Development of a Knowledge Based Environment Potential in Rosetta

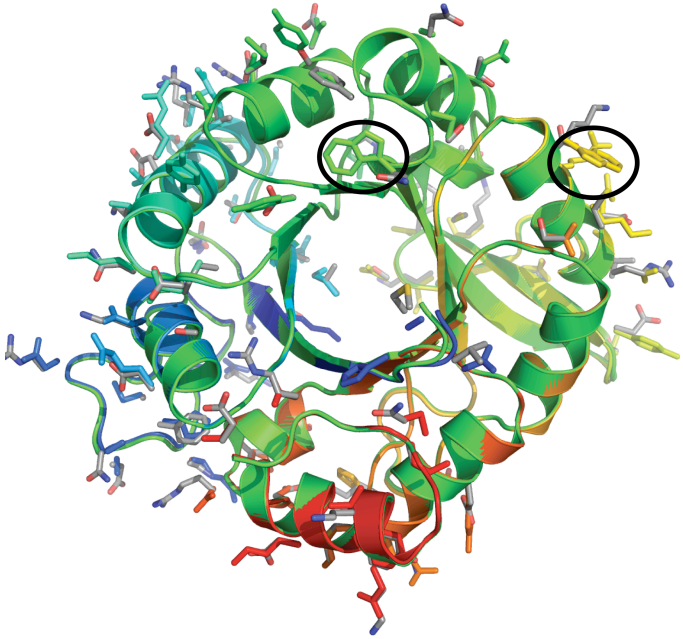
Sam DeLuca

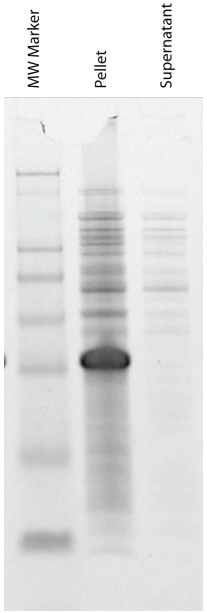
Vanderbilt University

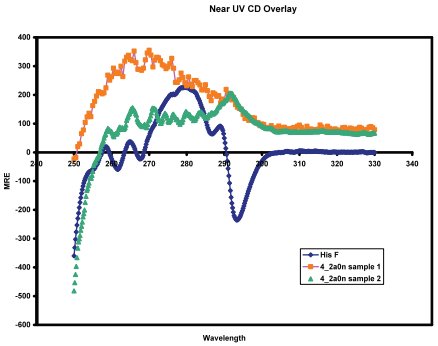
Outline

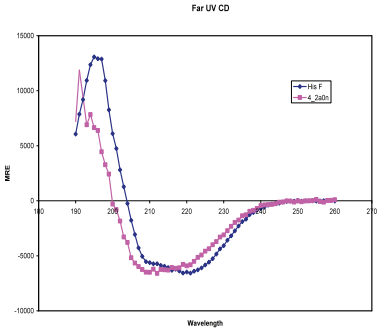
- State of Rosetta surface design
- The Rosetta Solvation model
- Estimating surface accessibility
- Development of a Knowledge Based Potential
- Weight optimization

Rosetta Design incorrectly designs surface Residues



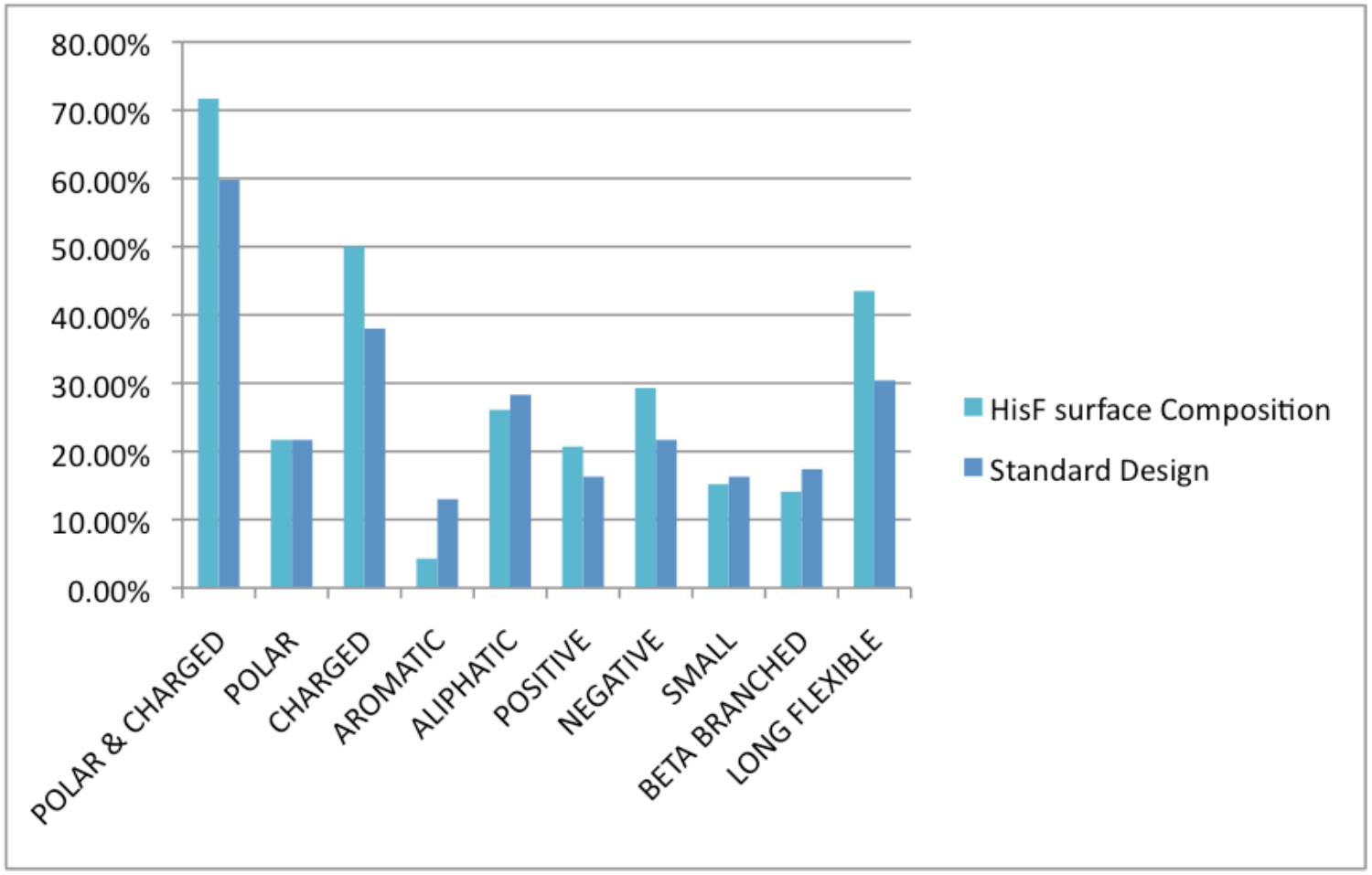






- Design was lowest scoring of 32 designs (8 rounds from 4 starting structures)
  - 2A0N and 1THF (Native and Relaxed Structures)
- 79 Mutations (after 3 removed from core)
  - 1THF (Relaxed)

The Standard Energy Function Introduces Biases



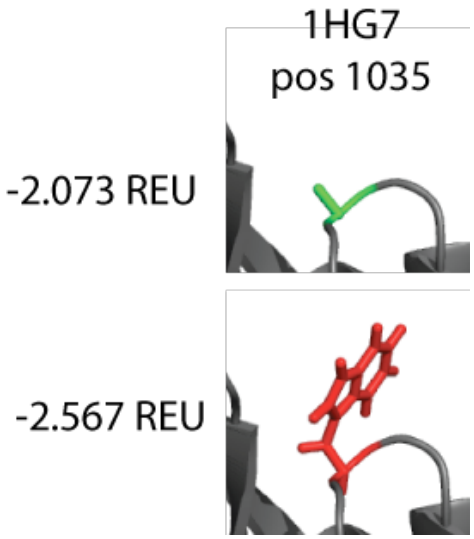
The Rosetta Solvation potential does not significantly penalize apolar surface residues.

- Lazaridis-Karplus solvation potential
- Combination of physical models and knowledge based potentials
- Assumes base reference energy of amino acid in solvation
- Change in energy due to solvent exclusion upon folding is measured

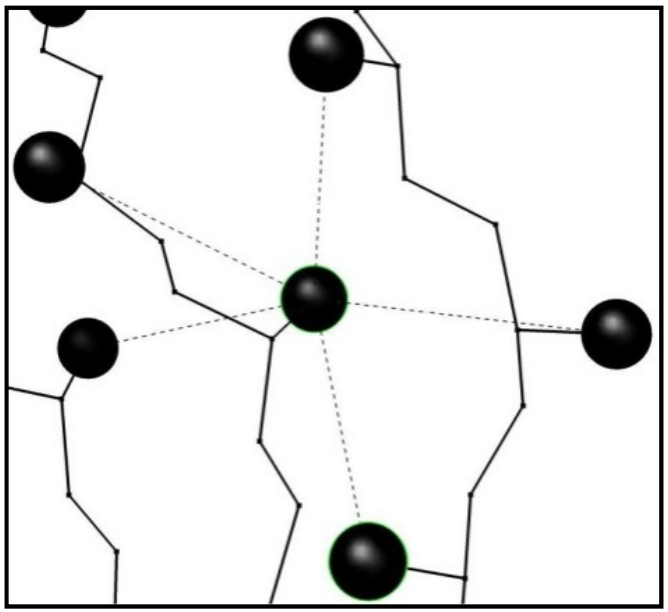
Lazaridis and Karplus. Effective energy function for proteins in solution. Proteins (1999) vol. 35 (2) pp. 133-152

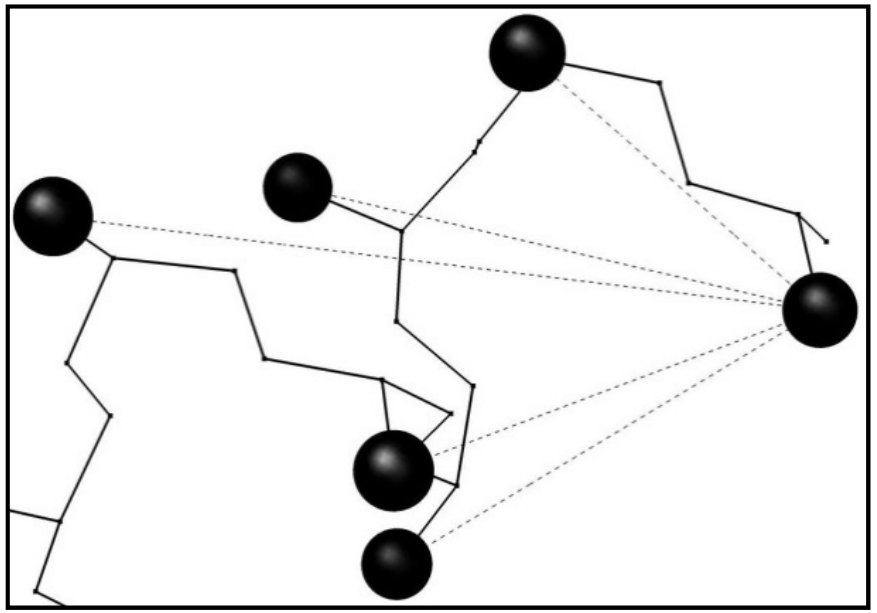
The Rosetta Solvation potential does not significantly penalize apolar surface residues.

- Surface residue selection is dominated by Reference energy and Solvation potential
- Desolvation does not significantly penalize apolar surface residues
- A term is needed to properly capture penalties beyond solvation



Distinguishing between Isotropic and Anisotropic positioning



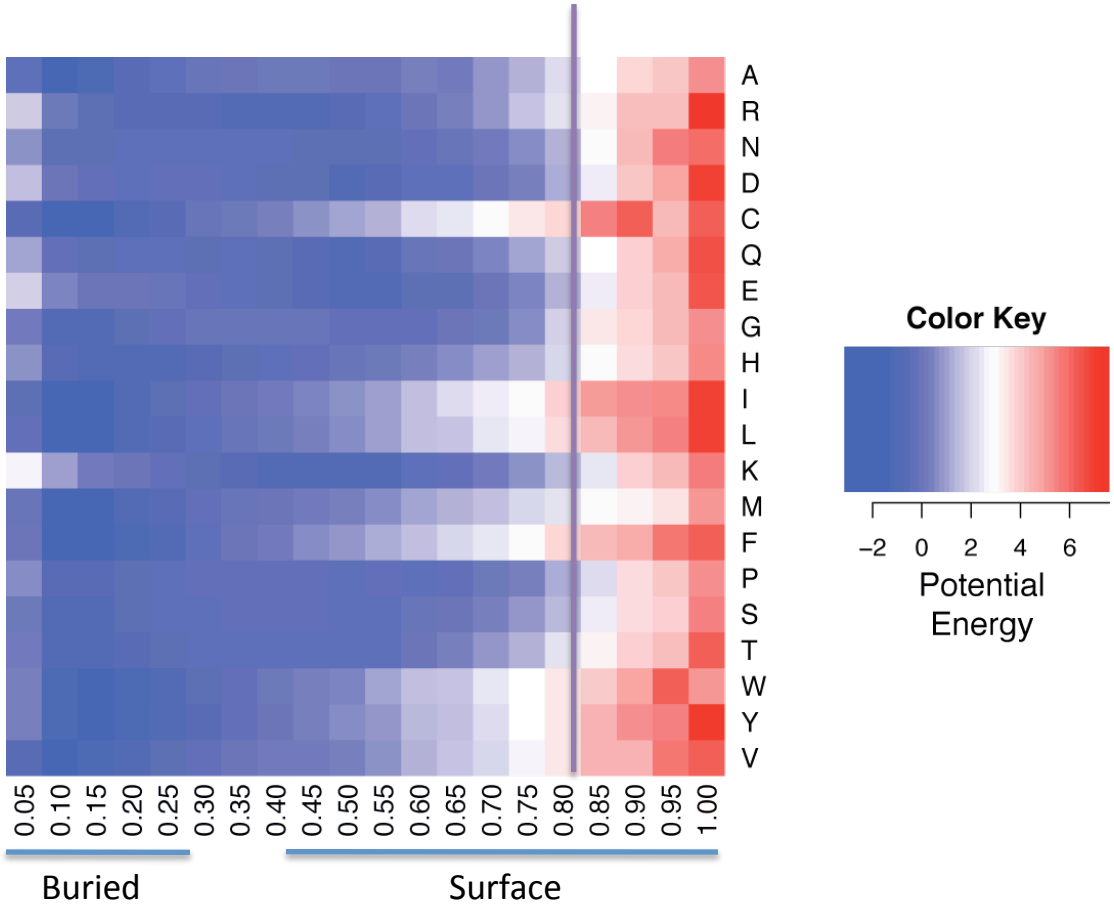


Durham et al. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. Journal of molecular modeling (2009) pp.

Development of a Knowledge Based Potential

- 1795 proteins
  - Resolution < 2.5Å
  - Sequence Homology < 25%
  - Sequence Length 40-30,000
- Potential energy was calculated as – ln(propensity(aa[j]) where j=burial level

Development of a Knowledge Based Potential



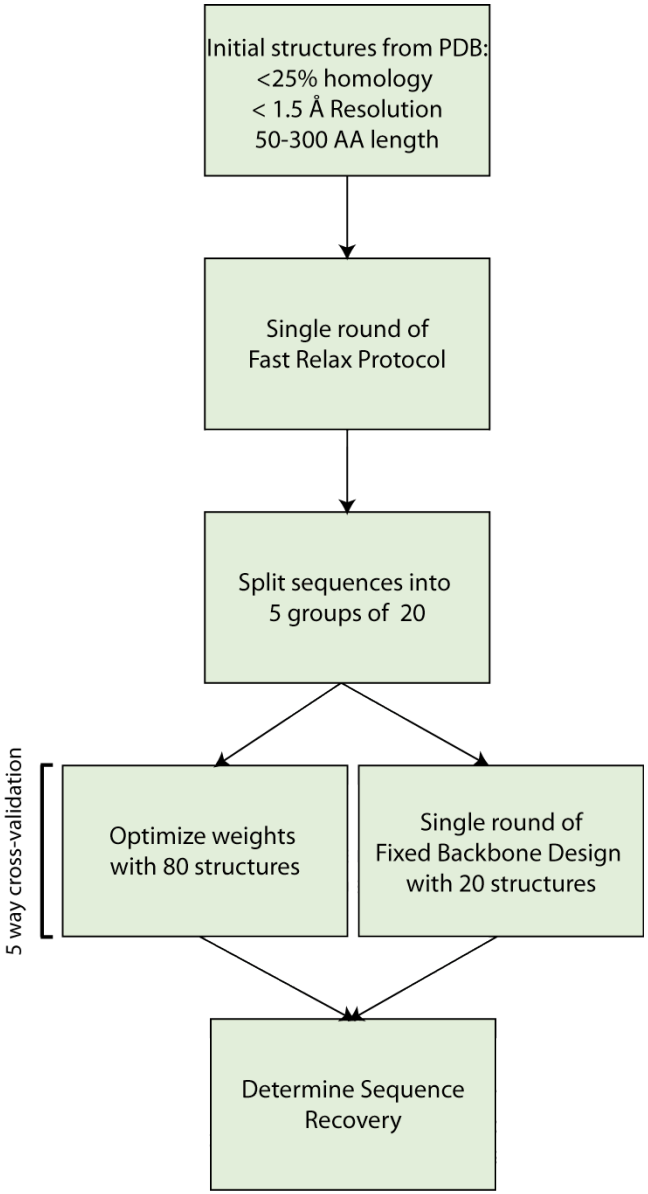
Rosetta 3.0 implementation

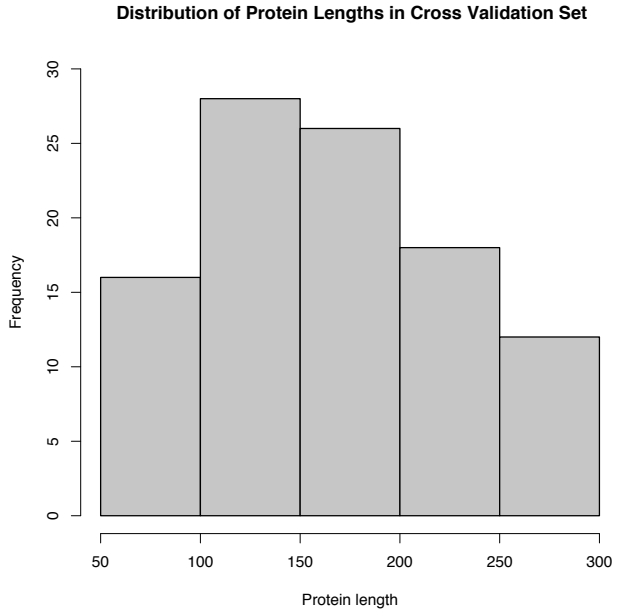
- The neighbor vector knowledge based potential is implemented as a long range one body rosetta energy scoring term.
  - neigh_vect
    - knowledge based potential energy
  - neigh_vect_raw
    - neighbor vector score
  - neigh_count
    - Weighted neighbor count

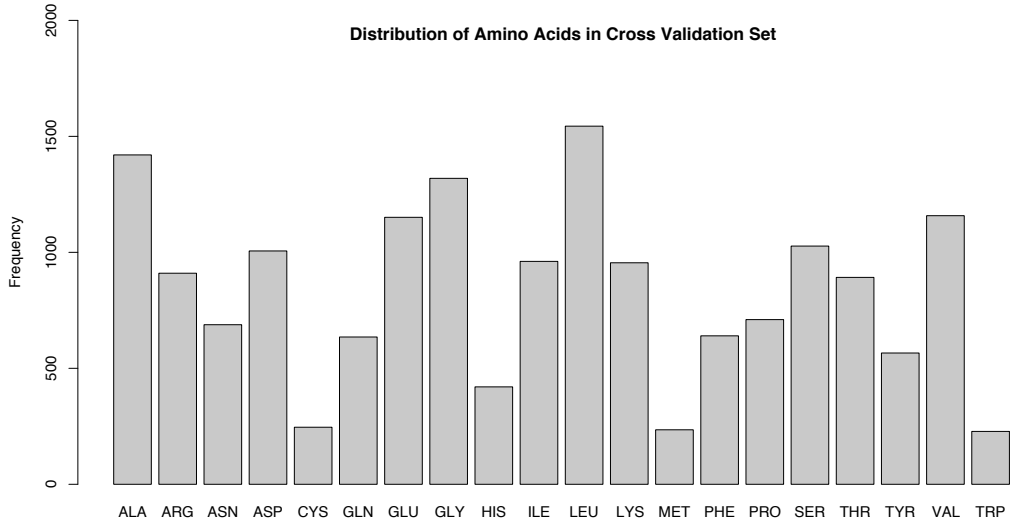
Weight optimization Protocol

- Rosettas optE_parallel application was used to optimize weights
- Ten iterations of optimization
- weights optimized to maximize native rotamer recovery

Cross Validation Protocol



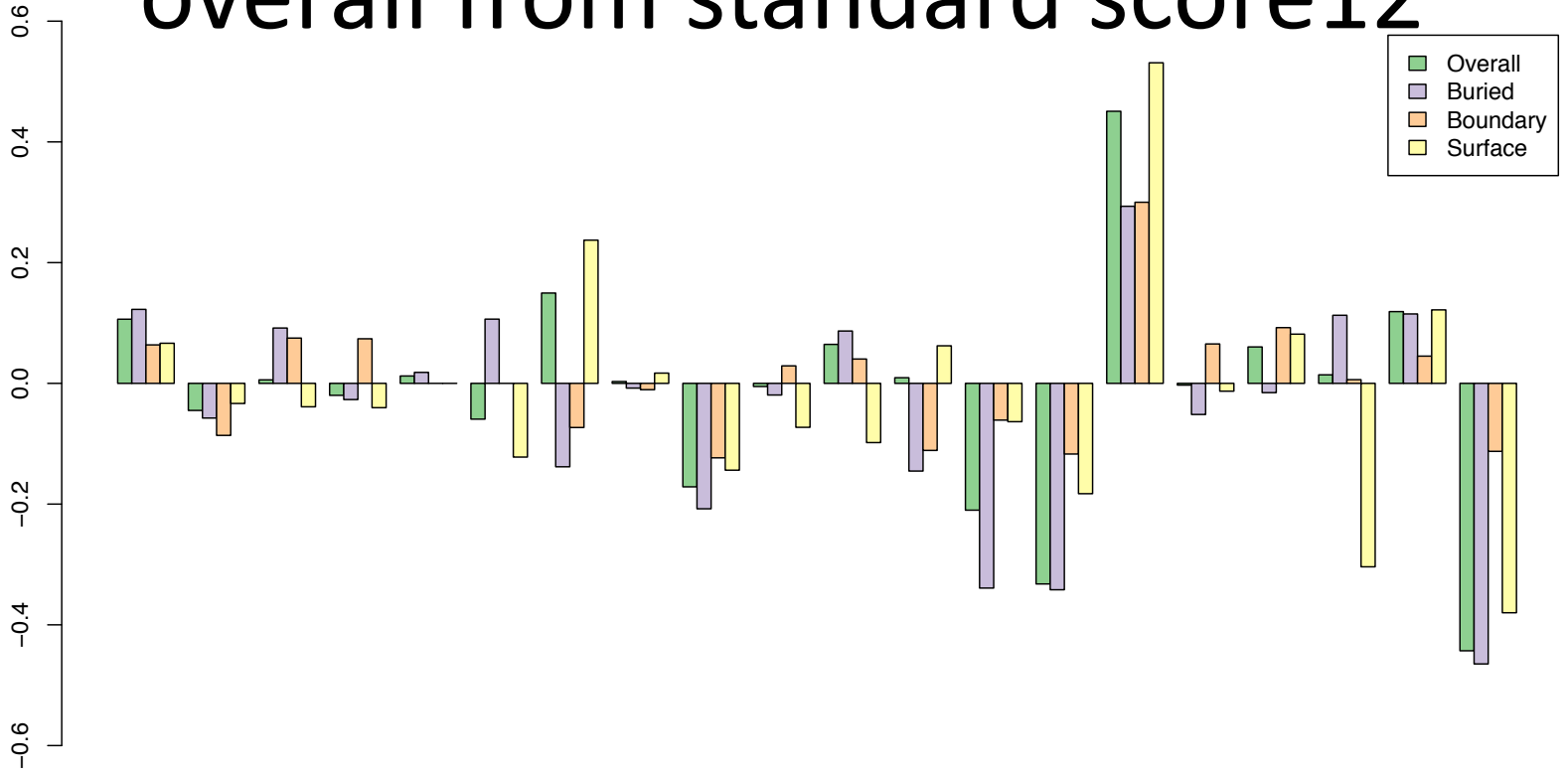




Cross Validation Statistics

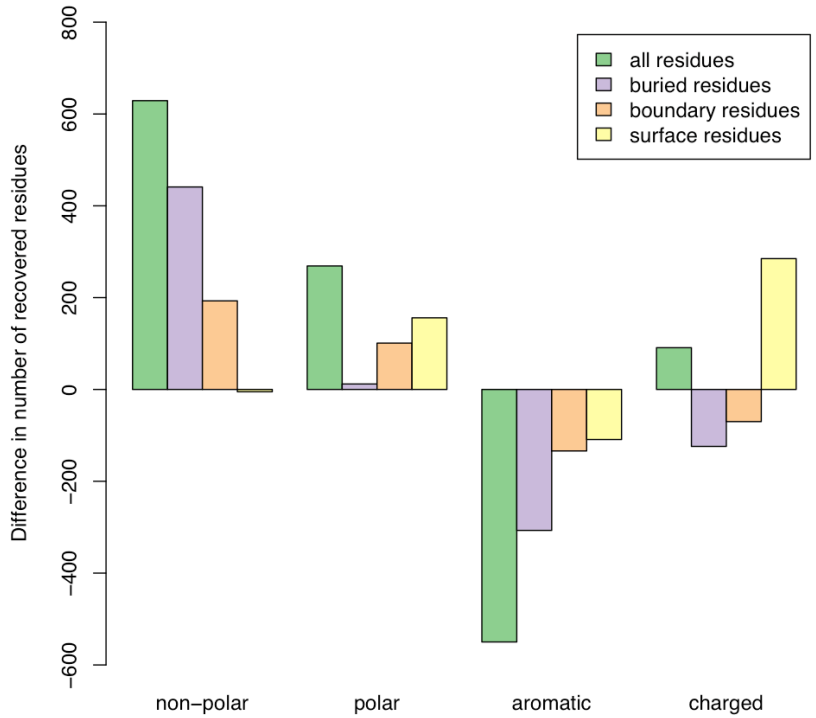
Scoring Term	Mean	Standard deviation
fa_atr	0.80	0.00
fa_rep	0.44	0.00
fa_sol	0.81	0.01
fa_intra_rep	0.00	0.00
pro_close	0.09	0.04
fa_pair	0.41	0.02
hbond_sr_bb	1.25	0.11
hbond_lr_bb	1.27	0.08
hbond_bb_sc	3.15	0.06
hbond_sc	1.56	0.04
dslf_ss_dst	1.09	0.09
dslf_cs_ang	0.93	0.15
dslf_ss_dih	1.10	0.09
dslf_ca_dih	0.95	0.12
fa_dun	0.83	0.02
p_aa_pp	1.13	0.02
neigh_vect	0.48	0.02
ref	1.00	0.00

Sequence Recovery is improved 2.7% overall from standard score12



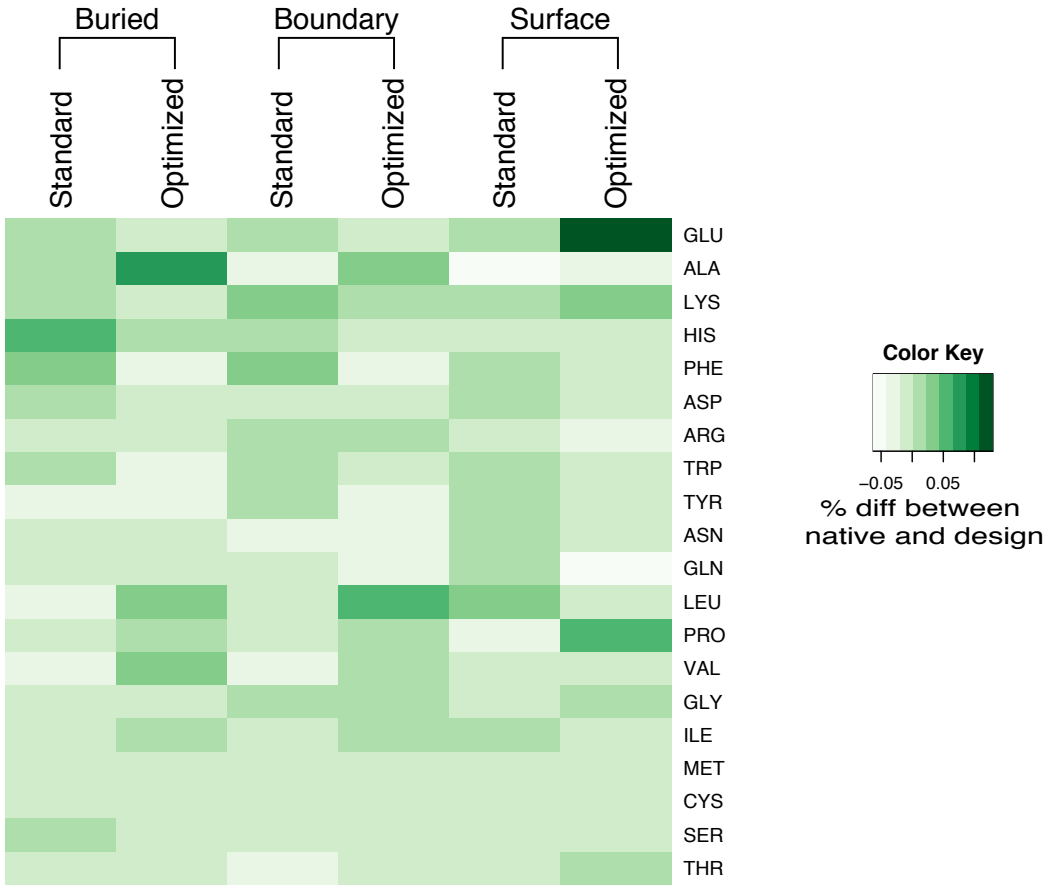
Overall	151	-41	4	-20	3	-38	174	4	-72	-5	100	9	-50	-213	320	-3	54	8	138	-101
Buried	75	-10	11	-4	3	10	-17	-3	-32	-12	86	-17	-39	-146	44	-15	-4	37	79	-66
Boundary	39	-15	9	11	0	0	-9	-4	-19	18	40	-13	-7	-50	45	19	24	2	31	-16
Surface	37	-16	-16	-27	0	-48	200	11	-21	-11	-26	39	-4	-17	231	-7	34	-31	28	-19

Sequence Recovery is improved in non-polar/charged residues

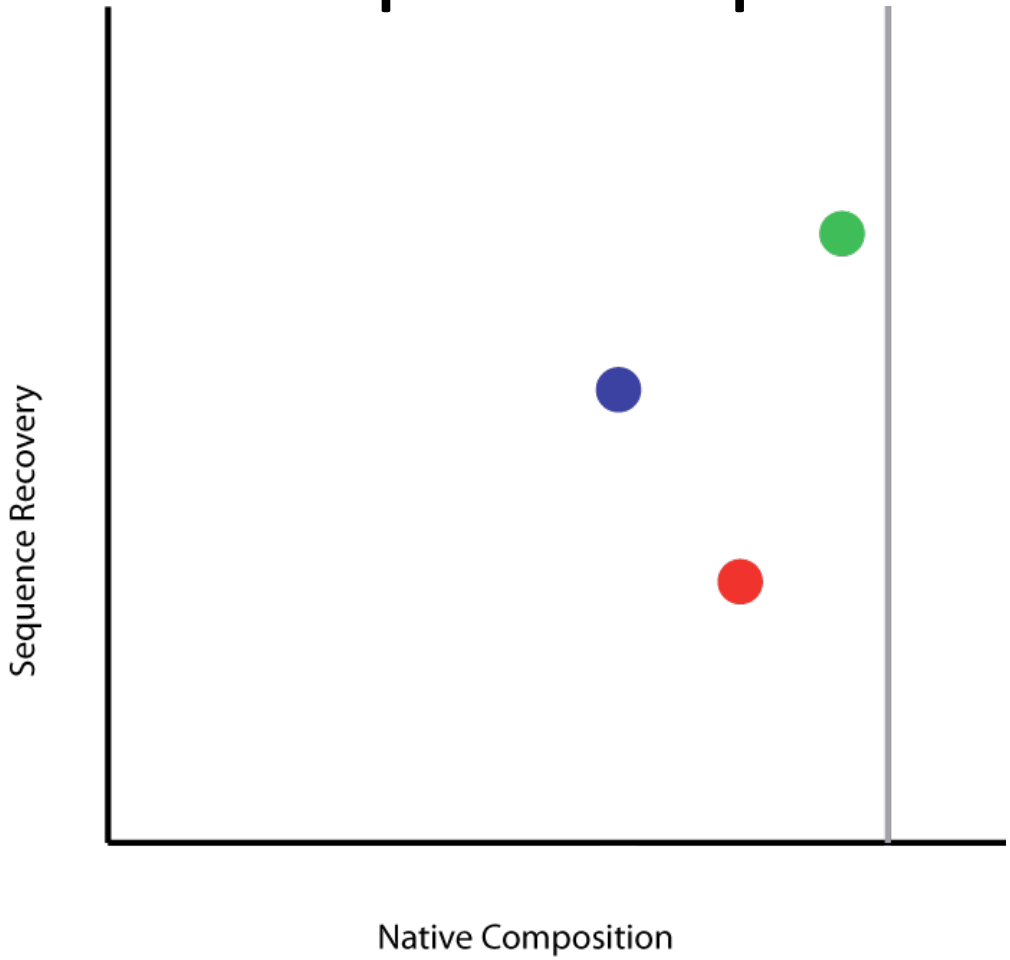


Overall	653	236	-524	64
Buried	449	15	-298	-164
Boundary	218	103	-119	-84
Surface	-14	118	-107	312

Sequence Composition



Reducing biases by further sampling of the sequence space



Acknowledgements

- Jens Meiler
- Brent Dorr
- Elizabeth Ashley Durham
- Steven Combs
- Kristian Kaufmann
- Gordon Lemmon
- The rest of the Meiler lab and Rosetta Community