Supplementary Material

Our protocol for extracting contacts from DOCK ensembles

DOCK ensembles of each ligand-protein system were obtained from the SB2012 dataset.[1] Some poses in the DOCK ensemble sample similar binding regions regions in cartesian space, which we call the binding patterns(BPs). These BPs, which encode the consensus among different poses, can be used to extract common contacts shared among different poses, thus reducing the redundancy of the original DOCK ensembles. This step also reduces ambiguity and noise through retaining only common contacts. In order to extract the BPs, it is first necessary to quantify the similarity among poses and then to cluster together the similar ones. To quantify the similarity between a pair of poses, we measure the closeness of ligand heavy atoms in cartesian space. The result is converted to a binary vector, called the spatial similarity vector (Eq. 1). To filter out the trivial BPs (in which two poses have only a few atoms close in Cartesian space), we evaluate the size of binding pattern of any pose pairs by computing the Hamming weight[2] (Eq. 3) of its spatial similarity vector. The ones with Hamming weight less than one-fifth of the size of its spatial similarity vector is defined as trivial BPs. Those pose pairs are removed. To further eliminate redundant pose pairs, we conduct clustering among those remaining pose pairs. First, we convert the similarity among pose pairs into a graph (G) where each similar pose pair is represented by vertices connected by one edge. The connectivity of this graph encodes the poses similarities while the similarity among group of poses is captured by the so-called sub-complete graph (where its vertices are all connected with each other). Therefore clustering similar poses is equivalent to identifying sub-complete graphs in G. The BP shared among similar poses in a cluster is identified using the effective spatial similarity vector (Eq. 4). The information in the BP is the set of contacts between ligand heavy atoms and any protein C_α atoms. We convert this information into MELD restraints. Two atoms are defined to be in contact if their distance is shorter than 5Å. The contacts of each BP are enforced separately in MELD. Each set of contacts is enforced at 70% confidence, and only one set is active at a time.

Measuring similarity among poses in the DOCK ensembles.

To measure the similarity between any two poses (i,j), we defined the spatial similarity vector. The vector consists of 0's and 1's, and its length equals the number of heavy atoms of the ligand.

$$\overrightarrow{V}_{\mathrm{i,j}} = [1, 1, 1, 1, 1, 0, ...0] \tag{1}$$

For k-th atom in poses i and j:

$$\overrightarrow{V}_{\mathrm{i,j}}[k] = \begin{cases} 1, & d_{\mathrm{i,j}}[k] \leq d_{\text{cut-off}} \\ 0, & d_{\mathrm{i,j}}[k] > d_{\text{cut-off}} \end{cases} \tag{2}$$

Therefore, given two poses i and j, the spatial similarity vector was generated by calculating the Euclidean distance of all corresponding heavy atoms ($d_{\mathrm{i,j}}$) and comparing them with the cut-off distances. The positions of 1's in the spatial similarity vector denote a subset of atoms in poses (i,j) that are close to each other. We called it the binding pattern (BP).

To quantify the similarity between any two poses (i,j), we defined a operation associated with spatial similarity vector, the Hamming weight, which is $W_H(\overrightarrow{V_{\mathrm{i,j}}})$

$$W_H(\overrightarrow{V}_{\mathrm{i,j}}) = \text{count of maximum consecutive 1's in } \overrightarrow{V}_{\mathrm{i,j}}. \tag{3}$$

To quantify the similarity among a group of similar poses (a.k.a BPs), we defined the effective spatial similarity vector $\overrightarrow{V}_{\mathrm{eff}}$

$$\overrightarrow{V}_{\mathrm{eff}} = \overrightarrow{V}_1 \cap ... \overrightarrow{V}_i ... \cap \overrightarrow{V}_n, \overrightarrow{V}_i \in \mathbf{V}; \tag{4}$$

where $\boldsymbol{V}$ consists of the spatial similarity vectors of all possible pose pairs in a cluster.

Additional push restraints to accelerate ligand unbinding

During the initial MELD x MD docking simulations, we observed that unbinding a ligand from its receptor is difficult, especially for charged ligands. Theoretically, we could overcome this difficulty by increasing the highest 'temperature' of our replica exchange ladder. However, to keep a good exchange probability, this solution required more replicas and added additional computational cost. Instead, we adopted a different strategy by adding push restraints between the center-of-mass of protein and the ligand carbon atoms. We made four copies of those push restraints and put them into four MELD restraint groups. These restraint groups split the unbinding path into four steps by incrementally increasing the R_2 of MELD's "harmonic distance restraints" in each restraint group. Therefore, as the "walker" climbed along the replica exchange ladder, the ligand was gradually pushed away from the receptor.

Details about unbinding push restraints are given here. The closest distances of any ligand carbon atoms from the protein center-of-mass (defined as the Cα atom closest to the protein's geometrical center-of-mass) is defined as $d_{\mathrm{com,crystal}}$. The largest distance the ligand can be pushed is set to 25Å, considering the protein's radius of gyration and the size of the solvent box. The unbinding path between protein and ligand is evenly split into four regions:

$$\text{unbinding path } = \begin{cases} [d_1, d_5] & d_1 = d_{\mathrm{closest}} \\ [d_2, d_5] & d_2 = d_1 + (d_5 - d_1)/4 \\ [d_3, d_5] & d_3 = d_2 + (d_5 - d_1)/4 \\ [d_4, d_5] & d_4 = d_3 + (d_5 - d_1)/4 \end{cases} \tag{5}$$

where $d_{\text{closest}} = d_{\text{com,crystal}} - 2\text{Å}$ and $d_5 = 25\text{Å}$.

Determining the alignment error due to receptor flexibility

We introduced more flexibility to MELD x MD binding simulations than was formerly available in rigid docking systems. This allowed us to more accurately capture the flexibility of the system, but it introduced uncertainty when the quality of the prediction was assessed.

A measure widely used to describe the quality of docking predictions is the ligand RMSD (LRMSD). This is the RMSD of the atoms of the ligand computed after aligning on the receptor. In rigid docking, the receptor usually does not change its shape, while in MELD x MD docking simulations the receptor atom positions fluctuate around their equilibrium position. This flexibility limits our ability to align the receptor with the static crystal structure, and intrinsically results in a higher value of LRMSD for our system. To quantify the influence of this intrinsic alignment error on the value we reported in the Results section, we performed MD simulations using Amber, in which we applied Cartesian restraints to the positions of heavy atoms of the protein residues within 5Å of any ligand heavy atoms and also to the ligand heavy atoms themselves. The restraint strength was 100 $\text{kcal}\,\text{mol}^{-1}\,\text{Å}^{-2}$. After minimization and equilibration, each system was run for 100 ns. The LRMSD of each frame in the trajectory was computed by aligning receptor C_α atoms against the native crystal structure. In Figure S1 we report the distribution of LRMSD from the native for the 30 systems we simulated in restrained runs.

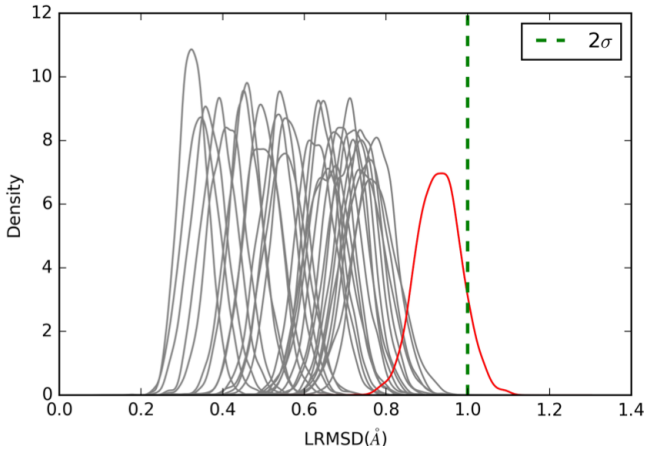

Figure S1: **The effect of receptor flexibility on the computation of LRMSD.** Here ligand atoms and receptor interface atoms are restrained in place, so the observed LRMSD comes only from the uncertainty in aligning the flexible receptor. Red indicates the system with the largest mean alignment error. The green vertical dashed line corresponds to the 2σ variance of the underlying error distribution.

As Figure S1 shows, even for almost perfect docking (where the interface RMSD is close to zero due to the Cartesian restraint), the LRMSD still indicates some errors. We need therefore to account for this uncertainty when we assess our results. This uncertainty in each system can be approximated using a Gaussian distribution. Considering the 2σ variance

of the system with the largest uncertainty (the red one in Fig S1), we chose 1.0Å as the alignment error.

Protocol to identify ligand-protein contacts in the MELD predicted and native poses

The Protein-Ligand Interaction Profiler[3] (PLIP, version 2.1.3) is used to report ligand-protein contacts from predicted poses. Even when sampling the native binding pose, the ligand and the binding site atoms can rearrange inside their thermal envelop. Similar (but not identical) sets of contacts can be established between the receptor and the ligand. For example, equivalent salt bridges can be established between a positively charged atom and any of the two oxygen of a de-protonated carboxy group. To account for this contact flexibility, when comparing our predictions with native structures, we describe intermolecular contacts as "residue to residue" rather than "atom to atom". Prior to the contact calculation, CPPTRAJ[4] is used to remove all solvents/ions sitting more than 5Å away from any ligand heavy atoms. DOCK predicted poses along with the receptors extracted from the SB2012 dataset are converted from mol2 to pdb format using Chimera[5] prior to using PLIP.

Protocol to run AutoDockFR docking simulation

The AutoDockFR[6,7] (a.k.a ADFRsuite, BUILD 5 ,Oct_28_19) is used to perform rigid as well as flexible receptor docking experiment. The crystal receptor structure of each complex, which is used for both types of docking studies, is extracted from the SB2012 dataset. A random unbound conformation of ligand used for docking is selected randomly among all frames in the highest temperature replica generated during the MELD x MD prediction. The AutoDock4 force field[8] is used to describe both the ligand and the receptor. Atom types and the Gasteiger partial charges are assigned to ligand and receptor atoms using the 'prepare_ligand' and 'prepare_receptor' tools. A grid box, which encompasses the native ligand pose (–boxMode ligand), with the default padding space of 8Å (except 1BB5, for which a 12Å is used in flexible receptor docking) and 0.375Å grid point spacing is computed using the AGFR tool inside the ADFRsuite. The ADFR tool inside the ADFRsuite is used to perform docking with default values for nbRuns (50) and maxEvals (2500000). For the flexible receptor docking simulations, protein residues that are in close contact (2.0Å) with the ligand in the crystal pose are chosen as flexible. The ligand root-mean-square distance cut-off values to determine success are 2.5Å[7] and 2.0Å for flexible and rigid receptor docking respectively.

Explicit solvent model is necessary for accurately modeling native pose

Water molecules play important roles in ligand-protein association. The displacement of water from the binding pocket can have important effects on the energetic of the the binding process. Water mediated interactions between ligand and protein are important for ligand recognition.[9] These effects are not captured by implicit solvent models. Explicit solvent model like TIP3P[10] better capture the unique effects that hydrogen bonds have on the behaviour of liquid water. This is highlighted by benchmarks of ligand-protein complexes stability using explicit TIP3P water or GB-neck2[11] solvent model.

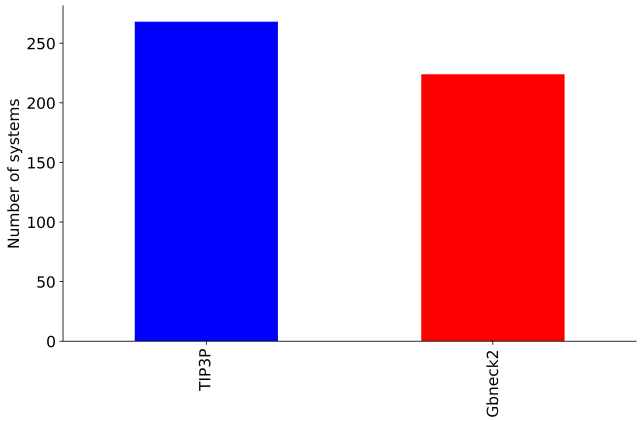

Figure S2: **The effect of solvent model on the stability of native pose.** The blue and red bars denote the number of stable ligand-protein complexes simulated using TIP3P or GB-neck2 solvent model.

We find that more ligand-protein complexes are stable in TIP3P explicit solvent than in GB-neck2 implicit solvent. Also, GB-neck2 is not parameterized for elements like Cl and Br which appear often in small molecule ligands.[11] With all those concerns in minds, we decided to simulate the ligand binding process using TIP3P explicit solvent – a first for MELDxMD based simulations.

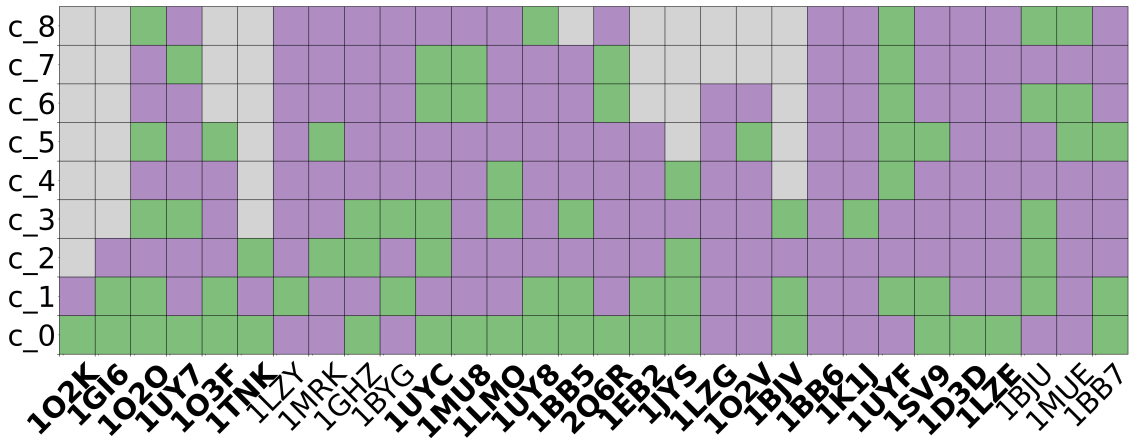

Figure S3: **MELD x MD is able to discover novel poses.** Indices of MELD predicted poses are labeled on y-axis in population descending order with invalid ones color as grey. DOCK-like (a.k.a one with LRMSD less than 3Å to any of the TOP30 DOCK poses) or novel ones are colored as green and pink respectively. Bold labels highlight the MELD x MD successes.

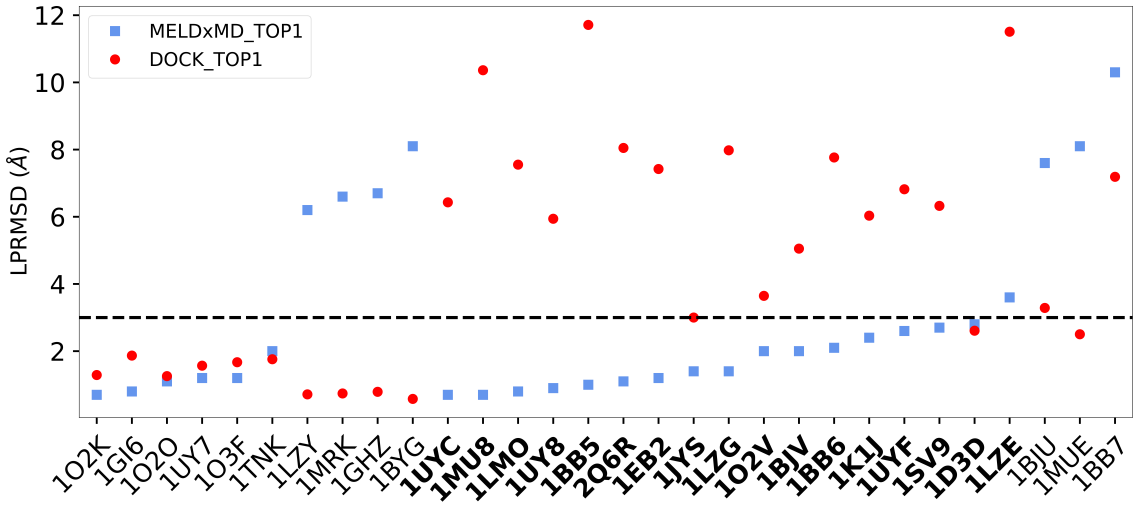

Figure S4: **LRMSD of TOP1 poses predicted by MELD x MD and DOCK.** LRMSD for MELD x MD TOP1 prediction (blue dot) and DOCK TOP1 pose (red dot) of all 30 testing ligand-protein complexes. The right most 20 systems are original DOCK failures. Bold are original DOCK failure cases that MELD x MD provides native-like pose at TOP1 level. The black dashed line represents a 3Å cut-off.

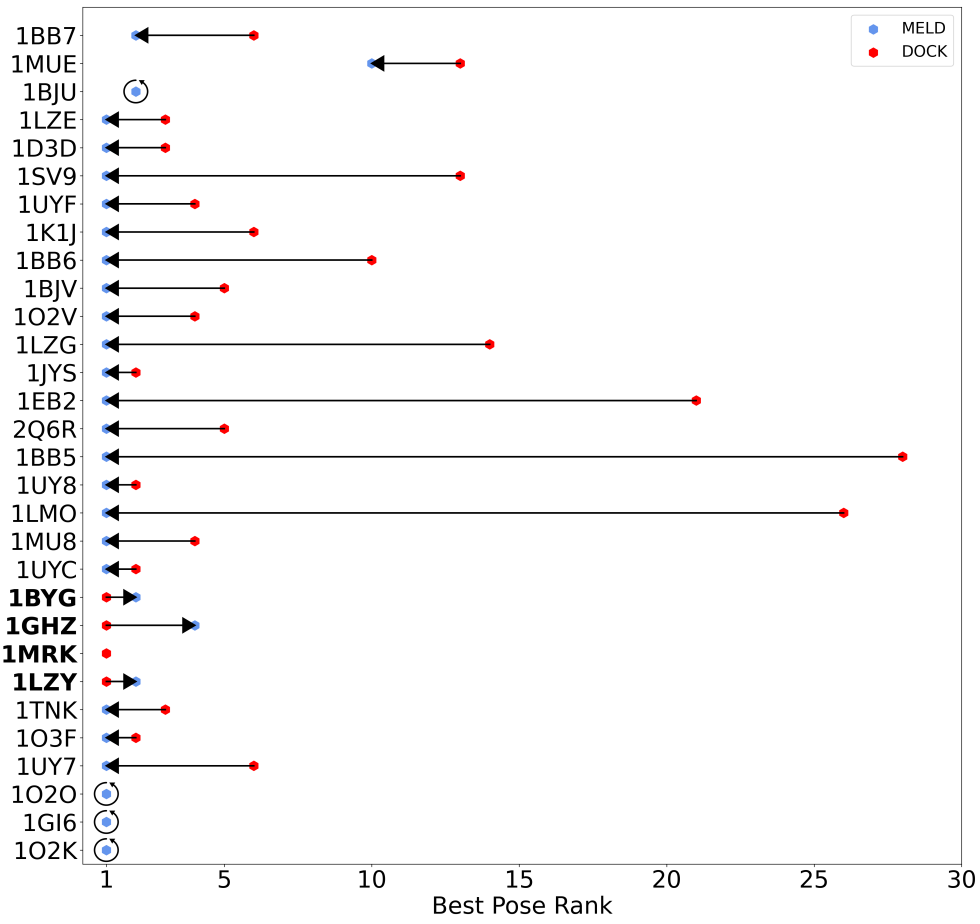

Figure S5: **MELD x MD improves the best pose ranking.** The ranking of native pose in MELD (blue dot) and most native-like one in DOCK (red dot) are given. A leftward arrow indicates an improvement in ranking by MELD. A cycle indicates a tie. A rightward arrow indicates a deterioration in ranking by MELD. No arrow for 1MRK as no native pose predicts by MELD x MD.

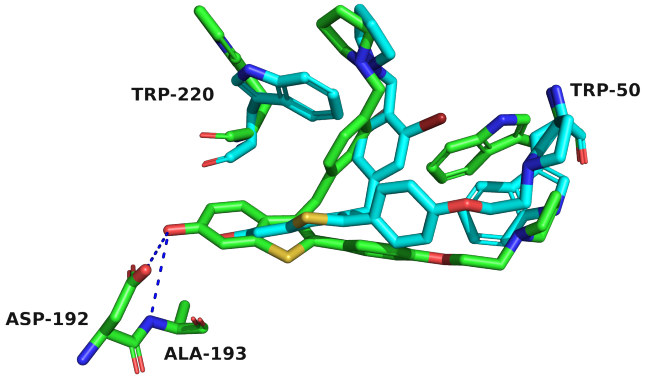

Figure S6: **Different sidechain conformations of TRP50 and TRP220 in native and MELD TOP1 poses of 1D3D.** Sidechain conformations of TRP50 and TRP220 in the MELD TOP1 pose (Cyan) deviate a lot from native ones (Green) causing close contacts and even steric clash with ligand in crystal conformation. It thus prevent key contacts to be formed (highlights the missing native hydrogen bonds as blue dash lines.)

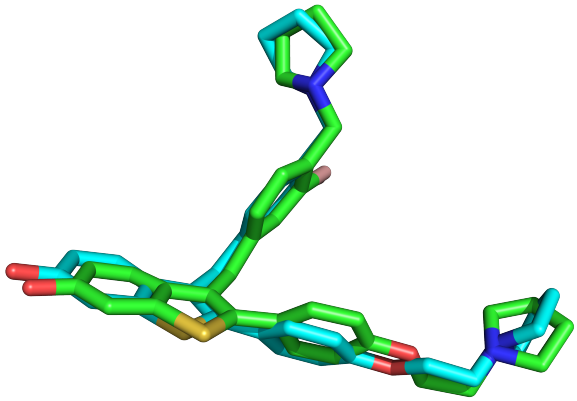

Figure S7: **Better structure exists in the MELD TOP1 ensemble of 1D3D.** A better pose (Cyan) in the MELD TOP1 ensemble is shown along with the native one (Green). Superposition is performed on protein backbone atoms.

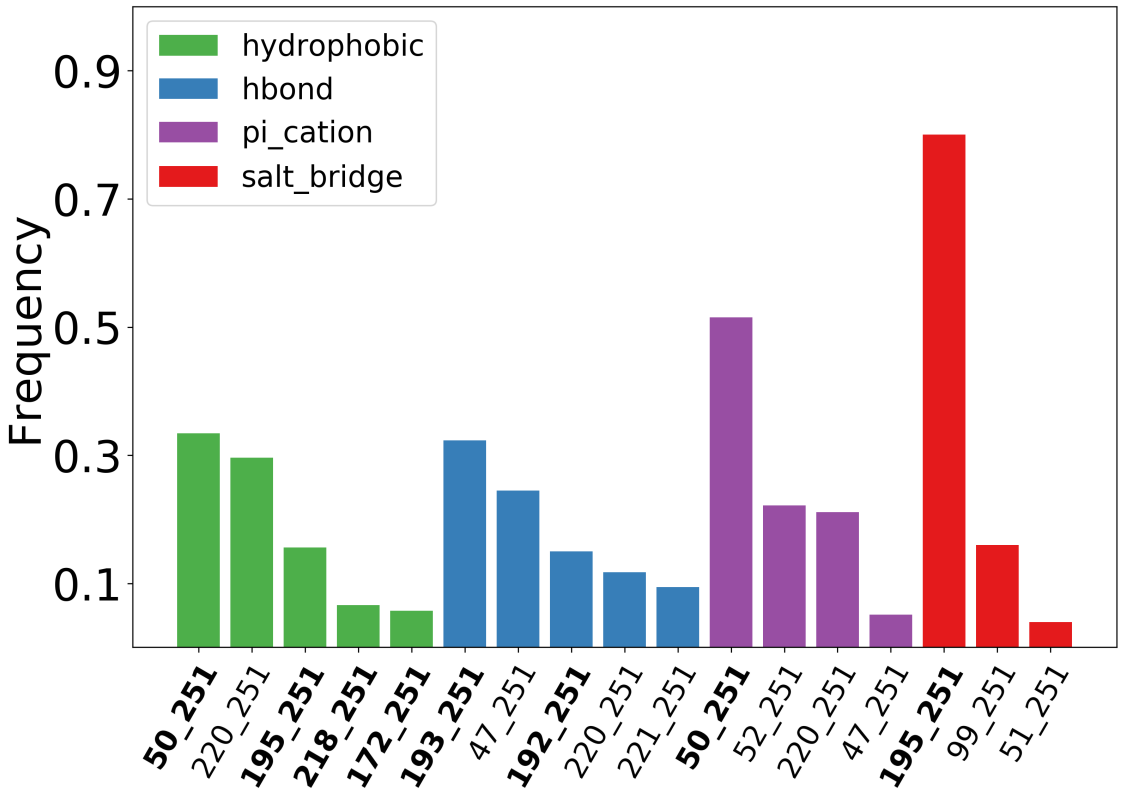

Figure S8: **Frequency of different contacts in the MELD TOP1 ensemble of 1D3D.** Contact frequency of the TOP5 hydrophobic (Green), hydrogen bond (Blue), pi-cation (Purple), salt bridge (Red) interactions are shown. The bar labels on the x-axis are the indices of the corresponding residues where the native ones are bold. Based on the frequency of different kinds of contacts, we are able to recover most of the native ones.

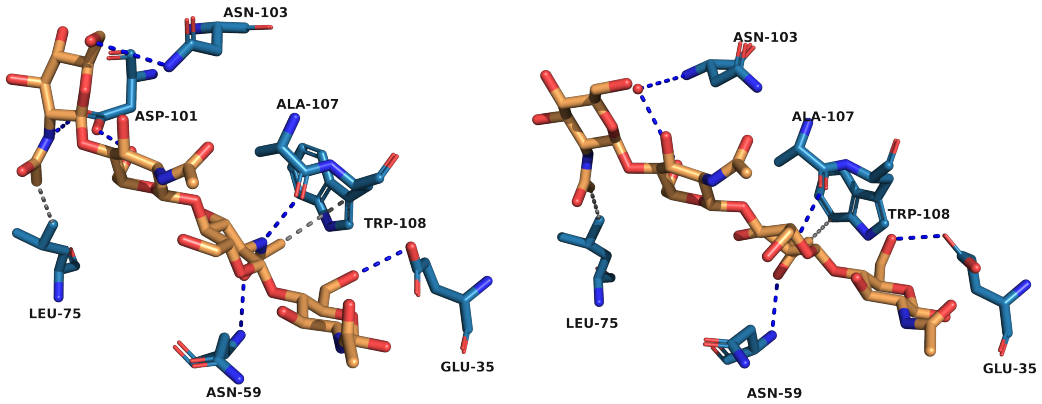

Figure S9: **Important ligand-protein contacts presents in native(Left Panel) and MELD x MD TOP1 prediction(Right Panel) for 1LZE.** The overall binding modes are similar indicating by the similar binding pattern consists of hydrogen bonds(blue dashed line) and hydrophobic interactions(gray dashed line) in the MELD x MD TOP1 prediction.

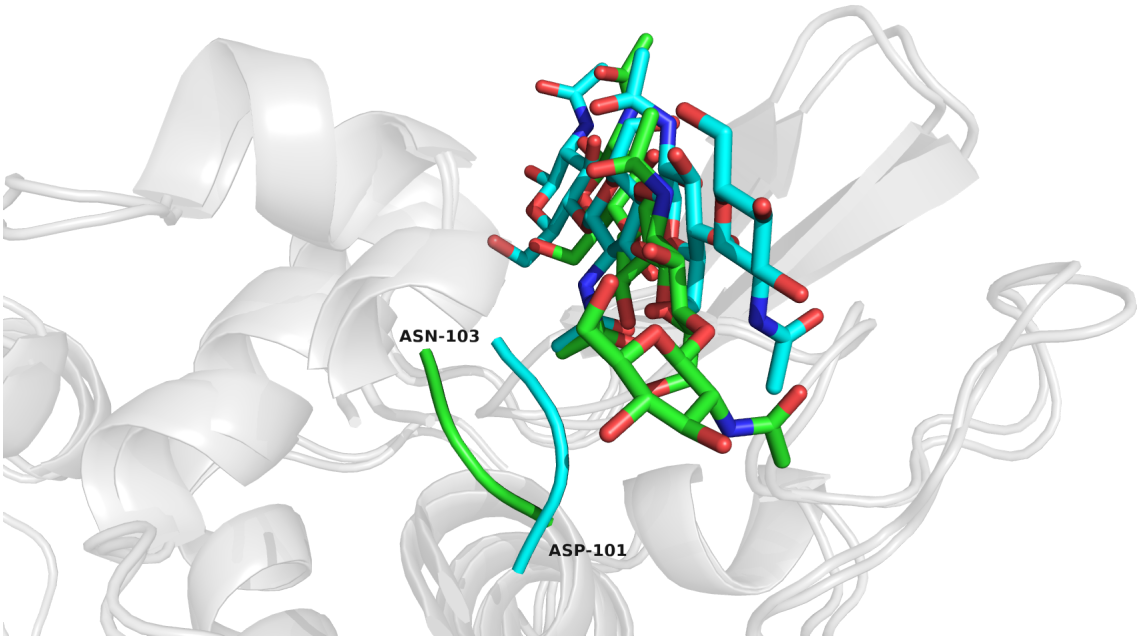

Figure S10: **Displaced loop in 1LZE 'push" ligand away from its crystal pose.** Ligand in crystal (green licorice) and MELD TOP1 prediction (cyan licorice) along with the displaced loop (ASP101 to ASN103) are highlighted.

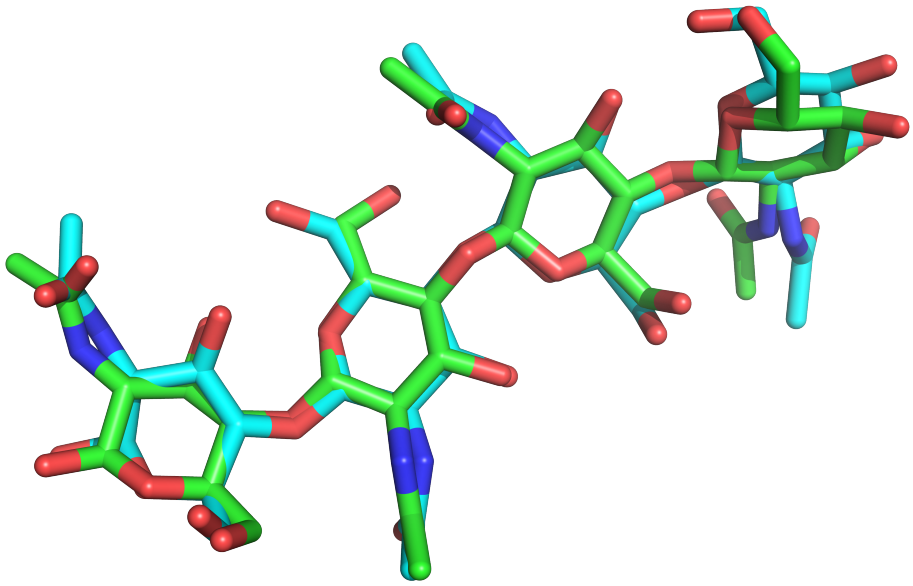

Figure S11: **Better structure exists in the MELD TOP1 ensemble for 1LZE.** Better pose (cyan, LRMSD: 0.9Å) in the MELD TOP1 ensemble is shown along with the native one (green). Superposition is performed on protein backbone atoms.

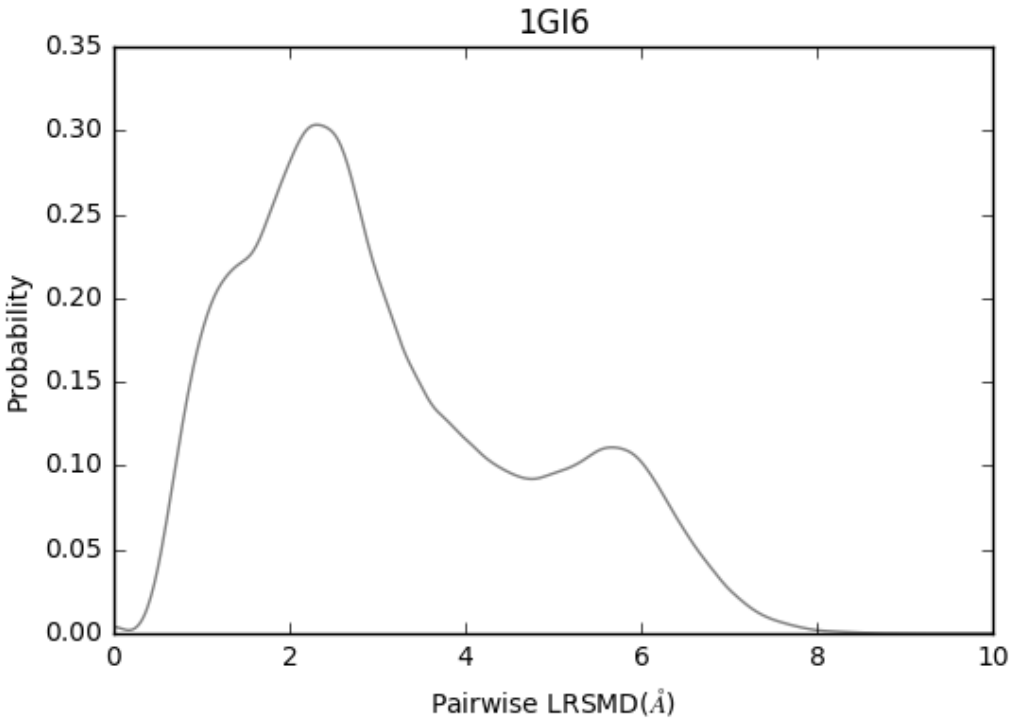

Figure S12: **Pairwise ligand RMSD probablility distribution of the most populated non-native cluster in 1GI6.** Two separated peaks located at about 2 Å and 6 Å indicate that this is a heterogeneous cluster.

Table S1: **Correct Predictions of the Native pose of 23/30 Complexes**[a]

	MELD TOP1		DOCK TOP1	
PDBID	**LRMSD(Å)**	**fTR %**	**LRMSD(Å)**	**fTR %**
1O2K	0.7	100	1.3	57
1GI6	0.8	100	1.9	42
1O2O	1.1	85	1.3	85
1UY7	1.2	100	1.6	83
1O3F	1.2	100	1.7	85
1TNK	2.0	50	1.8	0
1LZY	6.2	37	0.7	100
1MRK	6.6	83	0.7	66
1GHZ	6.7	20	0.8	100
1BYG	8.1	66	0.6	83
1UYC	0.7	100	6.4	50
1MU8	0.7	87	10.4	25
1LMO	0.8	100	7.6	0
1UY8	0.9	83	5.9	66
1BB5	1.0	77	11.7	33
2Q6R	1.1	77	8.0	44
1EB2	1.2	75	7.4	62
1JYS	1.4	100	3.0	50
1LZG	1.4	80	8.0	60
1O2V	2.0	71	3.6	85
1BJV	2.0	83	5.1	50
1BB6	2.1	63	7.8	63
1K1J	2.4	77	6.0	11
1UYF	2.6	66	6.8	50
1SV9	2.7	75	6.3	75
1D3D	2.8	33	2.6	44
1LZE	3.6	77	11.5	33
1BJU	7.6	42	3.3	57
1MUE	8.1	50	2.5	40
1BB7	10.3	57	7.2	57

[a] We report the RMSD of ligand (LRMSD, aligned on C_α atoms), the fraction of true (fTP) inter-molecular contacts for both MELD and DOCK TOP1 prediction.

Table S2: **Details of complexes and simulation cost**[a]

	# of Atoms		Simulation Time(ns)	Wall-Clock Time(hr)		
PDBID	**Protein**	**Ligand**	**MELD**	**MELD**	**Rigid Receptor**	**Flexible Receptor**
1O2K	3220	45	100	8.56	0.18	0.24
1GI6	3220	33	100	9.78	0.18	0.24
1O2O	3220	45	100	9.44	-	-
1UY7	3295	44	700	68.83	-	-
1O3F	3220	43	100	9.39	-	-
1TNK	3220	24	500	61.39	0.18	0.24
1LZY	1950	57	-	-	-	0.24
1MRK	3846	32	-	-	-	0.24
1GHZ	3220	31	-	-	0.18	0.24
1BYG	3939	62	-	-	0.18	-
1UYC	3295	48	100	10.22	-	-
1MU8	4034	53	600	91.67	0.18	-
1LMO	1956	57	300	33.33	-	-
1UY8	3280	44	800	81.78	-	-
1BB5	2030	84	100	12.83	-	-
2Q6R	3796	46	100	18.39	-	-
1EB2	3220	64	100	12.11	-	-
1JYS	3352	15	100	12.56	0.18	0.24
1LZG	1956	56	100	12.61	-	-
1O2V	3220	42	100	9.72	-	-
1BJV	3220	45	100	10.44	-	-
1BB6	1956	102	200	26	0.18	0.24
1K1J	3220	68	100	12.28	-	-
1UYF	3295	51	1000	98.89	-	-
1SV9	1851	29	100	8.78	-	-
1D3D	4035	75	100	15.44	-	-
1LZE	1957	111	100	13.89	-	-
1BJU	3220	34	-	-	-	-
1MUE	4028	49	-	-	-	-
1BB7	1956	75	-	-	-	-

[a] We reported the number of atoms in the complexes, the MELD simulation time needed to converge to native pose and the actual wall-clock time for MELD x MD, rigid receptor docking and flexible receptor docking using AutoDockFr. For rigid and flexible receptor docking, the average wall-clock time of computing grid and performing docking using a 12 cores Intel(R) Xeon(R) CPU E5-2665 across all docking systems was reported. For MELD, it was the wall-clock time used to converge to the native pose on 30 RTX5000 GPUs. For the failures case, a "-" was given to the wall-clock as well as simulation time of the corresponding method.

Investigating MELD x MD failures

Here we showed the population change of different poses overtime in the lowest 4 temperature replicas for all 30 protein-ligand complexes. In each panel, each line represents the population of a pose, identified by clustering the trajectory, over the time of simulation. The most populated pose is depicted as the solid black line. A converged simulation is identified as the one with a 'plateaued' solid black line. As it is shown in the Figure. S13 , most of the MELD x MD successes have converged with an exception of 1UYF. As for the MELD x MD failures, six cases (1GHZ,1MUE,1MRK,1BYG,1BB7,1BJU) might fail due to poor convergence. The rest case (1LZY) has converged. Therefore it is more likely failed due to force field flaws. We further validated this possible force field failure case (1LZY) using MELD competitive binding simulation[12] where only the information of native pose and the most populated non-native one identified previously were used. The result (Fig. S14) showed that the force field favors the non-native pose than the native one.

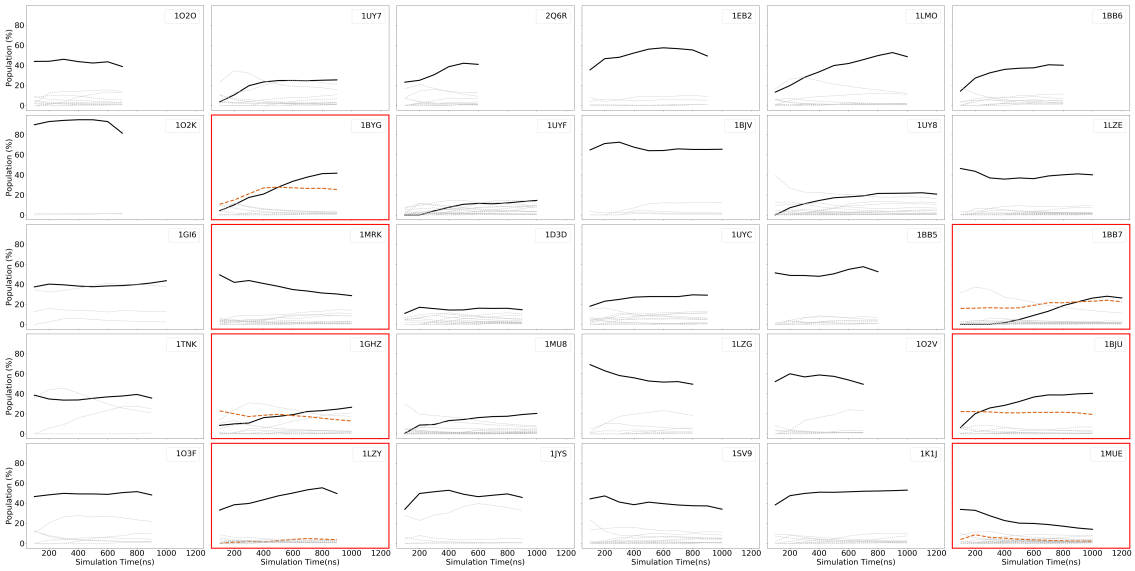

Figure S13: **Poses population change over time in lowest 4 temperature replicas for all 30 protein-ligand complexes.** The black panels are MELDxMD success cases with native pose shown as the solid black line, and the red panels are MELDxMD failures with native pose shown as the orange dashed line. For 1MRK, there is no orange line, since the closest cluster centroid is 3.3Å.

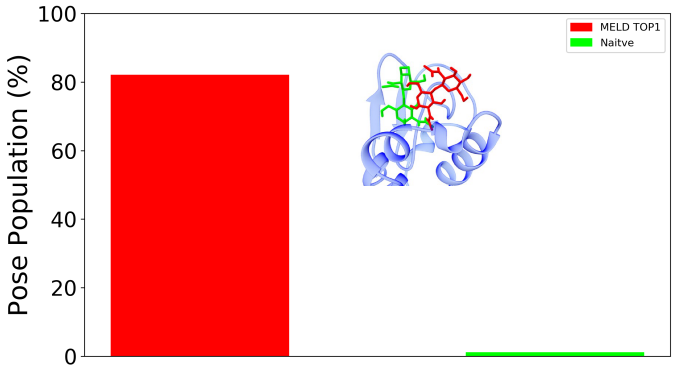

Figure S14: **MELD competitive binding simulation results of 1LZY.** The MELD TOP1 (Red) is sampled most of the time (∼80%) during the MELD relative binding simulation.

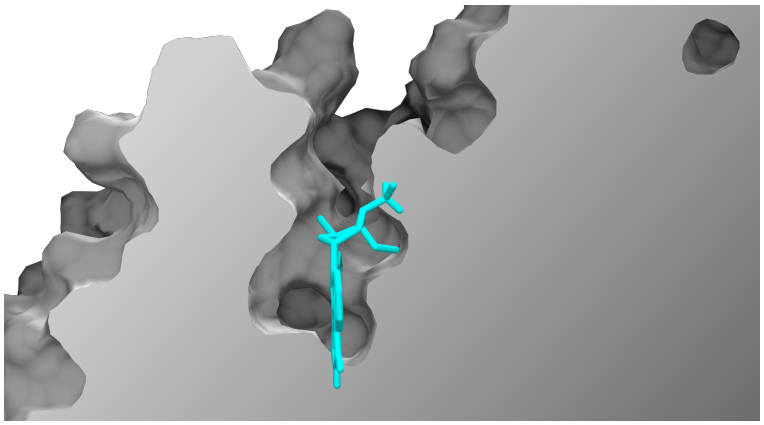

Figure S15: **Deep binding pocket in 1BYG**

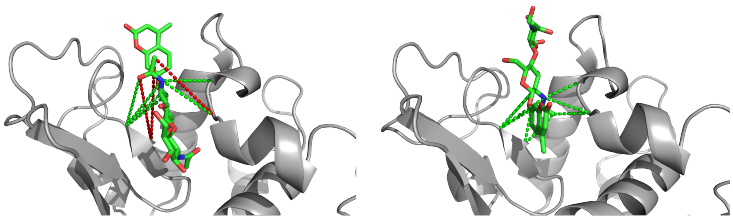

Figure S16: **MELD restraints failed to distinguish MELD TOP1 and native pose of 1BB7.** MELD restraints in the native restraint group are shown for both the MELD TOP1 pose (left panel) and the native pose (right panel). The satisfied and unsatisfied restraints are shown as green and red respectively for both MELD TOP1 pose (6/9 satisfied) and native pose (9/9 satisfied). Since each restraint group will enforce 6 restraints during the simulation, the native restraint group can not distinguish between these two poses.

Table S3: **Rescuing 3/7 original MELD x MD failures**[a]

	Rigid MELD x MD		**Original MELD x MD**	
	LRMSD(Å)	P(native)/P(non-native)	LRMSD(Å)	P(native)/P(non-native)
1GHZ	0.5	3.2	6.7	0.7
1BJU	1.3	3.3	7.6	0.5
1MUE	1.5	1.7	8.1	0.1
1MRK	3.5	-	6.6	-
1LZY	6.0	-	6.2	-
1BYG	7.6	0.3	8.1	0.6
1BB7	9.0	0.2	10.3	0.9

[a]We reported the RMSD of ligand (LRMSD, aligned on C_α atoms) of MELD TOP1 pose, the relative population of most populated native versus most populated non-native poses. A number bigger than 1 indicate a success. The bold PDBIDs represent the systems that we rescued (1GHZ,1BJU,1MUE) or improve to a near native pose (1MRK) by introducing more rigidity during MELD x MD simulation. For system no native-like pose (LRMSD less than 3Å) is predicted at TOP1 level, a "-" is given.

References

(1) Mukherjee, S.; Balius, T. E.; Rizzo, R. C. Docking validation resources: protein family and ligand flexibility experiments. *J. Chem. Inf. Model.* **2010**, *50*, 1986–2000.

(2) Thompson, T. M. *From error-correcting codes through sphere packings to simple groups*; American Mathematical Soc., 1983; Vol. 21.

(3) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: fully automated protein–ligand interaction profiler. *Nucleic Acids Res.* **2015**, *43*, W443–W447.

(4) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(5) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem* **2004**, *25*, 1605–1612.

(6) Zhao, Y.; Stoffler, D.; Sanner, M. Hierarchical and multi-resolution representation of protein flexibility. *Bioinformatics* **2006**, *22*, 2768–2774.

(7) Ravindranath, P. A.; Forli, S.; Goodsell, D. S.; Olson, A. J.; Sanner, M. F. AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility. *PLoS Comput. Biol* **2015**, *11*, e1004586.

(8) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem* **2007**, *28*, 1145–1152.

(9) Spyrakis, F.; Ahmed, M. H.; Bayden, A. S.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. The roles of water in the protein matrix: a largely untapped resource for drug discovery. *J. Med. Chem.* **2017**, *60*, 6781–6827.

(10) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(11) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved generalized born solvent model parameters for protein simulations. *J. Chem* **2013**, *9*, 2020–2034.