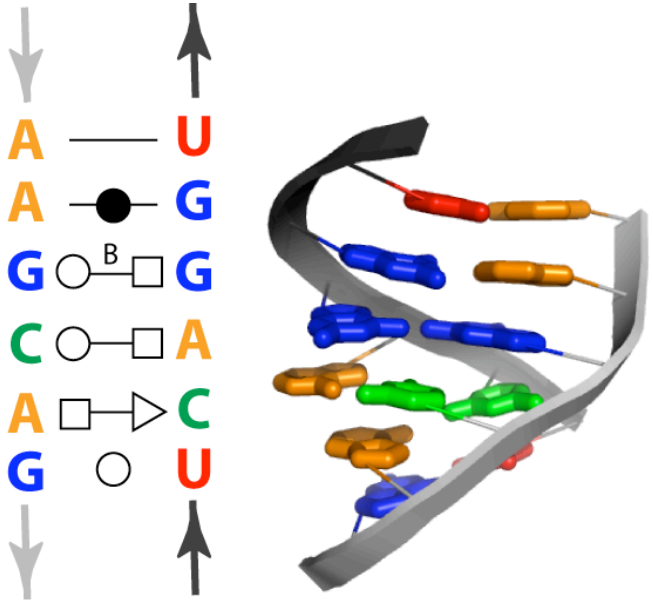


Little RNA puzzles

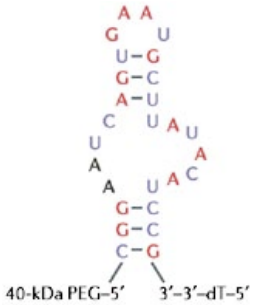
Rhiju Das, Kyle Beauchamp, Parin Sripakdeevong

Stanford Biochemistry + Biophysics

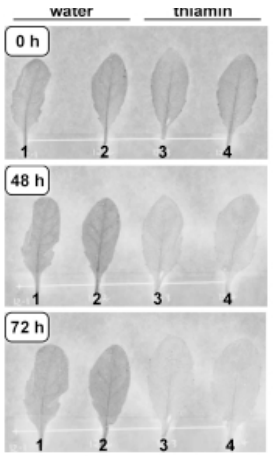
RosettaCon, August 4, 2009

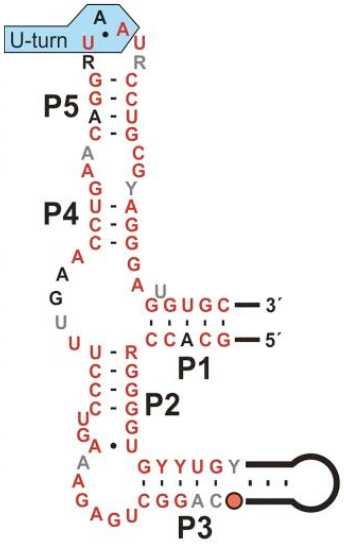
A flourishing RNA world

Engineered ribozymes and aptamers



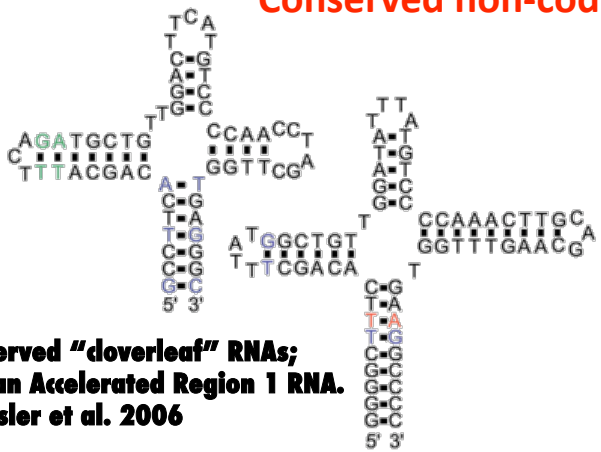
"Riboswitches"



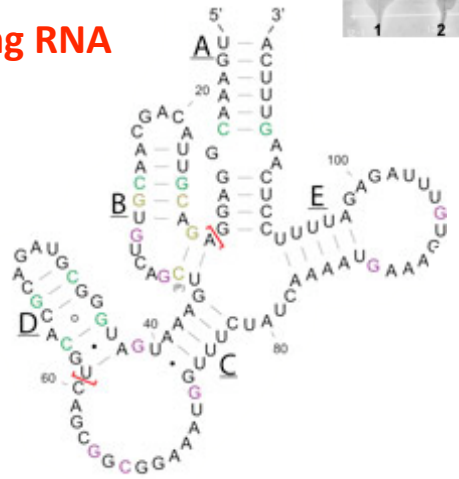


Breaker & colleagues, 2007

Conserved non-coding RNA

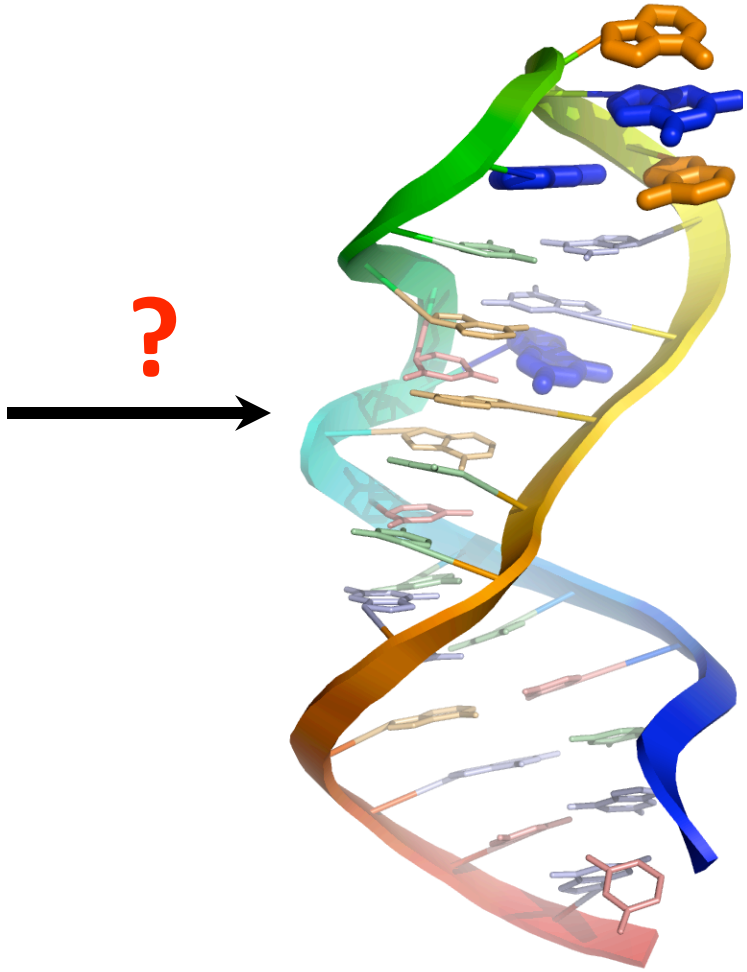


Conserved "cloverleaf" RNAs; Human Accelerated Region 1 RNA. Haussler et al. 2006



Goal: a predictive understanding of RNA self-assembly

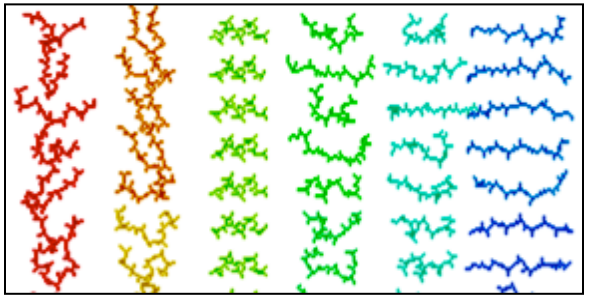
ugcuccu
aguacga
gaggacc
ggagug



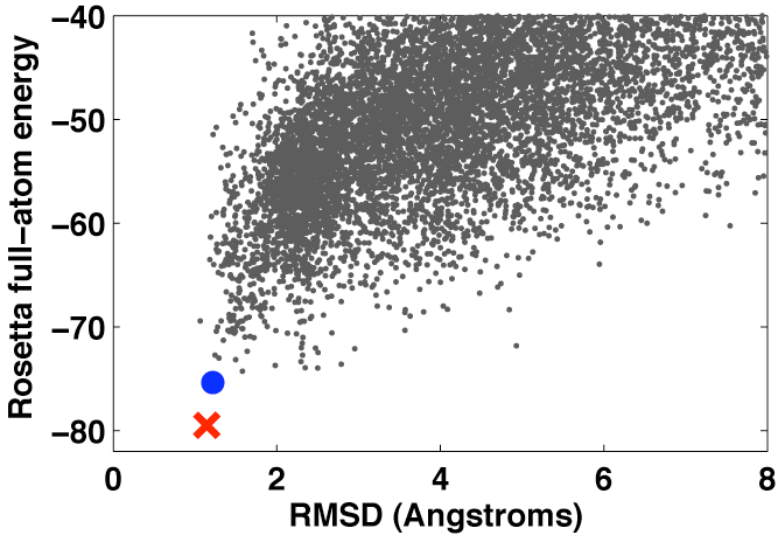
A Rosetta tour through the RNA world

AGCU

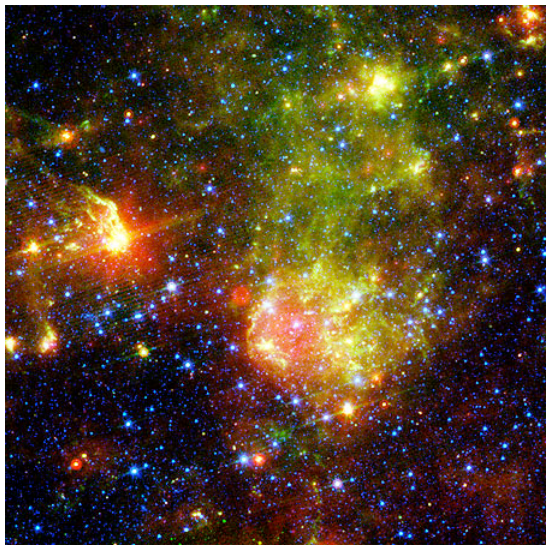
Why RNA prediction should be easy



Fragment assembly and all that



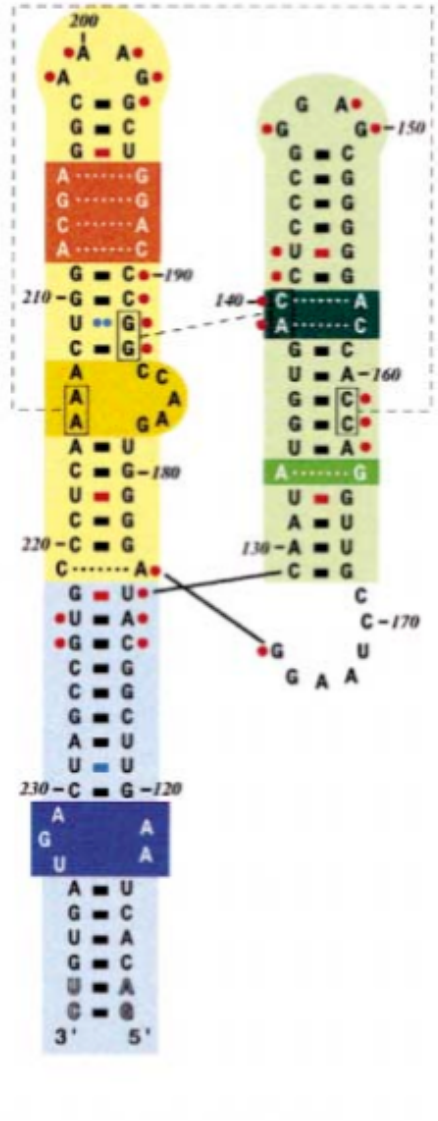
Achieving near-atomic accuracy



Beating the “astronomical” conformational sampling problem

Why RNA should be easy to model

GACACUAAGUUCGGCA
UCAAUAUGGUGACCUC
CCGGGAGCGGGGGACC
ACCAGGUUGCCUAGAG
GGGUGAACCGGCCCAG
GUCGGAAACGGAGCAG
GUCAAAACUCCCGUGC
UGAUCAGUAGUGU

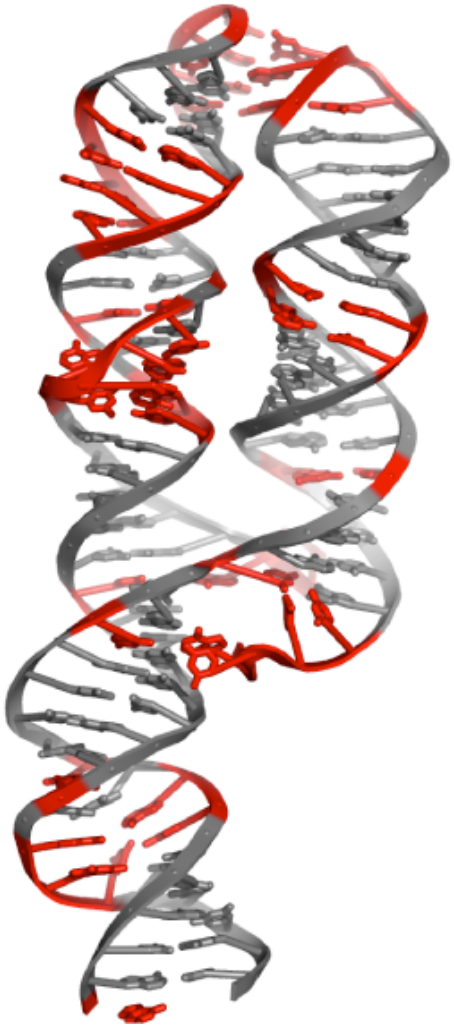


Signal Recognition Particle RNA
Oubridge et al., 2002

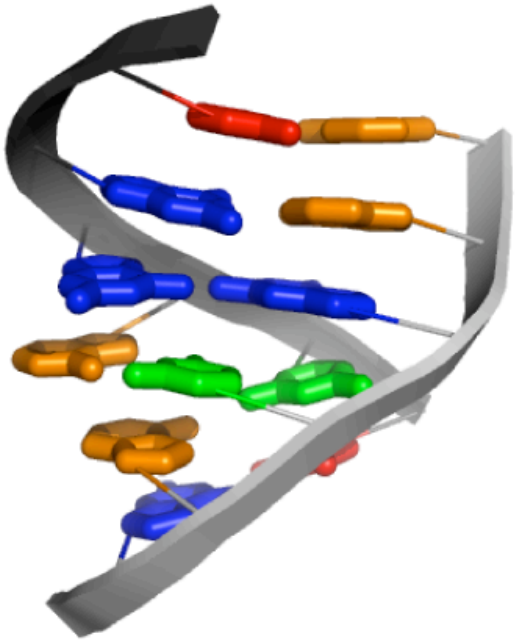
Why RNA should be easy to model

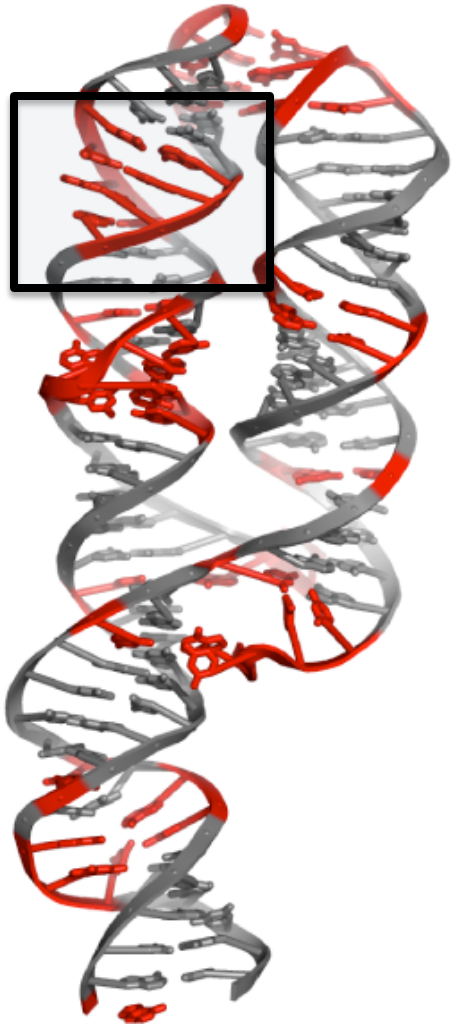
Canonical double helices

Non-canonical regions



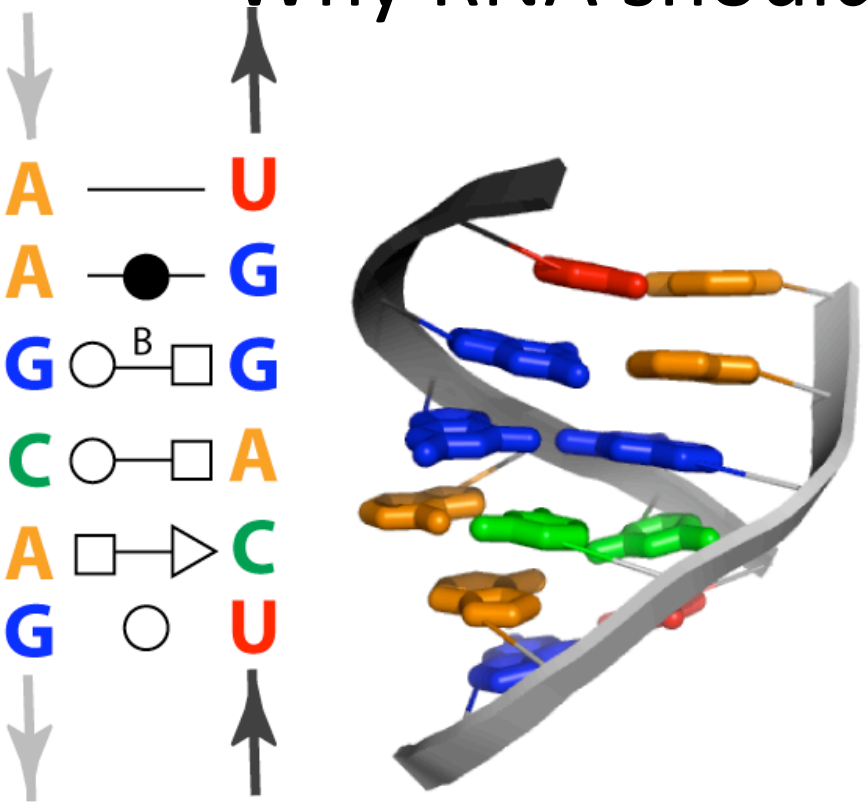
Why RNA should be easy to model

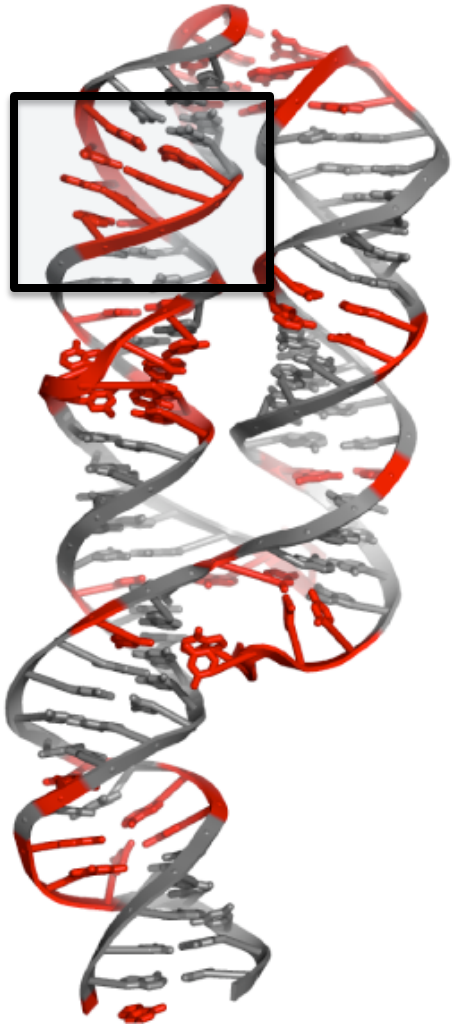




- Modules <20 residues
- A four letter alphabet.

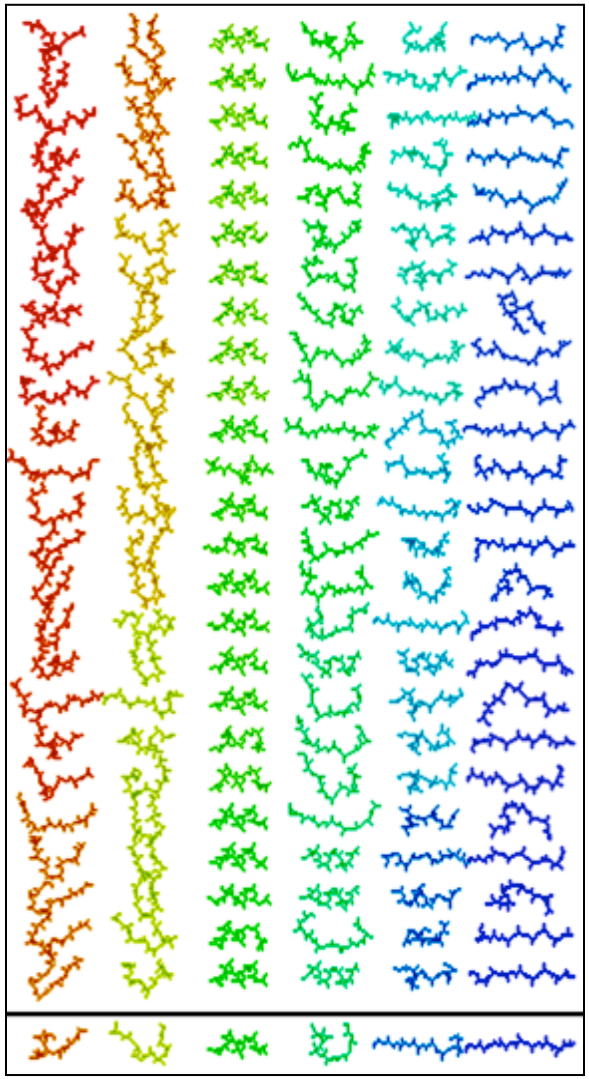
Why RNA should be easy to model

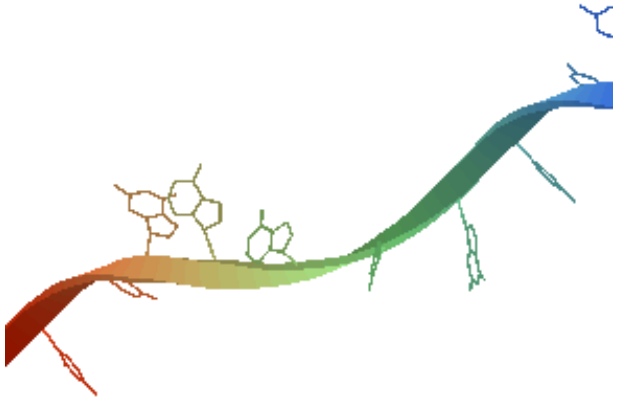




- Modules <20 residues
- A four letter alphabet.
- Stereotyped side-chain interactions (Leontis/Westhof classification)

De novo modeling

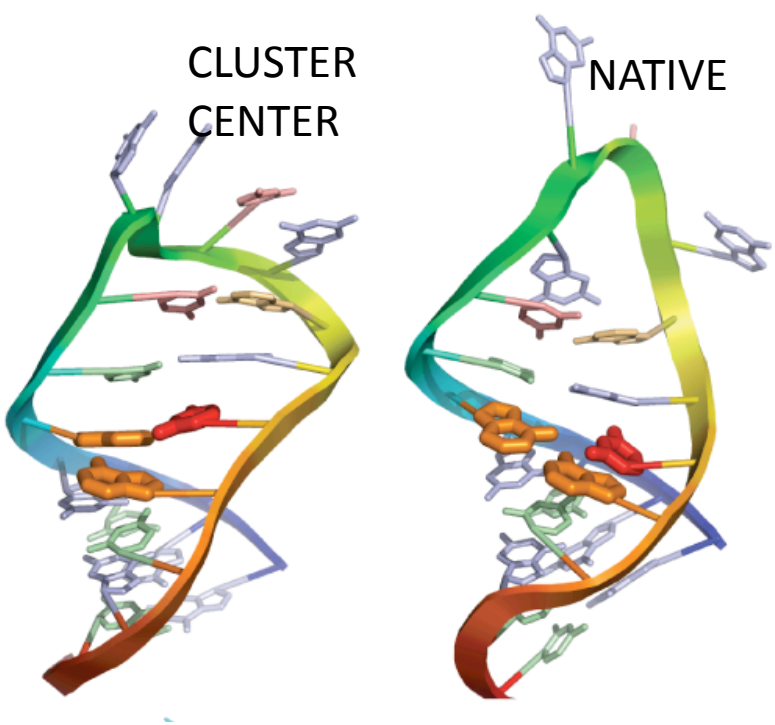




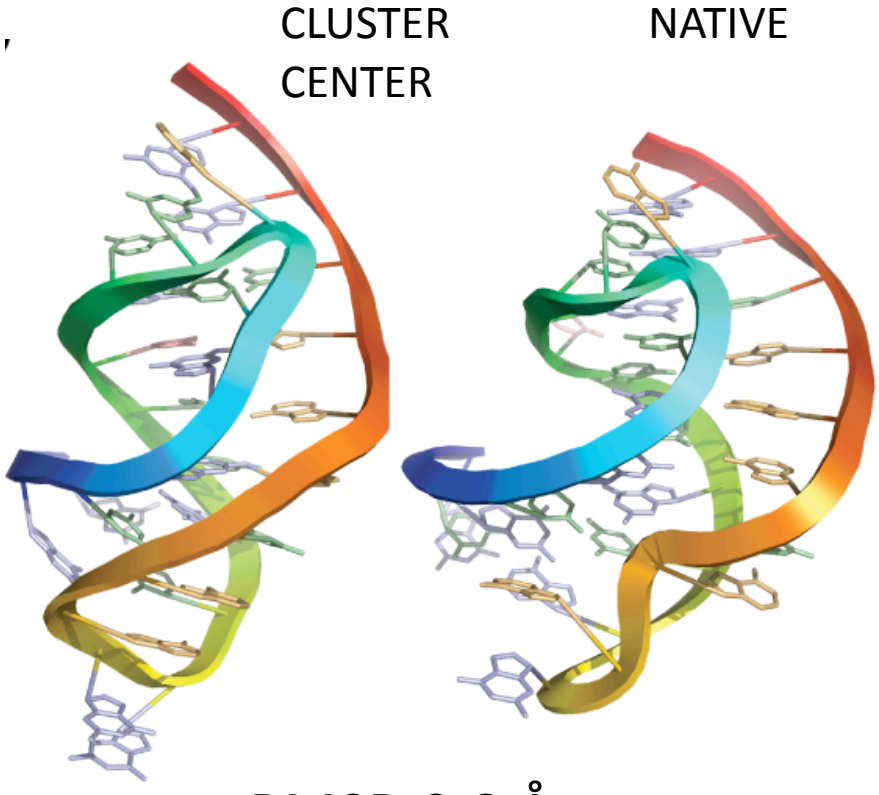
Fragment **A**ssembly of **RNA** (**FARNA**)

Energy: Just maximize base pairs and base stacks.

Low resolution modeling: OK up to a point



RMSD 4.0 Å
Stem loop in HIV-1 RNA



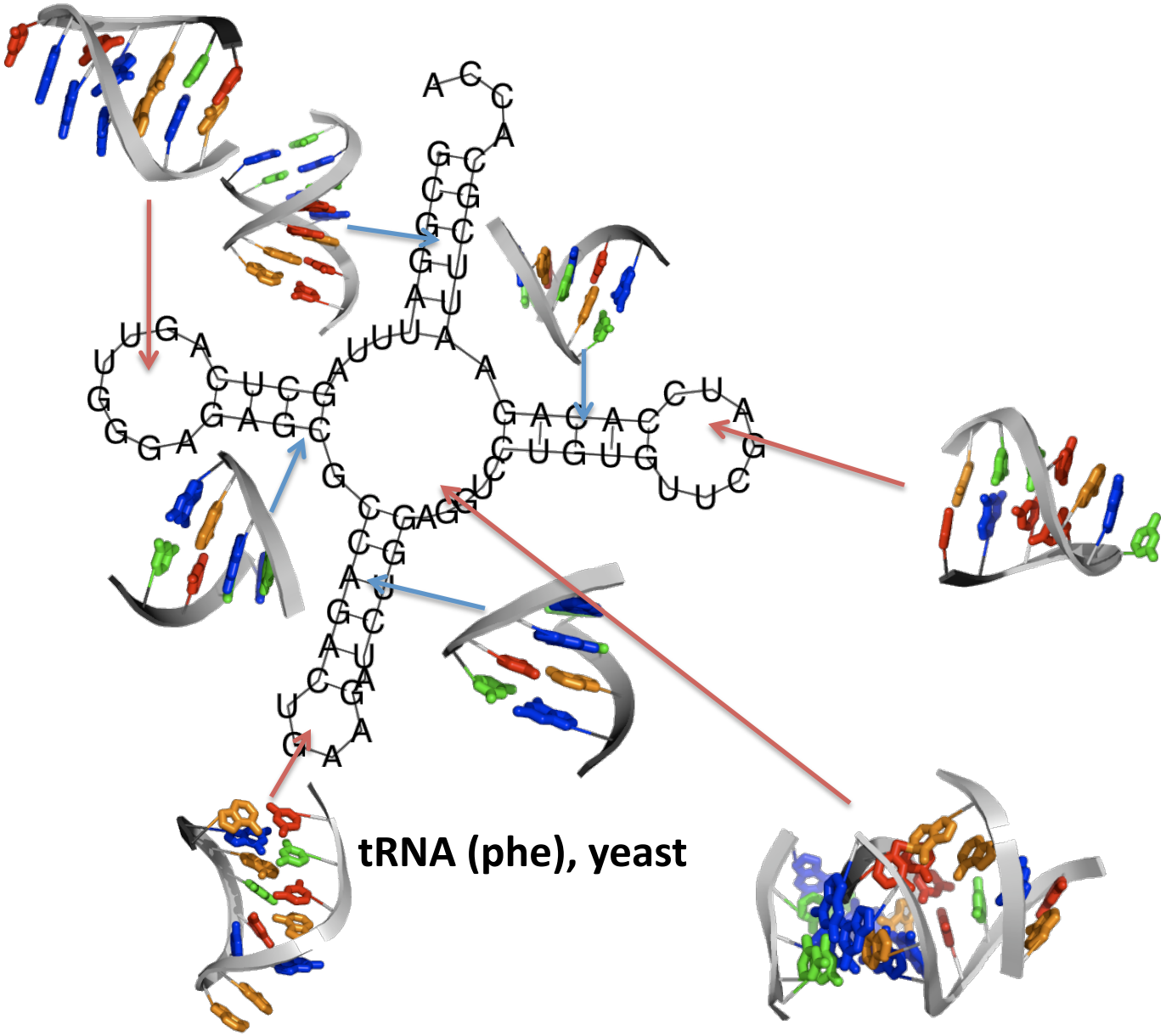
RMSD 3.8 Å
Viral pseudoknot

Similar accuracies now achieved with other approaches:

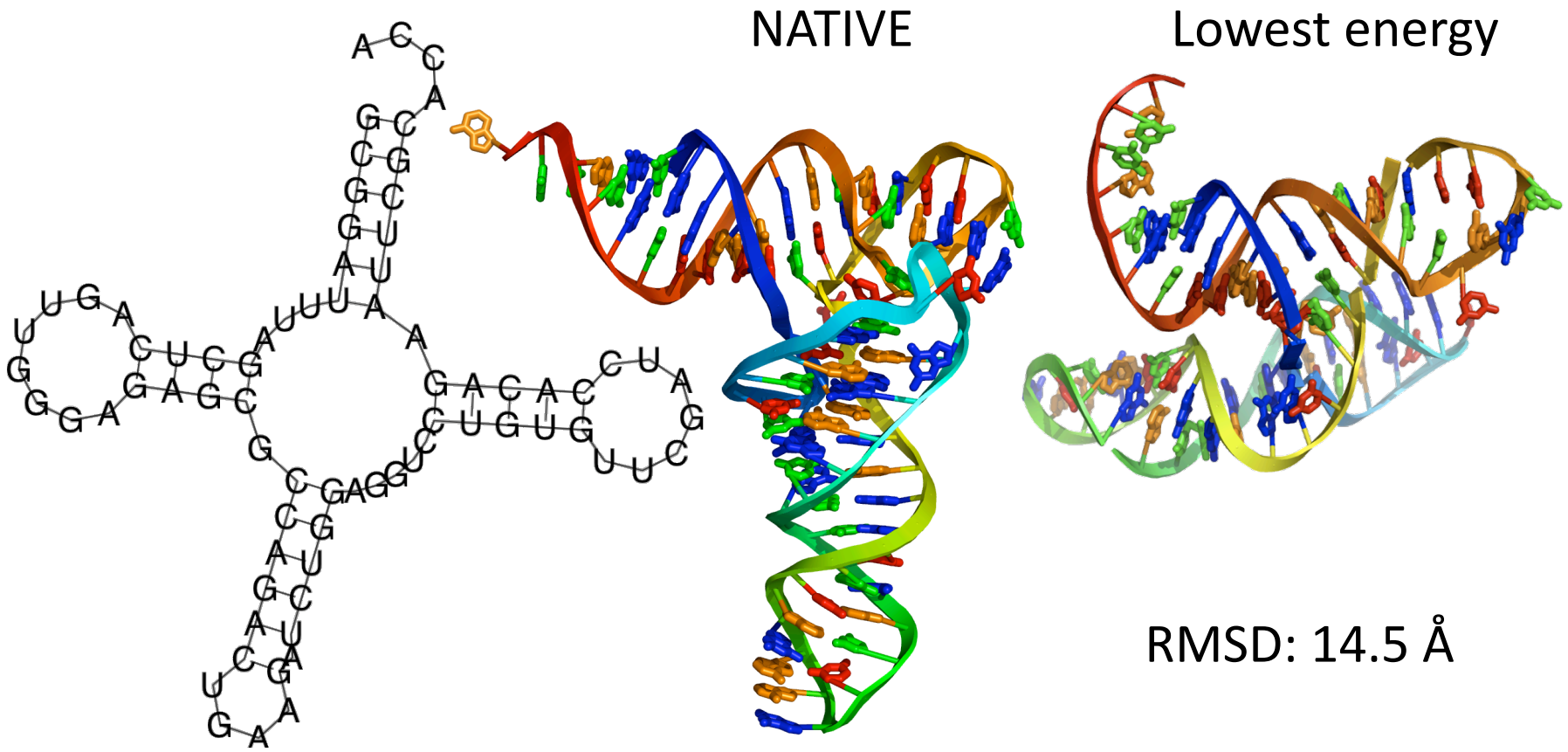
MC-SYM/MC-FOLD (Major + colleagues, 2008)

Discrete Molecular dynamics (Dokholyan, 2008)

Low resolution modeling: OK up to a point

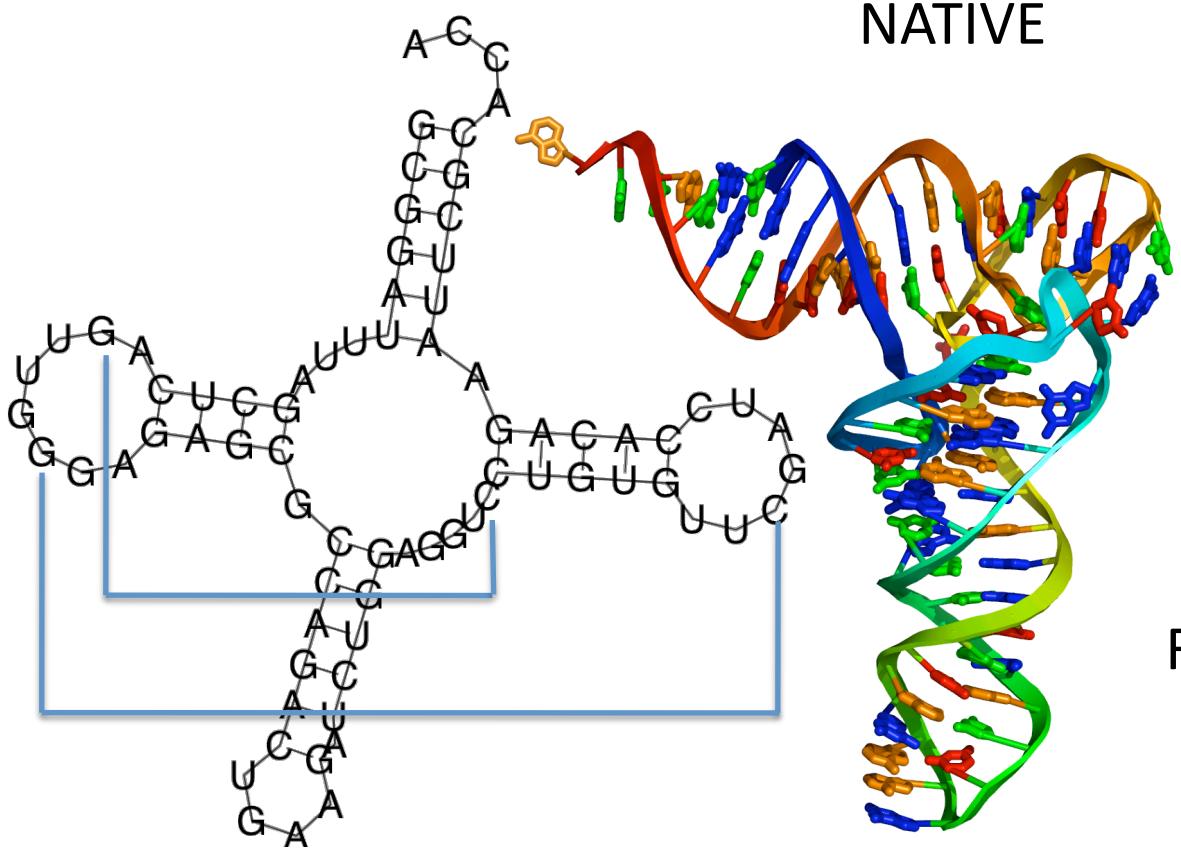


Low resolution modeling: OK up to a point



tRNA (phe), yeast

Low resolution modeling: OK up to a point



NATIVE

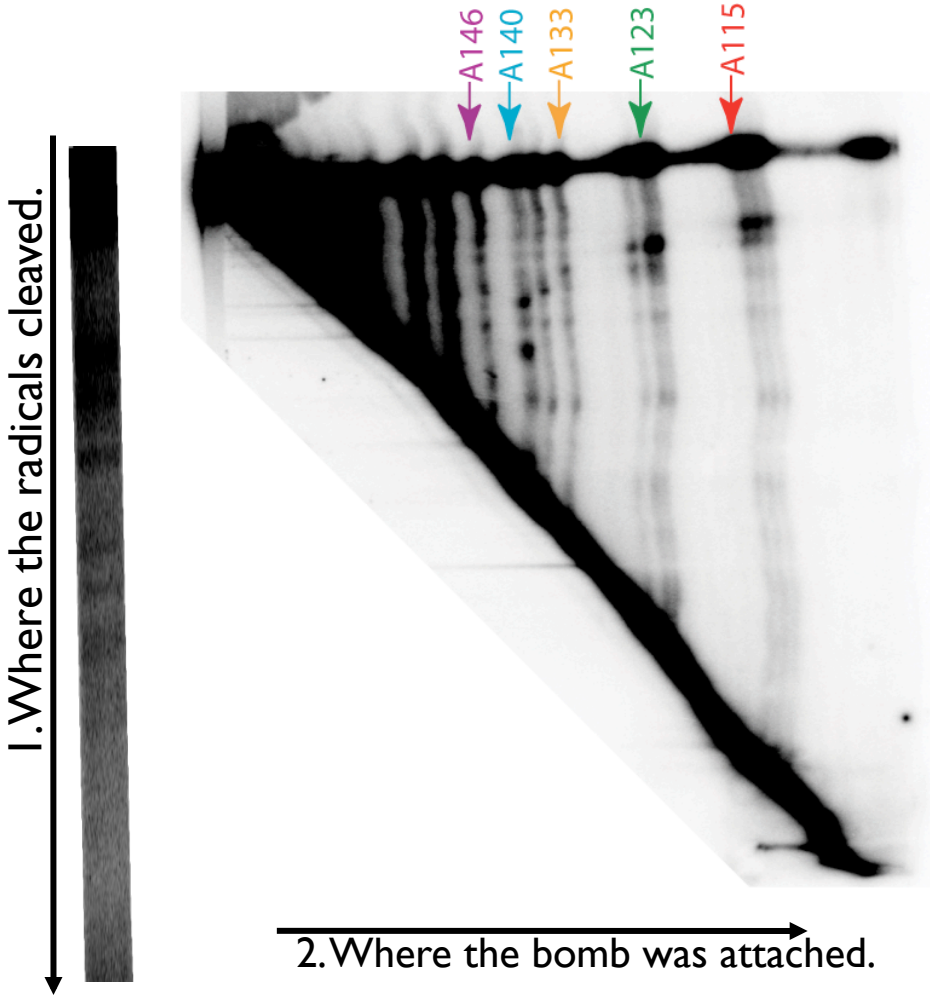
Lowest energy with
2 constraints

RMSD: 6.1 Å

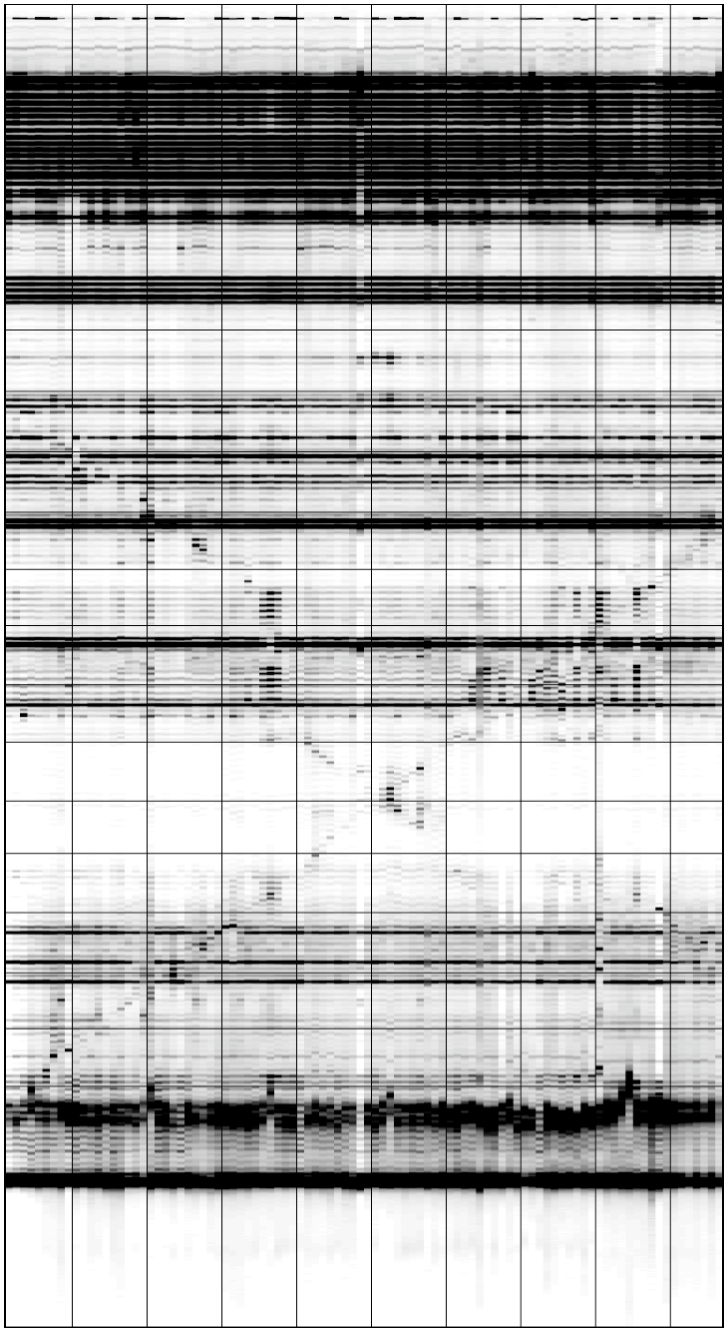
tRNA (phe), yeast

Choice 1. Data

Multiplexed •OH Cleavage Analysis



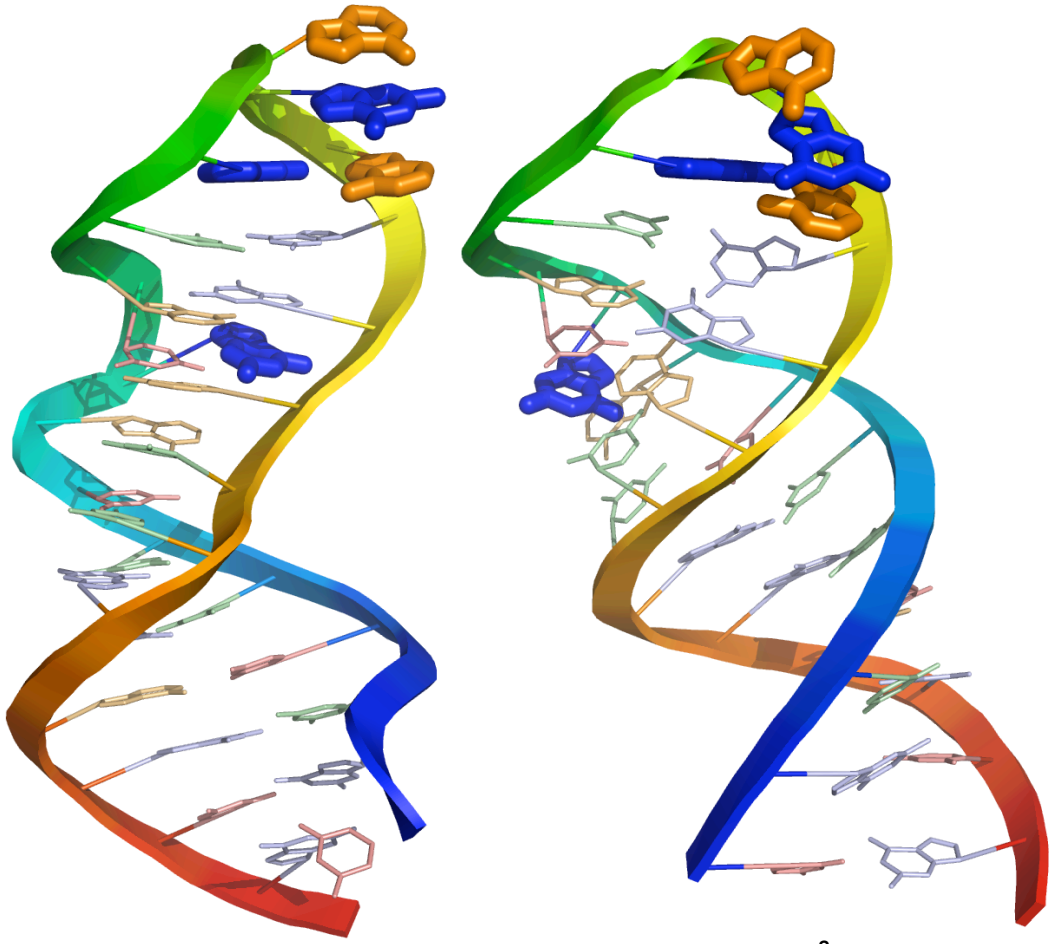
RNA tapestries



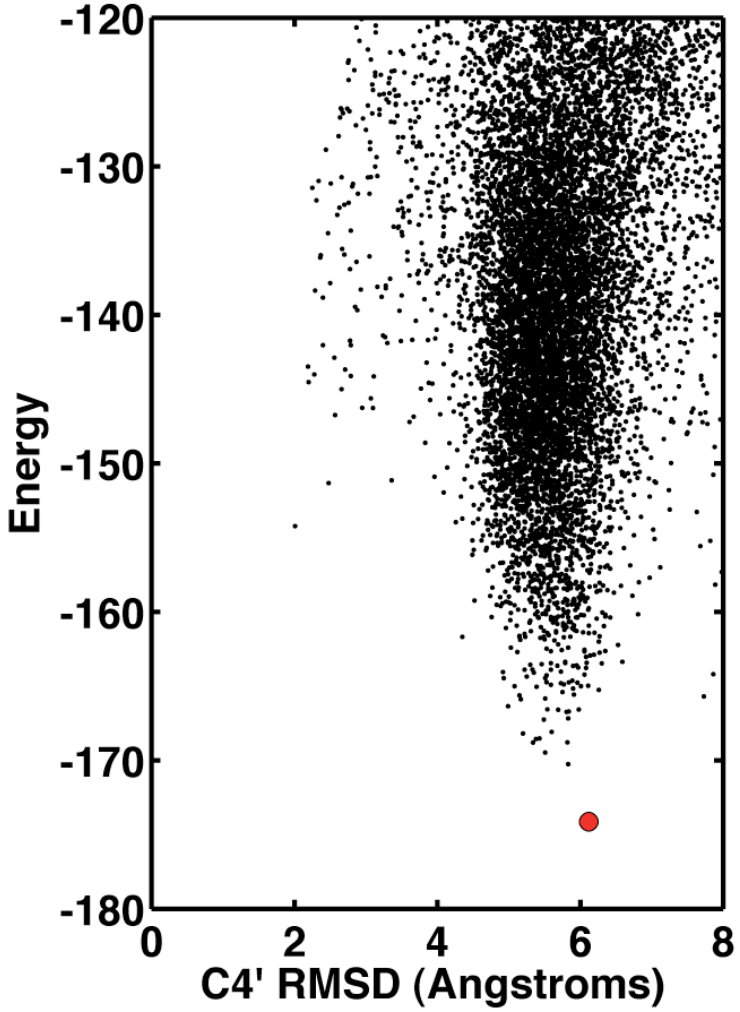
Success or failure? The sarcin/ricin loop

Native

FARNA Model

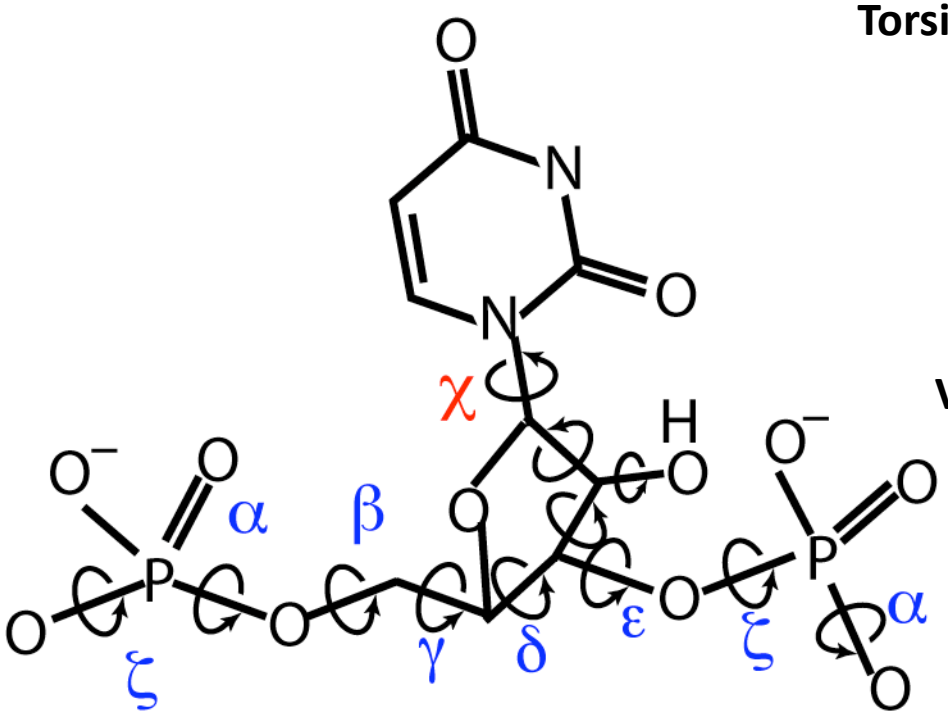


RMSD: 6.0 Å

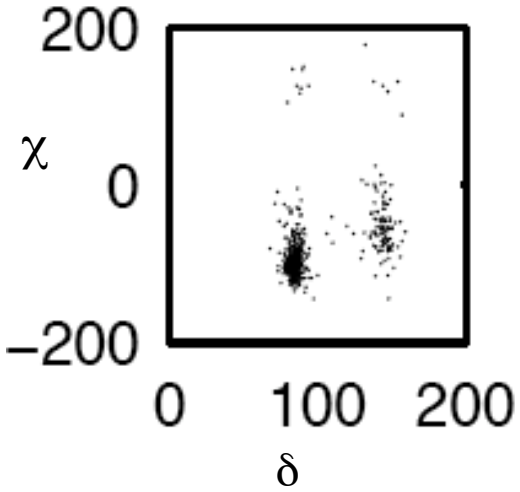


For a rigorous test, portions of the fragment library directly homologous to the sarcin/ricin loop have been excised.

All-atom refinement for RNA?

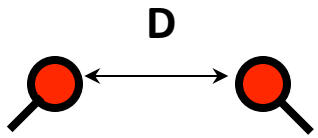


Torsional potential

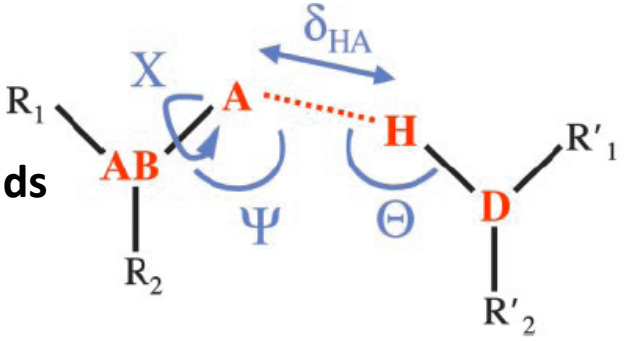


Van der waals packing

Desolvation



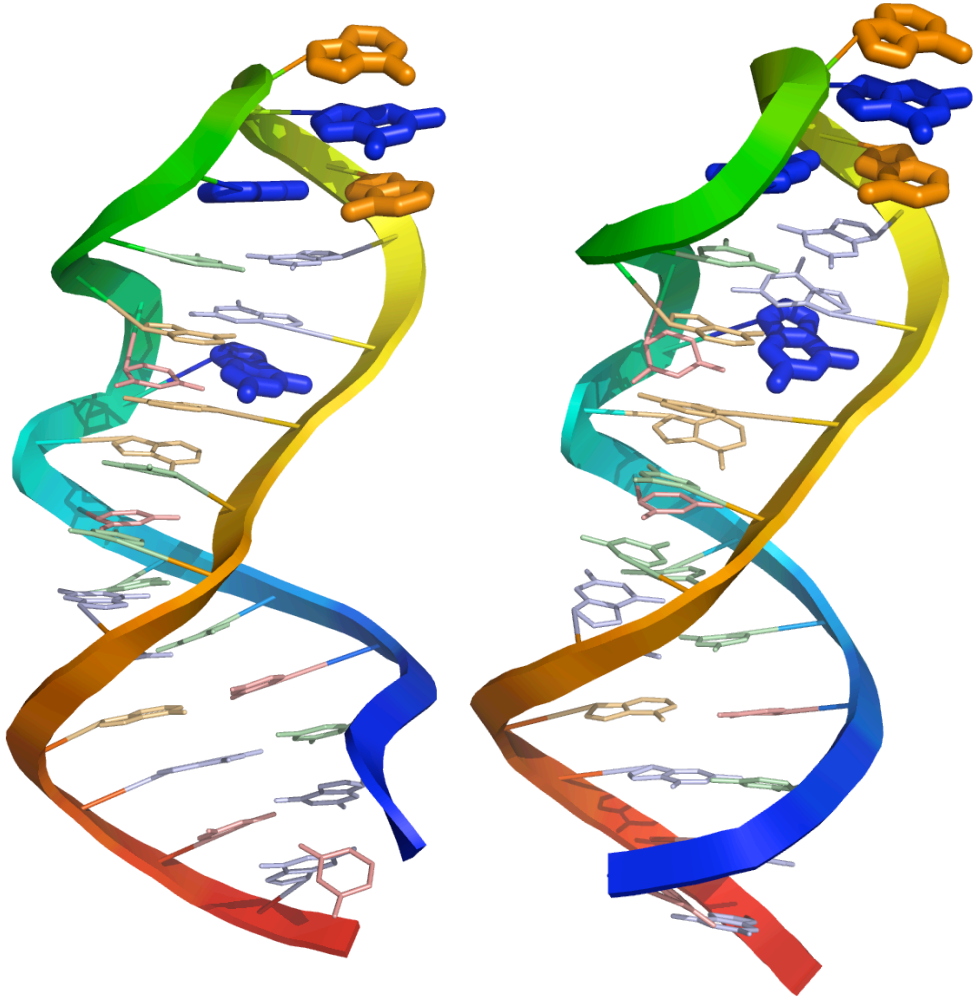
Hydrogen bonds



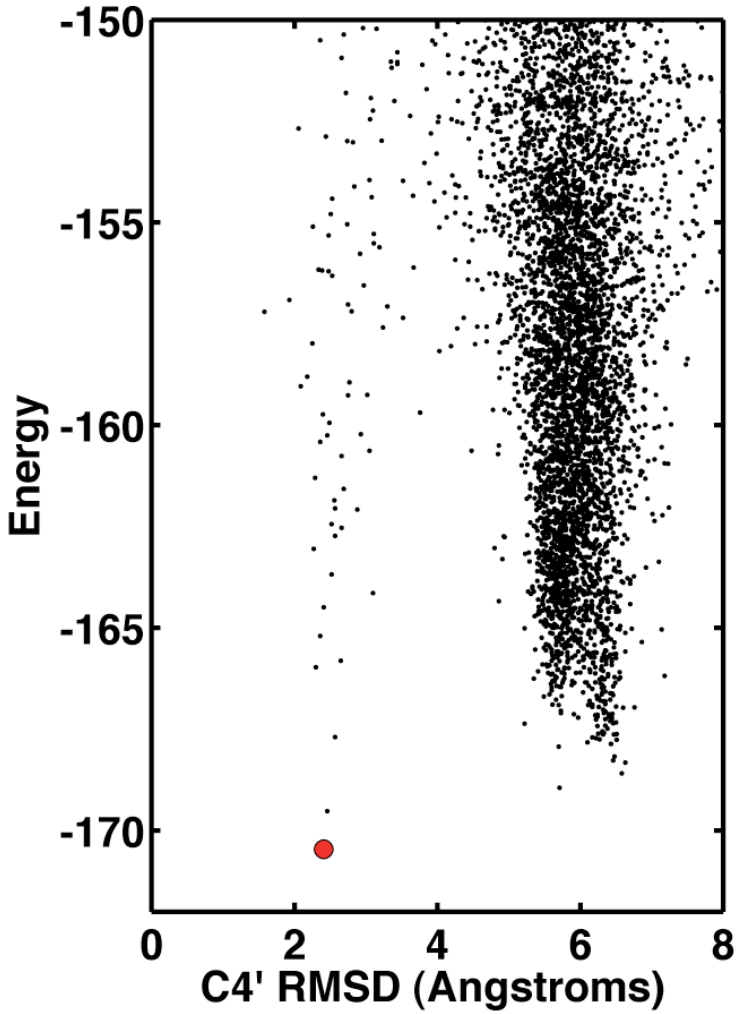
The power of all-atom refinement

Native

Model
[FARNA+ **all-atom refinement**]



RMSD 2.2 Å



For a rigorous test, portions of the fragment library directly homologous to the sarcin/ricin loop have been excised.

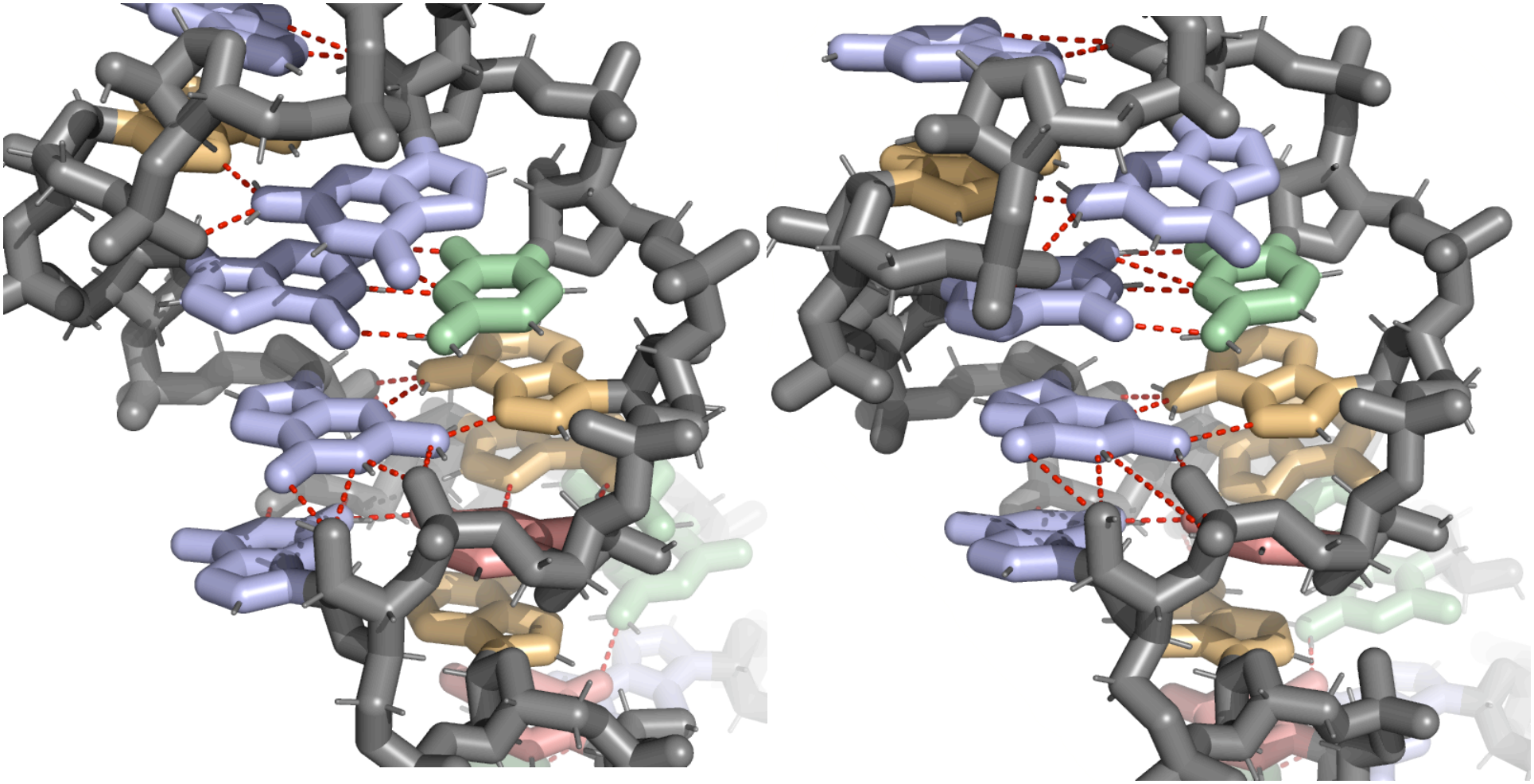
The power of all-atom refinement

Native

Model

Fragment assembly +

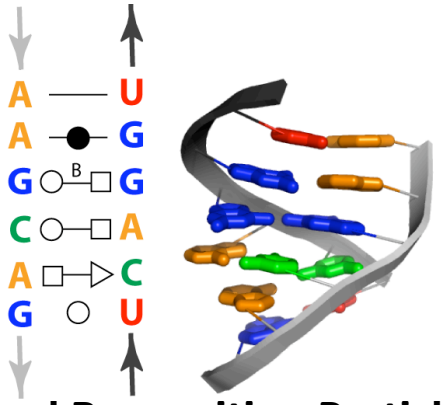
all-atom refinement



RMSD over non-helical region: 1.4 Å (C4´); 1.6 Å (all atoms)

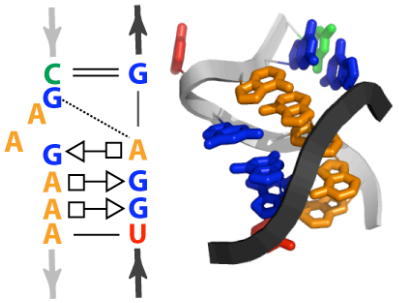
“Solving” the classic motifs

NATIVE



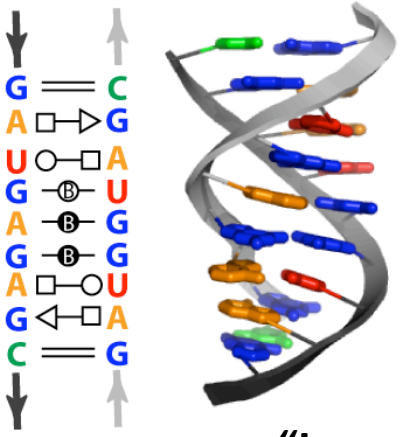
Signal Recognition Particle Domain

NATIVE



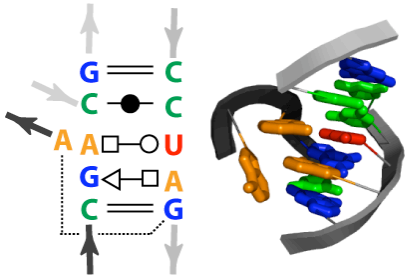
Kink/turn motif

NATIVE



“Loop E motif”

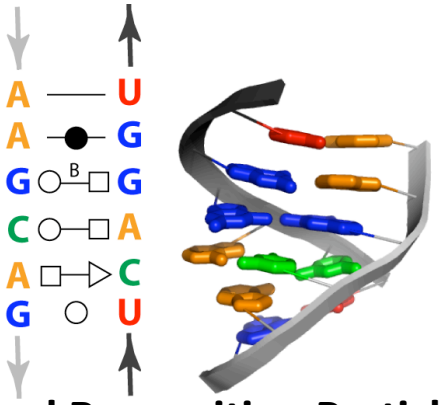
NATIVE



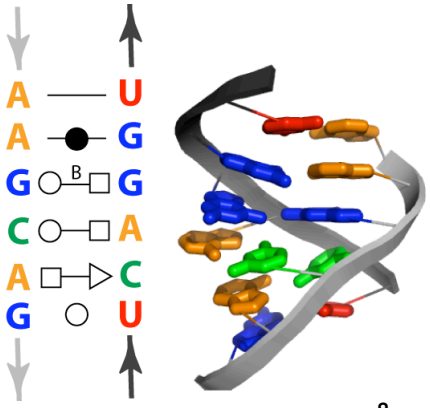
Hook/turn motif

“Solving” the classic motifs

NATIVE

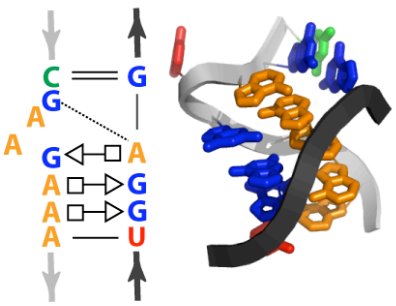


CLUSTER CENTER

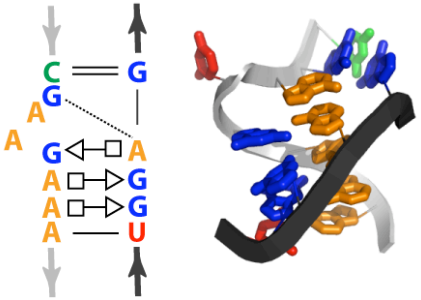


Signal Recognition Particle Domain RMSD 1.2 Å

NATIVE



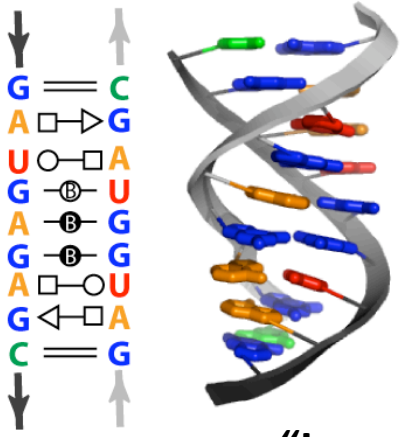
CLUSTER CENTER



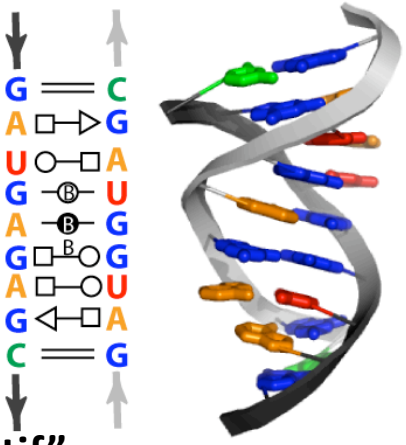
RMSD 1.43 Å

Kink/turn motif

NATIVE



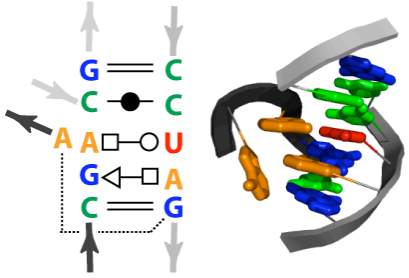
CLUSTER CENTER



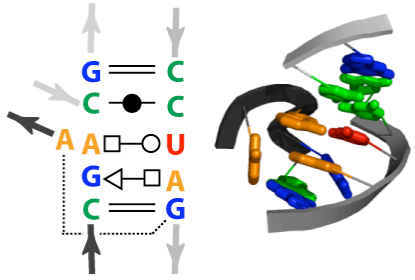
“Loop E motif”

RMSD 1.64 Å

NATIVE



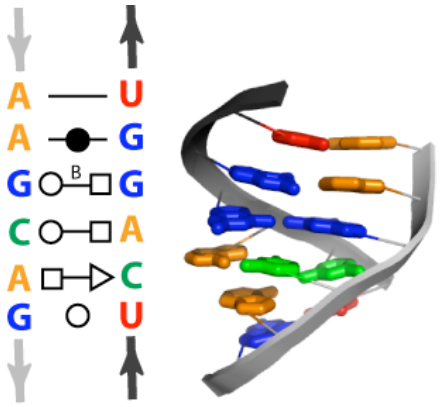
CLUSTER LOWEST ENERGY

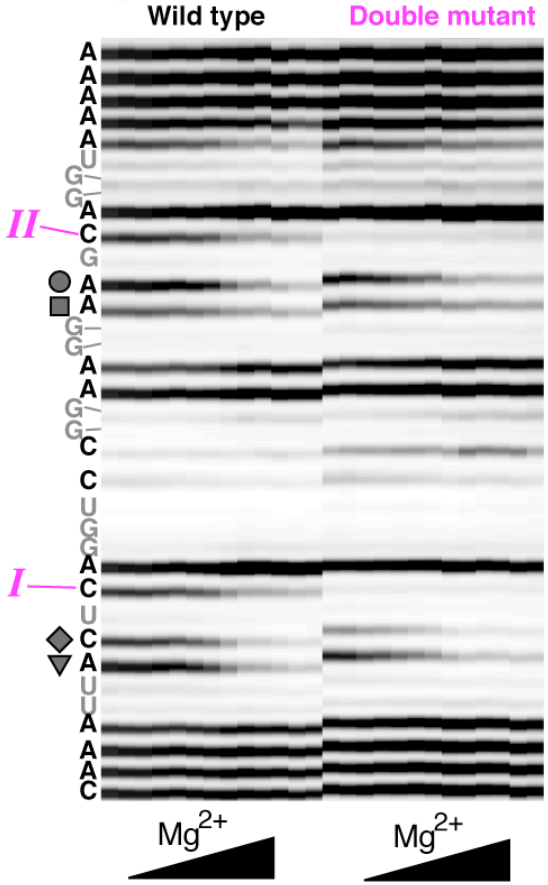


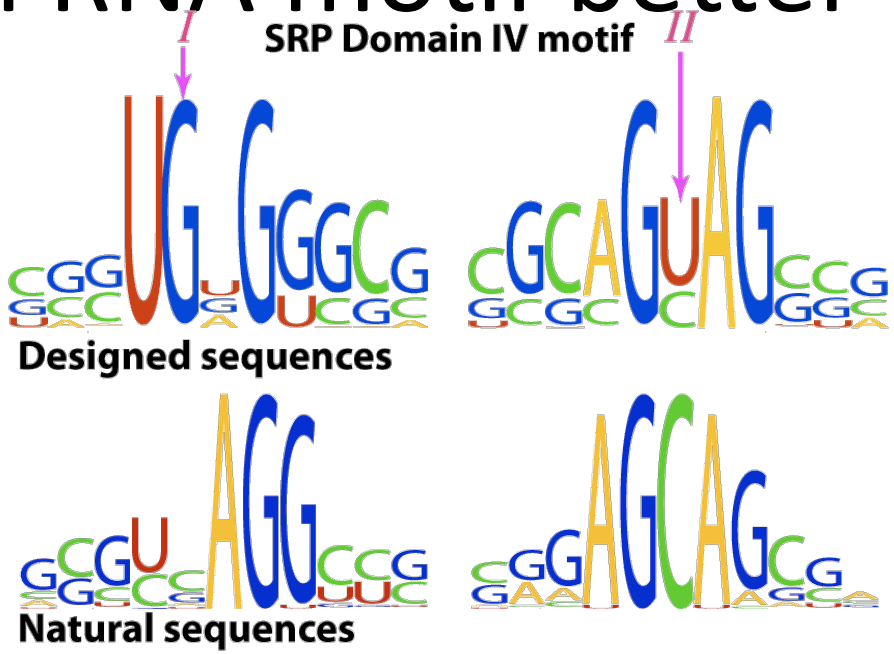
RMSD 2.06 Å

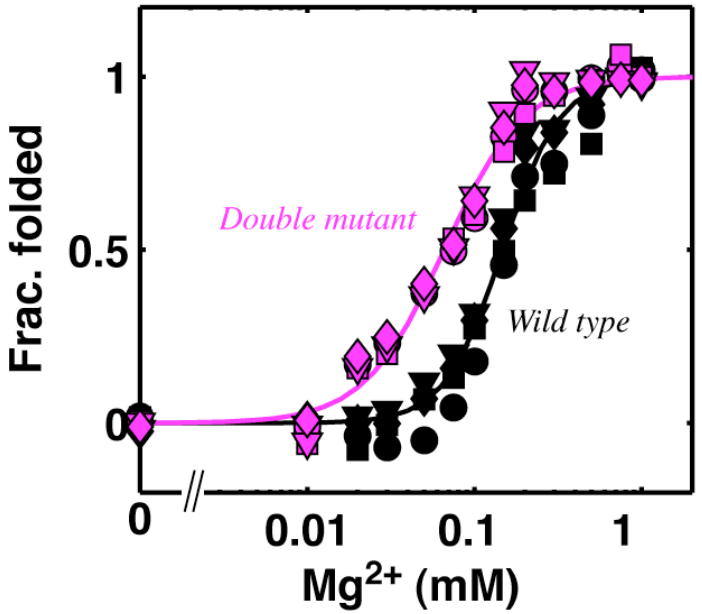
Hook/turn motif

Making an RNA motif better



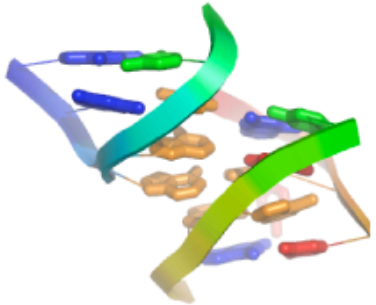


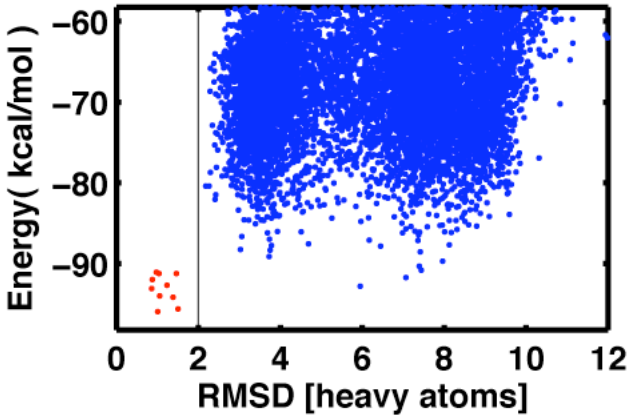




De novo modeling

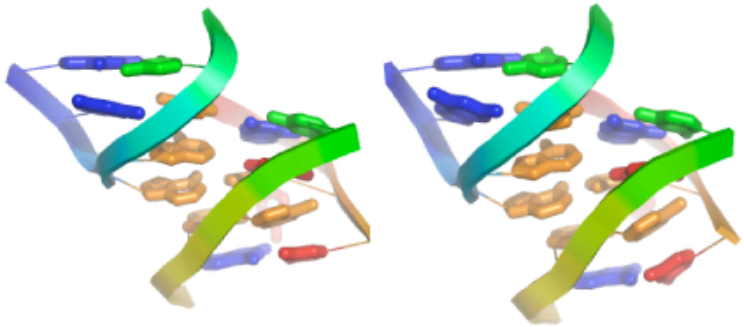
The biggest bottleneck: conformational sampling



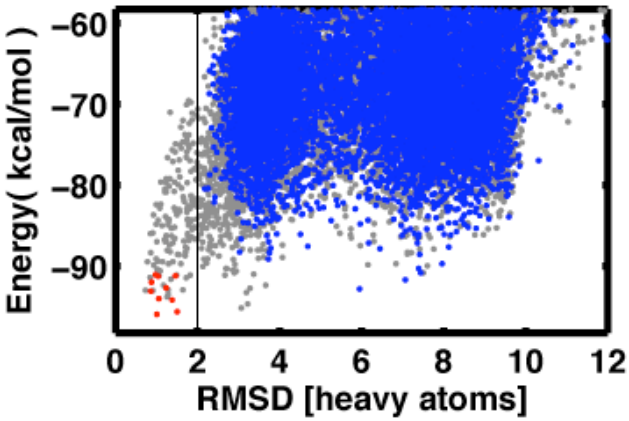


De novo modeling

The biggest bottleneck: conformational sampling



1.0 Å rmsd



A universal obsession

domains (<85 residues). The primary bottleneck to consistent high-resolution prediction appears to be conformational sampling.

Toward High-Resolution de Novo Structure Prediction for Small Proteins

Philip Bradley, Kira M. S. Misura, David Baker*

generated during CASP8. Developing more effective conformational sampling algorithms and protocols is a critical area for current research in protein structure prediction.

Structure prediction for CASP8 with all-atom refinement using Rosetta

Srivatsan Raman,[1] Robert Vernon,[1] James Thompson,[2] Michael Tyka,[1] Ruslan Sadreyev,[3] Jimin Pei,[3] David Kim,[1] Elizabeth Kellogg,[1] Frank DiMaio,[1] Oliver Lange,[1] Lisa Kinch,[3] Will Sheffler,[2] Bong-Hyun Kim,[4] Rhiju Das,[1] Nick V. Grishin,[3,4] and David Baker[1,2,3]*

to the functional loop regions. However, despite progress in loop prediction methods[1,2], design applications are limited by the difficulty in modeling purely local conformational moves and by the need for advances in sampling and evaluating loop conformations.

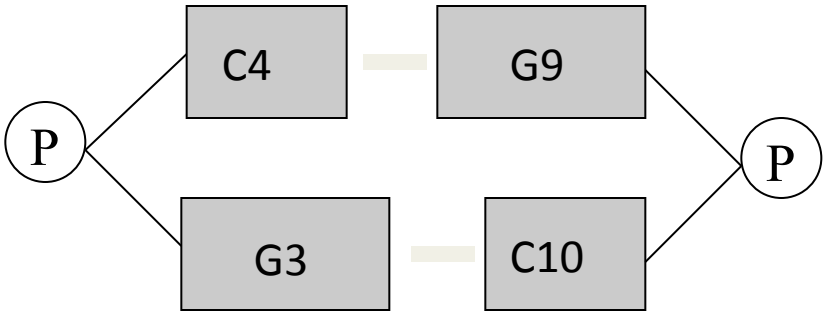
Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling

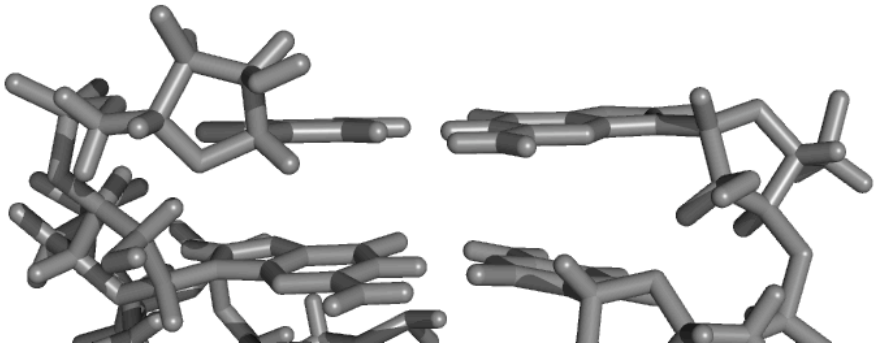
Daniel J Mandell[1,2], Evangelos A Coutsias[3] & Tanja Kortemme[1,2,4]

Step-by-step sampling

C A

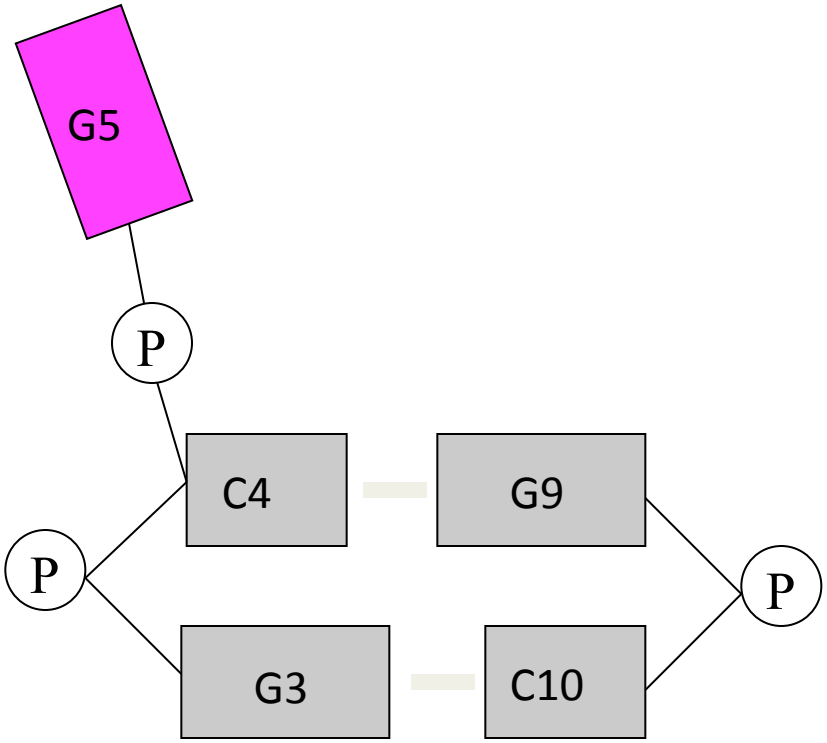
G A

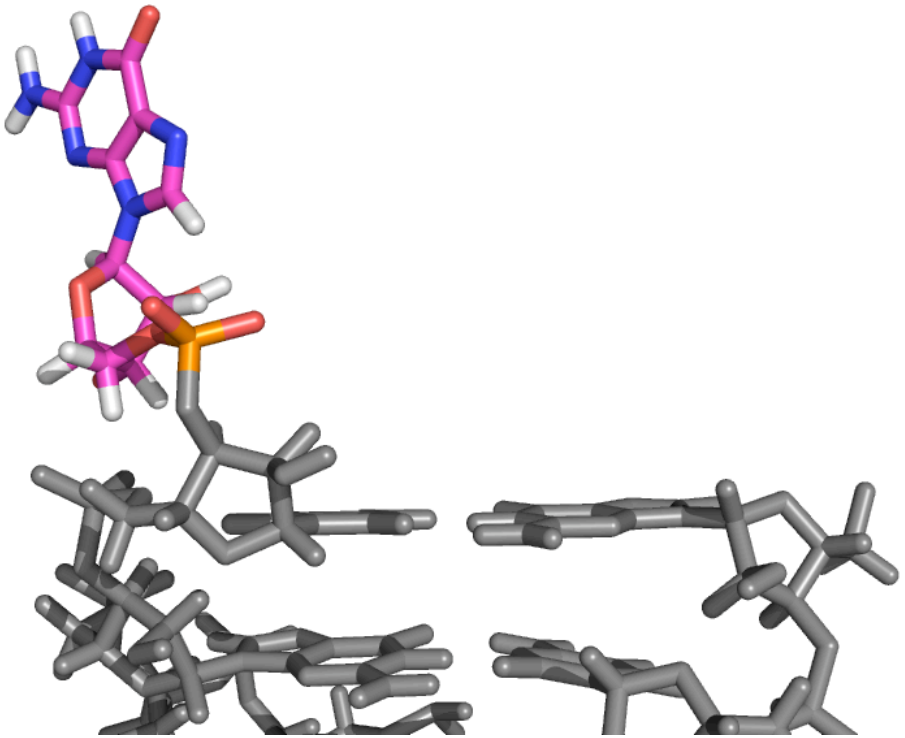




Parin Sripakdeevong

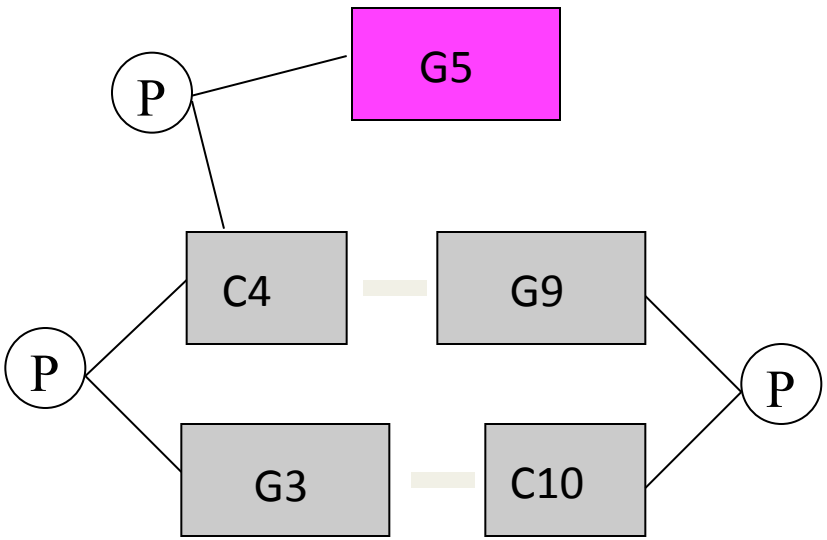
Step-by-step sampling

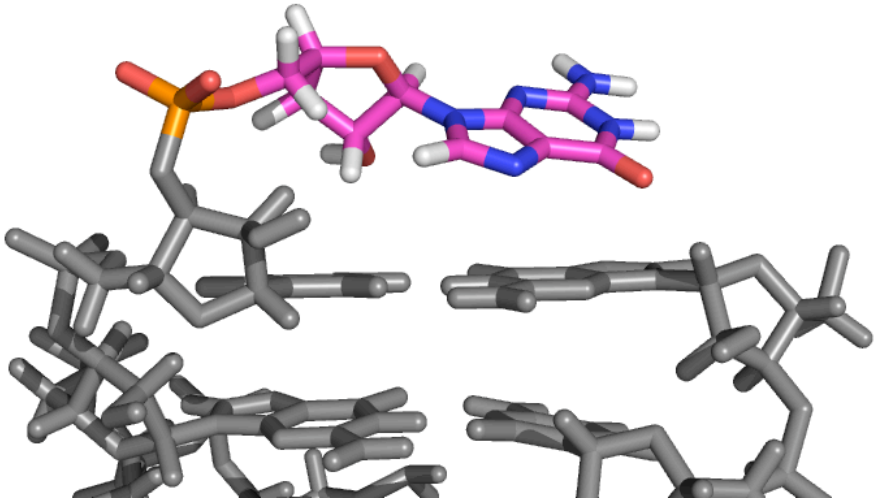




Parin Sripakdeevong

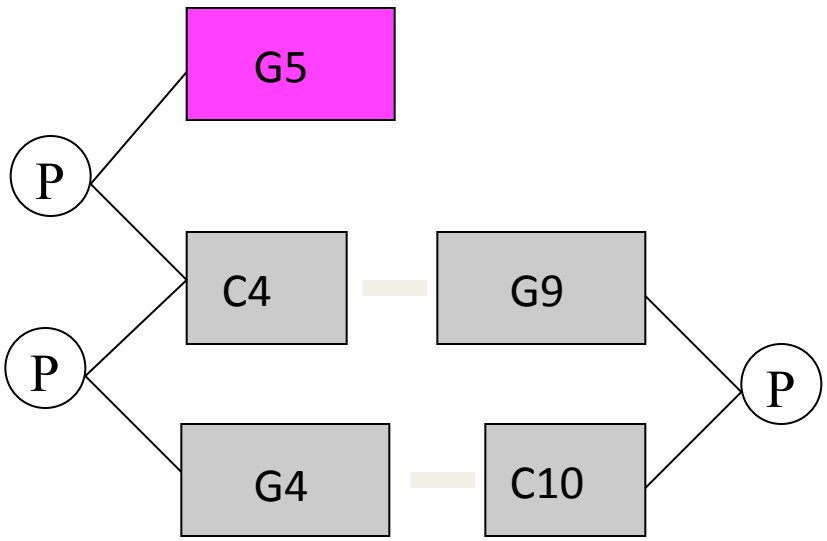
Step-by-step sampling

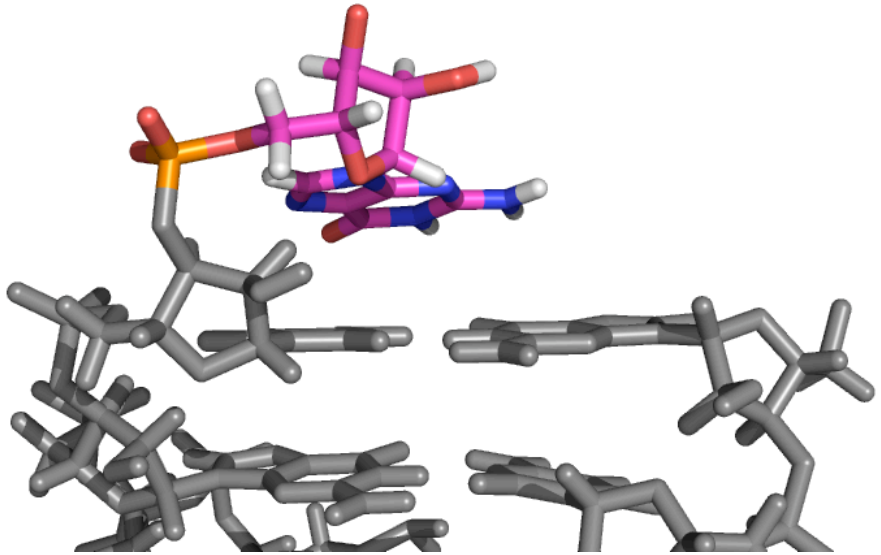




Parin Sripakdeevong

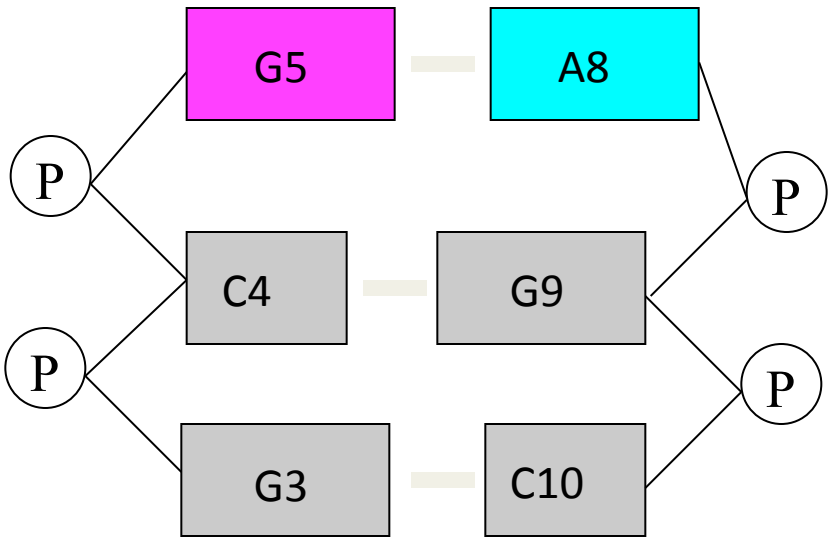
Step-by-step sampling

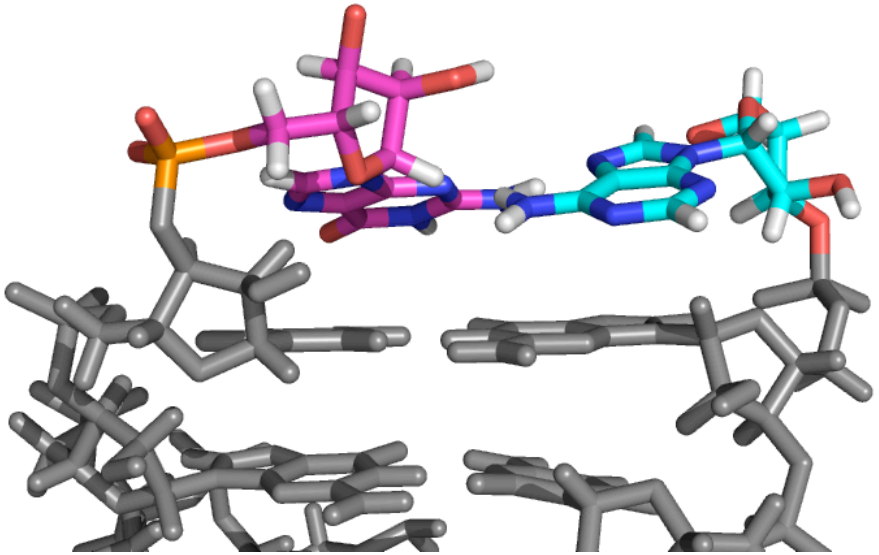




Parin Sripakdeevong

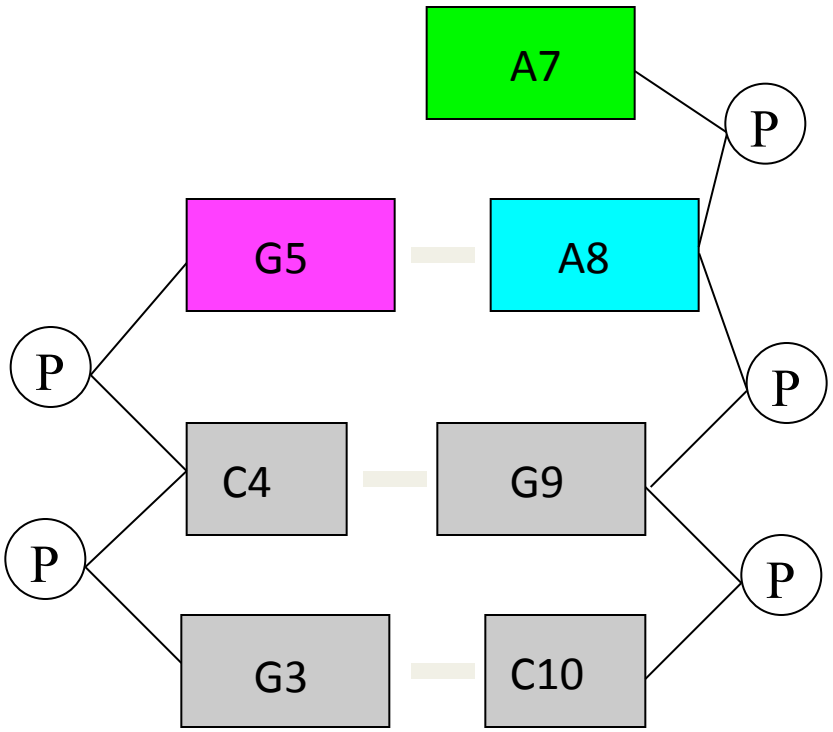
Step-by-step sampling

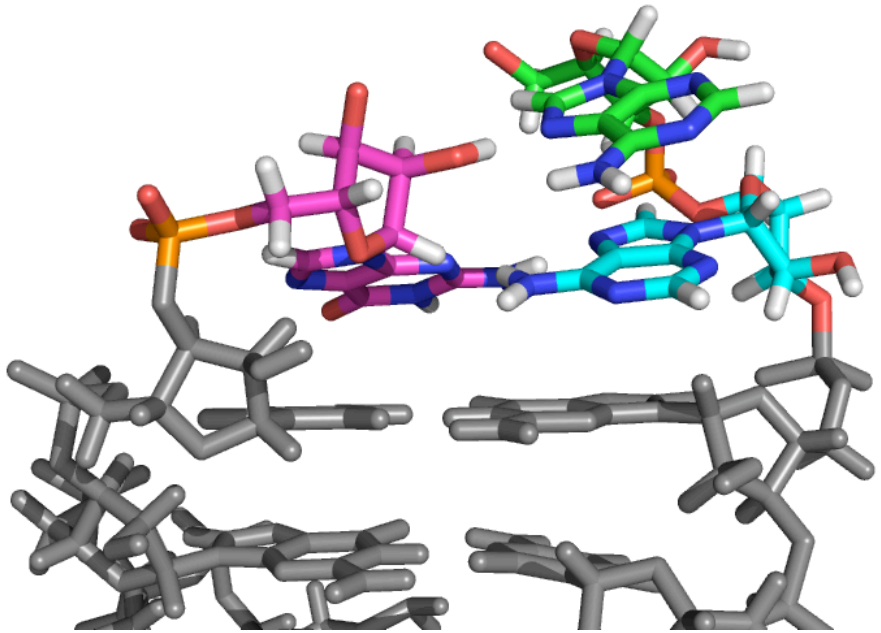




Parin Sripakdeevong

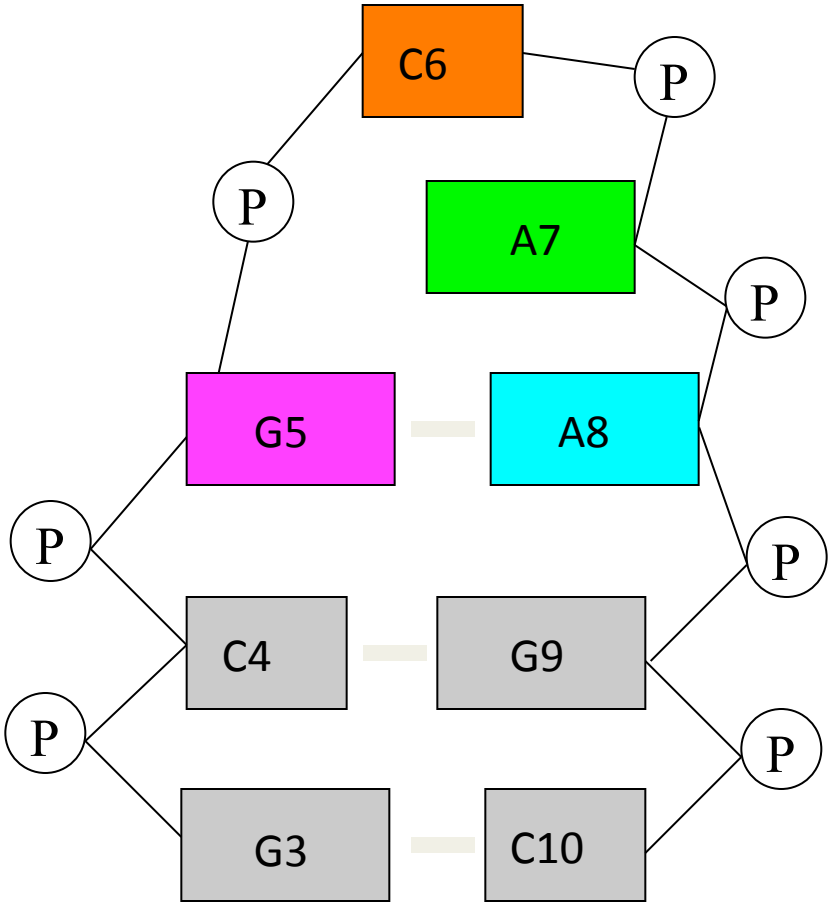
Step-by-step sampling

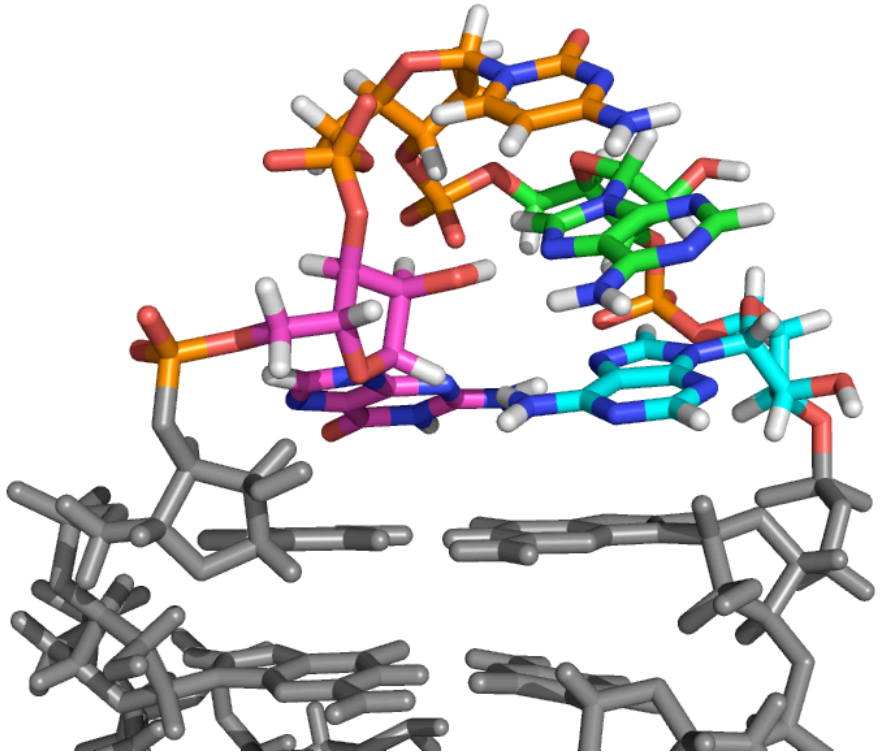




Parin Sripakdeevong

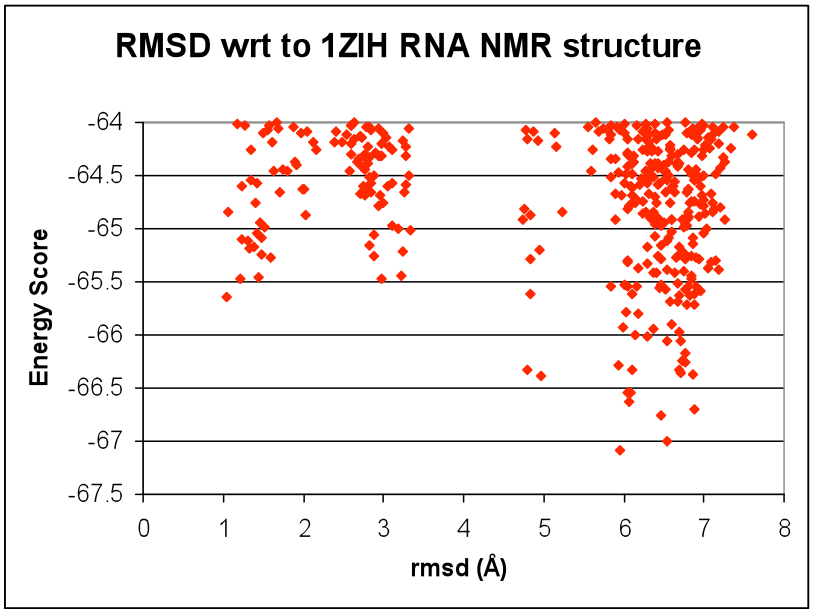
Step-by-step sampling





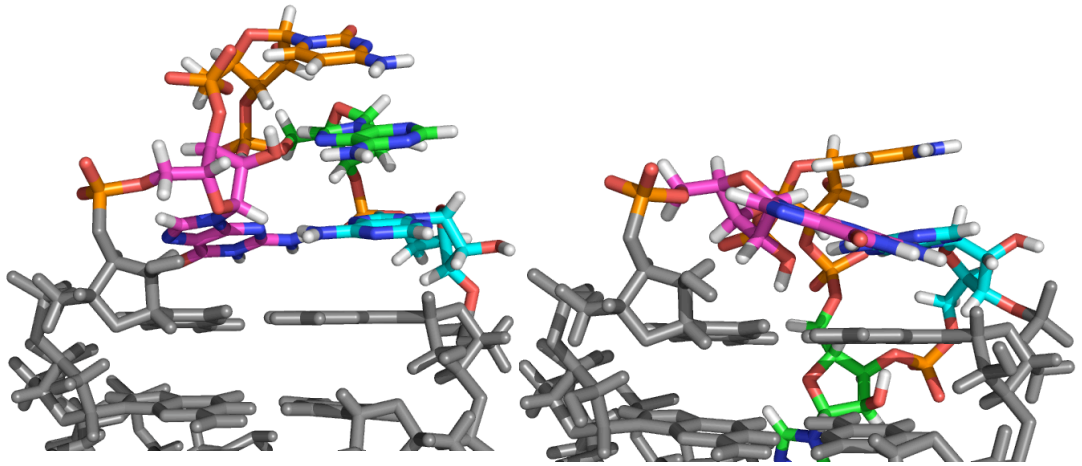
Parin Sripakdeevong

Step-by-step sampling



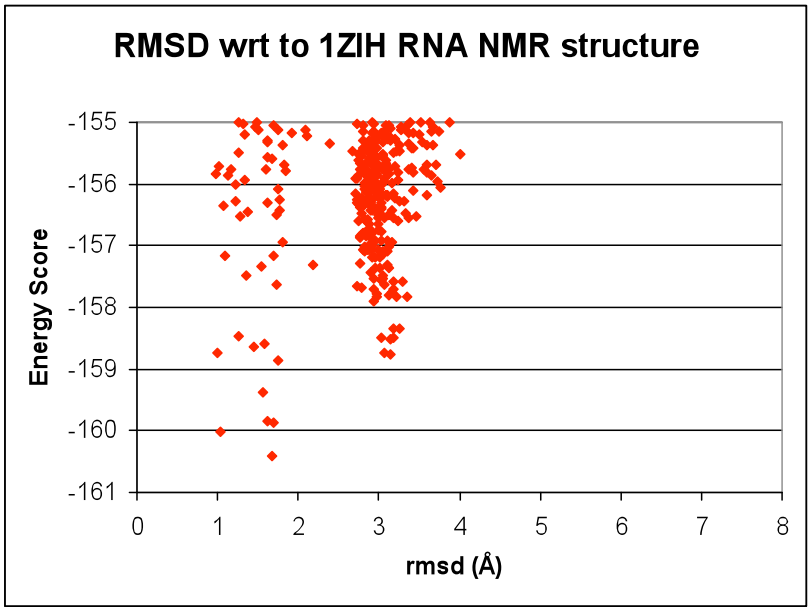
1ZIH NMR

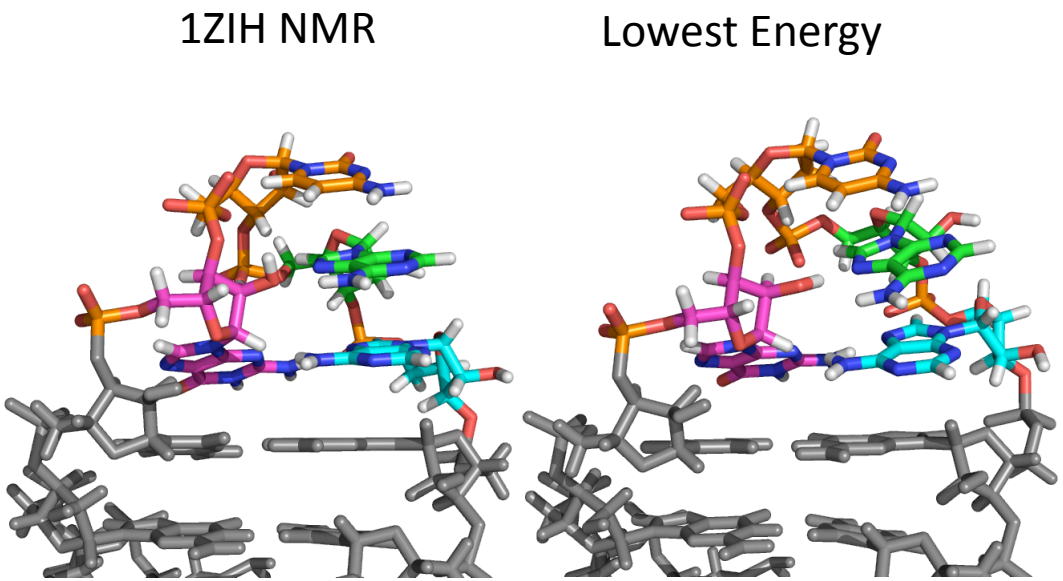
Lowest Energy



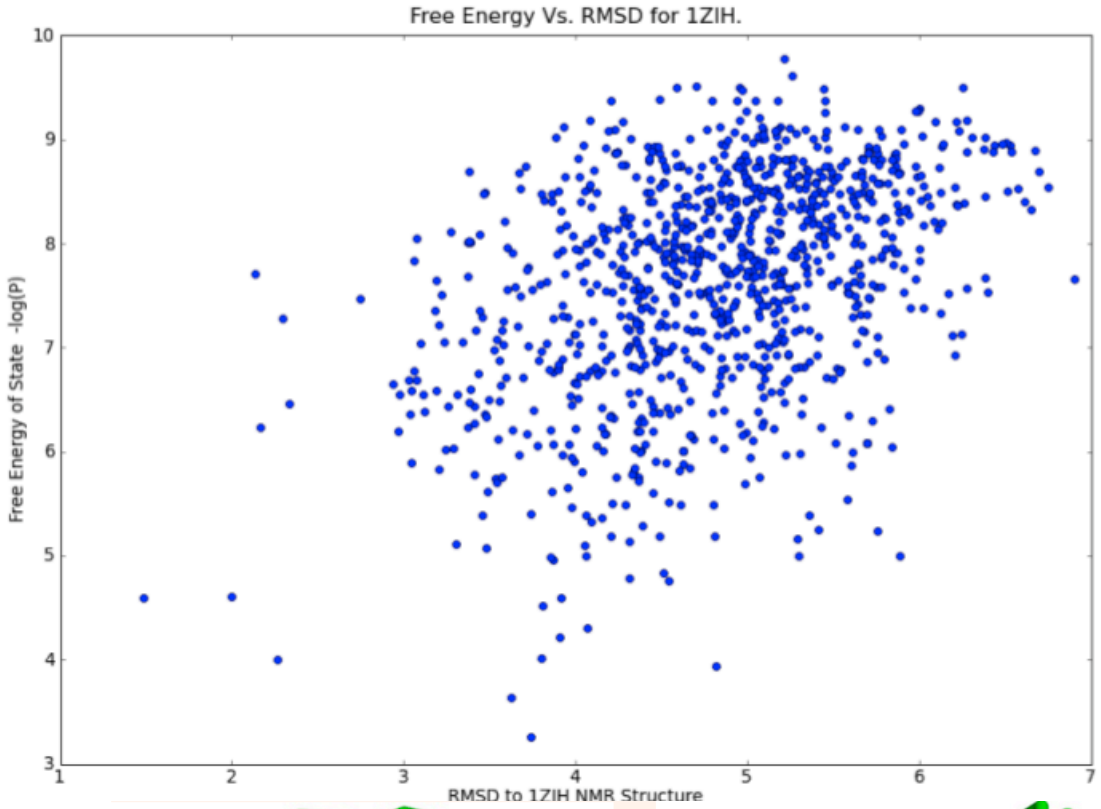
Parin Sripakdeevong

Step-by-step sampling



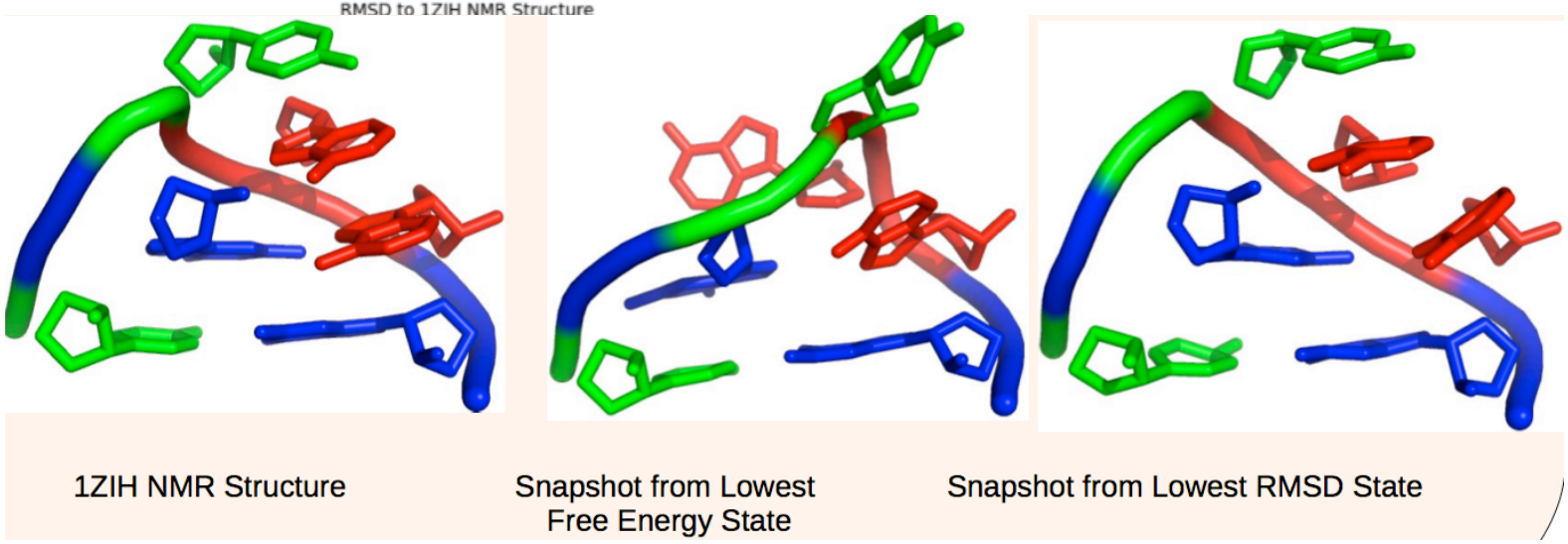


Parin Sripakdeevong



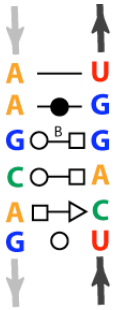
Super-fast MD

- Get free energy
- GPU's

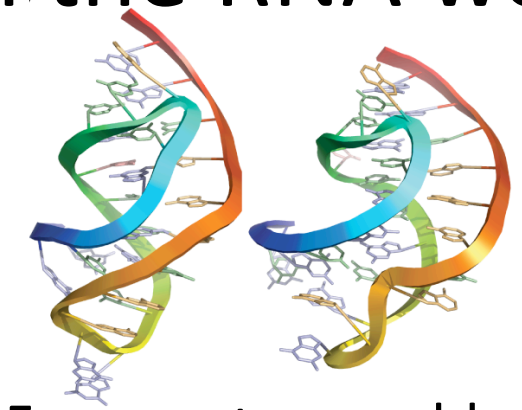


Kyle Beauchamp – molecular dynamics prodigy

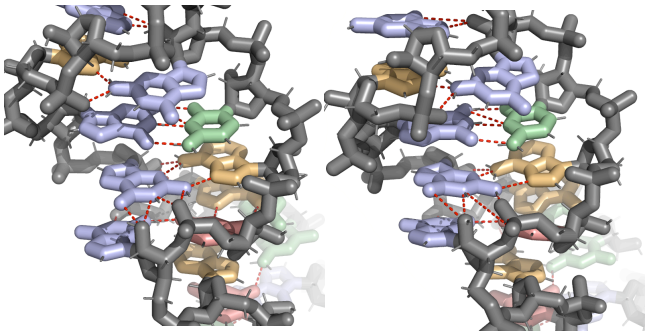
A Rosetta tour through the RNA world



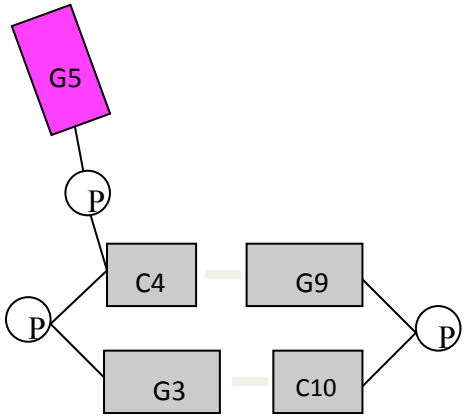
Why RNA prediction should be easy



Fragment assembly and all that



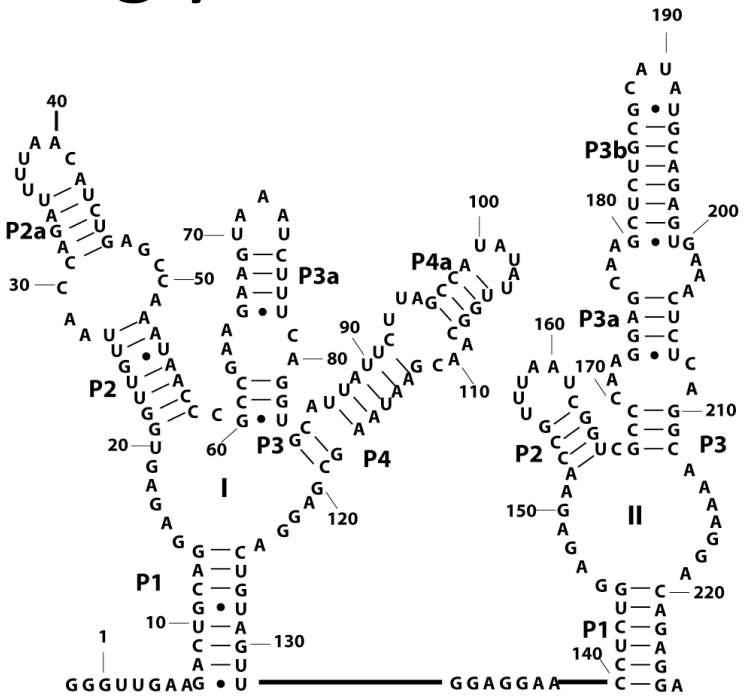
Achieving near-atomic accuracy



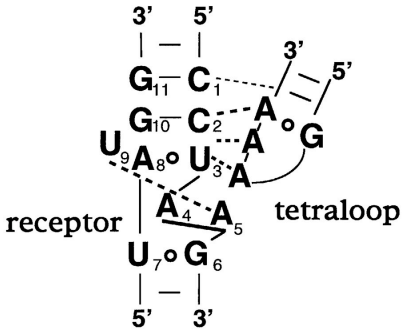
Beating the “astronomical” conformational sampling problem

Next up: **blind predictions**

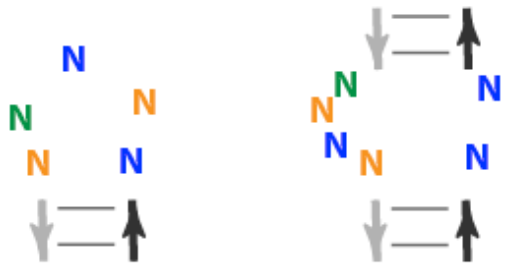
Modeling a big RNA: the glycine riboswitch



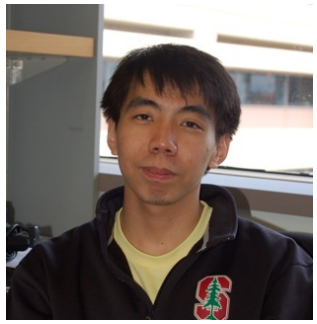
High resolution structure prediction



Predicting energy + structure for *every* loop



We thank:



- **Stanford Biochemistry Department** for adopting us
- John Karanicolas, David Baker (RNA hi-res)
- Vijay Pande and group (MD expertise)
- Adrien Treuille, Jee Lee + colleagues
- Andrew Leaver-Fay, Sergey Lyskov, **Rosetta community**

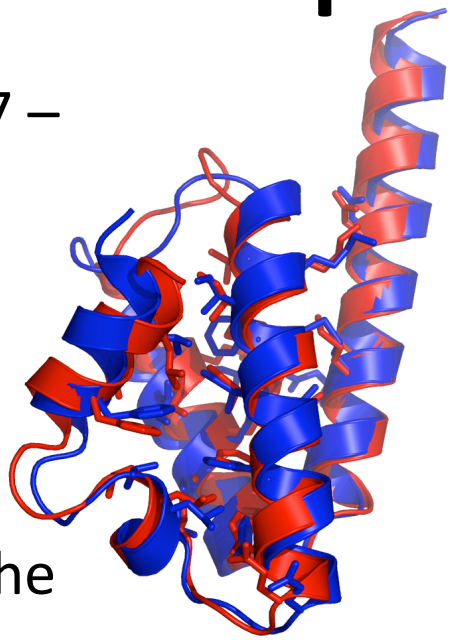
Jane Coffin Childs Foundation

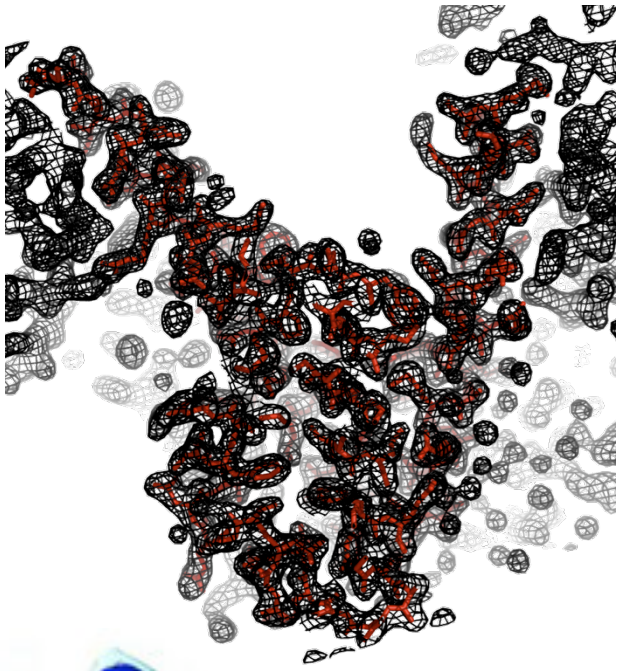
NATIONAL INSTITUTES OF HEALTH

BURROUGHS WELLCOME FUND

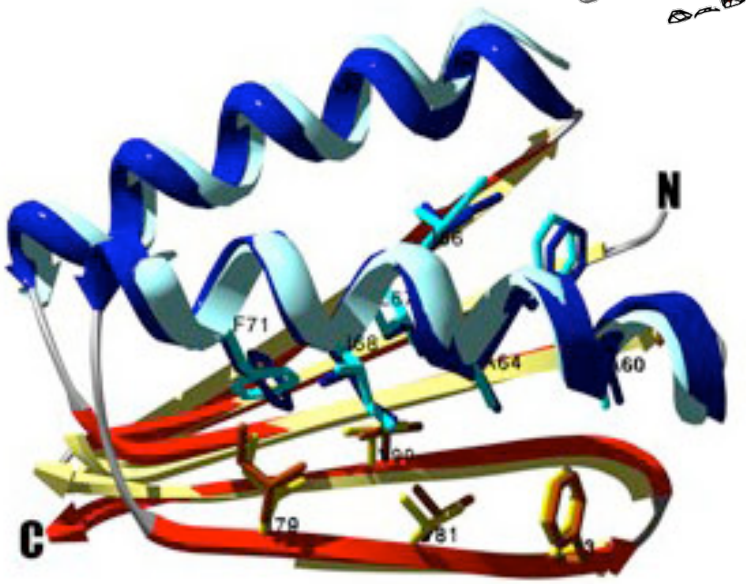
Inspiration from the **protein** world

- **Blind predictions** in CASP7 – some with near-atomic resolution agreement with **crystal structures**.





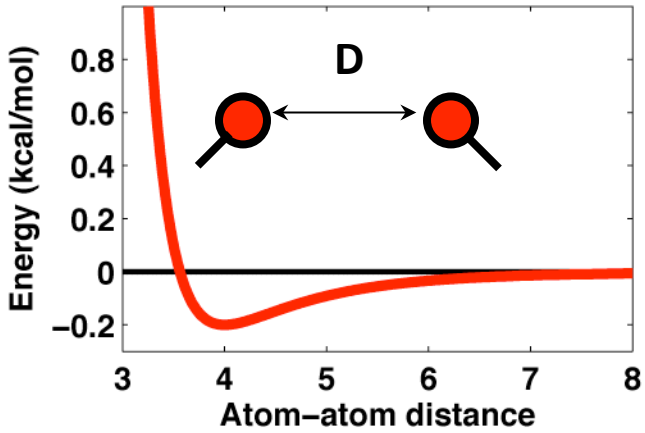
- **Practical connections**: engineering new proteins, the crystallographic phase problem, improving NMR models, more…



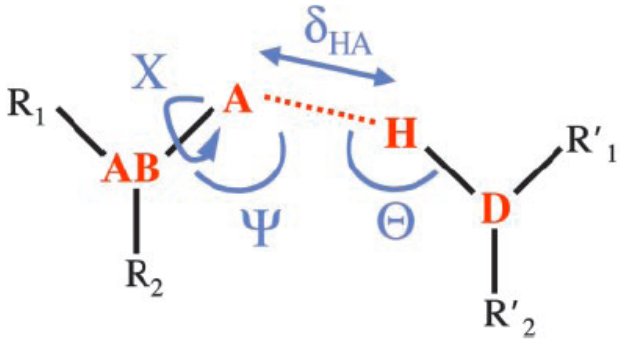
<u>The key:</u> Putting all the atoms in.

Ingredients of a high resolution potential

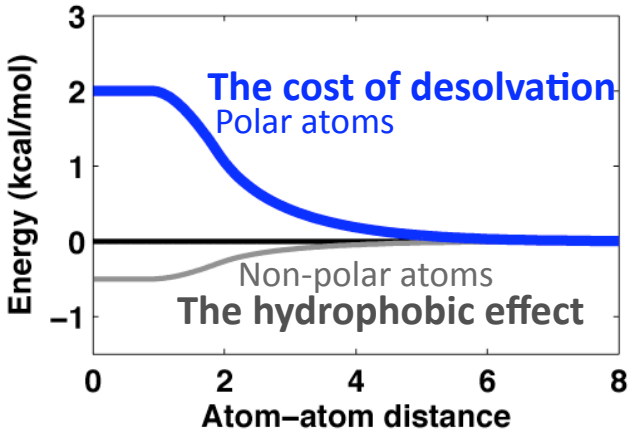
1. Van der waals packing



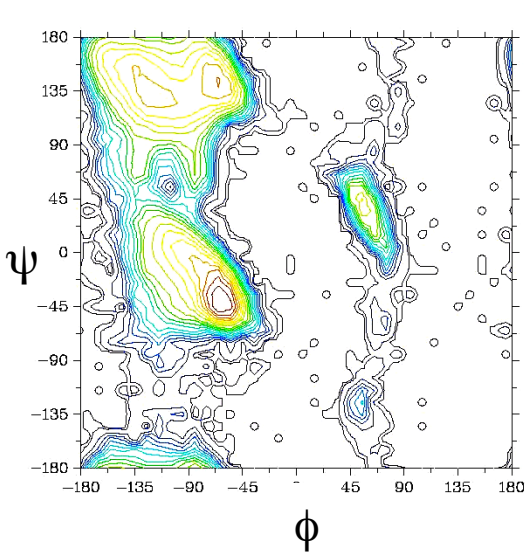
2. Hydrogen bonds

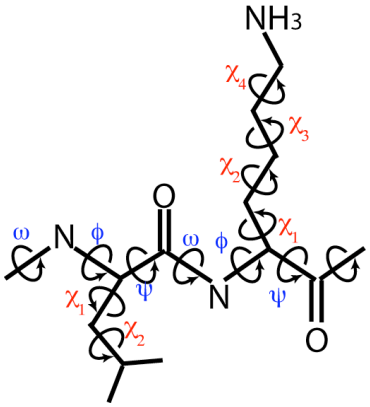


3. Manifestations of water



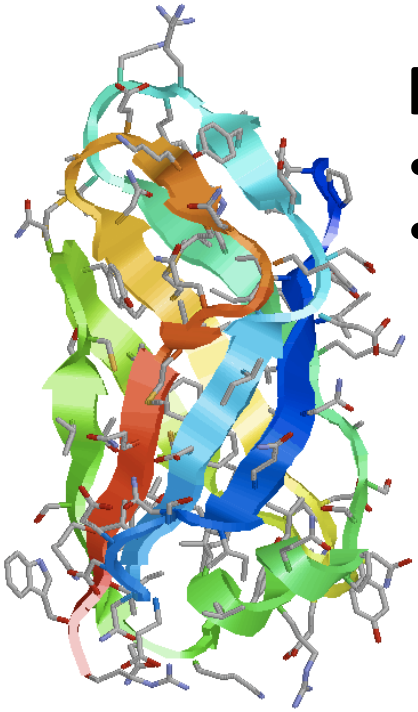
4. Torsional potential





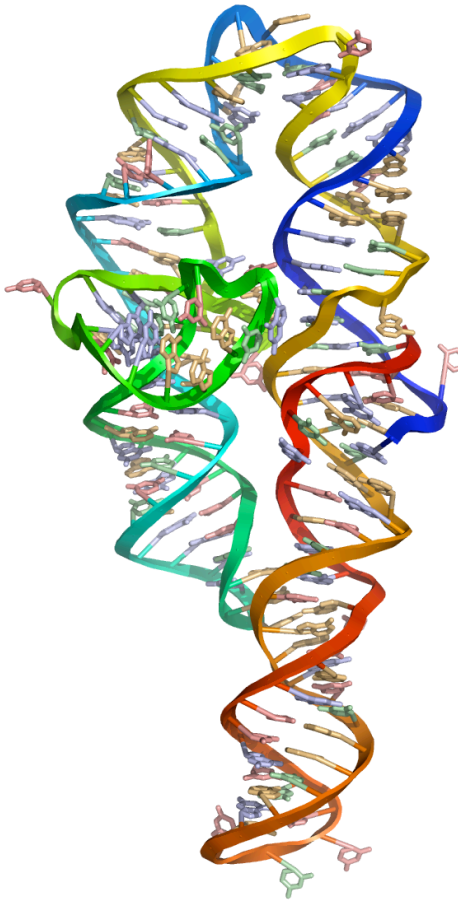
How can we visualize RNA?

Crystallography and NMR – powerful *when they work*



Proteins

- 50,000 structures
- 5000 “families”



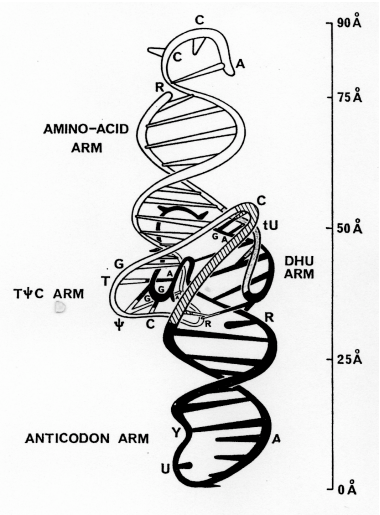
RNA

- ~1,000 structures
- 50 different ribozymes, riboswitches, etc.

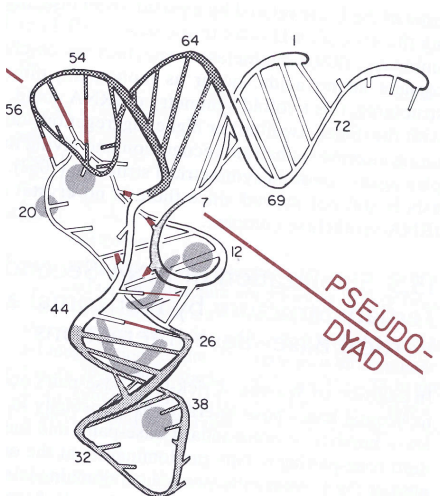
How can we visualize RNA?

Manual modeling, with clues from evolution

Michael Levitt's 1969 model of tRNA



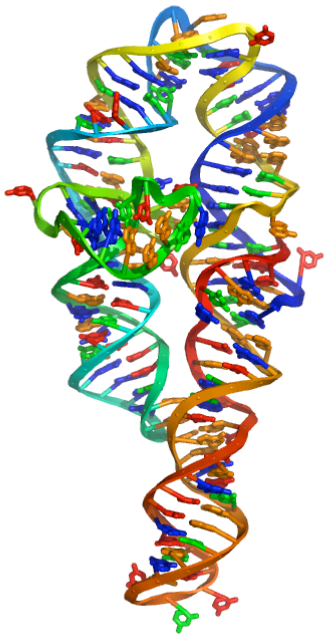
Model



Subsequently released crystal srtucture

Westhof & colleagues

Model



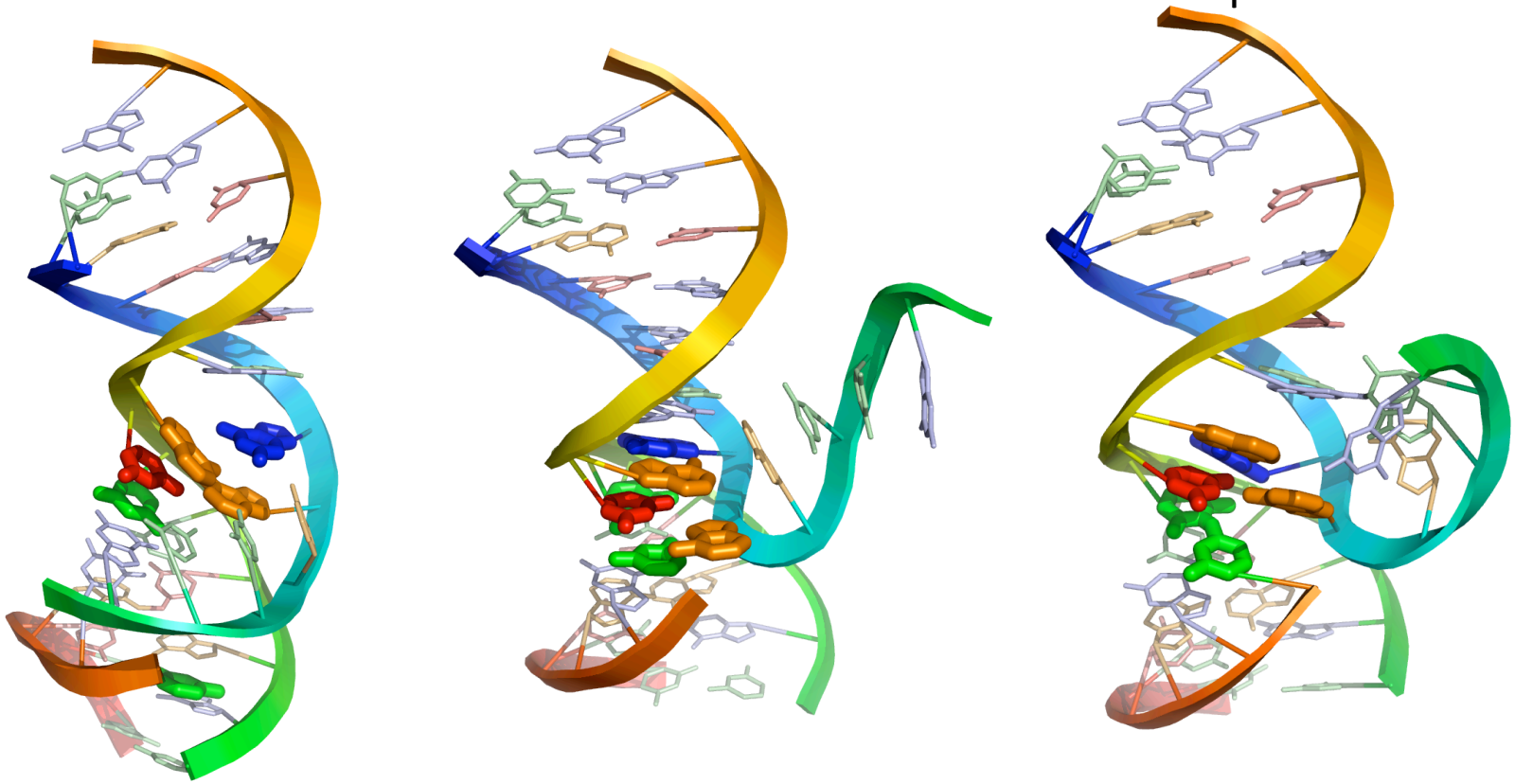
Subsequently released crystal srtucture

Hitting the limits of low resolution modeling

CLUSTER CENTER

NATIVE

A low RMSD conformation is sampled…



RMSD 10.5 Å

RMSD 3.8 Å

“Hook-turn” motif

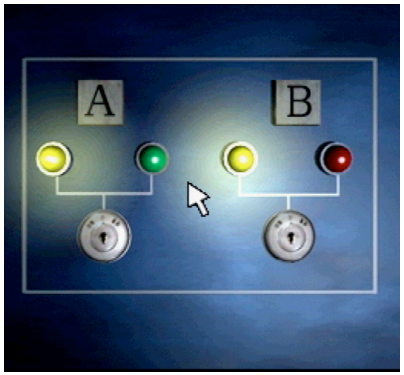
Can we do better?

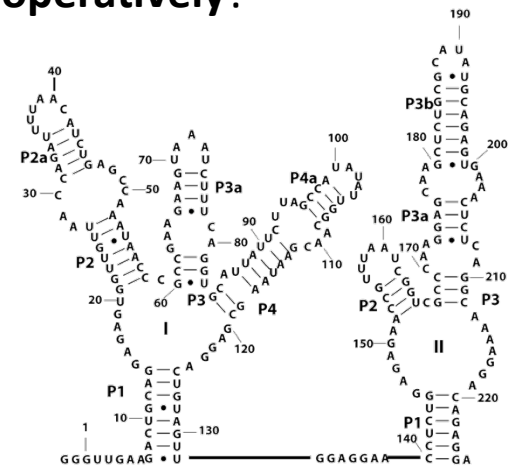
Truly *de novo*?

High res… 1–2 Å accuracy?

Ground-level physical questions in RNA

How do real riboswitches sense **multiple** ligands **cooperatively**?





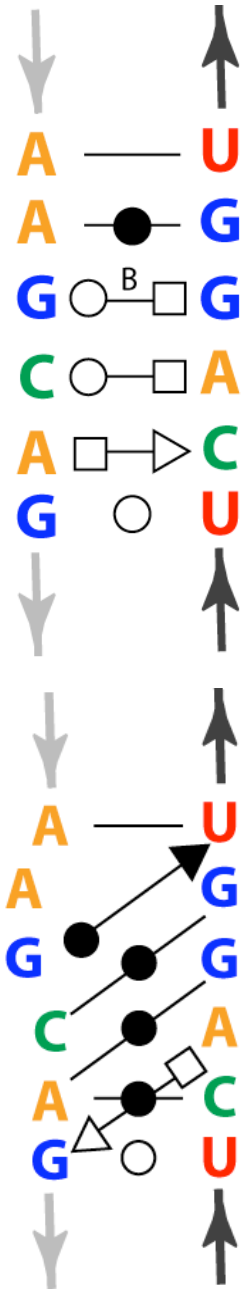
Why do RNA's misfold so frequently?

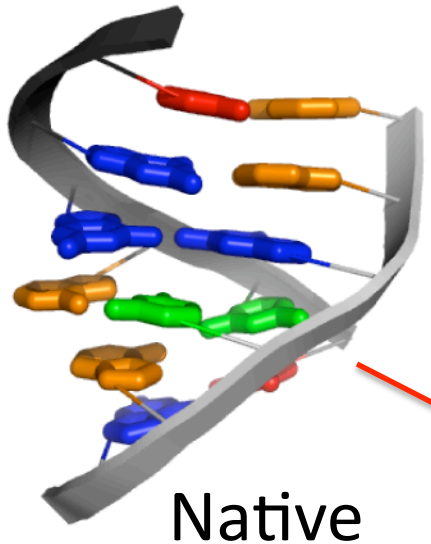
Do random RNA sequences have structure?

AAACGUUGCACUU AGCGAAUCAGUAA GGCAGUCG...

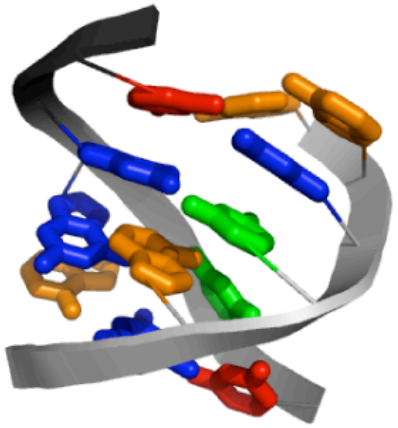
What is the full range of possibilities for RNA structure?

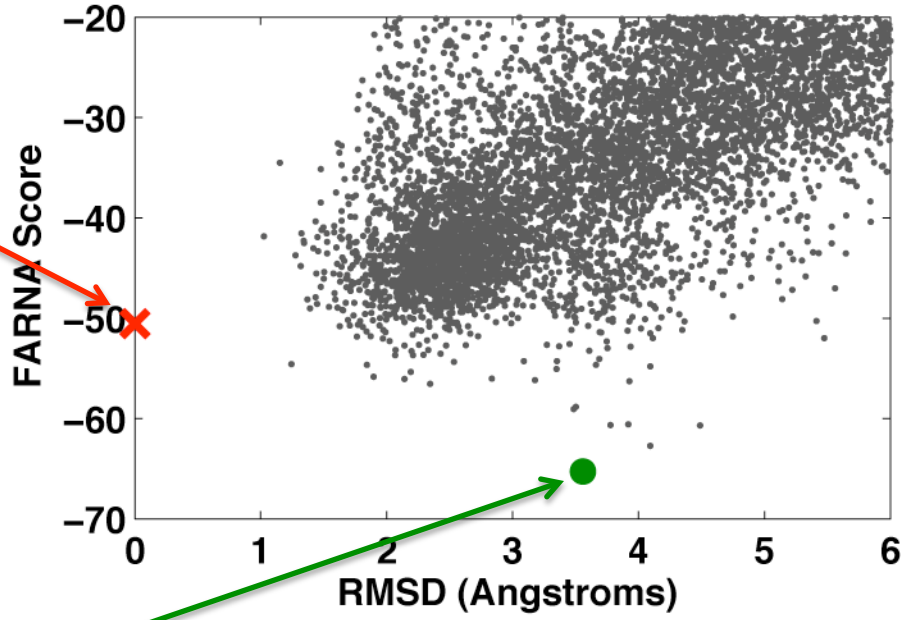
Choice 2. Hi res modeling.





Native

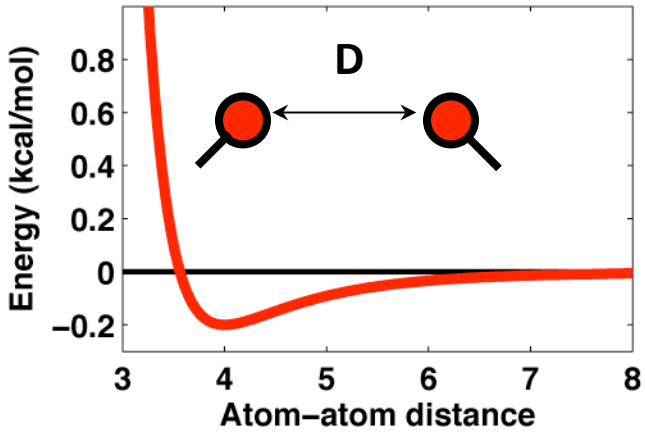




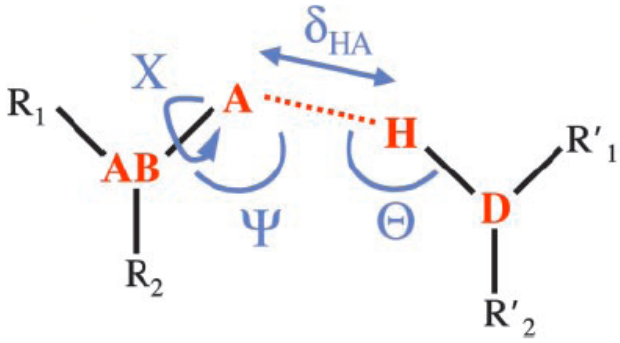
Best scoring by
fragment assembly score

Ingredients of a high resolution potential

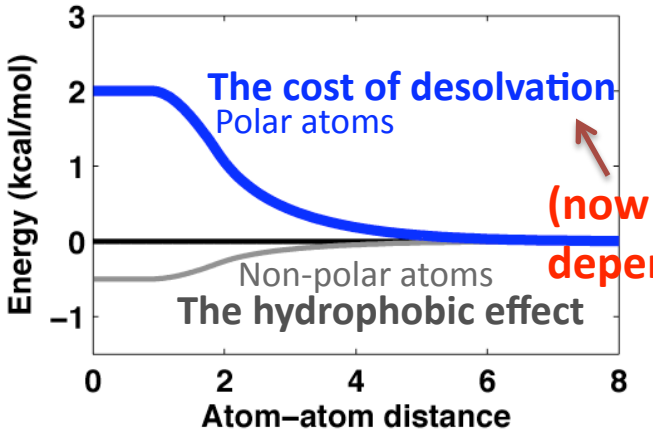
1. Van der waals packing



2. Hydrogen bonds

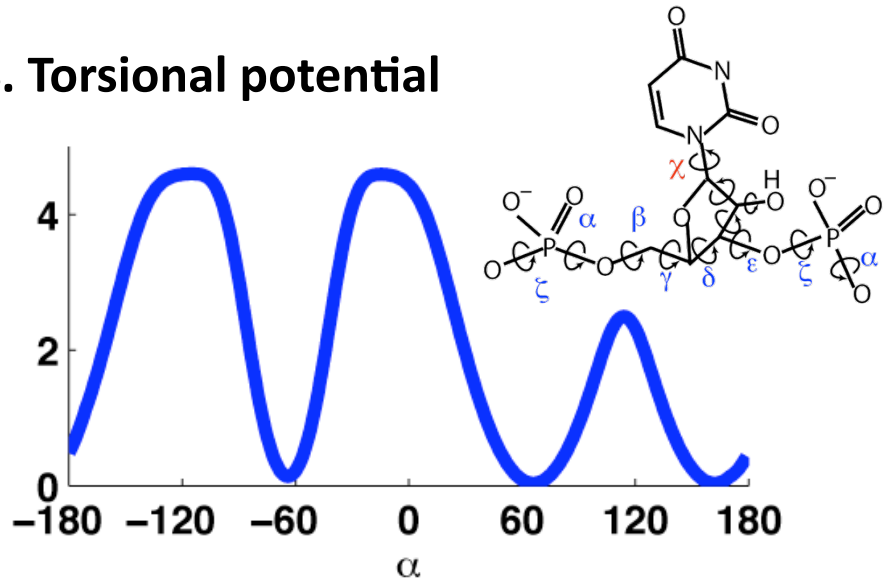


3. Manifestations of water



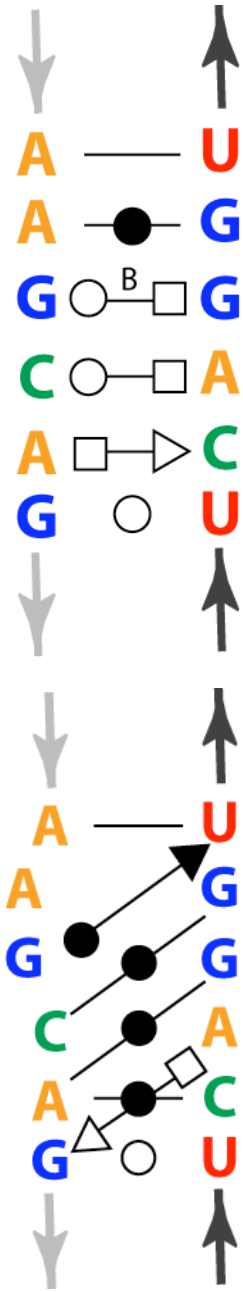
(now orientation-dependent)

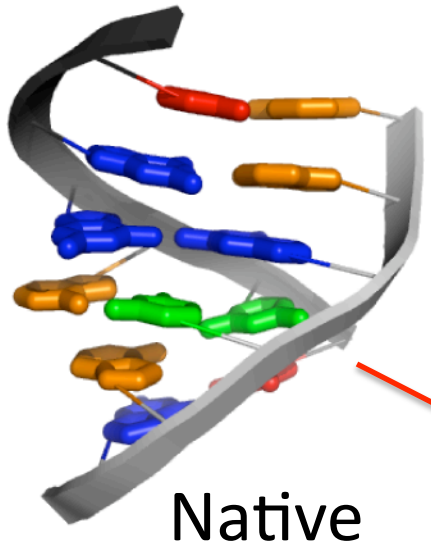
4. Torsional potential



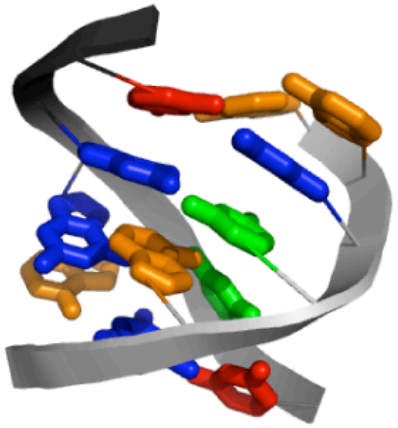
5. Electrostatic repulsion (screened)

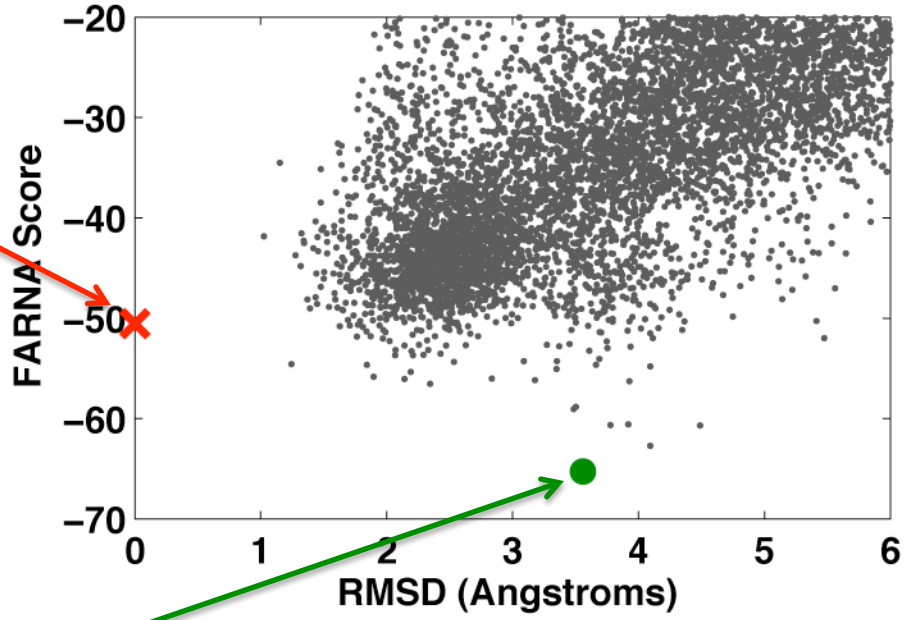
Choice 2. Hi res modeling.





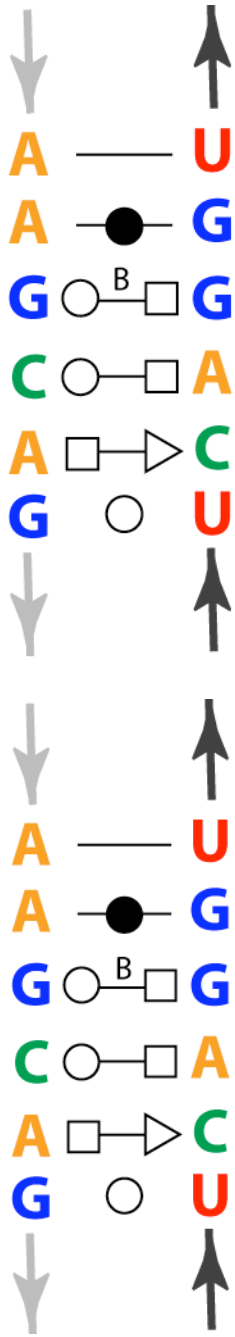
Native

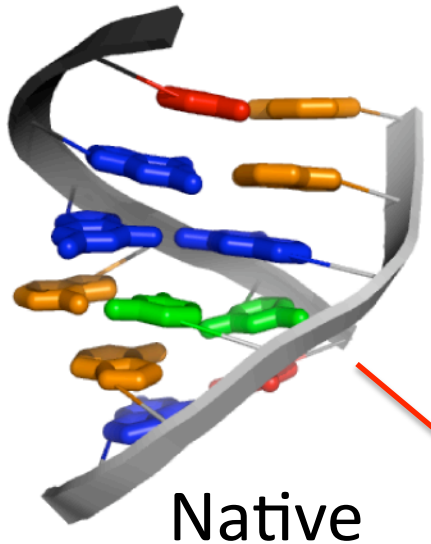




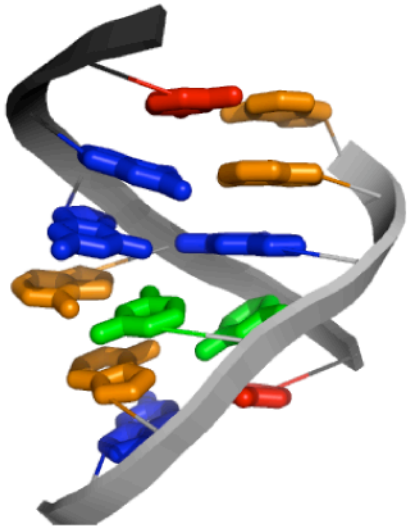
Best scoring by
fragment assembly score

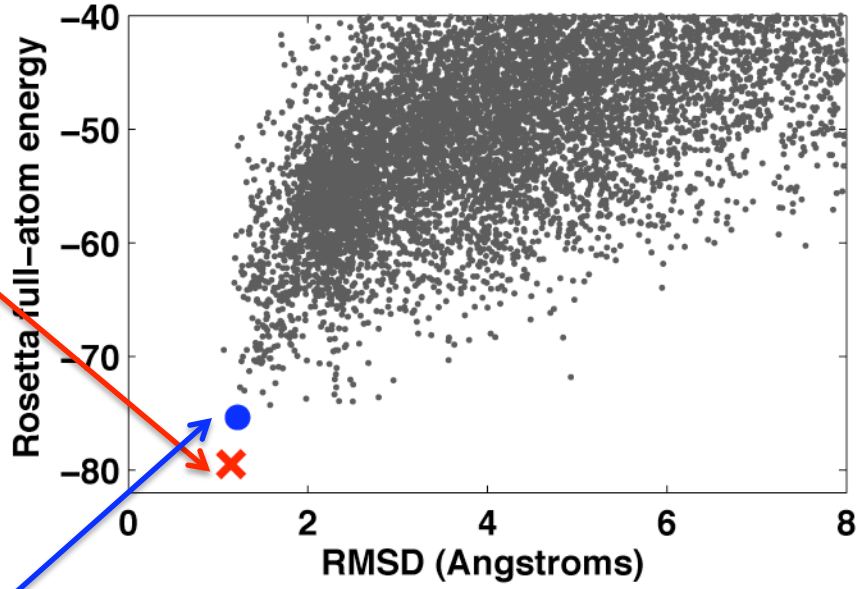
Achieving near-atomic accuracy





Native





Best scoring by
fragment assembly score