

Advances in RNA Structure Prediction

The current state and (near) future of the ViennaRNA Package

Ronny Lorenz

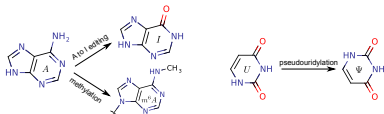
University of Vienna, Theoretical Biochemistry Group (TBI)

Vienna, Austria, February 22, 2021

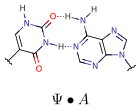
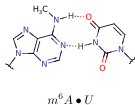
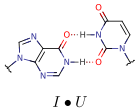
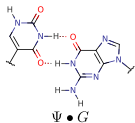
Modified Bases and Base Pairs

Modified Bases in RNA

Post-transcriptional RNA modifications (epitranscriptome):



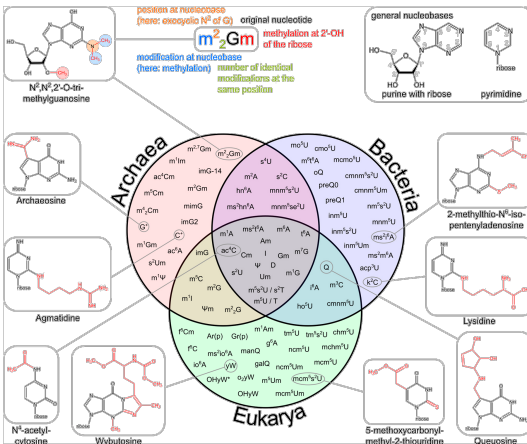
- Modomics Database¹ lists 172 different modified bases
- Commonly know modifications: *I*, Ψ , m^6A , m^1A , m^5C , ...
- Function and purpose of modifications still largely unknown
- Structural effects of base modifications:
 - correct folding of ncRNAs into functional structures (tRNA, rRNA, etc.)
 - 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



¹Boccaletto et al., "MODOMICS: a database of RNA modification pathways. 2017 update.", 2018, NAR 46.D1, D303–D307

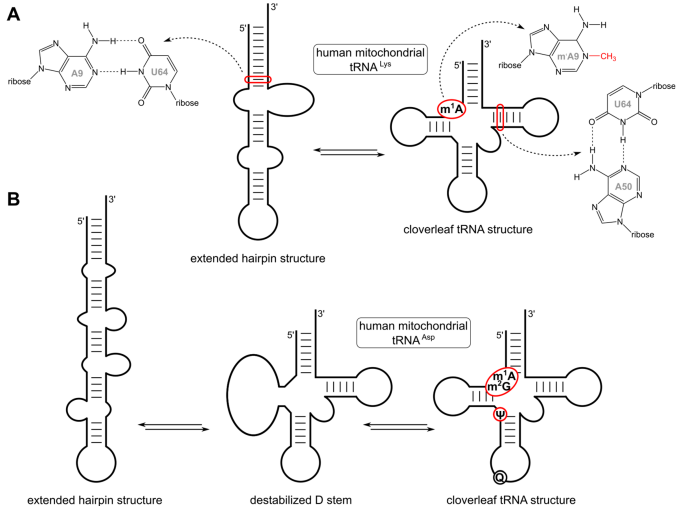
Modifications in tRNA²

- 93 known post-transcriptional modifications



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

Modifications in tRNA²



RNA Secondary Structure Prediction and Modified Bases

How to model modified bases in prediction algorithms?

Actual Requirements:

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

Obstacles:

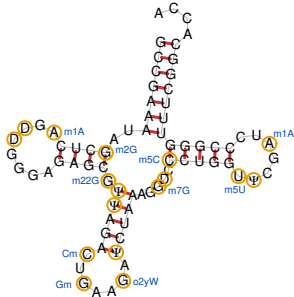
- 2D structure effects known only for a minority of modifications
- 3D effects either unknown or impossible to model
- Combinatorial explosion for energy parameters and pairing rules

Status quo:

- Some modifications prevent base pairing
- Stacking energies are available for $\Psi \bullet A$, $I \bullet U$, $I \bullet C$
- Some data available for (de-)stabilizing effects in literature

tRNA Secondary Structure Prediction

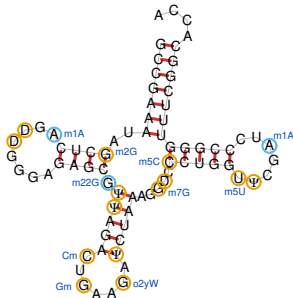
Example: human tRNA^{Phe}



- 17 out of 76 nucleotides are modified

tRNA Secondary Structure Prediction

Example: human tRNA^{Phe}

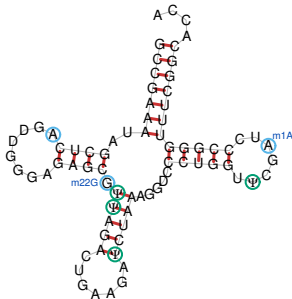


- 17 out of 76 nucleotides are modified
- Some modifications are known to block reverse transcriptase³

³Motorin et al., "Identification of modified residues in RNAs by reverse transcription-based methods.", 2007, Methods in Enzymology 425, 21–53

tRNA Secondary Structure Prediction

Example: human tRNA^{Phe}



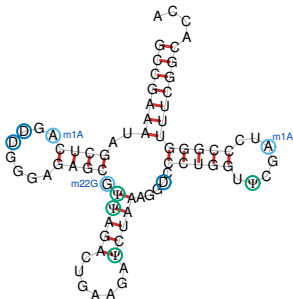
- 17 out of 76 nucleotides are modified
- Some modifications are known to block reverse transcriptase³
- $\Psi \bullet A$ Nearest Neighbor stacking parameters are available⁴

³Motorin et al., "Identification of modified residues in RNAs by reverse transcription-based methods.", 2007, Methods in Enzymology 425, 21–53

⁴Hudson et al., "Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides.", 2013, RNA 19.11, 1474–1482

tRNA Secondary Structure Prediction

Example: human tRNA^{Phe}



- 17 out of 76 nucleotides are modified
- Some modifications are known to block reverse transcriptase³
- Ψ•A Nearest Neighbor stacking parameters are available⁴
- Dihydrouridines (D) destabilize stacking⁵

³Motorin et al., "Identification of modified residues in RNAs by reverse transcription-based methods.", 2007, Methods in Enzymology 425, 21–53

⁴Hudson et al., "Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides.", 2013, RNA 19.11, 1474–1482

⁵Dalluge et al., "Conformational flexibility in RNA: the role of dihydrouridine.", 1996, NAR 24.6, 1073–1079

tRNA Secondary Structure Prediction

Pre-study on 606 sequences tRNAdb (RT-blocking modifications only)

Dataset	Size	Nucleotides			Performance w/o modification			Performance w/ modification		
		total	modified	RT blocking	PPV	TPR	cloverleaf	PPV	TPR	cloverleaf
Bacteria	139	10936	869	66	0.663	0.766	83/ 139	0.687	0.783	86/ 139
Archaea	76	5924	459	59	0.685	0.799	44/ 76	0.687	0.786	41/ 76
Eukaryotes (nuclear)	242	18841	2982	574	0.604	0.685	128/ 242	0.684	0.753	144/ 242
Eukaryotes (mitochondria)	111	7993	720	125	0.605	0.661	47/ 111	0.646	0.687	44/ 111
Eukaryotes (plastids)	38	2972	307	17	0.694	0.768	22/ 38	0.729	0.796	23/ 38
Eukaryotes (total)	391	29806	4009	716	0.613	0.687	197/ 391	0.678	0.739	211/ 391
tRNAdb (total)	606	46666	5337	841	0.635	0.719	324/ 606	0.681	0.755	338/ 606

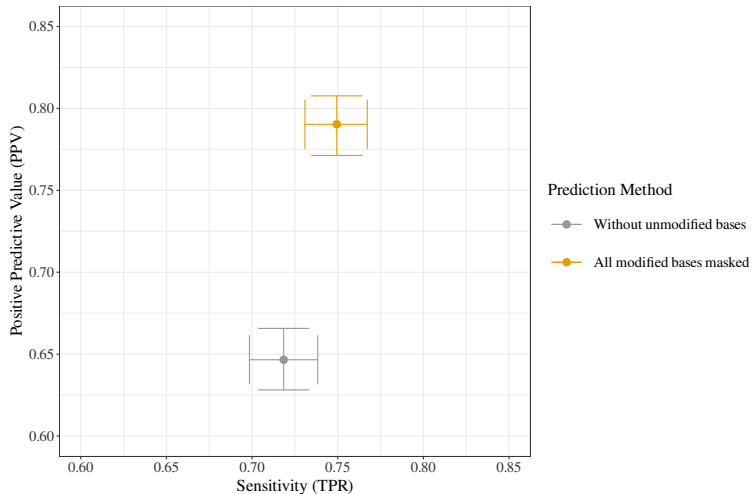
ViennaRNA's constraints framework to the rescue!

- RT-blocking modifications → hard constraints
- $\Psi \bullet A$ stacking energies → soft constraints
- Dihydrouridine (D) destabilization → soft constraints
 - C3'-endo sugar conformation is destabilized in favor of C2'-endo
 - more flexibility
 - promotes destacking
 - destabilization of 1.5 kcal/mol (mono), up to 5.3 kcal/mol (oligo)

This set of constraints on average already yields much better results!

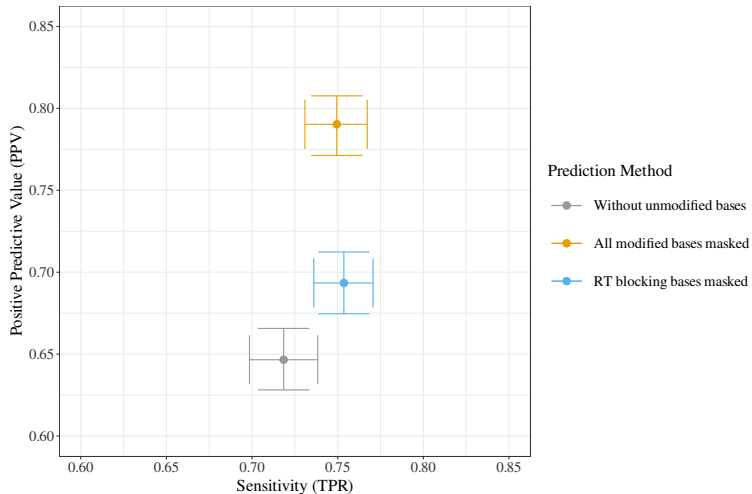
tRNA Secondary Structure Prediction

Performance on tRNADB data set (623 sequences)



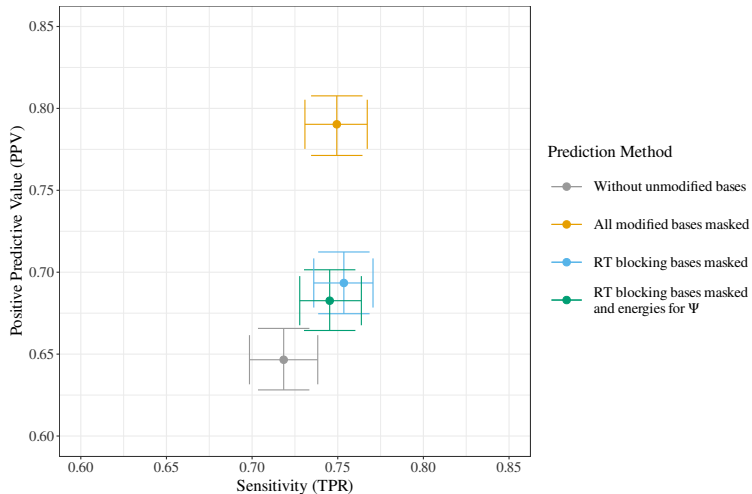
tRNA Secondary Structure Prediction

Performance on tRNAdb data set (623 sequences)



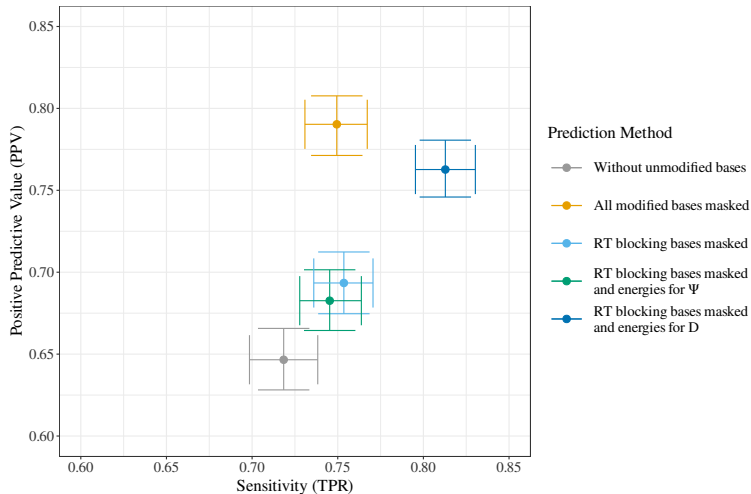
tRNA Secondary Structure Prediction

Performance on tRNADB data set (623 sequences)



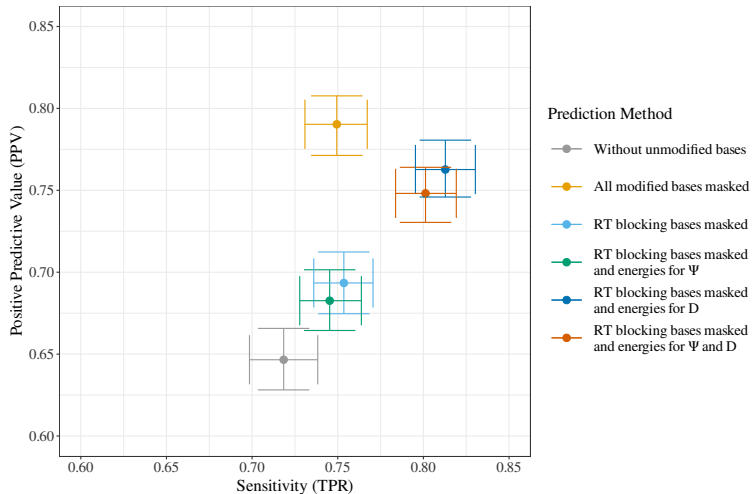
tRNA Secondary Structure Prediction

Performance on tRNADB data set (623 sequences)



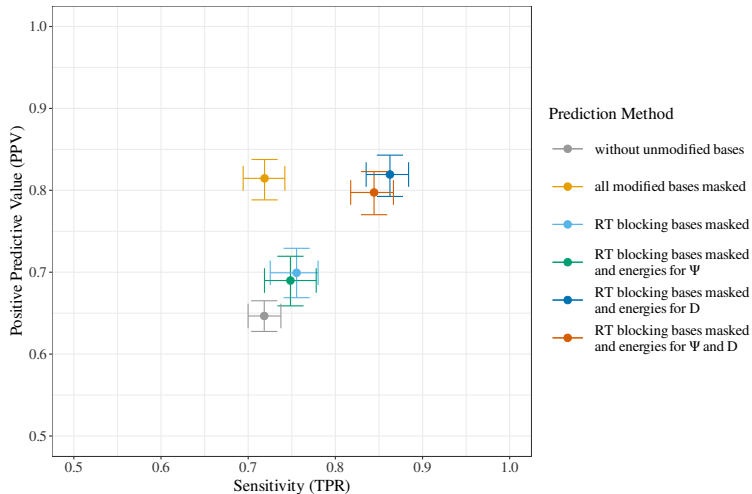
tRNA Secondary Structure Prediction

Performance on tRNADB data set (623 sequences)



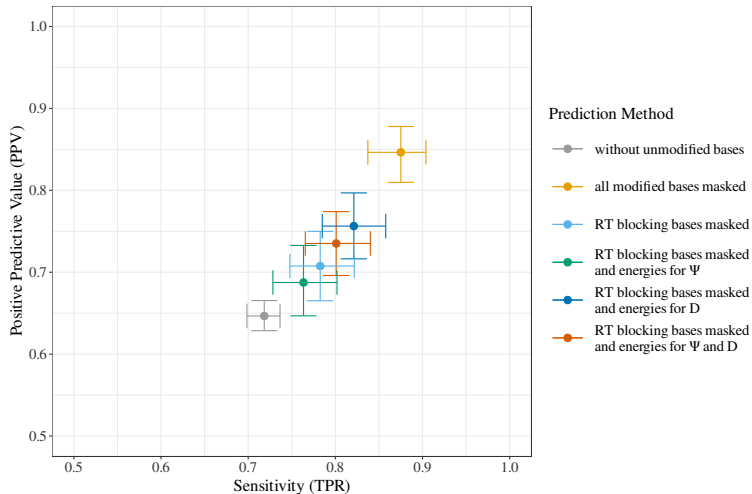
tRNA Secondary Structure Prediction

Performance on tRNADB data set (eucaryotes, 242 sequences)



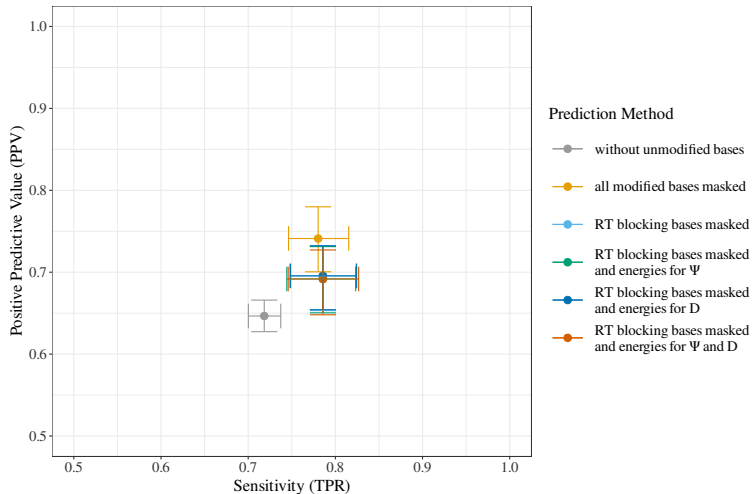
tRNA Secondary Structure Prediction

Performance on tRNADB data set (bacteria, 139 sequences)



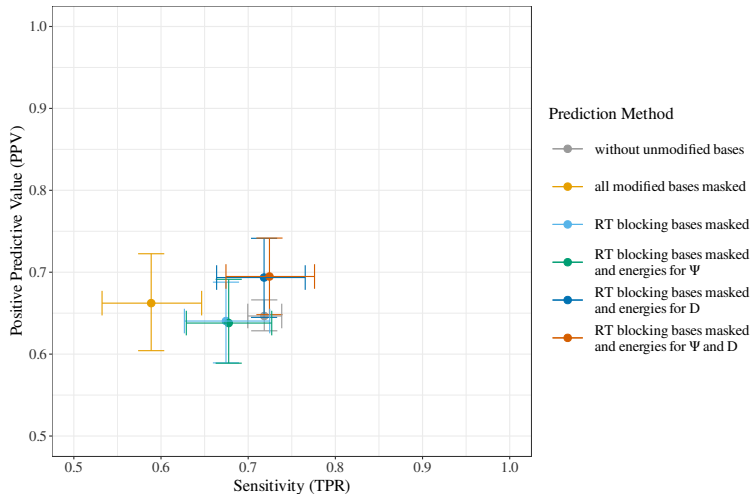
tRNA Secondary Structure Prediction

Performance on tRNADB data set (archaea, 76 sequences)



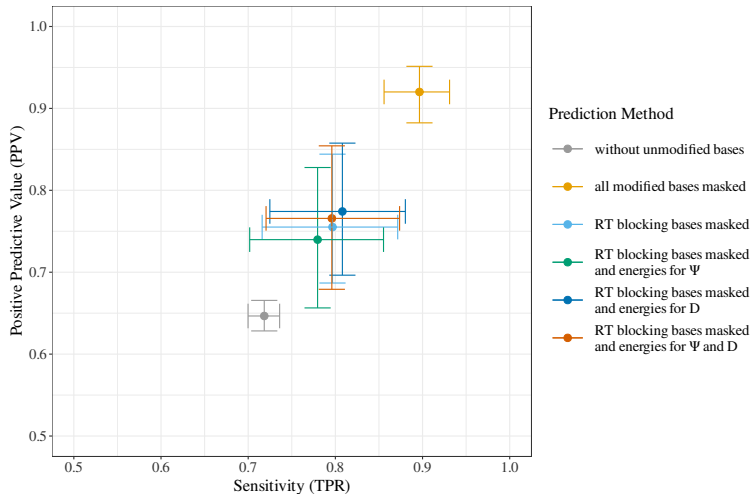
tRNA Secondary Structure Prediction

Performance on tRNADB data set (eucaryotes_mito, 111 sequences)



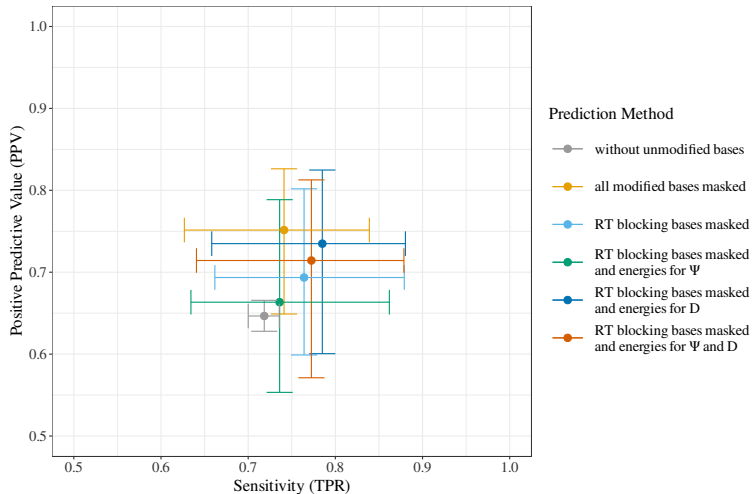
tRNA Secondary Structure Prediction

Performance on tRNADB data set (eucaryotes_plastids, 38 sequences)



tRNA Secondary Structure Prediction

Performance on tRNAdb data set (virus, 17 sequences)



Modified bases may heavily influence structure space

Takeaway Message:

- tRNAs require various modifications to adopt functional form
- Some can already be modeled through constraints
- Additional parameters do not necessarily increase performance
- Constraints become complex for more modifications and contexts
- Unrealistic to include full parameters with many modified bases
- No unique base annotation (tRNAdb⁶, RNAmod⁷, MODOMICS⁸)

Outlook:

- Gather more data on structural effects of modified bases
- Rule and energy parameter set for pairs with modified bases
- Define fallback-rules for missing data
- Full integration of modified bases in ViennaRNA Package

1 open PostDoc Position in the RNAdeco SFB Project

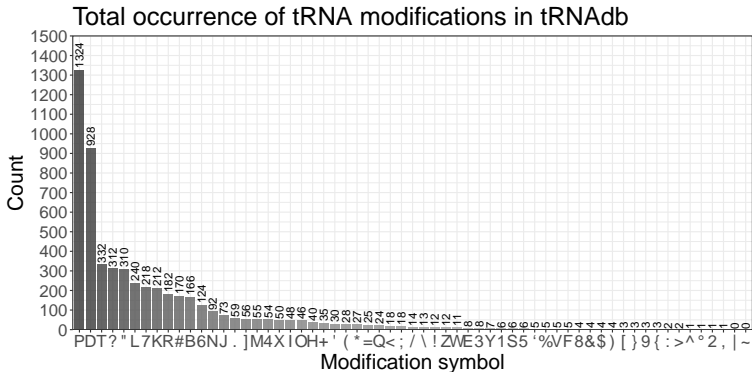
⁶Jühling et al., "tRNAdb 2009: compilation of tRNA sequences and tRNA genes.", 2009, NAR 37, D159–D162

⁷Liu et al., "RNAmod: an integrated system for the annotation of mRNA modifications", 2019, NAR 47, W548-W555

⁸Boccalletto et al., "MODOMICS: a database of RNA modification pathways. 2017 update", 2018, NAR 46, D303-D307

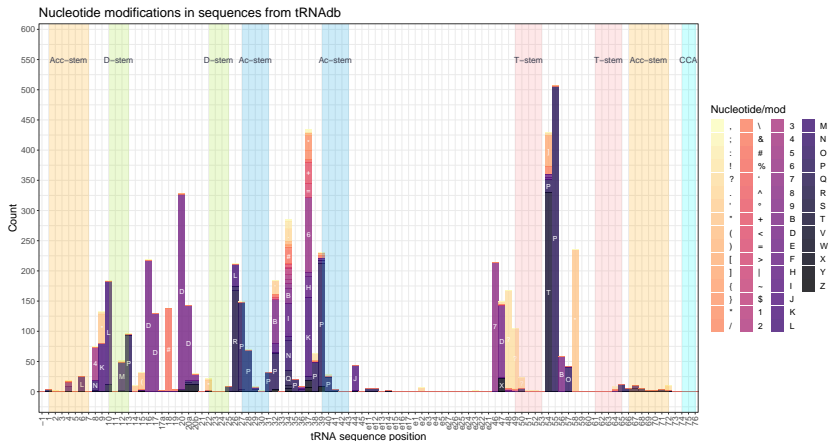
Modifications in tRNA

- Frequency of modifications in tRNAdb



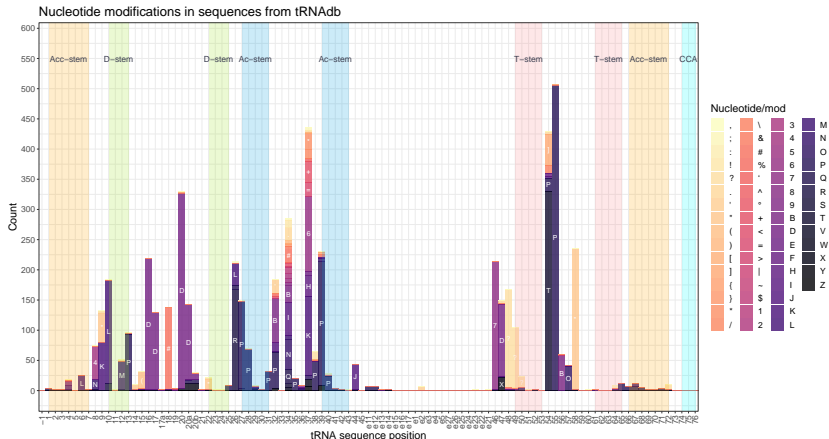
Modifications in tRNA

- 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?



Energy Parameters for Modified Bases

Where to get more NN parameters from?

- Typically obtained from UV-melting experiments
- More parameters to come from HRM fluorescence melting⁹
- In-silico parameter estimation using Rosetta-RECESS¹⁰

NN	RECCES	Expt. ¹¹	Diff.	NN	RECCES	Expt. ¹²	Diff.
5'AI 3'UC	-1.16 ± 0.09	-1.57 ± 0.44	0.41	5'AI 3'UU	-0.04 ± 0.10	-0.41 ± 0.47	0.37
5'AC 3'UI	-0.74 ± 0.13	-1.02 ± 0.40	0.28	5'UU 3'AI	-0.80 ± 0.08	0.43 ± 0.44	1.23
5'UI 3'AC	-0.82 ± 0.07	-0.96 ± 0.40	0.14	5'UI 3'AU	-0.02 ± 0.11	0.37 ± 0.39	0.39
5'UC 3'AI	-1.07 ± 0.09	-1.18 ± 0.44	0.11	5'AU 3'UI	-0.72 ± 0.06	-0.50 ± 0.44	0.22
5'GI 3'CC	-1.83 ± 0.10	-2.62 ± 0.40	0.79	5'GI 3'CU	-1.09 ± 0.06	-1.34 ± 0.33	0.25
5'GC 3'CI	-1.96 ± 0.09	-1.89 ± 0.31	0.07	5'GU 3'CI	-1.76 ± 0.11	-1.03 ± 0.30	0.73
5'CI 3'GU	-1.78 ± 0.16	-1.86 ± 0.31	0.08	5'CI 3'GU	-0.87 ± 0.14	-0.77 ± 0.39	0.10
5'CC 3'GI	-2.21 ± 0.05	-2.23 ± 0.40	0.02	5'CU 3'GI	-1.94 ± 0.13	-1.22 ± 0.37	0.72
5'II 3'CC	-1.03 ± 0.12	-	-	5'II 3'UU	0.07 ± 0.11	2.66 ± 0.88	2.59
5'IC 3'CI	-0.95 ± 0.13	-	-	5'IU 3'UI	-0.09 ± 0.09	3.58 ± 1.09	3.67
5'CI 3'IC	-0.71 ± 0.17	-	-	5'UI 3'IU	0.52 ± 0.14	2.23 ± 0.91	1.71

⁹Wang et al., "Assessment for Melting Temperature Measurement of Nucleic Acid by HRM", 2016, Journal of Analytical Methods in Chemistry, Volume 2016

¹⁰Chou et al., "Blind tests of RNA nearest-neighbor energy prediction", 2016, PNAS July 26, 113 (30) 8430-8435

¹¹D. J. Wright, C. R. Force, and B. M. Znosko, *Nucleic acids research*, vol. 46, no. 22, 2018

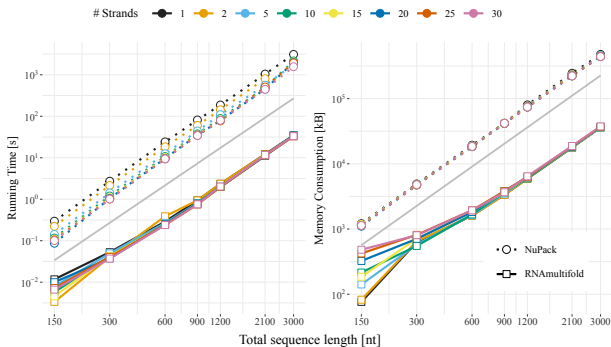
¹²D. J. Wright, J. L. Rice, D. M. Yanker, and B. M. Znosko, *Biochemistry*, vol. 46, no. 15, 2007

RNA-RNA Interactions

RNA-RNA interactions

ViennaRNA Package 2.5.0alpha2 contains RNAmultifold¹³

- Interaction of N RNAs with $n = n_1 + n_2 + \dots + n_N$
- Single or all permutations of a given complex
- All connected complexes up to N constituents
- Implements MFE, partition function, equilibrium concentrations
- $\mathcal{O}(n^3)$ base pair probabilities ($\mathcal{O}(n^2N)$ overhead)

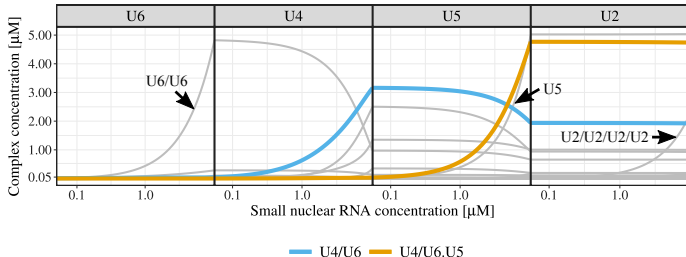


¹³Lorenz et al., 2020, "Efficient Computation of Base-pairing Probabilities in Multi-strand RNA Folding", Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 3: BIOINFORMATICS, 23-31, 2020, Valletta, Malta, & Lorenz et al., 2021, "Efficient Algorithms for Co-Folding of Multiple RNAs", LNCS, submitted

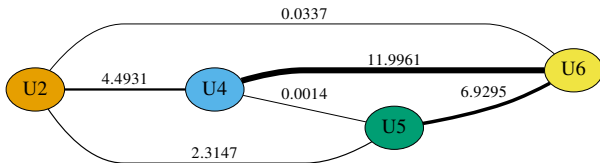
RNA-RNA interactions

Example: Splicosomal snRNA complex formation

- simplified model: no mRNA, proteins, modifications, etc.
- subsequent increase in concentration of U6, U4, U5 and U2



- Importance of binary interactions: $\Delta G_{A|B} = RT \ln Q - RT \ln Q_{A|B} \geq 0$



RNA-RNA interactions

Conclusion, Outlook, and Takeaway Message:

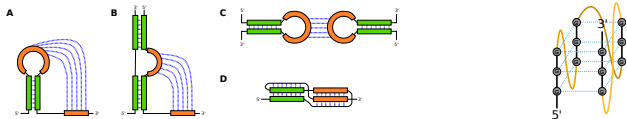
- RNAmultifold available in ViennaRNA Package 2.5.0a2
- Very fast NUPACK alternative
- Same model and parameters as for single sequences
- Merge process into mainline ViennaRNA in progress
- Suboptimal structure prediction still requires attention
- Re-use of DP matrices for different permutations in the future
- New benchmark against NUPACK 4 required¹⁴

¹⁴Fornace et al., "A Unified Dynamic Programming Framework for the Analysis of Interacting Nucleic Acid Strands: Enhanced Models, Scalability, and Speed", 2020, ACS Synth. Biol., 9, 2665-2678

RNA Structure Probing, Pseudoknots, and Structure Motifs

RNA Structure Probing, Pseudoknots, and Structure Motifs

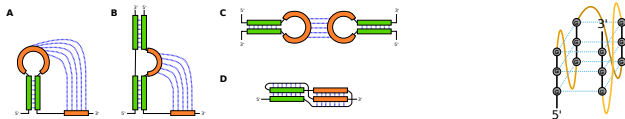
PaRNAssus: Deciphering Complex RNA Structure by Probing and Predictions



- Joint project between FWF (Austria) and ANR (France)
- Exp. probing at different conditions (ions, temperature, agents)
- Separate/Deconvolute (differential) probing signal
- Detection of higher-order structure motifs from probing signals
- Novel heuristics for PK and non-canonical structure prediction
- Implementation of selected already available PK grammars
- Refactoring of RNAPKplex for constraints support almost done
- Connect probing data and folding kinetics simulations

RNA Structure Probing, Pseudoknots, and Structure Motifs

PaRNAssus: Deciphering Complex RNA Structure by Probing and Predictions



- Joint project between FWF (Austria) and ANR (France)
- Exp. probing at different conditions (ions, temperature, agents)
- Separate/Deconvolute (differential) probing signal
- Detection of higher-order structure motifs from probing signals
- Novel heuristics for PK and non-canonical structure prediction
- Implementation of selected already available PK grammars
- Refactoring of `RNAPKplex` for constraints support almost done
- Connect probing data and folding kinetics simulations

2 open PhD Positions in Structural RNA Bioinformatics

Acknowledgements

- Christoph Flamm
- Ivo L. Hofacker
- Yann Ponty
- Bruno Sargueil
- Thomas Spicher
- Peter F. Stadler
- Yuliia Varenyk

Thank You for your attention!

tbi



ANR
FWF



2 open PhD Positions in Structural RNA Bioinformatics
1 open PostDoc Position in the RNAdeco SFB Project