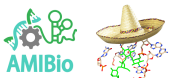
RNA *in silico* design with Infrared framework

Hua-Ting Yao

Theoretical Biochemistry Group (TBI), University of Vienna

Vienna, Austria

tbi



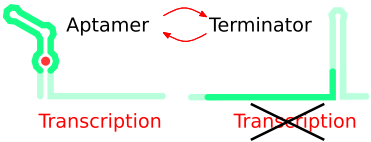
RNA (structural) design

- Finding RNA sequences with desired function

  → RNA vaccines, RNA drugs, therapies based on RNA . . .

- Structure ↔ Function

  → Off Riboswitch



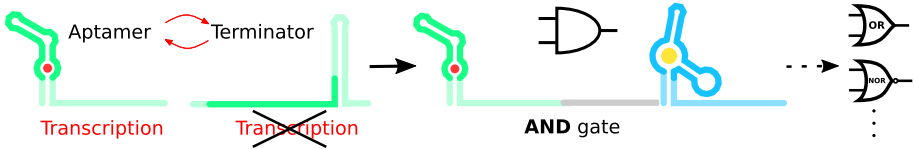
RNA (structural) design

- Finding RNA sequences with desired function

  → RNA vaccines, RNA drugs, therapies based on RNA . . .

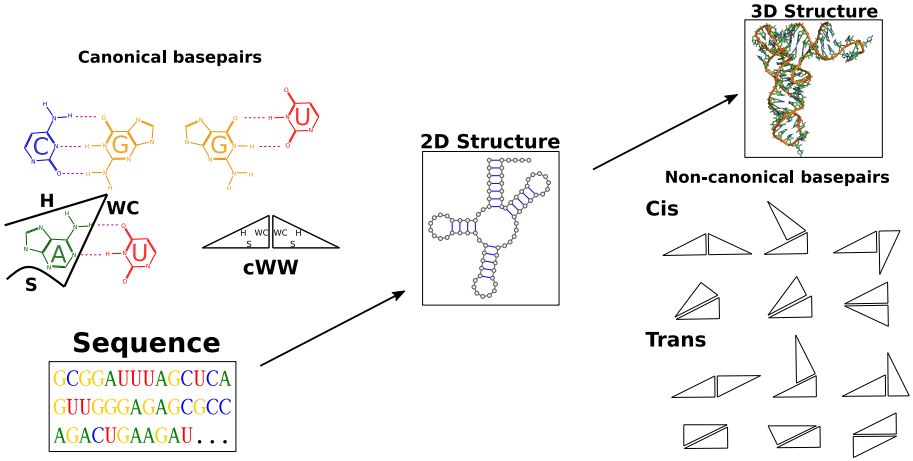
- Structure ↔ Function

  → Off Riboswitch → Riboswitch **AND** (G. Domin *et al.*, 2017)

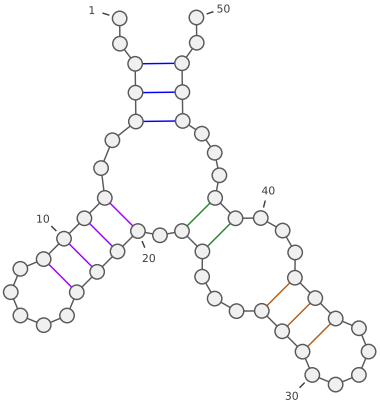


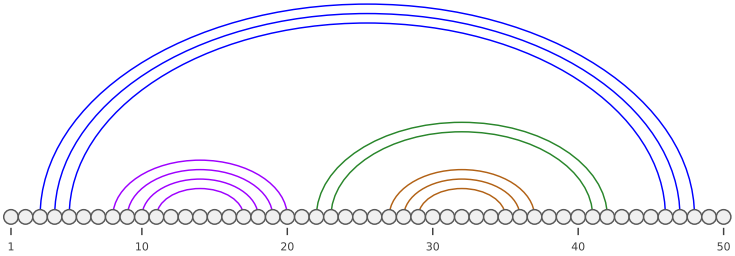
Outline

1. Brief introduction to RNA structural bioinformatics
2. Infrared in RNA design
3. RNAPONDD: a secondary structure design approach
4. Beyond secondary structure 1: small RNA - mRNA interaction
5. Beyond secondary structure 2: exonuclease-resistant RNA (xrRNA)


Canonical basepairs
H
WC
S
cWW
Sequence
GCGGAUUUAGCUCA
GUUGGGAGAGCGCC
AGACUGAAGAU...
2D Structure
3D Structure
Non-canonical basepairs
Cis
Trans

Representations of RNA structures

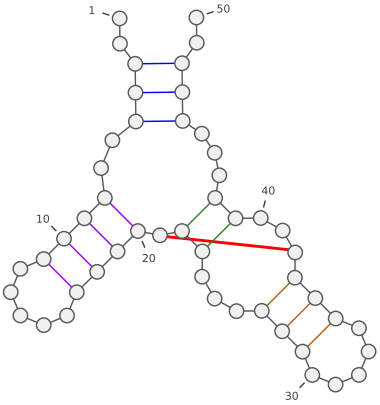


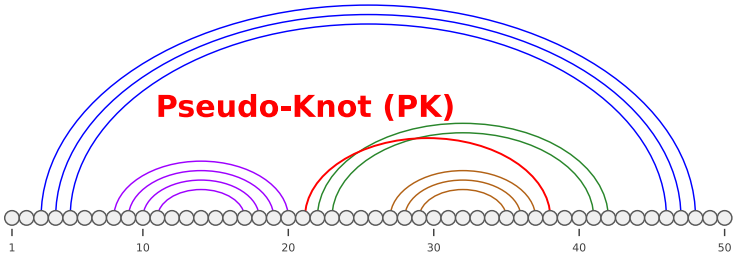


••(((••((((•••••))))•((•••(((•••••)))•••))•••)))••

Motzkin words $\rightarrow$ $\sim \frac{3^n}{n\sqrt{n}}$ secondary structures

Representations of RNA structures

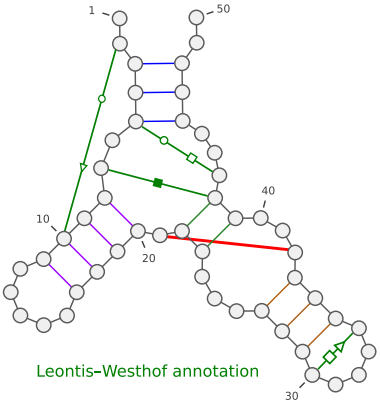




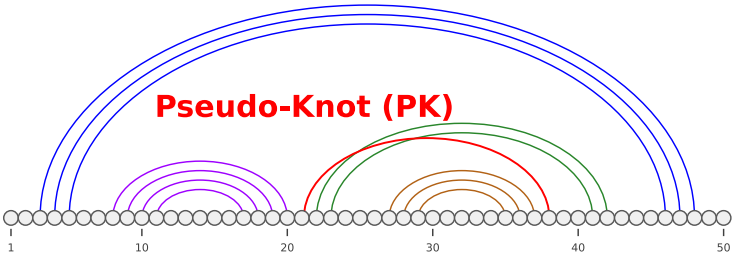
••(((••((((•••••))))[(((•••(((•••••)))]••)))•••)))••

Motzkin words $\rightarrow$ $\sim \frac{3^n}{n\sqrt{n}}$ secondary structures

Representations of RNA structures



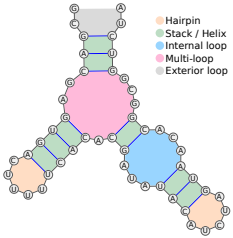
Leontis–Westhof annotation



••(((••((((•••••))))[(((•••(((•••••)))]••)))•••)))••

Motzkin words $\rightarrow$ $\sim \frac{3^n}{n\sqrt{n}}$ secondary structures

Thermodynamic model



- `ViennaRNA (1994 - )`

- Free energy $\mathcal{E}$

$$\mathcal{E}: \quad \Sigma^* \times \mathcal{S} \to \mathbb{R}, \quad (w, S) \mapsto \mathcal{E}(w, S)$$

w: sequence, S: structure

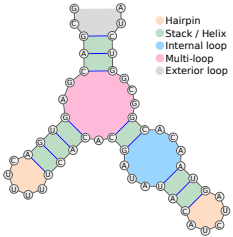
$$\mathcal{E}(w,S) = \Delta\Big(\quad\Big) + \Delta\Big(\quad\Big) + \Delta\Big(\quad\Big) + \cdots$$

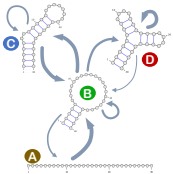
- Optimal conformation (Minimum Free-Energy, MFE)

  $\to$ secondary: $\mathcal{O}(n^3)$

  $\to$ pseudo-knotted: NP-hard

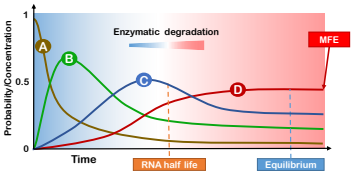
Thermodynamic model





A – Kinetic Landscape

Continuous-time Markov chain



B – Evolution of concentrations

(stolen from Y. Ponty)

- `ViennaRNA (1994 - )`

- Free energy $\mathcal{E}$

$$\mathcal{E}: \quad \Sigma^* \times \mathcal{S} \to \mathbb{R}, \quad (w, S) \mapsto \mathcal{E}(w, S)$$

w: sequence, S: structure

$$\mathcal{E}(w,S) = \Delta\Big(\quad \Big) + \Delta\Big(\quad \Big) + \Delta\Big(\quad \Big) + \cdots$$

- Optimal conformation (Minimum Free-Energy, MFE)

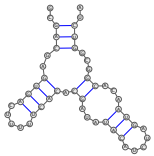
  $\rightarrow$ secondary: $\mathcal{O}(n^3)$

  $\rightarrow$ pseudo-knotted: NP-hard

- Boltzmann equilibrium, kinetic, ...

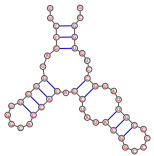
- Positive Design: Compatibility with one or few target structure(s)
  → optimize affinity towards given target(s) *i.e.* free-energy

w_1



-4.3

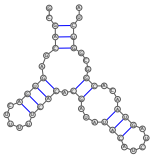
w_2



-6.7

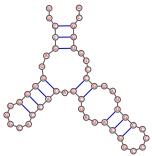
- Positive Design: Compatibility with one or few target structure(s)
  → optimize affinity towards given target(s) *i.e.* free-energy
  - Constraints: basepair compatibility, sequence pattern . . .
  - (Weight) Functions: free-energy, GC-content . . .
  - Sampling

w_1



-4.3

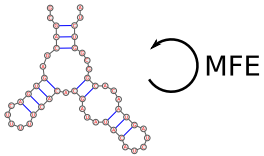
w_2



-6.7

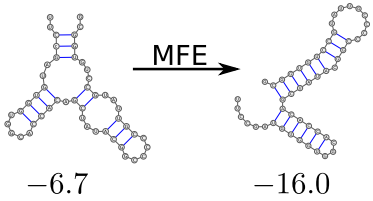
- Positive Design: Compatibility with one or few target structure(s)
  → optimize affinity towards given target(s) *i.e.* free-energy
  - Constraints: basepair compatibility, sequence pattern . . .
  - (Weight) Functions: free-energy, GC-content . . .
  - Sampling

w_1



- Negative Design: Avoidance of (exponential) unwanted structures
  → specificity towards given targets *i.e.* Minimum Free Energy (MFE)
  - Objective function: MFE, structure ensemble . . .
  - Optimization

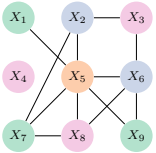
w_2



Infrared: Graph coloring as weighted CSP

Constraint satisfaction problem

- Variables $\mathcal{X} = \{X_1, \ldots, X_9\}$
- Domains $\mathcal{D} = \{D_1, \ldots, D_9\}$ with $D_i = \{1, 2, 3, 4\}$
- Constraints $\mathcal{C} = \{\text{Diff}(X_i, X_j) \mid \text{ edge } (i, j)\}$

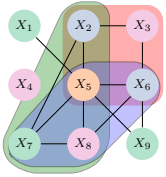


[H-T. Yao *et al.*, ALMOB 2024]

Infrared: Graph coloring as weighted CSP

Constraint satisfaction problem (weighted)

- Variables $\mathcal{X} = \{X_1, \ldots, X_9\}$
- Domains $\mathcal{D} = \{D_1, \ldots, D_9\}$ with $D_i = \{1, 2, 3, 4\}$
- Constraints $\mathcal{C} = \{\text{Diff}(X_i, X_j) \mid \text{edge } (i, j)\}$
- Functions $\mathcal{F} = \{\text{Card}, \text{Card}, \text{Card}\}$



3+4+4 = 11

Problems:

1. Optimization:

$$x^* = \underset{\text{valid } x}{\operatorname{argmax}} \sum_{F \in \mathcal{F}} \alpha_F F(x)$$

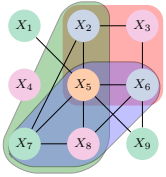
[H-T. Yao *et al.*, ALMOB 2024]

```
import infrared as ir

model = ir.Model(number=9,
                 domain=[1,2,3,4],
                 name='X')

model.add_constraints(Diff(i,j)
    for i, j in edges)

model.add_functions([Card(i, j, k, l)
    for i, j, k, l in cycles])

sol = ir.Optimizer(model).optimize()
```

Infrared: Graph coloring as weighted CSP

Constraint satisfaction problem (weighted)

- Variables $\mathcal{X} = \{X_1, \ldots, X_9\}$
- Domains $\mathcal{D} = \{D_1, \ldots, D_9\}$ with $D_i = \{1, 2, 3, 4\}$
- Constraints $\mathcal{C} = \{\mathsf{Diff}(X_i, X_j) \mid \text{edge } (i, j)\}$
- Functions $\mathcal{F} = \{\mathsf{Card}, \mathsf{Card}, \mathsf{Card}\}$



3+4+4 = 11

Problems:

1. Optimization:

$$x^* = \underset{\text{valid } x}{\operatorname{argmax}} \sum_{F \in \mathcal{F}} \alpha_F F(x)$$

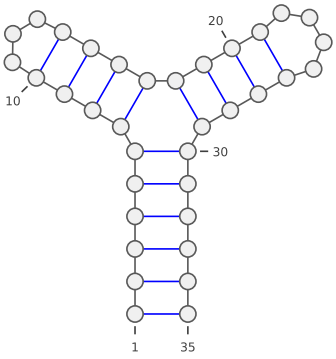
2. (Boltzmann) Sampling:

$$\mathbb{P}(x) \propto \exp\left(\sum_{F \in \mathcal{F}} \alpha_F F(x) \right)$$

[H-T. Yao *et al.*, ALMOB 2024]

```python
import infrared as ir

model = ir.Model(number=9,
                 domain=[1,2,3,4],
                 name='X')

model.add_constraints(Diff(i,j)
    for i, j in edges)

model.add_functions([Card(i, j, k, l)
    for i, j, k, l in cycles])

sol = ir.Optimizer(model).optimize()


sample = ir.Sampler(model).sample()
```

Single structure design



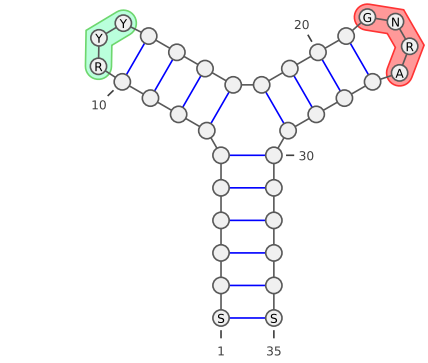

```
import infrared as ir
import infrared.rna as rna

model = ir.Model(35, 4)

target = "((((((((((...))))))((((....))))))))))"
model.add_constraints(rna.BPComp(i, j)
    for (i, j) in rna.parse(target))

sampler = ir.Sampler(model)
samples = [sampler.sample() for _ in range(1000)]
```

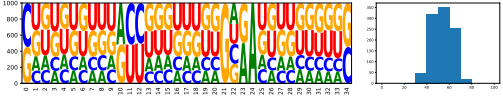
0 : A 1 : C 2 : G 3 : U

AU, CG, · · ·

[H-T. Yao *et al.*, RNA Folding, 2024 (Book chapter)]

Single structure design



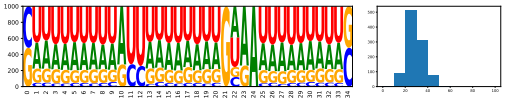


[H-T. Yao *et al.*, RNA Folding, 2024 (Book chapter)]

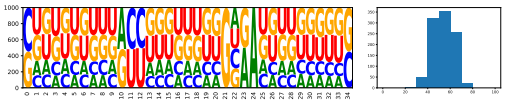
```
import infrared as ir
import infrared.rna as rna

model = ir.Model(35, 4)          # 0:A 1:C 2:G 3:U

target = "((((((((((...))))))(((....)))))))))"
model.add_constraints(rna.BPComp(i, j)
    for (i, j) in rna.parse(target))   # AU, CG, ···

# N:ACGU  S:CG  R:AG  Y:CU
iupac_seq = "SNNNNNNNNRYYNNNNNNNNGNRANNNNNNNNS"
for i, x in enumerate(iupac_seq):
    model.add_constraints(
        ir.ValueIn(i, rna.iupacvalues(x)))


sampler = ir.Sampler(model)
samples = [sampler.sample() for _ in range(1000)]
```

Control GC

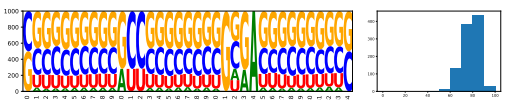
$\alpha = -1$



$\alpha = 0$



$\alpha = +1$

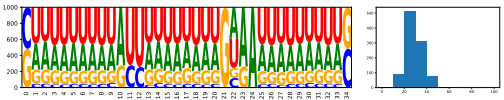


Method 1:

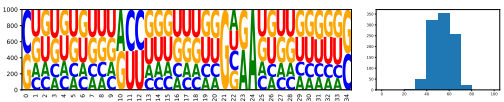
```
model.add_functions([rna.GCCont(i))     CG : 1  AU : 0
    for i in range(n)], 'gc')
model.set_feature_weight(α, 'gc')

sampler = ir.Sampler(model)
samples = [sampler.sample() for _ in range(1000)]
```

Control GC

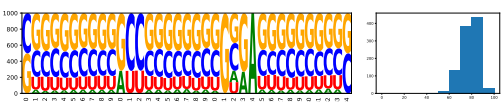
$\alpha = -1$



$\alpha = 0$



$\alpha = +1$



Method 1:

CG : 1 AU : 0

```
model.add_functions([rna.GCCont(i))
    for i in range(n)], 'gc')
model.set_feature_weight(α, 'gc')

sampler = ir.Sampler(model)
samples = [sampler.sample() for _ in range(1000)]
```

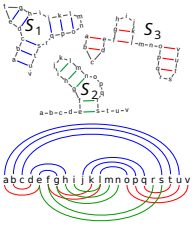
Method 2 (Targeted sampling):

```
sampler = ir.Sampler(model)
sampler.set_target( 0.75 * n, 0.01 * n, 'gc' )
samples = [sampler.targeted_sample()
    for _ in range(1000)]
```

Automatically update α

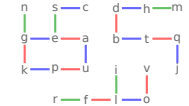
Weighted Constraint Satisfaction Problem (CSP)

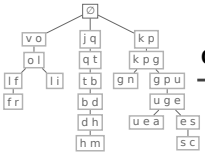
$(\mathcal{X}, \mathcal{D}, \mathcal{C}, \mathcal{F})$



Constraints $\mathcal{C}$: S_1, S_2, S_3

Functions $\mathcal{F}$: GC%, E_1, E_2, E_3

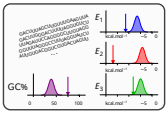




Tree decomposition

Treewidth w

Sampling



Optimization

GCCGCGGUAGCU · · ·

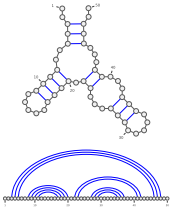
Time Complexity

$\mathcal{O}(d^{w+1} \cdot (|\mathcal{X}| + |\mathcal{C}| + |\mathcal{F}|))$

Optimization

Objective function

Positive Design



Target structure

Sampling

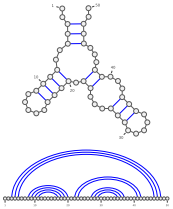
```
CGCCCUGCGG ... UUCUCGGGAC
GUUCGCAUUG ... GCUGUCGGAU
CCGGGAGGCG ... CCUUUCCCGC
CAGGCAGACC ... CGAGUGCCUU
UUCCCCGUUG ... GGGCUGGGCC
AUCGGGCUUU ... GCCGUCUGGU
UUCCAGAGCC ... CGGGCUGGUA
CACGGGGCGG ... GUUUCCCGUG
AAGGCCGGCC ... GGGGGGCCUA
UUGGGGCGCC ... ACGGCCCCAC
```

Optimal folding

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

Disruptive basepairs

RNAPOND: Motivation



Target structure

Sampling

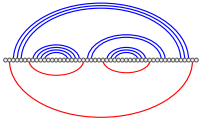
```
CGCCCUGCGG ... UUCUCGGGAC
GUUCGCAUUG ... GCUGUCGGAU
CCGGGAGGCG ... CCUUUCCCGC
CAGGCAGACC ... CGAGUGCCUU
UUCCCCGUUG ... GGGCUGGGCC
AUCGGGCUUU ... GCCGUCUGGU
UUCCAGAGCC ... CGGGCUGGUA
CACGGGGCGG ... GUUUCCCGUG
AAGGCCGGCC ... GGGGGGCCUA
UUGGGGCGCC ... ACGGCCCCAC
```

Optimal folding

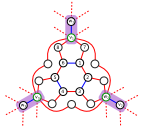
```
..(((..(((  ...  ))...)))..
(((((...))  ...  ..))))))..
....(((((  ...  ))))))....
.(((((((((  ...  )))))))).
.(((((...(  ...  )))))))).
.........(  ...  )))))))).
(((((((..(  ...  )..))))))
......(((  ...  .))))).....
(((((((((  ...  ..)))))))).
.(((((((((  ...  ....)))))).
.(((((((((  ...  .).)))))).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

Disruptive basepairs

- Forbid disruptive basepairs

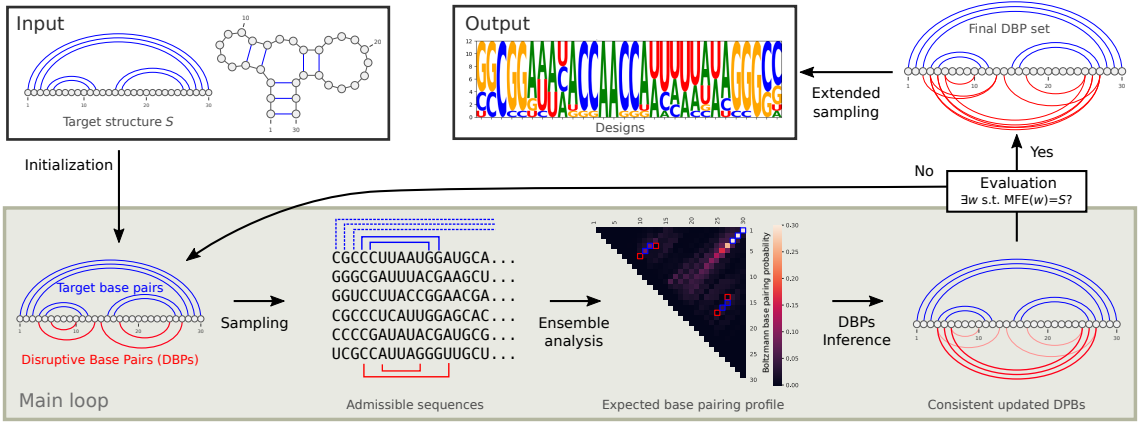


- Decision problem is NP-hard → Polynomial reduction from 3-SAT



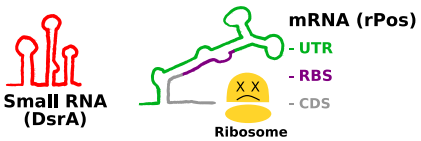
↔ clause

RNAPCND: Workflow



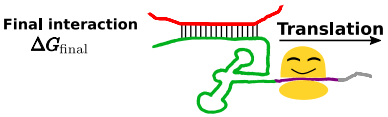
[H-T. Yao *et al.*, RECOMB'21]

Beyond 2D (I): small RNA - mRNA interaction

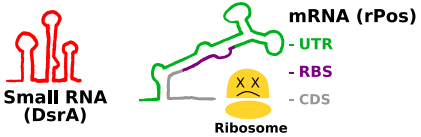


messenger RNA (mRNA):

1. **UTR**: Untranslated region
2. **RBS**: Ribosome binding site
3. **CDS: Coding sequence**



Beyond 2D (I): small RNA - mRNA interaction

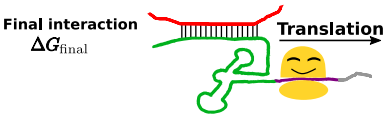


messenger RNA (mRNA):

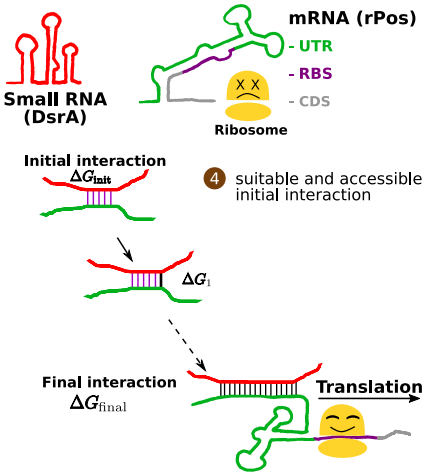
1. **UTR**: Untranslated region
2. **RBS**: Ribosome binding site
3. **CDS: Coding sequence**

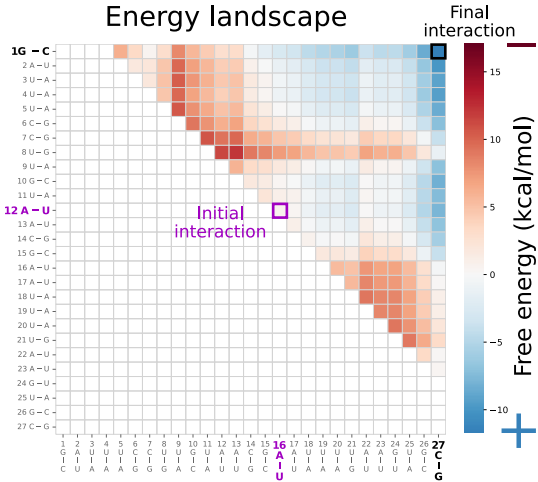
Design goals:

1. sRNA and mRNA should bind strongly
2. poor **RBS** accessibility w/o sRNA
3. in the bound state, **RBS** is highly accessible



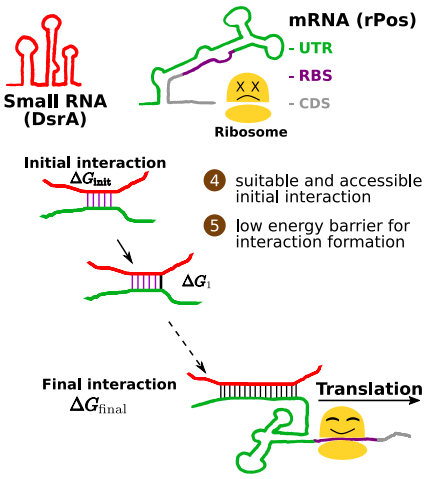
Beyond 2D (I): small RNA - mRNA interaction

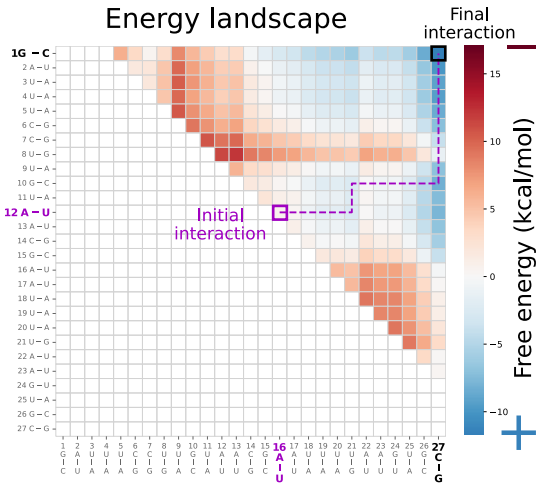




Computed with RRIkinDP (M. Waldl *et al*, 2024)

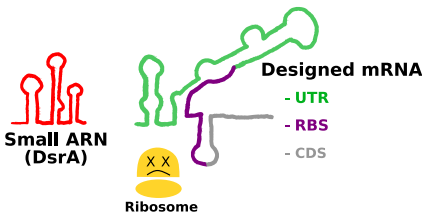
Beyond 2D (I): small RNA - mRNA interaction





Computed with RRIkinDP (M. Waldl *et al*, 2024)

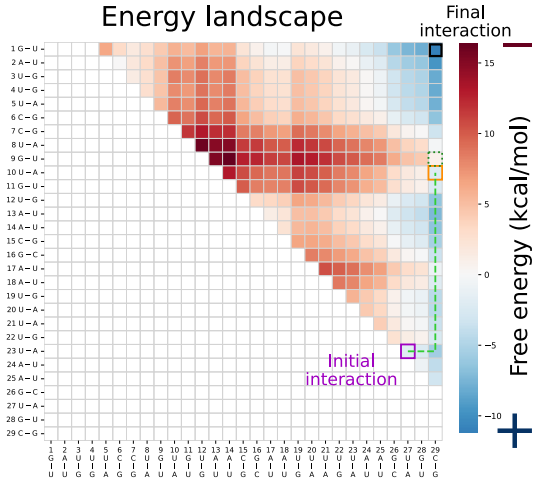
Beyond 2D (I): Design of interaction (equilibrium only)



Design w/o kinetic:

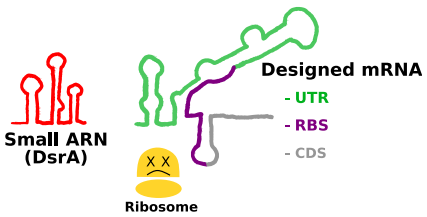
1. sRNA and mRNA should bind strongly $\approx -12.97\,\mathrm{kcal/mol}$
2. poor RBS accessibility w/o sRNA $\approx 12.29\,\mathrm{kcal/mol}$
3. in the bound state, RBS is highly accessible $\approx 5.01\,\mathrm{kcal/mol}$

V



[M. Waldl *et al.*, RNA Design, 2025 (Book chapter)]

Beyond 2D (I): Design of interaction with kinetic

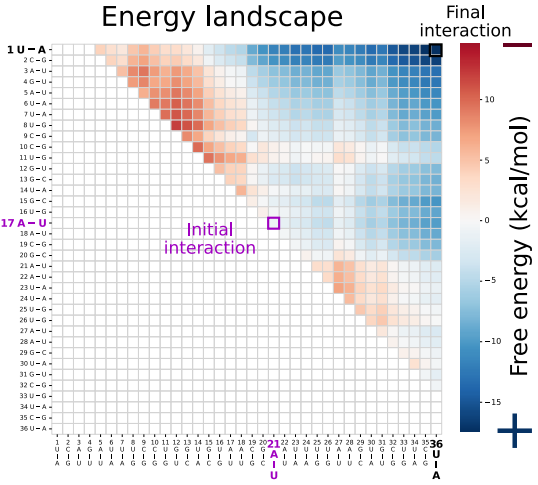


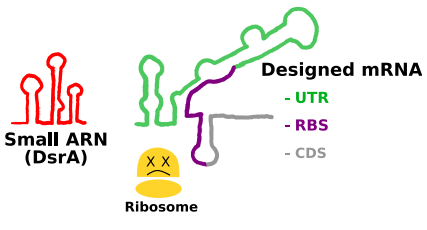
Design with kinetic:

1. sRNA and mRNA should bind strongly $\approx -17.00\,\mathrm{kcal/mol}$
2. poor RBS accessibility w/o sRNA $\approx 12.58\,\mathrm{kcal/mol}$
3. in the bound state, RBS is highly accessible $\approx 4.88\,\mathrm{kcal/mol}$

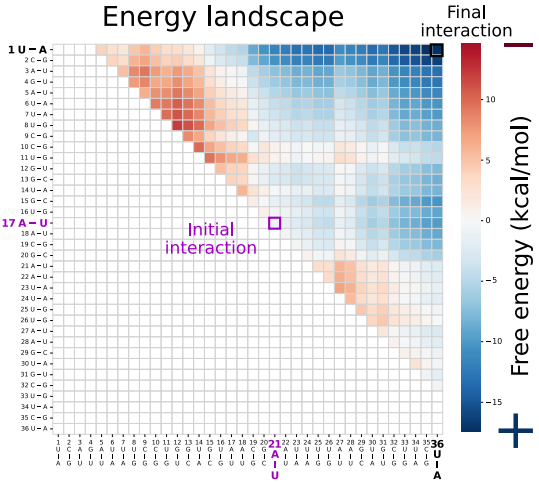
   V
4. suitable and accessible initial interaction
5. low energy barrier for interaction formation

[M. Waldl *et al.*, RNA Design, 2025 (Book chapter)]

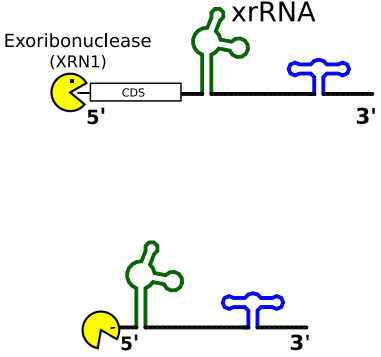


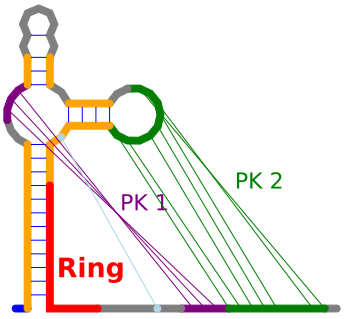


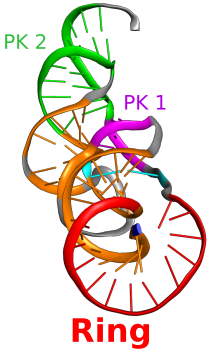
→System of 3 RNAs (molecules)



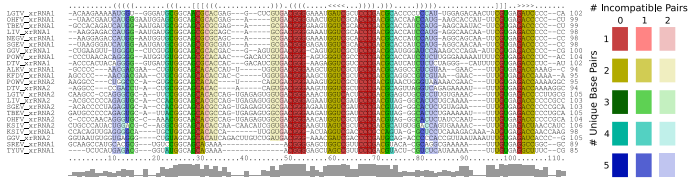
Beyond 2D (II): exonuclease-resistant RNA (xrRNA)

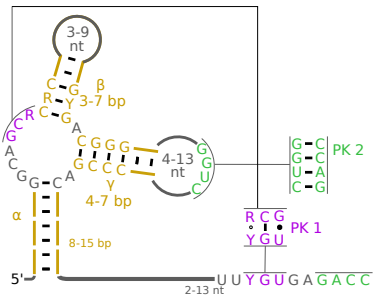


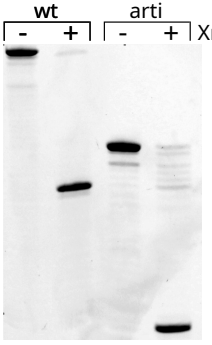




Beyond 2D (II): Design of xrRNA





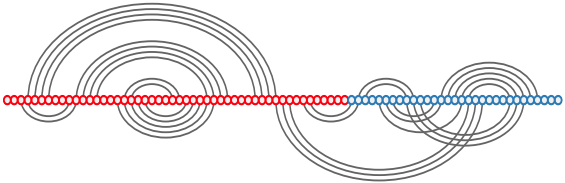


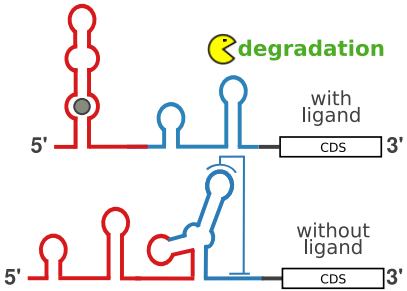
1. Sampling:
   - sequence conservation
   - varied length $\{A, C, G, U, -\}$
2. Optimization:
   - 2D: ensemble defect
   - (Future work) PK: CParty (L. Trinity *et al*, 2024)
3. Validation:
   - MD simulation
   - *in vitro* experiment (in collaboration with M. Mörl)

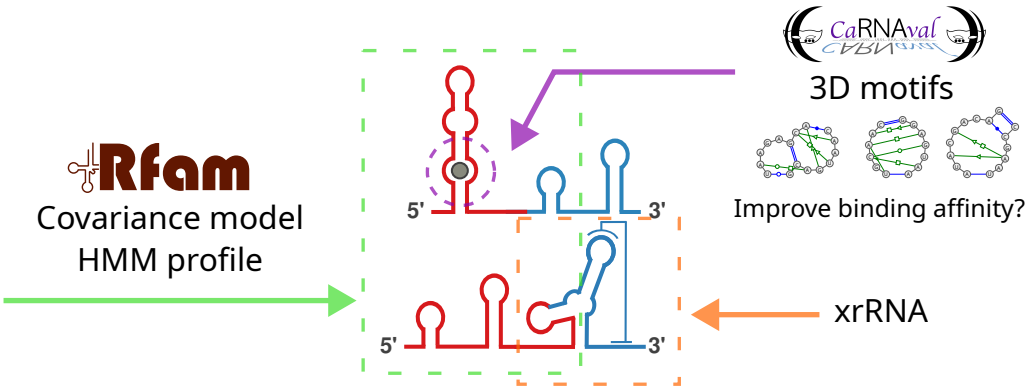
→ Toward *de novo* design

[Ongoing work with L. Sidl]

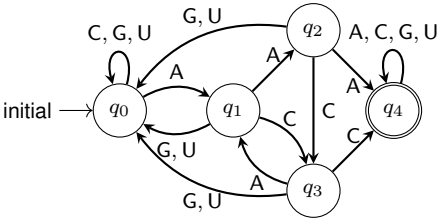
Beyond 2D (II): xrRNA-riboswitch






Rfam
Covariance model
HMM profile
5'
3'
5'
3'
CaRNAval
3D motifs
Improve binding affinity?
xrRNA
[Ongoing work with S. Gardelle]

Application of Infrared: Deterministic finite automaton



Sequence	States	
AGUACCGCU	0100134444	Accepted
AGUACAACU	0100131230	Rejected

Transition function $\delta : \mathcal{Q} \times \Sigma \to \mathcal{Q}$

Variables $\mathcal{X}$ $=$ $\{x_1, \ldots, x_n\}$ $\cup$ $\{y_0, y_1, \ldots, y_{n-1}, y_n\}$

Domains $\mathcal{D}$ $=$ $\Sigma \times \cdots \times \Sigma$ $\times$ $\{q_0\} \times \mathcal{Q} \times \cdots \times \mathcal{Q} \times \{q_4\}$

Transition constraint $\delta_{\text{const}}(x_i, y_i, y_{i-1}) = \begin{cases} \text{True} & \text{if } \delta(y_{i-1}, x_i) = y_i \\ \text{False} & \text{otherwise.} \end{cases}$

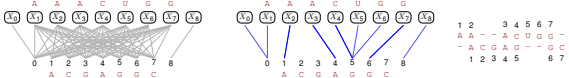
Potential applications in RNA design: insertion of protein binding site, control of repeated regions ...

Conclusion & Perspective

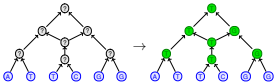
- Infrared: flexible and efficient framework for RNA design
- Design beyond secondary structure: PK, 3D, interaction
- Design with different conditions thanks to latest ViennaRNA supports
  $\rightarrow$ some modifications (Y. Varenyk *et al*, 2023), mono-valent salt concentration (H-T Yao *et al*, 2023)
- Architecture Improvement to support applications in RNA design, e.g. mRNA design

Infrared: Applications

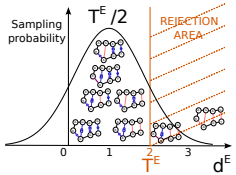
- Sequences (structure) alignment, $w \approx 3$



- Parsimony for phylogenetic reconstruction, $w \approx 4$

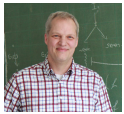


- Explore of 3D motif variations, $w \approx 3$ (T. Boury *et al.*, 2023)



Acknowledgment

R. Lorenz 

I. Hofacker 

P. Stadler 

Y. Ponty 

S. Will 

S. Berkemer 

M. Wolfinger 

L. Sidl 

J. Waldispühl 

B. Marchand 

M. Waldl 

S. Gardelle 

V. Reinharz