Supplementary Information for:

Incorporating protein flexibility and conformational energy penalties in docking screens to improve ligand discovery

Marcus Fischer[1,2,3], Ryan G. Coleman[1,3], James S. Fraser*[4] & Brian K. Shoichet*[1,2]

[1]Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA 94158, and [2] Faculty of Pharmacy, Donnelly Center, University of Toronto, 160 College St. Toronto Ontario M5S 3E1. [3]These authors contributed equally to this work. [4]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94158.

*Corresponding authors:
James Fraser: james.fraser@ucsf.edu, (415) 493-8421
Brian Shoichet: bshoichet@gmail.com, (416) 978-7224

Supplementary Text

Supplementary Methods

Docking Methods

For this project, ligands were protonated with EPIK[1], using 6 ±0.75 as the pH range, and only states that were at least 50% populated in that range were kept, as before[2]. Protonation states of top scoring molecules were manually checked using the pKa prediction tool built into Marvin from ChemAxon (Marvin version 5.5.1.0 (ChemAxon, 2011)), many molecules were discarded during this step.

Scoring is composed of precalculated grids that evaluate the various energy terms, van der Waals using AMBER[3], electrostatics using QNIFFT[4,5] (an adaptation of DELPHI), and ligand desolvation[6]. Additionally, the ability to save a few top poses of a molecule, a new feature in DOCK 3.7, is taken advantage of here, to calculate a more accurate estimate of the flexible occupancies as in equation 2 or using many poses as inputs to a black-box re-weighting scheme[7], as well as analyzing additional poses.

Receptor flexibility is taken into account in the following manner, an improvement over our own past methods[8], with a complete rewrite of the code and integrated into DOCK 3.7[9]. A few side-chains are designated as flexible, the rest are rigid. The reference (or invariant) state for van der Waals and ligand desolvation calculations is a receptor with the flexible side-chains removed entirely. The reference (or invariant) state for electrostatics is the receptor with the side-chains in the most occupied state. For each alternate position, flexible atom positions are evaluated with van der Waals and ligand desolvation grids. Since both the van der Waals and ligand desolvation terms are additive, this approximation is unlikely to lead to problems. For the electrostatic term for each flexible portion of the protein, the receptor is evaluated with the flexible portion in question and all other side-chains in the most occupied state. The electrostatic potential grid from this evaluation has the electrostatic potential grid from the invariant state subtracted from it, resulting in a grid that shows the change in electrostatic potential from moving the side-chain or loop to the new position.

Though electrostatics are not additive, the approximation here would be reasonable as long as two mobile, polar side-chains are not moving close together. In the CcP Gateless site, many of the mobile side-chains were within 4 Å of each other. This was tested, and ligand poses were at most 0.9 kcal/ mol different in electrostatics score as compared to computing the electrostatics score without any decomposition approximations. With the old-style electrostatics decomposition, electrostatic energy errors were much higher, up to 2.3 kcal/ mol[10]. The new-style electrostatics decomposition also predicted more 3/5 poses correctly (RMSD < 0.4Å) in the top pose and 5/5 correctly in the top 10, whereas the old-style electrostatics decomposition[10] only predicted 1/5 correctly in the top pose and only 3/5 correctly in the top 10 poses. For the two ligands, where the new method failed to identify the binding pose as the top pose, the correct poses were generated as the next rank in the list (for **4** with Δ(energy) = 1.4 kcal/mol, for **5** with Δ(energy) = 0.04 kcal/mol) (Suppl. Figure 4). For speed and memory usage, grids are trimmed to the minimum necessary for docking, which means that DOCK 3.7 can dock with 9 different copies of the 3 energy grids in less than half a gigabyte of memory. All grids are allocated dynamically so that only the input files must be changed, code does not need recompiled to run different numbers or combinations of grids.

In the CcP Gateless binding site, one loop from residues 186 to 194 was mobile and modeled in 3 positions; A, B, C. In one of these positions, a residue in the loop takes two conformations, so for technical reasons it was modeled as a separate loop even though only one residue moved (N193), resulting in 4 loop conformations; A, B, C and F. Two residues were also independently mobile, glutamic acid 199 and methionine 228, both modeled in 2 positions. Overall, there were 8 mobile conformations and therefore 16 total receptor combinations that were scored for every ligand pose.

After scoring the invariant receptor grid, if the energy for a given pose is reasonable, all energies across all grids are evaluated. As scoring each atom position on each grid is by far the most time consuming process in DOCK 3.7, there is only a linear response in time depending on how many grids are used. After each grid is scored, all possible combinations of scores are considered to see which pose has the best score. Additionally, each grid representing a flexible side-chain can contribute a penalty based on how occupied the residue is, according to $m\ k_B T \log_e(\text{occupancy})$, cf. equation 1, where m is the flexible multiplier, k_B is the Boltzmann

constant, T is the temperature and the occupancy is obtained through crystallography. Note, that in the case of one fully occupied conformation (*ln*(1)=0) no energy penalty will be applied while with lower occupancy the penalty increases.

As stated, the top 10 poses were saved, these will be the top 10 to any combination of receptor possibilities. Additionally, the top pose to each receptor possibility can be saved (here up to 16; 3 loop positions, loop C having a flexible sidechain, times 2 positions for Glu199 times 2 positions for Met228), which can be useful when examining docking results. Since many poses are saved to many receptor combinations, the prediction of the occupancy of the various loops and flexible side-chains could be made from this according to Boltzmann's law according to the following. For every pose energy the propensity = e^(-energy/k_BT) is computed, then these propensities these are used to compute the relative occupancy of one receptor combination to all receptor combinations, finally producing an occupancy for a given loop or residue.
To determine the ensemble energy of a given ligand, the energies from the best ten docked poses for that ligand were summed, plus at least one from all 16 possible combinations of receptor conformations. Using this limited number of poses appears sufficient as poses with much higher energies (>3 kcal/mol) will not change the propensities by more than 1%. Empirically, increasing sampling to 1000 ligand poses did not substantially change the results (Suppl. Table 4).

We used Pearson Correlation Coefficients (PCC) to quantify the agreement between our predictions of the flexible protein loop and the occupancies derived from the experiment. The PCC values are simply the Pearson Correlation Coefficient of the predicted loop propensities (for each loop, a number between 0 and 1, where the numbers across all loops sum to 1) and the refined occupancies of the loops (an identical range of numbers is possible).

We searched integer values from 0 to +4 by calculating new loop propensities for each of the five ligand complexes. For what we found to be the optimal *m* value of 2, the PCC between experimental and predicted occupancies rose to 0.83 – a small but significant difference from the 0.77 correlation found without any weighting, with a corresponding decrease in p-value from 0.0059 (*m*=1) to 0.0047 (*m*=2). When poses and

occupancies of 5 known ligands were examined, a weighting scheme using a flexible multiplier *m*=2 times the energy penalty proved to be better at reproducing the occupancies of the loops. This scheme favored higher occupancy states twice as much as dictated by occupancy alone, so predictions of loop and residue occupancies were made with the '1x' and '2x' weighting schemes.

The occupancy of loop conformation *B* reproducibly converges to 4% in the *apo* structure (Suppl. Figures 2 and 3) and lies below our expected imprecision threshold of ±10%. We modeled a range of reasonable occupancy values for the *B* loop (0.0000001 to 0.15), corresponding to energy penalties from ~9.6 to ~1.1 kcal/mol. An occupancy of <0.01 is not meaningful which sets an artificial penalty limit of (*m* *) 2.73 kcal/ mol. However, the Boltzmann weighting methods provides freedom to assign high penalties for artificially low occupancy values (e.g. *m* * 23 kcal/ mol for occ. of 10^{-10}). Upon calculating the PCC and p-value of experimental loop occupancy vs. predicted loop propensity, we found that statistically significant correlations (p-value < 0.05) were still obtained even after changing the input loop occupancy of loop *B* anywhere in the range from 0.0001 to 0.09 (Suppl. Figure 6C).

We implemented a version of the black-box re-weighting scheme of Ytreberg & Zuckreman[7]. First, we use the feature of DOCK 3.7 to extract the top 1000 poses, plus one pose for each receptor combination. In practice, these poses seem reasonably distributed amongst the receptor combinations. We use the nearest neighbors strategy described by first computing the root mean squared deviation (RMSD) between two poses. This RMSD is the ligand RMSD, corrected for symmetry by the Hungarian algorithm[11], added to the all atom protein RMSD to account for loop movements. This RMSD is used to compute nearest neighbors for each pose, and to find the radius of the hypersphere that contains the appropriate number of neighbors. To compute the black-box re-weight of each pose, we use the following equation:

$$\mathrm{BBRW(pose\ x)} = \frac{e^{-\mathrm{score(x)}/k_B*T}}{N_{dist}/R_{hyp}^{dof}} \quad \text{(Supplementary Equation 1)}$$

Where the score of any pose is score(x), N_{dist} is the chosen parameter value for the number of nearby points to use, R_{hyp} is the radius of the hypersphere for that pose x that contains that many points and dof is the degrees

of freedom or dimensionality parameter. Once the BBRW energy of each pose is computed, this is used in Equation 2 to compute loop propensities.

We investigated several parameters for N_{dist} and dof, dof only seemed to significantly change the resulting loop propensity when it was set very high (as it tends to drown out non-dominant loop propensities). The value for N_{dist} did change the results only insignificantly and overall the results agree within a reasonable margin with the raw energy scores used in Equation 2 for loop propensity calculations.

For the final screening against the fragment-like portion of ZINC[12] molecules were chosen from the top 750 ranking fragment “hits” to any receptor combination and many were eliminated due to known problems, for instance an incorrectly charged state due to EPIK[1,13] by manually checking each protonation and tautomerization state with ChemAxon's Marvin. Putative ligands were selected to bind to interesting receptor combinations or interesting predicted occupancies of receptor combinations, interesting in the sense of shedding light on the performance of our algorithm for cases with different degrees of difficulty.

For each ligand, we have both the predicted and experimentally determined numbers. In the p-value test, we permute the labels on these numbers and check the PCC value, the fraction of times the PCC is higher than the observed PCC for the original values is reported as the p-value - for statistical significance. A PCC of 0 indicates random agreement, 1 indicates perfect agreement between theory and experiment (predicting the mix of occupancies exactly), and -1 indicates anti-correlated agreement (predicting entirely loop C when the experimental occupancy was entirely loops A and B, for instance). This agreement allowed us to quantify the extent of the agreement and monitor it across different weighting schemes.

We calculated PCC for all energy multipliers m from -20 to 20 for (a) all nine new structures (Figure 5B; “Pearson all”) and (b) excluding the data with partial presence of MES or the 4th loop (Figure 5B; “Pearson 6”). To obtain p-values, we enumerated over all possible combinations of predictions. Results obtained by using our energy penalties to rank and purchase compounds prospectively fall within a narrow window of statistically relevant p-values < 0.01 and at the top of the PCC curve (Fig. 5B).

Of final note is that the computation of thousands of PCCs and p-values over many parameter sets took longer than the entire docking calculation.

Experimental Methods

Affinities were measured by fluorescence monitoring of the heme Soret band shift as before[2,14].

Crystallographic data were collected at the ALS Berkeley beamline 8.3.1 and processed using the xia2 pipeline[15]. Structures were solved by molecular replacement with Phaser[16] using the same model and R_{free} for all structures. Alternating cycles of refinement and model building were carried out in Refmac_5.5.0109[17] and Coot[18], respectively with the ligand being added at the late stages of refinement. Phenix loop occupancy refinement was applied starting with the equally weighted triple loop apo model for all ligand complexes.

Note, that occupancy is not an independent variable – it is highly correlated with the crystallographic B-factor. B-factors integrate conformational substates, lattice disorder, diffusion and atomic or coordinated molecular vibration. Differences in these variables may be a property of the crystal rather than the structure. Also, model errors, such as inclusion of irrelevant or exclusion of relevant conformational heterogeneity can lead to changes in the B-factor model which can again have an impact on refined occupancies and hence, the energy penalization.

The advanced search on the pdb used the following search query: “Has free ligands=no AND Resolution is between 0.0 and 1.5 AND Experimental Method is X-RAY and has Experimental Data AND Representative Structures at 100% Sequence Identity”. Query refinement on the pdb webpage provides useful options to follow particular links of interest (Suppl. Fig. 12). We found 24,864 *apo* structures, of which 13,373 have deposited electron density, which enables fitting multiple models and occupancy-based energy penalization. Of these, 2899 are high resolution (<1.8 Å), and 932 are very high resolution (<1.5 Å). Removing sequence identical ones reduces the number from 932 to 827 as mentioned in the main text.

Supplementary Results

The influence of temperature on modeled conformations

Admittedly, the 837 high-resolution pdb structures to which our methods is directly applicable drop to 51 if one restricts them to those collected at room temperature, as the apo-structure of CcP was. At cryo-temperatures, where the rest of the structures were determined, the residue and loop flexibility we observe and exploit here might disappear. To investigate this, we determined the apo structure of the CcP-gateless mutant not only at room temperature but also at cryogenic temperatures (Suppl. Table 1). The two structures are similar, and in the region of the loop and flexible residues the low temperature structure largely recapitulates the occupancies observed at room temperature (Suppl. Figure 3b). One key difference is that the low occupancy B conformation observed at room temperature disappears, consistent with its high-energy status (i.e., higher energy states are less populated at lower temperatures[19]. This leaves only the C and A conformations, whose weights change only very modestly from those at room temperature (Suppl. Figure 3b). Whereas this does suggest that fewer high-energy states will be observable for low temperature structures, it is the A and C conformations of the loop, and the alternate conformations of Asn193, Glu199, and Met228, that have been missed in previous docking studies, that remain present in the cryo-apo structure, and that we ultimately target in this study to find novel ligands. Thus, room temperature structures will more fully explore conformational heterogeneity present in protein structures[20], even cryo structures are likely to support enough conformations to support this analysis; this becomes ever more true as resolution rises[21,22].

Prospective prediction of side-chain flexibility

In addition to the more difficult case of predicting loop movement, we also turned to investigate the correspondence with the predictions of three residues, Asn193, Glu199 and Met228, that we had treated as flexible based on the *apo* density. The first two residues lie at the interface between bulk solvent and the cavity. The last sits near the ring system conserved in most ligands. Two residues were especially challenging: 1) Glu199 is poorly constrained, to a point where no well-defined electron density could be observed, 2) the Met228 residue is relatively unresponsive to binding compounds **1-5**, but is predicted to adjust its backbone carbonyl by 18° upon binding of ligand **9**. However, the Met228 conformation for all the prospective ligands

remained unchanged, making this prediction one of the two failures out of the total 27 modeled conformations (Suppl. Figure 7). The other misprediction was for ligand **13** where the crystal structure shows that an unmodeled tetrahedral water molecule coordinates two of the flexible residues (Asn193 and Glu199) and stabilizes the population of the dominant rotameric states (Suppl. Figure 7). It was this dataset in which the electron density of Glu199 was most clearly defined (Figure 3H). A noteworthy success was the prediction of the double Asn193 rotamer responding to ligand **14** binding (Suppl. Figure 7).

Statistical justification for m

Another reason why model choice on the basis of retrospective enrichment alone can be misleading is that although some ligands rank high for all loops (Suppl. Table 4), they show a clear preference for only one loop conformation (Fig. 4). For example, purchasing compound 7 on the basis of its A loop rank would have also lead to discovering 7 to bind to the CcP mutant – but for the wrong reasons, as experimental occupancies confirm clear preference for B and not A as predicted by docking propensities. Only flexible docking correctly predicted those occupancy profiles representing the protein response to ligand binding for the right reasons. The agreement of the blind loop predictions with the experimental data using the weighting multiplier m = 2 is statistically significant with p-values < 0.01 (Figure 5B) and suggests a physically meaningful choice of docking model and energy penalization. In contrast, we found anti-correlation for the highest enriching loop (Figure 5B, and Suppl. Figure 8), confirming its inherent bias.

Supplementary Discussion

The contribution of the multiple conformations to the discovery of the new, larger ligands, and the trade-off between ligand binding and loop conformational energy, may be understood by considering the ligands selecting the *C* conformation of the 186-194 loop. In the *B* conformation of this loop—which we have used in all previous studies against this cavity[2,14] — three internal hydrogen bonds are made between the loop and the rest of the protein (Asn193 to Gly178, Gly189 to Thr180, and Gly190 to Gly226). In the *C* conformation, one new bond is gained between Gly178 and Ala192. Perhaps more importantly, in the *B* conformation, the carbonyl backbone of Gly190 points into the already negatively charged binding site, causing additional

electrostatic strain and further explaining the low occupancy and high energy of this state (Suppl. Figure 1). An electrostatic assessment of the receptor desolvation of the binding site for the different loop conformation also finds that loop C confers the lowest receptor desolvation (using the Poisson Boltzmann program Qnifft[5], Suppl. Table 7). Of course, hydrogen-bond inventories don't linearly translate into energy differences, nor do calculations of electrostatic solvation; still, these analyses suggests that the ensemble-based energies derived from the apo-state occupancies are sensible. Had we not modeled this *C* conformation, we would not have scored these molecules favorably, not only because of steric changes in the pocket but also because they would have lost the opportunity for the polar interaction to the amide backbone of Gly191 in the docking. The same coin must be tendered to weight this interaction, however—had we included the *C* conformation, but not penalized it by its occupancy weight, we would have underweighted these hydrogen bonds between the ligand and the loop. This sort of trade-off played out in most of the new docked ligands and their structures.

Suppl. Table 1 Data collection and refinement statistics

compound	**1**	**2**	**3**	**4**
PDB ID	4NVF	4NVE	4NVD	4NVC
ZINC ID	1583444 (MES)	331902 (BZI)	331945	36634
Data collection				
Space group	P212121	P212121	P212121	P212121
Cell dimensions				
a, *b*, *c* (Å)	50.7, 74.3, 106.0	50.9, 73.5, 104.3	50.8, 73.6, 104.5	50.8, 73.9, 104.4
□□□ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	32.9-1.49 (1.53-1.49)*	38.83-1.54 (1.58-1.54)	31.5-1.3 (1.33-1.3)	31.5-1.6 (1.65-1.6)
R_{merge}	0.1 (0.50)	0.04 (0.386)	0.04 (0.47)	0.04 (0.1)
I/□*I*	7.6 (2.3)	18.0 (2.2)	15.4 (2.4)	20.0 (6.3)
Completeness (%)	99.7 (99.9)	93.5 (62.0)	99.5 (97.8)	84.8 (43.1)
Redundancy	3.9 (3.9)	3.7 (2.4)	3.9 (3.3)	3.6 (2.1)
Mosaicity	0.34	0.18	0.28	0.49
Wilson B	15.4	17.4	13.3	14.6
Refinement				
Resolution (Å)	32.9-1.49	38.8-1.54	31.5-1.3	31.5-1.6
No. reflections	65855 (4815)	54804 (2630)	96286 (6952)	44139 (1630)
R_{work}/ R_{free}	0.1628/0.1932	0.1451/0.1796	0.1299/0.1525	0.1423/0.1847
No. atoms				
Protein	2508	2421	2546	2454
Ligand/ion	55	52	54	52
Water	380	333	482	383
B-factors				
Protein	15.5	18.3	15.1	15.8
Ligand/ion	11.5	13.9	10.6	11.2
Water	28.4	28.3	34.1	30.2
R.m.s deviations				
Bond lengths (Å)	0.018	0.016	0.016	0.015
Bond angles (º)	1.75	1.46	1.60	1.43

*Highest resolution shell is shown in parenthesis.

compound	**5**	**APO-RT**	**6**	**7**
PDB ID	4NVB	4NVA	4NVG	4NVH
ZINC ID	8652421	-	6656163	4962659
Data collection				
Space group	P212121	P212121	P212121	P212121
Cell dimensions				
a, *b*, *c* (Å)	51.0, 74.7, 106.6	51.5, 76.5, 107.4	50.7, 70.4, 101.9	50.9, 74.7, 106.7
□□□ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	39.2-1.17 (1.2-1.17)	39.69-1.57 (1.61-1.57)	33.3-1.74 (1.79-1.74)	32.1-1.24 (1.27-1.24)
R_{merge}	0.056 (0.369)	0.057 (0.64)	0.061 (0.562)	0.039 (0.581)
I/□*I*	17.7 (2.1)	13.0 (2.1)	12.5 (2.3)	21.8 (2.3)
Completeness (%)	96.6 (71.8)	99.7 (99.8)	99.8 (99.9)	97.5 (83.7)
Redundancy	5.5 (2.0)	4.1 (4.1)	4.0 (4.2)	5.6 (2.4)
Mosaicity	0.10	0.10	0.49	0.13
Wilson B	9.7	19.0	22.1	12.2
Refinement				
Resolution (Å)	39.2-1.17	39.7-1.57	33.3-1.74	32.1-1.24
No. reflections	131945 (7157)	59784 (4393)	37920 (2758)	112699 (6989)
R_{work}/ R_{free}	0.1212 /0.1363	0.1244/0.1530	0.1859/0.2246	0.1285/ 0.1450
No. atoms				
Protein	2519	2463	2569	2632
Ligand/ion	57	43	81	57
Water	602	244	196	449
B-factors				
Protein	12.4	20.9	24.9	15.5
Ligand/ion	9.6	14.3	25.0	12.9
Water	29.1	36.3	33.3	29.0
R.m.s deviations				
Bond lengths (Å)	0.013	0.018	0.016	0.018
Bond angles (°)	1.42	1.59	1.95	1.56

compound	**8**	**9**	**10**	**11**
PDB ID	4NVI	4NVJ	4NVK	4NVL
ZINC ID	331160	13739037	1596053	34979991
Data collection				
Space group	P212121	P212121	P212121	P212121
Cell dimensions				
a, *b*, *c* (Å)	50.8, 73.5, 104.2	51.3, 73.1, 104.0	50.6, 71.6, 102.9	50.8, 73.6, 104.1
□□□ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	31.4-1.51 (1.55-1.51)	46-1.81(1.86-1.81) *	30.9-1.56 (1.60-1.56)	32.6-1.43 (1.47-1.43)
R_{merge}	0.043 (0.4)	0.074 (0.715)	0.042 (0.566)	0.038 (0.428)
I/□*I*	16.9 (2.2)	13.6 (2.0)	19.7 (2.1)	21.3 (2.2)
Completeness (%)	99.1 (94.7)	99.5 (99.3)	99.8 (99.8)	98.1 (85.1)
Redundancy	4.0 (2.6)	4.1 (4.2)	4.3 (3.6)	3.9 (2.5)
Mosaicity	0.31	0.18	0.30	0.252
Wilson B	17.7	20.9	19.1	14.1
Refinement				
Resolution (Å)	31.4-1.51	46-1.81	30.9-1.56	32.6-1.43
No. reflections	61321 (4240)	36134 (2613)	53815 (3925)	71072 (4463)
R_{work}/ R_{free}	0.1452/ 0.1780	0.1552/ 0.1837	0.1402/ 0.1737	0.1483/ 0.1754
No. atoms				
Protein	2638	2551	2677	2751
Ligand/ion	75	70	64	58
Water	338	129	301	426
B-factors				
Protein	19.4	24.7	21.8	14.7
Ligand/ion	23.0	22.1	19.2	17.6
Water	31.8	30.9	35.0	28.3
R.m.s deviations				
Bond lengths (Å)	0.017	0.017	0.019	0.019
Bond angles (°)	1.54	1.46	1.77	1.66

compound	**12**	**13**	**14**	**APO-cryo**
PDB ID	4NVM	4NVN	4NVO	4OQ7
ZINC ID	203341	519712	388812	-
Data collection				
Space group	P212121	P212121	P212121	P212121
Cell dimensions				
a, *b*, *c* (Å)	50.9, 73.7, 104.6	51.0, 73.4, 104.4	51.3, 75.2, 105.5	50.4, 70.3, 102.8
□□□ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	31.5-1.51 (1.55-1.51)	31.5-1.47 (1.51-1.47)	46.2-1.71 (1.75-1.71)	38.0-1.89 (1.94-1.89)
R_{merge}	0.044 (0.413)	0.046 (0.358)	0.057 (0.703)	0.071 (0.615)
I/□*I*	16.6 (2.2)	15.4 (2.2)	17.6 (1.9)	14.5 (2.1)
Completeness (%)	98.5 (90.8)	99.2 (93.5)	99.8 (99.7)	99.6 (99.0)
Redundancy	3.8 (2.5)	3.9 (2.4)	4.3 (4.3)	3.9 (3.6)
Mosaicity	0.29	0.27	0.22	0.424
Wilson B	17.2	16.3	20.2	22.0
Refinement				
Resolution (Å)	31.5-1.51	31.5-1.47	46.2-1.71	38.0-1.89
No. reflections	61457 (4115)	66685 (4587)	44741 (3250)	29781 (2137)
R_{work}/ R_{free}	0.1412/ 0.1726	0.1423/ 0.1697	0.1335/ 0.1685	0.1750/ 0.2168
No. atoms				
Protein	2645	2610	2686	2666
Ligand/ion	74	73	56	43
Water	345	395	217	194
B-factors				
Protein	18.0	17.2	21.2	26.0
Ligand/ion	16.7	19.5	15.0	20.5
Water	30.7	31.9	32.5	30.5
R.m.s deviations				
Bond lengths (Å)	0.018	0.019	0.018	0.015
Bond angles (°)	1.52	1.65	1.50	1.40

Suppl. Table 2 R-factors of crystallographic refinement of apo protein for different loop combinations shows that adding more loops improves the model statistically and combinations with the highest occupancy loop *C* rank higher than non-*C* models. Note that *B* alone ranks higher than *A* which is unexpected in terms of their relative occupancies but can be explained by the *B* loop conformation falling into the *C* loop density which in its absence in the model is assigned a higher relative occupancy than observed in the automatic refinement.

Loop	R free
ABC	0.1639
ABCD	0.1639
AC	0.1641
BC	0.1646
C	0.1648
AB	0.1659
B	0.1669
A	0.1695
D	0.1712

Suppl. Table 3 Pearson Correlation Coefficients between Retrospective Predictions of Loop Occupancies by equation 2 or Black-box re-weighting (BBRW)[7] with various parameters.

Mean (standard deviation)	1000 poses N_{dist} = 24 dof = 6	1000 poses N_{dist} = 10 dof = 6	1000 poses N_{dist} = 50 dof = 10	1000 poses N_{dist} = 50 dof = 1
Loop *B* Occupancy 0.000001 to 0.15	0.944 (0.031)	0.941 (0.056)	0.927 (0.037)	0.993 (0.004)
Flexible Weighting Multiplier *m* 0.0 to 4.0	0.942 (0.033)	0.973 (0.026)	0.921 (0.030)	0.994 (0.005)
Dock Score Weighting Multiplier *m* 0.1 to 3.0	0.919 (0.029)	0.902 (0.051)	0.919 (0.012)	0.995 (0.001)

See Supplementary Docking Methods for a description of the implementation of the BBRW algorithm. Here, 4 different parameter sets were used with the top 1000 poses to compute loop propensities and compare them to the ones computed with equation 2 across different parameters. N_{dist} is the chosen number of nearby points to examine, dof is the degrees of freedom / dimensionality constant used in the equation.

Suppl. Table 4 Pearson Correlation Coefficients between Prospective Predictions of Loop Occupancies by equation 2 or Black-box re-weighting[7] with various parameters.

Mean (standard deviation)	1000 poses N_{dist} = 24 dof = 6	1000 poses N_{dist} = 10 dof = 6	1000 poses N_{dist} = 50 dof = 10	1000 poses N_{dist} = 50 dof = 1
Loop *B* Occupancy 0.000001 to 0.15	0.965 (0.013)	0.899 (0.045)	0.992 (0.007)	0.996 (0.002)
Flexible Weighting Multiplier *m* 0.0 to 4.0	0.945 (0.027)	0.903 (0.048)	0.980 (0.014)	0.997 (0.002)
Dock Score Weighting Multiplier *m* 0.1 to 3.0	0.886 (0.076)	0.855 (0.028)	0.968 (0.039)	0.993 (0.006)

See Supplementary Docking Methods for a description of the implementation of the BBRW algorithm. Here, 4 different parameter sets were used with the top 1000 poses to compute loop propensities and compare them to the ones computed with equation 2 across different parameters. N_{dist} is the chosen number of nearby points to examine, dof is the degrees of freedom / dimensionality constant used in the equation.

Suppl. Table 5 Ranks of newly discovered ligands for the consensus model or for a docking screen against a model including only one of the three possible loops.

Compound ID	Loop State Xtal / Dock	Rank - Flexible	Rank - Just A	Rank - Just B	Rank - Just C
6	*C / B*	8	11	6	173
7	*B / B*	38	29	20	401
8	*B / B*	70	39	28	527
9	*C / C*	163	4403	3345	120
10	*AC / A*	322	97	176	5563
11	*A / A*	330	98	97	1571
12	*A / A*	433	131	174	1440
13	*AC / A*	526	170	501	13587
14	*BC / C*	556	220	193	592

Suppl. Table 6 Table of various performance metrics (EF1: enrichment factor at 1%, AUC: Area Under Curve, logAUC: adjusted log Area Under Curve) for the various loops and weightings.

Name	EF1	AUC	logAUC
All *m*=0	31.25	91.51	44.63
All *m*=1	31.25	91.15	41.39
All *m*=2	28.13	90.87	39.01
Loop *A*	37.50	91.70	46.14
Loop *B*	43.75	91.85	47.55
Loop *C*	15.63	89.42	31.64

Suppl. Table 7 Values of binding site receptor desolvation of the crystallographic conformations, measured using dielectric reaction field energies using Qnifft[5]. Desolvation was measured using an artifical ligand that fills the binding site.

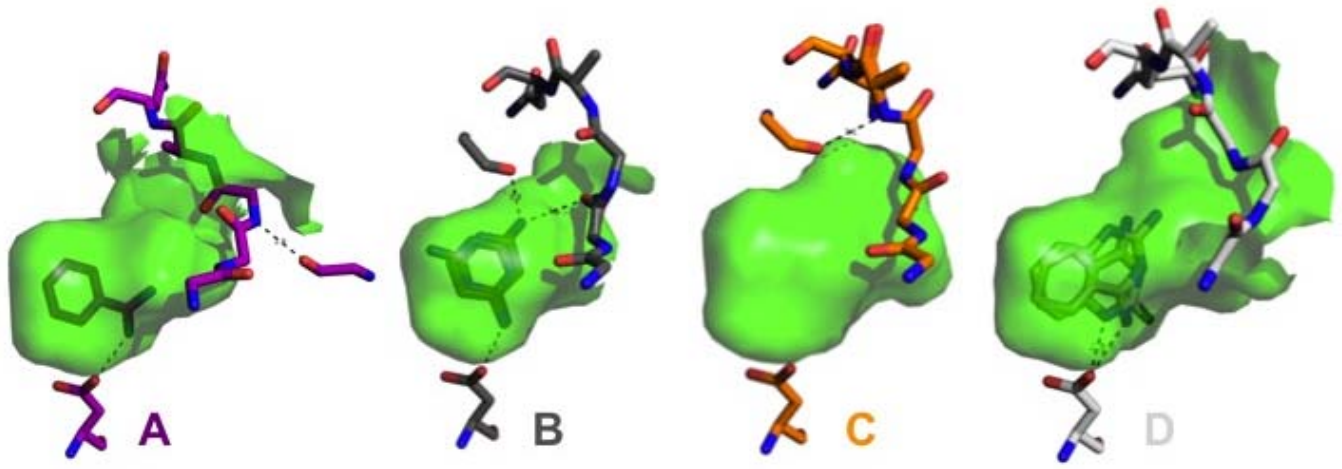
Loop	Binding Site Receptor Desolvation (kcal/mol)
Loop *A*	26.49
Loop *B*	23.23
Loop *C*	18.05
Loop *D*	27.04

Suppl. Table 8 Automatically refined ligand and MES occupancies for each crystal structure.

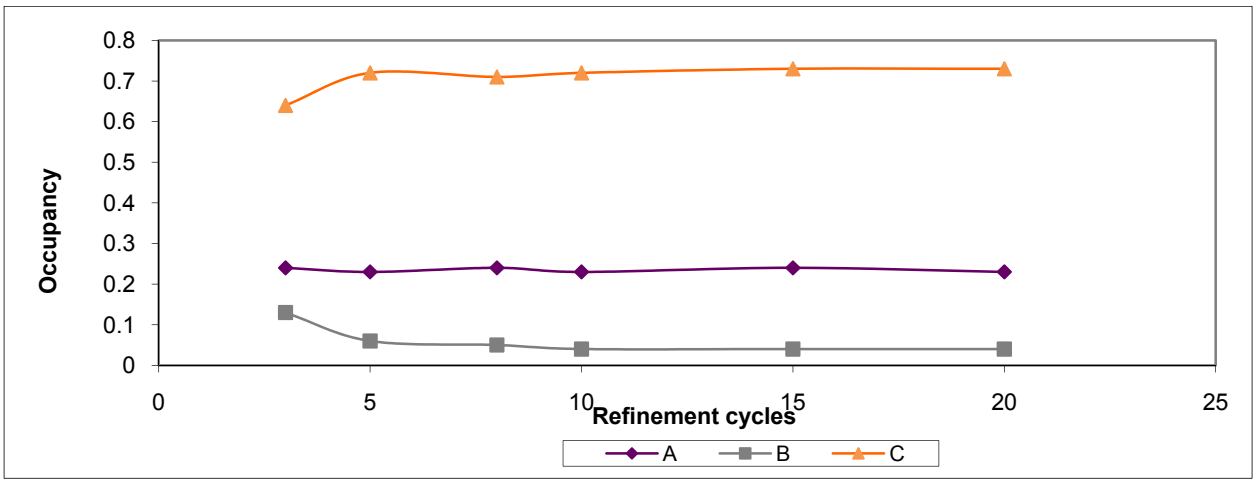
Compound #	Ligand occupancy	MES occupancy
6	0.51	0.42
7	1	-
8	0.84	-
9	0.46	0.35
10	1	-
11	1	-
12	0.68	0.27
13	1	-
14	1	-

Suppl. Table 9 Comparison of lists of ligands previously known for the CcP gateless cavity to those of a rigid-body docking study[2] to this work. Hdon, hydrogen-bond donor; Hacc, hydrogen-bond acceptor: Desol, desolvation: $MaxT_c$, maximum Tanimoto coefficient.

	xlogP	Mol Weight	Rot Bonds	Hdon	Hacc	Apolar Desol	Polar Desol	$MaxT_c$ to Known
Known	0.85	148.0	0.58	2.3	3.0	2.18	-13.64	1
Barelier et al.[2]	1.47	184.7	1.14	2.1	3.4	2.98	-14.73	0.39
This study	1.67	200.0	1.8	1.9	3.4	4.06	-25.03	0.36

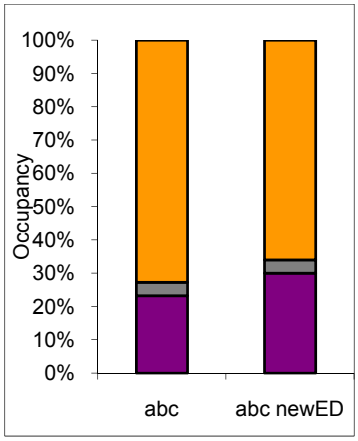


Suppl. Figure 1 Internal hydrogen bonds of the different loop conformations.
Loop *B* and *D* have one internal H-bond less suggesting an explanation for the low occupancy in the Apo state. The presence of the ligands favors the selection of those states via beneficial interactions, either directly or through water.

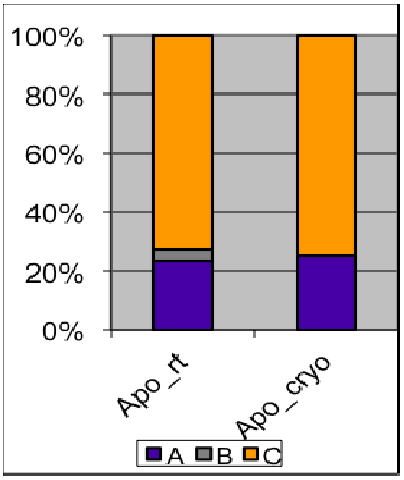


Suppl. Figure 2 Refined occupancies for loops *A, B and C* converge with increasing number of refinement cycles around 10 cycles.

(a)



(b)

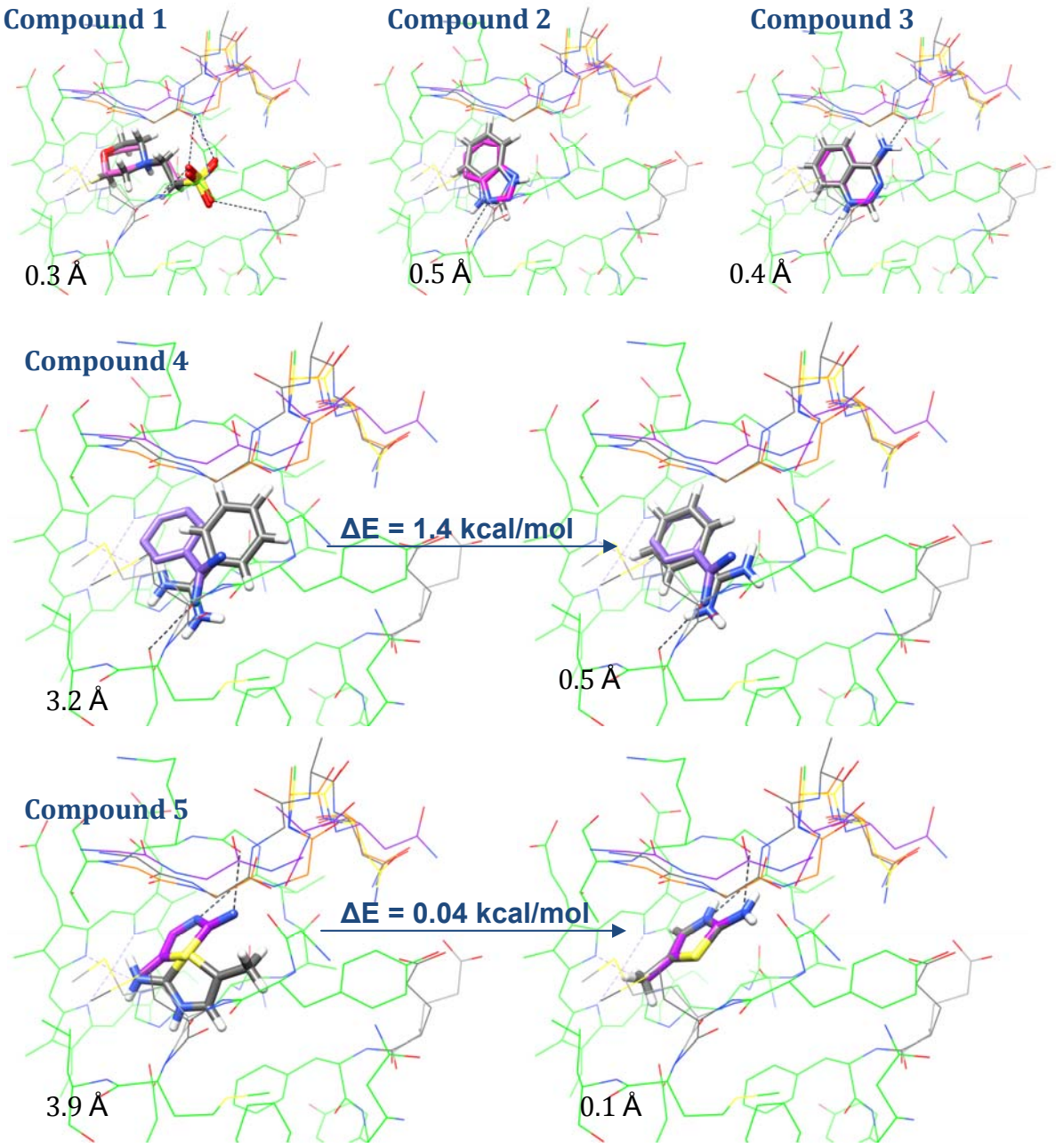


Suppl. Figure 3 Comparison of automatic occupancy refinement of the apo CcP Gateless mutant.

(a) Low occupancy of 4% is reproduced for another RT dataset. Values for A and C are in qualitative agreement after 10 rounds of phenix occupancy refinement.

	A	*B*	*C*
Apo	0.23	0.04	0.72
Apo newED	0.3	0.04	0.66

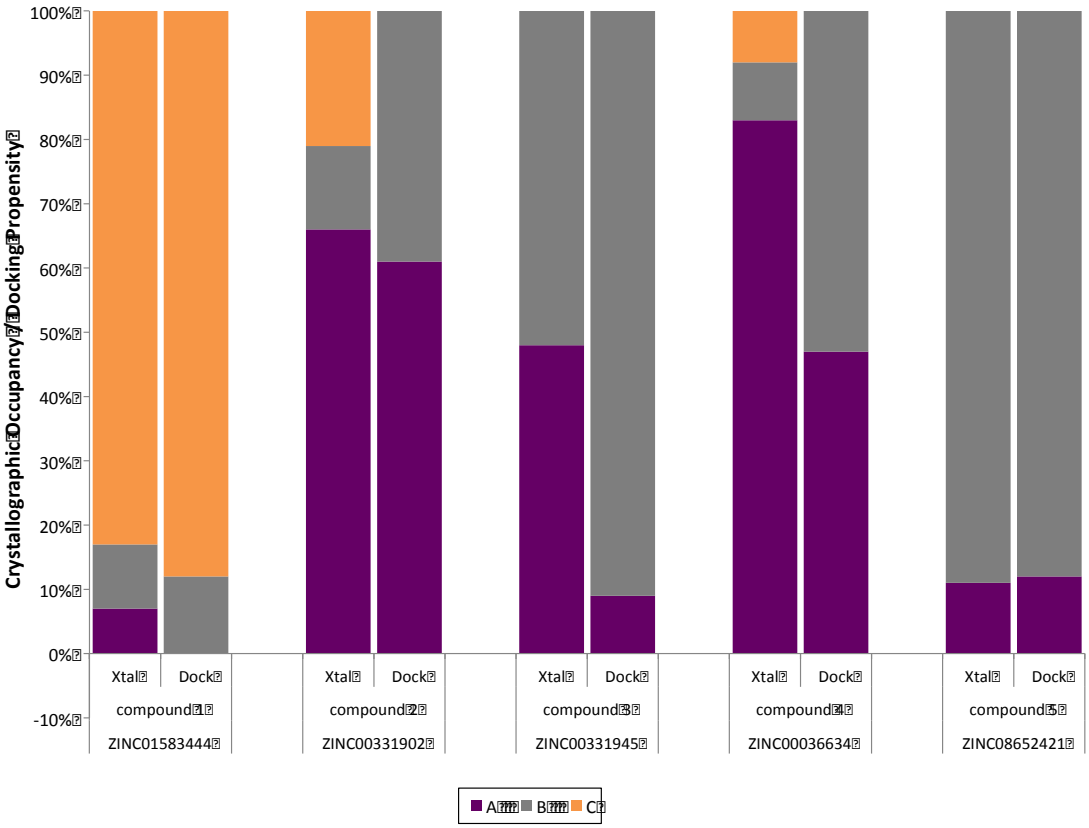
(b) Apo-RT and apo-cryo show a very similar distribution of states. Note that the occupancy for minor loop B disappears at cryo, which is why we used RT to be able to automatically assign a penalty.



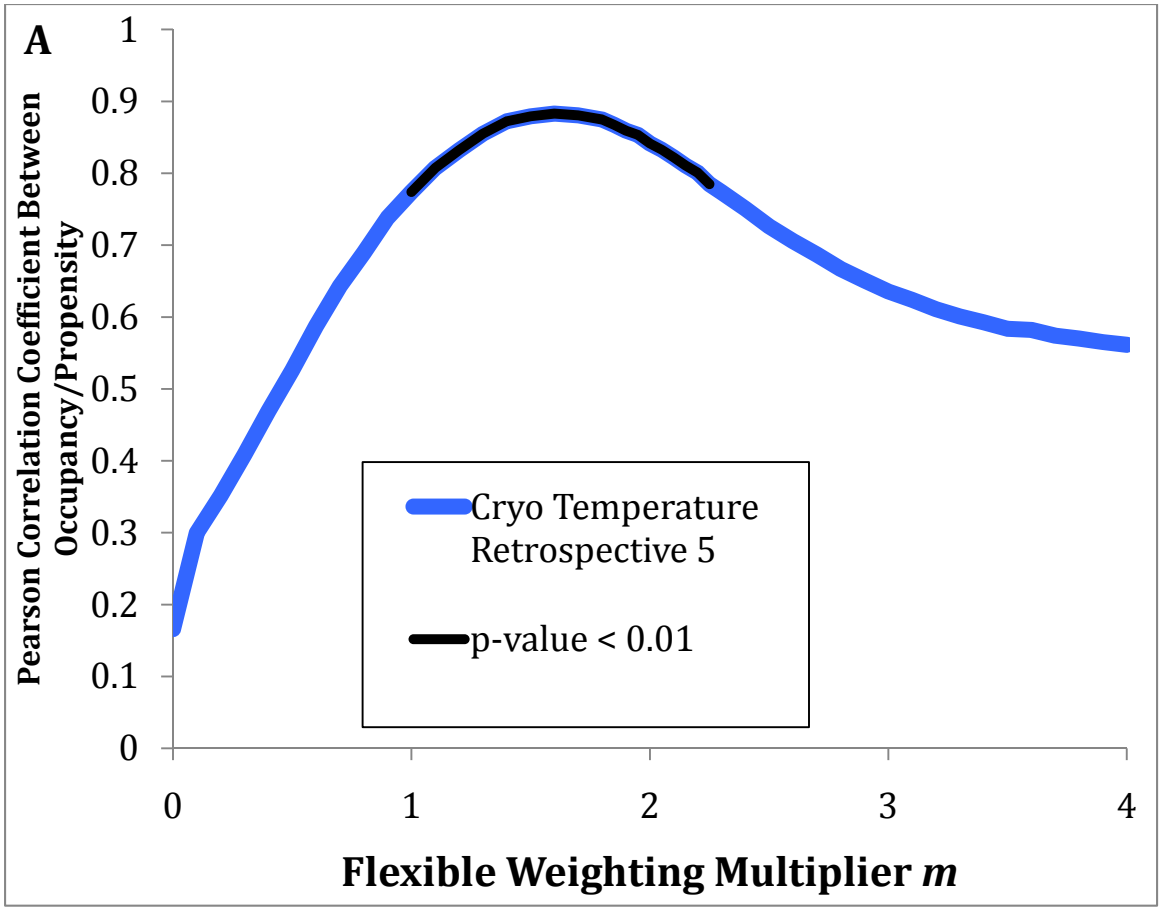
Suppl. Figure 4 Binding poses and RMSDs for retrospective flexible docking.

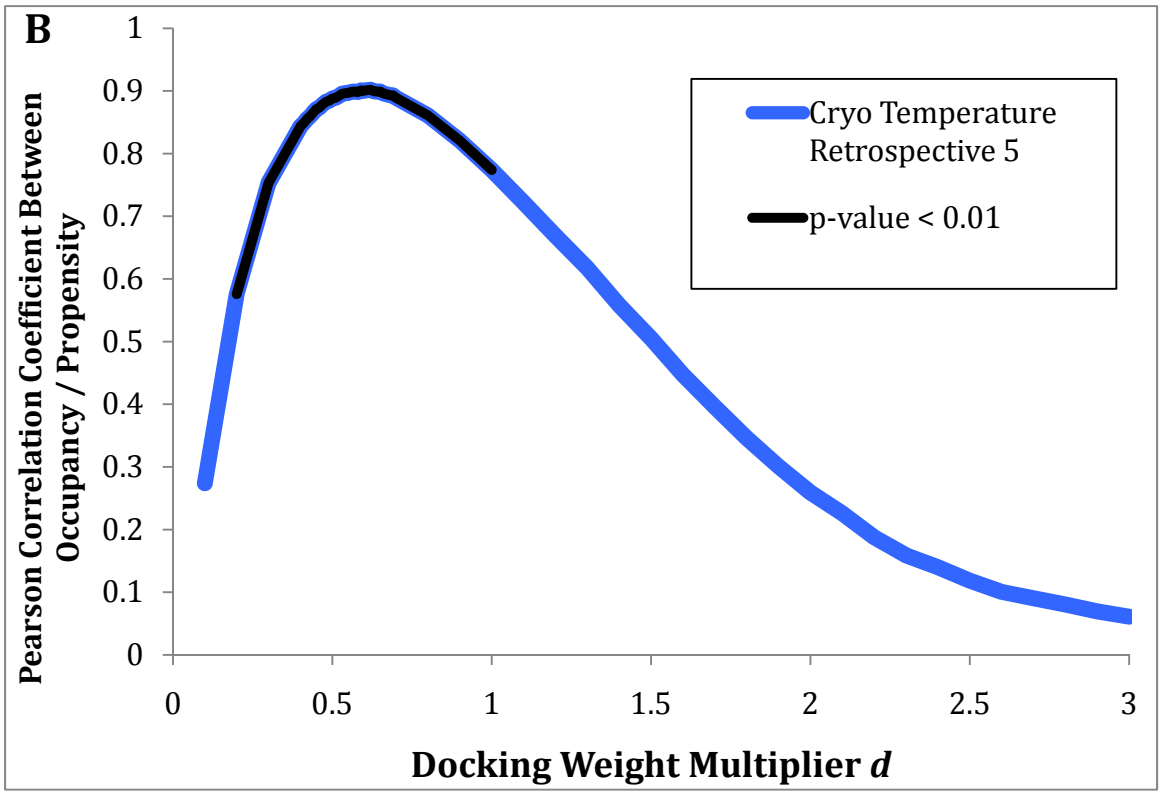
Shown are the top ranks with the exception of two compounds (**4** and **5**) where the correct pose was found further down the top 10 list (delta Energy ΔE).

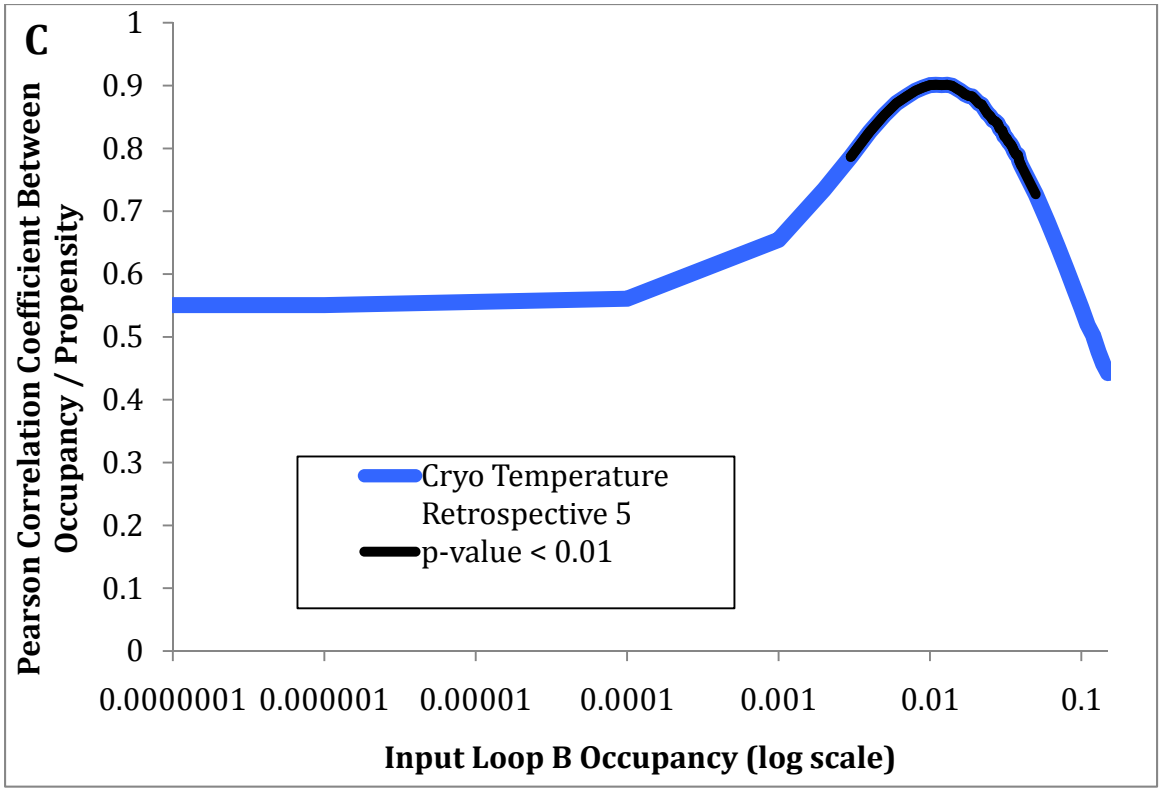
Docked compound in grey, experimental binding pose in purple. RMSDs between prediction and crystal structure are given in the bottom left for each compound.



Suppl. Figure 5 Loop Occupancy Propensity prediction with flexible weighting multiplier *m* = 1. Crystallographic occupancies on the left of each pair, the right is the predicted propensity from docking for *m*=1 (compare to Figure 2 for *m*=2). Here, the overall Pearson Correlation Coefficient of the occupancies/propensities of these ligands is 0.77, whereas for *m*=2 the coefficient is 0.83.


A
Pearson Correlation Coefficient Between Occupancy/Propensity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
Cryo Temperature Retrospective 5
p-value < 0.01
Flexible Weighting Multiplier m


B
Pearson Correlation Coefficient Between Occupancy / Propensity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.5
1
1.5
2
2.5
3
Cryo Temperature Retrospective 5
p-value < 0.01
Docking Weight Multiplier d

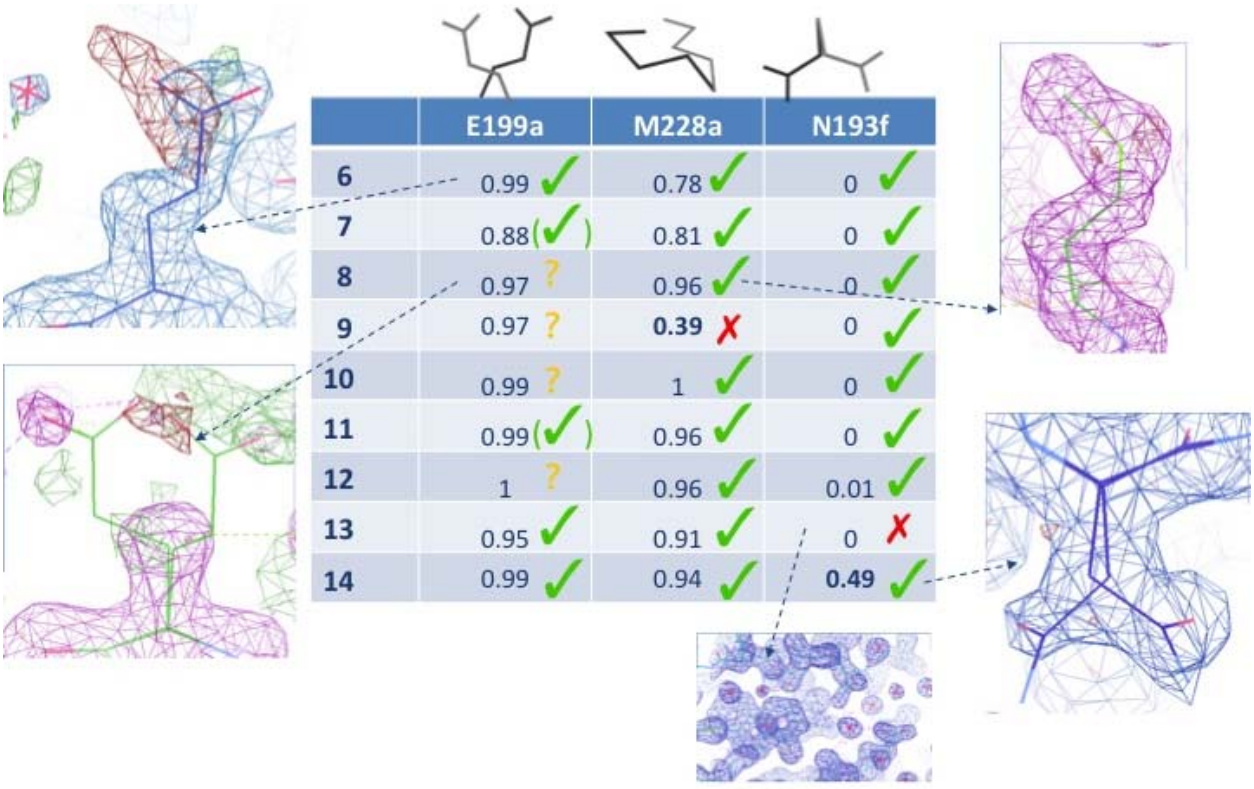


Suppl. Figure 6 Response of PCC to the crystallographic occupancy and the predicted DOCK loop propensity of the retrospective compounds **1-5**.

(a) PCC for crystal structures to docking predictions are shown. For *m*=1, the coefficient is 0.774, whereas for *m*=2 the coefficient is 0.841. The maximum coefficient is around *m* = 1.7, 0.883. Significant p-values less than 1% are highlighted in black.

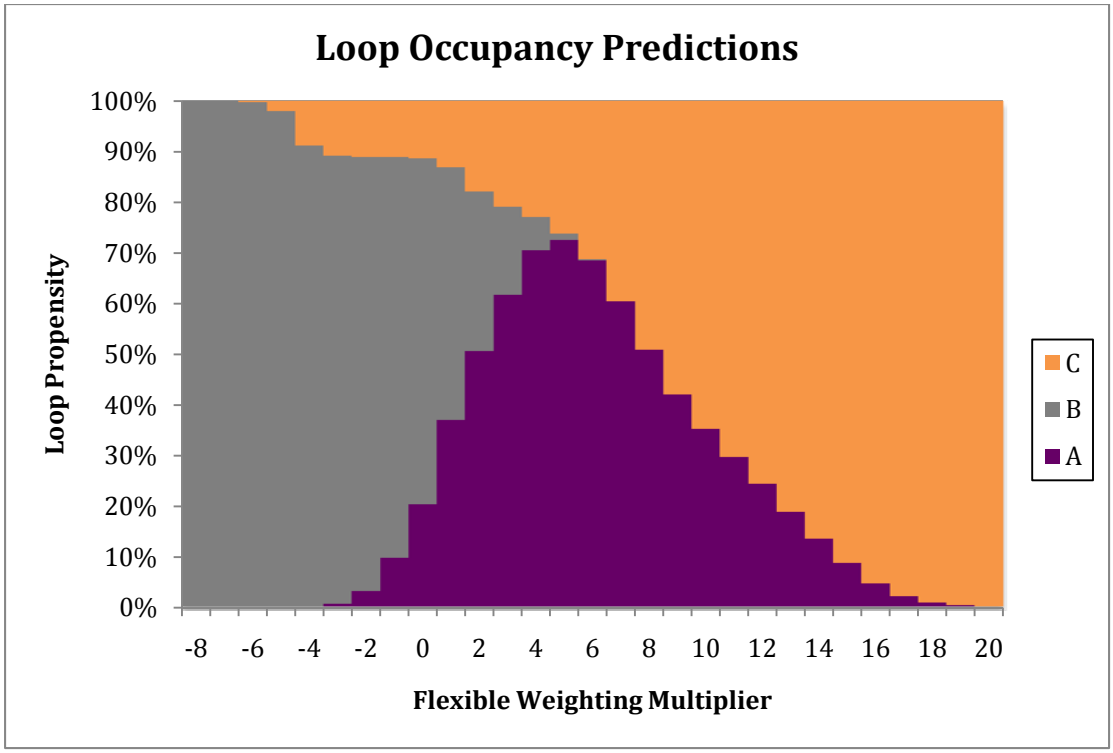
(b) PCC for structures are shown, varying the docking weight multiplier *d* from 0.1 to 3. Significant p-values less than 1% are highlighted.

(c) PCC for structures are shown, varying the input loop *B* occupancy from 0.000001 to 0.15, shown in log scale. Significant p-values less than 1% are highlighted. The maximum PCC of 0.901 is at 0.013 occupancy, at the 0.04 occupancy, the PCC is 0.772.

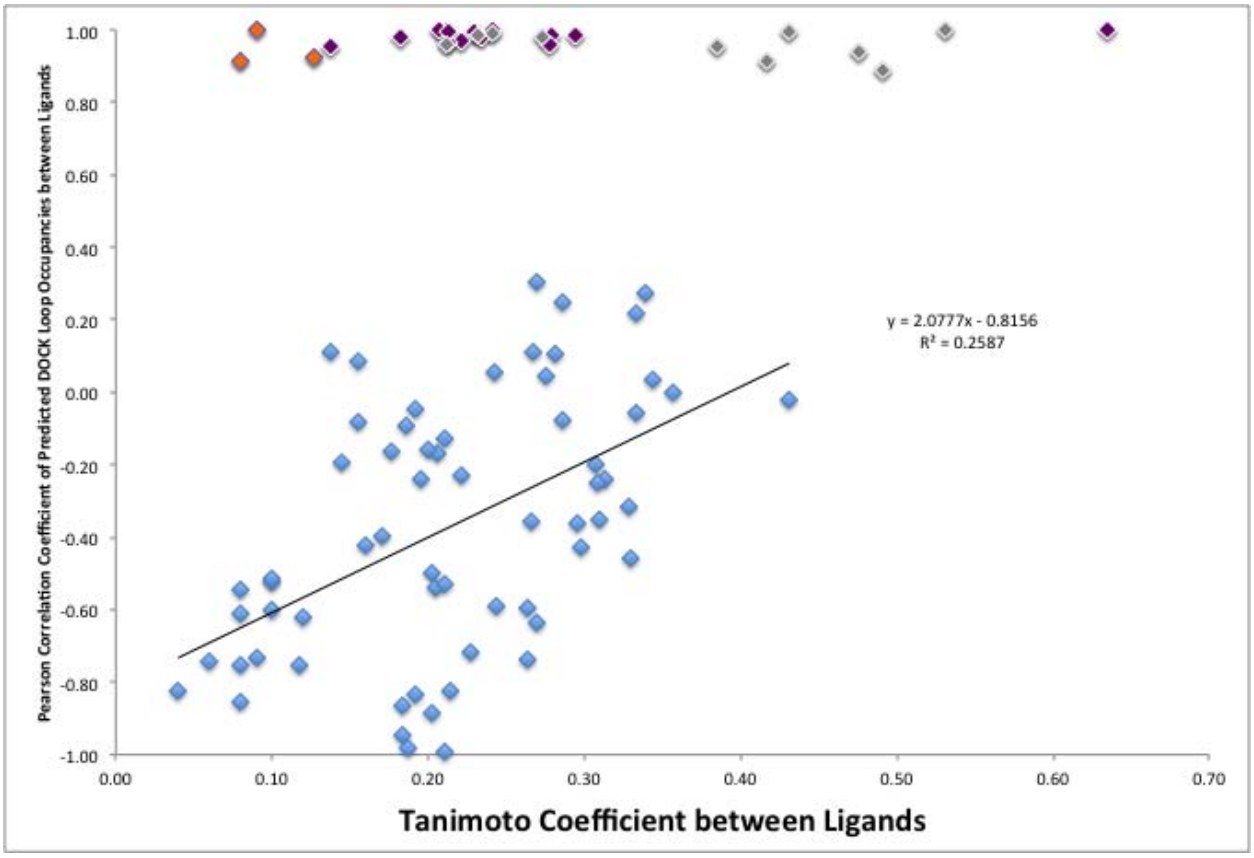


Suppl. Figure 7 Analysis of side-chain predictions vs. experimental data.

Green ‘check’ means correct; (‘check' in brackets) means predicted pose is plausibly the highest occupancy one but not exclusively; yellow ‘?’ means unable to determine pose as no density is present (high degree of positional uncertainty); red ‘X’ means failed prediction. Only two failures are obvious: the N193 F position of compound **13** and M228 for compound **9**. Note the difficulty in predicting the alternative state of N193 for compound **14** right, this is a noteworthy success.



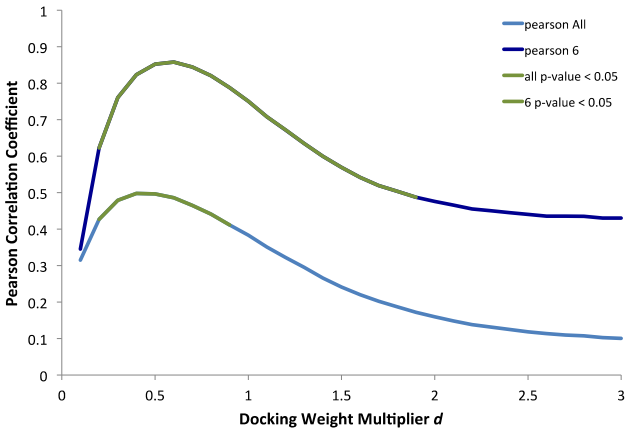
Suppl. Figure 8 Sum over all loop states the prospective ligands predict at each flexible weight multiplier.



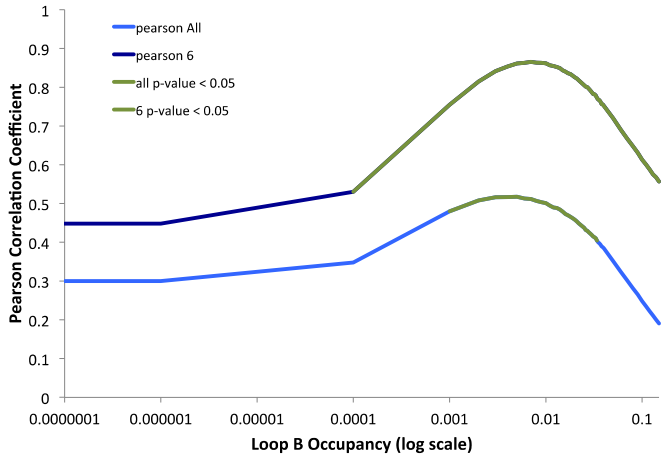
Suppl. Figure 9 A comparison of the chemical similarity of the 14 ligands from this paper and the correlation of their loop occupancies.

Data points near 1.0 PCC are colored according to the dominant loop pose of that ligand (*A*=purple, *B*=grey, *C*=orange). Note, that many dissimilar ligands by chemical similarity bind to the same major loop conformations (judged by occupancy distribution). There is a weak trend of increasing correlation with increasing Tanimoto. The R^2 for only the bottom cloud of blue points is 0.26.

A



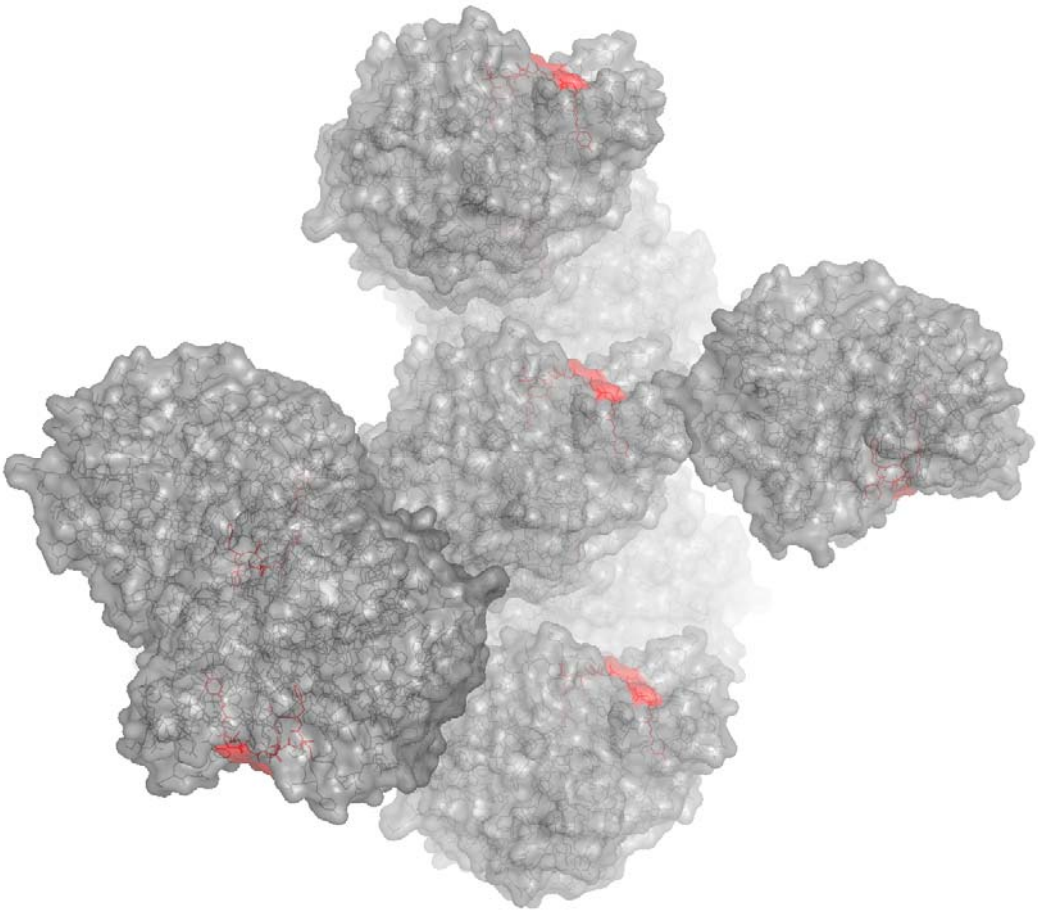
B



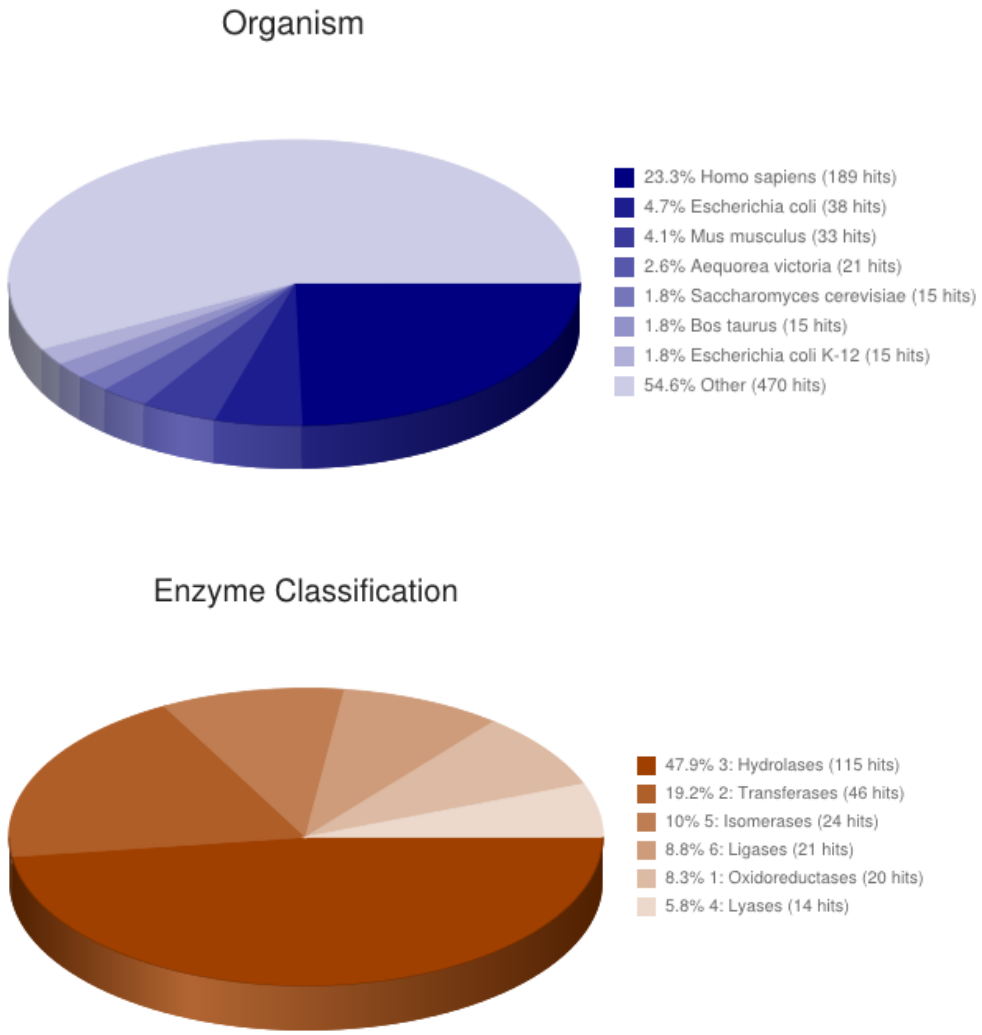
Suppl. Figure 10 Response of PCC between the crystallographic occupancy and the predicted DOCK loop propensity of the prospective compounds **6-14.**

(a) Coefficients for all 9 prospective compounds shown, along with the 6 without MES or a 4th loop, as in Figure 5. Significant p-values are highlighted. Flexible weighting multiplier *m*=1 was used here.

(b) Coefficients for all 9 prospective compounds as well as just the 6 without MES or a fourth loop are shown, varying the input loop B occupancy from 0.000001 to 0.15, shown in log scale. Significant p-values less than 5% are highlighted. Flexible weighting multiplier *m*=1 was used here.

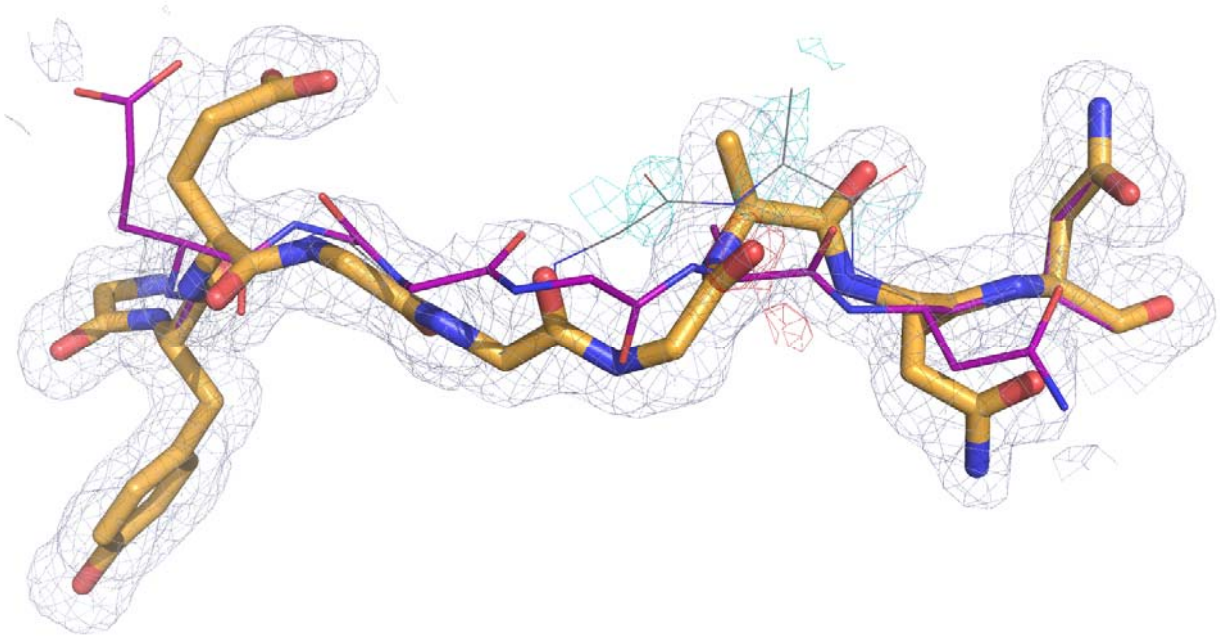


Suppl. Figure 11 Loop movement is unobstructed by crystal contacts.
The loop region (red) can move freely and therefore occupancies, used to derive energy penalties, are unbiased from crystal packing artifacts and reflective of the energies of the loop states within the crystal environment. Loop movement is also required to allow ligand access to the binding site. This is exemplified by the apo-RT structure and its crystallographic symmetry mates but the same packing occurs for all other complexes crystallized in the same space group.



Suppl. Figure 12 High-resolution apo structures in the PDB, susceptible to multi-conformer analysis, fall into multiple families (only the enzymes are shown here).

After performing the query search on the pdb website there are several useful, clickable options available that can be used to follow links of special interest. A separate file including a table with PDB-ID, CATH/ SCOP/ PFAM classification, paper title, macromolecule name, source, taxonomy id, biological process, molecular function, EC number and collection temperature (whenever available) can be downloaded for further inspection.



Suppl. Figure 13 Same Figure as Figure 1C but including the *A* loop (in addition to *C* loop) in refinement. Consequently the difference features (cyan) for the minor *B* loop excluded from refinement become more pronounced.

Supplementary references

1 Shelley, J. *et al.* Epik: a software program for pKa prediction and protonation state generation for drug-like molecules. *Journal of Computer-Aided Molecular Design* **21**, 681-691, (2007).

2 Barelier, S. *et al.* Roles for ordered and bulk solvent in ligand recognition and docking in two related cavities. *PLOS ONE*, (2013).

3 Meng, E. C., Shoichet, B. & Kuntz, I. D. Automated Docking with Grid-Based Energy Evaluation. *J. Comp. Chem.* **13**, 505-524, (1992).

4 Sharp, K. A. Polyelectrolyte electrostatics: Salt dependence, entropic, and enthalpic contributions to free energy in the nonlinear Poisson–Boltzmann model. *Biopolymers* **36**, 227-243, (1995).

5 Gallagher, K. & Sharp, K. Electrostatic Contributions to Heat Capacity Changes of DNA-Ligand Binding. *Biophysical Journal* **75**, 769-776, (1998).

6 Mysinger, M. M. & Shoichet, B. K. Rapid Context-Dependent Ligand Desolvation in Molecular Docking. *J. Chem. Inf. Model.* **50**, 1561-1573, (2010).

7 Ytreberg, F. M. & Zuckerman, D. M. A black-box re-weighting analysis can correct flawed simulation data. *Proceedings of the National Academy of Sciences* **105**, 7982-7987, (2008).

8 Wei, B. Q., Baase, W. A., Weaver, L. H., Matthews, B. W. & Shoichet, B. K. A Model Binding Site for Testing Scoring Functions in Molecular Docking. *J Mol Biol* **322**, 339-355, (2002).

9 Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J. & Shoichet, B. K. Ligand Pose and Orientational Sampling in Molecular Docking. *PLoS ONE* **In Press**, (2013).

10 Wei, B. Q., Weaver, L. H., Ferrari, A. M., Matthews, B. W. & Shoichet, B. K. Testing a flexible-receptor docking algorithm in a model binding site. *J Mol Biol* **337**, 1161-1182, (2004).

11 Brozell, S. *et al.* Evaluation of DOCK 6 as a pose generation and database enrichment tool. *Journal of Computer-Aided Molecular Design* **26**, 749-773, (2012).

12 Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J Chem Inf Model*, (2012).

13 Liao, C. & Nicklaus, M. C. Comparison of Nine Programs Predicting pKa Values of Pharmaceutical Substances. *Journal of Chemical Information and Modeling* **49**, 2801-2812, (2009).

14 Rocklin, G. J. *et al.* Blind prediction of charged ligand binding affinities in a model binding site. *J Mol Biol*, (2013).

15 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190, (2010).

16 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674, (2007).

17 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D* **67**, 355-367, (2011).

18 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **60**, 2126-2132, (2004).

19 Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proceedings of the National Academy of Sciences* **108**, 16247-16252, (2011).

20 Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669-673, (2009).

21 Minasov, G., Wang, X. & Shoichet, B. K. An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. *J Am Chem Soc* **124**, 5333-5340, (2002).

22 Chen, Y., Minasov, G., Roth, T. A., Prati, F. & Shoichet, B. K. The deacylation mechanism of AmpC beta-lactamase at ultrahigh resolution. *J Am Chem Soc* **128**, 2970-2976, (2006).