Supplementary Tables 1 and 2 show the individual weights of the energy terms obtained during the five way cross validation, and the averaged energy function used during benchmarking. The 5 weights obtained for the Solvation free energy potential and the NV environment KBP were averaged, and the average value shows a relatively low standard deviation. As seen in supplementary table 2, the standard deviations of the averaged reference energies were significantly higher. This is to be expected as the space of possible reference energies is much larger. Due to these large standard deviations, the reference energies were reoptimized rather than averaged in the final NV environment KBP energy function.

Supplementary Table 1: Energy function weights for the NV environment KBP

			Five way cross validation sets						
	ROSETTA Scoring term	Scoring tterm description	1	2	3	4	5	Mean	Standard Deviation
Free weights	fa_sol	Solvation Free energy Potential	0.558	0.569	0.563	0.547	0.585	0.564	0.014
	neigh_vect	NV environment KBP	1.025	1.013	0.996	1.059	0.978	1.014	0.030
Fixed weights	fa_atr	Attractive force	0.800					0.800	0.000
	fa_rep	Repulsive force	0.440					0.440	0.000
	fa_intra_rep	Intra-residue repulsive force	0.004					0.004	0.000
	pro_close	Proline closure bonus	1.000					1.000	0.000
	fa_pair	Pair energy	0.490					0.490	0.000
	hbond_sr_bb	Hydrogen bonding: short range backbone	0.5850					0.585	0.000
	hbond_lr_bb	Hydrogen bonding: long range backbone	1.170					1.170	0.000
	hbond_bb_sc	Hydrogren bonding: backbone-sidechain	1.170					1.170	0.000
	hbond_sc	Hydrogen bonding: sidechain-sidechain	1.100					1.100	0.000
	dslf_ss_dst	Disulfide sidechain distance	1.000					1.000	0.000
	dslf_cs_ang	Disulfide cystine sulfur angle	1.000					1.000	0.000
	dslf_ss_dih	Disulfide sidechain-sidechain dihederal	1.000					1.000	0.000
	dslf_ca_dih	Disulfide Cα-sidechain dihederal	1.000					1.000	0.000
	rama	Ramachandran score	0.200					0.200	0.000
	omega	Omega angle score	0.500					0.500	0.000
	p_aa_pp	Probability of an a	0.320					0.320	0.000
	fa_dun	dunbrack rotamer library	0.560					0.560	0.000

Supplementary Table 2: Reference energy weights for the NV environment KBP

Amino Acid	Five way cross validation sets							
	1	2	3	4	5	Mean	Standard Deviation	Re-optimized Energies
A	-0.313	-0.345	-0.332	-0.299	-0.311	-0.320	0.018	-0.280
C	-0.186	-0.267	-0.251	-0.203	-0.273	-0.236	0.039	-0.191
D	-0.051	-0.152	-0.117	-0.045	-0.158	-0.105	0.053	-0.089
E	-0.116	-0.225	-0.221	-0.134	-0.230	-0.185	0.055	-0.163
F	0.979	1.008	1.098	0.976	1.035	1.019	0.050	1.002
G	0.187	0.862	0.681	0.322	0.952	0.601	0.334	0.318
H	0.773	0.727	0.727	0.755	0.683	0.733	0.034	0.738
I	-0.087	-0.063	-0.066	-0.147	-0.034	-0.079	0.041	-0.089
K	-0.035	-0.113	-0.125	-0.046	-0.104	-0.085	0.041	-0.056
L	-0.288	-0.27	-0.302	-0.350	-0.286	-0.301	0.029	-0.295
M	-0.475	-0.472	-0.532	-0.514	-0.478	-0.494	0.027	-0.488
N	-0.523	-0.559	-0.549	-0.500	-0.582	-0.543	0.031	-0.532
P	-0.486	-0.556	-0.579	-0.488	-0.551	-0.532	0.042	-0.494
Q	-0.481	-0.582	-0.554	-0.491	-0.569	-0.536	0.046	-0.497
R	-0.280	-0.364	-0.324	-0.255	-0.389	-0.322	0.055	-0.294
S	-0.396	-0.436	-0.438	-0.373	-0.437	-0.416	0.029	-0.393
T	-0.312	-0.375	-0.373	-0.307	-0.368	-0.347	0.034	-0.332
V	-0.175	-0.166	-0.166	-0.217	-0.134	-0.1720	0.029	-0.176
W	1.430	1.505	1.529	1.469	1.426	1.472	0.045	1.474
Y	0.842	0.853	0.902	0.851	0.815	0.853	0.031	0.842

Supplementary Tables 3 and 4 show the standard deviations of PSSM and Sequence Recovery values produced by all 5 scoring functions generated by the 5 way cross validation. Comparison of the standard deviations with observed improvements in PSSM and sequence recovery suggests that recovery improvements greater than 1% are likely statistically significant.

Supplementary Table 3: Standard deviations for 100 protein benchmark set data

	Percent PSSM Recovery		Percent Sequence Recovery	
	Reference	NV environment KBP	Reference	NV Environment KBP
Buried	0.4%	0.3%	0.8%	0.2%
Boundary	0.6%	0.8%	0.3%	0.5%
Surface	0.4%	0.2%	0.4%	0.1%
Overall	0.2%	0.1%	0.3%	0.1%

Supplementary Table 4: Standard deviations for 42 protein benchmark set data

	Percent PSSM Recovery		Percent Sequence Recovery	
	Reference	NV environment KBP	Reference	NV environment KBP
Buried	0.7%	0.5%	1.2%	0.6%
Boundary	0.7%	0.5%	0.3%	0.6%
Surface	0.3%	0.4%	0.4%	0.4%
Overall	0.2%	0.3%	0.5%	0.3%