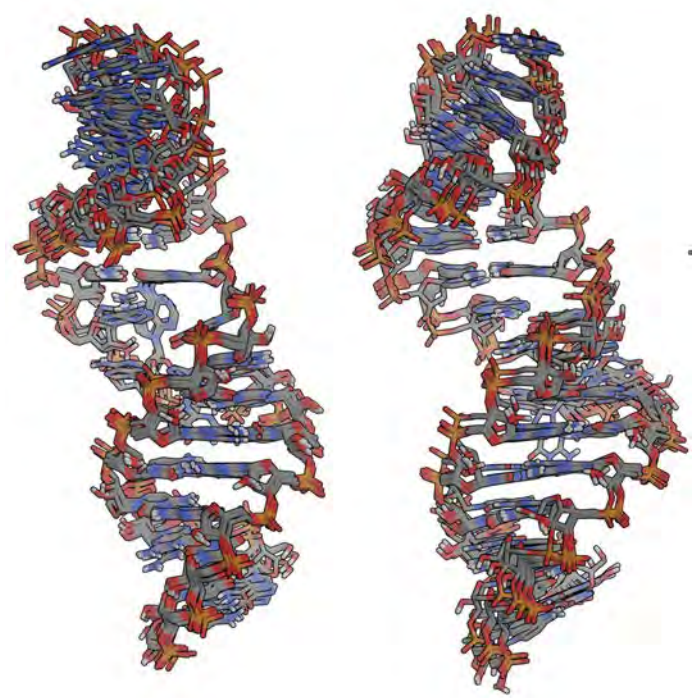
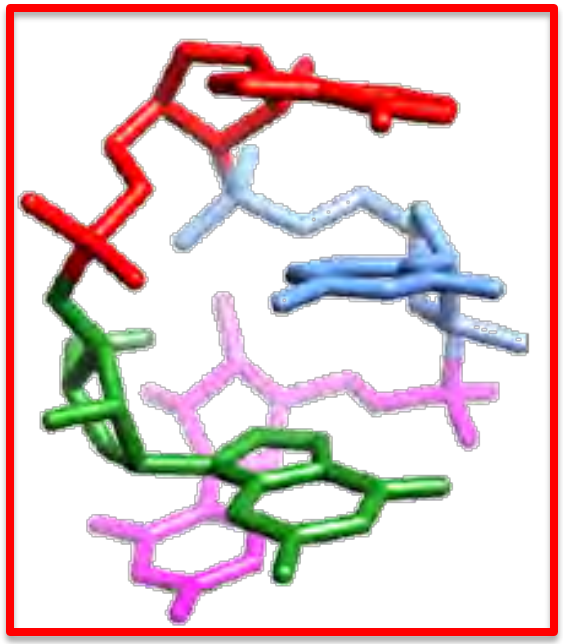
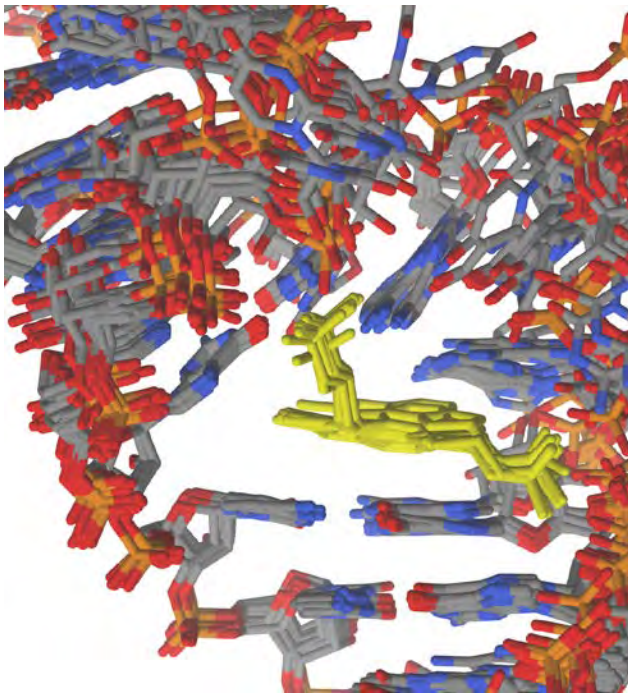
Blue Waters Symposia 2018

Using ensembles of molecular dynamics simulations to give insight into biomolecular structure, dynamics, and function







Thomas E. Cheatham III **tec3@utah.edu**

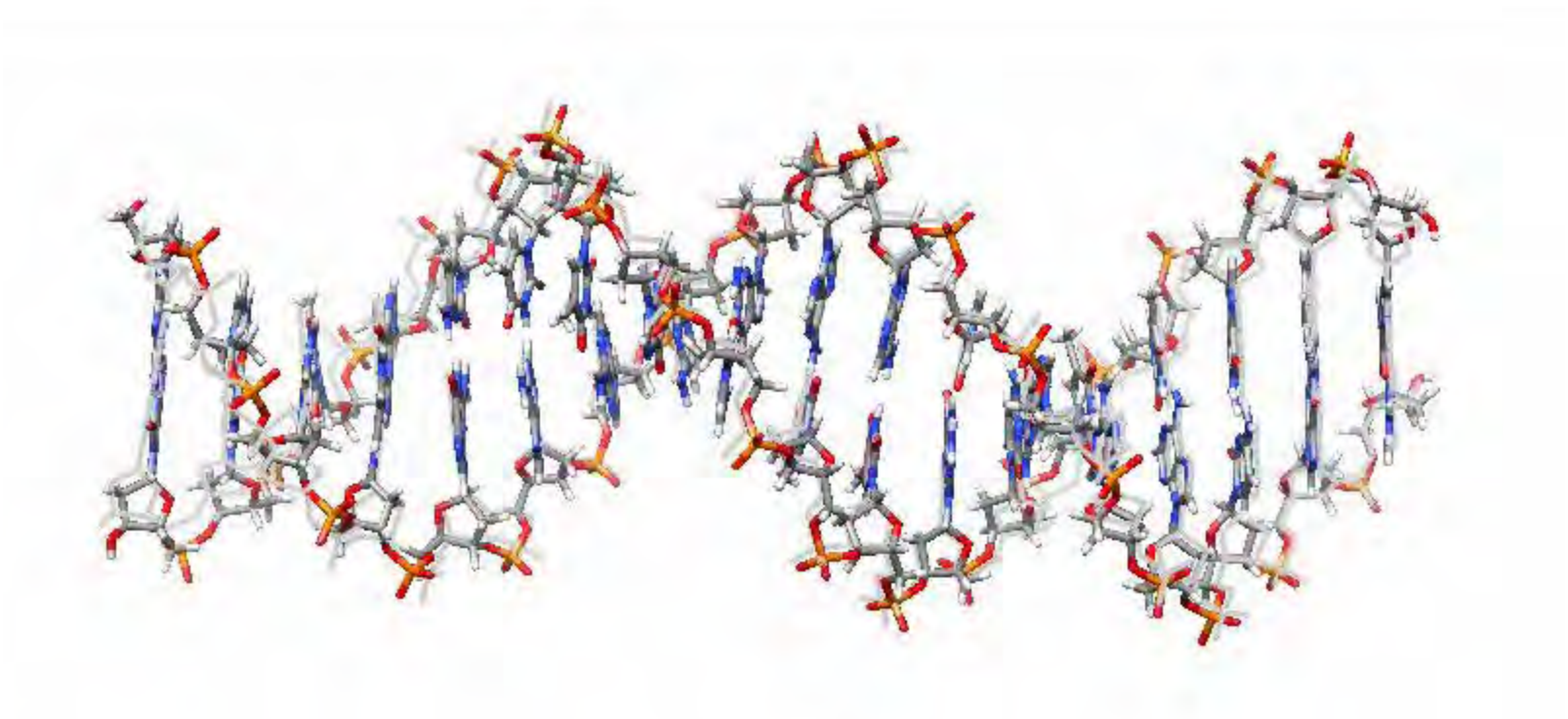
Professor of Medicinal Chemistry, College of Pharmacy

Director, Research Computing and CHPC,

University Information Technology, University of Utah

biomolecular simulation

…structure, dynamics, interactions, ΔG, sampling, force fields



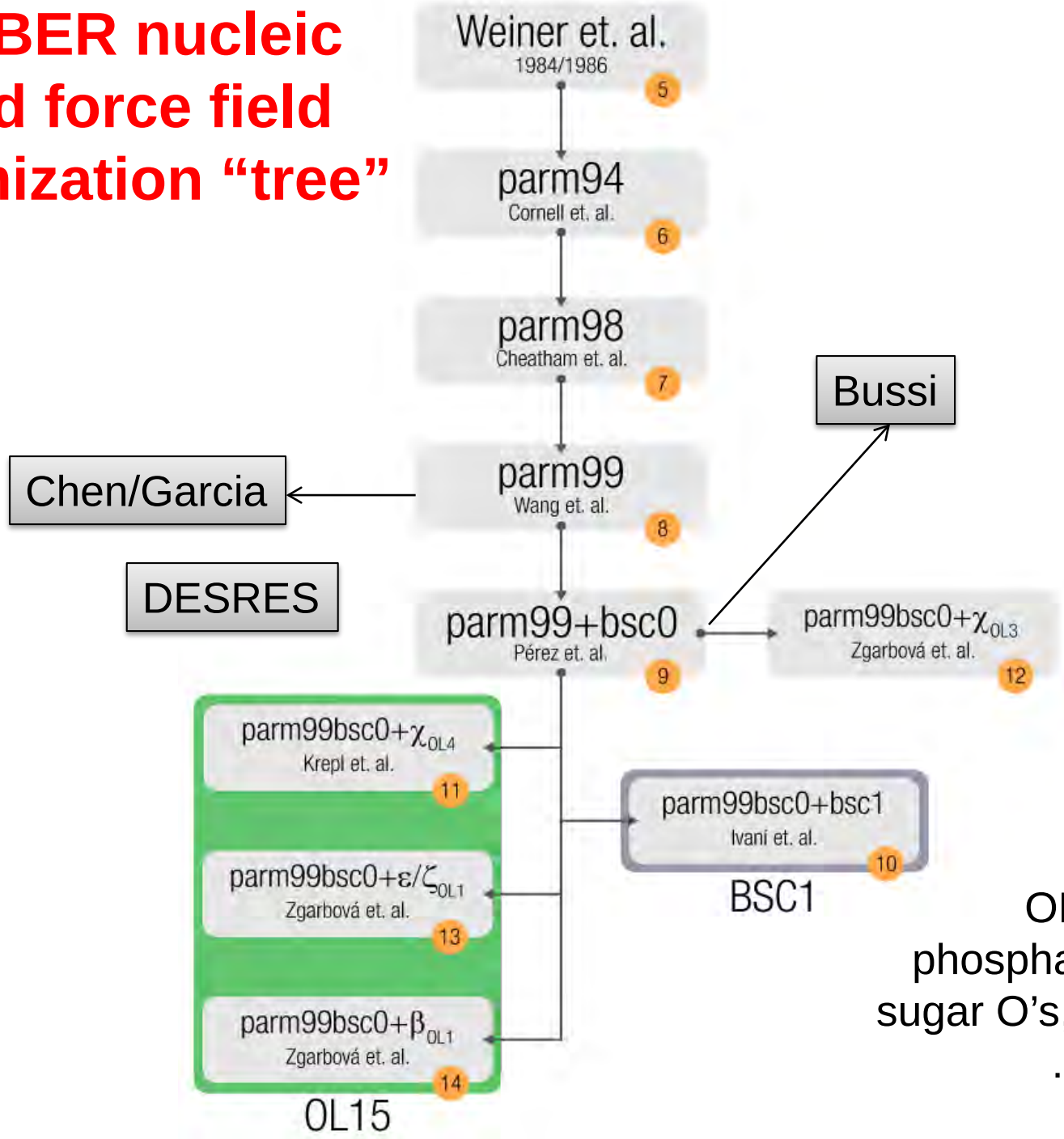
AMBER ff, MD on Anton1@PSC – data at 2 ns intervals, 10 ns running average, every 5th frame (~10 μs of MD shown).

reproducibility, convergence, agreement with experiment, new insight

Products

- 3 PRAC awards (2011-2018), 1 Ebola RAPID
- 50+ Cheatham group publications, 2013-6/2018
- GPU-accelerated Amber 14, Amber 16, Amber 18
- multi-dimensional replica exchange (M-REMD)
- 4 levels of parallelism in CPPTRAJ (molecular dynamics trajectory analyses – ensemble, file/analyses, OpenMP, CUDA) **[paper finally accepted]**
- method validation (Anton vs. AMBER vs. GROMACS vs. CHARMM)
- re-refined NMR structures, Mg-dependent structure
- hydrogen mass repartitioning
- reproducibility & convergence
- force field assessment / validation / optimization

AMBER nucleic acid force field optimization "tree"



OPC
phosphate mods
sugar O's, O2' mods
…

We can (using very long simulation or even better using M-REMD approaches) converge the conformational ensembles of various models:

- **duplexes**
- dinucleotides
- tetranucleotides
- tetraloops (UUCG, GNRA, …)
- mini-dumbells (CCTGCCTG, TTTATTTA)

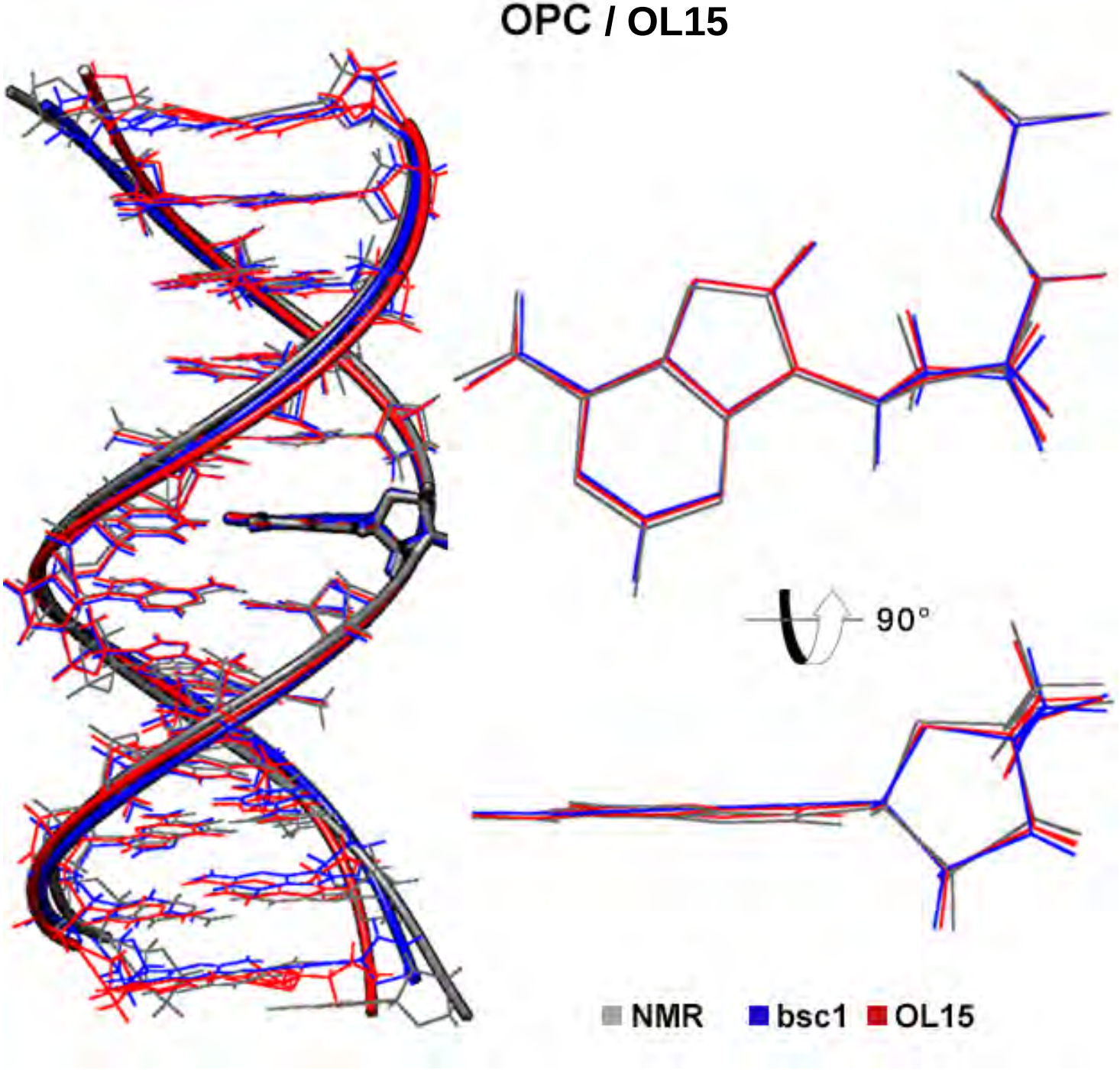
Root mean square (RMS) deviations (Å) of average structures from MD to NMR of the Dickerson dodecamer. The average structures from simulations were calculated over the full aggregated trajectories of each system (***100 independent MD trajectories, 11 µs, omit first 1 µs, aggregate – except C36 1.1 µs, omit first 200 ns***); the DDD NMR reference was an average of the models in the 1NAJ structure. RMS deviations were calculated over all heavy atoms of the internal eight base pairs.

	bsc0	bsc1	OL15	CHARMM36	CHARMM36-JC
TIP3P	1.00	0.64	0.53	1.29	1.30
OPC	0.91	0.61	0.44		

Wow!

Deviation to experiment!

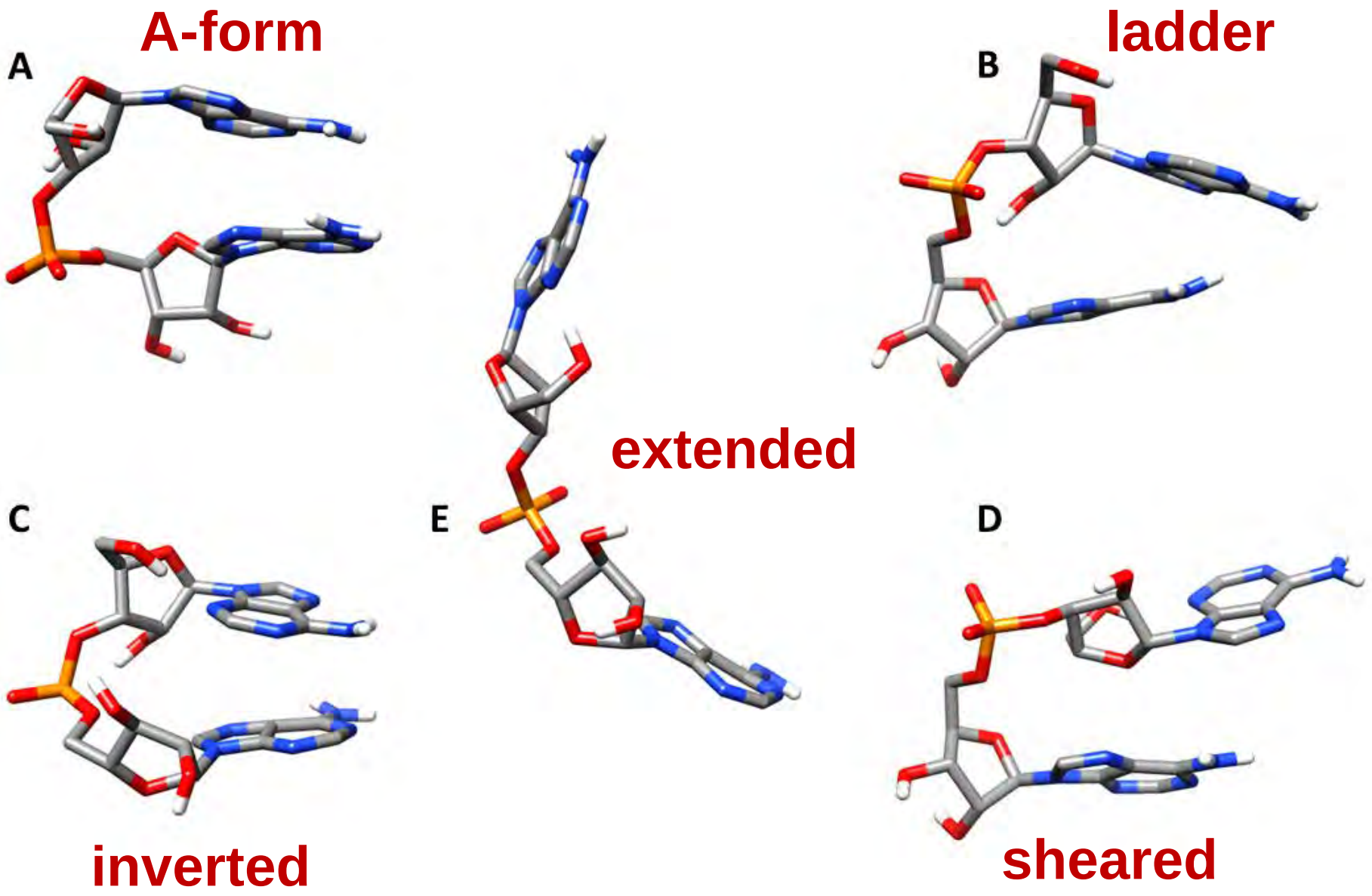
DNA duplex agreement to NMR, $d(CGCGAATTCGCG)_2$


OPC / OL15
90°
NMR
bsc1
OL15

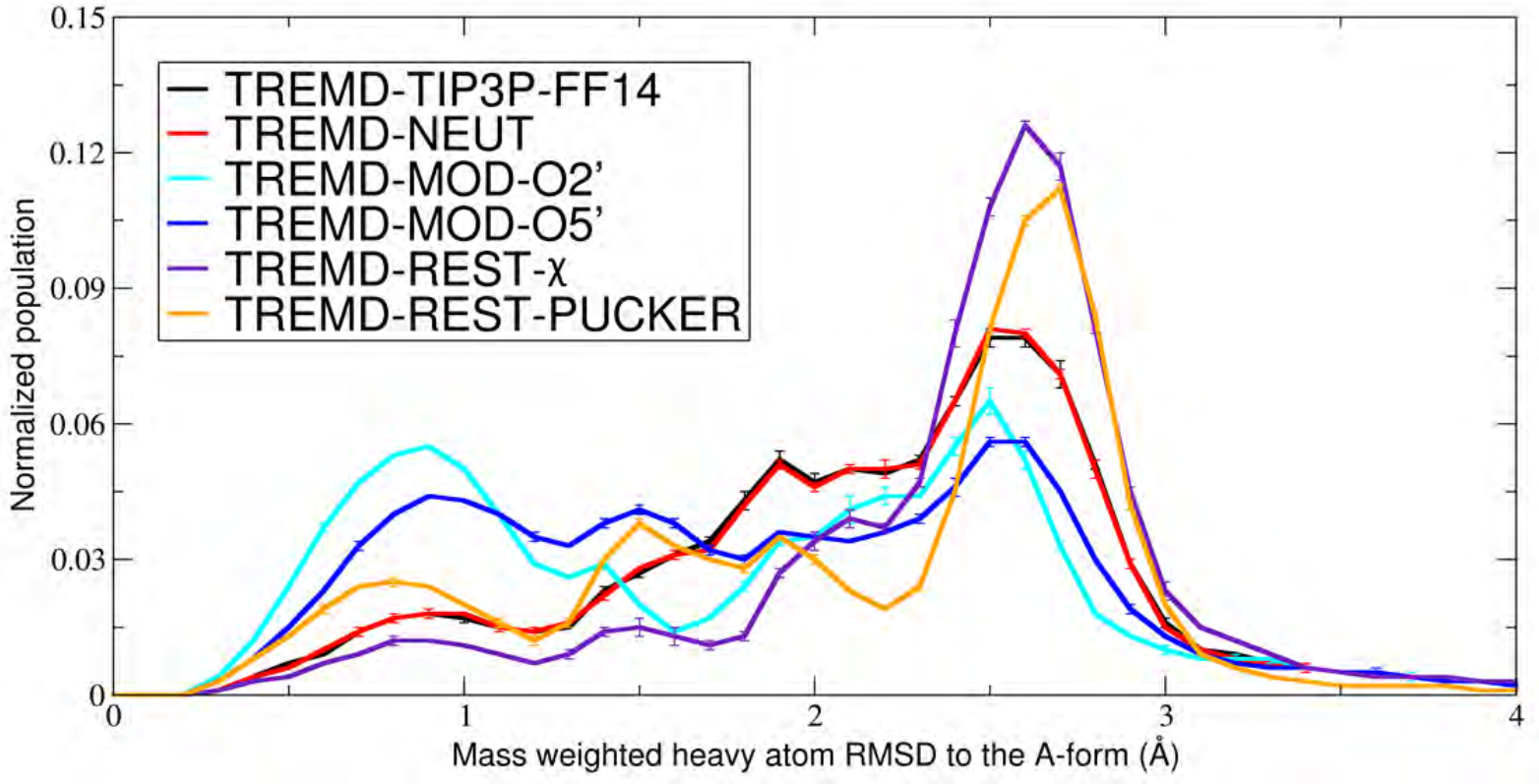
We can (using very long simulation or even better using M-REMD approaches) converge the conformational ensembles of various models:

- duplexes
- dinucleotides
- tetranucleotides
- tetraloops (UUCG, GNRA, …)
- mini-dumbells (CCTGCCTG, TTTATTTA)

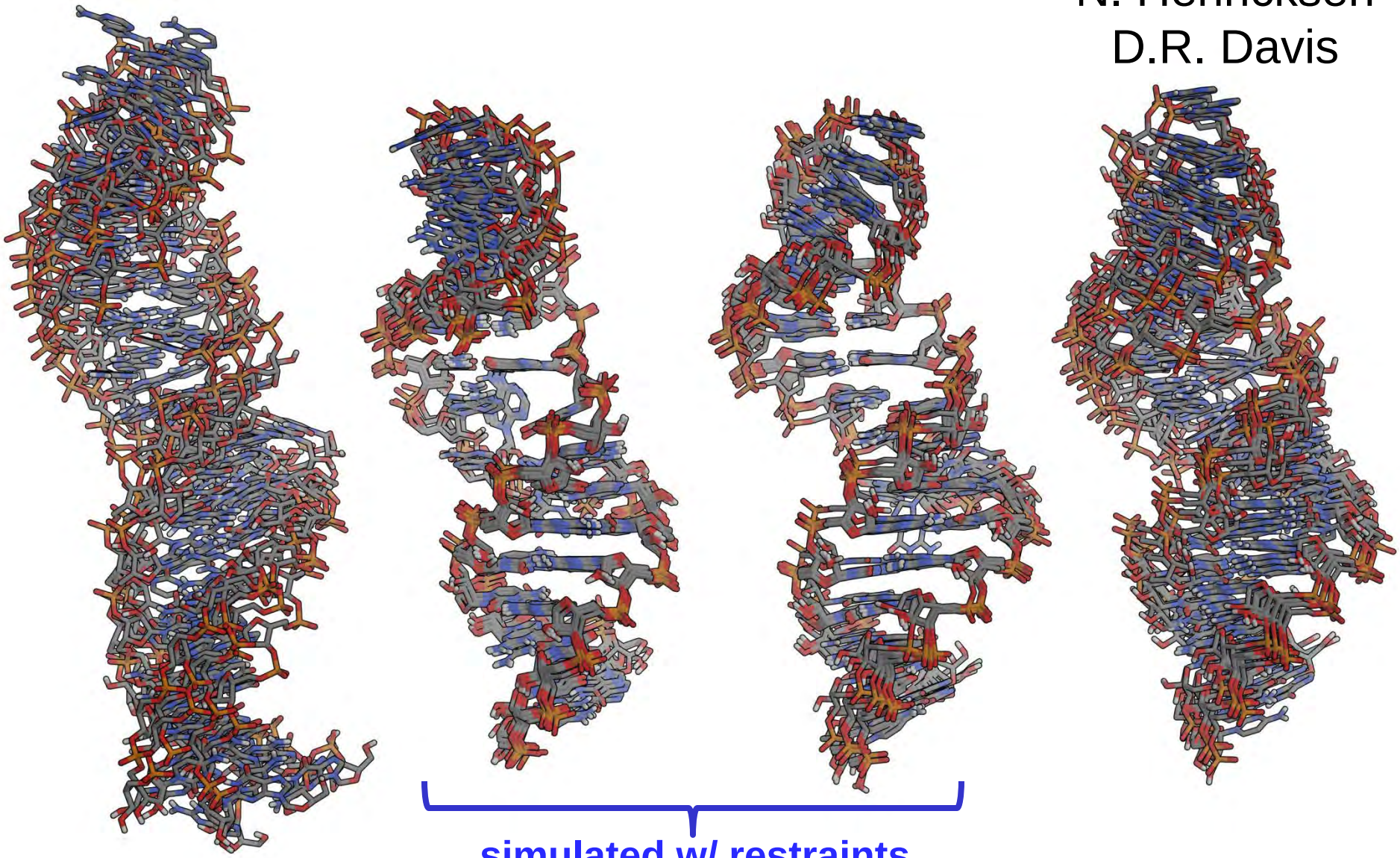
We can assess various force fields, re-weight to experimental observables, and parameter scan various changes to the underlying potentials to see influence on ensemble…


A-form
A
ladder
B
extended
E
C
D
inverted
sheared

Conformational cluster	Average RMSD (Å)	Average suite outliers (%)
A-form	1.2 ± 0.2	12.7 ± 2.6
Ladder	1.6 ± 0.3	15.4 ± 4.8
Sheared	2.4 ± 0.2	43.3 ± 5.7
Inverted	2.8 ± 0.2	39.2 ± 8.1
Extended	3.6 ± 0.5	43.0 ± 8.8



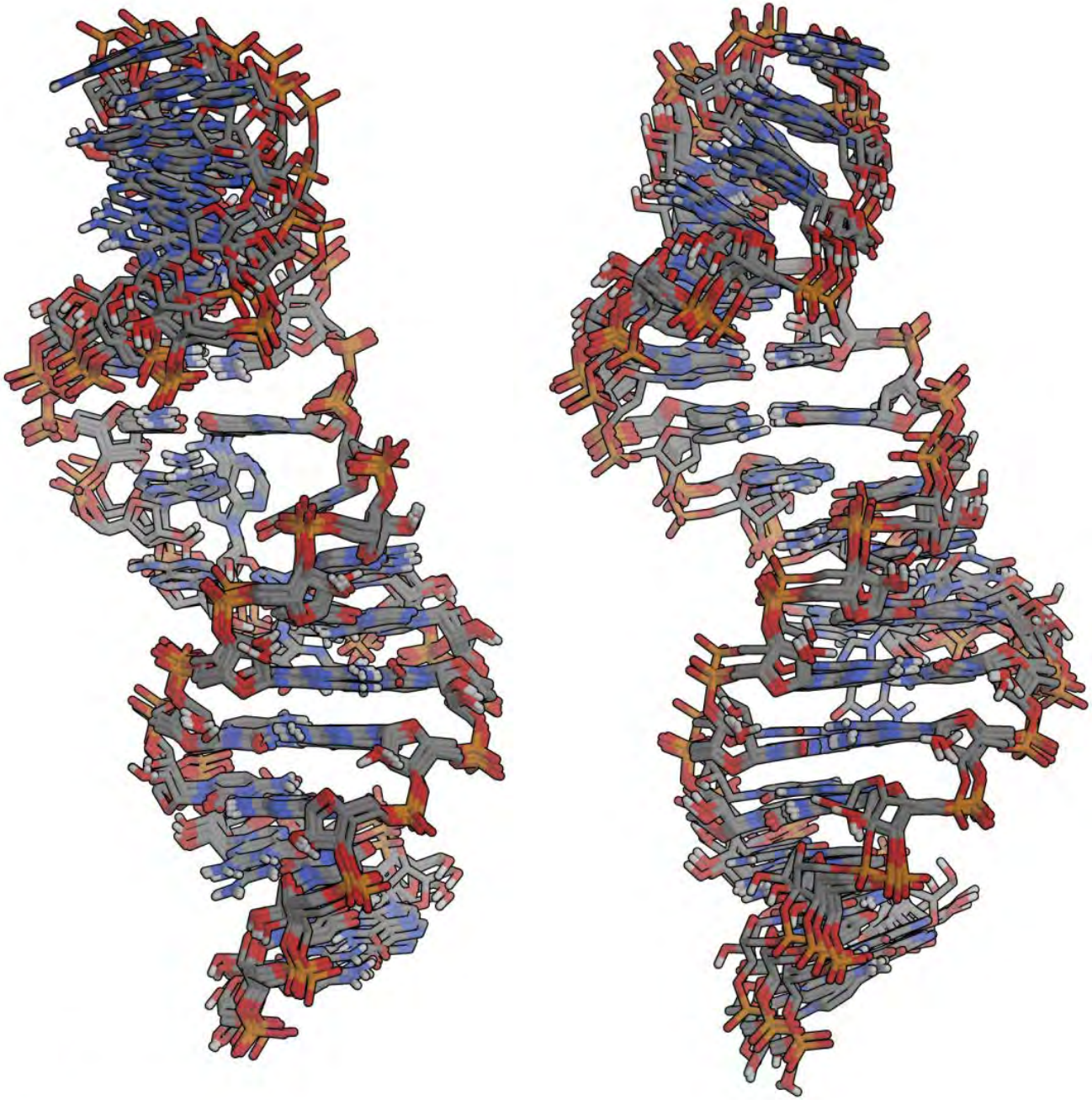
N. Henricksen
D.R. Davis



NMR: 1R2P

simulated w/ restraints,
modern force field,
explicit solvent

NMR:2F88



N. Henricksen
D.R. Davis

Key issues:

- Need long MD to expose problems with sampling, restraints, …

- Beware of bad NOEs

- RDCs are good to include if available

- automatic refinement is still a ways off

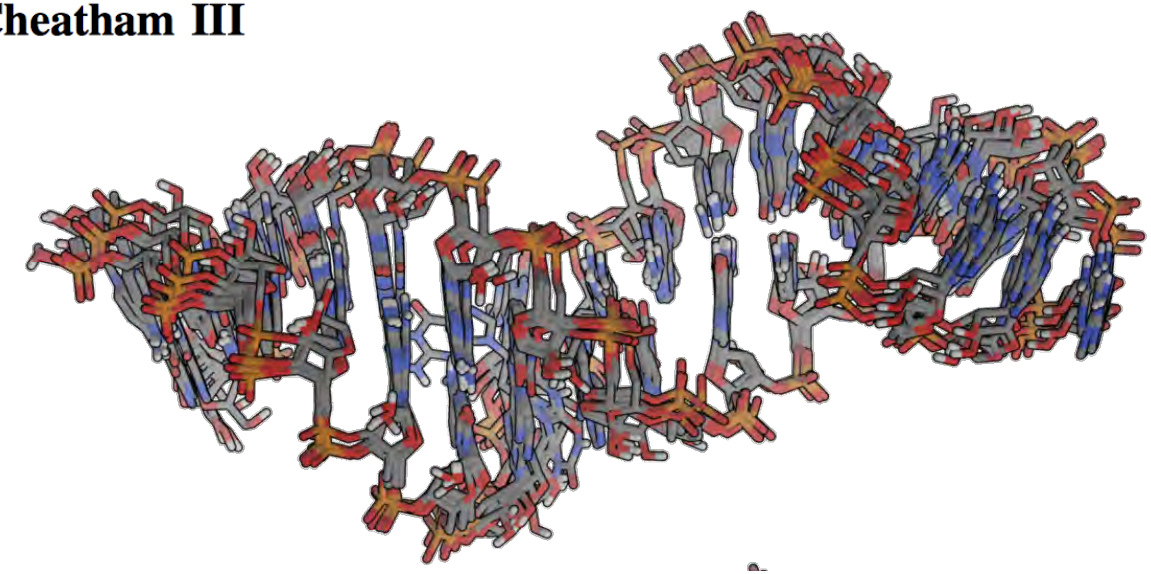
simulated w/ restraints, modern force field, explicit solvent

J Biomol NMR (2012) 53:321–339
DOI 10.1007/s10858-012-9642-5

ARTICLE

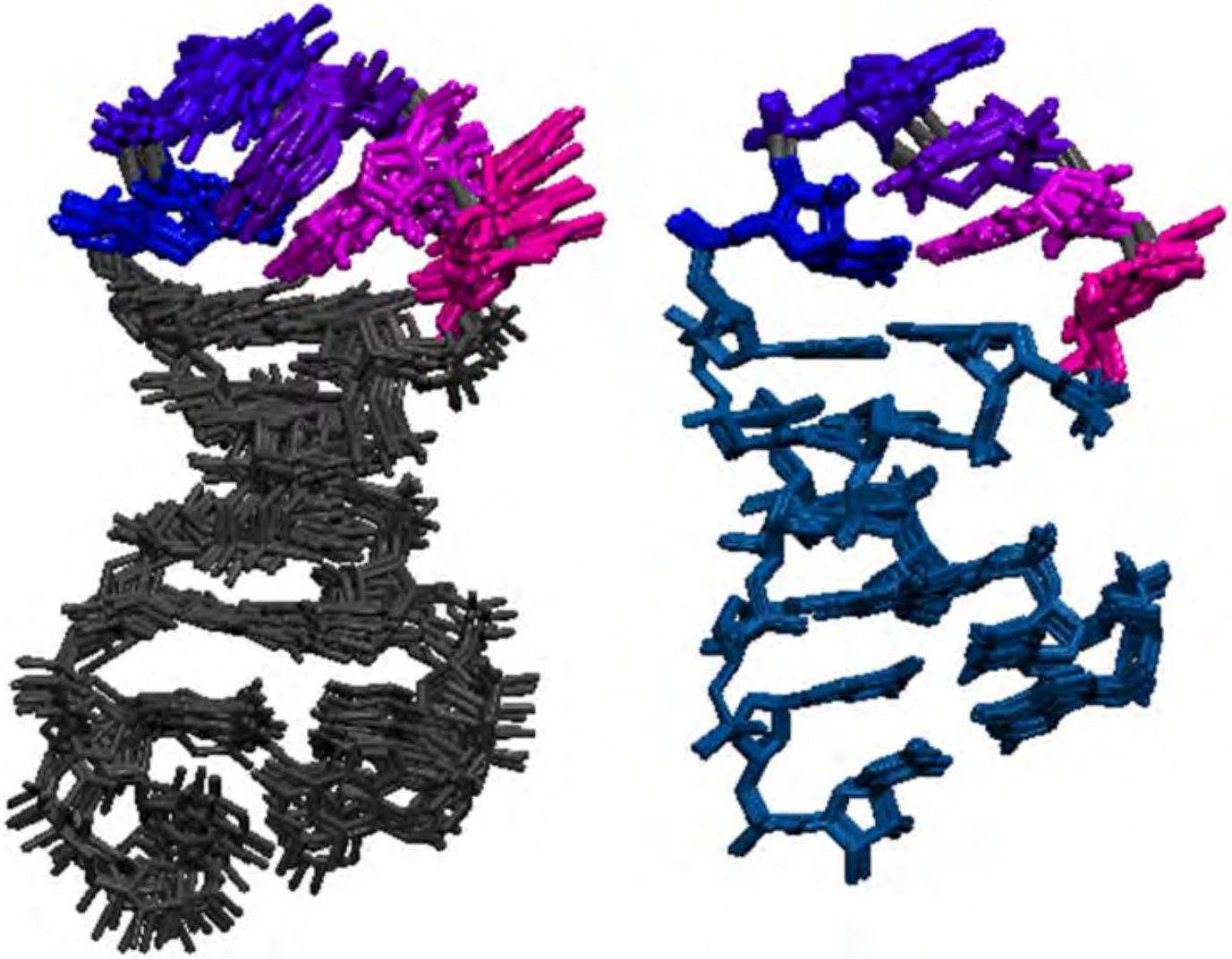
Molecular dynamics re-refinement of two different small RNA loop structures using the original NMR data suggest a common structure

Niel M. Henriksen · Darrell R. Davis · Thomas E. Cheatham III



decoy: 1TBK ↔ 1YN2 ± Mg^{2+}

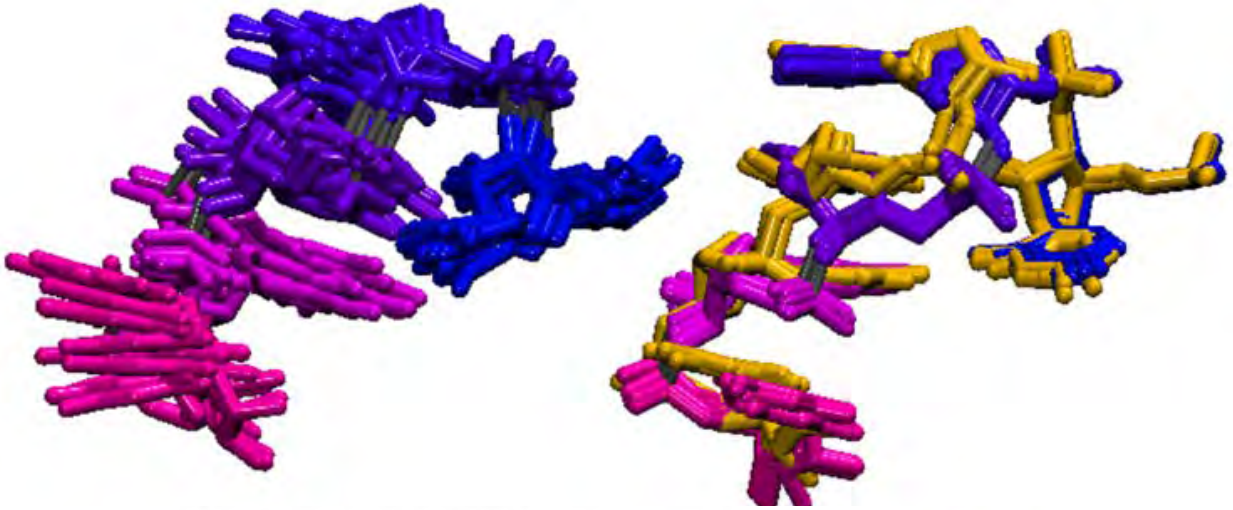
-Mg^{2+} deviates from NMR structure: re-refine…

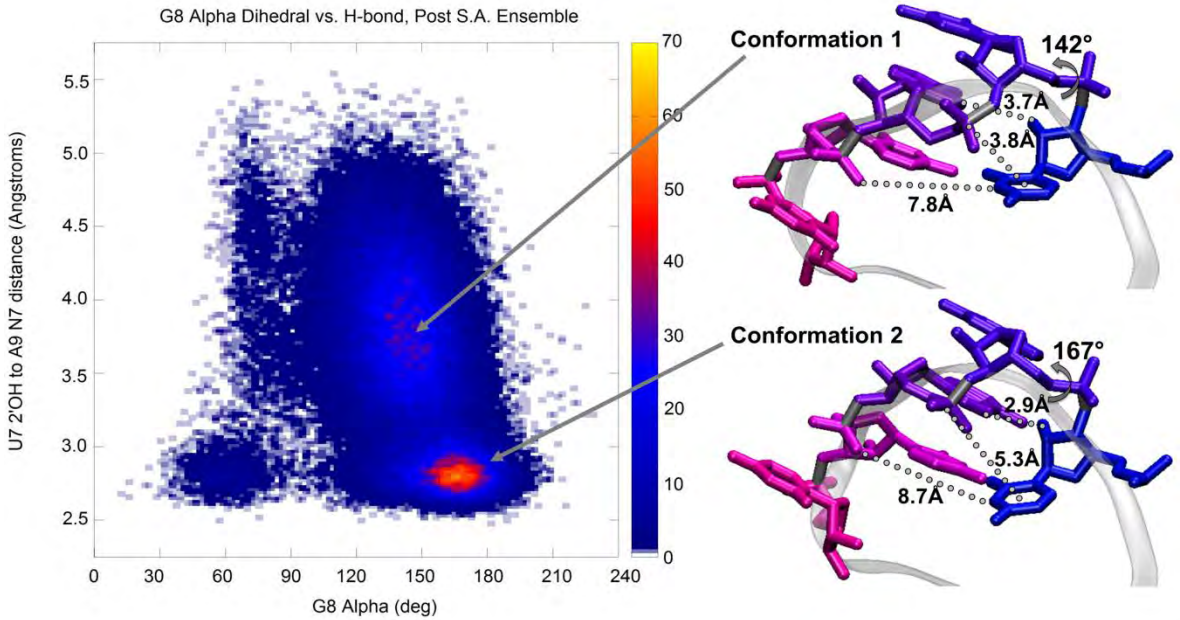


original NMR **re-refined NMR**

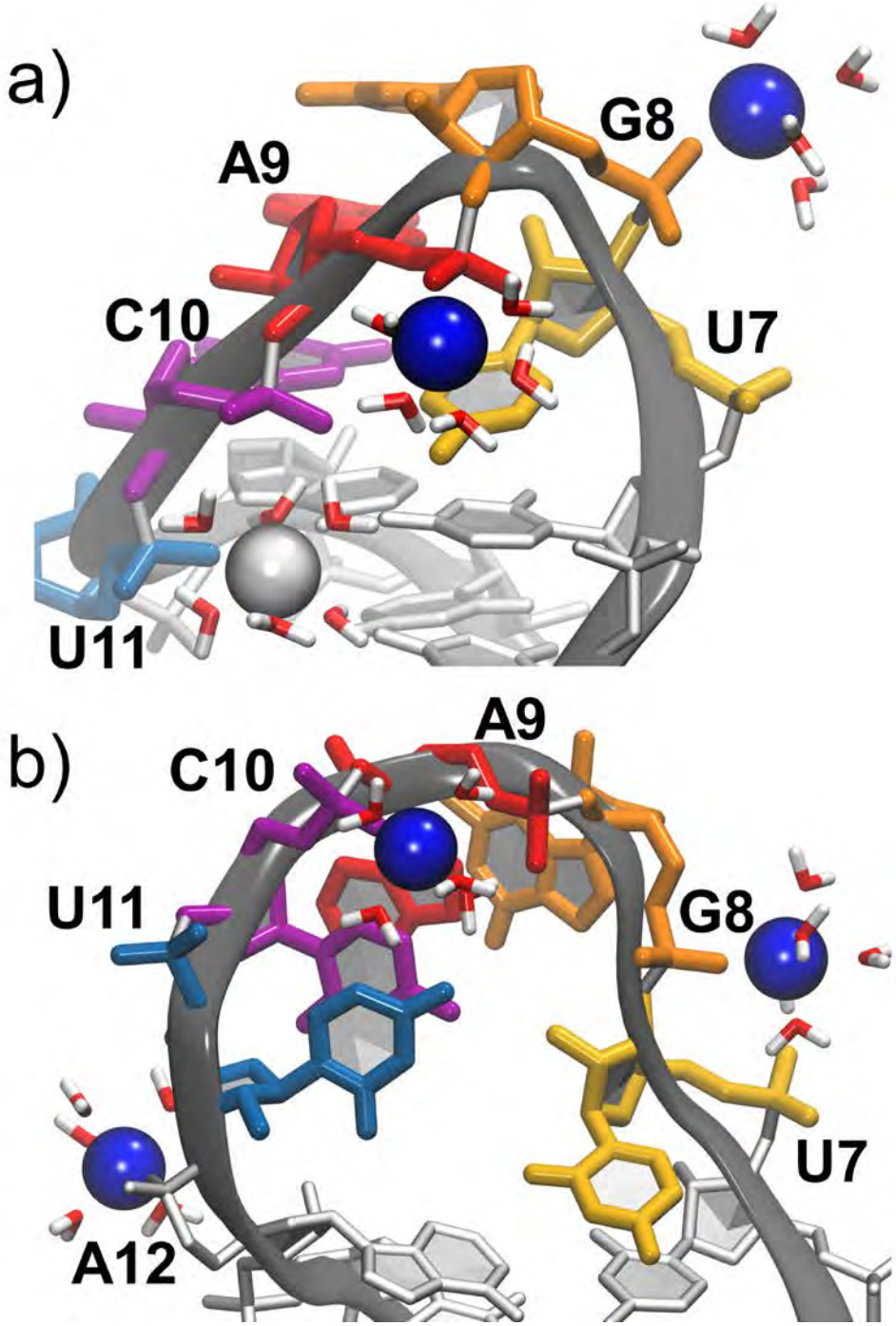
decoy: 1TBK ↔ 1YN2 ± Mg^{2+}

-Mg^{2+} deviates from NMR structure: re-refine…





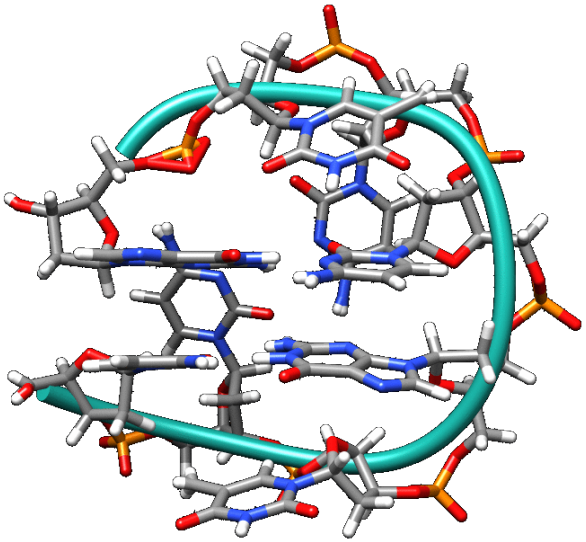
OK ☺

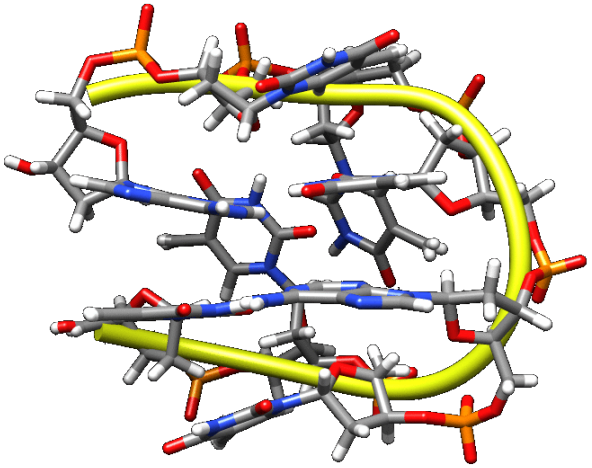


trapped for ms ☹

12-6-4 chelated ion affinity is 12-13.5 kcal/mol!

should the force field target the correct Mg^{2+} - water affinity?





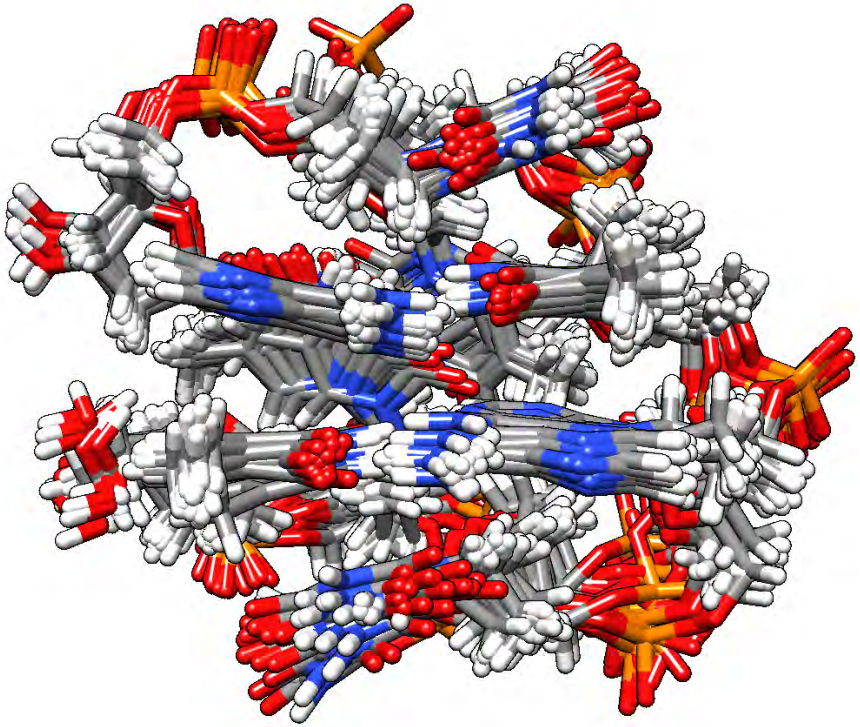
CCTGCCTG

TTTATTTA

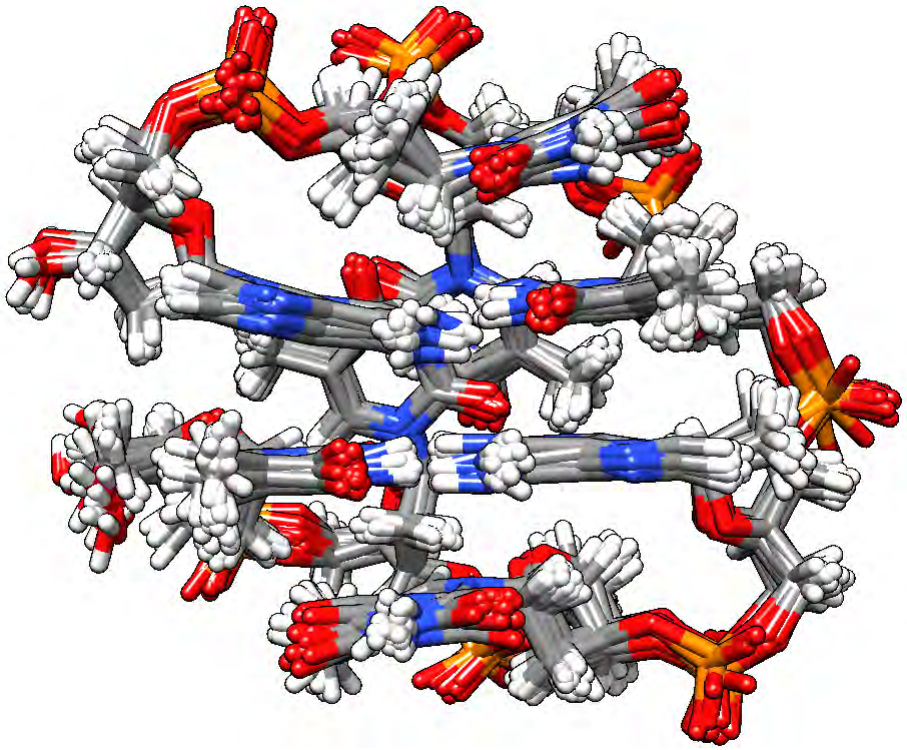
Pei Guo and Sik Lok Lam, JACS (2016)

NMR re-refinement

- Starting from each of the 20 conformations → re-refine with bsc1/OL15 and opc/opc3 – with original restraint file (264 bond and angle restraints)
- Run form 100 ns, extract representative conformation from most populated cluster.

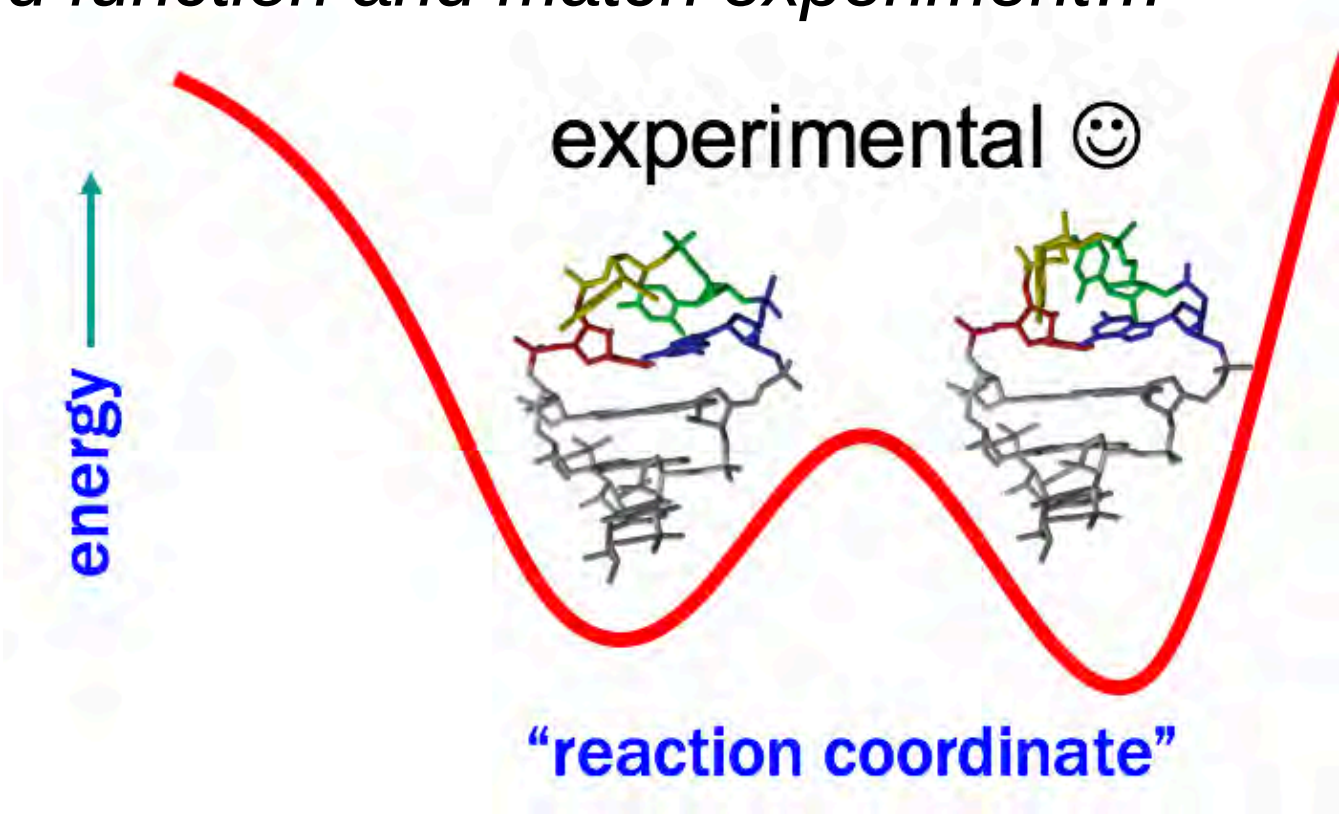


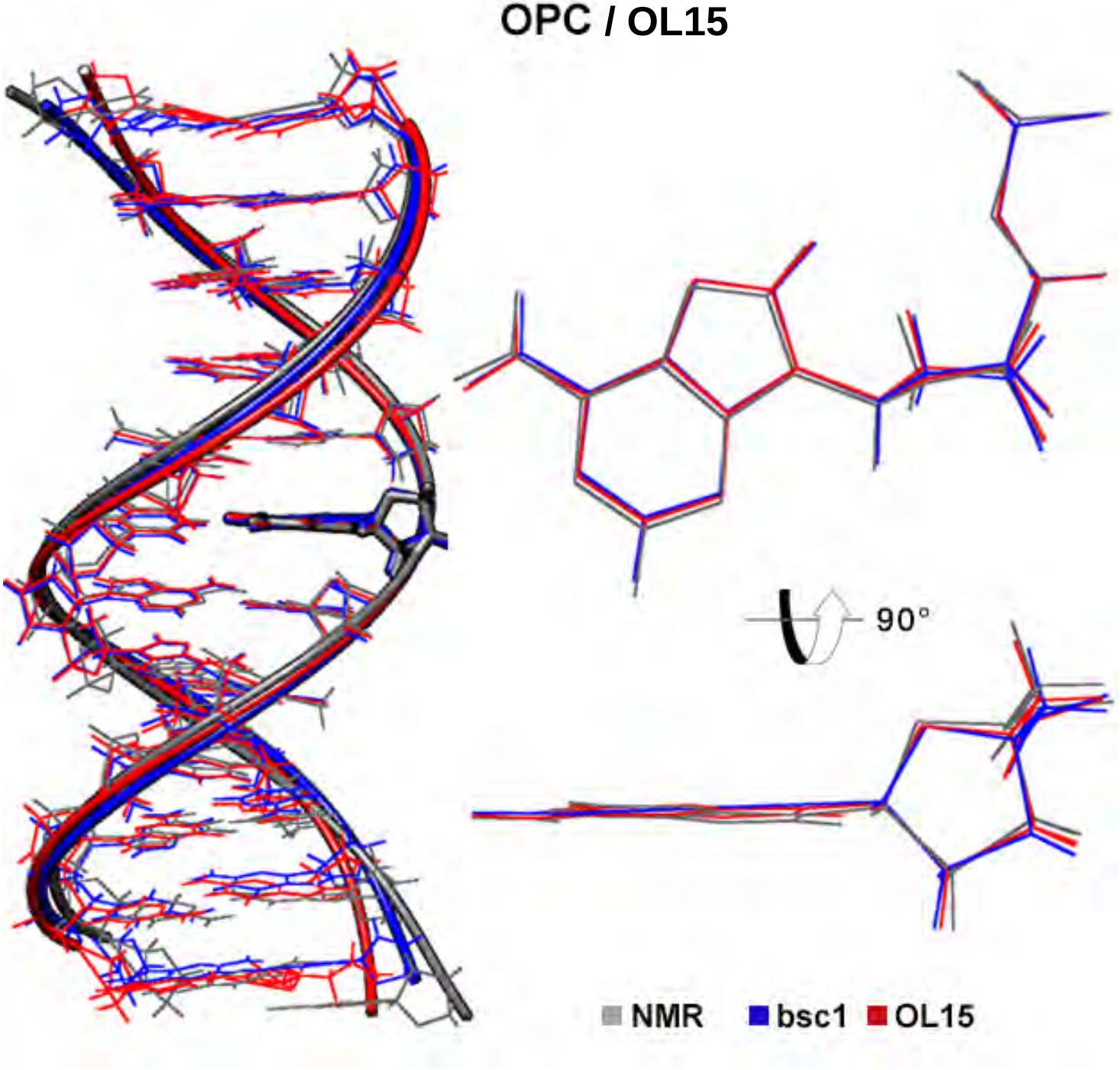
NMR original



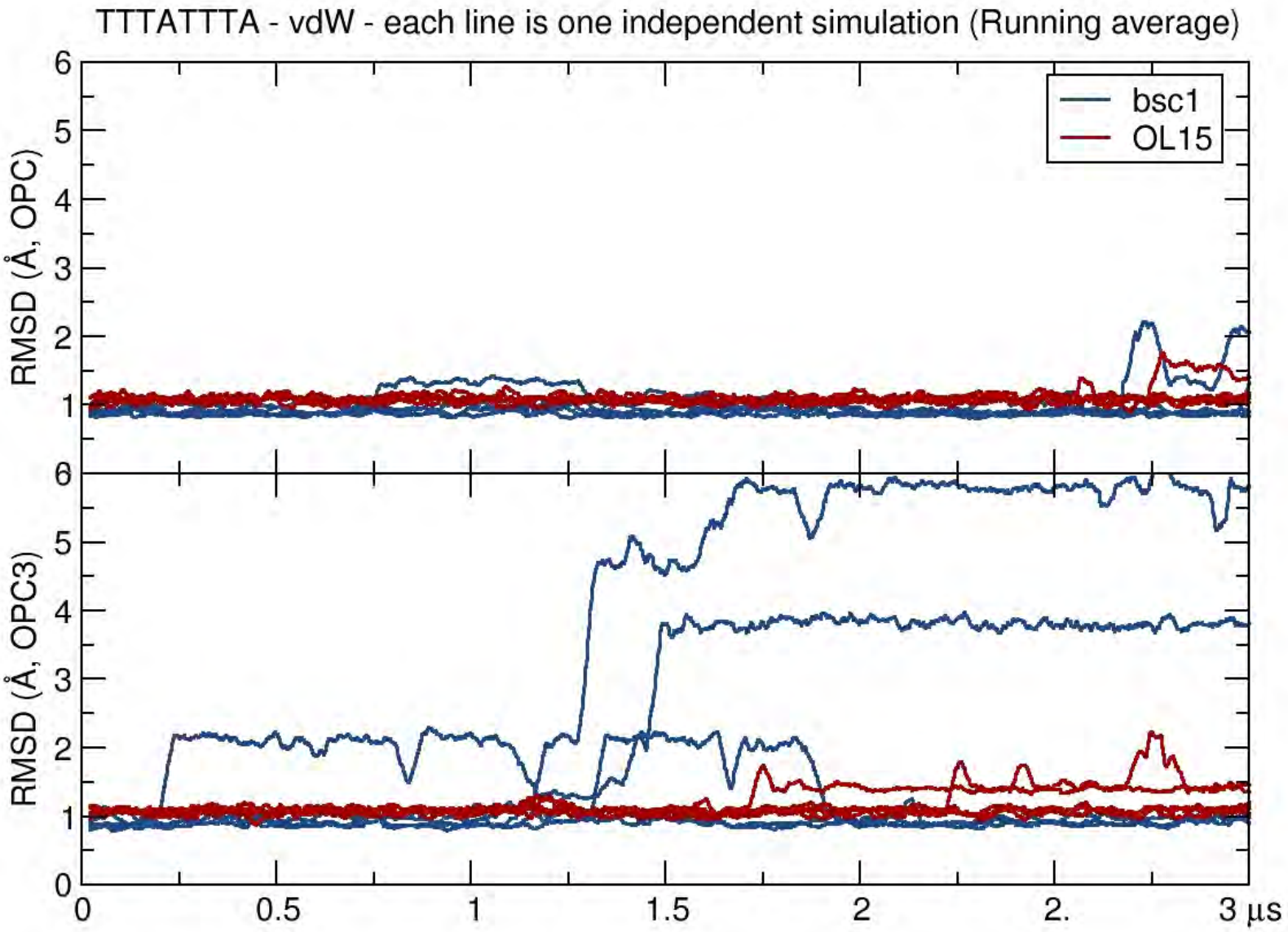
are the force fields reliable?
(free energetics, sampling, dynamics)

Short simulations stay near experimental structure; analyses can provide insight in structure, dynamics and function and match experiment…

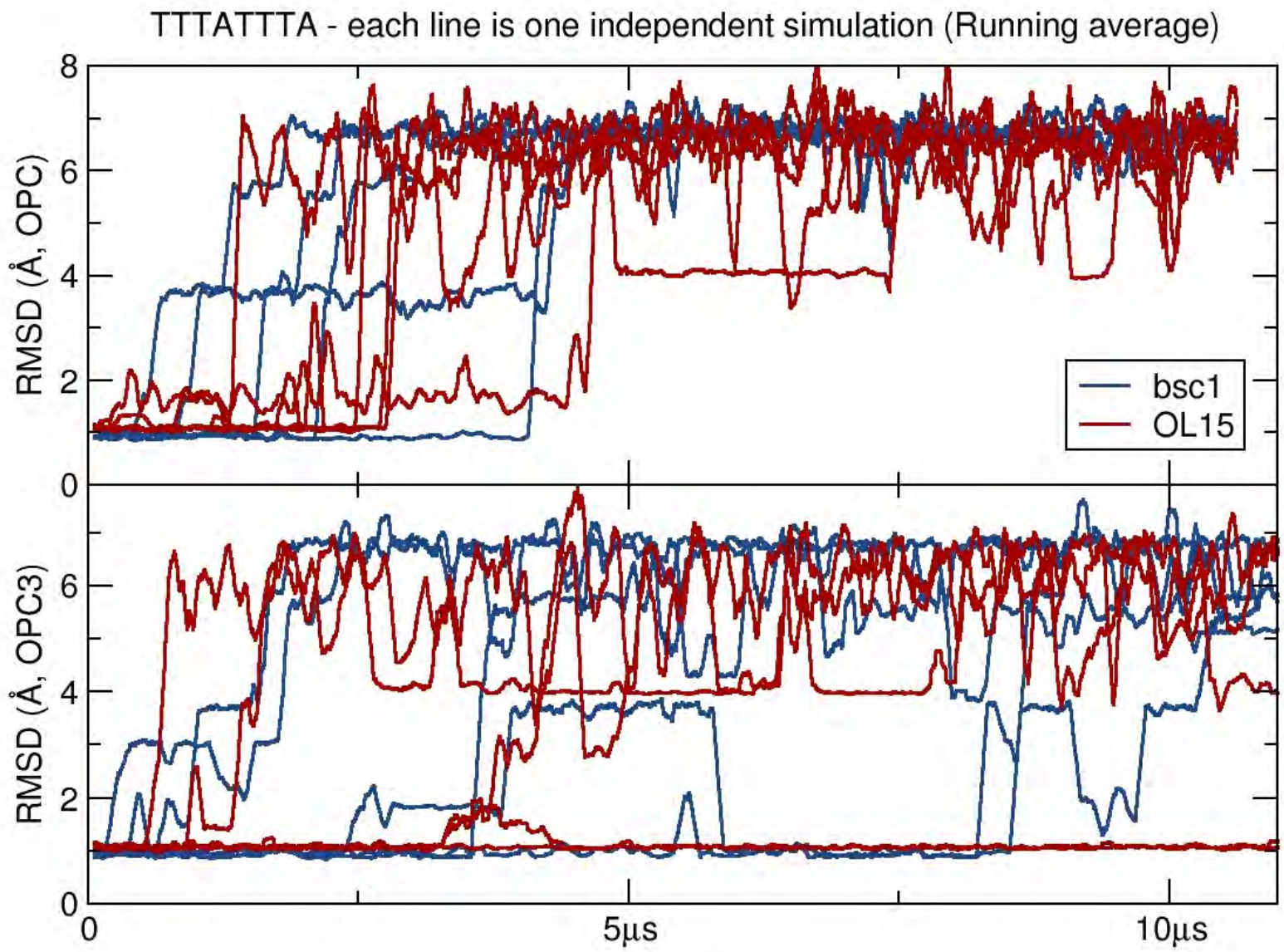



OPC / OL15
90°
NMR
bsc1
OL15

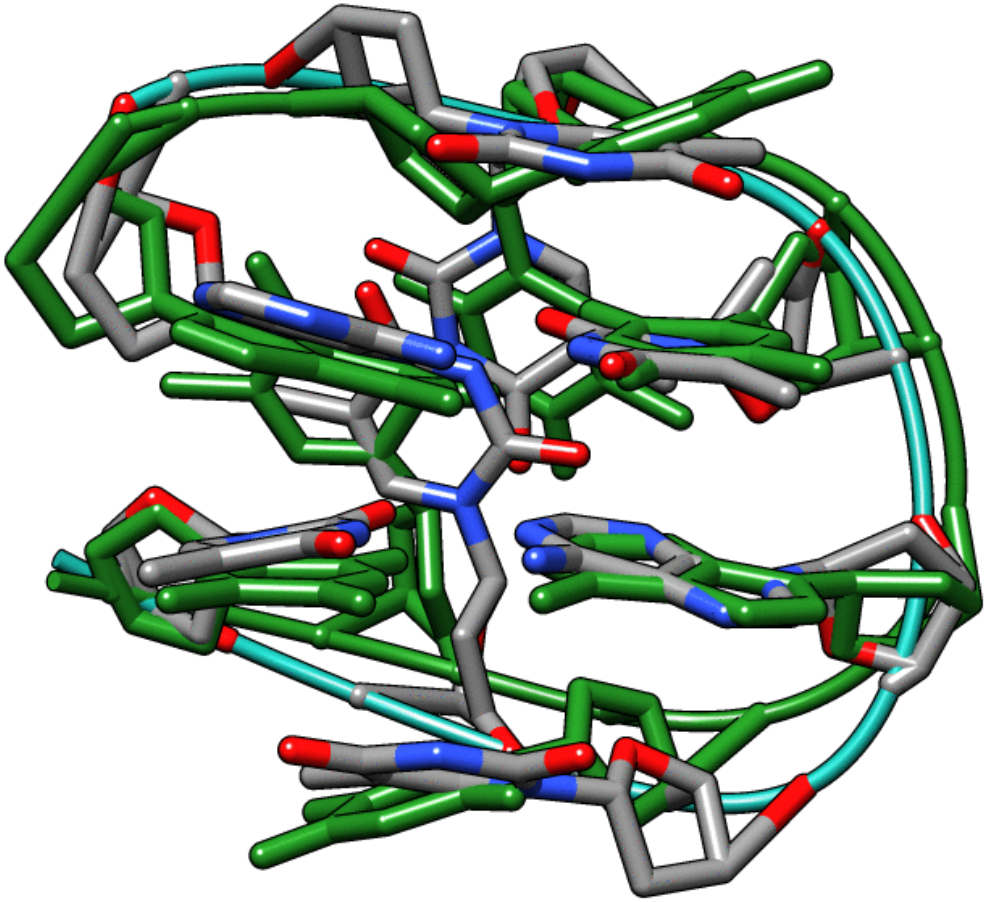
TTTATTTA


TTTATTTA - vdW - each line is one independent simulation (Running average)
RMSD (Å, OPC)
RMSD (Å, OPC3)
bsc1
OL15
0
1
2
3
4
5
6
0
0.5
1
1.5
2
2.
3 µs

TTTATTTA


TTTATTTA - each line is one independent simulation (Running average)
RMSD (Å, OPC)
RMSD (Å, OPC3)
0
2
4
6
8
0
5µs
10µs
bsc1
OL15

Folded d(TTTA)$_2$ from a 10 μs TREMD simulation using the OL15 + CG mods. Green is the NMR avg structure (RMSD difference ~1.7 Å using C1' of paired bases). …but, only 12% of population…

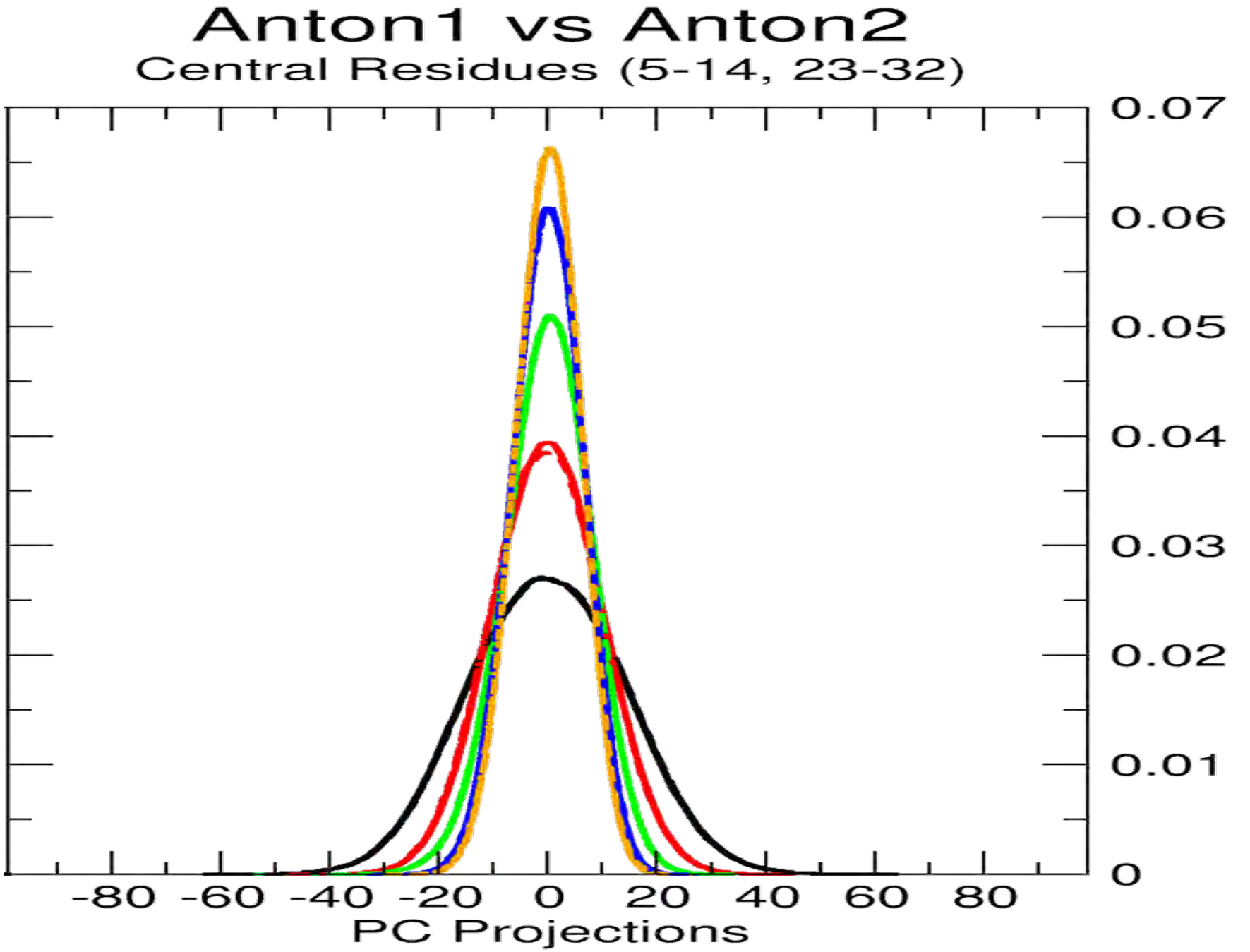


What did we learn?

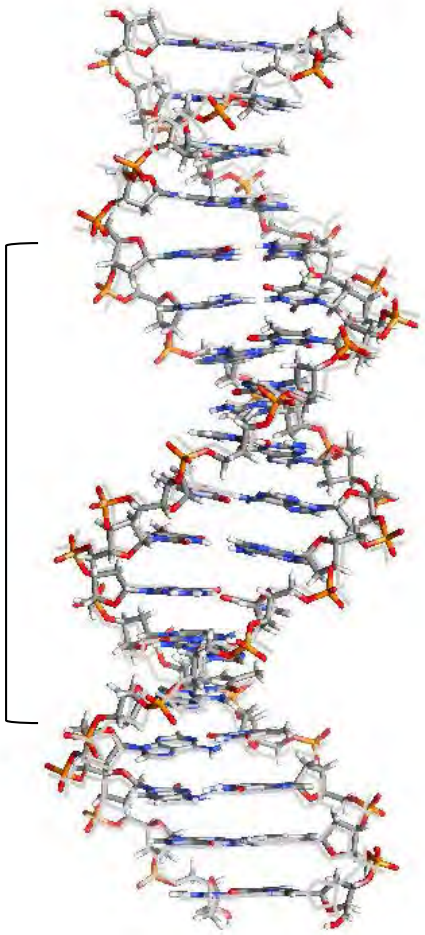
- Ensembles of independent simulations show similar **convergence** properties with respect to the structure and dynamics of the internal part of a DNA helix

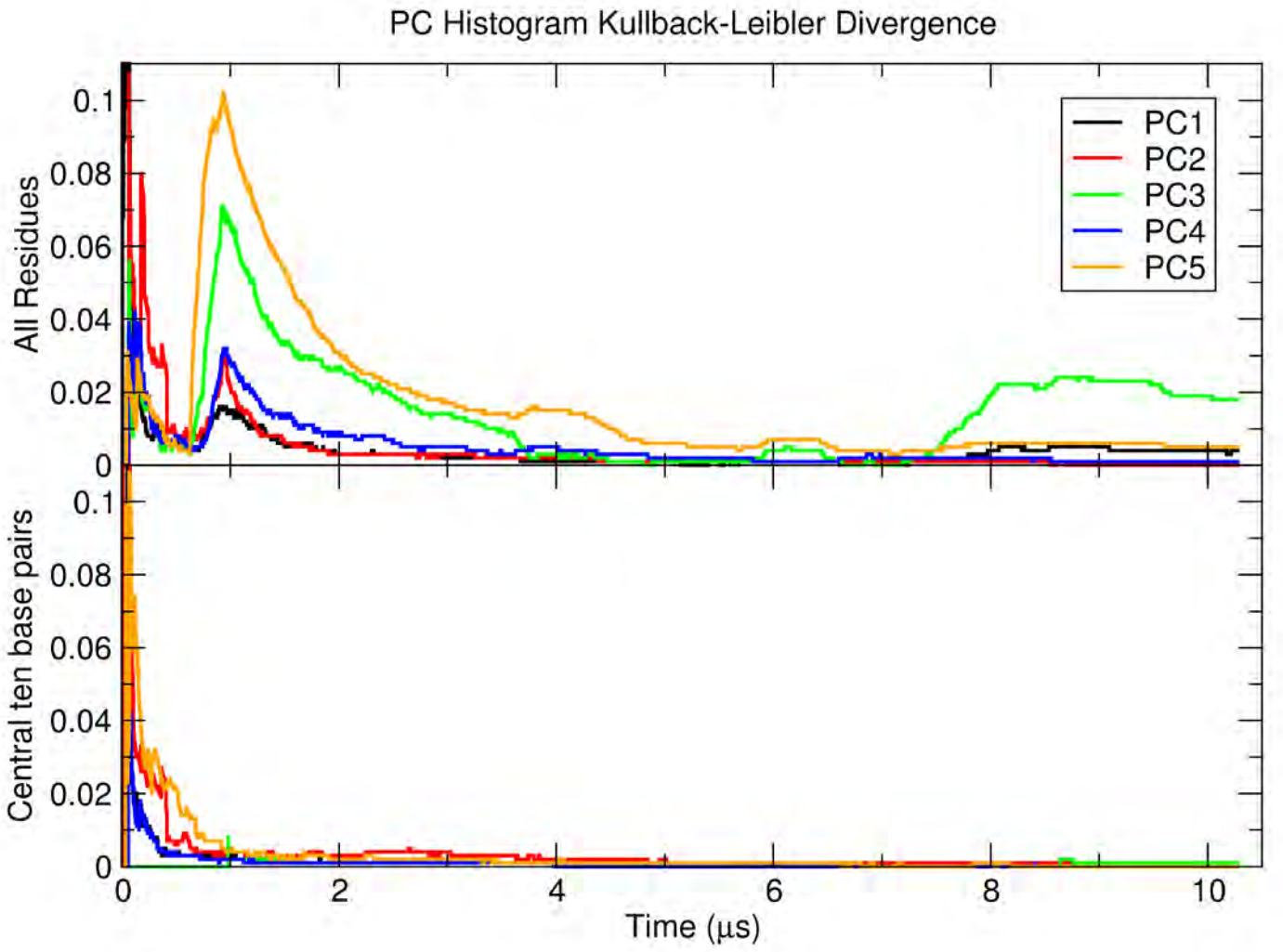
- Independent simulations (on special purpose hardware or GPUs or CPUs) give **reproducible** results

- We can converge conformational ensembles with M-REMD, however most over-populate anomalous conformations.

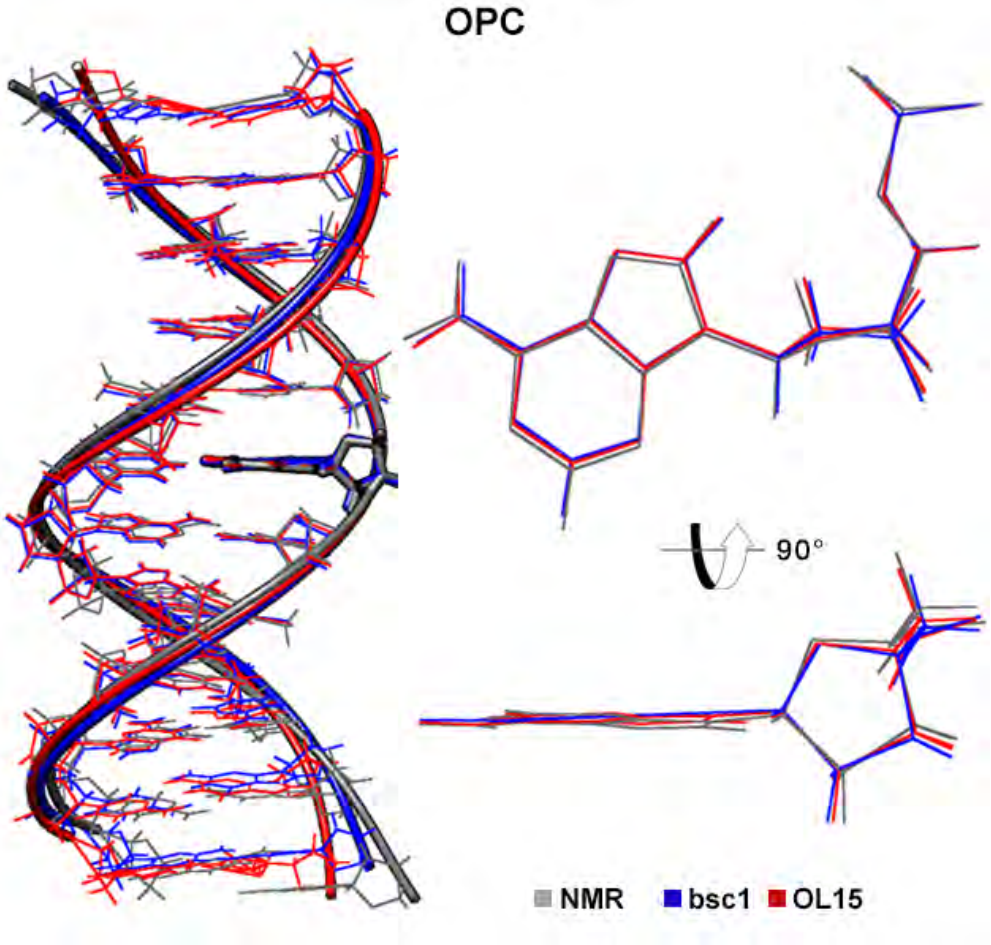

Anton1 vs Anton2
Central Residues (5-14, 23-32)
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-80 -60 -40 -20 0 20 40 60 80
PC Projections

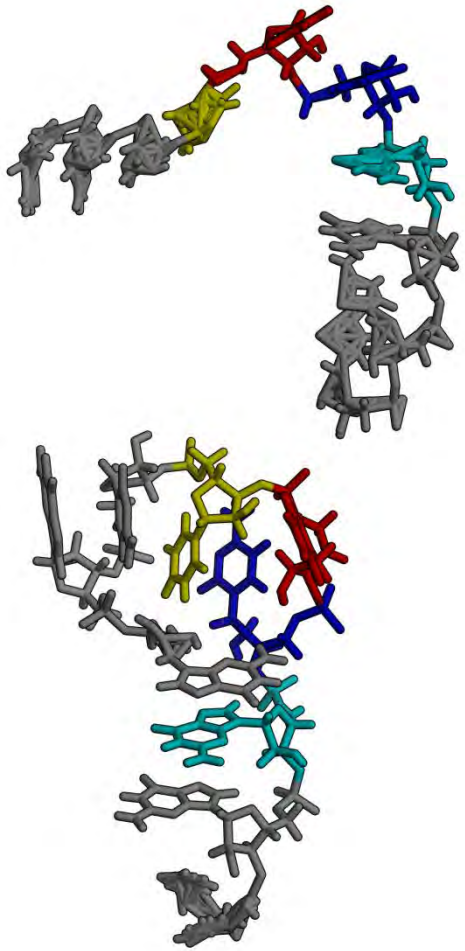
Test for convergence within and between simulations…



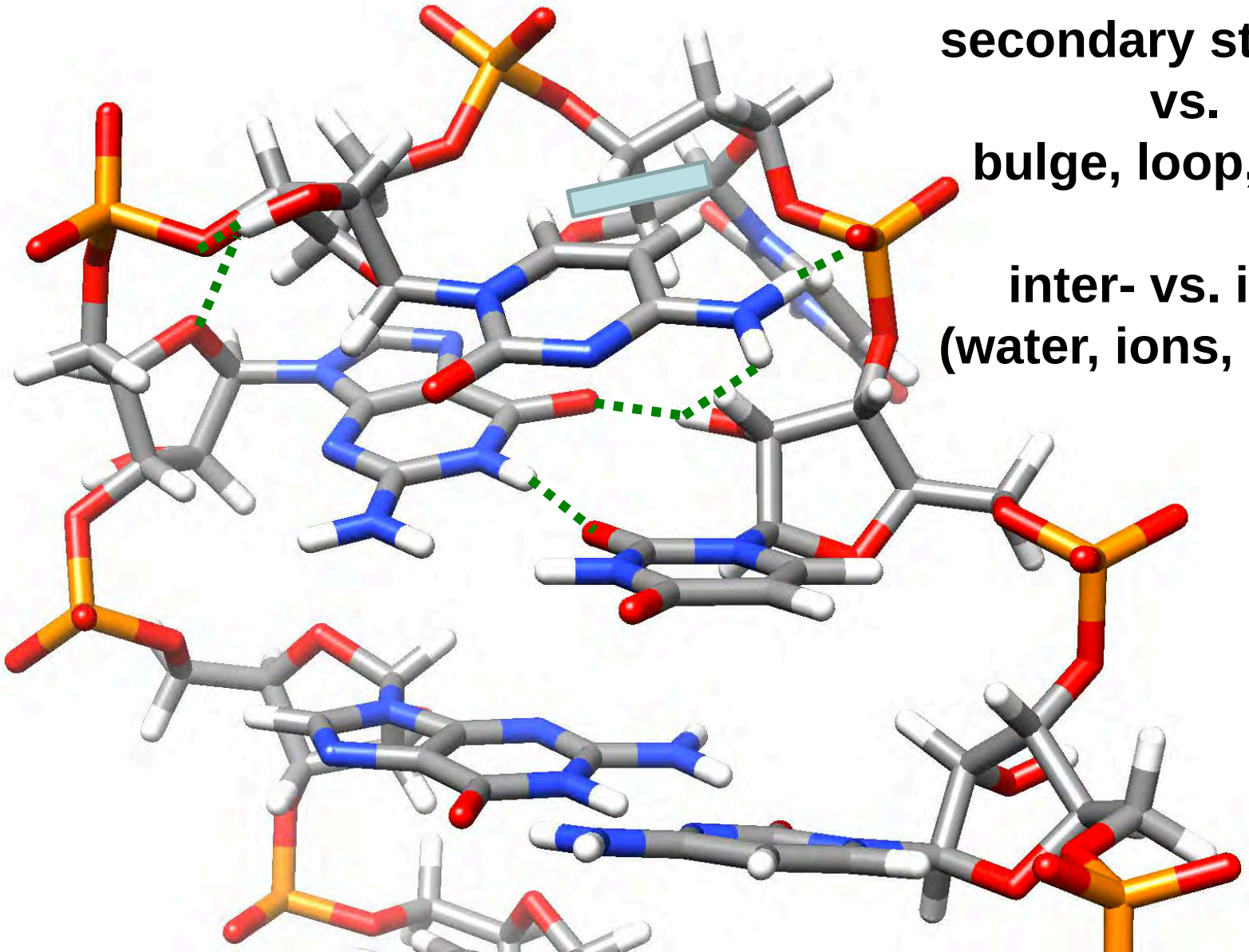


Why the drastic difference? (between DNA helix and RNA tetraloop)



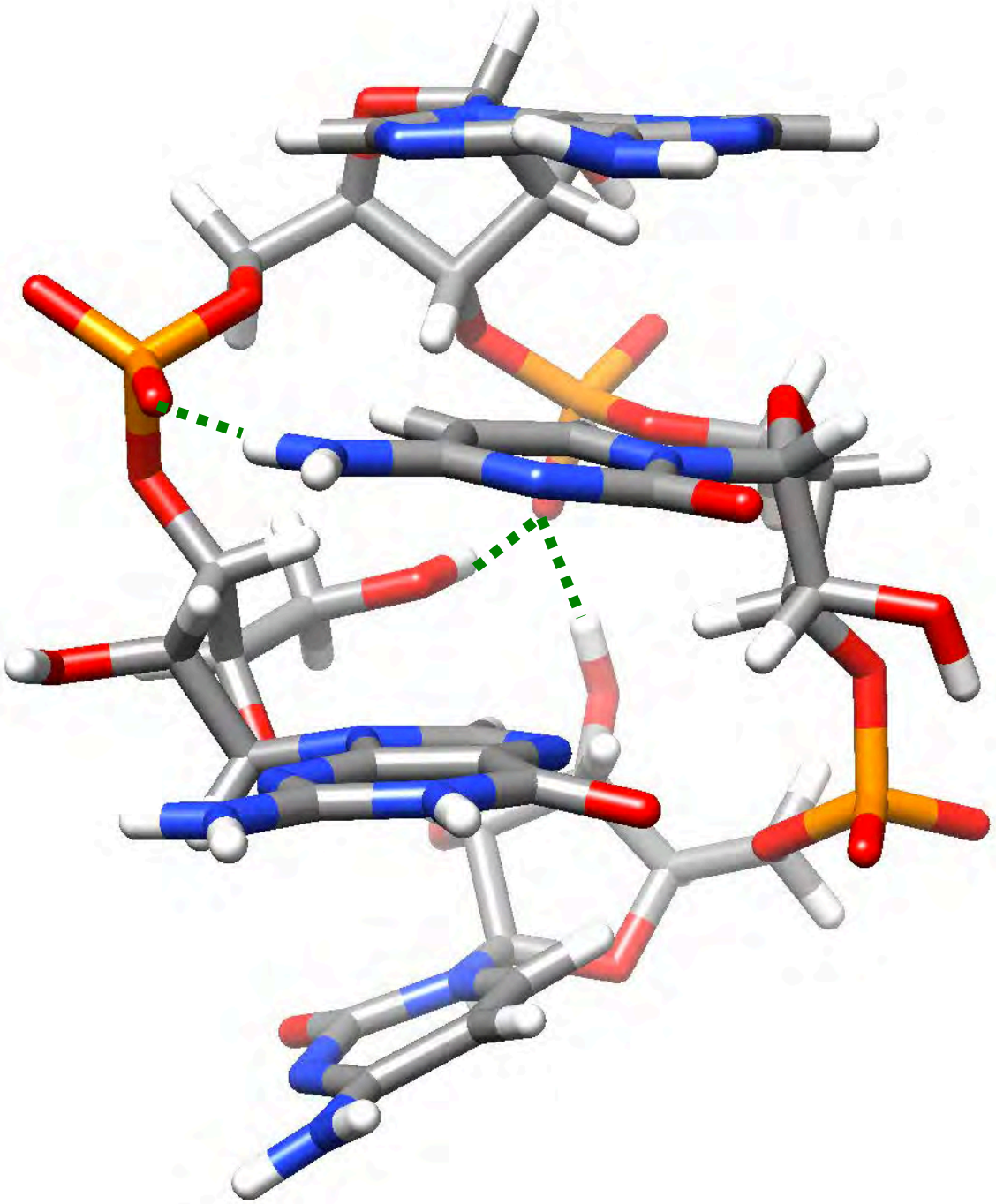


Why the drastic difference? "Balance"

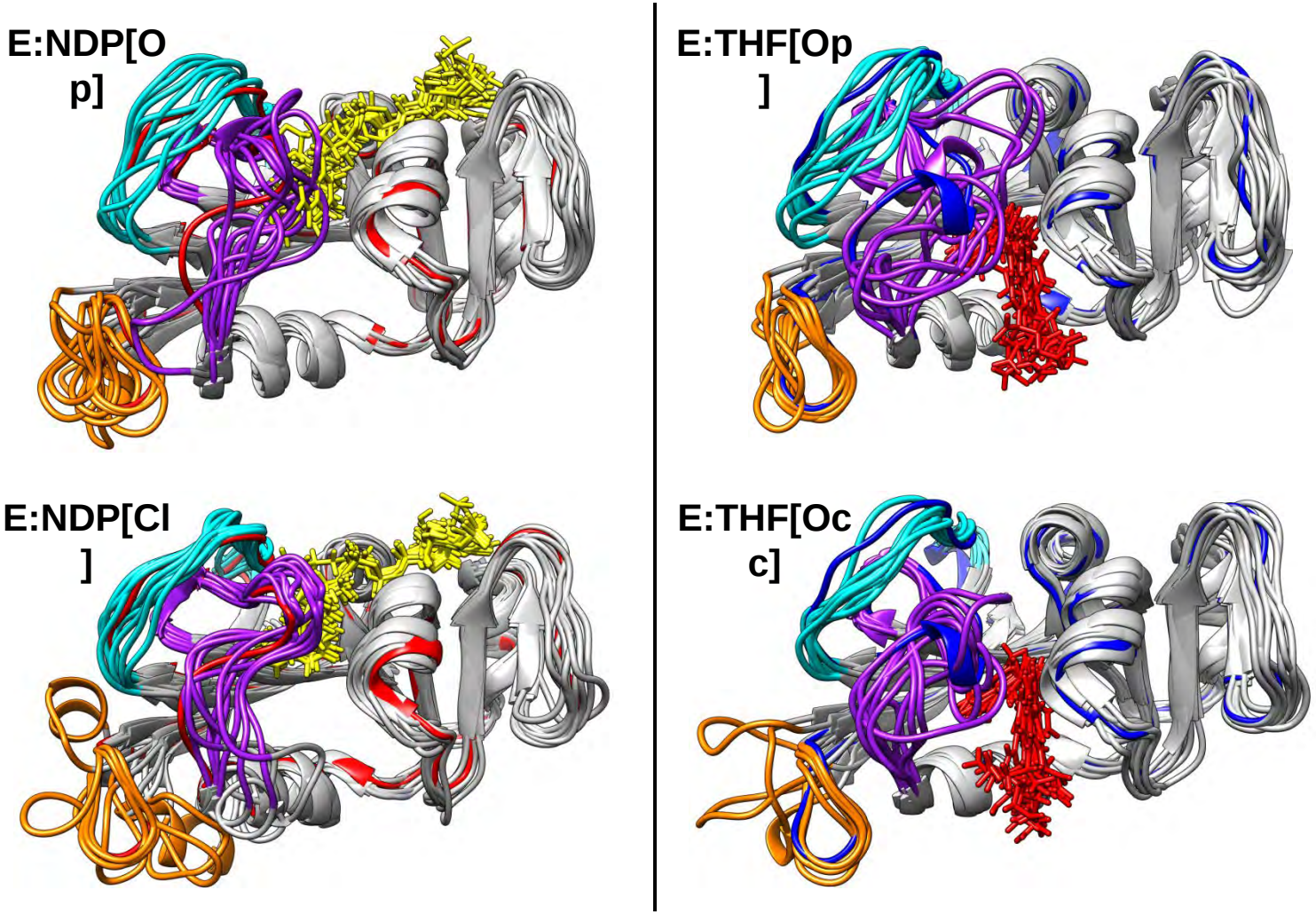


secondary structure
vs.
bulge, loop, 3^0, …

inter- vs. intra-
(water, ions, biomol)



Most-Sampled Enzyme Conformations

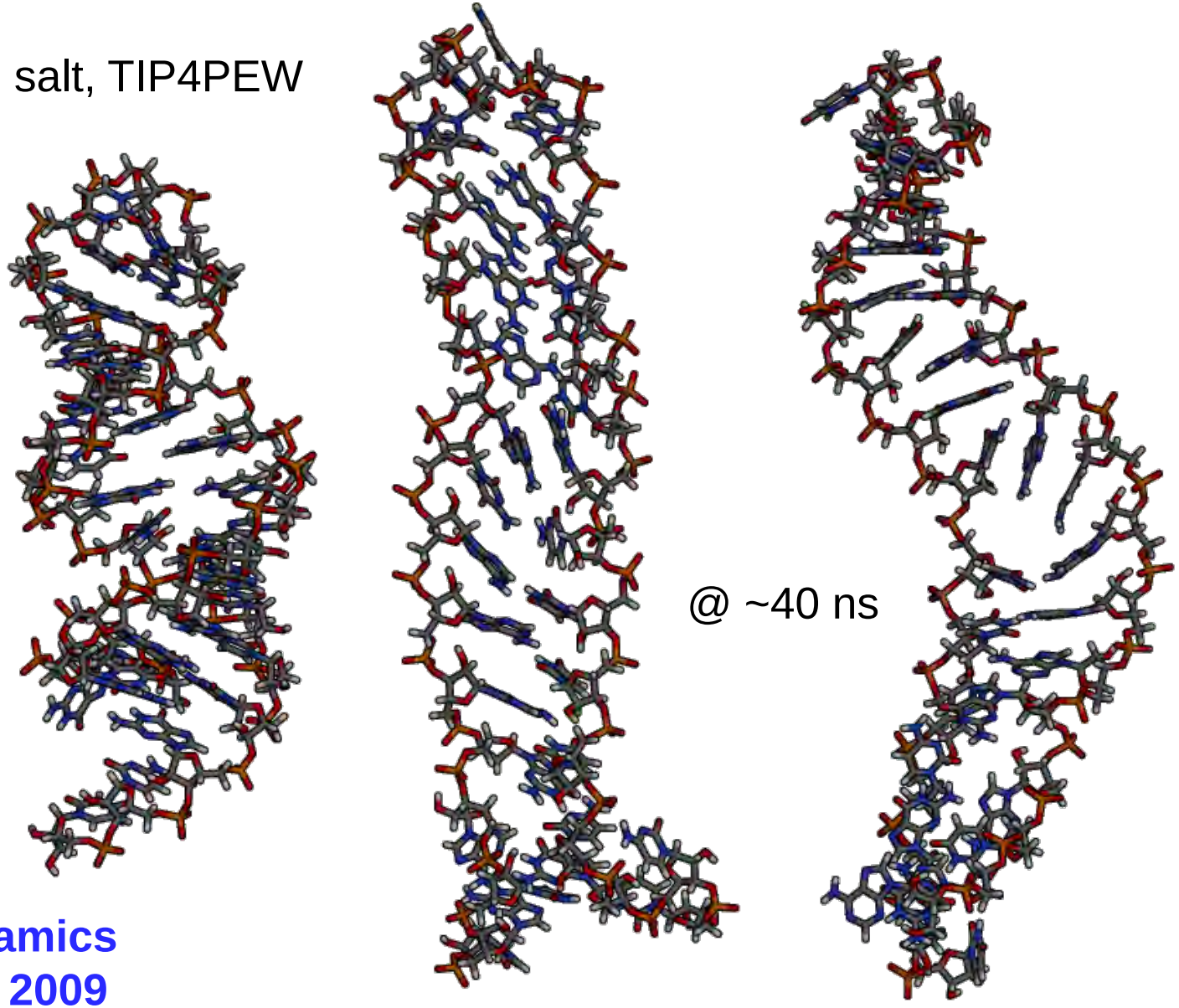


- Top cluster in each replica (10x replicas)
- Cluster w/ final 1 μs of each replica

Issues with balance and why I should know better?

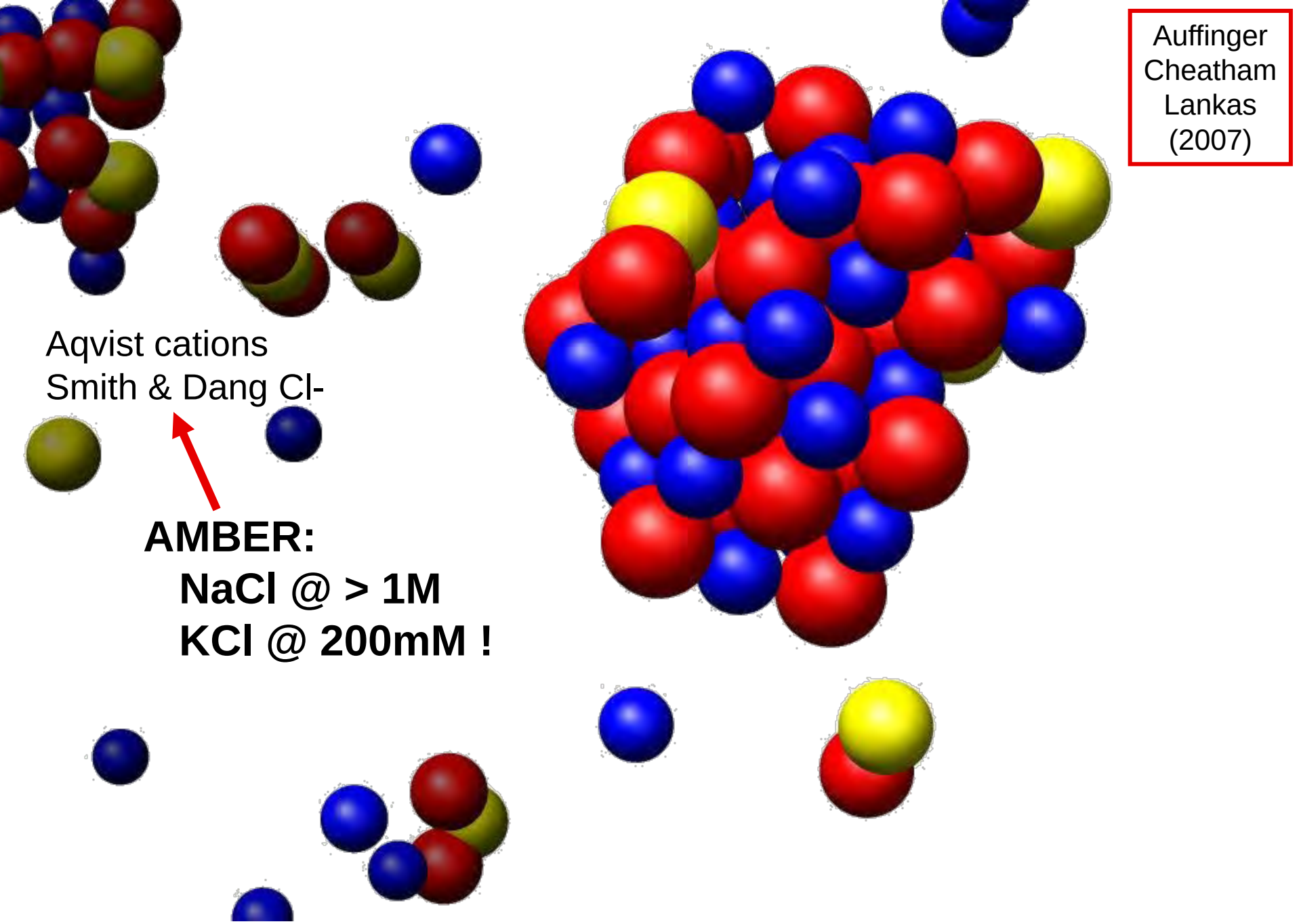
- RNA is sensitive to salt concentration & equilibration

ff99, NO salt, TIP4PEW

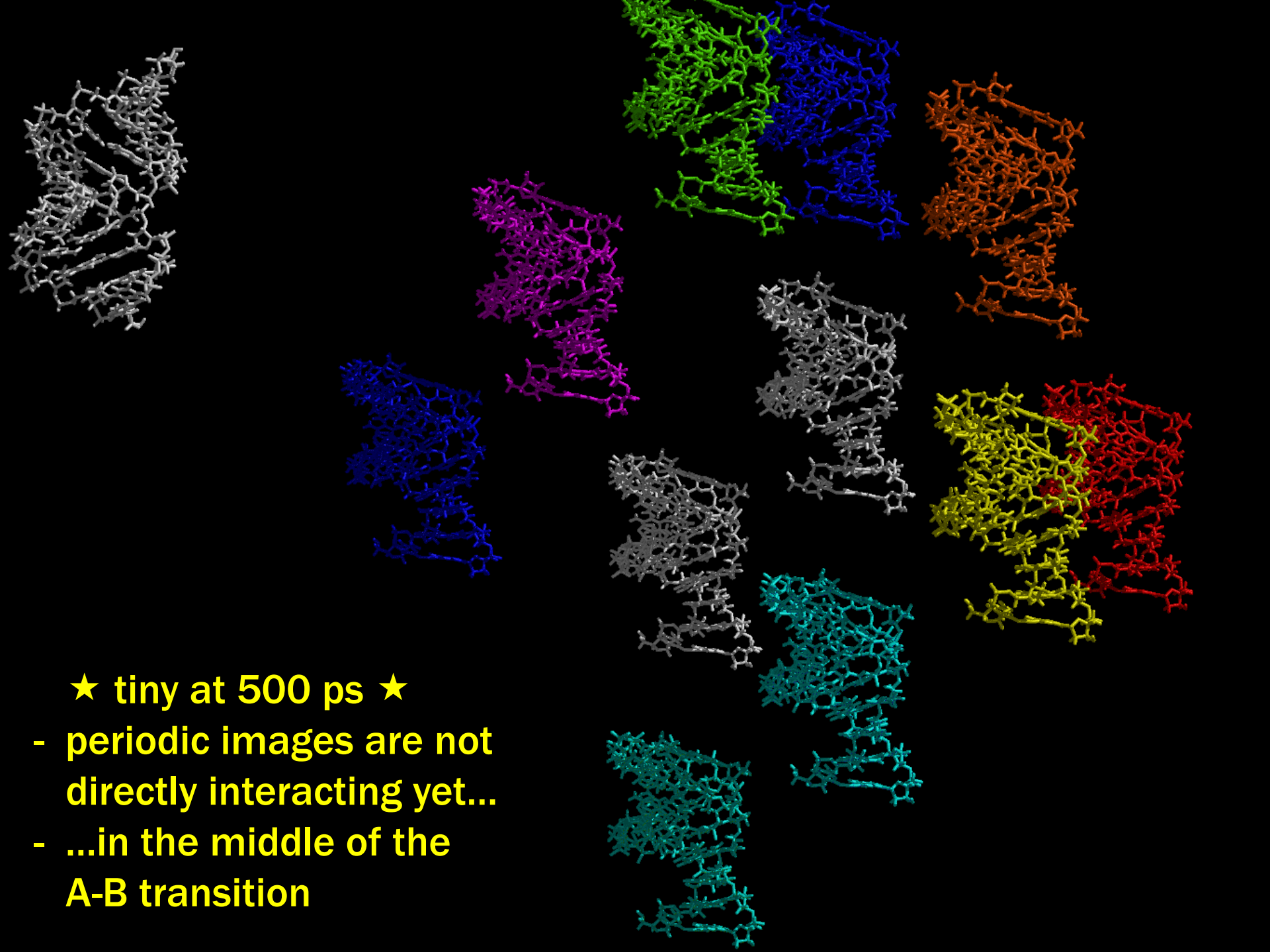


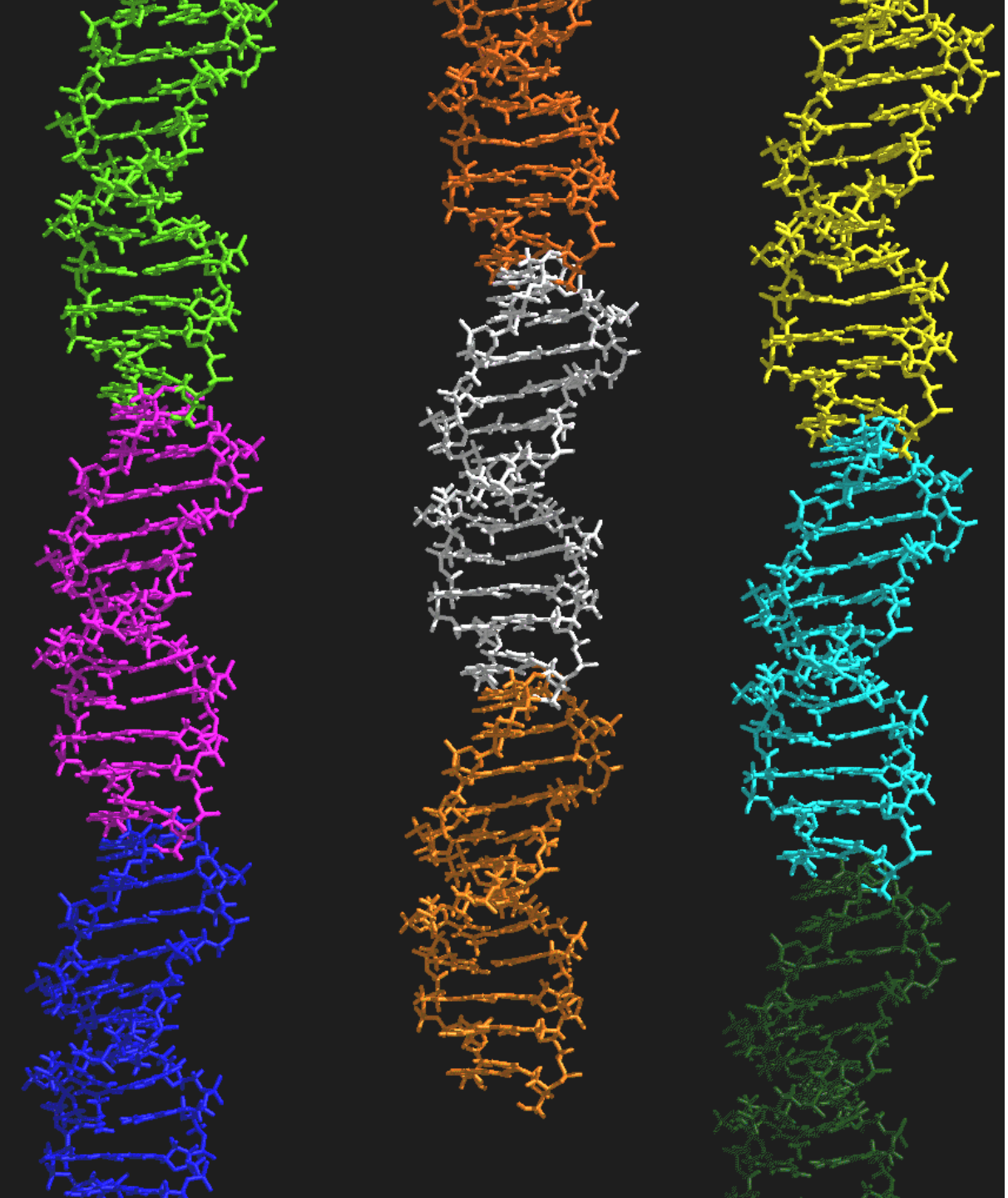
RNA dynamics
Telluride 2009

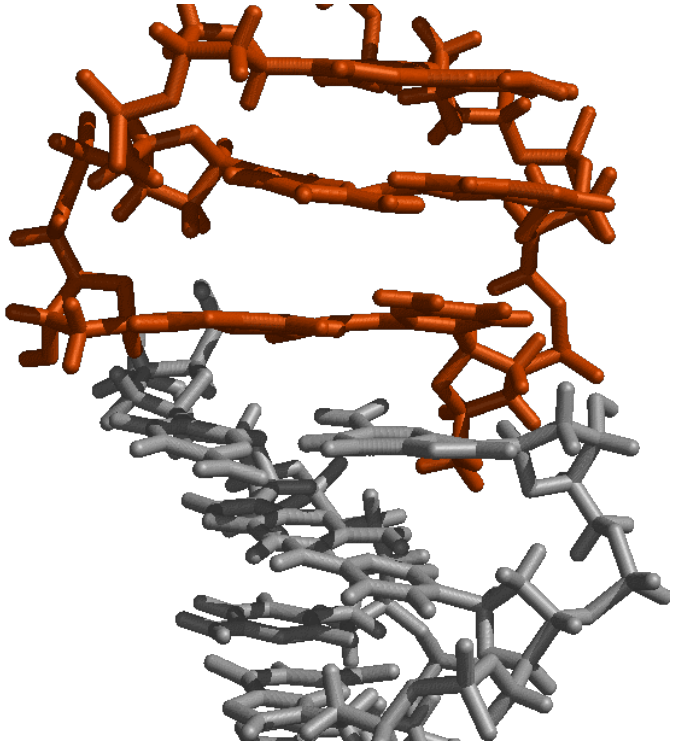
1MFY – influenza A C4 promoter (NMR)



[crystallization not seen with CHARMM all_27, Beglov&Roux < 4M]


★ tiny at 500 ps ★
- periodic images are not directly interacting yet...
- ...in the middle of the A-B transition

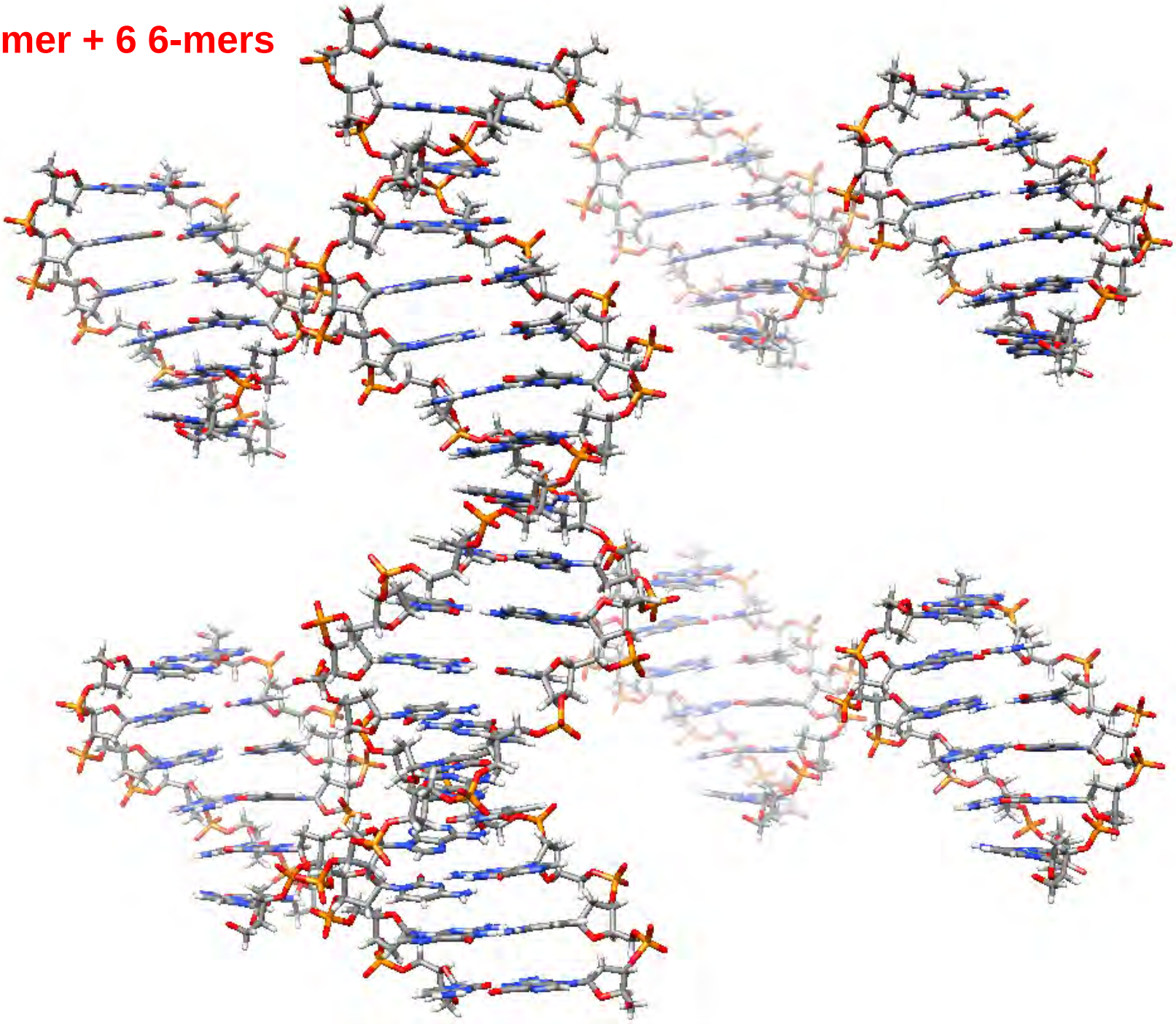


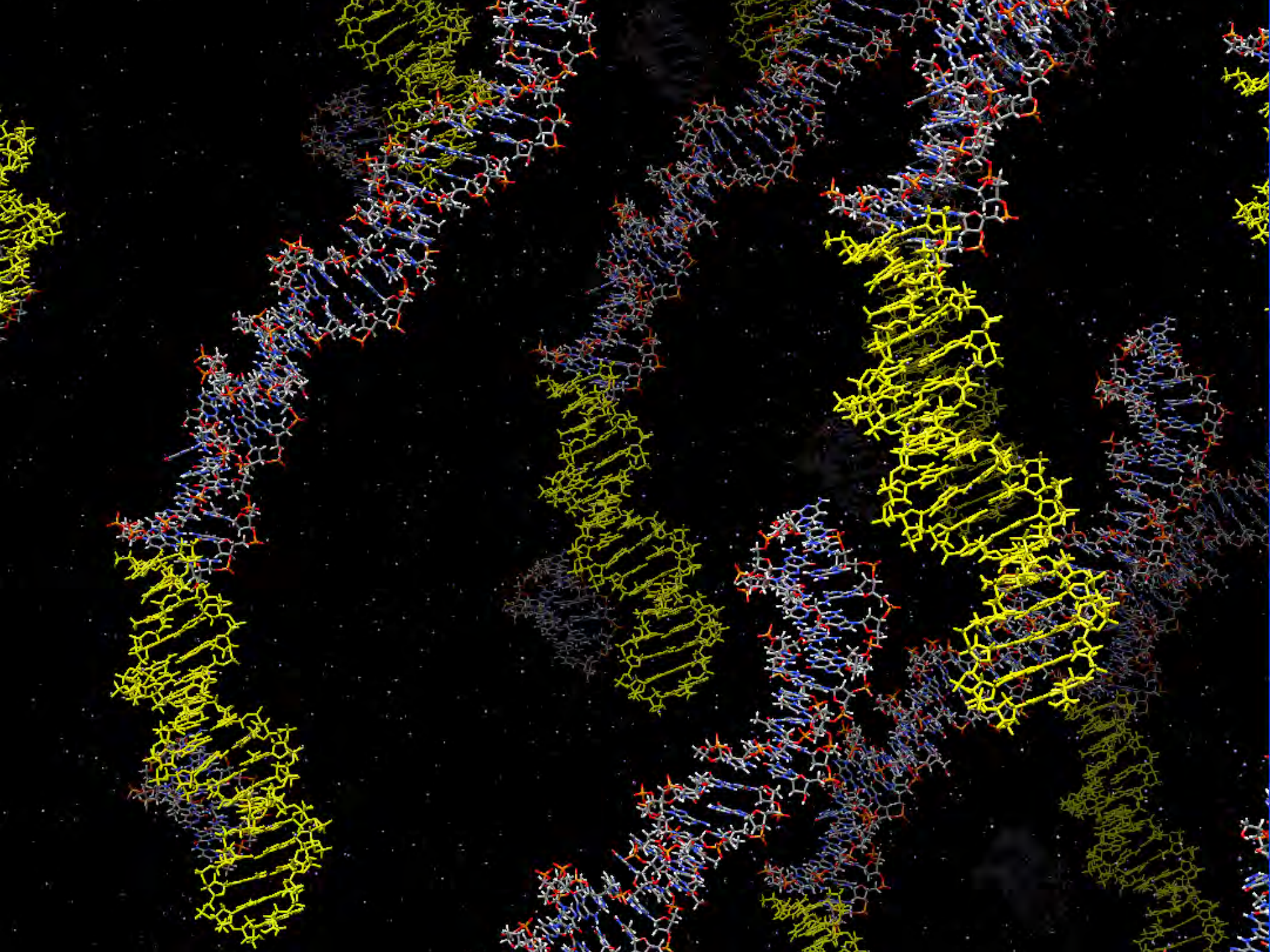


tiny at 40ns
Infinite crystal!

parm94 or parm99
at 1M or 4M salt

18-mer + 6 6-mers





What are some of the problems that still remain?

- Force field (mis)-balance
- Sampling – trapped conformations
- [force field / methods inter-operability]
- [judging convergence or overlap of independent simulations from different groups]
- [ion influences]

How to assess and improve? Model systems where we can (easily) get complete sampling… (di-, tetra-, …)

If force fields are ”broken”, can we still use them?

What if we bias with information from experiment?

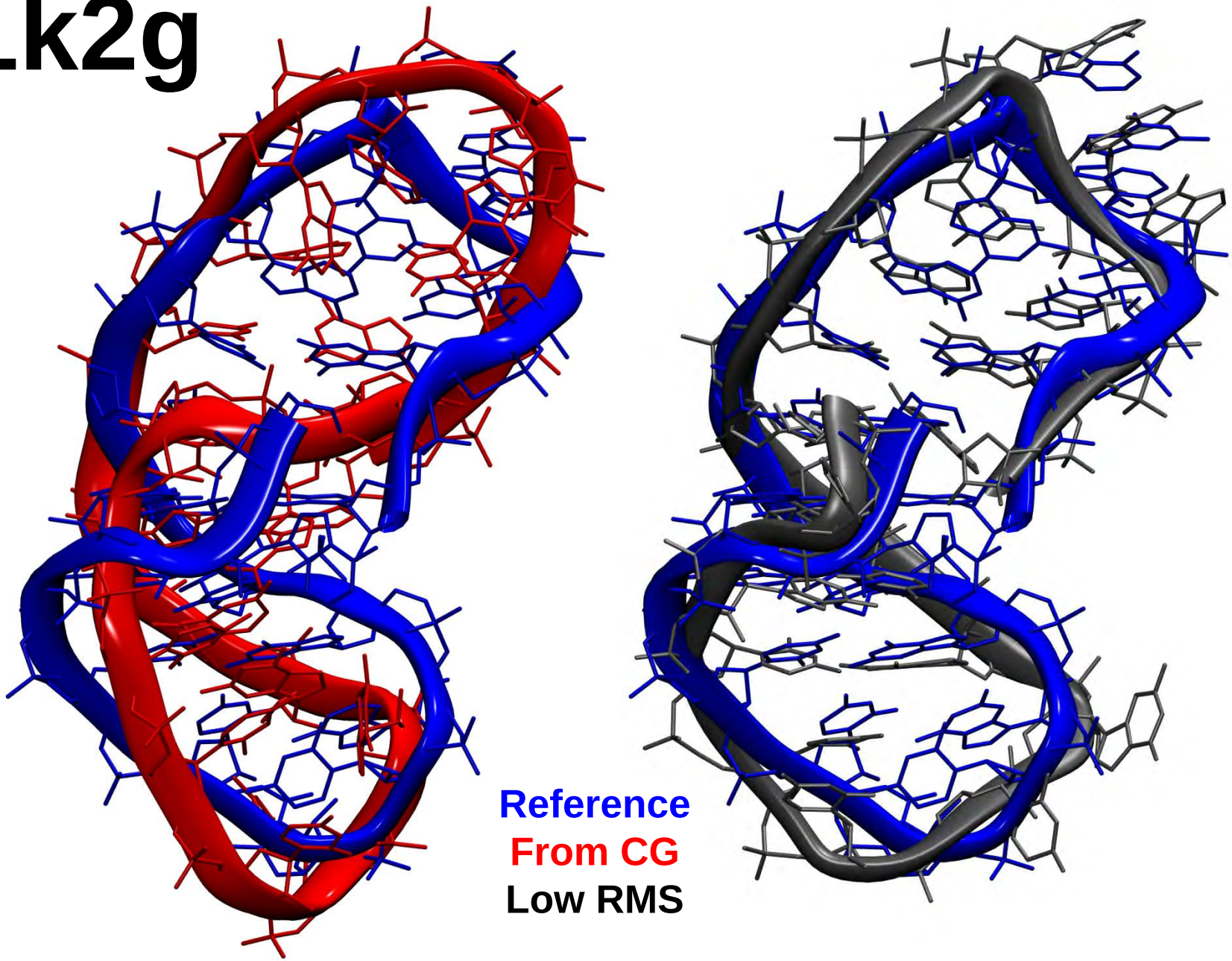
All Atom MD Refinements:

System	Lowest Starting RMSD to Native	Lowest Unrestrained RMSD	Lowest Restrained RMSD
1k2g	4.17	2.35	1.85
1a60	7.04	4.75	5.38
1evv	11.33	6.45	n/a
3pdr	17.11	18.17*	n/a

** this is the M-Box Riboswitch, crystalized with Mn^{2+} which wasn't included in the simulations (implicit solvent), so I'd expect this to do worse.*

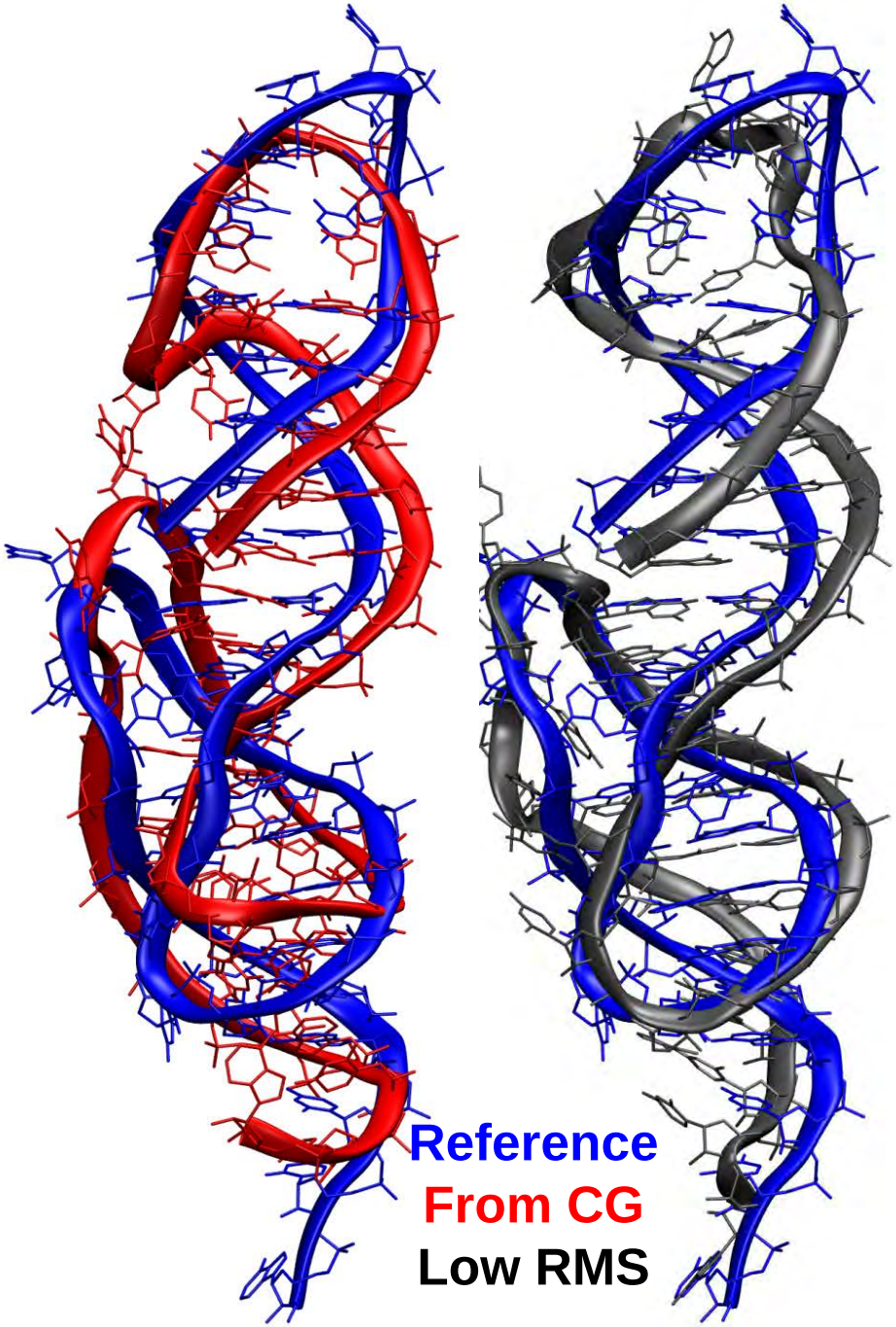
…refining Dokholyan CG structures with MD in implicit solvent…

1k2g

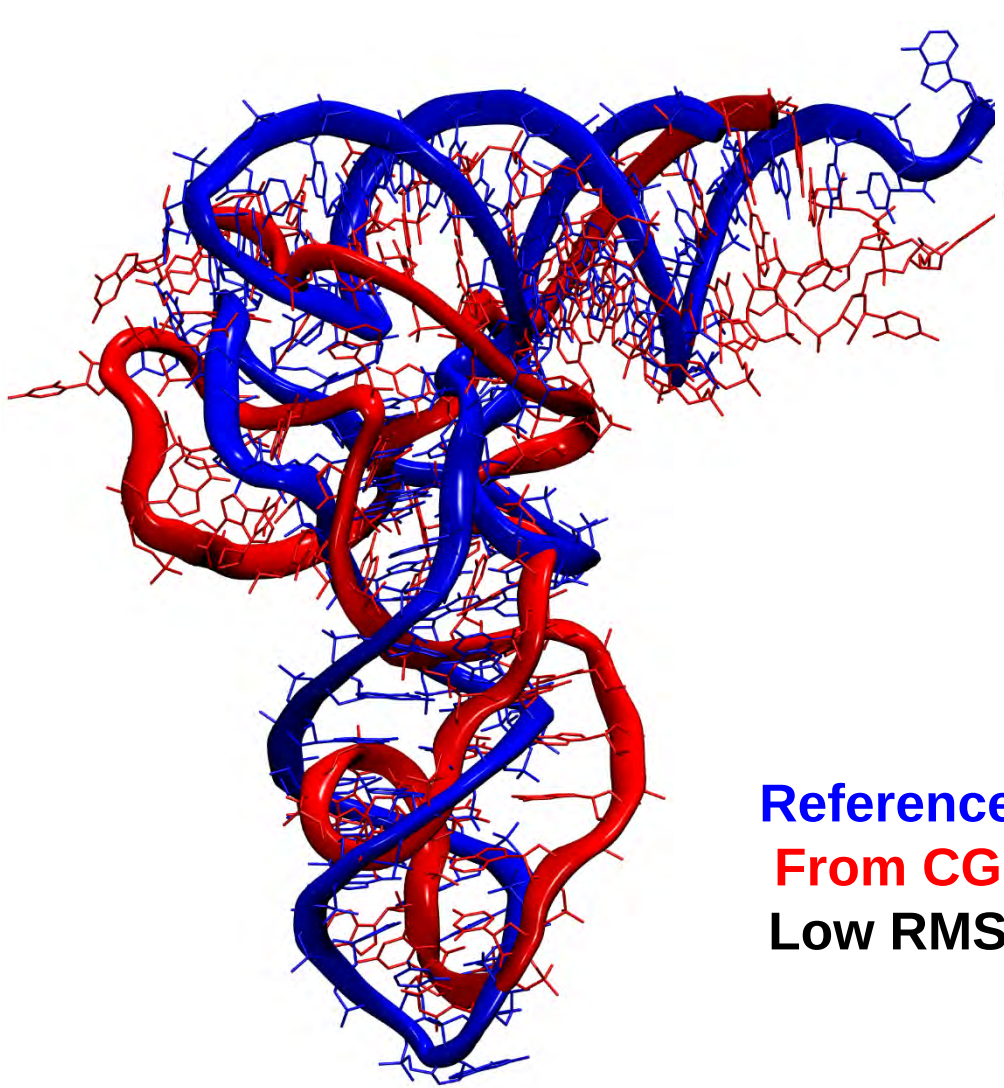


Reference
From CG
Low RMS

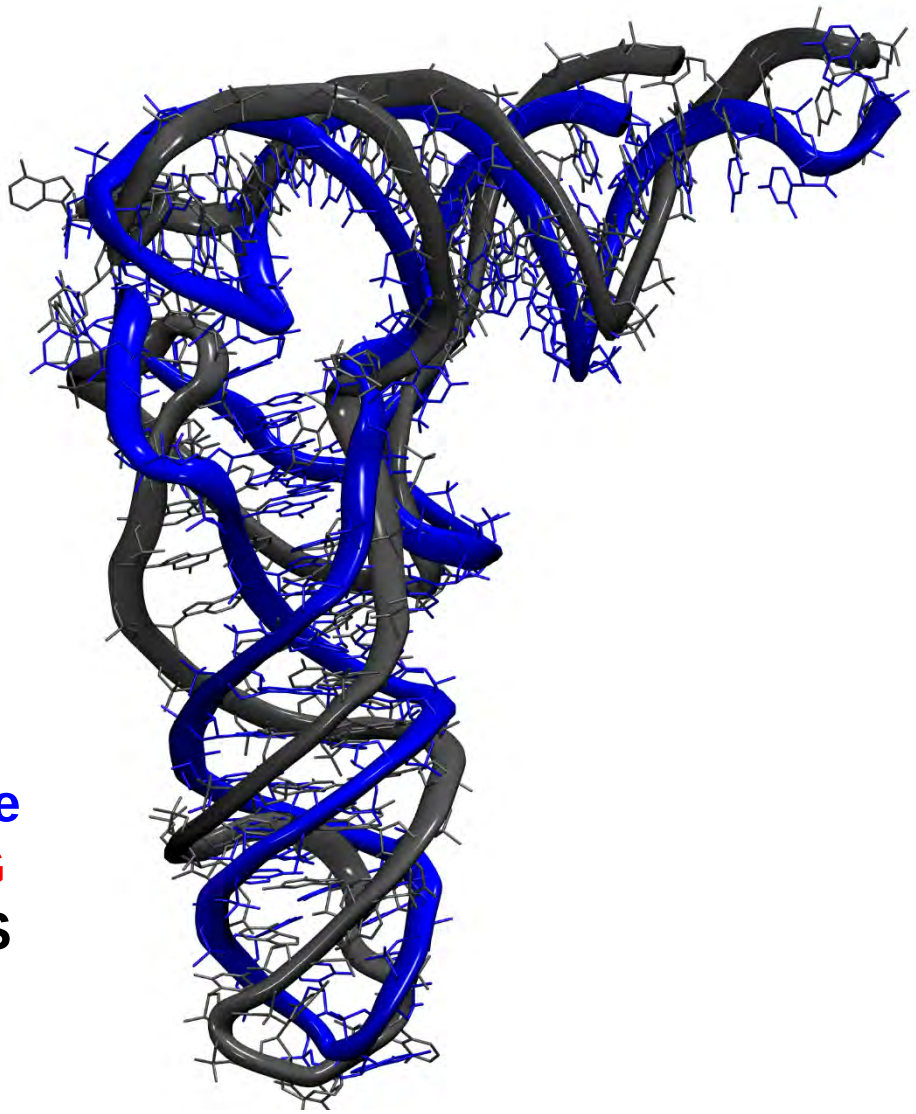
1a60



1evv



Reference
From CG
Low RMS



CPPTRAJ developments

levels of parallelism

(1) Ensemble processing (in || with MPI) – M-REMD

- convergence, reproducibility

(2) MPI over file / intra-file level parallelization

(3) OpenMP for computational intensive analyses

(4) CUDA for time-consuming distance calculations

- Supports general datasets: 1D, 2D, …
- Interactive analysis on large memory resources
- [energetic analyses]
- support for more file formats
- symmetric RMSD, atom map, multiple topologies

People: Niel Henriksen, Hamed Hayatshahi, Dan Roe, Julien Thibault, Kiu Shahrokh, Rodrigo Galindo, Christina Bergonzo, Sean Cornillie

$$$:

National Science Foundation
WHERE DISCOVERIES BEGIN

R01-GM098102:	"RNA-ligand interactions: sim. & experiment	~2015
R01-GM072049:	"P450 dehydrogenation mechanisms"	~2014
R01-GM081411:	"…simulation … refinement of nucleic acid"	~2013
NSF CHE-1266307	"CDS&E: Tools to facilitate deeper data analysis, …"	~2015
NSF "Blue Waters" PetaScale Resource Allocation for AMBER RNA		

Computer time:

PITTSBURGH SUPERCOMPUTING CENTER

D E Shaw Research

"Anton"
(3 past awards)

XSEDE
Extreme Science and Engineering Discovery Environment

XRAC MCA01S027
~10M core hours

BLUE WATERS
SUSTAINED PETASCALE COMPUTING

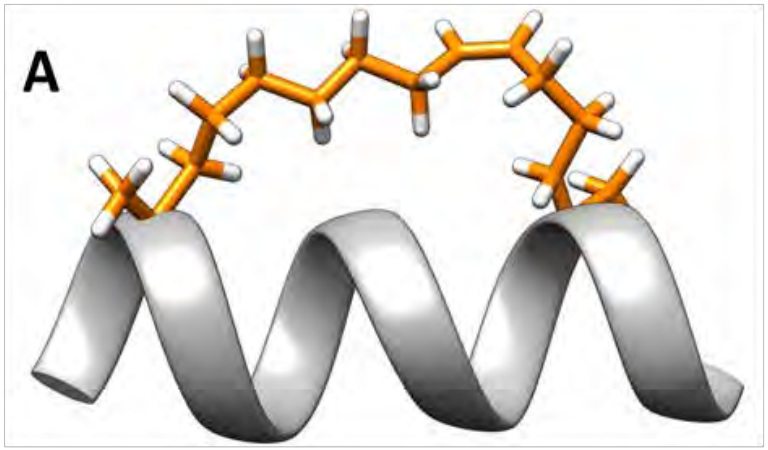
~12M GPU hours

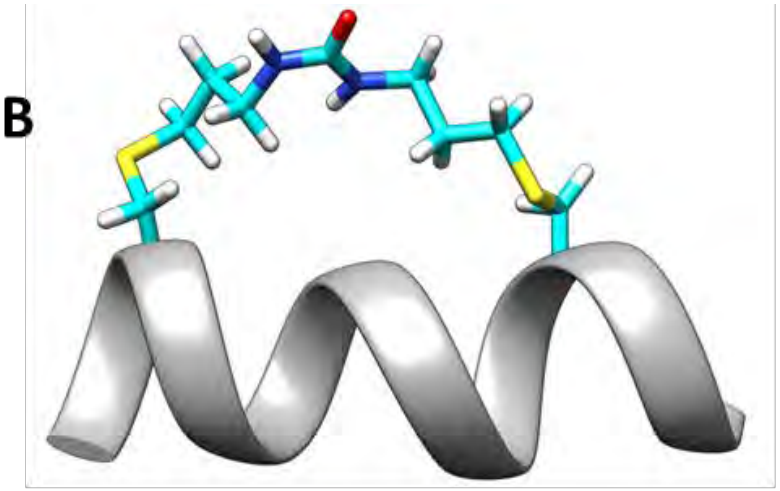
!!!

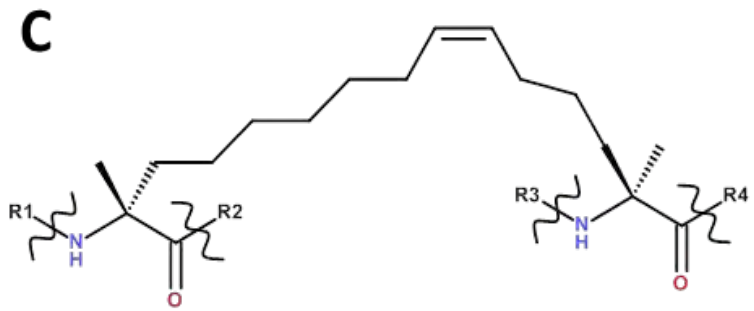
CHPC

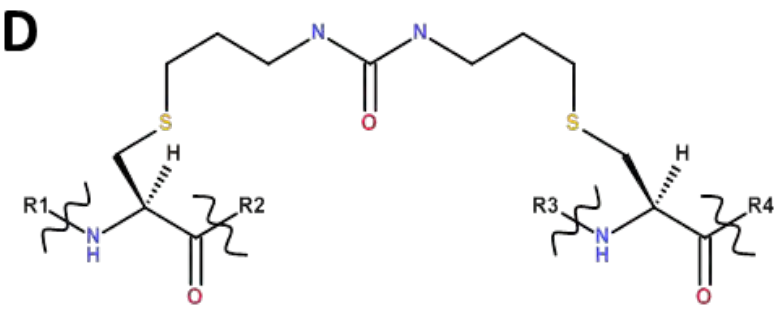
~3M hours

Attempting to stabilize CCmut3 (binds to BCR-CC)
Collaboration with Lim lab @ Utah



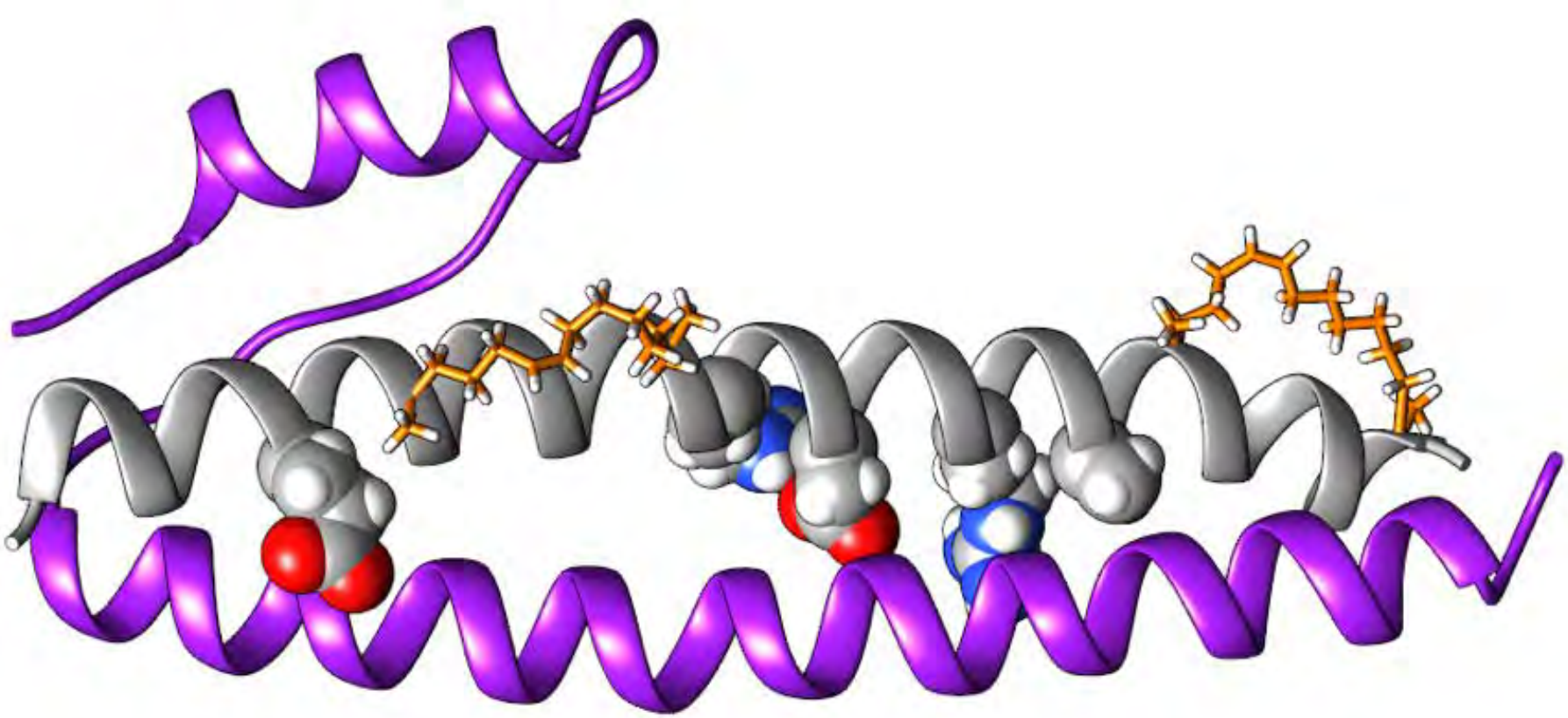




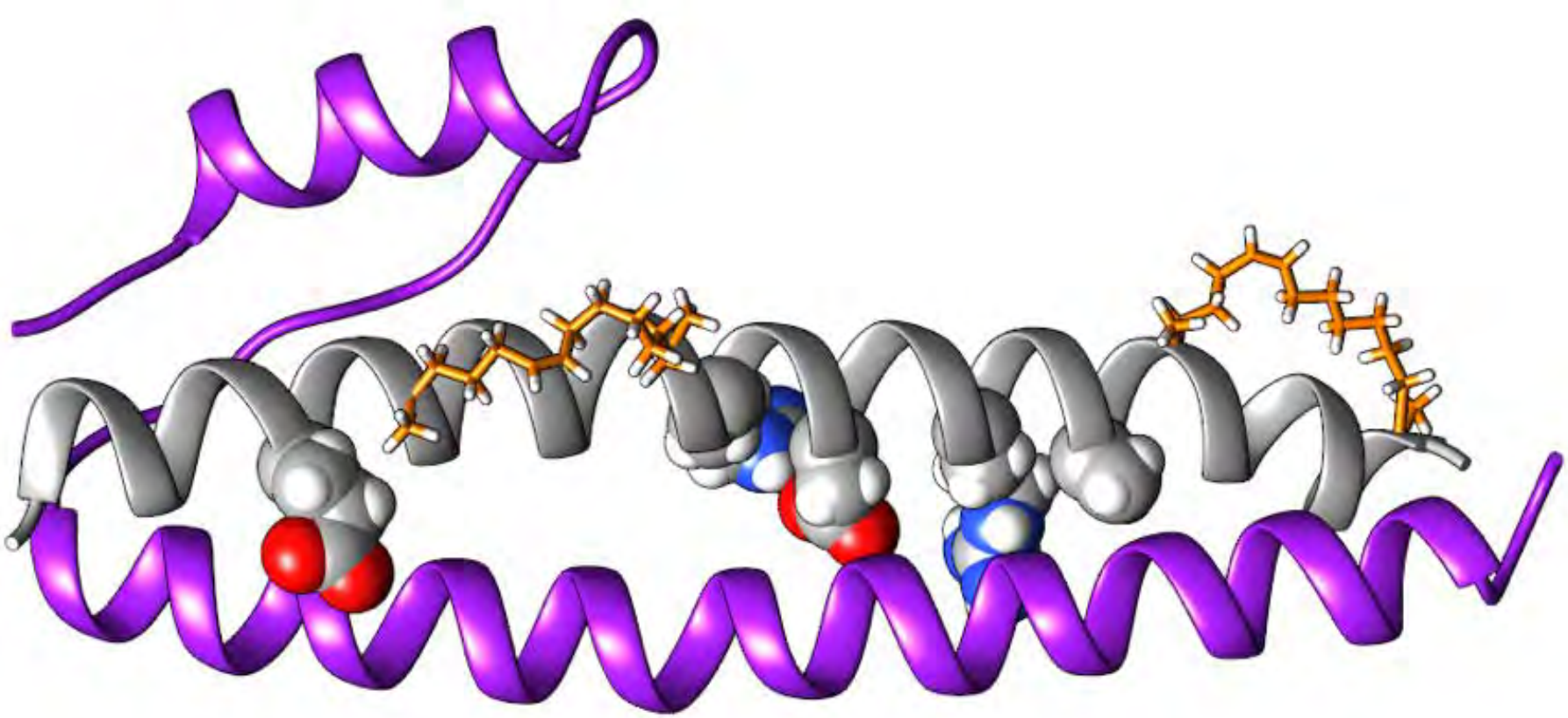


Stapled peptides of various flavors

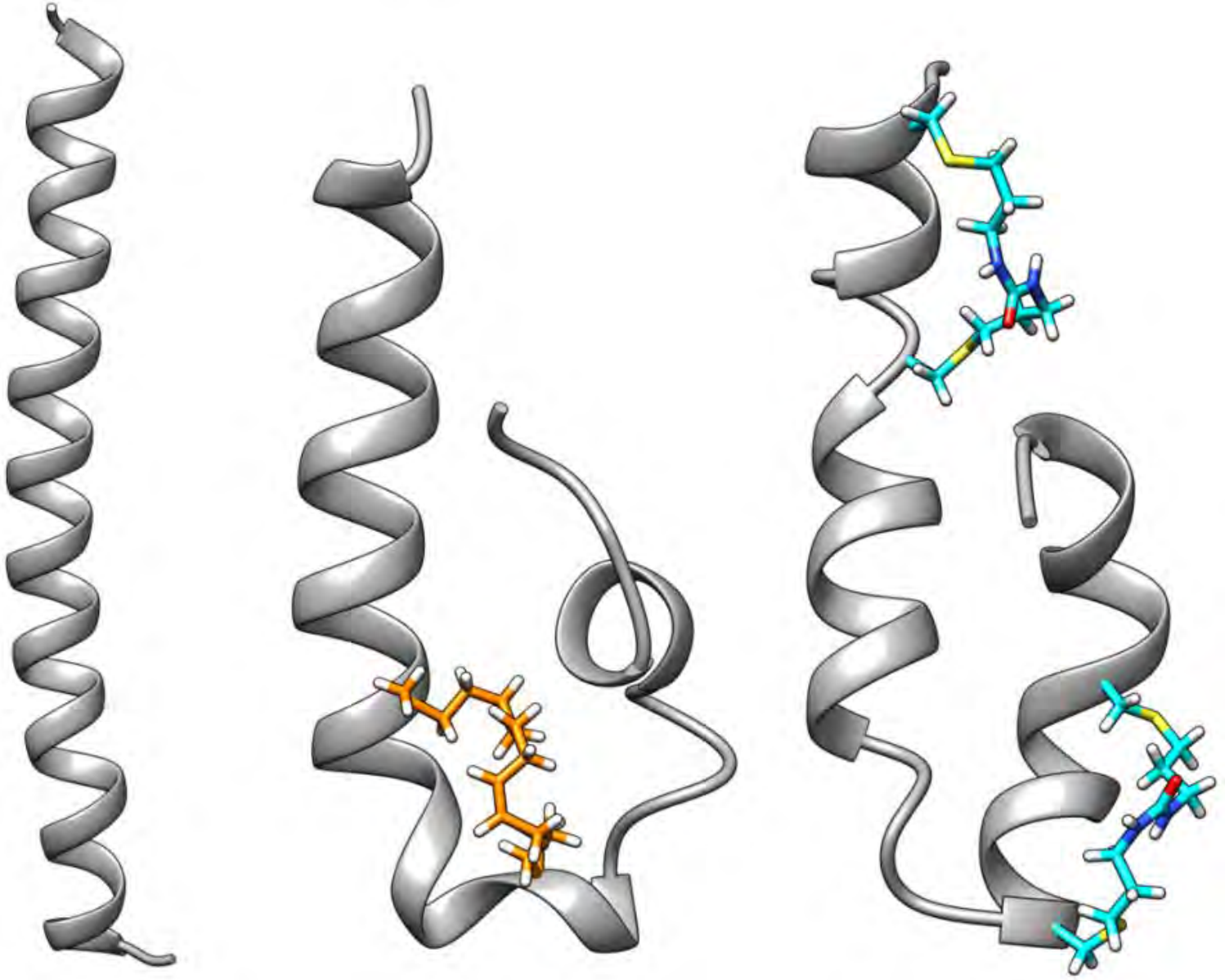
Stapled CCmut3 bound to BCR-CC generally more stable!
We can suggest best candidates (for staple location) to synthesize!!!



Stapled CCmut3 bound to BCR-CC generally more stable!
We can suggest best candidates (for staple location) to synthesize!!!



…but…



Free CCmut3
is ”stable”

Staples lead to “folding”, hydrophobic collapse,
enhanced susceptibility to proteolysis

People: Niel Henriksen, Hamed Hayatshahi, Dan Roe, Julien Thibault, Kiu Shahrokh, Rodrigo Galindo, Christina Bergonzo, Sean Cornillie, James Robertson

$$$:

National Science Foundation
WHERE DISCOVERIES BEGIN

R01-GM098102:	"RNA-ligand interactions: sim. & experiment	~2015
R01-GM072049:	"P450 dehydrogenation mechanisms"	~2014
R01-GM081411:	"…simulation … refinement of nucleic acid"	~2013
NSF CHE-1266307	"CDS&E: Tools to facilitate deeper data analysis, …"	~2015
NSF "Blue Waters" PetaScale Resource Allocation for AMBER RNA		

Computer time:

PITTSBURGH SUPERCOMPUTING CENTER

D E Shaw Research

"Anton"
(3 past awards)

XSEDE
Extreme Science and Engineering Discovery Environment

XRAC MCA01S027
~10M core hours

BLUE WATERS
SUSTAINED PETASCALE COMPUTING

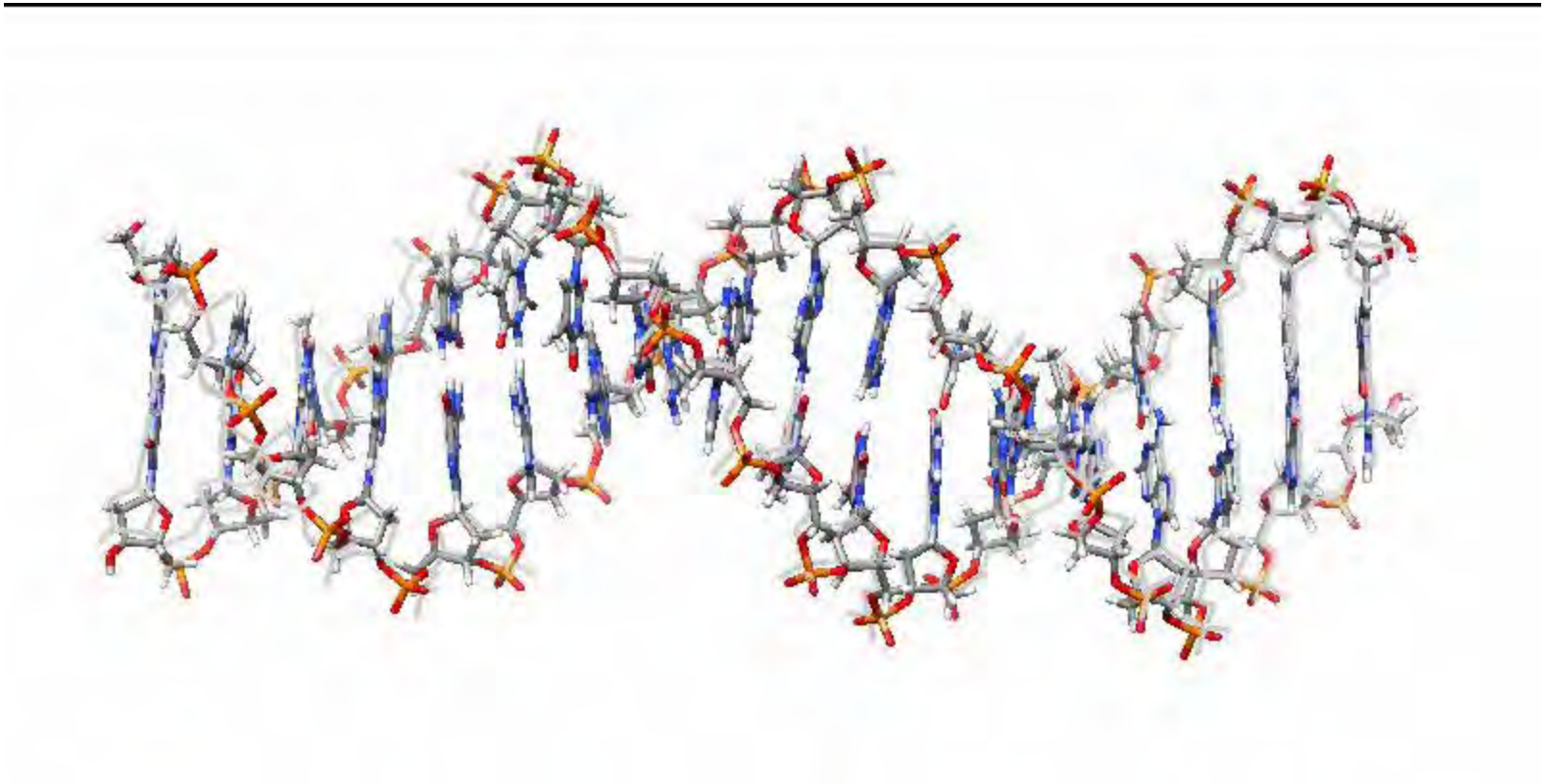
~12M GPU hours

!!!

CHPC

~3M hours

questions?



2 ns intervals, 10 ns running average, every 5th frame (~10 us).