Coarse Grain 3D RNA Structure Prediction

Peter Kerpedjiev, Ivo Hofacker

Institute for Theoretical Chemistry, University of Vienna

http://www.tbi.univie.ac.at/

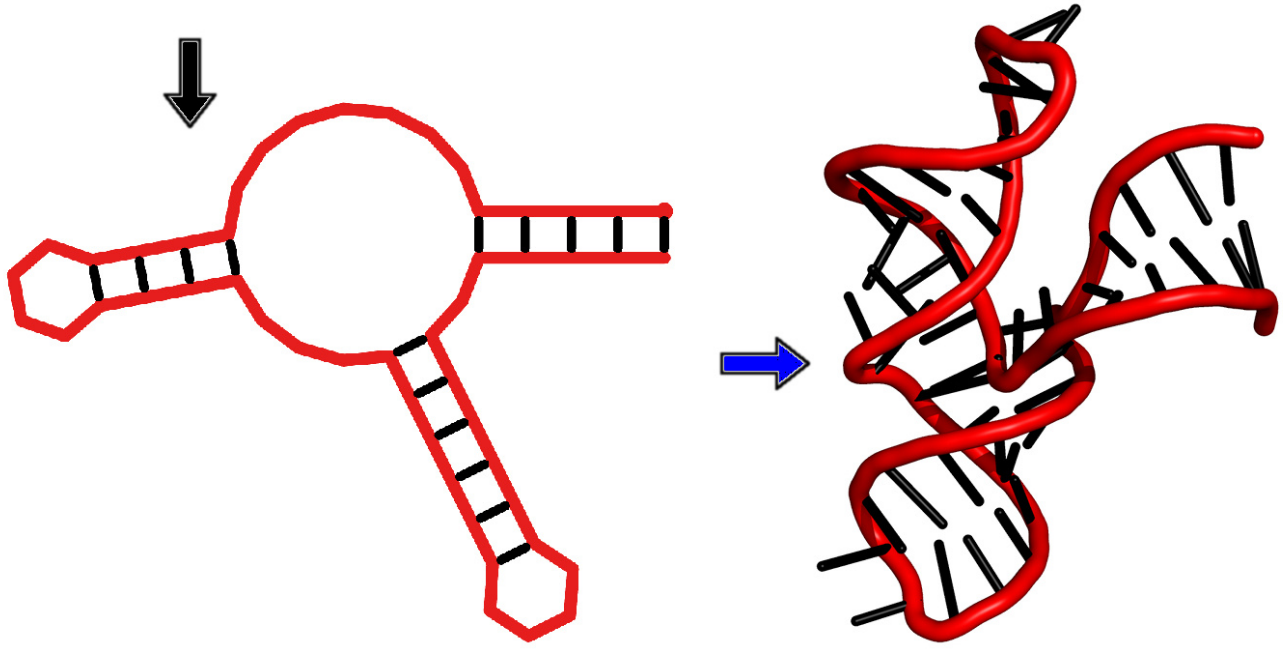
Benasque, July 2015

tbi

universität wien

Hierarchical RNA Folding

GCGCUCUGAUGAGGCCGCAAGGCCGAAACUGCCGCAAGGCAGUCAGCGC



- RNAs fold hierarchically
- To predict 3D structure, assume we know 2D structure

Sampling Approach

Sample structures rather than predicting the *optimal* structure.

Monte Carlo (MCMC) approach:

1. Create a model
2. Evaluate energy
3. Accept / Reject
4. Perturb model
5. Go to step 2

Coarse Graining

Conformation space is of RNA 3D structures is too large for efficient sampling.

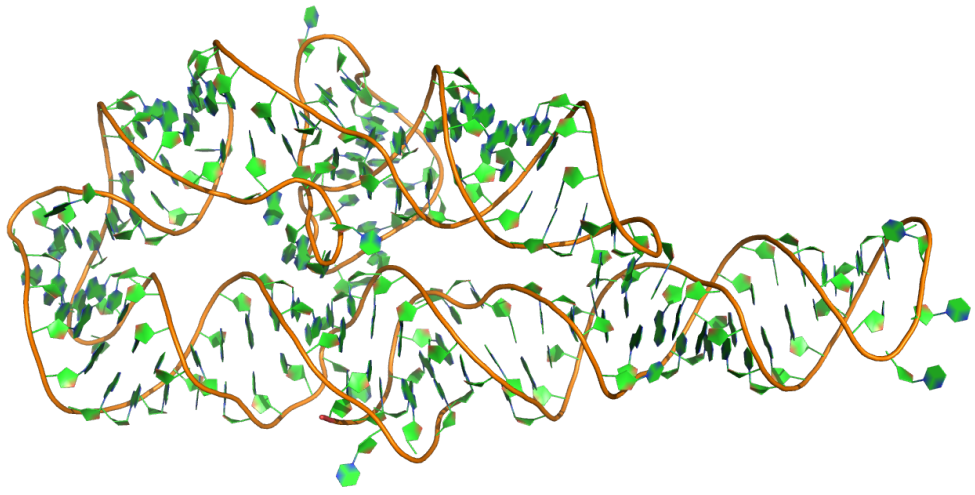
Solution: Coarse graining

Remove details from the models in order to

- make larger strides across the conformation space
- spend less effort evaluating each structure

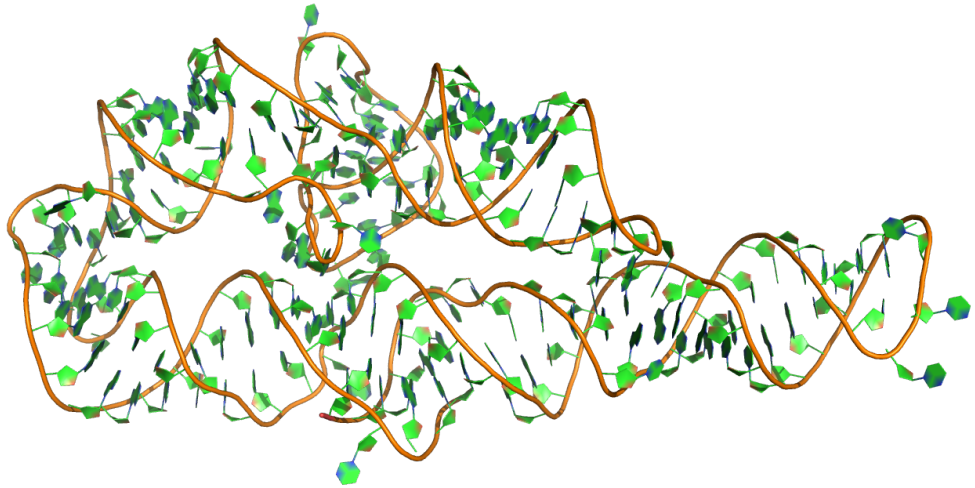
Coarse Graining

All-Atom Representation

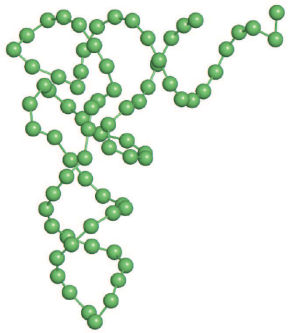


Coarse Graining

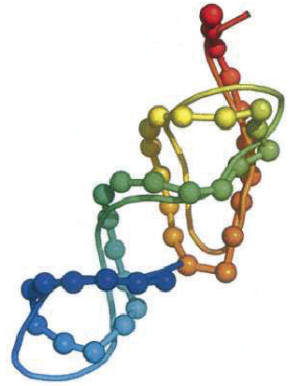
All-Atom Representation



Easiest: Coarse Graining with one (or few) points per Nucleotide

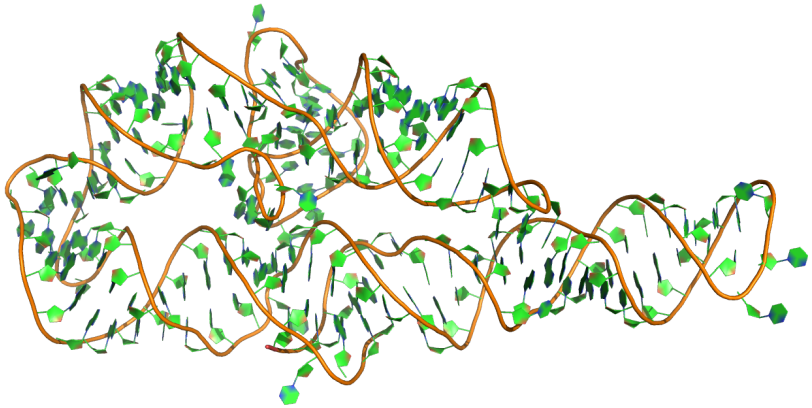


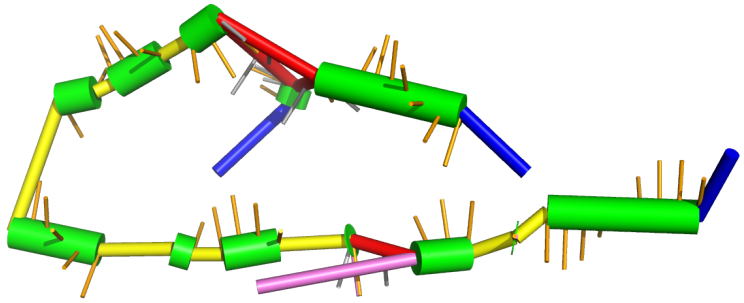
NAST (Jonikas et al. 2009)



DMD (Ding et al. 2008)

Ernwin Coarse Grained Representation

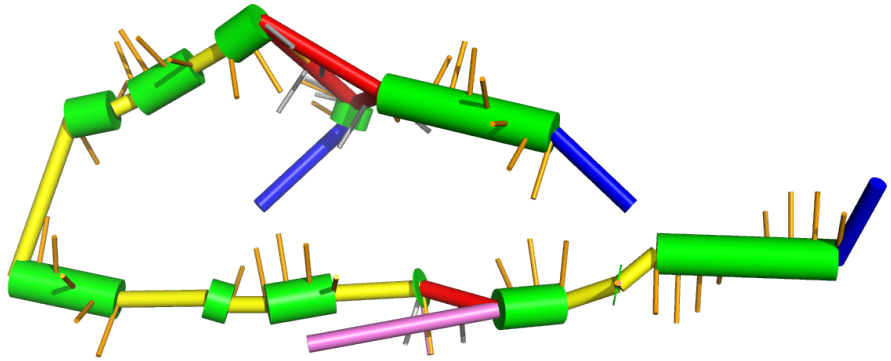




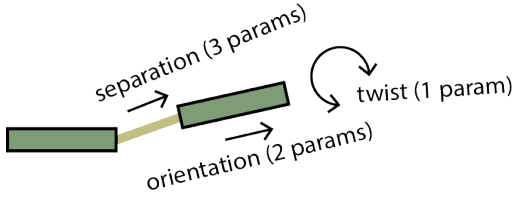
Helix based coarse graining with

- 6 parameters per interior loop
- 3 parameters per hairpin / exterior loop
- 1 parameter per helix

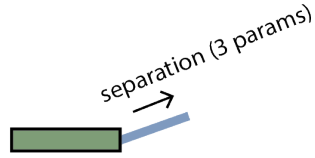
Coarse Grain Model Parameters



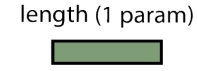
Interior/Multi Loops



Hairpin/Exterior Loops



Helices

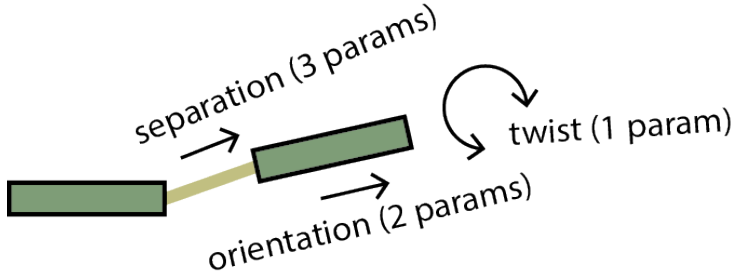


Structure Evaluation

Structure evaluation is split into two parts:

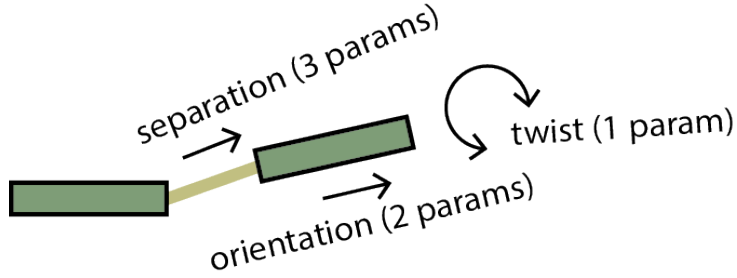
- Local structure:
  Relative orientation of adjacent helices
- Global structure:
  Avoid steric clashes
  Ensure compactness of structures
  Long-range interactions

Local Structure



- Learn frequency distribution for helix orientations dependent on loop size
- Used large number of Rosetta generated models with random sequences

Local Structure



- Learn frequency distribution for helix orientations dependent on loop size
- Used large number of Rosetta generated models with random sequences
- Locally correct structures can be now be sampled directly!
- MCMC only needed to include long range interactions

Non-local Energies

How to include non-local energies?

- Knowledge based potentials common approach for global properties and long-range interactions
- Given feature X with a reference (desired) frequency distribution $P(X)$

$$E_X = -c \ln \frac{P(X)}{Q(X)}$$

  with $Q(X)$ the background (expected) distribution.
- Used in the accept/reject step of MCMC

Non-local Energies

How to include non-local energies?

- Knowledge based potentials common approach for global properties and long-range interactions
- Given feature X with a reference (desired) frequency distribution $P(X)$

$$E_X = -c \ln \frac{P(X)}{Q(X)}$$

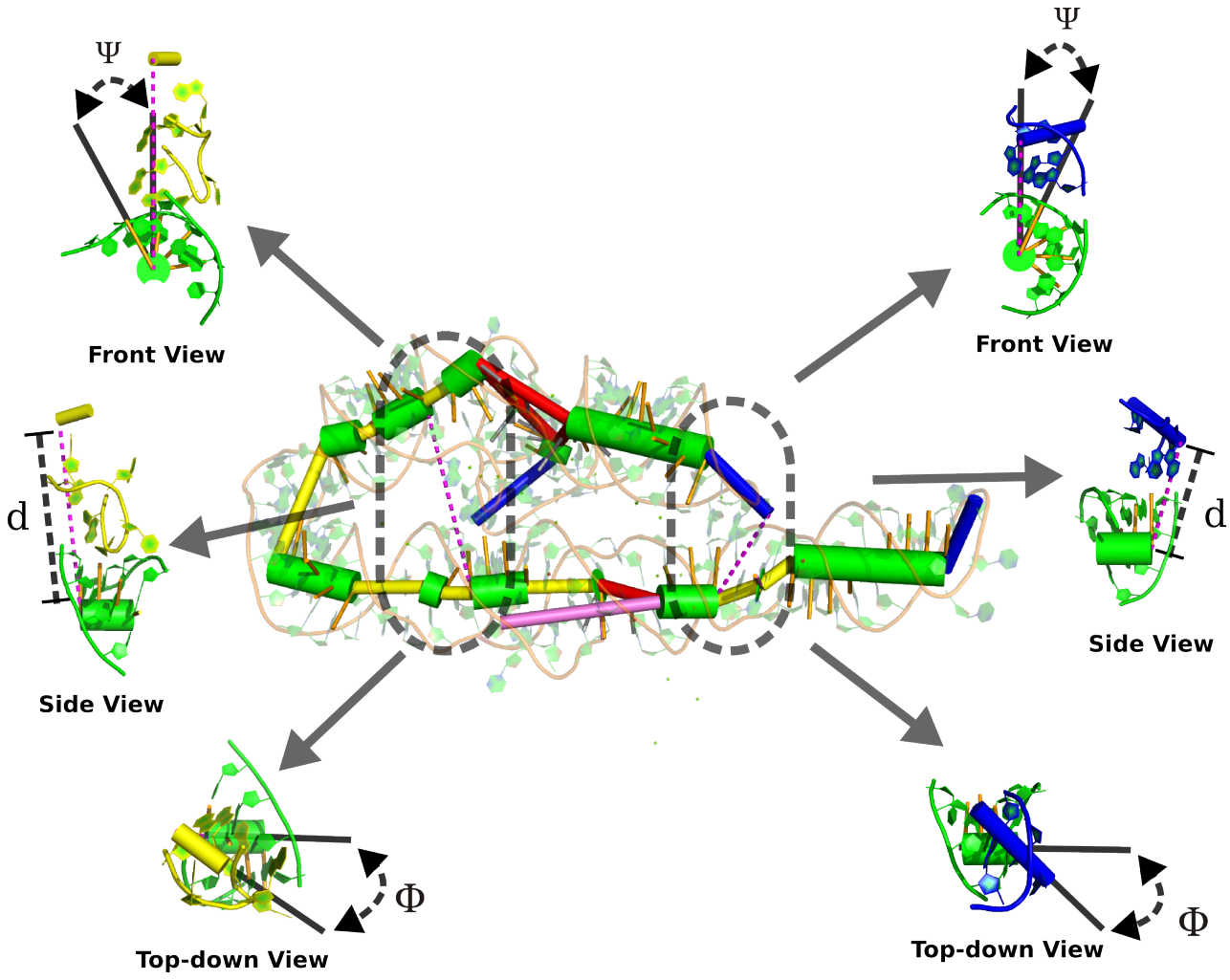
  with $Q(X)$ the background (expected) distribution.
- Used in the accept/reject step of MCMC
- Instead of a fixed $Q(X)$, we sample $Q(X)$ during the simulation
  Known as reference ratio method (Hammelryck, 2010)

Non-local Energies

Currently, we use 4 non-local energy terms:

- Clash detection — uses 1 virtual atom for each nucleotide
- Radius of gyration — ensures compactness of structures
- Loop-loop — distance between two hairpin loops
- A-minor interaction — unpaired A interacting with the minor groove of a helix (hairpin and interior loop version)

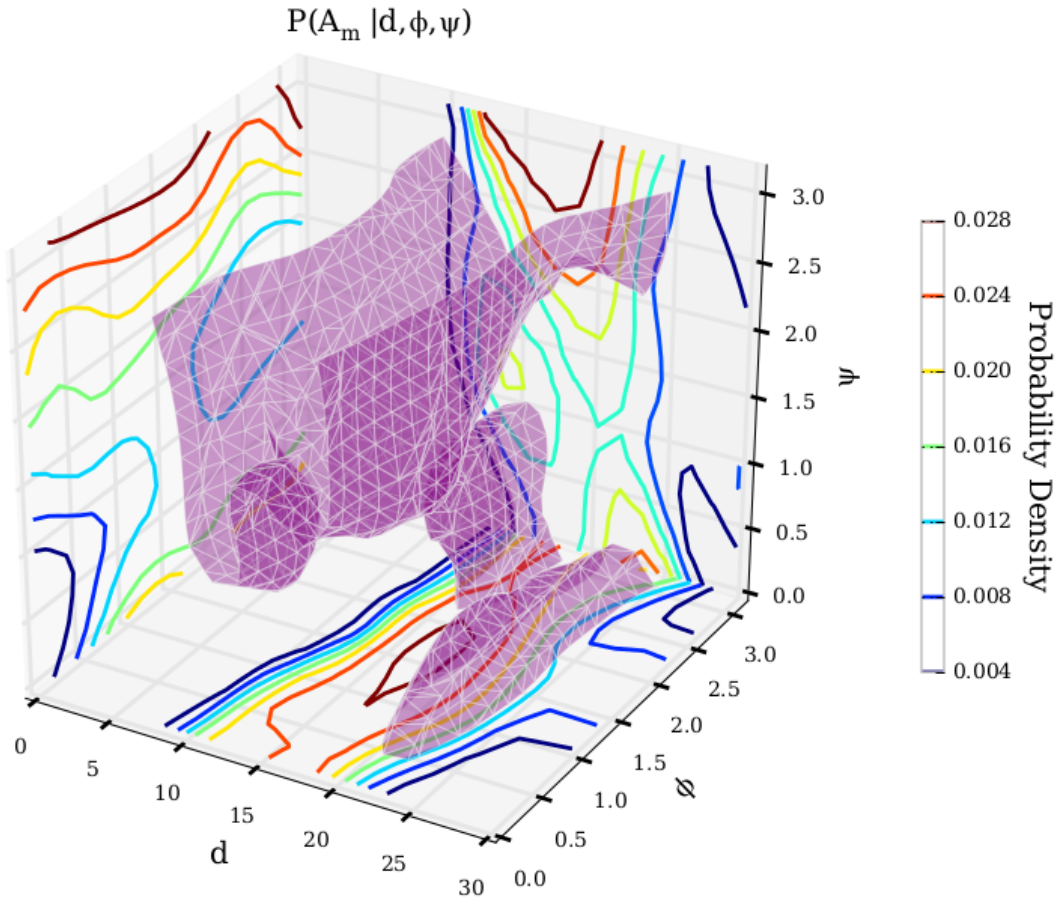
A-minor Interaction



The parameters d, Φ, and Ψ describe how the donor loop (blue) is oriented with respect to the receptor stem (green).

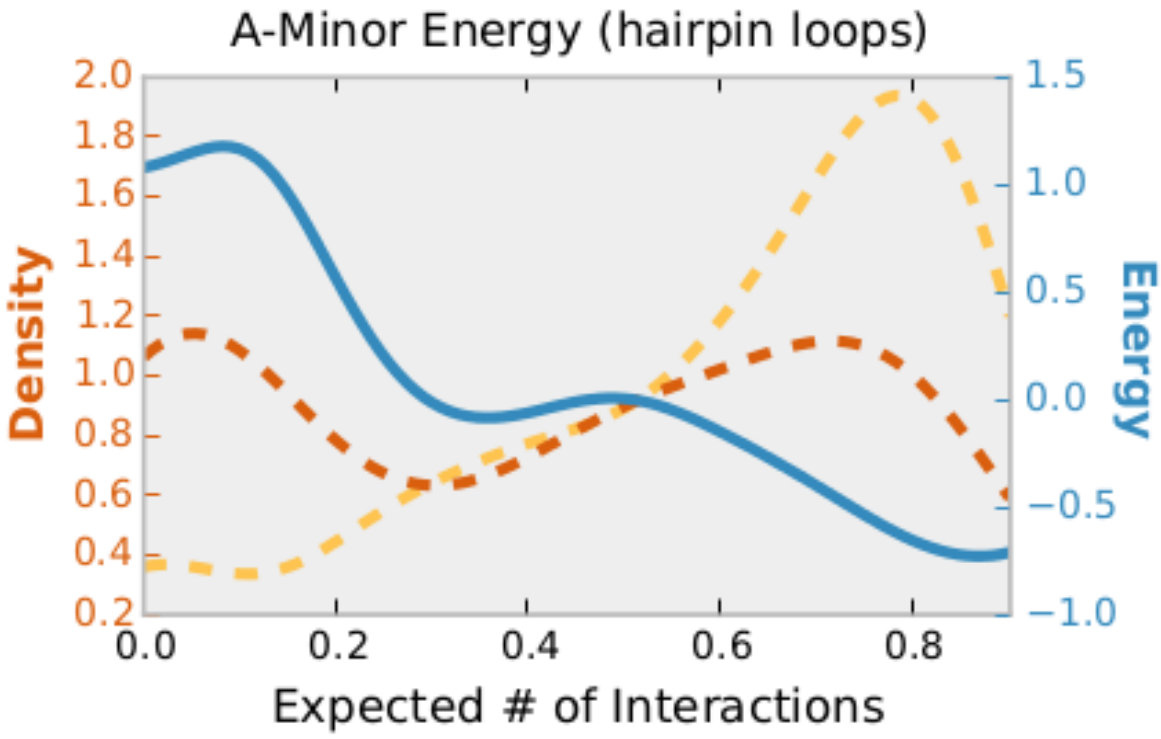
A-minor Interaction

Probability Density of the A-Minor Energy

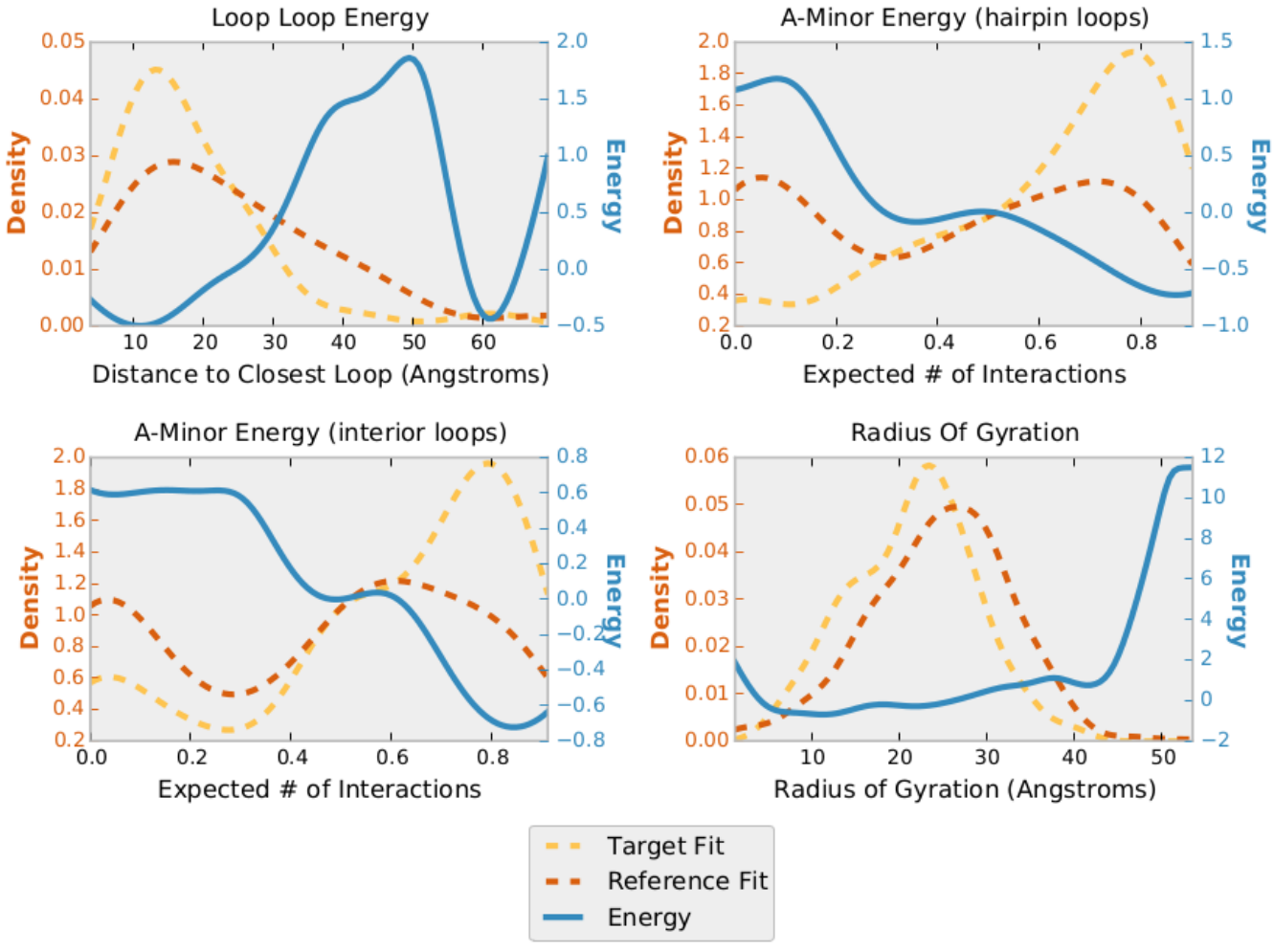


Expected vs. Sampled

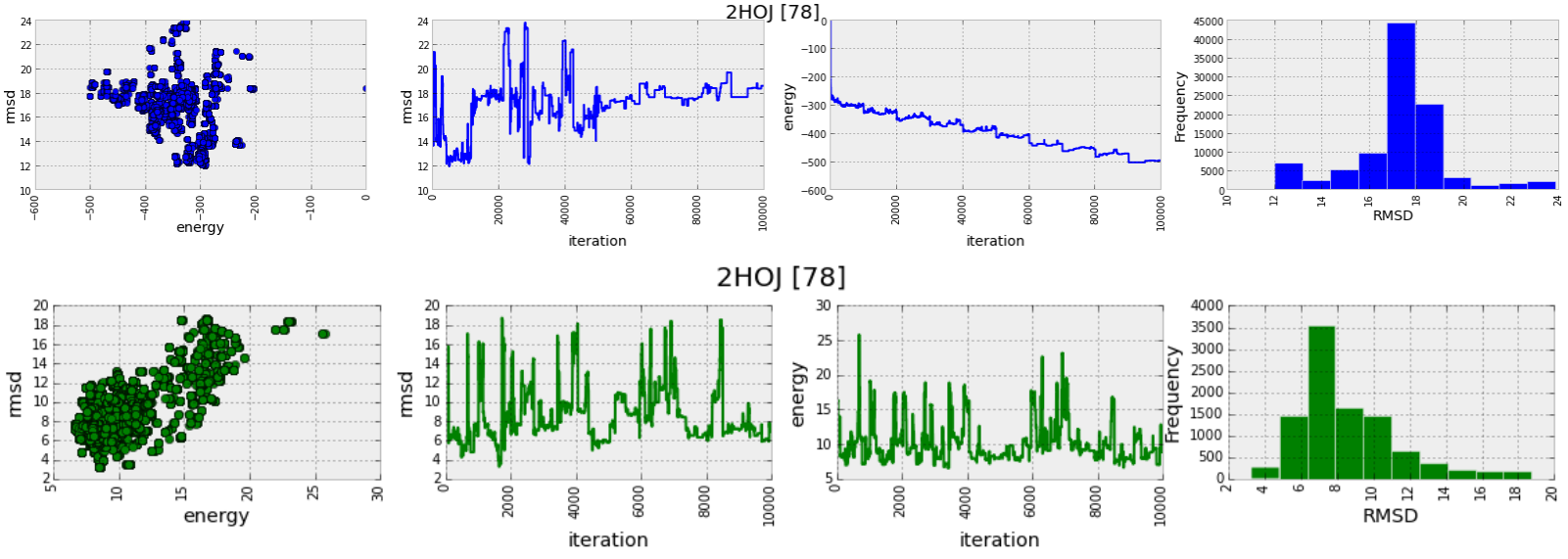
How many interactions do we expect a loop to be involved in?



Other Energies

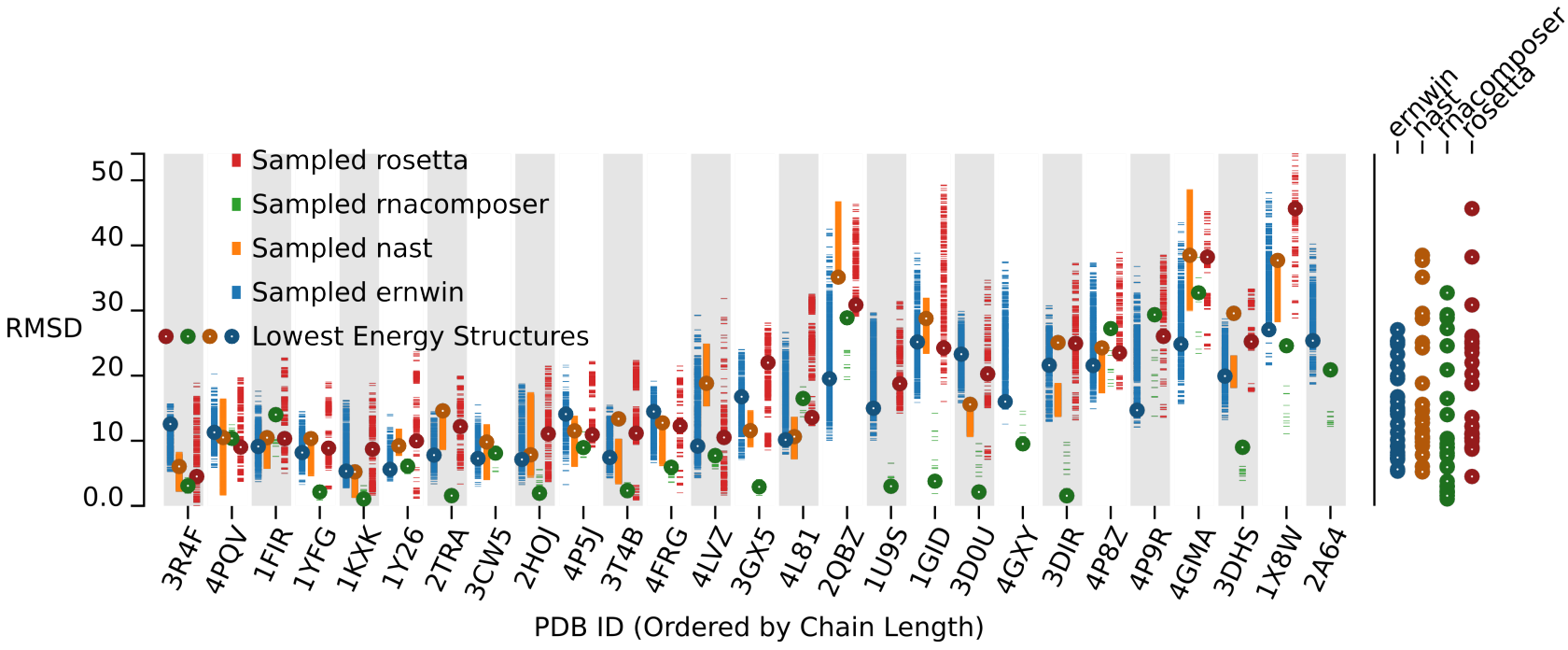


How Useful is An Energy?

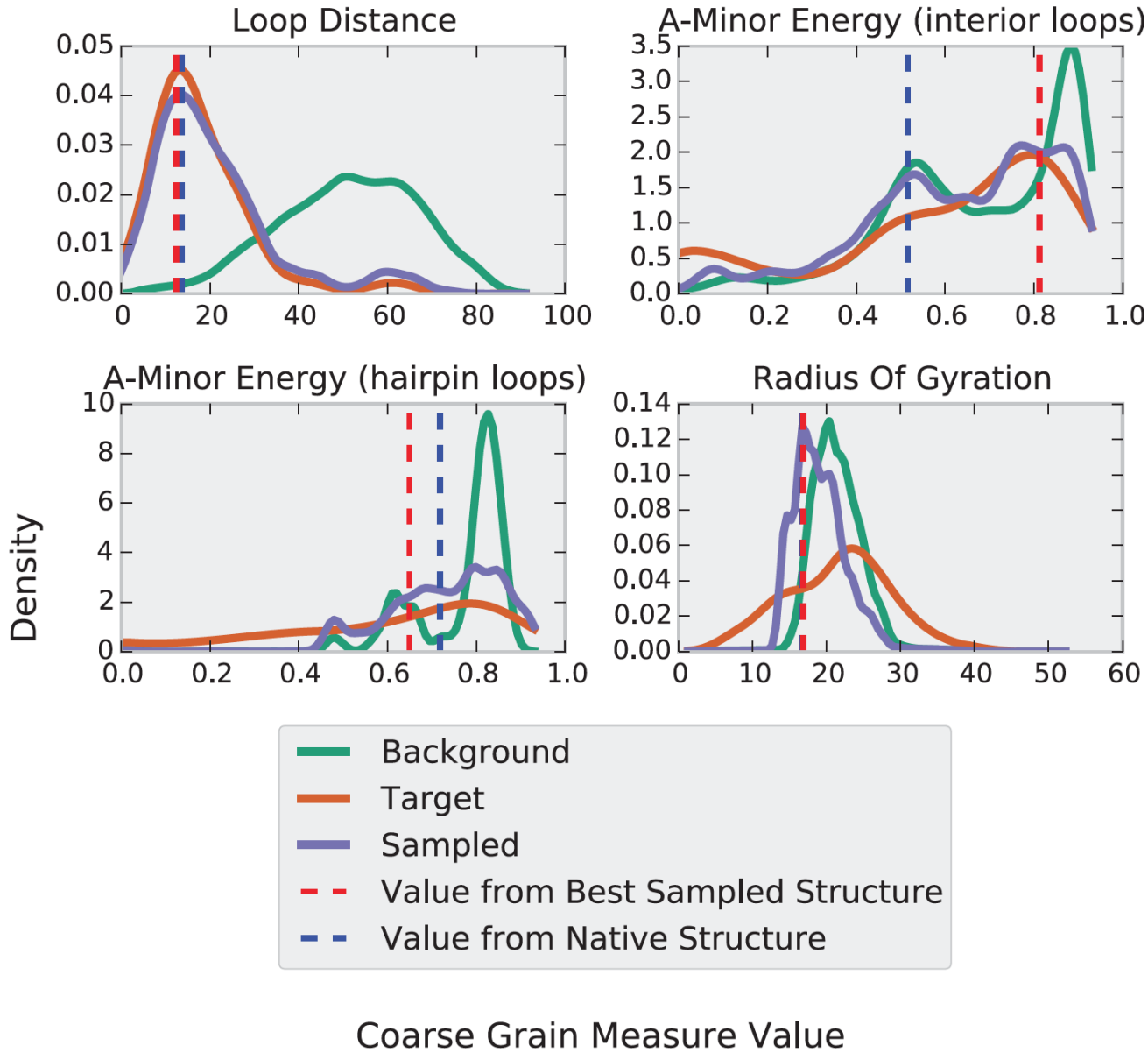


The ernwin energy has a better profile than the Rosetta energy.

Prediction Accuracy and Comparisons



Sampling Coarse-Grain Features



Summary

- Introduced highly coarse-grained helix based structure model
- Allows more complete sampling of configuration space
- Proposed long-range energies over a coarse-grain model
- Direct sampling of local structure and knowledge based potentials for long-range interactions

Details: Kerpedjiev et al., *RNA* **21**, pp 1110-1121, 2015

Code at `https://github.com/pkerpedjiev/ernwin`

Limitations and Improvements

- Conversion from coarse grained $\rightarrow$ atomic resolution and refinement
- Better long-range energy terms
- More sequence dependence
- Incorporation of interior loop motifs
- Incorporation of multi-loop motifs

Including 3D Motifs

Can we use 3D motif prediction to improve ernwin?

- Run JAR3D on each interior loop
  returns a list motifs and PDB ids for each motif
- Choose loop conformation from one of the JAR3D instances
- Compare predictions with/out JAR3D

Unfortunately, no JAR3D predictions for multi-loops yet.

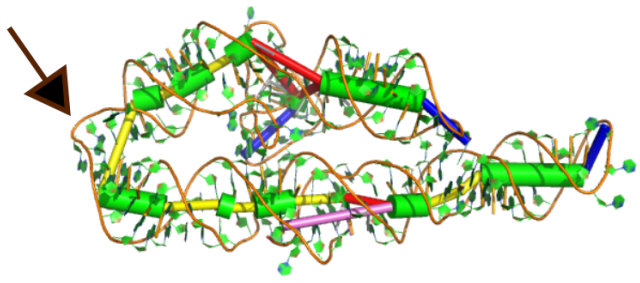
Including Motif Predictions from JAR3D

Name	Length	Int. Loops (found)	MCC Original	MCC JAR3D	MCC Change	RMSD Change
1GID_A	158	8(5)	0.67	0.79	0.12	-10.07
1X8W_A	242	10(6)	0.64	0.69	0.05	-2.94
4GMA_Z	192	8(3)	0.68	0.71	0.04	-3.79
3T4B_A	83	2(1)	0.89	0.91	0.02	-1.92
3D0U_A	161	7(0)	0.75	0.76	0.01	-5.52
4GXY_A	161	10(5)	0.70	0.71	0.01	0.33
4LVZ_A	89	2(1)	0.82	0.83	0.01	-1.74
2TRA_A	73	0(0)	0.90	0.91	0.01	-0.80
2HOJ_A	78	3(0)	0.90	0.91	0.01	-0.05
4L81_A	96	2(2)	0.82	0.83	0.00	-1.98
1Y26_X	71	1(0)	0.98	0.97	0.00	1.00
3DHS_A	215	6(3)	0.75	0.74	0.00	0.60
4P5J_A	83	0(0)	0.86	0.85	0.00	0.15
3CW5_A	75	0(0)	0.91	0.90	-0.01	0.29
1U9S_A	155	8(3)	0.80	0.79	-0.01	-2.01
1KXK_A	70	4(2)	0.88	0.86	-0.02	0.37
4P9R_A	189	3(3)	0.72	0.70	-0.02	-1.16
4PQV_A	68	0(0)	0.87	0.84	-0.03	0.15
3DIR_A	172	6(3)	0.81	0.76	-0.05	0.99
3GX5_A	94	4(2)	0.88	0.79	-0.09	1.38

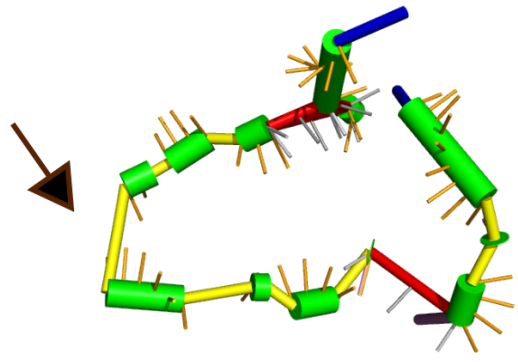
MCCs computed by converting 3D structures into contact maps

Correctly Predicted Kink-Turn Motif

Struct Name	Struct Length	Interior Loops	-MCC Change	RMSD Change
1GID_A	158	3	-0.12	-10.07



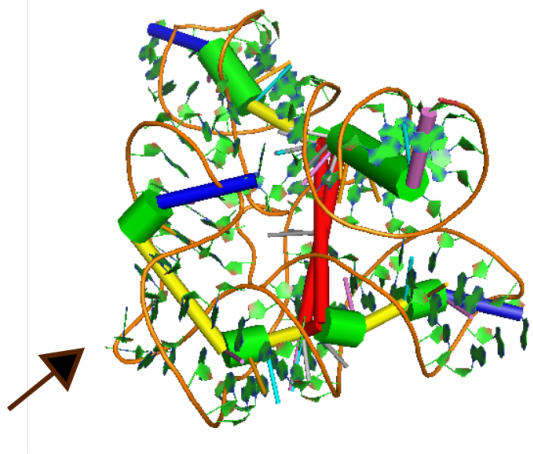
Crystal Structure



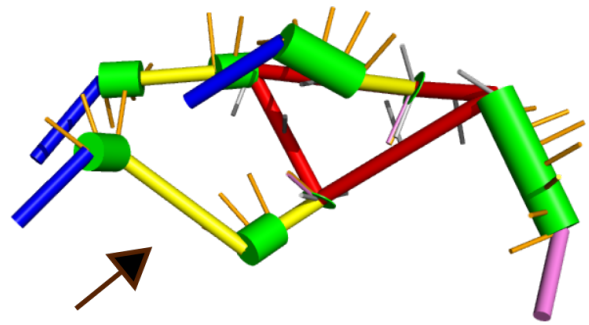
Best Prediction

Incorrectly Predicted Interior Loop

Struct Name	Struct Length	Interior Loops	-MCC Change	RMSD Change
3GX5_A	94	2	0.09	1.38



Crystal Structure



Best Prediction