Supplementary Information (SI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2025

Supplementary Information: Preserving structural integrity: Fold reproducibility in computational design of proteins non-homologous to wild-type sequence

Bondeepa Saikia and Anupaul Baruah
Department of Chemistry, Dibrugarh University, Dibrugarh-786004, India

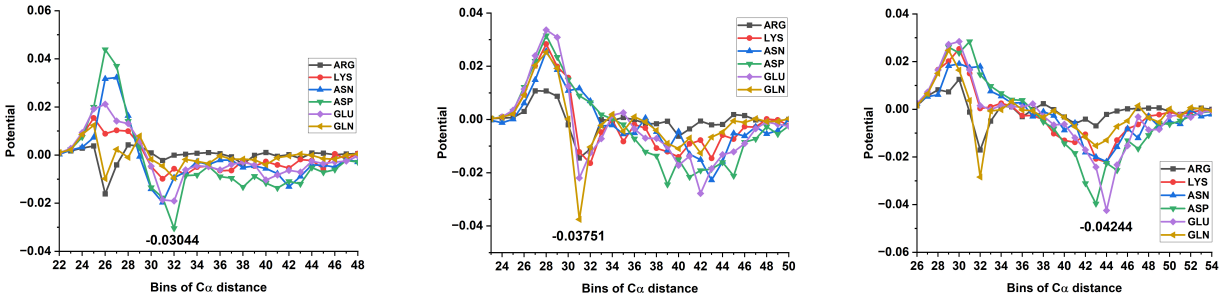
Comparison with RosettaDesign

A comparison of our method with RosettaDesign is carried out and is tabulated in Table S1. From the table, it is clear that our model and developed statistical potential are simple and coarse-grained which significantly reduces computational cost. While RosettaDesign uses Rosetta energy function, a composite energy function that merges physical modeling with statistical preferences. Again Rosetta uses all atom model while our model is coarse-grained. Given the fundamental differences between the two methods, a direct comparison between the sequences generated by our approach and those from RosettaDesign would be inappropriate, as the two methods operate at different levels of design complexity and goal. Nevertheless, we have designed three sequences using RosettaDesign web server (http://rosettadesign.med.unc.edu/) for each three target structures and are represented by RDesign_1, RDesign_2 and RDesign_3. The RMSDs, sequence similarity with the wild-type sequences are also calculated and are provided in Fig. S2, Fig. S3 and Fig. S4. Interestingly, while these sequences achieve 100% recognition of the target secondary structure content, their sequence similarity to the wild-type sequences is significantly higher than that of the sequences designed using our model. The 100% recognition of secondary structure is attributed to the use of the native backbone during RosettaDesign. In contrast, our Cα-based coarse-grained model allows greater sequence diversity, making it better suited for exploring non-homologous yet fold-compatible sequences.

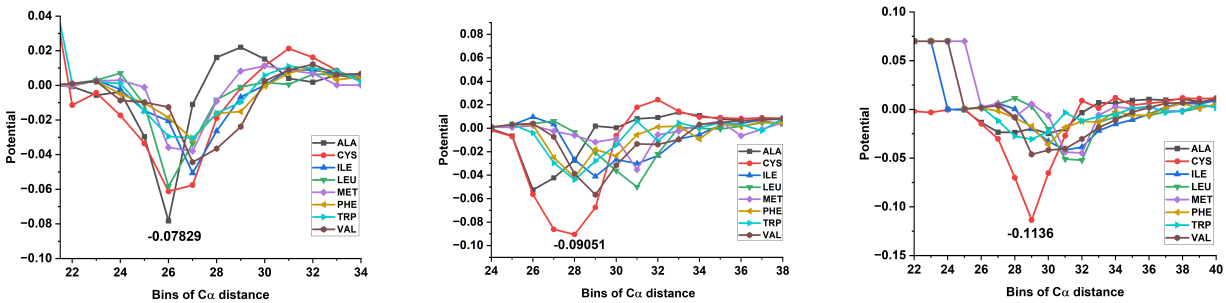
Table S1. Comparison of our design method with RosettaDesign

Feature	**RosettaDesign (web server)**	**Our Model**
Objective	Focuses on stabilizing a given protein fold generating sequences compatible with a fixed backbone.	Design non-homologous sequences with low sequence similarity to wild-type, while preserving the native fold.
Input	All-atom or backbone PDB structure.	Cα backbone coordinates of the target protein.
Potential	The energy function consists of (i) a Lennard–Jones potential, (ii) the Lazaridis-Karplus implicit solvation model, (iii) an explicit orientation dependent hydrogen bonding term, (iv) torsion potential, (v) a unique reference value for each amino acid type, (vi) electrostatic interactions.	Cα distance based one body and two body statistical potentials.
Design method	Side-chain rotamer optimization by Dunbrack's backbone dependent rotamer library, energy minimization using Monte Carlo simulated annealing starting from very high temperature to 0°. Starting from a random sequence, single amino acid substitutions are accepted via Metropolis criterion.	Monte Carlo simulation incorporating random point mutations, single amino acid substitutions are accepted via Metropolis criterion.
Sequence variability	Tends to produce sequences with amino acid frequencies comparable to those found in naturally occurring proteins. Yield sequences with moderate similarity to wild type	Specifically aims for non-homologous sequences that fold into native-like structure.

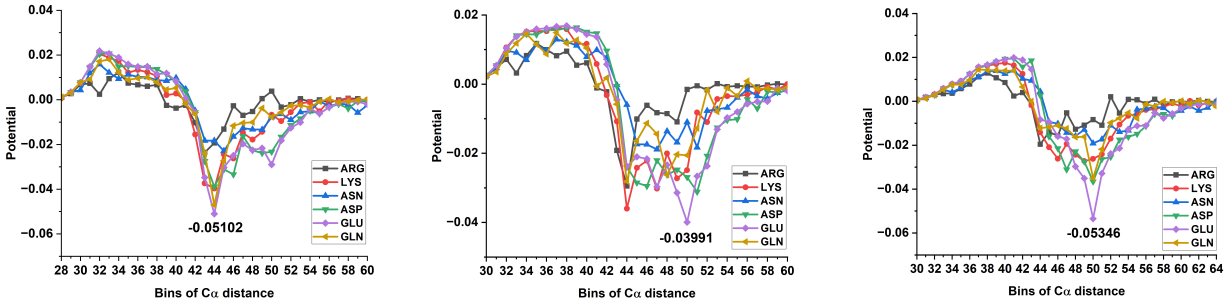
Following are the plots of potential versus bins of Cα distances for different nearest neighbor distances, NN_i, i=2, 3, 4, 6, 7, 8 and 9.



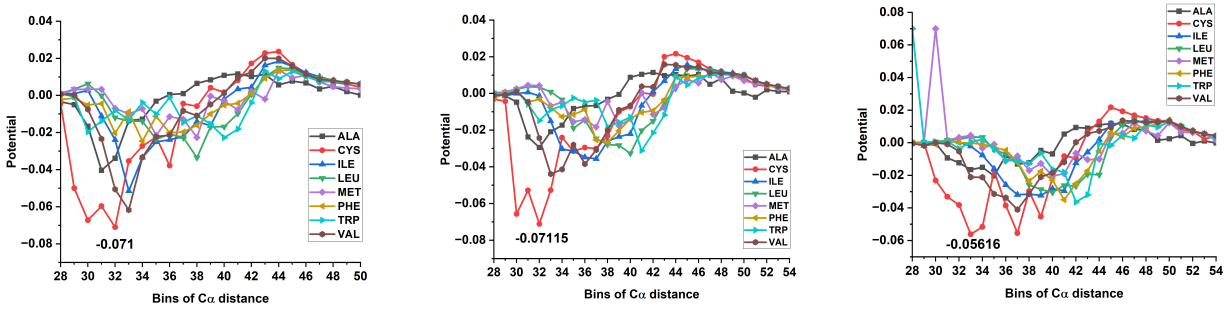
(a) NN_2-hydrophilic (b) NN_3-hydrophilic (c) NN_4-hydrophilic



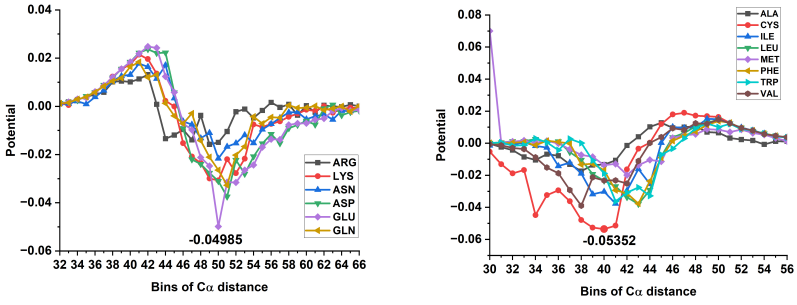
(d) NN_2-hydrophobic (e) NN_3-hydrophobic (f) NN_4-hydrophobic



(g) NN_6-hydrophilic (h) NN_7-hydrophilic (i) NN_8-hydrophilic

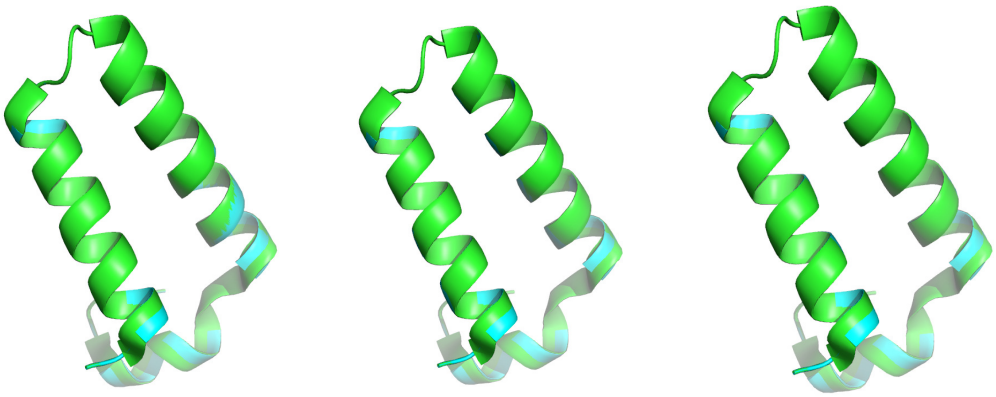


(j) NN_6-hydrophobic (k) NN_7-hydrophobic (l) NN_8-hydrophobic



(m) NN_9-hydrophilic (n) NN_9-hydrophobic

Fig. S1. Plot of potential versus bins of Cα distances (each bin is of 0.2 Å size) for hydrophilic and hydrophobic amino acid residues.

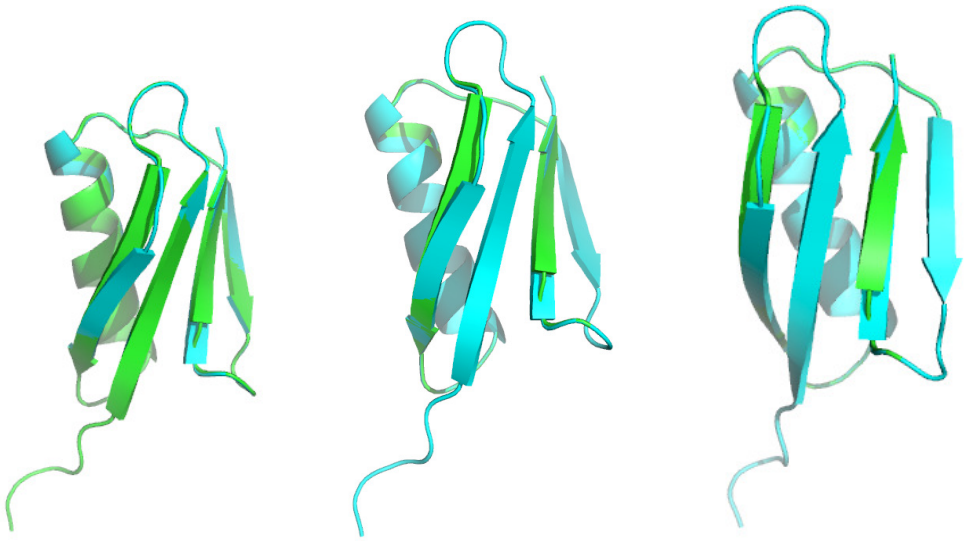


(a) 6FM8, RDesign_1, RMSD=0.007 Å, **Sim=20%**

(b) 6FM8, RDesign_2, RMSD=0.007 Å, **Sim=20%**

(c) 6FM8, RDesign_3, RMSD=0.003 Å, **Sim=16%**

Fig. S2. The three-dimensional structures of the target protein, 6FM8 (colored green) and the structure of the protein designed by RosettaDesign (colored cyan) and their corresponding RMSDs and sequence similarity to the wild-type sequences.



(a) 2IGD, RDesign_1, RMSD=0.008 Å, **Sim=36.06%**

(b) 2IGD, RDesign_2, RMSD=0.008 Å, **Sim=36.06%**

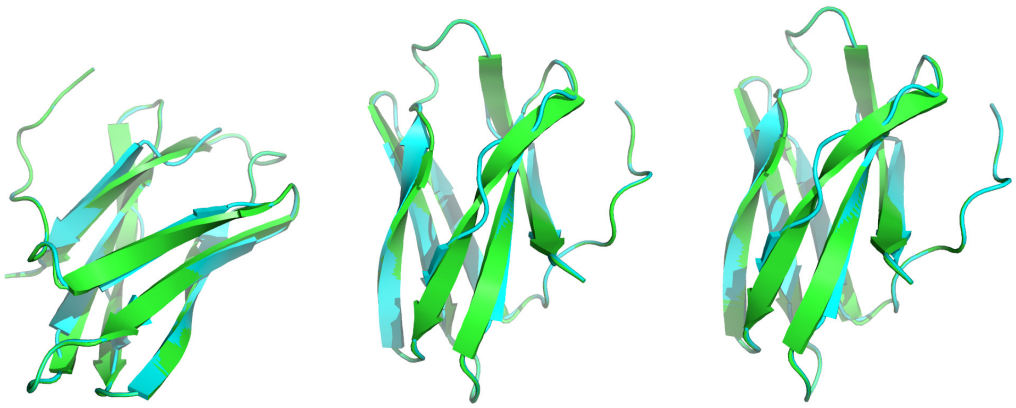
(c) 2IGD, RDesign_3, RMSD=0.008 Å, **Sim=36.06%**

Fig. S3. The three-dimensional structures of the target protein, 2IGD (colored green) and the structure of the protein designed by RosettaDesign (colored cyan) and their corresponding RMSDs and sequence similarity to the wild-type sequences.

The PDB IDs of the 556 proteins which are used to generate non-native conformational ensemble are provided in Table S2 to enhance transparency and reproducibility.

Table S2. PDB IDs of 556 proteins considered for non-native conformation generation

PDB IDs										
1A3H.pdb	1A8Q.pdb	1AAJ.pdb	1ACX.pdb	1AD6.pdb	1AJM.pdb	1AKO.pdb	1ALY.pdb	1ANU.pdb	1ARL.pdb	1BEC.pdb
1BEO.pdb	1BG5.pdb	1BGF.pdb	1BHE.pdb	1BI5.pdb	1BKB.pdb	1BM8.pdb	1BOL.pdb	1BQC.pdb	1BUO.pdb	1BYI.pdb
1BYW.pdb	1C25.pdb	1C3G.pdb	1C44.pdb	1CDY.pdb	1CEM.pdb	1CQY.pdb	1CRZ.pdb	1CWY.pdb	1CZ1.pdb	1D2P.pdb
1DAB.pdb	1DHN.pdb	1DIV.pdb	1DIX.pdb	1DJA.pdb	1DQ0.pdb	1DSL.pdb	1DTZ.pdb	1DUA.pdb	1DUS.pdb	1DVO.pdb
1E5W.pdb	1EDG.pdb	1EDQ.pdb	1EHD.pdb	1EIA.pdb	1EOV.pdb	1EQ6.pdb	1ESR.pdb	1EW4.pdb	1EYH.pdb	1F00.pdb
1F1S.pdb	1FBN.pdb	1FHL.pdb	1FNF.pdb	1FSF.pdb	1FUS.pdb	1FYX.pdb	1G2U.pdb	1G62.pdb	1G8A.pdb	1GBS.pdb
1GH2.pdb	1GLN.pdb	1GQN.pdb	1GQZ.pdb	1GVP.pdb	1GYD.pdb	1GYV.pdb	1H09.pdb	1H2P.pdb	1H3Q.pdb	1H4A.pdb
1HE9.pdb	1HH2.pdb	1HKA.pdb	1HOE.pdb	1I5P.pdb	1I60.pdb	1IAD.pdb	1IFG.pdb	1IJB.pdb	1ILK.pdb	1IO2.pdb
1IS1.pdb	1IV8.pdb	1J2V.pdb	1J7G.pdb	1J7X.pdb	1K40.pdb	1K89.pdb	1KIV.pdb	1KLO.pdb	1KOE.pdb	1KS9.pdb
1KUU.pdb	1LLN.pdb	1LMI.pdb	1LNS.pdb	1LPL.pdb	1LWB.pdb	1LXA.pdb	1MF7.pdb	1MIL.pdb	1MIX.pdb	1MSC.pdb
1MUF.pdb	1MZL.pdb	1N7O.pdb	1N81.pdb	1ND7.pdb	1NG6.pdb	1NI5.pdb	1NOA.pdb	1O5T.pdb	1OJQ.pdb	1OLR.pdb
1OOI.pdb	1OTP.pdb	1P3C.pdb	1P4X.pdb	1P7N.pdb	1P7S.pdb	1PCL.pdb	1PDB.pdb	1PJB.pdb	1PRZ.pdb	1Q2Y.pdb
1Q50.pdb	1QAU.pdb	1QCX.pdb	1QJ9.pdb	1QMT.pdb	1QQF.pdb	1QQH.pdb	1QTF.pdb	1QTO.pdb	1QTS.pdb	1QW2.pdb
1QXX.pdb	1QZM.pdb	1R8N.pdb	1RC9.pdb	1RCB.pdb	1RL0.pdb	1S29.pdb	1S7I.pdb	1SBX.pdb	1SCZ.pdb	1SRV.pdb
1SUR.pdb	1T2I.pdb	1TJE.pdb	1TQG.pdb	1UB9.pdb	1UKF.pdb	1ULZ.pdb	1UOK.pdb	1UX5.pdb	1V05.pdb	1VCC.pdb
1VGJ.pdb	1VGP.pdb	1VLS.pdb	1VSM.pdb	1WGB.pdb	1WHI.pdb	1WS6.pdb	1WU3.pdb	1WWB.pdb	1WWI.pdb	1X0M.pdb
1X1H.pdb	1X3O.pdb	1XEU.pdb	1XIX.pdb	1XT0.pdb	1Y2Q.pdb	1YQ8.pdb	1YUW.pdb	1YW5.pdb	1YZF.pdb	1ZEQ.pdb
1ZHV.pdb	1ZLB.pdb	2A6Z.pdb	2AH5.pdb	2ARA.pdb	2B0A.pdb	2B8I.pdb	2BFH.pdb	2BK8.pdb	2C08.pdb	2C6U.pdb
2C83.pdb	2CG7.pdb	2CKX.pdb	2CW4.pdb	2CWY.pdb	2CYG.pdb	2DP9.pdb	2DYI.pdb	2E8F.pdb	2EDM.pdb	2EHG.pdb
2ERF.pdb	2F82.pdb	2FB9.pdb	2FBO.pdb	2FC3.pdb	2FCB.pdb	2FD5.pdb	2FPH.pdb	2FRG.pdb	2G5X.pdb	2G69.pdb
2GGO.pdb	2H2C.pdb	2H85.pdb	2H8E.pdb	2HC8.pdb	2HRZ.pdb	2I6V.pdb	2ICT.pdb	2IE8.pdb	2IN0.pdb	2J6B.pdb
2J9V.pdb	2JA4.pdb	2JA9.pdb	2LAO.pdb	2NSC.pdb	2NSN.pdb	2NWD.pdb	2O6Q.pdb	2OF3.pdb	2OHE.pdb	2OKT.pdb
2ORX.pdb	2OV1.pdb	2OZF.pdb	2P25.pdb	2P52.pdb	2PCY.pdb	2PET.pdb	2PKO.pdb	2PLC.pdb	2PLF.pdb	2PND.pdb
2PNE.pdb	2PPN.pdb	2Q34.pdb	2Q98.pdb	2Q9V.pdb	2QEV.pdb	2QHT.pdb	2QT4.pdb	2QZQ.pdb	2R0S.pdb	2R6Q.pdb
2RA1.pdb	2RH3.pdb	2RKQ.pdb	2SGA.pdb	2SIL.pdb	2TGI.pdb	2V14.pdb	2V6C.pdb	2VAJ.pdb	2VQ4.pdb	2W0G.pdb
2W0I.pdb	2X8X.pdb	2X9Z.pdb	2Y0R.pdb	2YGS.pdb	2YVN.pdb	2YWX.pdb	3A4C.pdb	3A7L.pdb	3AAP.pdb	3ACP.pdb
3B02.pdb	3BJA.pdb	3BWT.pdb	3C12.pdb	3CAF.pdb	3CH7.pdb	3CO1.pdb	3CSR.pdb	3DJ9.pdb	3DJN.pdb	3DPA.pdb
3DQP.pdb	3E9U.pdb	3ENU.pdb	3EZM.pdb	3FH2.pdb	3FHF.pdb	3FK5.pdb	3FS5.pdb	3H6J.pdb	3HAK.pdb	3HVM.pdb
3HVV.pdb	3ILS.pdb	3IPZ.pdb	3IV4.pdb	3JV1.pdb	3JZ9.pdb	3JZZ.pdb	3K1D.pdb	3KB5.pdb	3KJE.pdb	3KT9.pdb
3KVD.pdb	3LD1.pdb	3LIG.pdb	3LJ8.pdb	3M66.pdb	3M70.pdb	3MTV.pdb	3MX7.pdb	3MZZ.pdb	3N0K.pdb	3ND2.pdb
3NSM.pdb	3O5P.pdb	3ONJ.pdb	3OZH.pdb	3PBH.pdb	3PFM.pdb	3PM2.pdb	3PTJ.pdb	3QPA.pdb	3R6D.pdb	3RA2.pdb
3RD2.pdb	3RDJ.pdb	3RFF.pdb	3RFW.pdb	3RZY.pdb	3S0A.pdb	3SEB.pdb	3SH4.pdb	3SO8.pdb	3T5B.pdb	3T7F.pdb
3TOW.pdb	3TUA.pdb	3UAF.pdb	3UFC.pdb	3VFI.pdb	3VOR.pdb	3VPY.pdb	3WA1.pdb	3WFV.pdb	3WP5.pdb	3WY8.pdb
3WZF.pdb	4A02.pdb	4ACJ.pdb	4BGC.pdb	4BKD.pdb	4BPF.pdb	4CFI.pdb	4CU2.pdb	4CYA.pdb	4DNI.pdb	4DNU.pdb
4DOM.pdb	4E2U.pdb	4ETX.pdb	4EVM.pdb	4FCU.pdb	4FEI.pdb	4GN2.pdb	4IPC.pdb	4IZE.pdb	4J4R.pdb	4J5Q.pdb
4JJO.pdb	4JP2.pdb	4JP6.pdb	4K3C.pdb	4KK4.pdb	4L5Q.pdb	4L9R.pdb	4LON.pdb	4MHP.pdb	4MJA.pdb	4MOA.pdb
4N6T.pdb	4N74.pdb	4NFB.pdb	4O6G.pdb	4O7Q.pdb	4O8S.pdb	4OZS.pdb	4OZX.pdb	4OZY.pdb	4P5U.pdb	4PA1.pdb
4PBO.pdb	4PEP.pdb	4PLQ.pdb	4PMH.pdb	4PQD.pdb	4PS6.pdb	4RSX.pdb	4RXV.pdb	4RZ9.pdb	4TX7.pdb	4U3H.pdb
4U64.pdb	4W98.pdb	4WJS.pdb	4WZ0.pdb	4XIO.pdb	4XP8.pdb	4YPC.pdb	4YTE.pdb	4Z3I.pdb	4ZGI.pdb	4ZM7.pdb
4ZOU.pdb	4ZPC.pdb	4ZQA.pdb	4ZZF.pdb	5AVG.pdb	5AVH.pdb	5AY9.pdb	5BTH.pdb	5BWY.pdb	5C3T.pdb	5D6B.pdb
5DZ9.pdb	5E0Z.pdb	5E10.pdb	5E1Y.pdb	5EFX.pdb	5EHA.pdb	5F6A.pdb	5F70.pdb	5G51.pdb	5GLX.pdb	5GMB.pdb
5GUJ.pdb	5GY3.pdb	5H0Q.pdb	5H28.pdb	5H3X.pdb	5H4E.pdb	5HD3.pdb	5HD9.pdb	5HFD.pdb	5HU4.pdb	5IHW.pdb
5ILU.pdb	5IPW.pdb	5K6F.pdb	5KVR.pdb	5M4S.pdb	5MGV.pdb	5NJO.pdb	5ORI.pdb	5ORM.pdb	5OUW.pdb	5P70.pdb
5RPE.pdb	5TUN.pdb	5UVR.pdb	5VTL.pdb	5WFT.pdb	5XCY.pdb	5XEF.pdb	5XI8.pdb	5XPW.pdb	5XUO.pdb	5Y33.pdb
5Y8E.pdb	5YDN.pdb	5YNS.pdb	5YO6.pdb	5YSX.pdb	5YWZ.pdb	5Z2D.pdb	5Z67.pdb	5Z7Q.pdb	5ZA4.pdb	5ZHG.pdb
6A41.pdb	6A9Y.pdb	6AMN.pdb	6AZ5.pdb	6BIO.pdb	6CVA.pdb	6D0A.pdb	6DGA.pdb	6DR3.pdb	6E7E.pdb	6ELM.pdb
6EXX.pdb	6F2Q.pdb	6F4C.pdb	6F4M.pdb	6FHZ.pdb	6GGP.pdb	6GPM.pdb	6GV5.pdb	6IGV.pdb	6IOJ.pdb	6IY4.pdb
6JCC.pdb	6JL7.pdb	6JOY.pdb	6JV8.pdb	6JZA.pdb	6KD0.pdb	6KNE.pdb	6L27.pdb	6L7Q.pdb	6LG3.pdb	6M4C.pdb
6NZS.pdb	6Q3V.pdb	6SWI.pdb	6SYG.pdb	6TF5.pdb	6TG6.pdb	6TYY.pdb	6UFW.pdb	6WES.pdb	6WIN.pdb	6XP8.pdb
6ZM8.pdb	7BFY.pdb	7BXY.pdb	7CHR.pdb	7CNB.pdb	7CSN.pdb	7DKR.pdb	7DMF.pdb	7DMS.pdb	7DU0.pdb	7EP8.pdb
7ERU.pdb	7EYK.pdb	7KPH.pdb	7RFD.pdb	7W3W.pdb	8OHM.pdb					



(a) 1HOE, RDesign_1, RMSD=0.004 Å, **Sim=36.48%**

(b) 1HOE, RDesign_2, RMSD=0.006 Å, **Sim=41.89%**

(c) 1HOE, RDesign_3, RMSD=0.004 Å, **Sim=40.54%**

Fig. S4. The three-dimensional structures of the target protein, 1HOE (colored green) and the structure of the protein designed by RosettaDesign (colored cyan) and their corresponding RMSDs and sequence similarity to the wild-type sequences.