SI. Testing Inhomogeneous Solvation Theory in Structure-Based Ligand Discovery

Trent E. Balius,[(a)#] Marcus Fischer,[(a)#‡] Reed M. Stein,[(b)] Thomas B. Adler,[(a)] Crystal N. Nguyen,[(c)] Anthony Cruz,[(d,e)] Michael K. Gilson,[(c)] Tom Kurtzman,[(d,e,f)] and Brian K. Shoichet*[(a)]

a. University of California, San Francisco, Department of Pharmaceutical Chemistry, San Francisco, California, 94158, United States of America

b. University of California, San Francisco, Graduate Program in Pharmaceutical Sciences and Pharmacogenomics, San Francisco, California, 94158, United States of America

c. University of California, San Diego, Skaggs School of Pharmacy and Pharmaceutical Sciences, La Jolla, CA, 92093, United States of America

d. Lehman College Department of Chemistry, 250 Bedford Park Blvd West Bronx New York, 10468, United States of America

e. Ph.D. Program in Chemistry, The Graduate Center of the City University of New York, 365 5th Avenue, New York New York, 10016, United States of America

f. Ph.D. Program in Biochemistry, The Graduate Center of the City University of New York, 365 5th Avenue, New York New York, 10016, United States of America

‡ Present address: Departments of Chemical Biology and Therapeutics, Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, United States.

# contributed equally

* bshoichet@gmail.com -- to whom correspondence should be addressed.

Key words: ordered water, Inhomogeneous Solvation Theory, ligand discovery, molecular docking, structure-based drug design

Supplemental material.

Section S1. Retrospective docking.

Enrichment. We quantified enrichment by calculating the area under the curve (AUC) and the log-adjusted AUC (logAUC) values with respect to the receiver operator characteristic (ROC) curves: ligand and property-matched decoys (PMD) were generated based on actives using the DUD-E method.(1) Enrichment studies were performed on 25+1 systems: CcP-ga consisting of 46 ligands and 3,338 decoys and 25 DUD-E systems (AA2AR, ACES, ADA, AMPC, CXCR4, EGFR, FA10, FABP4, GLCM, HIVPR, HMDH, HS90A, ITAL, KIT, KITH, LCK, NRAM, PARP1, PLK1, PPARA, PTN1, PUR2, SRC, THRB, and TRY1) consisting of 6571 ligands and 397,864 decoys in total. See ref(1) for more details of the DUD-E benchmark set.

Pose reproduction. We post-processed the ligands from our enrichment calculations and compared their poses to the crystallographic conformations. All crystal complexes were aligned into the docking frame using UCSF Chimera. DOCK6.6 was used to calculate the symmetry-corrected root mean square deviation (RMSD) using the Hungarian algorithm.(2, 3) We looked at two measures of pose fidelity: (1) average RMSD; and (2) the percent docking success (# of poses < RMSD threshold / # molecules × 100).

Section S2. GIST grids and how to combine them.

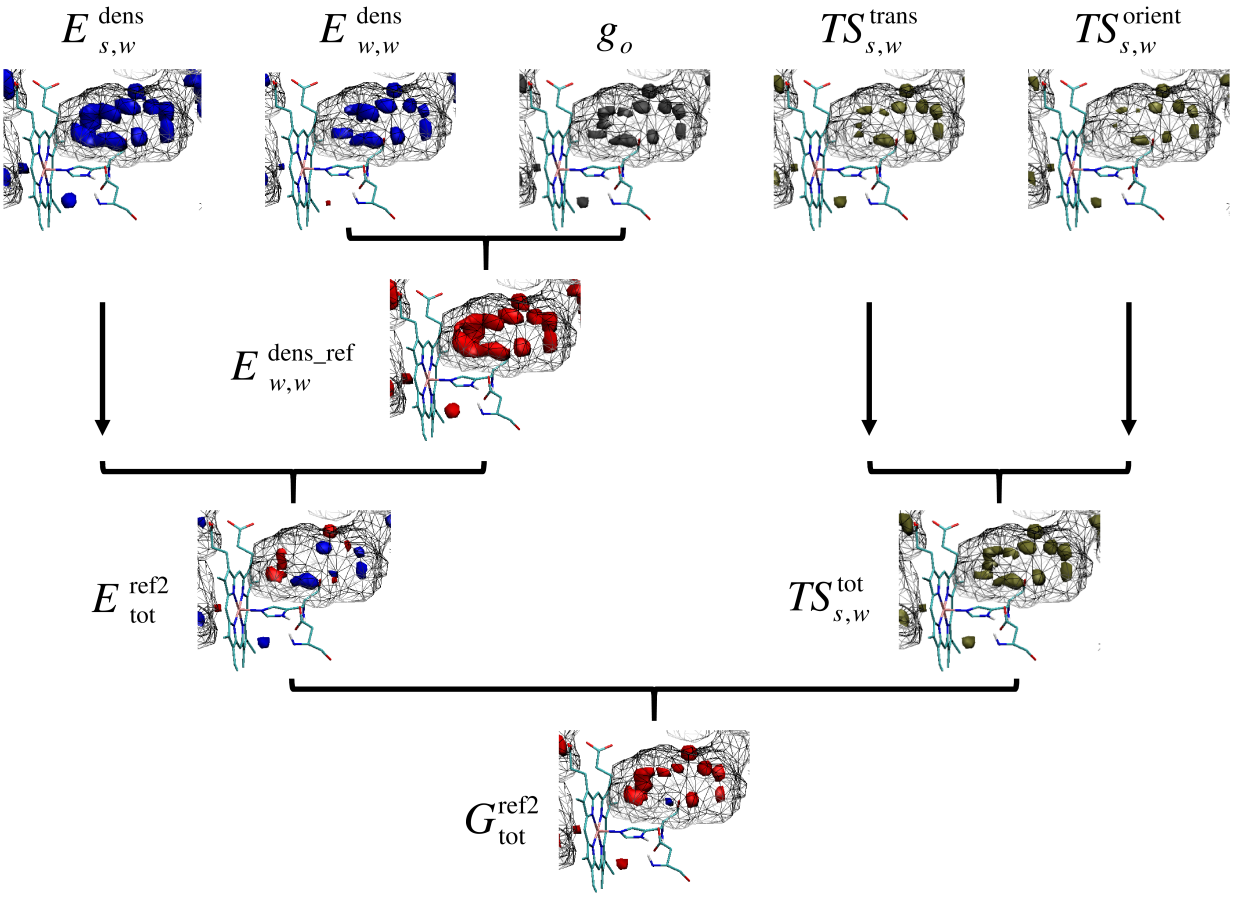
In docking, two tasks are performed: sampling and scoring. In this paper the objective is to improve the scoring aspect by adding a receptor desolvation ($E_{rec,desol}$) term to the DOCK scoring function (eq 1, main document). The receptor desolvation term is estimated by using GIST grids (see **Section S3**). Here, we focus on how to generate GIST grids for use in docking by combining the five GIST components that are output by the Cpptraj program (cf. Ambertools14):

- Enthalpy between solvent (water) and solute (receptor) ($E_{s,w}^{\mathrm{dens}}$);
- Enthalpy of water with water ($E_{w,w}^{\mathrm{dens}}$), also called the two-body term;
- Translational entropy between water and receptor ($TS_{s,w}^{\mathrm{trans}}$);

- Orientational entropy between water and receptor ($TS_{s,w}^{\text{orient}}$);
- Density of water in the context of the receptor (g_o).

The four energy values are in kcal/mol/Å^3. The density is unitless (density/bulk density). The GIST nomenclature has undergone a development over time, particularly whether the enthalpies are to be scaled by one-half, as discussed previously,(4-6) and here. The GIST grids used here are obtained using Amber14 and Ambertools14.(7)

We combine the GIST terms (outlined above) in four physically meaningful ways to be used in docking. There are two issues to explore regarding this new GIST term: (1) the best way to combine the GIST components (discussed here and in **Section S4**); and (2) the best scaling factor to bring the GIST term into balance with the other scoring function terms (discussed in **Sections S3 and S5**).



Schema S1. Illustration of how the GIST grids are combined in this work. For enthalpy and free energy contributions > 0.5 kcal/mol/Å^3, regions are coloured red. For the case < -0.5 kcal/mol/Å^3, the regions appear blue. Tan colored are regions with entropy contributions > 0.5 kcal/mol/Å^3. Regions of water density g_o > 6.0 units (6 times that of bulk) are displayed in grey.

To estimate the free energy difference of water transfer (desolvation), we need to subtract the energy of water in bulk from the energy on the surface of the protein. This is done by referencing the water-water term to bulk (eq S1):

$$E^{\text{dens\_ref}}_{w,w}(i) = E^{\text{dens}}_{w,w}(i) + 0.3184 \times g_o(i) \qquad \text{(Equation S1)}$$

Here, the *i* refers to a grid position, a voxel. The constant was calculated using two parameters (taken from the Amber manual): mean energy, C_{bulk} = - 9.533 kcal/mol/water, and number density,

C_{num_dens} = 0.0334 waters / Å^3. $C_{bulk} \times C_{num\_dens}$ = - 0.3184 kcal/mol/ Å^3.

In this study, we include displacement from all voxels: both high and low occupied sites. In previous IST displacement studies(4) voxels only received a score if the density was above a cutoff. This ignores contributions from low density regions that may have a considerable contribution (see **Section S8**). Also in prior work,(4) the energy normalized to density (eq S2) was used.

$$E_{w,w}^{\text{norm\_ref}}(i) = \frac{E_{w,w}^{\text{dens}}(i)}{0.0334 \times g_o(i)} + 9.533 \qquad \text{(Equation S2)}$$

The normalized value is the average energy per water in the voxel and thus the units of normalized energies ($E_{w,w}^{\text{norm\_ref}}$) are in kcal/mol/water. Although we did consider the normalized grid (preliminary enrichment experiments yielded poor results), we chose to use the referenced grid (eq S1). The units also indicate that the un-normalized grids are more compatible with our scoring function.

The GIST grids may be combined to produce the total enthalpy grid (eq S3) and the total free energy grid (eq S4).

$$E_{tot}^{\text{ref}}(i) = E_{s,w}^{\text{dens}}(i) + E_{w,w}^{\text{dens\_ref}}(i) \qquad \text{(Equation S3)}$$

$$G_{tot}^{\text{ref}}(i) = E_{s,w}^{\text{dens}}(i) + E_{w,w}^{\text{dens\_ref}}(i) - \left(TS_{s,w}^{\text{orient}}(i) + TS_{s,w}^{\text{trans}}(i)\right) \qquad \text{(Equation S4)}$$

In addition, we scaled the water-water term by two (eqs S5 and S6, and **Schema S1**).

$$E_{tot}^{\text{ref2}}(i) = E_{s,w}^{\text{dens}}(i) + 2 \times E_{w,w}^{\text{dens\_ref}}(i) \qquad \text{(Equation S5)}$$

$$G_{tot}^{\text{ref2}}(i) = E_{s,w}^{\text{dens}}(i) + 2 \times E_{w,w}^{\text{dens\_ref}}(i) - \left(TS_{s,w}^{\text{orient}}(i) + TS_{s,w}^{\text{trans}}(i)\right) \qquad \text{(Equation S6)}$$

In-house Python scripts were used to combine grids and are available at docking.org/~tbalius/Code.

In eqs S5-S6, the factor of two results from every water interacting with every other water. Each water involved in the interaction retains half the energy (eq S7).

$$E_k = \frac{1}{2} \sum_{\substack{l \in W \\ l \neq k}} E_{k,l} \qquad \text{(Equation S7)}$$

Here, k and l denote waters and W is the set of all waters. The water-water term in eqs S5 and S6 has the

full interaction energy at every voxel.

Section S3. GIST Displacement Algorithm.

To estimate the cost of desolvating the receptor upon binding, we first identify the voxels displaced by the ligand ($V = \{\mathrm{v}_i \mid \mathrm{v}_i \in \mathrm{ligand}\}$). A voxel is considered to be displaced if it is contained within the van der Waals radius of an atom during the docking calculation. We sum up the energies of those voxels (eq S8) and multiply the sum by the volume of the voxel (vol = 0.125 Å^3) to get a value in kcal/mol.

$$E_{rec,desol} = \alpha \times vol \times \sum_{\mathrm{v}_i \in V} E_{GIST}(\mathrm{v}_i) \qquad \text{(Equation S8)}$$

Here, α is a scaling factor. The algorithm is made available in the source code of the new release of the DOCK3.7 program.

To make estimating the GIST component fast and compatible with DOCK 3.7, some approximations were made. Double counting occurs only rarely when non-connected parts of the molecules overlap (**Figure S1**, right panel). We determined that there was very good agreement between the GIST energies calculated with double-counting during docking and the exact GIST energies calculated by a rescoring procedure (**Figure S1**, left panel).

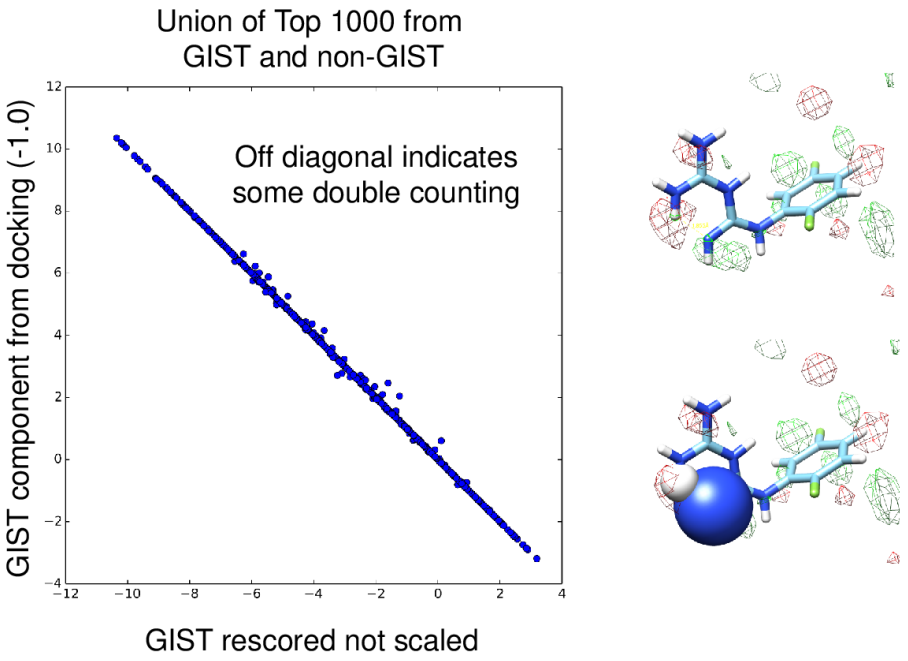


Figure S1. GIST value calculation during docking is a good approximation. The left panel shows a correlation between the top scoring molecules from two screens, where the poses and scores are taken from the virtual docking screen with the GIST term. The GIST component is taken from the screening results (y-axis) and from rescoring the poses. The right panel shows a molecule for which double counting has occurred.

Section S4. Comparison of GIST combinations.

We explored which of the four combinations of the GIST components (discussed above) is best for estimating receptor desolvation during docking. We performed retrospective tests on the four GIST grids, Enthalpy1 (eq S3), Free Energy1 (eq S4), Enthalpy2 (eq S5), and Free Energy2 (eq S6), used to estimate the desolvation component (where $\alpha = 1$ in eq S8).

For each GIST grid we ran ten docking calculations to obtain a mean value and standard deviation. Because DOCK is deterministic, we modified our sampling (by perturbing the spheres used to orient the molecules into the binding site during docking) to obtain different results. Ten runs were used to better gauge the confidence in our results in the same way as performing a wet lab experiment in

triplicate.

Here, the Enthalpy2 (eq S8) performed the best with logAUC of 57.46 (**Figure S2** and **Table S1**) followed by Free Energy1 (eq S7) as the second best with logAUC of 56.08. The Enthalpy2 grids were used for the remainder of this study.

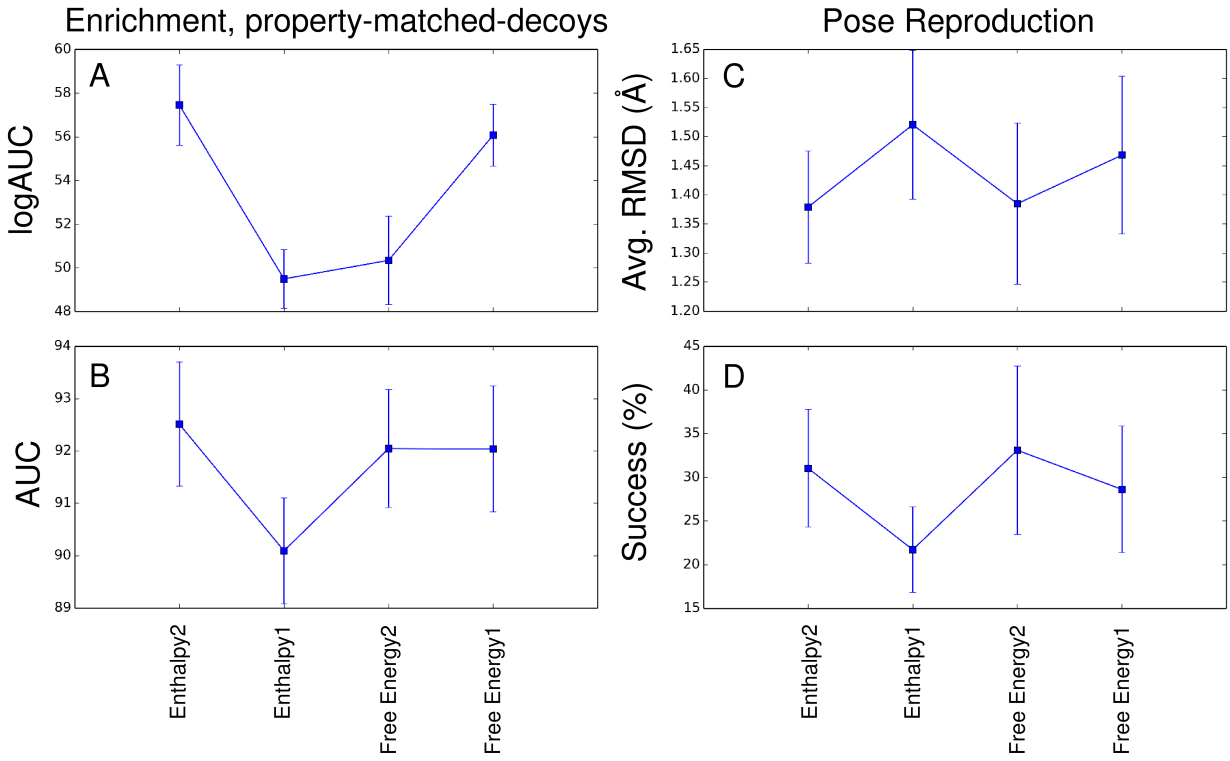


Figure S2. Comparison of GIST combinations. CcP-ga docking enrichment values (panels A and B) and pose reproduction (panels C and D) shown using different combinations of the GIST grids incorporated into the DOCK3.7 scoring function. The error bars are generated by running DOCK3.7 ten times with modified sampling.

Table S1. Comparison of GIST combinations.

	LogAUC		AUC		avg RMSD (Å)		success (%) [a]	
	mean	std	mean	std	mean	std	mean	std
Enthalpy2	57.46	1.84	92.51	1.19	1.38	0.10	31.03	6.72
Enthalpy1	49.50	1.34	90.09	1.01	1.52	0.13	21.72	4.89
Free Energy2	50.35	2.02	92.05	1.13	1.38	0.14	33.10	9.66
Free Energy1	56.08	1.42	92.04	1.20	1.47	0.14	28.62	7.24

[a] Success percent of systems with RMSD less than 1.0 Å

Section S5. Retrospective analysis for CcP-ga.

Next, we explored what the best scaling factor (α in eq S8) is for weighting the receptor desolvation term in the DOCK3.7 scoring function (main text eq 1). All other terms in eq 1 (besides $E_{rec,desol}$) have scaling factors of one.

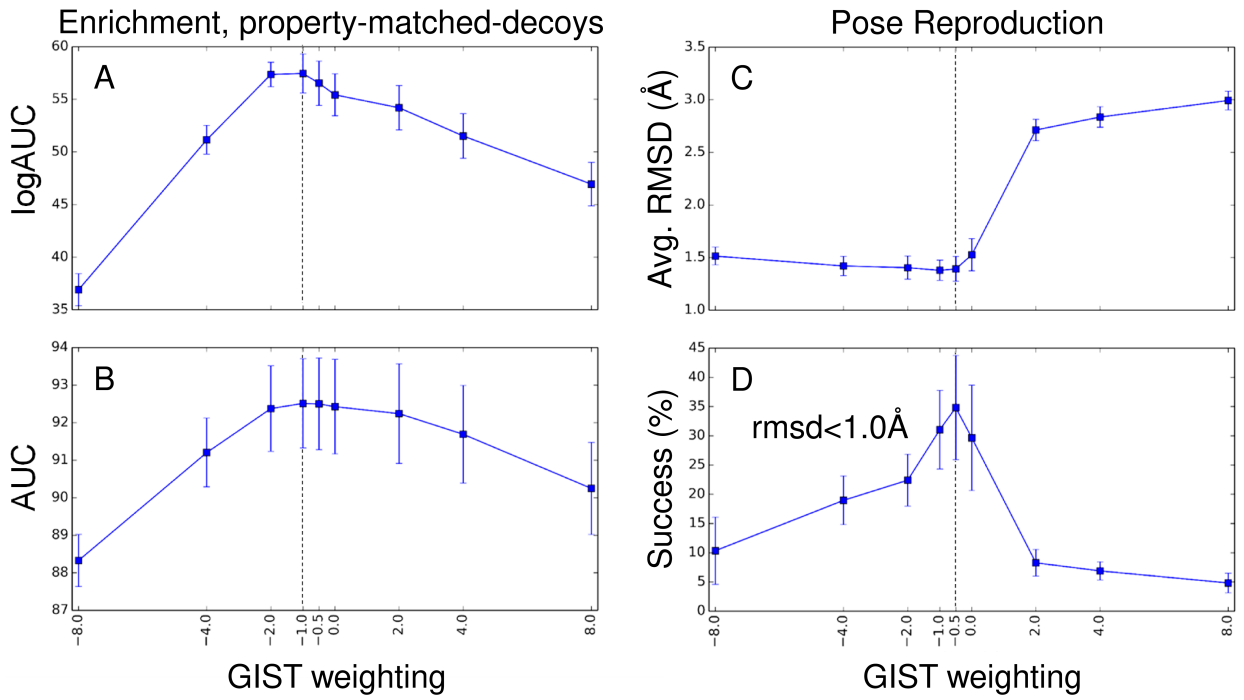


Figure S3. Examination of weighting factors for the receptor desolvation (GIST) term in the

DOCK3.7 scoring function. Retrospective analysis of CcP-ga is shown. (A, B) Enrichment analysis. Panel (A) shows logAUC. Panel (B) shows the AUC. (C, D) Pose reproduction analysis. Panel (C) shows RMSD averaged over all ligands. Panel (D) shows the success rate (number of ligand with RMSD <1.0 Å). The blue squares represent the mean of 10 docking runs and the error bars show the standard deviation indicating the variance in distribution of values.

Table S2. CcP-ga retrospective analysis for changing weight of GIST component.

GIST scale (α)	**logAUC**		**AUC**		**avg RMSD (Å)**		**success (%)** [a]	
	mean	std	mean	std	mean	std	mean	std
-8.0	36.91	1.52	88.33	0.69	1.51	0.08	10.34	5.77
-4.0	51.16	1.38	91.20	0.91	1.42	0.09	18.97	4.15
-2.0	57.36	1.16	92.38	1.14	1.40	0.11	22.41	4.43
-1.0 (full GIST)	57.46	1.84	92.51	1.19	1.38	0.10	31.03	6.72
-0.5	56.54	2.10	92.50	1.22	1.39	0.12	34.83	8.92
0.0 (non-GIST)	55.43	2.00	92.43	1.26	1.53	0.15	29.66	9.02
2.0	54.20	2.11	92.24	1.33	2.71	0.10	8.28	2.29
4.0	51.52	2.12	91.69	1.30	2.84	0.10	6.90	1.54
8.0	46.94	2.07	90.25	1.23	2.99	0.09	4.83	1.69

[a] Success percent of systems with RMSD less than 1.0 Å

Section S6. GIST convergence analysis.

To gauge if we ran the simulations long enough, the full simulation was divided into ten 5ns sub-trajectories and GIST grids were generated for each for comparison. First, we calculated the second-norm between pairs of GIST grids to quantify how similar the corresponding voxels are to one another between two grids; second, we docked to the different GIST grids (as the receptor desolvation component of the scoring function in eq 1) and quantified the variability in enrichment (logAUC).

Sub-trajectory GIST grids were compared to the full simulation GIST grid (**Figure S4**, top panel), and to neighboring sub-trajectory GIST grids (**Figure S4**, bottom). The oscillating behavior in both curves (**Figure S4**) indicates convergence.

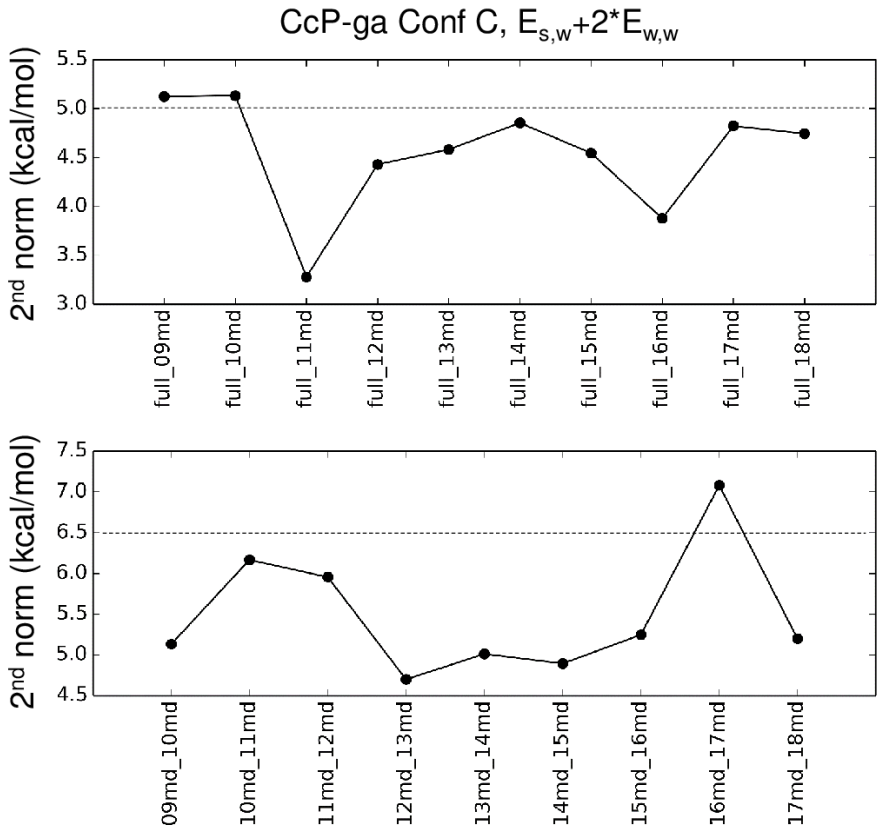


Figure S4. Comparison of GIST grids from sub-trajectories to gauge convergence. The combined GIST grid of solute-water enthalpy and water-water enthalpy scaled by two are evaluated here. Top, each sub-trajectory is compared to the full simulation. Bottom, each sub-trajectory is compared to its immediate neighbors.

We examined the variance of docking performance when using the sub-trajectory GIST grids (0.19 logAUC units, **Table S3**). As a control, we looked at the variance by modifying the sampling (1.84 logAUC units, **Table S3**). When compared to the modified sampling, the sub-trajectory docking varied little (9.6 times less). These data show that docking with the GIST grids of the 5 ns long simulations gave very similar docking results as the full 50 ns simulation (differing at most by 0.36 logAUC units).

Table S3. Spread of retrospective docking enrichments with modified sampling and GIST grids. logAUC subtraj analysis (convergences).

Trajectory	Spheres	mean	std	max	min	diff
Sub [a]	original	58.51	0.19	58.76	58.12	0.64
Full [b]	original	58.40	--	--	--	--
Full	modified [c]	57.46	1.84	62.24	55.16	7.08

[a] 10 GIST grids generated from 5 ns sub-trajectory;

[b] One GIST grid from the 50 ns trajectory;

[c] 10 perturbed spheres

Section S7. Retrospective analysis for 25 DUD-E systems.

These data are also discussed in the result section of the main text.

When comparing GIST to no-GIST results across the 25 DUD-E systems, GIST performs worse (average logAUC difference is -1.33, Table S4), unlike CcP-ga which performs best with a weighting of -1.0. However, when we lower the weighting of GIST component to -0.5 the results got slightly better than the no-GIST enrichments (avg. ΔlogAUC = 0.28, Table S4). When examining the GIST grids, we observed extrema of very high energies at specific voxels. For example, ADA had the most extreme voxel of any system with a value of -119.73 kcal/mol/Å^3 that if displaced would penalize the score by +14.97 kcal/mol. Such a large penalty seems to be unreasonable in the contexed of our scoring. Thus, we truncate these peaks to ±3.0 kcal/mol/Å^3 (which remains a high value, 5 to 19 fold higher than the standard deviation (σ) of the 210,000 voxels in the grid). This truncation impacts only 0.03% of the voxels, ranging from 17 to 88 for the favorable water voxels and 0 to 10 for unfavorable voxels. When truncation of extrema is combined with a weighting of -0.5 there is an additional improvement of GIST compared with no-GIST (avg. ΔlogAUC = 0.53, Table S4, Figure S5). AA2AR and AMPC both change classification from same to better when truncated grids are used, FXA likewise shifts but this is due to very slight change in logAUC. We believe that the extrema are artificially high due to the following: (1) The simulations are run with the protein's heavy atoms strongly restrained (5 kcal/mol/Å^2). Since waters interact with the restrained atoms, their densities and energies are more concentrated than if the residue/atoms could move. The waters that are interacting with a moving atom would also move smearing the water's densities and energies across more voxels. (2) Entropy is neglected and the positions that have the highest energies are also those position where the waters are most frozen, so there is likely an entropic cost to having the water there.

Table S4. DUD-E evaluation of GIST contribution on enrichment calculations.

Analysis of different weighting factors on enrichments.[a]				
	better	Same	worse	avg. ΔlogAUC
weight: -0.5	10	9	6	0.28
weight: -1.0	8	5	12	-1.33
weight: -2.0	5	4	16	-6.55
weight: -0.5, truncate 3.0	13	6	6	0.53
Weight: -1.0, truncate 3.0	11	3	11	-0.39
[a] Each row sums to the 25 systems.				

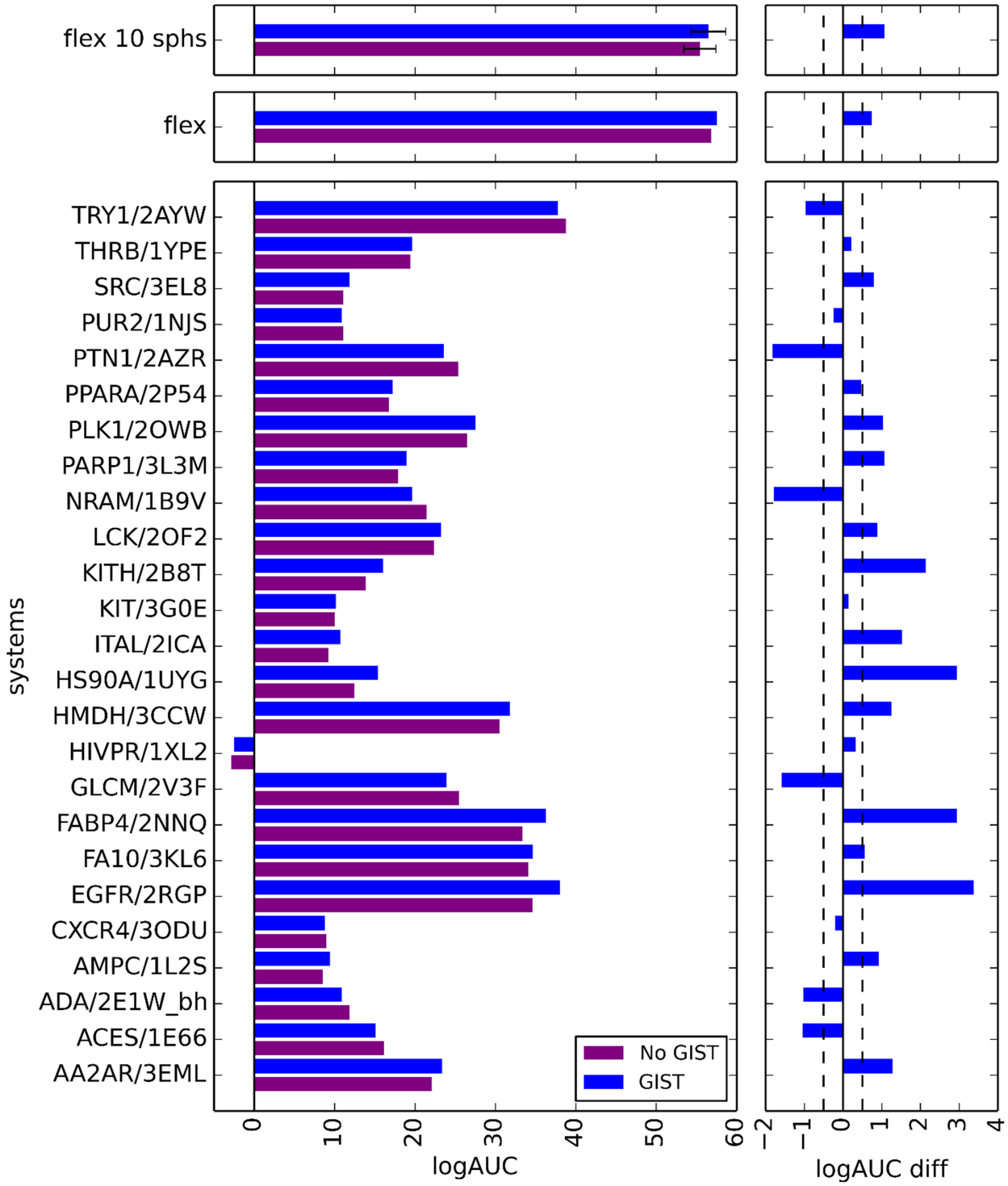


Figure S5. Enrichment analysis of CcP-ga and 25 DUD-E systems. Bar graphs of logAUC values for six docking types are shown: non-GIST in purple and GIST in blue (with the GIST component weighted

by -0.5 and the GIST grids truncated at 3.0 kcal/mol/Å^3 results). The bottom panels show the total enrichment values for No-GIST and GIST, while the top two panels show the difference (GIST - non-GIST). CcP results are shown for 10 perturbed results (error bars show standard deviation as an indication of the distribution of the results) and for the original sphere set. ADA was prepared by hand. All other systems were prepared with an automated procedure.

Section S8. Binding site analysis.

We examine the CcP-ga closed binding site to understand the nature of solvent in the site. In **Figure S6** the enthalpy with water-water term scaled by two (Enthalpy2, eq S5) is shown. The regions of unfavorable energy for waters (>1.0 kcal/mol/Å^3) are shown in red, which are favorable to displace according to the GIST scoring function. The favorable regions for water (>-1.0 kcal/mol/Å^3) are shown in blue, which are unfavorable to displace according the GIST scoring function. The favorable site (s1) proximal to Asp233, is the most favorable water location in the site. The region closest to the heme has two unfavorable water locations (s2 and s3) (**Figure S6A**). There is also an unfavorable location (s4) proximal to Gly178. Finally, there is a region close to the cavity entrance that encompasses three additional favorable water locations (s5, s6, and s7). Decreasing the cutoff value to 0.01 kcal/mol/Å^3 reveals the irregular shapes of the hydration sites (**Figure S6B**). Note that the majority of the solvation energy is concentrated at these seven sites. However, just accounting for the most intense sites (as WaterMap does) will neglect the lower magnitude regions, which do add up (-1.47, and +2.42, **Table S5**) and contribute to the score.

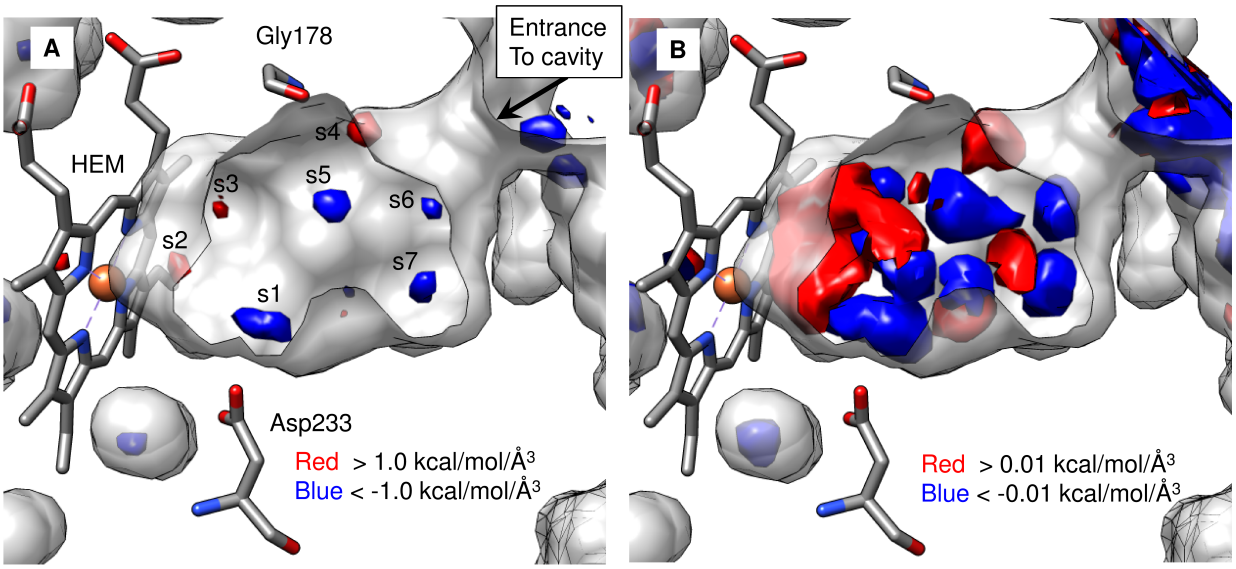


Figure S6. Hydration of CcP-ga with the GIST enthalpy grid. A. Here, GIST enthalpy grids with a cutoff of 1.0 kcal/mol/Å^3 are shown. The only opening to the closed cavity is indicated by an arrow. Seven hydration sites are indicated, s1 though s7. **B.** The cutoff value is decreased to 0.01.

Table S5. Site energetics of subregions.[a]

Subsite name	Energies (kcal/mol)
s1	-4.27
s2	2.58
s3	1.63
s4	1.67
s5	-2.36
s6	-2.20
s7	-1.22
Sum positive	5.88
Sum negative	-10.05
Whole site positive	8.30
Whole site negative	-11.52
Total	-3.22

Remainder positive	-1.47
Remainder negative	2.42
[a] Sites are spheres with a radius of 1.4Å located at the centers of intensities of the energies.	

Section S9. Prospective testing.

The behavior of the 17 tested molecules (cf. **Table S6**) is presented in the following, including ranks and energies. Ligand occupancies are presented in **Table S7**; for compound **14**, MES was not completely removed from the binding site and its partial occupancy is shown in **Figure S7**. Ligand efficacy is determined from the affinity (**Figure S8)** and ranges from -1.0 to -0.22 (**Table S6**). The ligands that make water-mediated interactions with Asp233 on average bind more weakly than the molecules that bind with a direct electrostatic interaction (**Table S8**).

From among those molecules substantially changing rank or pose due to including GIST, 17 were purchased for experimental testing. Compounds **3** to **14** were acquired and tested because their ranks improved with GIST, while compounds **15** to **17** were acquired and tested because of better ranks without the GIST term (**Table 1**). Molecules that ranked higher by GIST scored more favorably than without GIST by up to -1.8 kcal/mol, but could also be more unfavorable by as much as +2.0 kcal/mol out of a total docking score that ranged from -42.8 to -35.4 kcal/mol among the top-scoring 1000 molecules of VS1 (**Table S6**). The observation that GIST can improve ranks while reducing scores reflects its global effects on other high-ranking molecules that were affected more substantially still, emphasizing the role of decoy molecules in docking. For molecules whose rank was substantially better without GIST, the GIST term ranged from 8.1 to 8.7 kcal/mol (unfavorable), showing that GIST strongly disfavored these otherwise high-ranking molecules. We also looked for molecules where a substantial pose change occurred between the two scoring functions (e.g. compounds **1** and **2**, **Table 1**). Finally, we considered

implicit water-mediated interactions to be favorable regions in the GIST grid within hydrogen-bonding distance to ligand and protein, though no explicit water molecules were used. This occurred with compounds **3**, **4**, **5**, and **6 (Table 1**). We now consider the 14 molecules prioritized by including GIST (pro-GIST), and then turn to those 3 prioritized by excluding GIST (anti-GIST).

Intriguingly, GIST penalties on these deprioritized molecules, at around +8 kcal/mol, had a much stronger impact on reducing their ranks than favorable GIST energies had on improving them; as with most scoring terms in docking, deprioritizing decoys is as or even more important than highly scoring what turns out to be true ligands.

Table S6. Detailed properties of selected molecules.

Cmpd #	Name	Rank1 GIST	Rank2 Non-GIST	Δlogrank	RMSD [a]	tc_lig	tc_knb	gist1	Energy1	Energy2	Screen	k_d (μM) [b]	Ligand Efficiency (kcal/mol/atom) [c]	Notes [d]
3	ZINC000004705523	13	249	1.28	0	0.24	0.22	-1.67	-42.78	-41.11	1	3472±172	-0.28	WM
6	ZINC000019439634	91	355	0.59	0	0.27	0.16	0.86	-39.63	-40.5	1	3435±860	-0.28	WM
9	ZINC000020357620	98	745	0.88	0	0.36	0.21	-0.65	-39.56	-38.91	1	522±21	-0.41	
4	ZINC000006869116	112	464	0.62	0	0.29	0.17	0.6	-39.36	-39.96	1	809.7±99	-0.38	WM
12	ZINC000002389932	118	645	0.74	0	0.57	0.545	-0.02	-39.27	-39.25	1	619±63	-0.49	
13	ZINC000039212696	147	1462	1	0	0.32	0.17	-1.82	-38.98	-37.16	1	n.d.		
11	ZINC000000161834	358	1212	0.53	0	0.44	0.42	0.28	-37.39	-37.68	1	1.3±0.03	-1.00	
1	ZINC000002564381	490	180	-0.43	3.21	0.54	0.21	1.46	-36.74	-41.77	1	n.d.		F
8	ZINC000042684308	601	1916	0.5	0	0.21	0.19	0.04	-36.23	-36.26	1	1962±554	-0.41	
--	ZINC000095079390	615	2612	0.63	0	0.25	0.18	-0.96	-36.16	-35.2	1	n.a.		
2	ZINC000006557114	664	740	0.05	3.17	0.59	0.33	2.03	-35.94	-38.94	1	154±19	-0.43	F
5	ZINC000006855945	869	2550	0.47	0	0.24	0.21	-0.07	-35.35	-35.28	1	1606±287	-0.35	WM
7	ZINC000001827502	5	19	0.58	0	0.36	0.21	2.12	-46.33	-48.44	2	113.7±20.05	-0.41	
14	ZINC000000112552	747	4380	0.77	0	0.44	0.19	0.01	-40.06	-40.07	2	29.6±2.5	-0.62	
10	ZINC000074543029	1128	4923	0.64	0	0.24	0.2	0.46	-39.35	-39.8	2	~712±231	-0.36	
17	ZINC000022200625	6000	577	-1.02	0	0.26	0.19	8.09	-35.84	-43.92	2	n.d.		AG
15	ZINC000002534163	9487	906	-1.02	0	0.3	0.23	8.56	-34.66	-43.22	2	NB		AG
16	ZINC000000156254	14828	1657	-0.95	0	0.26	0.17	8.7	-33.39	-42.09	2	5464±2694 (NB)	-0.22	AG

[a] RMSD uses the Hungarian algorithm

[b] **n.a.,** not available - molecule not in assayable form. **n.d**., not determinable - compound interference with absorbance peaks. **NB**, non-binder. "~", assay interference of compound **10** before saturation was reached.

[c] $\Delta G = -RT\ln(Kd)$ and Ligand Efficiency calculated using heavy atoms count.

[d] WM = water mediated, F = poses changes, AG = Anti-GIST (the molecule or its pose was preferred in the screen without GIST)

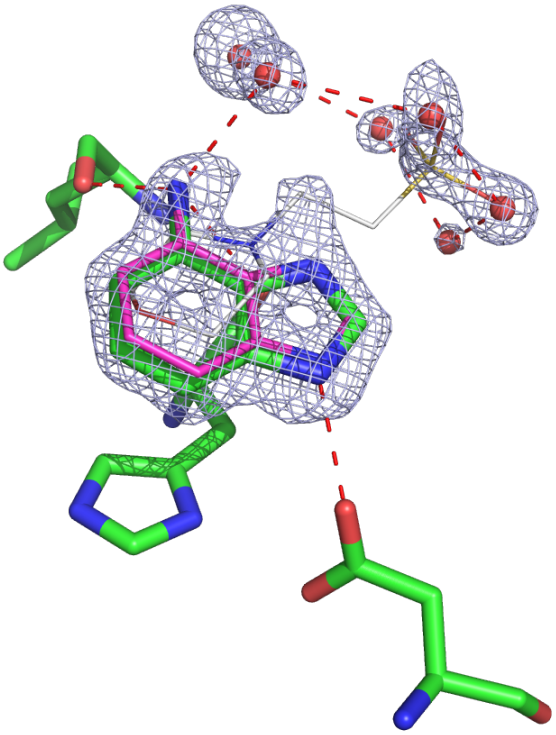


Figure S7. Compound **14** refined to 73% in the presence of 26% MES from the crystallization buffer

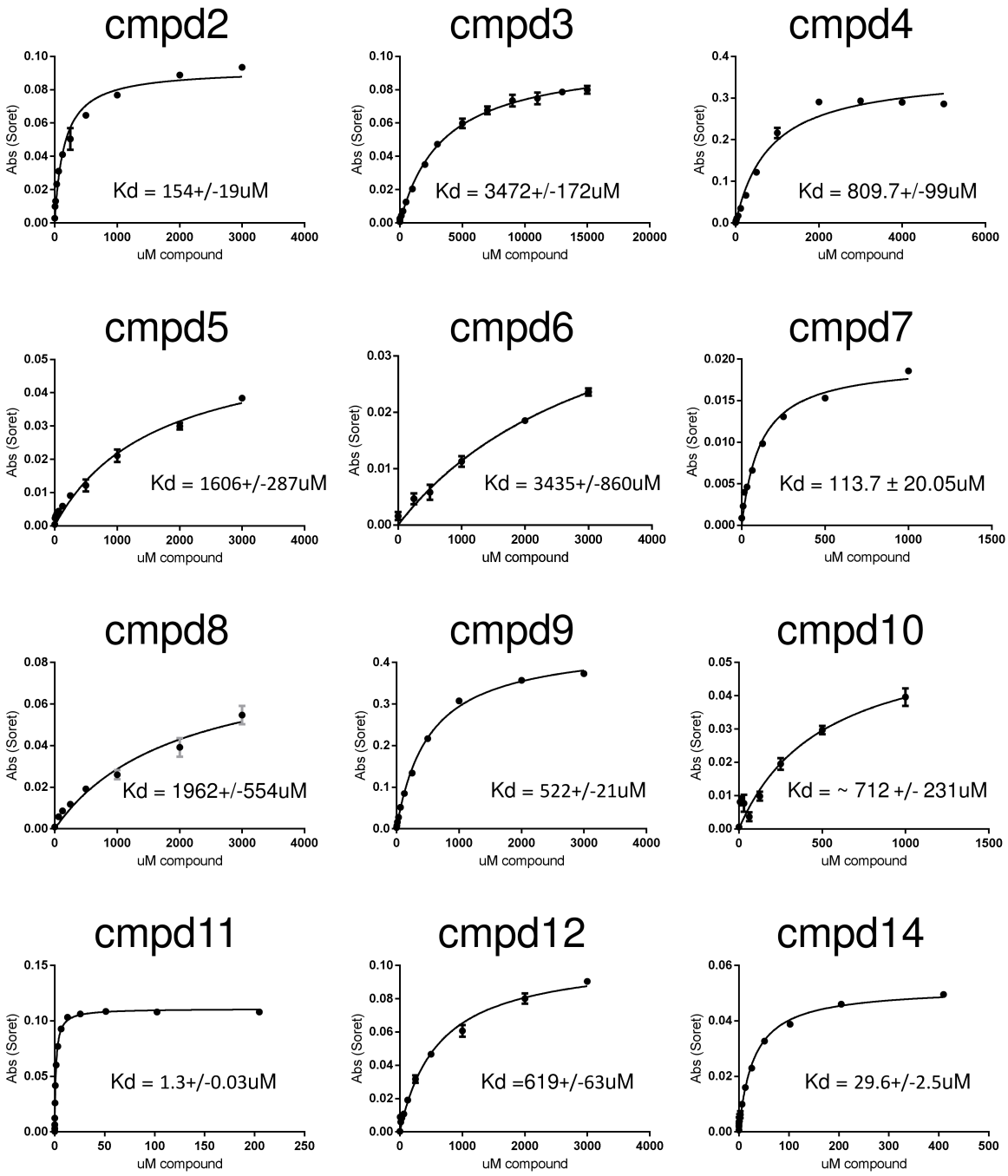


Figure S8. Ligand binding curves. The Soret band shift is shown as a function of ligand concentration (μM).

Table S7. Ligand occupancies after automatic refinement in Phenix.

Cmpd #	**Ligand occupancy**
1	0.88
2	0.90 (one conformation modeled)
3	0.93
8	0.92
9	0.90
10	0.92
11	0.87
12	0.93
14	0.73 (+MES @ 0.26)

While the occupancy for the major pose of **2** refined to 90%, the alternative pose would sterically clash with a nearby protein loop that has insufficient electron density to allow explicit modeling of alternative conformations.

Table S8. Comparison of affinities for compounds that interact via a water molecule versus direct interaction with the anionic D233 residue. [a]

WM			**NonWM**	
Cmpd #	***Affinity*** *(μM)*		***Cmpd #***	***Affinity*** *(μM)*
1	n.d.		**2**	154
3	3472		7	114
4	810		**8**	1962
5	1606		**11**	1
6	3435		**12**	619
9	522		**14**	30
10	712			
average	1759.5		*average*	480
median	**1208**		*median*	**134**

[a] Bold compound numbers indicate that a crystal

structure is available beyond the docking prediction.

Section S10. Timings.

The GIST-scoring algorithm is more time- and memory-intensive than trilinear interpolation, which is used in the other scoring components. To determine how GIST affects the speed of docking calculations, we ran one set of ligands from each system ten times on the same, dedicated machine (**Table S9**). This results in a 1.5 to 16.4 times (on average six-fold) slowdown in runtime. However, we anticipate that using good GIST approximations will result in no slowdown and little impact on docking quality.

Table S9. DOCK3.7 run time slowdown with GIST referenced to non-GIST.

PDB code	DUD-E name	Avg number of heavy atoms	Slowdown [a]
1B9V	NRAM	25.34	3.87
1E66	ACES	29.48	2.21
1L2S	AMPC	20.19	5.21
1NJS	PUR2	33.29	7.30
1UYG	HS90A	27.95	4.67
1XL2	HIVPR	41.06	4.37
1YPE	THRB	34.88	7.97
2AYW	TRY1	33.66	16.40
2AZR	PTN1	39.93	12.97
2B8T	KITH	30.24	3.14
2E1W	ADA	24.77	3.78
2ICA	ITAL	36.38	13.27
2NNQ	FABP4	30.30	4.27
2OF2	LCK	34.70	9.34
2OWB	PLK1	33.08	6.76
2P54	PPARA	32.18	2.92

2RGP	EGFR	31.39	4.45
2V3F	GLCM	27.26	1.15
3CCW	HMDH	36.66	4.05
3EL8	SRC	34.62	4.43
3EML	AA2AR	31.97	2.65
3G0E	KIT	38.77	2.44
3KL6	FA10	33.52	9.94
3L3M	PARP1	30.30	5.34
3ODU	CXCR4	26.67	5.16
	CcP-ga	12.01	1.51
Average		31.18	5.75
[a] Slowdown = (timing from GIST docking) / (timings from non-GIST docking)			

Section S11. Supplemental Methods.

Experimental affinities and structures. The protein was purified and crystallized as described.(8) To reach high ligand occupancies, crystals were transferred into increasing ligand concentrations up to 100 mM (compound solubility permitting) and soaked for several minutes in each drop containing 25% 2-Methyl-2,4-pentanediol (MPD) as a cryoprotectant.

Diffraction images of flash-frozen crystals were collected at beamline 8.3.1. at the Advanced Light Source, Berkeley CA, and processed automatically with the Xia2 pipeline.(9) Initial phases were obtained by Phaser molecular replacement(10) using a model structure lacking several flexible residues and the loop region (residues 186-194). To avoid bias these regions were also excluded from early rounds of refinement using phenix.refine.(11) The ligand and binding site water molecules were only added in the final stage of crystallographic refinement and their occupancies were set to a value below 1 to automatically refine to their final values via phenix.refine (**Table S10**) without manual intervention. Ligand restraint dictionaries were generated from SMILES strings via phenix.elbow,(12) using either automatic or CSD-Mogul geometry optimization. Composite 2mFo-DFc OMIT maps(13) excluding the

ligand fraction were calculated using phenix.composite_omit_map and converted to 2mFo-DFc FFT maps in ccp4 format in order to generate figures using PyMOL.(14)

Crystallographic models were tested with phenix, Coot(15) and the PDB validation tool(16) before depositing the protein-ligand complexes at the PDB as 5U60 (**1**), 5U5W (**2**), 5U5Z (**3**), 5U61 (**8**), 5U5Y (**9**), 5UG2 (**10**), 5U5X (**11**), 5U5U (**12**), 5U5V (**14**) **(Table S10**).

Experimental affinities were measured at least in duplicate by monitoring the shift of the heme Soret band on ligand binding and plotted using a one-site binding least squares fitting method (GraphPad Prism 6.03(17)).

Table S10. Crystallographic statistics.

PDB code	5U60	5U5W	5U5Z	5U61	5U5Y	5UG2	5U5X	5U5U	5U5V
Compound#	1	2	3	8	9	10	11	12	14
Wavelength	1.1158	1.1158	1.1158	1.1158	1.1158	1.1158	1.1158	1.1158	1.1158
Resolution range	45.81 - 1.5 (1.554 - 1.5)	53.52 - 1.29 (1.336 - 1.29)	46.04 - 1.26 (1.305 - 1.26)	53.42 - 1.222 (1.266 - 1.222)	43.47 - 1.3 (1.346 - 1.3)	59.74 - 1.34 (1.388 - 1.34)	46.05 - 1.55 (1.605 - 1.55)	46.04 - 1.33 (1.378 - 1.33)	46.08 - 1.222 (1.266 - 1.222)
Space group	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	50.8941 73.372 105.123 90 90 90	51.0492 75.0215 107.049 90 90 90	51.2511 73.2386 104.829 90 90 90	51.0321 74.782 106.831 90 90 90	51.2595 74.6629 106.937 90 90 90	51.309 72.8483 104.41 90 90 90	51.0058 74.9346 107.047 90 90 90	51.0072 74.9196 106.941 90 90 90	51.0884 74.9366 106.764 90 90 90
Total reflections	740999 (41375)	1083127 (31765)	1022166 (19420)	1084592 (9315)	1052981 (33957)	965518 (42080)	580786 (14686)	1032044 (43192)	1084769 (9132)
Unique reflections	63621 (6113)	99797 (7275)	90443 (3425)	87495 (1446)	93027 (5497)	87217 (7599)	56073 (3567)	90240 (6497)	103584 (2653)
Multiplicity	11.6 (6.8)	10.9 (4.4)	11.3 (5.7)	12.4 (6.4)	11.3 (6.2)	11.1 (5.5)	10.4 (4.1)	11.4 (6.6)	10.5 (3.4)
Completeness (%)	0.99 (0.98)	0.95 (0.71)	0.84 (0.33)	0.72 (0.12)	0.91 (0.55)	0.97 (0.87)	0.92 (0.60)	0.94 (0.69)	0.85 (0.22)
Mean I/sigma(I)	13.63 (0.57)	13.43 (0.70)	14.11 (0.85)	28.35 (3.30)	14.44 (0.60)	14.03 (0.47)	10.60 (0.57)	13.17 (0.67)	17.30 (0.96)
Wilson B-factor	20.93	13.39	14.06	11.53	13.5	16.2	20.85	15.06	11.82
R-merge	0.1043 (3.688)	0.116 (2.812)	0.07535 (1.697)	0.04183 (0.3266)	0.09983 (2.758)	0.09216 (3.552)	0.124 (2.452)	0.1087 (3.791)	0.05207 (0.4428)
R-meas	0.1091 (4.001)	0.1218 (3.2)	0.0788 (1.866)	0.04363 (0.3536)	0.1046 (3.012)	0.09658 (3.93)	0.1303 (2.809)	0.1138 (4.117)	0.05449 (0.5199)
CC1/2	0.999 (0.288)	0.998 (0.192)	0.999 (0.513)	1 (0.938)	0.999 (0.494)	0.999 (0.39)	0.998 (0.305)	0.999 (0.41)	0.999 (0.882)
CC*	1 (0.669)	0.999 (0.567)	1 (0.824)	1 (0.984)	1 (0.813)	1 (0.749)	1 (0.683)	1 (0.763)	1 (0.968)
Reflections used in refinement	63621 (5754)	99808 (6979)	90444 (3381)	87495 (1444)	92342 (5184)	87216 (6629)	56078 (3487)	90242 (6337)	103584 (2653)
Reflections used for R-free	3102 (284)	5010 (339)	4552 (164)	4371 (79)	4651 (257)	4320 (318)	2801 (158)	4424 (310)	5162 (126)
R-work	0.1537	0.1439 (0.3639)	0.1408	0.1204	0.1641	0.1542	0.1588	0.1720	0.1366 (0.2268)

	(0.3624)		(0.2815)	(0.1547)	(0.3734)	(0.3830)	(0.3149)	(0.3398)	
R-free	0.1904 (0.4373)	0.1729 (0.3800)	0.1664 (0.3304)	0.1408 (0.1818)	0.1932 (0.3793)	0.1743 (0.4293)	0.1968 (0.3770)	0.1923 (0.3571)	0.1564 (0.2534)
CC(work)	0.975 (0.713)	0.969 (0.597)	0.976 (0.773)	0.976 (0.960)	0.973 (0.743)	0.976 (0.741)	0.976 (0.766)	0.975 (0.727)	0.972 (0.937)
CC(free)	0.970 (0.697)	0.968 (0.618)	0.968 (0.748)	0.962 (0.971)	0.965 (0.728)	0.970 (0.738)	0.960 (0.789)	0.962 (0.695)	0.963 (0.867)
Number of non-hydrogen atoms	2730	2850	2763	2869	2815	2816	2705	2766	2960
Macromolecules	2449	2479	2415	2477	2430	2364	2400	2381	2543
Ligands	12	12	12	9	11	12	8	9	10
Protein residues	290	290	290	290	290	290	290	290	290
RMS(bonds)	0.017	0.013	0.012	0.008	0.013	0.007	0.014	0.009	0.011
RMS(angles)	1.42	1.38	1.35	1.32	1.37	1.07	1.22	1.16	1.34
Ramachandran favored (%)	99	99	99	99	99	99	100	99	99
Ramachandran allowed (%)	0.99	0.98	0.67	0.66	0.67	0.69	0.34	1.4	0.63
Ramachandran outliers (%)	0	0	0	0	0	0	0	0	0
Rotamer outliers (%)	0.39	0.38	0.39	0	0	0.8	0	0	0.38
Clashscore	1.03	0.2	0.83	0.61	1.66	0.85	0.42	0.63	1.18
Average B-factor	26.27	18.62	18.97	15.75	19.54	22.63	25.06	19.54	16.5
Macromolecules	25.61	17.49	17.99	14.47	18.3	20.68	24.22	18.24	15.24
Ligands	22.35	19.08	23.38	17.05	16.6	18.41	19.54	13.67	15.69
Solvent	34.82	28.32	27.72	25.72	29.47	35.28	33.99	29.64	26.32

Preparing the receptor for MD. The protein preparation is described in the main text's method section, but further details are explained here. The proteins were assigned FF12SB (CcP-ga protein) or FF14SB (all DUD-E proteins) force field parameters. At the time CcP-ga simulations were run, the FF14SB parameters were not yet released. The proteins were placed in a box of TIP3P waters such that every atom of protein was 10 Å from the boundary of the box. The number of waters are presented in **Table S11**. For CcP-ga (4NVA, the apo structure), ten crystallographic waters were retained for the simulation. No crystallographic waters were retained for the simulations of the DUD-E systems. For CcP-ga, use of these crystallographic waters alters the GIST grids, particularly for occluded water locations. Some cofactors and structural ions were kept and disulfide bonds were defined (**Table S11**). Tutorials, which describe (1) running Molecular dynamics for GIST grid generation; and (2) docking with GIST grids, are available at [http://wiki.docking.org/index.php/DOCK_3.7_with_GIST_tutorials].

For CcP-ga, the heme force field was downloaded from the web.(18) The heme parameters were originally prepared for hemoglobin and myoglobin, and thus needed to be adapted for Cytochrome c Peroxidases. The heme parameters were modified by adding a positive charge to the iron (iron Fe III has a 1.25 charge).(19, 20) Amber preparation (prep and frcmod) files for the heme are available at docking.org/~tbalius/Code.

Table S11. CcP-ga and DUD-E number of residues, waters, cofactors and ions in the simulations

Protein name	PDB code	Residues	Waters	Atoms	Ions / cofactor / disulfides / capping groups [a]
CcP-ga	4NVA (closed)	290	11,013	4614	Heme
AA2AR	3EML	290	14514	4569	Disulfides, caps
ACES	1E66	532	16481	8346	Disulfides
ADA	2E1W	349	9775	5536	ZN
AMPC	1L2S	358	12080	5581	
CXCR4	3ODU	306	15546	4988	Disulfides, caps
EGFR	2RGP	257	12374	4120	Caps
FA10	3KL6	282	13069	4331	Disulfides
FABP4	2NNQ	131	5372	2059	
GLCM	2V3F	497	14611	7765	Disulfides, caps
HIVPR	1XL2	198	7841	3128	
HMDH	3CCW	842	36285	12608	
HS90A	1UYG	209	8014	3295	
ITAL	2ICA	179	6917	2901	
KIT	3G0E	332	13892	5298	
KITH	2B8T	206	11994	3290	
LCK	2OF2	271	12925	4392	
NRAM	1B9V	391	11140	5979	Disulfides, Ca ion
PARP1	3L3M	348	12689	5510	
PLK1	2OWB	294	16083	4828	
PPARA	2P54	267	11020	4282	
PTN1	2AZR	297	12120	4811	
PUR2	1NJS	200	9464	3056	
SRC	3EL8	263	9783	4200	Caps
THRB	1YPE	250	8567	4023	Disulfides, caps
TRY1	2AYW	223	8042	3221	Disulfides

[a] NME and ACE were added to cap breaks (missing residues).

Docking. Scripts and programs in the DOCK3.7 distribution(21) were used to prepare the receptors and ligand databases for docking and to carry out the library screens. Blastermaster.py was used to prepare the protein: hydrogens were added with *Reduce*,(22) spheres were generated with *sphgen*(23) and by converting the crystallographic ligand atoms to spheres (spheres are used to orient molecules into the binding site); electrostatic grids were generated by solving the Poisson-Boltzmann equation with the *Qnifft* program;(24) van der Waals grids were calculated using *Chemgrid*,(25) the ligand desolvation grids were produced with *solvmap*,(26) all distributed within the DOCK3.7 program suite. A GIST component to the scoring function was integrated in a new release of DOCK3.7 (**Section S3** and **Figure S1**). Default parameters were otherwise used for docking. CcP-ga was prepared as a flexible receptor with 16 different conformations, as described.(8) All other systems used a single receptor conformation. To use GIST, proteins were aligned using Chimera(27) into the simulation's frame of reference before DOCK preparation.

Enrichment calculations. LogAUC is described in Mysinger and Shoichet(26). We specify a lower bound of 0.001 FPR to avoid infinitely negative values of log(0). The maximum area under the curve is 3, we then convert this value to a percent (maximum area) and subtract the area under the random curve. Thus, LogAUC ranges from -14.5 to 85.5 where 0 is random and anything above 0 is better than random, and blow, worse. Note that these values will change for other lower bounds (the lambda parameter in Mysinger et al.). The CcP-ga ligand datebases where generated as described below at ph4, while the DUD-E databases were obtained from the Autodude webpage (21). Protein structures were prepared for docking described above (docking section).

Database generation. The databases were generated using the DOCK3.7 ligand generation pipeline. ChemAxon (molconvert) was used to generate a 3D molecule from SMILES. The protonated states of the ligands are generated using Marvin of ChemAxon. Protonation states of the molecule were generated at pH 4.0 (greater than 20% occupancy). AMSOL7.1(28) was used to calculate the partial charges and per atom decomposition of ligand desolvation, Openeye Omega was used to generate an ensemble of conformations of each ligand. These conformations are stored in db2 format using the db2

generation program distributed with DOCK 3.7. Ligand databases downloaded from ZINC15 used the same pipeline but were generated at pH 6.4.

References.

1. Mysinger MM, Carchia M, Irwin JJ, & Shoichet BK (2012) Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of Medicinal Chemistry* 55(14):6582-6594.
2. Allen WJ & Rizzo RC (2014) Implementation of the Hungarian Algorithm to Account for Ligand Symmetry and Similarity in Structure-Based Design. *Journal of Chemical Information and Modeling* 54(2):518-529.
3. Brozell SR*, et al.* (2012) Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J. Comput. Aided Mol. Des.* 26(6):749-773.
4. Nguyen CN, Cruz A, Gilson MK, & Kurtzman T (2014) Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J Chem Theory Comput* 10(7):2769-2780.
5. Nguyen CN, Young TK, & Gilson MK (2012) Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J Chem Phys* 137(4):044101.
6. Ramsey S*, et al.* (2016) Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *Journal of Computational Chemistry* 37(21):2029-2037.
7. Case DA*, et al.* (2014) AMBER 14 (University of California, San Francisco).
8. Fischer M, Coleman RG, Fraser JS, & Shoichet BK (2014) Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat. Chem.* 6(7):575-583.
9. Winter G (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* 43(1):186-190.
10. McCoy AJ*, et al.* (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40(4):658-674.
11. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D* 66(2):213-221.
12. Moriarty NW, Grosse-Kunstleve RW, & Adams PD (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallographica Section D* 65(10):1074-1080.
13. Terwilliger TC*, et al.* (2008) Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallographica Section D* 64(5):515-524.
14. Delano WL (2002) The PyMOL Molecular Graphics System).
15. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of

Coot. *Acta Crystallographica Section D* 66(4):486-501.
16. Berman H, Henrick K, & Nakamura H (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* 10(12):980-980.
17. Götz AW*, et al.* (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* 8(5):1542-1555.
18. Bryce R (AMBER parameter database.
19. Banci L, Carloni P, & Savellini GG (1994) Molecular Dynamics Studies on Peroxidases: A Structural Model for Horse Radish Peroxidase and a Substrate Adduct. *Biochemistry* 33(41):12356-12366.
20. Rocklin GJ*, et al.* (2013) Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. *Journal of Molecular Biology* 425(22):4569-4583.
21. Coleman RG, Carchia M, Sterling T, Irwin JJ, & Shoichet BK (2013) Ligand Pose and Orientational Sampling in Molecular Docking. *PLoS ONE* 8(10):e75992.
22. Word JM, Lovell SC, Richardson JS, & Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation1. *Journal of Molecular Biology* 285(4):1735-1747.
23. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, & Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology* 161(2):269-288.
24. Sharp KA (1995) Polyelectrolyte electrostatics: Salt dependence, entropic, and enthalpic contributions to free energy in the nonlinear Poisson–Boltzmann model. *Biopolymers* 36(2):227-243.
25. Meng EC, Shoichet BK, & Kuntz ID (1992) Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry* 13(4):505-524.
26. Mysinger MM & Shoichet BK (2010) Rapid Context-Dependent Ligand Desolvation in Molecular Docking. *J. Chem. Inf. Model.* 50(9):1561-1573.
27. Pettersen EF*, et al.* (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25(13):1605-1612.
28. Hawkins GD*, et al.* (2004) AMSOL7.1 (University of Minnesota, Minneapolis).