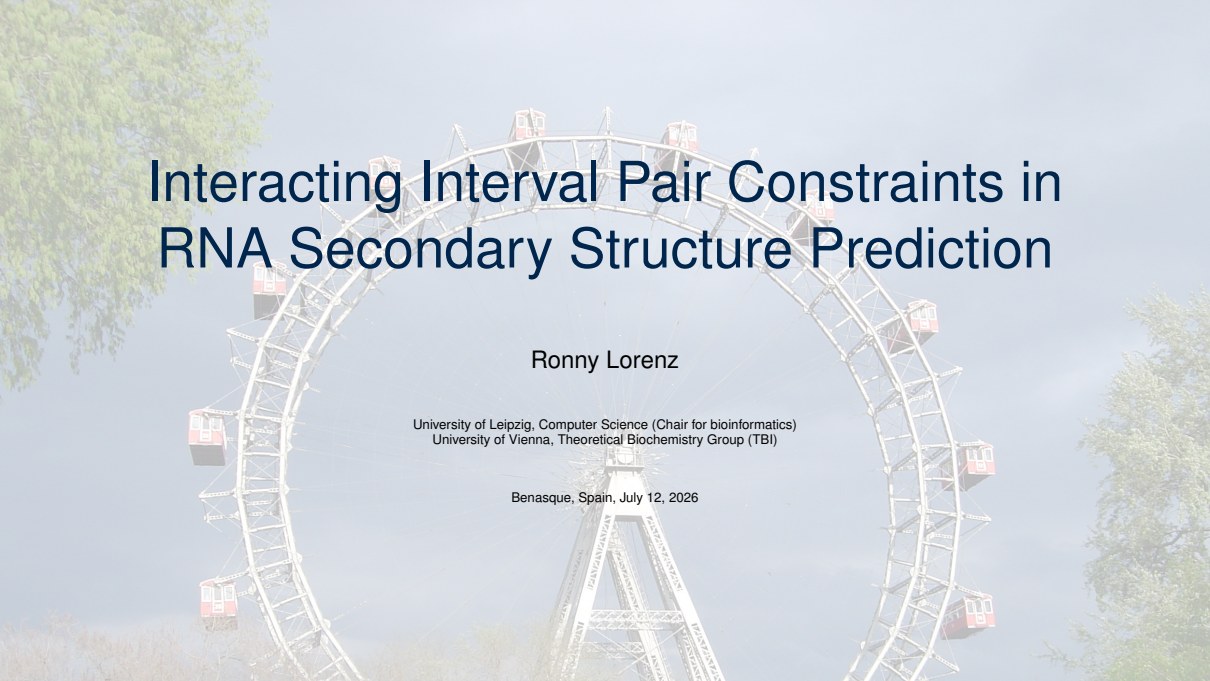




Interacting Interval Pair Constraints in RNA Secondary Structure Prediction

Ronny Lorenz

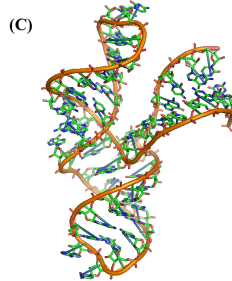
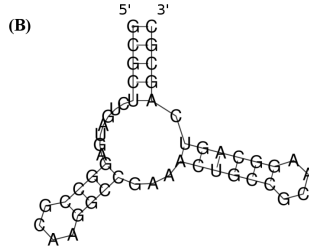
University of Leipzig, Computer Science (Chair for bioinformatics)
University of Vienna, Theoretical Biochemistry Group (TBI)

Benasque, Spain, July 12, 2026

RNA Secondary Structures

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

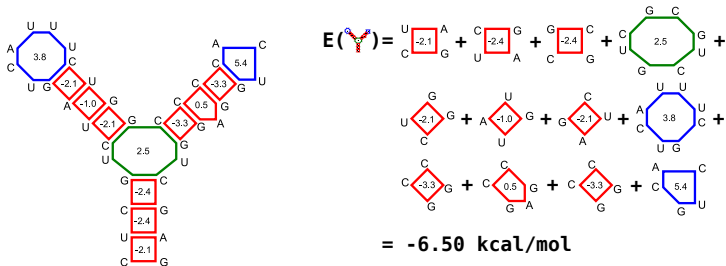
(A) 5' - GCGCUCUGAUGAGGCCGCAAGGCCGAAACUGCCGCAAGGCAGUCAGCGC - 3'



Secondary Structure :

- Captures majority of stabilizing interactions
- Many thermodynamic properties can be predicted efficiently
- Starting point for 3D structure modeling
- Good abstraction to perform coarse grained folding kinetics simulation

RNA Secondary Structures - Nearest Neighbor Energy Model



- 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] Abhinav Mittal, Douglas H Turner, and David H Mathews (2024). "NNDB: An Expanded Database of Nearest Neighbor Parameters for Predicting Stability of Nucleic Acid Secondary Structures". In: *Journal of Molecular Biology*, p. 168549. DOI: 10.1016/j.jmb.2024.168549

RNA Secondary Structures - Statistical Thermodynamics

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

Efficient ($\mathcal{O}(n^3)$) dynamic programming algorithms are available to compute e.g.

- Minimum free energy $MFE = \min_s E(s)$
- Partition function Z
- Probabilities p_{kl} for base pairs (k, l) , i.e. $p_{kl} = p(F)$ with $F = (k, l)$
- Accessibilities q_{il}^μ , i.e. $p(F)$ with $F = [i, \dots, i + l - 1]$ is unpaired
- ...

RNA Secondary Structures - Statistical Thermodynamics

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

Efficient ($\mathcal{O}(n^3)$) dynamic programming algorithms are available to compute e.g.

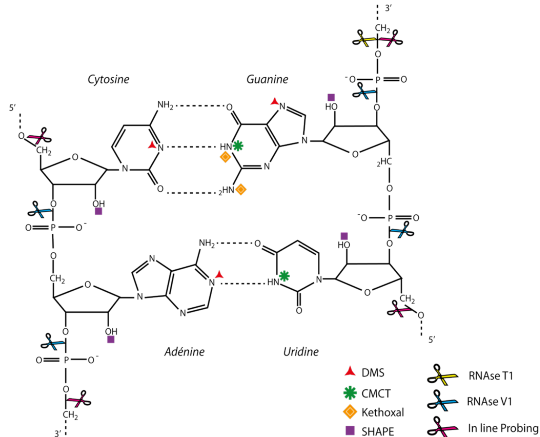
- Minimum free energy $MFE = \min_s E(s)$
- Partition function Z
- Probabilities p_{kl} for base pairs (k, l) , i.e. $p_{kl} = p(F)$ with $F = (k, l)$
- Accessibilities q_{il}^μ , i.e. $p(F)$ with $F = [i, \dots, i + l - 1]$ is unpaired
- ...

However, (as for all models) predictions are far from perfect

- (longer) RNAs might never reach equilibrium
- uncertainty in free energy parameters
- too many simplifying model assumptions

Structure Probing Data - Methods

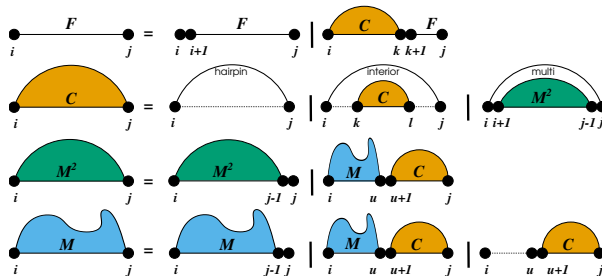
- SHAPE, DMS, CMCT, Enzymatic cleavage, Lead probing, ...
- Usually combined with high-throughput sequencing
- Experiments yield 1-D reactivity profiles



Structure Probing Data - 2D Modeling

Integration of reactivity data into Nearest Neighbor Model

- Deigan [2], Kertesz [3], Zarringhalam [4], Washietl [5], Eddy [6]
- Convert reactivity into per nt. unpaired and/or paired scores (soft constraints)



[2] Katherine E. Deigan et al. (2009). "Accurate SHAPE-directed RNA structure determination". In: *PNAS* 106, pp. 97–102. DOI: 10.1073/pnas.0806929106

[3] M. Kertesz et al. (2010). "Genome-wide measurement of RNA secondary structure in yeast". In: *Nature* 467.7311, pp. 103–107

[4] Kourosh Zarringhalam et al. (2012). "Integrating Chemical Footprinting Data into RNA Secondary Structure Prediction". In: *PLOS ONE* 7.10

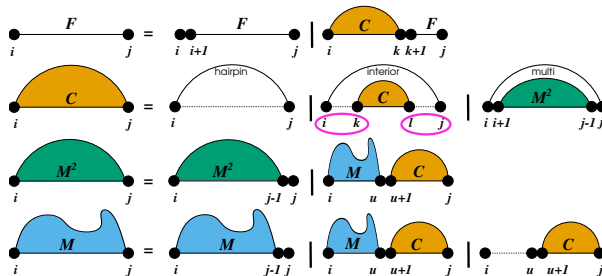
[5] Stefan Washietl et al. (2012). "RNA folding with soft constraints: reconciliation of probing data and thermodynamics secondary structure prediction". In: *Nucleic Acids Res* 40.10, pp. 4261–4272. DOI: 10.1093/nar/gks009

[6] Sean R. Eddy (2014). "Computational analysis of conserved RNA secondary structure in transcriptomes and genomes". In: *Annu Rev Biophys* 43, pp. 433–456

Structure Probing Data - 2D Modeling

Integration of reactivity data into Nearest Neighbor Model

- **Deigan** [2], Kertesz [3], Zarrinhalam [4], Washietl [5], Eddy [6]
- Convert reactivity into per nt. unpaired and/or paired scores (soft constraints)



[2] Katherine E. Deigan et al. (2009). "Accurate SHAPE-directed RNA structure determination". In: *PNAS* 106, pp. 97–102. DOI: 10.1073/pnas.0806929106

[3] M. Kertesz et al. (2010). "Genome-wide measurement of RNA secondary structure in yeast". In: *Nature* 467.7311, pp. 103–107

[4] Kourosh Zarrinhalam et al. (2012). "Integrating Chemical Footprinting Data into RNA Secondary Structure Prediction". In: *PLOS ONE* 7.10

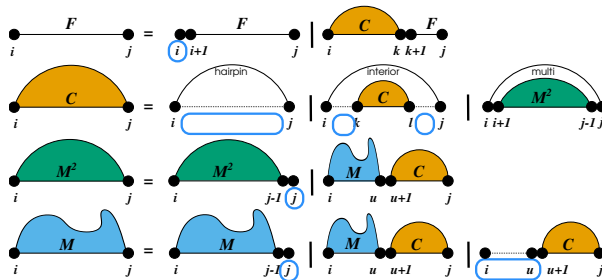
[5] Stefan Washietl et al. (2012). "RNA folding with soft constraints: reconciliation of probing data and thermodynamics secondary structure prediction". In: *Nucleic Acids Res* 40.10, pp. 4261–4272. DOI: 10.1093/nar/gks009

[6] Sean R. Eddy (2014). "Computational analysis of conserved RNA secondary structure in transcriptomes and genomes". In: *Annu Rev Biophys* 43, pp. 433–456

Structure Probing Data - 2D Modeling

Integration of reactivity data into Nearest Neighbor Model

- Deigan [2], Kertesz [3], Zarringhalam [4], **Washietl** [5], Eddy [6]
- Convert reactivity into per nt. unpaired and/or paired scores (soft constraints)



[2] Katherine E. Deigan et al. (2009). "Accurate SHAPE-directed RNA structure determination". In: *PNAS* 106, pp. 97–102. DOI: 10.1073/pnas.0806929106

[3] M. Kertesz et al. (2010). "Genome-wide measurement of RNA secondary structure in yeast". In: *Nature* 467.7311, pp. 103–107

[4] Kourosh Zarringhalam et al. (2012). "Integrating Chemical Footprinting Data into RNA Secondary Structure Prediction". In: *PLOS ONE* 7.10

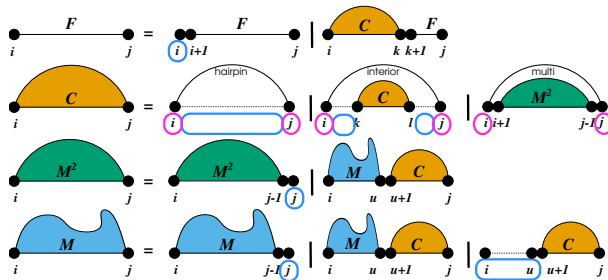
[5] Stefan Washietl et al. (2012). "RNA folding with soft constraints: reconciliation of probing data and thermodynamics secondary structure prediction". In: *Nucleic Acids Res* 40.10, pp. 4261–4272. DOI: 10.1093/nar/gks009

[6] Sean R. Eddy (2014). "Computational analysis of conserved RNA secondary structure in transcriptomes and genomes". In: *Annu Rev Biophys* 43, pp. 433–456

Structure Probing Data - 2D Modeling

Integration of reactivity data into Nearest Neighbor Model

- Deigan [2], Kertesz [3], **Zarrinhalam** [4], Washietl [5], **Eddy** [6]
- Convert reactivity into per nt. unpaired and/or paired scores (soft constraints)



[2] Katherine E. Deigan et al. (2009). "Accurate SHAPE-directed RNA structure determination". In: *PNAS* 106, pp. 97–102. DOI: 10.1073/pnas.0806929106

[3] M. Kertesz et al. (2010). "Genome-wide measurement of RNA secondary structure in yeast". In: *Nature* 467.7311, pp. 103–107

[4] Kourosh Zarrinhalam et al. (2012). "Integrating Chemical Footprinting Data into RNA Secondary Structure Prediction". In: *PLOS ONE* 7.10

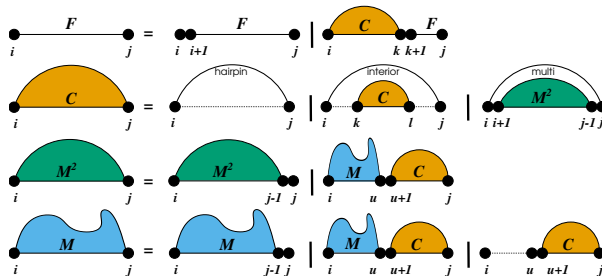
[5] Stefan Washietl et al. (2012). "RNA folding with soft constraints: reconciliation of probing data and thermodynamics secondary structure prediction". In: *Nucleic Acids Res* 40.10, pp. 4261–4272. DOI: 10.1093/nar/gks009

[6] Sean R. Eddy (2014). "Computational analysis of conserved RNA secondary structure in transcriptomes and genomes". In: *Annu Rev Biophys* 43, pp. 433–456

Structure Probing Data - 2D Modeling

Integration of reactivity data into Nearest Neighbor Model

- Deigan [2], Kertesz [3], Zarrinhalam [4], Washietl [5], Eddy [6]
- Convert reactivity into per nt. unpaired and/or paired scores (soft constraints)

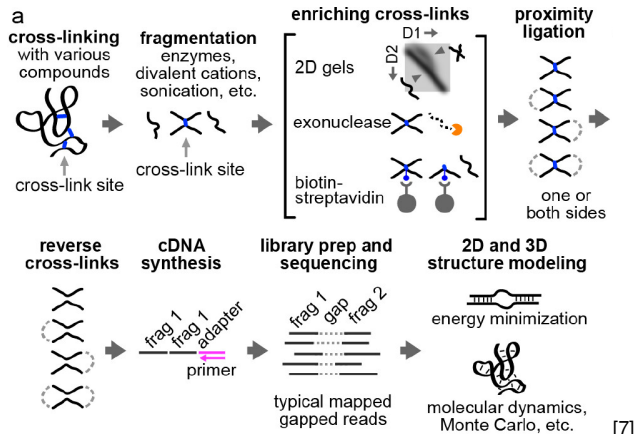


What other data sources can be integrated into secondary structure prediction?

- [2] Katherine E. Deigan et al. (2009). "Accurate SHAPE-directed RNA structure determination". In: *PNAS* 106, pp. 97–102. DOI: 10.1073/pnas.0806929106
- [3] M. Kertesz et al. (2010). "Genome-wide measurement of RNA secondary structure in yeast". In: *Nature* 467.7311, pp. 103–107
- [4] Kourosh Zarrinhalam et al. (2012). "Integrating Chemical Footprinting Data into RNA Secondary Structure Prediction". In: *PLOS ONE* 7.10
- [5] Stefan Washietl et al. (2012). "RNA folding with soft constraints: reconciliation of probing data and thermodynamics secondary structure prediction". In: *Nucleic Acids Res* 40.10, pp. 4261–4272. DOI: 10.1093/nar/gks009
- [6] Sean R. Eddy (2014). "Computational analysis of conserved RNA secondary structure in transcriptomes and genomes". In: *Annu Rev Biophys* 43, pp. 433–456

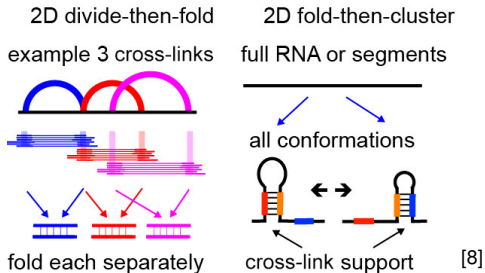
RNA Crosslinking - Protocols

- CLASH, LIGR-seq, SPLASH, PARIS, COMRADES, RIC-seq, etc.
- High-throughput sequencing yields split-reads, i.e. pairs of sequence intervals
- Interval pairs indicate base pairing



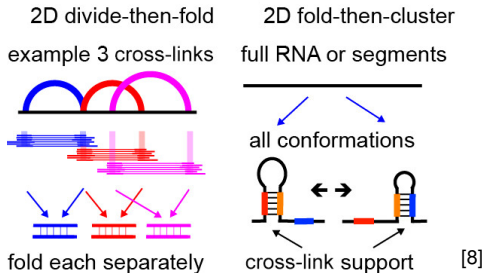
RNA Crosslinking - 2D Modeling

State of the art 2D structure reconstruction with cross-linking data



RNA Crosslinking - 2D Modeling

State of the art 2D structure reconstruction with cross-linking data



Drawbacks :

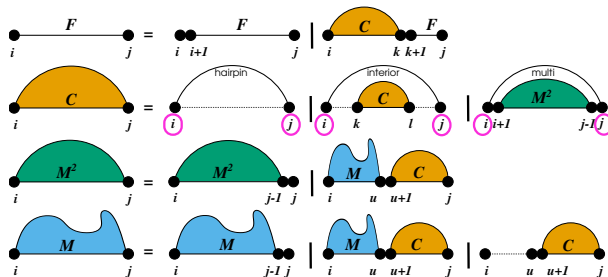
- assumes too many interactions between interval pair (I, I')
- helices might be incompatible
- sampling might miss important structures

RNA Crosslinking - 2D Modeling via base pair soft constraints

Heuristic approach mapping interval probabilities to base pairs ^[9]:

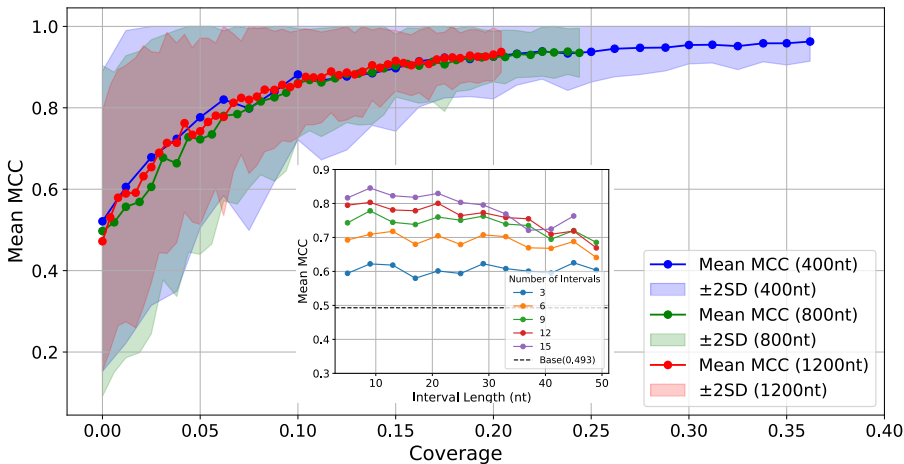
$$\Gamma_{I,I'} = \min\left\{-RT \ln \frac{p(I,I')}{1-p(I,I')}, 0\right\}$$

$$\hat{\Gamma}_{ij} = \begin{cases} \lambda \Gamma_{I,I'}, & \text{with } \lambda = 1/\ell \text{ if } i \in I \wedge j \in I' \\ 0 & \text{otherwise} \end{cases}$$



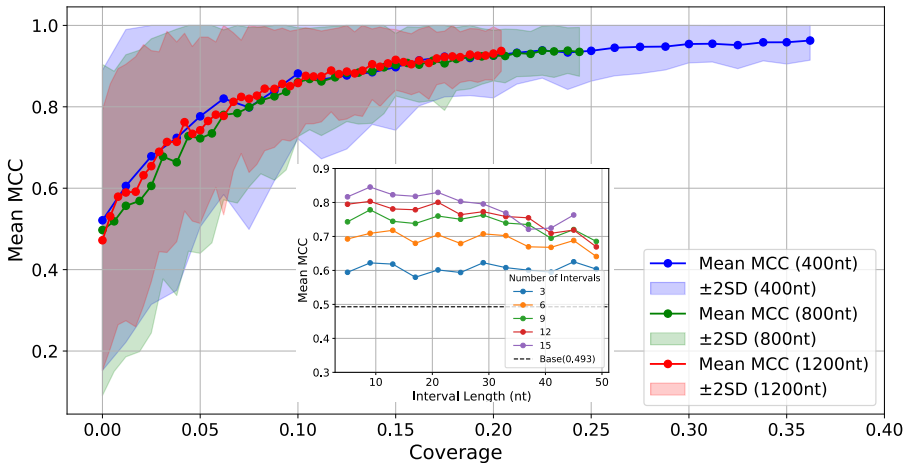
RNA Crosslinking - 2D Modeling via base pair soft constraints

Results on simulated data ($\Gamma(I, I') = 5 \text{ kcal} \cdot \text{mol}^{-1}$, $|I| = |I'| = 5 \text{ nt}$)



RNA Crosslinking - 2D Modeling via base pair soft constraints

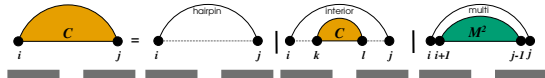
Results on simulated data ($\Gamma(I, I') = 5 \text{ kcal} \cdot \text{mol}^{-1}$, $|I| = |I'| = 5 \text{ nt}$)



Rather apply bonus energy to structures connecting (I, I') exactly once!

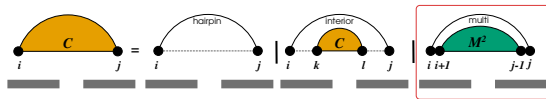
RNA Crosslinking - 2D Modeling via interacting interval pair constraints

- Assume interval pairs do not fully overlap, i.e. $(I, I') \cap (J, J') = (\emptyset, \emptyset) \vee (\emptyset, K') \vee (K, \emptyset)$
- Add $\Gamma(I, I')$ to innermost base pair $(i, j) \in s$ connecting (I, I')



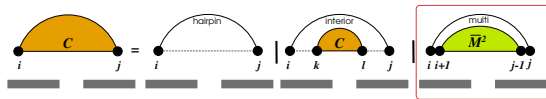
RNA Crosslinking - 2D Modeling via interacting interval pair constraints

- Assume interval pairs do not fully overlap, i.e. $(I, I') \cap (J, J') = (\emptyset, \emptyset) \vee (\emptyset, K') \vee (K, \emptyset)$
- Add $\Gamma(I, I')$ to innermost base pair $(i, j) \in s$ connecting (I, I')



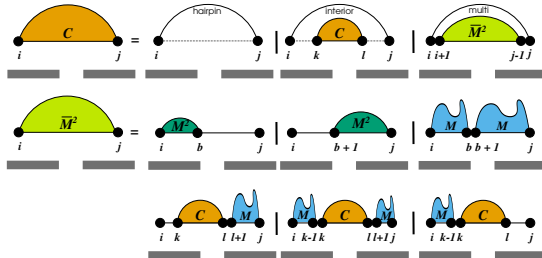
RNA Crosslinking - 2D Modeling via interacting interval pair constraints

- Assume interval pairs do not fully overlap, i.e. $(I, I') \cap (J, J') = (\emptyset, \emptyset) \vee (\emptyset, K') \vee (K, \emptyset)$
- Add $\Gamma(I, I')$ to innermost base pair $(i, j) \in s$ connecting (I, I')



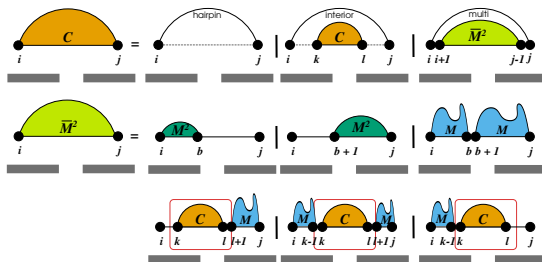
RNA Crosslinking - 2D Modeling via interacting interval pair constraints

- Assume interval pairs do not fully overlap, i.e. $(I, I') \cap (J, J') = (\emptyset, \emptyset) \vee (\emptyset, K') \vee (K, \emptyset)$
- Add $\Gamma(I, I')$ to innermost base pair $(i, j) \in s$ connecting (I, I')



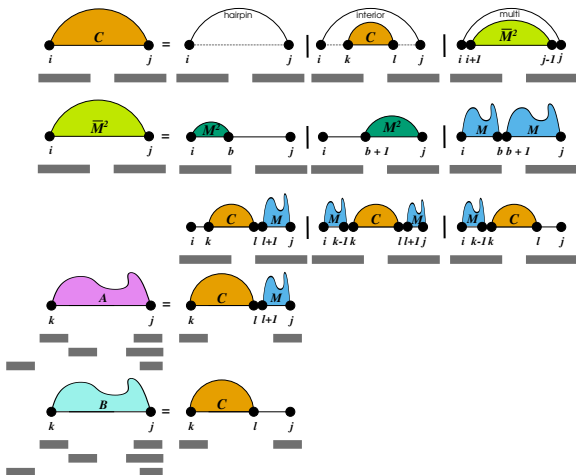
RNA Crosslinking - 2D Modeling via interacting interval pair constraints

- Assume interval pairs do not fully overlap, i.e. $(I, I') \cap (J, J') = (\emptyset, \emptyset) \vee (\emptyset, K') \vee (K, \emptyset)$
- Add $\Gamma(I, I')$ to innermost base pair $(i, j) \in s$ connecting (I, I')



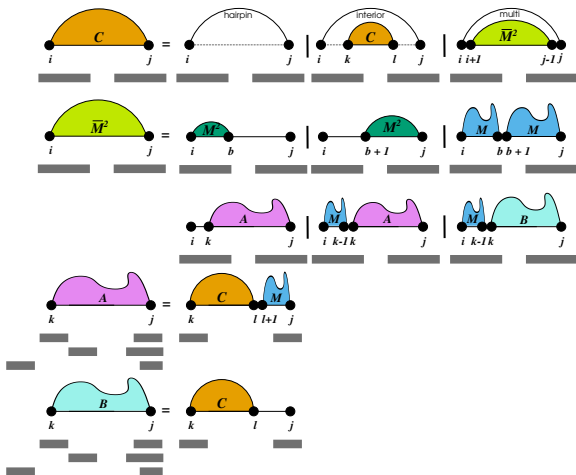
RNA Crosslinking - 2D Modeling via interacting interval pair constraints

- Assume interval pairs do not fully overlap, i.e. $(I, I') \cap (J, J') = (\emptyset, \emptyset) \vee (\emptyset, K') \vee (K, \emptyset)$
- Add $\Gamma(I, I')$ to innermost base pair $(i, j) \in s$ connecting (I, I')



RNA Crosslinking - 2D Modeling via interacting interval pair constraints

- Assume interval pairs do not fully overlap, i.e. $(I, I') \cap (J, J') = (\emptyset, \emptyset) \vee (\emptyset, K') \vee (K, \emptyset)$
- Add $\Gamma(I, I')$ to innermost base pair $(i, j) \in s$ connecting (I, I')



RNA Crosslinking - 2D Modeling via interacting interval pair constraints

- Assume interval pairs do not fully overlap, i.e. $(I, I') \cap (J, J') = (\emptyset, \emptyset) \vee (\emptyset, K') \vee (K, \emptyset)$
- Add $\Gamma(I, I')$ to innermost base pair $(i, j) \in s$ connecting (I, I')

Outside computation, i.e. base pair probabilities, is “trivial” and left as an exercise for the reader! ^[10]

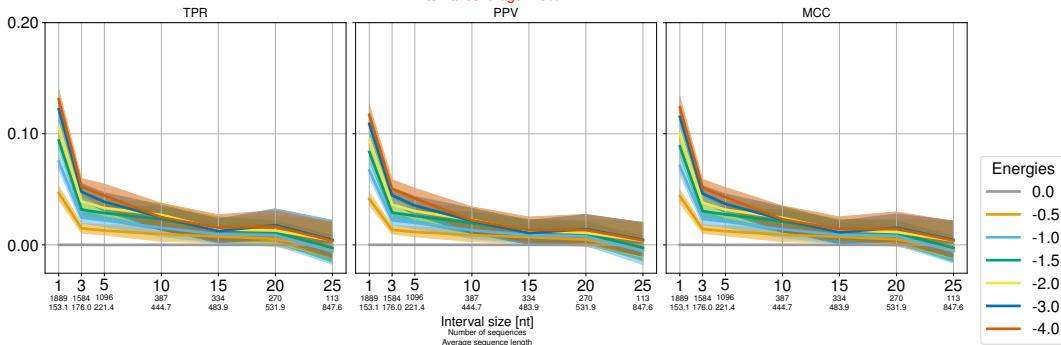
[10] *or the readers favorite AI tool*

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[11], 1920 pk-free sequence/structure pairs from RNAstrand ^[12]

Performance Benchmark (gain over default algorithm)

Interval coverage: 20%



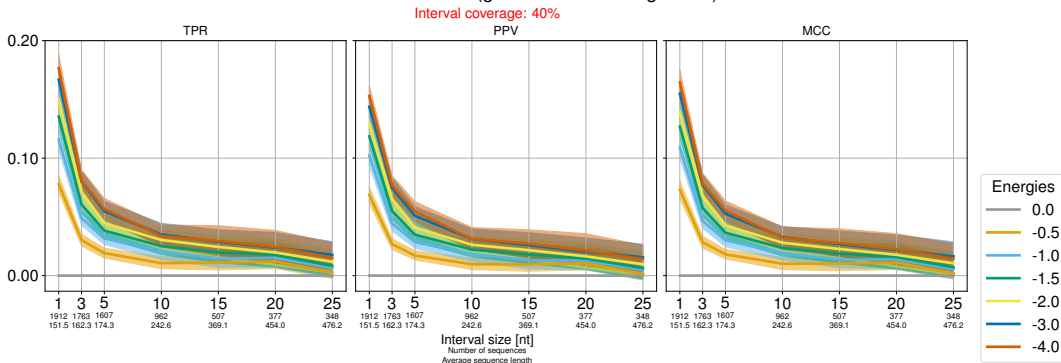
[11] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

[12] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[11], 1920 pk-free sequence/structure pairs from RNAstrand ^[12]

Performance Benchmark (gain over default algorithm)



[11] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

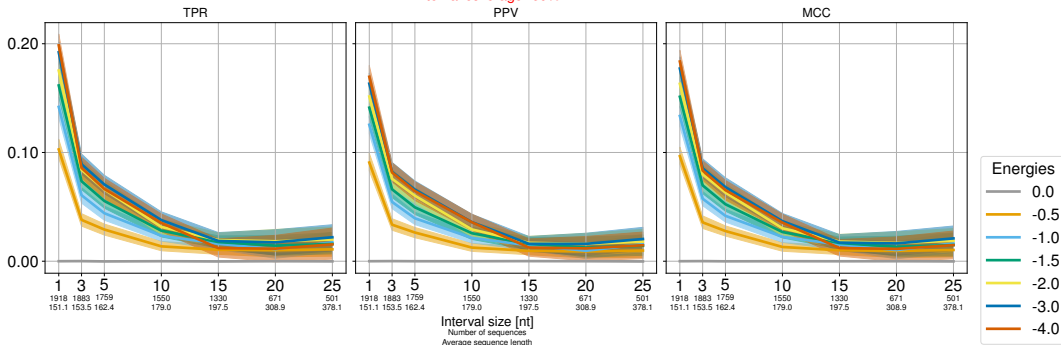
[12] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[11], 1920 pk-free sequence/structure pairs from RNAstrand ^[12]

Performance Benchmark (gain over default algorithm)

Interval coverage: 60%



[11] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

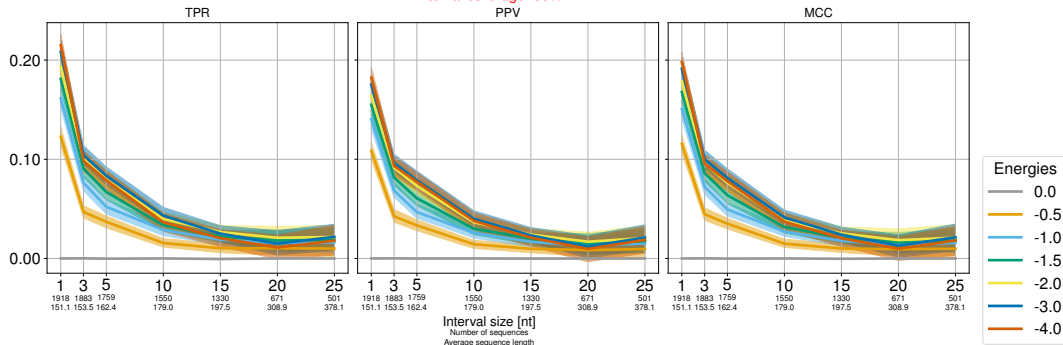
[12] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[11], 1920 pk-free sequence/structure pairs from RNAstrand ^[12]

Performance Benchmark (gain over default algorithm)

Interval coverage: 80%



[11] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

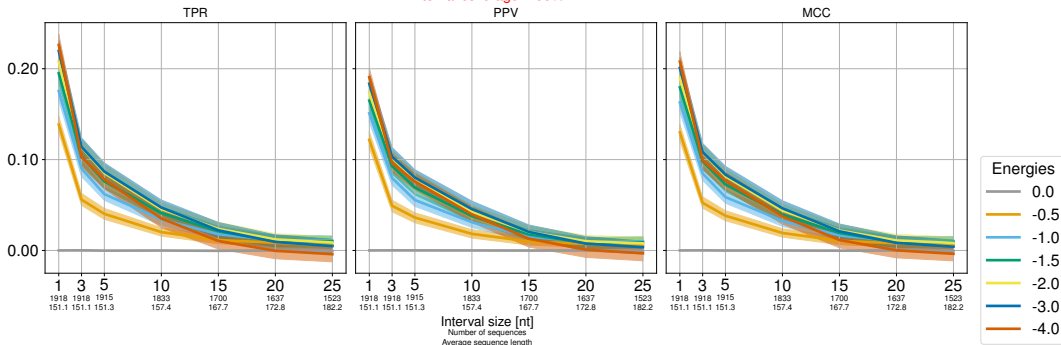
[12] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[11], 1920 pk-free sequence/structure pairs from RNAstrand ^[12]

Performance Benchmark (gain over default algorithm)

Interval coverage: 100%



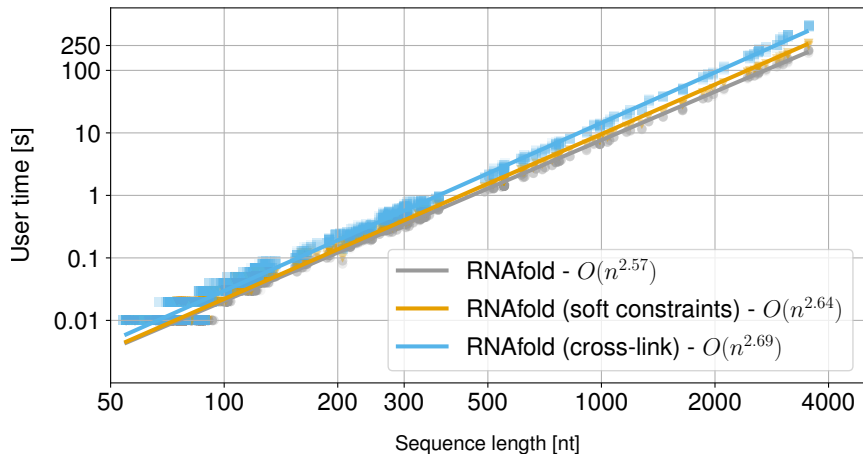
[11] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

[12] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[11], 1920 pk-free sequence/structure pairs from RNAstrand ^[12]

Runtime comparison (MFE + Partition function + Base Pair Probabilities)



[11] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

[12] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

Outlook

Experimental data in structure prediction

- generally increases accuracy in RNA structure modeling
- crosslinking is a rich source of 2D information
- compatible with secondary structure folding grammar

ViennaRNA Package

- multiple probing data conversion strategies
- combination of multiple probing data
- crosslinking data support with next release (v2.8)

Future work

- analysis and pre-processing of crosslinking data
- crosslinking data for inter-molecular interaction?
- crosslinking data in pseudo-knot predictions?
- novel crosslinking protocols that yield shorter intervals?

Acknowledgements

Uni Leipzig

- Sarah v. Loehneysen
- Maria Waldl
- Peter F. Stadler + group

TBI Vienna

- Maximilian Faissner
- Paulína Holotová
- Yingjie Pan
- Leonhard Sidl
- Denis Skibinski
- Thomas Spicher
- Cristian A. Velandia-Huerto
- Hua-Ting Yao
- Ivo L. Hofacker

Thank You for your attention!



UNIVERSITÄT
LEIPZIG

tbi



universität
wien

Characteristics of real RNA cross-linking data sets

Characteristics of RNA-RNA interactions in selected protocols and data sets ^[13].

Method	δ	α	Species	Mapping	Length
CLASH	10-40	NA	yeast	snoRNA-rRNA	< 6857
	16-30	81		snoRNA-snoRNA	89-490
COMRADES	NA	< 1000	zika virus	RNA genome	10807
KARR-Seq	79	14-164	human	18S rRNA/mRNA	< 4067
PARIS	23	> 200	human, mouse	Genome	< 2.49×10^8
SPLASH	NA	300	human, yeast	Transcriptome	< 3353

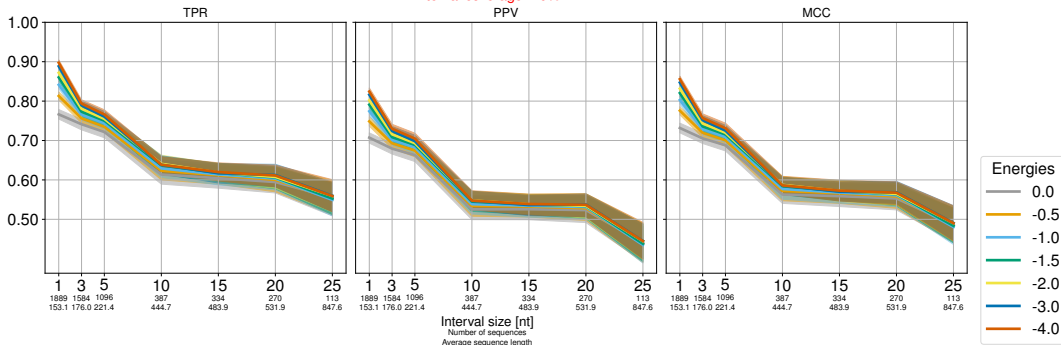
Table: Size of interacting interval pairs (δ) and intervening distances (α) reported in nucleotides (nt) for mapped chimeric reads in published datasets for selected RNA cross-linking methods.

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[14], 1920 pk-free sequence/structure pairs from RNAstrand ^[15]

Performance Benchmark

Interval coverage: 20%



[14] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

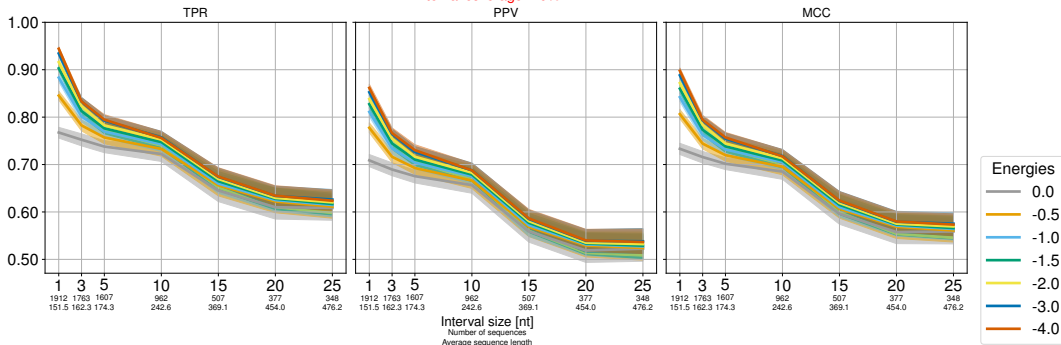
[15] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[14], 1920 pk-free sequence/structure pairs from RNAstrand ^[15]

Performance Benchmark

Interval coverage: 40%



[14] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

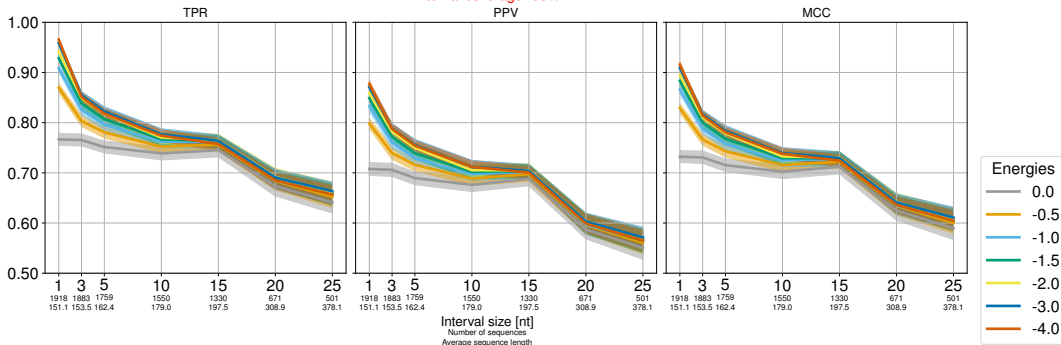
[15] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[14], 1920 pk-free sequence/structure pairs from RNAstrand ^[15]

Performance Benchmark

Interval coverage: 60%



[14] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

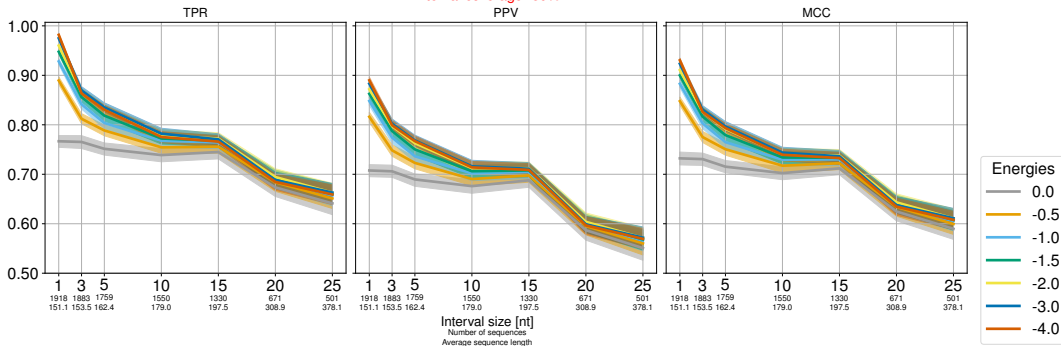
[15] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[14], 1920 pk-free sequence/structure pairs from RNAstrand ^[15]

Performance Benchmark

Interval coverage: 80%



[14] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

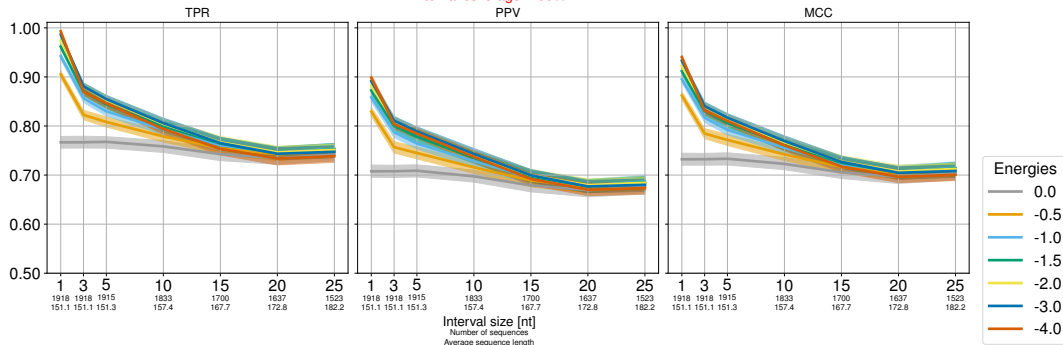
[15] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[14], 1920 pk-free sequence/structure pairs from RNAstrand ^[15]

Performance Benchmark

Interval coverage: 100%



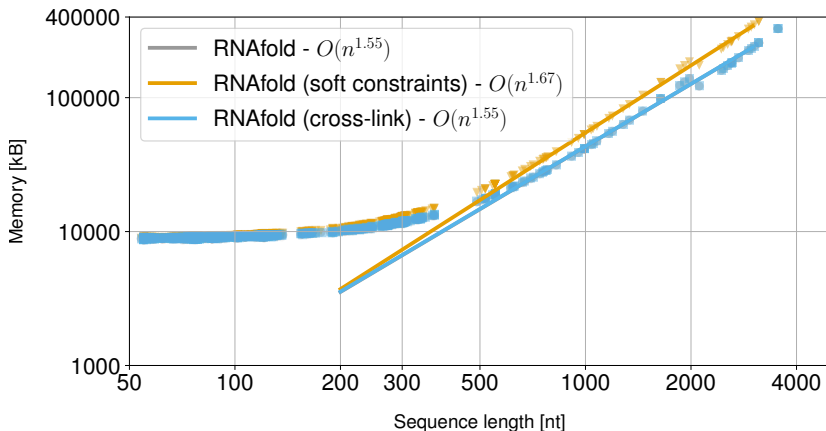
[14] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

[15] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340

RNA Crosslinking - 2D Modeling via interacting interval pair constraints

Artificial interval pairs ^[14], 1920 pk-free sequence/structure pairs from RNAstrand ^[15]

Peak memory consumption (MFE + Partition function + Base Pair Probabilities)



[14] $(I, I') \cap (J, J') = (\emptyset, \emptyset)$

[15] Mirela Andronescu et al. (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC bioinformatics* 9.1, p. 340