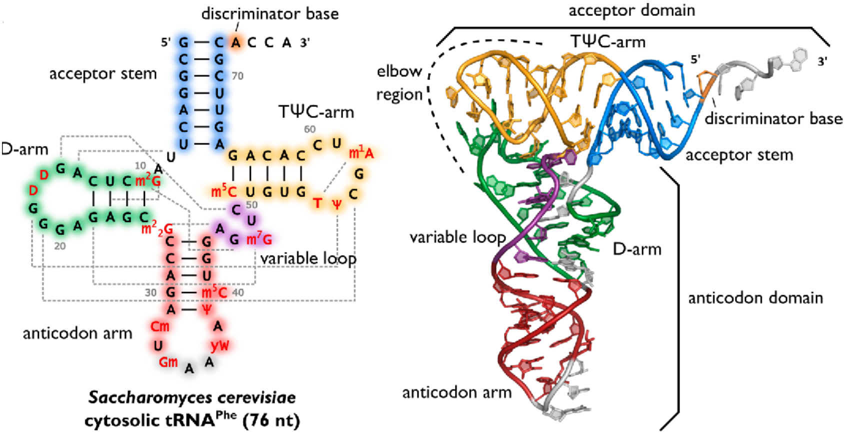
The effect of modified nucleotides on secondary structure prediction in tRNA

Yuliia Varenyk
TBI

Bled Winterseminar

February 13, 2023

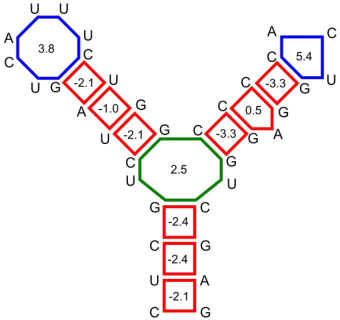
Modifications in tRNA sequences

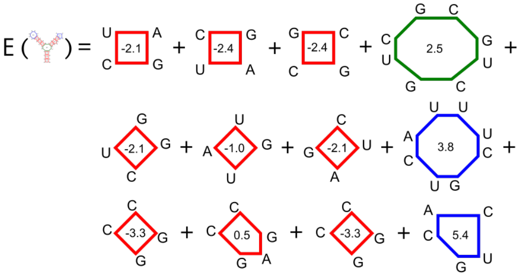


[1]

[1] Lorenz, C., Lünse, C.E., & Mörl, M. (2017). tRNA Modifications: Impact on Structure and Thermal Adaptation. Biomolecules, FIG1

RNA secondary structure prediction - Nearest Neighbor energy model





= -6.50 kcal/mol

- Secondary structures can be uniquely decomposed into loops.
- Contributions of a base pair only depends on neighboring pairs.
- Each loop is assigned a free energy contribution[2].

---

[2] Turner et al., "NNDB: The nearest neighbor parameter database for predicting stability of nucleic acid secondary structure.", 2009, NAR 38, D280–D282

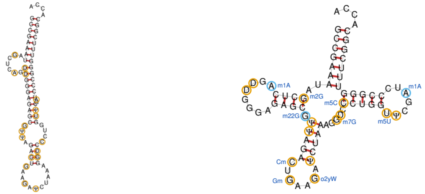
Images by Ronny Lorenz

RNA modifications & secondary structure prediction

Traditional methods lack the means for handling modifications in RNA structure prediction.

- Implementations are limited to the standard RNA nucleotide alphabet AUGC, no complete NN energy parameter set for other nucleotides is available .
- `ViennaRNA Package`[3] provides a plug-in mechanism to adjust free energy parameters and to influence the candidate structure space.

Example: human tRNA Phe (tdbR00000103). 17 out of 76 nucleotides are modified.



Structure prediction as unmodified sequence

Structure prediction including modifications

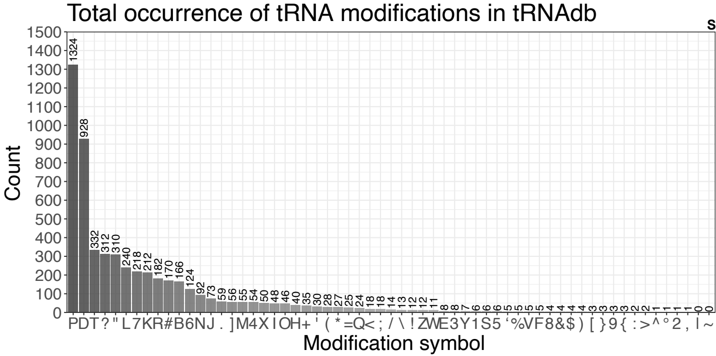
---

[3] Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C. et al. (2011). ViennaRNA Package 2.0. Algorithms Mol Biol

Databases of RNA modifications

- `MODOMICS` [Dunin-Horkawicz et al., 2006, Boccaletto et al., 2018]
- **Transfer RNA database (`tRNAdb`) [Jühling et al., 2009]**
- RNA Modification Database (`RNAMDB`) [Cantara et al., 2011]
- Small Subunit rRNA Modification Database (`SSUmods`) [McCloskeyand Rozenski, 2005]

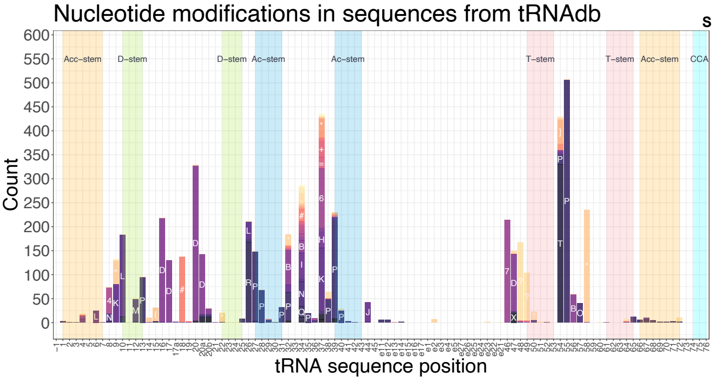
Modifications in `tRNAdb`



Symbol	Modification
P	pseudouridine
D	dihydrouridine
T	ribosylthymine
?	5-methylcytidine
”	1-methyladenosine
L	N2-methylguanosine
7	7-methylguanosine
K	1-methylguanosine
R	N2,N2-dimethylguanosine
#	2’-O-methylguanosine
B	2’-O-methylcytidine
6	N6-threonylcarbamoyladenosine
N	modified uridine
J	2’-O-methyluridine
.	unknown nucleotide
]	1-methylpseudouridine
M	N4-acetylcytidine
4	4-thiouridine
X	3-(3-amino-3-carboxypropyl)uridine

Location of modified bases in tRNA sequences (`tRNAdb`)

Nucleotide modifications in sequences from tRNAdb



Symbol	Modification
P	pseudouridine
D	dihydrouridine
T	ribosylthymine
?	5-methylcytidine
”	1-methyladenosine
L	N2-methylguanosine
7	7-methylguanosine
K	1-methylguanosine
R	N2,N2-dimethylguanosine
#	2’-O-methylguanosine
B	2’-O-methylcytidine
6	N6-threonylcarbamoyladenosine
N	modified uridine
J	2’-O-methyluridine
.	unknown nucleotide
]	1-methylpseudouridine
M	N4-acetylcytidine
4	4-thiouridine
X	3-(3-amino-3-carboxypropyl)uridine

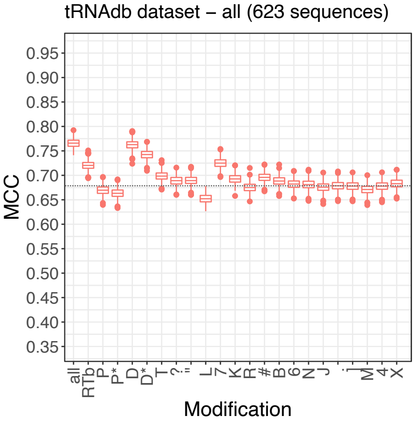
Evaluation of RNA secondary structure prediction (MFE)

- Matthews Correlation Coefficient

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

- Assessment of influence of each individual modification on the prediction results:
  - ◇ apply constraints for individual modification;
  - ◇ compare reference structure and prediction result with modification constraints (MCC);
  - ◇ bootstrap sample of tRNA dataset (1000 repeats);
  - ◇ analyze obtained sampling distribution.

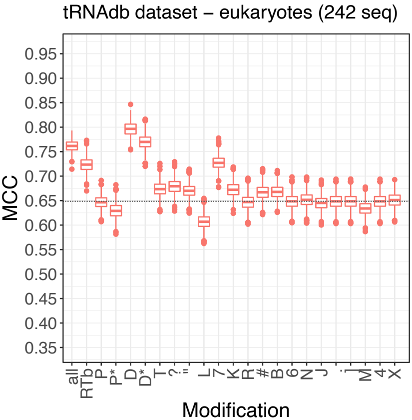
Structure prediction of sequences from `tRNAdb`

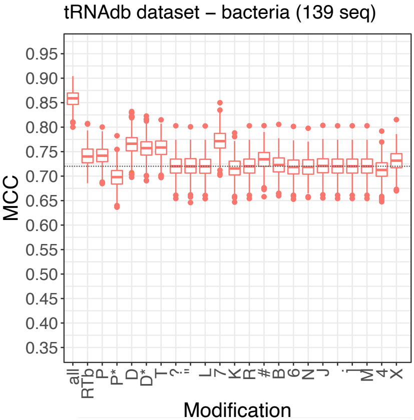

tRNAdb dataset – all (623 sequences)

Symbol	Modification
P	pseudouridine
D	dihydrouridine
T	ribosylthymine
?	5-methylcytidine
”	1-methyladenosine
L	N2-methylguanosine
7	7-methylguanosine
K	1-methylguanosine
R	N2,N2-dimethylguanosine
#	2’-O-methylguanosine
B	2’-O-methylcytidine
6	N6-threonylcarbamoyladenosine
N	modified uridine
J	2’-O-methyluridine
.	unknown nucleotide
]	1-methylpseudouridine
M	N4-acetylcytidine
4	4-thiouridine
X	3-(3-amino-3-carboxypropyl)uridine

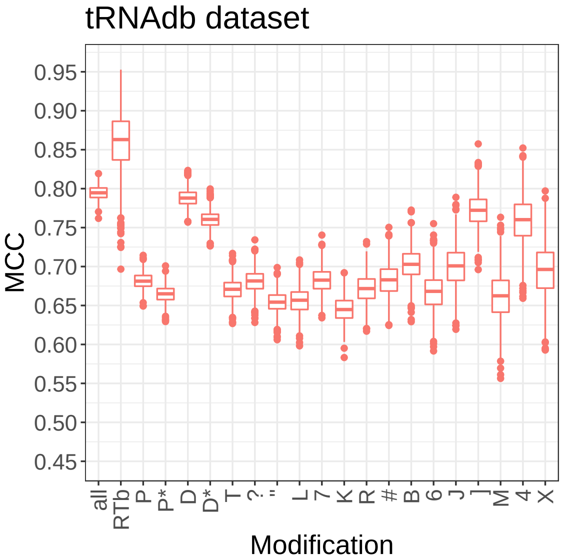
- **all** - mask all modified nucleotides to stay unpaired
- **RTb** - mask nucleotides to stay unpaired if their modification is known to block RT
- **symbol** - mask certain specified nucleotide to stay unpaired
- **symbol*** - apply energy parameter corrections

Structure prediction of sequences from `tRNAdb`

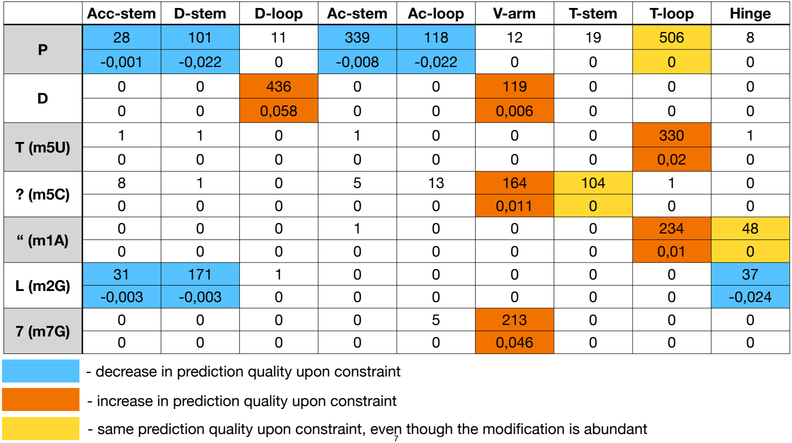

tRNAdb dataset – eukaryotes (242 seq)


tRNAdb dataset – bacteria (139 seq)

Structure prediction of sequences from `tRNAdb` - only sequences that contain the modification included



Prediction quality upon position-dependent constraints



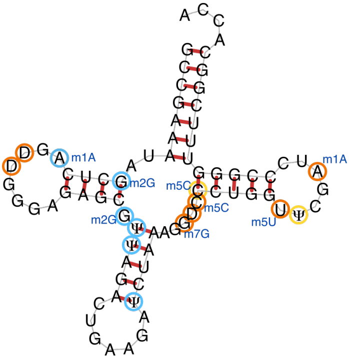
	Acc-stem	D-stem	D-loop	Ac-stem	Ac-loop	V-arm	T-stem	T-loop	Hinge
P	28	101	11	339	118	12	19	506	8
	-0,001	-0,022	0	-0,008	-0,022	0	0	0	0
D	0	0	436	0	0	119	0	0	0
	0	0	0,058	0	0	0,006	0	0	0
T (m5U)	1	1	0	1	0	0	0	330	1
	0	0	0	0	0	0	0	0,02	0
? (m5C)	8	1	0	5	13	164	104	1	0
	0	0	0	0	0	0,011	0	0	0
“ (m1A)	0	0	0	1	0	0	0	234	48
	0	0	0	0	0	0	0	0,01	0
L (m2G)	31	171	1	0	0	0	0	0	37
	-0,003	-0,003	0	0	0	0	0	0	-0,024
7 (m7G)	0	0	0	0	5	213	0	0	0
	0	0	0	0	0	0,046	0	0	0

🟦 - decrease in prediction quality upon constraint

🟧 - increase in prediction quality upon constraint

🟨 - same prediction quality upon constraint, even though the modification is abundant

Prediction quality upon position-dependent constraints



🟦 - decrease in prediction quality upon constraint

🟧 - increase in prediction quality upon constraint

🟨 - same prediction quality upon constraint, even though the modification is abundant

8

Conclusions

Observations

- Identification of modifications of potential interest (e.g. 7-methylguanosine in eukaryotes and bacteria).
- Prediction improvement using constraints.

Outlook

- Further investigation of the structural effects of modifications positionwise.
- Investigation of modification effects in rRNA, mRNA.
- Assembly of database of structural effects of nucleotide modifications from available literature, partner groups of RNA DECO.

Thank you!

Ivo Hofacker

Ronny Lorenz

Thomas Spicher

TBI group

RNA Deco

Bled Winterseminar