Computational Estimation of Energy Parameters for modified nucleotides

–

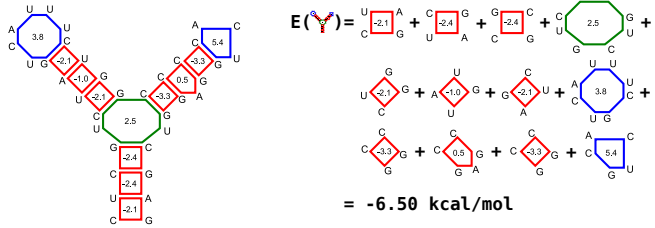
A tRNA use case

Thomas Spicher

TBI Vienna
University of Vienna

Computational Approaches to RNA Structure and Function
Benasque Science Center
Jul 05 - Jul 17, 2026

RNA Secondary Structures - Nearest Neighbor Energy Model

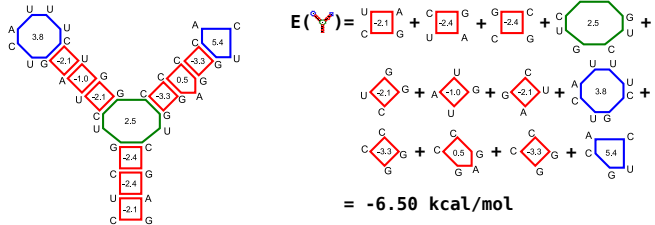


- Decomposition of Secondary structures s into loops L
- Contributions of a base pair depend on neighboring pairs
- Each loop L is assigned a free energy contribution E_L[1]

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

[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



- 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]

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

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

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

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

[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 Modifications - Status Quo of the ViennaRNA Package [2]

Built-in modifications:

- inosine (I)
- pseudouridine (Ψ)
- m^6A
- 7-deaza-adenosine ($7DA$)
- purine (a.k.a. nebularine)
- dihydrouridine (D) (in silico)

Available through JSON files:

- $m^1\Psi$ (in silico)
- m^5C (in silico)

Prevent base pairing (hard constraints):

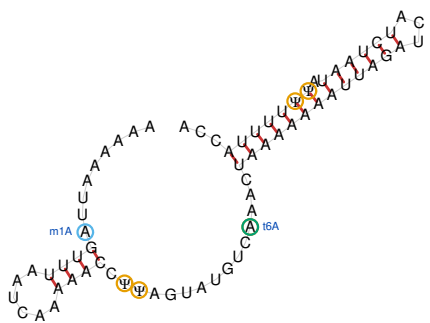
- m^1A
- m^1G
- m^2_2G (Wobble pairing allowed)
- m^3U
- m^6_2A
- $acp^3\Psi$

Energy parameters are incomplete and mostly restricted to base pair stacking

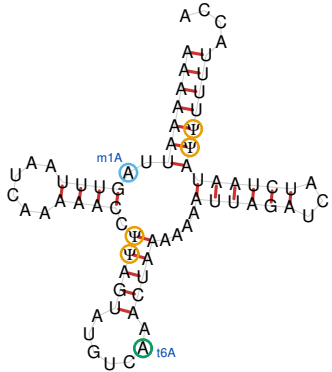
[2] Varenyk, Y., Spicher, T., Hofacker, I.L, Lorenz, R., "Modified RNAs and predictions with the ViennaRNA Package", Bioinformatics, Nov. 2023

Using existing NN parameters on tRNA

Default parameters



Modifications included



Known parameters Prevent base pairing Unknown parameters

Aedes albopictus Mitochondrial $tRNA^{Asp}$

tRNA Databases

- tRNAdb 2009[3]
- t-psi-c database[4]
- Modomics database[5]

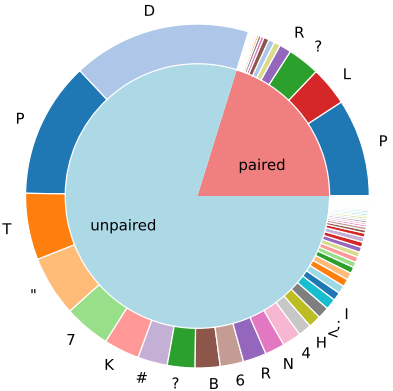
- 874 sequences with at least one annotated modifications
- Aligned with infernal, according to the tRNAdb alignment (Sprinzl)

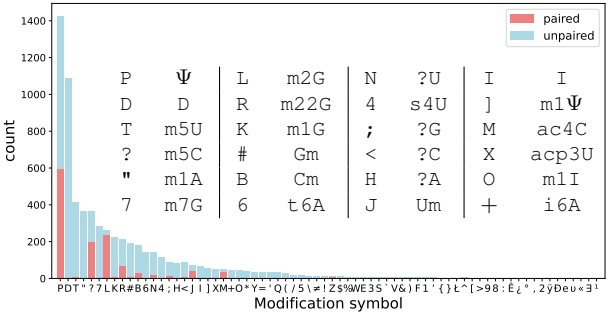
[3] Jühling, Frank, et al. "tRNAdb 2009: compilation of tRNA sequences and tRNA genes." Nucleic acids research 37.suppl 1 (2009).

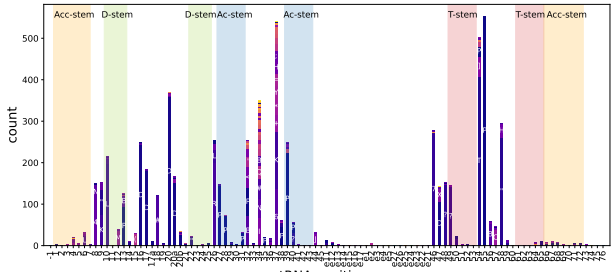
[4] Sajek, Marcin Piotr, et al. "T-psi-C: user friendly database of tRNA sequences and structures." Nucleic acids research 48.D1 (2020).

[5] Cappannini, Andrea, et al. "MODOMICS: a database of RNA modifications and related information. 2023 update." Nucleic Acids Research 52.D1 (2024).

Modifications Overview

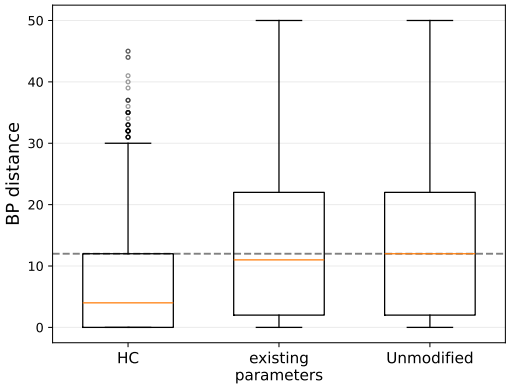


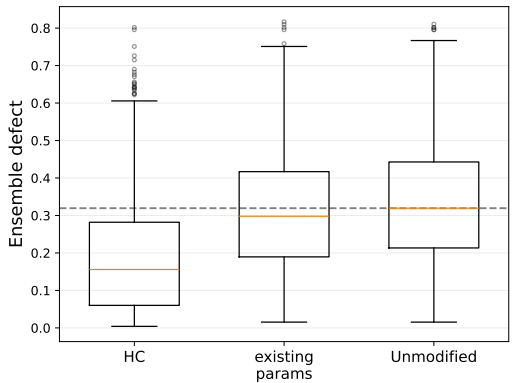




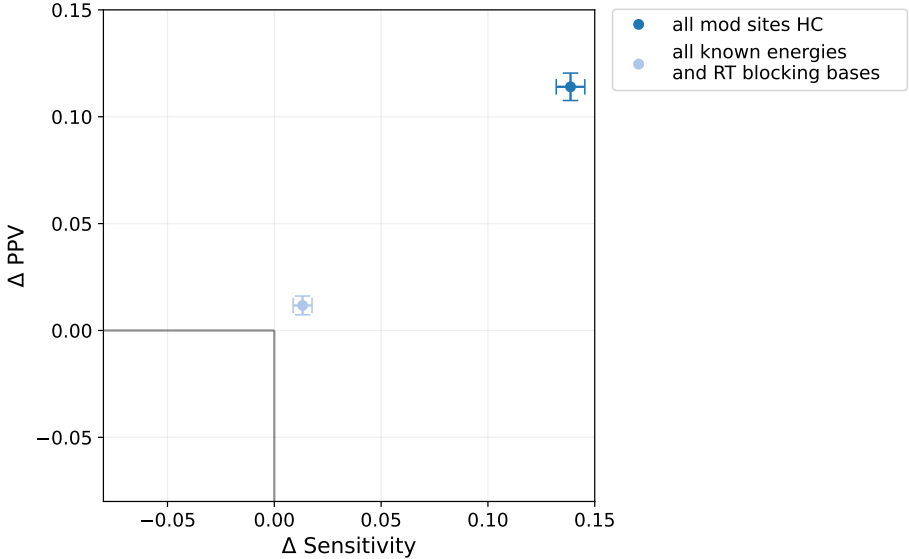
How well are predictions performing on our dataset?

- Base pair distance
  - Minimum number of changes from MFE to target
- Ensemble defect
  - Average dissimilarity of the ensemble to a given target structure

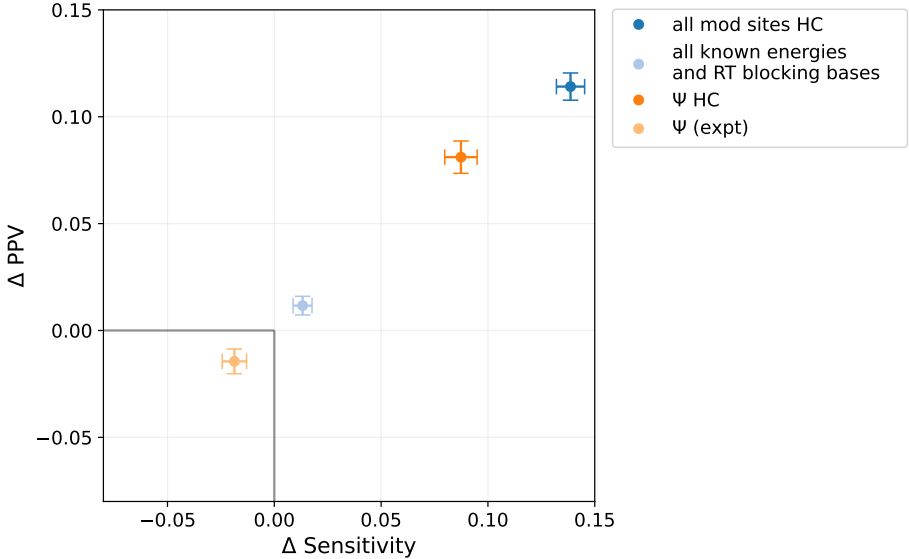




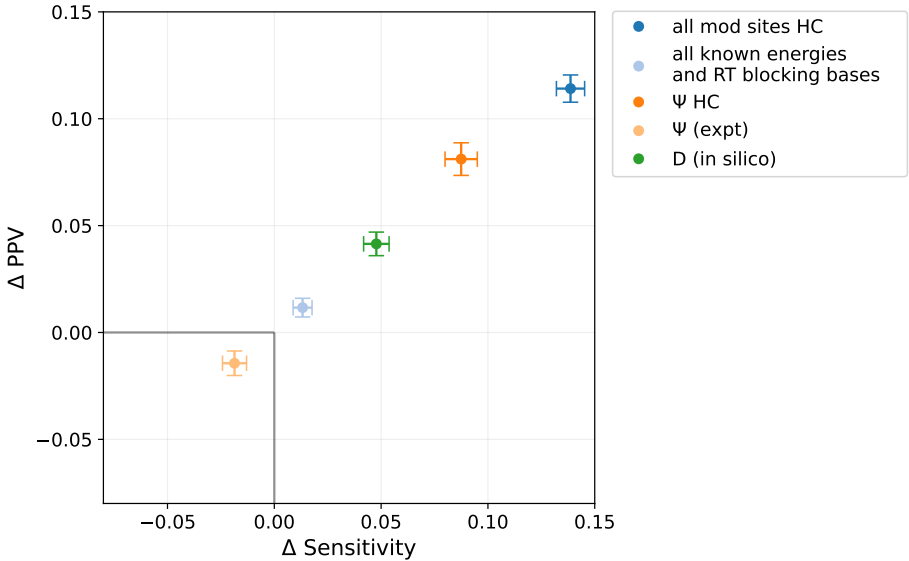
How well are predictions performing on our dataset?



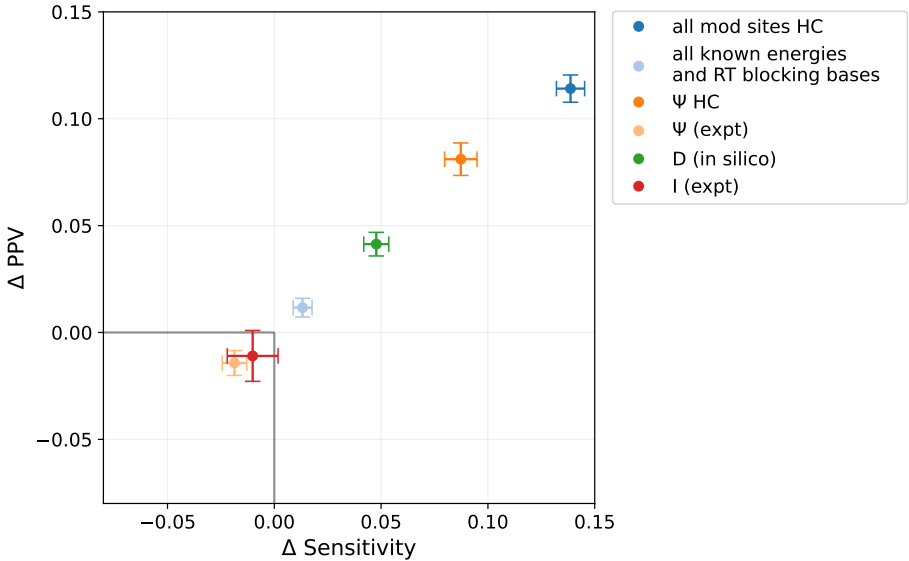
How well are predictions performing on our dataset?



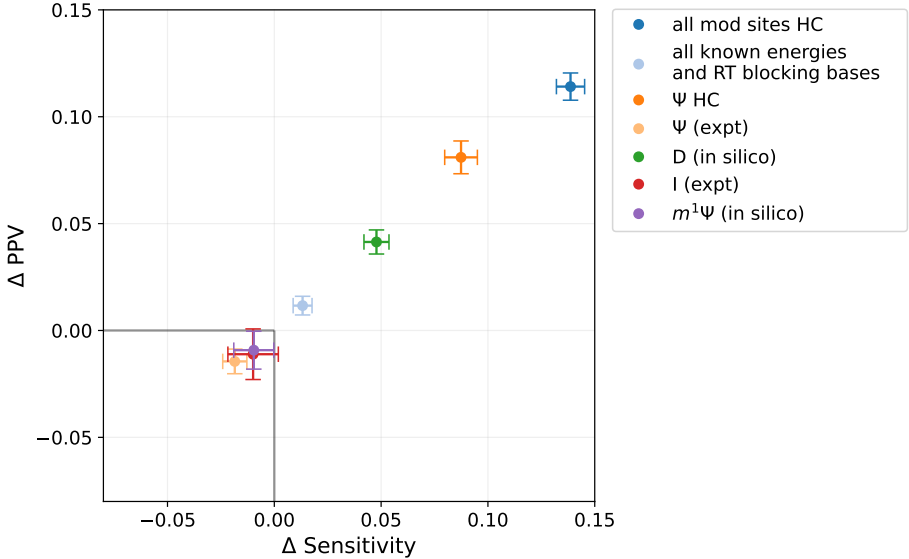
How well are predictions performing on our dataset?



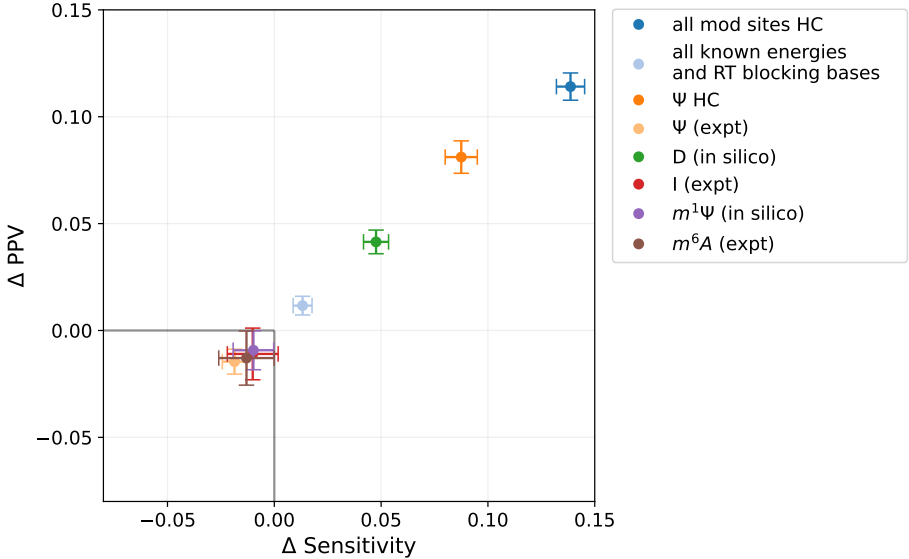
How well are predictions performing on our dataset?



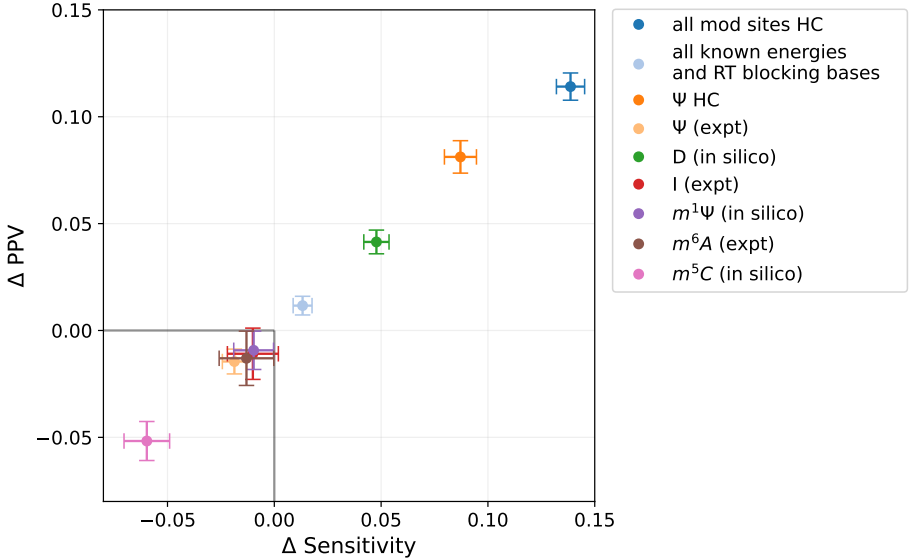
How well are predictions performing on our dataset?



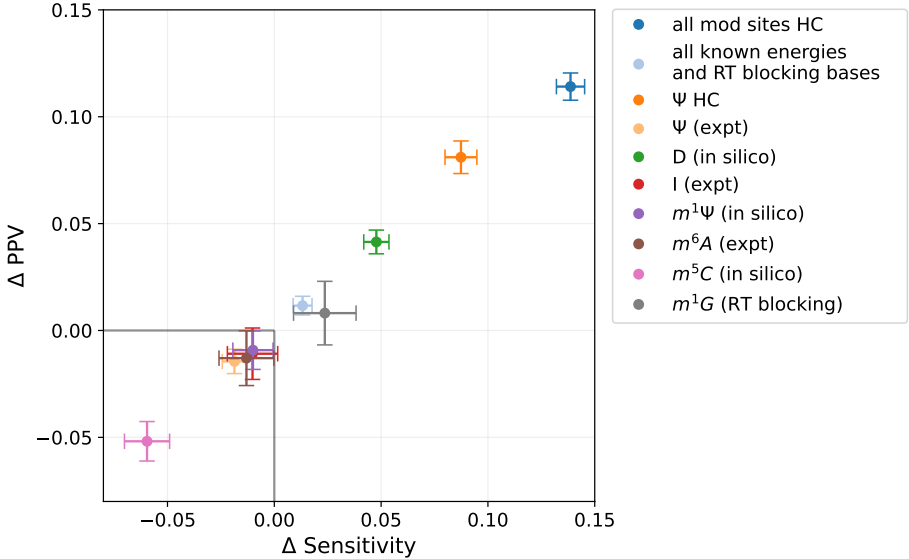
How well are predictions performing on our dataset?



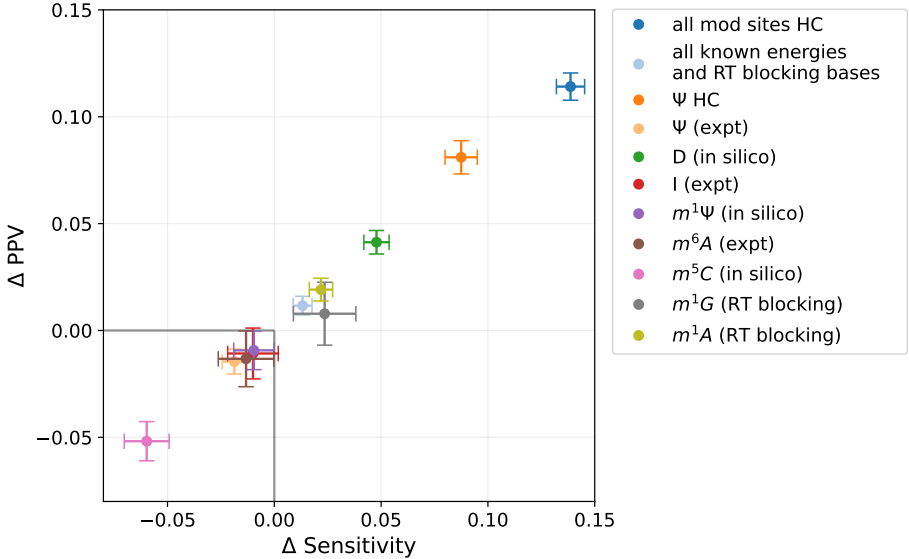
How well are predictions performing on our dataset?



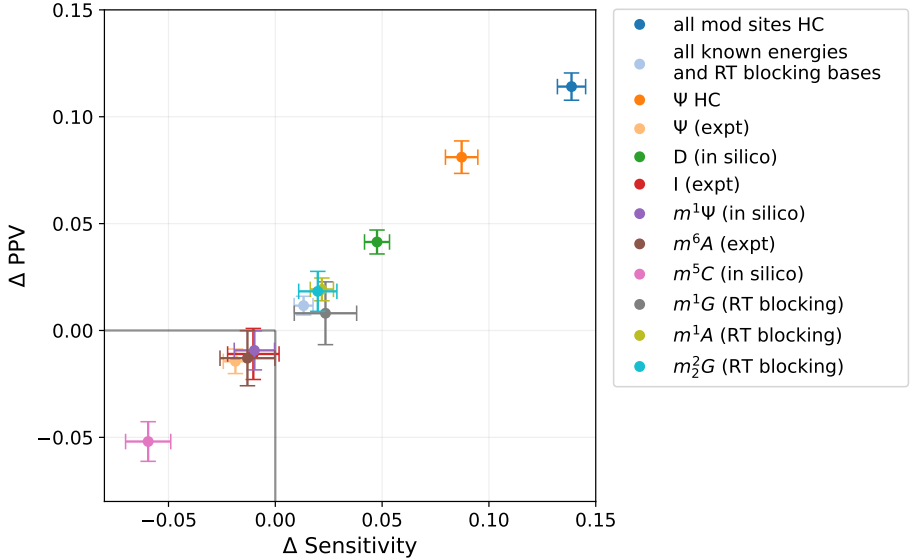
How well are predictions performing on our dataset?



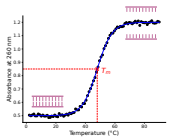
How well are predictions performing on our dataset?



How well are predictions performing on our dataset?



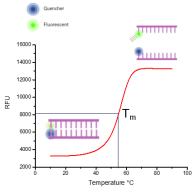
How can we extract energy parameters for modified bases?



Experiments

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

How can we extract energy parameters for modified bases?



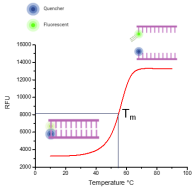
Experiment

- Fluorophore-quencher approach[6]
  - up to 100 times less material
  - Drastic reduction of the costs
- Availability of modifications can be a problem

[6] Khosravi, Spicher, et al. "Systematic measurement of thermodynamic nearest-neighbor parameters for Inosine-containing double-stranded RNAs using a fluorophore-quencher-based approach" (submitted).

How can we extract energy parameters for modified bases?

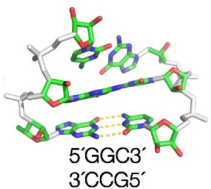
Experiment



- Fluorophore-quencher approach[6]
  - up to 100 times less material
  - Drastic reduction of the costs
- Availability of modifications can be a problem

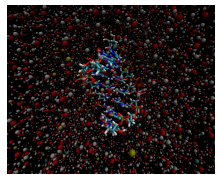
Computational methods

Coarse grained modeling



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

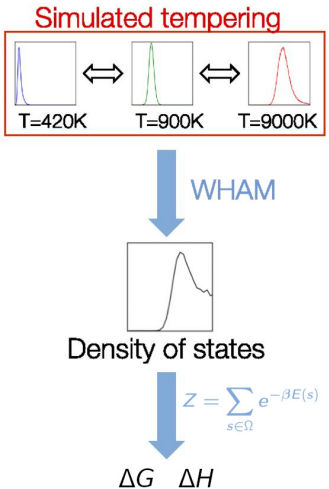
Molecular dynamics



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

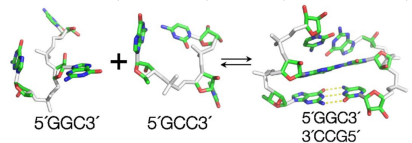
[6] Khosravi, Spicher, et al. "Systematic measurement of thermodynamic nearest-neighbor parameters for Inosine-containing double-stranded RNAs using a fluorophore-quencher-based approach" (submitted).

RECCES[7] Simulation



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

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

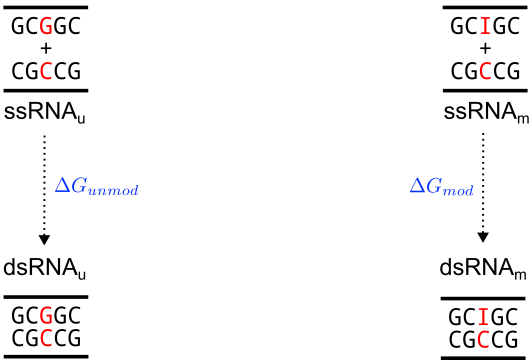


$$\Delta G\begin{pmatrix}5'\text{GGC}\\3'\text{CCG}\end{pmatrix} =$$

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

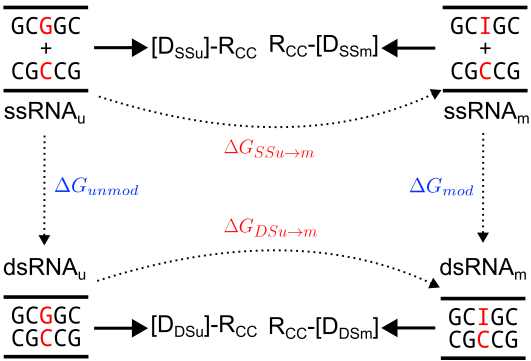
[7] Chou et al., "Blind tests of RNAnearest-neighbor energy prediction", 2016 Proceedings of the National Academy of Sciences, 113(30), 8430–8435.

Thermodynamic cycle from Transformato[8]



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

Thermodynamic cycle from Transformato[8]



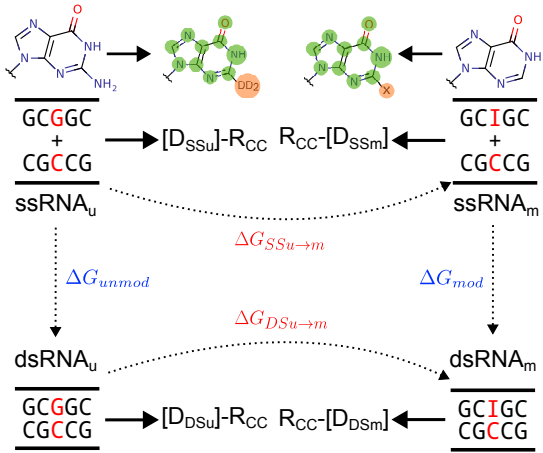
- Construction of an alchemical path

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

calculated by the Multistate Bennett Acceptance Ratio

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

Thermodynamic cycle from Transformato[8]



- Construction of an alchemical path
- Maximum common substructure

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

calculated by the Multistate Bennett Acceptance Ratio

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

RMSD for the different stacking parameter sets

Dihydrouridine (D)

	unmodified	MD
MD	1.353	
RECCES	0.846	0.848

Pseudouridine (Ψ)

	unmodified	expt	RECCES
expt	1.066		
RECCES	1.073	1.895	
MD	0.119	1.034	1.106

N1-Methylpseudouridine ($m^1\Psi$)

	unmodified	expt	MD Dutta
expt	0.520		
MD Dutta	0.632	0.429	
RECCES	0.665	0.910	0.966

Inosine (I)

	unmodified (G)	expt UV	expt FQ	RECCES
expt UV	1.052			
expt FQ	0.821	0.594		
RECCES	0.950	0.680	0.480	
MD	0.954	0.487	0.332	0.402

N6-methyladenosine (m^6A)

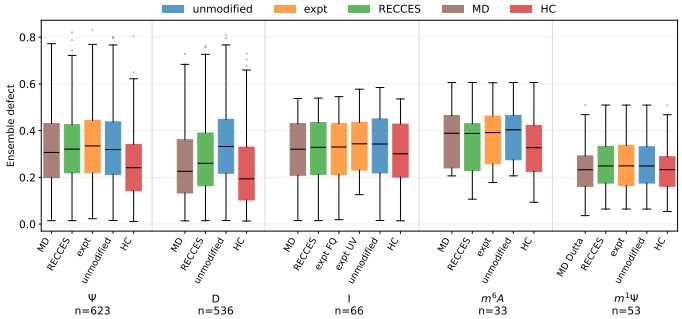
	unmodified	expt	RECCES
expt	0.499		
RECCES	1.497	1.055	
MD	0.283	0.505	1.393

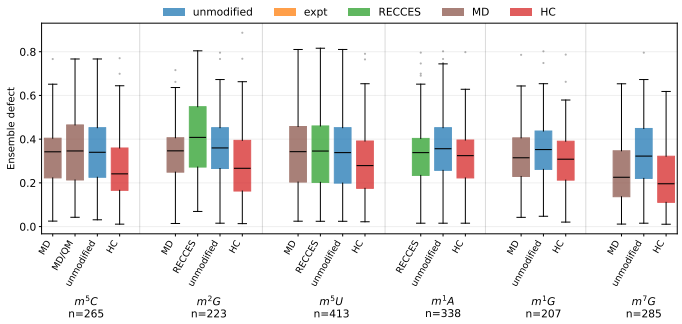
N2-methylguanosine (m^2G)

	unmodified	RECCES
RECCES	2.202	
MD	0.492	1.975

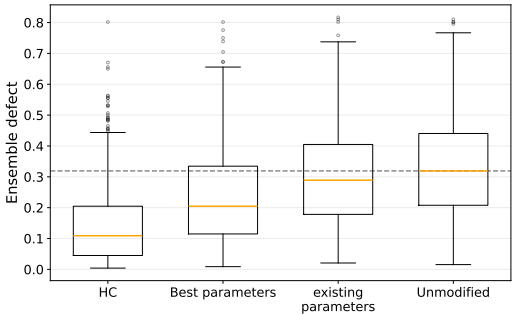
5-methyluridine (m^5U)

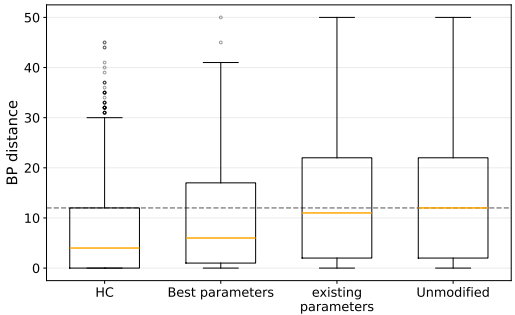
	unmodified	RECCES
RECCES	0.861	
MD	0.481	1.016


unmodified
expt
RECCES
MD
HC
Ensemble defect
0.0
0.2
0.4
0.6
0.8
MD
RECCES
expt
unmodified
HC
Ψ
n=623
D
n=536
expt FQ
expt UV
I
n=66
m^6A
n=33
MD Dutta
$m^1\Psi$
n=53


unmodified
expt
RECCES
MD
HC
Ensemble defect
0.0
0.2
0.4
0.6
0.8
MD
MD/DM
unmodified
HC
RECCES
m^5C
n=265
m^2G
n=223
m^5U
n=413
m^1A
n=338
m^1G
n=207
m^7G
n=285

How good could predictions possibly get?





Conclusion

- We can now fold RNA secondary structures with modified nucleotides
  - On tRNA, using the modification information improves the prediction
  - Some modifications appear to be there for the correct folding
- The improvement is smaller than we hoped
  - Many modifications have no parameters at all
  - The parameters that exist are mostly incomplete
  - Larger errors than the canonical Turner set for experimental and computational data
- The gap is closing: m^6A essentially complete, preprints for Ψ and $m^1\Psi$
- Folding with modifications lets us ask which ones matter — in tRNA and beyond
  - So far: base-pair-blocking modifications and D dominate tRNA folding
  - But most of the information is still missing

Acknowledgments

Ivo Hofacker
Ronny Lorenz
Katrin Gutenbrunner
Yuliia Varenyk

Thank you for your attention