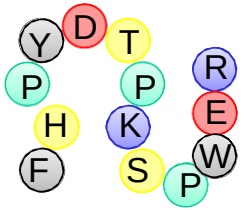
Protein Design

Gustavo Araiza
Rosetta Workshop
December 2022

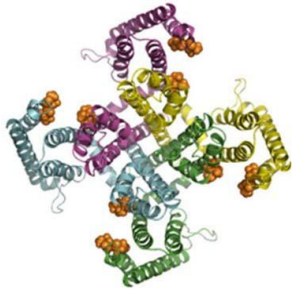
Protein Design is the Inverse Protein Folding Problem



Protein Folding Problem

Sequence ⇌ Structure

Inverse Folding Problem

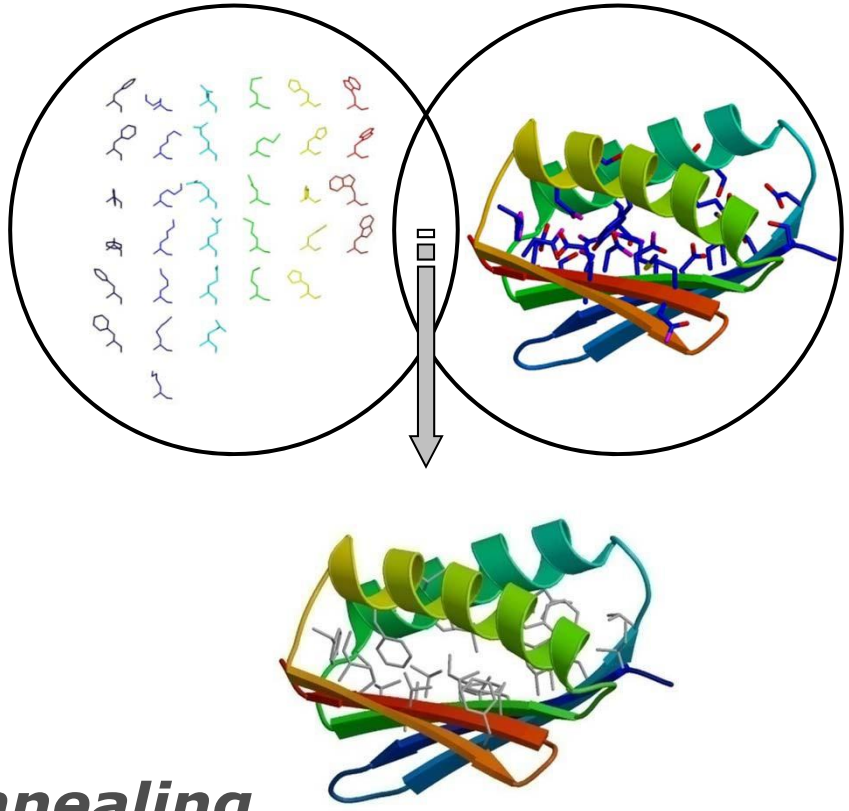


Given a protein fold – which primary sequence(s) can fold into it?

Protein Design Uses the Rosetta Energy Function and Local Rotamer Libraries

Local Rotamer Bias

Approximate interactions between sidechains using the distribution of sidechain conformations seen in known protein structures

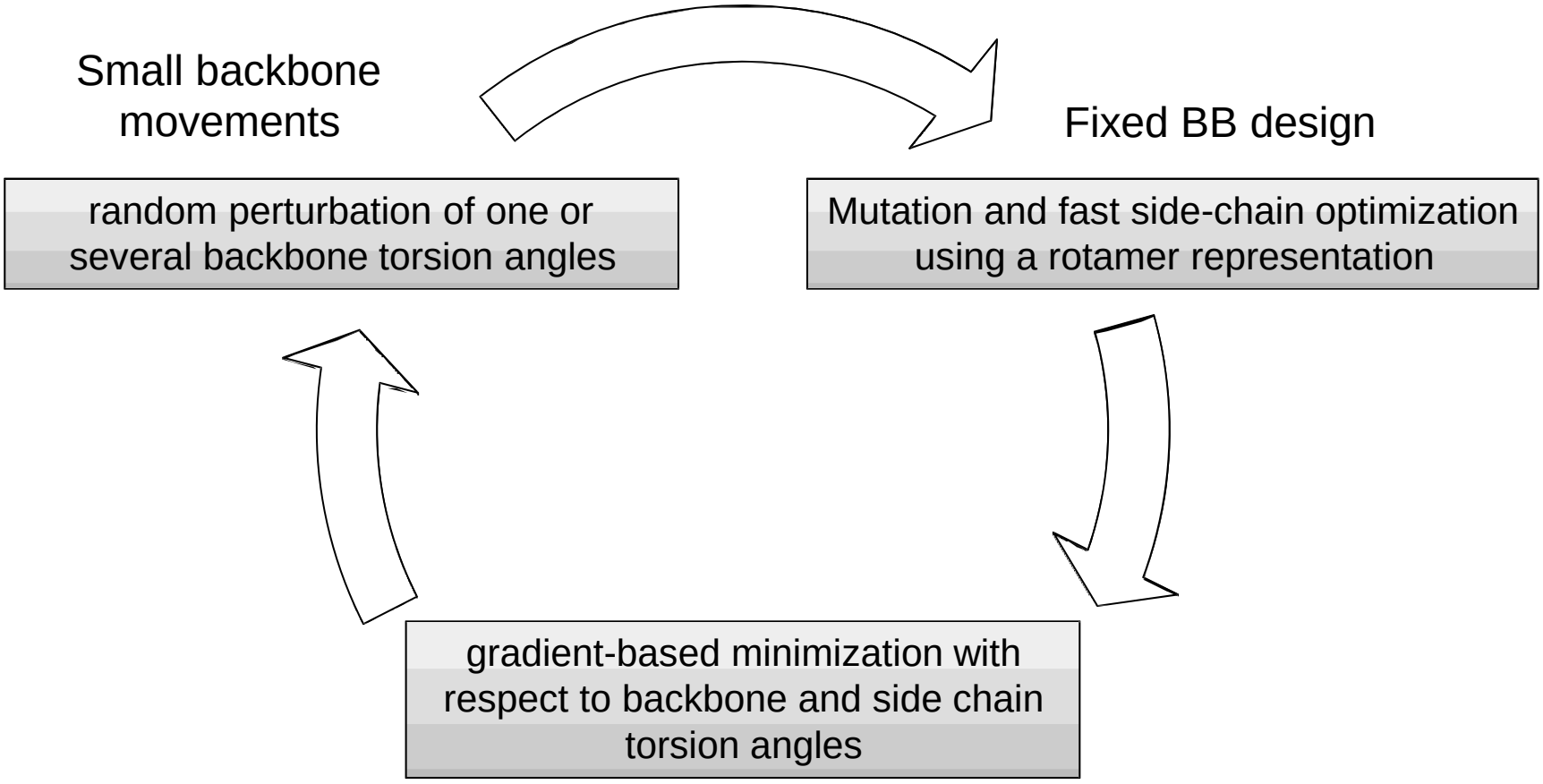


Energy function

- ***VDW interactions***
- ***solvation***
- ***hydrogen bonding potential***
- **elec *interactions***
- ***rotamer probability***

Simulated Annealing Monte Carlo optimization

Iterative Optimization of Sequence and Conformation in Rosetta



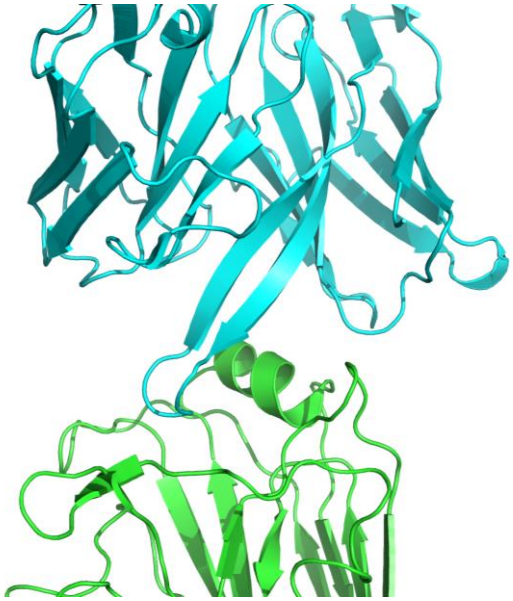
Single-state design

Also known as redesign
for antibodies: computational affinity maturation

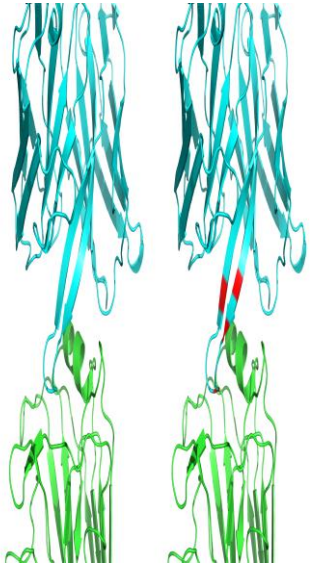
Goal: take an existing antibody-antigen complex and optimize the antibody sequence for tighter binding

Antibody

Antigen

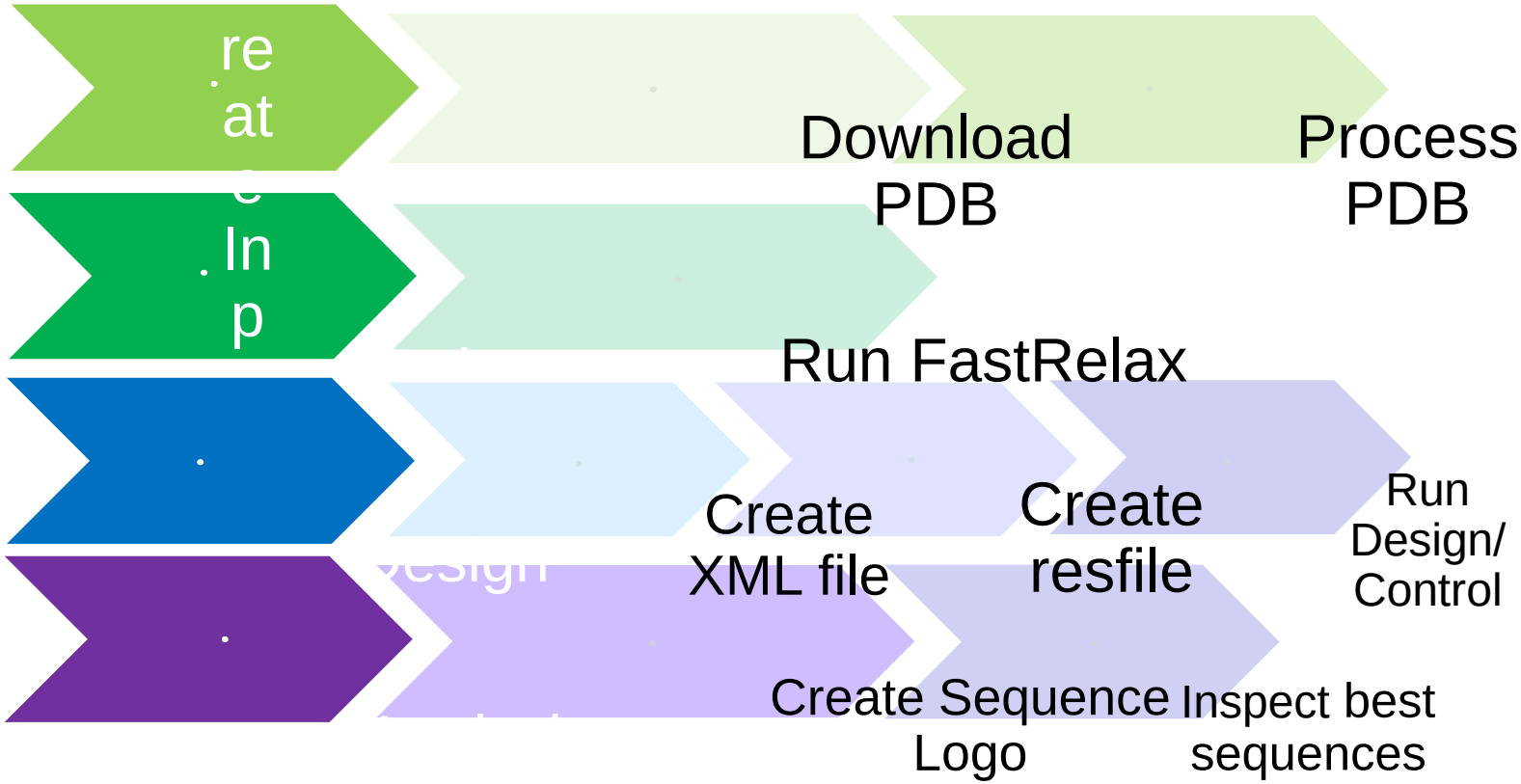


Design



PDB ID 4yk4

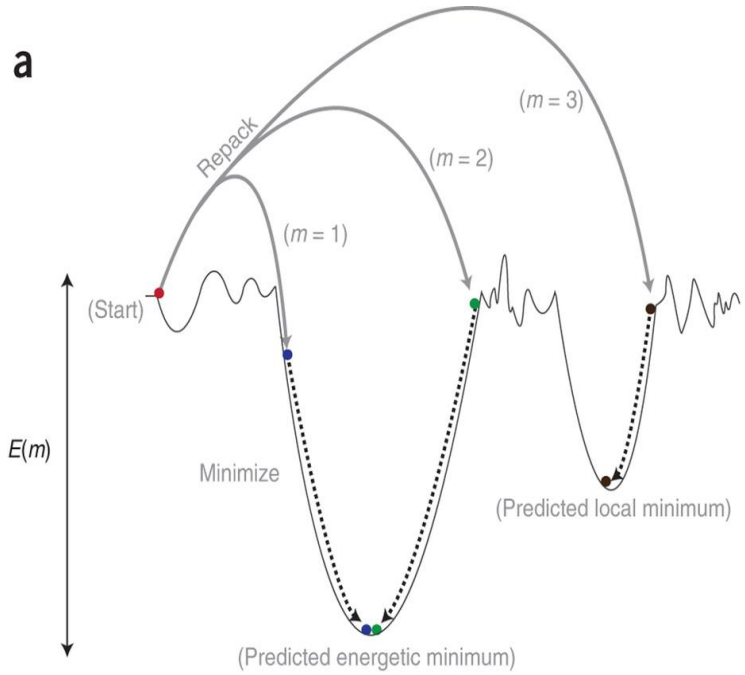
Single-state design protocol overview



Run FastRelax

FastRelax is designed to optimize the protein backbone/side chains to model at an energy minimum

Helps relieve clashes that may introduce artifacts into design



Benchmarks show as little as 1 Å backbone movement can completely change sampled sequences

`-constrain_relax_to_start_coords`

keeps things close to inputs

Combs, *et al*, Nat. Prot. 2013

Create XML file

Please open

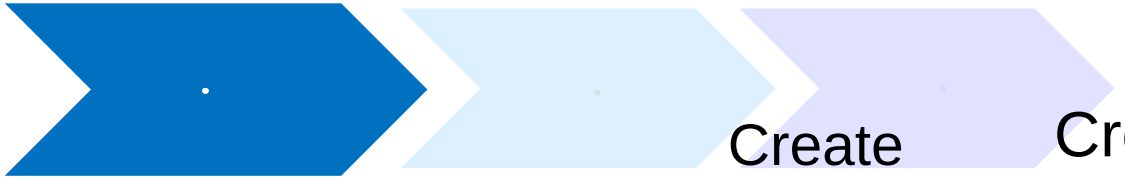
protein_design/single_state_design/input_files/design.xml

Where should you start looking?

```
<PROTOCOLS>
    Run the design protocol
        <Add mover="design" />

    Calculate interface metrics for the final sequence
        <Add mover="analyze" />
</PROTOCOLS>
```

```
<PackRotamerMover name="design" scorefxn="REF2015"
                        task_operations="ifcl,rrf" />
```




```
Design and repack residues based on resfile
<ReadResfile name="rrf" filename="4HKX.resfile"/>
```

```
NATRO
start
30 H ALLAA
31 H ALLAA
…
152 L ALLAA
155 L ALLAA
…
333 A NATAA
334 A NATAA
```

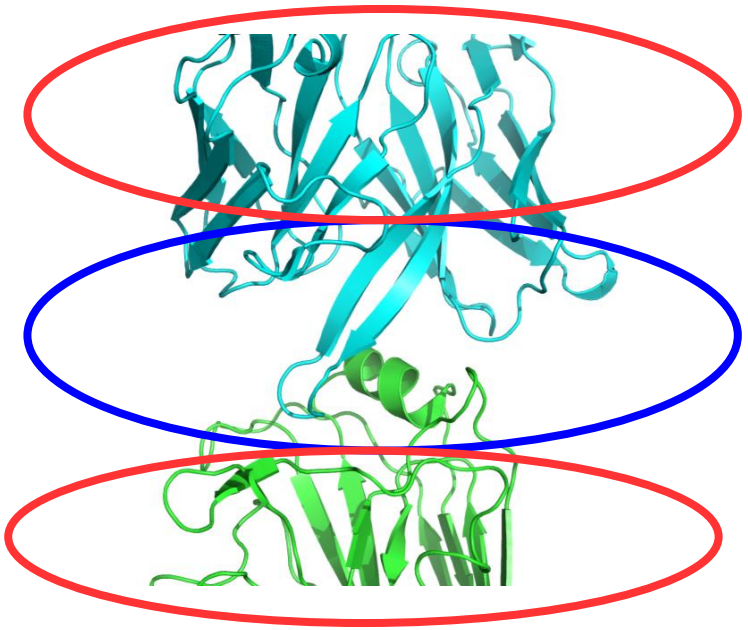
Do only design the residues specified below (the interface)

Free sequence design on heavy and light chain interface residues

Repack the antigen, do not mutate

Autogenerate resfile: Use the python script protein_design/scripts/define_interface.py

Interface: any atom within a cutoff (default 5 Å) of any atom in the opposing chain.



Not Interface

Interface

Not Interface

- **Total score:** score of the entire complex

- **Binding energy:** difference in energy between the bound and unbound partners

- **Binding density:** Binding energy divided by the buried surface area. Prevents a low binding energy by increasing buried surface area.

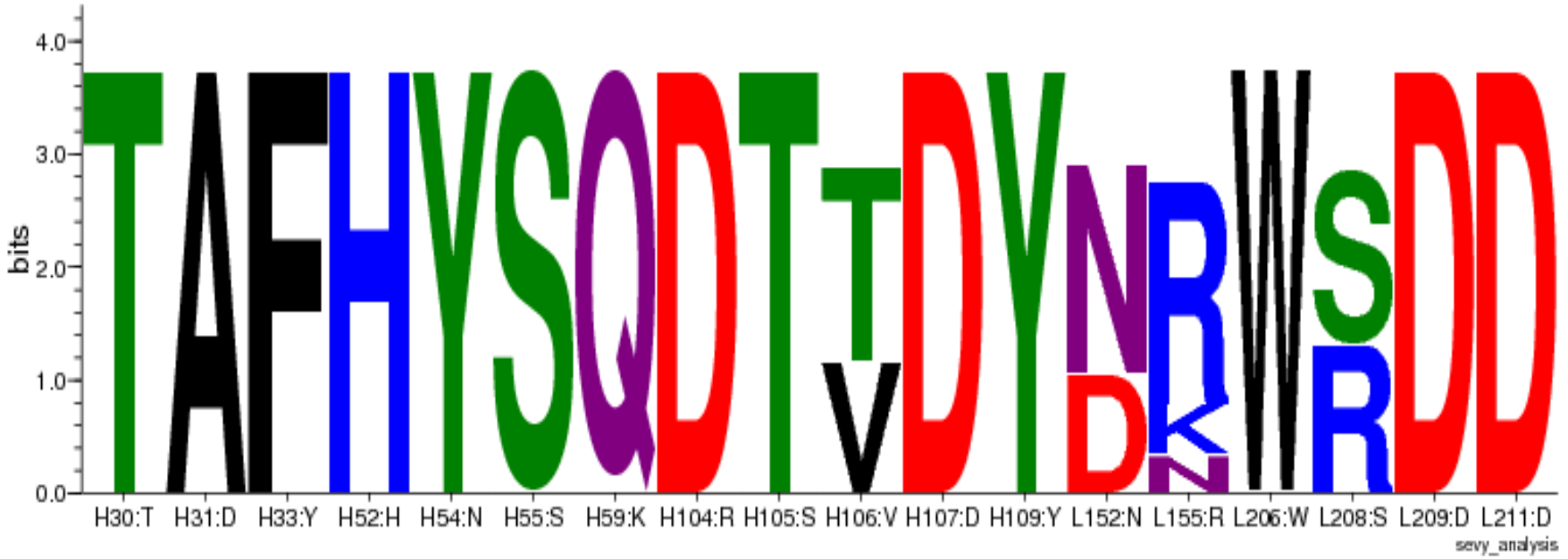
```
<InterfaceAnalyzerMover name="analyze" scorefxn="REF2015"
  packstat="0" pack_input="0" pack_separated="1"
  fixedchains="H,L" />
```

- packstat: activates packstat calculation (packing statistics, Rosetta holes); can be slow so it defaults to off
- fixedchains: comma-delimited list of chain ids to define a group in the interface.
- pack_separated: repack the exposed interfaces when calculating binding energy? Usually a good idea.
- pack_input: prepack before separating chains when calculating binding energy? Useful if these are non-Rosetta inputs

Create Sequence Logo

Useful to quickly see which residues are being designed, and what amino acids are being put there

Made by WebLogo application through
protein_design/scripts/design_analysis.py



http://weblogo.berkeley.edu/

Use the python script located in protein_design/scripts/PerResidueEnergies.py

Plots relative per residue energies.
(input pose – design pose)
Values smaller than zero indicate improvements relative to the input pose

Red: Mutations

Create Sequence Logo

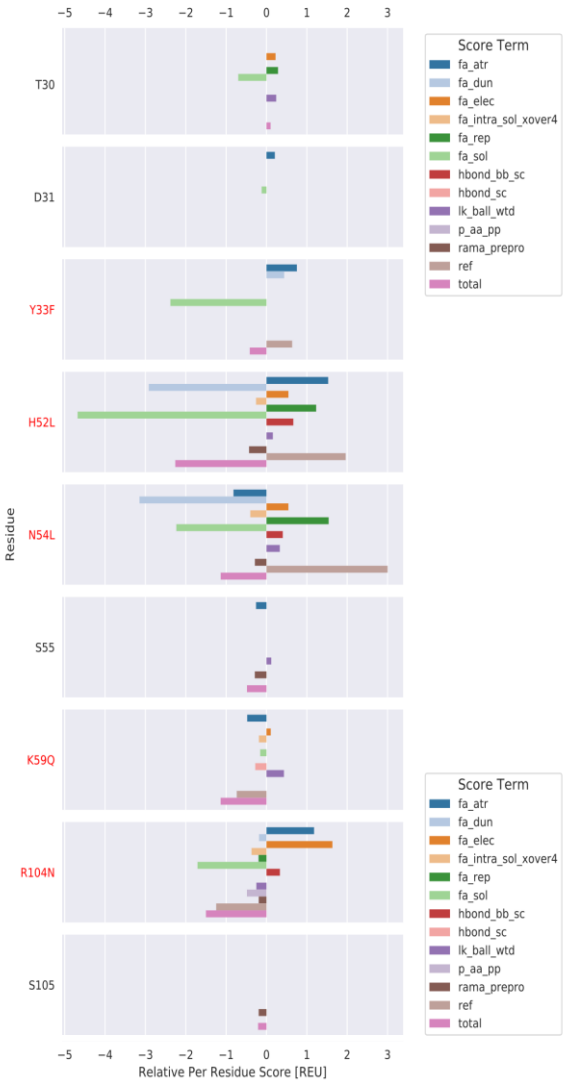
Inspect best sequences

- Position specific score changes for each Rosetta scoring term
- REF 2015 scoring terms:

The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design
Alford et al (2017)

- Made by supplementary script through

protein_design/scripts/PerResidueEnergies.py



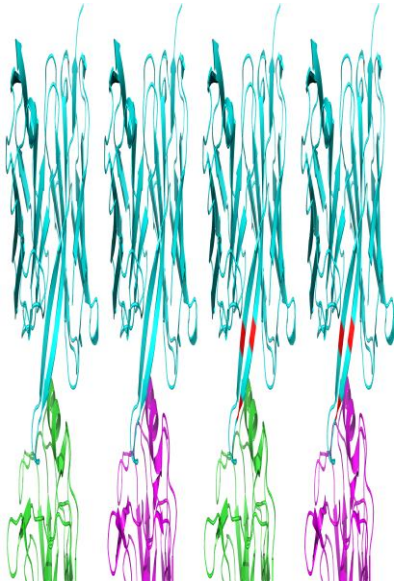
Multistate design

Multistate design: Optimize ONE sequence for low energy in MULTIPLE structures (states)

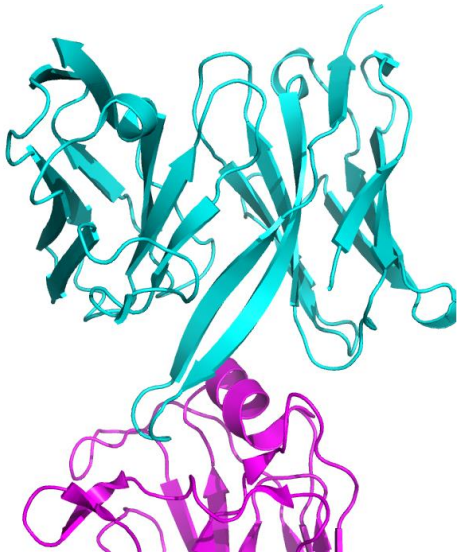
Redesign an antibody to recognize multiple targets

Antibody

Antigen 1



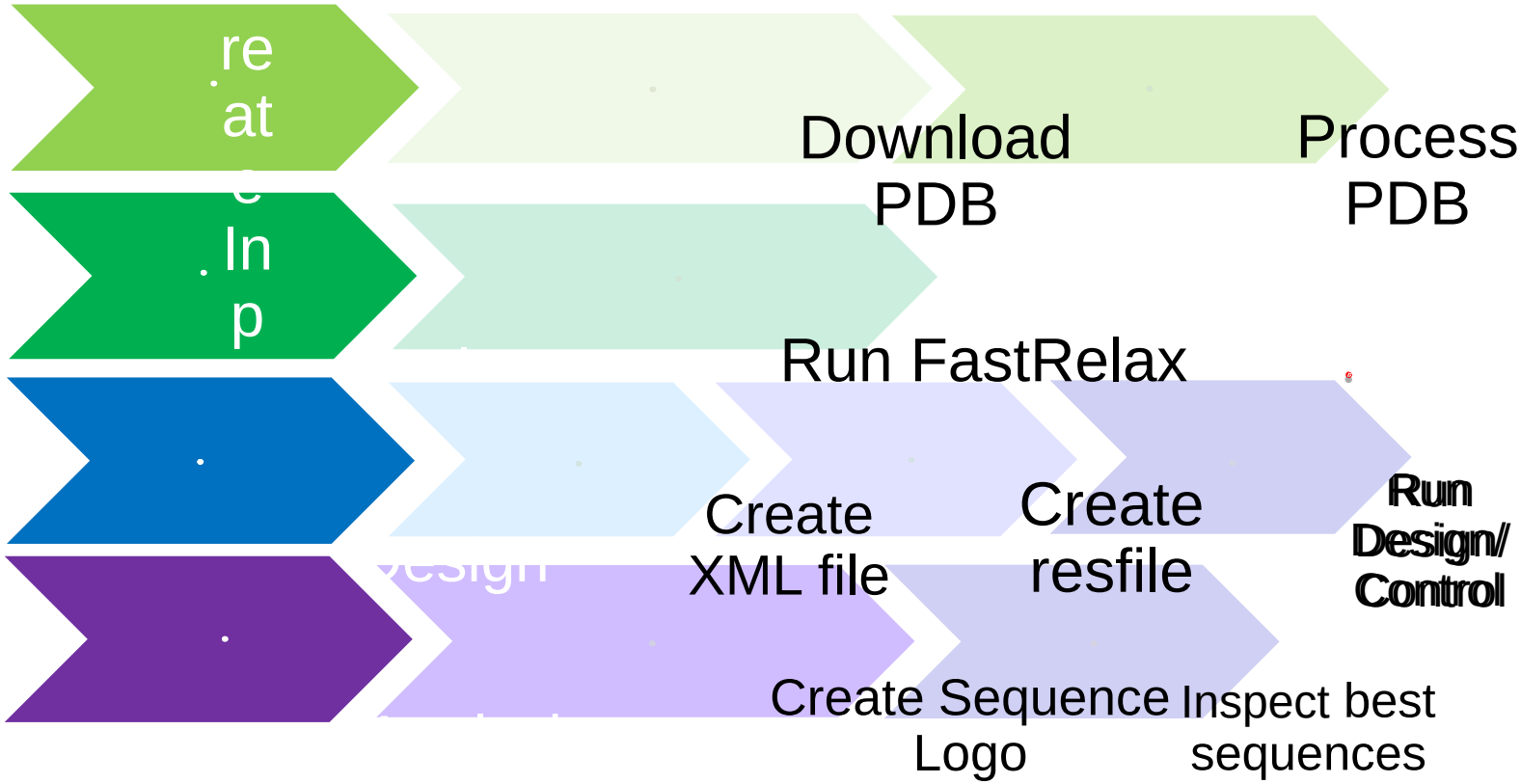
PDB ID 4yk4



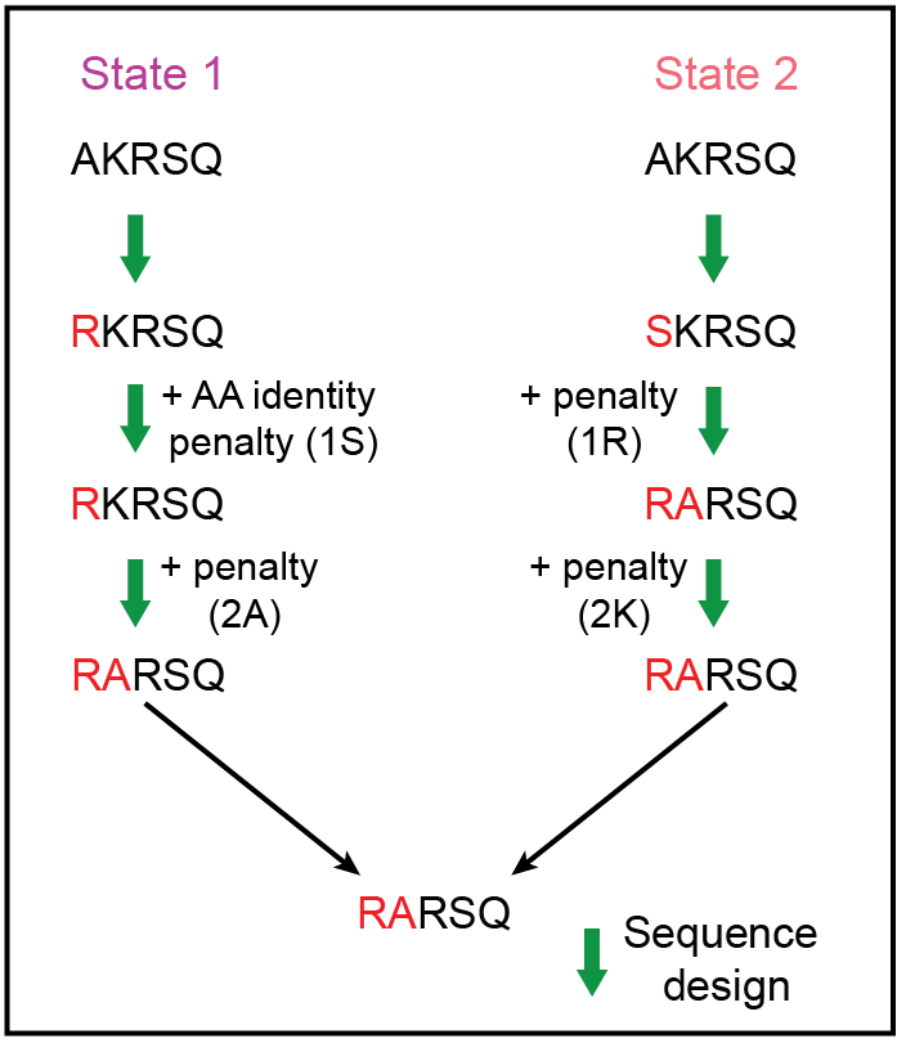
Still the Same Antibody

Antigen 2

Single-state design protocol overview



REstrained CONvergence in MSD (RECON)



Sevy, A. M., Jacobs, T. M., Crowe, J. E. & Meiler, J. *PLoS Comput. Biol.* **11,** e1004300 (2015).

Create XML file

Please open
protein_design/multi_state_design/input_files/design.xml

```
<PROTOCOLS>
    Run four rounds of design
    <Add mover=msd1 />
    <Add mover=msd2 />
    <Add mover=msd3 />
    <Add mover=msd4 />

    Find a consensus sequence
     <Add mover=finish />

    Calculate interface metrics
     <Add mover=analyze />
</PROTOCOLS>
```

Multiple design operations with gradually forcing the design to one consensus sequence

- Differ in constraint weight

Agree on the final consensus sequence (if yet unclear)

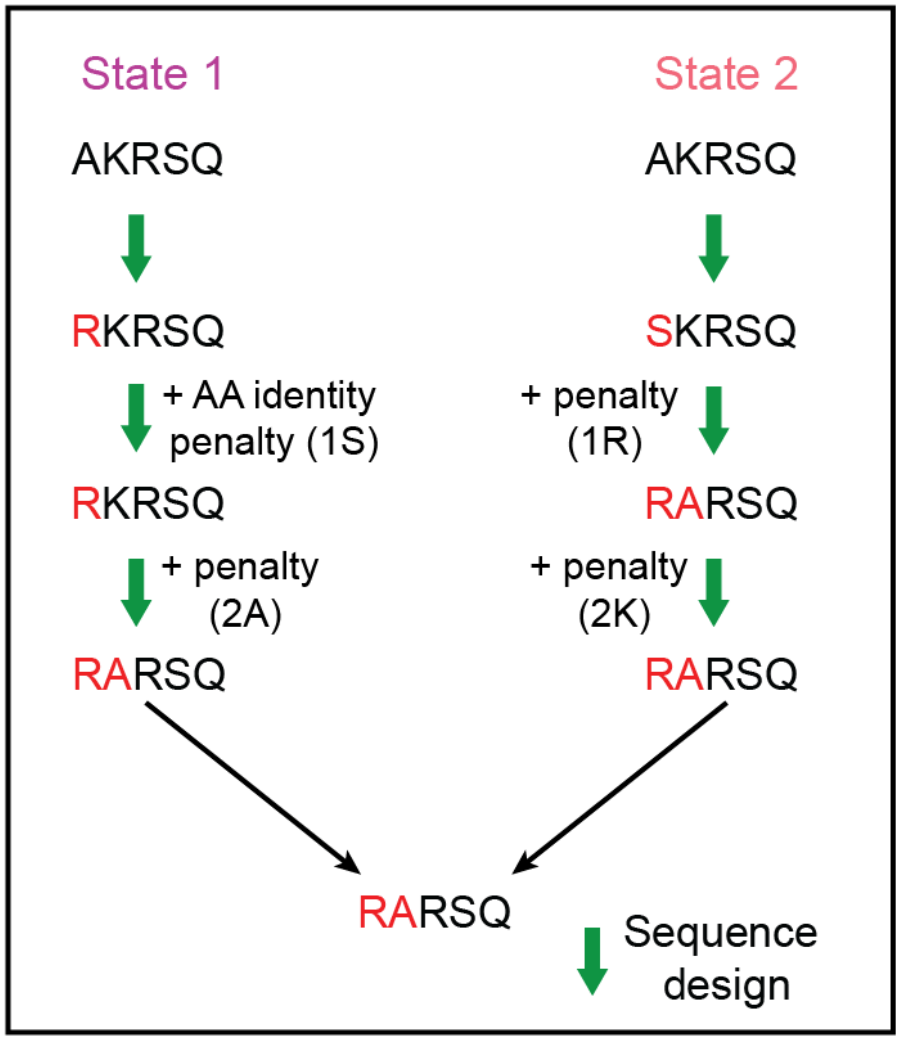
REstrained CONvergence in MSD (RECON)

MSDMover 1

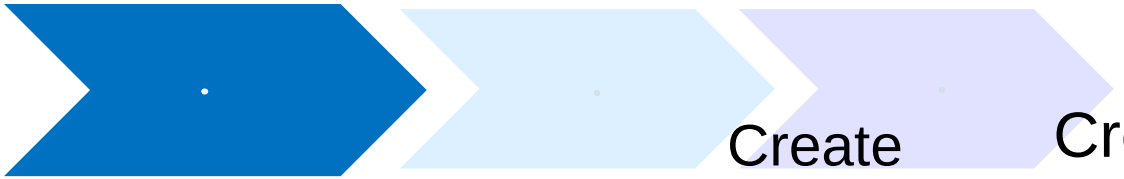
MSDMover 2

MSDMover 3

FindConsensusSequence



Sevy, A. M., Jacobs, T. M., Crowe, J. E. & Meiler, J. *PLoS Comput. Biol.* **11,** e1004300 (2015).



State 1		State 2	(State 3 ...)
NATRO		NATRO	
start		start	
30 H ALLAA	⟷	30 H ALLAA	
31 H ALLAA	⟷	31 H ALLAA	
…		…	
152 L ALLAA	⟷	152 L ALLAA	
155 L ALLAA	⟷	156 L ALLAA	
…		…	
332 A NATAA	⟷	332 A NATAA	
334 A NATAA	↘	333 A NATAA	
		334 A NATAA	

Designed residues must match 1-to-1 … but structurally, not numerically

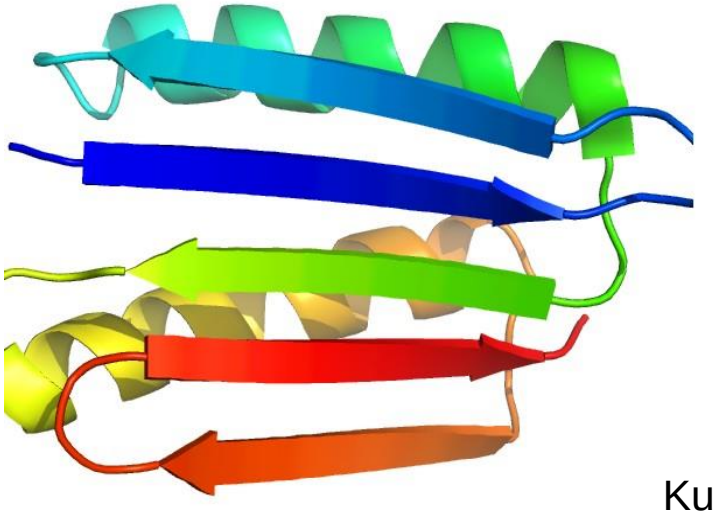
States can differ in the number of residues being repacked

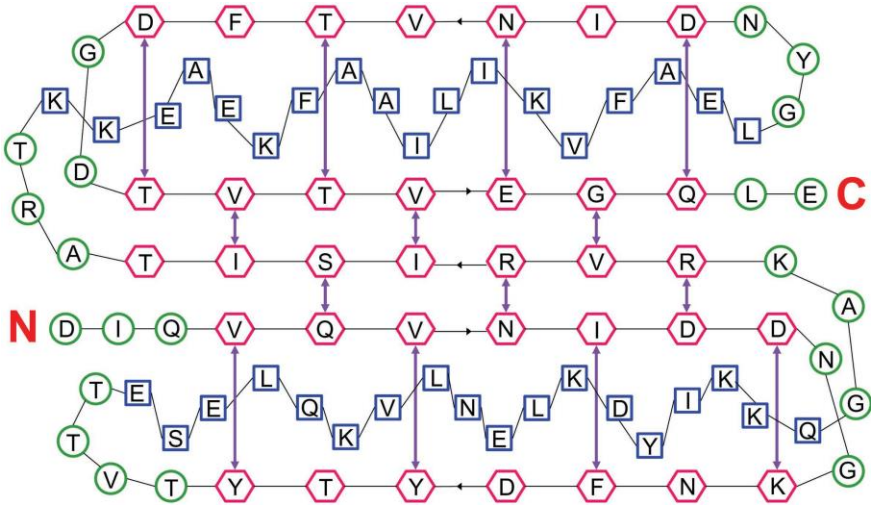
Rosetta Protein Design Applications

De Novo Design of a Novel Fold

Top7 “back-of-the envelope” drawn topology not found in the PDB at time of design

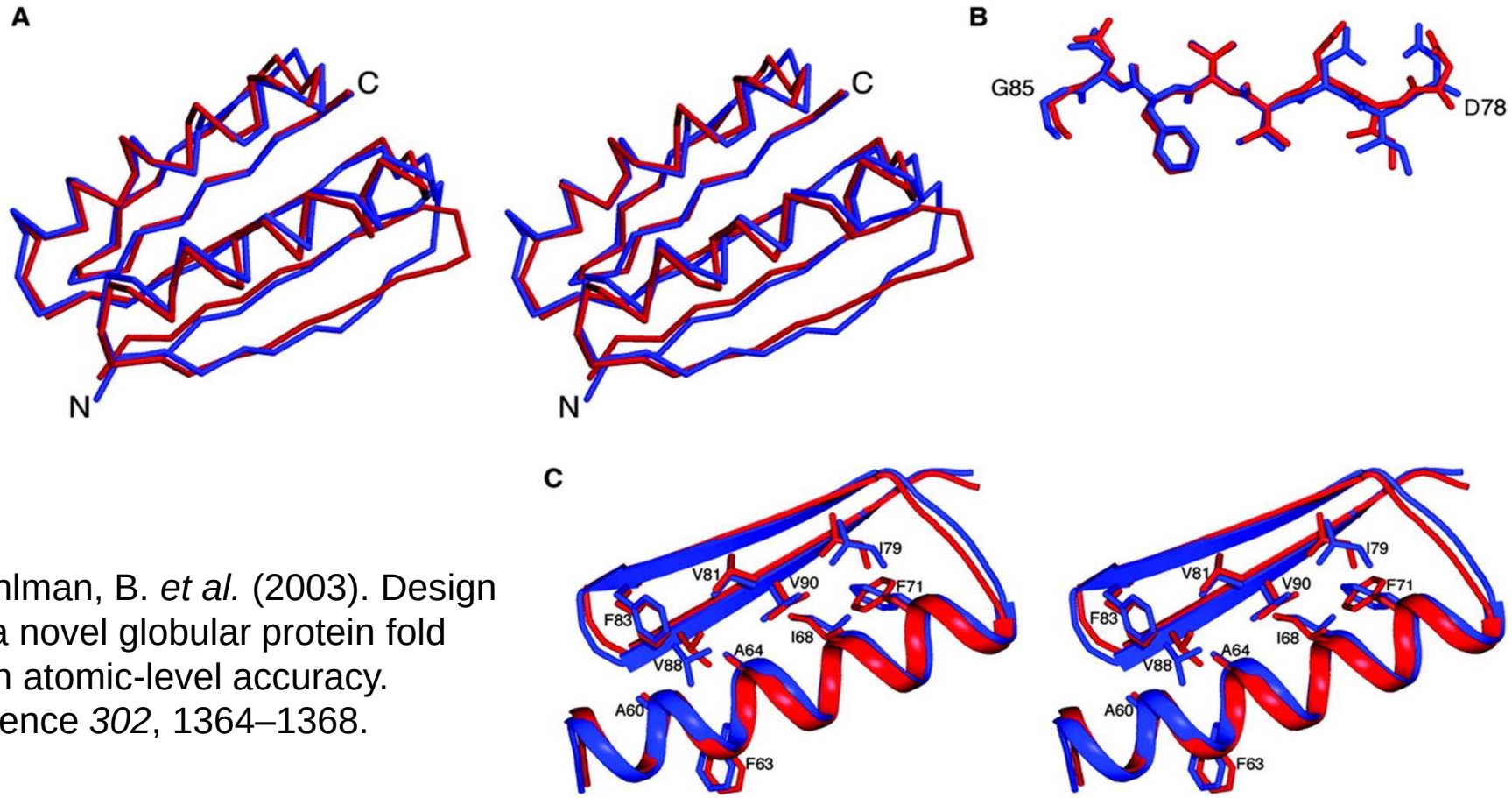
Iterative fixed backbone design + backbone perturbations





Kuhlman, B. *et al.* (2003). Design of a novel globular protein fold with atomic-level accuracy. Science *302*, 1364–1368.

Atomic Level Accuracy of Design (blue) to X-ray structure (red)



Kuhlman, B. *et al.* (2003). Design of a novel globular protein fold with atomic-level accuracy. Science *302*, 1364–1368.

Design of Protein-Ligand Interfaces

RosettaMatch to identify stable backbone

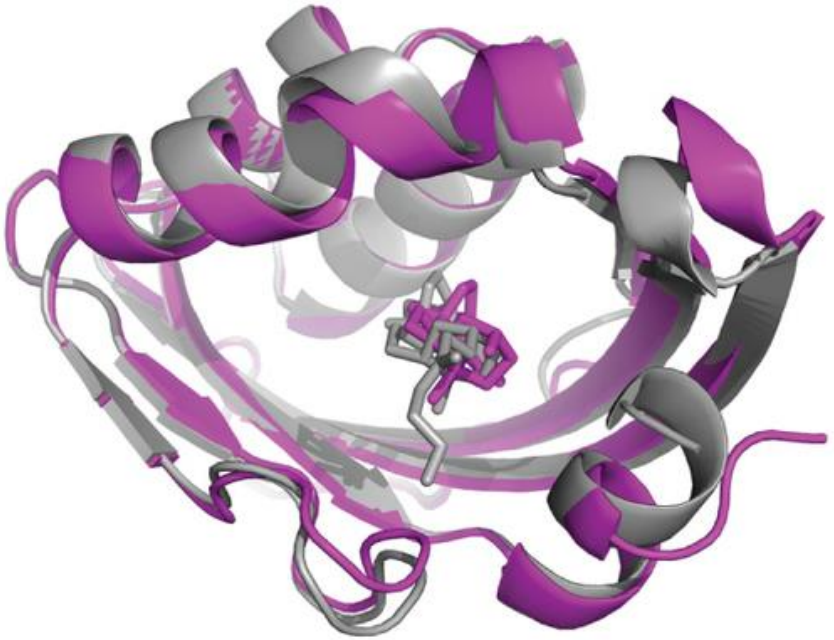
Sequence Design:

Round 1 = maximize binding affinity for ligand

Round 2 = protein stabilization

Computational model = grey

X-ray structure = purple



Tinberg, C.E., Khare, S.D., Dou, J., Doyle, L., Nelson, J.W., Schena, A., Jankowski, W., Kalodimos, C.G., Johnsson, K., Stoddard, B.L., Baker, D. (2013). Computational design of ligand-binding proteins with high affinity and selectivity. Nature *501* 212-216

Additional Design Applications

- **Novel Enzyme Design – RosettaMatch and RosettaDesign**

Siegel, J.B. *et al*. (2010). Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science *329*, 309–313

- **Novel Protein Therapeutic Design**

Fleishman, S.J. *et al.* (2011). Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science *332*, 816–821.

- **Design of a thermally stabilized enzyme**

Korkegian, A., Black, M.E., Baker, D., and Stoddard, B.L. (2005). Computational thermostabilization of an enzyme. Science *308*, 857–860.

- **Design of self-assembling proteins as nanomaterials**

King, N.P., Sheffler, W., Sawaya, M.R., Vollmar, B.S., Sumida, J.P., Andre, I., Gonen, T., Yeates, T.O., Baker, D. (2012). Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. Science *336* 1171-1174

Additional Design Applications

•**Design of symmetric superfolds to understand protein folding evolution**
Fortenberry, C. *et al.* (2011). Exploring symmetry as an avenue to the computational design of large protein domains. J. Am. Chem. Soc. *133*, 18026–18029.

•**Rational epitope design**
Wu, X., et al. (2010). Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. Science *329*, 856– 861.

•**Rational vaccine design**
Jardine, J., et al. (2013). Rational HIV Immunogen Design to Target Specific Germline B Cell Receptors. Science.