

# Modeling RNA-RNA interaction formation on direct paths

Maria Waldl

Department of Theoretical Chemistry, University of Vienna  
Währingerstrasse 17, 1090 Vienna, Austria  
`maria@tbi.univie.ac.at`

Bled, February 13th, 2023

# Kinetic features of fixed candidate structures

# Kinetic features of fixed candidate structures

## kinetic models

- ▶ explore full structure space
- ▶ coarse graining
- ▶ computationally costly

## thermodynamic models

- ▶ established tools
- ▶ efficient
- ▶ no kinetics

# Kinetic features of fixed candidate structures

## kinetic models

- ▶ explore full structure space
- ▶ coarse graining
- ▶ computationally costly

## thermodynamic models

- ▶ established tools
- ▶ efficient
- ▶ no kinetics

⇒ thermodynamic structure candidate

Kinetically favorable path to forming this structure?

# Kinetic features of fixed candidate structures

## kinetic models

- ▶ explore full structure space
- ▶ coarse graining
- ▶ computationally costly

## thermodynamic models

- ▶ established tools
- ▶ efficient
- ▶ no kinetics

⇒ thermodynamic structure candidate

Kinetically favorable path to forming this structure?

biological relevant vs. non-functional interactions

# Kinetic features of fixed candidate structures

## kinetic models

- ▶ explore full structure space
- ▶ coarse graining
- ▶ computationally costly

## thermodynamic models

- ▶ established tools
- ▶ efficient
- ▶ no kinetics

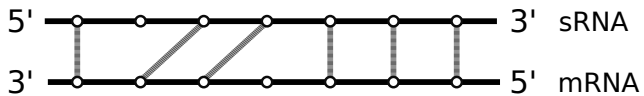
⇒ thermodynamic structure candidate

Kinetically favorable path to forming this structure?

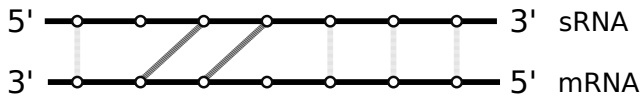
biological relevant vs. non-functional interactions

kinetic features for improve interaction predictions

## Interaction formation along direct paths

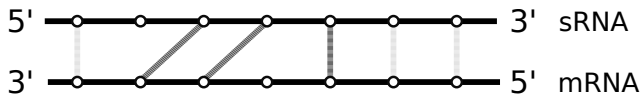


## Interaction formation along direct paths

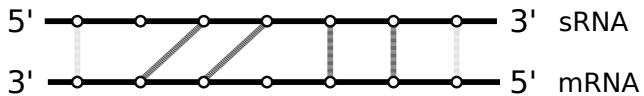




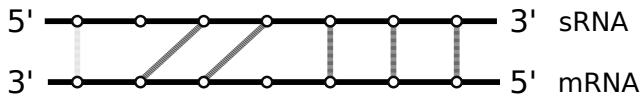
## Interaction formation along direct paths



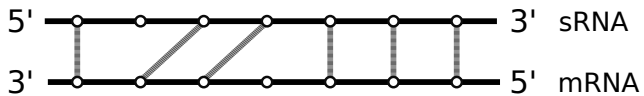
## Interaction formation along direct paths



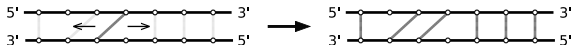
## Interaction formation along direct paths



## Interaction formation along direct paths

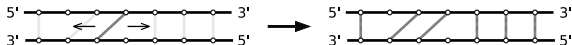


# Interaction formation along direct paths



⇒ structures on direct paths

# Interaction formation along direct paths



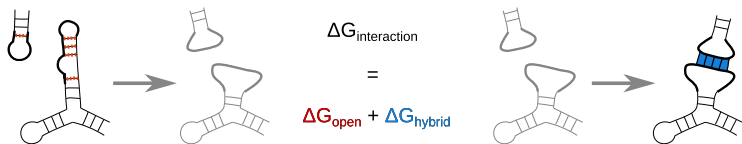
⇒ structures on direct paths

- ▶ fixed set of base pairs
- ▶ only consecutive substructures
- ▶ move set: base pair opening and closing

# Free energy of interaction structures



# Free energy of interaction structures



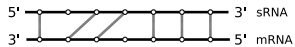
## intramolecular folding

fast (instant)  
unpaired probability  
of interaction sites

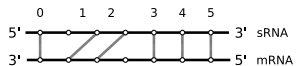
slow (minimal)  
removal of conflict base pairs



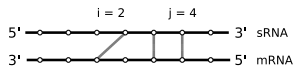
# Energy landscape



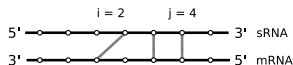
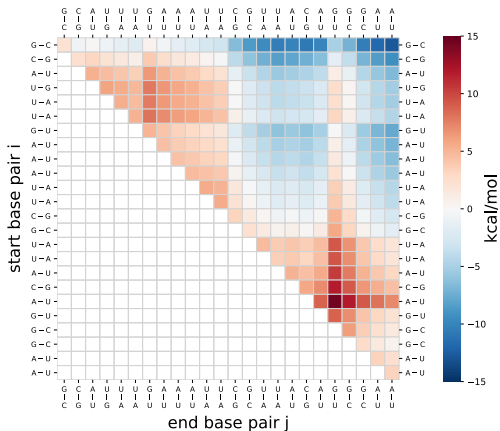
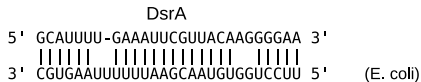
# Energy landscape



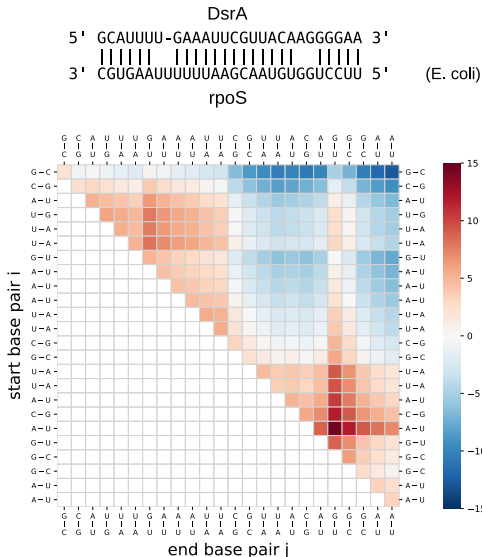
# Energy landscape



# Energy landscape



# Energy landscape



- ▶ seed stability
- ▶ seed accessibility
- ▶ local minima
- ▶ barriers on paths
- ▶ ...

## DP algorithm for minimal folding barriers

$$B_s(i, j) = \min \begin{cases} \max \begin{cases} E(i, j) \\ B_s(i + 1, j) \end{cases} & \text{if } i < s \\ \max \begin{cases} E(i, j) \\ B_s(i, j - 1) \end{cases} & \text{if } j > s \\ E(s, s) & \text{if } s = i = j \end{cases}$$

$$E(i, j) = E^{\text{hybrid}}(i, j) + E^{\text{unpaired}}(i, j)$$

$$E^{\text{hybrid}}(i, j) = \min \begin{cases} E^{\text{hybrid}}(i + 1, j) + L(i, i + 1) & \text{if } i < j \\ E_{\text{init}} & \text{if } i = j \end{cases}$$

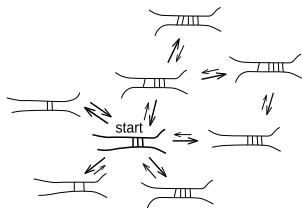
## Detailed kinetics from master equation

$$\frac{P_i(t)}{dt} = \sum_{i \neq j} (P_j(t)k_{ji} - P_i(t)k_{ij})$$

- ▶ set of structures  
⇒ states  $\Omega$
- ▶ move set  $M$   
⇒ neighborhood relation
- ▶ energy function  $E$
- ▶ free energy difference  $\Delta G^\ddagger$   
⇒ folding rate  $k_{ij}$

$$k_{ij} = \begin{cases} k_0 & \text{if } \Delta G^\ddagger \leq 0, \\ k_0 e^{\frac{-\Delta G^\ddagger}{RT}} & \text{otherwise} \end{cases}$$

## Detailed kinetics from master equation



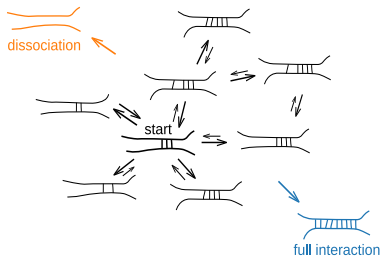
$$\frac{P_i(t)}{dt} = \sum_{i \neq j} (P_j(t)k_{ji} - P_i(t)k_{ij})$$

- ▶ set of structures  
⇒ states  $\Omega$
- ▶ move set  $M$   
⇒ neighborhood relation
- ▶ energy function  $E$
- ▶ free energy difference  $\Delta G^\ddagger$   
⇒ folding rate  $k_{ij}$

$$k_{ij} = \begin{cases} k_0 & \text{if } \Delta G^\ddagger \leq 0, \\ k_0 e^{\frac{-\Delta G^\ddagger}{RT}} & \text{otherwise} \end{cases}$$



# Detailed kinetics from master equation

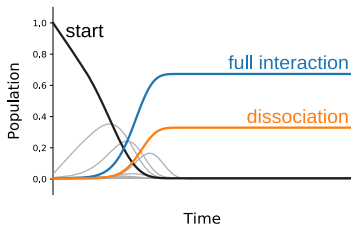
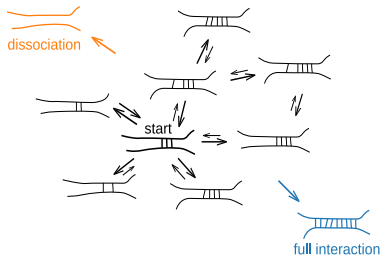


$$\frac{P_i(t)}{dt} = \sum_{i \neq j} (P_j(t)k_{ji} - P_i(t)k_{ij})$$

- ▶ set of structures  
⇒ states  $\Omega$
- ▶ move set  $M$   
⇒ neighborhood relation
- ▶ energy function  $E$
- ▶ free energy difference  $\Delta G^\ddagger$   
⇒ folding rate  $k_{ij}$

$$k_{ij} = \begin{cases} k_0 & \text{if } \Delta G^\ddagger \leq 0, \\ k_0 e^{\frac{-\Delta G^\ddagger}{RT}} & \text{otherwise} \end{cases}$$

# Detailed kinetics from master equation

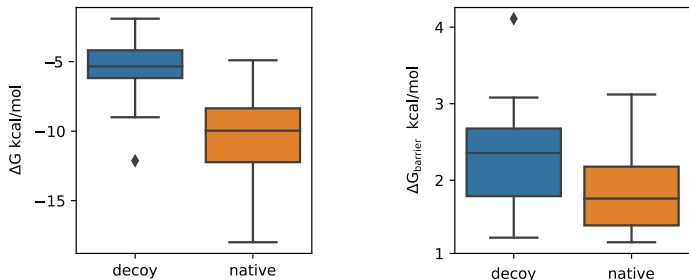


$$\frac{P_i(t)}{dt} = \sum_{i \neq j} (P_j(t)k_{ji} - P_i(t)k_{ij})$$

- ▶ set of structures  
⇒ states  $\Omega$
- ▶ move set  $M$   
⇒ neighborhood relation
- ▶ energy function  $E$
- ▶ free energy difference  $\Delta G^\ddagger$   
⇒ folding rate  $k_{ij}$

$$k_{ij} = \begin{cases} k_0 & \text{if } \Delta G^\ddagger \leq 0, \\ k_0 e^{-\frac{\Delta G^\ddagger}{RT}} & \text{otherwise} \end{cases}$$

# Do kinetic features provide new information?



(72 sRNA + mRNA pairs from *E. coli* )

# Increased prediction performance with kinetics

E. coli data set:

- ▶ native sRNA + native mRNA (104 pairs)
- ▶ native sRNA + 4 x shuffled mRNA (420 pairs)

⇒ thermodynamic interaction prediction

⇒ compute kinetic features

⇒ train ML classifier

# Increased prediction performance with kinetics

E. coli data set:

- ▶ native sRNA + native mRNA (104 pairs)
- ▶ native sRNA + 4 x shuffled mRNA (420 pairs)

⇒ thermodynamic interaction prediction

⇒ compute kinetic features

⇒ train ML classifier

linear discriminator, 10 fold validation

no overlap in sRNAs between test and training set

features	MCC	recall	precision
$E_{\text{therm}}$	0.53	0.50	0.76
$E_{\text{therm}}$ + seed	0.60	0.52	0.86
$E_{\text{therm}}$ + detailed dynamics	0.61	0.58	0.81
$E_{\text{therm}}$ + seed + detailed dynamics	0.65	0.65	0.84
$E_{\text{therm}}$ + seed + detailed dynamics + barrier	0.71	0.67	0.87

## Summary

- ▶ **computational model** for interaction formation
- ▶ **efficient computation** methods for features
- ▶ **benchmark** prediction capabilities of features
- ▶ test **mechanistic** hypothesises

⇒ stable seed interaction

⇒ fast formation of full interaction

## Outlook

- ▶ easy to use prediction tool

In collaboration with ...

## Vienna

- ▶ Ivo L. Hofacker
- ▶ Irene K. Beckmann
- ▶ and the TBI Team

## Paris

- ▶ Sebastian Will

## Freiburg

- ▶ Rolf Backofen
- ▶ Martin Raden

Funding: FWF (I-2874-N28, DK RNA Biology, F 80 RNAdeco)



universität  
wien

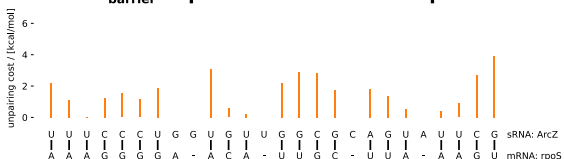


FWF

# seed accessibility and folding barrier example



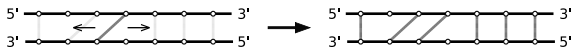
**$E_{\text{barrier}}$  dependent on start base pair**



**Cost for keeping the pairing positions of the interaction unpaired in the intramolecular structure**



# Dsra mechanism



## Example energy landscape (E. coli: DsrA rpoS)

Free energy  $\Delta G$  in kcal/mol for each interaction substructure from base pair  $i$  to base pair  $j$

