Predicting Kinetics of RNA–RNA Interaction

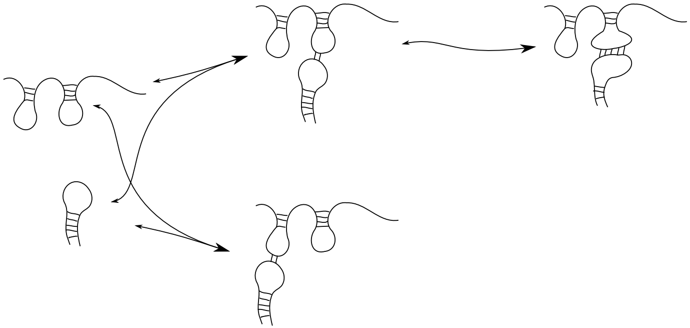
Sebastian Will[1] and Rolf Backofen[2]

[1]École Polytechnique - Institute Polytechnique de Paris

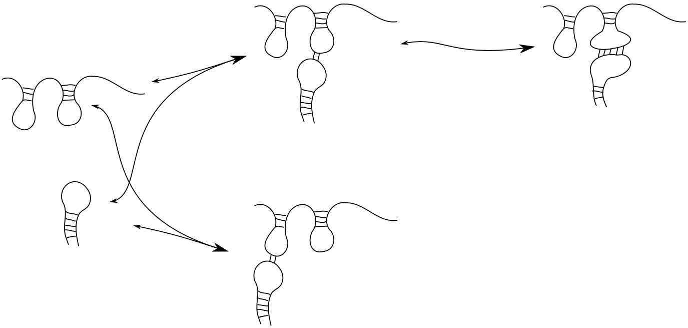
[2] University of Freiburg

Benasque'22

RNA–RNA interaction is a dynamic process



RNA–RNA interaction is a dynamic process



In principle, the interaction process can be studied:

1. *Model* as **Continuous-time Markov Process** (by **States** and **Transition rates**)
2. *Solve* **Master Equation**

Final objective: kinetics of interaction process

Definition (Markov process)

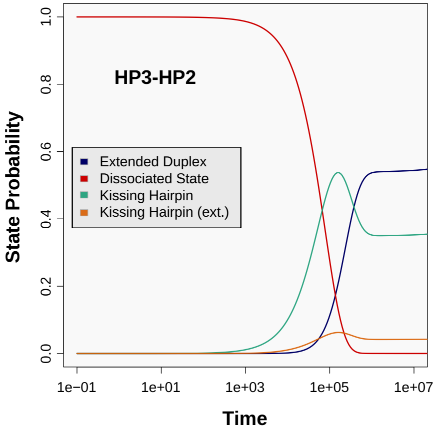
The **Markov process** $(\mathcal{X}, R, P_0)$ is the stochastic process governed by the **master equation**

$$\frac{dP}{dt} = RP,$$

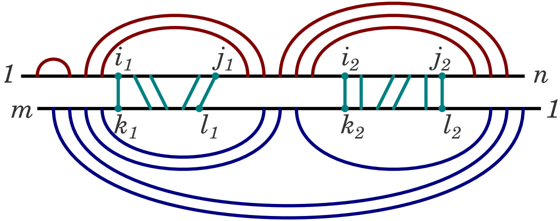
Solve as

$$P_t = \exp(tR)P_0.$$

$$\boxed{P_t(x) = \text{probability of } x \text{ at } t}$$

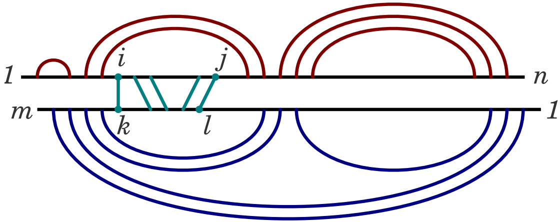


Joint (secondary) structures



Hybridization sites i_1, j_1, k_1, l_1 and i_2, j_2, k_2, l_2

Joint (secondary) structures



Single hybridization site i, j, k, l

E(**joint structure**) =
E(**structure 1**) + E(**hybridization**) + E(**structure 2**)[1]

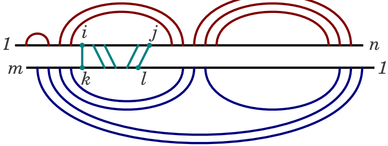
[1] $+E_{\text{duplex init}}$ if hybridization is non-empty

The joint secondary structure landscape

Given two RNAs A and B.

Energy landscape $\mathcal{L}^{\mathsf{j}} = (\mathcal{X}, \mathcal{N}, \mathsf{E})$

- state space $\mathcal{X}$ = set of joint secondary structures (of A and B)
- neighborhood $\mathcal{N}$ defined by single base pair moves;
  i.e. neighbors differ by 1 base pair
- energy function E on jss



The joint structure state space explodes quickly

Example: CGCAAUGCGAAUGCC and CGCGAUUCG

```
CGCAAUGCGAAUGCC
.((.........)).
.((..(....).)).
.((.(......))).
.((.((....)))).
.(((.......))).
.(((.(...).))).
.(((.(....)))).
......(......).
......((....)).
..((.......))..
.......(....)..
....((....))...
((.....))......
(((...)))......
       ⋮
```

×

```
CGCGAUUCG
(.......)
((.....))
..(.....)
..((...))
.(.....).
...(...).
.........
```

×

```
CGCAAUGCGAAUGCC&CGCGAUUCG
.......(((((((.&.)).)))))
.......(((((((.&.))).))))
........((((((.&.)).)))).
(((...((....)).&.))).....
(((............&.))).....
......(((......&)))......
........((((((.&.))).))).
((.((((((......&))).)))))
.((...))((((((.&.)).)))).
.....((((......&)))).....
((...((((......&))))...))
((.((((((......&)))).))))
.((...))((((((.&.))).))).
.......(((((...&....)))))
            ⋮
```

The joint structure state space explodes quickly

Example: `CGCAAUGCGAAUGCC` and `CGCGAUUCG`

```
CGCAAUGCGAAUGCC
.((.........)).
.((..(....).)).
.((.(......))).
.((.((....)))).
.(((.......))).
.(((.(...).))).
.(((.(....)))).
......(......).
......((....)).
..((.......))..
.......(....)..
....((....))...
((.....))......
(((...)))......
       ⋮
```

×

```
CGCGAUUCG
(.......)
((.....))
..(.....)
..((...))
.(.....).
...(...).
.........
```

×

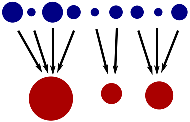
```
CGCAAUGCGAAUGCC&CGCGAUUCG
.......(((((((.&.)).)))))
.......(((((((.&.))).))))
........((((((.&.)).)))).
(((...((....)).&.))).....
(((............&.))).....
......(((......&)))......
........((((((.&.))).))).
((.((((((......&))).)))))
.((...))((((((.&.)).)))).
.....((((......&)))).....
((...((((......&))))...))
((.((((((......&)))).))))
.((...))((((((.&.))).))).
.......(((((...&....)))))
            ⋮
```

⇒ **Apply coarse graining (here, even several levels)**

Energy landscapes and general coarse graining

Coarse graining acts on landscapes $\mathcal{L}$ and yields macro-landscape $\mathcal{L}'$.

Coarse graining assigns states $x \in \mathcal{X}$ to macro-states $\alpha \in \mathcal{X}'$.

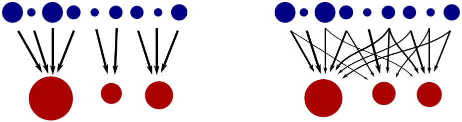


Energy landscapes and general coarse graining

Coarse graining acts on landscapes $\mathcal{L}$ and yields macro-landscape $\mathcal{L}'$.

Coarse graining assigns states $x \in \mathcal{X}$ to macro-states $\alpha \in \mathcal{X}'$.

discrete



continuous

Definition (Coarse graining matrix)

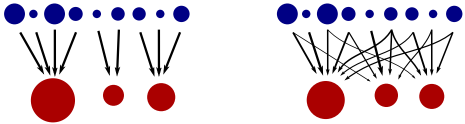
$\mathbf{C} \in \mathbb{R}^{|\mathcal{X}'| \times |\mathcal{X}|}$; $\quad \mathbf{C}_{\alpha x} = Pr[\, x \text{ assigned to } \alpha \,]$

Energy landscapes and general coarse graining

Coarse graining acts on landscapes $\mathcal{L}$ and yields macro-landscape $\mathcal{L}'$.

Coarse graining assigns states $x \in \mathcal{X}$ to macro-states $\alpha \in \mathcal{X}'$.

discrete *continuous*



Definition (Coarse graining matrix)

$\mathbf{C} \in \mathbb{R}^{|\mathcal{X}'|\times|\mathcal{X}|}$; $\mathbf{C}_{\alpha x} = Pr[\,x \text{ assigned to } \alpha\,]$

Remarks:

- in discrete case, $\mathbf{C} \in \{0,1\}^{|\mathcal{X}'|\times|\mathcal{X}|}$.
- columns of $\mathbf{C}$ sum to 1 (stochastic)

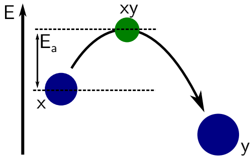
General coarse graining II

(Re)define **energy landscape** as $\mathcal{L} =: (\mathcal{X}, \tilde{Z}, Z)$, where

- $\mathcal{X}$ is a set of states x
- Z_x are Boltzmann weights of states $\exp(-\mathrm{E}(x)/RT)$
- $\tilde{Z}_{xy}$ are Boltzmann weights of transition states

Then,

- states x and y are connected if $Z_{xy} > 0$.
- rate (constant) from y to x: $r_{x \leftarrow y} = \tilde{Z}_{xy}/Z_y$ **(Arrhenius rate).**



$$r_{x \leftarrow y} = \exp(-\mathrm{E}_a/RT) = \exp(-(\mathrm{E}_{xy} - \mathrm{E}_x)/RT)$$

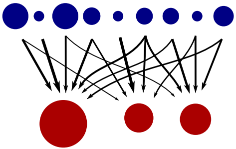
Example:

Metropolis rates

$\mathrm{E}_{xy} = max(E_x, E_y)$

General coarse graining III

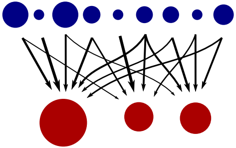
General coarse graining: $\mathcal{L} = (\mathcal{X}, \tilde{Z}, Z) \longrightarrow_{\mathbf{C}} \mathcal{L}' = (\mathcal{X}', \tilde{Z}', Z')$



CG matrix $\mathbf{C} \in \mathbb{R}^{|\mathcal{X}'|\times|\mathcal{X}|}$

General coarse graining III

General coarse graining: $\mathcal{L} = (\mathcal{X}, \tilde{Z}, Z) \longrightarrow_{\mathbf{C}} \mathcal{L}' = (\mathcal{X}', \tilde{Z}', Z')$



CG matrix $\mathbf{C} \in \mathbb{R}^{|\mathcal{X}'| \times |\mathcal{X}|}$

Then:

- CG by $\mathbf{C}$ determines state weights:

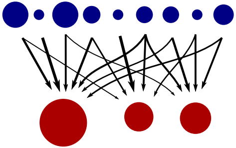
$$Z' = \mathbf{C}Z \qquad \equiv \qquad Z'_\alpha = \sum_{x \in \mathcal{X}} \mathbf{C}_{\alpha x} Z_x$$

- CG by $\mathbf{C}$ determines transition weights:

$$\tilde{Z}'_{\alpha\beta} = \sum_{x,y \in \mathcal{X}} \mathbf{C}_{\alpha x} \cdot \mathbf{C}_{\beta y} \cdot \tilde{Z}_{xy} \qquad \text{(canonical CG)}$$

General coarse graining III

General coarse graining: $\mathcal{L} = (\mathcal{X}, \tilde{Z}, Z) \longrightarrow_{\mathbf{C}} \mathcal{L}' = (\mathcal{X}', \tilde{Z}', Z')$



CG matrix $\mathbf{C} \in \mathbb{R}^{|\mathcal{X}'| \times |\mathcal{X}|}$

Then:

- CG by **C** determines state weights:

$$Z' = \mathbf{C}Z \qquad \equiv \qquad Z'_\alpha = \sum_{x \in \mathcal{X}} \mathbf{C}_{\alpha x} Z_x$$

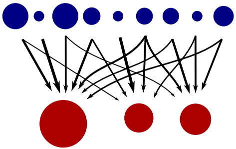
- CG by **C** determines transition weights:

$$\tilde{Z}'_{\alpha\beta} = \sum_{x,y \in \mathcal{X}} \mathbf{C}_{\alpha x} \cdot \mathbf{C}_{\beta y} \cdot \tilde{Z}_{xy} \qquad (\textbf{canonical CG})$$

discrete case: $r_{\alpha \leftarrow \beta} = \sum_{x \in \alpha, y \in \beta} Pr[y \mid \beta] r_{x \leftarrow y}$

General coarse graining III

General coarse graining: $\mathcal{L} = (\mathcal{X}, \tilde{Z}, Z) \longrightarrow_{\mathbf{C}} \mathcal{L}' = (\mathcal{X}', \tilde{Z}', Z')$



CG matrix $\mathbf{C} \in \mathbb{R}^{|\mathcal{X}'| \times |\mathcal{X}|}$

Then:

- CG by **C** determines state weights:

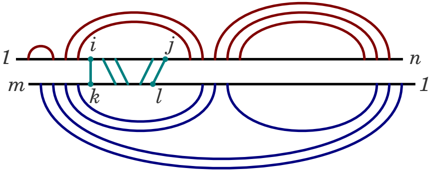
$$Z' = \mathbf{C}Z \qquad \equiv \qquad Z'_\alpha = \sum_{x \in \mathcal{X}} \mathbf{C}_{\alpha x} Z_x$$

- CG by **C** determines transition weights:

$$\tilde{Z}'_{\alpha\beta} = \sum_{x,y \in \mathcal{X}} \mathbf{C}_{\alpha x} \cdot \mathbf{C}_{\beta y} \cdot \tilde{Z}_{xy} \qquad \text{(canonical CG)}$$

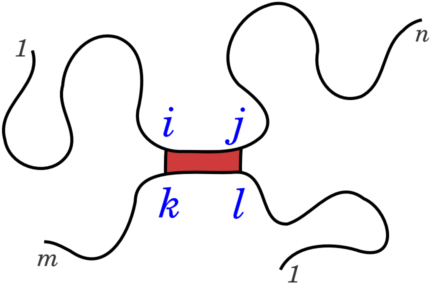
general case: $r_{\alpha \leftarrow \beta} = \sum_{x,y \in \mathcal{X}} \mathbf{C}_{\alpha x} \cdot \mathbf{C}_{\beta y} Pr[y \mid \beta] r_{x \leftarrow y}$

RNAup-like model



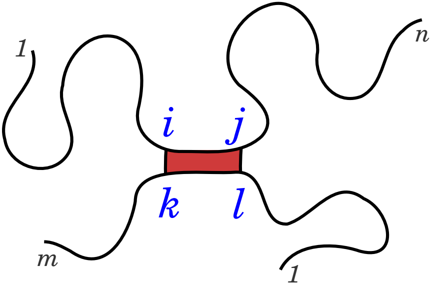
joint structure
with hybridization site (i,j,k,l)

RNAup-like model

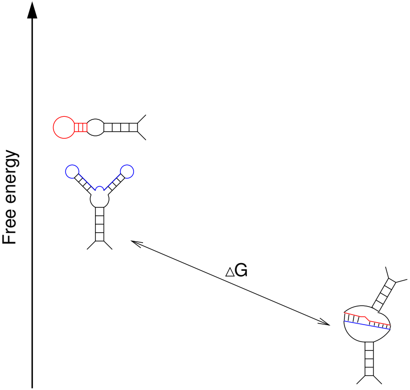


hybdridization site state (i,j,k,l)

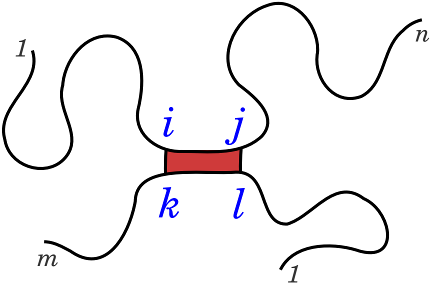
RNAup-like model



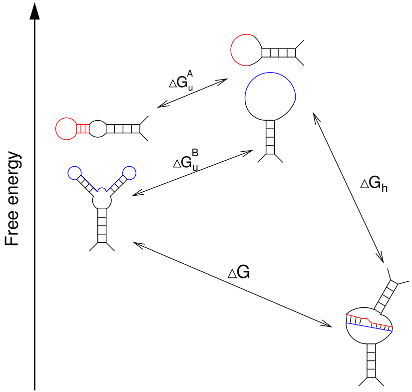
hybdridization site state (i,j,k,l)



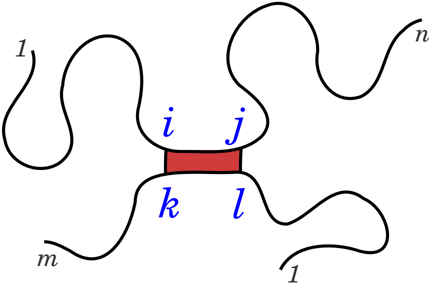
RNAup-like model



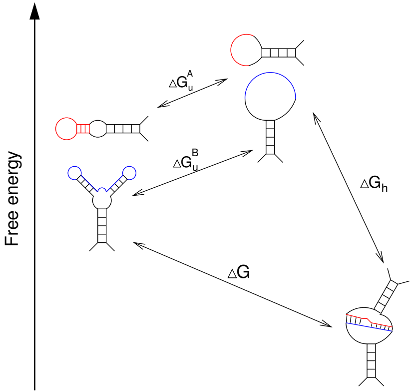
hybdridization site state (i,j,k,l)



RNAup-like model

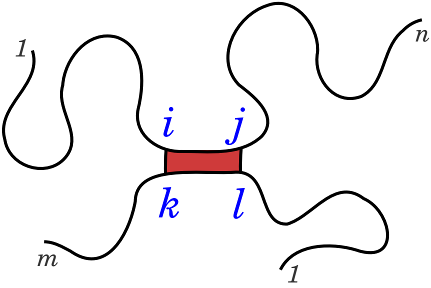


hybdridization site state (i,j,k,l)

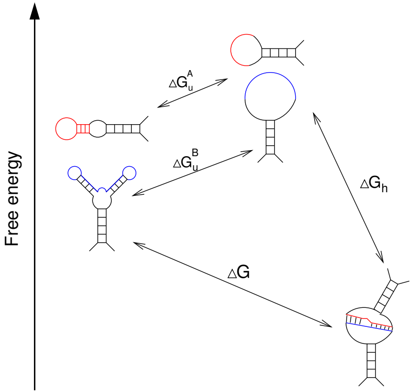


$$\Delta G(i,j,k,l) := \Delta G_u^A(i,j) + \Delta G_u^B(k,l) + \Delta G_h(i,j,k,l)$$

RNAup-like model

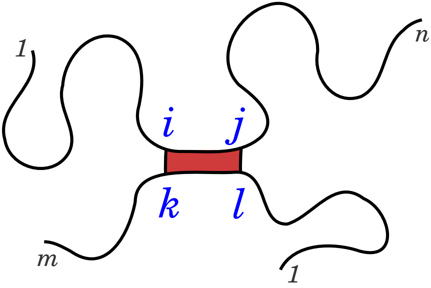


hybdridization site state (i,j,k,l)

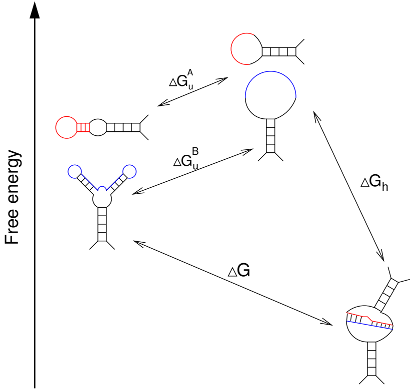


$$Z^{\mathrm{h}}(i,j,k,l) := \sum_{\text{joint structure } x \text{ with site } (i,j,k,l)} \exp(-E(x)/RT)$$

RNAup-like model

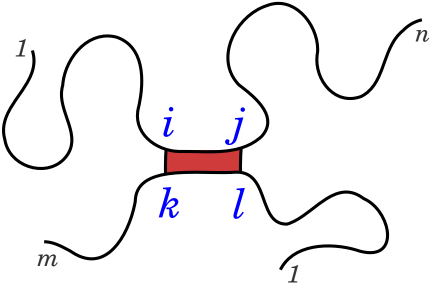


hybdridization site state (i,j,k,l)

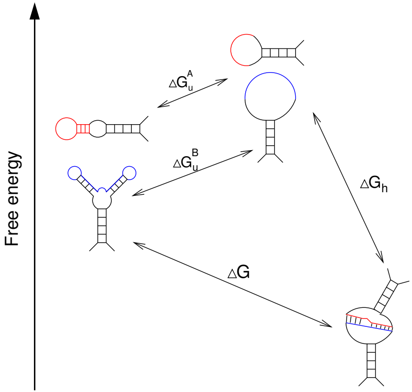


$$Z^{\mathrm{h}}(i,j,k,l) := Z^{A}(i,j) \cdot Z^{B}(k,l) \cdot Z^{\mathrm{hyb}}(i,j,k,l)$$

RNAup-like model



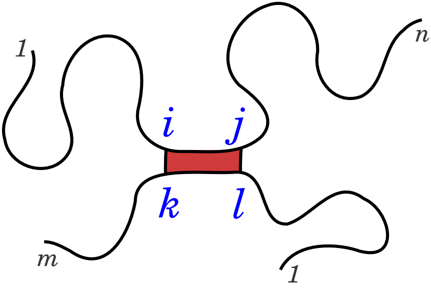
hybdridization site state (i,j,k,l)

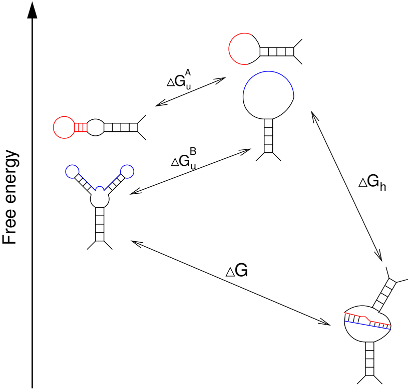


$$Z^{\mathrm{h}}(i,j,k,l) := Z^{A}(i,j) \cdot Z^{B}(k,l) \cdot Z^{\mathrm{hyb}}(i,j,k,l)$$

Hybridization site states **coarse grain** joint structures: $Z^{\mathrm{h}} = \mathbf{C}^{\mathrm{h}} Z^{\mathrm{j}}$

RNAup-like model



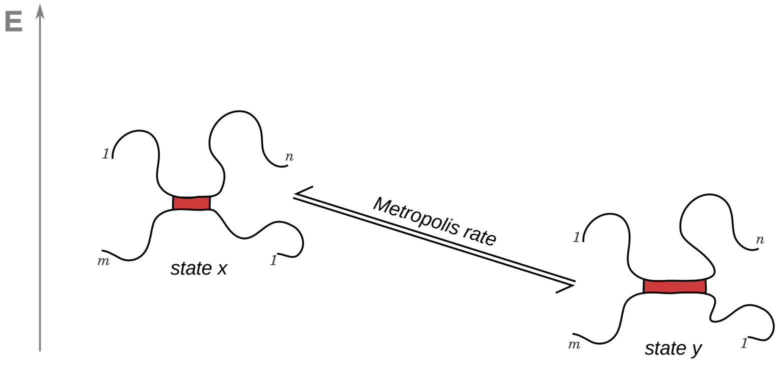


hybdridization site state (i,j,k,l)

$$Z^{\mathrm{h}}(i,j,k,l) := Z^{A}(i,j) \cdot Z^{B}(k,l) \cdot Z^{\mathrm{hyb}}(i,j,k,l)$$

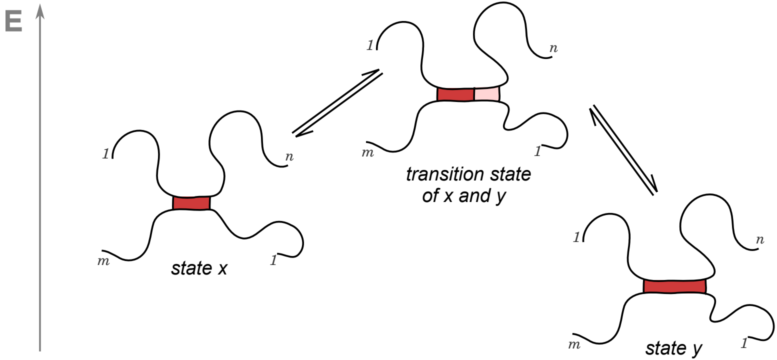
+ Fewer States **+ Efficient** **+ Complex interactions**

Transition rates for the hybridization site states



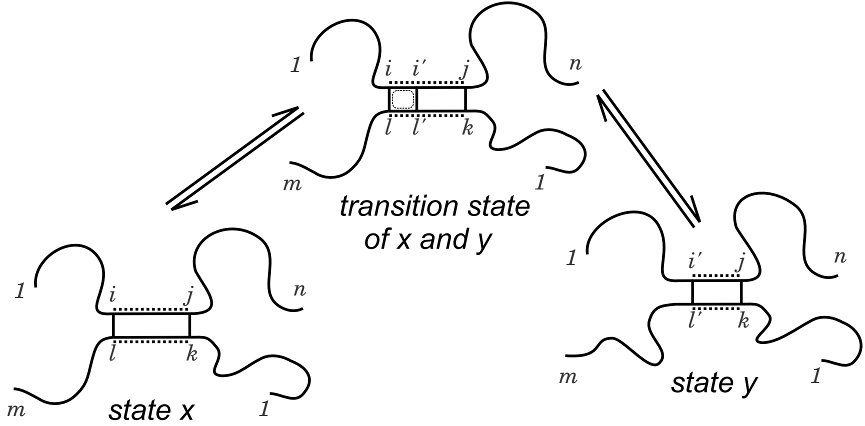
Transition rates defined by weights of (complex) transition states

Transition rates for the hybridization site states

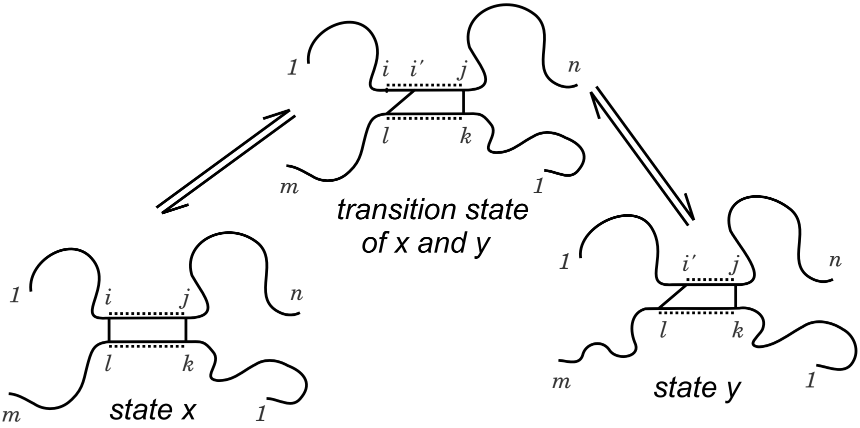


Transition rates defined by weights of (complex) transition states

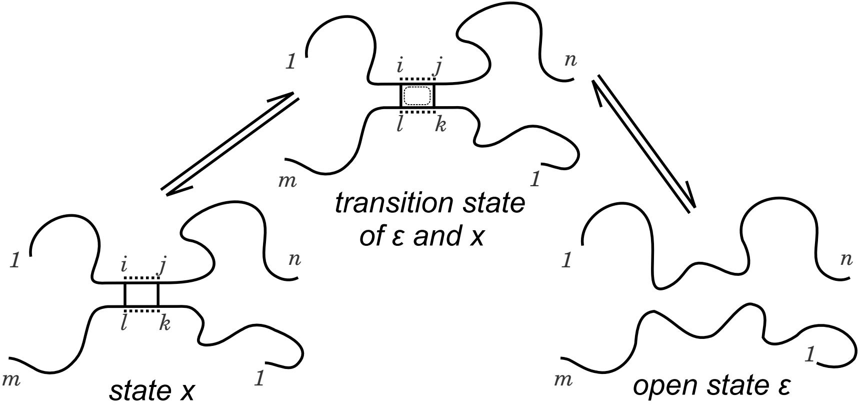
Grow and Shrink Moves



Shift Moves


1
i
i′
j
n
l
k
m
1
transition state
of x and y
1
i
j
n
l
k
m
state x
1
1
i′
j
n
l
k
m
state y
1

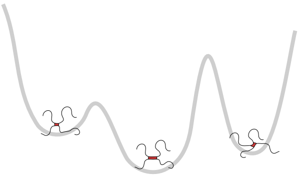
Association and Dissociation



Further coarse graining

Continous CG with 'traditional' discrete CG as pre-processing

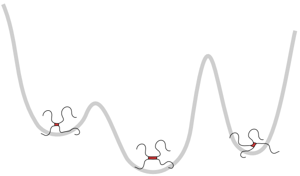
1. **Discrete:** partition into *gradient basins*



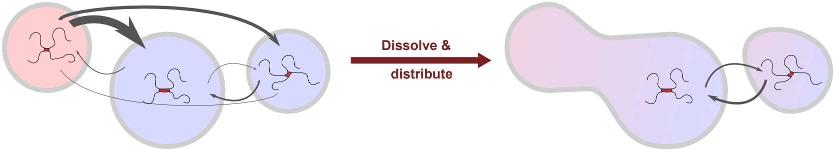
Further coarse graining

Continous CG with 'traditional' discrete CG as pre-processing

1. **Discrete:** partition into *gradient basins*



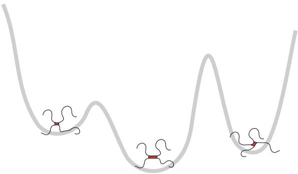
2. **Continous:** dissolve 'shallow' basins; distribute proportional to **rate** (or equivalently, transition state weight)



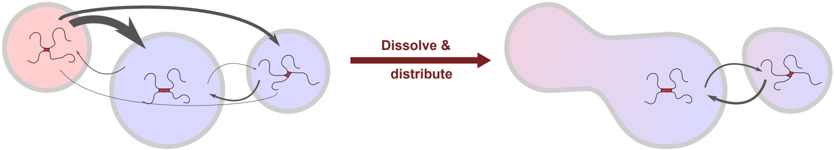
Further coarse graining

Continous CG with 'traditional' discrete CG as pre-processing

1. **Discrete:** partition into *gradient basins*



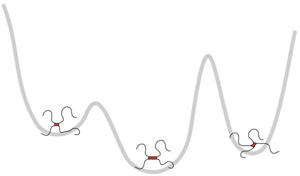
2. **Continous:** dissolve 'shallow' basins; distribute proportional to **rate** (or equivalently, transition state weight)



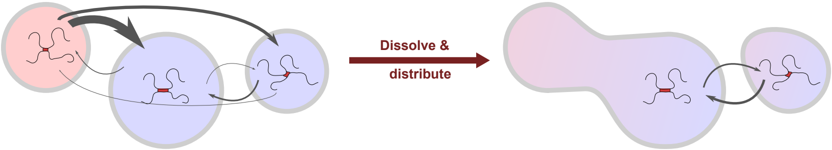
Further coarse graining

Continous CG with 'traditional' discrete CG as pre-processing

1. **Discrete:** partition into *gradient basins*



2. **Continous:** dissolve 'shallow' basins; distribute proportional to **rate** (or equivalently, transition state weight)

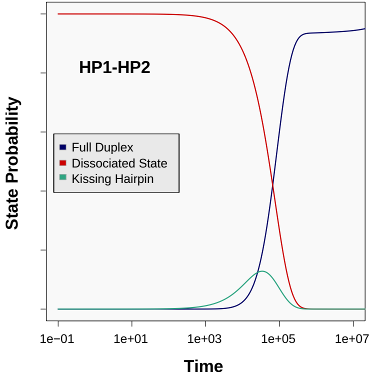


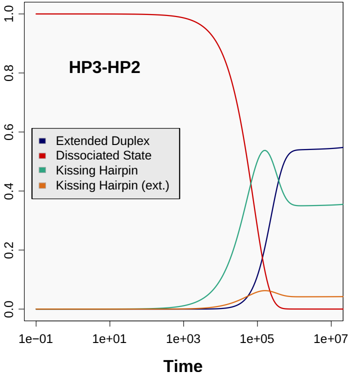
Both steps fit the concept of **general coarse graining** w/ corresp. matrices $\mathbf{C}^{gb}$ and $\mathbf{C}^{c}$; landscapes $\mathcal{L}^{gb}$ and $\mathcal{L}^{c}$

Quick validation w/ FRET experiments

- kinetic mutation study: Salim et al., Biophys J, 2012
- **3 RNA fragments**: **HP1**, **HP2** from DsrA–rpoS; **HP3** = HP1-variant
- **HP1–HP2** can form kissing hairpin *and* full duplex
- **HP3–HP2** *cannot* form complete duplex

side remark: reproducing the results works only at *correct temperature*





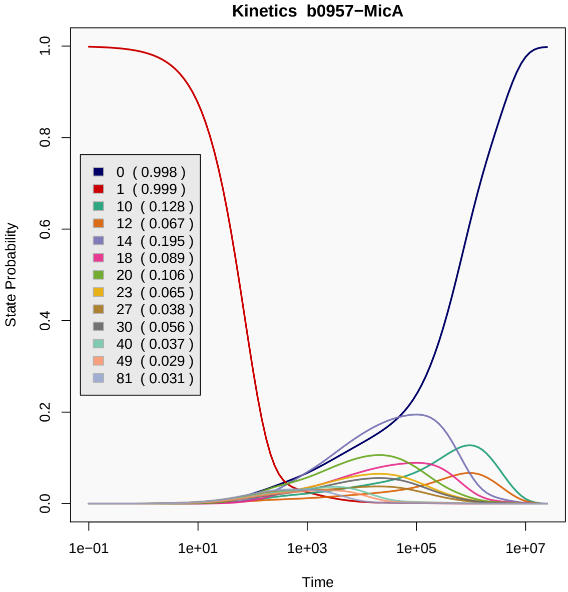
Example: E. coli MicA–ompA

```
>Mic A
GAAAGACGCGCAUUUGUUAUCAUCAUCCCUGAAUUCAGAGAUGAAAUUUUGGCCACUCACGAGUGGCCUUUU
>ompA 5'UTR
CUUUUUUUUCAUAUGCCUGACGGAGUUCACACUUGUAAGUUUUCAACUACGUUGUAGACUUUACAUCGCCAG
GGGUGCUCGGCAUAAGCCGAAGAUAUCGGUAGAGUUAAUAUUGAGCAGAUCCCCCGGUGAAGGAUUUAACCG
UGUUAUCUCGUUGGAGAUAUUCAUGGCGUAUUUUGGAUGAUAACGAGGCGCAAAAAAUGAAAAAGACAGCUA
UCGCGAUUGCAGUGGCACUGGCUGGUUUCGCUACCGUAGCGCAGGCCGCUCCGAAAGAUAACACCUGGUACA
CUGGUGCUAAAC
```

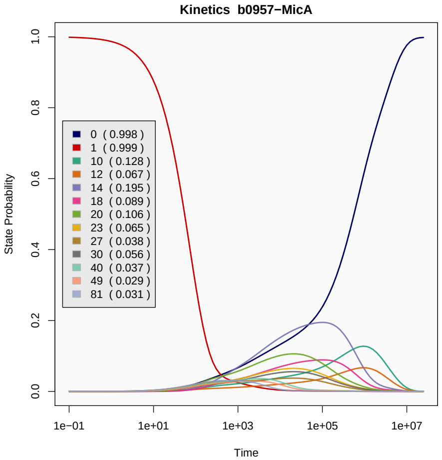
- **Enumerate** hybridization site states — 416992 states
- **discrete coarse graining** — 40772 gradient basins
- **continuous coarse graining** — 255 continuous macro states
- **Solve** macro-transition system

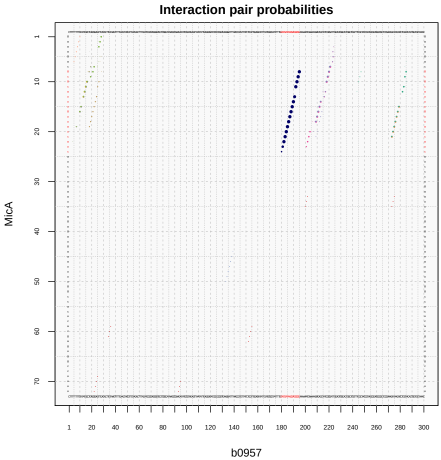
Total computation time: *several minutes*

Interpretation of results (MicA–ompA)



Interpretation of results (MicA–ompA)



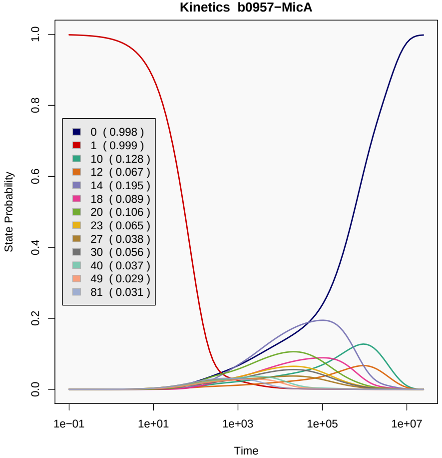


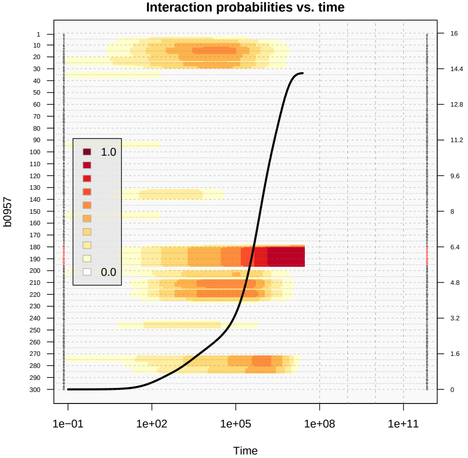
Probabiltities of interaction base pairs:

$$Pr[(i,j) \mid t] = \sum_{\substack{x^{\mathrm{c}} \in \mathcal{X}^{\mathrm{c}}, x^{\mathrm{gb}} \in \mathcal{X}^{\mathrm{gb}} \\ x^{\mathrm{h}} \in \mathcal{X}^{\mathrm{h}}, x^{\mathrm{j}} : (i,j) \in x^{\mathrm{j}}}} Pr[\alpha \mid t] \cdot \mathbf{C}^{\mathrm{c}}_{x^{\mathrm{c}} x^{\mathrm{gb}}} \cdot \mathbf{C}^{\mathrm{gb}}_{x^{\mathrm{gb}} x^{\mathrm{h}}} \cdot \mathbf{C}^{\mathrm{h}}_{x^{\mathrm{h}} x^{\mathrm{j}}}$$

Interaction probability of mRNA positions: $Pr[i \mid t] = \sum_{\text{sRNA pos. } j} Pr[(i,j) \mid t]$

Interpretation of results (MicA–ompA)





Probabiltities of interaction base pairs:

$$Pr[(i,j) \mid t] = \sum_{\substack{x^{c}\in\mathcal{X}^{c},x^{gb}\in\mathcal{X}^{gb} \\ x^{h}\in\mathcal{X}^{h},x^{j}:(i,j)\in x^{j}}} Pr[\alpha \mid t] \cdot \mathbf{C}^{c}_{x^{c}x^{gb}} \cdot \mathbf{C}^{gb}_{x^{gb}x^{h}} \cdot \mathbf{C}^{h}_{x^{h}x^{j}}$$

Interaction probability of mRNA positions: $Pr[i \mid t] = \sum_{\text{sRNA pos. } j} Pr[(i,j) \mid t]$

Take home

- **Kinetics model** for fairly complex RNA–RNA interaction
- Kinetic analysis of sRNA–mRNA 5'UTR interaction "in minutes"
- Tailored coarse graining enables **feasible computation**
- Procedure for **continuous coarse graining**
- Generalized coarse graining offers **unified perspective**: from discrete CG (e.g. gradient basin) to continuous CG (e.g. Stadler&Stadler, 2010)
- ... and allows **'back propagation'**
- Interpretation at **base pair resolution**
- **[WIP]** software will be made availabe