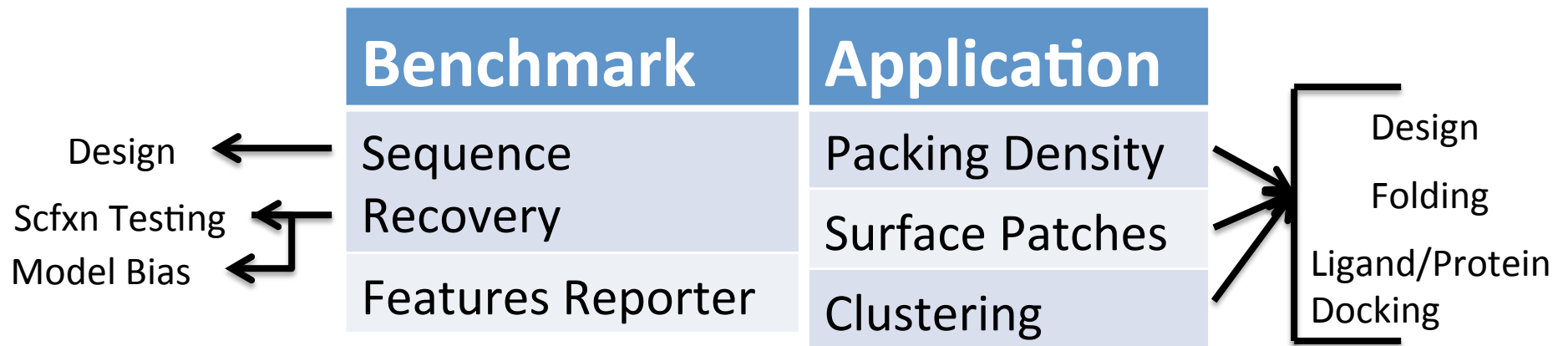


Model Validation Using a Set of Tools

Steven Combs

Model Validation Is Split into Two Classes



Look at your models!!!!!!!!!!!!!!

Sequence Recovery Identifies Problems with Score Functions

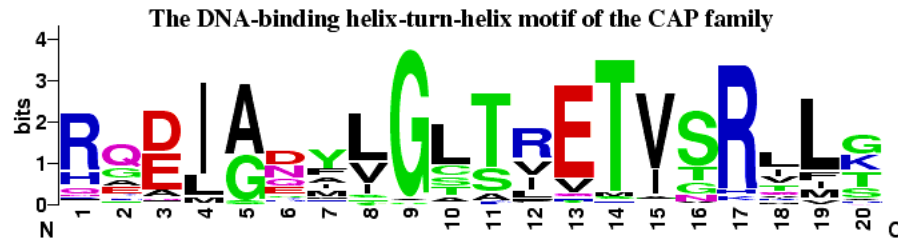
- Stable fold is close to optimal
- May only be important in benchmark studies
- Goal is to be close to native with enough divergence

```
sequence_recovery.default.linuxgccrelease -native_pdb_list  
<native_list> -redesign_pdb_list <designed_pdb_list>
```

Scorefunction	Core	Surface	Overall
Score12' no relax	52.1%	25.2%	36.0%
Score12' relax	66.0%	34.6%	47.6%

Native protein sequences are close to optimal for their structures Brian Kuhlman and David Baker 2000 PNAS

Sequence Logos Identify Sequence Bias in Design



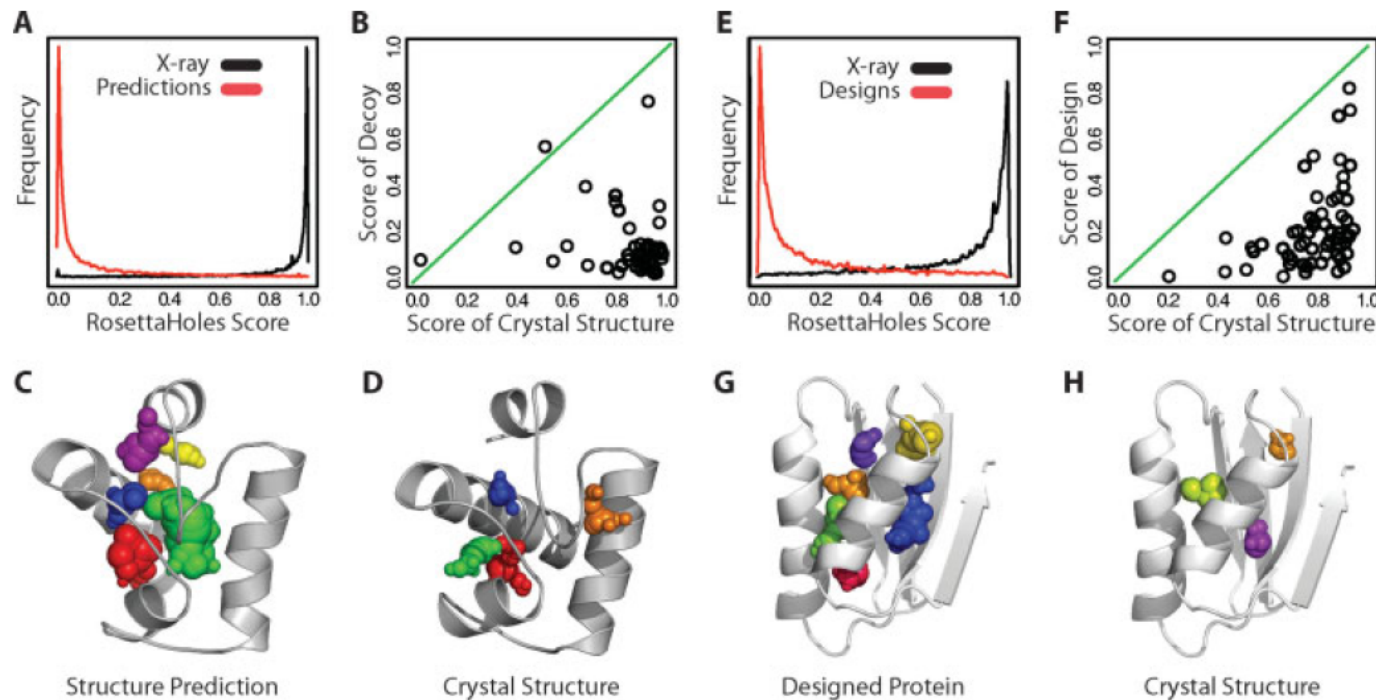
- Identify conserved regions
- More precise description than a consensus sequence
- Height indicates sequence conservation
 - Height within stack indicates relative frequency

Packstats Quantitates Packing Density

- Packstats = packing density
 - Range 0 to 1
 - Average ~0.80

Scorefunction	Packstats
Score12' no relax	0.53
Score12' relax	0.59

`packstats.default.linuxgccrelease -s <pdb_file>`

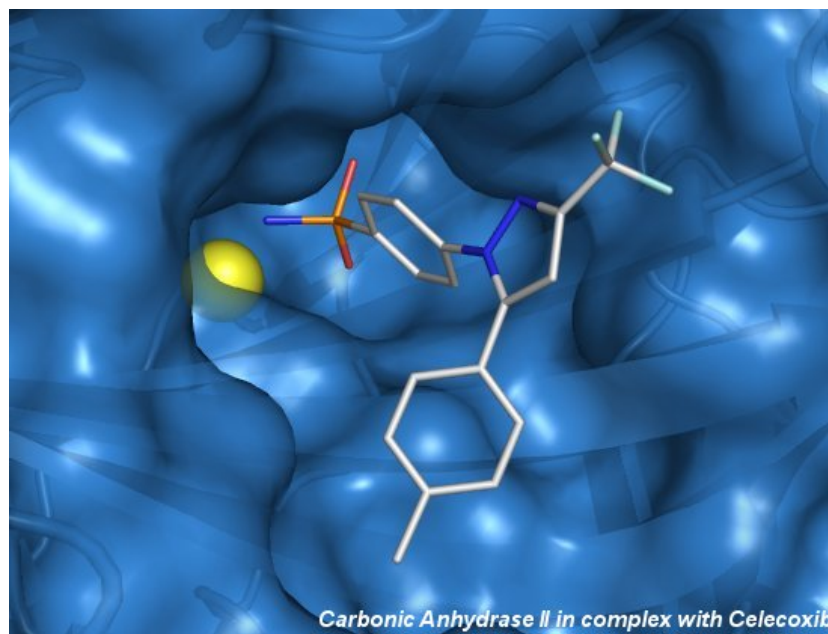


Sheffler, W.; Baker, D. *Protein Science* **2009**, *18*(1), 229-39.

Interface Analyzer Identifies Packing and Unsatisfied Hbonds

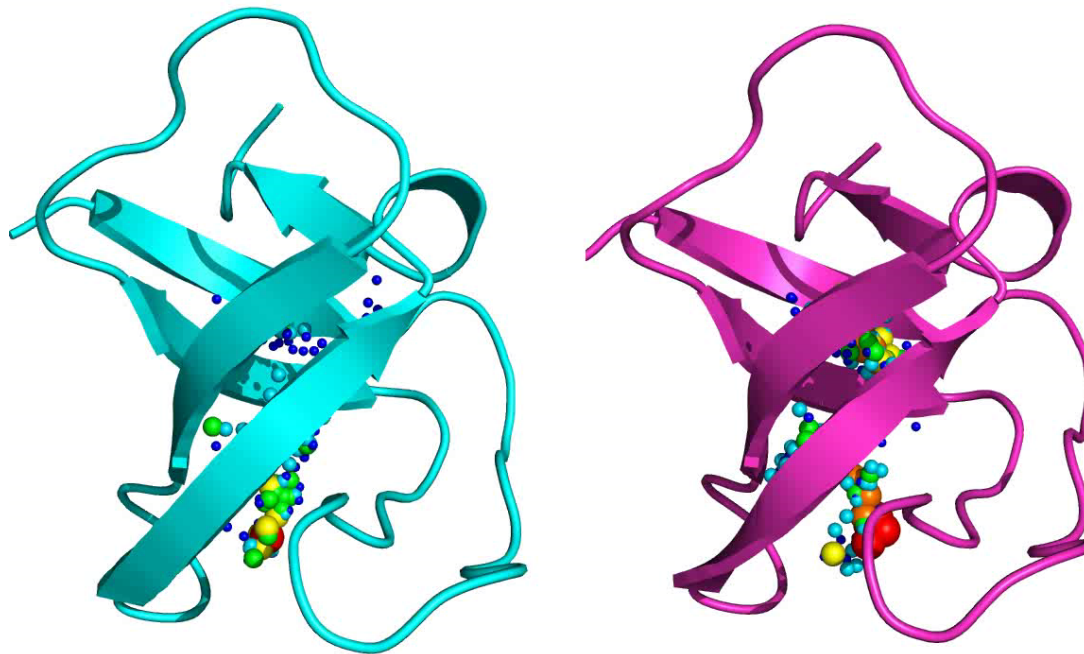
- Ligand/Protein – Protein docking
- Specific to interface

`InterfaceAnalyzer.default.linuxgccrelease -database rosetta_database/ -l designed -compute_packstat -overwrite -out:file:score_only interface_analyzer.txt`



RosettaHoles Is a Visual Tool for Packstats

- Developed for model verification
- Allows for visualization of voids created from poor packing of residues



Steps to get RosettaHoles to Work

- Email Will Sheffler and ask him to include the application
- Download Dalphaballs
 - https://simtk.org/project/xml/downloads.xml?group_id=212
- Command: `holes.default.linuxgccrelease -l pdb.list -make_pdb -make_voids -holes:dalphaball DAlphaBall/src/DAlphaBall.icc -out:[file:silent_struct_type] binary`
- Use pymol to visualize the holes
 - `pymol pymol_rosettaholes.py <holes_generated_pdb>`
 - In Pymol type:
 - `showpacking`
 - `select cavities, resn cav`
 - `color bfactor`

A Large Scale Redesign of Proteins

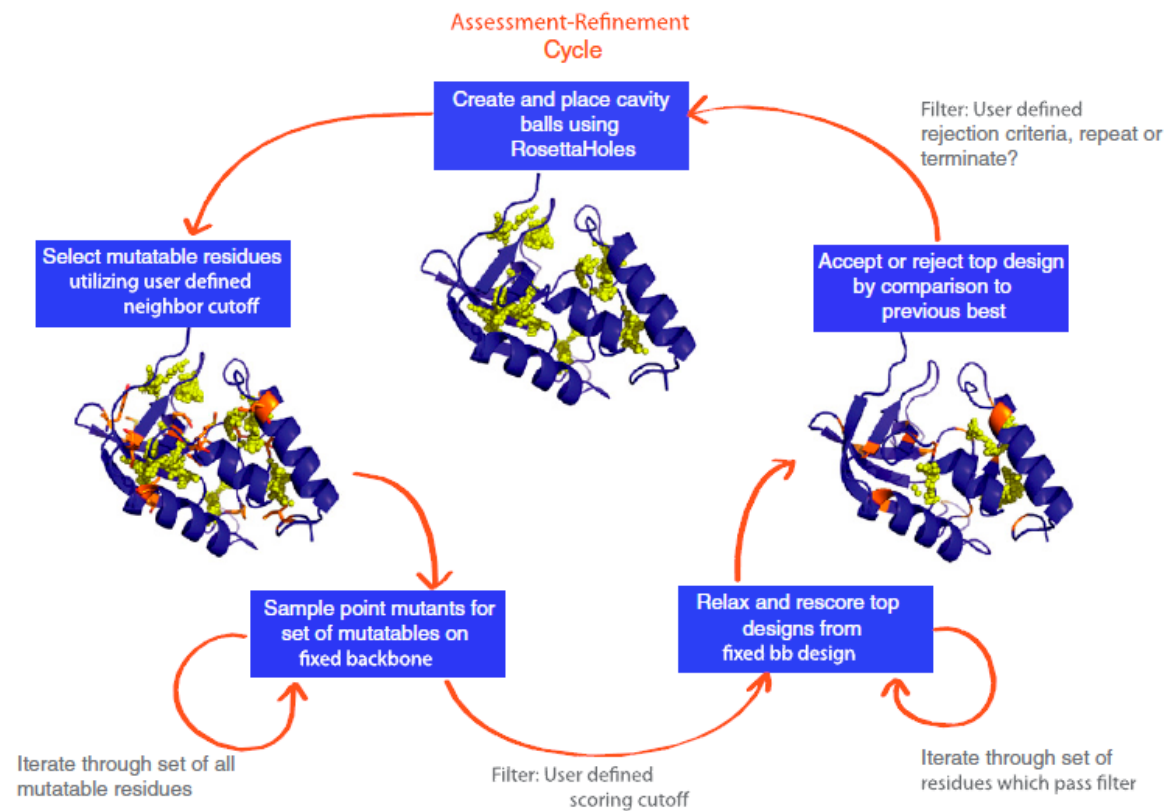
- Dantas et. al redesigned 9 proteins in Rosetta
 - 6 of the 9 proteins were successful (folded)
 - 3 of the proteins were unsuccessful (aggregated, unfolded, destabilized)

Table 2. Summary of experimental results

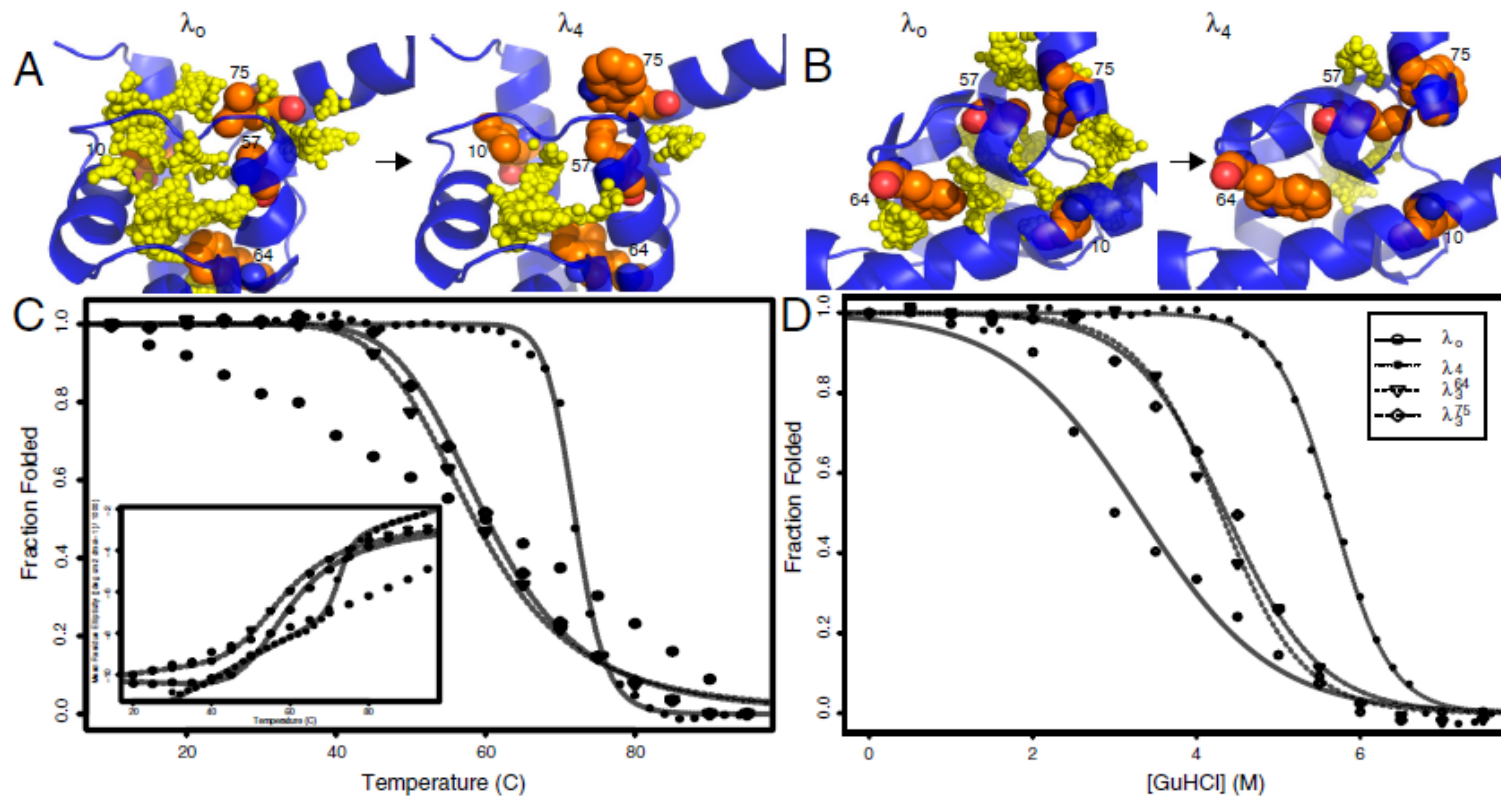
Redesigned proteins	CD spectra	Gu-HCl melt	Temperature melt	1D ¹ H NMR	Verdict
src SH3	Random-coil	Non-cooperative	Non-cooperative	Sharp-lines; no dispersion	Unfolded
Tenascin	β -Sheet like WT	Aggregated unable to determine (UTD)			Aggregated
λ -Repressor	α -Helical like WT	Cooperative stability < WT	Non-cooperative	Broad-lines; weak dispersion	Destabilised less rigid
Acylphosphatase	α/β -Like WT	Cooperative stability = WT	Cooperative $T_m > WT$	Broad-lines; strong dispersion	Stable less rigid
Immunophilin FKBP12	α/β -Like WT	Cooperative stability = WT	UTD	UTD	Stable multimeric
Ribosomal S6	α/β -Like WT	Does not denature	UTD	UTD	Stabilised multimeric
Protein L 1	α/β -Like WT	Cooperative stability < WT	Cooperative $T_m > WT$	Sharp lines; strong dispersion	Destabilised well-folded
Protein L 2	α/β -Like WT	Cooperative stability = WT	Cooperative $T_m > WT$	Sharp lines; strong dispersion	Stable well-folded
RNA-binding U1A	α/β -Like WT	Cooperative stability > WT	Cooperative $T_m > WT$	Sharp lines; strong dispersion	Stabilised well-folded
Procarboxypeptidase	α/β -Like WT	Cooperative stability > WT	Cooperative $T_m > WT$	Sharp lines; strong dispersion	Stabilised well-folded

Dantas, G., et al., *A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins.* Journal of molecular biology, 2003. **332**(2): p. 449-60

Rescue of Two Failed Designs Via RosettaVIP



Rescue of Two Failed Designs Via RosettaVIP



HotPatch Server Identifies Hydrophobic Patches



HotPatch Web Gateway

Statistical Analysis of Unusual Patches on Protein Surfaces

VERSION 4.0

HotPatch finds unusual patches on the surface of proteins, and computes just how unusual they are (**patch rareness**), and how likely each patch is to be of functional importance (**functional confidence (FC)**). The statistical analysis is done by comparing your protein's surface against the surfaces of a large set of proteins whose functional sites are known.

INPUT: Upload your PDB file as input:
(Maximum uploaded file size: 1.5 MB.)

PROTEIN FUNCTION TYPE: If you already know your protein's function type, you can compare against just a **sub-population** of proteins with that particular type of functional site.

(If you don't know your protein's function type, or you just want to compare it against proteins in general, choose "**Generic**".)

Choose One Protein Function:

☐ Generic Functional Site [gf] info?	☐ Protein Binding (Non-Catalytic) [pb] info?
☐ Catalytic (General) Site [cg] info?	☒ Oligomeric Interfaces [ol] info?
☐ Hydrolases [hy] info?	☐ Metal Ion-Binding Sites [mp] info?
☐ Proteases [pr] info?	☐ Interact w/DNA or RNA (catalytic or not) [dn] info?
☐ Transferases [tr] info?	☐ Interact w/Small Molecules (catalytic or not) [sm] info?
☐ Kinases [ki] info?	☐ Interact w/Lipids (catalytic or not) [lp] info?
☐ Oxidoreductases [or] info?	☐ Interact w/Carbohydrates (catalytic or not) [cm] info?

SURFACE PROPERTIES: HotPatch can find statistically significant patches with high values of particular properties on the surfaces of proteins.

Choose One or More Properties: (If you choose more than one property, you get multiple email responses.)

☒ Neural Network combination info?	☐ Concavity info?
☐ Electrostatic Potential info?	☐ Negative Potential info?
☒ Hydrophobicity info?	☐ Hydrophilicity info?
☐ Surface Roughness info?	

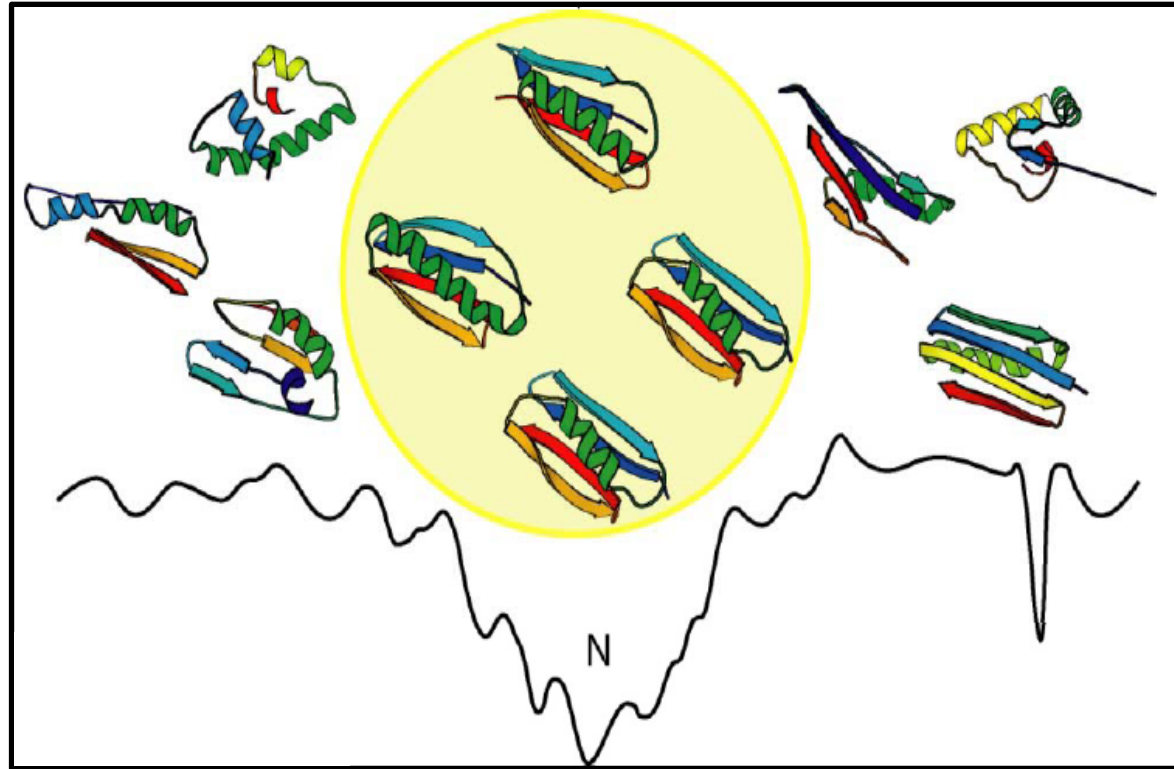
VISUALIZATION SCRIPT: Optionally, HotPatch can also write a script that will display the patches on the structure, when the script is loaded into some common molecular visualization programs. If you want a script, choose what type.

- ☐ None ☐ RasMol
☒ PyMol ☐ DS Viewer (formerly WebLab Viewer)

<http://hotpatch.mbi.ucla.edu/>

Pymol Session Demonstrating Hotpatch

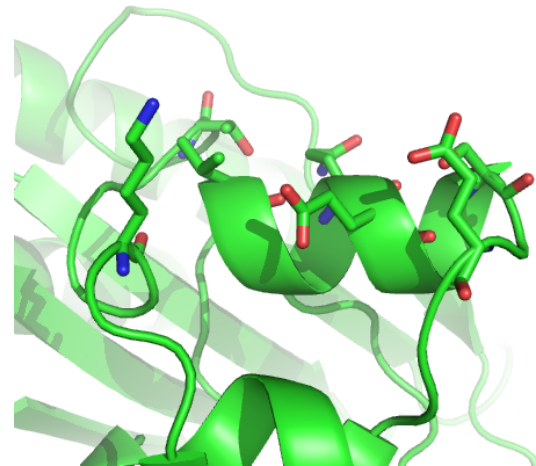
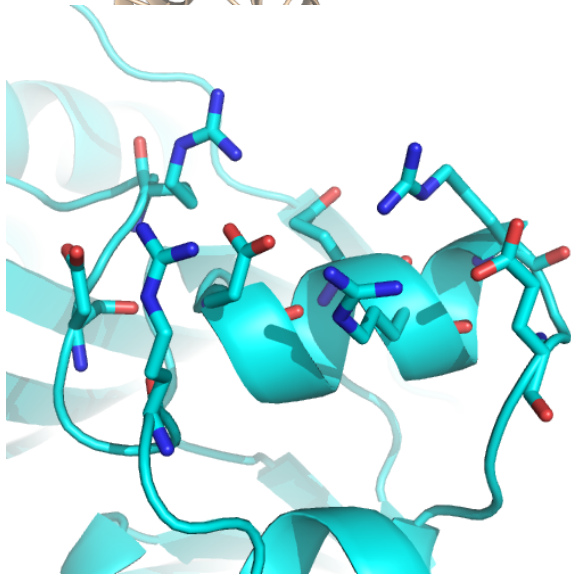
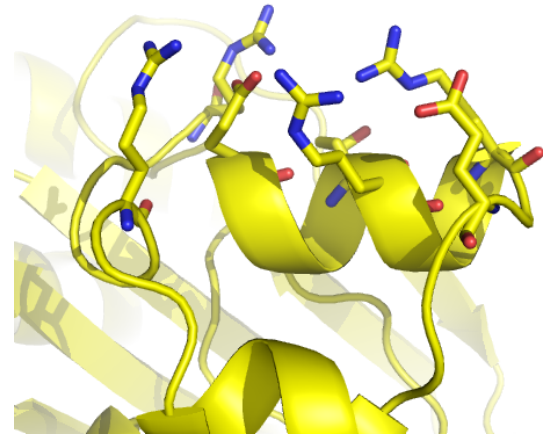
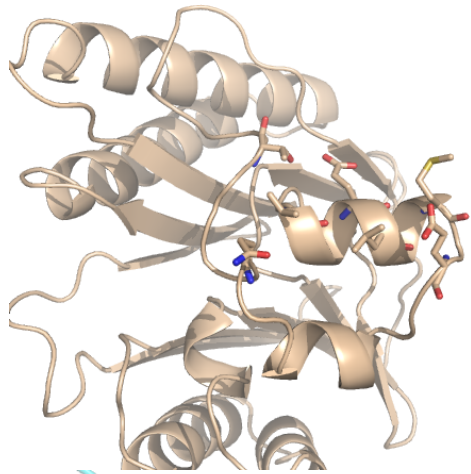
Clustering Alleviates Coarse Score Function Problems



- Rosetta Scoring function is too coarse to sample deep minimum
- Many models have similar folds
- Take largest clusters for analysis

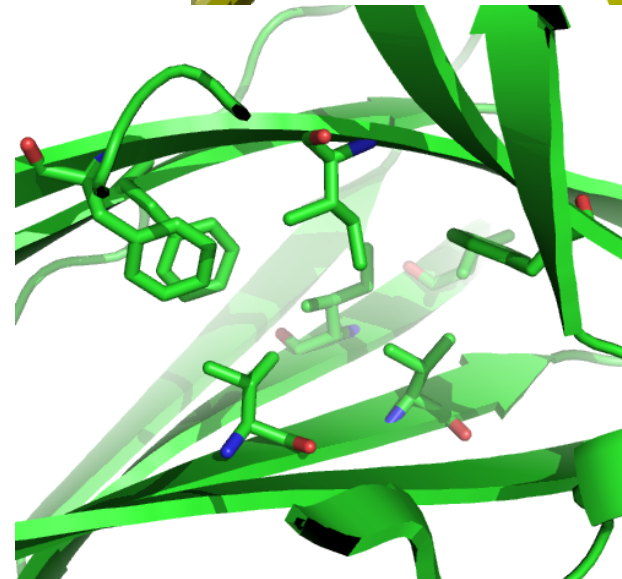
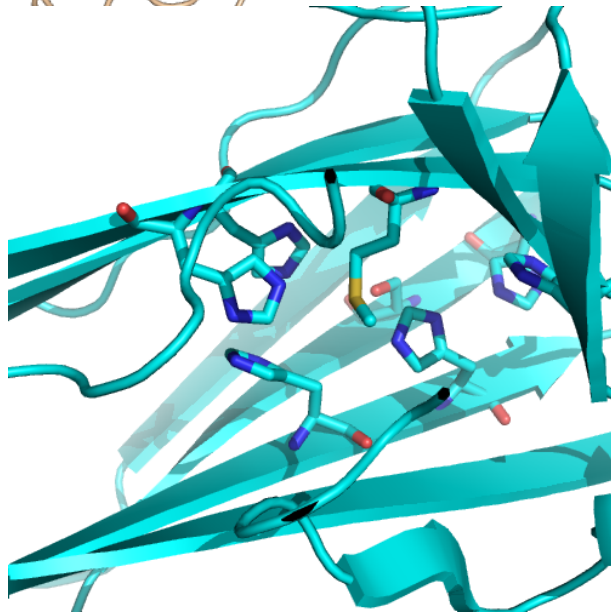
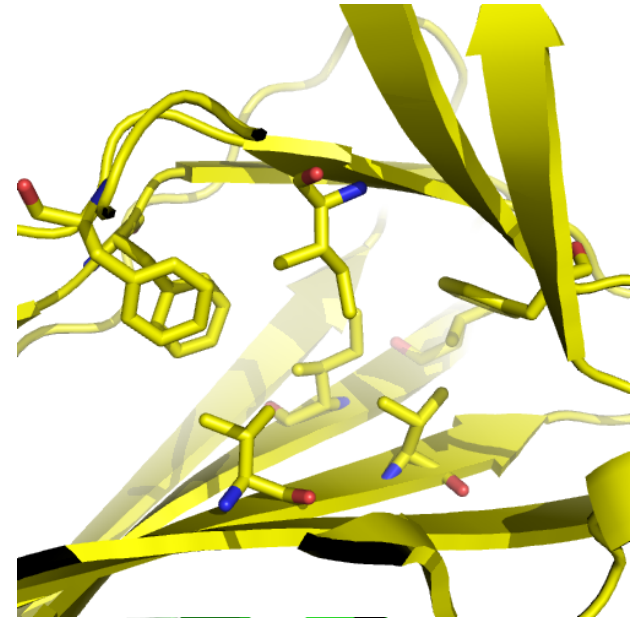
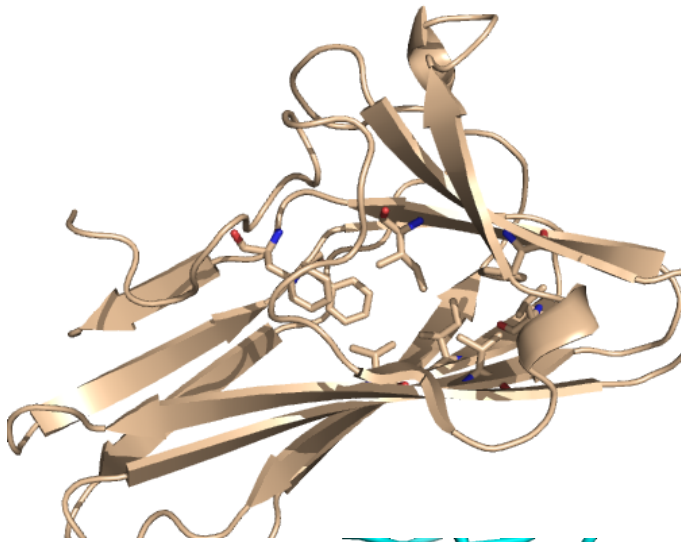
Looking at Models Identifies Problems:

Case 1



Looking at Models Identifies Problems:

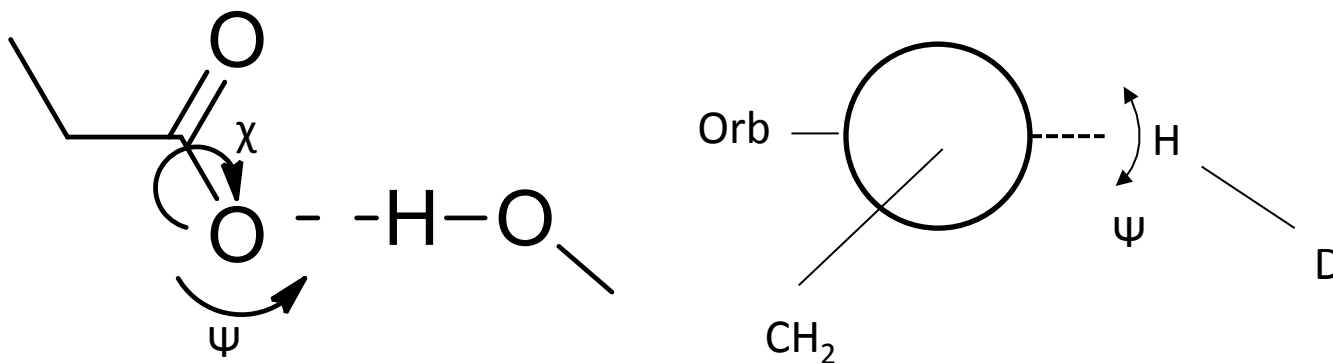
Case 2



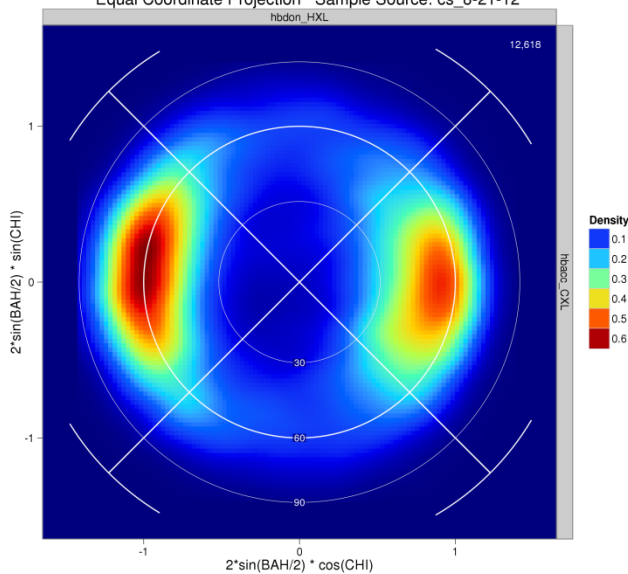
Features Reporters Identify Sampling and Scoring Biases

- Set of related databases
 - Allow for easy statistic generation
 - Residue Type
 - Residue Conformation
 - Number hydrogen bonds, cation-pi, pi-pi, salt bridges
 - Geometrical arrangement of interactions

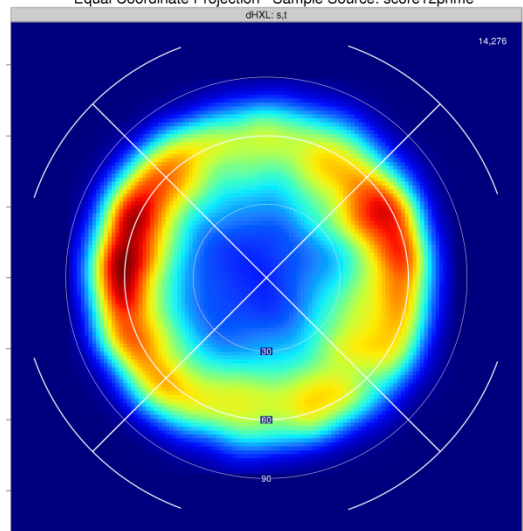
Features Reporters Identify Sampling and Scoring Biases



Hydrogen Bonds chi vs BAH Angles with Sequence Separation > 5
Sidechain Hydroxyl Donors to Sidechain Carboxyl Acceptors
Equal Coordinate Projection Sample Source: cs_8-21-12



HBonds chi vs BAH Angles Sequence Separation > 5 Don:dHXL: s,t Acc:aCXL: d,e
Equal Coordinate Projection Sample Source: score12prime



Hydrogen Bonds chi vs BAH Angles with Sequence Separation > 5
Sidechain Hydroxyl Donors to Sidechain Carboxyl Acceptors
Equal Coordinate Projection Sample Source: ref12_AOH_fix_noOpi

