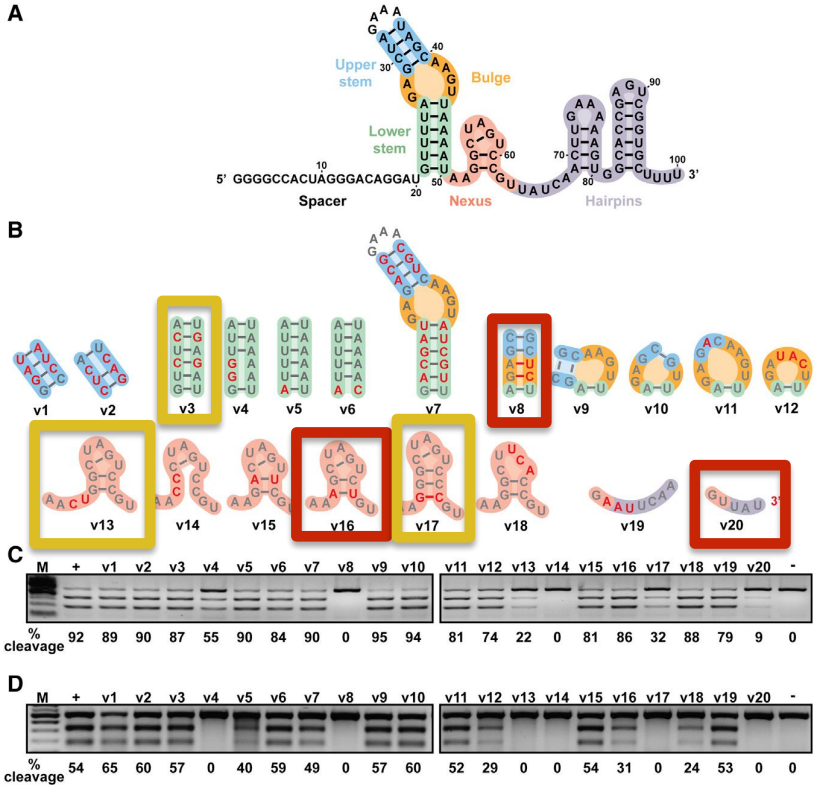
# gRNA/Cas9 Molecular Dynamics

**Zhewei Chen**

**Ch121 Final Project**

**04/24/2019**

# gRNAs have structure and function




Mutations boxed in red cause complete loss of gRNA activity

Mutations boxed in yellow significantly decreases gRNA activity

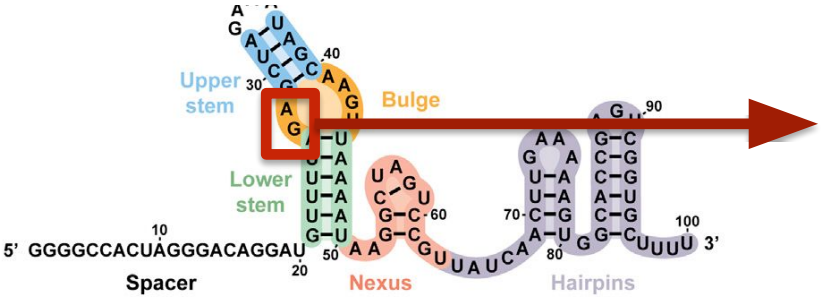
Nexus region and bulge loops are most important gRNA binding motifs

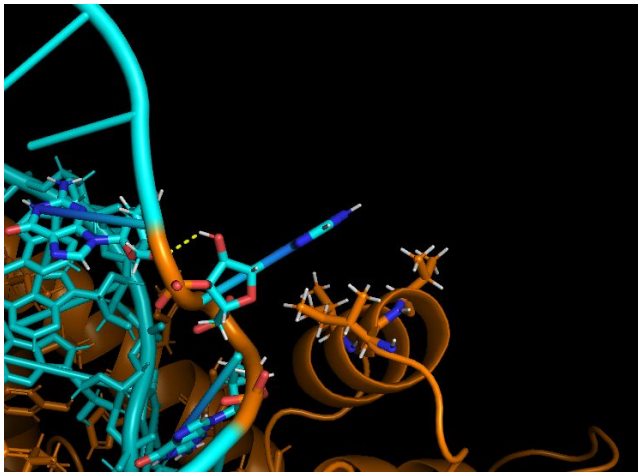
Cas9 handle stem needed for dCas9 binding

Terminator hairpins needed for termination and maybe function?

From Briner et al 2014, "Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality"

# But are there salt bridges? In dCas9 handle?



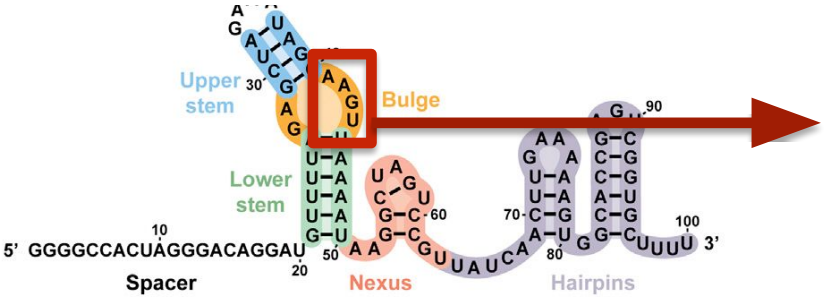



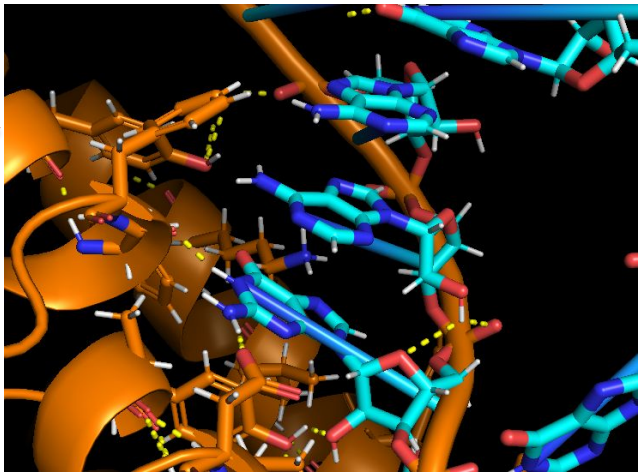
Left bulge is important for function, but does not have any apparent salt bridges with dCas9.

No interactions with the protein at all. The nearby side chains are all non-polar.

Yellow dashed lines = all polar contacts between residues 4A away from the left bulge of gRNA

# But are there salt bridges? In dCas9 handle?



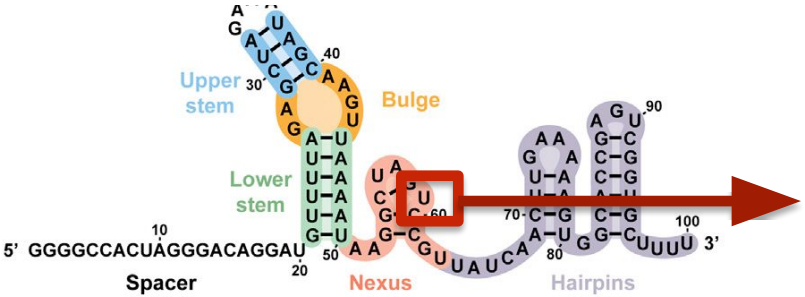



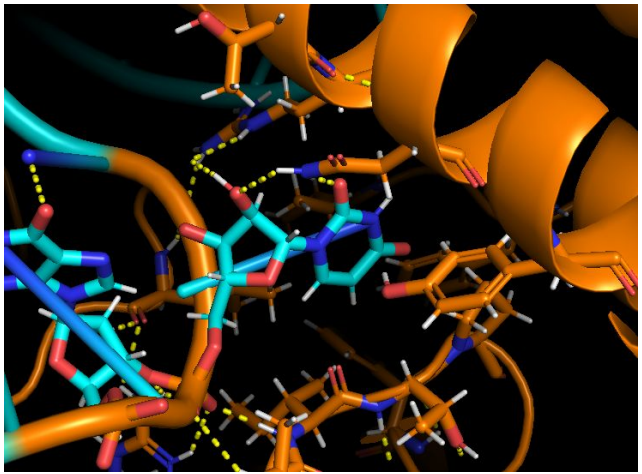
Right side bulge is important for function, but does not have any apparent salt bridges with dCas9.

Most salt bridges involve contact with the phosphate backbone of the gRNA

Yellow dashed lines =
all polar contacts between residues 4A away from the left bulge of gRNA

# But are there salt bridges? In Nexus domain?



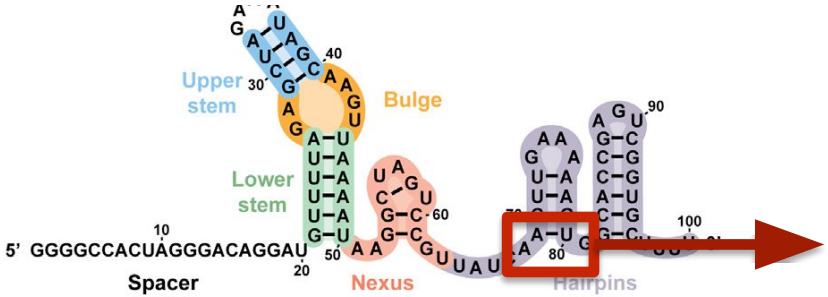
Nexus loop is a very conserved domain. If there are a salt bridges, it should be here.

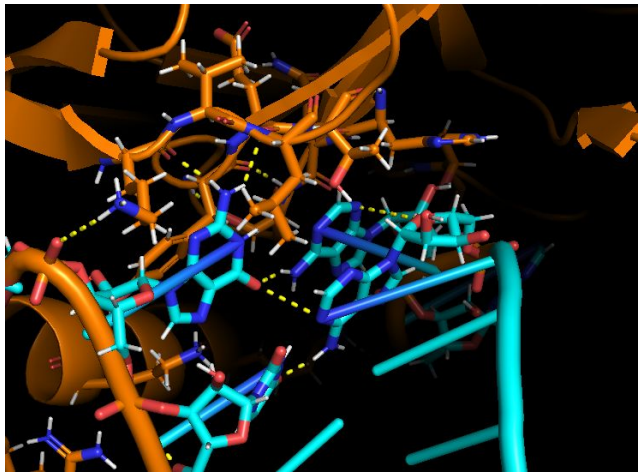
Here a single nucleotide fits inside some kind of binding pocket. However, this pocket is mostly non-polar.

Again, most salt bridges involve contact with the phosphate backbone of the gRNA.

Yellow dashed lines =
all polar contacts between residues 4A away from the left bulge of gRNA

# But are there salt bridges? Near the Terminator stem?



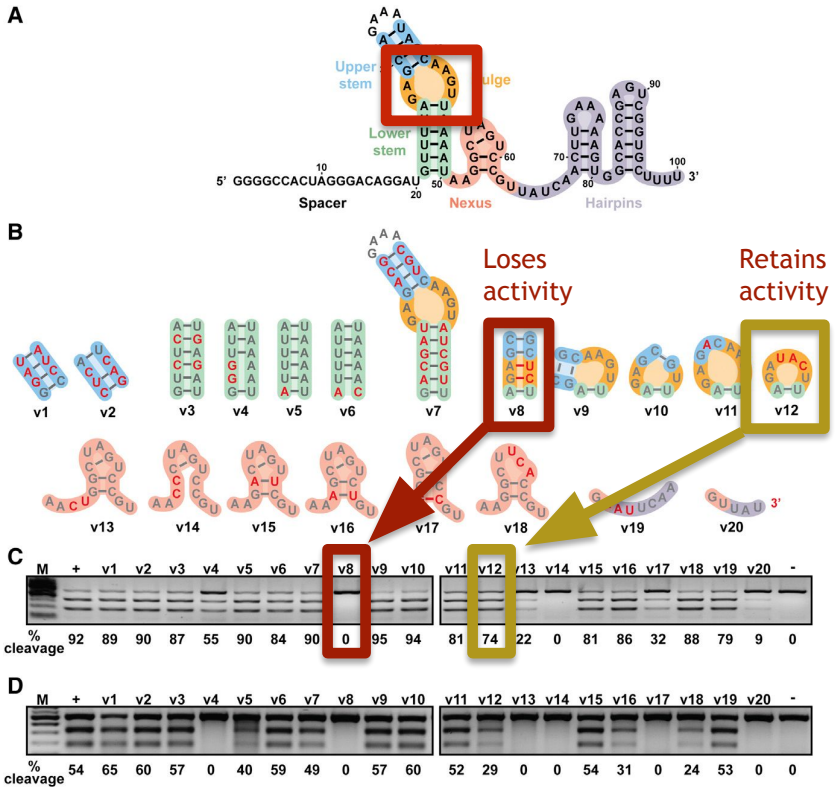



The terminator loop is important for function. Breaking this loop enables the cgRNA activity switch.

Again, no significant hydrogen bonding with the protein. :??

Yellow dashed lines =
all polar contacts between residues 4A away from the left bulge of gRNA

# Build some homology models



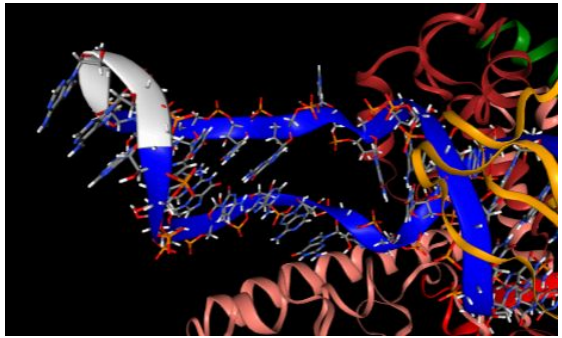

## Approach

1) Introduce mutations to gRNA via RNA homology modeling

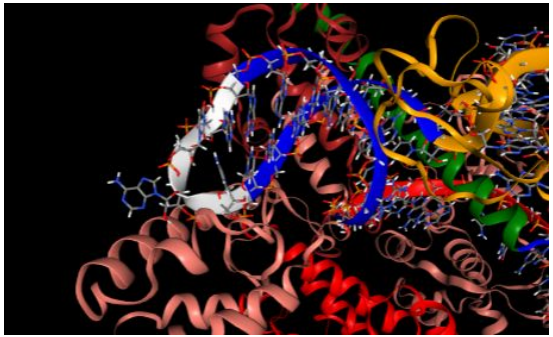
2) See how residues and domains move before and after mutations

3) If important interactions exist, we expect to see some diffusional shift during molecular dynamics simulation

# Build some homology models



Unmodified gRNA



gRNA with deleted upper dCas9 handle

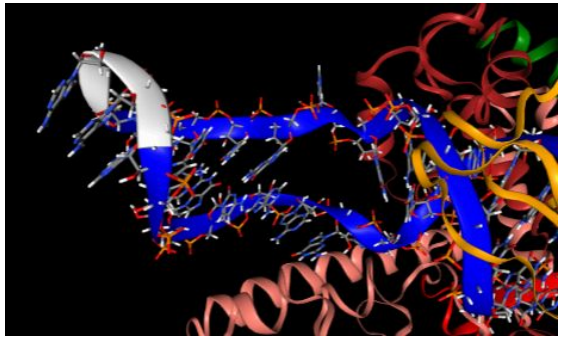
**Approach**

1) Introduce mutations to gRNA via RNA homology modeling

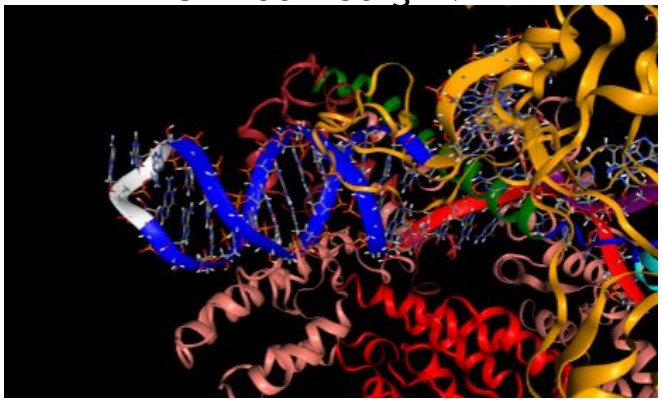
2) See how residues and domains move before and after mutations

3) If important interactions exist, we expect to see some diffusional shift during molecular dynamics simulation

# Build some homology models



Unmodified gRNA



gRNA with deleted bulges

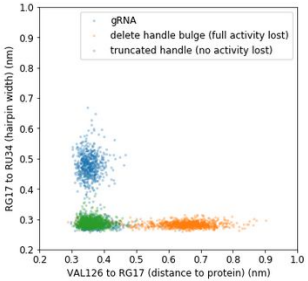
**Approach**

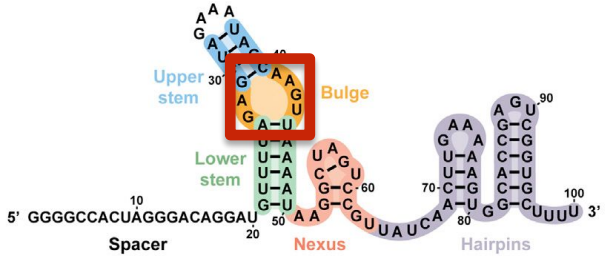
1) Introduce mutations to gRNA via RNA homology modeling

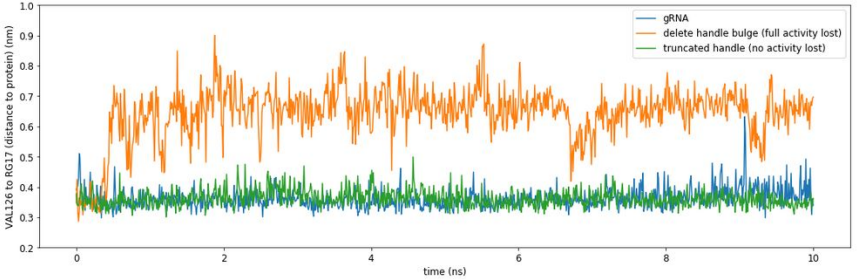
2) See how residues and domains move before and after mutations

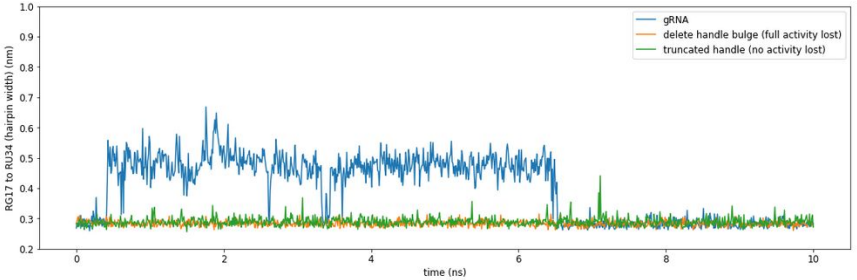
3) If important interactions exist, we expect to see some diffusional shift during molecular dynamics simulation

# If we mutate, does it wiggle?



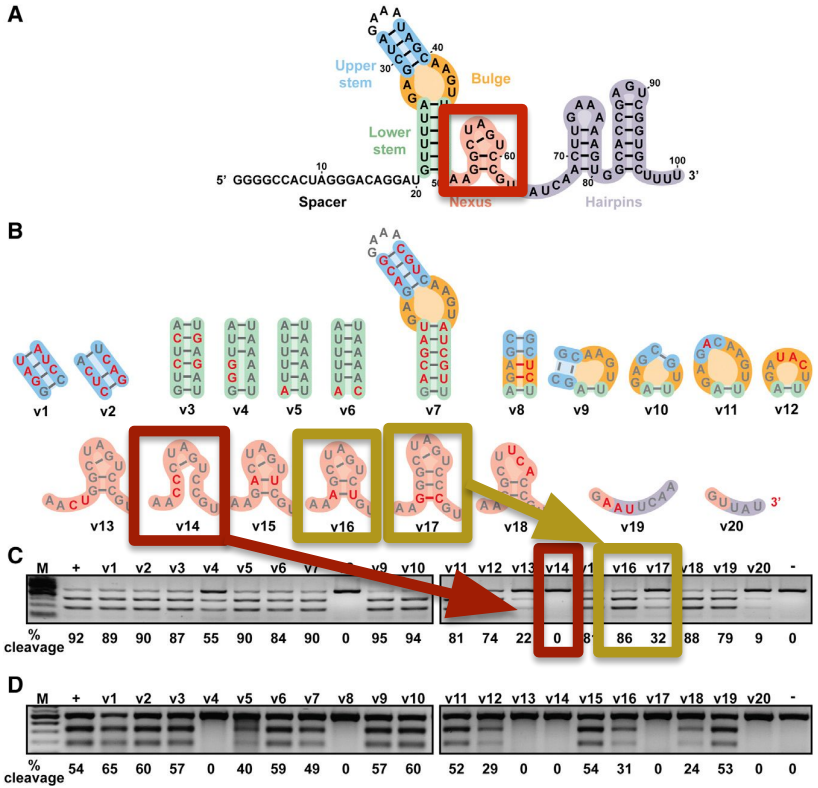




Deleting the bulge causes gRNA to wiggle more?

Truncated handle and regular gRNA do no wiggle much

Unclear if this would cause gRNA to fall out binding pocket in a longer trajectory simulation

# Build some homology models




## Approach

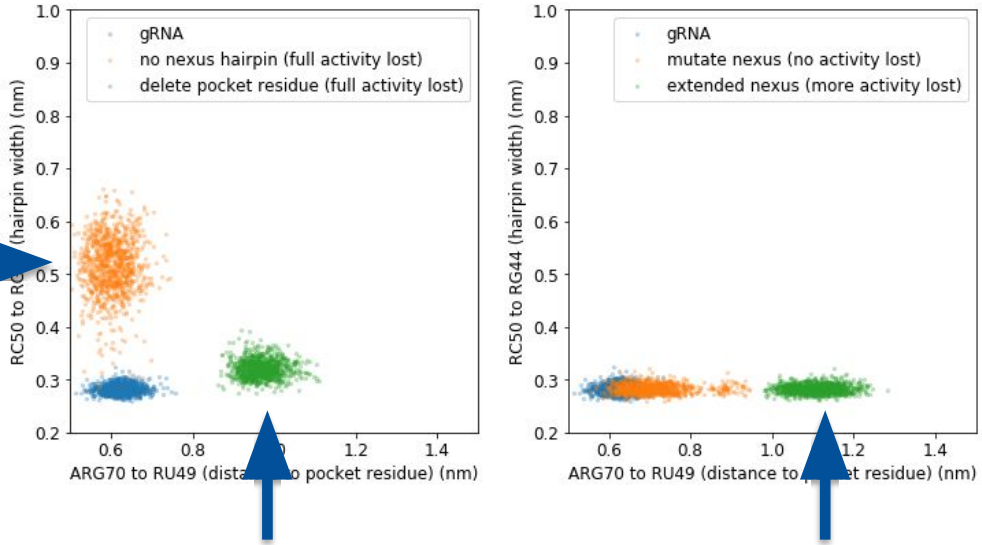
1) Introduce mutations to gRNA via RNA homology modeling

2) See how residues and domains move before and after mutations

3) If important interactions exist, we expect to see some diffusional shift during molecular dynamics simulation

# If we mutate, does it wiggle?

Hairpin starts to unfold, as expected

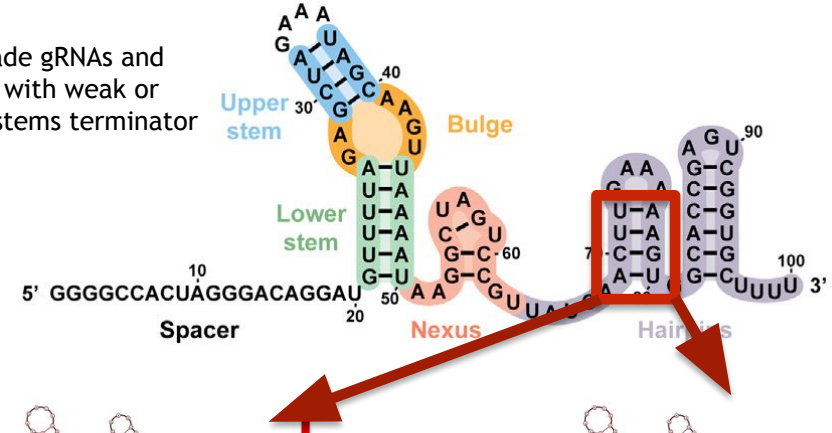



Still unclear how this affects protein activity?

Distance to pocket residue starts to increase as with mutations that disrupt function

# Build some homology models

Also made gRNAs and cgRNAs with weak or strong stems terminator stems




cgRNA
(retains activity)

cgRNA no stem
(loses activity)

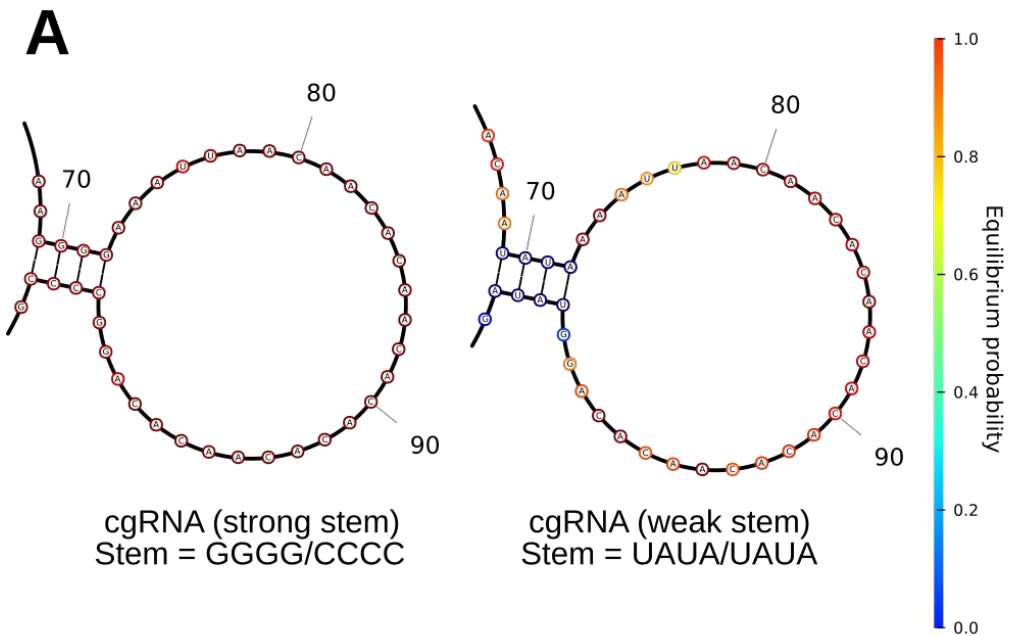
**Approach**

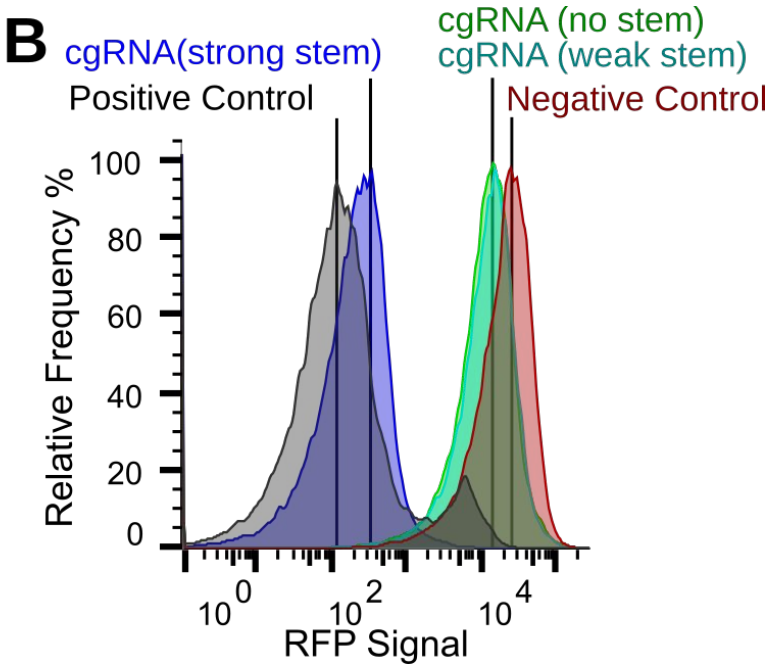
1) Introduce mutations to gRNA via RNA homology modeling

2) See how residues and domains move before and after mutations

3) If important interactions exist, we expect to see some diffusional shift during molecular dynamics simulation

# Novel structure function in S. Pyogenes Terminator

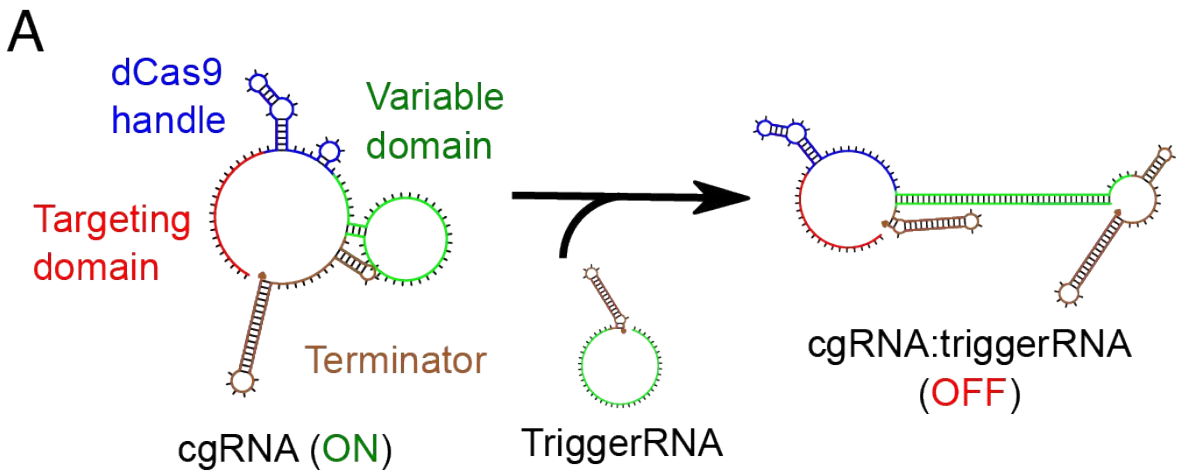


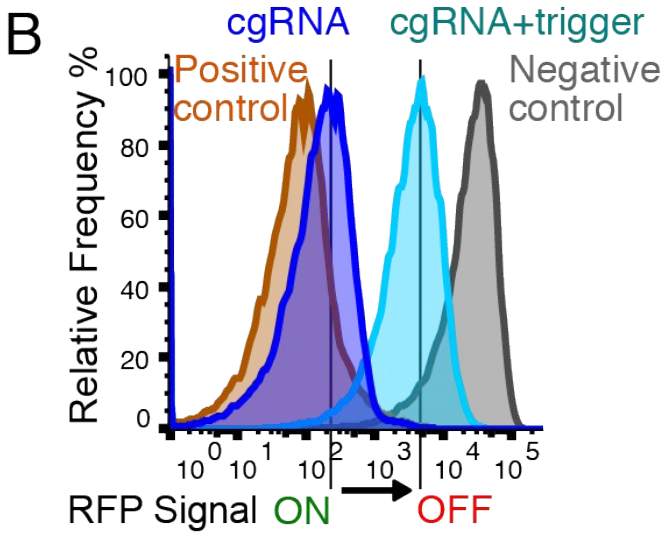




More RFP signal means gRNA loses activity

From scRNA_Cas9_20190211ZC

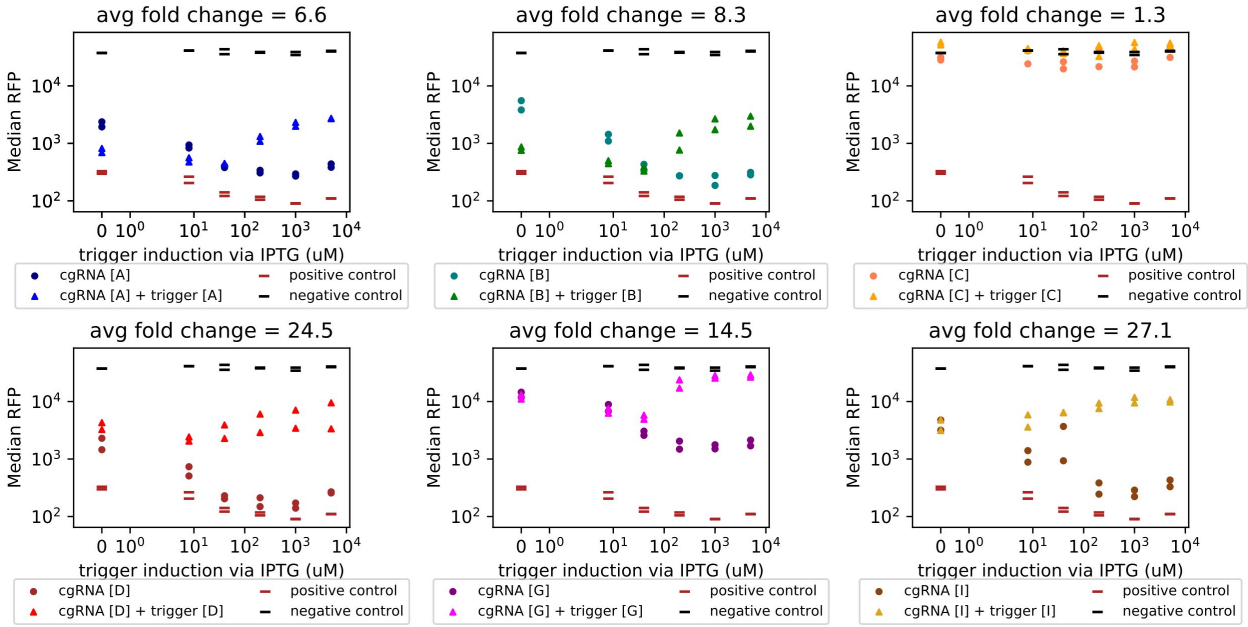
# Novel activity switch mechanism for CRISPR gRNAs




From one of my experiments long long ago

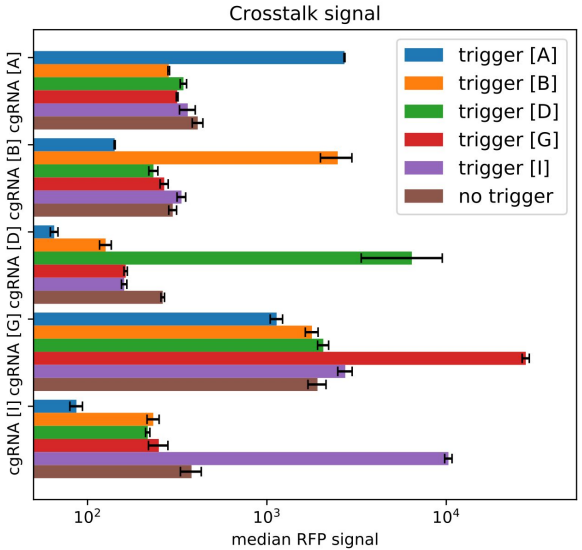
# Orthogonal sets of RNA regulators can be generated

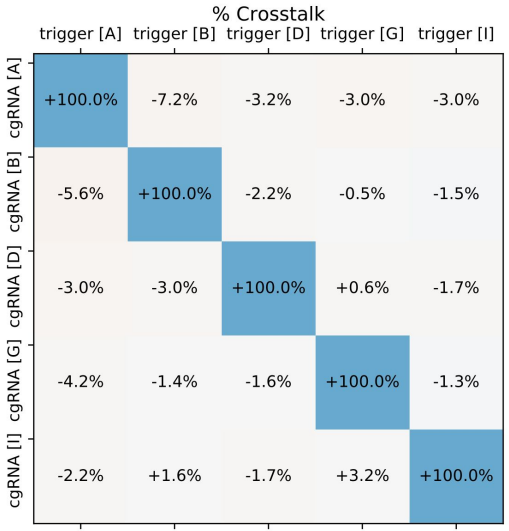



cgRNAs are dose responsive and orthogonal to each other

From some experiment long long ago

# Orthogonal sets of RNA regulators can be generated

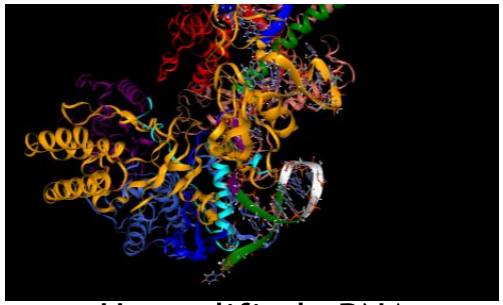


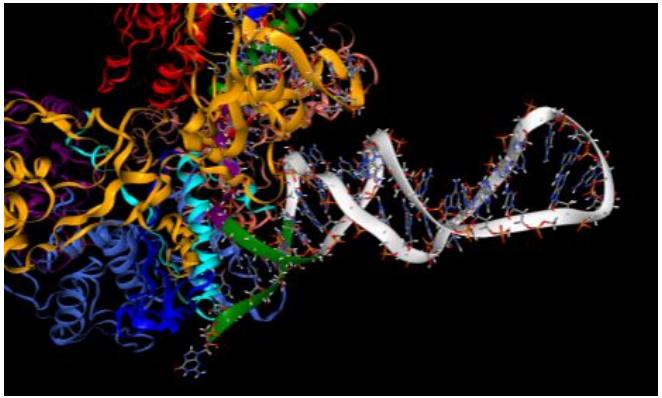

cgRNAs are dose responsive and orthogonal to each other

From some experiment long long ago

# Can also make cgRNAs via Rosetta



Unmodified gRNA



cgRNA

**Approach**

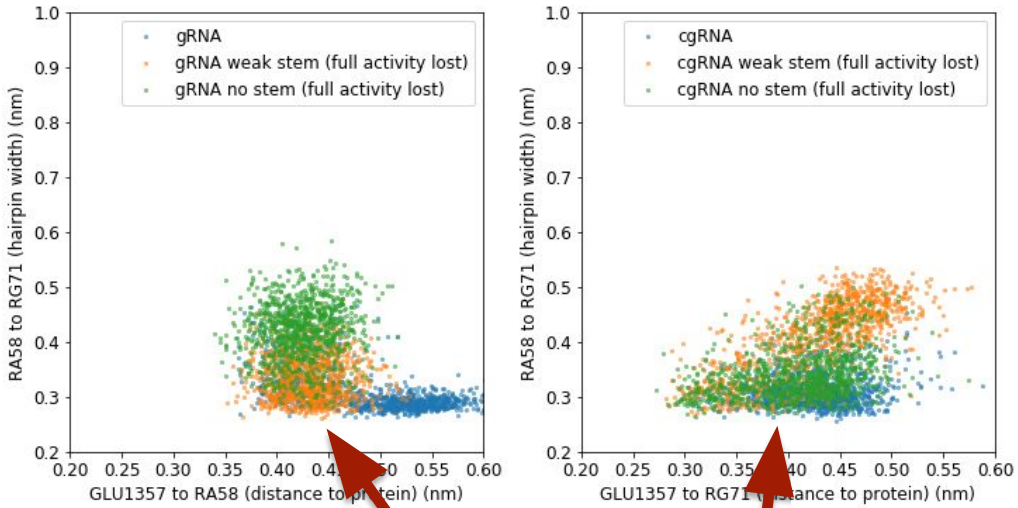
1) Introduce mutations to gRNA via RNA homology modeling

2) See how residues and domains move before and after mutations

3) If important interactions exist, we expect to see some diffusional shift during molecular dynamics simulation

# If we mutate, does it wiggle?

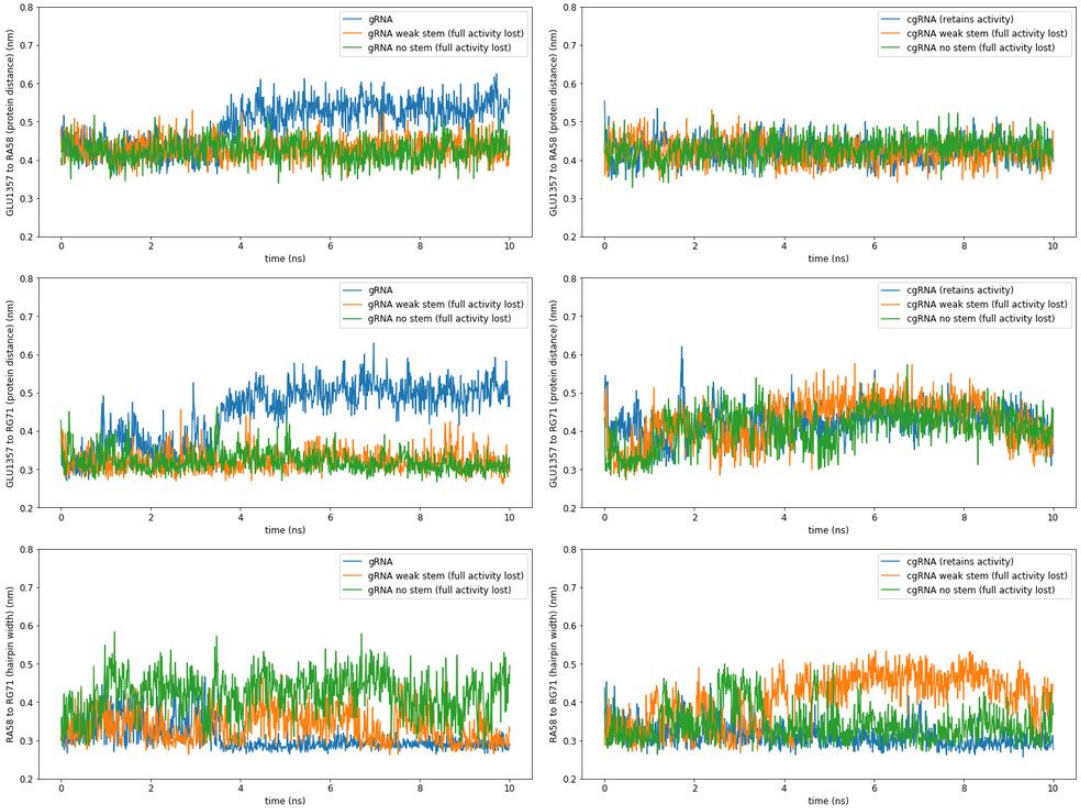
Hairpin width increases with hairpin breaking mutations (expected)




Distance to protein decreases for loss of function mutations :?

Molecular dynamics is not really informative on why cgRNA work

# If we mutate, does it wiggle?




No interesting correlation with binding to protein surface

Hairpin width increases with hairpin breaking mutations (expected)

# Challenges

- Hard to simulate trajectories long enough to see something interesting
- Most simulation effort wasted on water vibrations
- Initial states are usually stuck in local energy minima

# Possible next steps

**Coarse graining with cgmartini and elastic network**

Current system:

Total # of atoms = ~220k

Run time = 1 ns/hr

Coarse grained:

Total # of atoms = ~220k/50 = 4400 atoms?

Run time = 1ns/hr * 50 = 40ns/hr = ~1us per day

**Adaptive weighted histogram sampling**

Try to bias trajectories to sample under represented ensemble conformations

Can provide more efficient sample of reaction coordinate

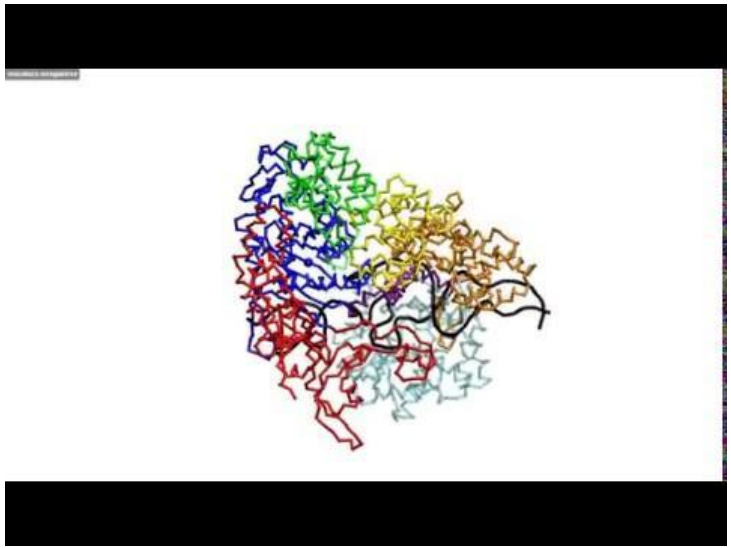
http://courses.theophys.kth.se/SI3450/awh.pdf

**Gaussian Accelerated molecular dynamics**

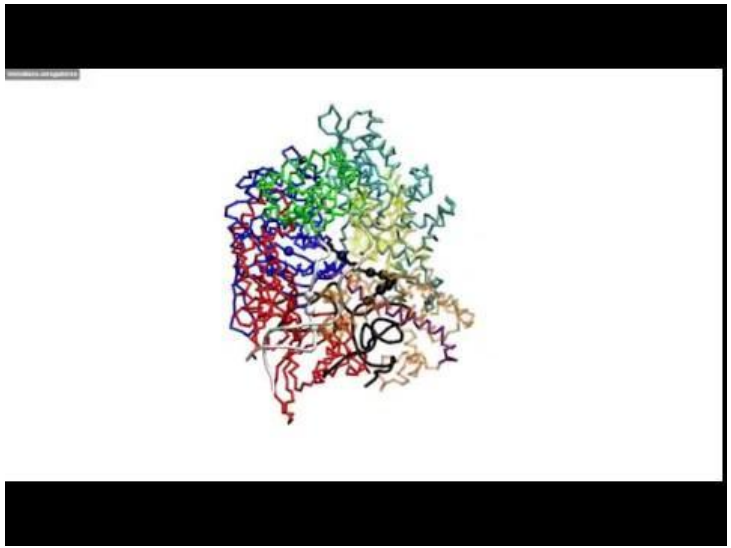
Similar to AWH but only implement in Amber MD

Tries to use boost potentials to escape energies wells and do better sampling

# Coarse grain gRNA + Cas9 dynamics



Cas9 —> Cas9:gRNA



Cas9:gRNA —> Cas9:gRNA:DNA

Can be done with limited compute resources! :)
But results may not always be accurate

Results from https://doi.org/10.1002/prot.25229

# Or accelerated full atom MD?

conformational transition

Inactive state of Cas9 (4UN3.pdb)

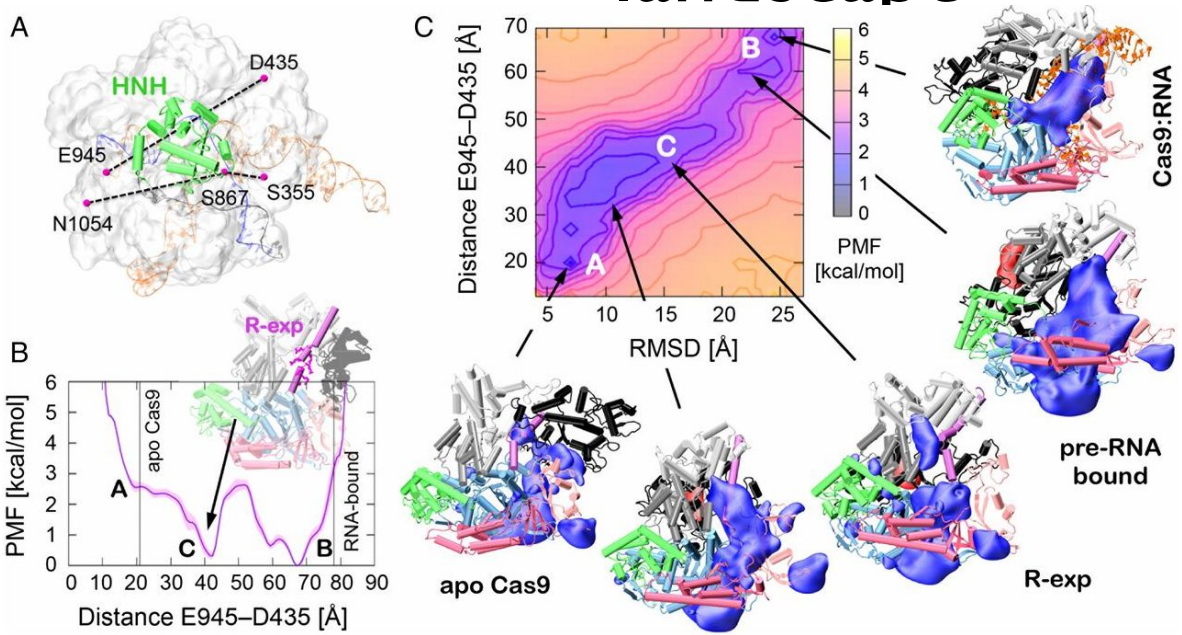
Cas9 <— Cas9:gRNA

Cas9:gRNA:DNA —> Cas9 + DNA cleaved

Gaussian accelerated full atom trajectories run for 15us
—> too long to be practical biomolecule engineering

Results from https://doi.org/10.1073/pnas.1707645114

# Mapping out the RNA docking energy landscape




If we could do gaMD, we can sample out the energy landscape for gRNA/Cas9 docking —> use information to design better cgRNAs

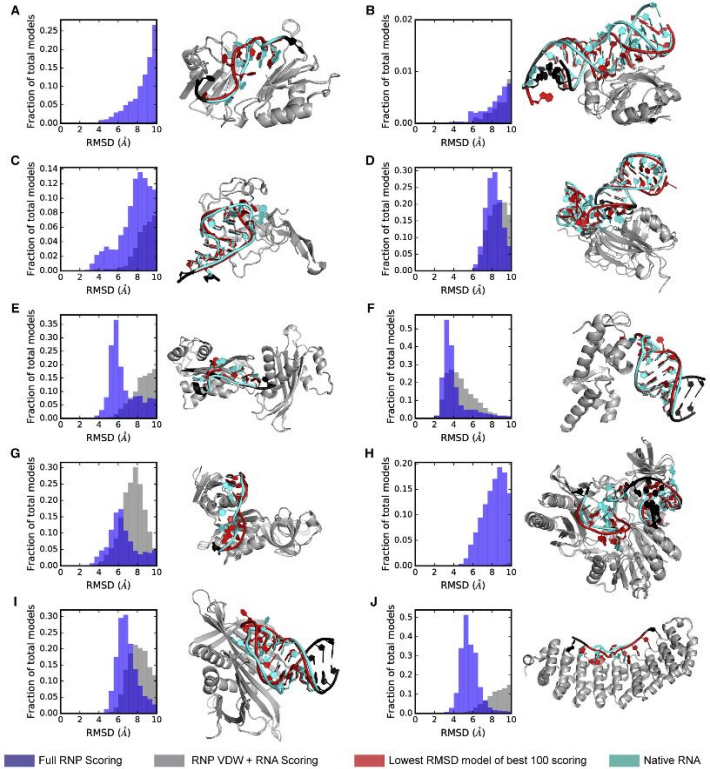
Results from https://doi.org/10.1002/prot.25229

# Actual Next Steps

**Still need good starting states to do good reaction coordinate scans**

**Solution →**
**Use Rosetta RNA_denovo to sample for good RNA/protein docks of the Apo Cas9**

**But Problem →**
**RNA_denovo still requires good initial positions for building RNA/protein docks**




Results from https://doi.org/10.1016/j.str.2018.10.001

# Actual Next Steps

**Still need good starting states to do good reaction coordinate scans**

**Solution →**
**Use Rosetta RNA_denovo to sample for good RNA/protein docks of the Apo Cas9**

**But Problem →**
**RNA_denovo still requires good initial positions for building RNA/protein docks**

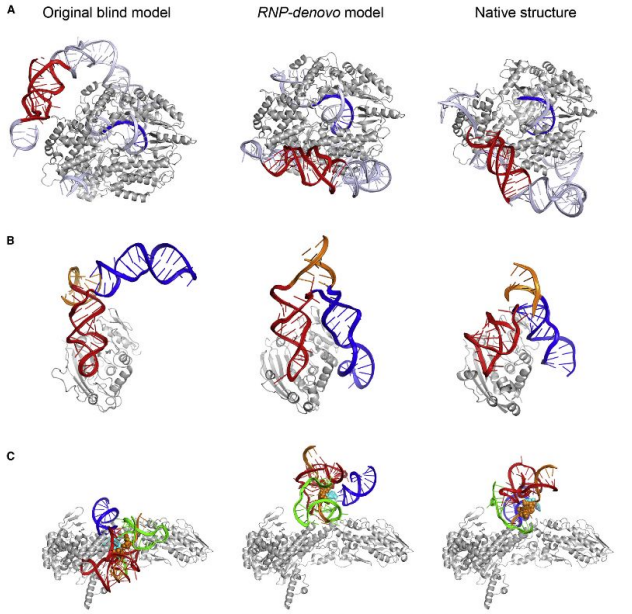



**Figure 5. Revisiting Three Past Modeling Challenges with the Rosetta *RNP-denovo* Fold-and-Dock Method**
(A) The best of the previously selected blind human telomerase core RNP models built without FRET data (Parks et al., 2017) (left; RMSD over select pseudoknot residues = 78.8 Å; pseudoknot RNA motif colored red, template RNA colored blue, other modeled RNA colored light blue, protein colored gray), the best *RNP-denovo* fold-and-dock model by RMSD of the top 100 scoring models built without FRET (middle; RMSD over select pseudoknot residues = 15.0 Å), and the cryo-EM structure of human telomerase (Nguyen et al., 2018) (right).
(B) The best of the previously submitted ten RNA methyltransferase CAPRI T33 models (left; RMSD = 31.0 Å; RNA colored blue, red, and orange; protein colored gray), the best of the top 100 scoring *RNP-denovo* fold-and-dock models (middle; RMSD = 13.6 Å), and the best T34 model, which closely resembles the crystal structure (right; interface RMSD to crystal structure = 1.5 Å).
(C) The previously published human spliceosomal C complex model (Anokhina et al., 2013) (left; RMSD over key active site residues [shown as spheres] = 34.5 Å, U2 RNA colored green, U5 RNA colored blue, U6 RNA colored red, intron colored orange, 5′ exon colored cyan, protein colored gray), the best RMSD model of the top 100 scoring *RNP-denovo* fold-and-dock models of the human C complex (middle; RMSD over key active site residues = 8.0 Å), and the cryo-EM structure of the yeast C complex (right).
See also Figure S5.

Results from https://doi.org/10.1016/j.str.2018.10.001

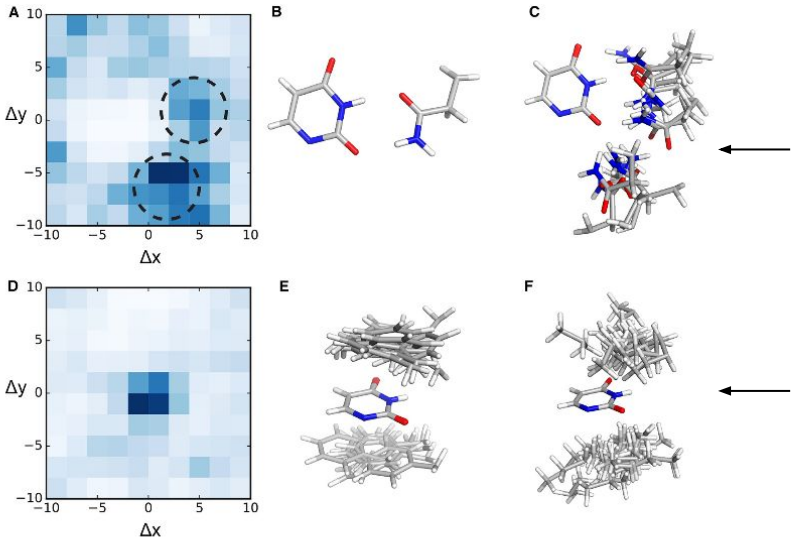
# CNN recommender system for RNA/protein docking sites

Possible solution →

**Train Convolution Neural Network (CNN) on PME electrostatics and learn about about good surface binding for protein/RNA**

Why should it work? →

**Statistical energy potential in RNA_denovo can pick up useful chemical features not represented in Rosetta force field model**




Pseudo base pairing between gluamine and uracil

Pi-Pi stacking of aromatic residues

**Figure 2. Statistical RNA-Protein Potential in Rosetta**

(A) The distribution of glutamine side-chain centroids around uracil in the plane of the base ($0 < |z| < 3$ Å), from the non-redundant set of RNA-protein crystal structures from the PDB (darker blue represents higher frequency). Two major hotspots are circled. Distributions of all protein side chains around all four RNA bases are shown in Figures S2A–S2D.

(B) A pseudo-pair between glutamine and uracil.

(C) Conformations from the two major hotspots circled in (A) show that the interactions between glutamine and uracil are not highly stereotyped.

(D) The distribution of phenylalanine side-chain centroids around uracil above and below the plane of the base ($3 < |z| < 6.5$ Å; darker blue represents higher frequency). Distributions of all protein side chains around all four RNA bases are shown in Figures S2E–S2H.

(E) Representative conformations from the hotspot in (D) show stereotyped stacking interactions.

(F) Conformations of valine around uracil also reveal frequent stacking interactions.

See also Figures S2–S4.

Results from https://doi.org/10.1016/j.str.2018.10.001