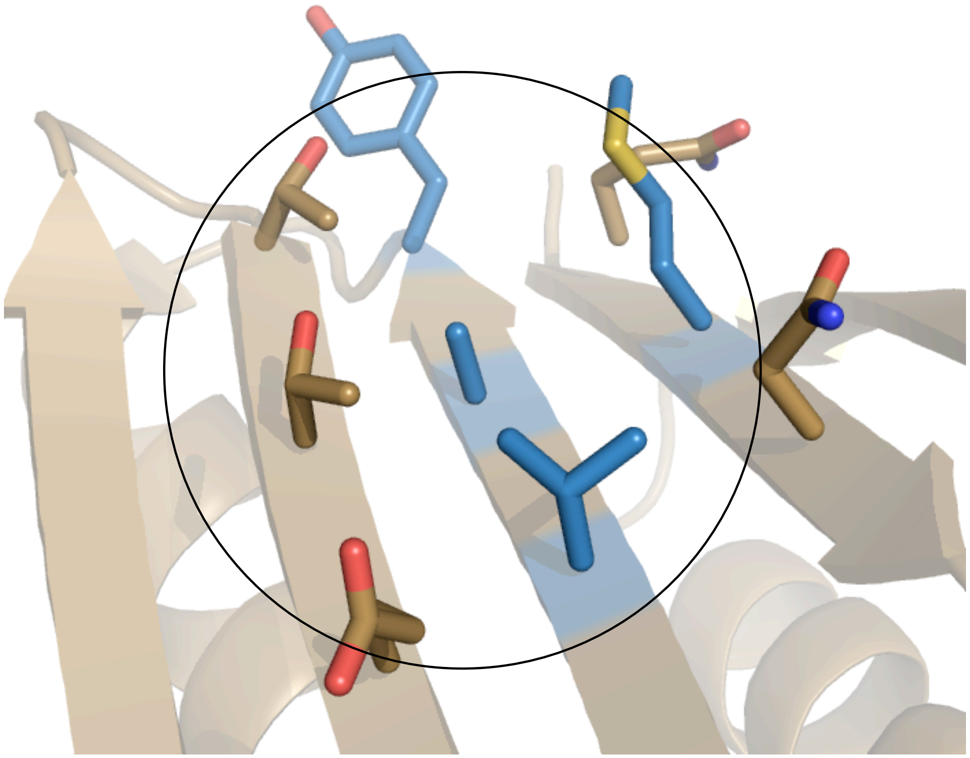
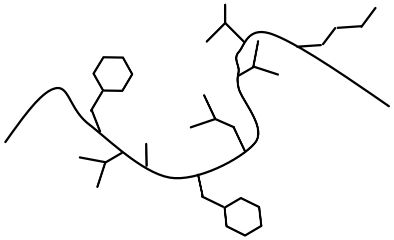


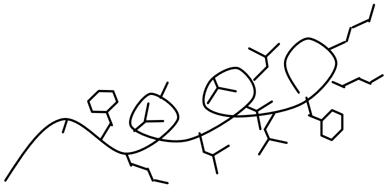
Simultaneously optimizing stability and solubility in protein design simulations

Ron Jacak, Kuhlman Lab
August 5, 2009

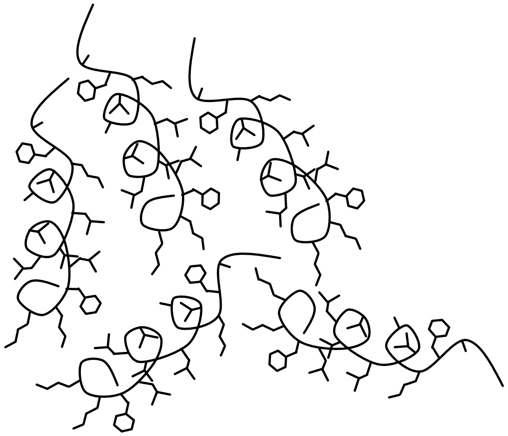
The balance between stability and solubility



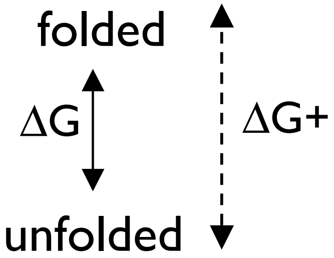
unfolded state



folded (target) state



aggregated state

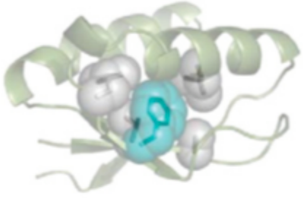


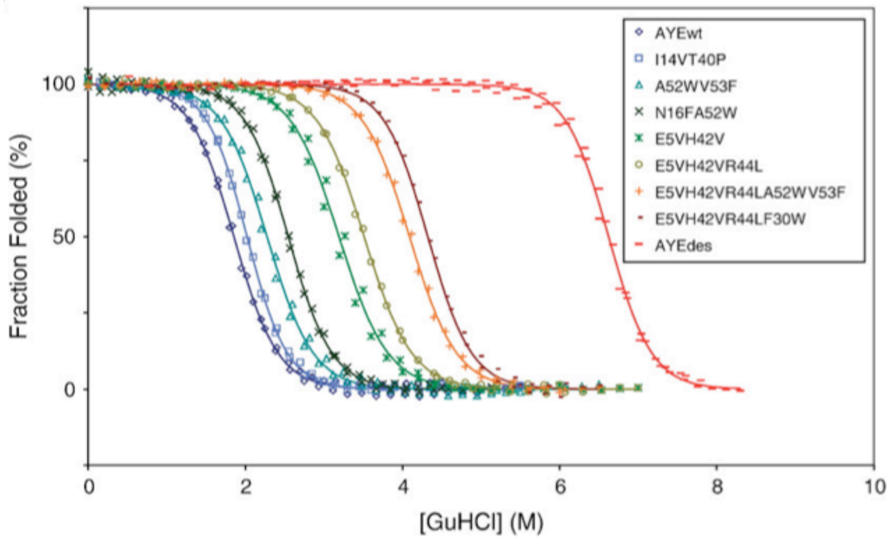
folded

ΔG

ΔG+

unfolded

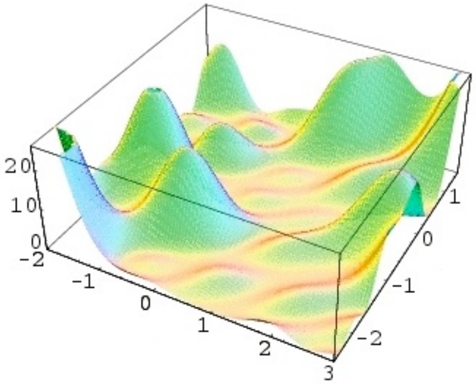




Dantas et al. JMB, 2007

Can the Rosetta energy function be trained to design for stability and solubility?

Optimizing energy function weights



Weights optimized using the optE protocol developed by Andrew Leaver-Fay

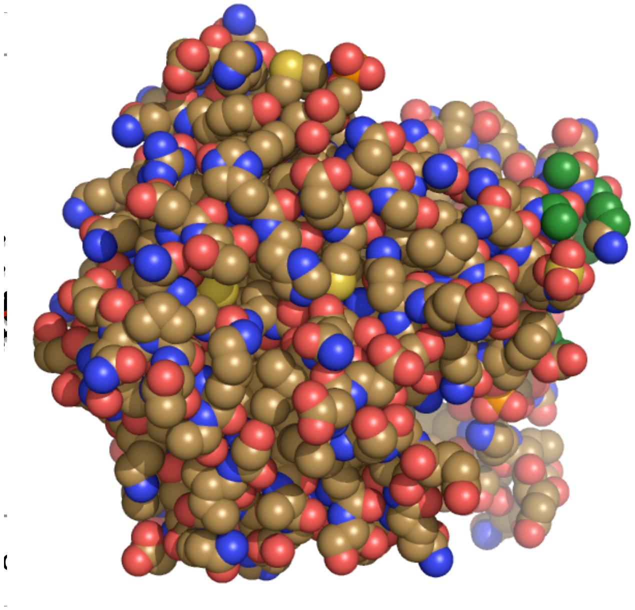
Native sequence recovery dataset:
60 PDBs of various folds

$\Delta\Delta G$ dataset:
Yin et al, Structure 2007, Guerois et al, JMB 2002
Total: 1280 mutants, 70 proteins
Randomly divided into training/test sets of 640 mutants each

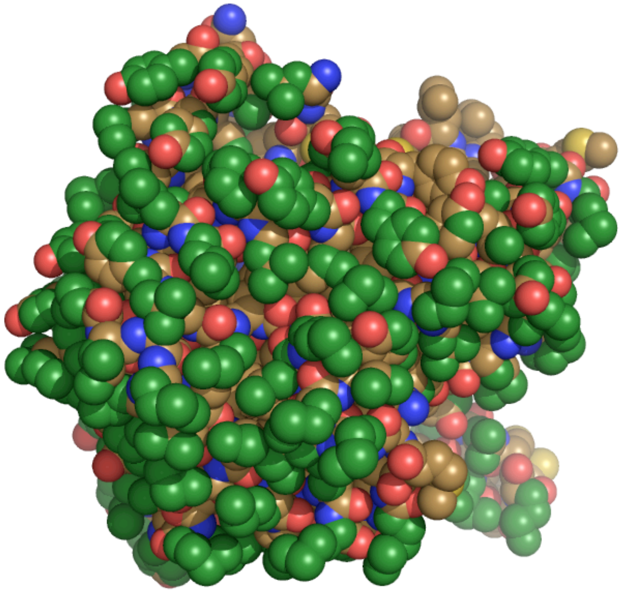
Mutants with $\Delta\Delta G > 10$ excluded; fa_rep capped at 10

Optimizing for predicting changes in stability

1NFP (flavoprotein)



wild type, 1NFP

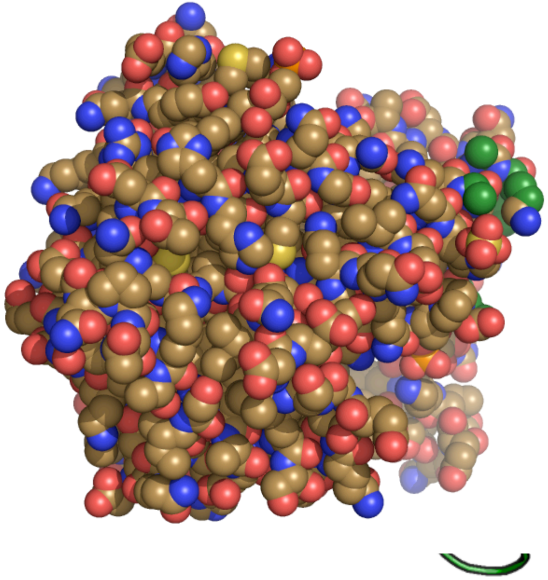


stability-optimized
score12 redesign

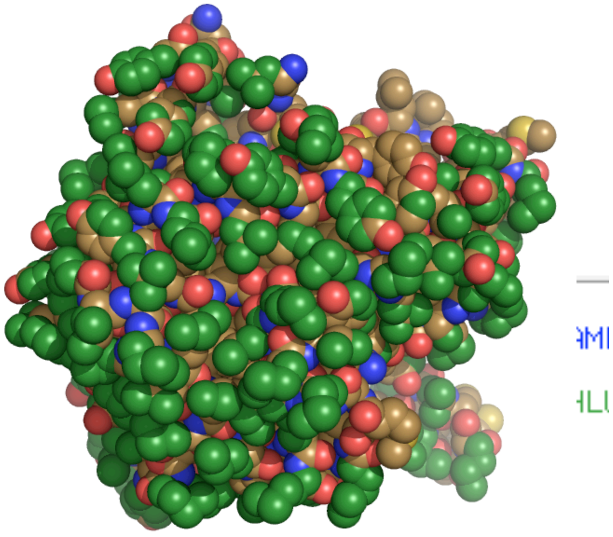
Overall sequence recovery, test set: 16.5%

ddG mutation correlation, R: 0.683

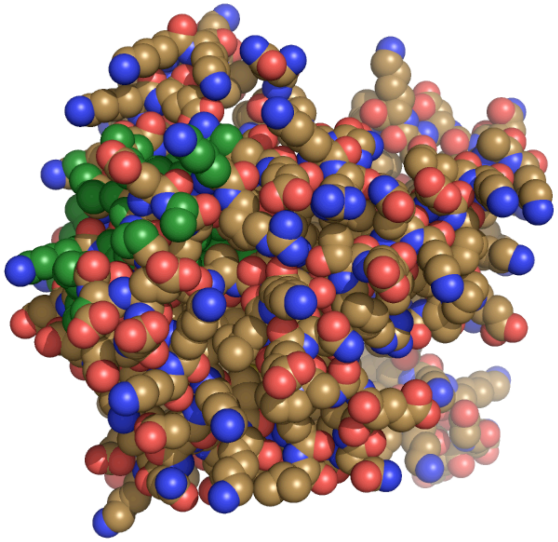
Rosetta trained for native sequence recovery, not stability



wild type, 1NFP

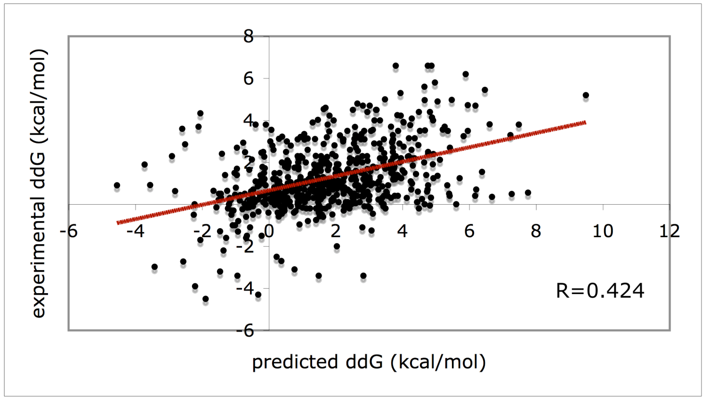


stability-optimized score12 redesign



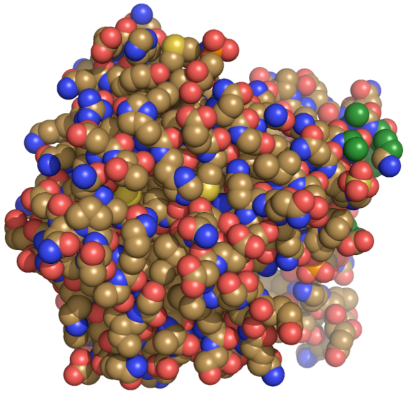
reweighted score12 redesign

ddG mutation correlation, R: 0.424

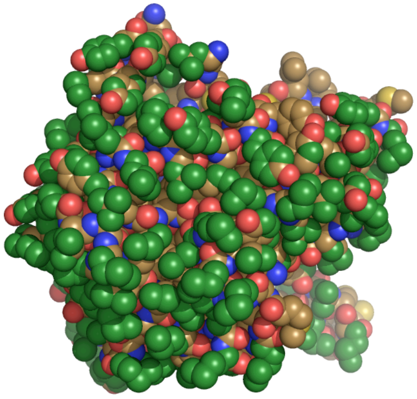


criteria optimized	energy function	test set average sequence recovery (%)	ddG correlation, R	test set average largest patch size (A2)
natives	---	---	---	439
ddG	refE	16.5	0.683	3715
NATAA	refE	37.7	0.424	697

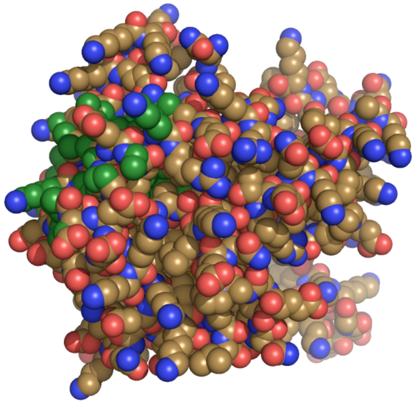
Optimizing for stability and sequence recovery



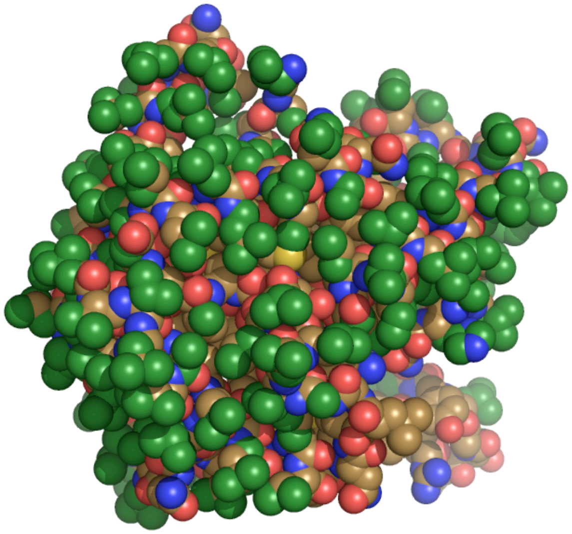
wild type, 1NFP



stability optimized redesign

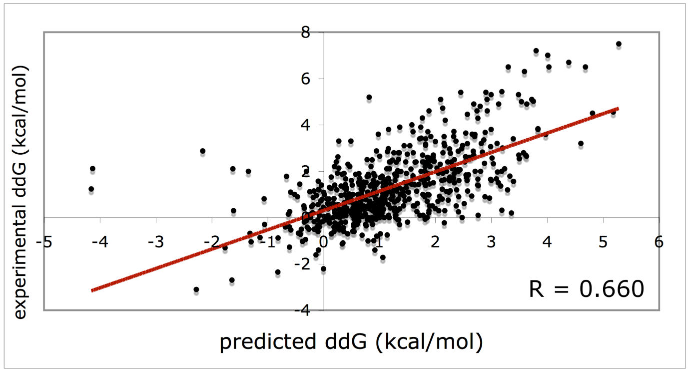


sequence recovery optimized redesign



stability and recovery optimized redesign

ddG mutation correlation, R: 0.660

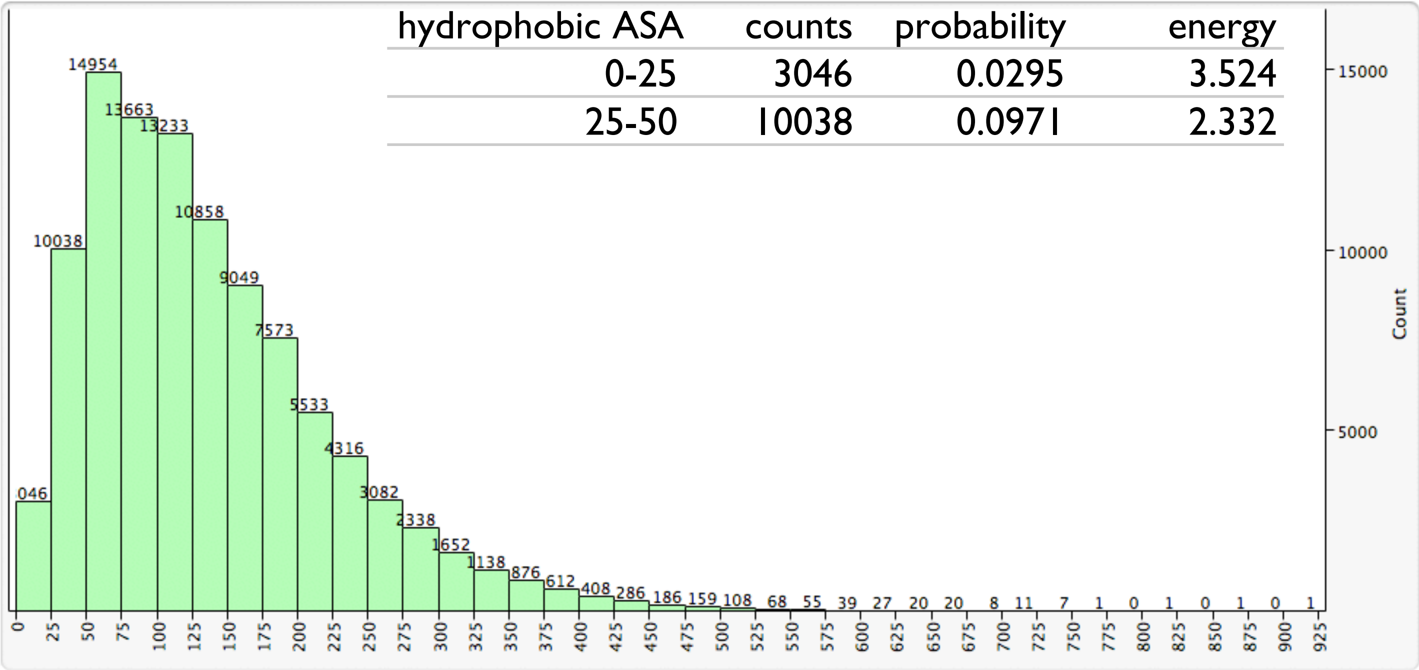


criteria optimized	energy function	test set average sequence recovery (%)	ddG correlation, R	test set average largest patch size (A2)
natives	---	---	---	439
ddG	refE	16.5	0.683	3715
NATAA	refE	37.7	0.424	697
NATAA+ddG	refE	34.1	0.660	1422

How can the Rosetta energy function be changed to prevent hydrophobic patches?

Development of a solubility score for Rosetta

Average exposed hydrophobic surface area within 10Å of a surface residue - “hydrophobic patch” size

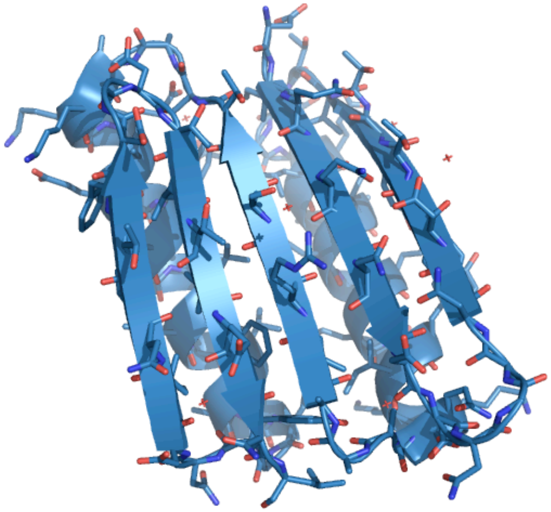


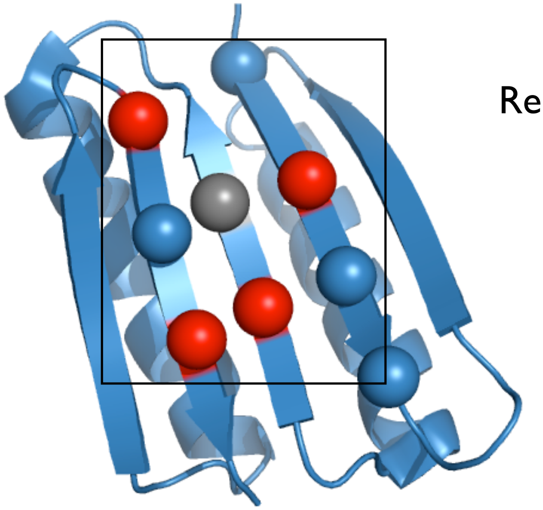
hydrophobic ASA	counts	probability	energy
0-25	3046	0.0295	3.524
25-50	10038	0.0971	2.332

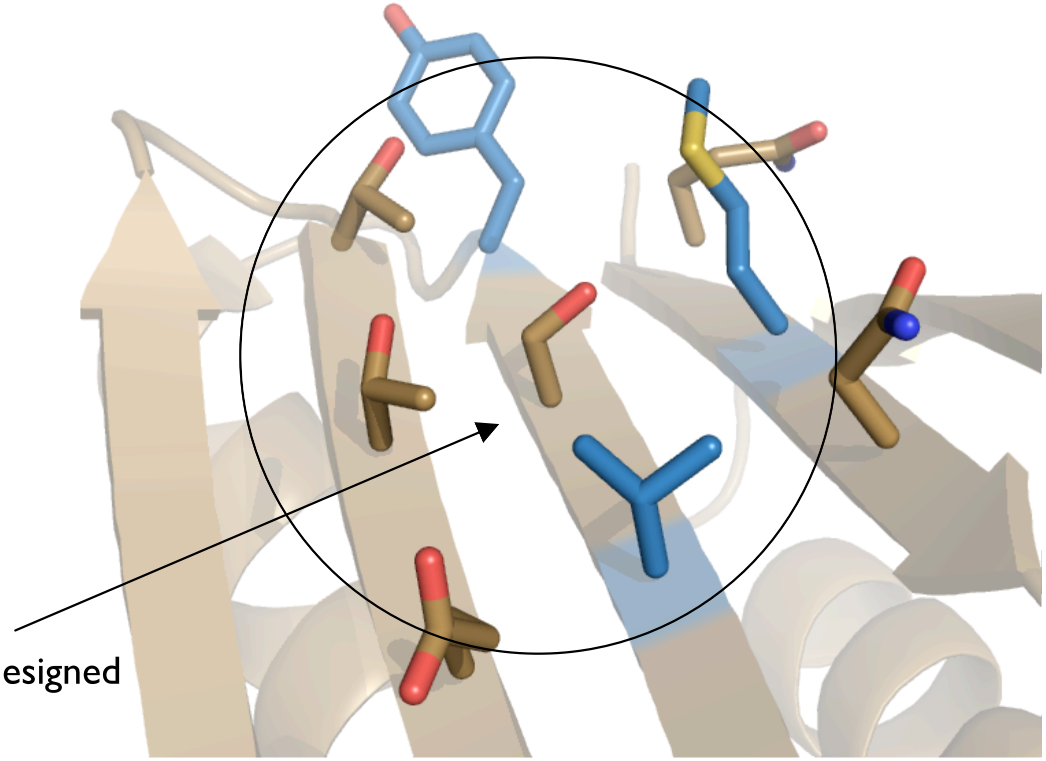
total hydrophobic ASA, set of 1300 monomeric proteins

$$E_{\text{surface}} = \sum_{\text{all surface res}} -\ln[\, P(\text{total hASA within } 10\text{Å})\,]$$

Evaluating changes in surfaceE

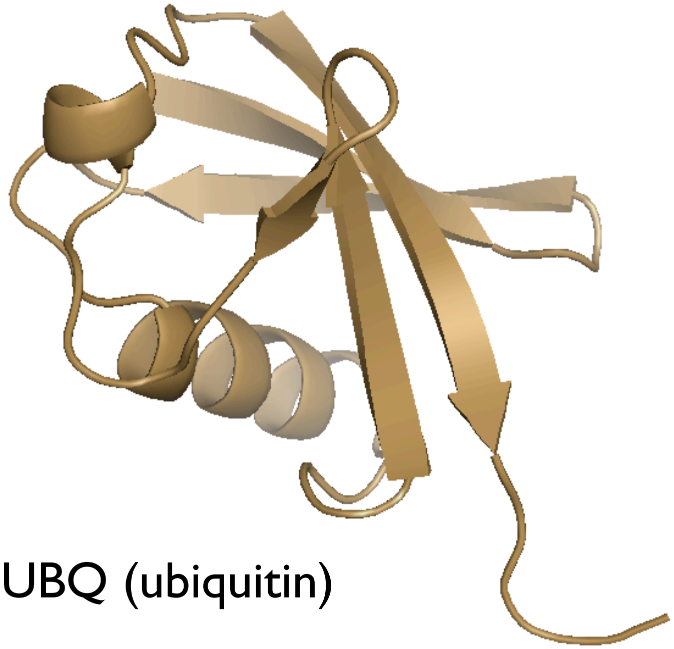






Residue being designed

Redesigns with surfaceE


1UBQ (ubiquitin)

wild type total energy	-137.6
wild type surface energy	117.3

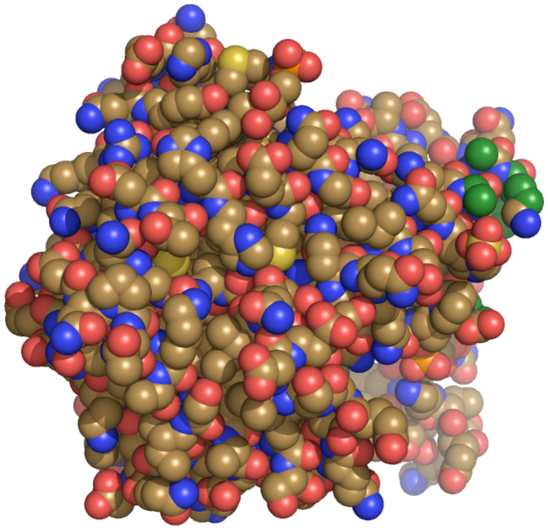
redesigns w/o surfaceE term

total energy	-167.0
surface energy	142.8
run time	17'43"

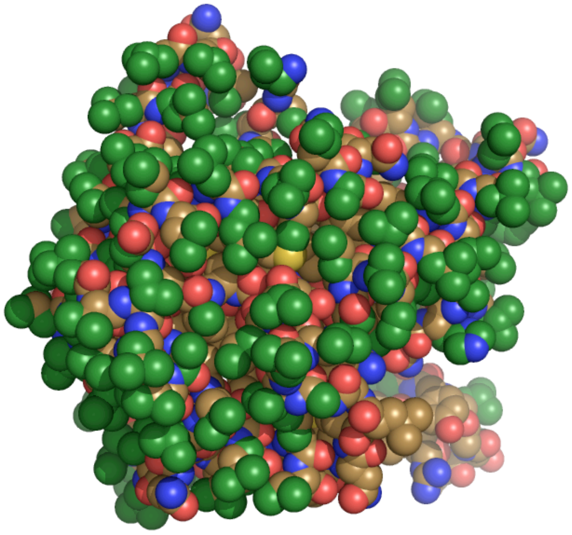
redesigns w surfaceE term

total energy	-158.7
surface energy	101.3
run time	22'30"

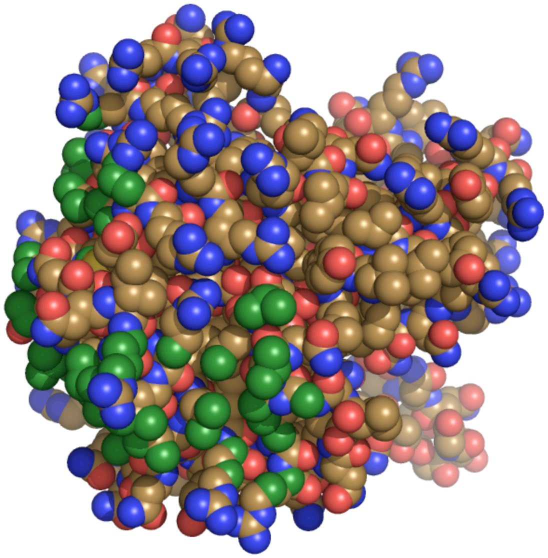
An optimized surfaceE energy function



wild type, 1NFP



stability and recovery optimized redesign



stability and recovery optimized redesign

criteria optimized	energy function	test set average sequence recovery (%)	ddG correlation, R	test set average largest patch size (A2)
natives	---	---	---	439
ddG	refE	16.5	0.683	3715
NATAA	refE	37.7	0.424	697
NATAA+ddG	refE	34.1	0.660	1422
NATAA+ddG	refE + surfaceE	32.7	0.659	451

Conclusion / Future Directions

- Designing with the standard Rosetta energy function gives proteins with large hydrophobic patches on the surface
- Using the surface energy term during design leads to designs with more native-like surfaces
- Native sequence recovery and ddG recovery when using the surfaceE term are comparable to other energy functions

- Moving Rosetta away from reference energies
  - Explicit unfolded state model based on fragments

Acknowledgements

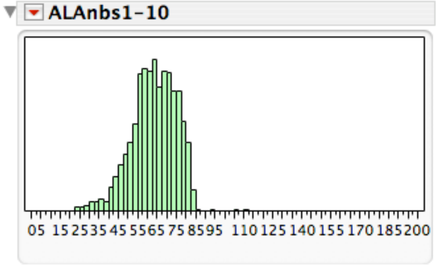
- Brian Kuhlman
- Kuhlman Lab
    - Doug Renfrew
    - Steven Lewis
- Andrew Leaver-Fay
- UNC Research Computing
    - Edward Dale

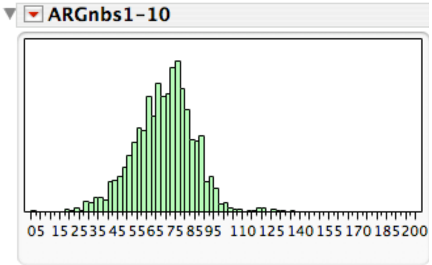
UNC Department of Biochemistry and Biophysics

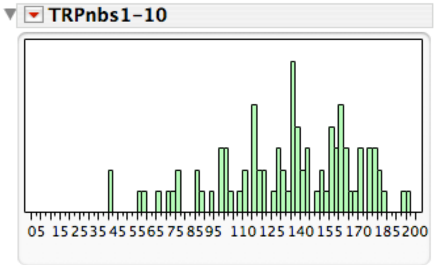
UNC Biophysics Program

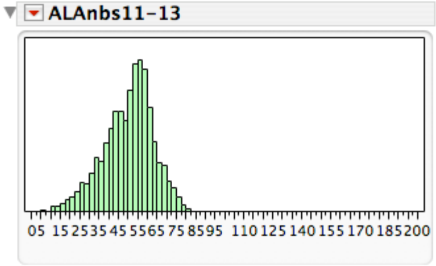
Average residue hydrophobic-ASA

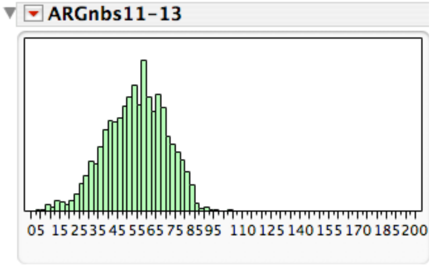
```
REM  Relative accessibilities read from external file "/Users/ronj/Projects/naccess2.1.1/standard.data"
REM  File of summed (Sum) and % (per.) accessibilities for
REM RES _ NUM      All-atoms   Total-Side   Main-Chain    Non-polar    All polar
REM                ABS   REL    ABS   REL    ABS   REL    ABS   REL    ABS   REL
RES PRO A   9   154.71 113.7  81.88  68.3  72.83 448.8  83.28  68.9  71.44 470.3
RES PHE A  10   135.42  67.9 105.01  64.0  30.42  86.0 105.37  63.8  30.05  87.8
RES GLU A  11   143.47  83.3 123.95  92.0  19.52  52.0  51.13  84.8  92.35  82.5
RES LEU A  12    56.83  31.8  52.68  37.3   4.15  11.1  52.68  37.0   4.15  11.4
```

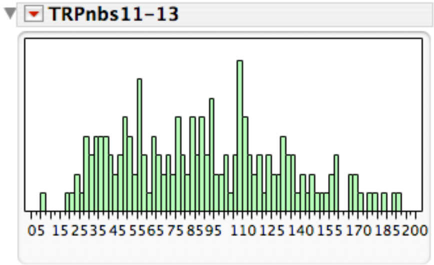


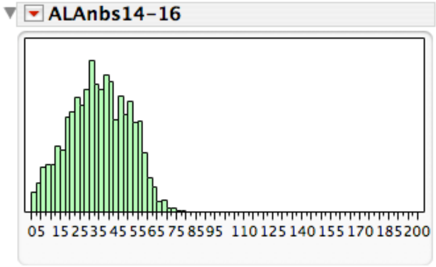


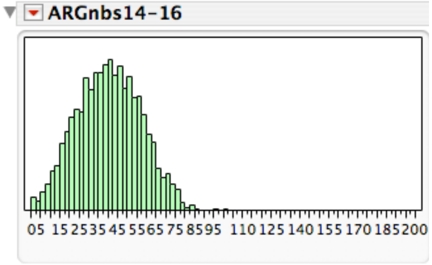


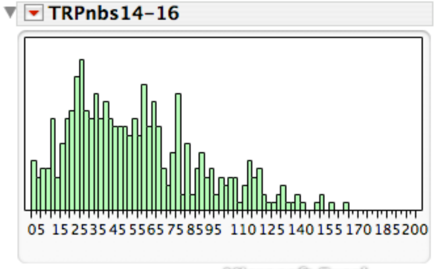












aa	<=10nb	<=13nb	<=16nb
VAL	96.397	69.151	46.306
ILE	112.334	78.985	51.073
LEU	116.176	83.098	52.291
MET	121.9	92.632	58.118
PHE	134.411	93.386	55.33
TRP	134.156	89.43	54.756
TYR	110.859	80.149	53.252
GLY	37.228	30.468	23.055
ALA	63.475	49.998	35.127
PRO	93.007	71.169	51.464
CYS	77.296	49.337	30.761
HIS	86.739	65.524	45.676
SER	46.35	36.242	26.126
ASN	35.207	25.759	18.425
ASP	41.575	30.469	21.423
GLN	51.777	39.345	29.211
GLU	58.815	43.727	31.905
THR	68.096	52.354	36.944
ARG	69.341	54.594	40.726
LYS	98.204	79.184	61.568