Supplementary Information

Table S1 Co-crystallized PDB structures used as reference to prepare protein-ligand binding poses for binding affinity calculations in Stage 1a.

Reference PDB	Target ligand
2P83	BACE_145, BACE_146
2QZL	BACE_20
3DV1	BACE_19, BACE_147, BACE_156
3DV5	BACE_4, BACE_5, BACE_11, BACE_21, BACE_22, BACE_23, BACE_24, BACE_25, BACE_26, BACE_27, BACE_28, BACE_29, BACE_30, BACE_31, BACE_32, BACE_33, BACE_34, BACE_35, BACE_36, BACE_37, BACE_38, BACE_39, BACE_87, BACE_88, BACE_90, BACE_91, BACE_92, BACE_93, BACE_94, BACE_95, BACE_143, BACE_148, BACE_149, BACE_150, BACE_151, BACE_153, BACE_157, BACE_158
3K5C	BACE_12, BACE_13, BACE_14, BACE_15, BACE_16, BACE_89, BACE_96, BACE_97, BACE_98, BACE_99, BACE_100, BACE_101, BACE_102, BACE_103, BACE_104, BACE_105, BACE_106, BACE_107, BACE_108, BACE_109, BACE_110, BACE_111, BACE_112, BACE_113, BACE_114, BACE_115, BACE_116, BACE_117, BACE_118, BACE_119, BACE_120, BACE_121, BACE_122, BACE_123, BACE_124, BACE_125, BACE_126, BACE_127, BACE_128, BACE_129, BACE_130, BACE_131, BACE_132, BACE_133, BACE_134, BACE_135, BACE_136, BACE_137, BACE_138, BACE_139, BACE_140, BACE_141, BACE_144, BACE_154, BACE_155
4DPF	BACE_10
4DPI	BACE_6, BACE_7, BACE_8, BACE_9, BACE_40, BACE_41, BACE_42, BACE_43, BACE_44, BACE_45, BACE_46, BACE_47, BACE_48, BACE_49, BACE_50, BACE_51, BACE_52, BACE_53, BACE_54, BACE_55, BACE_56, BACE_57, BACE_58, BACE_59, BACE_60, BACE_61, BACE_62, BACE_63, BACE_64, BACE_65, BACE_66, BACE_67, BACE_68, BACE_69, BACE_70, BACE_71, BACE_72, BACE_73, BACE_74, BACE_75, BACE_76, BACE_77, BACE_78, BACE_79, BACE_80, BACE_81, BACE_82, BACE_83, BACE_84, BACE_85, BACE_86
4GMI	BACE_1, BACE_142, BACE_152

Table S2 Co-crystallized PDB structures used as reference to prepare starting protein-ligand binding poses for binding affinity calculations in Stage 2. PDBs starting with 'BA' are the co-crystal structures released by GC4 organizers for Stage 2.

Reference PDB	Target ligand
3DV1	BACE_147
3DV5	BACE_21, BACE_22, BACE_24, BACE_25, BACE_26, BACE_28, BACE_32, BACE_33, BACE_34, BACE_35, BACE_36, BACE_37, BACE_148, BACE_149, BACE_150, BACE_151
BA01	BACE_1, BACE_152, BACE_156
BA02	BACE_143
BA04	BACE_4, BACE_23, BACE_38
BA05	BACE_5, BACE_27, BACE_29, BACE_30, BACE_31, BACE_39
BA06	BACE_6, BACE_44, BACE_54
BA07	BACE_7, BACE_40
BA08	BACE_8, BACE_49, BACE_72, BACE_73, BACE_74
BA09	BACE_9, BACE_41, BACE_42, BACE_43, BACE_45, BACE_46, BACE_50, BACE_51, BACE_52, BACE_53, BACE_55, BACE_56, BACE_57, BACE_58, BACE_59, BACE_60, BACE_61, BACE_62, BACE_63, BACE_64, BACE_65, BACE_66, BACE_67, BACE_68, BACE_69, BACE_70, BACE_71, BACE_78, BACE_79, BACE_80, BACE_81, BACE_82, BACE_83, BACE_84, BACE_85, BACE_86
BA10	BACE_10, BACE_47, BACE_48, BACE_75, BACE_76, BACE_77
BA11	BACE_11, BACE_87, BACE_88, BACE_90, BACE_91, BACE_92, BACE_93, BACE_94, BACE_95, BACE_153
BA12	BACE_12, BACE_97, BACE_98, BACE_99, BACE_106, BACE_107, BACE_108, BACE_113, BACE_115, BACE_116, BACE_118, BACE_119, BACE_120, BACE_121, BACE_132, BACE_133, BACE_134, BACE_135, BACE_136, BACE_137, BACE_140, BACE_157, BACE_158
BA13	BACE_13, BACE_126
BA14	BACE_14, BACE_100, BACE_101, BACE_102, BACE_103, BACE_104, BACE_105, BACE_114, BACE_117, BACE_122, BACE_123, BACE_124, BACE_127, BACE_128, BACE_129, BACE_130, BACE_131, BACE_144, BACE_145, BACE_146, BACE_154, BACE_155
BA15	BACE_15, BACE_89, BACE_96, BACE_109, BACE_110, BACE_111, BACE_112, BACE_125, BACE_138, BACE_139, BACE_141
BA16	BACE_16
BA18	BACE_142
BA19	BACE_19
BA20	BACE_20

Table S3 Performance of our ligand similarity-based protocol in D3R pose prediction challenge. The table shows the mean and median RMSD[a] for both the 'best' pose and the 'top' pose for our submissions in Stage 1a and Stage 1b, reported as **1a**. We have also reported the number of structures (out of 20) we were able to predict within an accuracy of 1, 2 and 3 Å RMSD. The 'best' pose refers to the pose with the lowest RMSD and the 'top' pose to the pose which our protocol ranked as the best. The reported RMSD values were all calculated by GC4 organizers and are reported in Å.

Docking	RMSD of Best Pose					RMSD of Top Pose	
Prediction	Mean	Median	# <1	# <2	# <3	Mean	Median
Ligand similarity Stage 1a	1.60	1.79	2	12	20	1.87	1.88
Ligand similarity Stage 1b	1.32	1.25	4	18	20	1.33	1.28

Table S4 RMSD (in Å) of the ligand conformer closest to the crystal conformation generated by OMEGA. High values of the 'lowest RMSD' indicates that the native crystal pose was never sampled by OMEGA. The 'lowest RMSD' also sets the limit for the maximum pose prediction accuracy which can be achieved by our protocol, since HYBRID does not do any additional minimization during docking. The RMSD is based on only heavy atoms.

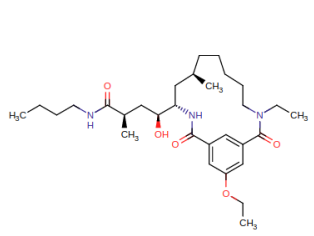
Ligand	Lowest RMSD	Ligand	Lowest RMSD
BACE_1	0.93	BACE_11	1.18
BACE_2	0.94	BACE_12	0.81
BACE_3	0.83	BACE_13	0.95
BACE_4	0.77	BACE_14	1.16
BACE_5	0.97	BACE_15	0.93
BACE_6	0.64	BACE_16	1.25
BACE_7	0.79	BACE_17	0.85
BACE_8	0.68	BACE_18	0.95
BACE_9	0.96	BACE_19	0.89
BACE_10	0.70	BACE_20	1.31

Table S5 Co-crystal PDB structures used as reference for hand-engineered predictions in Stage 1a.

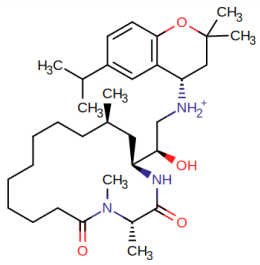
Reference PDB	Target ligand
3DV5	BACE_3
3DV5	BACE_5
4DPF	BACE_6
3K5C	BACE_15
3K5C	BACE_16
2QZL	BACE_20

Table S6 The initial and average distances (in Å) representing the interactions occurring between Asp32 and the BACE_18 ligand and Asp228 and the BACE_18 ligand. The average distances were calculated for the last 10 ns of MD simulations. As shown on Fig 7, the distances were calculated for complexes: A) BACE_18 with protonated Asp32 and deprotonated Asp228 (32p1, 228d1), B) BACE_18 with deprotonated Asp32 and deprotonated Asp228 (32d, 228d), C) BACE_18 with protonated Asp32 and protonated Asp228 (32p1, 228p1), and D) BACE_18 with protonated Asp32 and protonated Asp228 (32p2, 228p2). C and D are different by the orientation of the hydrogens and the oxygen (OD2) of the carboxyl group on their Asp32 and Asp228. The protonation states of the four complexes were held fixed during MD. For complex A: i) the reported distance between Asp32 and the BACE_18 ligand is calculated between the hydrogen connected to OD2 of Asp32 and the closest oxygen of BACE_18, and ii) the reported distance between Asp228 and BACE_18 is calculated between OD2 of Asp228 and the closest hydrogen of BACE_18. For complex B: i) the reported distance between Asp32 and the BACE_18 ligand is calculated between OD1 of Asp32 and the closest hydrogen of BACE_18, and ii) the reported distance between Asp228 and BACE_18 is calculated between OD2 of Asp228 and the closest hydrogen of BACE_18. For complex C: i) the reported distance between Asp32 and the BACE_18 ligand is calculated between the hydrogen connected to OD2 of Asp32 and the closest oxygen of BACE_18, and ii) the reported distance between Asp228 and BACE_18 is calculated between the hydrogen connected to OD2 of Asp228 and the closest oxygen of BACE_18. For complex D: i) the reported distance between Asp32 and the BACE_18 ligand is calculated between the hydrogen connected to OD2 of Asp32 and the closest oxygen of BACE_18, and ii) the reported distance between Asp228 and BACE_18 is calculated between the hydrogen connected to OD2 of Asp228 and the closest oxygen of BACE_18.

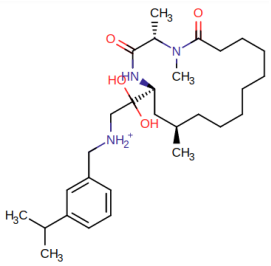
	A: 32p1, 228d1	B: 32d, 228d	C: 32p1, 228p1	D: 32p2, 228p2
Initial distance between Asp32 and BACE_18	3.31	1.62	3.31	4.05
Average distance between Asp32 and BACE_18 during MD	2.58 ± 0.04	1.66 ± 0.01	5.22 ± 0.07	4.20 ± 0.08
Initial distance between Asp228 and BACE_18	1.38	1.38	2.27	1.40
Average distance between Asp228 and BACE_18 during MD	2.49 ± 0.05	1.15 ± 0.12	2.26 ± 0.06	4.31 ± 0.07



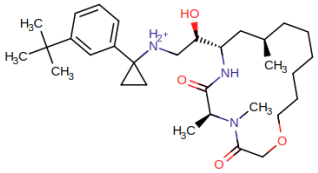
BACE_1



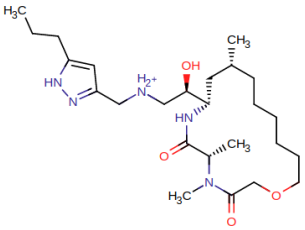
BACE_2



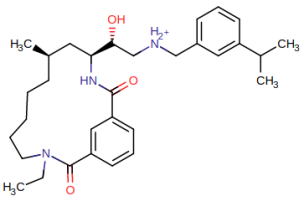
BACE_3



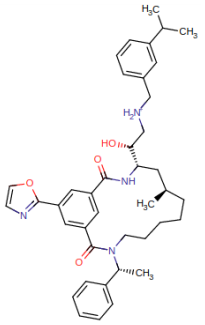
BACE_4



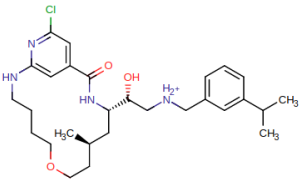
BACE_5



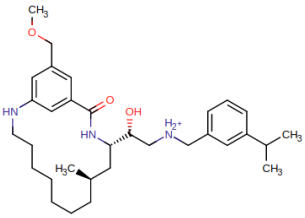
BACE_6



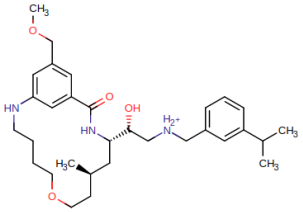
BACE_7



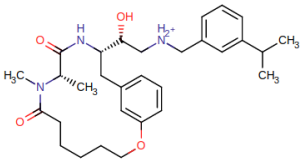
BACE_8



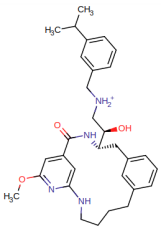
BACE_9



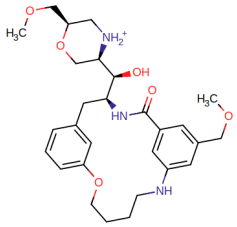
BACE_10



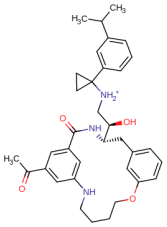
BACE_11



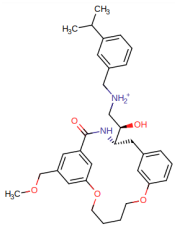
BACE_12



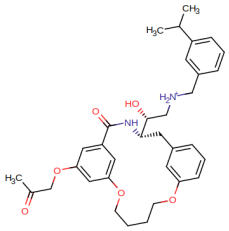
BACE_13



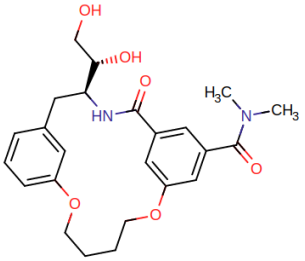
BACE_14



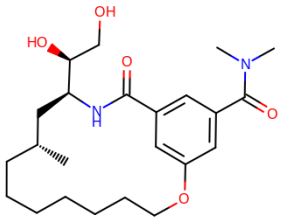
BACE_15



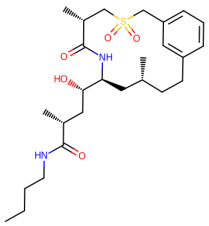
BACE_16



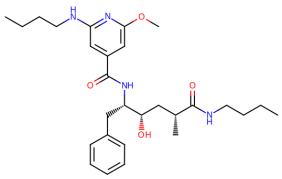
BACE_17



BACE_18



BACE_19



BACE_20

Fig. S1 Structures of the 20 ligands involved in the pose prediction challenge for the target BACE-1.

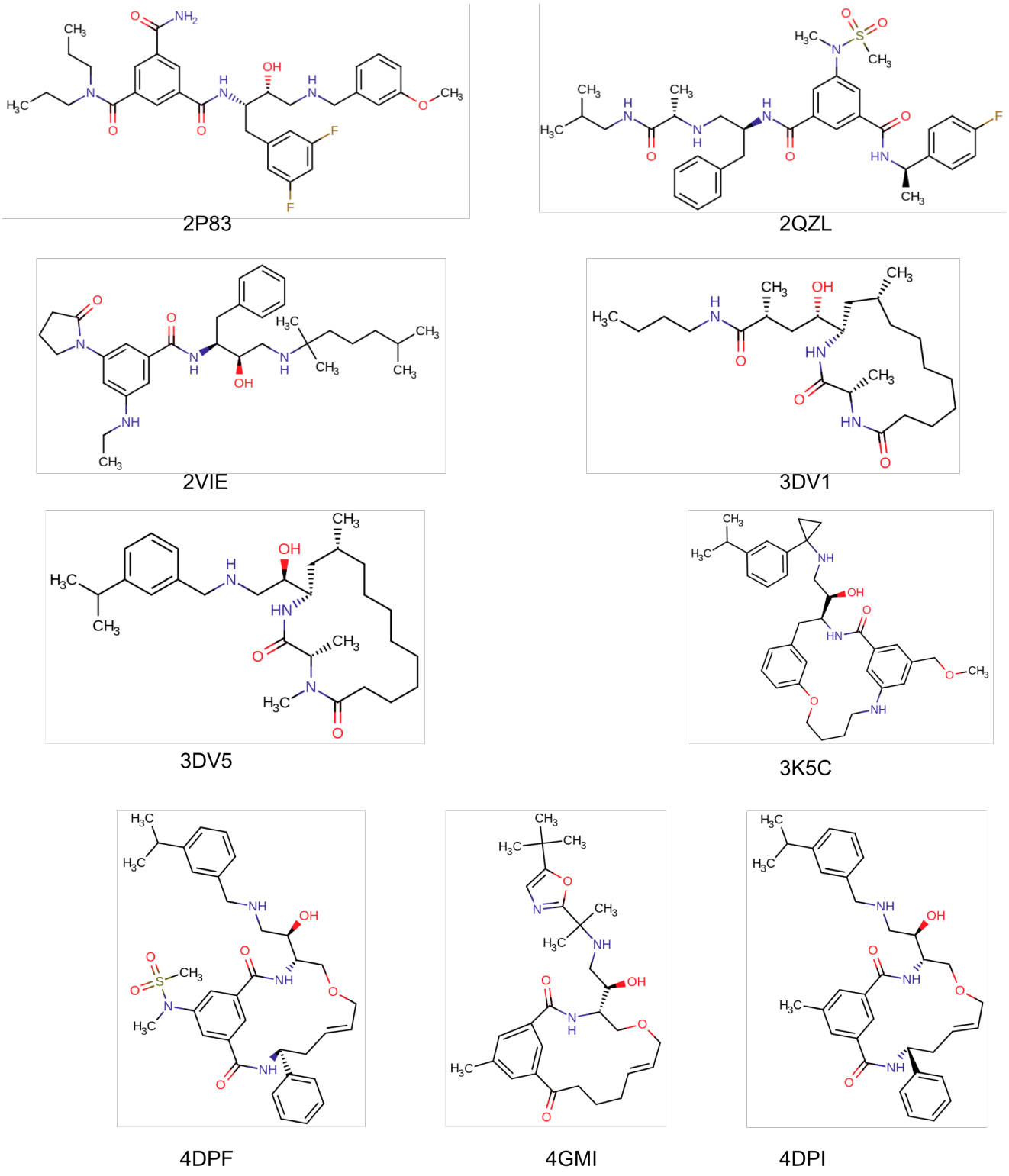


Fig. S2 Structures of the ligands used as references to generate the protein-ligand binding poses for the pose prediction and the affinity ranking challenges. The PDB code corresponding to the BACE-1 co-crystal structure for each ligand is reported under its 2D structure.

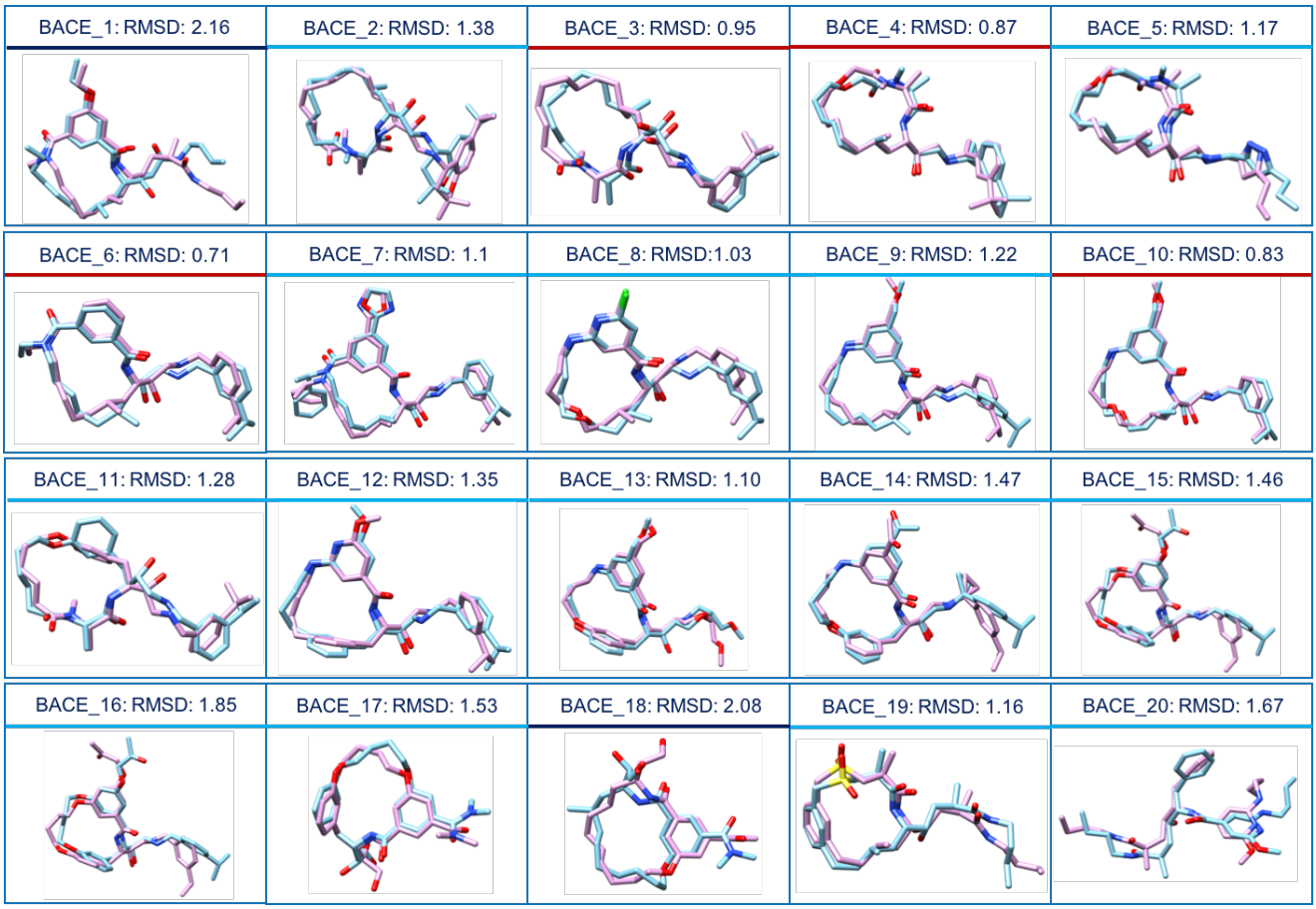


Fig. S3 Comparison of our Stage 1b 'best' poses (in blue) before MD simulations with the corresponding crystal pose (in pink). RMSD values are reported for each pose. Compounds underscored by a red line have RMSD values less than 1 Å, those underscored by a light blue line have RMSD values less than 2 Å and those underscored by a navy blue line have RMSD values higher than 2 Å. We were able to predict 18 out of 20 ligands within an acceptable RMSD accuracy of 2 Å.

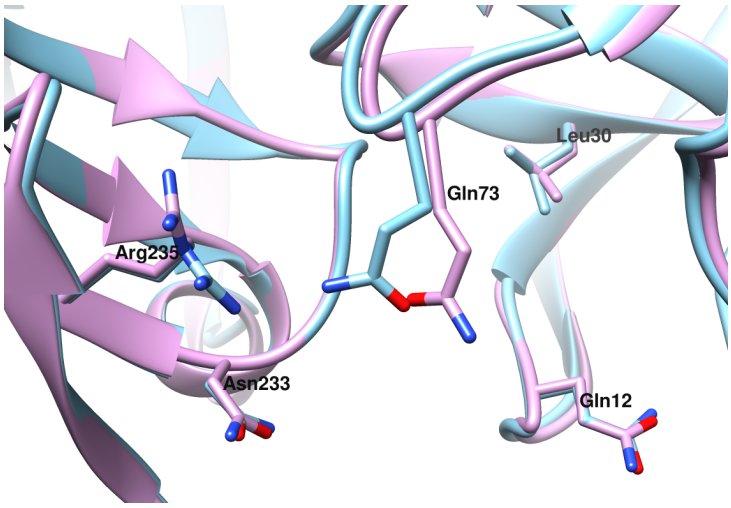


Fig. S4 Differences in the binding site of the proteins 3DV5 (in blue) and BA04 (in pink) used for cross-docking in Stage 1a and self-docking in Stage 1b respectively for the ligand BACE_4. Residues within 5 Å of the ligand and with different side-chain orientations in the two proteins are Gln12, Leu30, Gln73, Asn233 and Arg235. For the residues Gln12 and Asn233, the side-chains differs by 180° rotation of the side-chain dihedral angle. Our docking accuracy improved from 2.23 Å in the cross-docking stage to 0.87 Å in the self-docking stage due to the presence of correct side-chain orientations.

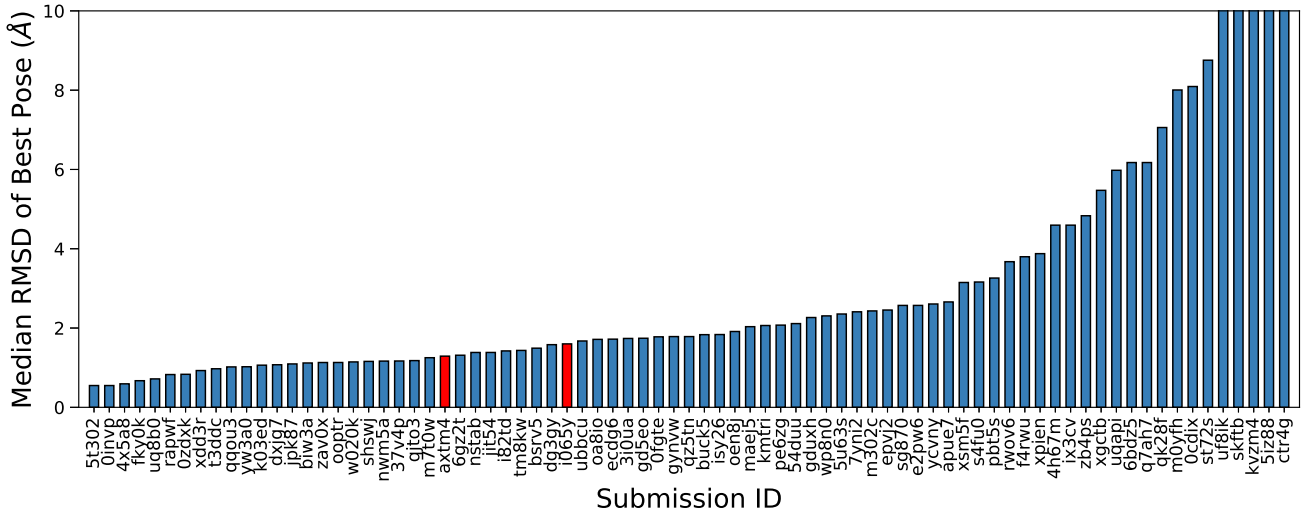


Fig. S5 Ranking of our submissions (red bars) in pose prediction challenge in Stage 1a. The ranking is based on median RMSD of the 'best' pose. The last four submissions had median RMSD > 10 Å. Submission i065y corresponds to our HYBRID-based pose prediction protocol and axtm4 corresponds to our best prediction which includes hand-engineered poses for selected ligands (see Table S5) and poses predicted with our standard HYBRID-based method for the rest of the ligands. Our ligand similarity-based protocol performed reasonably well in Stage 1a of D3R GC4.

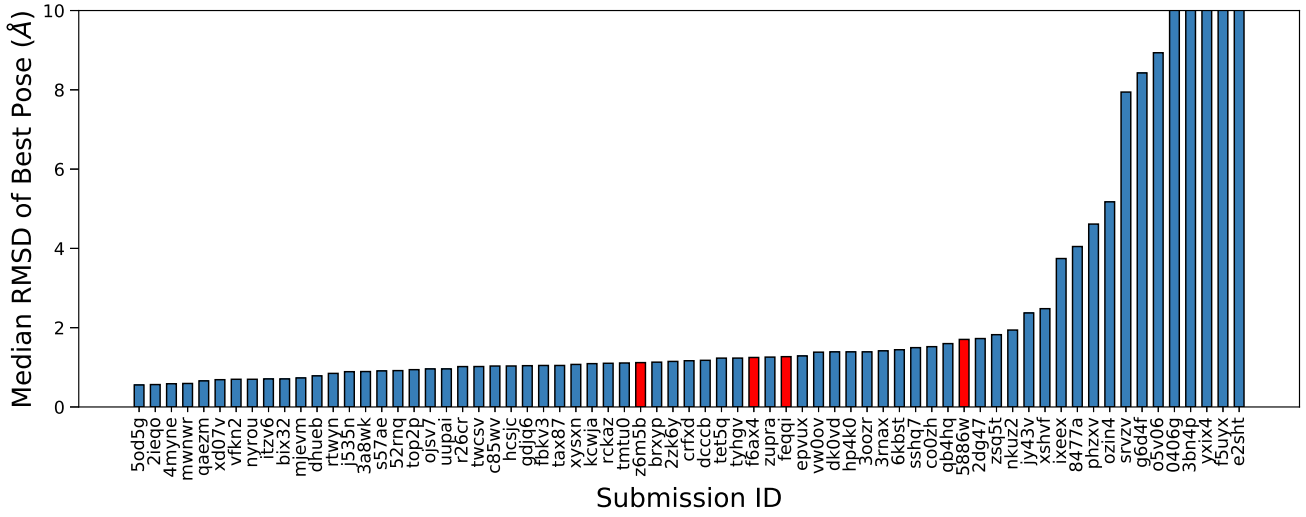


Fig. S6 Ranking of our submissions (red bars) in pose prediction challenge in Stage 1b. The ranking is based on median RMSD of the 'best' pose. The last four submissions had median RMSD > 10 Å. Submissions f6ax4, 5886w, feqqi and z6m5b correspond to our predictions generated using our HYBRID-based protocol, our HYBRID-based protocol at the end of 14 ns of explicit solvent MD simulations, AutoDock Vina and DOCK 6 respectively. Our HYBRID-based protocol performed averagely in this stage.

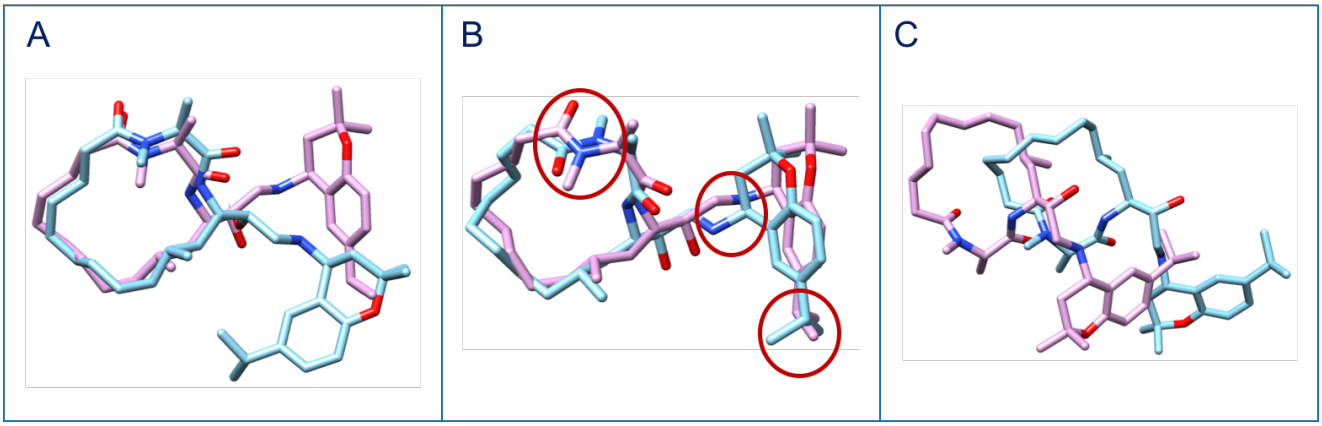


Fig. S7 Comparison of the BACE_2 co-crystal structure with our predicted 'best' pose in Stage 1b. A) Superimposition of the BACE_2 co-crystal structure (in pink) and the BACE_2 co-crystal structure after 14 ns of MD (in blue). B) Superimposition of the BACE_2 crystal structure (in pink) and our BACE_2 'best' ranked pose before MD. The circled regions highlight the conformational differences between the two structures. C) Superimposition of the BACE_2 co-crystal structure (in pink) and our Stage 1b 'best' pose after 14 ns MD simulation (in blue).

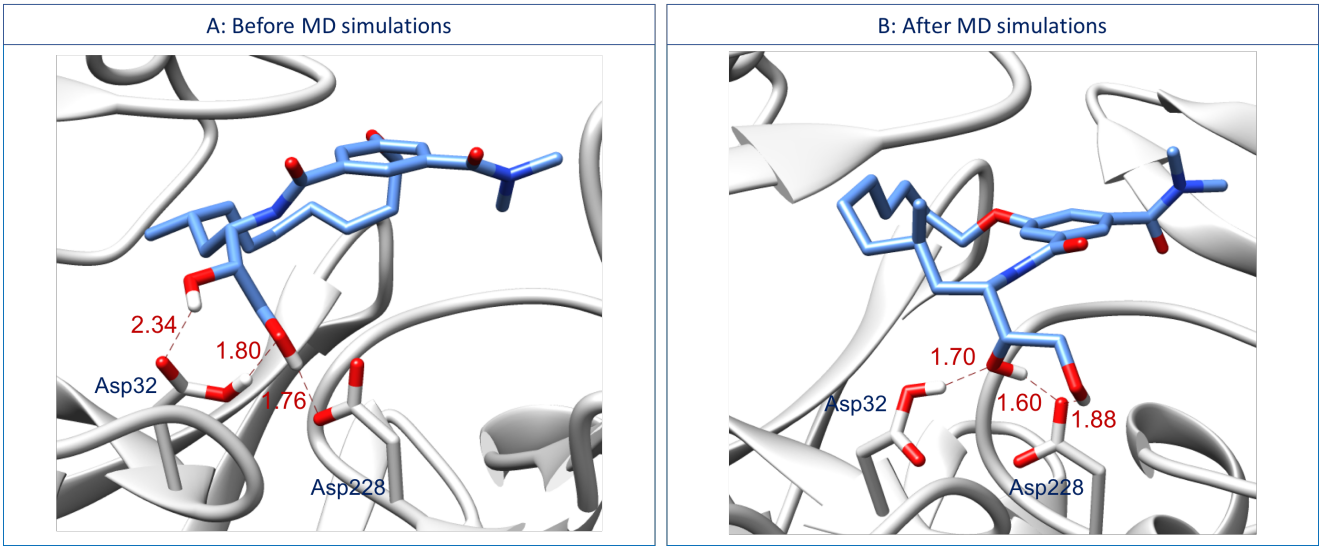


Fig. S8 Potential hydrogen-bonding interactions occurring between the ligand and the aspartyl dyad Asp32/Asp228 in the BACE_18 active site. A) BACE_18 complex submitted in the pose prediction challenge (Stage 1b) before MD simulations and B) BACE_18 complex submitted in the pose prediction challenge (Stage 1b) after 14 ns of MD simulations. The interactions between the aspartyl dyad residues and the BACE_18 ligand before MD are different than those observed after MD; this difference is mainly due to the conformational changes of the ligand and the side-chains of Asp32 and Asp228 during the MD simulations.

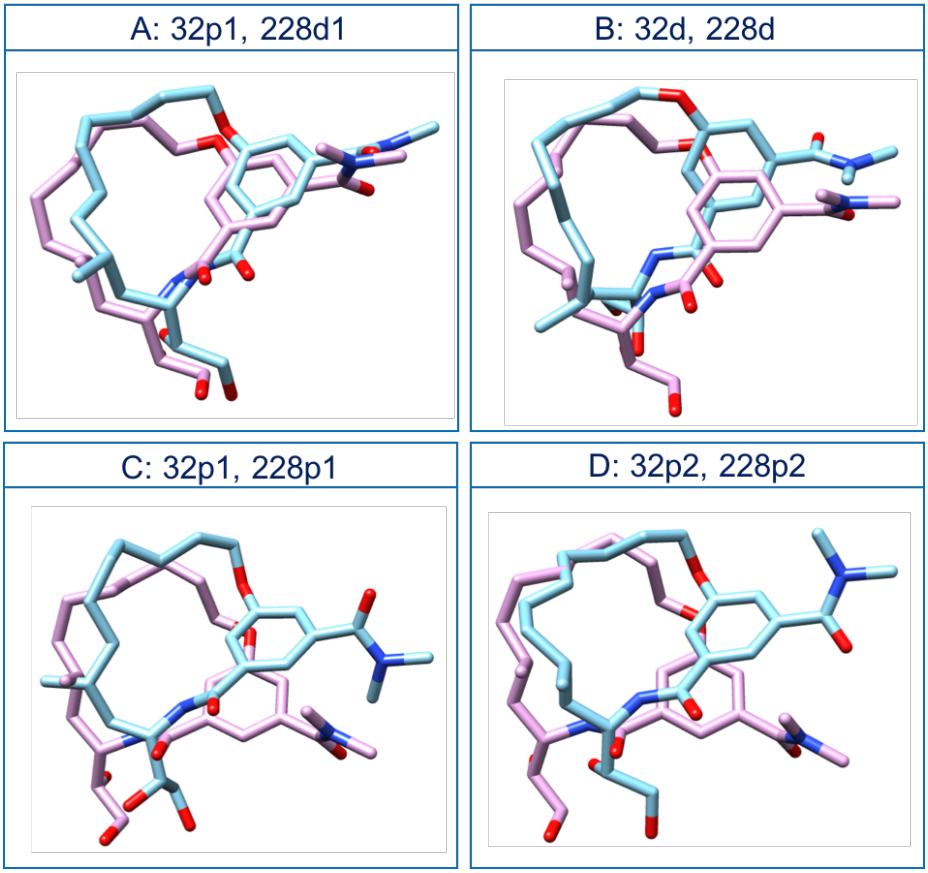


Fig. S9 Superimposition of the released BACE_18 structure (in pink) and the equilibrated structures of the predicted binding poses for the BACE_18 ligand (in blue) with different protonation states of the aspartyl dyad- A) BACE_18 with protonated Asp32 and deprotonated Asp228 (32p1, 228d1), B) BACE_18 with deprotonated Asp32 and deprotonated Asp228 (32d, 228d), C) BACE_18 with protonated Asp32 and protonated Asp228 (32p1, 228p1), and D) BACE_18 with protonated Asp32 and protonated Asp228, (32p2 228p2). The predicted binding poses are extracted at the end of a 14 ns MD trajectory performed on the 4 BACE-1-ligand complexes. The four ligand poses A, B, C, and D after MD are different.

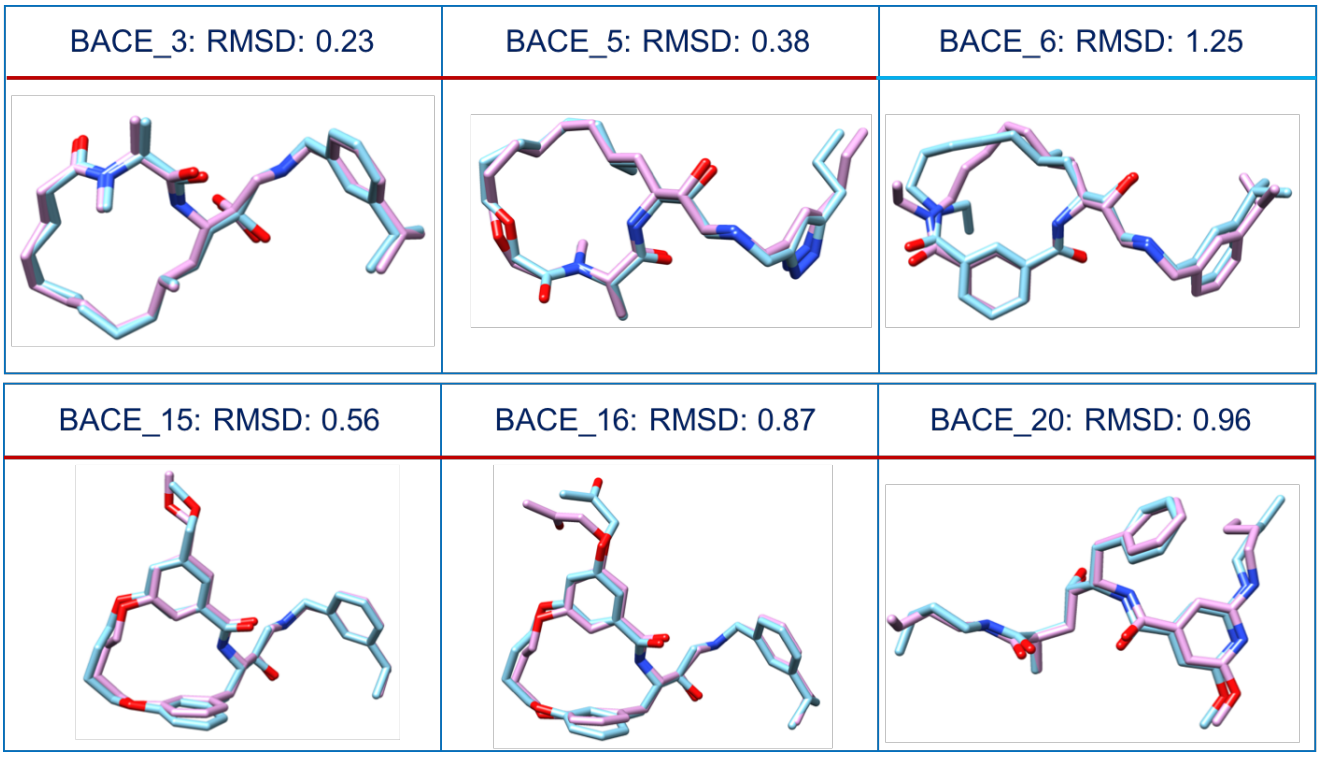


Fig. S10 Comparison of our Stage 1a hand-engineered poses (in blue) with the corresponding crystal poses (in pink). RMSD values are reported for each pose. Compounds underscored by red and blue lines have RMSD values less than 1 Å and less than 2 Å respectively. The hand-engineered poses did significantly better than those from our HYBRID-based protocol.

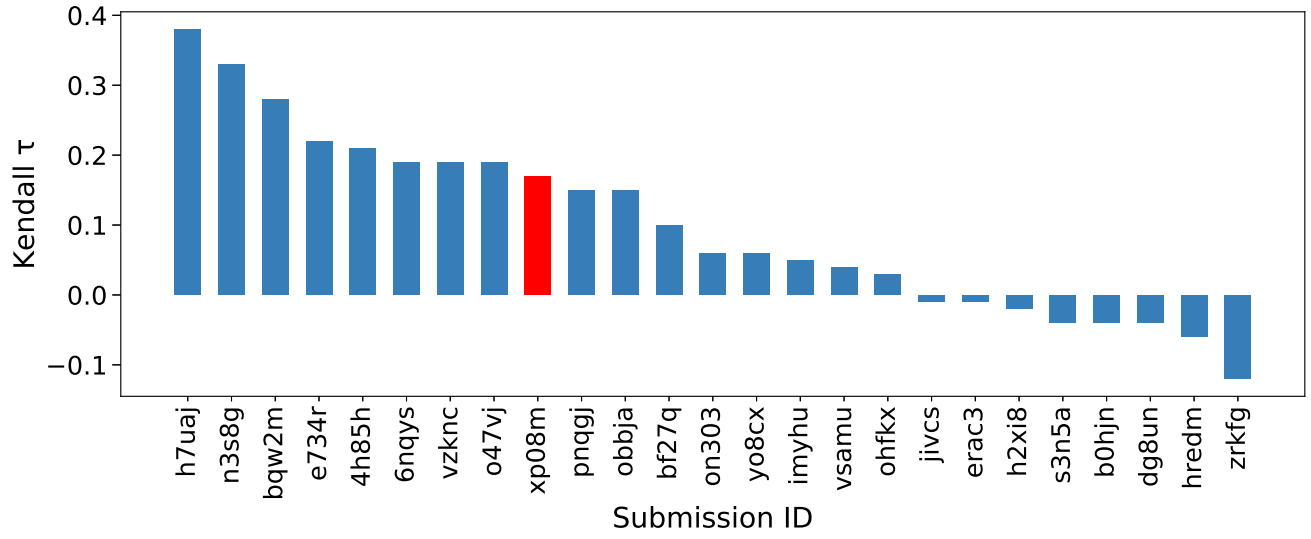


Fig. S11 BACE-1 Kendall's τ correlation coefficient scores between the predicted structure-based affinity ranking and the experimental affinity ranking in Stage 1a. The red bar corresponds to the ranking of our structure-based affinity ranking submission, xp08m, based on MM-GBSA calculations. The blue bars represent the submissions of the other participants of the D3R GC4. The evaluation of all the affinity ranking submissions with different metrics was done by the D3R GC4 organizers and could be found on the D3R website (https://drugdesigndata.org).

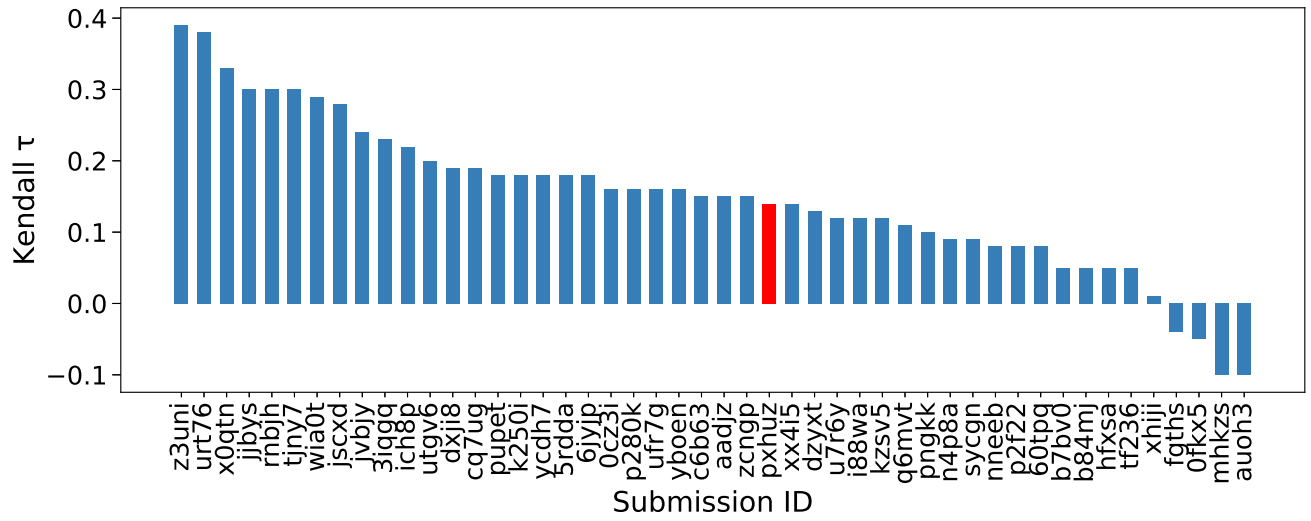


Fig. S12 BACE-1 Kendall's τ correlation coefficient scores between the predicted structure-based affinity ranking and the experimental affinity ranking in Stage 2. The red bar corresponds to the ranking of our structure-based affinity ranking submission, pxhuz, based on MM-GBSA calculations. The blue bars represent the submissions of the other participants of the D3R GC4. The evaluation of all the affinity ranking submissions with different metrics was done by the D3R GC4 organizers and could be found on the D3R website (https://drugdesigndata.org).

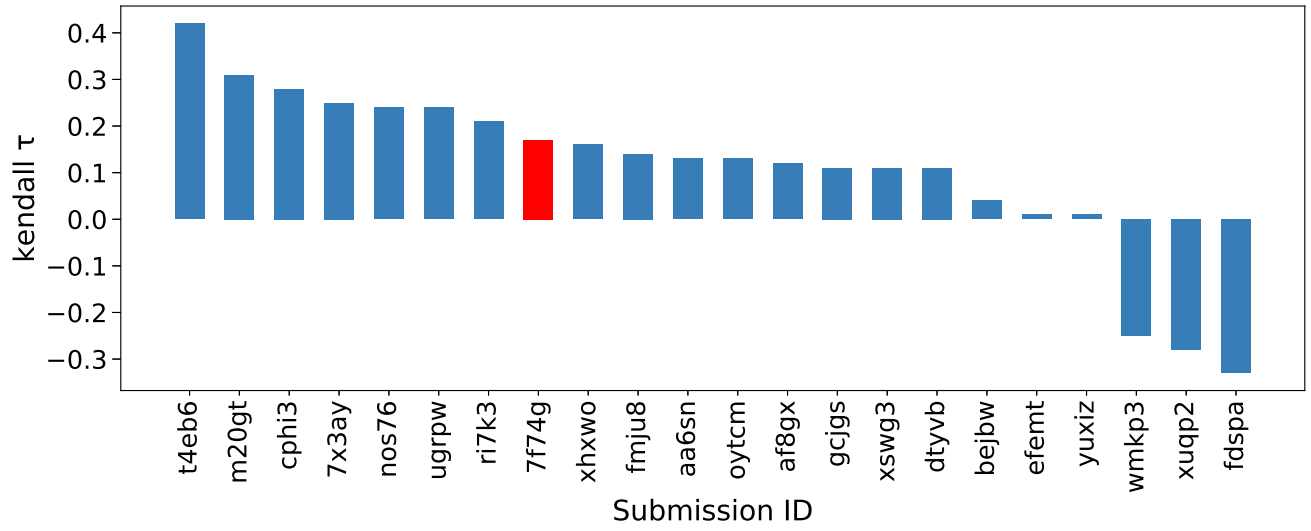


Fig. S13 BACE-1 Kendall's τ correlation coefficient scores between the predicted relative binding free energies and the experimental binding affinities in Stage 1a. The red bar corresponds to the ranking of our relative affinity ranking submission, 7f74g, based on MM-GBSA calculations. The blue bars represent the submissions of the other participants in D3R GC4. The evaluation of all the affinity ranking submissions with different metrics was done by the D3R GC4 organizers and could be found on the D3R website (https://drugdesigndata.org).

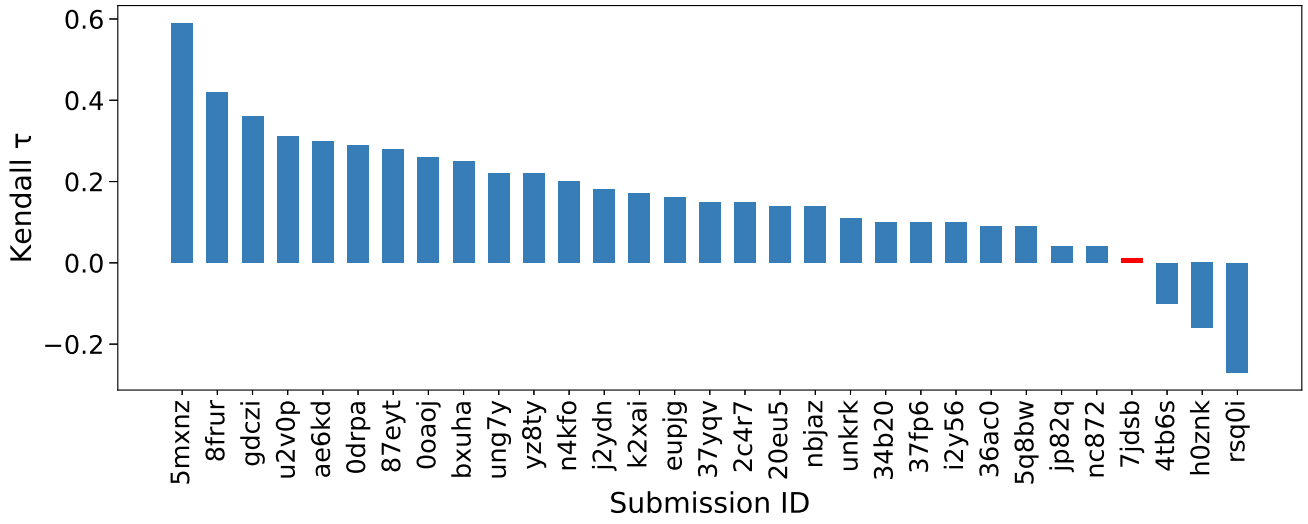


Fig. S14 BACE-1 Kendall's τ correlation coefficient scores between the predicted relative binding free energies and the experimental binding affinities in Stage 2. The red bar corresponds to the ranking of our relative affinity ranking submission, 7jdsb, based on MM-GBSA calculations. The blue bars represent the submissions of the other participants in D3R GC4. The evaluation of all the affinity ranking submissions with different metrics was done by the D3R GC4 organizers and could be found on the D3R website (https://drugdesigndata.org).