Incorporating RNA Modifications into ViennaRNA's Predictive Models

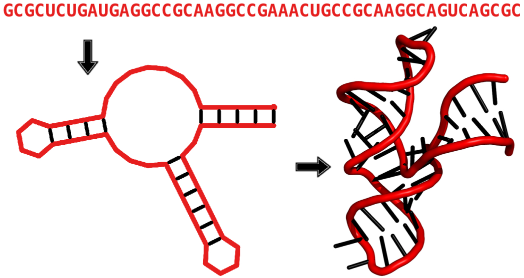
Michael T. Wolfinger[1,2,3]
Thomas Spicher[2]

[1] Research Group Bioinformatics and Computational Biology, University of Vienna, Austria
[2] Department of Theoretical Chemistry, University of Vienna, Austria
[3] RNA Forecast, Vienna, Austria

Hinxton, 2 April 2025

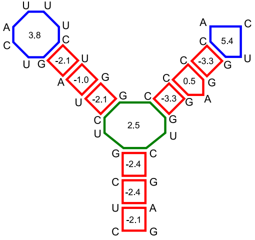
The RNA Folding Problem

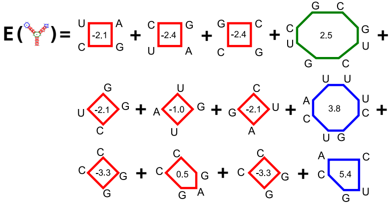
Hierarchical folding: Secondary structure forms first then helices arrange to form tertiary structure



- Secondary structure captures the majority of stabilizing
- Convenient and biologically useful description
- Efficient algorithms for many thermodynamic properties
- High prediction accuracy for small RNAs
- Good basis for 3D structure modeling and RNA folding kinetics

RNA Secondary Structures - Nearest Neighbor Energy Model





= -6.50 kcal/mol

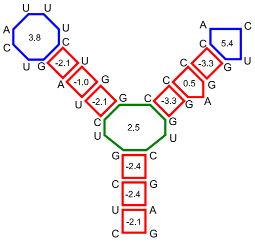
- Decomposition of Secondary structures s into loops L
- Contributions of a base pair depend on neighboring pairs
- Each loop L is assigned a free energy contribution E_L[1]

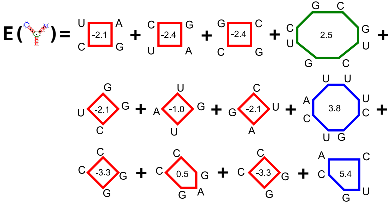
$$E(s) \approx \sum_{L \in s} E_L$$

---

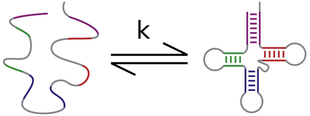
[1] Mittal et al., "NNDB: An Expanded Database of Nearest Neighbor Parameters for Predicting Stability of Nucleic Acid Secondary Structures", 2024, J. Mol. Biol., 168549

RNA Secondary Structures - Nearest Neighbor Energy Model





= -6.50 kcal/mol



- Decomposition of Secondary structures s into loops L
- Contributions of a base pair depend on neighboring pairs
- Each loop L is assigned a free energy contribution E_L[1]

$$E(s) \approx \sum_{L \in s} E_L$$

$$\Delta G = G(folded) - G(unfolded)$$

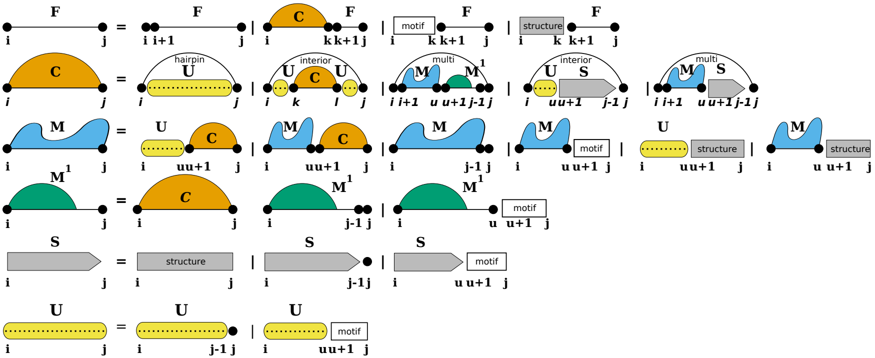
$$\Delta G = \Delta H - T\Delta S$$

[1] Mittal et al., "NNDB: An Expanded Database of Nearest Neighbor Parameters for Predicting Stability of Nucleic Acid Secondary Structures", 2024, J. Mol. Biol., 168549

RNA Secondary Structure Prediction

Nussinov decomposition scheme:

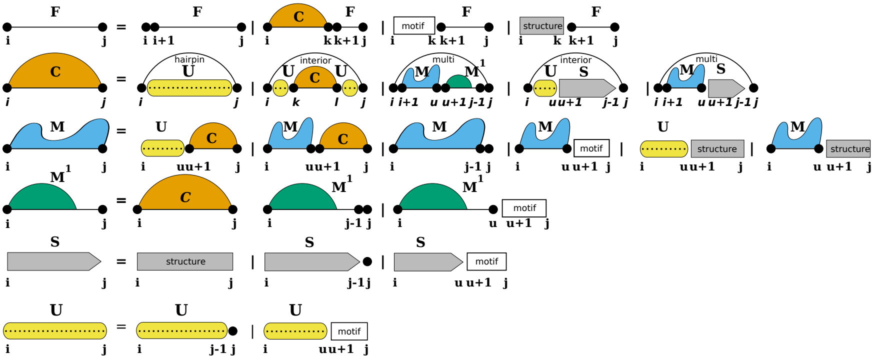
Full decomposition scheme:



RNA Secondary Structure Prediction

Nussinov decomposition scheme:

Full decomposition scheme:

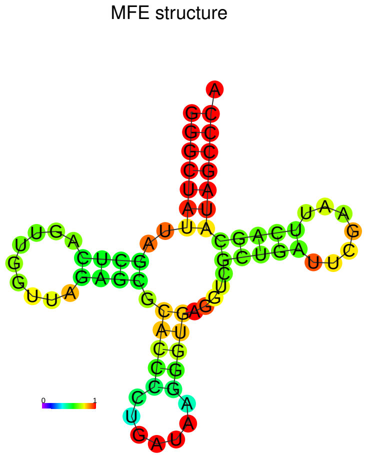


Availability

Implemented in `ViennaRNA` Package including `libRNA` and scripting language interface

ViennaRNA Secondary Structure Prediction

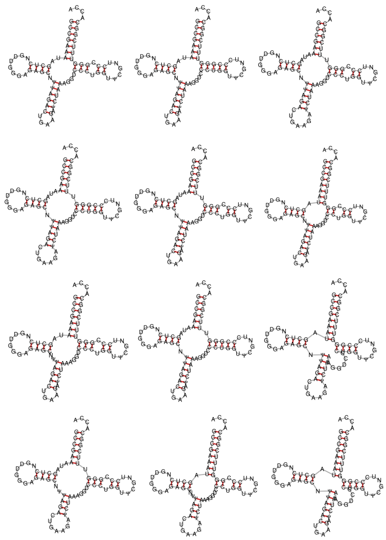
- Minimum free energy (MFE) structure



ViennaRNA Secondary Structure Prediction

- Minimum free energy (MFE) structure
- Suboptimal secondary structures

suboptimal secondary structures



ViennaRNA **Secondary Structure Prediction**

- Minimum free energy (MFE) structure
- Suboptimal secondary structures
- Partition function Z

$$p(s) \propto e^{-\beta E(s)} \quad \text{with} \quad \beta = \frac{1}{kT}$$

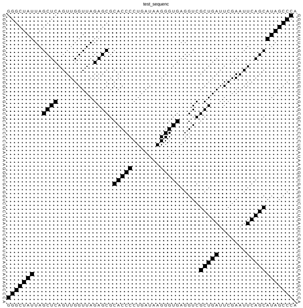
$$Z = \sum_{s} e^{-\beta E(s)} \equiv \sum_{E} g(E) e^{-\beta E}$$

- Probability $p(\mathcal{F})$ of feature $\mathcal{F}$

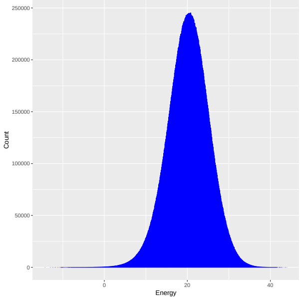
$$p(\mathcal{F}) = \frac{Z_{\mathcal{F}}}{Z} \quad \text{with} \quad Z_{\mathcal{F}} = \sum_{s | F \in s} e^{-\beta E(s)}$$

- Base Pair Probabilities, Accessibility, etc.

base pair probability dot plot



density of states (DoS)



ViennaRNA Secondary Structure Prediction

- Minimum free energy (MFE) structure
- Suboptimal secondary structures
- Partition function Z

$$p(s) \propto e^{-\beta E(s)} \quad \text{with} \quad \beta = \frac{1}{kT}$$

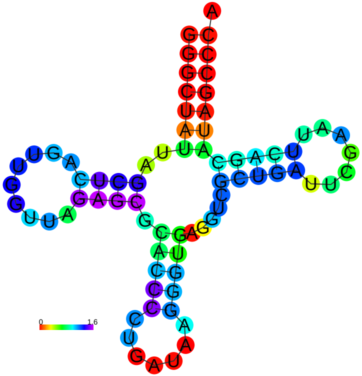
$$Z = \sum_s e^{-\beta E(s)} \equiv \sum_E g(E) e^{-\beta E}$$

- Probability $p(\mathcal{F})$ of feature $\mathcal{F}$

$$p(\mathcal{F}) = \frac{Z_{\mathcal{F}}}{Z} \quad \text{with} \quad Z_{\mathcal{F}} = \sum_{s|F\in s} e^{-\beta E(s)}$$

- Base Pair Probabilities, Accessibility, etc.
- Global / local reliability measures

positional ('Shannon') entropy



$$PE(i) = -\sum_k p_{ik} \ln p_{ik}$$

ViennaRNA Secondary Structure Prediction

- Minimum free energy (MFE) structure
- Suboptimal secondary structures
- Partition function Z

$$p(s) \propto e^{-\beta E(s)} \quad \text{with} \quad \beta = \frac{1}{kT}$$

$$Z = \sum_s e^{-\beta E(s)} \equiv \sum_E g(E) e^{-\beta E}$$

- Probability $p(\mathcal{F})$ of feature $\mathcal{F}$

$$p(\mathcal{F}) = \frac{Z_{\mathcal{F}}}{Z} \quad \text{with} \quad Z_{\mathcal{F}} = \sum_{s|F \in s} e^{-\beta E(s)}$$

- Base Pair Probabilities, Accessibility, etc.
- Global / local reliability measures
- **Hard constraints** limit candidate search space
- **Soft constraints** change candidate evaluation

standard NN energy model

$$E(s) \approx \sum_{L \in s} E_L$$

soft constraints allow for adding pseudo energy terms $E_L^{m_i}$

$$E(s) = \sum_{L \in s} E_L + \Delta\Delta E_L^m$$

$$\Delta\Delta E_L^m = \sum_i (E_L^{m_i} - E_L)$$

`ViennaRNA` Secondary Structure Prediction

- Minimum free energy (MFE) structure
- Suboptimal secondary structures
- Partition function Z

$$p(s) \propto e^{-\beta E(s)} \quad \text{with} \quad \beta = \frac{1}{kT}$$

$$Z = \sum_{s} e^{-\beta E(s)} \equiv \sum_{E} g(E) e^{-\beta E}$$

- Probability $p(\mathcal{F})$ of feature $\mathcal{F}$

$$p(\mathcal{F}) = \frac{Z_{\mathcal{F}}}{Z} \quad \text{with} \quad Z_{\mathcal{F}} = \sum_{s|F \in s} e^{-\beta E(s)}$$

- Base Pair Probabilities, Accessibility, etc.
- Global / local reliability measures
- **Hard constraints** limit candidate search space
- **Soft constraints** change candidate evaluation

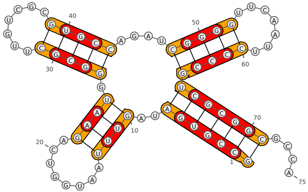
standard NN energy model

$$E(s) \approx \sum_{L \in s} E_L$$

soft constraints allow for adding pseudo energy terms $E_L^{m_i}$

$$E(s) = \sum_{L \in s} E_L + \Delta\Delta E_L^m$$

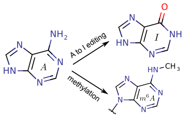
$$\Delta\Delta E_L^m = \sum_{i} (E_L^{m_i} - E_L)$$

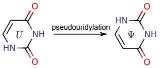


$$\Delta\, G(i) = m * ln(reactivity[i] + 1) + b$$

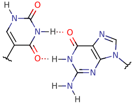
Modified Bases in RNA

Post-transcriptional RNA modifications (epitranscriptome):

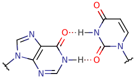




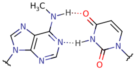
- `Modomics` Database[2] lists 335 different modified bases
- Commonly know modifications: I, Ψ, m^6A, m^1A, m^5C, $\ldots$
- Function and purpose of modifications still largely unknown
- Structural effects of base modifications:
  - correct folding of ncRNAs into functional structures (tRNA, rRNA, etc.)
  - regulation of protein binding sites (mRNAs, lncRNAs)
  - regulation of RNA-RNA binding sites (siRNA, miRNA)
  - Modifications may change pairing partner preference
  - Modifications may (de-)stabilize loop formation



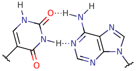
$\Psi \bullet G$



$I \bullet U$



$m^6A \bullet U$

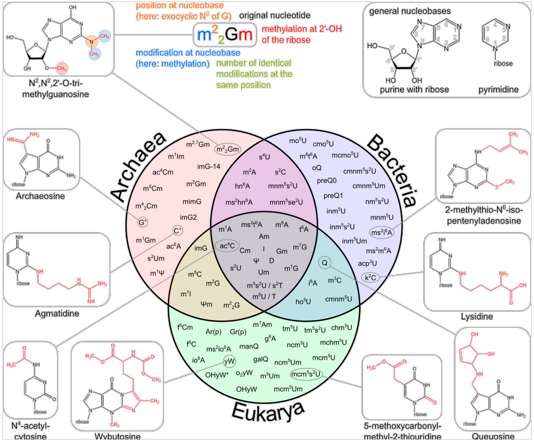


$\Psi \bullet A$

[2] Cappannini et al., “MODOMICS: a database of RNA modifications and related information.”, 2024, NAR 52.D1, D239–D244

Modifications in tRNA[3]

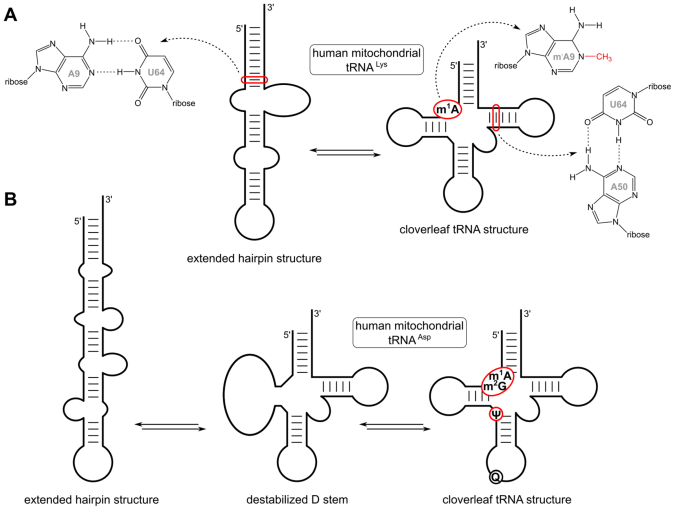
93 known post-transcriptional modifications



- Modifications can be subtle from the RNA structure perspective
- Some are essential to induce structural domain rearrangements

[3] Lorenz, Lünse, and Mörl, “tRNA Modifications: Impact on Structure and Thermal Adaptation”, Biomolecules 2017, 7, 35

Modifications in tRNA[3]



[3] Lorenz, Lünse, and Mörl, “tRNA Modifications: Impact on Structure and Thermal Adaptation”, Biomolecules 2017, 7, 35

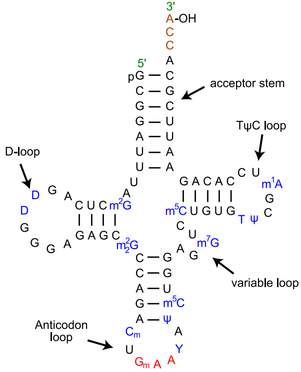
How to model modified bases in prediction algorithms?

Actual Requirements:

- Enhanced Nucleotide Alphabet
- Additional base pairing rules
- Corresponding energy parameters

Obstacles:

- 2D structure effects are known only for a few modifications
- 3D effects either are unknown or impossible to model
- Combinatorial explosion for energy parameters / pairing rules



RNA Modifications - Status Quo of the `ViennaRNA` Package

Built-in modifications:

- inosine (I)
- pseudouridine (Ψ)
- m^6A
- m^5C
- $7DA$
- purine (a.k.a. nebularine)
- dihydrouridine (D)

Available through `JSON` files:

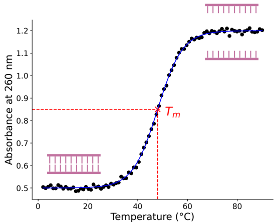
- $m^1\Psi$
- m^5C

Prevent base pairing (hard constraints):

- m^1A
- m^1G
- m_2^2G
- m^3U
- m_2^6A
- $acp^3\Psi$

Energy parameters are incomplete and mostly restricted to base pair stacking

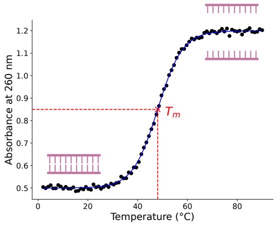
How To Obtain Energy Parameters for Modified Bases?



Experiments

- UV melting approach
- expensive and time-consuming
- Availability of modifications can be a problem

How To Obtain Energy Parameters for Modified Bases?

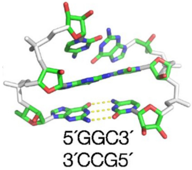


Experiments

- UV melting approach
- expensive and time-consuming
- Availability of modifications can be a problem

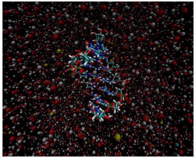
Computational methods

Coarse grained modeling



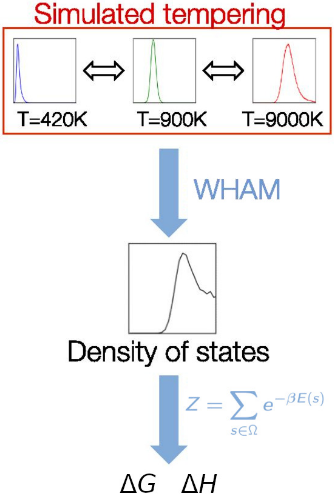
- No solvent
- Less degrees of freedom
- Lower computational costs

Molecular dynamics



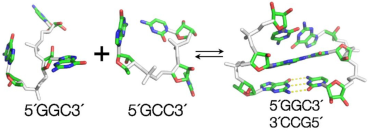
- Solvent
- Higher computational costs
- Fewer approximations
- Complex to set-up

RECCES[4] Simulation



- Monte Carlo simulation
- Temperature changes during the simulation
  - To overcome energy barriers
- Energy function
  - Combination of Physics- and knowledge-based terms

$$\Delta G = G(folded) - G(unfolded)$$



$$\Delta G\begin{pmatrix}\text{5'GGC}\\ \text{3'CCG}\end{pmatrix} = G\begin{pmatrix}\text{5'GGC}\\ \text{3'CCG}\end{pmatrix} - G(\text{GGC}) - G(\text{CCG})$$

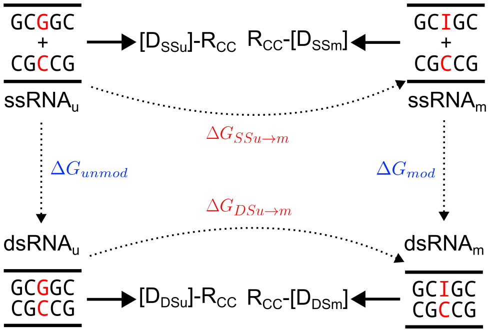
---

[4] Chou et al., "Blind tests of RNAnearest-neighbor energy prediction", 2016 Proceedings of the National Academy of Sciences, 113(30), 8430-8435.

Transformato[5] Thermodynamic Cycle

[5] Karwounopoulos et al. "Relative binding free energy calculations with transformato: A molecular dynamics engine-independent tool." Frontiers in Molecular Biosciences 9 (2022)

Transformato[5] Thermodynamic Cycle

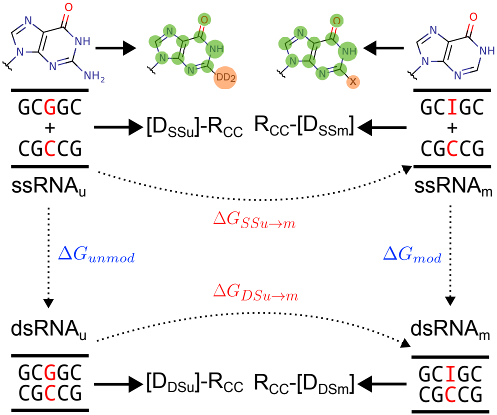


- Construction of an alchemical path

$$\Delta\Delta G_{unmod\rightarrow mod} = \Delta G_{mod} - \Delta G_{unmod}$$
$$= \Delta G_{DSu\rightarrow m} - \Delta G_{SSu\rightarrow Sm}$$

[5] Karwounopoulos et al. "Relative binding free energy calculations with transformato: A molecular dynamics engine-independent tool." Frontiers in Molecular Biosciences 9 (2022)

Transformato[5] Thermodynamic Cycle

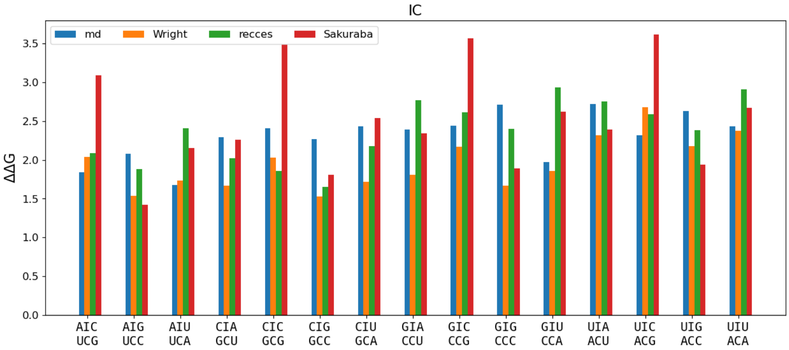


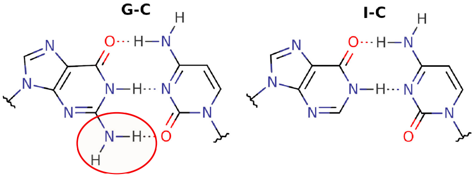
- Construction of an alchemical path
- Maximum common substructure

$$\Delta\Delta G_{unmod\to mod} = \Delta G_{mod} - \Delta G_{unmod}$$
$$= \Delta G_{DSu\to m} - \Delta G_{SSu\to Sm}$$

[5] Karwounopoulos et al. "Relative binding free energy calculations with transformato: A molecular dynamics engine-independent tool." Frontiers in Molecular Biosciences 9 (2022)

GC - IC


IC
md
Wright
recces
Sakuraba
ΔΔG
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
AIC UCG
AIG UCC
AIU UCA
CIA GCU
CIC GCG
CIG GCC
CIU GCA
GIA CCU
GIC CCG
GIG CCC
GIU CCA
UIA ACU
UIC ACG
UIG ACC
UIU ACA

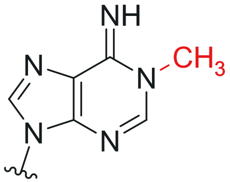

G-C
I-C

Aedes albopictus Mitochondrial tRNA^{Asp}

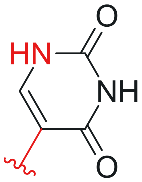
AAAAAAUU"GUUUAAUCAAAAACCPPAGUAUGUC6AACUAAAAAAAAUUAGAUCAUCUAAUAPPUUUUACCA

AAAAAAUUAGUUUAAUCAAAAACCUUAGUAUGUCAAACUAAAAAAAAUUAGAUCAUCUAAUAUUUUUUACCA

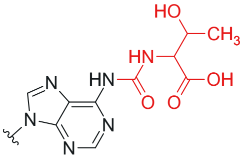
": 1-methyladenosine
P: pseudouridine
6: N6-threonylcarbamoyladenosine



m^1A



Ψ



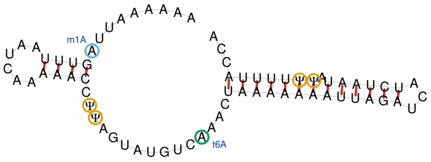
t^6A

Aedes albopictus Mitochondrial tRNA^Asp

AAAAAAUU"GUUUAAUCAAAAACCPPAGUAUGUC6AACUAAAAAAAAUUAGAUCAUCUAAUAPPUUUUACCA

AAAAAAUUAGUUUAAUCAAAAACCUUAGUAUGUCAAACUAAAAAAAAUUAGAUCAUCUAAUAUUUUUUACCA

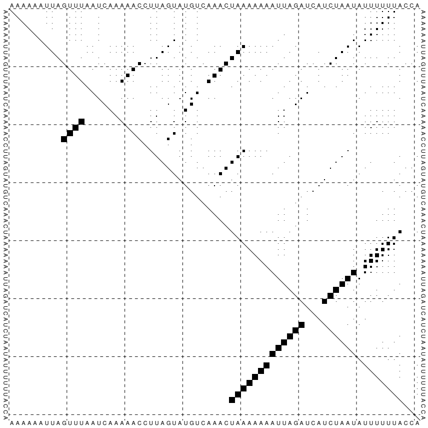
": 1-methyladenosine
P: pseudouridine
6: N6-threonylcarbamoyladenosine



Known parameters

Prevent base pairing

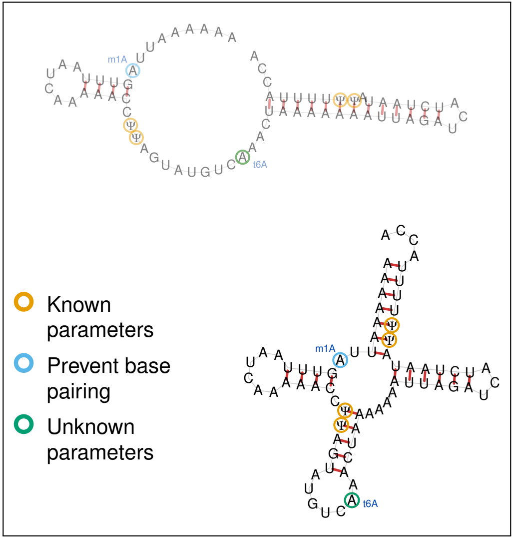
Unknown parameters

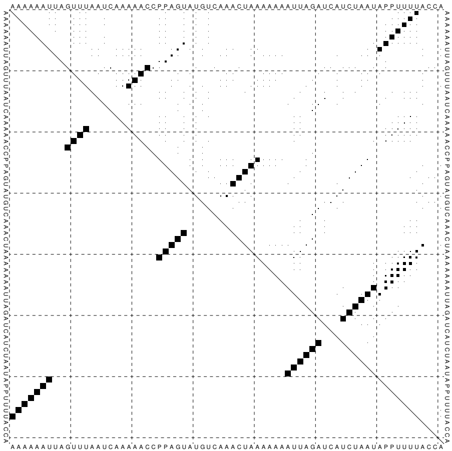


Aedes albopictus Mitochondrial tRNA^{Asp}

AAAAAAUU"GUUUAAUCAAAAACCPPAGUAUGUC6AACUAAAAAAAUUAGAUCAUCUAAUAPPUUUUACCA
AAAAAAUUAGUUUAAUCAAAAACCUUAGUAUGUCAAACUAAAAAAAUUAGAUCAUCUAAUAUUUUUUACCA

": 1-methyladenosine
P: pseudouridine
6: N6-threonylcarbamoyladenosine

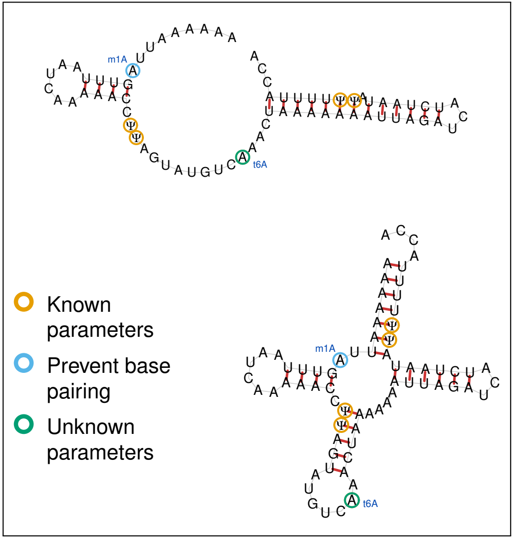


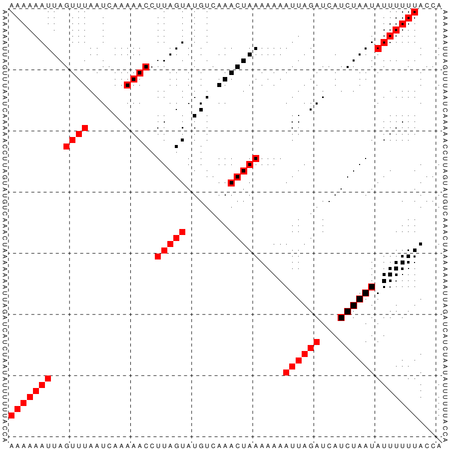


Aedes albopictus Mitochondrial $tRNA^{Asp}$

AAAAAAUU"GUUUAAUCAAAAACCPPAGUAUGUC6AACUAAAAAAAUUAGAUCAUCUAAUAPPUUUUACCA
AAAAAAUUAGUUUAAUCAAAAACCUUAGUAUGUCAAACUAAAAAAAUUAGAUCAUCUAAUAUUUUUUACCA

": 1-methyladenosine
P: pseudouridine
6: N6-threonylcarbamoyladenosine





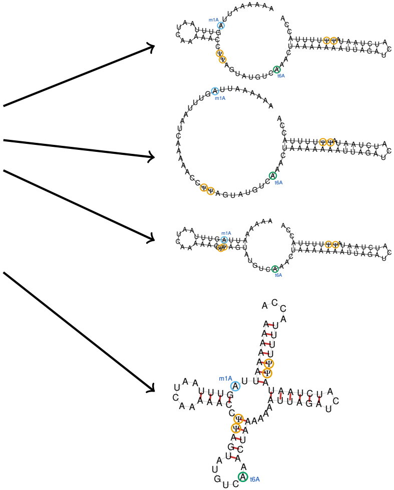
Aedes albopictus Mitochondrial tRNA^{Asp}

AAAAAAUU"GUUUAAUCAAAAACCPPAGUAUGUC6AACUAAAAAAAAUUAGAUCAUCUAAUAPPUUUUACCA
AAAAAAUUAGUUUAAUCAAAAACCUUAGUAUGUCAAACUAAAAAAAAUUAGAUCAUCUAAUAUUUUUUACCA

": 1-methyladenosine
P: pseudouridine
6: N6-threonylcarbamoyladenosine

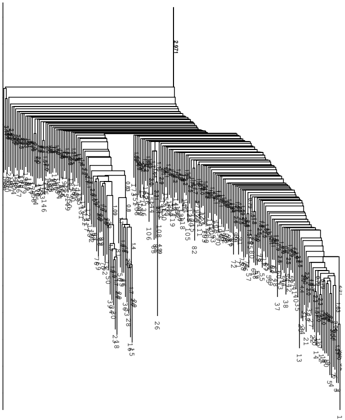
```
AAAAAAUUAGUUUAAUCAAAAACCUUAGUAUGUCAAACUAAAAAAAAUUAGAUCAUCUAAUAUUUUUUACCA  -5.10
.........((((......))))................(((((((((((((...)))))).)))))))...  -5.10
.......................................(((((((((((((...)))))).)))))))...  -5.10
.....((((((((......))))..))))..........(((((((((((((...)))))).)))))))...  -4.60
.........((((......)))).................((((((((((((....))))).)))))))...  -4.40
........................................((((((((((((....))))).)))))))...  -4.30
(((((((..((((......)))).(((((.......)))))....((((((...)))))))))))))....  -4.10
.........((((......)))).(((((.......)))))....((((((...))))))...........  -4.00
......(((((((...................)))))))....((((((...))))))...........  -4.00
......(((((((...((...((....)).))..)))))))....((((((...))))))...........  -4.00
........................(((((.......)))))....((((((...))))))...........  -3.90
```

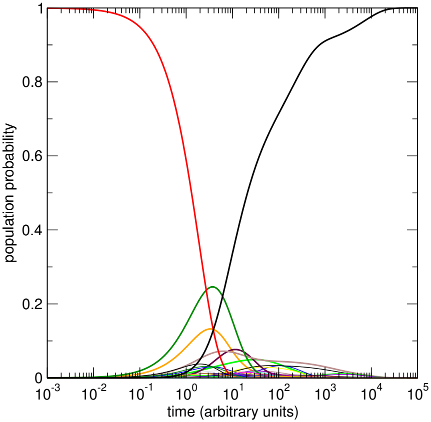
suboptimals of unmodified tRNA



Aedes albopictus Mitochondrial tRNA^{Asp} Folding Dynamics

Energy landscape[6] and folding dynamics[7] of tRNA^{Asp} **with modifications**





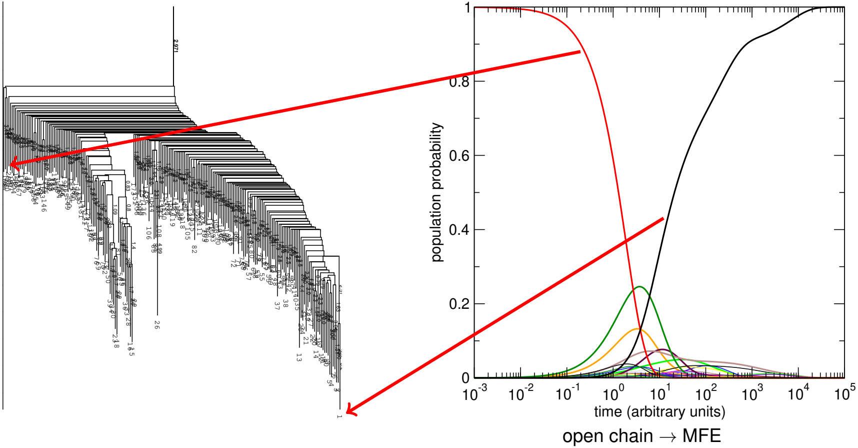
open chain → MFE

[6] Flamm et al., “Barrier trees of degenerate landscapes”, 2002, Z. Phys. Chem. 216: 155-73

[7] Wolfinger et al., “Efficient Computation of RNA Folding Dynamics”, 2004, J.Phys.A: Math.Gen. 37(17):4731-41

Aedes albopictus Mitochondrial $tRNA^{Asp}$ Folding Dynamics

Energy landscape[6] and folding dynamics[7] of $tRNA^{Asp}$ **with modifications**



[6] Flamm et al., “Barrier trees of degenerate landscapes”, 2002, Z. Phys. Chem. 216: 155-73

[7] Wolfinger et al., “Efficient Computation of RNA Folding Dynamics”, 2004, J.Phys.A: Math.Gen. 37(17):4731-41

Modified Bases Significantly Influence Structure Space

Takeaway Message:

- We employ MC and MD approaches to compute missing energy parameters
- Additional parameters do not necessarily increase performance
- Constraints become complex for more modifications and contexts
- Unrealistic to include full parameters with many modified bases
- Base annotation is not unique (`tRNAdb`[8], `RNAmod`[9], `MODOMICS`[10])

`ViennaRNA` Package:

- Can handle RNA modifications via the soft constraints framework
- New parameters - even partial - can easily be added as they become available
- Software is freely available for academic and commercial use

---

[8] Jühling et al., "tRNAdb 2009: compilation of tRNA sequences and tRNA genes.", 2009, NAR 37, D159–D162

[9] Liu et al., "RNAmod: an integrated system for the annotation of mRNA modifications", 2019, NAR 47, W548-W555

[10] Boccaletto et al., "MODOMICS: a database of RNA modification pathways. 2017 update", 2018, NAR 46, D303-D307

Acknowledgements

- Thomas Spicher
- Yuliia Varenyk
- Ronny Lorenz
- Ivo L. Hofacker

michaelwolfinger.com