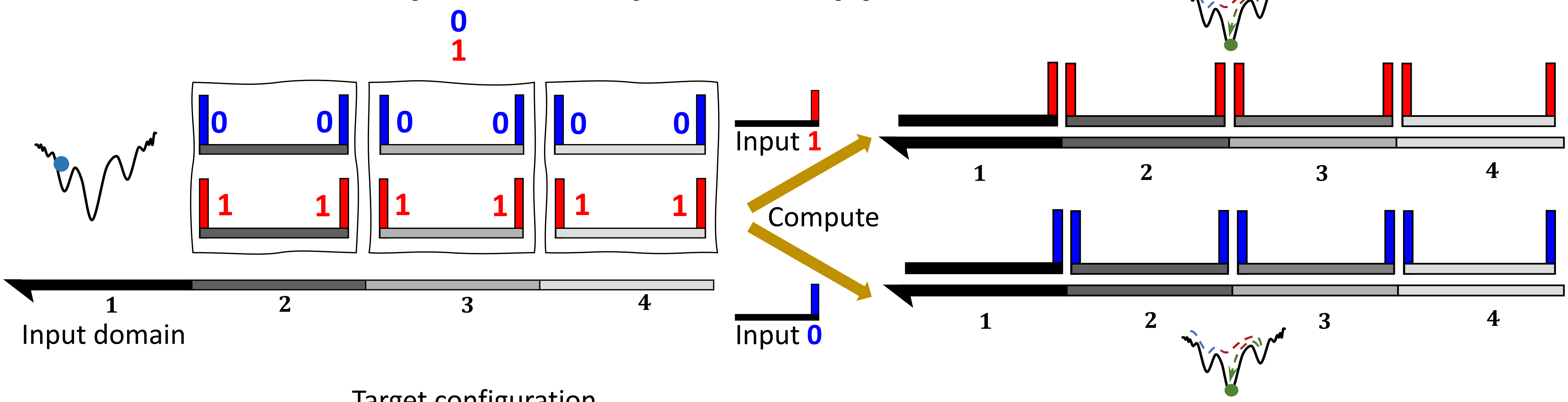


(2) strands in excess over scaffold $\Longrightarrow$ scaffold gets covered

(1) algorithmic error $\Longrightarrow$ error correction by displacement (enthalpic gain)

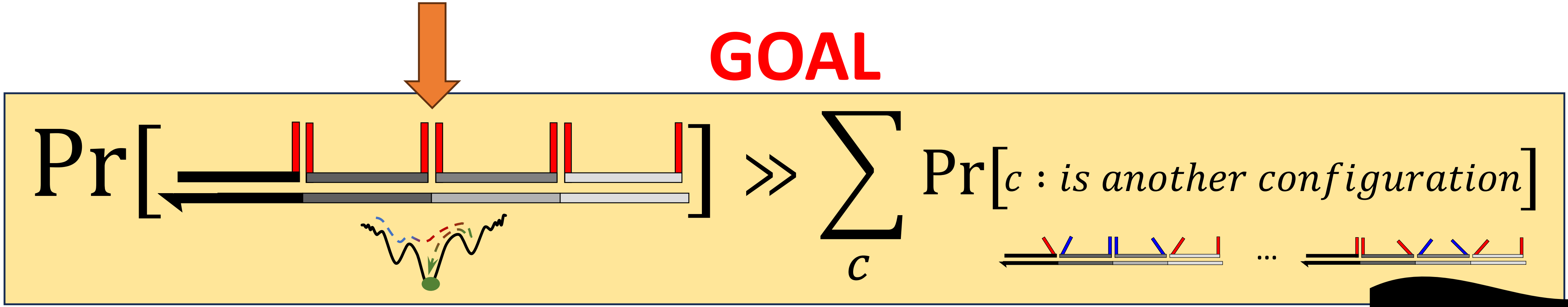
(3) weak strand-strand binding $\Longrightarrow$ no off-scaffold structures

Scaffolded DNA Computer example: Bit-Copy



Target configuration

GOAL

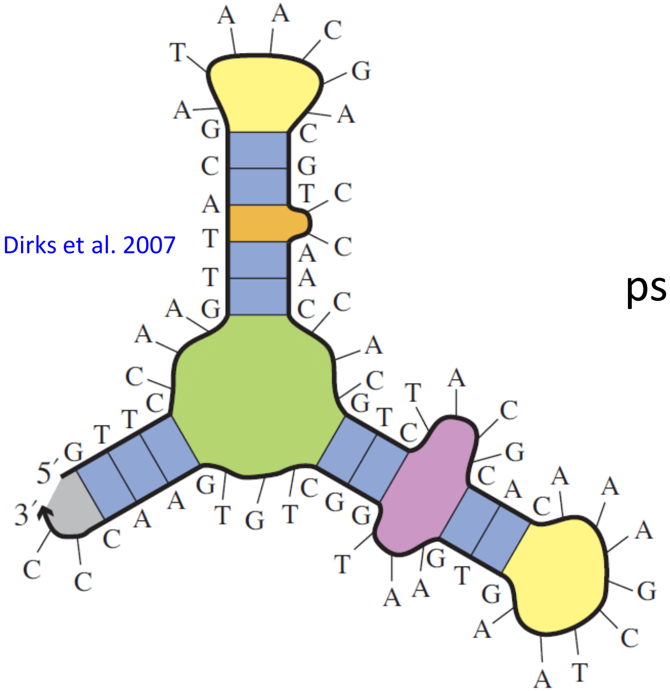


$$\Pr[\quad] \gg \sum_{c} \Pr[c : is\ another\ configuration]$$

At equilibrium

Secondary structure

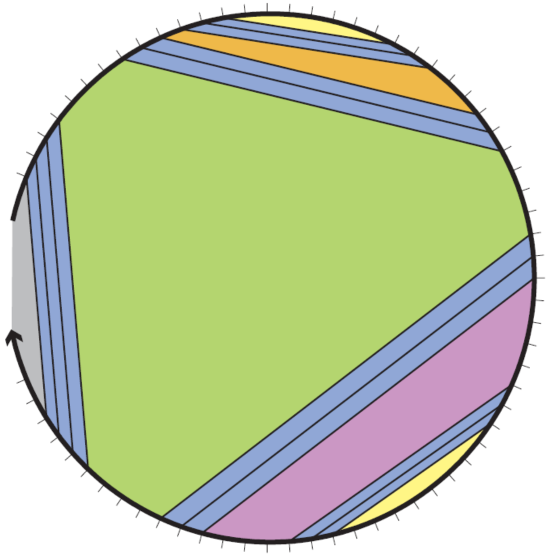
Single stranded DNA



Dirks et al. 2007

Secondary structure

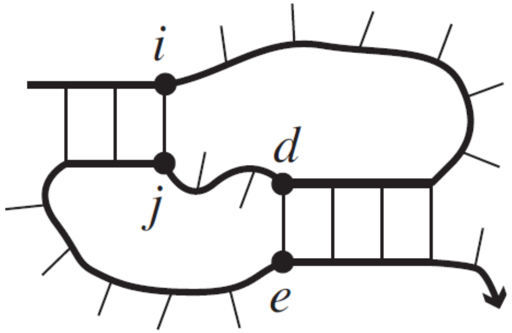
pseudoknot-free



Dirks et al. 2007

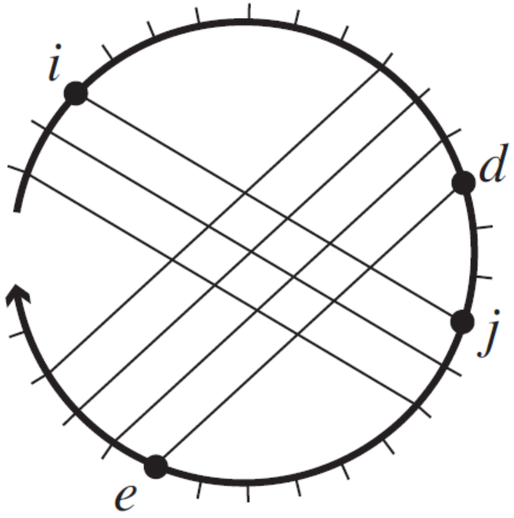
Polymer graph representation

NP – Hard



Dirks et al. 2007

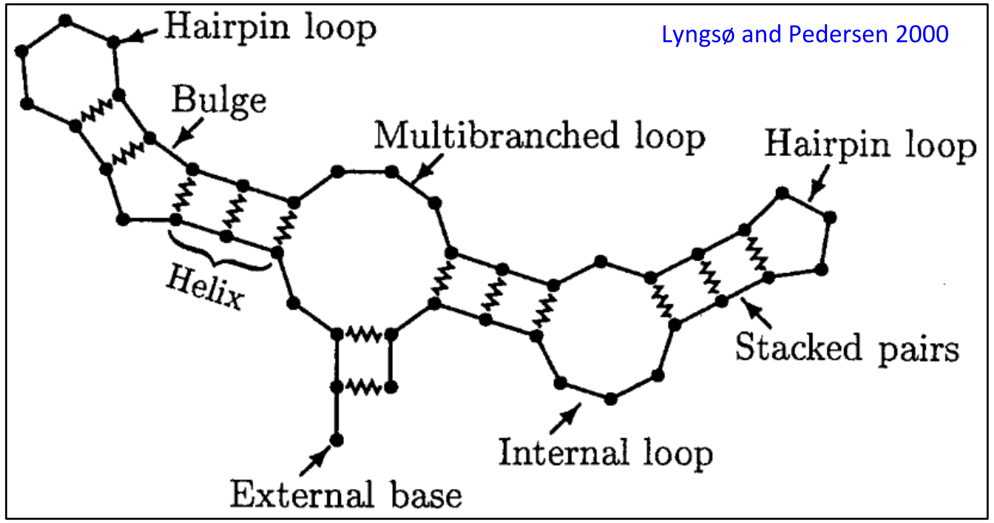
pseudoknotted



Dirks et al. 2007

Energy models, Minimum Free Energy and Partition Function

Single stranded system

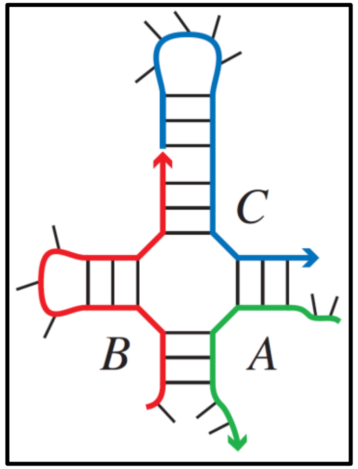


Lyngsø and Pedersen 2000

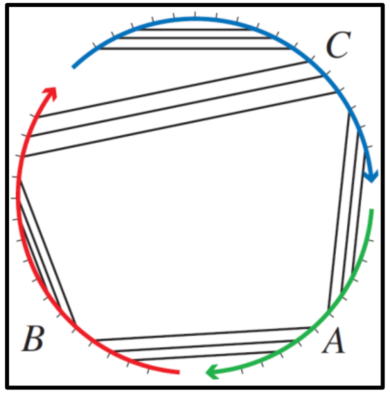
$$\overbrace{\Delta G(S)}^{\text{Free energy}} = \sum_{l\in S} \overbrace{\Delta G(l)}^{\text{Loop energy}}$$

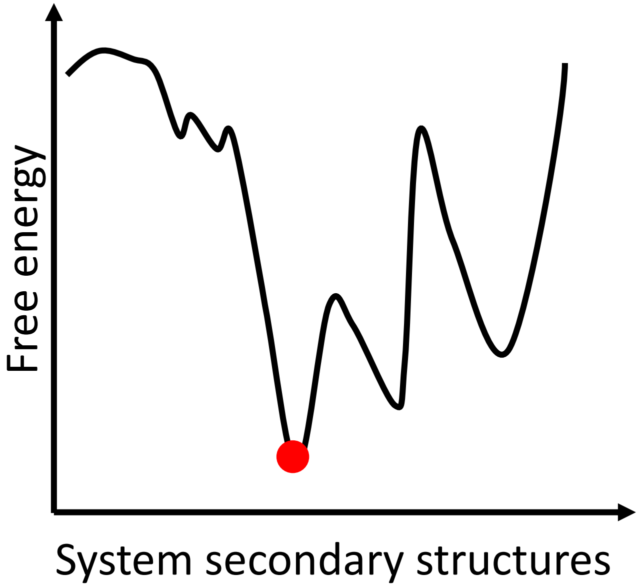
$$\overbrace{\Delta G(S)}^{\text{Free energy}} = \sum_{l\in S} \overbrace{\Delta G(l)}^{\text{Loop energy}} + \overbrace{(l-1) * \Delta G^{\text{assoc}}}^{\text{Entropic cost}}$$

Multi stranded system of l strands



Dirks et al. 2007





$$\boxed{\mathrm{MFE} = \min_{S\in\Omega} \Delta G(S)}$$

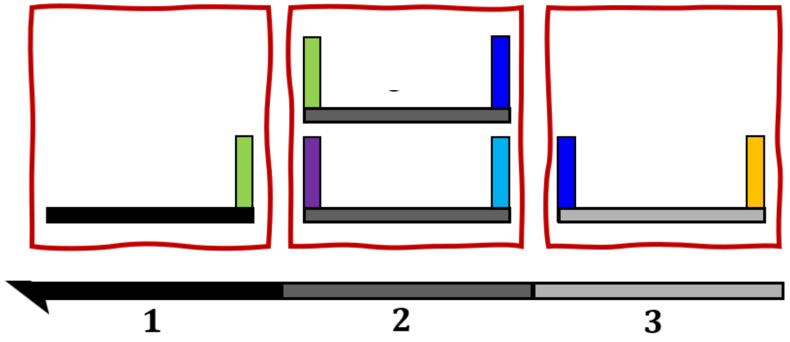
Minimum Free Energy

Boltzmann weighted sum

$$\boxed{Q = \sum_{S\in\Omega} e^{-\Delta G(S)/k_{\mathrm{B}}T}} \Rightarrow \boxed{\Pr[S] = \frac{e^{-\Delta G(S)/k_{\mathrm{B}}T}}{Q}}$$

Partition Function

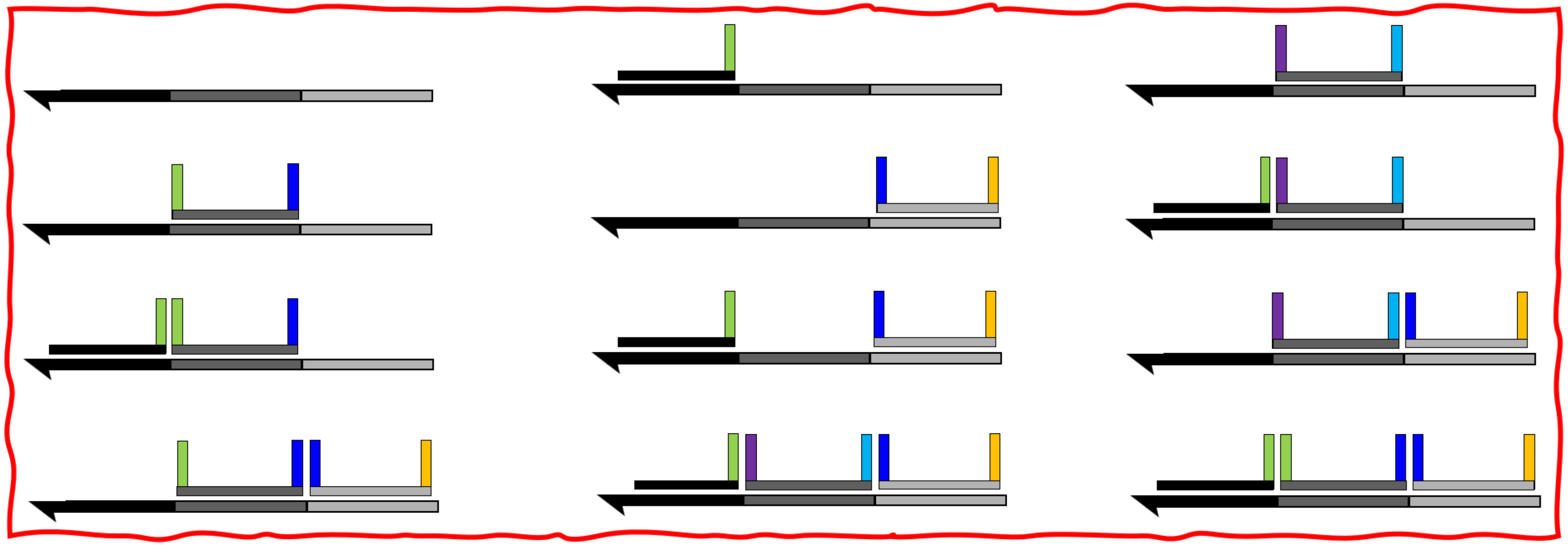
DOMAIN BASED ensemble of secondary structures for Scaffolded DNA Computer



Secondary structure

⇳

Scaffolded DNA Computer configuration



Scaffolded DNA Computer configuration space

Ω

- How many different configurations we will have?

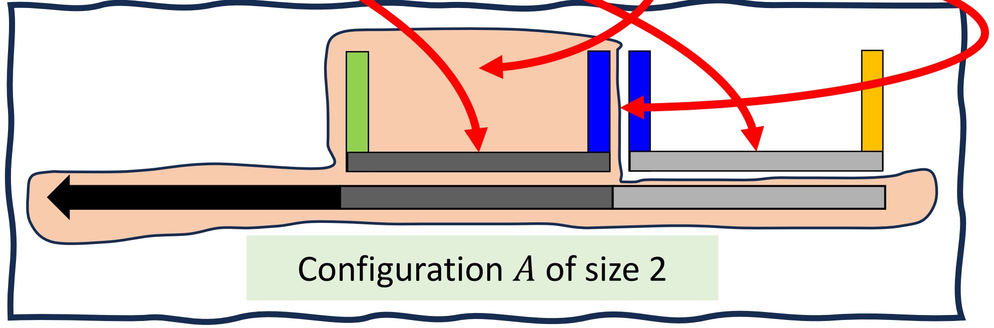
(Exponential in the # of scaffold domains)

$$|\Omega| = (\boldsymbol{k} + \boldsymbol{1})^{\boldsymbol{N}}$$

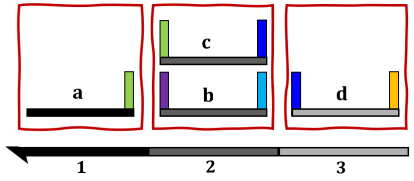
DOMAIN BASED energy model for the Scaffolded DNA Computer

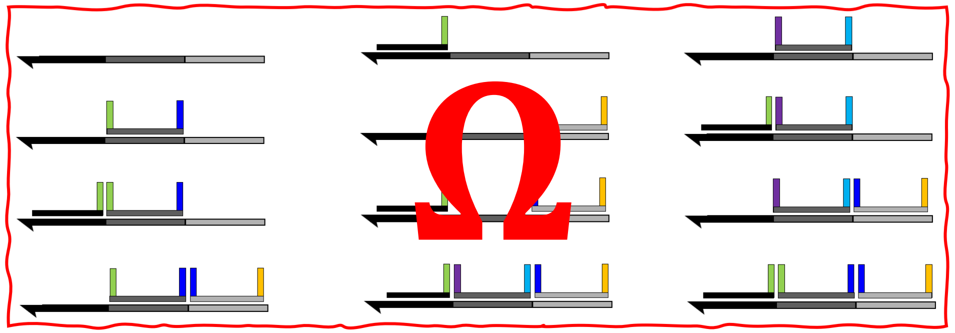
$$\Delta G(X) = \overbrace{\sum_{s\in X}\Delta G(M(s))}^{\text{Middle domain binding term}} + \overbrace{\sum_{s_i,s_{i+1}\in X}\Delta G\big(R(s_i),L(s_{i+1})\big)}^{\text{Toehold domain matching term}} + \overbrace{l * \Delta G^{\text{assoc}}}^{\text{Entropic cost term}}$$

X is a configuration of size l

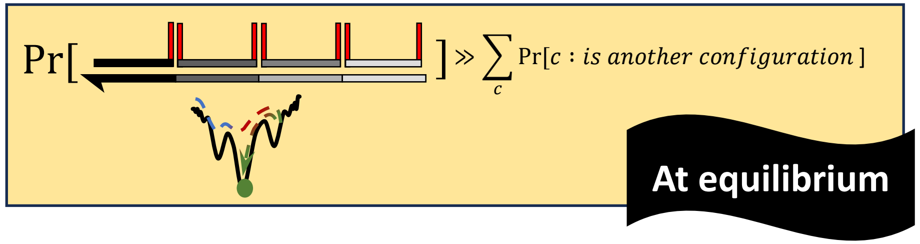


Configuration A of size 2





GOAL



Pr[ ] ≫ $\sum_{c}$ Pr[c : *is another configuration*]

At equilibrium

$$\text{MFE} = \min_{X\in\Omega}\{\Delta G(X)\}$$

$$Q = \sum_{X\in\Omega} e^{-\Delta G(X)/k_{\text{B}}T}$$

$$\Pr[\text{target}] = \frac{e^{-\overset{\text{MFE}}{\boxed{\Delta G(\text{target})}}/k_{\text{B}}T}}{Q}$$

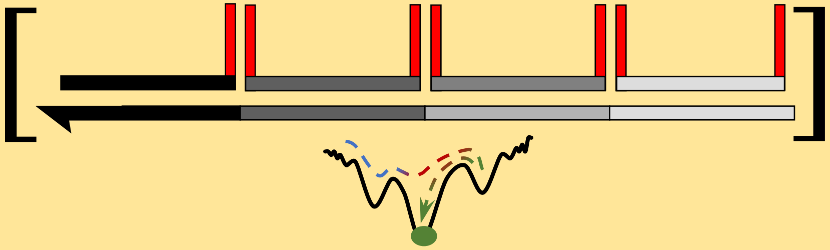
Computational complexity of Minimum Free Energy and the Partition Function

Input Type	MFE	Partition Function
Single Strand	$O(n^3)$ Nussinov et al.	$O(n^3)$ McCaskill
Multiple Strands, Bounded ($\leq s$)	?	$O(n^3)(s-1)!$ NUPACK
Multiple Strands, Unbounded	$NP - \text{Complete}$ Condon et al.	?

n bases, s strands

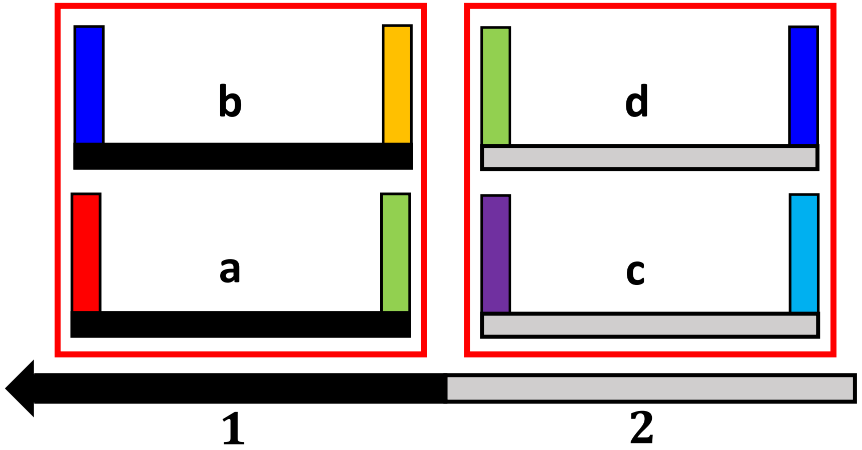
At equilibrium

GOAL

$\mathrm{Pr}[$  $] \gg \sum_{c} \mathrm{Pr}[c : \textit{is another configuration}]$

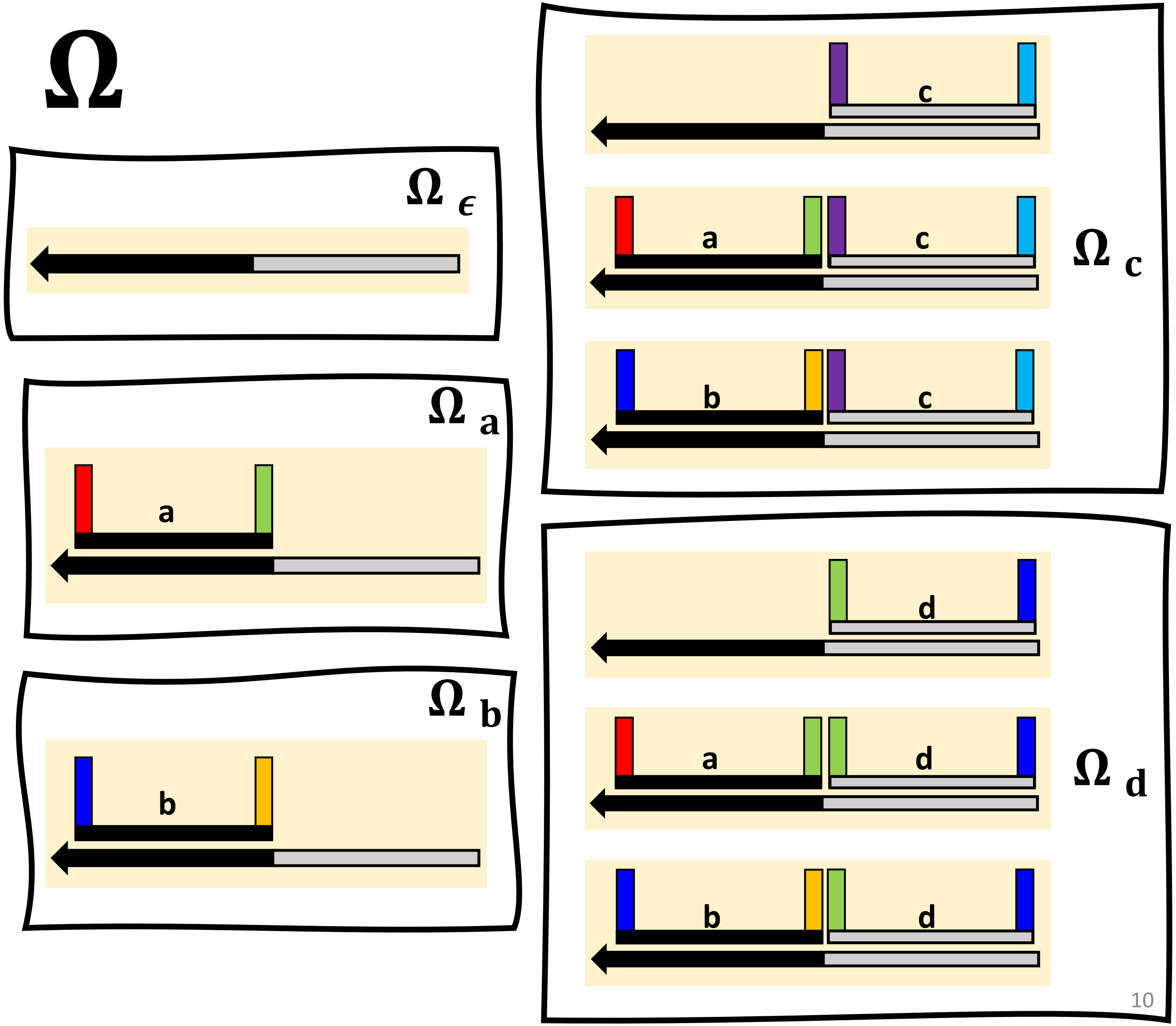
Efficiently

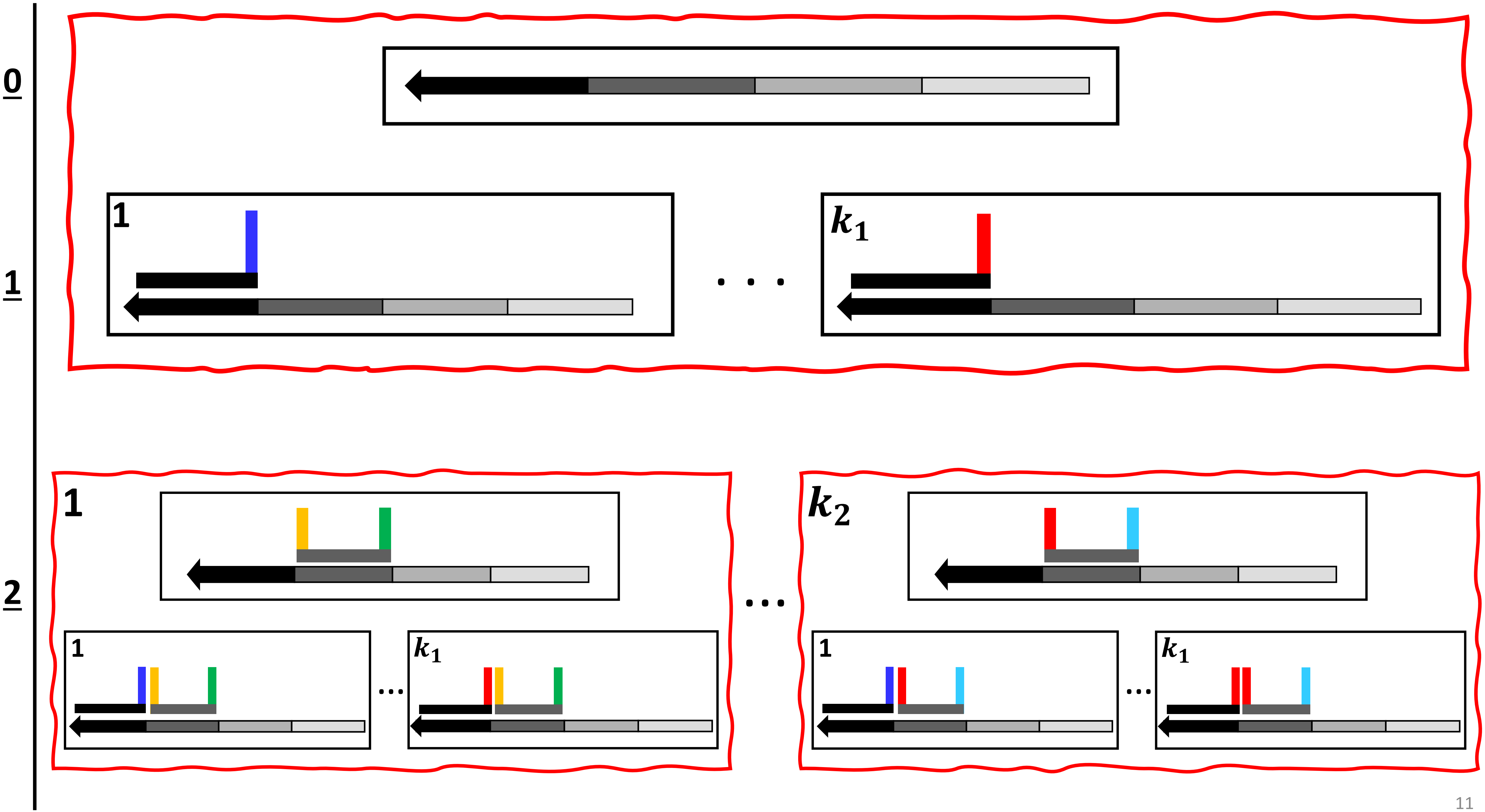
Is there recursive way to build these classes?

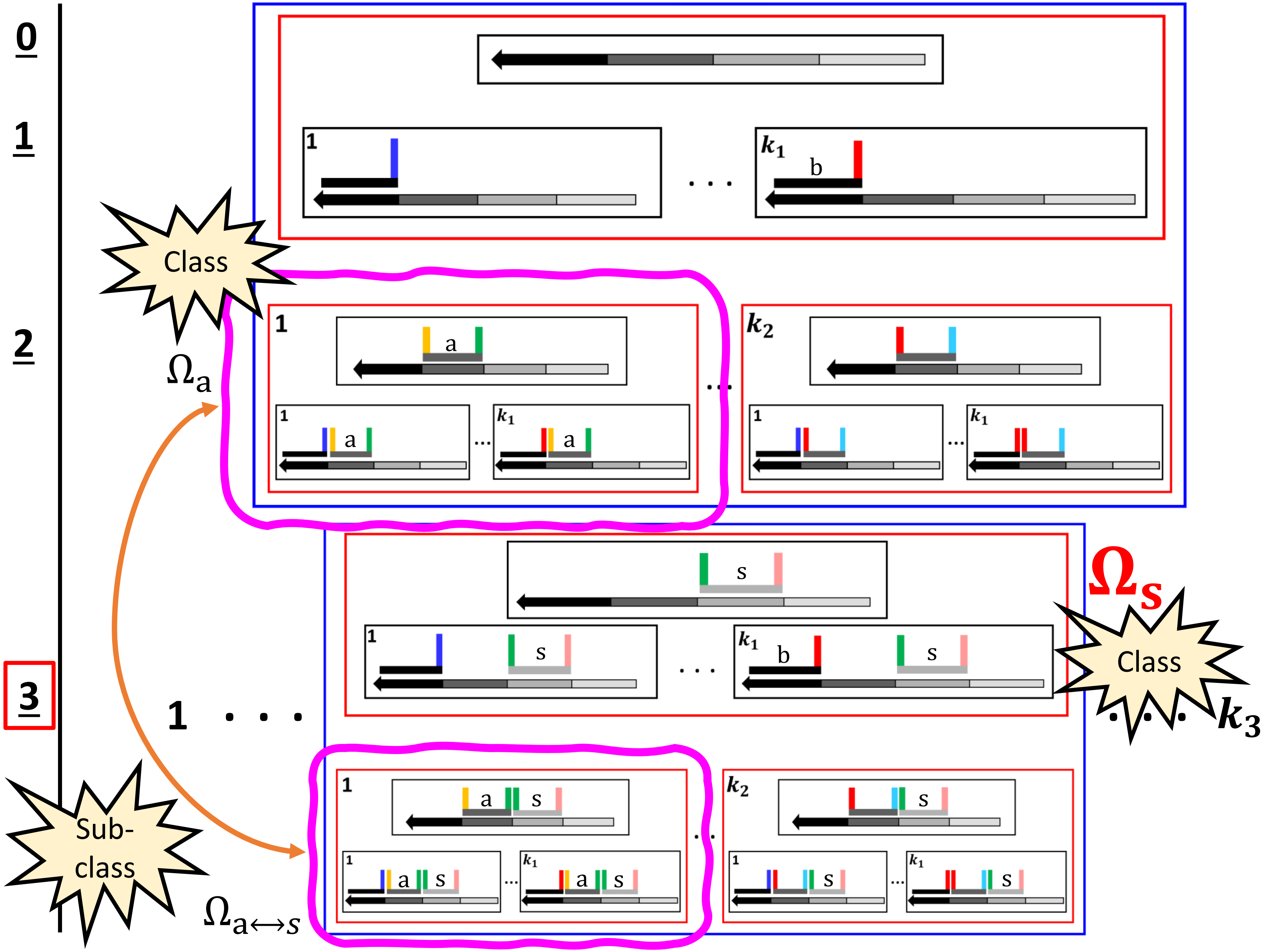


$$|\Omega| = (\boldsymbol{k} + \mathbf{1})^{N} = \mathbf{9}$$

$$Q = \sum_{X \in \boldsymbol{\Omega}} e^{-\Delta G(X)/k_{\mathrm{B}}T}$$



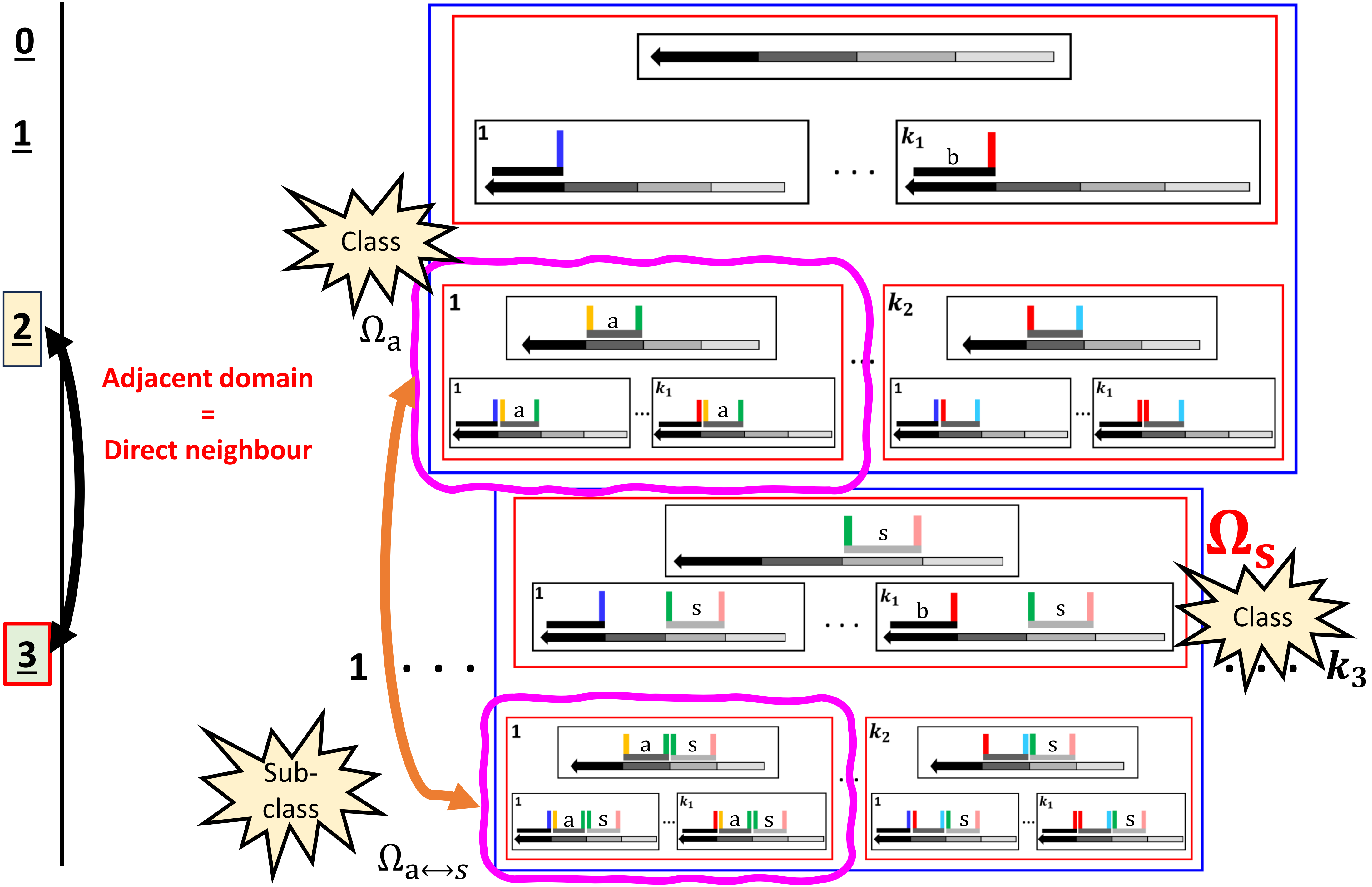

0
1
2
1
k_1
. . .
1
k_2
...
1
k_1
...
1
k_1
...



Layer = Domain

There is a 1-1 correspondence between each **class** of higher layers and a **sub-class** of any class in the current layer.

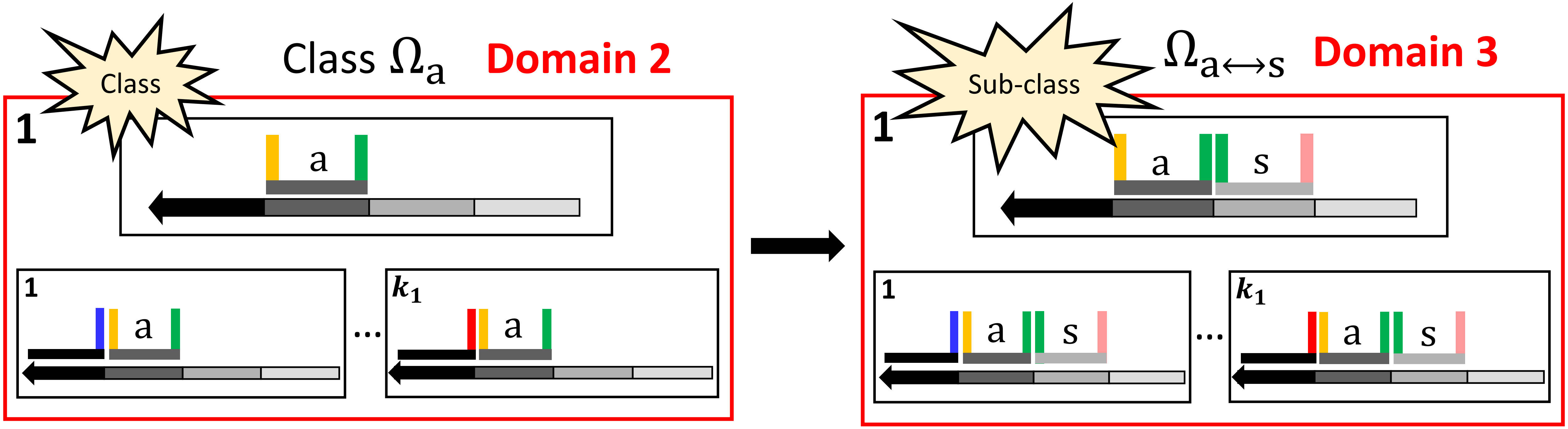
$$\begin{tikzcd} \Omega_a \arrow[d, leftrightarrow] \\ \Omega_{a\leftrightarrow s} \subset \boldsymbol{\Omega_s} \end{tikzcd}$$



0
1
2
3
1
k₁
b
. . .
Class
Ωa
1
a
1
a
…
k₁
a
…
k₂
1
…
k₁
Adjacent domain
=
Direct neighbour
s
1
s
. . .
k₁
b
s
Ωs
Class
k₃
1 . . .
1
a
s
1
a
s
…
k₁
a
s
…
k₂
s
1
s
…
k₁
s
Sub-
class
Ωa↔s

Can we use this recursive construction to propagate information through this hierarchy?

1 - Propagate information from only the previous layer (toehold matching possibility).



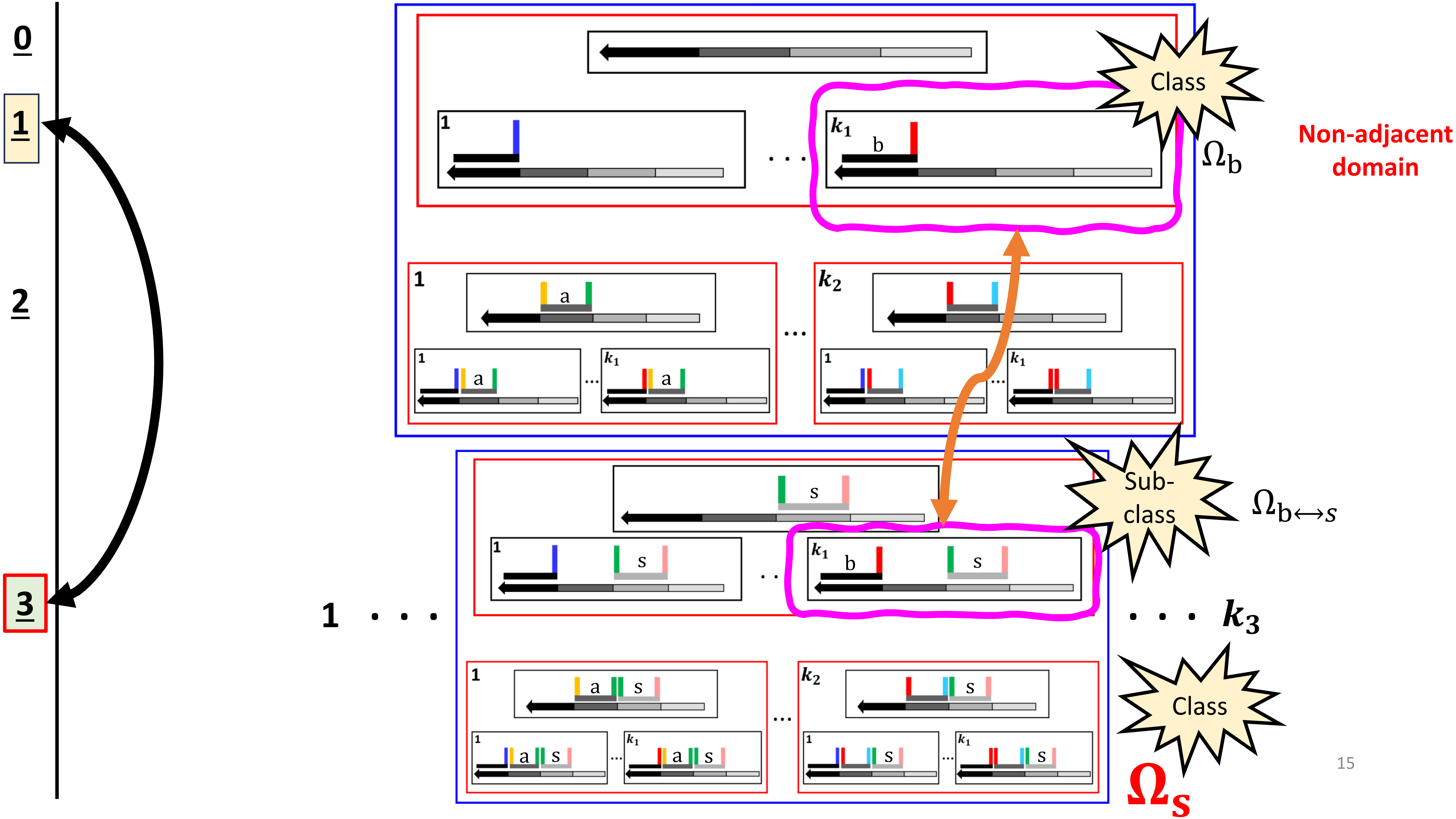
Class Ω_a **Domain 2**

$\Omega_{\mathrm{a}\leftrightarrow\mathrm{s}}$ **Domain 3**

Information given: $Q(\Omega_{\mathbf{a}})$

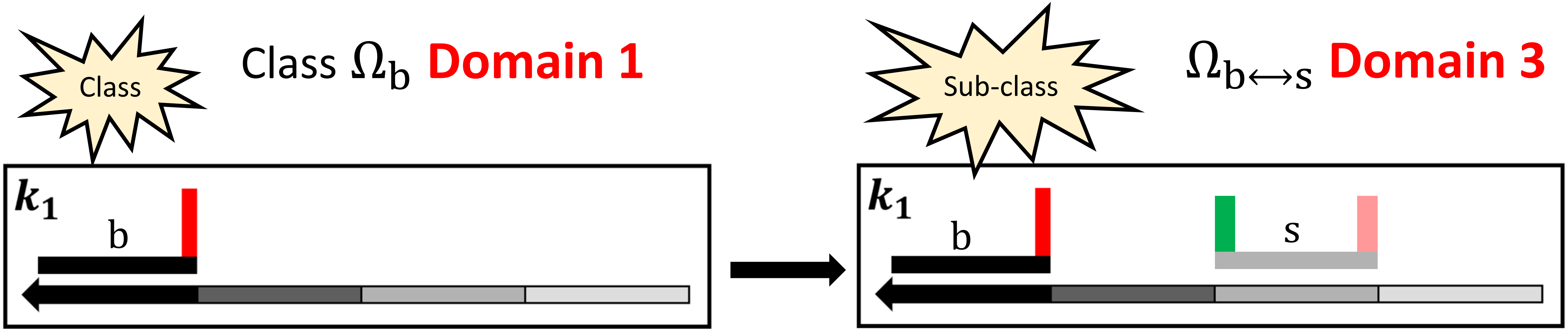
Information needed: $Q(\Omega_{\mathrm{a}\leftrightarrow\mathrm{s}})$

$$Q(\Omega_{\mathrm{a}\leftrightarrow\mathrm{s}}) = \ldots = \overbrace{e^{\frac{-\Delta G(M(s))+\Delta G^{assoc}}{k_B T}}}^{\text{Middle domain binding and the entropic cost term}} * \overbrace{e^{\frac{-\Delta G(R(a),L(s))}{k_B T}}}^{\text{Toehold matching possibility term}} * Q(\Omega_{\boldsymbol{a}})$$


0
1
2
3
1
b
k_1
a
k_2
Class
Ω_b
Non-adjacent domain
s
Sub-class
$\Omega_{b\leftrightarrow s}$
1 . . .
. . . k_3
Class
Ω_s

Can we use this recursive construction to propagate information through this hierarchy ?

2 - Propagate information from all other prior layers (No toehold matching possibility).



Class Ω_b **Domain 1**

$\Omega_{\mathrm{b}\leftrightarrow\mathrm{s}}$ **Domain 3**

Information given: $Q(\Omega_{\boldsymbol{b}})$

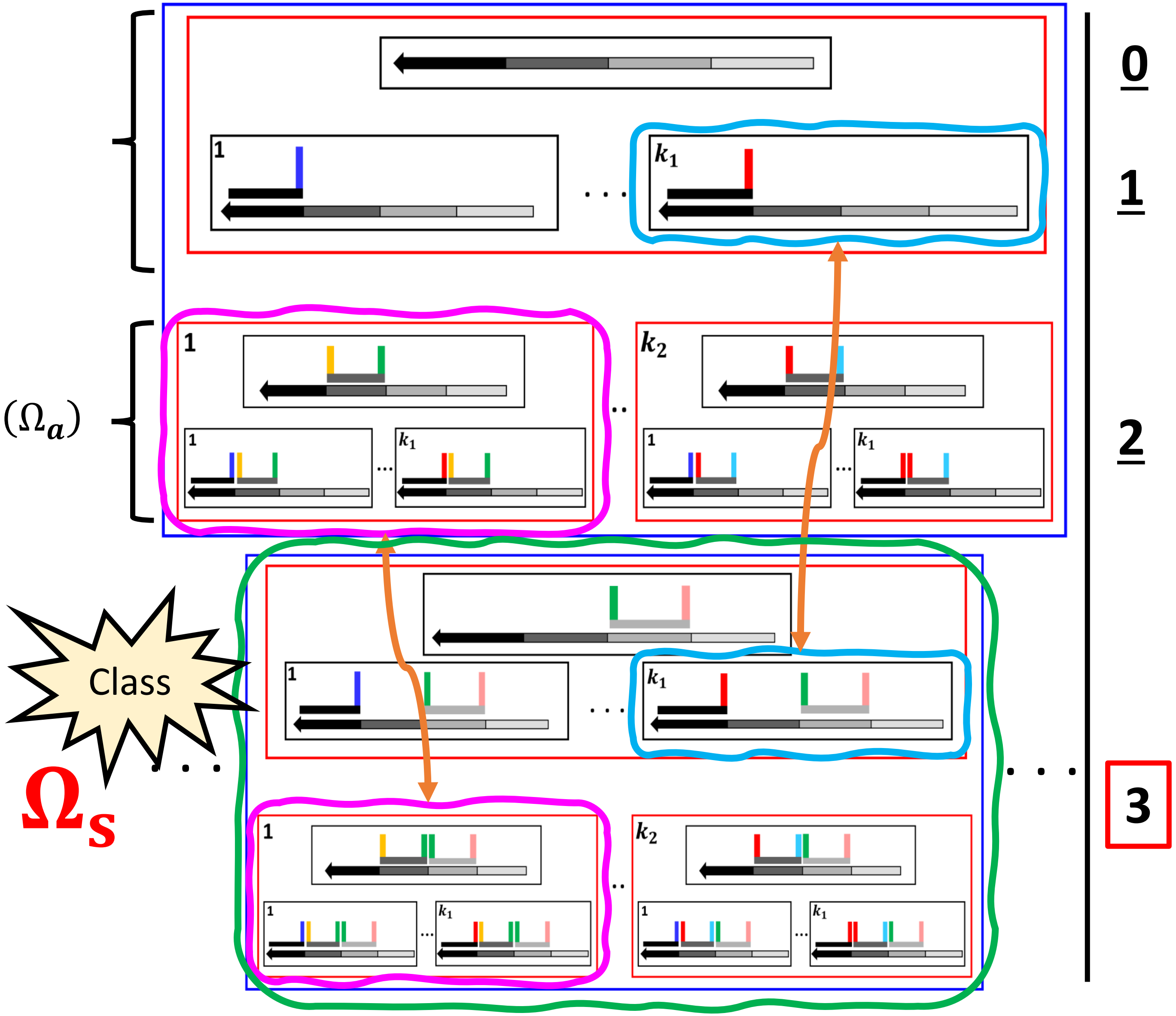
Information needed: $Q(\Omega_{\mathrm{b}\leftrightarrow\mathrm{s}})$

$$Q(\Omega_{\mathrm{b}\leftrightarrow\mathrm{s}}) = . . . = e^{\overbrace{\frac{-\Delta G(M(s))+\Delta \mathrm{G}^{\mathrm{assoc}}}{k_B T}}^{\text{Middle domain binding and the entropic cost term}}} * Q(\Omega_{\boldsymbol{b}})$$

$$\sum_b Q(\Omega_{b\leftrightarrow s}) = e^{\overbrace{\frac{-\Delta G(M(s))+\Delta G^{assoc}}{k_B T}}^{\text{Middle domain binding and the entropic cost}}} * Q(\Omega_{\boldsymbol{b}})$$

$$+$$

$$\sum_a Q(\Omega_{a\leftrightarrow s}) = e^{\frac{-\Delta G(M(s))+\Delta G^{assoc}}{k_B T}} * \overbrace{e^{\frac{-\Delta G(R(a),L(s))}{k_B T}}}^{\text{Toehold matching possibility}} * Q(\Omega_{\boldsymbol{a}})$$

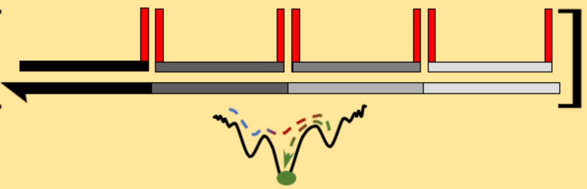


▶ **Theorem 2.** There is an

for a 1D SDC of length N w

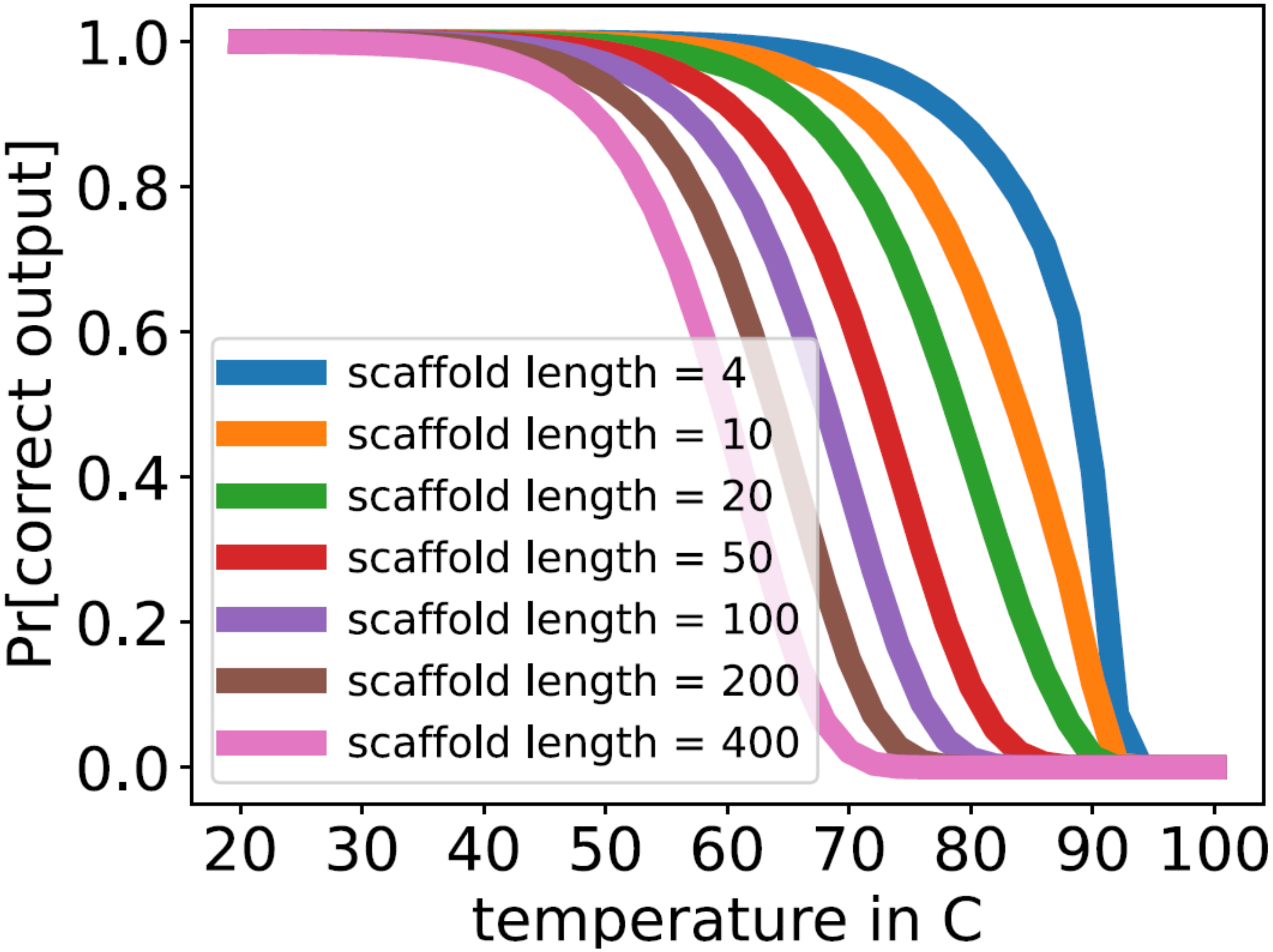
$$O(Q_s) = e^{\frac{-\Delta G(M(s))+\Delta G^{\text{assoc}}}{k_BT}} * \left[\sum \left(e^{\frac{-\Delta G(R(s'),}{k_BT}}\right.\right.$$

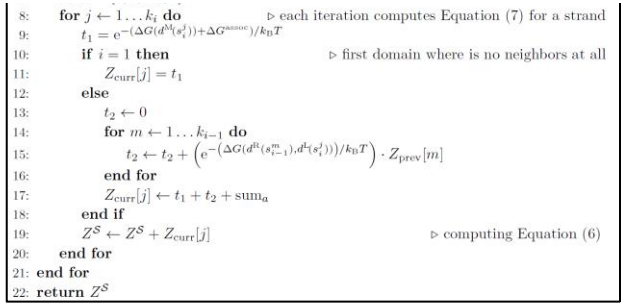
GOAL

Pr[] $\gg \sum_{c}$ Pr[c : *is another configuration*]

At equilibrium

Efficiently





```
    for $j \leftarrow 1 \dots k_i$ do                ▷ each iteration computes Equation (7) for a strand
        $t_1 = \mathrm{e}^{-(\Delta G(d^{\mathrm{M}}(s_i^j))+\Delta G^{\mathrm{assoc}})/k_{\mathrm{B}}T}$
        if $i = 1$ then                              ▷ first domain where is no neighbors at all
            $Z_{\mathrm{curr}}[j] = t_1$
        else
            $t_2 \leftarrow 0$
            for $m \leftarrow 1 \dots k_{i-1}$ do
                $t_2 \leftarrow t_2 + \left(\mathrm{e}^{-\left(\Delta G(d^{\mathrm{R}}(s_{i-1}^m), d^{\mathrm{L}}(s_i^j))\right)/k_{\mathrm{B}}T}\right) \cdot Z_{\mathrm{prev}}[m]$
            end for
            $Z_{\mathrm{curr}}[j] \leftarrow t_1 + t_2 + \mathrm{sum}_a$
        end if
        $Z^{\mathcal{S}} \leftarrow Z^{\mathcal{S}} + Z_{\mathrm{curr}}[j]$          ▷ computing Equation (6)
    end for
end for
return $Z^{\mathcal{S}}$
```

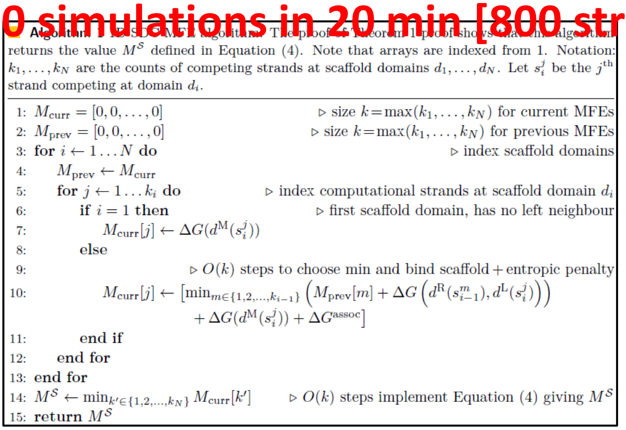
Benefits of Domain Based models !!

n to determine the domain-level MFE for

nds competing at each scaffold domain.

Previous: 1 simulation of length 13

Now: 280 simulations in 20 min [800 strands]

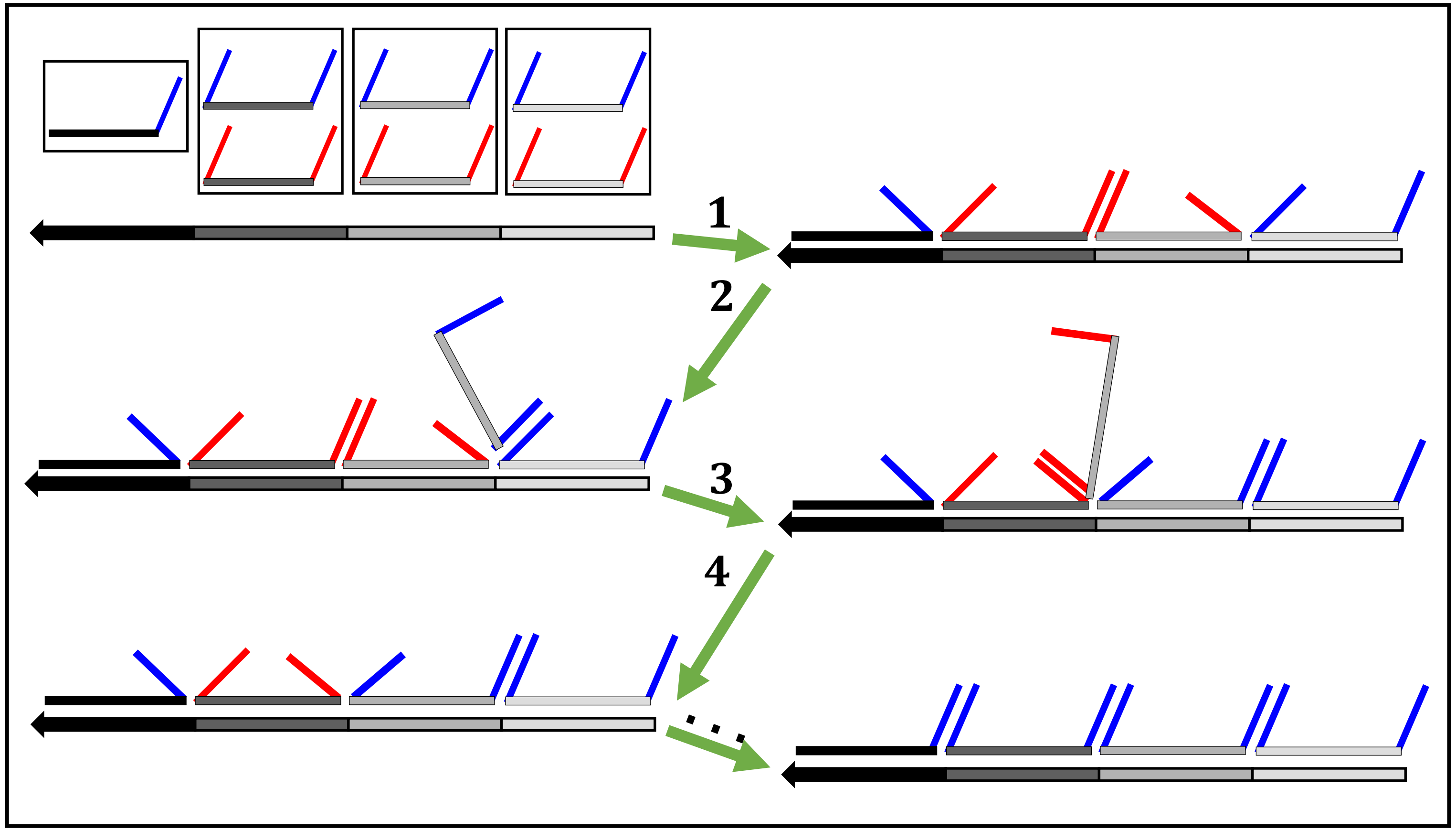


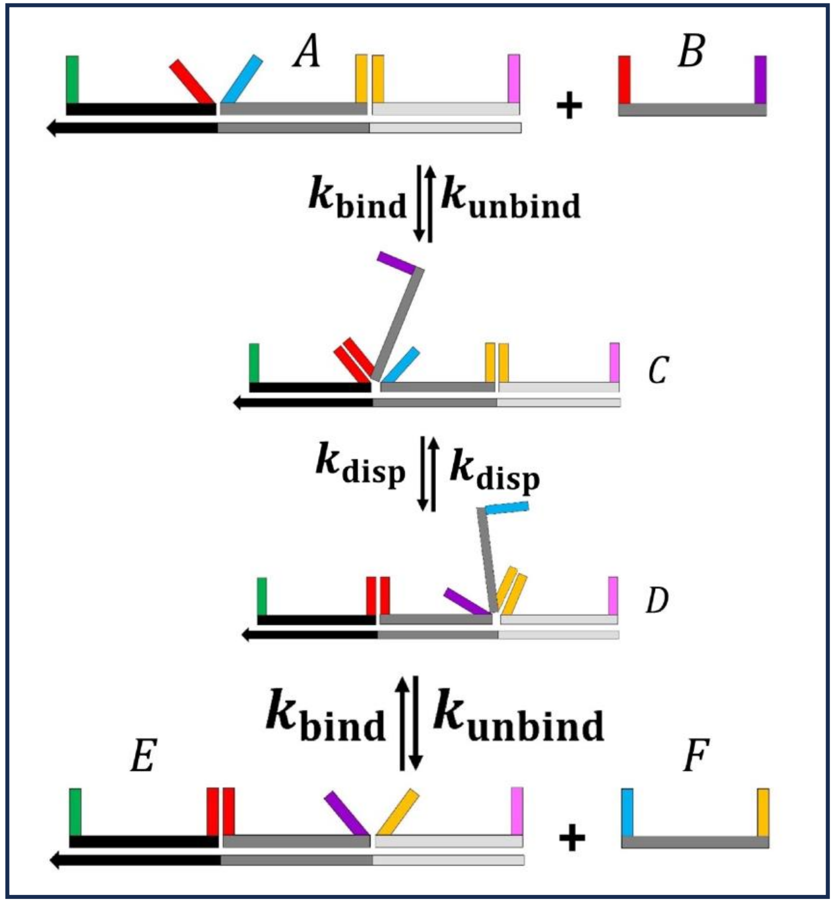
Algorithm 1 1D SDC MFE algorithm. The proof of Theorem 1 proof shows that the algorithm returns the value $M^{\mathcal{S}}$ defined in Equation (4). Note that arrays are indexed from 1. Notation: $k_1, \dots, k_N$ are the counts of competing strands at scaffold domains $d_1, \dots, d_N$. Let s_i^j be the j^{th} strand competing at domain d_i.

```
$M_{\mathrm{curr}} = [0, 0, \dots, 0]$              ▷ size $k = \max(k_1, \dots, k_N)$ for current MFEs
$M_{\mathrm{prev}} = [0, 0, \dots, 0]$              ▷ size $k = \max(k_1, \dots, k_N)$ for previous MFEs
for $i \leftarrow 1 \dots N$ do                      ▷ index scaffold domains
    $M_{\mathrm{prev}} \leftarrow M_{\mathrm{curr}}$
    for $j \leftarrow 1 \dots k_i$ do                ▷ index computational strands at scaffold domain $d_i$
        if $i = 1$ then                              ▷ first scaffold domain, has no left neighbour
            $M_{\mathrm{curr}}[j] \leftarrow \Delta G(d^{\mathrm{M}}(s_i^j))$
        else
                                                     ▷ $O(k)$ steps to choose min and bind scaffold + entropic penalty
            $M_{\mathrm{curr}}[j] \leftarrow [\min_{m \in \{1,2,\dots,k_{i-1}\}} \left(M_{\mathrm{prev}}[m] + \Delta G\left(d^{\mathrm{R}}(s_{i-1}^m), d^{\mathrm{L}}(s_i^j)\right)\right)$
                $+ \Delta G(d^{\mathrm{M}}(s_i^j)) + \Delta G^{\mathrm{assoc}}]$
        end if
    end for
end for
$M^{\mathcal{S}} \leftarrow \min_{k' \in \{1,2,\dots,k_N\}} M_{\mathrm{curr}}[k']$      ▷ $O(k)$ steps implement Equation (4) giving $M^{\mathcal{S}}$
return $M^{\mathcal{S}}$
```

Kinetic model for Scaffolded DNA computer

- The kinetic model for Scaffolded DNA Computer is a continuous-time Markov chain (CTMC) that satisfies detailed balance.

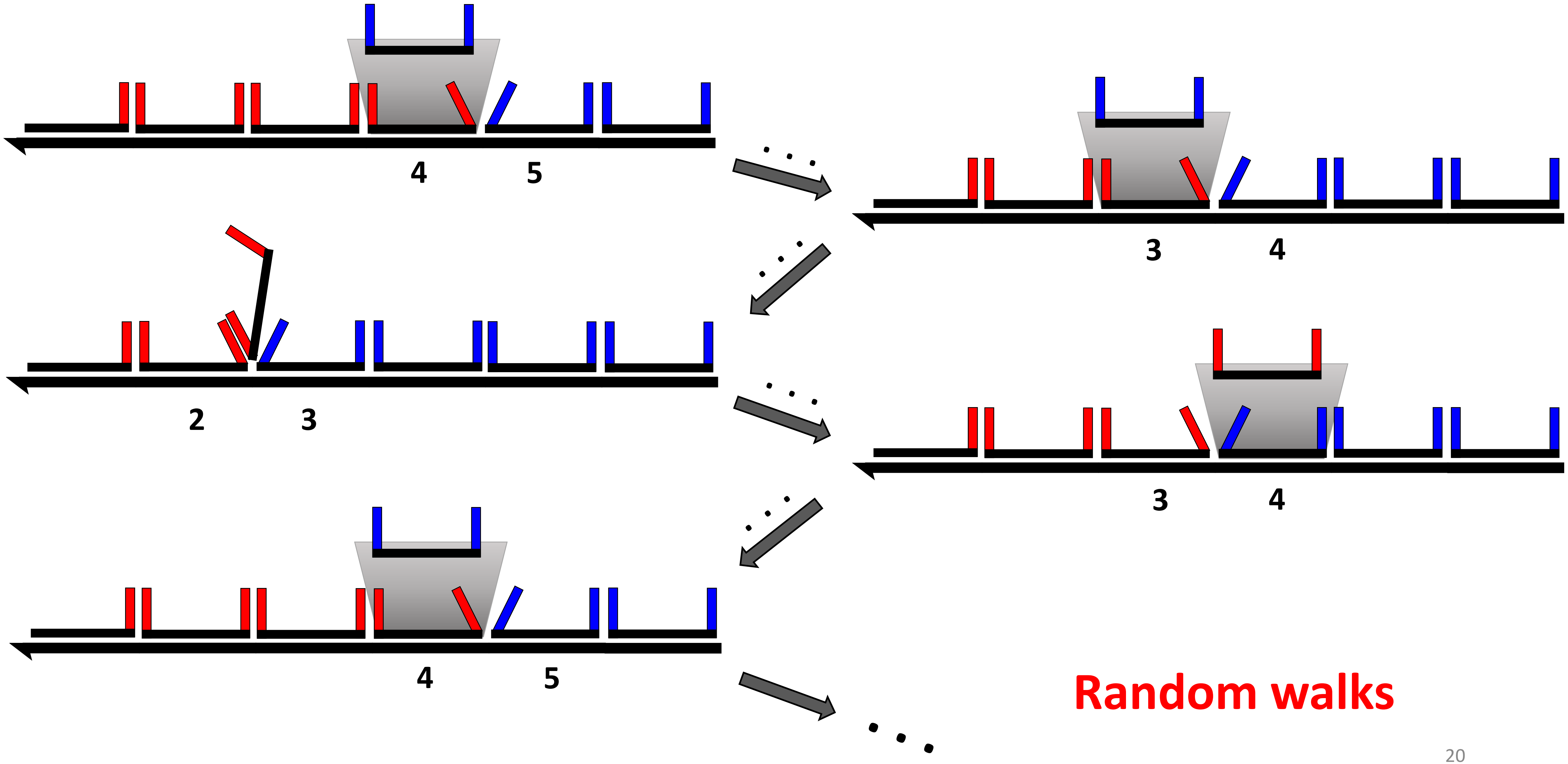




GOAL

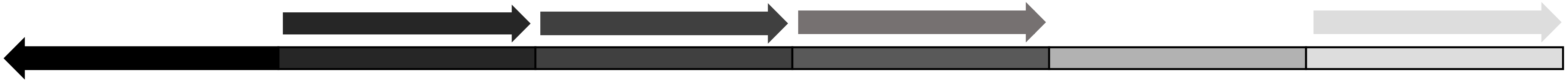
- Understanding system kinetics.
- Ability to propose some ideas that may help in speeding up the system.

Possible kinetic scenario with the Scaffolded DNA computer

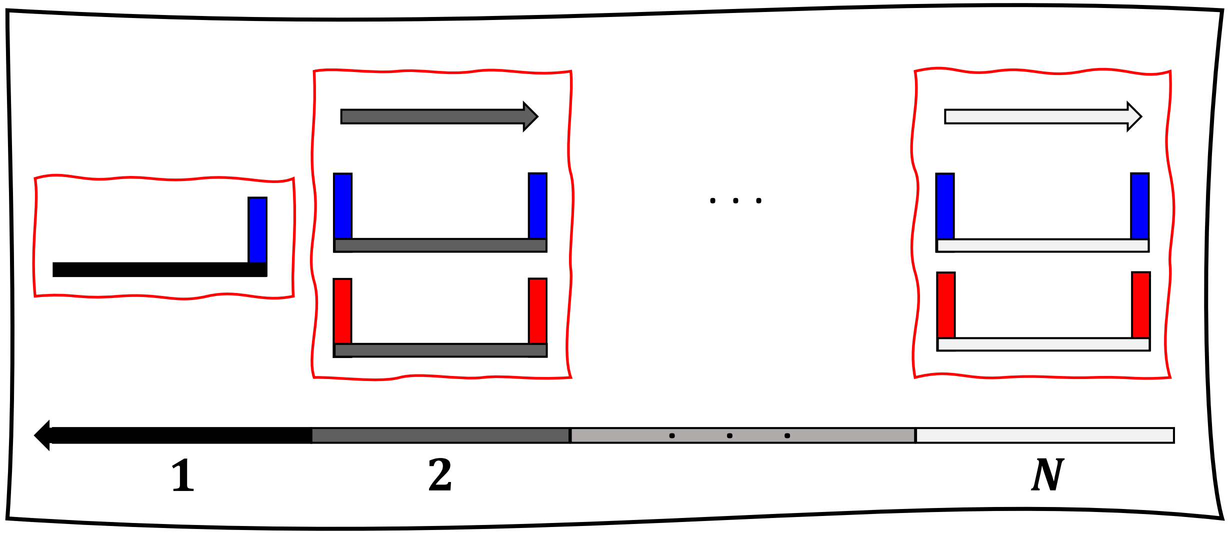


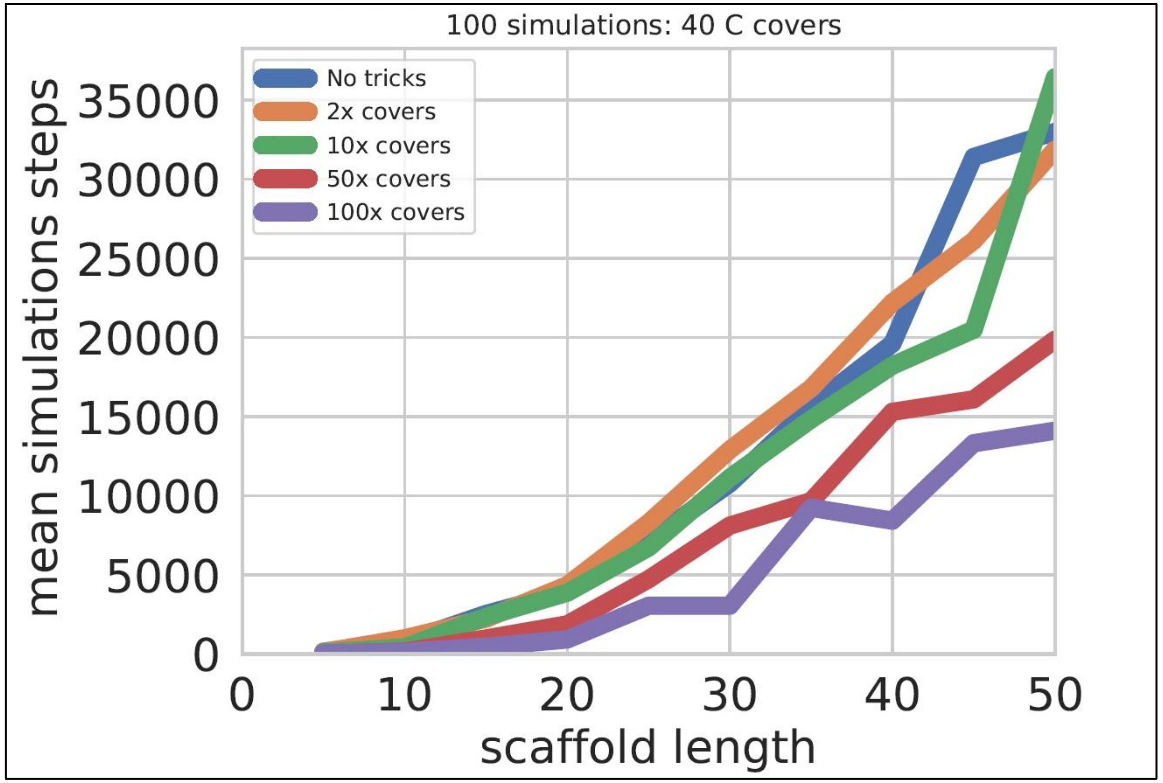
Random walks

Proposal 1: Covers

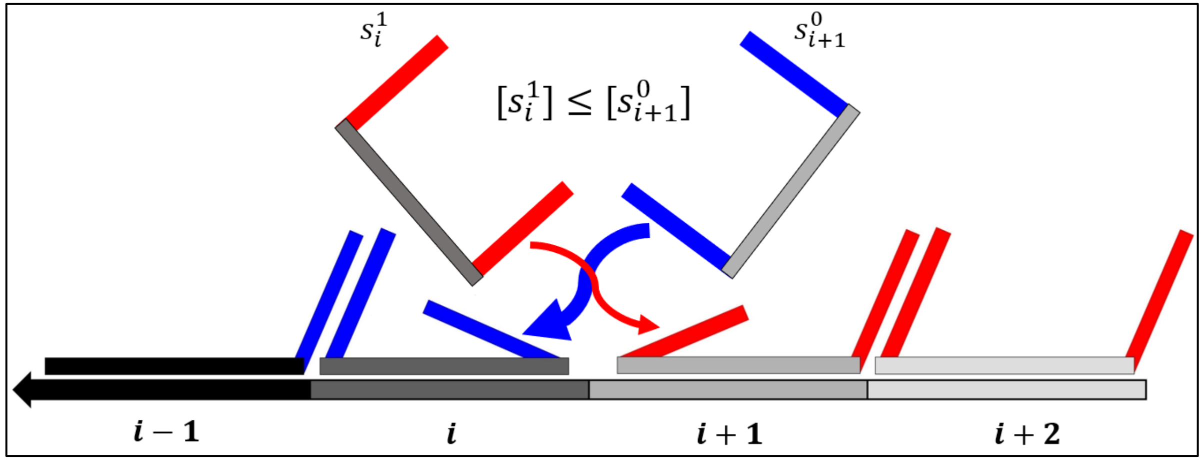


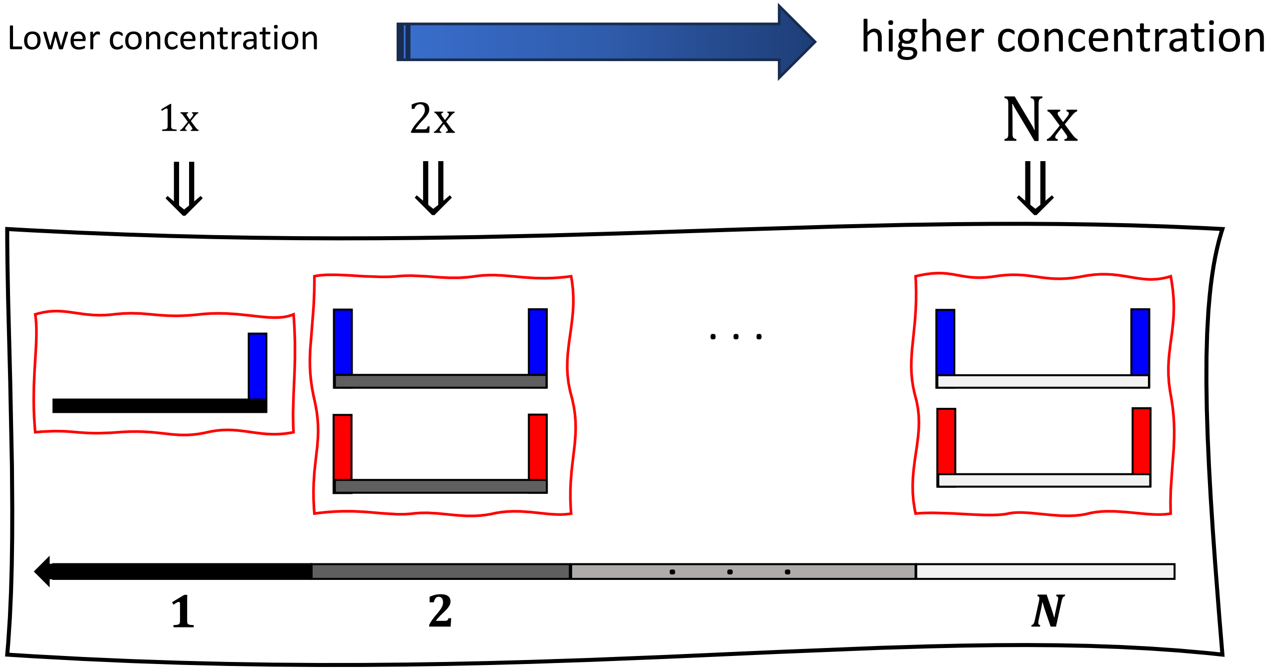
Scaffold with covers (here at the second, third, fourth, and sixth scaffold domains)

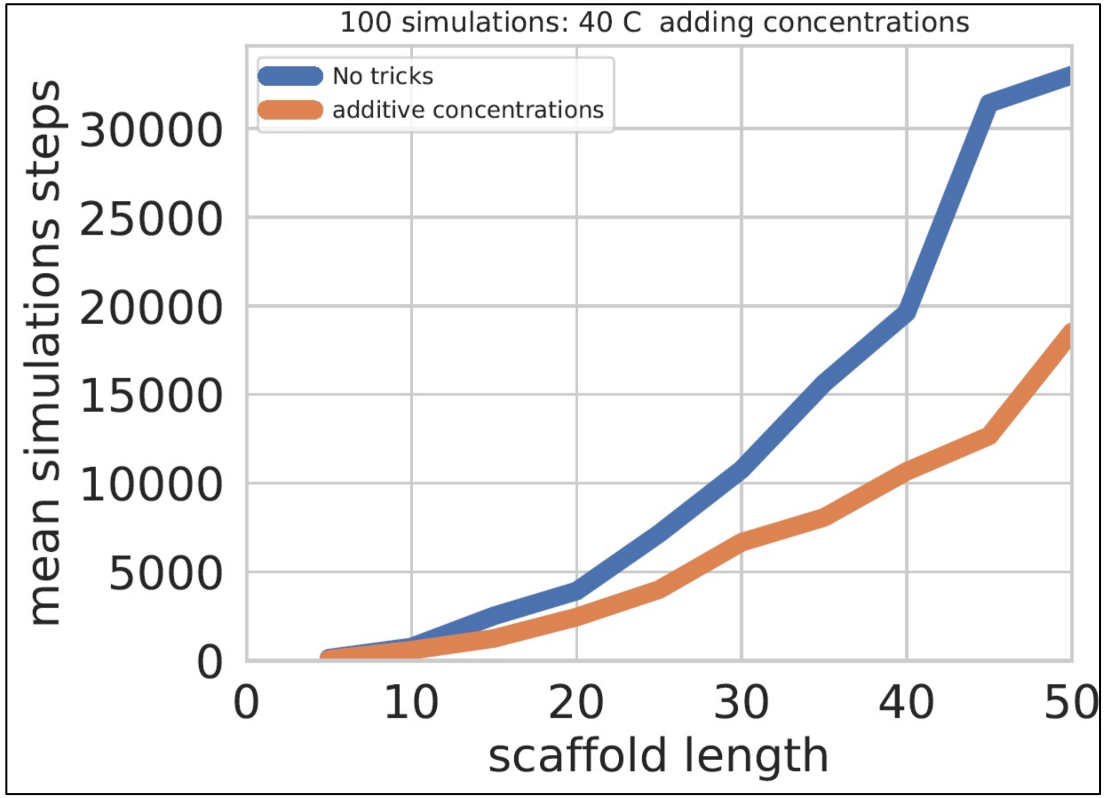




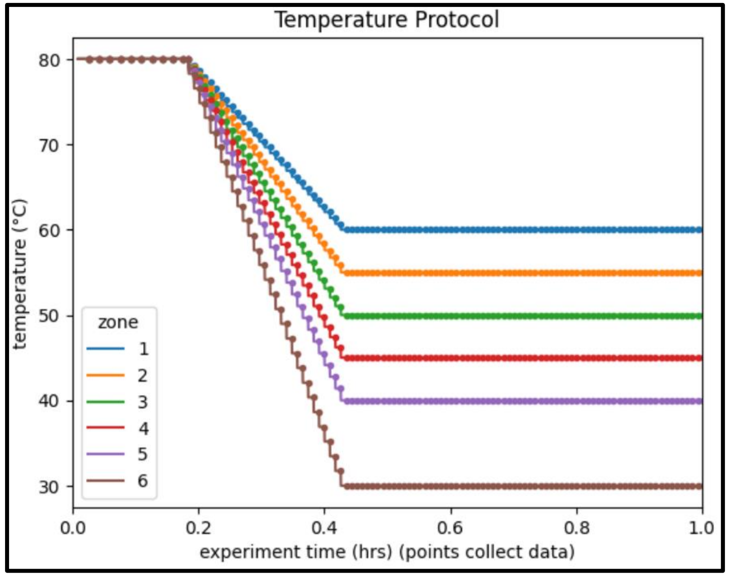
Proposal 2: Monotonically increasing competing strands concentrations along the scaffold

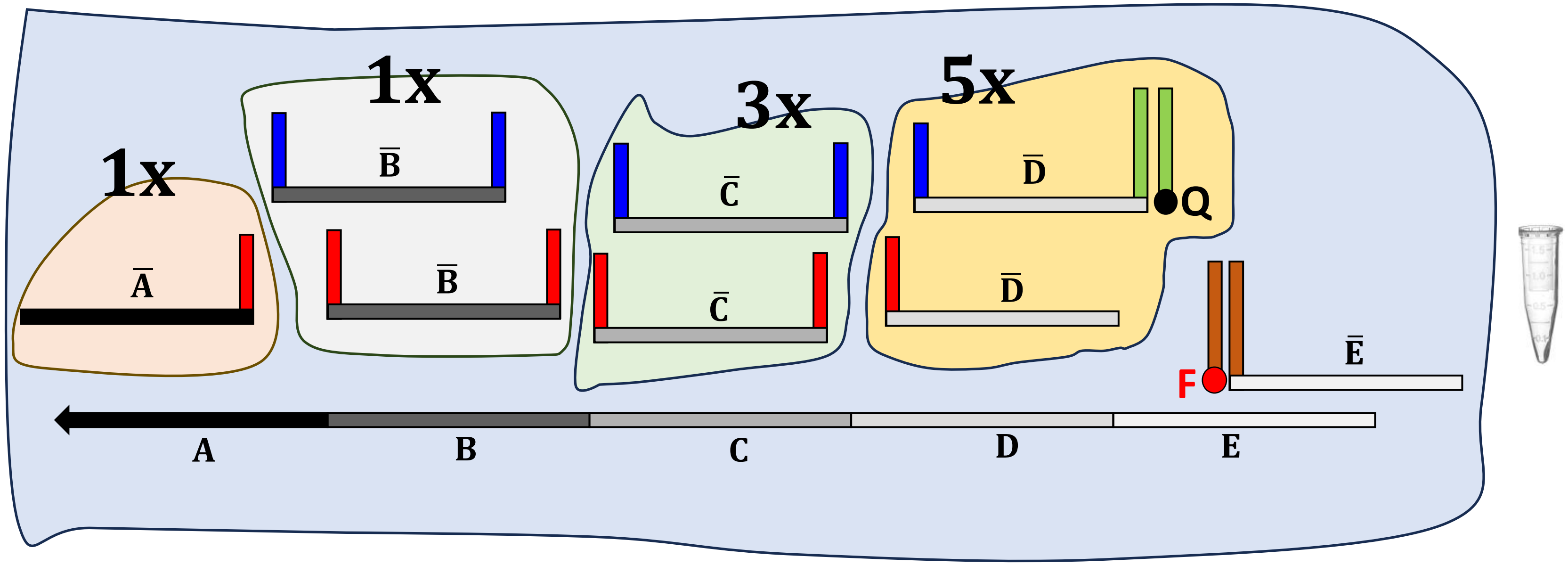


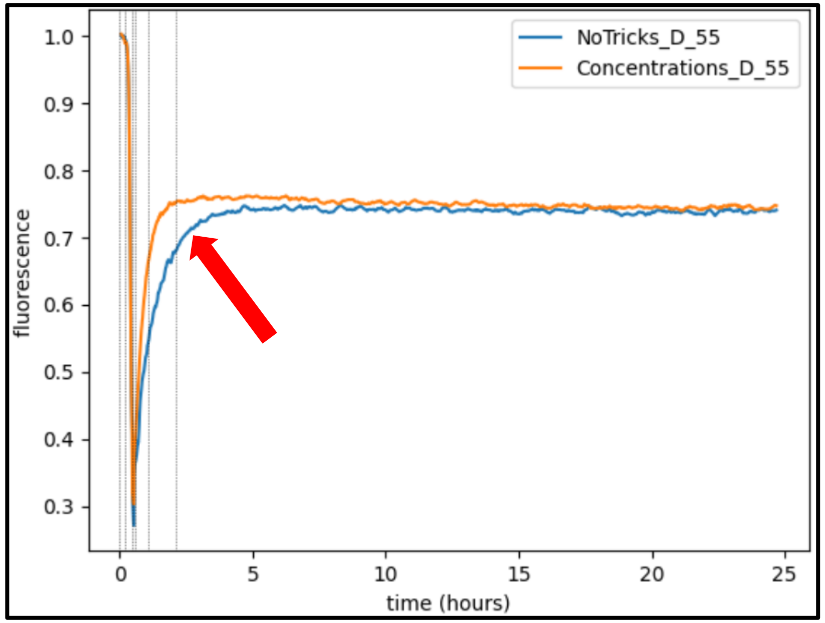


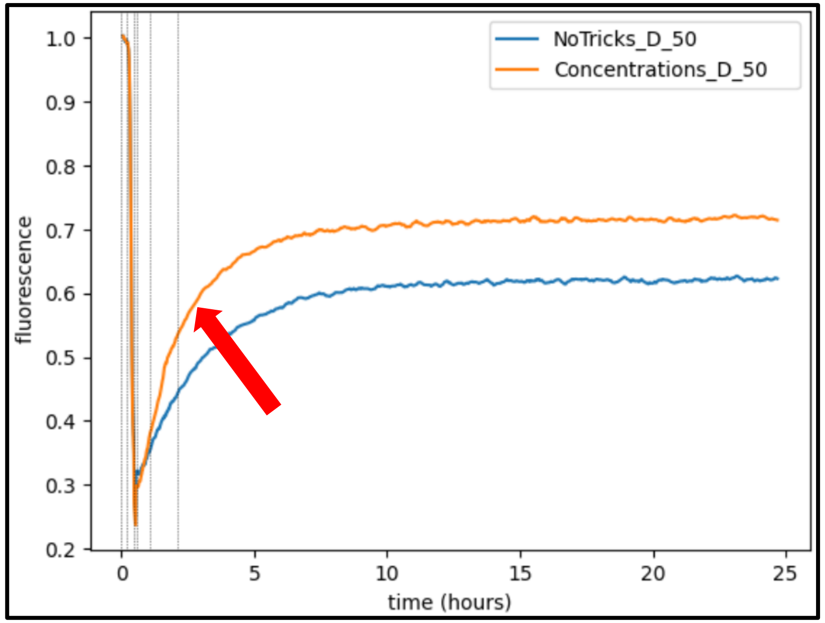


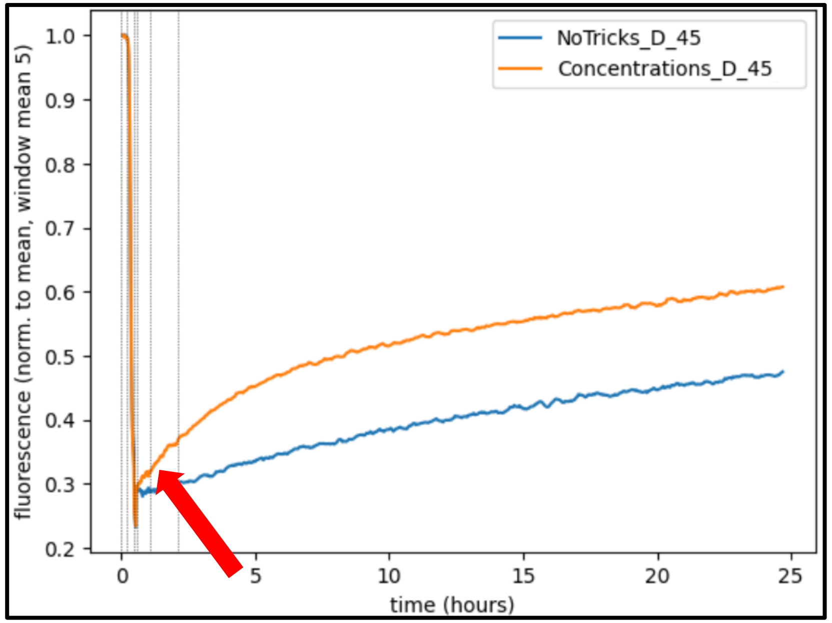
Concentration trick experiment [LATE INPUT] with concentrations [1x, 1x, 3x, 5x]

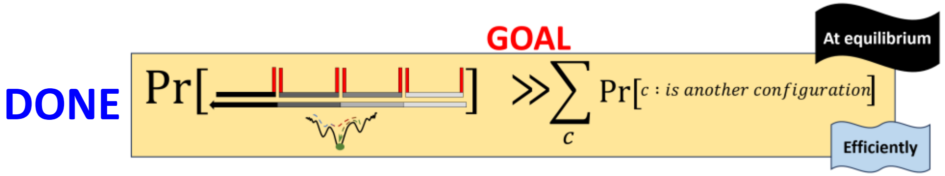











DONE

GOAL

At equilibrium

$$\Pr[\quad] \gg \sum_{c} \Pr[c : \textit{is another configuration}]$$

Efficiently

Conclusions

- Our polynomial time algorithms for MFE and Partition Function give some evidence that Scaffolded DNA Computer is thermodynamically favourable.

- The Scaffolded DNA Computer kinetic simulator confirms our intuition about the tricks that we think it will speed up the system.

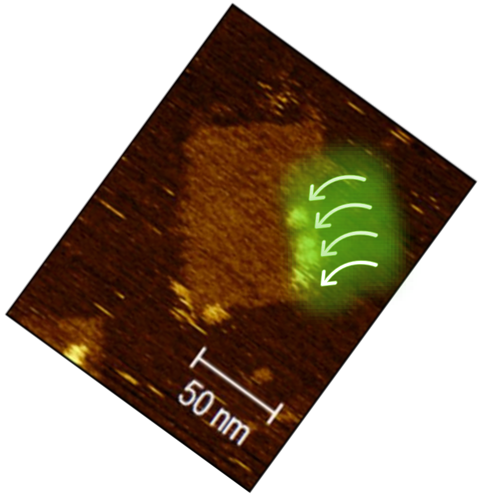
- Our preliminary experiments are promising with respect to the proposed tricks.

GOAL

In progress

- Understanding system kinetics.
- Ability to propose some ideas that may help in speeding up the system.

Future Work

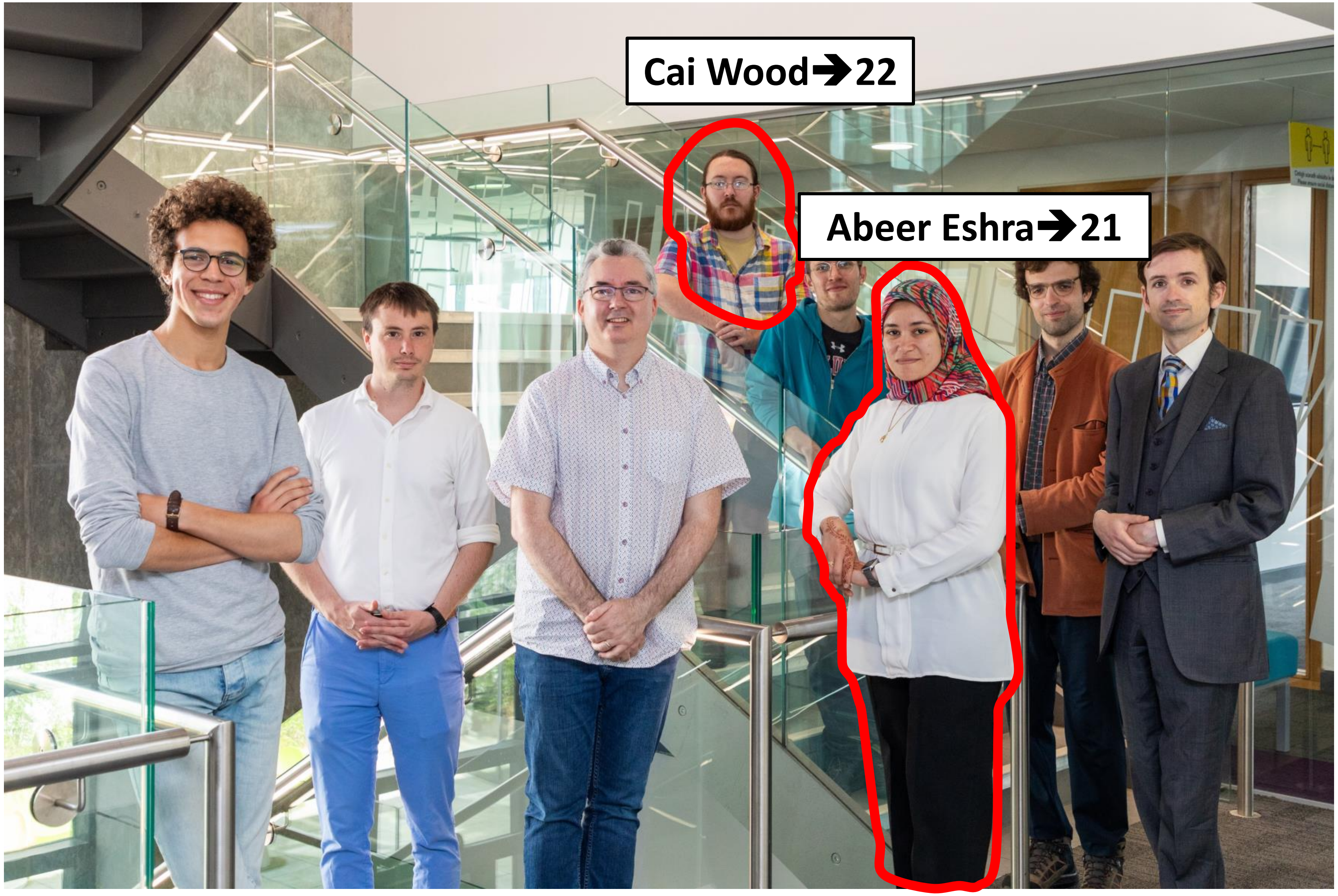


- Extending the work to the 2D case of "Algorithmic DNA Origami".

- Experimentally testing our tricks with bigger systems.

- Looking for other fast thermodynamic prediction algorithms for other engineered multistranded and/or pseudoknotted systems?
  - DNA strand displacement circuits
  - DNA tile-based self-assembly systems
  - DNA origami systems

Thanks



Hamilton Institute

Maynooth University
National University of Ireland Maynooth

dna.hamilton.ie

We're hiring!
Postdoc, PhD

sfi Science Foundation Ireland

erc

Acknowledgments: Dave Doty, David Soloveichik, Eric Winfree and Boya Wang.

European Innovation Council

Funded by the European Union