

# Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery

Marcus Fischer<sup>1,2†</sup>, Ryan G. Coleman<sup>1†</sup>, James S. Fraser<sup>3\*</sup> and Brian K. Shoichet<sup>1,2\*</sup>

**Proteins fluctuate between alternative conformations, which presents a challenge for ligand discovery because such flexibility is difficult to treat computationally owing to problems with conformational sampling and energy weighting. Here we describe a flexible docking method that samples and weights protein conformations using experimentally derived conformations as a guide. The crystallographically refined occupancies of these conformations, which are observable in an apo receptor structure, define energy penalties for docking. In a large prospective library screen, we identified new ligands that target specific receptor conformations of a cavity in cytochrome c peroxidase, and we confirm both ligand pose and associated receptor conformation predictions by crystallography. The inclusion of receptor flexibility led to ligands with new chemotypes and physical properties. By exploiting experimental measures of loop and side-chain flexibility, this method can be extended to the discovery of new ligands for hundreds of targets in the Protein Data Bank for which similar experimental information is available.**

In their native states proteins fluctuate among multiple conformations, and recent evidence from NMR spectroscopy<sup>1,2</sup> and crystallography<sup>3–7</sup> suggests apo proteins may transiently populate the conformations adopted in ligand complexes. It is tempting to wonder whether these conformations may be used prospectively to address two long-standing problems in exploiting protein flexibility in ligand discovery<sup>8</sup>: sampling protein states and weighting these states relative to one another<sup>9,10</sup>.

Sampling protein conformations for ligand discovery is challenging because of the many degrees of freedom available to folded proteins. Conformational changes often involve not only rotamer transitions, but also coordinated loop and main-chain movements. The different internal energies of these conformations affect ligand-binding affinity and, if unaccounted for, high-energy decoy conformations may dominate the docking.

Two strategies have been introduced to model protein flexibility in docking screens for new ligands. ‘Soft docking’<sup>11</sup> reduces the steric component of the scoring function and can identify ligands that might be accommodated by certain protein rearrangements. This, however, can increase docking false positives<sup>9</sup>. A related method averages several structures to represent multiple conformations<sup>12</sup>. This also reduces the number of states, but suffers from an unphysical averaging of energies, which reduces predictive success<sup>12</sup>.

A second strategy explicitly represents, and docks into, multiple receptor conformations<sup>13–16</sup>. These conformations may be sampled in different ligand complexes<sup>12,17–21</sup> or calculated using molecular dynamics<sup>22–26</sup>, elastic network models and related techniques<sup>27</sup>. Whereas the restriction to experimentally determined conformations ensures accessible states, it limits their number and remains biased to known structures. To calculate alternative conformations from simulations escapes such biases, but struggles to access states separated by barriers of higher energy. Neither approach easily assigns energy penalties to the different

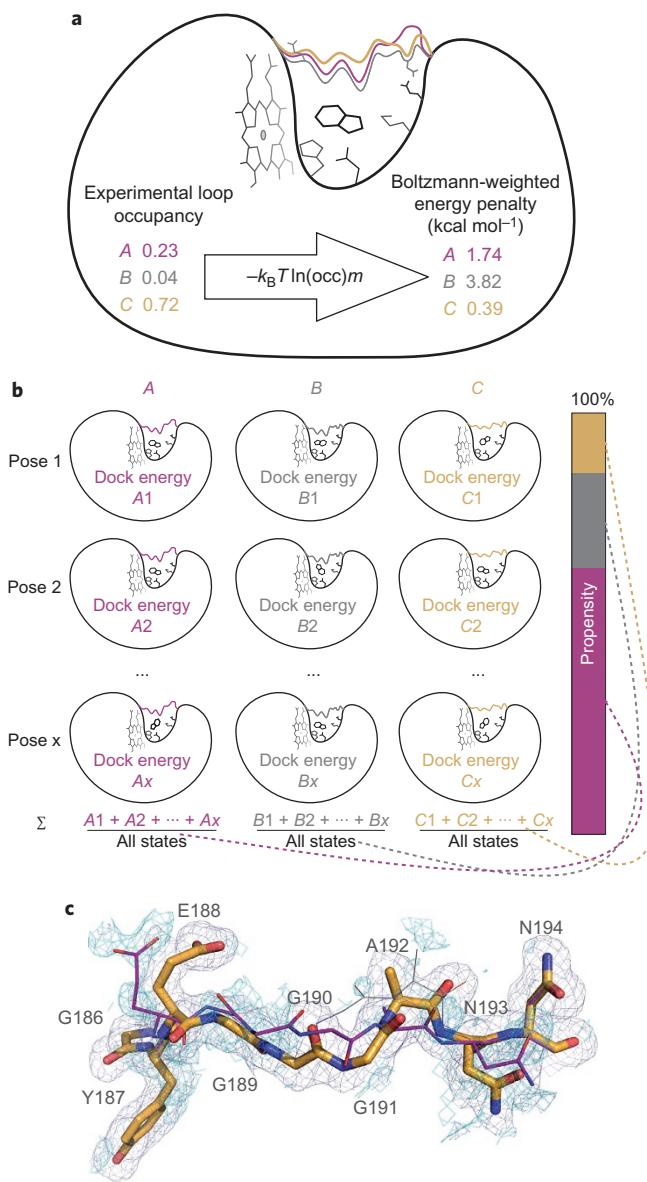
conformations, and several studies have found that using too many conformations in flexible docking can reduce the enrichment of known ligands over decoy molecules<sup>9,28–32</sup>.

Recent advances in crystallographic refinement offer the opportunity to model higher-energy conformational states using direct experimental observations<sup>3–5,33–35</sup>. Such alternative conformations can be discovered in weak electron-density features and reliably modelled at lower occupancies than the dominant conformation<sup>3,34,35</sup>. A liability of this approach is its inability to identify confidently the conformations present at less than ~10% of the ground state, or no more than about 2 kcal mol<sup>-1</sup> higher in energy at room temperature. It can represent coordinated transitions as easily as changes in side-chain rotamers, and the relative conformational energies emerge directly from crystallographic occupancies.

Here we explore the use of multiple conformations present in the electron-density map of an apo cavity site in cytochrome c peroxidase (CcP)<sup>36–39</sup> in docking screens. The substitution Trp191 → Gly in CcP creates an enclosed anionic cavity of about 200 Å<sup>3</sup>, which has been studied as a model site for ligand binding. In the variant studied here, residues 192–193 have been deleted, and the substitution Pro190 → Gly introduced, which increases the flexibility of the cavity’s gating loop. In the room-temperature structure that we determined to 1.57 Å resolution, one loop and three side chains of this ‘gateless’ cavity occupy multiple states in the electron density. We docked 583,363 compounds against 16 energy-weighted conformations of the cavity. To limit the calculation cost, we modified the treatment of ligand–protein electrostatic interaction energies to decompose them into an additive function<sup>40</sup>. This allowed us to sample 16 receptor states with only a 2.4-fold speed cost compared with that of a single structure. From the flexible docking screen, 15 new compounds were chosen to test, and ten of these were confirmed to bind. The crystal structures of nine of them were determined, which allowed us to compare predicted and

<sup>1</sup>Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California 94158, USA, <sup>2</sup>Faculty of Pharmacy, Donnelly Center, University of Toronto, 160 College St, Toronto, Ontario M5S 3E1 Canada, <sup>3</sup>Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California 94158, USA; <sup>†</sup>These authors contributed equally to this work.

\*e-mail: [james.fraser@ucsf.edu](mailto:james.fraser@ucsf.edu); [bshoichet@gmail.com](mailto:bshoichet@gmail.com)



**Figure 1 | Experimental occupancies of apo loop conformations set the penalties for docking.** **a**, From experimental loop occupancies to docking penalties. Flexible loop (in colours) and side-chain conformations of the apo CcP gateless protein are assigned Boltzmann-weighted energy penalties based on their crystallographic occupancy (here  $m = 2$ ). **b**, From docking energies to loop propensities. The Boltzmann sum of the energies of all  $x$  poses for a ligand to different loops A, B and C are calculated. The result is expressed as a percentage, which indicates the predicted preference of the ligand to bind to a particular loop conformation and can be compared to the experimental occupancies. **c**, Electron density shows evidence for three conformations of the apo loop, with the missing conformation of loops A (purple sticks) and B (grey lines) when only loop C (orange sticks) is included in the refinement shown as blue ( $2\text{mFo}-\text{DFc}$ ,  $1\sigma$ ) and cyan ( $\text{Fo}-\text{Fc}$ ,  $+1.5\sigma$ ). The stick radius is according to relative occupancies (see Fig. 2a). See Supplementary Fig. 13 for a more-pronounced difference cyan density for loop B when including A in addition to C in the refinement.

observed ligand poses and loop structures. The potentials for a broad application of this method are considered.

## Results

### From crystallographic occupancies to Boltzmann-weighted energy penalties.

Our first goal was to convert crystallographic

occupancies for the flexible 186–194 loop into Boltzmann-weighted energy penalties for docking (Fig. 1 and Supplementary Fig. 2). This loop adopts three conformations in previously determined ligand-bound structures<sup>39</sup>. These three states, which we designate as A, B and C, were combined into a multiconformer loop model. Occupancy refinement of these conformations, using the apo structure electron density, improved the agreement between the model and experimental data as judged by  $R_{\text{free}}$  values that ranged from 0.1639 to 0.1695 (Supplementary Table 2) and qualitatively fit the density (Fig. 1c).

We used the refined apo occupancies of each loop conformation to assign energy penalties to each conformation (see Fig. 1) using equation (1):

$$\text{energy penalty (conformation } A) = -k_B T \ln(\text{occ}(A))m \quad (1)$$

with  $k_B$  = Boltzmann constant,  $T$  = temperature (K),  $\text{occ}$  = occupancy,  $m$  = flexible weighting multiplier (see below). The occupancy of loop B dropped below 10% (Fig. 1), which we consider the imprecision of the refinement approach. Although refinement parameters can affect the occupancy, our procedure converged to 4% after ten refinement cycles, remained stable thereafter (Supplementary Fig. 2) and could be reproduced from another dataset (Supplementary Fig. 3). Also, the loop can move freely and is unobstructed by crystal contacts (Supplementary Fig. 11). In the Supplementary Information, we address the robustness and dependence of the results on the exact numerical occupancy value.

**Retrospective testing and integration of conformational weights and docking scores.** To test the usefulness of these energy penalties, we retrospectively docked five known cavity ligands with the new scoring function (compounds 1–5 (Table 1)) and recovered experimental poses for three of them with an average root-mean-square deviation (RMSD) of 0.4 Å (Supplementary Fig. 4).

A more stringent test compares not only the docked and observed ligand poses, but also the predicted and observed ensemble of protein conformations associated with each pose. If the energy weightings of the apo loop conformations are correct, we can combine them with docking scores to predict the distribution of conformations favoured for each ligand complex. Predicted loop propensities (analogous to experimental occupancies) were calculated, using equation (2), as the Boltzmann sum of the energy of all ligand poses bound to a specific loop conformation  $X = (A, B \text{ or } C)$  over the Boltzmann sum of the energy of all the poses generated to any loop (Supplementary Methods and Supplementary Table 4):

$$\text{propensity}_{\text{ligand } Z, \text{loop } X} = \frac{\sum_{x \in \text{loop } X} e^{\text{dock energy}(x, Z) / kT}}{\sum_{y \in \text{all states}} e^{\text{dock energy}(y, Z) / kT}} \quad (2)$$

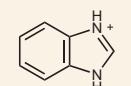
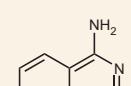
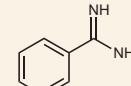
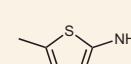
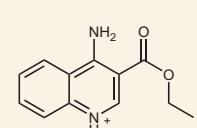
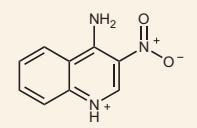
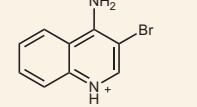
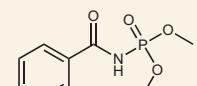
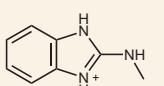
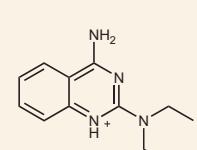
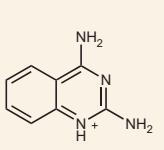
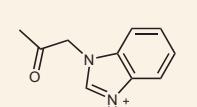
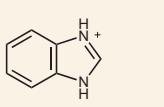
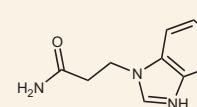
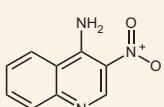
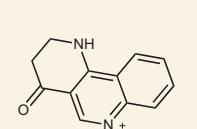
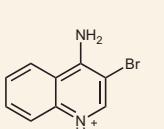
Docking energies are computed by DOCK3.7<sup>41</sup> according to equation (3), which integrates the receptor energy penalties from equation (1):

$$\begin{aligned} \text{dock energy (loop } X, \text{ ligand } Z) &= \text{energy penalty(loop } X) \\ &+ \sum_{\text{atom } z \in \text{ligand } z} V\text{dw}(z) + \text{elstat}(z) + \text{ligand desol}(z) \end{aligned} \quad (3)$$

Here the energy penalty is from equation (1),  $V\text{dw}(z)$  is the van der Waals energy of each ligand atom<sup>42</sup>,  $\text{elstat}(z)$  is the corresponding electrostatics energy<sup>43</sup> and ligand desol(z) is the ligand desolvation energy. The resulting propensities are expressed as a percentage, with all propensities for a ligand summing to 100% (Fig. 1b).

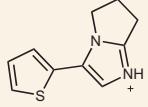
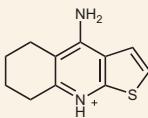
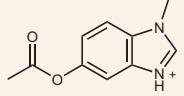
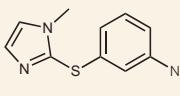
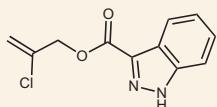
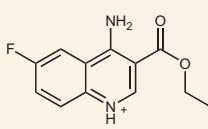
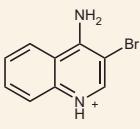
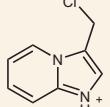
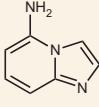
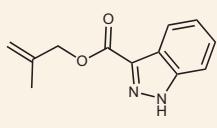
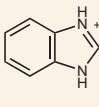
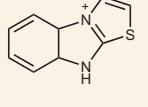
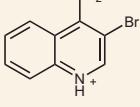
With the predicted loop occupancies for the holo complexes calculated from the apo state docking and loop propensities, we refined the observed loop occupancies against five ligand-complex datasets, determined to between 1.2 and 1.7 Å (Supplementary Table 1). This

**Table 1 | CcP ligands: previously discovered (1–5) and found by flexible docking (6–20).**

Rank	ZINC	No.	Compound	Ligand RMSD (Å)*	Loop state† Xtal / DOCK (PCC‡)	Affinity (μM) / ligand efficiency (kcal mol <sup>-1</sup> hac <sup>-1</sup> )	Max $T_c$ to known ligands§	Closest known ligand§	Closest known affinity   (μM)
n.a.	01583444	1		n.a.	C (n.a.)	n.d.	1	Is known	n.a
n.a.	00331902	2		n.a.	A (n.a.)	33   (0.68)	1	Is known	n.a
n.a.	00331945	3		n.a.	AB (n.a.)	3   (0.69)	1	Is known	n.a
n.a.	00036634	4		n.a.	A (n.a.)	9   (0.77)	1	Is known	n.a
n.a.	08652421	5		n.a.	B (n.a.)	106   (0.78)	1	Is known	n.a
8	06656163	6		0.88	C / B (-0.78)	71 ± 10 (0.35)	0.48		19 (ref. 37)
38	04962659	7		0.48	B / B (0.85)	8   (0.35) (ref. 37)	1	Was known* (ref. 37)	8 (ref. 37)
70	00331160	8		0.48	B / B (1.00)	19   (0.54) (ref. 37)	1	Was known* (ref. 37)	19 (ref. 37)
163	13739037	9		0.98	C / C (0.88)	n.b.d. <4 mM	0.29		288 (ref. 37)
322	01596053	10		0.79	AC / A (0.41)	22 ± 10 (0.4)	0.53		41 (ref. 39)
330	34979991	11		1.19	A / A (0.99)	46 ± 8 (0.46)	0.28		33 (ref. 39)
433	00203341	12		1.49	A / A (0.62)	28 ± 7 (0.44)	0.30		8 (ref. 37)
526	00519712	13		0.54	AC / A (0.61)	7 ± 0.6 (0.47)	0.36		19 (ref. 37)

Continued

**Table 1 | (continued)**

Rank	ZINC	No.	Compound	Ligand RMSD (Å)*	Loop state† Xtal / DOCK (PCC‡)	Affinity (μM) / ligand efficiency (kcal mol⁻¹ hac⁻¹)	Max $T_c$ to known ligands§	Closest known ligand§	Closest known affinity   (μM)
556	00388812	14		0.55	B/C (0.49)	23 ± 4 (0.49)	0.23		64 (ref. 37)
67	70974074	15		n.a.	/C	n.r.	0.27		203 (ref. 37)
84	16210033	16		n.a.	/C	n.b.d.	0.31		33 (ref. 39)
263	19702757	17		n.a.	/B	n.d.	0.35		19 (ref. 37)
355	12546268	18		n.a.	/A	396 ± 42 (0.42)	0.37		14 (ref. 37)
401	16207704	19		n.a.	/C	n.b.d.	0.31		33 (ref. 39)
487	01648614	20		n.a.	/A	n.r.	0.20		19 (ref. 37)

\*Ligand heavy atom RMSD between docked pose and crystal pose. †Dominant loop state listed unless two states within 10% of maximum, when both are listed. ‡PCC of loop occupancies versus predicted loop propensities from docking. §Extended connectivity fingerprints (ECFP4), ligand with this Tanimoto shown at the right. ||Determined previously; references in parenthesis.  
n.a., not applicable; n.b.d., no binding detectable (because precipitation >500 μM if not specified otherwise); n.d., not determined (insufficient solubility); n.r., not reproducible in repetitive runs.

yields a distribution of loop conformations for complexes (Supplementary Fig. 5). Overall, conformation-weighted docking correctly predicted the dominant loop state for all five ligands (Supplementary Fig. 5), with an overall Pearson correlation coefficient (PCC) of 0.77 (for three loop states for all five ligands) (Supplementary Fig. 6).

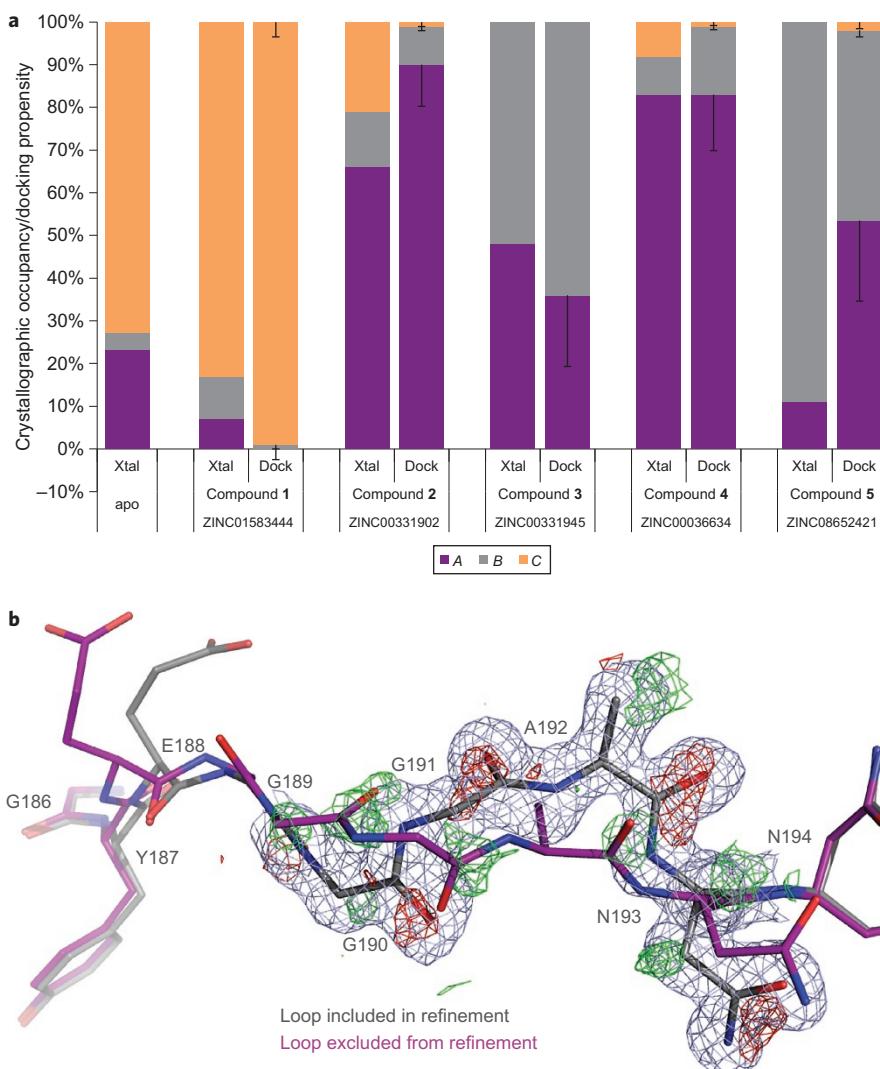
In calculating the propensity of the loop conformation for the individual ligand complexes, we add the ligand docking-energy score to the occupancy-based loop energies (equation (3)). As these loop energies are rarely on the same scale as the docking score, the combination of the two terms can be optimized. We investigated weighting the conformation energies ( $m$  in equation (1)). To reduce the dangers of overfitting, we established retrospectively a single-variable weighting term ( $m$ ) of 2 (performance was judged by the increase in statistical significance over other integers (Fig. 2, Supplementary Figs 5 and 6a, and Supplementary Methods).

As experimental occupancies can be imprecise and covary with B-factors, which model the fall-off of the density from the mean position, we investigated how the results depended on refined loop occupancies (see Supplementary Methods). The effect on the retrospective propensities was minor and correlations between the predictions and experiments remained significant (Supplementary Figs 6c and 10b). This gives confidence that the energy penalization is not overly sensitive to the input occupancy of low-occupancy states like

loop  $B$ , which is comforting in terms of the expected error in the determination and refinement of experimental occupancies.

**Prospective docking for new ligands that complement the different receptor conformations.** Fortified by these results, we used this energy-penalized ensemble of flexible states for the prospective docking of 583,363 fragments from the ZINC database<sup>44</sup>. From the top 0.1% of the highest-ranking molecules, 15 were chosen for experimental testing (compounds 6–20 (Table 1)). As is common in selecting docked molecules to test experimentally from the docking hit list<sup>45,46</sup>, we eliminated compounds that had problems with protonation or tautomerization states, and selected several molecules for chemical novelty, including some uncharged molecules. We particularly sought molecules predicted to bind to different protein–receptor conformations.

On testing, nine of the 15 compounds had  $K_d$  values between 7 and 400 μM as measured by the haem Soret-band shift (Table 1). The ligand efficiencies were between 0.35 and 0.54 kcal mol<sup>-1</sup> per heavy atom count (hac). For eight of these nine molecules we determined the X-ray crystal structures, and we also determined a structure for a tenth molecule for which we had been unable to determine an affinity. The nine ligand crystal structures recapitulated the predicted docked ligand poses (Fig. 3), with a mean RMSD of 0.82 Å (Table 1). Counting ligand binding by affinity

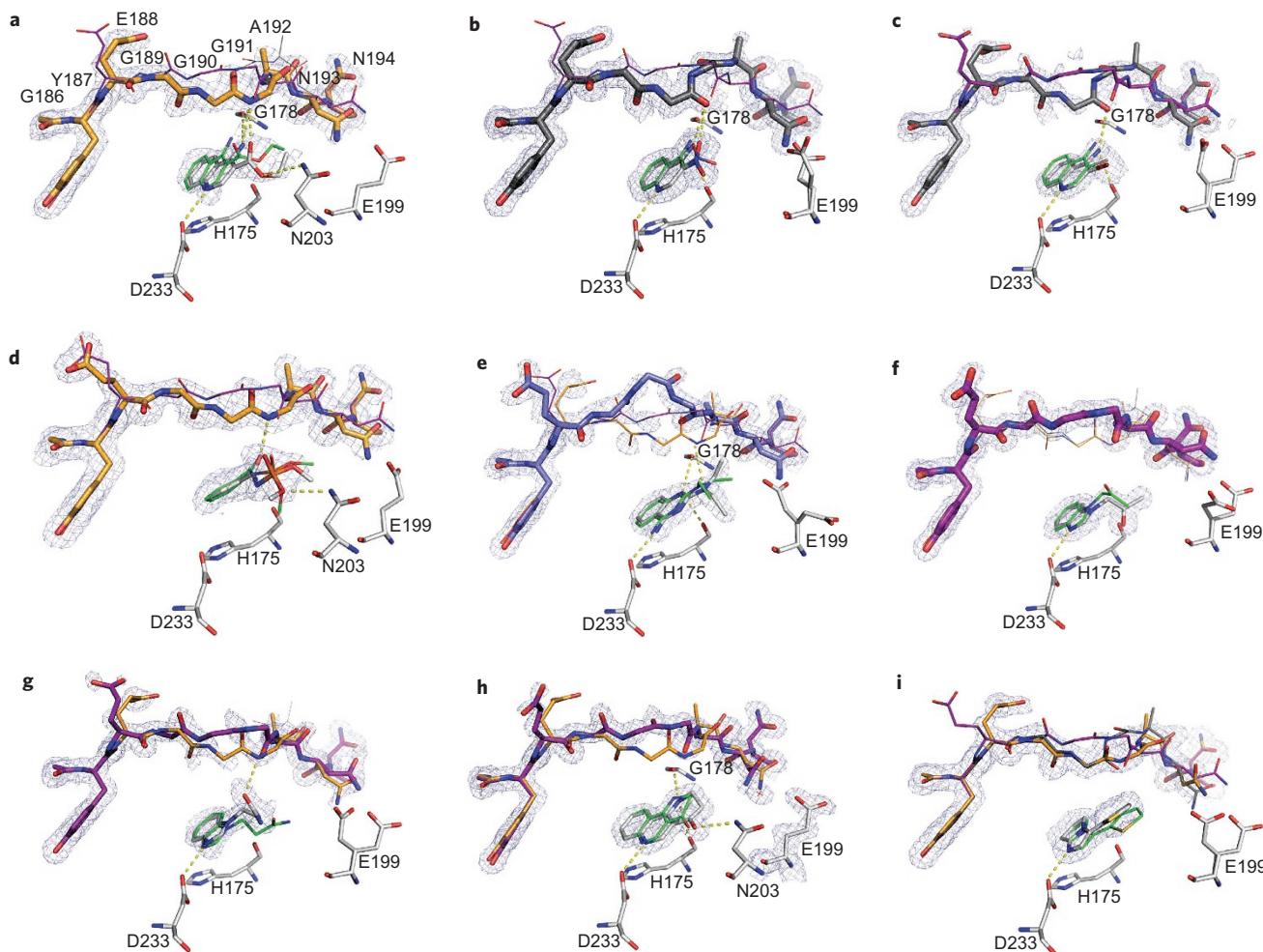


**Figure 2 | Predicting loop occupancies in holo complexes.** **a**, Experimental occupancies of the three flexible loop (186–194) conformations A, B and C are depicted as a percentage (Xtal). These can then be compared easily to their predicted docking propensities (dock) for compounds **1–5**<sup>44</sup>. Using Boltzmann-energy penalties and a multiplier  $m = 2$  results in close agreement of the major loop conformation between prediction (dock) and experiment (Xtal), with a PCC of 0.83 and a  $P$  value  $<0.01$ . Error bars for docking propensities are derived by using any flexible weighting multiplier  $m$  between 1.0 and 3.0 and taking the standard deviation. The error bars are symmetric in both the up and down direction; the up direction has been omitted for clarity. **b**, The difference electron-density map of CcP and compound **5** around the backbone carbonyls (red mesh for the main loop in grey sticks and green for the new loop conformation) provides evidence for the presence of a second loop conformation (purple) even at very low levels of around 10%. Resolution is 1.2 Å; 2mFo-DFc map (blue mesh) rendered at  $1\sigma$  and mFo-DFc (red and green mesh) map at  $2.6\sigma$ .

and by crystallography, the total hit rate was 67% (10/15). The new hits differed from previously known ligands: the highest pairwise topological similarity to these, using ECFP4-based Tanimoto<sup>47</sup> coefficients, was 0.36 (Supplementary Table 9), and new ligands were, on average, 52 Da heavier (from 148 to 200 Da) than the previously known ligands, and also larger than those discovered in an earlier rigid-body docking study<sup>37</sup>. A role of human selection of compounds, in both the previous docking studies and in this one, cannot be entirely controlled for.

When writing this manuscript we discovered that two of the 15 molecules (**7** and **8**) had been found independently in work that was then unpublished<sup>37</sup>; we do not count these as novel molecules. They do, nevertheless, illustrate the strength of the method. Of all the new ligands, **7** and **8** most closely resemble the earlier series of cavity ligands and, indeed, were predicted and observed to prefer the *B* loop conformation, which has been previously targeted in the older rigid-docking method.

**Ligands select the predicted protein loop and side-chain conformations.** A crucial point is the ability of the method to anticipate the protein conformational response to the new ligands. In seven of nine new holo structures, the predicted loop occupancies and conformations corresponded qualitatively to the observed ones, with at least the dominant loop being predicted correctly and frequently to the approximate ratios of the ensemble (Figs 3 and 4). For instance, compound **7** was experimentally observed to bind to loop *B* at 79% (with 20% occupancy of loop *A*), and was predicted to prefer loop *B* at 95% with a 5% loop *A* contribution. For compound **8** the prediction of loop conformations *A* and *B* at 33% and 67%, respectively, agrees quantitatively with the observed holo occupancies at 38% and 62%, respectively. The experimental occupancy of the *C* loop in the complex with compound **9** was 68% instead of the 100% predicted. Compound **11** bound primarily to loop *A* with an occupancy of 84% and the remaining occupancy was split



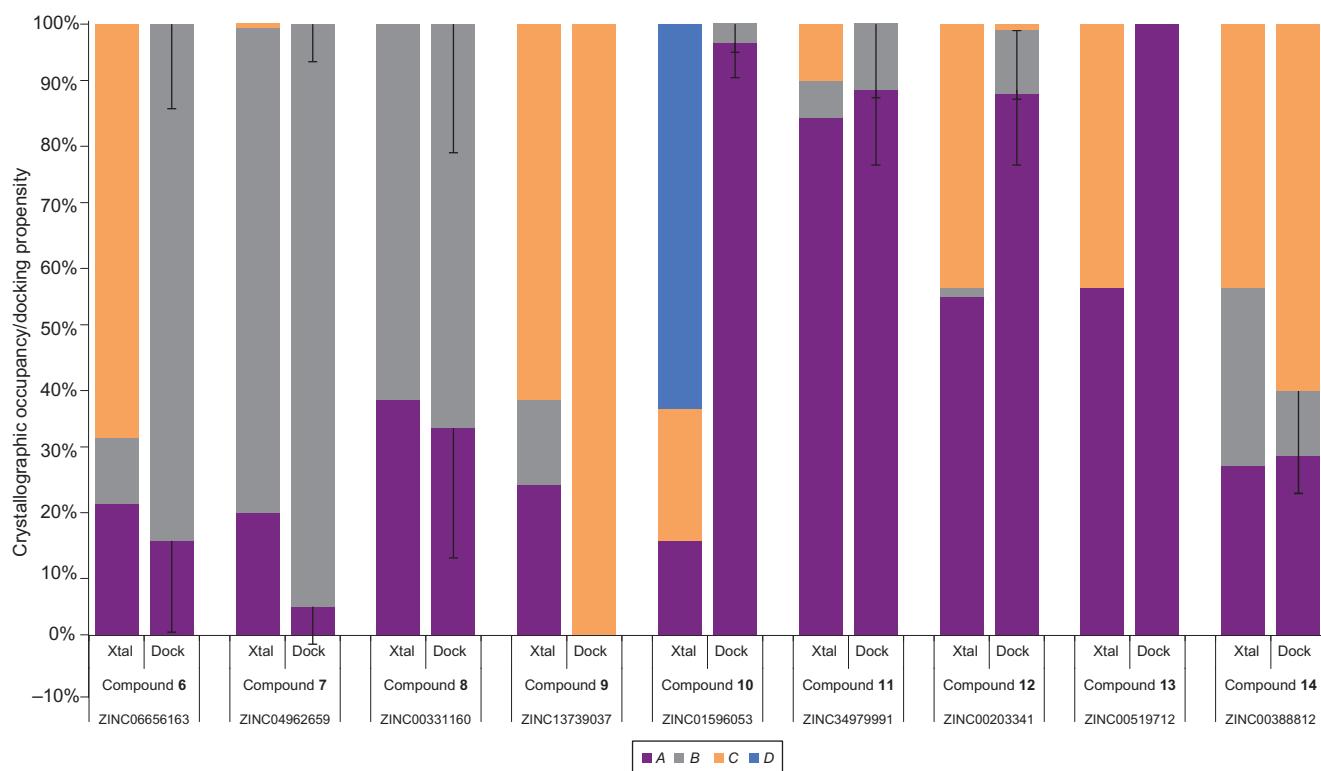
**Figure 3 | Experimental binding poses versus prospective docking predictions.** The electron density at  $1\sigma$  is shown for ligand and loop conformations.

The loop-stick thickness corresponds to experimentally observed occupancies; the colouring is as before with loop conformation A in purple, B in grey, C in orange and D (for compound 10) in blue. **a–i**, Superposition of the ligand poses, with experimental in grey versus docked in green: compound **6** (**a**), **7** (**b**), **8** (**c**), **9** (**d**), **10** (**e**), **11** (**f**), **12** (**g**), **13** (**h**) and **14** (**i**). For clarity, co-crystallized MES for compounds **6**, **9** and **12** is omitted.

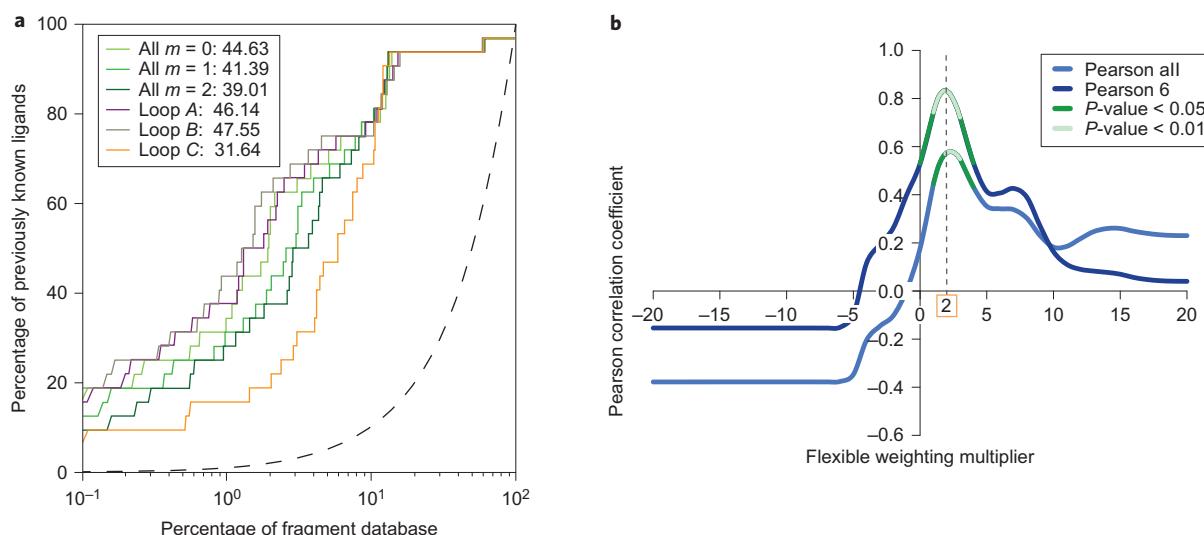
between the additional loops; the prediction was 89% for loop A, which corresponds to a PCC of 0.99. Compound 13 was among the few compounds predicted to bind to one loop conformation exclusively, the A loop. Whereas automatic occupancy refinement suggests a presence of loop C, this may be misled by nearby water molecules, as visual inspection of the electron density seems consistent with only the single A loop being present (Fig. 3). Compound 14 was chosen to bind to an ensemble of loop conformations: 29% loop A, 11% loop B and 60% loop C. The refined experimental occupancy values were consistent with these predictions, with 27% loop A, 30% loop B and 43% loop C, which correspond to a PCC of 0.84. For ligand 10, a fourth loop conformation was found that had not been modelled previously. To check for this conformation, D, in the apo structure, it was included as a fourth loop in the original model and refined as before, but could not be observed at any reliable level; this represents a false negative of our method (also see Supplementary Fig. 1 and Supplementary Table 7). Meanwhile, the prediction of side-chain conformations conformed to those observed crystallographically with only two failures out of the total 27 modelled conformations (Supplementary Methods and Supplementary Fig. 7). Finally, for several complexes the presence of 2-(*N*-morpholino)ethanesulfonic acid (MES), which itself is a weak ligand, prevented a full analysis of the results (Supplementary Table 8).

**Overcoming the bias of known structures: the correct model does not necessarily result in the best retrospective enrichment.** It is important to understand whether flexible receptor docking improved the results over standard rigid docking. We first investigated the ranks of the new ligands against both our fully flexible model and any individual loop model. No single model would have ranked all these ligands in the top 0.1% of the database as the flexible docking had (Supplementary Table 5). Had we docked prospectively against all conformations and combined the top-ranking ligands, the hits against the high-energy, low-occupancy B conformation would have dominated, as in previous screens<sup>37,39</sup>, and the calculation would have taken sevenfold longer.

This bias emerges even more strongly in retrospective screens of all previously known ligands. Such retrospective enrichment is widely used to judge docking performance and to select receptor structures for prospective docking. On that basis we would have selected the B loop, which dominated enrichment plots, and discarded the other conformers (Fig. 5a and Supplementary Table 6). Choosing the best enriching structure in retrospective studies for prospective screens seems intuitive, but it is biased by the binding of most known ligands to the B loop; they were, in fact, discovered by docking to that loop conformation. Compared with ligands that we ourselves had discovered against the single B-loop conformation, the new ligands that bind to the C- and A-loop conformations are



**Figure 4 | Predicting loop occupancies in bound complexes.** The loop occupancies and predicted propensities for compounds **6–14** are shown, with crystallographic occupancies at the left and predicted loop propensities (with  $m=2$ ) at the right for each pair. Error bars for the loop propensity represent the standard deviation of the occupancies by varying the flexible weighting multiplier  $m$  from 1 to 3. The major loop conformation is predicted correctly for all cases but compound **10**, which prefers a fourth loop (*D*) conformation that was not modelled; compounds **6** and **12** have a larger *C* loop presence because of the partial presence of MES. The error bars are symmetric in both the up and down direction; the up direction has been omitted for clarity.



**Figure 5 | Enrichment alone cannot distinguish first-in-class from best-in-class.** **a**, Retrospective enrichment of known ligands against decoys (adjusted log AUC (area under the curve)) for loops *A*, *B* and *C* individually (coloured as purple, grey and orange, respectively) and combined with different multipliers (tones of green). The adjusted area under the log ROC plot (ROC = receiver operating characteristic) is shown in the key (see Supplementary Table 6 for other common performance metrics). **b**, Pearson correlation of experimental and predicted loop occupancies with statistically significant areas highlighted in green (dark green for  $P$  values  $<0.05$ , light green for  $P < 0.01$ ). Pearson all is for all nine compounds; Pearson 6 is the correlation only considering compounds **7**, **8**, **9**, **11**, **13** and **14**, for which the results are not distorted by a partial presence of MES.

more diverse and are larger (Supplementary Table 9); they would not have been discovered using the highest enriching model alone (Supplementary Table 4) or, if so, for the wrong reasons (Fig. 5b and Supplementary Fig. 8).

It is appropriate to ask whether this method (dependent as it is on experimental density features) can be used on biorelevant targets. On examining the Protein Data Bank (PDB)<sup>48</sup> for projects that our flexible docking method could be applied to, we found

827 unique proteins with electron-density maps determined to  $<1.5$  Å resolution, a level substantially more conservative than the 2 Å we estimate required for confident occupancy fitting (Supplementary Methods and Supplementary Fig. 12). Although only 51 of these were determined at room temperatures, as the apo cavity structure was, analysis of an apo cavity structure determined at cryogenic temperatures suggests that much of the flexibility exploited here remains even at these lower temperatures, but there were also important differences (described more fully in the Supporting Information). Although room-temperature structures more fully explore conformational heterogeneity present in protein structures<sup>4</sup>, probably even cryogenic structures have enough conformations to support this analysis.

## Discussion

Three principal observations emerge from this study. First, partial occupancy conformations, apparent in electron density from room-temperature crystal structures, enable the modelling of alternative protein conformations in molecular docking. These features not only illuminate conformations accessible to ligands, but their occupancies provide energy weights for the docking scoring function, which prevents domination by higher-energy conformations. Multiple conformations may be represented with only a modest impact on the docking calculation time. Second, exploitation of these conformations enables the prospective prediction of ligands with new chemotypes and new physical properties, with close correspondence between the predicted ligand poses and protein-loop conformations and those subsequently determined by X-ray crystallography. Finally, there are over 800 unique proteins in the PDB, each with the requisite density maps, to which this method could be applied today (see Supplementary Data: possible protein targets).

We were surprised at the high correspondence between the loop propensities and ligand geometries from predictions and those in the X-ray structures of the new complexes. With the exception of the complexes with compound **6** and with compound **10**, the observed loops and residue conformations matched well those predicted, as judged by their relative occupancies and the correct prediction of the major loop conformation (Figs 2 and 4). For six of the nine structures, the occupancies not only corresponded qualitatively, but did so quantitatively as well, with PCCs greater than 0.6 (compounds **7–9** and **11–13**). This suggests that experimental conformational energy weights and docking scores may be in at least a qualitative balance, and may be combined pragmatically. Indeed, the method predicts loop occupancies 30% better than a naive method, which presumes all states to be equi-energetic (Fig. 5b). To check whether the loop-propensity prediction could have been achieved using only ligand similarity, we compared the topological similarity of the 14 known ligands with the correlation of their loop occupancies. Many topologically dissimilar ligands bound to the same major loop conformations (Supplementary Fig. 9), which suggests an advantage of structure-based methods over similarity-based methods alone.

Certain caveats merit discussion. Only a narrow range of conformations above the ground state can be observed reliably in this method. Even here, the *D* conformation of the 186–194 loop, observed in the complexes of ligand **10**, was unanticipated because it was not observed in the apo structure. Whereas the conformational occupancies and the docking propensities were, overall, in balance, there is no fundamental reason why they should be in balance, or that the weighting found here should extend to other systems. The weights assigned based on the occupancies may be converted into energies, but the docking scores, even when physics-based, leave out important terms and make substantial approximations. As docking scoring functions develop to better model physical forces, these terms should come into the balance more reliably than we, perhaps fortuitously, found here.

## Conclusions

Notwithstanding the caveats discussed above, this method had important successes. Partial occupancy modelling enabled the representation of alternative, energy-weighted protein conformations that could be integrated with molecular docking scores. This prevented domination by higher-energy conformations in the docking, which might fit ligands better, but at the cost of higher internal energies. Although the number of conformational states grows exponentially, the multiconformer receptor potentials could be recombined in a way that leads to only a modest impact on the docking calculation time. Exploitation of the new conformations illuminated ligands with new chemotypes and new physical properties, and we observed a close correspondence between the predicted ligand poses and protein-loop conformations with those determined subsequently by X-ray crystallography. There are well-over 800 unique proteins to which this method could be applied today.

## Methods

The protein was purified and crystallized as described<sup>39</sup>, with the exception of the apo protein that was crystallized in 100 mM KPi, pH 6.0. A loop model was generated from three main loop conformations (residues 186–194) observed in holo complexes<sup>39</sup> with compounds **4**, **5** and apo for loops *A*, *B* and *C*, respectively, and subjected to occupancy refinement (strategy = individual\_adp + occupancies) within PHENIX.REFINE<sup>33</sup> in which ten cycles were found to result in sufficient convergence of the loop occupancy (Supplementary Fig. 2). These models are deposited at the PDB as 4NVA–4NVO and 4OQ7 (Supplementary Table 1). Experimental affinities were measured by fluorescence monitoring of the haem Soret-band shift as before<sup>37,39</sup>.

**Flexible receptor preparation.** DOCK 3.7<sup>41</sup> uses physics-based scoring that consists of van der Waals<sup>42</sup> and ligand-desolvation terms<sup>43</sup>, combined with interaction electrostatics using a probe-charge implementation of the Poisson–Boltzmann (PB) equation. The first two components of this score can be broken down independently of the atom, so the receptor can be separated into invariant and flexible parts, with separate scoring grids constructed and then used during docking. For PB electrostatics, the scoring cannot be deconstructed into separate protein components as easily. Here we use QNIFFT<sup>43,50</sup> on separate but complete receptor conformations. To use these during docking, the PB map of each receptor conformation is compared with the PB of the most-occupied receptor conformation, and the difference maps for the overall conformation are used in docking to construct the overall electrostatic score. This results in a much better approximation to the PB map of a single conformation (see Supplementary Methods). Given a structure with defined flexible regions and occupancies, docking preparation takes place automatically. Both the scripts to do so and the DOCK3.7 code itself are available, without charge, for academic research at <http://dock.compbio.ucsf.edu/DOCK3.7/>

**Flexible receptor docking.** Several changes were made to the DOCK 3.7 code to enable flexible docking. Each ligand pose is scored against each part of the receptor conformation (here, two residues with two positions each and three loops plus a loop with a residue moved, plus an invariant grid), and nine grids were scored for each ligand pose. The scores were assembled into the  $2 \times 2 \times (3+1) = 16$  possible cavity conformations and the top score for each was saved, as were the top ten overall poses to any conformation to calculate the receptor conformational propensities. Additional analyses were performed with an implementation of the black-box reweighting algorithm (BBRW)<sup>51</sup> in place of equation (2) (Supplementary Table 3). The code for equation (2) and the BBRW algorithm is included in the DOCK 3.7 distribution (<http://dock.compbio.ucsf.edu/DOCK3.7/>). For the screen of the 583,363 ZINC<sup>44</sup> fragments, flexible docking took 1,516 core hours spread across 850 nodes, or less than two hours of wall time. Docking a single cavity conformation took 630 hours, only a 2.4-fold computation cost versus a 16-fold increase in conformations sampled.

**Accession codes.** Crystal structures are available at the PDB (Supplementary Table 1). The structures have the following primary accession codes, in which the numbers in parenthesis designate the ligand that is bound to CcP, and ‘apo\_RT’ and ‘apo\_cryo’ refer to apo forms of CcP at room temperature and at cryogenic temperature: 4NVA (apo\_RT), 4NVB (5), 4NVC (4), 4NVD (3), 4NVE (2), 4NVF (1), 4NVG (6), 4NVH (7), 4NVI (8), 4NVJ (9), 4NVK (10), 4NVL (11), 4NVM (12), 4NVN (13), 4NVO (14) and 4OQ7 (apo\_cryo).

Received 16 December 2013; accepted 11 April 2014;  
published online 25 May 2014

## References

1. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nature Chem. Biol.* **5**, 808–814 (2009).

2. Koveal, D., Clarkson, M. W., Wood, T. K., Page, R. & Peti, W. Ligand binding reduces conformational flexibility in the active site of tyrosine phosphatase related to biofilm formation A (TpbA) from *Pseudomonas aeruginosa*. *J. Mol. Biol.* **425**, 2219–2231 (2013).
3. Burnley, B. T., Pavel, V. A., Paul, D. A. & Piet, G. Modelling dynamics in protein crystal structures by ensemble refinement. *eLife Sci.* **1**, e00311 (2012).
4. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).
5. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl Acad. Sci. USA* **108**, 16247–16252 (2011).
6. Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* **280**, 558–563 (1979).
7. Rauh, D., Klebe, G. & Stubbs, M. T. Understanding protein–ligand interactions: the price of protein flexibility. *J. Mol. Biol.* **335**, 1325–1341 (2004).
8. Jorgensen, W. L. Rusting of the lock and key model for protein–ligand binding. *Science* **254**, 954–955 (1991).
9. Nicholls, A. The character of molecular modeling. *J. Comput. Aid. Mol. Des.* **26**, 103–105 (2012).
10. Barril, X. & Fradera, X. Incorporating protein flexibility into docking and structure-based drug design. *Expert Opin. Drug Disc.* **1**, 335–349 (2006).
11. Jiang, F., Lin, W. & Rao, Z. SOFTDOCK: understanding of molecular recognition through a systematic docking study. *Protein Eng.* **15**, 257–263 (2002).
12. Cosconati, S. *et al.* Protein flexibility in virtual screening: the BACE-1 case study. *J. Chem. Inf. Model.* **52**, 2697–2704 (2012).
13. Rueda, M., Totrov, M. & Abagyan, R. ALiBERO: evolving a team of complementary pocket conformations rather than a single leader. *J. Chem. Inf. Model.* **52**, 2705–2714 (2012).
14. Hritz, J., de Ruiter, A. & Oostenbrink, C. Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking. *J. Med. Chem.* **51**, 7469–7477 (2008).
15. Richter, L. *et al.* Diazepam-bound GABA receptor models identify new benzodiazepine binding-site ligands. *Nature Chem. Biol.* **8**, 455–464 (2012).
16. Corbeil, C. R. & Moitessier, N. Docking ligands into flexible and solvated macromolecules. 3. Impact of input ligand conformation, protein flexibility, and water molecules on the accuracy of docking programs. *J. Chem. Inf. Model.* **49**, 997–1009 (2009).
17. 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).
18. Clauben, H., Buning, C., Rarey, M. & Lengauer, T. FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **308**, 377–395 (2001).
19. An, J. *et al.* A Novel small-molecule inhibitor of the avian influenza H5N1 virus determined through computational screening against the neuraminidase. *J. Med. Chem.* **52**, 2667–2672 (2009).
20. Brooijmans, N. & Humbel, C. Chemical space sampling by different scoring functions and crystal structures. *J. Comput. Aid. Mol. Des.* **24**, 433–447 (2010).
21. Newman, J., Dolezal, O., Fazio, V., Caradoc-Davies, T. & Peat, T. The DINGO dataset: a comprehensive set of data for the SAMPL challenge. *J. Comput. Aid. Mol. Des.* **26**, 497–503 (2012).
22. Zhong, S. *et al.* Identification and validation of human DNA ligase inhibitors using computer-aided drug design. *J. Med. Chem.* **51**, 4553–4562 (2008).
23. Amaro, R. E. *et al.* Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA* **105**, 17278–17283 (2008).
24. Cheltsov, A. V. *et al.* Vaccinia virus virulence Factor N1L is a novel promising target for antiviral therapeutic intervention. *J. Med. Chem.* **53**, 3899–3906 (2010).
25. Sato, T. *et al.* Identification of novel drug-resistant EGFR mutant inhibitors by *in silico* screening using comprehensive assessments of protein structures. *Bioorg. Med. Chem.* **20**, 3756–3767 (2012).
26. Rogers, K. E. *et al.* Novel cruzain inhibitors for the treatment of Chagas' disease. *Chem. Biol. Drug Des.* **80**, 398–405 (2012).
27. Kumar, A. & Zhang, K. J. Computational fragment-based screening using RosettaLigand: the SAMPL3 challenge. *J. Comput. Aid. Mol. Des.* **26**, 603–616 (2012).
28. Bottegoni, G., Kufareva, I., Totrov, M. & Abagyan, R. Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. *J. Med. Chem.* **52**, 397–406 (2008).
29. Armen, R. S., Chen, J. & Brooks, C. L. An evaluation of explicit receptor flexibility in molecular docking using molecular dynamics and torsion angle molecular dynamics. *J. Chem. Theory Comput.* **5**, 2909–2923 (2009).
30. Dietzen, M., Zotenko, E., Hildebrandt, A. & Lengauer, T. On the applicability of elastic network normal modes in small-molecule docking. *J. Chem. Inf. Model.* **52**, 844–856 (2012).
31. Vinh, N., Simpson, J., Scammells, P. & Chalmers, D. Virtual screening using a conformationally flexible target protein: models for ligand binding to p38alpha MAPK. *J. Comput. Aid. Mol. Des.* **26**, 409–423 (2012).
32. Barril, X. & Morley, S. D. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* **48**, 4432–4443 (2005).
33. Afonine, P. V. *et al.* Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr. D* **66**, 1153–1163 (2010).
34. Lang, P. T. *et al.* Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci.* **19**, 1420–1431 (2010).
35. Shapovalov, M. V. & Dunbrack, R. L. Statistical and conformational analysis of the electron density of protein side chains. *Proteins: Struct. Funct. Bioinf.* **66**, 279–303 (2007).
36. Rosenfeld, R. J., Hays, A.-M. A., Musah, R. A. & Goodin, D. B. Excision of a proposed electron transfer pathway in cytochrome c peroxidase and its replacement by a ligand-binding channel. *Protein Sci.* **11**, 1251–1259 (2002).
37. Barelier, S. *et al.* Roles for ordered and bulk solvent in ligand recognition and docking in two related cavities. *PLoS ONE* **8**, e69153 (2013).
38. Brenk, R., Vetter, S. W., Boyce, S. E., Goodin, D. B. & Shoichet, B. K. Probing molecular docking in a charged model binding site. *J. Mol. Biol.* **357**, 1449–1470 (2006).
39. Rocklin, G. J. *et al.* Blind prediction of charged ligand binding affinities in a model binding site. *J. Mol. Biol.* **425**, 4569–4583 (2013).
40. 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).
41. Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J. & Shoichet, B. K. Ligand pose and orientational sampling in molecular docking. *PLoS ONE* **8**, e75992 (2013).
42. Meng, E. C., Shoichet, B. & Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comp. Chem.* **13**, 505–524 (1992).
43. 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).
44. 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.* **52**, 1757–1768 (2012).
45. Carlsson, J. *et al.* Ligand discovery from a dopamine D3 receptor homology model and crystal structure. *Nature Chem. Biol.* **7**, 769–778 (2011).
46. Mysinger, M. M. *et al.* Structure-based ligand discovery for the protein–protein interface of chemokine receptor CXCR4. *Proc. Natl Acad. Sci. USA* **109**, 5517–5522 (2012).
47. Rogers, D. J. & Tanimoto, T. T. A computer program for classifying plants. *Science* **132**, 1115–1118 (1960).
48. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acid. Res.* **28**, 235–242 (2000).
49. Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* **50**, 1561–1573 (2010).
50. Gallagher, K. & Sharp, K. Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys. J.* **75**, 769–776 (1998).
51. Ytreberg, F. M. & Zuckerman, D. M. A black-box re-weighting analysis can correct flawed simulation data. *Proc. Natl Acad. Sci. USA* **105**, 7982–7987 (2008).

## Acknowledgements

We thank G. Rocklin for enlightening discussions, P. Afonine for advice and bug fixing of the occupancy refinement in PHENIX, A. Doak for protein preparation and H. Lin, J. Karpik and T. Baliaus for reading this manuscript. This work was supported by US National Institutes of Health grants GM59957 (B.K.S.), DP5OD009180 (J.S.F.) and NRSA F32GM096544 (R.G.C.).

## Author contributions

M.F., R.G.C., J.S.F. and B.K.S. designed the study and wrote the paper, M.F. performed all experiments and refined structures with the assistance of J.S.F. R.G.C. wrote the computer code and performed all computational work.

## Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.S.F. and B.K.S.

## Competing financial interests

The authors declare no competing financial interests.

## Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery

Marcus Fischer<sup>1,2,3</sup>, Ryan G. Coleman<sup>1,3</sup>, James S. Fraser<sup>\*4</sup> & Brian K. Shoichet<sup>\*1,2</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA 94158, and <sup>2</sup>Faculty of Pharmacy, Donnelly Center, University of Toronto, 160 College St. Toronto Ontario M5S 3E1.

<sup>3</sup>These authors contributed equally to this work. <sup>4</sup>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<sup>1</sup>, using  $6 \pm 0.75$  as the pH range, and only states that were at least 50% populated in that range were kept, as before<sup>2</sup>. 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<sup>3</sup>, electrostatics using QNINFF<sup>4,5</sup> (an adaptation of DELPHI), and ligand desolvation<sup>6</sup>. 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<sup>7</sup>, as well as analyzing additional poses.

Receptor flexibility is taken into account in the following manner, an improvement over our own past methods<sup>8</sup>, with a complete rewrite of the code and integrated into DOCK 3.7<sup>9</sup>. 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<sup>10</sup>. 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<sup>10</sup> 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 =  $e^{(-\text{energy}/k_B T)}$  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  $\pm 10\%$ . We modeled a range of reasonable occupancy values for the *B* loop (0.0000001 to 0.15), corresponding to energy penalties from  $\sim 9.6$  to  $\sim 1.1$  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 & Zuckerman<sup>7</sup>. 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<sup>11</sup>, 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:

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

Where the score of any pose is  $\text{score}(x)$ ,  $N_{\text{dist}}$  is the chosen parameter value for the number of nearby points to use,  $R_{\text{hyp}}$  is the radius of the hypersphere for that pose  $x$  that contains that many points and  $\text{dof}$  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  $N_{\text{dist}}$  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  $N_{\text{dist}}$  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<sup>12</sup> 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<sup>1,13</sup> 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<sup>2,14</sup>.

15. Structures were solved by molecular replacement with Phaser<sup>16</sup> using the same model and R<sub>free</sub> for all structures. Alternating cycles of refinement and model building were carried out in Refmac\_5.5.0109<sup>17</sup> and Coot<sup>18</sup>, 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<sup>19</sup>). 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<sup>20</sup>, even cryo structures are likely to support enough conformations to support this analysis; this becomes ever more true as resolution rises<sup>21,22</sup>.

### *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<sup>2,14</sup>—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<sup>5</sup>, 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

<b>compound</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
PDB ID	4NVF	4NVE	4NVD	4NVC
ZINC ID	1583444 (MES)	331902 (BZI)	331945	36634
<b>Data collection</b>				
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
$\alpha, \beta, \gamma$ (°)	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)
$R_{\text{merge}}$	0.1 (0.50)	0.04 (0.386)	0.04 (0.47)	0.04 (0.1)
$I/\sigma I$	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
<b>Refinement</b>				
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)
$R_{\text{work}}/R_{\text{free}}$	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 (°)	1.75	1.46	1.60	1.43

\*Highest resolution shell is shown in parenthesis.

<b>compound</b>	<b>5</b>	<b>APO-RT</b>	<b>6</b>	<b>7</b>
PDB ID	4NVB	4NVA	4NVG	4NVH
ZINC ID	8652421	-	6656163	4962659
<b>Data collection</b>				
Space group	P212121	P212121	P212121	P212121
Cell dimensions				
<i>a, b, c</i> (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
$\alpha, \beta, \gamma$ (°)	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)
$R_{\text{merge}}$	0.056 (0.369)	0.057 (0.64)	0.061 (0.562)	0.039 (0.581)
$\text{I}/\sigma\text{I}$	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
<b>Refinement</b>				
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)
$R_{\text{work}}/R_{\text{free}}$	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

<b>compound</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>
PDB ID	4NVI	4NVJ	4NVK	4NVL
ZINC ID	331160	13739037	1596053	34979991
<b>Data collection</b>				
Space group	P212121	P212121	P212121	P212121
Cell dimensions				
<i>a, b, c</i> (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
$\alpha, \beta, \gamma$ (°)	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)
$R_{\text{merge}}$	0.043 (0.4)	0.074 (0.715)	0.042 (0.566)	0.038 (0.428)
$l/\sigma l$	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
<b>Refinement</b>				
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)
$R_{\text{work}}/R_{\text{free}}$	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

<b>compound</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>APO-cryo</b>
PDB ID	4NVM	4NVN	4NVO	4OQ7
ZINC ID	203341	519712	388812	-
<b>Data collection</b>				
Space group	P212121	P212121	P212121	P212121
Cell dimensions				
<i>a, b, c</i> (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
$\alpha, \beta, \gamma$ (°)	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)
$R_{\text{merge}}$	0.044 (0.413)	0.046 (0.358)	0.057 (0.703)	0.071 (0.615)
$\text{I}/\sigma\text{I}$	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
<b>Refinement</b>				
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)
$R_{\text{work}}/R_{\text{free}}$	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
<i>ABC</i>	0.1639
<i>ABCD</i>	0.1639
<i>AC</i>	0.1641
<i>BC</i>	0.1646
<i>C</i>	0.1648
<i>AB</i>	0.1659
<i>B</i>	0.1669
<i>A</i>	0.1695
<i>D</i>	0.1712

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

Mean (standard deviation)	1000 poses $N_{\text{dist}} = 24$ $\text{dof} = 6$	1000 poses $N_{\text{dist}} = 10$ $\text{dof} = 6$	1000 poses $N_{\text{dist}} = 50$ $\text{dof} = 10$	1000 poses $N_{\text{dist}} = 50$ $\text{dof} = 1$
Loop <i>B</i> 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 <i>m</i> 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 <i>m</i> 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.  $N_{\text{dist}}$  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<sup>7</sup> with various parameters.

Mean (standard deviation)	1000 poses $N_{\text{dist}} = 24$ $\text{dof} = 6$	1000 poses $N_{\text{dist}} = 10$ $\text{dof} = 6$	1000 poses $N_{\text{dist}} = 50$ $\text{dof} = 10$	1000 poses $N_{\text{dist}} = 50$ $\text{dof} = 1$
Loop <i>B</i> 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 <i>m</i> 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 <i>m</i> 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.  $N_{\text{dist}}$  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
<b>6</b>	<i>C / B</i>	8	11	6	173
<b>7</b>	<i>B / B</i>	38	29	20	401
<b>8</b>	<i>B / B</i>	70	39	28	527
<b>9</b>	<i>C / C</i>	163	4403	3345	120
<b>10</b>	<i>AC / A</i>	322	97	176	5563
<b>11</b>	<i>A / A</i>	330	98	97	1571
<b>12</b>	<i>A / A</i>	433	131	174	1440
<b>13</b>	<i>AC / A</i>	526	170	501	13587
<b>14</b>	<i>BC / C</i>	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<sup>5</sup>. Desolvation was measured using an artificial ligand that fills the binding site.

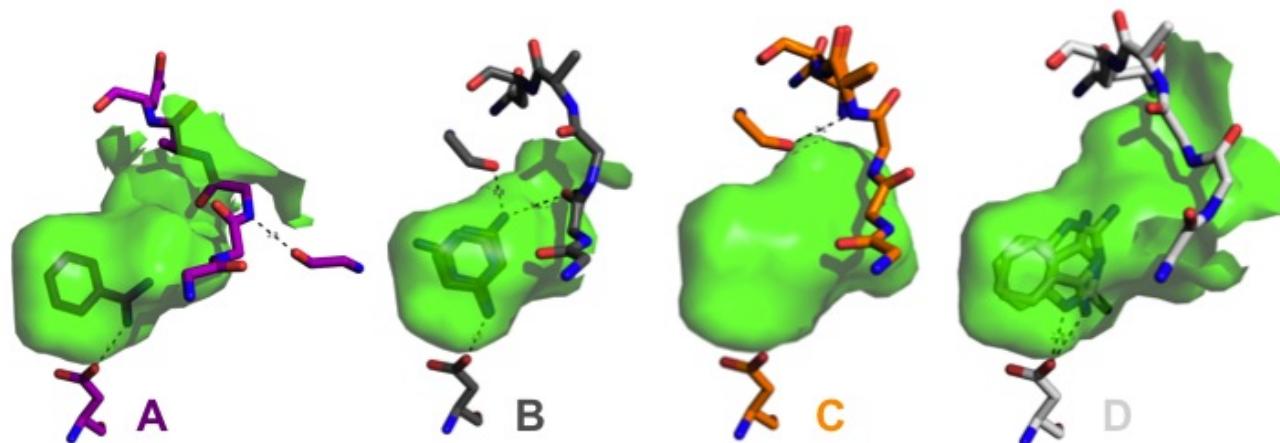
Loop	Binding Site Receptor Desolvation (kcal/mol)
Loop A	26.49
Loop B	23.23
Loop C	18.05
Loop D	27.04

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

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

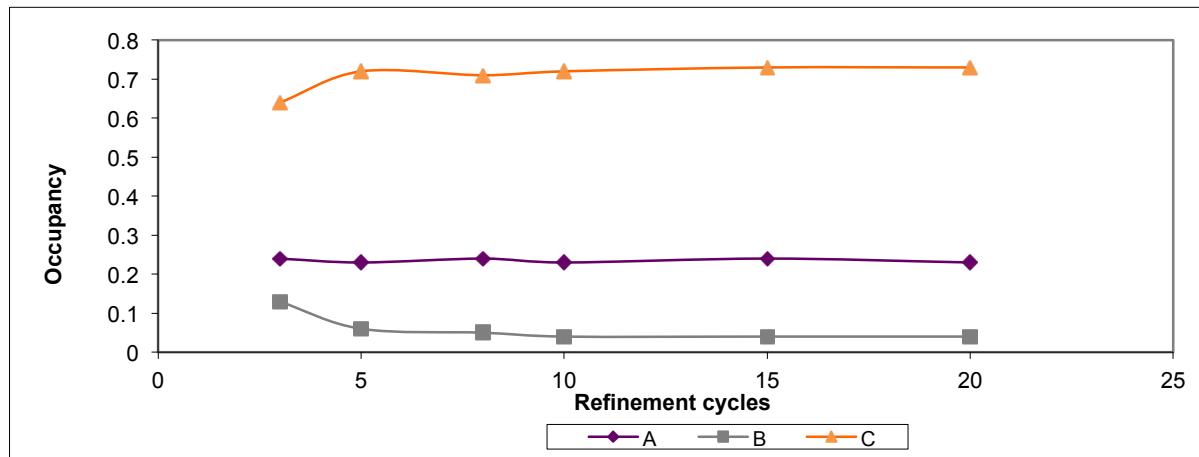
**Suppl. Table 9** Comparison of lists of ligands previously known for the CcP gateless cavity to those of a rigid-body docking study<sup>2</sup> to this work. Hdon, hydrogen-bond donor; Hacc, hydrogen-bond acceptor; Desol, desolvation; MaxT<sub>c</sub>, maximum Tanimoto coefficient.

	xlogP	Mol Weight	Rot Bonds	Hdon	Hacc	Apolar Desol	Polar Desol	MaxT <sub>c</sub> to Known
Known	0.85	148.0	0.58	2.3	3.0	2.18	-13.64	1
Barelier et al. <sup>2</sup>	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

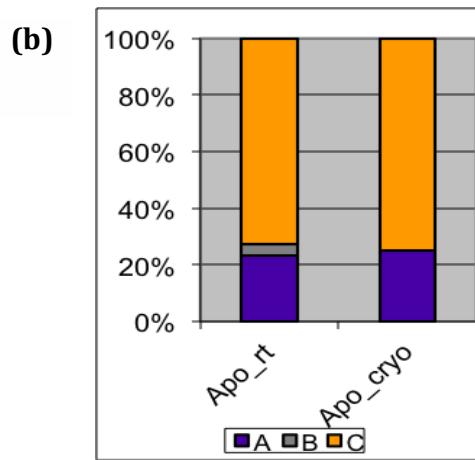
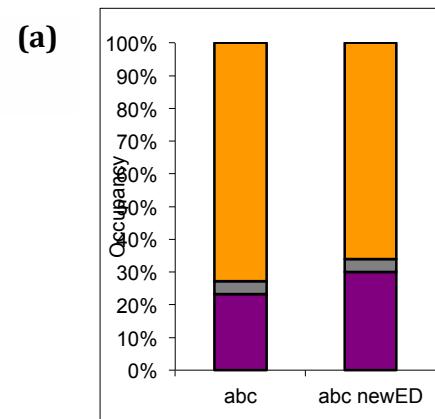


**Suppl. Figure 1** Internal hydrogen bonds of the different loop conformations.

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.

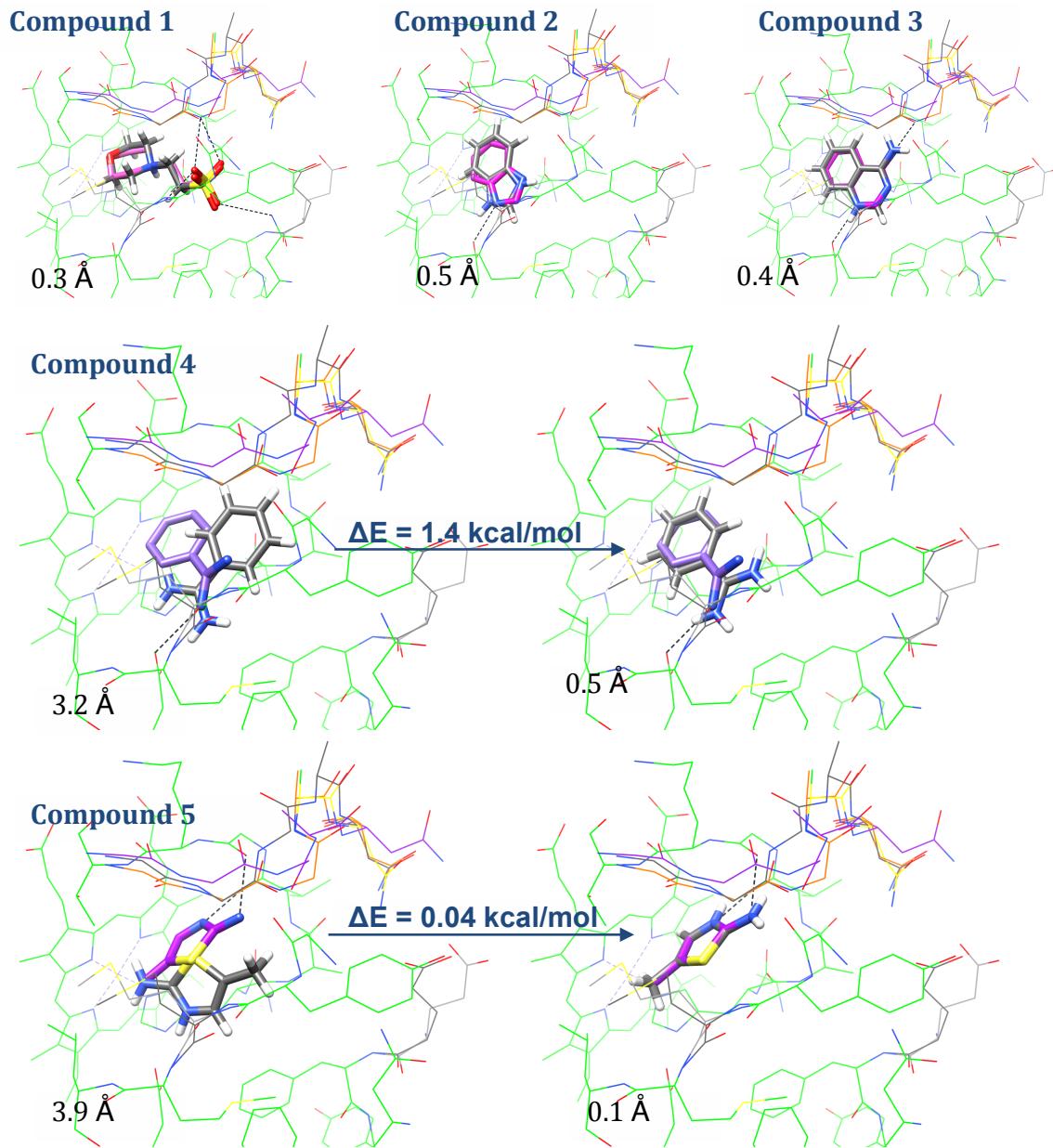


**Suppl. Figure 3** Comparison of automatic occupancy refinement of the apo CcP Gateless mutant.

(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

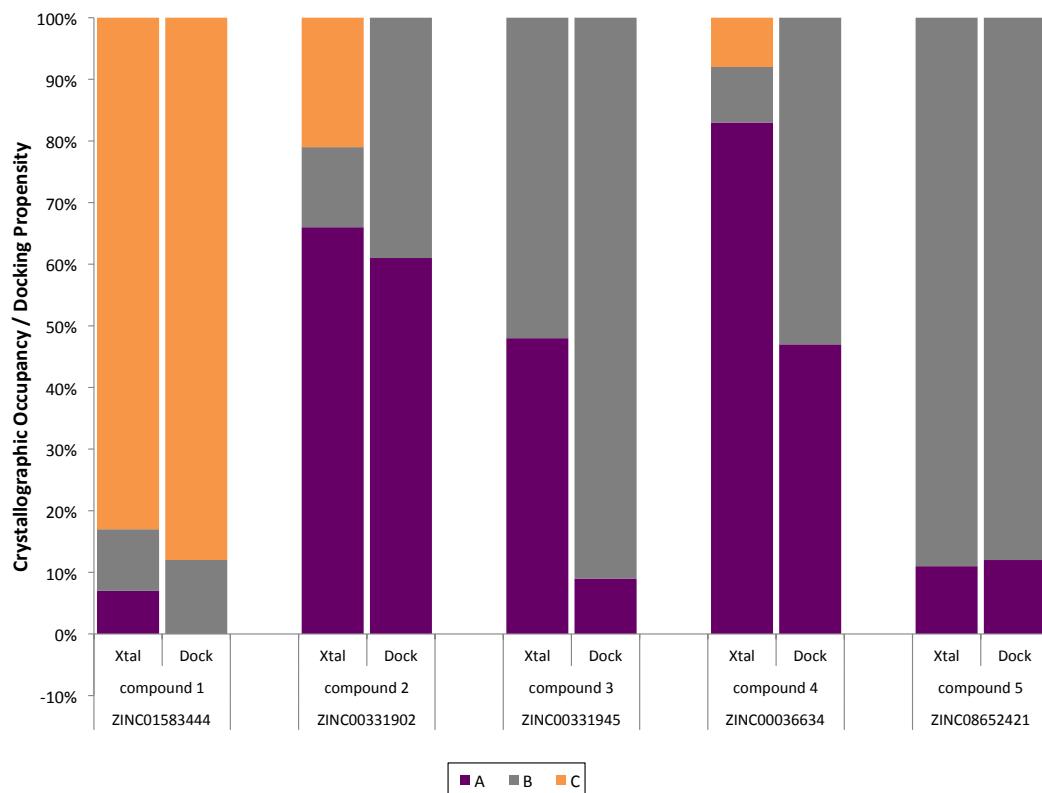
(b) Apo-RT and apo-cryo show a very similar distribution of states. Note that the occupancy for minor loop B disappears at cryo, which is why we used RT to be able to automatically assign a penalty.



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

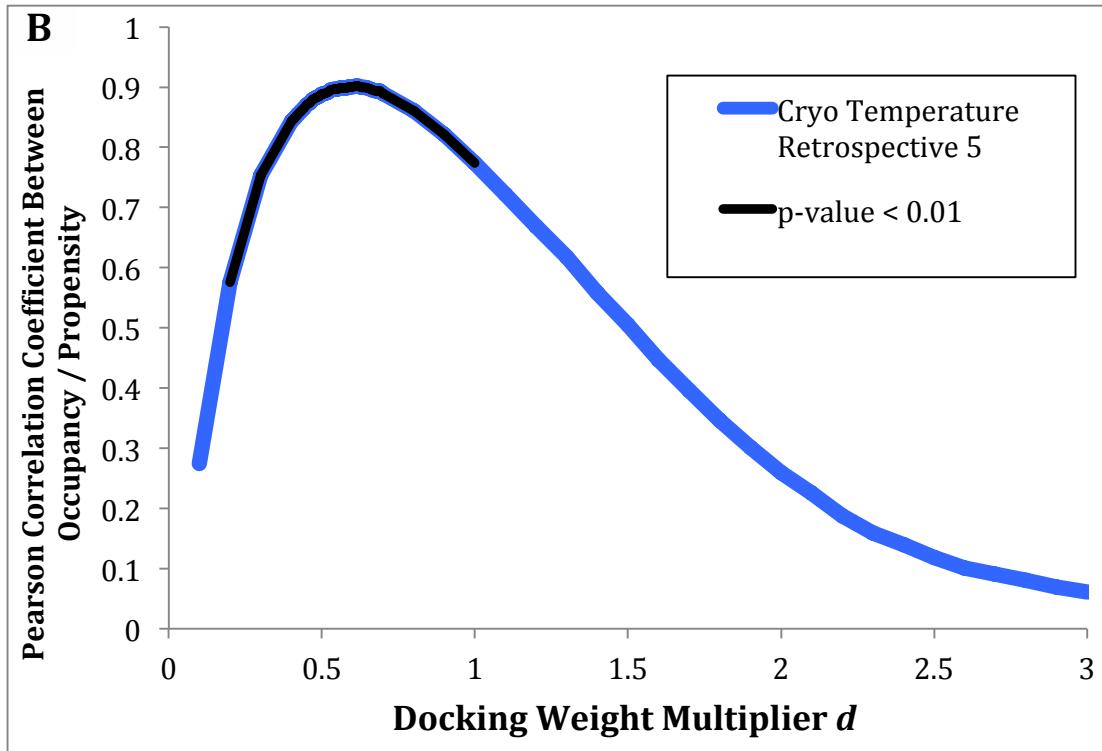
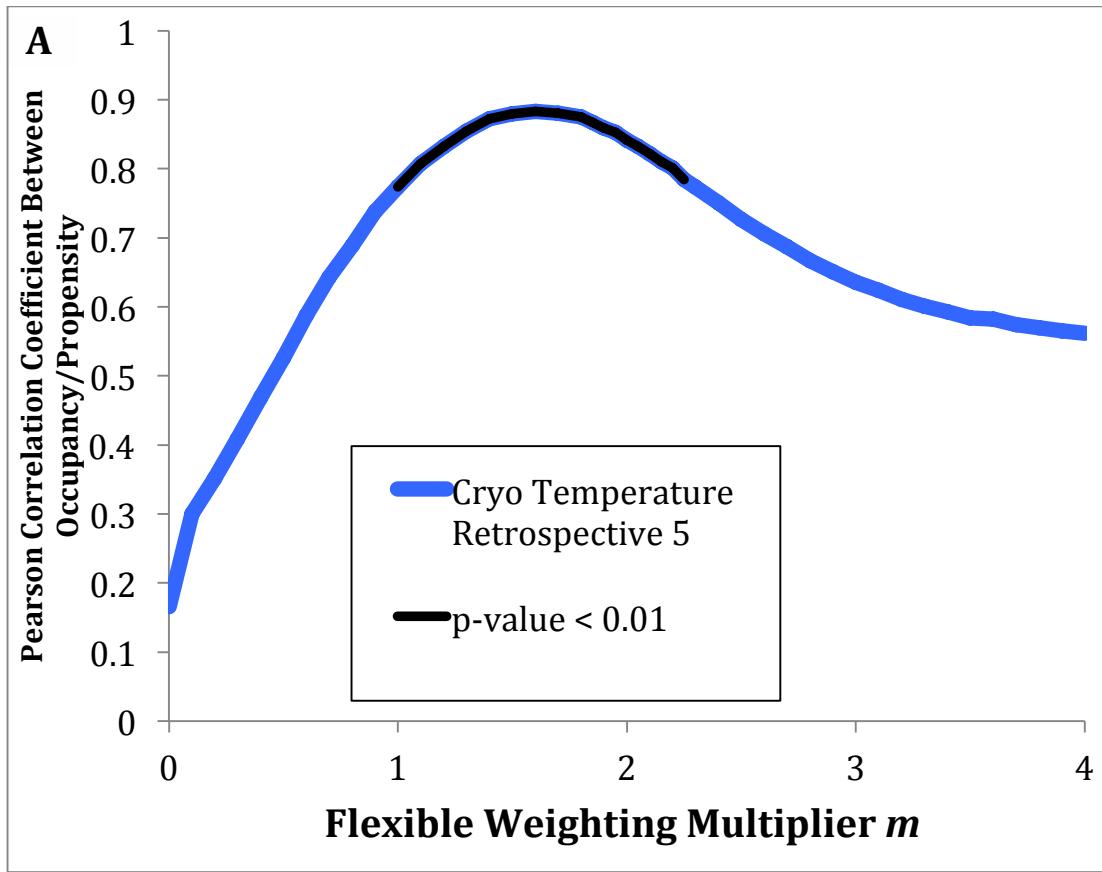
Shown are the top ranks with the exception of two compounds (**4** and **5**) where the correct pose was found further down the top 10 list (delta Energy  $\Delta E$ ).

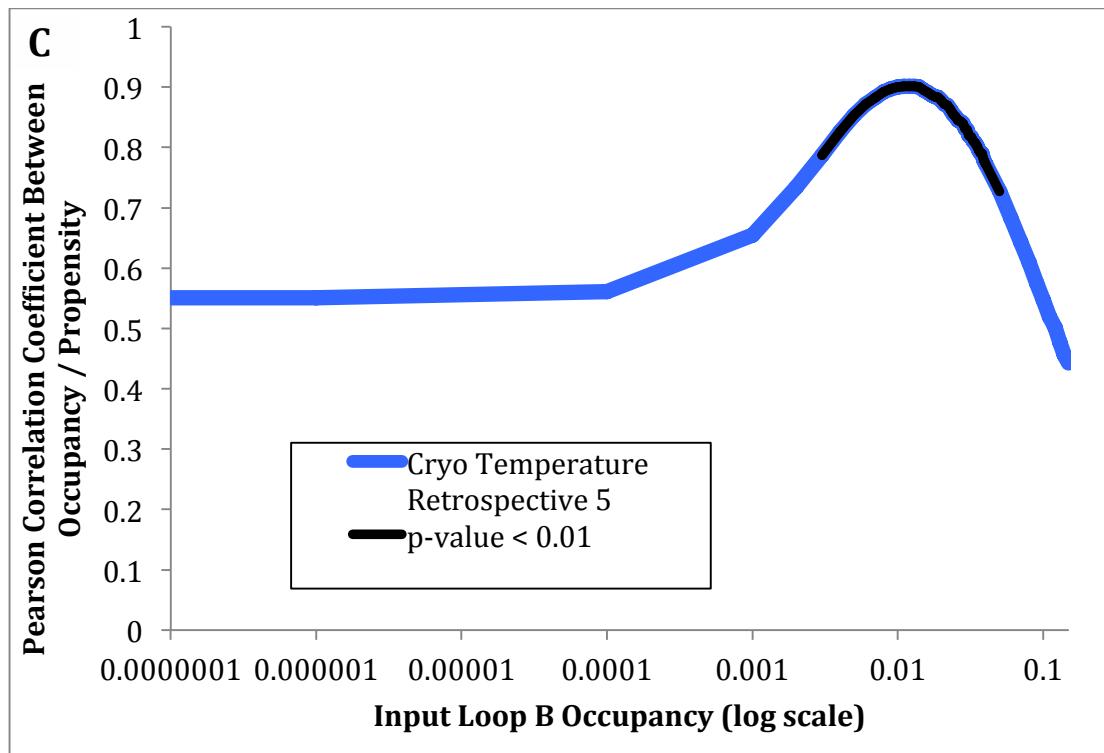
Docked compound in grey, experimental binding pose in purple. RMSDs between prediction and crystal structure are given in the bottom left for each compound.



**Suppl. Figure 5** Loop Occupancy Propensity prediction with flexible weighting multiplier  $m = 1$ .

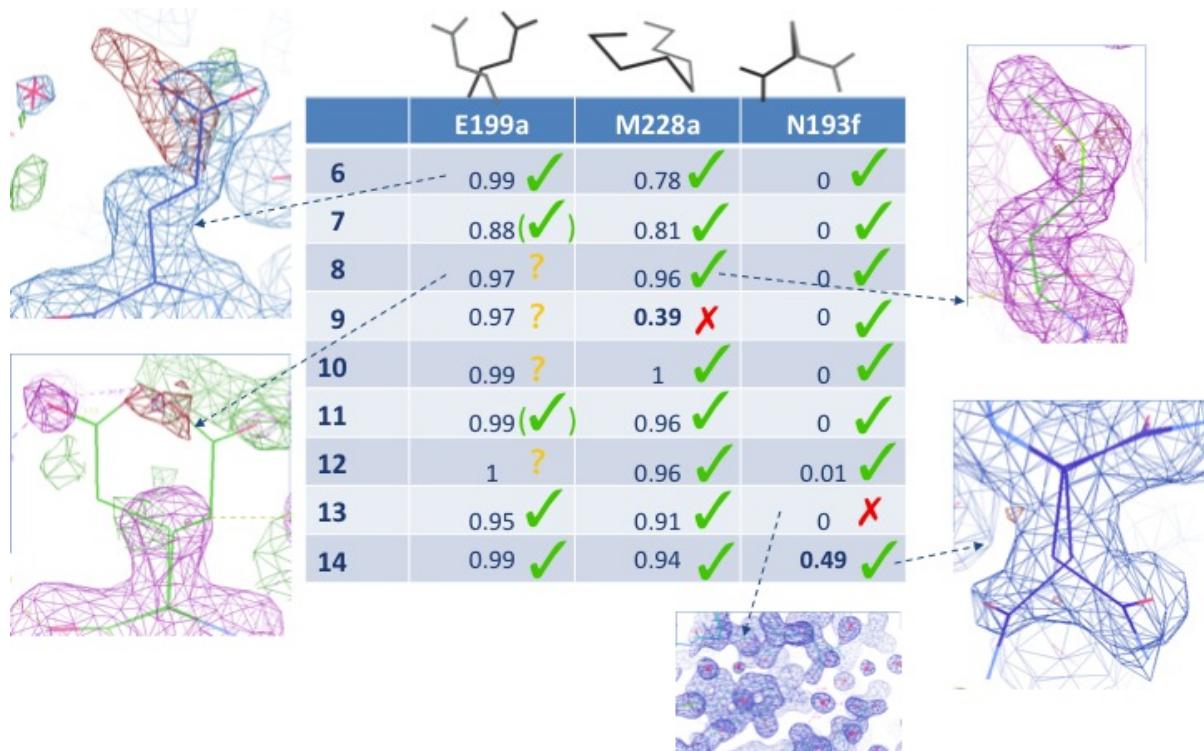
Crystallographic occupancies on the left of each pair, the right is the predicted propensity from docking for  $m=1$  (compare to Figure 2 for  $m=2$ ). Here, the overall Pearson Correlation Coefficient of the occupancies/propensities of these ligands is 0.77, whereas for  $m=2$  the coefficient is 0.83.





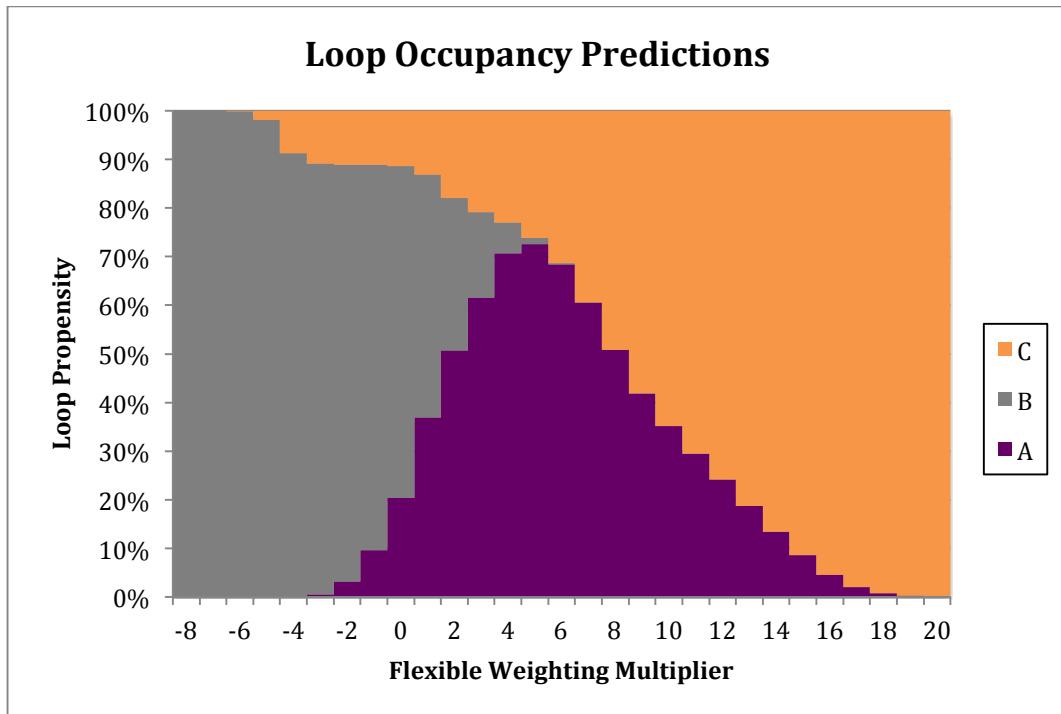
**Suppl. Figure 6** Response of PCC to the crystallographic occupancy and the predicted DOCK loop propensity of the retrospective compounds **1-5**.

- (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.

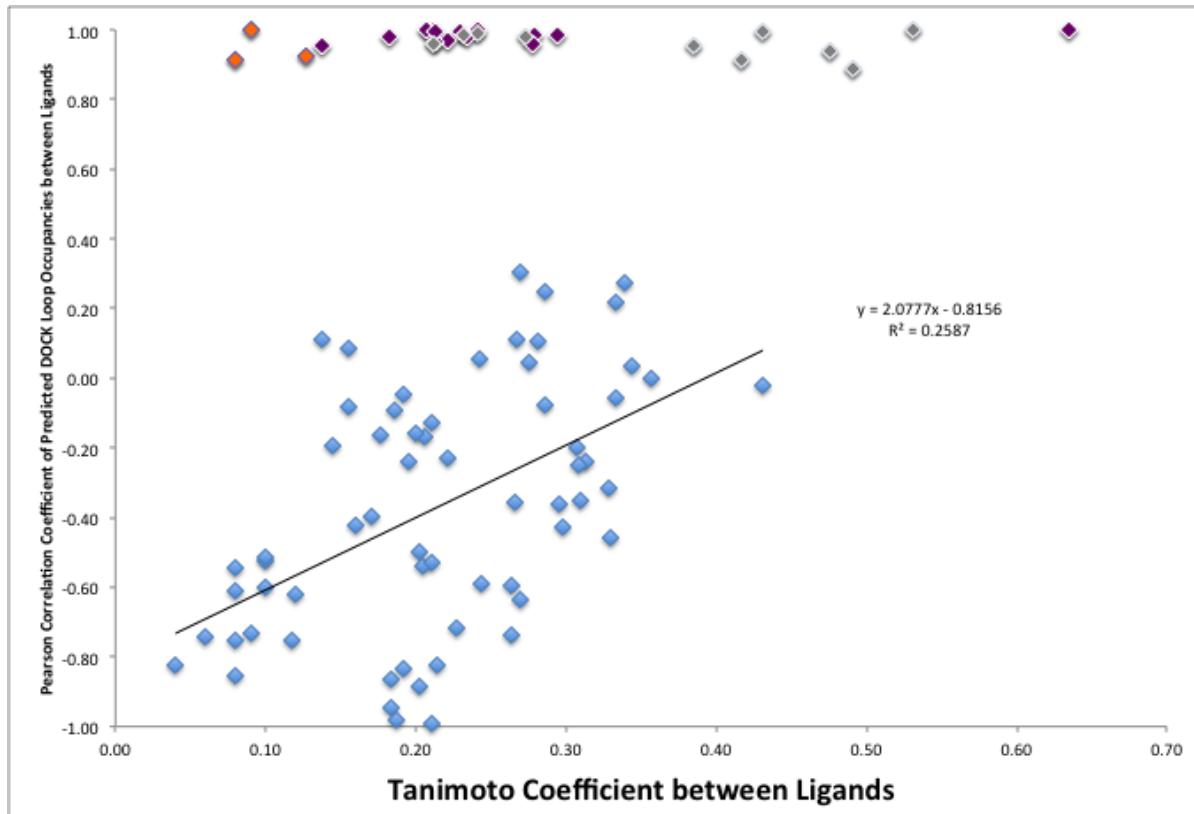


**Suppl. Figure 7** Analysis of side-chain predictions vs. experimental data.

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.

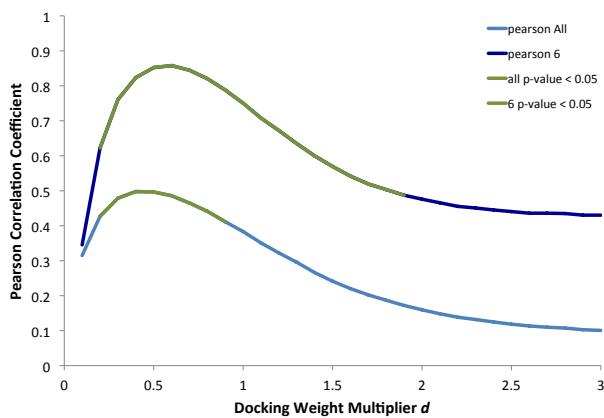
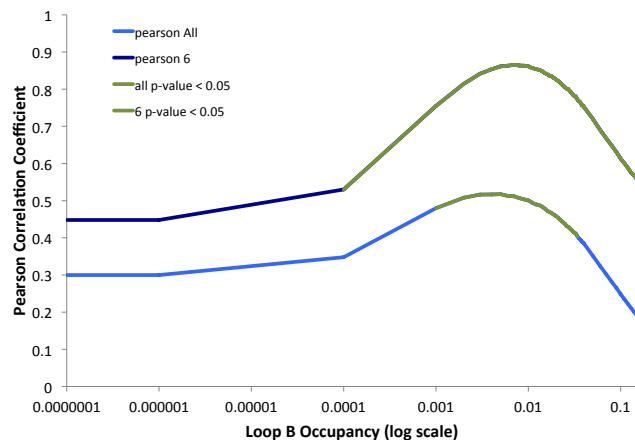


**Suppl. Figure 8** Sum over all loop states the prospective ligands predict at each flexible weight multiplier.



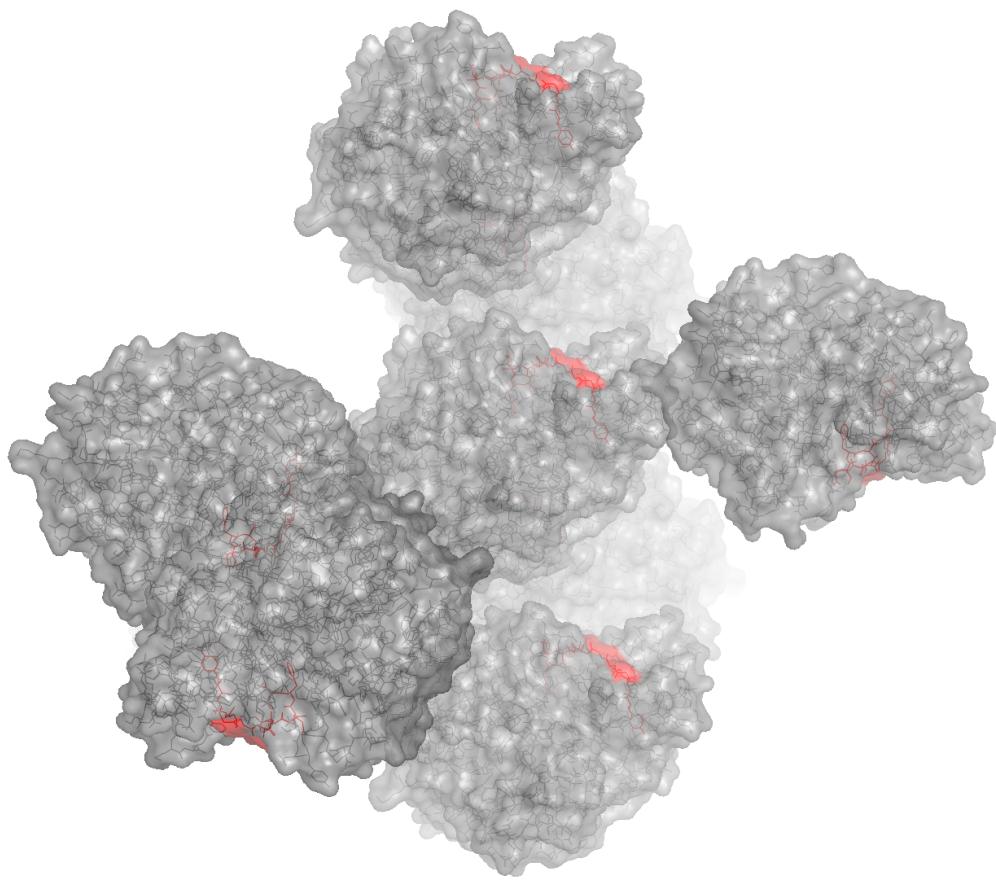
**Suppl. Figure 9** A comparison of the chemical similarity of the 14 ligands from this paper and the correlation of their loop occupancies.

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.

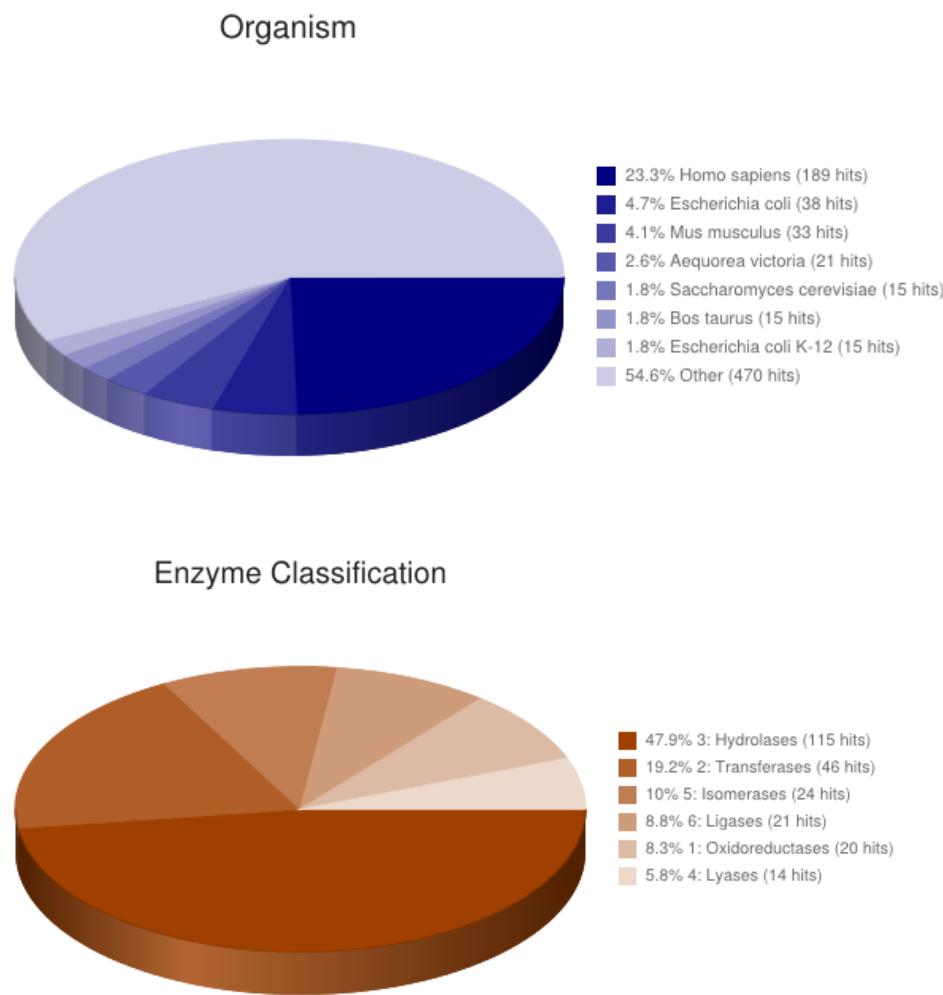
**A****B**

**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.

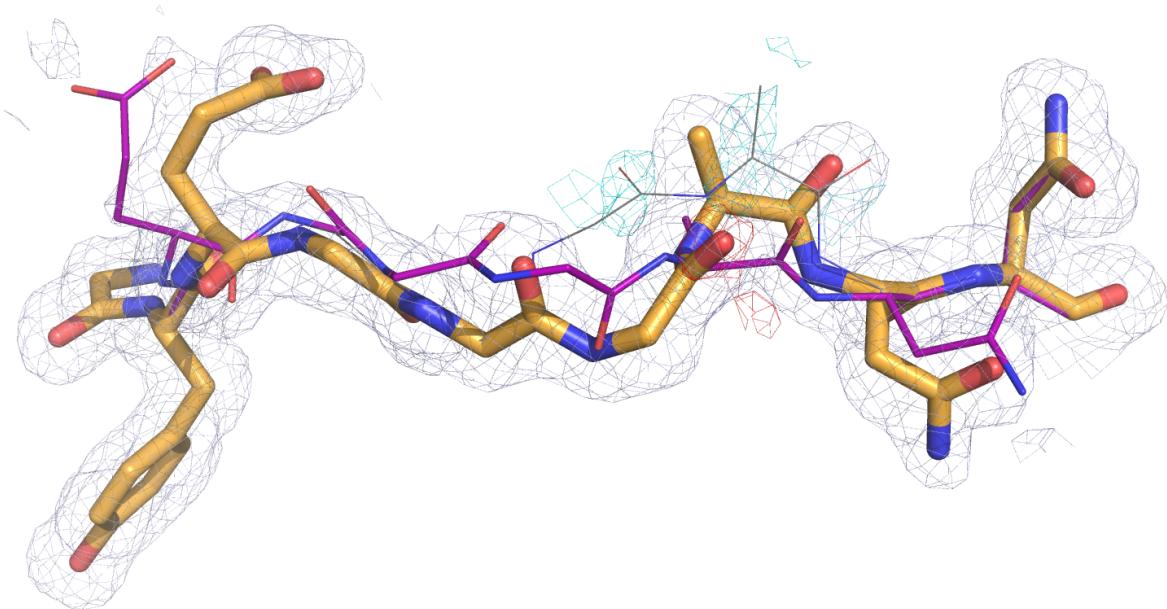
**Suppl. Figure 11 Loop movement is unobstructed by crystal contacts.**

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).