Supporting materials for the article

Self Complementarity within Proteins: Bridging the Gap between Binding and Folding

by
Sankar Basu[†], Dhananjay Bhattacharyya[‡] and Rahul Banerjee[†]*

[†]Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata – 700 064, India

[‡]Biophysics Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata – 700 064, India

Dataset S1. The database (DB2) used for the computation of electrostatic complementarity. PDB ID and chain identifier (underscored) for each polypeptide chain is given under its respective protein class. This database is a subset of a larger database (**DB1**) discussed in detail in a previous report **(1)**. As in **DB1**, in case of multiple occupancies of atoms the highest occupancy was used in the calculation and the first conformer for equal occupancies. Proteins with metal ions as an integral part of their structure (given in parentheses) were determined by viewing the structures in RasMol (http://www.umass.edu/microbio/rasmol) and also taking into account the buffering conditions whilst crystallization.

all α:

1C02_B, 1C1K_A, 1EL4_A, 1ELK_B, 1EYH_A, 1EZJ_A (Ca^{+2}), 1G5N_A (Ca^{+2}), 1G8E_A, 1I8O_A, 1JKV_A (Ca^{+2}, Mn^{+3}), 1LJ9_A, 1MOJ_A (Fe^{+3}), 1MXR_B (Fe^{+3}, Hg^{+2}), 1OOH_B, 1OXJ_A, 1Q08_A (Zn^{+2}), 1Q5Z_A, 1QYZ_A, 1R1T_B, 1R5Z_B, 1R7J_A, 1SZH_A, 1TFJ_A, 1U2W_B (Zn^{+2}), 1U84_A, 1UCR_B, 1WPA_A, 1WXC_A, 1ZKR_A, 2A61_A, 2AIB_A, 2B8I_A, 2BA2_C, 2CWL_A, 2D48_A, 2D4X_A, 2DPO_A, 2E1N_B, 2ERB_A, 2F2B_A, 2FBQ_A, 2FD5_A, 2FMM_E, 2FP1_A, 2FU4_A, 2GXG_A, 2HRA_A, 2I3F_B, 2IC6_B, 2IGP_A, 2IMI_B, 2NZ7_B, 2O37_A, 2O70_F, 2OEB_A, 2OQG_C, 2PQR_B, 2QRW_K, 2QSA_A, 2RKN_A (Zn^{+2}), 3B9W_A, 3BS7_B, 3C1D_B, 3CHM_A, 3DHZ_B (Fe^{+2}).

all β:

1DJR_H, 1DMH_B (Fe^{+3}), 1EZG_B, 1F7D_A, 1I0R_B, 1I4U_A, 1IBY_B (Cu^{+2}), 1J71_A, 1K12_A (Ca^{+2}), 1KJL_A, 1KT6_A, 1KZQ_B, 1LR5_C (Zn^{+2}), 1LSL_A, 1LXZ_A, 1LYQ_B, 1NTV_A, 1NYK_B, 1NYW_B, 1OU8_B, 1P3C_A, 1PXD_A, 1QB5_E, 1QXM_A, 1R77_B, 1R8O_A, 1ROC_A, 1RU4_A (Ca^{+2}), 1S1D_A (Ca^{+2}), 1SE0_A (Zn^{+2}), 1T2W_C, 1TH7_L, 1V8H_A, 1VM9_A, 1WD3_A, 1WLG_B, 1WLI_B, 1WS7_A, 1XU1_B, 1YFQ_A (Ca^{+2}), 1YOA_A, 1Z3T_A, 1ZVT_B, 2A6V_A, 2AYD_A (Zn^{+2}), 2D37_A, 2DCY_C, 2DP9_A, 2DUR_A (Ca^{+2}), 2EI5_A, 2ERF_A, 2ET1_A (Mn^{+2}), 2F6E_A, 2G9F_A, 2GGV_B, 2GH2_A (Fe^{+3}), 2HAL_A, 2ICC_A, 2IQY_A (Ca^{+2}), 2O8L_A, 2OFZ_A, 2OR7_B, 2PA7_A, 2PCX_A (Zn^{+2}), 2R5O_A, 2RJ2_A, 3BOV_A, 3BU1_A, 3D9X_C, 3F5R_A.

α | β:

1DJ0_A, 1EKG_A, 1EQ6_A, 1F1U_A, 1F46_B, 1F5M_B, 1G1T_A (Ca^{+2}), 1G2R_A, 1IQZ_A, 1J1G_A, 1J1Y_A, 1J27_A, 1J3W_B, 1JD5_A (Zn^{+2}), 1JYH_A, 1JYO_B, 1K2A_A, 1KJO_A (Zn^{+2}, Ca^{+2}), 1KQB_C, 1KUX_A, 1LN4_A, 1LO7_A, 1LQP_B, 1M4D_A, 1M4J_B, 1NE9_A, 1NWA_A, 1NWW_B, 1OZ9_A, 1PC4_A, 1PPV_B (Mg^{+2}, Mn^{+2}), 1R29_A, 1R45_A, 1RC9_A, 1RKI_A, 1RWZ_A, 1S5U_E, 1TA8_A, 1TKE_A, 1TP6_A, 1TR0_I, 1TU1_A, 1TUA_A,

1TUH_A, 1TUV_A, 1WRI_A, 1XBI_A, 1XMT_A, 1XPP_D, 1Y60_D, 1YAR_D, 1YGT_A, 1YPY_A, 1YWM_A, 1Z4R_A, 1ZDY_A, 1ZHQ_G, 1ZHX_A, 1ZT3_A, 2A15_A, 2AAL_A, 2ACY_A, 2B18_A, 2BBE_A, 2CVE_A, 2DT4_A, 2E11_B, 2E12_A, 2E8G_B, 2EA1_A, 2FHZ_A, 2FJR_B, 2FL4_A, 2FVV_A, 2FYG_A (Zn^{+2}), 2G64_A (Zn^{+2}), 2G8O_A, 2GEB_A (Ca^{+2}), 2GH5_B, 2GHT_A, 2GMN_B (Zn^{+2}), 2GU3_A, 2GWM_A, 2H2Z_A, 2I2Q_A, 2I7D_A, 2IF6_A, 2IG8_B, 2NR7_A, 2NTT_B, 2O28_A, 2OH5_A, 2P0W_A, 2P2O_B, 2P3H_A, 2PLQ_A, 2PU3_A, 2Q0L_B, 2QKP_B, 2Z51_A, 3BJK_C, 3BK8_A, 3C8I_B, 3COU_A, 3DA4_A, 3DHA_A (Zn^{+2}).

α + β:

1C7N_C, 1DGF_B, 1DJE_A, 1DQZ_B, 1EQC_A, 1EU8_A, 1F8M_B, 1G5T_A, 1G8A_A, 1I2A_A, 1I2K_A, 1I5G_A, 1I6W_B, 1I9C_B, 1IM5_A (Zn^{+2}), 1INL_C, 1IZC_A, 1JAY_B, 1JJF_A, 1JKE_C (Zn^{+2}), 1JX6_A (Ca^{+2}), 1K4M_B, 1K7C_A, 1LOK_A (Zn^{+2}), 1LUA_B, 1LWD_B, 1N3Y_A, 1NF9_A, 1O04_G, 1OOE_B, 1OZN_A, 1P6O_B (Zn^{+2}, Ca^{+2}), 1Q7L_A (Zn^{+2}), 1RKU_B, 1SG0_B (Zn^{+2}), 1SW5_B, 1T1V_B, 1TJY_A, 1TP9_B, 1TVP_A, 1UB3_B, 1UC7_A, 1UEH_B, 1UG6_A, 1UK8_A, 1V2X_A, 1V37_A, 1VFL_A (Zn^{+2}), 1WN5_B, 1WO8_A, 1WOU_A, 1WPN_A, 1WX1_B, 1XBY_A, 1XQ6_B, 1XS5_A, 1XX1_B, 1Y1P_A, 1Y1T_A, 1Y6V_A, 1YB6_A, 1YD0_A, 1YN9_B, 1YQZ_B, 1YZX_B, 1ZL0_A, 1ZOI_C, 1ZR6_A (Zn^{+2}), 1ZX0_C, 1ZZG_A, 1ZZW_B, 2A14_A, 2A1I_A, 2ACF_C, 2AEU_A, 2AGD_B, 2AP1_A (Zn^{+2}), 2AVD_B, 2B1K_A, 2B3F_B, 2B82_A, 2CYG_A, 2E5F_A, 2FB9_A, 2FLA_A, 2FQX_A, 2FY6_A, 2G84_A (Zn^{+2}), 2GSO_A (Zn^{+2}), 2H1C_A, 2H1V_A, 2HC9_A (Zn^{+2}), 2HL6_B, 2HOR_A, 2HPJ_A, 2HRC_B, 2HU9_A, 2HXU_A (Mg^{+2}), 2HY5_B, 2HZL_B, 2I49_A, 2I56_C (Zn^{+2}), 2NZX_C, 2OUI_B, 2P51_A (Mg^{+2}), 2PS1_A, 2QDX_A, 2QUL_B (Mn^{+2}), 2REM_C, 2RHJ_A, 2YW3_D, 2Z1Y_B, 2Z6R_B, 3B4U_B, 3B8X_B, 3BJE_B (Ca^{+2}), 3BOE_A, 3C9H_B, 3CQ5_B, 3D2H_A, 3DAF_A (Fe^{+2}), 3E96_B, 3ES7_B (Mg^{+2}), 3EVF_A.

multidomain:

1DLJ_A, 1EEJ_B, 1F60_A, 1FP2_A, 1FW1_A, 1G8L_B, 1HM9_B (Ca^{+2}), 1HYO_B (Ca^{+2}), 1IG3_A, 1JAK_A, 1K7I_A (Zn^{+2}, Ca^{+2}), 1KTB_A, 1LBV_A, 1M1H_A, 1OWL_A, 1P1M_A, 1RA0_A (Fe^{+3}), 1RQP_C, 1SZN_A, 1T0F_A, 1T2D_A, 1T4B_B, 1U60_C, 1VBK_A, 1YU0_A (Ca^{+2}), 1Z6F_A, 1ZJC_A (Co^{+2}), 2CZ1_B, 2GA2_A (Mn^{+2}), 2GDQ_B, 2GGO_A, 2GZ1_B, 2GZJ_F (Zn^{+2}), 2PUZ_B (Fe^{+3}), 3ETJ_A (Mg^{+2}).

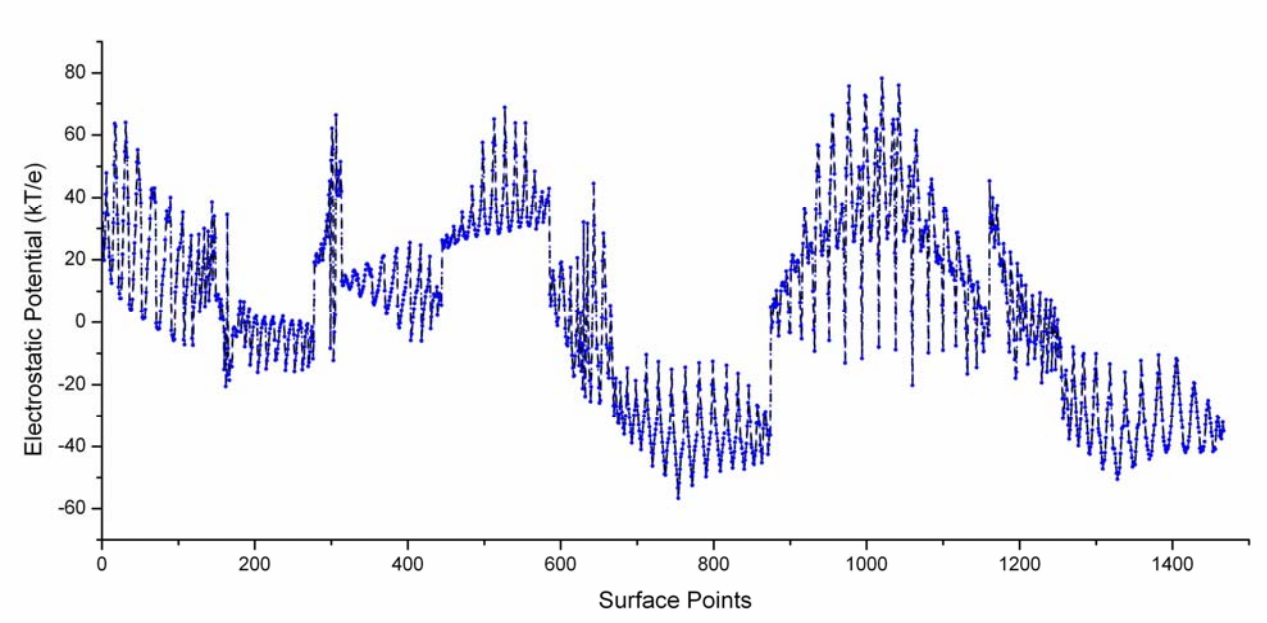

Figure S1. Electrostatic potentials computed by linearized Poisson-Boltzmaan equation (at zero ionic strength) and non-linear PBE at physiological counter-ionic strength (0.15 M NaCl) are virtually identical. Potentials calculated on individual surface points of a completely buried asparagine (58-Asn: 2HAQ) plotted in **blue:** LPBE; **black:** nonlinear PBE. The plot shows practically identical values in potential (obtained from the two methods) realized due to the atoms of the selected residue. Similar agreement has also been obtained for potentials realized due to the rest of the charged protein atoms.

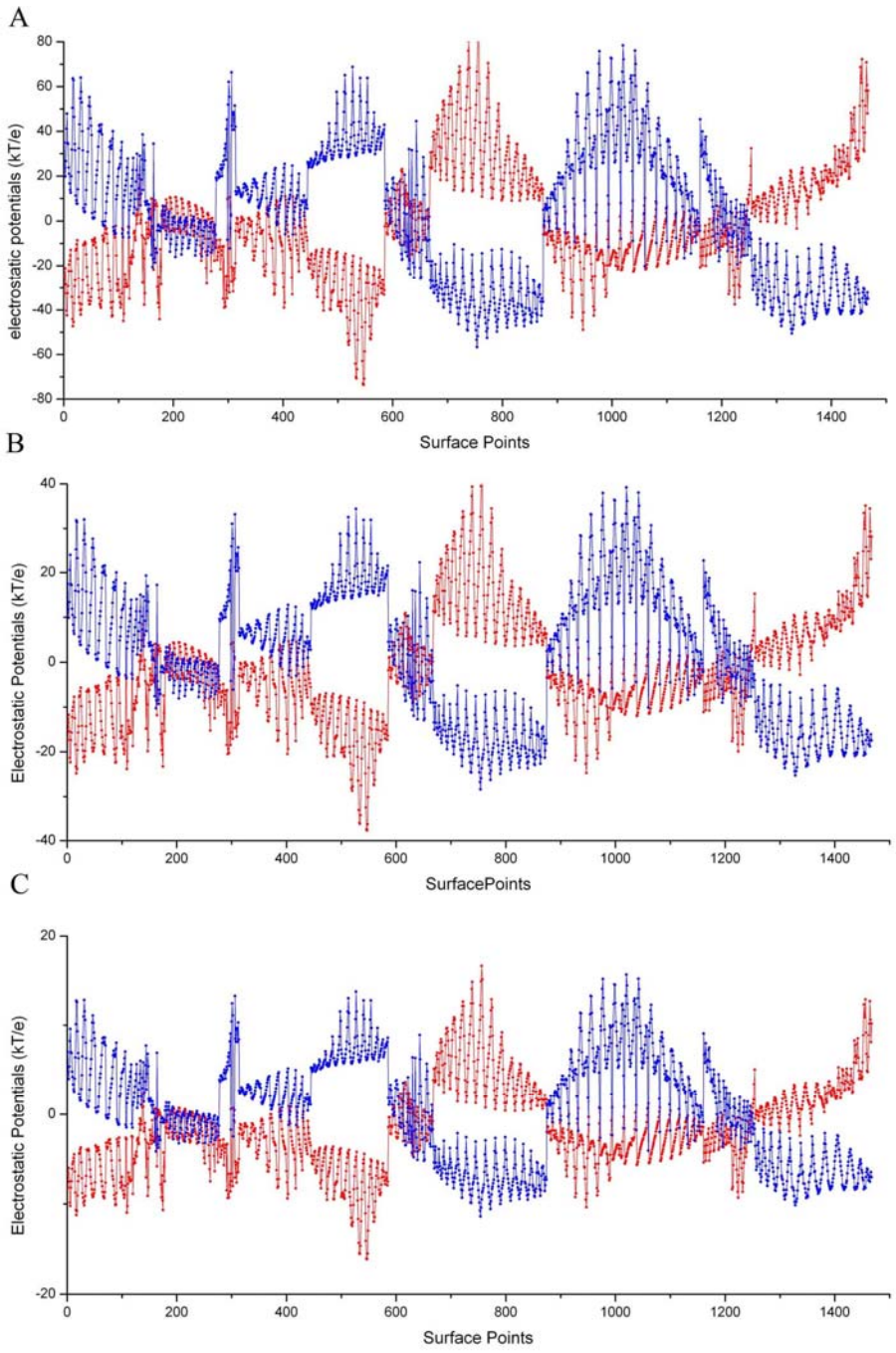

Figure S2. Variation in the internal dielectric of the continuum does not alter electrostatic complementarity for interior residues. Electrostatic potentials computed at the commonly used ranges of internal dielectrics: ε_p = 2 (A), 4 (B), 10 (C) for a completely buried asparagine (58-Asn: 2HAQ) (**blue:** due to the atoms of the selected residue, **red:** due to the rest of the charged protein atoms). As expected, change in ε_p changes merely the scale of the potentials thereby leading to conserved E_m values.

Table S1. Effect of large dielectric on E_m. Root mean square deviations of different E_m measures (E_m^{all}, E_m^{sc}, E_m^{mc}) calculated at (large) internal dielectric (ε_p) of 20, 40 respectively (tabulated in each entry separated by a comma) with respect to those calculated at the preferred value of $\varepsilon_p = 2$ for residues distributed in different burial bins: (bin1: $0.00 \leq Bur \leq 0.05$; bin2: $0.05 < Bur \leq 0.15$; bin3: $0.15 < Bur \leq 0.30$).

Residue	**bin1**			**bin2**			**bin3**		
	E_m^{all}	E_m^{sc}	E_m^{mc}	E_m^{all}	E_m^{sc}	E_m^{mc}	E_m^{all}	E_m^{sc}	E_m^{mc}
ALA	0.009, 0.017	0.010, 0.020	0.009, 0.017	0.017, 0.034	0.020, 0.042	0.017, 0.033	0.016, 0.038	0.015, 0.036	0.022, 0.050
VAL	0.008, 0.016	0.007, 0.016	0.008, 0.015	0.011, 0.025	0.011, 0.025	0.012, 0.024	0.014, 0.033	0.015, 0.032	0.015, 0.033
LEU	0.008, 0.019	0.009, 0.019	0.008, 0.015	0.012, 0.030	0.013, 0.029	0.011, 0.026	0.014, 0.036	0.012, 0.031	0.014, 0.034
ILE	0.008, 0.017	0.007, 0.016	0.007, 0.014	0.013, 0.031	0.012, 0.028	0.011, 0.027	0.014, 0.033	0.014, 0.031	0.014, 0.031
PHE	0.007, 0.014	0.005, 0.011	0.009, 0.016	0.011, 0.025	0.010, 0.021	0.015, 0.031	0.012, 0.029	0.012, 0.029	0.011, 0.025
TYR	0.006, 0.012	0.007, 0.014	0.007, 0.013	0.012, 0.026	0.013, 0.027	0.012 0.023	0.014, 0.035	0.013, 0.027	0.018, 0.04
TRP	0.007, 0.014	0.007, 0.015	0.007, 0.014	0.009, 0.024	0.011, 0.025	0.010, 0.023	0.016, 0.038	0.018, 0.039	0.017, 0.038
CYS	0.004, 0.009	0.006, 0.013	0.007, 0.014	0.008, 0.018	0.010, 0.020	0.014, 0.031	0.008, 0.016	0.016, 0.032	0.012, 0.026
MET	0.004, 0.008	0.005, 0.011	0.007, 0.014	0.006, 0.013	0.007, 0.014	0.014, 0.029	0.007, 0.015	0.015, 0.032	0.009, 0.020
SER	0.007, 0.013	0.007, 0.014	0.007, 0.015	0.012, 0.028	0.015, 0.032	0.013, 0.028	0.014, 0.033	0.014, 0.032	0.018, 0.041
THR	0.006, 0.012	0.007, 0.014	0.008, 0.015	0.013, 0.026	0.016, 0.032	0.013, 0.025	0.013, 0.031	0.010, 0.024	0.018, 0.042
ASN	0.009, 0.013	0.010, 0.016	0.009, 0.014	0.011, 0.024	0.014, 0.030	0.013, 0.027	0.011, 0.027	0.009, 0.022	0.015, 0.034
GLN	0.012, 0.018	0.015, 0.025	0.017, 0.025	0.011, 0.021	0.011, 0.023	0.016 0.029	0.010, 0.025	0.009, 0.022	0.016, 0.035
ASP	0.011, 0.02	0.011, 0.021	0.008, 0.015	0.012, 0.028	0.015, 0.035	0.009, 0.020	0.017, 0.036	0.012, 0.024	0.023, 0.049
GLU	0.013, 0.023	0.012, 0.021	0.009, 0.017	0.013, 0.029	0.014, 0.029	0.017, 0.033	0.013, 0.030	0.015, 0.036	0.015, 0.030
LYS	0.029, 0.035	0.030, 0.040	0.029, 0.033	0.012, 0.029	0.026, 0.055	0.004, 0.009	0.021, 0.050	0.006, 0.015	0.047, 0.096
ARG	0.013, 0.026	0.018, 0.039	0.005, 0.009	0.018, 0.042	0.027, 0.070	0.009, 0.017	0.022, 0.057	0.010, 0.021	0.033, 0.094
PRO	0.013, 0.022	0.010, 0.019	0.015, 0.022	0.013, 0.028	0.016, 0.035	0.014, 0.027	0.017, 0.041	0.014, 0.031	0.017, 0.039
HIS	0.007, 0.014	0.010, 0.019	0.007, 0.014	0.010, 0.022	0.012, 0.026	0.010, 0.023	0.012, 0.03	0.012, 0.026	0.017, 0.038

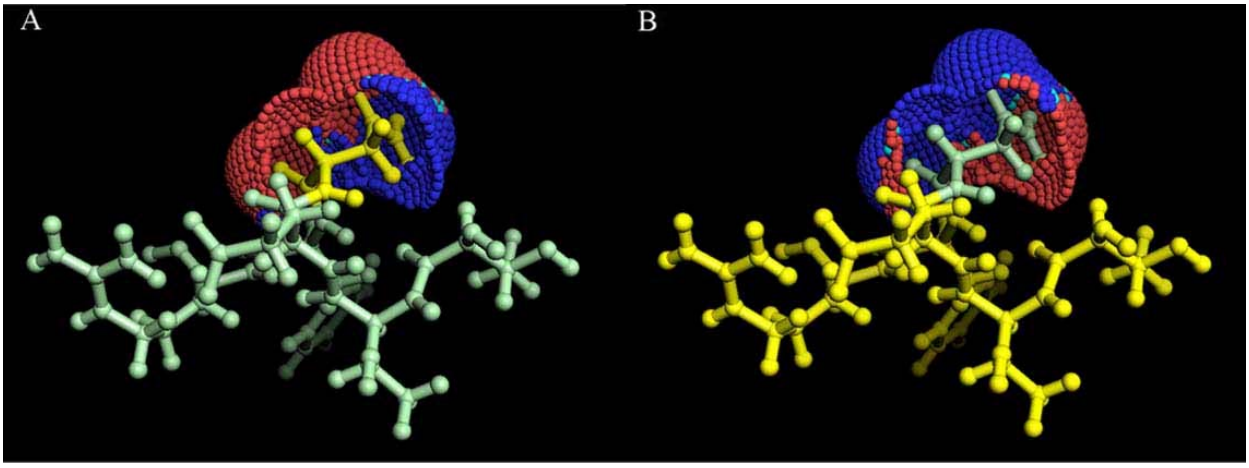

Figure S3. Molecular surface of an individual buried residue (target) exhibiting anti-correlated (complementary) electrostatic potentials. The figure shows the van der Waals surface (displayed as non-bonded spheres) of a completely buried asparagine (58-Asn from 2HAQ) along with its own atoms accompanied by a few more residues along the polypeptide chain as representative of the 'rest of the protein atoms' (sticks). Atoms (along with their interconnecting bonds) are colored by 'bright yellow' when 'charged' and by 'pale green' when 'uncharged'. Surface coloring follows standard conventions where patches of positive potentials are colored by 'blue', negative potentials by 'red' and neutral (0.0 ± 0.5 kT/e) by 'cyan'. (A) shows potentials realized due to the charged atoms of the residue itself whereas (B) shows potentials due to the rest of the charged protein atoms. Figure constructed by **PyMol** [http://www.pymol.org/].

Table S2. Assessment of the relative contributions of main chain and side chain atoms on electrostatic complementarity. Average E_m ($\overline{E_m}$) and their standard deviations (in parentheses) for different residues ($0.00 \leq$ Bur ≤ 0.05, see **Main Text**) calculated from different combinations of atomic sets and surfaces: **S1:** main chain atoms (target) versus main chain atoms ('rest of the protein') on main chain surface (of the target residue); **S2:** main chain atoms (target) versus main chain atoms (rest) on side chain surface (target). **S3:** side chain atoms (target) versus side chain atoms (rest) on side chain surface (target); **S4:** side chain atoms (target) versus all atoms (rest) on side chain surface (target).

Residue	$\overline{E_m}$			
	S1	**S2**	**S3**	**S4**
ALA	0.63 (0.25)	0.43 (0.22)	-0.04 (0.24)	-0.08 (0.18)
VAL	0.65 (0.22)	0.44 (0.23)	-0.02 (0.25)	0.15 (0.19)
LEU	0.65 (0.22)	0.42 (0.24)	0.01 (0.26)	0.14 (0.21)
ILE	0.65 (0.23)	0.43 (0.22)	-0.01 (0.28)	0.14 (0.24)
PHE	0.63 (0.24)	0.36 (0.28)	0.13 (0.16)	0.22 (0.17)
TYR	0.61 (0.25)	0.33 (0.27)	0.25 (0.27)	0.43 (0.22)
TRP	0.59 (0.28)	0.28 (0.26)	0.30 (0.22)	0.42 (0.19)
SER	0.53 (0.28)	0.25 (0.34)	0.24 (0.37)	0.53 (0.28)
THR	0.55 (0.28)	0.30 (0.29)	0.16 (0.34)	0.45 (0.28)
CYS	0.58 (0.27)	0.36 (0.23)	0.06 (0.29)	0.18 (0.25)
MET	0.64 (0.23)	0.38 (0.22)	0.12 (0.22)	0.21 (0.17)
ASP	0.47 (0.30)	0.17 (0.28)	0.27 (0.41)	0.55 (0.33)
GLU	0.56 (0.28)	0.27 (0.28)	0.31 (0.42)	0.54 (0.37)
ASN	0.54 (0.27)	0.23 (0.30)	0.33 (0.38)	0.64 (0.23)
GLN	0.57 (0.29)	0.29 (0.28)	0.32 (0.37)	0.60 (0.24)
LYS	0.62 (0.24)	0.29 (0.23)	0.40 (0.37)	0.58 (0.24)
ARG	0.59 (0.25)	0.23 (0.22)	0.28 (0.30)	0.53 (0.21)
PRO	0.49 (0.26)	0.25 (0.28)	-0.15 (0.35)	-0.05 (0.28)
HIS	0.55 (0.29)	0.30 (0.28)	0.28 (0.36)	0.49 (0.31)

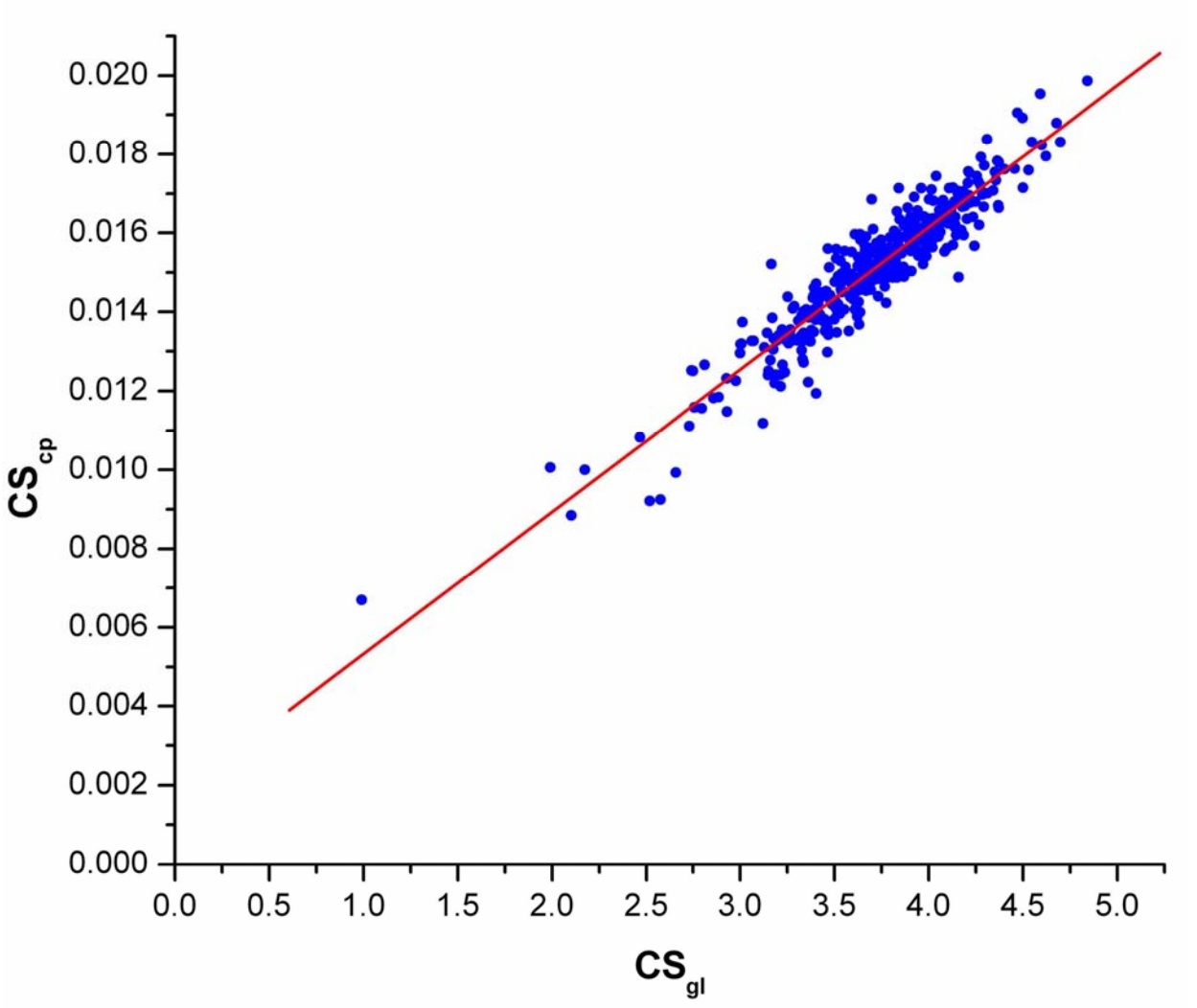

Figure S4. The two Complementarity Scores are linearly correlated. The figure shows the plot of CS_{gl} versus CS_{cp} computed for the 400 native folds (in **DB2**) linearly fitting to each-other ($R^2 = 0.94$).

Table S3. Performance of CS_{gl} and CS_{cp} in single decoy sets. Scores obtained by native and decoy structures have been tabulated for both functions in the decoy sets **(A)** Misfold **(2)** and **(B)** Pdberr **(3)** and sgpa **(4)**. For **(A)** Misfold, sequence of a native structure (e.g., 1BP2: 1st row) have been threaded onto the template from an unrelated fold (2PAZ) to generate the corresponding decoy (1BP2on2PAZ). In **(B)** MDC1 and MDC2 refer to the two molecular dynamic simulation snapshots of 2SGA. The decoy sets have been downloaded from the database '**Decoys 'R' Us**': [http://dd.compbio.washington.edu/].

(A).

Native	Decoy	Resolution / R-factor	CS_{gl}		CS_{cp}	
			Native	decoy	native	decoy
1BP2	1BP2on2PAZ	1.7/0.17	2.64	0.88	0.0113	0.0039
1FDX	1FDXon 5RXN	2.0/0.19	1.91	1.58	0.0071	0.0065
1HIP	1HIPon 2B5C	2.0/0.24	2.12	0.58	0.0097	0.0024
1LH1	1LH1on 2I1B	2.0/0.00	2.55	1.05	0.0111	0.0044
1P2P	1P2Pon1RN3	2.6/0.24	2.08	1.35	0.0085	0.0047
1PPT	1PPTon1CBH	1.4/0.00	2.37	0.94	0.0103	0.0033
1REI	1REIon5PAD	2.0/0.24	2.63	0.84	0.0108	0.0038
1RHD	1RHDon2CYP	2.5/0.00	1.58	0.84	0.0066	0.0035
1RN3	1RN3on1P2P	1.5/0.22	2.86	1.15	0.0120	0.0047
1SN3	1SN3on2CI2	1.2/0.19	2.13	0.60	0.0085	0.0032
1SN3	1SN3on2CRO	1.2/0.19	2.13	0.91	0.0085	0.0034
2B5C	2B5Con1HIP	2.0/0.16	3.48	0.97	0.0149	0.0034
2CDV	2CDVon2SSI	1.8/0.18	1.13	0.95	0.0047	0.0033
2CI2	2CI2on1SN3	2.0/0.20	4.09	1.05	0.0174	0.0045
2CI2	2CI2on2CRO	2.0/0.20	4.09	0.81	0.0174	0.0032
2CRO	2CROon1SN3	2.4/0.20	3.49	0.82	0.0130	0.0032
2CRO	2CROon2CI2	2.4/0.20	3.49	0.92	0.0130	0.0038
2CYP	2CYPon1RHD	1.7/0.22	2.84	0.84	0.0115	0.0033
2I1B	2I1Bon1LH1	2.0/0.17	2.90	0.96	0.0119	0.0034
2PAZ	2PAZon1BP2	1.6/0.18	3.21	1.42	0.0137	0.0053
2SSI	2SSIon2CDV	2.3/0.19	1.04	0.86	0.0046	0.0036
2TMN	2TMNon2TS1	1.6/0.18	2.71	0.96	0.0107	0.0037
2TS1	2TS1on2TMN	2.3/0.23	2.74	1.02	0.0114	0.0039
5PAD	5PADon1REI	2.8/0.00	2.26	1.06	0.0093	0.0042
5RXN	5RXNon1FDX	1.2/0.14	2.34	1.31	0.0117	0.0048

(B).

Native	Decoy	Chain length	CS_{gl}		CS_{cp}	
			Native	decoy	native	decoy
2F19	1F19	435	2.172	0.617	0.009	0.003
3HFL	2HFL	556	2.329	1.805	0.009	0.007
5FD1	2FD1	106	2.951	0.464	0.011	0.002
2SGA	MDC1	181	2.546	1.952	0.011	0.008
2SGA	MDC2	181	2.546	1.754	0.011	0.007

Table S4. Comparison in the performances of different knowledge-based scoring functions on single decoy sets. The functions include R_s, R_p **(5)**, RAPD, CDF **(6)**, Surfield **(7),** atomic knowledge based potential (AKBP) **(8)**, Residue Contact Potential (RCP) **(9)** along with the complementarity scores (CS_{gl}, CS_{cp}) developed in this study. The number of successful hits / total number of trials are tabulated.

Scoring Functions	Misfold	Pdberr and sgpa
R_s	24/24	5/5
R_p	20/24	5/5
RAPD	24/24	5/5
CDF	19/24	5/5
Surfield	23/23	-
AKBP	24/24	5/5
RCP	24/24	4/5
CS_{gl}	25/25	5/5
CS_{cp}	25/25	5/5

Table S5. Performance of CS_{gl} and CS_{cp} in multiple decoy sets of small proteins. Results tabulated for the decoy sets **(A)** 4-state reduced **(10)**, **(B)** Fisa **(11)**. **Resol / R** stands for resolution / R-factor of the native crystal structures whereas $\mathbf{N_{dec}}$ refers to the number of decoys. For **(B)** Fisa, there are 500 decoys for each native structure respectively. All proteins in Fisa belong to the all α class. $\mathbf{Z_{CS}}$ denotes the native Z-scores for the corresponding functions ($\boldsymbol{CS_{gl}}$ **/**$\boldsymbol{CS_{cp}}$**)**. The decoy sets have been downloaded from the database '**Decoys 'R' Us**': [http://dd.compbio.washington.edu/].

(A).

PDB ID	Length (aa)	class	Resol (Å) / R	N_{dec}	RMSD (Å) range of decoys	CS_{gl}		CS_{cp}	
						Rank	Z_{CS}	Rank	Z_{CS}
1CTF	68	α+β	1.70/0.17	630	1.3 - 9.1	1	7.9	1	7.1
1R69	63	All α	2.0/0.19	675	0.9 - 8.3	1	6.4	1	6.1
1SN3	65	α+β	1.2/0.19	660	1.3 - 9.1	1	5.6	1	4.7
2CRO	65	All α	2.4/0.20	674	0.8 - 8.3	1	5.9	1	5.0
3ICB	75	All α	2.3/0.18	653	0.9 - 9.4	1	6.7	1	4.9
4PTI	58	α+β	1.5/0.16	687	1.4 - 9.3	1	3.6	1	3.9
4RXN	54	All β	1.2/0.13	677	1.4 - 8.1	10	2.7	15	2.3

(B).

PDB ID	Length (aa)	Resol (Å) / R	RMSD (Å) range of decoys	CS_{gl}		CS_{cp}	
				Rank	Z_{CS}	Rank	Z_{CS}
1FC2	43	2.8/0.22	3.1 - 10.5	206	0.2	293	-0.2
1HDD-C	57	2.8/0.24	2.8 - 12.9	4	3.5	5	3.0
2CRO	65	2.4/0.20	4.3 - 12.6	1	7.2	1	6.3
4ICB	76	1.6/0.19	4.8 - 14.1	1	5.8	1	5.1

Table S6. Comparison in the performances of different knowledge based scoring functions on multiple decoy sets. The functions include DFIRE **(12)**, Rosetta **(13)**, ModPipe-Pair (MPP), ModPipe-Surf (MPS) **(14)**, TE13, LHL **(15)**, Force Model (FM) **(16)**, DOPE **(17)**, MJ **(18),** Surfield **(7),** R_s, R_p **(5)** along with the complementarity scores (CS_{gl}, CS_{cp}). All entries in the table refer to the rank of the native structure as detected by the corresponding method.

Decoy Set	PDB ID	DFIRE	Rosetta	MPP	MPS	TE13	LHL	FM	DOPE	MJ	Surfield	R_s	R_p	CS_{gl}	CS_{cp}
4state reduced	1CTF	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1R69	1	2	1	17	1	1	8	1	1	1	1	19	1	1
	1SN3	1	1	1	7	6	1	23	1	2	1	5	23	1	1
	2CRO	1	5	1	103	1	1	4	1	1	1	1	1	1	1
	3ICB	4	6	15	33	-	5	2	1	-	1	1	6	1	1
	4PTI	1	1	1	71	7	1	13	1	3	1	1	1	1	1
	4RXN	1	1	1	18	16	51	85	1	1	1	1	1	10	15
Fisa	1FC2	254	158	491	1	-	-	1	357	-	1	-	-	206	293
	1HDD	1	90	293	18	-	-	1	1	-	1	-	-	4	5
	2CRO	1	26	11	146	-	-	1	1	-	1	-	-	1	1
	4ICB	1	1	196	2	-	-	1	1	-	1	-	-	1	1

Table S7. Performance of CS_{gl} and CS_{cp} in decoy sets composed by homology modeling. Results tabulated for the decoy sets **(A)** Hg_structal **(19)**, **(B)** Ig_structal **(19)** and **(C)** Ig_structal_hires **(19)**. **Resol / R** stands for resolution / R-factor of the native crystal structure whereas $\mathbf{N_{dec}}$ refers to the number of decoys. For **(A)** Hg_structal, **(B)** Ig_structal and **(C)** Ig_structal_hires there are 29, 61 and 19 decoys for each native structure respectively. **(A)** Hg_structal is constituted of globins whereas **(B)** Ig_structal and **(C)** Ig_structal_hires contains immunoglobulins. $\mathbf{Z_{CS}}$ denotes the native Z-scores for the corresponding function ($\boldsymbol{CS_{gl}}$ /$\boldsymbol{CS_{cp}}$). The decoy sets have been downloaded from the database '**Decoys 'R' Us**': [http://dd.compbio.washington.edu/].

(A).

PDB ID	Length (aa)	Resol (Å) / R	RMSD (Å) range of decoys	CS_{gl}		CS_{cp}	
				Rank	Z_{CS}	Rank	Z_{CS}
1ASH	147	2.2/0.18	2.222 - 6.947	1	4.1	1	3.6
1BAB-B	146	1.5/0.16	0.702 - 6.920	1	3.4	1	3.1
1COL-A	197	2.4/0.18	12.399 - 30.284	1	4.8	1	4.6
1CPC-A	162	1.7/0.18	6.835 - 13.957	1	4.6	1	4.5
1ECD	136	1.4/0.00	1.471 - 6.188	1	3.8	1	4.1
1EMY	153	1.8/0.15	0.735 - 9.281	1	2.9	1	2.8
1FLP	142	1.5/0.17	1.734 - 7.227	1	4.1	1	4.1
1GDM	153	1.7/0.16	2.609 - 8.371	1	3.2	1	3.3
1HBG	147	1.5/0.15	2.050 - 6.896	1	3.8	1	4.1
1HBH-A	142	2.2/0.16	0.958 - 6.347	1	1.9	1	2.3
1HBH-B	146	2.2/0.16	1.024 - 7.330	1	2.3	1	2.4
1HDA-A	141	2.2/0.19	0.487 - 5.794	1	2.6	1	2.7
1HDA-B	145	2.2/0.19	0.545 - 5.644	1	2.9	1	2.9
1HLB	157	2.5/0.15	2.891 - 7.001	1	1.6	1	2.4
1HLM	158	2.9/0.19	2.973 - 8.737	20	-0.5	17	-0.3
1HSY	153	1.9/0.16	0.795 - 9.681	1	2.6	1	2.2
1ITH-A	141	2.5/0.15	1.638 - 6.071	1	4.5	1	4.5
1LHT	153	2.0/0.18	0.814 - 9.736	1	2.5	1	2.4
1MBA	146	1.6/0.19	1.829 - 7.314	1	3.9	1	3.8
1MBS	153	2.5/0.00	1.698 - 9.304	29	-1.2	30	-1.3
1MYG-A	153	1.8/0.20	0.479 - 9.562	1	2.6	1	2.5
1MYJ-A	153	1.9/0.21	0.623 - 7.944	1	2.9	1	3.0
1MYT	146	1.7/0.18	1.014 - 10.043	1	3.0	1	3.0
2DHB-A	141	2.8/0.00	0.648 - 6.358	14	-0.1	13	0.2
2DHB-B	146	2.8/0.00	0.858 - 7.062	13	0.2	12	0.3
2LHB	149	2.0/0.14	3.022 - 8.080	1	2.9	1	3.0
2PGH-A	141	2.8/0.15	0.707 - 6.485	16	-0.1	14	-0.1
2PGH-B	146	2.8/0.15	0.769 - 7.479	11	0.4	14	0.2
4SDH-A	145	1.6/0.16	2.273 - 6.429	1	3.4	1	3.3

(B).

PDB ID	Length (aa)	Resol (Å) / R	RMSD (Å) range of decoys	CS_{gl}		CS_{cp}	
				Rank	Z_{CS}	Rank	Z_{CS}
1ACY	232	3.0/0.21	1.167 - 4.455	1	5.7	1	5.7
1BAF	222	2.9/0.20	1.123 - 3.973	1	3.8	1	4.0
1BBD	231	2.8/0.19	1.610 - 4.537	1	3.5	1	4.2
1BBJ	221	3.1/0.18	0.797 - 4.047	1	2.3	1	2.3
1DBB	231	2.7/0.21	0.993 - 4.203	3	1.9	1	2.2
1DFB	231	2.7/0.18	1.383 - 4.826	2	1.8	3	1.8
1DVF	223	1.9/0.19	0.706 - 4.176	1	5.2	1	5.2
1EAP	225	2.5/0.19	1.557 - 4.537	1	4.9	1	4.7
1FAI	231	2.7/0.19	1.258 - 4.619	1	3.3	1	3.6
1FBI	229	3.0/0.19	1.322 - 4.993	2	2.5	1	1.9
1FGV	227	1.9/0.18	0.977 - 4.502	1	3.7	1	3.7
1FIG	227	3.0/0.22	1.423 - 4.388	19	0.5	12	0.8
1FLR	228	1.9/0.19	1.090 - 4.284	1	3.7	1	3.8
1FOR	225	2.8/0.17	0.998 - 4.200	1	2.6	1	2.9
1FPT	231	3.0/0.23	0.891 - 4.332	7	1.4	9	0.9
1FRG	233	2.8/0.19	0.952 - 4.134	1	5.5	1	5.7
1FVC	229	2.2/0.18	1.505 - 4.970	1	3.8	1	4.1
1FVD	227	2.5/0.17	0.798 - 4.137	1	3.3	1	3.7
1GAF	221	2.0/0.24	0.841 - 4.034	1	3.6	1	2.9
1GGI	226	2.8/0.18	0.978 - 4.247	1	3.2	1	3.5
1GIG	231	2.3/0.19	1.424 - 4.497	1	5.6	1	5.7
1HIL	233	2.0/0.19	0.910 - 4.303	1	5.2	1	5.2
1HKL	221	2.7/0.18	0.762 - 3.996	6	1.5	9	1.2
1IAI	228	2.9/0.21	1.069 - 4.594	5	1.7	5	1.5
1IBG	232	2.7/0.20	1.483 - 4.713	1	3.7	1	3.7
1IGC	227	2.6/0.16	0.960 - 4.237	12	0.9	14	0.8
1IGF	231	2.8/0.18	0.927 - 4.241	1	3.5	1	3.7
1IGI	231	2.7/0.17	1.579 - 4.266	1	3.7	1	4.0
1IGM	227	2.3/0.20	1.077 - 4.146	1	2.0	1	2.5
1IKF	233	2.5/0.16	1.330 - 5.105	1	4.6	1	4.8
1IND	222	2.2/0.18	1.183 - 4.213	1	3.7	1	3.5
1JEL	230	2.8/0.19	1.032 - 4.193	1	3.9	1	3.8
1JHL	224	2.4/0.21	1.100 - 4.083	1	3.7	1	3.7
1KEM	231	2.2/0.18	1.386 - 4.671	1	2.2	1	2.2
1MAM	227	2.5/0.21	1.473 - 4.170	1	3.3	1	3.8
1MCP	235	2.7/0.22	1.186 - 4.520	1	4.5	1	4.4
1MFA	229	1.7/0.16	2.053 - 4.805	1	4.6	1	4.7
1MLB	223	2.1/0.18	0.899 - 4.186	1	3.4	1	3.9

1MRD	225	2.4/0.19	1.311 - 4.150	1	3.8	1	3.5
1NBV	232	2.0/0.24	1.205 - 4.237	1	2.6	1	2.4
1NCB	227	2.5/0.16	1.167 - 4.385	1	3.1	1	3.1
1NGQ	229	2.4/0.19	1.407 - 4.065	1	3.8	1	3.9
1NMB	231	2.5/0.21	1.496 - 5.568	1	3.6	1	3.5
1NSN	224	2.9/0.19	1.163 - 4.033	15	0.8	19	0.5
1OPG	232	2.0/0.16	1.379 - 4.795	8	1.1	8	1.1
1PLG	228	2.8/0.16	1.036 - 4.168	1	4.3	1	3.9
1RMF	231	2.8/0.18	1.563 - 4.457	7	1.3	8	0.8
1TET	228	2.3/0.14	0.768 - 4.119	1	4.0	1	4.3
1UCB	228	2.5/0.20	1.183 - 4.346	1	2.9	1	3.2
1VFA	224	1.8/0.15	1.135 - 4.080	1	4.7	1	4.6
1VGE	231	2.0/0.18	1.619 - 5.004	1	5.4	1	5.2
1YUH	225	3.0/0.19	1.265 - 4.292	11	1.1	30	0.3
2CGR	228	2.2/0.21	0.925 - 4.226	1	3.7	1	4.3
2FB4	236	1.9/0.18	1.667 - 6.135	1	4.7	1	4.9
2FBJ	224	2.0/0.19	1.078 - 4.204	1	4.3	1	4.1
2GFB	227	3.0/0.21	1.639 - 4.123	1	4.1	1	4.2
3HFL	223	2.6/0.19	1.512 - 5.056	1	2.7	1	3.6
3HFM	220	3.0/0.24	1.029 - 3.993	5	1.5	8	1.0
6FAB	228	1.9/0.20	1.193 - 4.636	1	4.3	1	3.5
7FAB	219	2.0/0.16	1.845 - 4.547	1	4.7	1	4.8
8FAB	228	1.8/0.17	1.913 - 6.816	1	4.4	1	4.9

(C).

PDB ID	**Length (aa)**	**Resol (Å) / R**	**RMSD (Å) range of decoys**	***CS$_{gl}$***		***CS$_{cp}$***	
				Rank	**Z_{CS}**	**Rank**	**Z_{CS}**
1DVF	223	1.9/0.19	0.706 - 4.176	1	3.4	1	3.4
1FGV	227	1.9/0.18	0.977 - 4.502	1	2.9	1	2.9
1FLR	228	1.9/0.19	1.090 - 4.284	1	3.0	1	3.0
1FVC	229	2.2/0.18	1.505 - 4.970	1	2.9	1	2.9
1GAF	221	2.0/0.24	0.841 - 4.034	1	3.1	1	2.8
1HIL	233	2.0/0.19	0.910 - 4.303	1	3.6	1	3.5
1IND	222	2.2/0.18	1.183 - 4.213	1	3.1	1	2.9
1KEM	231	2.2/0.18	1.386 - 4.671	1	2.1	1	2.1
1MFA	229	1.7/0.16	2.053 - 4.805	1	3.3	1	3.3
1MLB	223	2.1/0.18	0.899 - 4.186	1	2.6	1	2.8
1NBV	232	2.0/0.24	1.205 - 4.237	1	2.3	1	2.1
1OPG	232	2.0/0.16	1.379 - 4.795	1	1.1	1	1.2
1VFA	224	1.8/0.15	1.135 - 4.080	1	3.2	1	3.2
1VGE	231	2.0/0.18	1.619 - 5.004	1	3.5	1	3.3
2CGR	228	2.2/0.21	0.925 - 4.226	1	2.9	1	3.1

2FB4	236	1.9/0.18	1.667 - 6.135	1	3.3	1	3.4
2FBJ	224	2.0/0.19	1.078 - 4.204	1	3.1	1	3.1
6FAB	228	1.9/0.20	1.193 - 4.636	1	2.8	1	2.4
7FAB	219	2.0/0.16	1.845 - 4.547	1	3.3	1	3.4
8FAB	228	1.8/0.17	1.913 - 6.816	1	3.2	1	3.5

Table S8. Performance of CS_{gl} and CS_{cp} in identifying the native crystal structures in the Rosetta all atom decoy set (20). N_{dec} and **Resol** stands for the number of decoys for each native structure and its crystallographic resolution. **RMSD** refers to the C^{α}-rms deviation of the decoy closest to the native in the set. Z_{CS} denotes the native Z-scores for the corresponding function (CS_{gl} /CS_{cp}). The dataset was downloaded from: http://trimer.tamu.edu/~daniel/decoys_11-14-01.tar.gz.

PDB ID	Length (aa)	class	Resol (Å)	N_{dec}	RMSD (Å)	CS_{gl}		CS_{cp}	
						Rank	Z_{CS}	Rank	Z_{CS}
1A32	65	All α	2.1	1610	0.90	1	5.3	1	5.0
1AIL	67	All α	1.9	1807	1.97	1	6.1	1	6.4
1AM3	57	All α	1.7	1898	1.80	1	7.4	1	6.5
1BQ9	53	All β	1.2	1825	2.79	1	7.1	1	6.8
1CC5	76	All α	2.5	1892	4.31	36	1.6	58	1.3
1CEI	85	All α	1.8	1897	4.57	1	8.6	1	8.5
1CSP	64	All β	2.45	1809	3.24	1	8.5	1	7.2
1CTF	67	α|β	1.70	1922	2.66	1	9.7	1	9.1
1DOL	62	α+β	2.4	1871	3.76	1	7.1	1	6.8
1HYP	75	All α	1.8	1893	4.05	1	6.3	1	6.1
1LFB	69	All α	2.8	1893	2.47	1	5.2	1	6.1
1MSI	60	All β	1.25	1894	5.40	1	10.6	1	10.1
1MZM	71	All α	1.78	1934	2.69	1	6.1	1	5.8
1ORC	56	α+β	1.54	1883	2.81	1	8.3	1	9.6
1PGX	57	α+β	1.66	1851	1.48	1	7.3	1	7.3
1PTQ	43	All β	1.95	1885	5.42	1	7.1	1	7.1
1R69	61	All α	2.00	1733	1.38	1	6.0	1	5.7
1TIF	59	α+β	1.80	1849	2.60	1	8.5	1	9.1
1TUC	61	All β	2.02	1894	4.48	1	6.6	1	6.6
1UTG	62	All α	1.34	1897	3.36	1	7.1	1	4.7
1VCC	77	α+β	1.60	1857	3.85	1	8.1	1	9.0
1VIF	48	All β	1.80	1896	0.44	1	5.7	1	6.4
2FXB	81	α|β	0.92	1800	5.48	2	3.3	1	3.7
5ICB	72	All α	1.50	1870	2.98	1	8.0	1	7.8
5PTI	55	α+β	1.00	1853	3.94	1	5.9	1	6.1

Table S9. Performance of CS_{gl} and CS_{cp} in CASP9 dataset (21). CASP9 experiment (conducted in July – August, 2010) consisted of 129 accepted targets out of which 18 were cancelled during the experiment. Of the remaining 111, 90 were X-ray crystal structures. T0543-2XRQ and T0605-3NMD were not considered in the calculation, the former due to its huge chain length (887 residues) and the latter being a single standalone helix. Results are tabulated for all other valid crystal structure targets. $\mathbf{N_{mod}}$ stands for the number of 'first models' used in the calculations. **Resol/$\mathbf{R_{obs}}$** represent the crystallographic resolution and R-factor (observed) of the native structure respectively. **RMSD** refers to the C^{α}-rms deviation of the model closest to the native in the set, calculated by Dali server **(24)**. $\mathbf{N_{res}}$ refers to the number of residues to be modeled for each target. **Rank** and $\mathbf{Z_{CS}}$ denotes the native rank and Z-scores for the corresponding function (CS_{gl} /CS_{cp}). For the target T0602-3NKZ, none of the models was superposible to the native by Dali server **(22)** (RMSD: N/A).

Target ID	N_{res} (aa)	PDB ID	Resol (Å) / R_{obs}	N_{mod}	RMSD (Å)	CS_{gl}		CS_{cp}	
						Rank	Z_{CS}	Rank	Z_{CS}
T0515	365	3MT1	2.50, 0.194	130	2.2	1	6.0	1	6.1
T0516	229	3NO6	1.65, 0.167	89	2.0	1	3.8	2	3.8
T0517	159	3PNX	1.92, 0.191	127	1.6	1	4.0	1	4.0
T0518	288	3NMB	2.40, 0.172	79	1.5	1	3.8	1	4.0
T0520	189	3MR7	2.60, 0.184	144	2.0	3	2.6	3	2.6
T0521	179	3MSE	2.10, 0.224	85	1.3	1	4.2	1	4.5
T0522	134	3NRD	2.06, 0.170	87	0.8	1	3.6	1	3.4
T0523	120	3MQO	1.70, 0.226	131	1.1	2	3.5	2	3.4
T0524	325	3MWX	1.45, 0.149	82	1.8	1	5.4	1	5.5
T0525	215	3MQZ	1.30, 0.150	82	2.1	1	5.1	1	5.1
T0526	290	3NRE	1.59, 0.162	124	2.0	1	4.4	1	4.3
T0527	142	3MR0	2.35, 0.224	85	1.9	1	3.5	1	3.6
T0528	388	3N0X	1.50, 0.149	89	2.5	1	5.8	1	5.7
T0529	569	3MWT	1.98, 0.183	50	1.3	1	2.4	1	2.4
T0530	115	3NPP	2.15, 0.181	83	1.6	1	3.8	2	3.5
T0532	506	3MX3	2.00, 0.167	50	1.9	1	3.0	1	3.0
T0534	384	3N8U	1.44, 0.168	126	2.7	1	3.7	1	3.7
T0537	381	3N6Z	1.30, 0.167	145	2.6	1	5.7	1	5.7
T0540	90	3MX7	1.76, 0.178	138	2.3	1	5.2	1	4.9
T0542	590	3N05	2.35, 0.233	50	1.5	1	2.7	1	2.7
T0547	611	3NZP	3.00, 0.183	50	2.8	1	2.0	1	2.0
T0548	106	3NNQ	2.69, 0.234	86	3.0	1	2.9	1	2.8
T0550	339	3NQK	2.61, 0.200	107	2.7	1	5.3	1	5.3
T0558	294	3NO2	1.35, 0.147	129	2.2	1	6.7	1	6.7
T0561	161	2XSE	1.90, 0.176	129	2.9	1	2.7	1	3.0
T0563	279	3ON7	2.20, 0.187	88	1.9	1	5.1	1	5.2
T0565	326	3NPF	1.72, 0.142	87	1.8	1	5.8	1	5.7

T0566	156	3N72	1.77, 0.191	140	2.1	1	3.2	1	3.7
T0567	145	3N70	2.80, 0.233	91	1.8	3	1.3	5	1.2
T0568	158	3N6Y	1.50, 0.182	114	2.5	1	3.8	1	4.1
T0570	258	3N70	2.80, 0.233	91	1.6	1	4.7	1	4.9
T0571	344	3N91	2.40, 0.183	108	2.3	1	3.9	1	3.8
T0573	311	3OOX	1.44, 0.170	92	1.8	1	4.0	1	4.2
T0574	126	3NRF	1.50, 0.186	113	2.2	1	4.4	1	4.4
T0575	216	3NRG	2.56, 0.209	93	1.6	3	2.7	1	2.8
T0576	172	3NA2	2.29, 0.190	110	2.1	1	5.9	1	6.0
T0578	164	3NAT	2.92, 0.206	107	2.1	1	3.4	1	3.4
T0580	105	3NBM	1.30, 0.134	141	1.3	1	3.1	1	3.0
T0581	136	3NPD	1.60, 0.162	93	1.8	1	3.3	1	3.3
T0582	222	3O14	1.70, 0.154	128	1.9	1	3.9	1	4.0
T0584	352	3NF2	2.20, 0.233	145	1.7	1	2.9	1	2.8
T0585	234	3NE8	1.24, 0.161	94	1.6	3	2.7	4	2.6
T0586	125	3NEU	1.58, 0.195	145	1.0	1	3.3	1	3.3
T0588	400	3NFV	1.95, 0.158	138	2.8	1	4.5	1	4.5
T0589	465	3NET	2.70, 0.219	87	1.8	1	3.8	1	3.8
T0591	406	3NRA	2.15, 0.156	88	1.8	1	3.8	1	3.8
T0592	144	3NHV	2.50, 0.199	141	1.3	1	3.1	1	2.8
T0593	208	3NGW	2.31, 0.187	88	2.2	1	3.8	1	3.9
T0594	140	3NI8	2.50, 0.226	141	1.4	1	3.3	1	3.4
T0596	213	3NI7	2.78, 0.250	143	1.3	7	1.7	7	1.8
T0597	429	3NIE	2.30, 0.225	86	1.9	1	2.8	1	2.7
T0598	161	3NJC	1.69, 0.188	135	2.5	7	1.5	5	1.4
T0599	399	3OS6	2.40, 0.173	89	1.3	1	3.5	1	3.5
T0601	449	3QTD	2.70, 0.227	82	1.4	1	3.4	1	3.4
T0602	123	3NKZ	2.11, 0.158	148	**N/A**	4	1.6	5	1.5
T0603	305	3NKD	1.95, 0.220	90	1.6	2	3.8	1	3.9
T0604	549	3NLC	2.15, 0.218	50	1.5	1	2.5	1	2.5
T0606	169	3NOH	1.60, 0.161	123	1.7	1	4.4	1	4.7
T0607	471	3PFE	1.50, 0.139	88	2.5	1	2.9	1	2.9
T0608	279	3NYY	1.60, 0.150	134	2.1	1	4.6	1	4.7
T0609	340	3OS7	1.80, 0.145	87	2.1	1	3.8	1	3.8
T0610	186	3OT2	1.96, 0.217	138	1.8	1	4.6	1	4.7
T0611	227	3NNR	2.49, 0.232	90	2.1	1	3.2	1	3.2
T0613	287	3OBI	1.95, 0.236	91	1.1	2	2.2	2	2.2
T0615	179	3NQW	2.90, 0.213	92	1.9	2	2.3	3	2.2
T0616	103	3NRT	2.54, 0.201	104	3.2	2	2.8	2	2.6
T0617	148	3NRV	2.00, 0.198	91	1.7	1	3.2	1	3.2
T0618	182	3NRH	1.80, 0.198	133	2.9	1	2.9	1	2.8
T0619	111	3NRW	1.70, 0.187	148	1.1	1	3.4	1	3.6
T0620	312	3NR8	2.80, 0.216	85	1.2	1	3.9	1	4.1
T0621	172	3NKG	2.00, 0.172	65	2.9	1	3.5	1	3.4
T0622	138	3NKL	1.90, 0.169	151	1.6	2	2.6	2	2.3

T0623	220	3NKH	2.50, 0.158	88	1.8	1	3.7	1	3.9
T0624	81	3NRL	1.90, 0.219	96	2.2	2	2.6	1	2.9
T0625	233	3ORU	1.11, 0.127	123	1.9	1	3.6	1	3.5
T0626	283	3O1L	2.20, 0.175	91	1.2	1	5.0	1	5.3
T0627	261	3OQL	2.54, 0.178	142	2.1	10	0.7	10	0.8
T0628	295	3NUW	2,09, 0.174	142	2.8	1	4.1	1	4.1
T0629	216	2XGF	2.20, 0.182	95	2.1	1	5.1	1	4.8
T0632	168	3NWZ	2.57, 0.233	90	1.0	3	2.5	1	3.1
T0634	140	3N53	2.20, 0.252	91	1.3	4	2.4	4	2.3
T0635	191	3N1U	1.80, 0.178	85	0.6	2	3.3	2	3.6
T0636	336	3P1T	2.60, 0.202	87	1.5	1	5.5	1	5.5
T0638	269	3NXH	2.58, 0.247	90	1.6	1	3.5	1	3.3
T0639	128	3NYM	1.90, 0.175	81	1.3	1	3.9	1	3.8
T0640	250	3NYW	2.16, 0.236	90	2.0	1	4.6	1	4.6
T0641	296	3NYI	1.90, 0.171	89	1.8	1	4.2	1	4.2
T0643	83	3NZL	1.20, 0.150	134	2.1	1	3.2	1	3.4

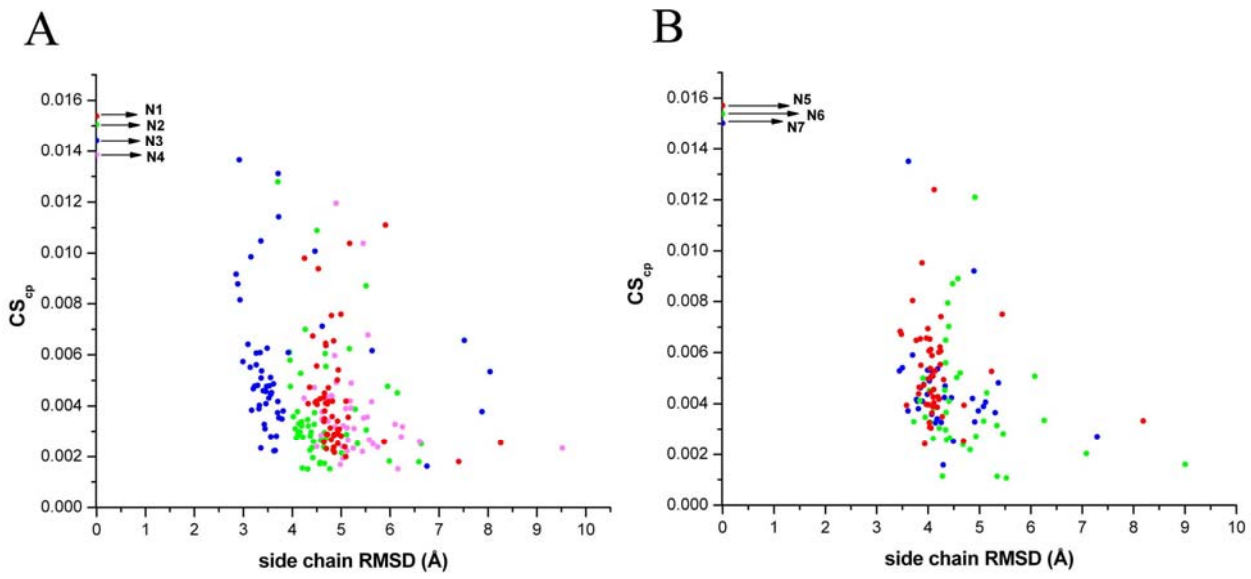

Figure S5. Complementarity Scores fall with increased deviations from the native side chain coordinates although with substantial scatter. CS_{cp} values plotted as a function of side chain rmsd's for 7 CASP9 targets (native and models). In **(A)** N1, N2, N3, N4 correspond to the native crystal structures of the CASP9 targets: T0610 (3OT2), T0526 (3NRE), T0580 (3NBM), T0593 (3NGW) and their corresponding models plotted in red, green, blue and light magenta respectively. In **(B)** N5, N6, N7 correspond to native CASP9 targets: T0640 (3NYW), T0582 (3O14), T0573 (3OOX), and their corresponding models plotted in red, green, and blue respectively. Models which were non-superposible to the native (by Dali server **(22)**) were excluded from the calculations. A similar pattern was obtained for CS_{gl}.

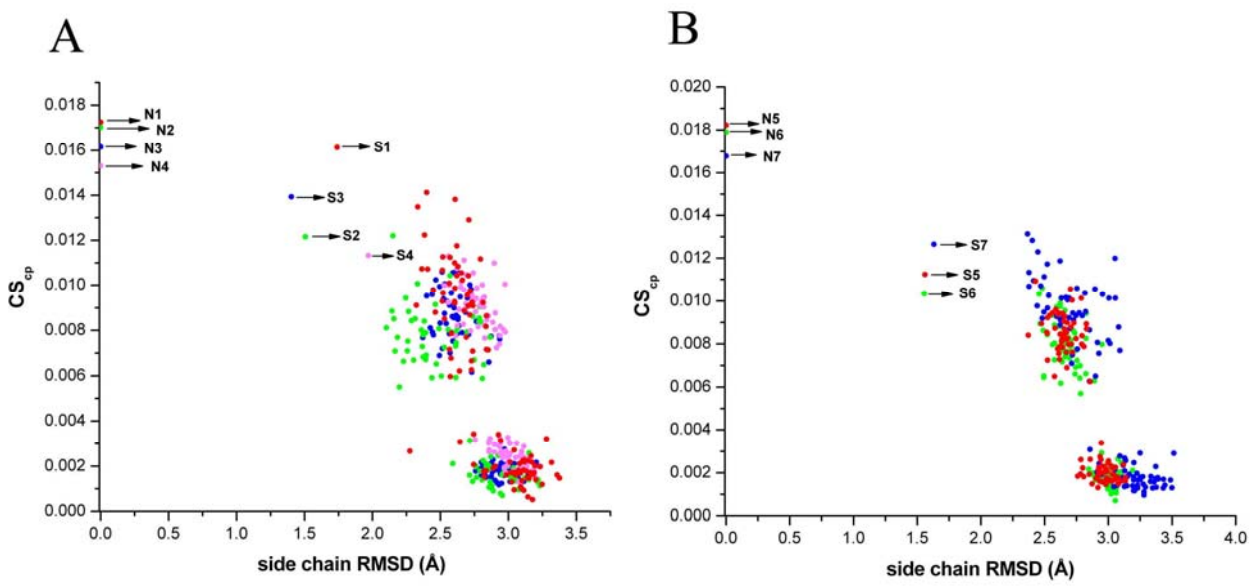

Figure S6. Complementarity Scores drop with increased errors incorporated in a native structure with backbone fixed and randomization of the side chain conformers. The figure plots the CS_{cp} values as a function of side chain rmsd's for 7 globular proteins from **DB2** and their models (See **Main Text**). The models for each native included 50 structures with randomized conformers, the same 50 structures, energy minimized subsequent to randomization of the side chain torsions and a unique solution generated by SCWRL4.0 **(23)** upon threading. In **(A)** N1, N2, N3, N4 and S1, S2, S3, S4 correspond to the native structures and the unique solutions generated by SCWRL4.0 respectively for 2O37 (red), 1DJR (green), 1C02 (blue), 1TUA (light magenta). In **(B)** N5, N6, N7 and S5, S6, S7 correspond to the native structures and the unique solutions generated by SCWRL4.0 respectively for 3CHM (red), 2O70 (green), 2PQR (blue). In most cases (except for 2PQR and 2O70), the models derived from SCWRL4.0 (c) gave values closest to the native, significantly distinguished from the other two clusters. The cluster corresponding to the energy minimized structures which gave superior scores relative to the cluster corresponding to the randomized conformers. Similar patterns have been obtained for CS_{gl}.

Dataset S2. Description of the 100 pairs of polypeptide chains (low sequence identity, same fold) used in the calculation from the PREFAB 4.0 database (24). PDB ID of each pair (separated by a front slash) along with their chain identifier (underscored) is given. The protein class, Fold and sequence identity (%) upon structural alignment (by Dali server **(22)**) are also provided in parentheses separated by comma. For polypeptide chains with their N and C terminal domains falling under two distinct folds both of which being separately used in the calculation involving different pairs, the relevant stretches of residues are also mentioned.

1BPW_A/1EUH_A (α|β, ALDH-like, 30)**,** 2CPL/2NUL (all β, Cyclophilin-like, 30), 1AMY/1JDC (α|β, TIM-barrel, 30), 1HYT/1EZM (α+β, Zincin-like, 30), 1PSR_A/4ICB (all α, EF hand-like, 30), 1AUK_A/1FSU (α|β, Alkaline phosphatase-like, 30), 1C8N_C/1SMV_A (all β, Nucleoplasmin-like, 29), 1A7S/2CGA_B (all β, Trypsin-like serine proteases, 29), 1AQU_A/1FML_B (α|β, P-loop containing nucleoside triphosphate hydrolases, 29), 1NOX/1BKJ_A (α+β, FMN-dependent nitroreductase-like, 29), 1G4U_S/1YTW (α|β, Phosphotyrosine protein phosphatases II, 29), 1AC5/1CPY (α|β, alpha/beta-Hydrolases, 29), 1CG5_B/1GCW_C (all α, Globin-like, 28), 1HTN/1RTM_1 (α+β, C-type lectin-like, 28), 1ZIN/5UKD_A (α|β, P-loop containing nucleoside triphosphate hydrolases, 28), 1PVL/7AHL_A (all β, Leukocidin-like, 28), 2DKB/2OAT_C (α|β, PLP-dependent transferase-like, 28), 1B5L/1WU3 (all β, 4-helical cytokines, 28), 1BBH_A/1JAF_B (all α, Four-helical up-and-down bundle, 27), 1CPC_A/1ALL_A (all α, Globin-like, 27), 1A3K/1LCL (all β, glucanases, 27), 1BDB/1YBV_A (α|β, NADP-binding Rossmann-fold domains 27), 1D4T_A/1JWO_A (α+β, SH2-like, 27), 1PNE/3NUL (α+β, Profilin-like, 27), 1NBC_A/1TF4_A: 461-605 (all β, Common fold of diphtheria toxin 27), 1IMB_A/1JP4_A (Multidomain, Carbohydrate phosphatase, 27), 1BBZ_A/1CKA_A (all β, SH3-like barrel, 26), 1HMT/1LFO (all β, Lipocalins, 26), 1NWP_A/1QHQ_A (all β, Cupredoxin-like, 26), 1BXK_A/1XEL (α|β, NADP-binding Rossmann-fold domains 26), 1DB1_A/2LBD (all α, Nuclear receptor ligand-binding domain, 26), 1FMK/1HCL (α+β, Protein kinase-like, 26), 1MUP/1BJ7 (all β, Lipocalins, 25), 1BEC/1TVD_A (all β, Immunoglobulin-like beta-sandwich, 25), 1CPT/1F4T_A (all α, Cytochrome P450, 25), 1DPE/1JEV_A (α|β, Periplasmic binding protein-like II, 25), 1ATG/1WOD (α|β, Periplasmic binding protein-like II 24), 1GCA/2DRI (α|β, Periplasmic_binding protein-like I, 24), 1OVA_A/1SEK (Multidomain, Serpins, 24), 1D3G_A/2DOR_A (α|β, TIM-barrel, 24), 1RHS/1E0C_A (α|β, Rhodanese, 24), 1VFR_A/1NOX (α+β, FMN-dependent nitroreductase-like, 24), 1PSR_A/1CLL (all α, EF Hand-like, 24), 1ERV/1THX (α|β, Thioredoxin, 23), 1DCI_A/1EF8_A (α|β, Crotonase, 23), 1F7C_A/1TX4_A (all α, GTPase activation domain, 23), 1CLC/1TF4_A: 1-460 (all α, alpha/alpha toroid, 23), 1G55_A/6MHT_A (α|β, SAM-dependent methyltransferases, 23), 1AQT/1E79_H (all β, Epsilon subunit of F1F0-ATP synthase N-terminal domain, 22), 1FIT/1KPF (α+β, HIT-like, 22), 1CIP_A/1HUR_A (α|β, P-loop containing nucleoside triphosphate hydrolases, 22), 1H8U/1RTM (α+β, C-type lectin-like, 21), 1DUN/1DUP_A (all β, beta-clip, 21), 1AH7/1CA1 (all α, Phospholipase C, 21), 1GKY/1NKS_A (α|β, P-loop containing nucleoside triphosphate hydrolases, 21), 1FDR/1FNC_A: 155-314 (Multidomain, Ferredoxin reductase-like, 21), 1ENY/1YBV_A (α|β, NADP-binding Rossmann-fold domains, 21), 1CY5/3YGS_P (all α, DEATH domain, 20), 1D4V_B/2TNF_A (all β, TNF-like, 21), 1AMP/1XJO (α|β, Phosphorylase, 20), 1BKR_A/1AOA (all α, CH domain-like, 20), 1DCE_B/1FT1_B (all α, alpha-alpha toroid, 19), 1DCF_A/3CHY(α|β, Flavodoxin-like, 19), 1H75_A/1KTE (α|β,

Thioredoxin fold, 19), 1BSV_A/1XEL (α|β, NADP-binding Rossmann-fold domains, 19), 1PDR/1KWA_A (all β, PDZ domain-like, 19), 1BQG/2MNR (α+β, Enolase N-terminal domain-like, 19), 1COZ_A/1IHO_A (α|β, Adenine nucleotide alpha hydrolase-like, 19), 1DAA_A/1EKF_A (Multidomain, D-aminoacid aminotransferase-like PLP-dependent enzymes, 18), 1HLB/2FAL (all α, Globin-like, 18), 1FAO_A/1MAI (all β, PH domain-like barrel, 18), 1NEU/1TVD_A (all β, Immunoglobulin-like beta-sandwich, 17), 1GAK_A/1LIS (all α, Fertilization-protein, 17), 1PLC/1RCY (all β, Cupredoxin-like, 17), 1CNT_3/1LKI (all α, 4-helical cytokines, 16), 1FNC_A: 19-154/2PIA (all β, Reductase, 16), 1PTY/1YTW (α|β, Phosphotyrosine protein phosphatases II16), 1AWC_A/2HTS (all α, DNA/RNA-binding 3-helical bundle, 16), 1MFA/1NEU (all β, Immunoglobulin-like beta-sandwich, 15), 1AM2/1AT0 (all β, Hedgehog C-terminal Hog auto processing domain, 14), 1RCB/1LKI (all α, 4-helical cytokines, 14), 1KUH/1IAE (α+β, Zincin-like, 14), 1MPG/1MUN (all α, DNA-glycosylase, 14), 1QFO_A/1CCZ_A (all β, Immunoglobulin-like beta-sandwich, 13), 1GSA/2HGS_A (α|β, Pre-ATP Grasp-domain, 13), 1FKN_A/1HVC (all β, Acid proteases, 13), 1DAP_A/1HYH_A (α|β, NADP-binding Rossmann-fold domains, 13), 1ASH/2FAL (all α, Globin-like, 12), 1JFU_A/1PRX_A (α|β, Thioredoxin, 12), 1FDS/1HE2_A (α|β, NADP-binding Rossmann-fold domains, 12), 1AX8/1RCB (all α, 4-helical cytokines, 12), 1HAN/1CJX_A (α+β, Glyoxalase, 12), 153L/1CNS_A (α+β, Lysozyme-like, 12), 1BI0/1SMT_A (all α, DNA/RNA-binding 3-helical bundle11), 1H97_A/2HBG (all α, Globin-like, 11), 1BOB_A/1QSM (α+β, Acyl-CoA N-acyltransferases, 11), 1FOF_A/1E25_A (Multidomain, β-lactamase, 11), 1AMJ/1ZYM (α|β, The "swivelling" beta/beta/alpha domain, 9), 1FR9_A/1FGX_A (α|β, Nucleotide-diphospho-sugar transferases 7), 1FLT_X/1QA9_A (all β, Immunoglobulin-like beta-sandwich, 6)

Table S10. Performance of CS_{gl} and CS_{cp} in detecting cross threaded sequences with low identity (≤ 30%) and same fold (from the PREFAB 4.0 database (24)), amongst random decoys. The first column tabulates PDB ID, chain identifier (underscored) of the two polypeptide chain separated by a front slash followed by their sequence identity (%) and rms deviations (Å) upon structural alignment respectively (in parentheses). N1 and N2 refer to the two native structures respectively. 1on2 and 2on1 represent the cross-threaded structures, where, for example, '1on2' refers to the sequence of N1 threaded onto the backbone of N2. Z_{CS} denotes the native Z-scores for the corresponding function (CS_{gl}/CS_{cp}).

PDB ID (seq id(%), rmsd(Å))	CS_{gl}						CS_{cp}					
	N1	N2	1on2	2on1	Random Decoys	Z_{CS}	N1	N2	1on2	2on1	Random Decoys	Z_{CS}
1BPW/1EUH (30, 1.7)	3.33	3.94	1.28	1.49	0.80 (± 0.12)	4.9	0.0132	0.0159	0.0050	0.0060	0.0030 (± 0.0006)	4.2
2CPL/2NUL (30, 1.8)	4.30	4.21	1.42	1.45	0.76 (± 0.14)	4.8	0.0175	0.0164	0.0061	0.0055	0.0031 (± 0.0007)	4.7
1AMY/1JDC (30, 2.0)	3.90	3.93	1.07	1.22	0.71 (± 0.15)	2.9	0.0160	0.0153	0.0041	0.0044	0.0026 (± 0.0006)	2.8
1HYT/1EZM (30, 2.1)	3.63	3.94	1.14	1.40	0.75 (± 0.16)	3.3	0.0140	0.0158	0.0050	0.0057	0.0032 (± 0.0006)	3.6
1PSR/4ICB (30, 2.1)	3.84	3.00	1.52	0.91	0.79 (± 0.15)	2.8	0.0151	0.0125	0.0053	0.0035	0.0030 (± 0.0007)	2.0
1AUK/1FSU (30, 2.4)	3.52	3.53	1.34	0.84	0.72 (± 0.14)	2.6	0.0134	0.0137	0.0048	0.0037	0.0030 (± 0.0005)	2.5
1C8N/1SMV (29, 1.5)	3.62	3.24	1.20	1.10	0.72 (± 0.13)	3.3	0.0143	0.0130	0.0048	0.0045	0.0030 (± 0.0005)	3.3
1A7S/2CGA (29, 1.7)	3.80	3.21	1.04	1.45	0.70 (± 0.15)	3.6	0.0147	0.0137	0.0042	0.0058	0.0029 (± 0.0006)	3.5
1AQU/1FML (29, 2.0)	4.96	3.99	1.41	1.15	0.74 (± 0.14)	3.9	0.0183	0.0157	0.0054	0.0044	0.0031 (± 0.0005)	3.6
1NOX/1BKJ (29, 2.1)	3.91	3.46	1.00	1.07	0.79 (± 0.15)	1.6	0.0152	0.0135	0.0041	0.0043	0.0030 (± 0.0006)	2.0
1G4U/1YTW (29, 2.5)	3.49	3.91	1.22	1.45	0.76 (± 0.15)	3.8	0.0147	0.0159	0.0046	0.0059	0.0029 (± 0.0007)	3.4
1AC5/1CPY (29, 2.5)	3.34	2.22	0.74	1.23	0.74 (± 0.13)	1.9	0.0130	0.0091	0.0030	0.0049	0.0027 (± 0.0007)	1.8
1CG5/1GCW (28, 1.5)	4.01	3.62	1.13	1.35	0.78 (± 0.16)	2.9	0.0161	0.0140	0.0050	0.0064	0.0031 (± 0.0007)	3.7
1HTN/1RTM (28, 1.6)	2.56	3.13	1.43	0.77	0.70 (± 0.15)	2.7	0.0108	0.0117	0.0059	0.0031	0.0027 (± 0.0007)	2.6
1ZIN/5UKD (28, 1.8)	4.04	3.81	1.70	1.45	0.72 (± 0.14)	6.1	0.0160	0.0145	0.0071	0.0057	0.0032 (± 0.0005)	6.4
1PVL/7AHL (28, 1.9)	3.44	3.40	1.34	1.32	0.75 (± 0.14)	4.1	0.0144	0.0130	0.0052	0.0052	0.0031 (± 0.0007)	3.0
2DKB/2OAT (28, 2.0)	3.37	3.65	1.18	0.92	0.71 (± 0.15)	2.3	0.0137	0.0144	0.0049	0.0036	0.0030 (± 0.0006)	2.1
1B5L/1WU3 (28, 2.1)	3.44	3.90	1.82	1.32	0.74 (± 0.13)	6.4	0.0142	0.0151	0.0071	0.0053	0.0031 (± 0.0005)	6.2
1BBH/1JAF (27, 1.3)	3.53	3.37	1.44	1.63	0.72 (± 0.15)	5.4	0.0147	0.0139	0.0053	0.0061	0.0031 (± 0.0005)	5.2
1CPC/1ALL (27, 1.3)	3.79	4.19	1.61	1.24	0.84 (± 0.15)	3.9	0.0150	0.0177	0.0065	0.0053	0.0031 (± 0.0006)	4.6
1A3K/1LCL	3.78	4.24	0.80	1.23	0.76	1.8	0.0140	0.0164	0.0035	0.0045	0.0029	1.8

(27, 1.5)					(± 0.14)						(± 0.0006)	
1BDB/1YBV (27, 1.6)	3.68	3.90	1.49	1.24	0.76 (± 0.13)	4.3	0.0147	0.0164	0.0057	0.0048	0.0030 (± 0.0006)	3.8
1D4T/1JWO (27, 1.7)	3.86	2.23	1.40	1.79	0.71 (± 0.15)	5.9	0.0156	0.0091	0.0054	0.0073	0.0031 (± 0.0006)	5.4
1PNE/3NUL (27, 1.8)	4.00	4.01	1.28	1.90	0.75 (± 0.12)	7.0	0.0159	0.0147	0.0054	0.0075	0.0028 (± 0.0006)	6.1
1NBC/1TF4 (27, 2.1)	4.29	4.26	1.08	0.95	0.69 (± 0.13)	2.5	0.0156	0.0167	0.0042	0.0036	0.0029 (± 0.0006)	1.8
1IMB/1JP4 (27, 2.6)	3.76	3.89	1.29	1.01	0.71 (± 0.16)	2.8	0.0152	0.0147	0.0053	0.0040	0.0029 (± 0.0007)	2.5
1BBZ/1CKA (26, 1.4)	2.38	3.78	1.64	1.17	0.70 (± 0.14)	5.0	0.0098	0.0180	0.0062	0.0041	0.0027 (± 0.0006)	4.1
1HMT/1LFO (26, 1.7)	2.91	2.96	0.98	0.91	0.68 (± 0.13)	2.1	0.0123	0.0110	0.0042	0.0041	0.0028 (± 0.0006)	2.3
1NWP/1QHQ (26, 1.8)	3.42	3.97	1.43	1.17	0.71 (± 0.13)	4.5	0.0135	0.0161	0.0061	0.0047	0.0026 (± 0.0004)	5.6
1BXK/1XEL (26, 2.1)	3.68	4.04	1.13	1.03	0.73 (± 0.13)	2.7	0.0142	0.0163	0.0045	0.0040	0.0030 (± 0.0005)	2.5
1DB1/2LBD (26, 2.1)	3.81	3.53	1.45	1.69	0.78 (± 0.13)	6.1	0.0157	0.0148	0.0057	0.0071	0.0031 (± 0.0005)	6.6
1FMK/1HCL (26, 2.5)	3.28	3.82	1.41	1.59	0.77 (± 0.14)	5.2	0.0132	0.0152	0.0053	0.0063	0.0029 (± 0.0006)	4.8
1MUP/1BJ7 (25, 1.8)	2.48	3.51	1.21	0.79	0.74 (± 0.14)	1.9	0.0105	0.0137	0.0051	0.0034	0.0029 (± 0.0007)	1.9
1BEC/1TVD (25, 1.9)	3.18	3.84	1.53	1.38	0.71 (± 0.14)	5.3	0.0124	0.0147	0.0059	0.0054	0.0028 (± 0.0006)	4.8
1CPT/1F4T (25, 2.2)	3.52	3.85	1.56	1.04	0.75 (± 0.17)	3.2	0.0139	0.0153	0.0063	0.0043	0.0029 (± 0.0006)	4.0
1DPE/1JEV (25, 6.7)	4.26	4.01	1.11	1.03	0.78 (± 0.15)	1.9	0.0174	0.0161	0.0044	0.0040	0.0029 (± 0.0007)	1.9
1ATG/1WOD (24, 1.8)	3.51	3.80	1.09	1.32	0.78 (± 0.12)	3.5	0.0145	0.0144	0.0043	0.0056	0.0031 (± 0.0005)	3.7
1GCA/2DRI (24, 1.9)	3.87	3.70	1.36	1.54	0.77 (± 0.14)	4.9	0.0159	0.0146	0.0054	0.0060	0.0032 (± 0.0005)	5.2
1OVA/1SEK (24, 2.1)	3.68	3.26	1.63	1.17	0.69 (± 0.15)	4.7	0.0146	0.0126	0.0064	0.0047	0.0028 (± 0.0005)	5.5
1D3G/2DOR (24, 2.3)	3.12	3.81	1.12	0.86	0.71 (± 0.12)	2.3	0.0124	0.0151	0.0046	0.0036	0.0030 (± 0.0005)	2.2
1RHS/1E0C (24, 2.3)	3.59	3.89	1.21	1.18	0.76 (± 0.13)	3.3	0.0143	0.0155	0.0047	0.0045	0.0029 (± 0.0005)	3.4
1VFR/1NOX (24, 2.6)	4.12	3.91	1.06	1.30	0.71 (± 0.13)	3.6	0.0168	0.0152	0.0042	0.0054	0.0030 (± 0.0006)	3.0
1PSR/1CLL (24, 3.5)	3.84	3.00	1.67	0.77	0.72 (± 0.14)	3.6	0.0151	0.0125	0.0077	0.0033	0.0029 (± 0.0006)	4.3
1ERV/1THX (23, 1.8)	4.37	4.21	1.12	1.01	0.74 (± 0.13)	2.5	0.0185	0.0156	0.0044	0.0038	0.0027 (± 0.0006)	2.3
1DCI/1EF8 (23, 2.0)	3.43	3.52	1.42	1.50	0.76 (± 0.14)	5.0	0.0138	0.0136	0.0060	0.0057	0.0029 (±0.0007)	4.2
1F7C/1TX4 (23, 2.4)	4.27	4.23	1.82	1.60	0.79 (±0.16)	5.8	0.0160	0.0181	0.0071	0.0064	0.0031 (±0.0006)	6.1
1CLC/1TF4 (23, 2.5)	4.09	4.44	0.95	1.38	0.78 (± 0.12)	3.2	0.0164	0.0172	0.0040	0.0054	0.0031 (± 0.0005)	3.2
1G55/6MHT (23, 2.6)	4.23	3.85	1.16	1.11	0.74 (± 0.12)	3.3	0.0151	0.0150	0.0046	0.0044	0.0028 (± 0.0006)	2.8

1AQT/1E79 (22, 1.7)	3.34	2.53	1.45	1.81	0.76 (± 0.15)	5.8	0.0137	0.0096	0.0056	0.0065	0.0031 (± 0.0005)	5.9
1FIT/1KPF (22, 2.3)	4.70	3.70	1.08	1.22	0.75 (± 0.14)	2.9	0.0180	0.0146	0.0043	0.0045	0.0032 (± 0.0005)	2.4
1CIP/1HUR (22, 2.4)	3.82	3.31	1.10	1.22	0.71 (± 0.14)	3.2	0.0154	0.0135	0.0045	0.0047	0.0030 (± 0.0005)	3.2
1H8U/1RTM (21, 1.9)	3.56	3.41	1.02	1.40	0.76 (± 0.12)	3.8	0.0152	0.0146	0.0044	0.0056	0.0032 (± 0.0006)	3.0
1DUN/1DUP (21, 1.9)	4.07	3.68	1.72	1.31	0.75 (± 0.15)	5.1	0.0154	0.0136	0.0064	0.0059	0.0028 (± 0.0007)	4.8
1AH7/1CA1 (21, 2.1)	4.04	3.74	0.95	1.17	0.72 (± 0.12)	2.8	0.0169	0.0144	0.0040	0.0043	0.0029 (± 0.0005)	2.5
1GKY/1NKS (21, 3.2)	3.52	3.74	0.90	0.97	0.67 (± 0.13)	2.0	0.0148	0.0144	0.0041	0.0037	0.0027 (± 0.0006)	2.0
1FDR/1FNC (21, 2.4)	3.94	3.71	1.01	0.76	0.67 (± 0.12)	1.8	0.0159	0.0146	0.0043	0.0035	0.0029 (± 0.0006)	1.7
1ENY/1YBV (21, 2.2)	3.11	3.89	1.31	1.24	0.73 (± 0.10)	5.5	0.0122	0.0164	0.0052	0.0050	0.0029 (± 0.0005)	4.4
1D4V/2TNF (21, 1.7)	3.24	3.04	1.12	1.22	0.77 (± 0.14)	2.9	0.0131	0.0162	0.0046	0.0045	0.0031 (± 0.0006)	2.4
1CY5/3YGS (20, 1.5)	3.83	3.93	1.57	1.86	0.77 (± 0.13)	7.3	0.0156	0.0157	0.0065	0.0079	0.0032 (± 0.0005)	8.0
1AMP/1XJO (20, 1.9)	3.69	3.80	1.12	1.01	0.71 (± 0.12)	3.0	0.0150	0.0146	0.0045	0.0042	0.0027 (± 0.0003)	2.8
1BKR/1AOA (20, 2.1)	4.86	3.60	1.50	1.28	0.66 (± 0.15)	4.9	0.0193	0.0153	0.0057	0.0053	0.0026 (± 0.0007)	4.1
1DCE/1FT1 (19, 2.0)	3.60	3.41	1.35	1.37	0.77 (± 0.11)	5.4	0.0148	0.0136	0.0056	0.0056	0.0031 (± 0.0005)	5.0
1DCF/3CHY (19, 2.0)	3.18	4.44	1.12	1.23	0.74 (± 0.13)	3.3	0.0127	0.0171	0.0040	0.0045	0.0027 (± 0.0006)	2.6
1H75/1KTE (19, 2.0)	3.82	3.94	1.04	1.75	0.75 (± 0.11)	5.9	0.0154	0.0163	0.0042	0.0073	0.0031 (± 0.0005)	5.3
1BSV/1XEL (19, 2.3)	3.76	4.04	1.00	1.51	0.79 (± 0.10)	4.6	0.0147	0.0163	0.0040	0.0061	0.0029 (± 0.0005)	4.3
1PDR/1KWA (19, 2.4)	3.62	2.85	1.34	0.75	0.73 (± 0.10)	3.2	0.0140	0.0103	0.0056	0.0033	0.0027 (± 0.0006)	2.9
1BQG/2MNR (19, 2.4)	3.34	3.87	1.08	1.36	0.70 (± 0.12)	4.3	0.0132	0.0152	0.0044	0.0057	0.0030 (± 0.0005)	4.1
1COZ/1IHO (19, 3.1)	3.97	3.67	0.81	1.29	0.72 (± 0.13)	2.5	0.0143	0.0144	0.0036	0.0053	0.0031 (± 0.0005)	2.7
1DAA/1EKF (18, 2.1)	4.11	3.37	1.08	0.96	0.73 (± 0.12)	2.4	0.0159	0.0135	0.0043	0.0038	0.0030 (± 0.0005)	2.1
1HLB/2FAL (18, 2.5)	2.38	3.61	1.25	1.35	0.69 (± 0.17)	3.6	0.0094	0.0147	0.0049	0.0052	0.0030 (± 0.0005)	4.1
1FAO/1MAI (18, 3.1)	3.79	3.47	0.93	1.38	0.76 (± 0.12)	3.3	0.0155	0.0155	0.0039	0.0056	0.0029 (± 0.0006)	3.1
1NEU/1TVD (17, 2.0)	3.76	3.82	0.78	1.17	0.69 (± 0.15)	1.9	0.0149	0.0149	0.0034	0.0047	0.0030 (± 0.0006)	1.8
1GAK/1LIS (17, 2.0)	4.03	4.33	1.32	0.94	0.71 (± 0.12)	3.5	0.0171	0.0177	0.0054	0.0035	0.0030 (± 0.0005)	2.9
1PLC/1RCY (17, 2.5)	3.79	3.97	1.03	1.10	0.73 (± 0.15)	2.3	0.0146	0.0147	0.0040	0.0040	0.0028 (± 0.0006)	2.0
1CNT/1LKI (16, 2.4)	3.75	3.69	1.06	1.30	0.68 (± 0.15)	3.3	0.0161	0.0150	0.0044	0.0053	0.0030 (± 0.0006)	3.1
1FNC/2PIA	3.71	3.62	1.01	0.91	0.71	1.7	0.0146	0.0149	0.0045	0.0036	0.0029	1.9

(16, 2.5)					(± 0.15)						(± 0.0006)	
1PTY/1YTW (16, 2.8)	4.01	3.92	0.78	1.31	0.72 (± 0.13)	2.5	0.0163	0.0162	0.0029	0.0053	0.0027 (± 0.0005)	2.8
1AWC/2HTS (16, 2.8)	4.13	4.92	1.27	0.76	0.67 (± 0.11)	3.1	0.0163	0.0198	0.0052	0.0032	0.0026 (± 0.0005)	3.2
1MFA/1NEU (15, 1.8)	3.90	3.76	0.93	1.07	0.73 (± 0.15)	2.3	0.0157	0.0149	0.0040	0.0043	0.0028 (± 0.0005)	2.7
1AM2/1AT0 (14, 2.1)	3.21	2.30	1.24	0.85	0.75 (± 0.12)	2.4	0.0133	0.0094	0.0053	0.0038	0.0031 (± 0.0005)	2.9
1RCB/1LKI (14, 2.8)	2.47	3.70	0.94	1.23	0.74 (± 0.11)	3.1	0.0093	0.0150	0.0035	0.0052	0.0028 (± 0.0005)	3.1
1KUH/1IAE (14, 2.9)	3.92	3.90	0.69	0.87	0.67 (± 0.10)	1.1	0.0146	0.0167	0.0034	0.0035	0.0029 (± 0.0005)	1.1
1MPG/1MUN (14, 3.7)	3.89	3.91	0.95	0.79	0.66 (± 0.10)	2.1	0.0160	0.0160	0.0038	0.0032	0.0025 (± 0.0005)	2.0
1QFO/1CCZ (13, 2.2)	4.11	4.08	1.36	1.34	0.70 (± 0.13)	5.0	0.0168	0.0172	0.0057	0.0059	0.0029 (± 0.0005)	5.8
1GSA/2HGS (13, 3.5)	3.72	3.33	0.91	0.68	0.71 (± 0.12)	0.7	0.0150	0.0135	0.0034	0.0029	0.0026 (± 0.0005)	1.1
1FKN/1HVC (13, 3.8)	3.97	3.67	0.80	1.24	0.74 (± 0.10)	2.3	0.0158	0.0152	0.0037	0.0046	0.0028 (± 0.0005)	2.7
1DAP/1HYH (13, 4.6)	3.19	3.89	1.27	1.01	0.76 (± 0.12)	3.2	0.0115	0.0151	0.0051	0.0040	0.0029 (± 0.0005)	3.3
1ASH/2FAL (12, 1.8)	4.10	3.65	0.89	1.43	0.75 (± 0.13)	3.2	0.0158	0.0152	0.0032	0.0060	0.0027 (± 0.0006)	3.2
1JFU/1PRX (12, 2.3)	3.66	3.60	1.01	1.00	0.70 (± 0.13)	2.3	0.0155	0.0153	0.0041	0.0040	0.0030 (± 0.0005)	2.1
1FDS/1HE2 (12, 2.5)	3.85	3.88	0.80	1.23	0.70 (± 0.15)	2.2	0.0148	0.0150	0.0035	0.0046	0.0029 (± 0.0005)	2.3
1AX8/1RCB (12, 2.6)	3.86	2.47	1.64	0.90	0.71 (± 0.13)	4.3	0.0159	0.0093	0.0062	0.0037	0.0031 (± 0.0006)	3.1
1HAN/1CJX (12, 2.9)	3.64	4.08	0.85	0.70	0.65 (± 0.11)	1.1	0.0151	0.0162	0.0034	0.0030	0.0026 (± 0.0005)	1.2
153L/1CNS (12, 3.8)	3.98	3.51	0.79	0.65	0.67 (± 0.11)	0.5	0.0160	0.0151	0.0028	0.0028	0.0026 (± 0.0005)	0.4
1BI0/1SMT (11, 2.2)	3.13	3.27	1.29	1.02	0.70 (± 0.11)	4.1	0.0121	0.0142	0.0052	0.0043	0.0028 (± 0.0005)	3.9
1H97/2HBG (11, 2.4)	4.25	3.99	0.83	1.40	0.76 (± 0.15)	2.4	0.0167	0.0152	0.0034	0.0054	0.0029 (± 0.0006)	2.5
1BOB/1QSM (11, 3.4)	3.84	3.20	1.03	0.89	0.69 (± 0.14)	1.9	0.0150	0.0131	0.0039	0.0039	0.0029 (± 0.0005)	2.0
1FOF/1E25 (11, 3.4)	3.73	3.90	0.86	0.96	0.67 (± 0.12)	2.0	0.0154	0.0152	0.0036	0.0039	0.0026 (± 0.0005)	2.3
1AMJ/1ZYM (9, 3.7)	3.92	1.94	1.09	1.95	0.78 (± 0.15)	4.9	0.0158	0.0084	0.0042	0.0072	0.0032 (± 0.0006)	4.2
1FR9/1FGX (7, 3.9)	4.40	3.63	0.94	0.91	0.67 (± 0.12)	2.1	0.0179	0.0145	0.0037	0.0037	0.0027 (± 0.0005)	2.0
1FLT/1QA9 (6, 2.4)	2.97	2.30	0.86	0.75	0.69 (± 0.14)	1.0	0.0127	0.0090	0.0034	0.0031	0.0026 (± 0.0005)	1.3

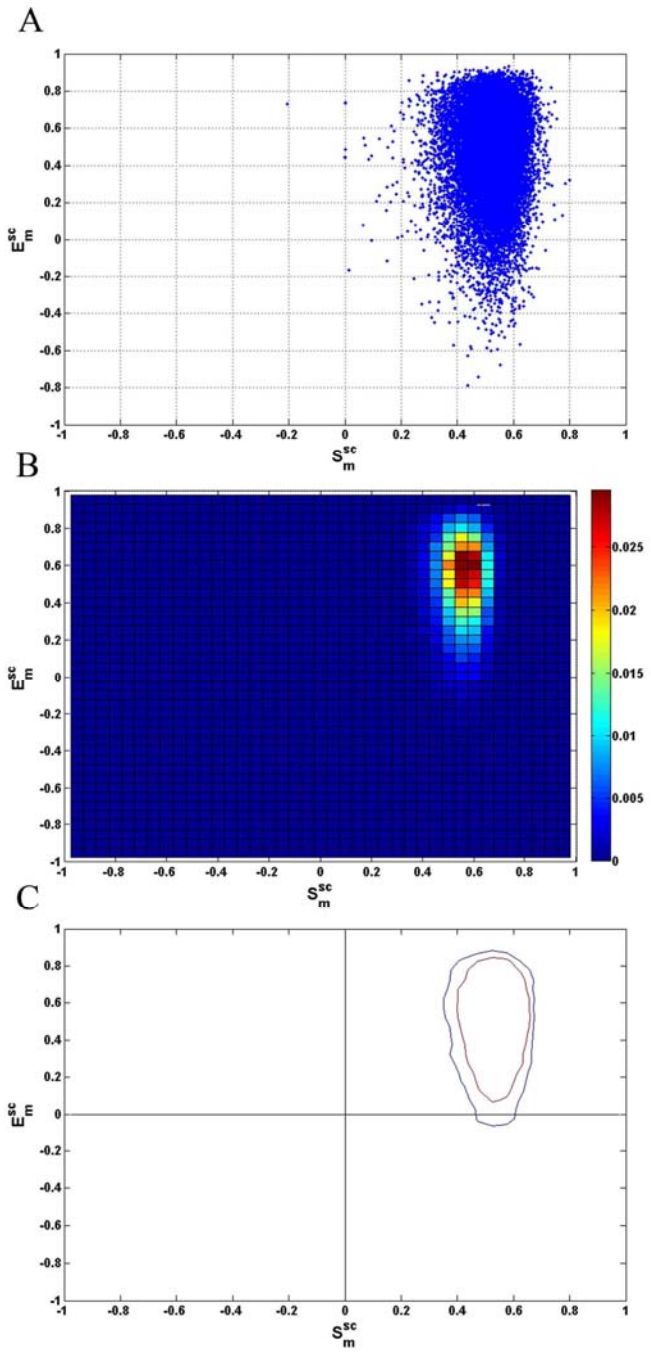

Figure S7. CP1. The Complementarity plot for burial bin 1. All plots span -1.0 to 1.0 in both the X (S_m^{sc}) and Y (E_m^{sc}) axes. (A) shows the distribution of (23850) residues in the first burial bin ($0.0 \leq \text{Bur} \leq 0.05$) from **DB2.** (B) shows the probabilities to locate a point in each grid (P_{grid}) of the plot with a color bar corresponding to the probability values and (C) the contour plot based on contour levels ($P_{grid} \geq 0.005$ for the first contour level and ≥ 0.002 for the second) (see **Main Text).**

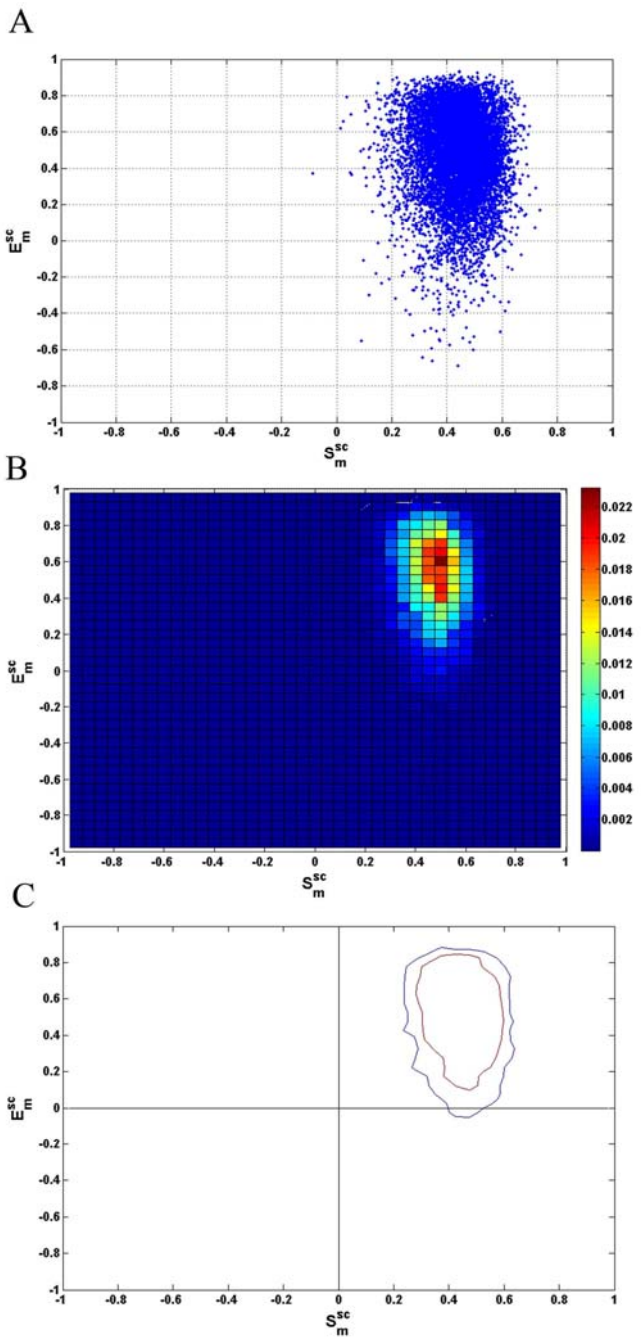

Figure S8. CP2. The Complementarity plot for burial bin 2. All plots span -1.0 to 1.0 in both the X (S_m^{sc}) and Y (E_m^{sc}) axes. (A) shows the distribution of (10624) residues in the second burial bin ($0.05 < \text{Bur} \leq 0.15$) from **DB2.** (B) shows the probabilities to locate a point in each grid (P_{grid}) of the plot with a color bar corresponding to the probability values and (C) the contour plot based on contour levels ($P_{grid} \geq 0.005$ for the first contour level and ≥ 0.002 for the second) (see **Main Text).**

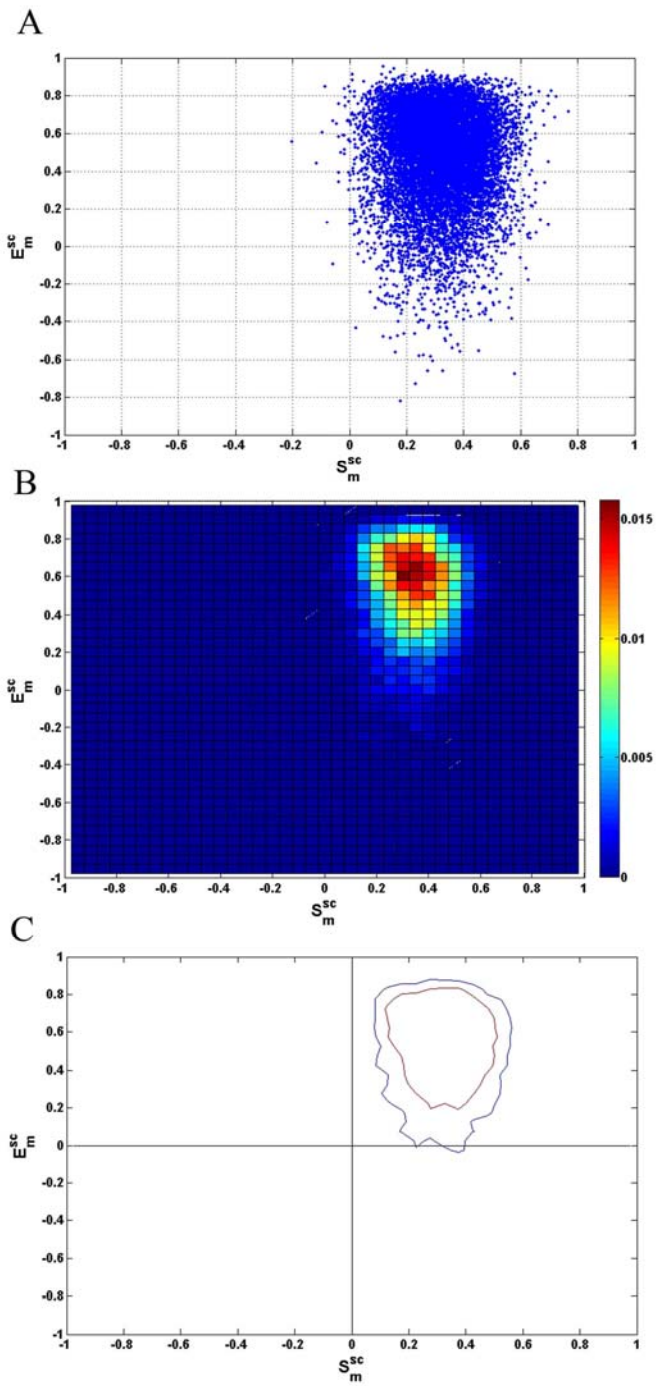

Figure S9. CP3. The Complementarity plot for burial bin 3. All plots span from -1.0 to 1.0 in both the X (S_m^{sc}) and Y (E_m^{sc}) axes. (A) shows the distribution of (13255) residues in the third burial bin ($0.15 < \text{Bur} \leq 0.30$) from **DB2.** (B) shows the probabilities to locate a point in each grid (P_{grid}) of the plot with a color bar corresponding to the probability values and (C) the contour plot based on contour levels ($P_{grid} \geq 0.005$ for the first contour level and ≥ 0.002 for the second) (see **Main Text).**

Table S11. Distribution of residues from Superseded and Upgraded pairs of structures plotted in CP1 (0.0 ≤ Bur ≤ 0.05), CP2 (0.05 < Bur ≤ 0.15) and CP3 (0.15 < Bur ≤ 0.30). Residues possibly in error in the superseded PDB files were detected by side chain rms deviation (method 1) **(A)** and deviation in (χ_1) torsion angles (method 2) **(B)** (see **Main Text).** The 1st and 2nd rows for each pair correspond to the upgraded and superseded PDB files respectively. The resolution, crystallographic R-factor and chain identifier (underscored) as was mentioned in the PDB (ftp://ftp.wwpdb.org/pub/pdb/data/status/obsolete.dat) are given in parentheses below each structure. All metal ions (as given in parentheses) were considered in the calculation. The number of residues from each PDB file falling in the three burial bins is given along with their distribution in 'probable', 'less probable' and 'improbable' regions in the appropriate CP (in parentheses separated by comma). For the 1st method (rmsd), all residues of the superseded structure, '2FD1' were considered to be in error since it was non-superposable onto its upgraded partner '5FD1' by Dali server **(24)**. For 2HFL, 3HFL and 1YQV (the same antibody-antigen complex, solved at gradually improved resolutions), all chains (L, H, Y) were considered in the calculations.

(A)

Pair No.	Pair	Bin1 0.0 ≤ Bur ≤ 0.05	Bin2 0.05 < Bur ≤ 0.15	Bin3 0.15 < Bur ≤ 0.30
1	1ABE_A (1.7, 0.137)	26 (21, 3, 2)	3 (3, 0, 0)	8 (5, 3, 0)
	1ABP (NA, NA)	23 (5, 3, 15)	3 (1, 0, 2)	6 (3, 0, 3)
2	2AMG_A (Ca^{+2}) (2.0, 0.178)	15 (13, 0, 2)	-	2 (1, 0, 1)
	1AMG (Ca^{+2}) (2.2, 0.174)	14 (11, 1, 2)	-	3 (0, 1, 2)
3	3HFL (2.65, 0.196)	23 (19, 2, 2)	7 (3, 2, 2)	11 (4, 3, 4)
	2HFL (NA, NA)	25 (11, 2, 12)	6 (3, 2, 1)	13 (3, 6, 4)
4	1YQV (1.7, 0.195)	35 (28, 2, 5)	9 (6, 1, 2)	10 (7, 1, 2)
	2HFL (NA, NA)	35 (14, 4, 17)	9 (3, 2, 4)	11 (4, 4, 3)
5	2F19 (2.8, 0.182)	27 (17, 5, 5)	9 (6, 2, 1)	18 (8, 4, 6)
	1F19 (NA, NA)	20 (1, 0, 19)	18 (1, 2, 15)	13 (3, 3, 7)
6	5FD1 (1.9, 0.215)	21 (17, 1, 3)	11 (7, 2, 2)	17 (12, 3, 2)

	2FD1 (NA, NA)	4 (0, 1, 3)	10 (3, 0, 7)	11 (2, 4, 5)
7	3BJL_A (2.3, 0.165)	6 (3, 1, 2)	1 (1, 0, 0)	2 (1, 1, 0)
	1BJL (NA, NA)	2 (1, 0, 1)	1 (0, 1, 0)	4 (0, 1, 3)
8	2DTR_A (Co^{+2}) (1.9, 0.171)	5 (5, 0, 0)	2 (1, 1, 0)	2 (1, 1, 0)
	1DTR (2.8, 0.210)	5 (4, 0, 1)	-	4 (1, 2, 1)
9	7FAB_L (2.0, 0.169)	1 (1, 0, 0)	1 (0, 1, 0)	1 (1, 0, 0)
	1FAB_L (NA, NA)	2 (2, 0, 0)	2 (1, 0, 1)	1 (0, 1, 0)
10	1A45_A (2.3, 0.186)	14 (11, 2, 1)	1 (0, 1, 0)	2 (2, 0, 0)
	2GCR (NA, NA)	11 (8, 1, 2)	-	5 (0, 1, 4)
11	3EFZ_A (2.08, 0.227)	2 (2, 0, 0)	-	1 (1, 0, 0)
	2IJP (2.14, 0.237)	2 (0, 1, 1)	-	1 (1, 0, 0)
12	5PTP_A (Ca^{+2}) (1.34, 0.152)	10 (8, 2, 0)	2 (2, 0, 0)	7 (4, 3, 0)
	1PTP (Mg^{+2}) (NA, NA)	9 (6, 0, 3)	2 (2, 0, 0)	6 (1, 2, 3)
13	4AAH_A (Ca^{+2}) (2.4, 0.152)	14 (12, 0, 2)	1 (1, 0, 0)	3 (2, 1, 0)
	3AAH_A (Ca^{+2}) (2.4, 0.201)	12 (5, 1, 6)	2 (2, 0, 0)	4 (0, 2, 2)
14	6LDH_A (2.0, 0.202)	6 (4, 1, 1)	5 (1, 0, 4)	5 (2, 0, 3)
	2LDH (NA, NA)	4 (2, 1, 1)	4 (0, 2, 2)	6 (2, 2, 2)
15	3DFR_A (1.7, 0.152)	16 (16, 0, 0)	6 (6, 0, 0)	7 (5, 2, 0)
	1DFR (NA, NA)	10 (1, 3, 6)	6 (0, 0, 6)	6 (0, 4, 2)
16	5CPV_A (Ca^{+2}) (1.6, 0.187)	3 (2, 0, 1)	-	2 (2, 0, 0)
	1CPV (Ca^{+2}) (1.85, NA)	1 (0, 0, 1)	-	-

17	3F70_A (2.1, 0.203)	23 (18, 2, 3)	2 (1, 0, 1)	10 (6, 0, 4)
	3DBB_A (2.15, 0.212)	31 (25, 3, 3)	3 (2, 0, 1)	7 (1, 4, 1)
18	4CAA (2.9, 0.199)	14 (11, 2, 1)	2 (2, 0, 0)	2 (2, 0, 0)
	1CT3_A (3.1, 0.189)	15 (9, 4, 2)	1 (1, 0, 0)	1 (1, 0, 0)
19	3CPA_A (Zn^{+2}) (2.0, NA)	5 (5, 0, 0)	-	2 (1, 0, 1)
	1CPA (Zn^{+2}) (NA, NA)	3 (1, 0, 2)	2 (2, 0, 0)	1 (0, 1, 0)
20	1P2Z_A (2.2, 0.178)	88 (71, 11, 6)	35 (24, 4, 7)	61 (36, 17, 8)
	1DHX_A (2.9, 0.225)	74 (19, 19, 36)	38 (12, 3, 23)	56 (13, 23, 20)

(B)

Pair No.	**Pair**	**Bin1 $0.0 \leq Bur \leq 0.05$**	**Bin2 $0.05 < Bur \leq 0.15$**	**Bin3 $0.15 < Bur \leq 0.30$**
1	1ABE_A (1.7, 0.137)	29 (22, 6, 1)	9 (9, 0, 0)	16 (10, 5, 1)
	1ABP (NA, NA)	26 (7, 2, 17)	9 (5, 0, 4)	12 (7, 2, 3)
2	2AMG_A (Ca^{+2}) (2.0, 0.178)	9 (7, 1, 1)	1 (1, 0, 0)	3 (2, 0, 1)
	1AMG (Ca^{+2}) (2.2, 0.174)	9 (6, 1, 2)	2 (0, 0, 2)	1 (0, 0, 1)
3	3HFL (2.65, 0.196)	22 (15, 1, 6)	14 (7, 2, 5)	23 (8, 5, 10)
	2HFL (NA, NA)	23 (8, 2, 13)	12 (5, 2, 5)	18 (6, 9, 3)
4	1YQV (1.7, 0.195)	17 (15, 1, 1)	11 (6, 2, 3)	15 (12, 1, 2)
	2HFL (NA, NA)	18 (5, 2, 11)	10 (3, 4, 3)	11 (4, 7, 0)

5	2F19 (2.8, 0.182)	19 (12, 4, 3)	13 (8, 2, 3)	22 (8, 3, 11)
	1F19 (NA, NA)	15 (1, 3, 11)	14 (1, 1, 12)	10 (3, 2, 5)
6	5FD1 (1.9, 0.215)	13 (12, 0, 1)	7 (5, 1, 1)	8 (6, 2, 0)
	2FD1 (NA, NA)	-	5 (1, 0, 4)	2 (0, 1, 1)
7	3BJL_A (2.3, 0.165)	2 (1, 1, 0)	2 (1, 0, 1)	2 (1, 1, 0)
	1BJL (NA, NA)	-	3 (0, 0, 3)	2 (1, 0, 1)
8	2DTR_A (Co^{+2}) (1.9, 0.171)	4 (4, 0, 0)	8 (6, 2, 0)	6 (4, 2, 0)
	1DTR (2.8, 0.210)	4 (2, 1, 1)	5 (3, 2, 0)	4 (0, 3, 1)
9	7FAB_L (2.0, 0.169)	3 (3, 0, 0)	1 (1, 0, 0)	-
	1FAB_L (NA, NA)	2 (1, 0, 1)	2 (1, 1, 0)	-
10	1A45_A (2.3, 0.186)	7 (5, 2, 0)	1 (0, 1, 0)	7 (5, 1, 1)
	2GCR (NA, NA)	6 (2, 0, 4)	3 (1, 0, 2)	5 (3, 0, 2)
11	3EFZ_A (2.08, 0.227)	1 (1, 0, 0)	-	2 (1, 0, 1)
	2IJP (2.14, 0.237)	1 (0, 0, 1)	-	2 (1, 0, 1)
12	5PTP_A (Ca^{+2}) (1.34, 0.152)	8 (7, 1, 0)	3 (3, 0, 0)	11 (9, 2, 0)
	1PTP (Mg^{+2}) (NA, NA)	8 (5, 1, 2)	1 (1, 0, 0)	10 (4, 2, 4)
13	4AAH_A (Ca^{+2}) (2.4, 0.152)	13 (12, 0, 1)	3 (3, 0, 0)	2 (1, 1, 0)
	3AAH_A (Ca^{+2}) (2.4, 0.201)	11 (3, 2, 6)	5 (5, 0, 0)	1 (0, 1, 0)
14	6LDH_A (2.0, 0.202)	2 (1, 1, 0)	2 (1, 0, 1)	4 (2, 0, 2)

	2LDH (NA, NA)	1 (1, 0, 0)	-	2 (1, 0, 1)
15	3DFR_A (1.7, 0.152)	7 (7, 0, 0)	10 (8, 1, 1)	13 (11, 2, 0)
	1DFR (NA, NA)	4 (0, 1, 3)	15 (5, 3, 7)	9 (1, 4, 4)
16	5CPV_A (Ca^{+2}) (1.6, 0.187)	3 (2, 1, 0)	1 (1, 0, 0)	5 (5, 0, 0)
	1CPV (Ca^{+2}) (1.85, NA)	2 (2, 0, 0)	2 (2, 0, 0)	1 (0, 1, 0)
17	3F70_A (2.1, 0.203)	4 (4, 0, 0)	2 (0, 1, 1)	11 (5, 1, 5)
	3DBB_A (2.15, 0.212)	5 (3, 1, 1)	3 (1, 1, 1)	6 (3, 3, 0)
18	4CAA (2.9, 0.199)	9 (5, 1, 3)	10 (7, 3, 0)	10 (6, 4, 0)
	1CT3_A (3.1, 0.189)	12 (6, 3, 3)	7 (5, 0, 2)	6 (4, 2, 0)
19	3CPA_A (Zn^{+2}) (2.0, NA)	8 (7, 1, 0)	3 (2, 0, 1)	7 (5, 0, 2)
	1CPA (Zn^{+2}) (NA, NA)	7 (3, 0, 4)	5 (2, 1, 2)	5 (2, 1, 2)
20	1P2Z_A (2.2, 0.178)	58 (45, 8, 5)	20 (13, 2, 5)	49 (26, 16, 7)
	1DHX_A (2.9, 0.225)	35 (8, 9, 18)	17 (8, 0, 9)	29 (7, 8, 14)

Table S12. Markedly different distributions of points from superseded and upgraded structures in the Complementarity Plots. Frequencies of residues with positional errors from superseded pdb files along with frequencies of the same residues from upgraded structures, in the different regions of the complementarity plots. The same residue from the two sets could lie in different burial bins. χ^2 (df = 3-1: probable, less probable, improbable; $\chi^2_{0.05}$ = 5.991) between expected (from **DB2**) and observed (superseded and upgraded) distributions have been given. **(A)** tabulates the results for CP1 ($0.0 \leq Bur \leq 0.05$) while (B) and (C) tabulate the same for CP2 ($0.05 < Bur \leq 0.15$) and CP3 ($0.15 < Bur \leq 0.3$). Errors were detected by side chain rmsd's (method 1, see **Main Text).**

(A)

$0.0 \leq Bur \leq 0.05$					
	Total Count of points	Occupancy (%)			χ^2
		Probable	Less Probable	Improbable	
Expected	-	82.1	9.2	8.7	-
Superseded	302	41.4	14.6	44.0	503.950
Upgraded	354	80.2	9.6	10.2	1.093

(B)

$0.05 < Bur \leq 0.15$					
	Total Count of points	Occupancy (%)			
		Probable	Less Probable	Improbable	χ^2
Expected	-	76.1	13.9	10.0	-
Superseded	107	30.8	11.2	58.0	275.308
Upgraded	97	66.0	14.4	19.6	10.242

(C)

0.15 < Bur ≤ 0.30					
	Total Count of points	Occupancy (%)			
		Probable	Less Probable	Improbable	χ^2
Expected	-	70.7	16.8	12.5	-
Superseded	158	22.2	38.6	39.2	187.782
Upgraded	161	56.5	24.2	19.3	15.736

Table S13. Markedly different distributions of points from superseded and upgraded structures in the Complementarity Plots. Frequencies of residues with positional errors from superseded pdb files along with frequencies of the same residues from upgraded structures, in the different regions of the complementarity plots. The same residue from the two sets could lie in different burial bins. χ^2 (df = 3-1: probable, less probable, improbable; $\chi^2_{0.05}$ = 5.991) between expected (from **DB2**) and observed (superseded and upgraded) distributions have been given. **(A)** tabulates the results for CP1 (0.0 ≤ Bur ≤ 0.05) while (B) and (C) tabulate the same for CP2 (0.05 < Bur ≤ 0.15) and CP3 (0.15 < Bur ≤ 0.3). Errors were detected by deviation in χ_1 torsions (method 2, see **Main Text**).

(A)

0.0 ≤ Bur ≤ 0.05					
	Total Count of points	Occupancy (%)			χ^2
		Probable	Less	Improbable	
Expected	-	82.1	9.2	8.7	-
Superseded	189	33.3	14.8	51.9	465.746
Upgraded	228	77.6	12.7	9.7	3.860

(B)

0.05 < Bur ≤ 0.15					
	Total Count of points	Occupancy (%)			
		Probable	Less Probable	Improbable	χ^2
Expected	-	76.1	13.9	10.0	-
Superseded	120	40.8	12.5	46.7	181.115

Upgraded	121	67.8	14.0	18.2	9.205

(C)

0.15 < Bur ≤ 0.30					
	Total Count of points	Occupancy (%)			
		Probable	Less Probable	Improbable	χ^2
Expected	-	70.7	16.8	12.5	-
Superseded	136	34.6	33.8	31.6	88.351
Upgraded	216	58.8	21.3	19.9	16.410

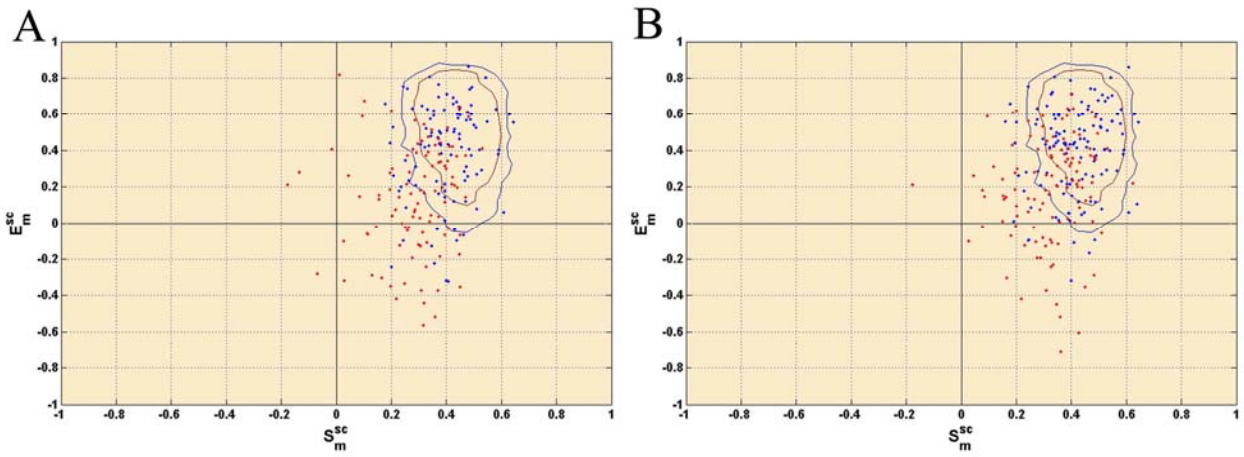

Figure S10. Distribution of points in CP2. Distribution of points corresponding to residues with coordinate errors in the superseded (plotted in blue) and the same residues in the upgraded structures (red) for CP2 (0.05 < Bur ≤ 0.15). In **(A)** the erroneous residues have been detected by method 1 (side chain rmsd) and in **(B)** by method 2 (χ_1 torsion) (see **Main Text**).

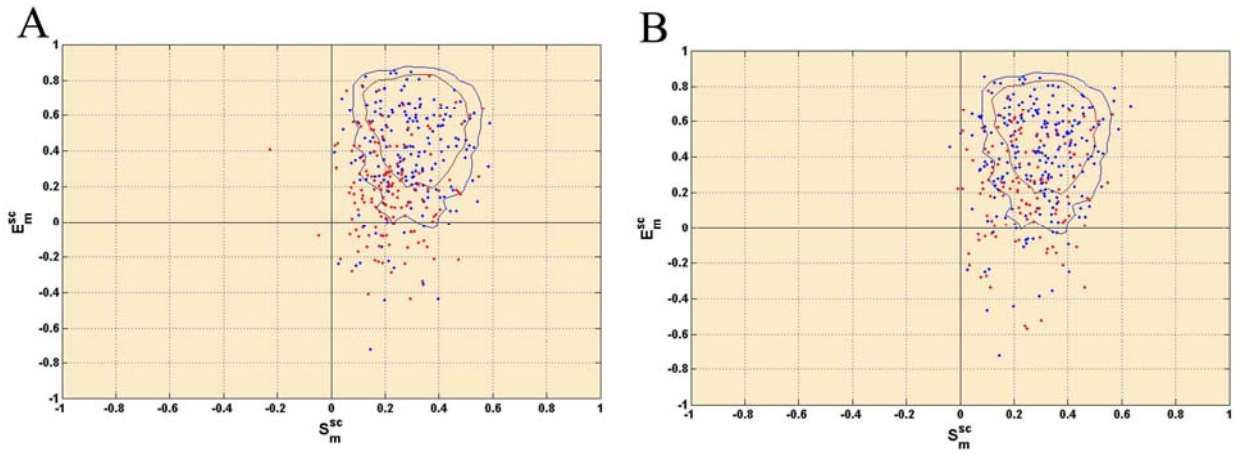

Figure S11. Distribution of points in CP3. Distribution of points corresponding to residues with coordinate errors in the superseded (plotted in blue) and the same residues in the upgraded structures (red) for CP2 ($0.15 < \text{Bur} \leq 0.3$). In **(A)** the erroneous residues have been detected by method 1 (side chain rmsd) and in **(B)** by method 2 (χ_1 torsion) (see **Main Text**).

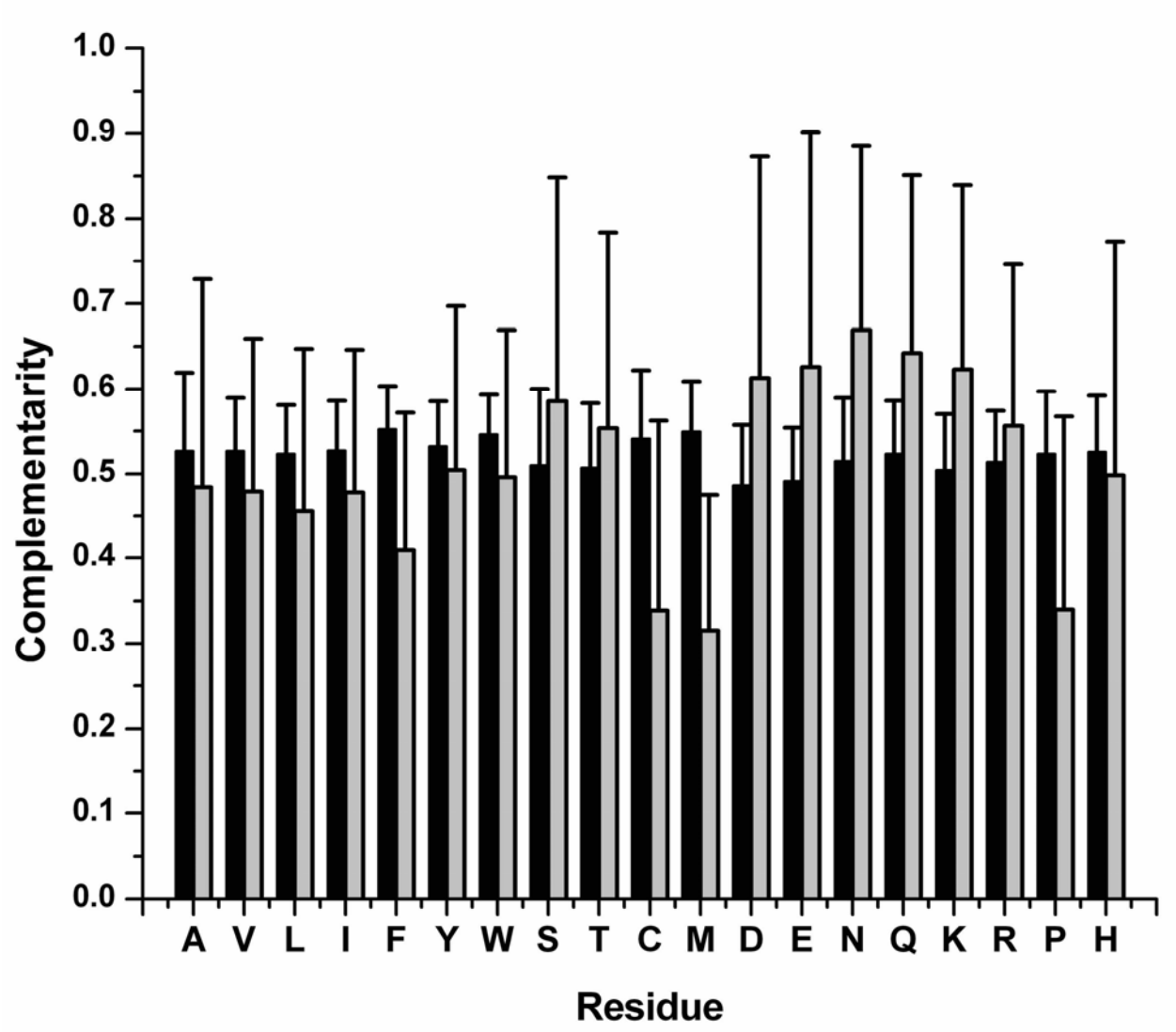

Figure S12. Trends in surface and electrostatic complementarities for completely buried amino acid side chains. Mean $\overline{S_m^{sc}}$ (black), $\overline{E_m^{sc}}$ (gray), plotted as filled thick bars along with their standard deviations (represented by error bars) for different residues in the 1st burial bin ($0.0 \leq \text{Bur} \leq 0.05$).

Table S14. Comparison of the values between electrostatic and shape complementarities. Mean $\overline{E_m^{sc}}$, $\overline{S_m^{sc}}$ and their standard deviations (in parentheses) tabulated for different amino acid residues distributed in three burial bins (bin1: $0.00 \leq \text{Bur} \leq 0.05$; bin2: $0.05 < \text{Bur} \leq 0.15$; bin3: $0.15 < \text{Bur} \leq 0.30$) where 'Bur' stands for the burial ratio (see **Main Text**).

Residue	$\overline{E_m^{sc}}$			$\overline{S_m^{sc}}$		
	bin1	**bin2**	**bin3**	**bin1**	**bin2**	**bin3**
ALA	0.48 (0.25)	0.46 (0.25)	0.50 (0.26)	0.53 (0.09)	0.43 (0.12)	0.30 (0.15)
VAL	0.48 (0.18)	0.45 (0.17)	0.46 (0.18)	0.53 (0.06)	0.45 (0.08)	0.35 (0.11)
LEU	0.46 (0.19)	0.41 (0.20)	0.43 (0.21)	0.52 (0.06)	0.45 (0.08)	0.34 (0.10)
ILE	0.48 (0.17)	0.43 (0.18)	0.46 (0.18)	0.53 (0.06)	0.46 (0.08)	0.35 (0.10)
PHE	0.41 (0.16)	0.40 (0.18)	0.42 (0.18)	0.55 (0.05)	0.49 (0.07)	0.40 (0.10)
TYR	0.50 (0.19)	0.48 (0.20)	0.45 (0.20)	0.53 (0.05)	0.46 (0.07)	0.36 (0.09)
TRP	0.50 (0.17)	0.46 (0.20)	0.47 (0.18)	0.55 (0.05)	0.48 (0.06)	0.38 (0.09)
SER	0.59 (0.26)	0.54 (0.28)	0.54 (0.29)	0.51 (0.09)	0.41 (0.12)	0.28 (0.14)
THR	0.55 (0.23)	0.54 (0.24)	0.52 (0.26)	0.51 (0.08)	0.42 (0.10)	0.30 (0.12)
CYS	0.35 (0.22)	0.32 (0.20)	0.29 (0.23)	0.54 (0.08)	0.45 (0.12)	0.34 (0.14)
MET	0.32 (0.16)	0.28 (0.15)	0.27 (0.15)	0.55 (0.06)	0.46 (0.08)	0.33 (0.11)
ASP	0.61 (0.26)	0.63 (0.22)	0.63 (0.23)	0.49 (0.07)	0.39 (0.09)	0.26 (0.10)
GLU	0.63 (0.28)	0.61 (0.22)	0.57 (0.22)	0.49 (0.06)	0.39 (0.08)	0.27 (0.09)
ASN	0.67 (0.22)	0.63 (0.27)	0.59 (0.25)	0.52 (0.07)	0.43 (0.09)	0.31 (0.11)
GLN	0.64 (0.21)	0.63 (0.19)	0.53 (0.23)	0.52 (0.06)	0.42 (0.08)	0.30 (0.10)
LYS	0.62 (0.22)	0.56 (0.25)	0.51 (0.23)	0.50 (0.07)	0.42 (0.08)	0.30 (0.10)
ARG	0.56 (0.19)	0.60 (0.19)	0.58 (0.19)	0.51 (0.06)	0.43 (0.08)	0.30 (0.10)
PRO	0.34 (0.23)	0.38 (0.21)	0.40 (0.22)	0.52 (0.07)	0.43 (0.10)	0.31 (0.12)
HIS	0.50 (0.28)	0.50 (0.25)	0.46 (0.29)	0.53 (0.07)	0.45 (0.08)	0.34 (0.10)

References

1. Basu, S., D. Bhattacharya, and R. Banerjee. 2011. Mapping the distribution of packing topologies within protein interiors shows predominant preference for specific packing motifs. BMC Bioinformatics. 12:195.
2. Holm, L., and C. J. Sander. 1992. Evaluation of protein models by atomic solvation preference. J Mol Biol. 225:93-105.
3. Branden, C. I., and T. A. Jones. 1990. Between objectivity and subjectivity. Nature. 343:687-689.
4. Avbelj, F., J. Moult, D. H. Kitson, M. N. James, and A. T. Hagler. 1990. Molecular dynamics study of the structure and dynamics of a protein molecule in a crystalline ionic environment, Streptomyces griseus protease A. Biochemistry. 29:8658-8676.
5. Bahadur, R. P., and P. Chakrabarti. 2009. Discriminating the native structure from decoys using scoring functions based on the residue packing in globular proteins. BMC Structural Biology. 9:76.
6. Samudrala, R., and J. Moult. 1998. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J. Mol. Biol. 275:895-916.
7. Arab, S., M. Sadeghi, C. Eslahchi, H. Pezeshk, and A. Sheari. 2010. A pairwise residue contact area-based mean force potential for discrimination of native protein structure. BMC Bioinformatics. 11:16.
8. Lu, H., and J. Skolnick. 2001. A distance-dependent atomic knowledge based potential for improved protein structure selection. Proteins. 44:223-232.
9. Skolnick, J., A. Kolinski, A. Ortiz. 2000. Derivation of protein-specific pair potentials based on weak sequence fragment similarity. Proteins. 38:3-16.
10. Park, B., and M. Levitt. 1996. Energy functions that discriminate X-ray and near-native folds from well-constructed decoys. J Mol Biol, 258:367-392.
11. Simons, K. T., C. Kooperberg, E. Huang, and D. Baker. 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J. Mol. Biol. 268:209-225.
12. Zhang, C., S. Liu, H. Zhou, and Y. Zhou. 2004. An accurate, residue-level, pair potential of mean force for folding and binding based on the distance-scaled, ideal-gas reference state. Protein Sci. 13:400-411.
13. Misura, K. M., D. Chivian, C. A. Rohl, D. E. Kim, and D. Baker. 2006. Physically realistic homology models built with ROSETTA can be more accurate than their templates. Proc. Natl. Acad. Sci. USA. 103:5361-5366.
14. Melo, F., R. Sanchez, and A. Sali. 2002. Statistical potentials for fold assessment. Protein Sci. 11:430-448.
15. Li, X., C. Hu, and J. Liang. 2003. Simplicial edge representation of protein structures and alpha contact potential with confidence measure. Proteins. 53:792-805.
16. Mirzaie, M., C. Eslahchi, H. Pezeshk, and M. Sadeghi. 2009. A distance-dependent atomic knowledge-based potential and force for discrimination of native structures from decoys. Proteins. 77:454-463.
17. Shen, M. Y., and A. Sali. 2006. Statistical potential for assessment and prediction of protein structures. Protein Sci. 15:2507-2524.

18. Miyazawa, S., and R. L. Jernigan. 1996. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J. Mol. Biol. 256:623-644.
19. Samudrala, R., and M. Levitt. 2000. Decoys 'R' Us: a database of incorrect conformation to improve protein structure prediction. Protein Sci. 9:1399-1401.
20. Tsai, J., R. Bonneau, A. V. Morozov, B. Kuhlman, C. A. Rohl, and D. Baker. 2003. An improved protein decoy set for testing energy functions for protein structure prediction. Proteins, 53:76-87.
21. Moult, J., K. Fidelis, A. Kryshtafovych, and A. Tramontano. 2011. Critical assessment of methods of protein structure prediction (CASP)—Round IX Proteins 79 (Suppl 10):1–5.
22. Holm, L., and P. Rosenstrom. 2010. Dali server: conservation mapping in 3D. Nuc. Acids. Res. 38:W545-549.
23. Krivov, G. G., M. V. Shapovalov, and R. L. Dunbrack. 2009. Improved prediction of protein side-chain conformations with SCWRL4. Proteins. 77:778-795.
24. Edgar, R. C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nuc. Acids. Res. 32:1792-1797.