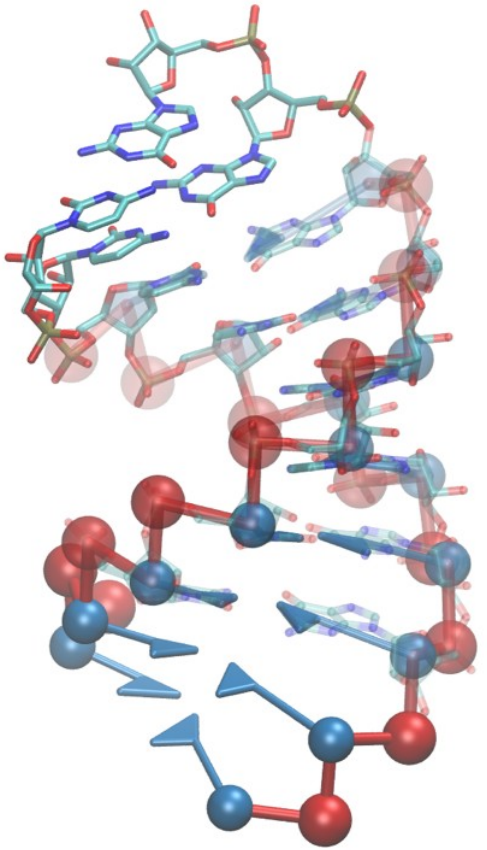
Conference on Atomistic Simulations of Biomolecules

ICTP, Trieste, March 2017

Introducing atomistic details into a base-centered representation of RNA

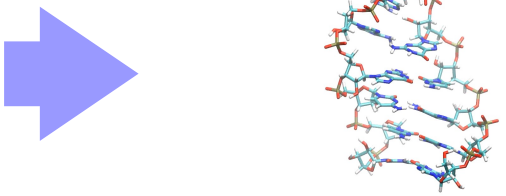


Simón Poblete, Sandro Bottaro
and Giovanni Bussi.

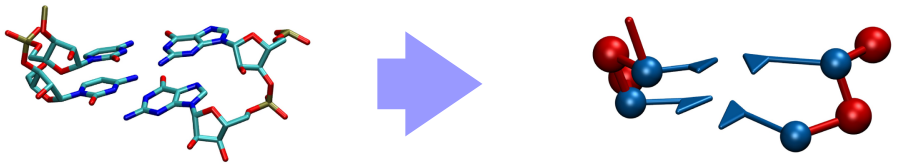
Molecular and Statistical Biophysics - SISSA, Trieste, Italy

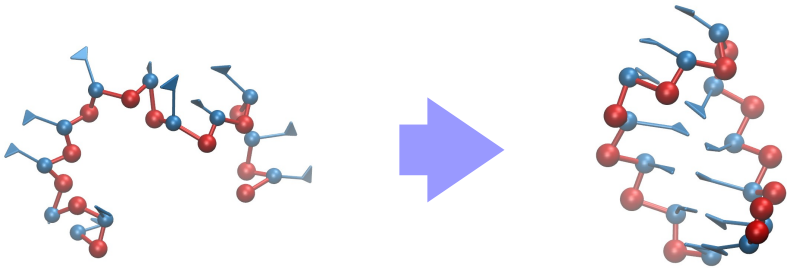
My talk in a nutshell

1) We want to predict 3D structure of RNA fragments from their sequence.

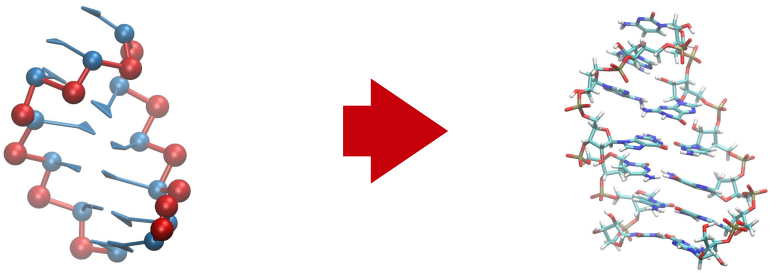
GGGCGCAAGCCU 

2) We use a simplified, coarse-grained model for folding the structure.





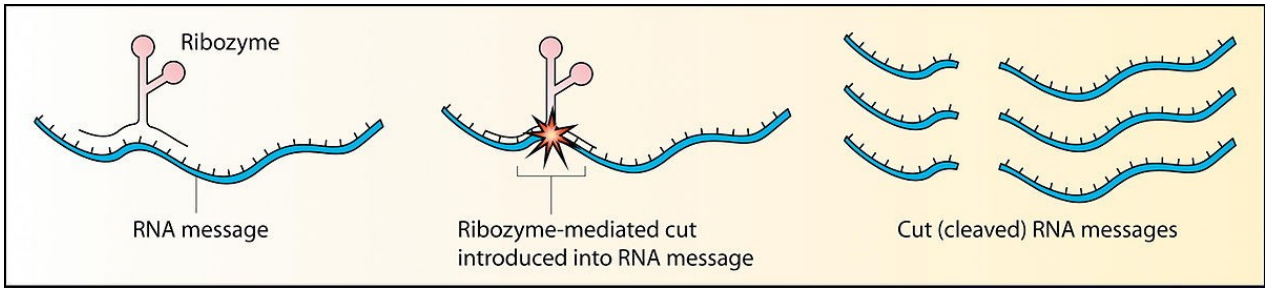
3) Once we have the final prediction, we would like to reinsert the atomistic details.

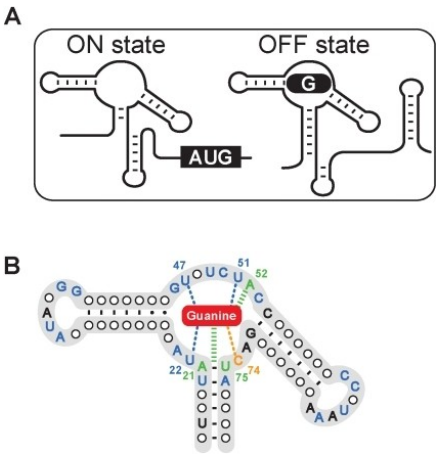


Introduction

RNA has a considerable biological importance :

- Transcription
- Translation
- Catalytic role (ribozymes)
- Regulatory function (miRNAs, riboswitches, ...)

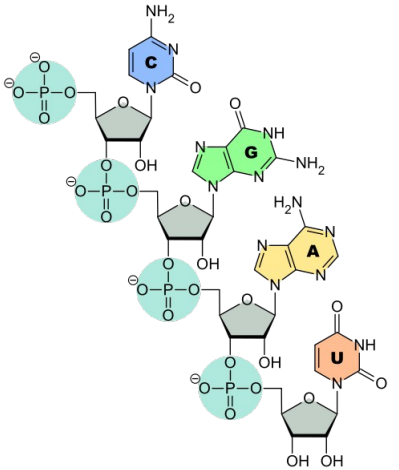


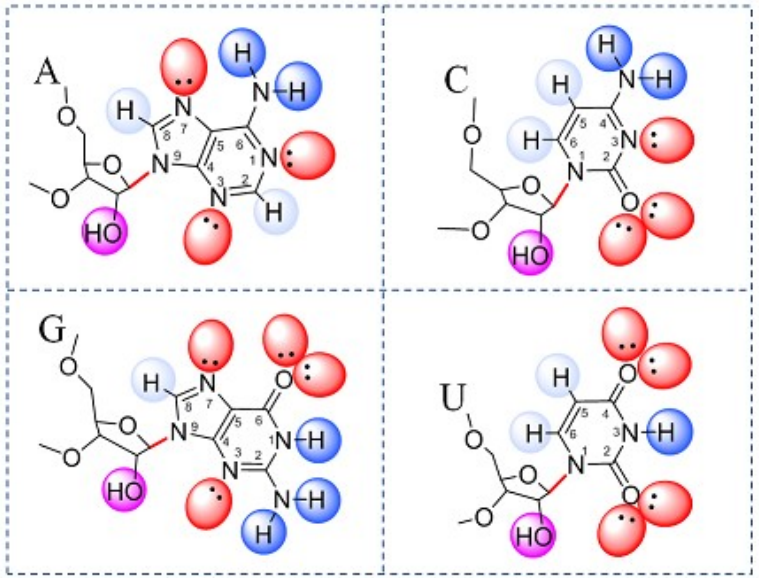


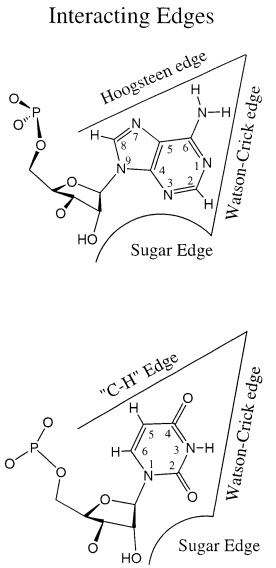
Structure determines function!

Something about RNA...

- Structure: Sugar, Base and Phosphate group.
- Differences with DNA:
  - Ribose sugar instead of deoxyribose : more reactive.
  - Uracil instead of Thymine.
- Rich set of base-base interactions





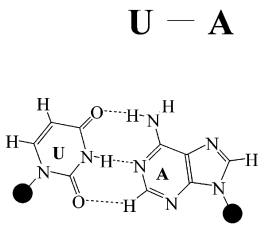


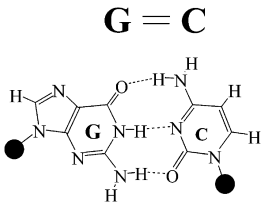
Leontis, Stombaugh and Westhof, Nucleic Acids Res. (2002)

Something about RNA...

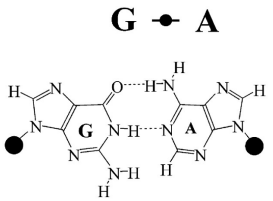
Rich set of base-base interactions

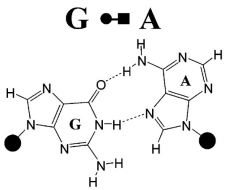
No.	Glycosidic bond orientation	Interacting edges		Abbreviation	Symbol	Triangle abstraction
		NT1	NT2			
1	*Cis*	Watson–crick	Watson–crick	cWW		
2	*Trans*	Watson–crick	Watson–crick	tWW		
3	*Cis*	Watson–crick	Hoogsteen	cWH		
		Hoogsteen	Watson–crick	cHW		
4	*Trans*	Watson–crick	Hoogsteen	tWH		
		Hoogsteen	Watson–crick	tHW		
5	*Cis*	Watson–crick	Sugar edge	cWS		
		Sugar edge	Watson–crick	cSW		
6	*Trans*	Watson–crick	Sugar edge	tWS		
		Sugar edge	Watson–crick	tSW		
7	*Cis*	Hoogsteen	Hoogsteen	cHH		
8	*Trans*	Hoogsteen	Hoogsteen	tHH		
9	*Cis*	Hoogsteen	Sugar edge	cHS		
		Sugar edge	Hoogsteen	cSH		
10	*Trans*	Hoogsteen	Sugar edge	tHS		
		Sugar edge	Hoogsteen	tSH		
11	*Cis*	Sugar edge (priority)	Sugar edge	cSs		
		Sugar edge	Sugar edge (priority)	csS		
12	*Trans*	Sugar edge (priority)	Sugar edge	tSs		
		Sugar edge	Sugar edge (priority)	tsS		

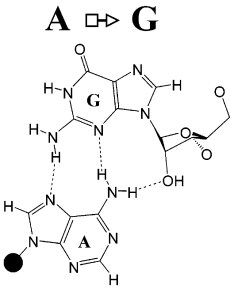




Canonical







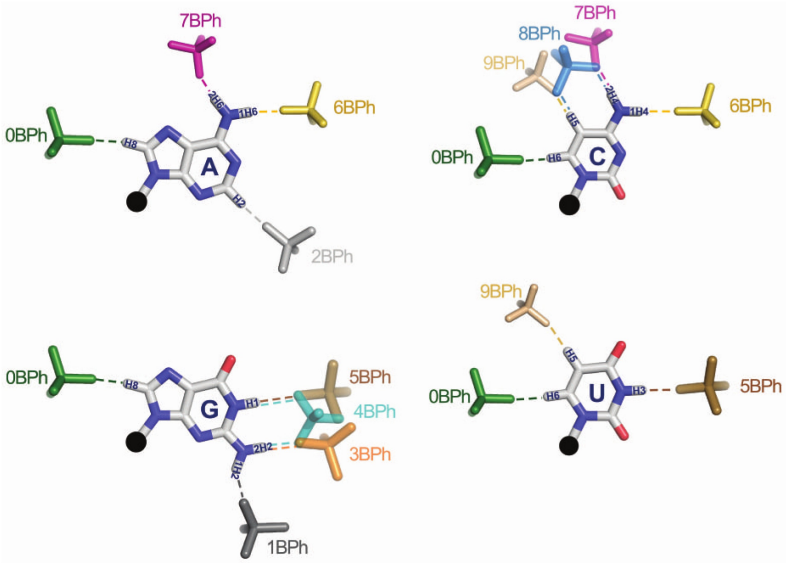
cWW cWH tHS

Non-canonical

Leontis, Stombaugh and Westhof, Nucleic Acids Res. (2002)

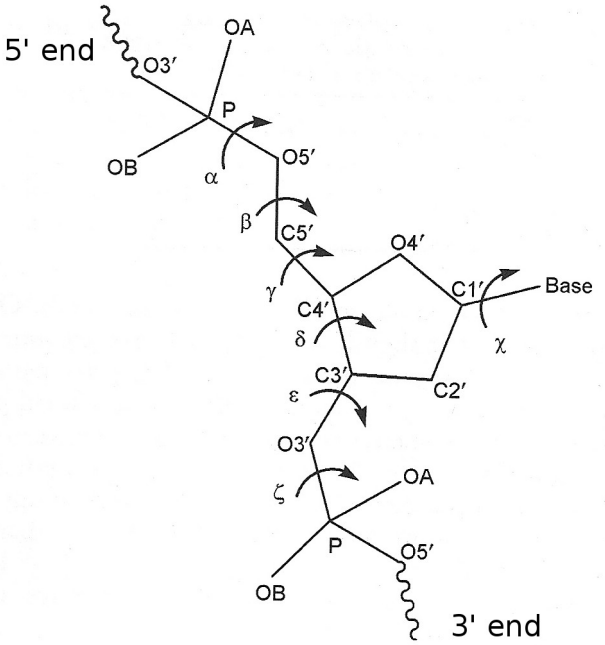
Something about RNA...

Also, interactions between bases and phosphate groups



Zirbel, Sponer, Sponer, Stombaugh and Leontis, Nucleic Acids Res. (2009).

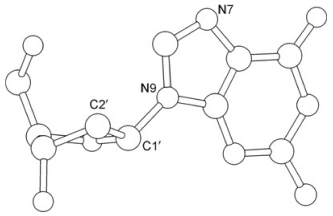
Backbone conformation defined by 6 (7) dihedral angles

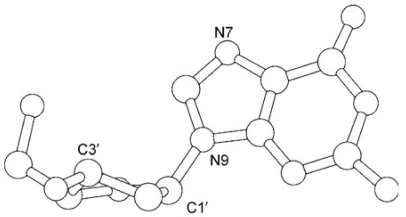


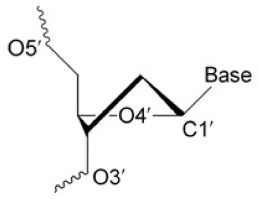
Richardson et al., RNA (2008).

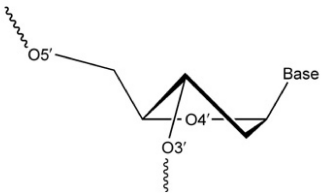
Something about RNA

Sugar pucker







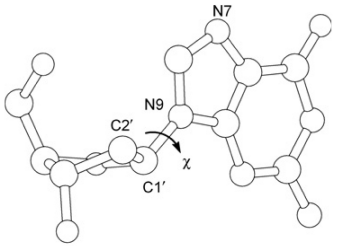


C2' endo

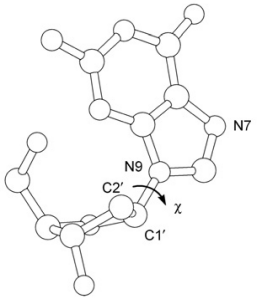
C3' endo

Glycosidic bond angle

anti



syn



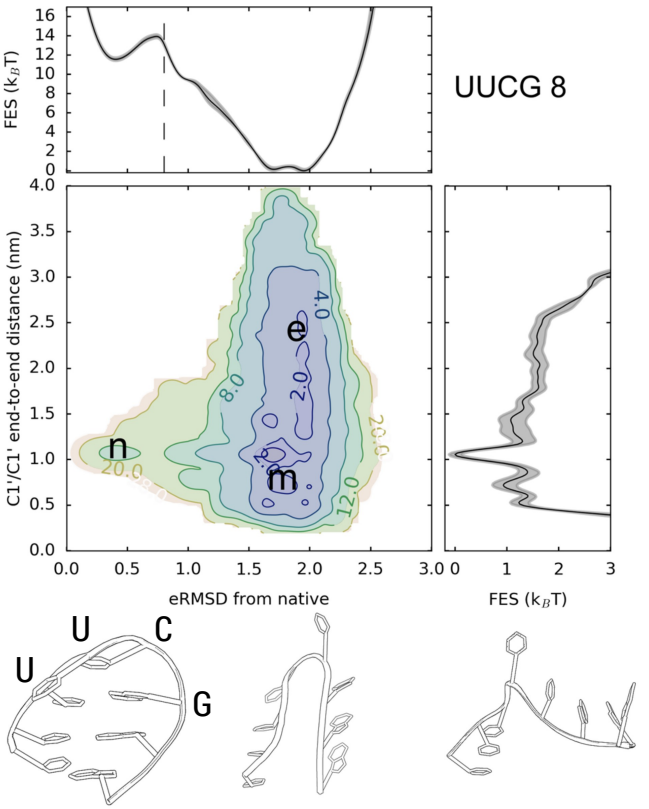
high-anti (intermediate)

Motivation

- Structure prediction is an important, but tough task. Brute-force Atomistic MD might be extremely inneficient.
- RNA atomistic force fields still require refinement.

UUCG tetraloop :

Misfolded structures are more stable than native

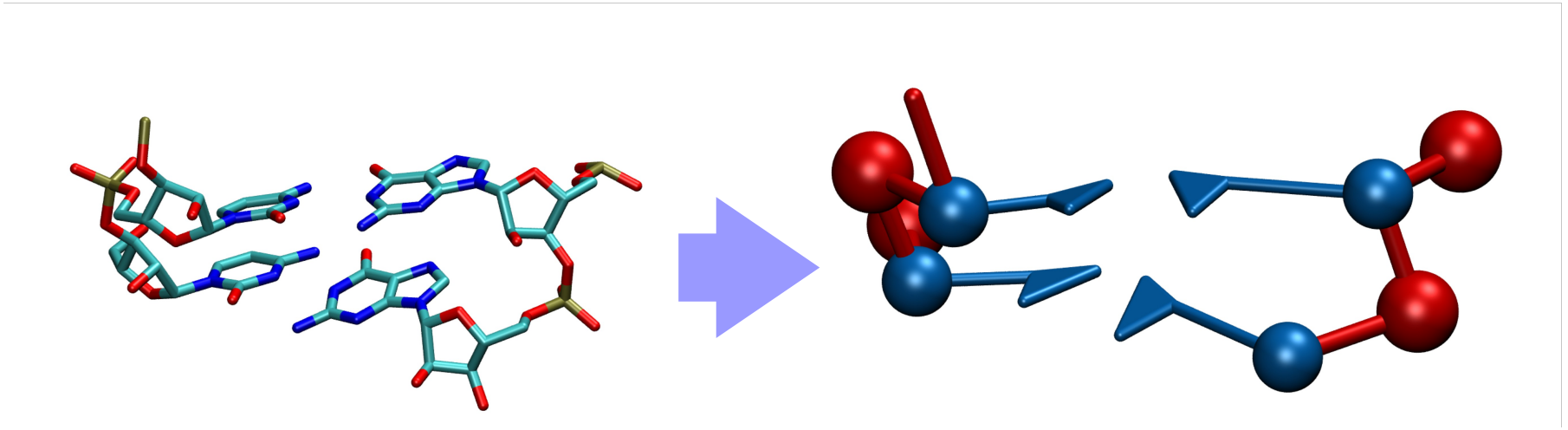


Bottaro, Banas, Sponer and Bussi, J. Phys. Chem. Lett. (2016)

Coarse-grained model

A simpler, faster representation, with effective interactions.

SPQR model : a nucleoside is an anisotropic particle; phosphate group is a point particle.



Thoroughly description of canonical and non-canonical base pairs.

Base-phosphate interactions

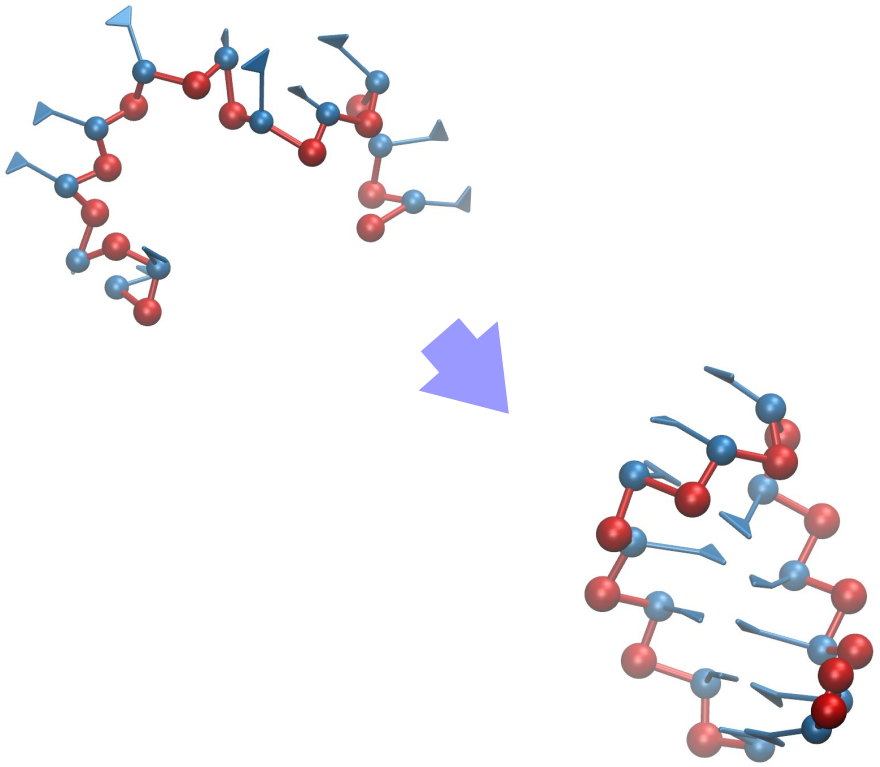
Puckers and glycosidic bond angles specified

Poblete, Bottaro and Bussi, submitted.

Coarse-grained model

A simpler, faster representation, with effective interactions.

SPQR model : a nucleoside is an anisotropic particle; phosphate group is a point particle.

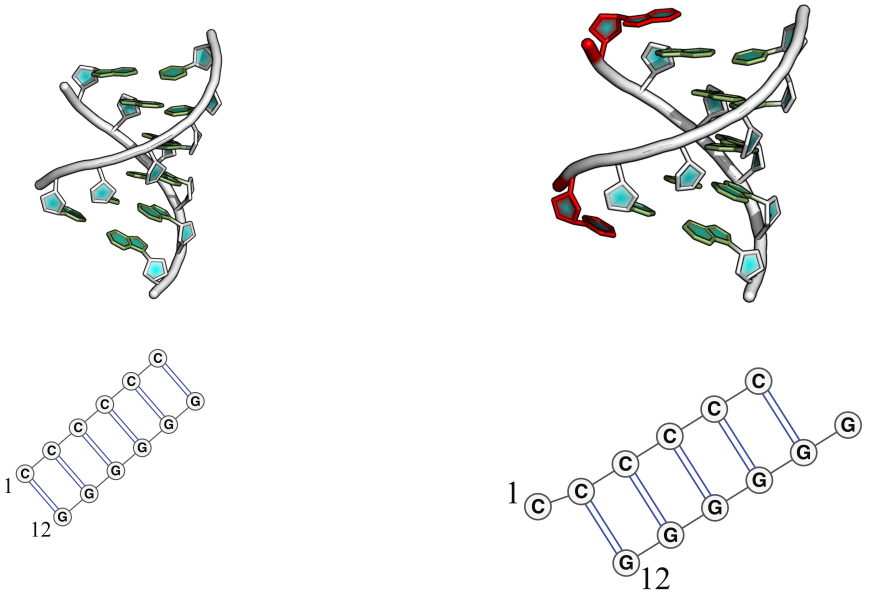


Simulated annealing to predict structures of minimum energy

GCAA tetraloop
(pdb 1ZIH)

Backmapping

- Once the structure is predicted, we wish to reintroduce atomistic detail.
- Minimize “distance” between atomistic and predicted structure.
- RMSD : Misleading, not suitable to compare nucleic acids structure

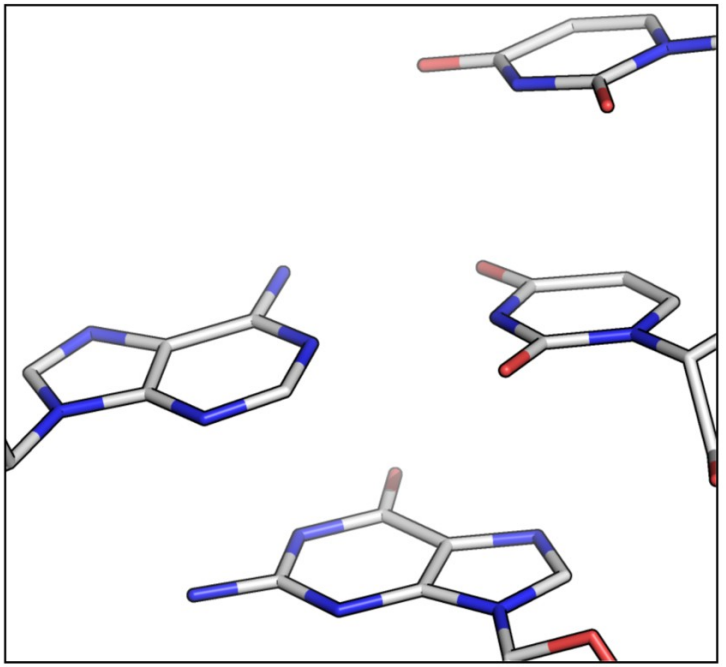


1.9 A (backbone-only 1.4 A, bases-only 1.9)

Bottaro, Di Palma and Bussi, Nucleic Acids Res. (2014).

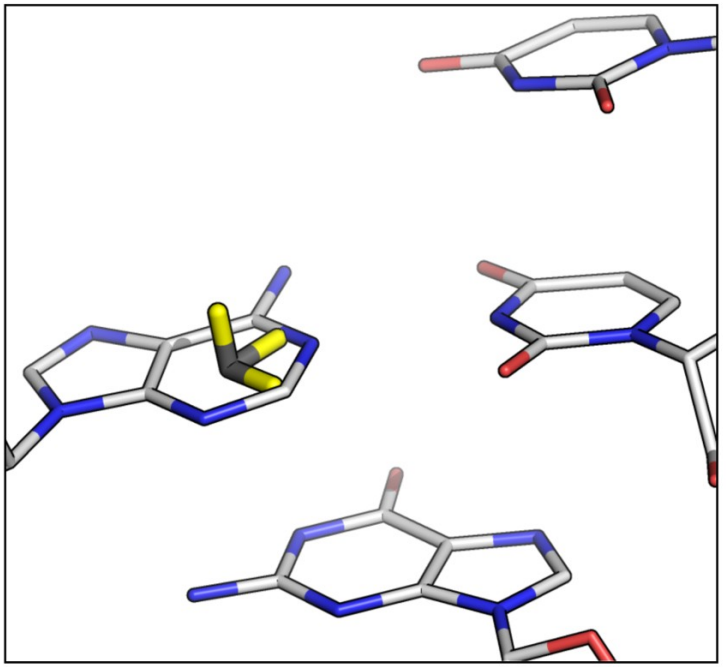
ε RMSD

A simplified geometrical description of base arrangement



ε RMSD

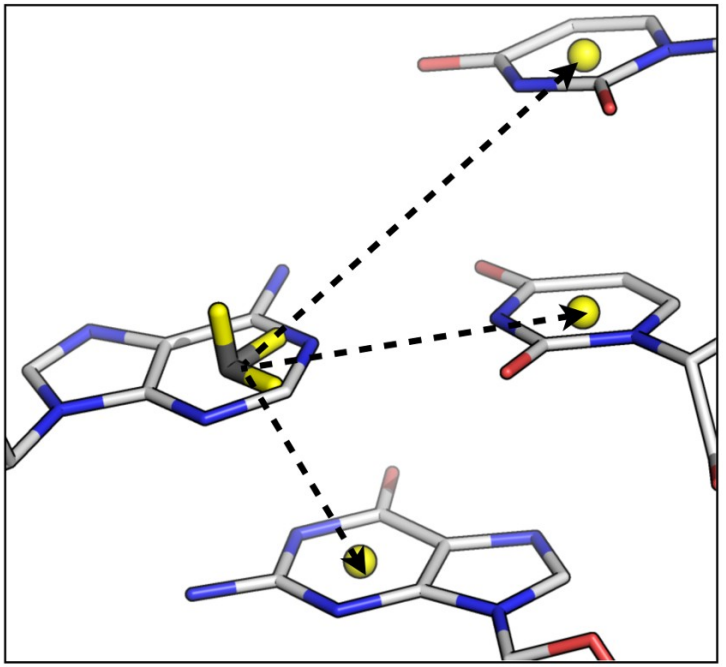
A simplified geometrical description of base arrangement



Sit on a base

ε RMSD

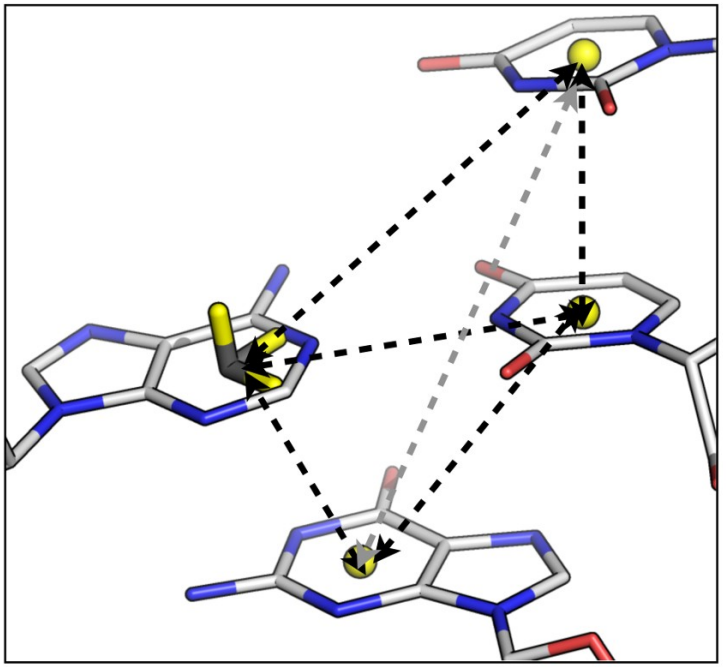
A simplified geometrical description of base arrangement



Calculate the distance vectors from here

ℰRMSD

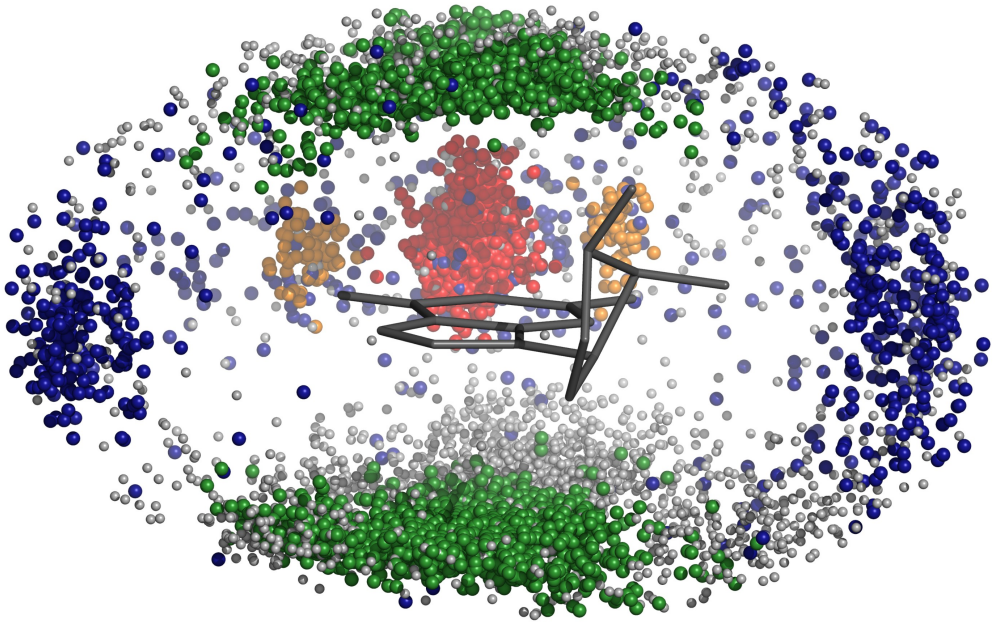
A simplified geometrical description of base arrangement



Repeat the procedure inside a cutoff distance

εRMSD

A simplified geometrical description of base arrangement



Ellipsoidal distance within a cutoff,

Captures better base pairing and stacking,

And describe better the nucleobase volume

Data from large ribosomal subunit
PDB:1S72, Klein, Moore, and Steitz, J. Mol. Biol. (2004).

Bottaro, Di Palma and Bussi, Nucleic Acids Res. (2014).
Bottaro, Di Palma and Bussi, RNA and Disease (2015).

$\mathcal{E}$RMSD

With this collection of vectors, we introduce the $\mathcal{E}$RMSD, a more specific distance between RNA structures.

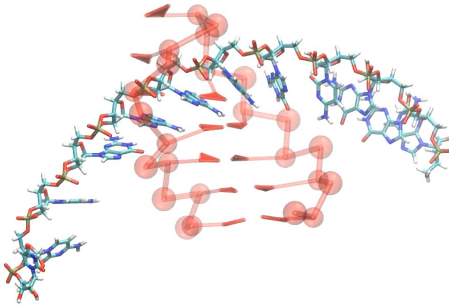
$$\mathbf{G}(\tilde{\mathbf{r}}) = \begin{pmatrix} \sin(\gamma\tilde{r})\frac{\tilde{r}_x}{\tilde{r}} \\ \sin(\gamma\tilde{r})\frac{\tilde{r}_y}{\tilde{r}} \\ \sin(\gamma\tilde{r})\frac{\tilde{r}_z}{\tilde{r}} \\ 1+\cos(\gamma\tilde{r}) \end{pmatrix} \times \frac{\Theta(\tilde{r}_{\text{cutoff}} - \tilde{r})}{\gamma}$$

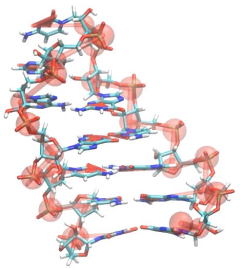
$$\gamma = \pi / \tilde{r}_{\text{cutoff}}$$

$$\mathcal{E}\text{RMSD} = \sqrt{\frac{1}{N}\sum_{j,k} |\mathbf{G}(\tilde{\mathbf{r}}^{\alpha}_{jk}) - \mathbf{G}(\tilde{\mathbf{r}}^{\beta}_{jk})|^2}$$

Backmapping procedure:

Steered MD, minimizing the ERMSD with respect to a reference structure (typically, predicted with our coarse-grained method)

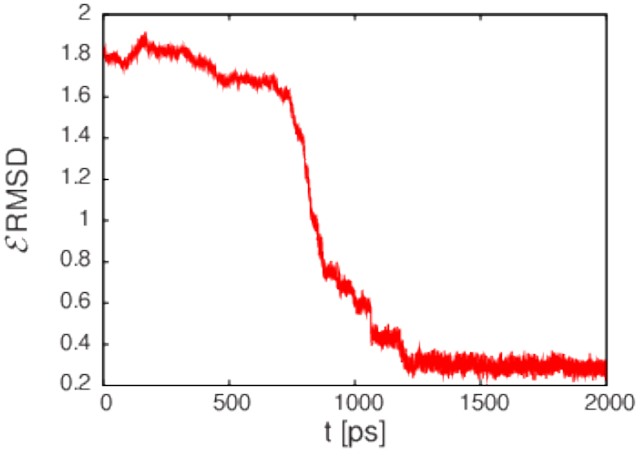




Short simulation of 3 ns

Explicit solvent, GROMACS 4.6.7 (*), harmonic constraint implemented in PLUMED(**)

$$V(\vec{s},t) = \frac{1}{2}\kappa(t)(\vec{s} - \vec{s}_0(t))^2$$



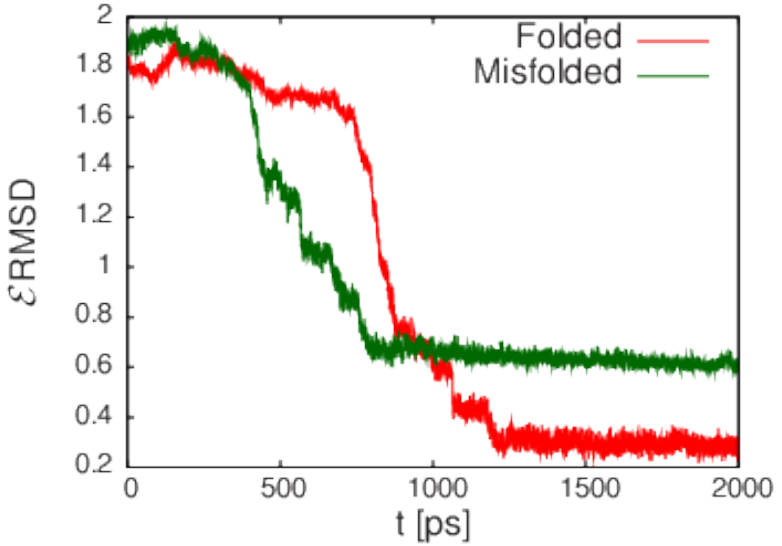
* www.gromacs.org, Amber99 (parmbsc0+OL3), TIP3P water, Na+ ions (Joung and Cheatham, J. Phys. Chem. B (2008)).

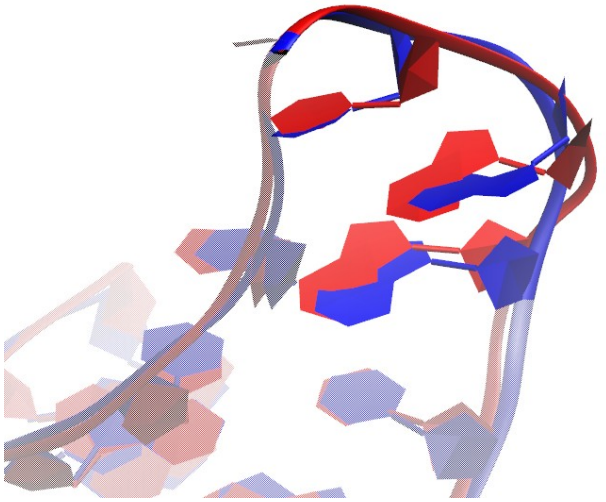
** www.plumed.org

Backmapping procedure:

Steered MD, minimizing the $\mathcal{E}$RMSD with respect to a reference structure (typically, predicted with our coarse-grained method).

Misfolded cases of bases not flipped are detected





Backmapping: test

What happens with the backbone conformations?

We performed a test on :

Five crystal structures to their native state.

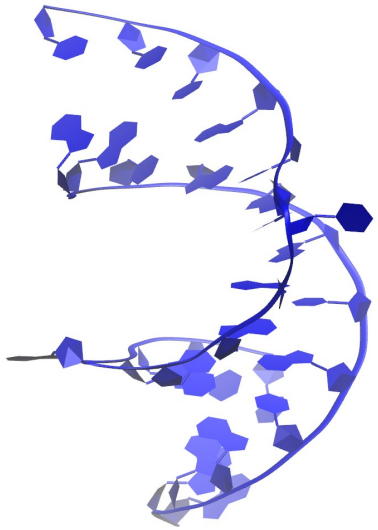
Three predicted structures from **SPQR** CG annealing.

Ran 6 independent simulations. We chose the 50 lowest $\mathcal{E}$RMSD configurations, and analyze their pucker and glycosidic bond angle using **dangle**.

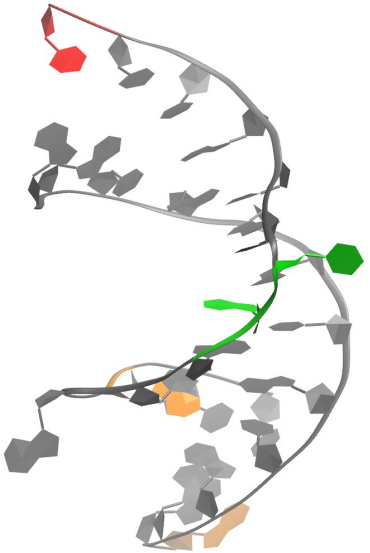
http://kinemage.biochem.duke.edu/software/dangle.help.html

Backmapping: test

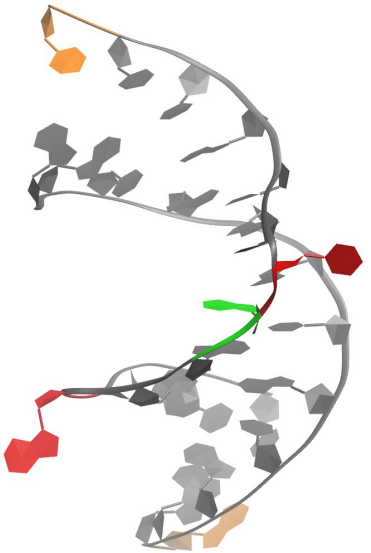
Backmapping on the crystal structures



Native



Sugar puckers



Glycosidic bond angle conformation

In the lowest 50 $\mathcal{E}$RMSD configurations, right conformation found in:

More than 80%

Between 50% and 80%

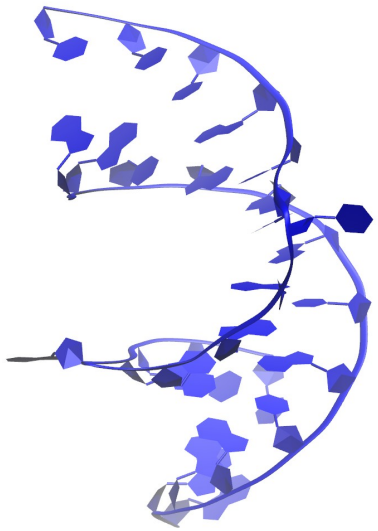
Less than 50%

Well predicted, uncommon cases

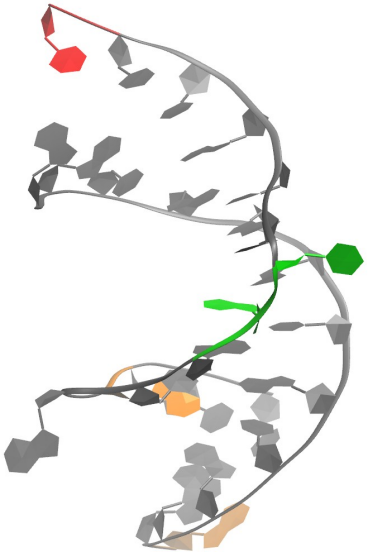
PDB:1csl

Backmapping: test

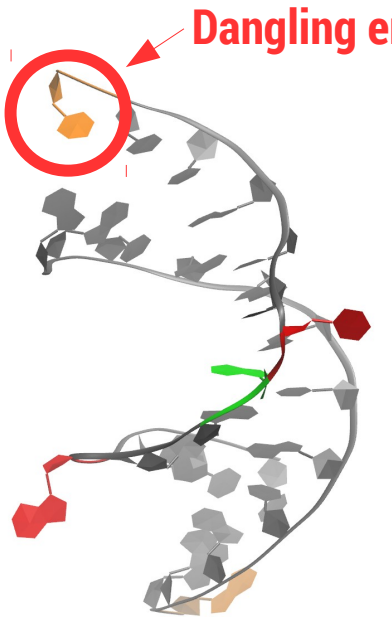
Backmapping on the crystal structures



Native



Sugar puckers



Glycosidic bond angle conformation

In the lowest 50 $\mathcal{E}$RMSD configurations, right conformation found in:

More than 80%

Between 50% and 80%

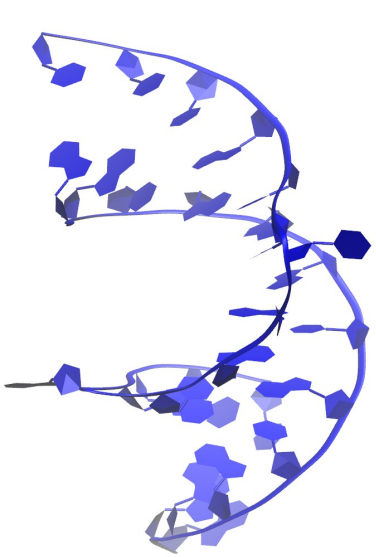
Less than 50%

Well predicted, uncommon cases

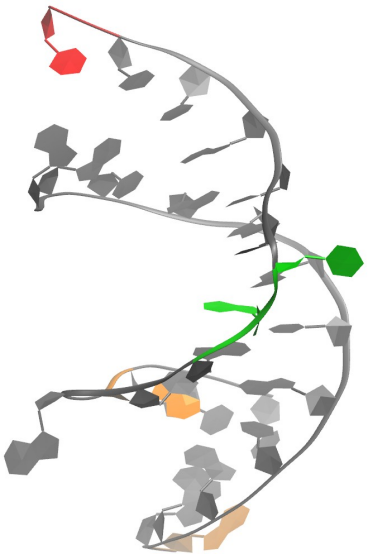
PDB:1csl

Backmapping: test

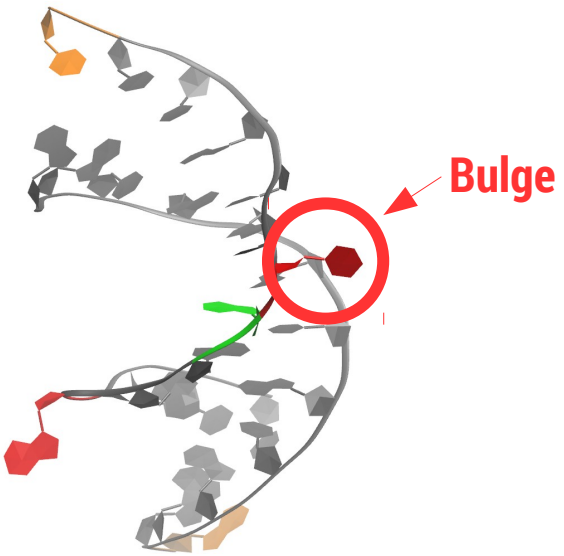
Backmapping on the crystal structures



Native



Sugar puckers

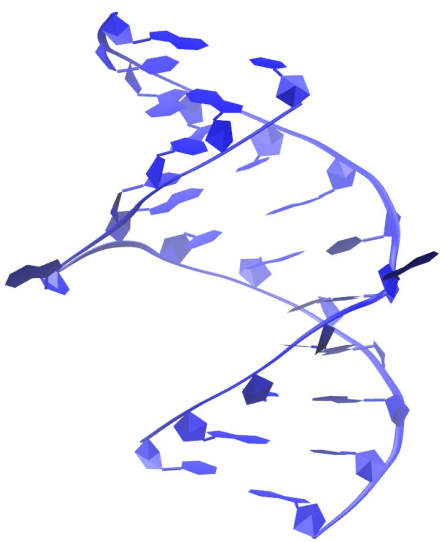


Glycosidic bond angle conformation

PDB:1csl

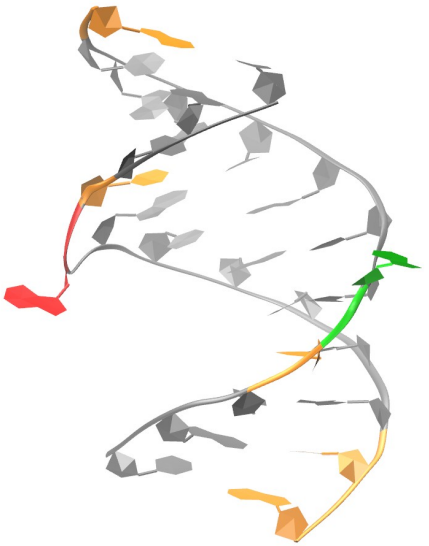
Backmapping: test

Backmapping on the crystal structures



Native

Sugar puckers



Glycosidic bond angle conformation

In the lowest 50 $\mathcal{E}$RMSD configurations, right conformation found in:

More than 80%

Between 50% and 80%

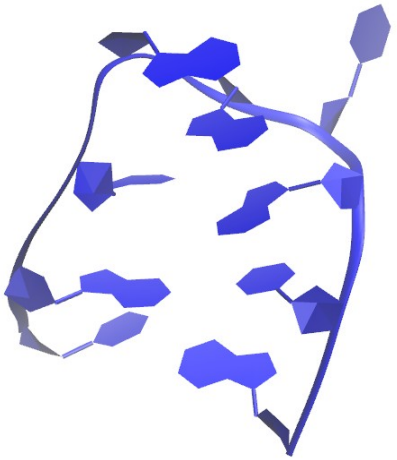
Less than 50%

Well predicted, uncommon cases

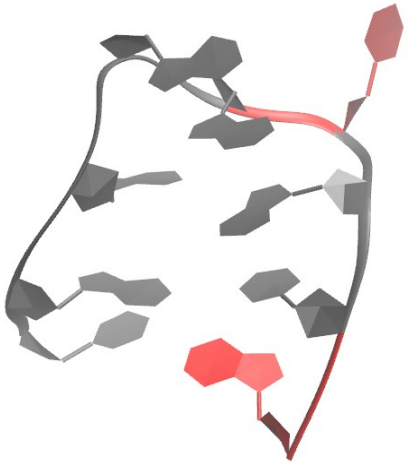
PDB:1i9x

Backmapping: test

Backmapping on the crystal structures



Native



Sugar puckers

Glycosidic bond angle conformation

In the lowest 50 $\mathcal{E}$RMSD configurations, right conformation found in:

More than 80%

Between 50% and 80%

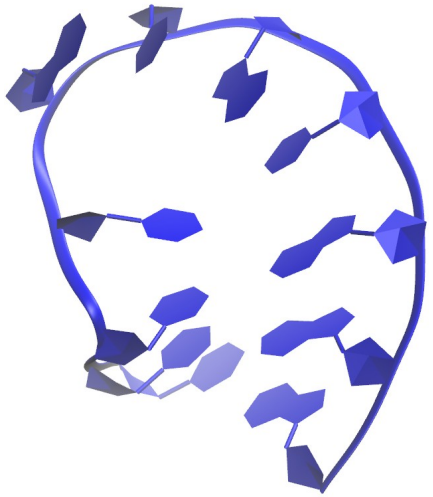
Less than 50%

Well predicted, uncommon cases

PDB:1xjr

Backmapping: test

Backmapping on the crystal structures



Native

Sugar puckers

Glycosidic bond angle conformation

In the lowest 50 $\mathcal{E}$RMSD configurations, right conformation found in:

More than 80%

Between 50% and 80%

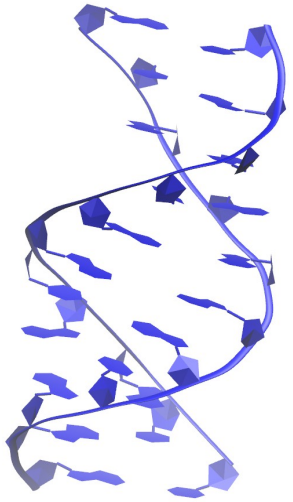
Less than 50%

Well predicted, uncommon cases

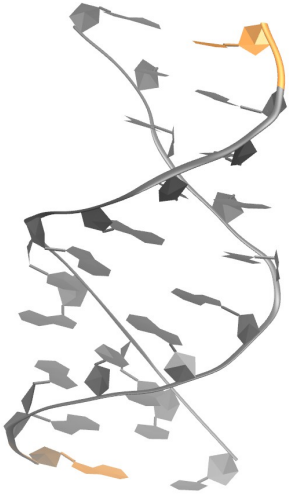
PDB:2gdi

Backmapping: test

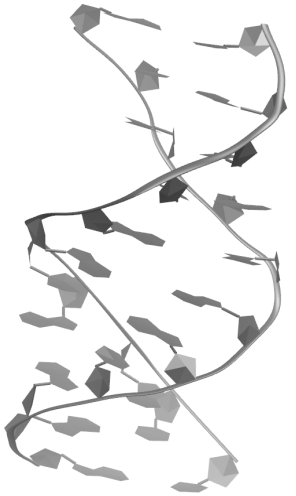
Backmapping on the crystal structures



Native



Sugar puckers



Glycosidic bond angle conformation

In the lowest 50 $\mathcal{E}$RMSD configurations, right conformation found in:

More than 80%

Between 50% and 80%

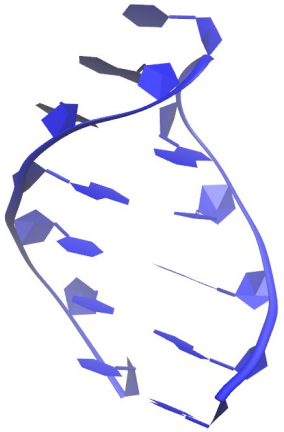
Less than 50%

Well predicted, uncommon cases

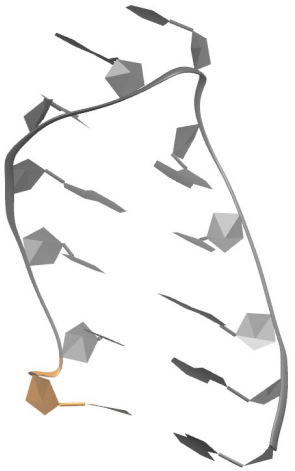
PDB:354d

Backmapping

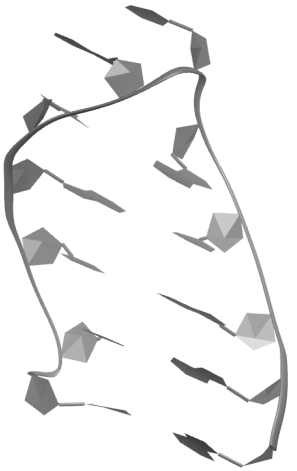
Backmapping on the predicted structures from coarse-grained model.



Native



Sugar puckers



Glycosidic bond angle conformation

In the lowest 50 $\mathcal{E}$RMSD configurations, right conformation found in:

More than 80%

Between 50% and 80%

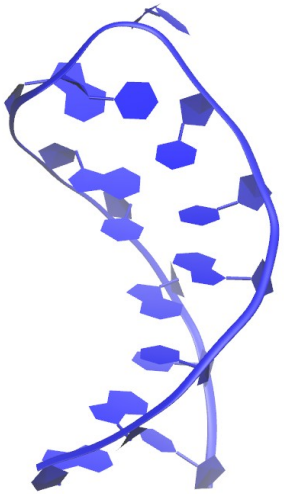
Less than 50%

Well predicted, uncommon cases

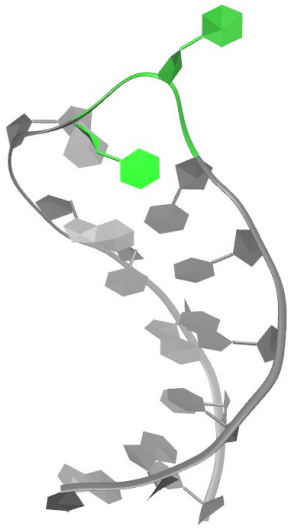
PDB:1zih

Backmapping

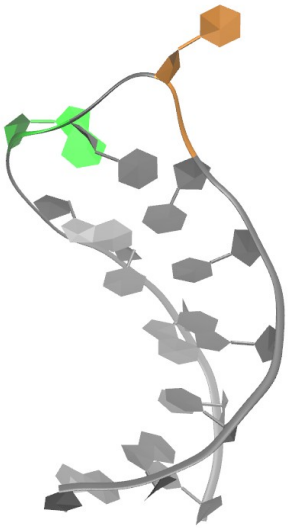
Backmapping on the predicted structures from coarse-grained model.



Native



Sugar puckers



Glycosidic bond angle conformation

In the lowest 50 $\mathcal{E}$RMSD configurations, right conformation found in:

More than 80%

Between 50% and 80%

Less than 50%

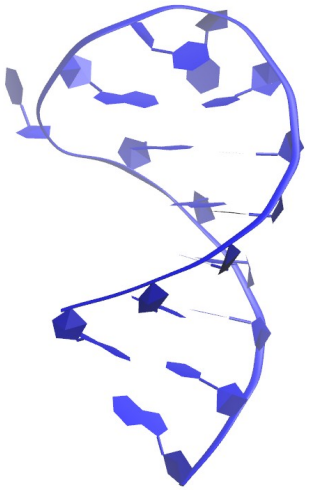
Well predicted, uncommon cases

PDB:2koc
(UUCG tetraloop)

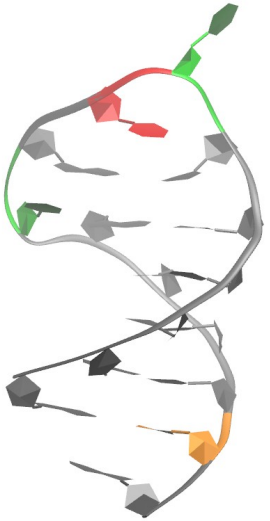
"Flipped" G and C, with not typical (A-form) glycosidic and pucker conformations, respectively.

Backmapping

Backmapping on the predicted structures from coarse-grained model.



Native



Sugar puckers

Glycosidic bond angle conformation

In the lowest 50 $\mathcal{E}$RMSD configurations, right conformation found in:

More than 80%

Between 50% and 80%

Less than 50%

Well predicted, uncommon cases

PDB:2l6i

Conclusions

- Introduced a backmapping procedure for nucleic acids, based on εRMSD, more suitable metric for RNA structures.
- Good results, but particularly difficult for bulges, dangling ends and end nucleotides in general.
- Effective in cases where bases are “flipped”.

Acknowledgements

European Research Council.