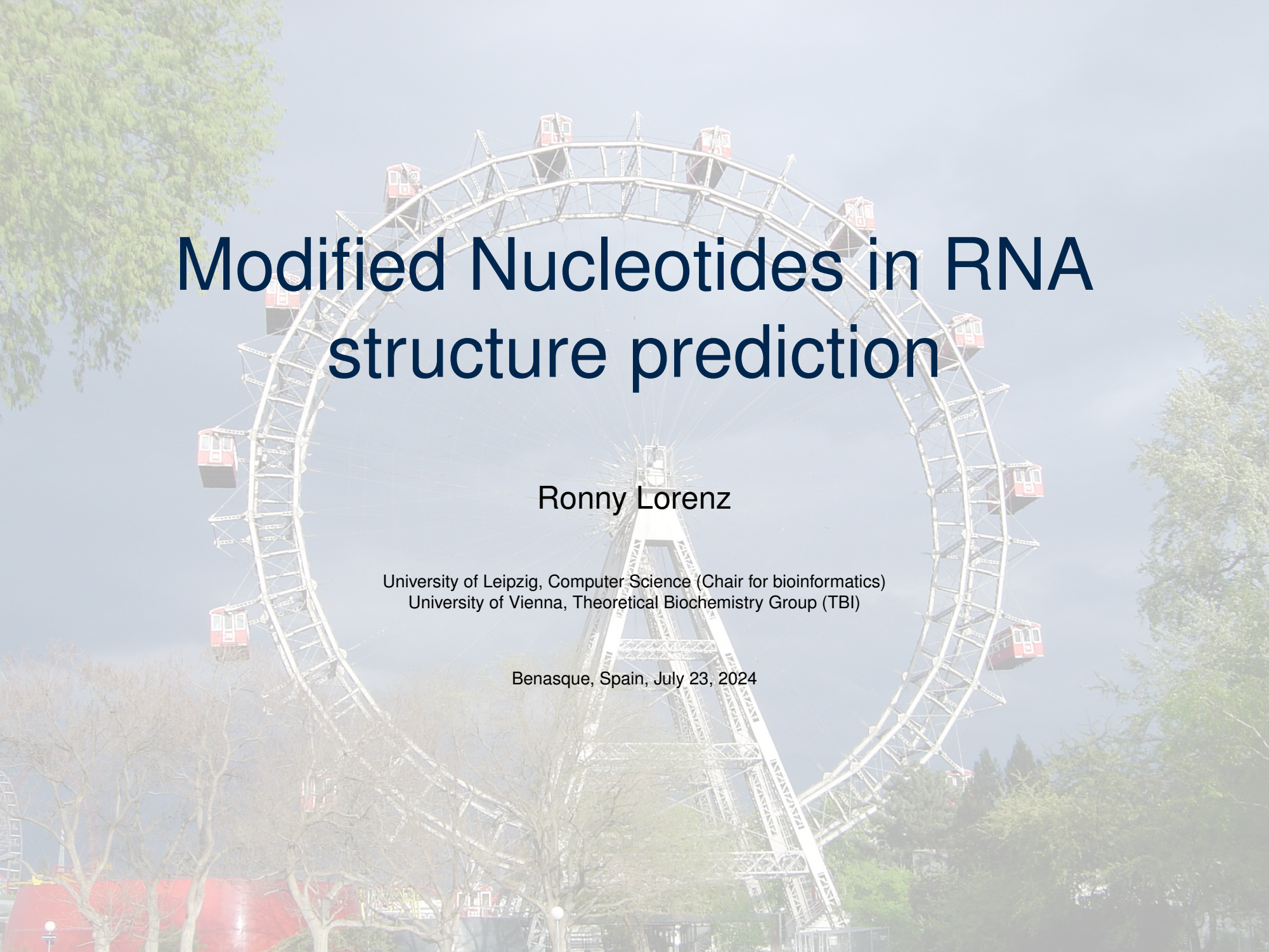


Modified Nucleotides in RNA structure prediction

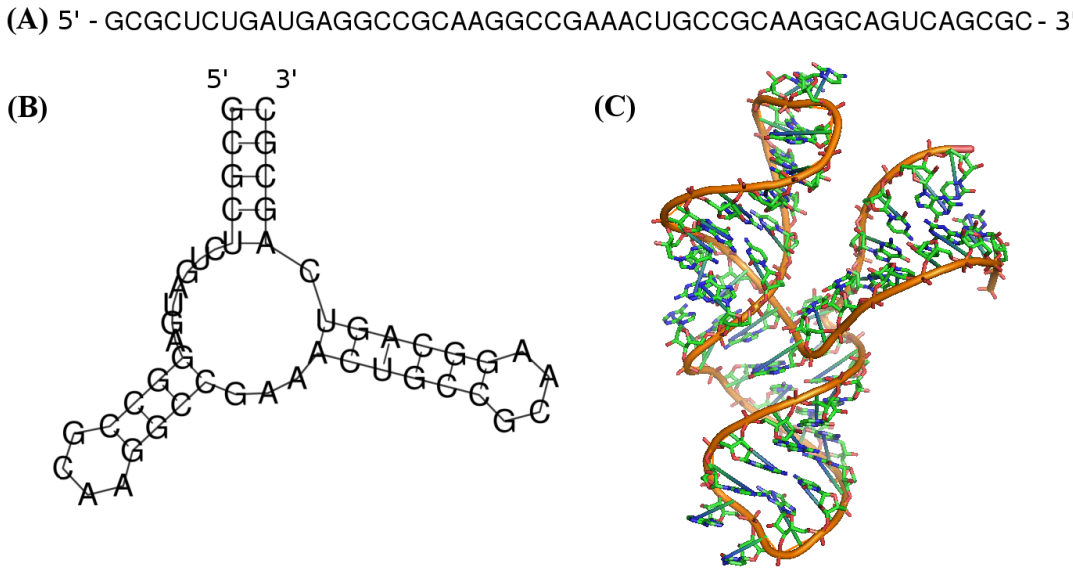
Ronny Lorenz

University of Leipzig, Computer Science (Chair for bioinformatics)
University of Vienna, Theoretical Biochemistry Group (TBI)

Benasque, Spain, July 23, 2024

RNA Secondary Structures

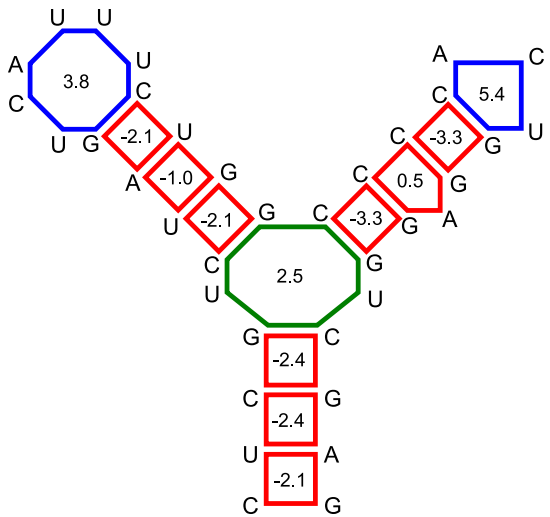
RNAs fold hierarchical (A) $\rightarrow$ (B) $\rightarrow$ (C)

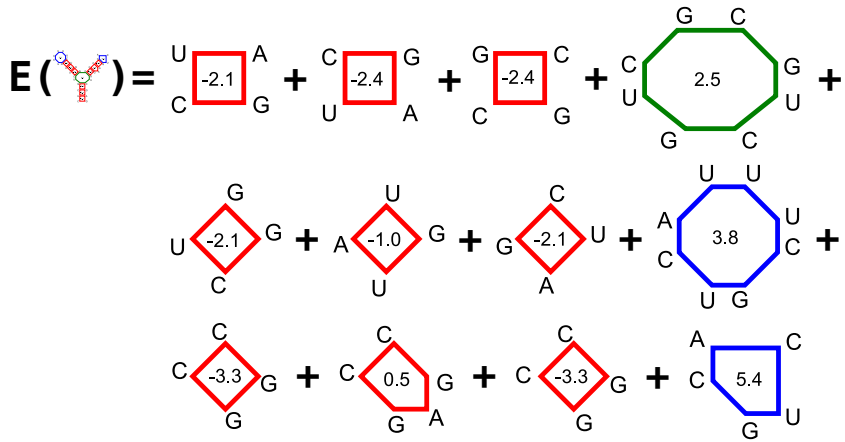


Secondary Structure (B):

- Set of nested base pairs
- Captures majority of stabilizing interactions
- Many thermodynamic properties can be predicted efficiently
- Very good prediction accuracy for small RNAs
- Starting point for 3D structure modeling
- Good abstraction to perform folding kinetics simulation

RNA Secondary Structures - Nearest Neighbor Energy Model





= -6.50 kcal/mol

- Secondary structures s can be uniquely decomposed into loops L
- Contributions of a base pair only depends 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] *A. Mittal et al. "NNDB: An Expanded Database of Nearest Neighbor Parameters for Predicting Stability of Nucleic Acid Secondary Structures". In: Journal of Molecular Biology (2024), p. 168549.* DOI: 10.1016/j.jmb.2024.168549

RNA Secondary Structures - Statistical Thermodynamics

$$
\begin{aligned}
p(F) &\propto e^{-\beta E(F)}, \quad \text{with} \quad \beta = \frac{1}{RT} \\
&= \frac{1}{Q} \sum_{s|F \in s} e^{-\beta E(s)}, \quad \text{with} \quad Q = \sum_{s} e^{-\beta E(s)}
\end{aligned}
$$

Efficient dynamic programming algorithms are available to compute

- Minimum free energy $MFE = \min_s E(s)$
- Partition function Q
- Probabilities p_{kl} for base pairs (k,l), i.e. $p_{kl} = p(F)$ with $F = (k,l)$
- Suboptimal and locally stable structures, ligand binding, ...

Implementations for different input data types[2]

- single sequences
- (multiple) interacting sequences
- sequence alignments (consensus structure)

---

[2] *R. Lorenz et al. "ViennaRNA Package 2.0". In: Algorithms for Molecular Biology 6 (2011), pp. 1–14.* DOI: 10.1186/1748-7188-6-26

RNA Secondary Structures - Statistical Thermodynamics

$$
\begin{aligned}
p(F) &\propto e^{-\beta E(F)}, \quad \text{with} \quad \beta = \frac{1}{RT} \\
&= \frac{1}{Q} \sum_{s|F \in s} e^{-\beta E(s)}, \quad \text{with} \quad Q = \sum_{s} e^{-\beta E(s)}
\end{aligned}
$$

Efficient dynamic programming algorithms are available to compute

- Minimum free energy $MFE = \min_s E(s)$
- Partition function Q
- Probabilities p_{kl} for base pairs (k,l), i.e. $p_{kl} = p(F)$ with $F = (k,l)$
- Suboptimal and locally stable structures, ligand binding, ...

Implementations for different input data types[2]

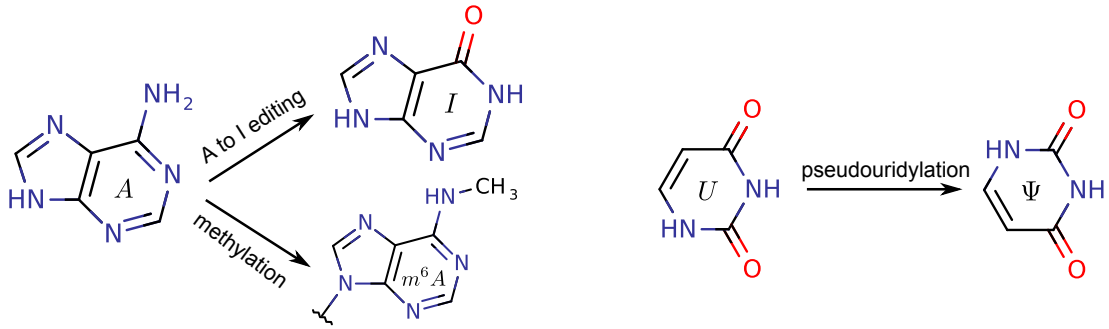
- single sequences
- (multiple) interacting sequences
- sequence alignments (consensus structure)

Implementations are restricted to unmodified bases!

[2] *R. Lorenz et al. "ViennaRNA Package 2.0". In: Algorithms for Molecular Biology 6 (2011), pp. 1–14.* DOI: 10.1186/1748-7188-6-26

Modified Bases in RNA

Post-transcriptional RNA modifications (epitranscriptome):

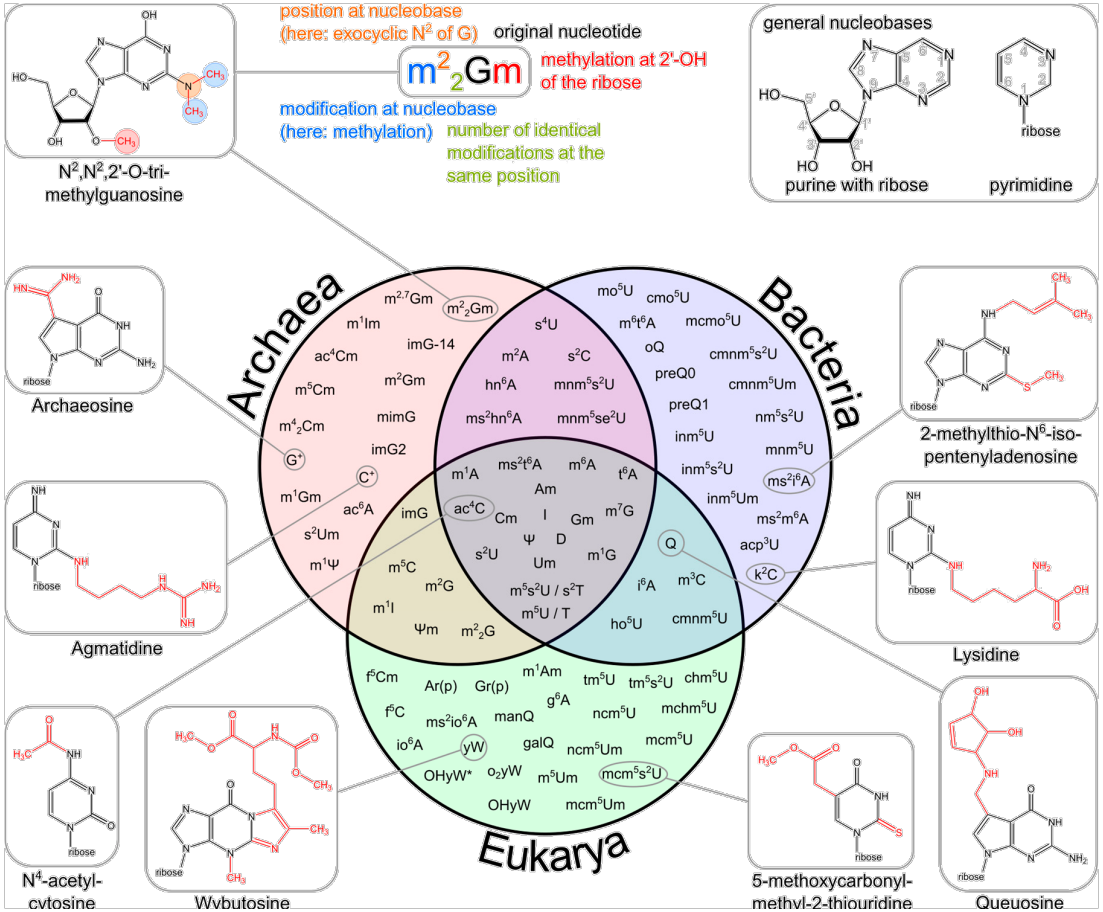


- `Modomics` Database[3]: > 170 distinct modified bases
- Well known modifications: I, Ψ, m^6A, m^1A, m^5C, $\ldots$
- Function and purpose of modifications still largely unknown
- Structural effects of base modifications:
  1. correct folding into functional structures (tRNA, rRNA, etc.)
  2. regulation of protein binding sites (mRNAs, lncRNAs)
  3. regulation of RNA-RNA binding sites (siRNA, miRNA)
  4. Modifications may change pairing partner preference
  5. Modifications may (de-)stabilize loop formation

[3] *A. Cappannini et al. "MODOMICS: a database of RNA modifications and related information. 2023 update".* In: Nucleic Acids Research 52.D1 (2024), pp. D239–D244. DOI: 10.1093/nar/gkad1083

Modifications in tRNA[4]

- 93 known post-transcriptional modifications



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

[4] *C. Lorenz et al. "tRNA modifications: impact on structure and thermal adaptation". In: Biomolecules 7.2 (2017), p. 35.* DOI: 10.3390/biom7020035

How to incorporate modified bases in predictions?

Obstacles:

- 2D structure effects known only for a minority of modifications
- 3D effects may yield non-Nearest-Neighbor dependencies
- Combinatorial explosion for energy parameters and pairing rules

Possible Solutions:

- Re-implement algorithms with
  1. Enhanced nucleotide alphabet
  2. Additional pairing rules
  3. More energy parameters
- Use constraints[5]
  1. Hard constraints: pairing rules
  2. Soft constraints: supplement (partial) energy parameters

[5] *R. Lorenz et al. "RNA folding with hard and soft constraints".* In: Algorithms for Molecular Biology *11.1 (2016), pp. 1–13.* DOI: 10.1186/s13015-016-0070-z

How to incorporate modified bases in predictions?

Obstacles:

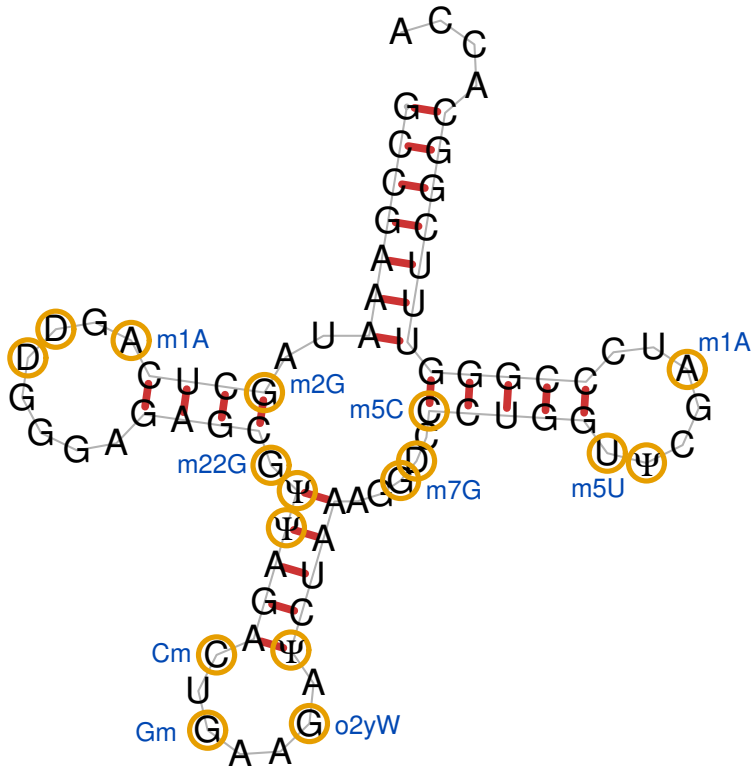
- 2D structure effects known only for a minority of modifications
- 3D effects may yield non-Nearest-Neighbor dependencies
- Combinatorial explosion for energy parameters and pairing rules

Possible Solutions:

- Re-implement algorithms with
  1. Enhanced nucleotide alphabet
  2. Additional pairing rules
  3. More energy parameters
- **Use constraints**[5]
  1. Hard constraints: pairing rules
  2. Soft constraints: supplement (partial) energy parameters

[5] *R. Lorenz et al. "RNA folding with hard and soft constraints".* In: Algorithms for Molecular Biology *11.1 (2016), pp. 1–13.* DOI: 10.1186/s13015-016-0070-z

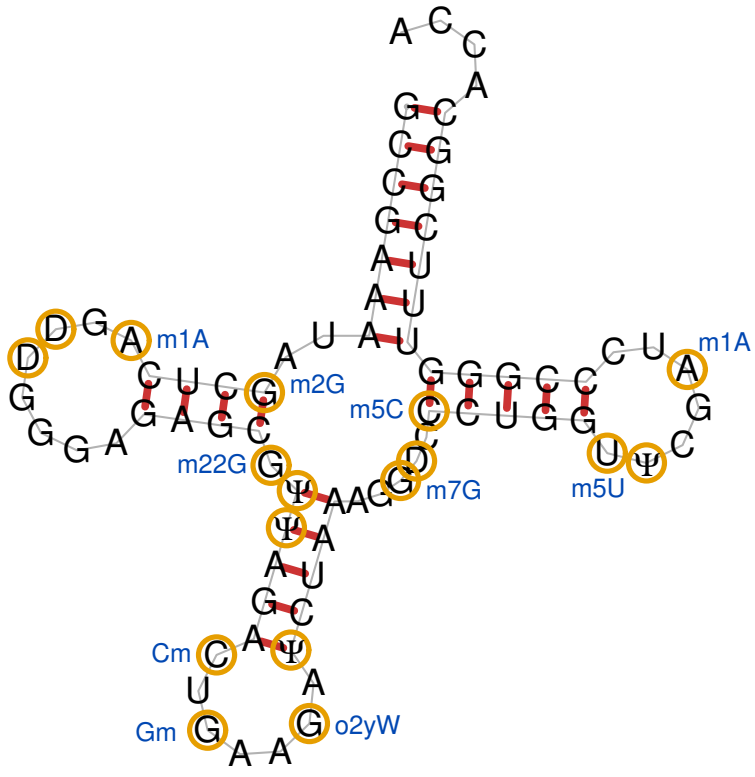
Example: human tRNA Phe (tdbR00000103)[6]



- 17 out of 76 nucleotides are modified
- Predicted ground state (MFE) is **not** cloverleaf structure

[6] *F. Jühling et al. "tRNAdb 2009: compilation of tRNA sequences and tRNA genes". In: Nucleic Acids Research 37.suppl_1 (2009), pp. D159–D162*

Example: human tRNA Phe (tdbR00000103)[6]



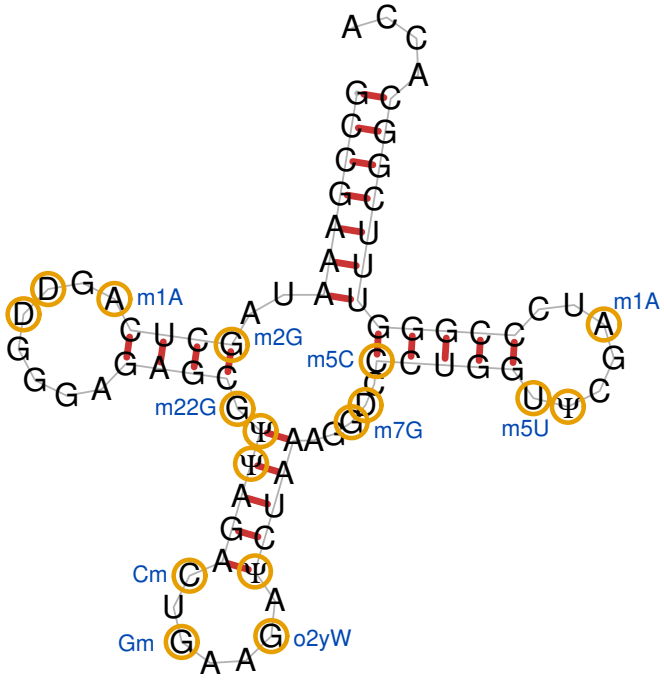
- 17 out of 76 nucleotides are modified
- Predicted ground state (MFE) is **not** cloverleaf structure

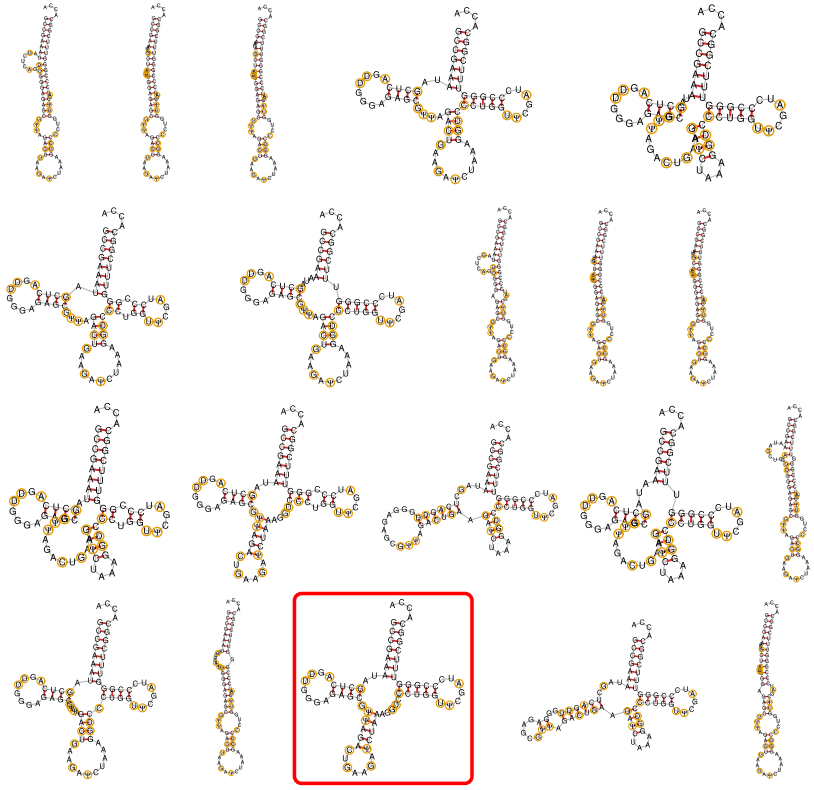
What about suboptimal structures $S^{\delta} = \{s \in \Omega \mid E(s) \leq E_{\text{MFE}} + \delta\}$?

[6] *F. Jühling et al. "tRNAdb 2009: compilation of tRNA sequences and tRNA genes". In: Nucleic Acids Research 37.suppl_1 (2009), pp. D159–D162*

Example: human tRNA Phe (tdbR00000103)

1. Suboptimals for *unmodified sequence*

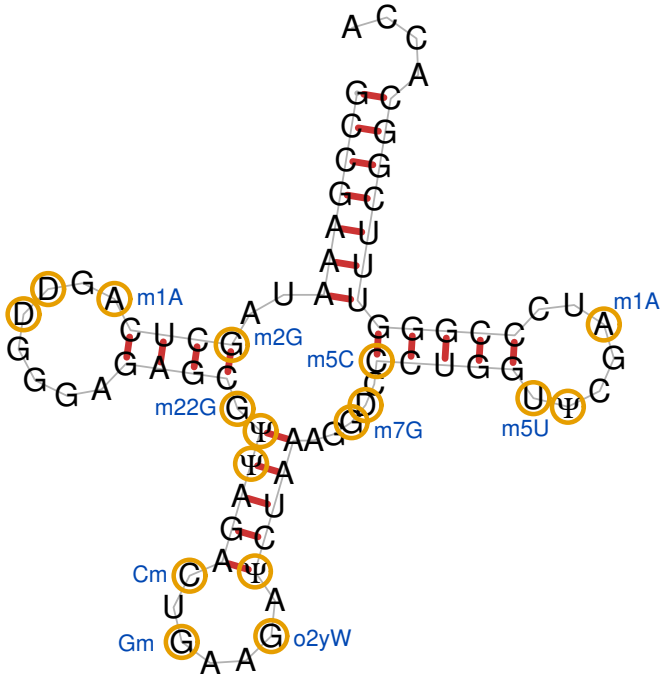


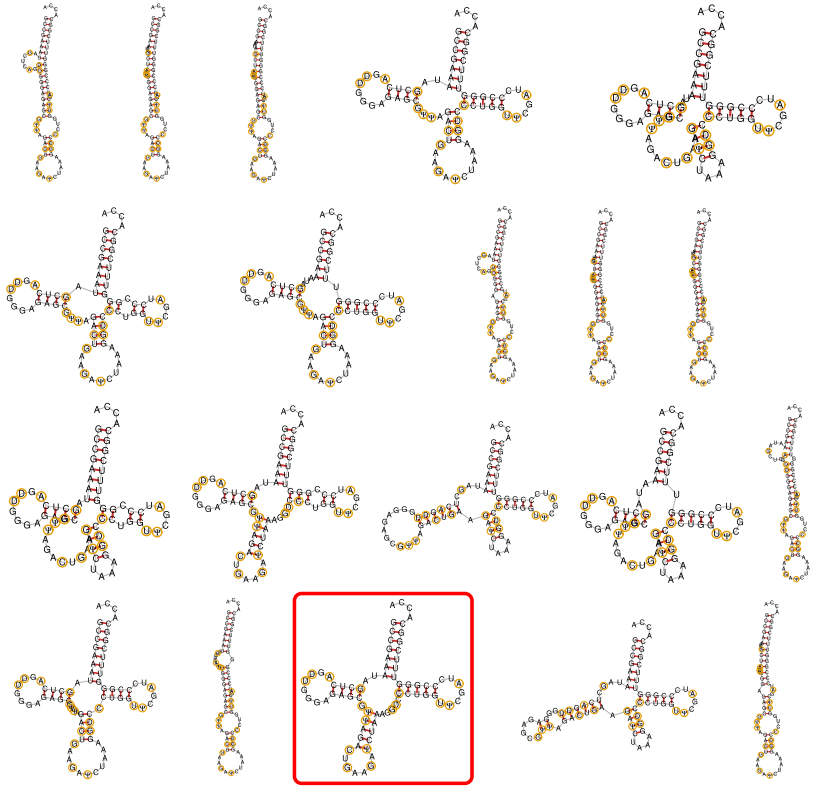


- Correct cloverleaf structure at position 18
- $\bar{d}^{3}_{\mathrm{BP}} = \frac{1}{|S^3|}\sum_{s\in S^3} d_{\mathrm{BP}}(s_{\mathrm{ref}}, s) = 21.55$ bp, $E_{\mathrm{ref}} - E_{\mathrm{MFE}} = 1.7$ kcal/mol
- Most predictions are not cloverleaf shaped

Example: human tRNA Phe (tdbR00000103)

1. Suboptimals for *unmodified sequence*



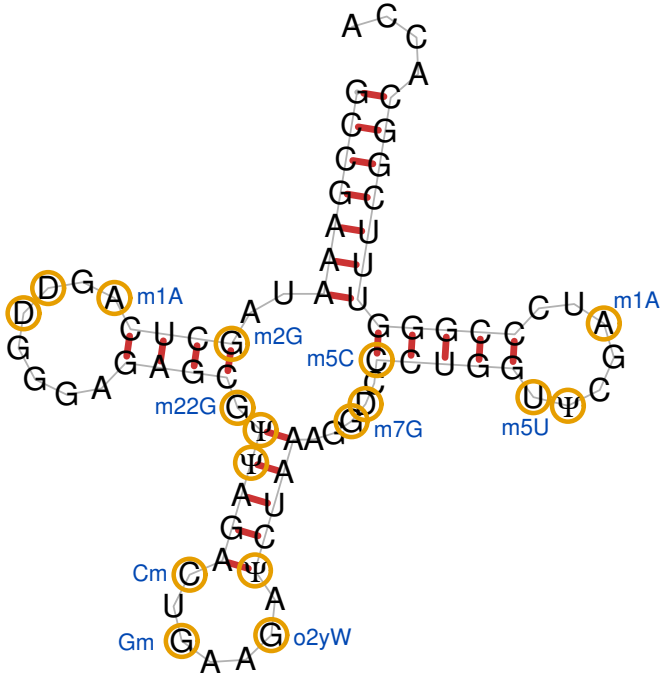


- Correct cloverleaf structure at position 18
- $\bar{d}^{3}_{\mathrm{BP}} = \frac{1}{|S^3|}\sum_{s\in S^3} d_{\mathrm{BP}}(s_{\mathrm{ref}}, s) = 21.55$ bp, $E_{\mathrm{ref}} - E_{\mathrm{MFE}} = 1.7$ kcal/mol
- Most predictions are not cloverleaf shaped

How about masking modified bases?

Example: human tRNA Phe (tdbR00000103)

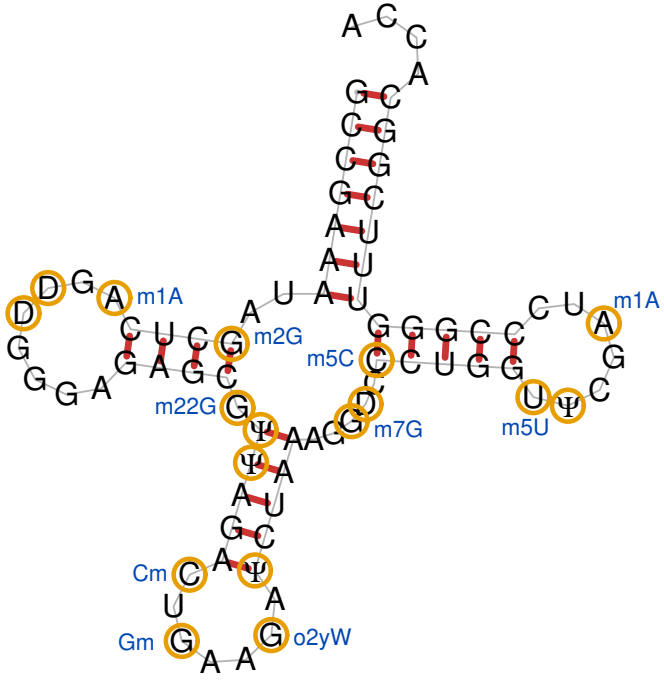
2. Suboptimals for sequence with *all modified bases masked*



- Modified bases are excluded from pairing
- Correct cloverleaf structure not among the solution set
- $\bar{d}^3_{\mathrm{BP}} = 16.7$ bp, anticodon arm is missing

Example: human tRNA Phe (tdbR00000103)

2. Suboptimals for sequence with *all modified bases masked*

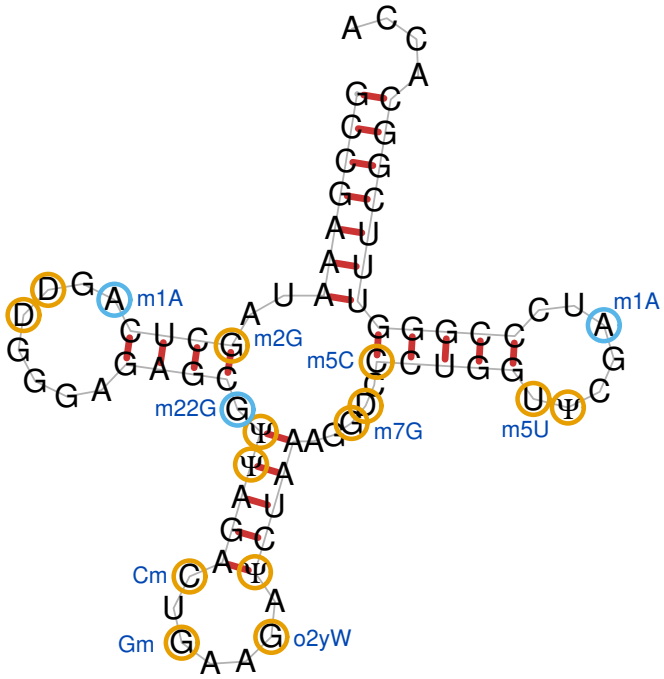


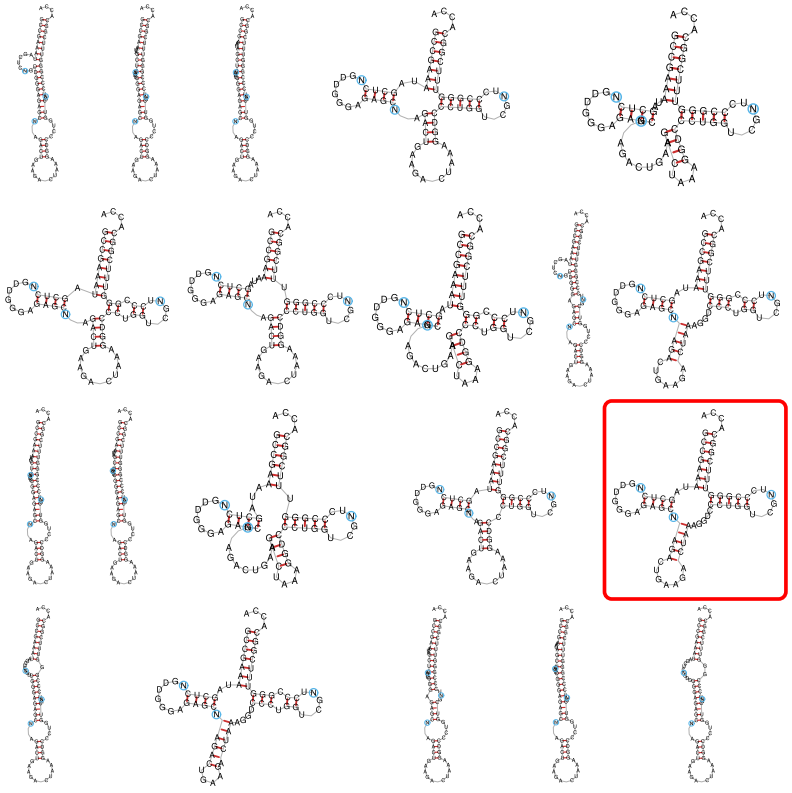
- Modified bases are excluded from pairing
- Correct cloverleaf structure not among the solution set
- $\bar{d}^3_{\mathrm{BP}} = 16.7$ bp, anticodon arm is missing

Some modified bases are required to pair!

Example: human tRNA Phe (tdbR00000103)

3. Suboptimals with *masked bases* for RT blocking modifications[7]



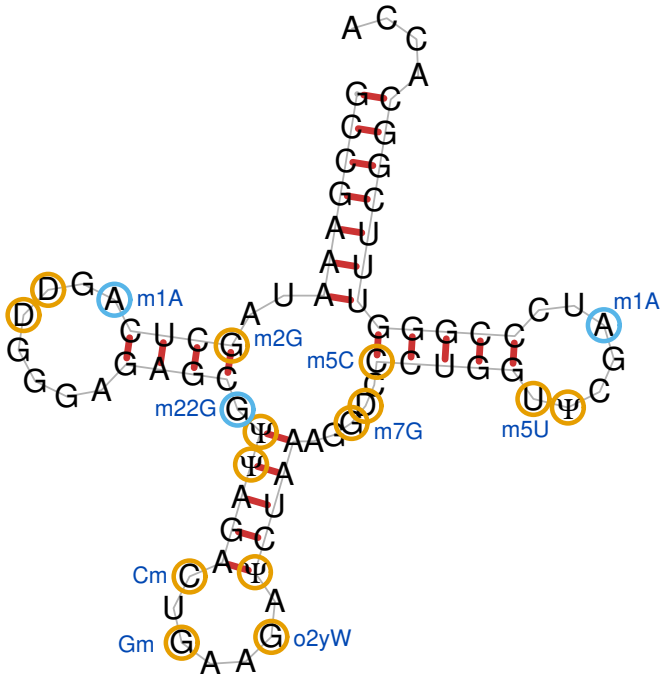


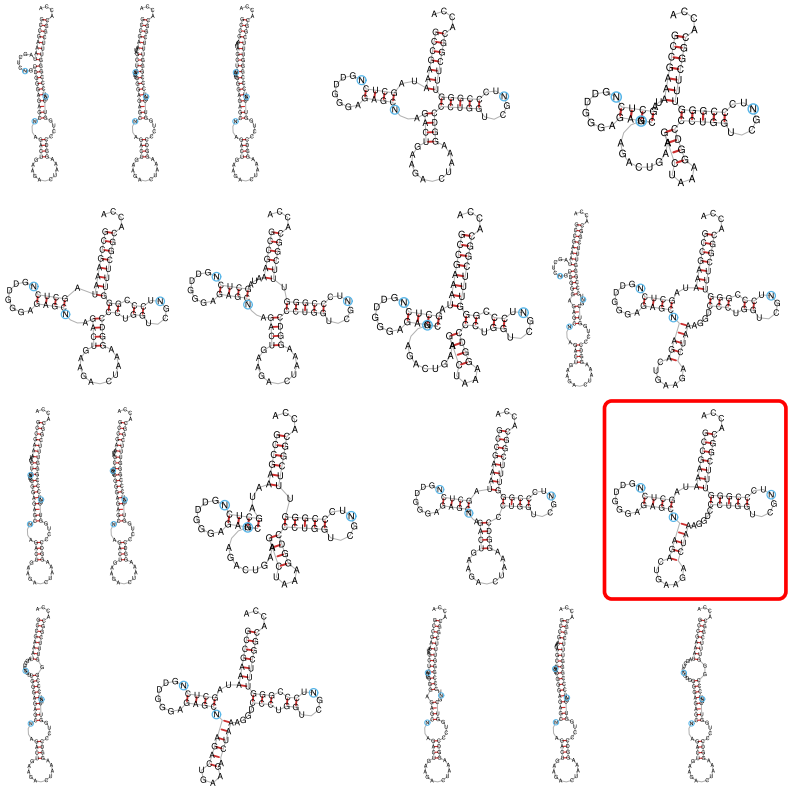
- Correct cloverleaf structure at position 15
- $\bar{d}^3_{\mathrm{BP}} = 19.7$ bp, $E_{\mathrm{ref}} - E_{\mathrm{MFE}} = 1.6$ kcal/mol
- Still, most predicted structure are not tRNA-like!

[7] *Y. Motorin et al. "Identification of modified residues in RNAs by reverse transcription-based methods". In: Methods in Enzymology 425 (2007), pp. 21–53.* DOI: 10.1016/S0076-6879(07)25002-5

Example: human tRNA Phe (tdbR00000103)

3. Suboptimals with *masked bases* for RT blocking modifications[7]





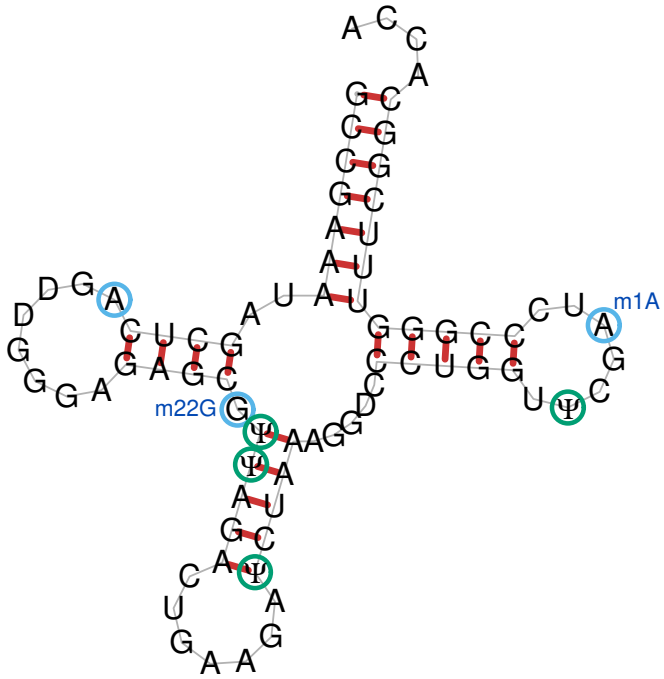
- Correct cloverleaf structure at position 15
- $\bar{d}^3_{\mathrm{BP}} = 19.7$ bp, $E_{\mathrm{ref}} - E_{\mathrm{MFE}} = 1.6$ kcal/mol
- Still, most predicted structure are not tRNA-like!

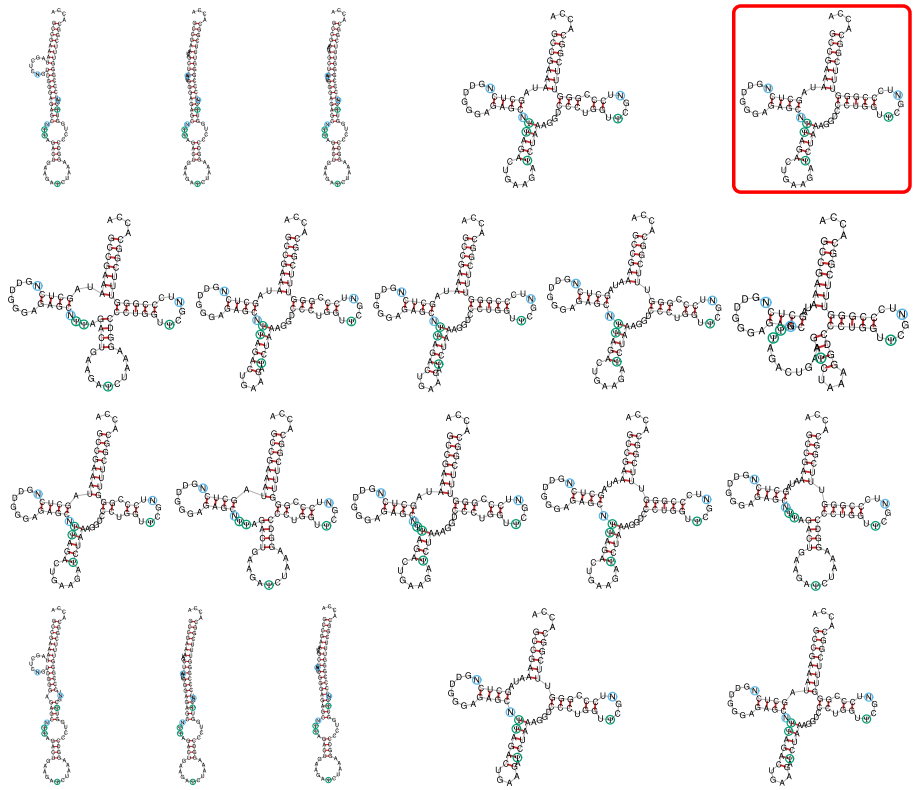
Ψ-A stacking energies are available in literature!

[7] *Y. Motorin et al. "Identification of modified residues in RNAs by reverse transcription-based methods". In: Methods in Enzymology 425 (2007), pp. 21–53.* DOI: 10.1016/S0076-6879(07)25002-5

Example: human tRNA Phe (tdbR00000103)

4. Suboptimals with RT masking and stacking energies for Ψ-A[8]



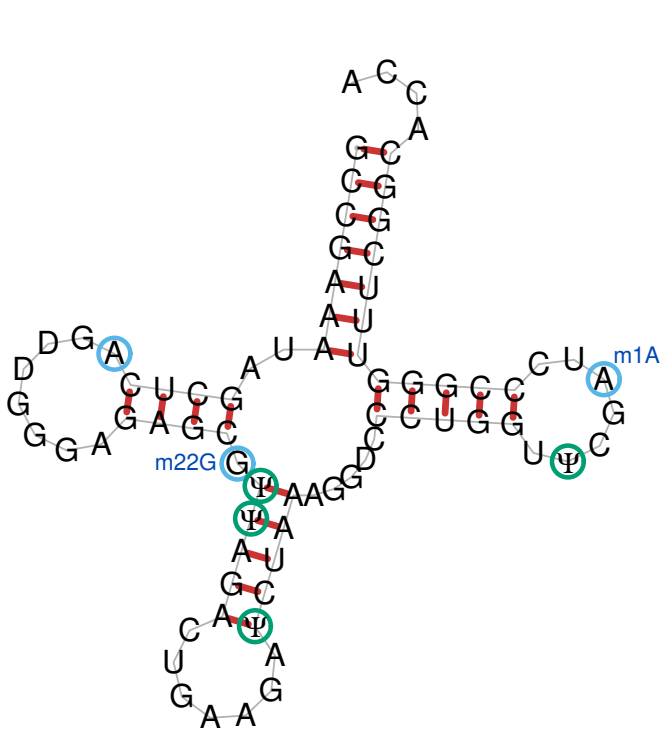


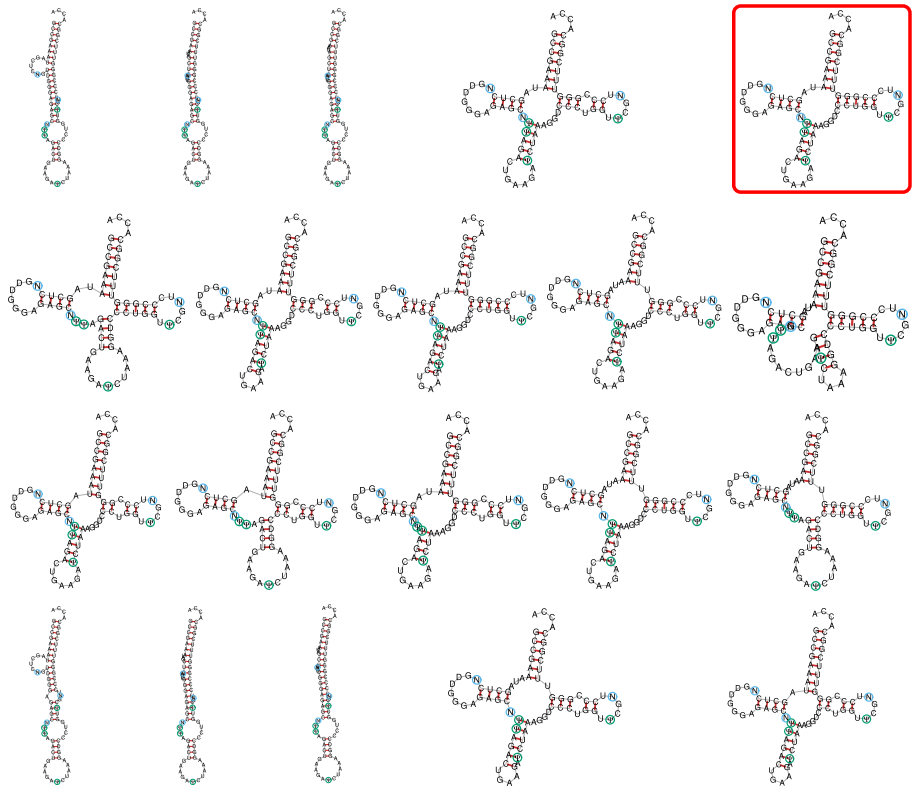
- Correct cloverleaf structure at position 5
- $\bar{d}^{3}_{\mathrm{BP}} = 15.5$ bp, $E_{\mathrm{ref}} - E_{\mathrm{MFE}} = 0.83$ kcal/mol
- Can we do better?

[8] *G. A. Hudson et al. "Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides". In: RNA 19.11 (2013), pp. 1474–1482.* DOI: 10.1261/rna.039610.113

Example: human tRNA Phe (tdbR00000103)

4. Suboptimals with RT masking and stacking energies for Ψ-A[8]





- Correct cloverleaf structure at position 5
- $\bar{d}^{3}_{\mathrm{BP}} = 15.5$ bp, $E_{\mathrm{ref}} - E_{\mathrm{MFE}} = 0.83$ kcal/mol
- Can we do better?

Dihydrouridine (D) adds flexibility to the D-loop!

[8] *G. A. Hudson et al. "Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides". In: RNA 19.11 (2013), pp. 1474–1482.* DOI: 10.1261/rna.039610.113

Example: human tRNA Phe (tdbR00000103)

No (stacking) energies with dihydrouridine (D) are available

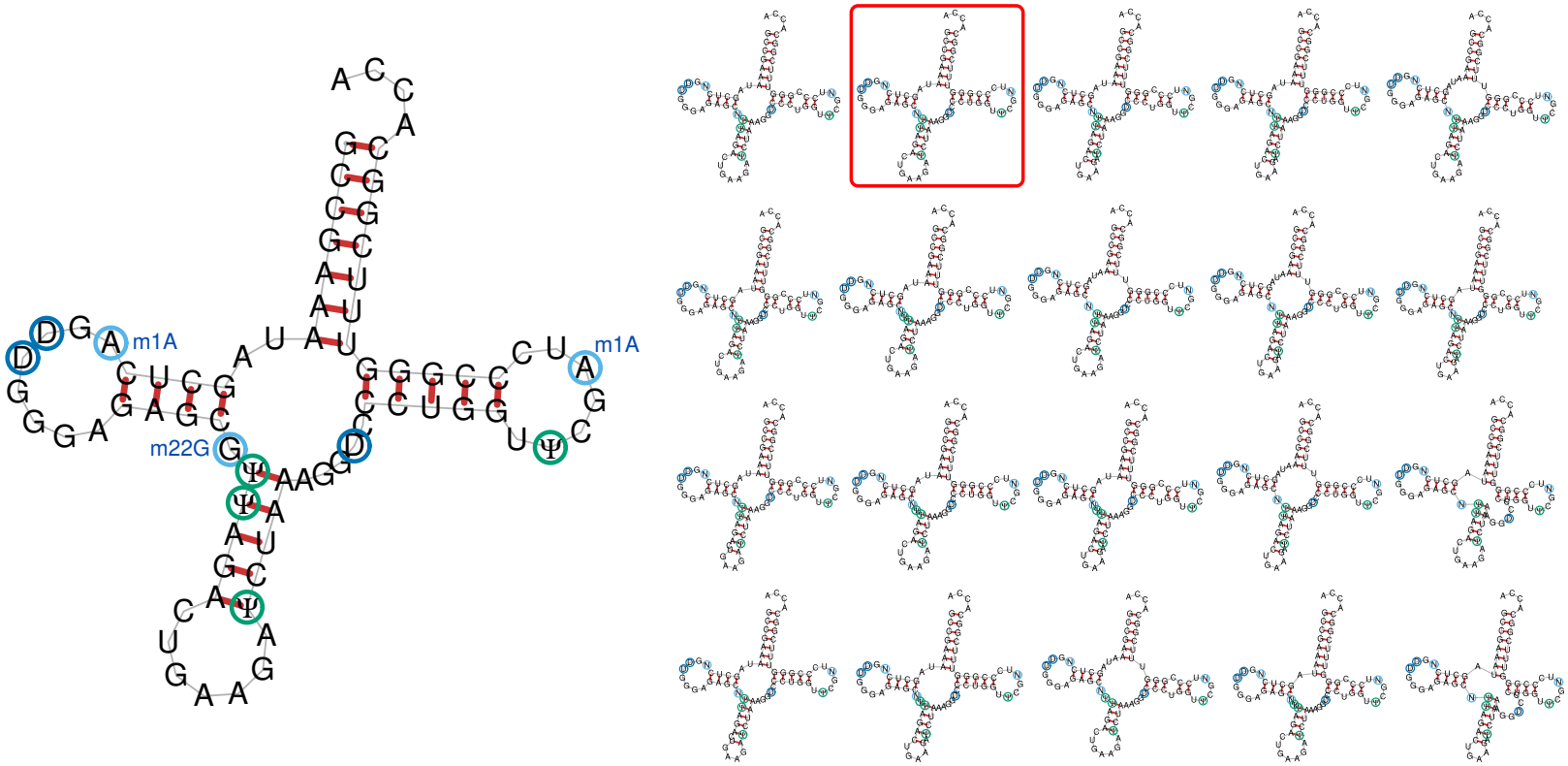
Structural effects of dihydrouridine[9]

- C3'-endo sugar conformation is destabilized in favor of C2'-endo
- more flexibility
- promotes destacking
- destabilization is somwhere in the range of 1.5 kcal/mol

[9] *J. J. Dalluge et al. "Conformational flexibility in RNA: the role of dihydrouridine".* In: Nucleic Acids Research *24.6 (1996), pp. 1073–1079.* DOI: *10.1093/nar/24.6.1073*

Example: human tRNA Phe (tdbR00000103)

5. Suboptimals with RT masking and energies for Ψ-A and D[9]

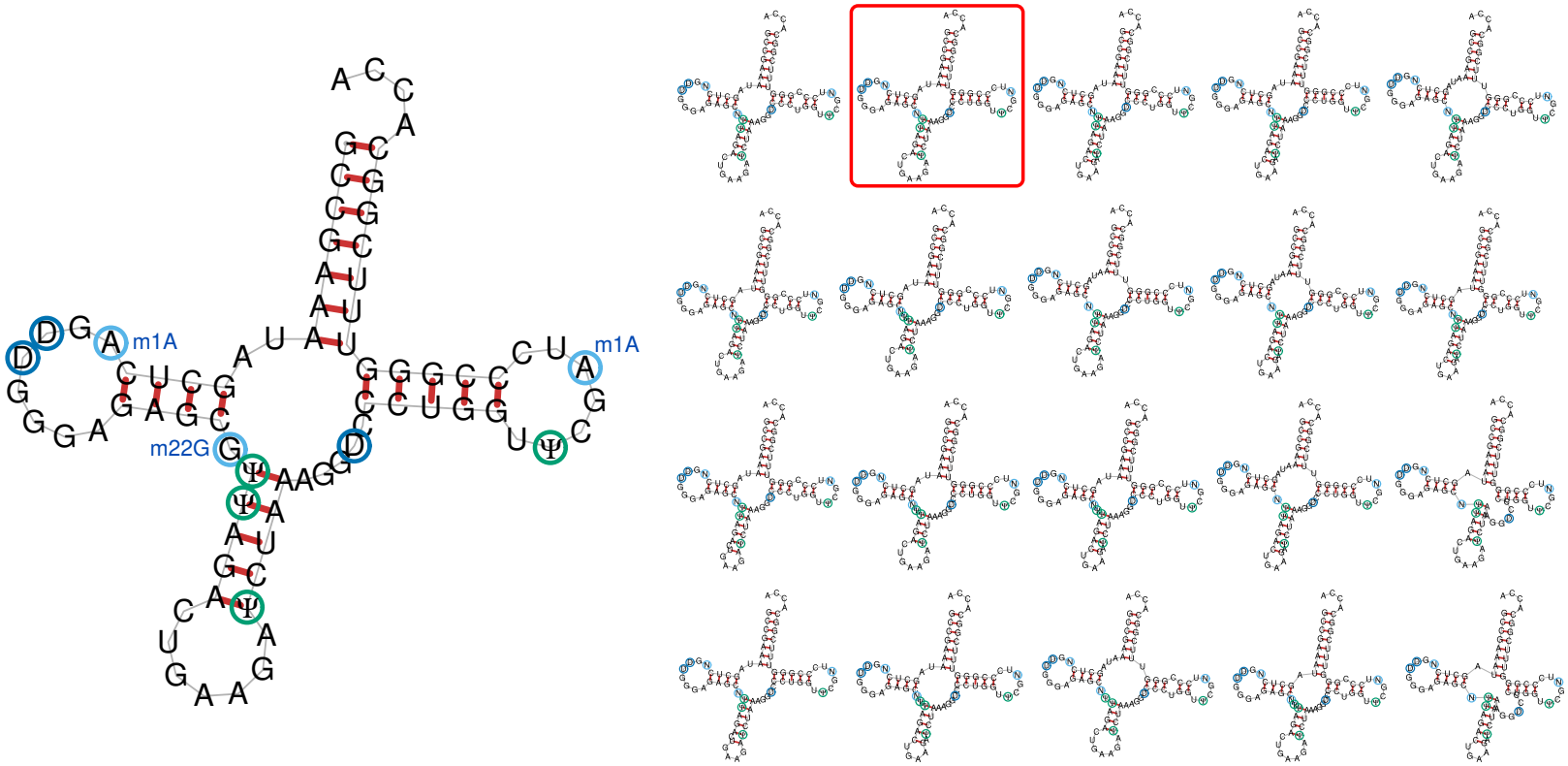


- Correct cloverleaf structure at position 2
- $\bar{d}^3_{\mathrm{BP}} = 6.92$ bp, $E_{\mathrm{ref}} - E_{\mathrm{MFE}} = 0.2$ kcal/mol

[9] *J. J. Dalluge et al. "Conformational flexibility in RNA: the role of dihydrouridine". In: Nucleic Acids Research 24.6 (1996), pp. 1073–1079.* DOI: 10.1093/nar/24.6.1073

Example: human tRNA Phe (tdbR00000103)

5. Suboptimals with RT masking and energies for Ψ-A and D[9]

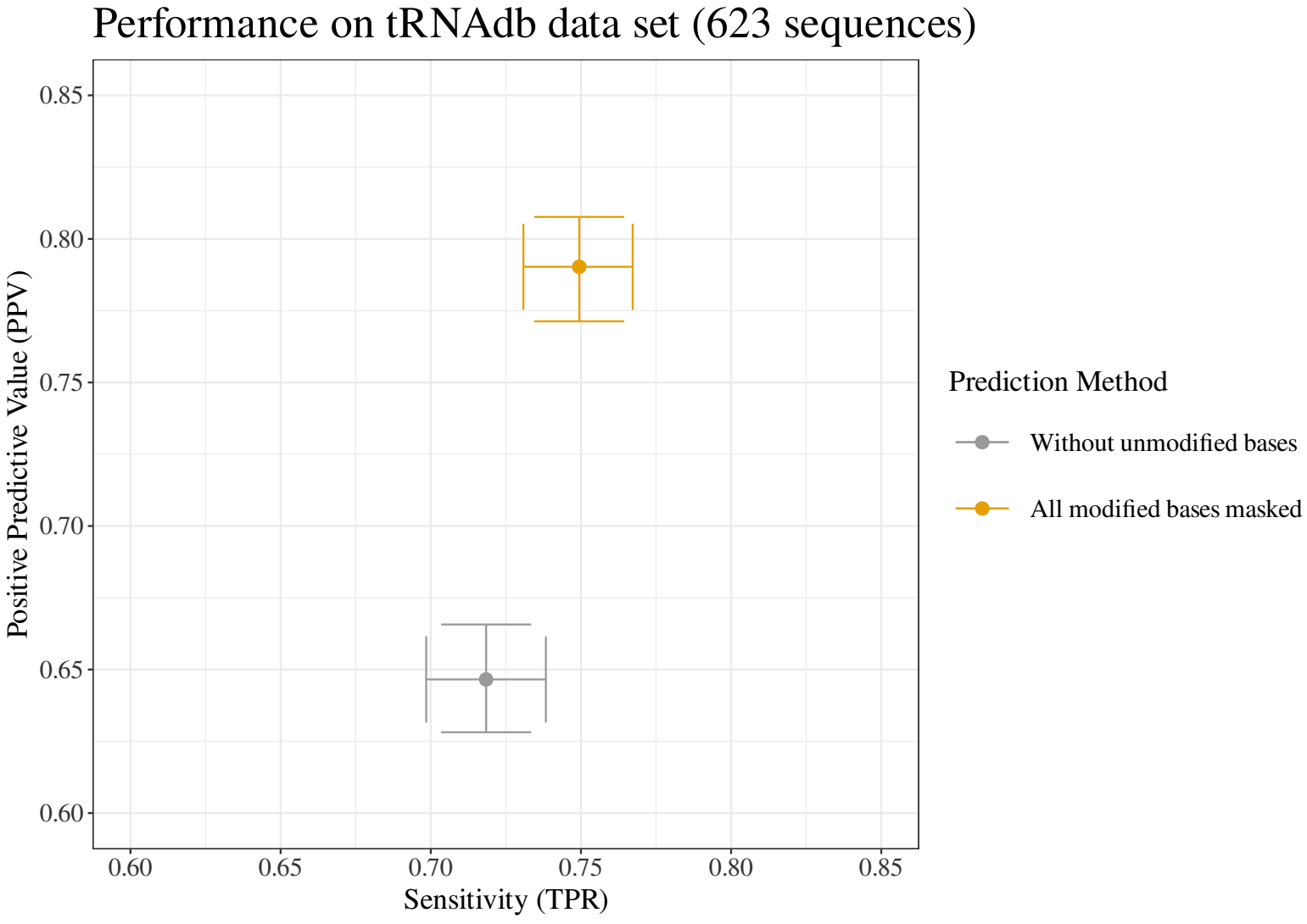


- Correct cloverleaf structure at position 2
- $\bar{d}^{3}_{\mathrm{BP}} = 6.92$ bp, $E_{\mathrm{ref}} - E_{\mathrm{MFE}} = 0.2$ kcal/mol

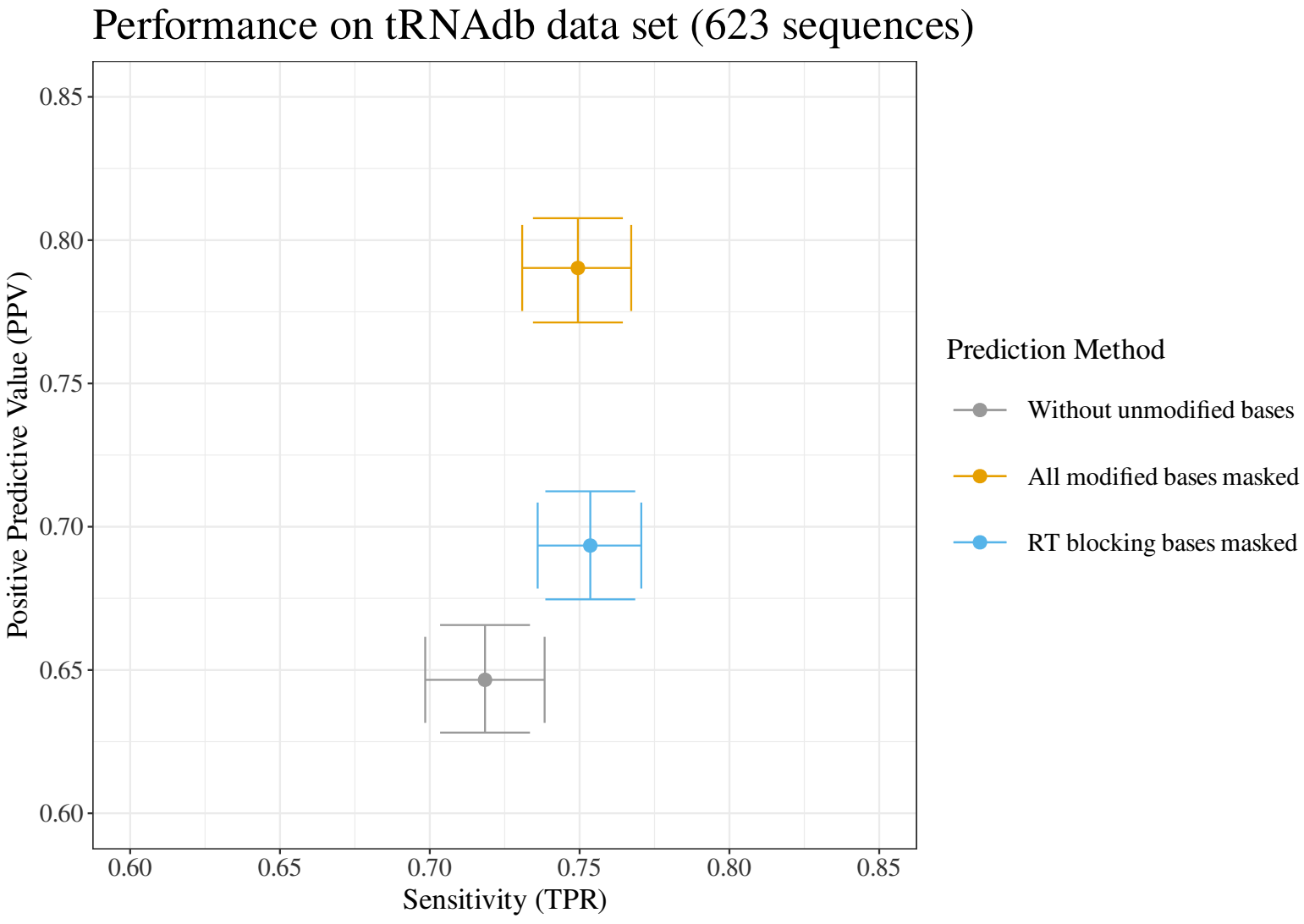
How do these constraints affect prediction for other tRNAs?

[9] *J. J. Dalluge et al. "Conformational flexibility in RNA: the role of dihydrouridine".* In: Nucleic Acids Research *24.6 (1996), pp. 1073–1079.* DOI: 10.1093/nar/24.6.1073

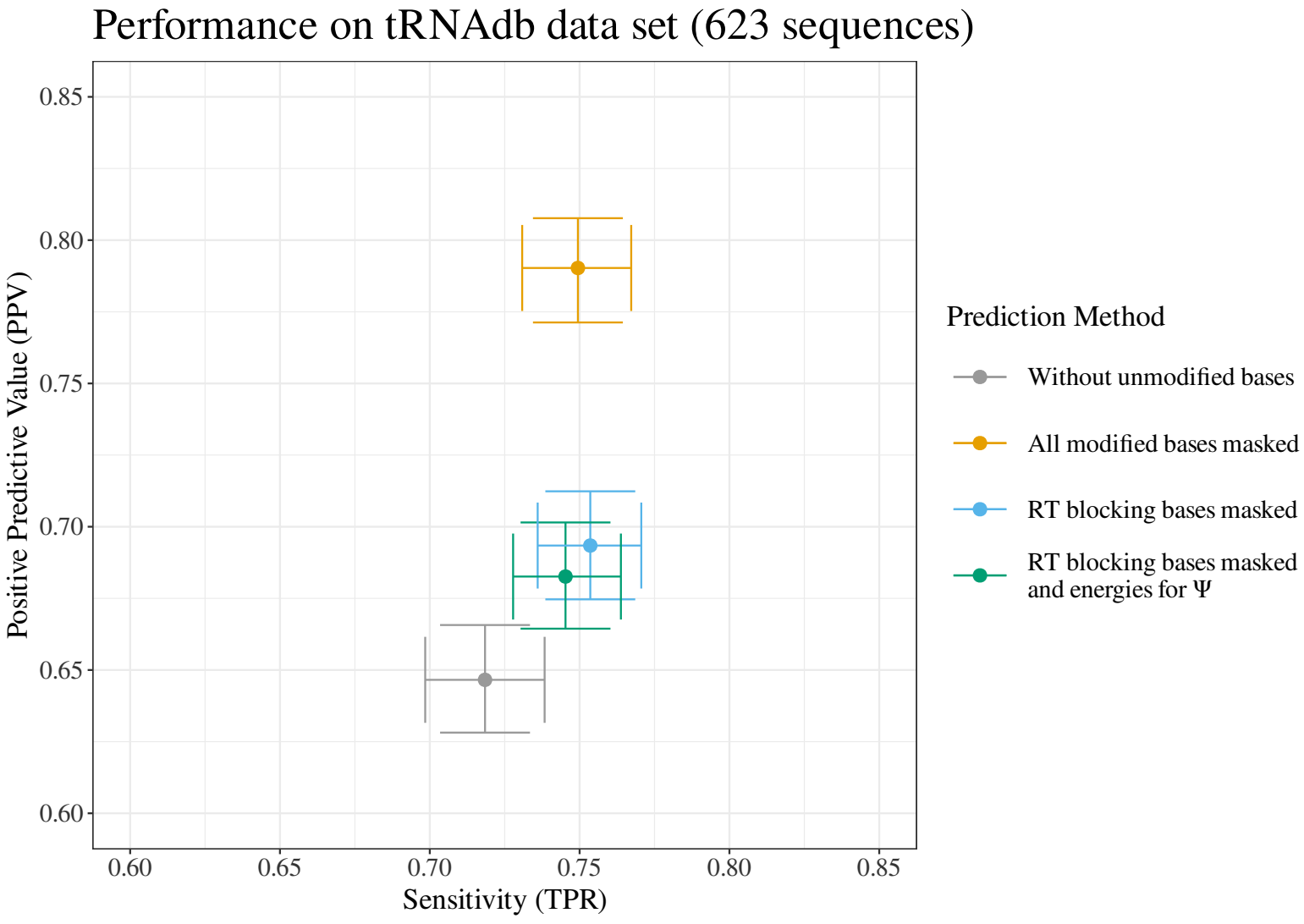
Overall prediction improvement



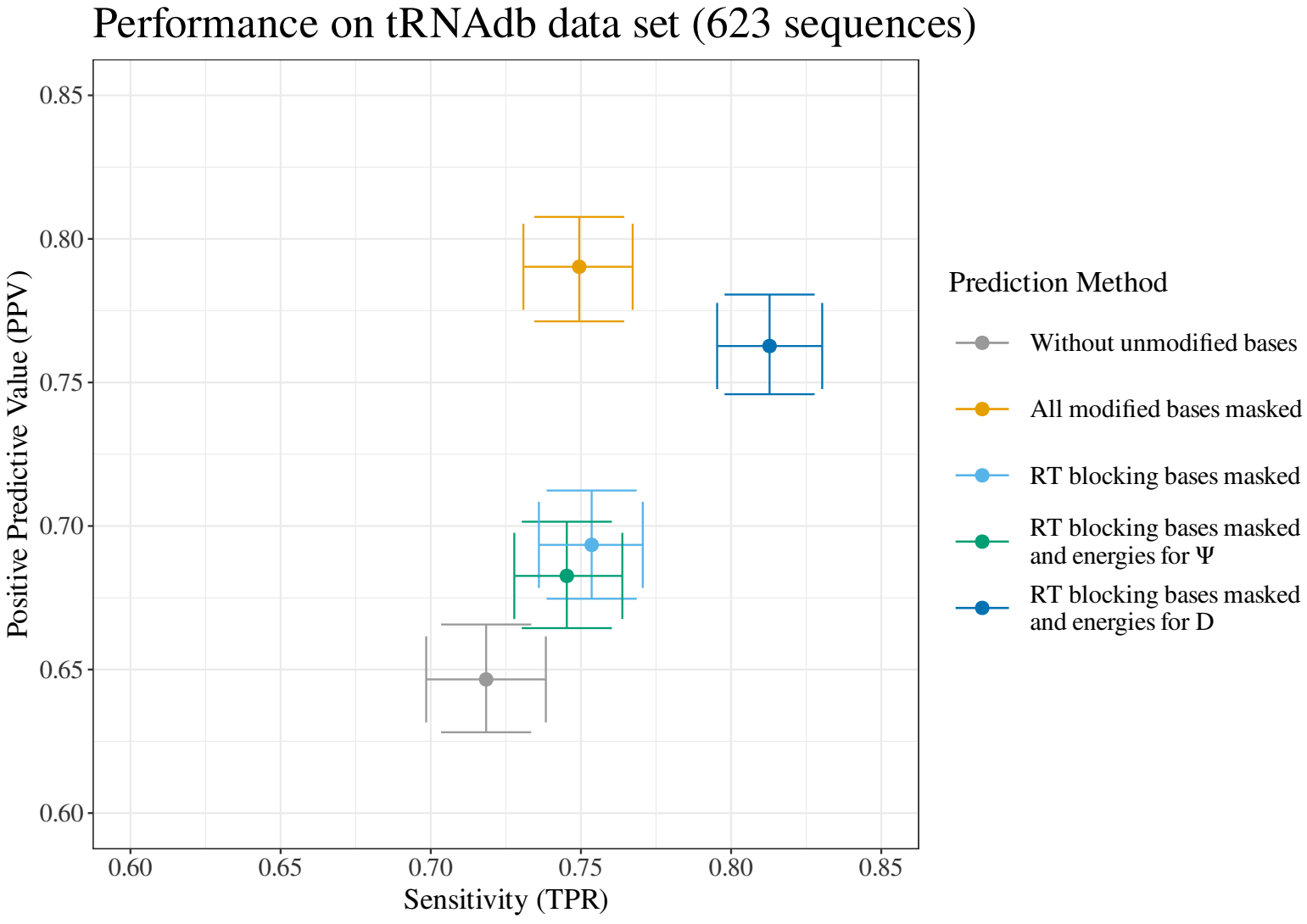
Overall prediction improvement



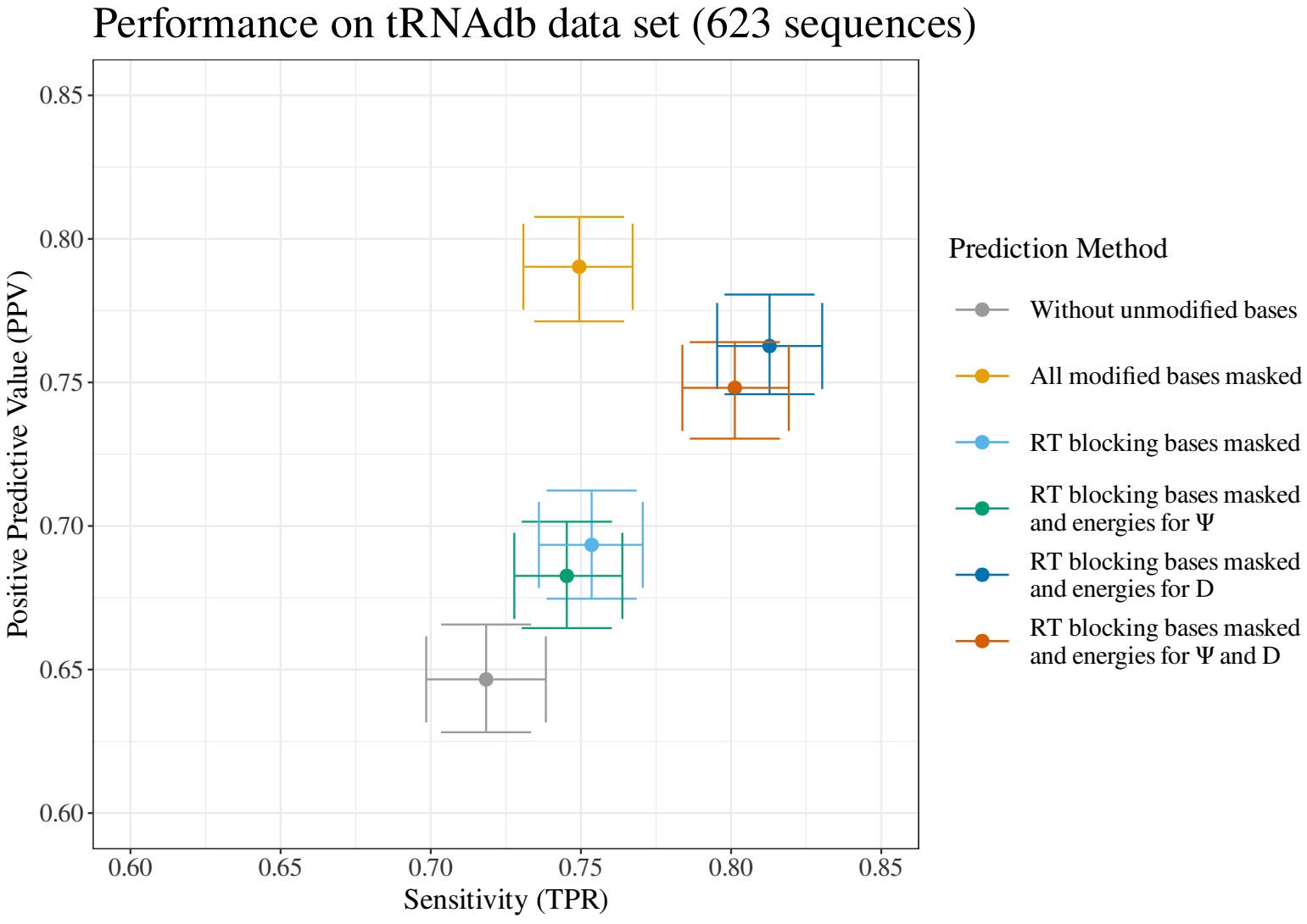
Overall prediction improvement



Overall prediction improvement



Overall prediction improvement



Secondary Structures with Modified Bases

Implemented into the ViennaRNA Package[10]:

- Includes stacking parameters:

One-letter-code	Modified base
7	7-deaza-adenonsine (7DA)
I	Inosine
6	N6-methyladenosine (m^6A)
P	Pseudouridine
9	Purine (a.k.a. nebularine)
D	Dihydrouridine (Rosetta-RECESS)

- Extension for arbitrarily many additional modifications
- Available from command line and API
- Easy extension with more data using JSON parameter files (stacking, terminal, mismatch, dangles, fallback, pairing partners)

[10] *Y. Varenyk et al. "Modified RNAs and predictions with the ViennaRNA Package". In: Bioinformatics 39.11 (2023), btad696.* DOI: *10.1093/bioinformatics/btad696*

Secondary Structures with Modified Bases

Implemented into the ViennaRNA Package[10]:

- Includes stacking parameters:

One-letter-code	Modified base
7	7-deaza-adenonsine (7DA)
I	Inosine
6	N6-methyladenosine (m^6A)
P	Pseudouridine
9	Purine (a.k.a. nebularine)
D	Dihydrouridine (Rosetta-RECESS)

- Extension for arbitrarily many additional modifications
- Available from command line and API
- Easy extension with more data using JSON parameter files (stacking, terminal, mismatch, dangles, fallback, pairing partners)

Sometimes including modified base yields worse predictions!

[10] *Y. Varenyk et al. "Modified RNAs and predictions with the ViennaRNA Package". In: Bioinformatics 39.11 (2023), btad696.* DOI: *10.1093/bioinformatics/btad696*

Outlook

Typically parameters are obtained from UV-melting experiments (expensive)

Outlook

Typically parameters are obtained from UV-melting experiments (expensive)

Parameters:

- Add (more) in-silico predicted parameters[11][12]
- Add parameters obtained from flourescence melting/titration[13]
- Add modifications known to stay unpaired (literature search, experts knowledge)

Implementation:

- Comparative structure prediction
- Multiple interacting RNAs
- Pairs between different modified bases

---

[11] *F.-C. Chou et al. "Blind tests of RNA nearest-neighbor energy prediction". In: Proceedings of the National Academy of Sciences 113.30 (2016), pp. 8430–8435.* DOI: *10.1073/pnas.152333511*

[12] *M. C. Hopfinger et al. "Predictions and analyses of RNA nearest neighbor parameters for modified nucleotides". In: Nucleic Acids Research 48.16 (2020), pp. 8901–8913.* DOI: *10.1093/nar/gkaa654*

[13] *J. Wang et al. "Assessment for melting temperature measurement of nucleic acid by HRM". In: Journal of analytical methods in chemistry 2016.1 (2016), p. 5318935.* DOI: *10.1155/2016/5318935*

Acknowledgements

TBI Vienna

- Yuliia Varenyk
- Thomas Spicher
- Ivo L. Hofacker

MedUni Vienna

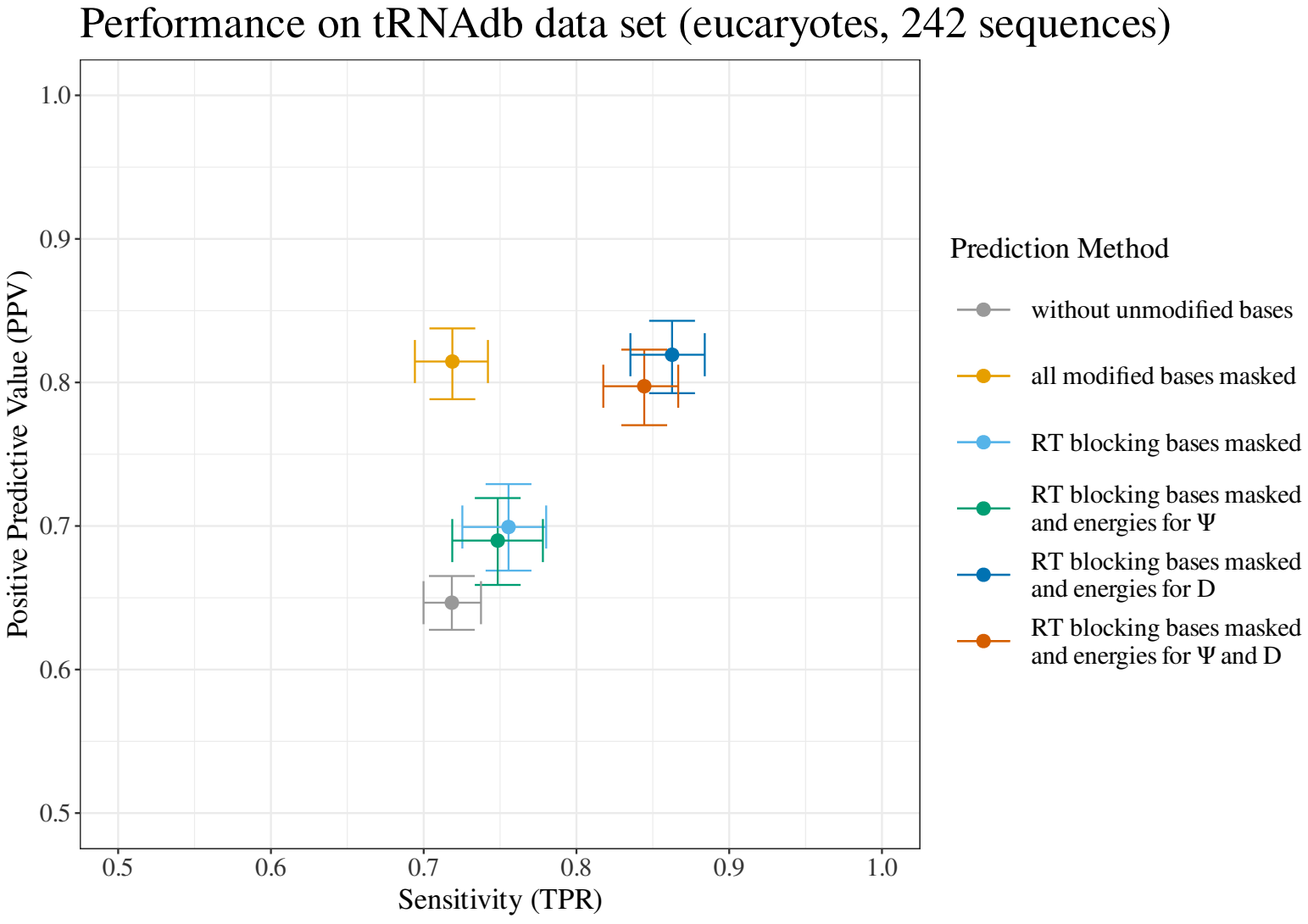
- Michael Jantsch
- Hamid Reza Mansouri Khosravi

Uni Leipzig

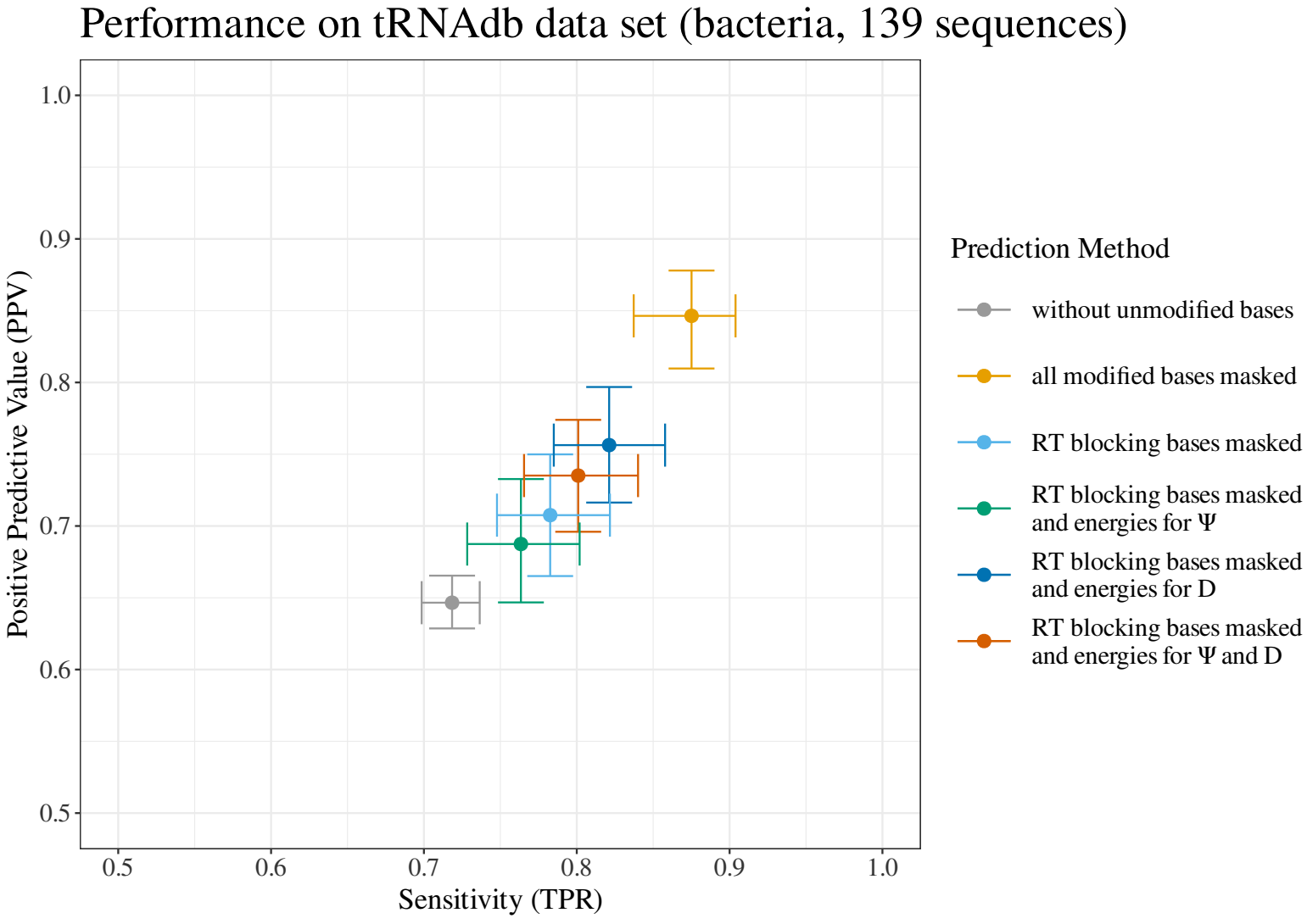
- Peter F. Stadler + group

Thank You for your attention!

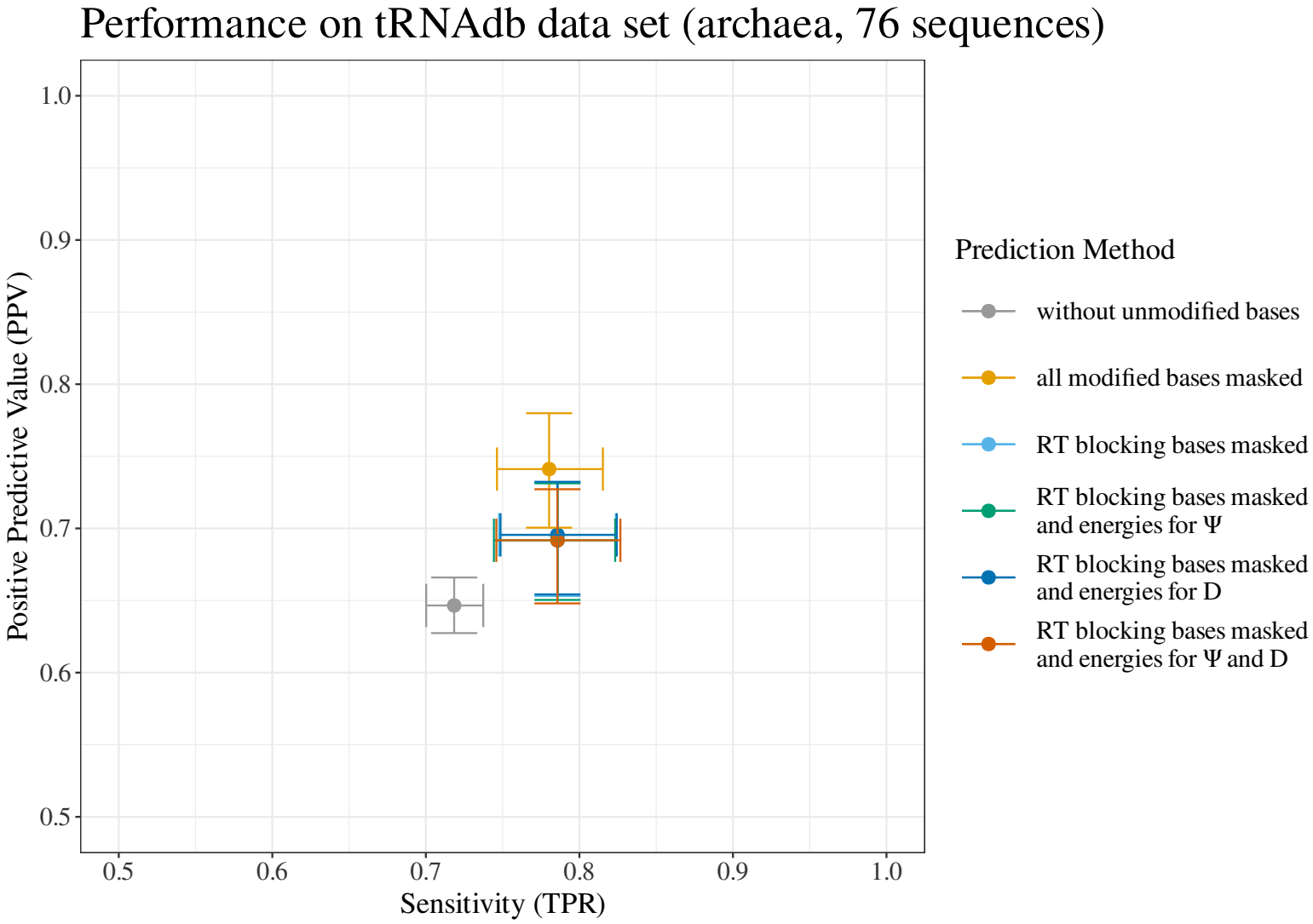
tRNA Secondary Structure Prediction



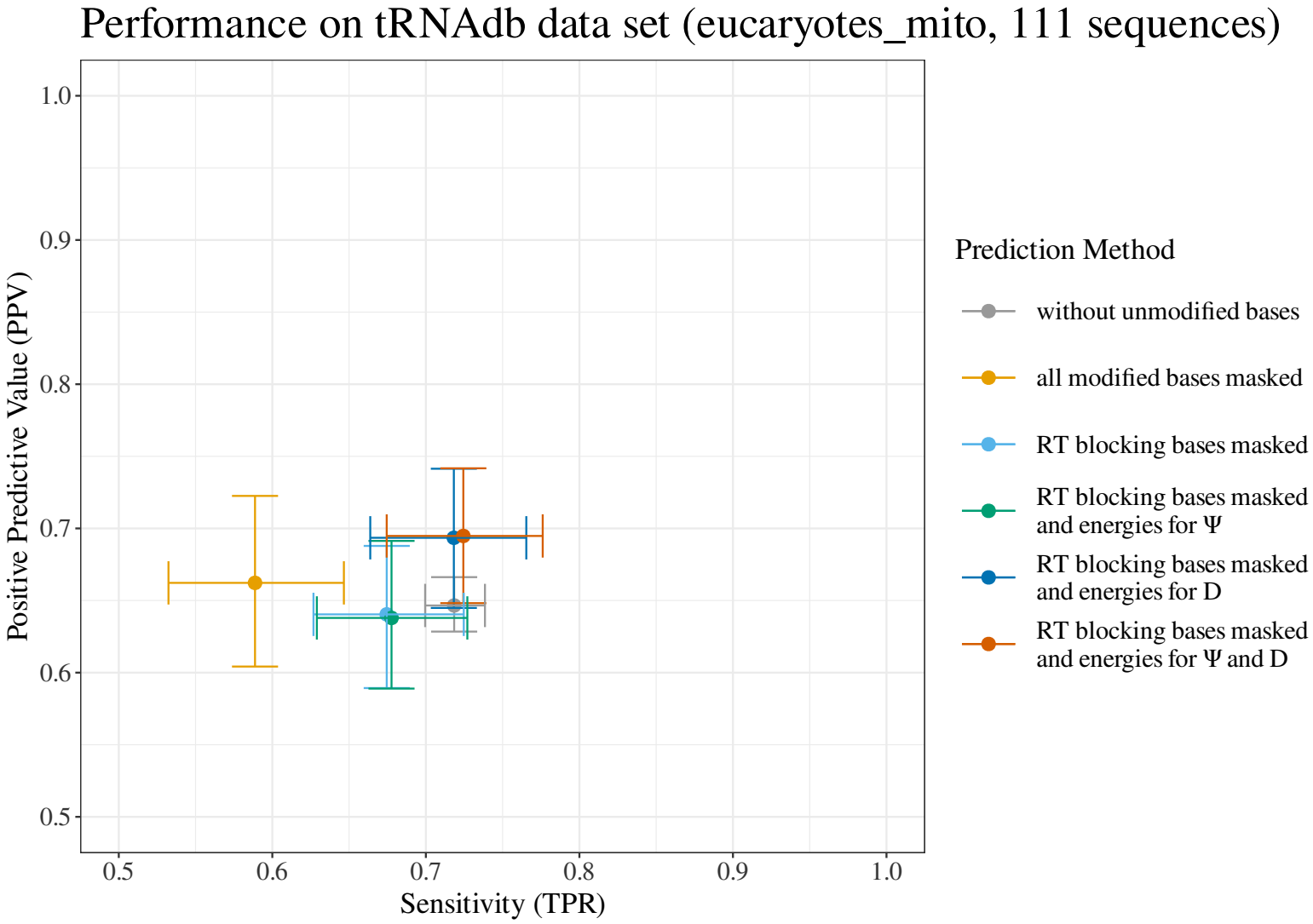
tRNA Secondary Structure Prediction



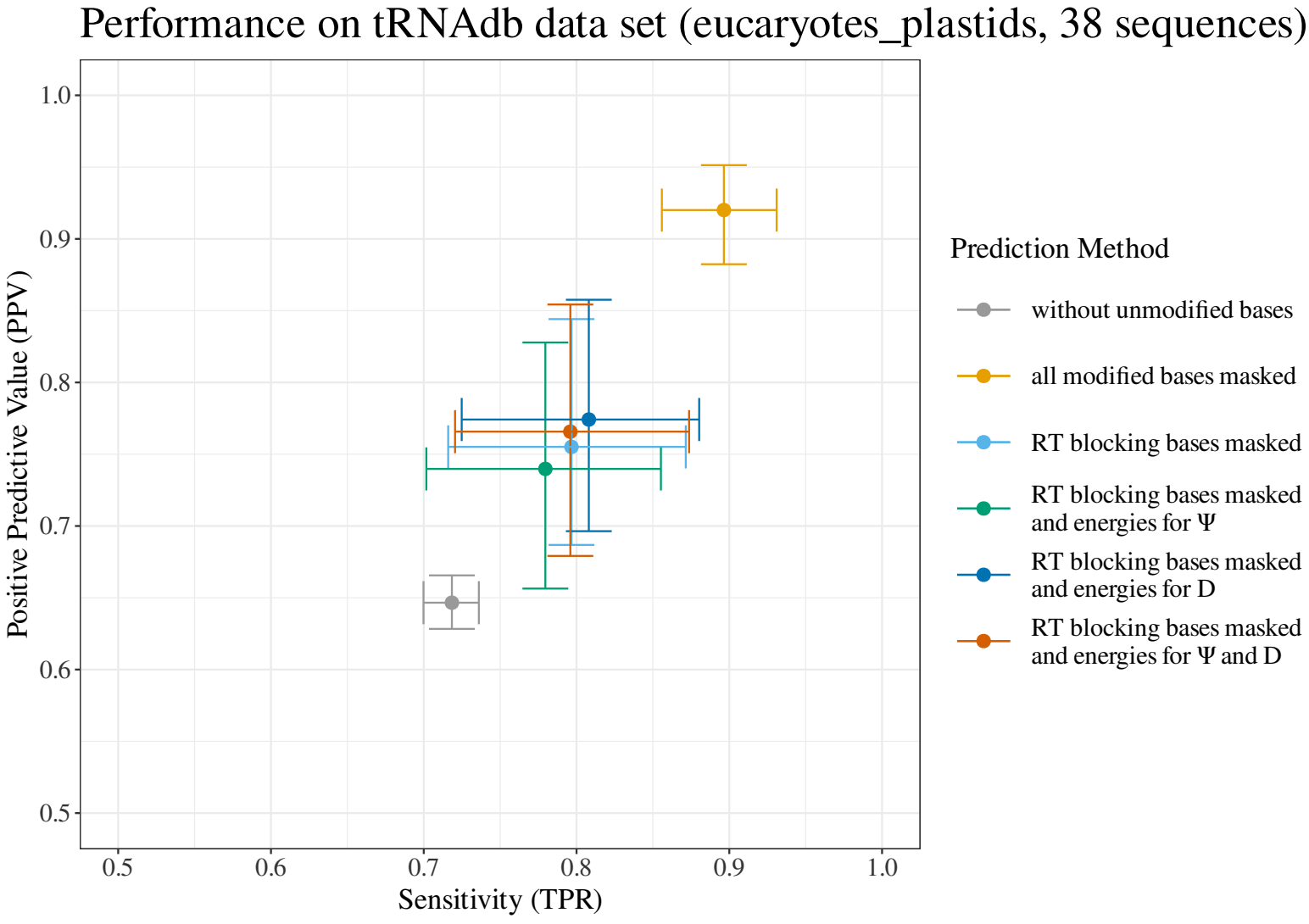
tRNA Secondary Structure Prediction



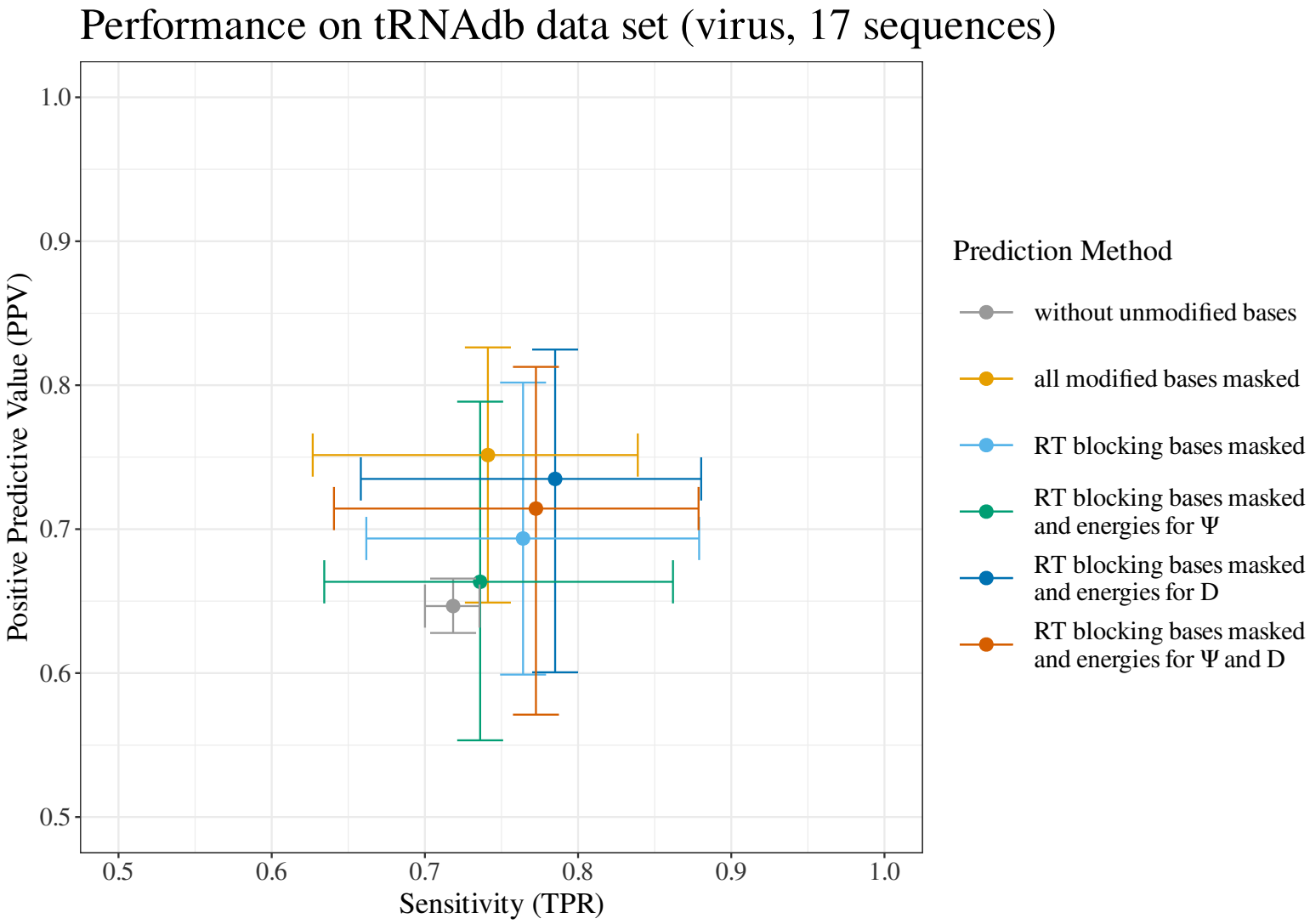
tRNA Secondary Structure Prediction



tRNA Secondary Structure Prediction

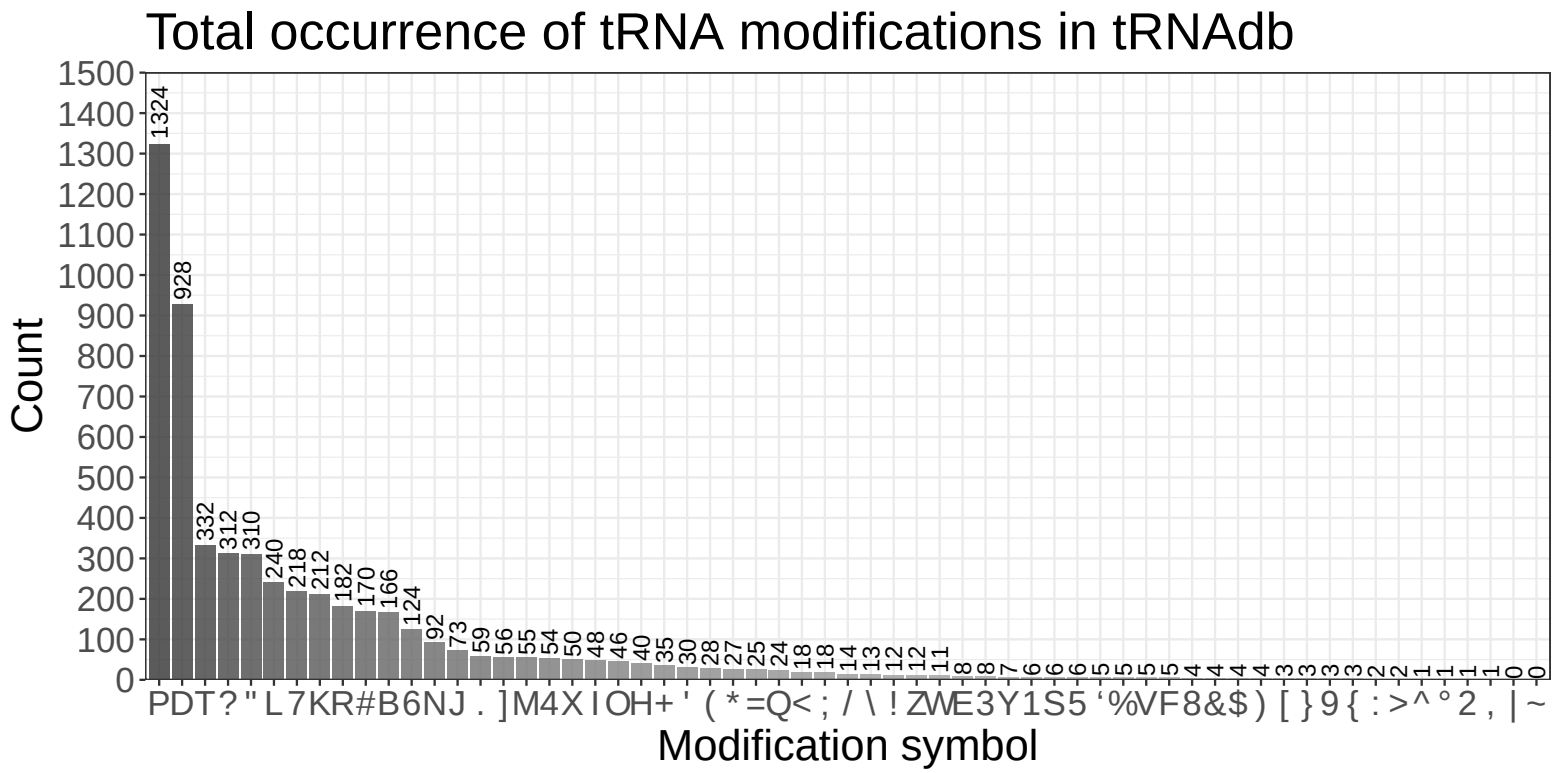


tRNA Secondary Structure Prediction



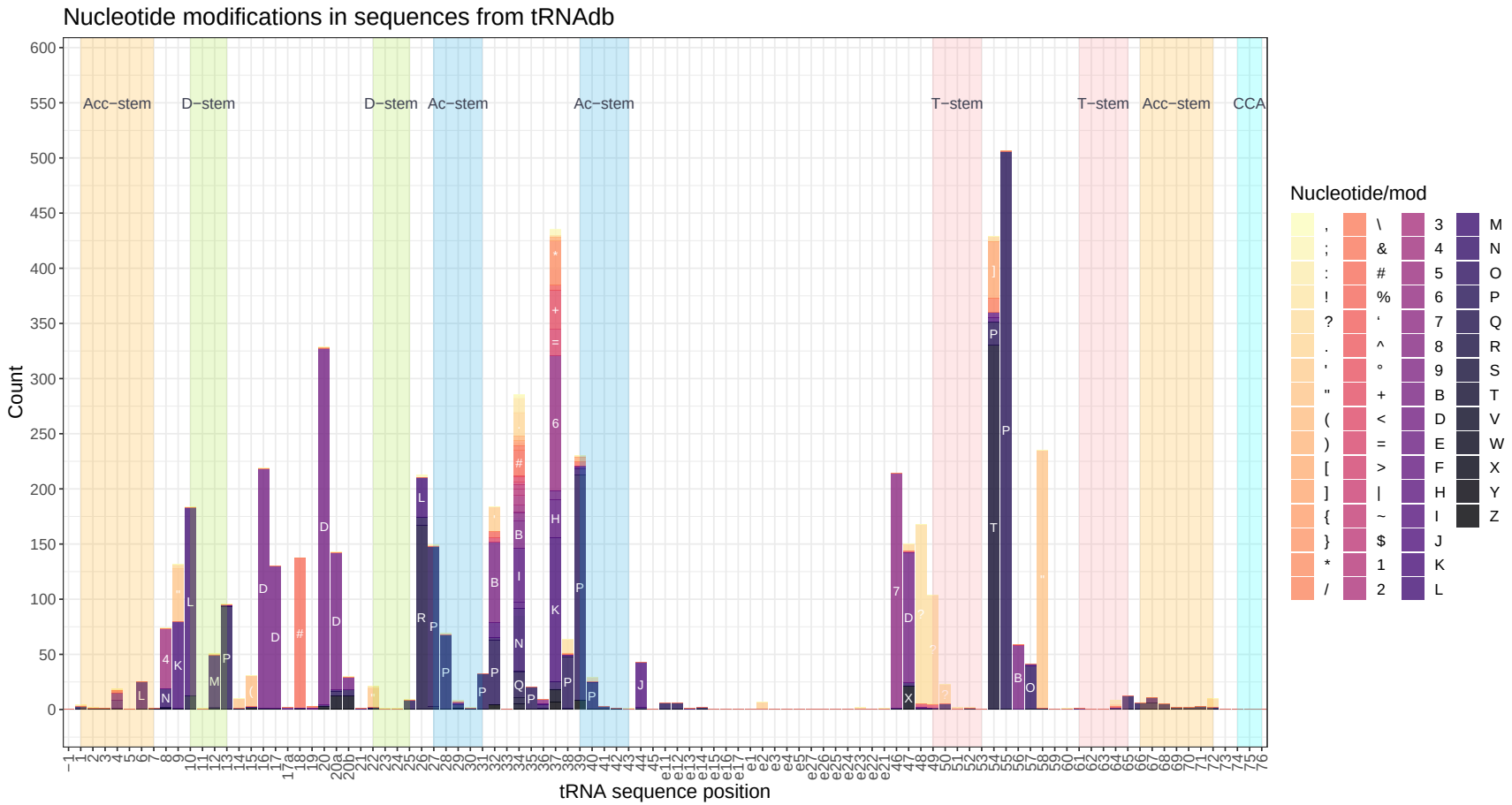
Modifications in tRNA

- Frequency of modifications in `tRNAdb`



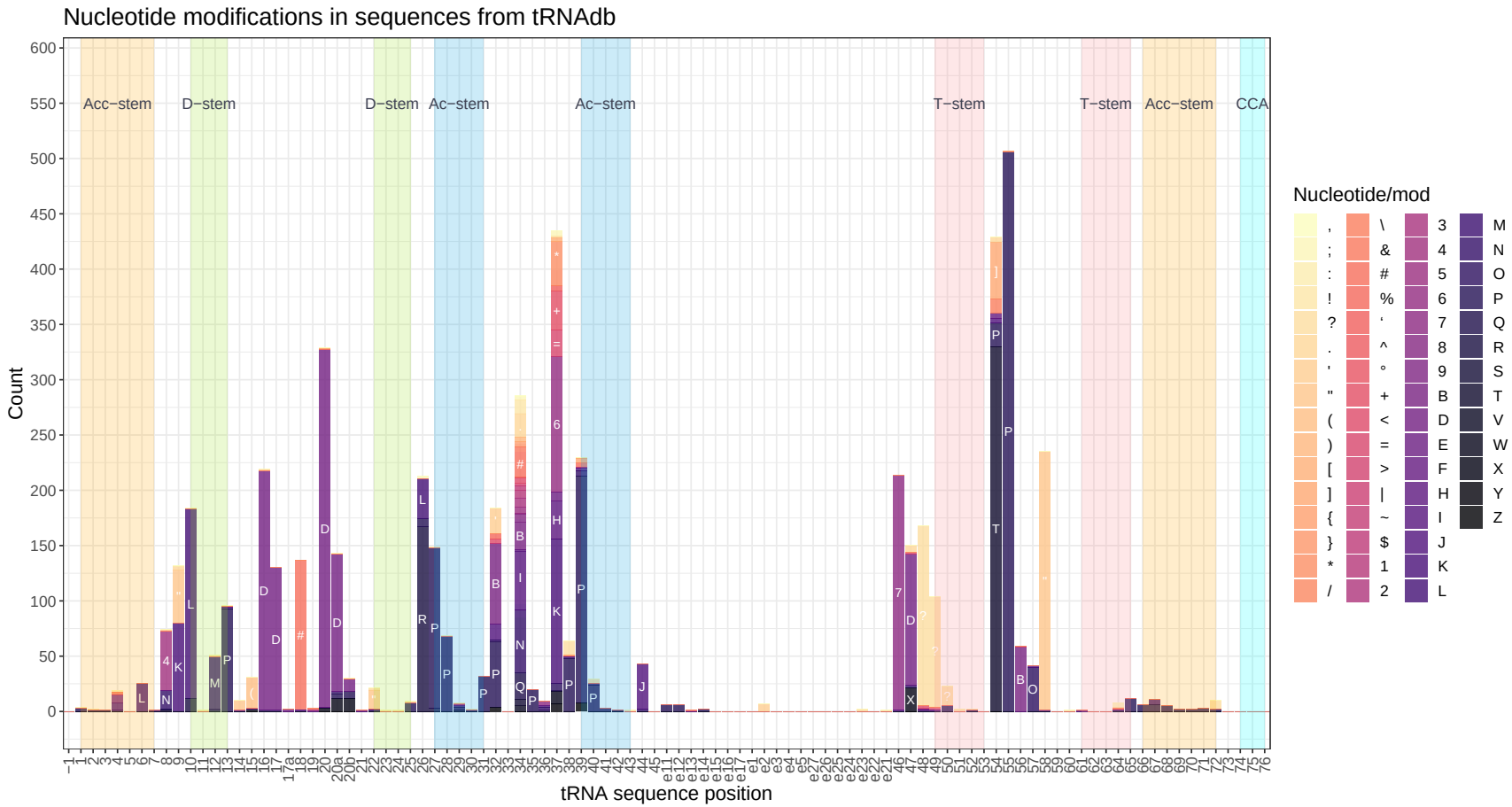
Modifications in tRNA

- Frequency of modifications in `tRNAdb`
- Which modifications can be found where?



Modifications in tRNA

- Frequency of modifications in `tRNAdb`
- Which modifications can be found where?
- Which modifications might induce structural rearrangements?



Nearest Neighbor Parameters for Modified Bases

Base pair stacking and helix ends

- I-U (Wright et al. 2007, partial, UV-melting)
- I-U (Jolley et al. 2015, partial, in-silico)
- I-C, D-U, iG-iC (Chou et al. 2016, partial, in-silico)
- I-C, iG-iC, D-U, I-U (Hopfinger et al. 2020, partial, in-silico)
- I-C (Wright et al. 2018, partial, UV-melting)
- Ψ-A (Hudson et al. 2013, partial, UV-melting)
- Ψ-A (Deb et al. 2019, partial, in-silico, NMR)
- Ψ-$\{A, U, C, G\}$ (Kierzek et al. 2014, partial, UV-melting, CD, NMR)
- m^6A-U (Roost et al. 2015, partial, UV-melting, NMR)
- modified U-$\{A, U, C, G\}$, I-$\{A, C, U\}$ (Vendeix et al. 2009, partial, in-silico)
- P-U (Jolley et al. 2017, partial, UV-melting)
- 7DA-U (Richardson et al. 2016, partial, UV-melting)

Other data on structural effects

- Dihydrouridine (Dalluge et al. 1996, Dyubankova et al. 2015, NMR)
- 2-thiouridine (s^2U), Ψ, dihydrouridine (Sipa et al. 2007, partial, CD)
- tRNA modifications (Lorenz C. et al. 2017, meta)
- m^6A, m^1A, m^5C, Ψ, I (Harcourt et al. 2017, meta)