Supporting Information

FTFlex: Accounting for binding site flexibility to improve fragment-based identification of druggable hot spots

Laurie E. Grove[1], David R. Hall[2], Dmitri Beglov[3], Sandor Vajda[3], and Dima Kozakov[3]

[1]Department of Sciences, Wentworth Institute of Technology, Boston, MA, USA
[2]Acpharis Inc., Holliston, MA, USA
[3]Structural Bioinformatics Laboratory, Department of Biomedical Engineering, Boston University, Boston, MA, USA

To whom correspondence may be addressed. Email: midas@bu.edu
To whom correspondence may be addressed. Email: vajda@bu.edu

SUPPLEMENTAL METHODS

The FTMap algorithm

The FTmap mapping algorithm has been developed for the identification of 'hot spots' of proteins, regions that are most likely to participate in molecular interactions and hence are prime targets for drug design (Brenke et al., 2009). FTFlex is an extension of FTMap, which is also the first step of the FTFlex. FTMap consists of four steps as follows.

(1) *Sampling the protein surface on a grid.* The rotational/translational space of each probe is systematically sampled on a grid around the fixed protein, consisting of 0.8 Å translations and of 500 rotations at each location. The energy function includes a stepwise approximation of the Van der Waals energy with attractive and repulsive contributions, and an electrostatics solvation term based on Poisson-Boltzmann continuum model with the dielectric constants of e = 4 and e = 80 for the protein and the solvent, respectively.

(2) *Refinement of probe-protein complexes.* The 2000 complexes are refined by off-grid energy minimization during which the protein atoms are held fixed while the atoms of the probe molecules are free to move. The energy function includes the bonded and van der Waals terms of the CHARMM potential and an electrostatics/solvation term based on the Analytic Continuum Electrostatic (ACE) model, as implemented in CHARMM.

(3) *Clustering of low energy probe poses.* The minimized probe conformations are grouped into clusters using a simple greedy algorithm and a 4 Å RMSD clustering radius. Clusters with less than 10 members are excluded from consideration. The retained clusters are ranked on the basis of their Boltzmann averaged energies. Six clusters with the lowest average energies are retained for each probe.

(4) *Identification of hot spots as consensus sites binding several probe clusters.* To determine the hot spots, FTMap finds consensus sites (CSs), i.e., regions on the protein where clusters of different probes overlap. Therefore the probe clusters are clustered again using the distance between the centers of mass of the cluster centers as the distance measure and 4Å as the clustering radius. The consensus sites are ranked based on the number of their clusters, with duplicate clusters of the same type also considered in the count. The largest CS defines the most important hot spot, with smaller CSs identifying secondary hot spots that generally also contribute to ligand binding.

The FTFlex algorithm

The goal of FTFlex is to expand the FTMap algorithm to flexible proteins by accounting for the potential ligand induced conformational changes while preserving the high numerical efficiency of the approach. The feasibility of developing such an algorithm is suggested by a number of observations. First, mapping finds the approximate location of the binding site when applied to unbound protein structures, even in the case of substantial deviations between unbound and bound states. Thus, we can restrict considerations to some neighborhood of the hot spot regions from the initial mapping, which substantially simplifies the conformational search. Second, the major

reason for underestimating the importance of a hot spot when mapping an unbound structure is that the site becomes well formed only in the presence of a bound ligand, changing the conformation of adjacent side chains and increasing the size of the binding site. The opposite case, i.e., when the site is more open in the unbound structure and it closes down on the bound ligand, does not substantially affect the mapping results. The algorithm, identifies the potentially flexible side chains around an initial hot spot, generates their alternative conformers, and adjusts the side chains one-by-one to generate a modified structure with a maximally opened binding site. Steps of the FTFlex algorithms are as follows.

(1) *Identification of the binding site residues to be considered.* First, the ligand binding site is determined from a preliminary mapping calculation. The site with the highest hit rate of probe clusters is taken to be the binding site. Probe clusters within the binding site are used to identify all residues within a 5 Å radius. Next, any residue within 5 Å of the binding site is filtered based on hydrophobicity and cavity size restrictions. A previously developed hydrophobicity potential is used to calculate the hydrophobicity value for each of the residues (Chuang et al., 2008). Only residues with at least a hydrophobicity value of 3.00 are retained. A cavity measure, developed for use in FTMap (Brenke et al., 2009), is also calculated for each residue, and only residues with a cavity measure of at least 60% of the maximum value are retained. For each of the selected residues, possible rotamers are determined from a previously developed rotamer library (Beglov et al., 2012). This library was developed from a set of 4184 protein structures from the PDB by extracting and clustering the end group positions for each of the 18 amino acids using a 1Å clustering radius; glycine and alanine were not included. For each cluster, we define a population value by counting the number of end group positions within the 1Å radius around the cluster center.

(2) *Selecting movable side chains and their rotamers*. To determine the potential rotamers of the selected side chains we perform energy minimization starting from each rotamer in the above rotamer library, to reach local minima (Beglov et al., 2012). The side chains are considered one-by-one in the minimizations, with all others being fixed in their unbound state. The energy minimizations are performed using the CHARMM force field with the Analytical Continuum Electrostatics (ACE) solvation model (Brooks et al., 1983). Final rotamers are then clustered based on a 1 Å endpoint radius. For each cluster the average Boltzmann energy is calculated. Additionally, the population of each cluster is calculated as a sum of the population values of the original end groups (by definition, each representing a cluster of end group positions) that converged to the particular cluster after minimization. Thus, all possible rotamers have an associated energy and population. Side chains that are predicted to have multiple conformers with low energy and/or high population are considered to be movable. An analysis of our data shows that the movable residues are largely restricted to three categories. These movable residues include the bulky aromatics: Tyr, Phe, Trp, and His, the long, flexible amino acids: Lys, Arg, and Met, and Gln, Glu, Asp, and Asn . Hence, to improve efficiency FTFlex only considers rotamers of these residues.

The rotamers generated for the movable residues are filtered based on energy and population restrictions. Based on the analysis of rotamers predicted, a threshold of low energy and high population (greater than 10,000) was established. A threshold of 10,000 was determined based on analysis of the population counts for different residues. Rotamer clusters with populations over 10,000 were found to include all high population clusters, while

eliminating low population rotamers that had a population of less than 10% of the total. To avoid redundancy, rotamers that represent the apo conformation are also identified and ignored by FTFlex. To identify rotamers that are close to the apo conformation, a root mean square deviation (RMSD) value was determined for each residue type by averaging over all RMSDs between the apo-like rotamer and the original residue conformation. Calculation of the RMSD only took into account the atoms of the side chain; backbone atoms were excluded. The resulting average RMSD plus one standard deviation was used as the cutoff of apo-like residues.

(3) *Selection of final rotamers.* Since the goal of FTFlex is to open up binding pockets that may be partially closed in the ligand-free structure, we determine which of the selected rotamers yields the most improvement with regards to creating a larger binding site. This improvement is defined as an increase in the probe cluster count in the ligand binding site based on comparison to the initial mapping results. For this comparison, the rotamers of each movable residue are individually inserted back into the original apo structure. This modified structure is mapped using the FTMap algorithm to determine if there is an increase in the probe cluster count. The rotamer that yields the best improvement is used to generate the final modified structure. In cases where there are multiple movable residues, the rotamers of each residue are tested individually. For each residue the rotamer that yields the most improved mapping results is selected. These select rotamers are then substituted back into the apo structure to yield the modified apo structure. Mapping of the modified apo structure of each target using FTMap provides the final probe cluster count.

The FTFlex server workflow

The FTFlex server operates using the following steps. First, the protein structure of interest is mapped. The user must select the pdb file of interest and upload this file to the server. FTFlex maps the corresponding structure using the FTMap server. The initial mapping results are then returned to the user. The user must then determine the region on the protein surface that they would like to explore further. FTFlex will only search for flexible residues within the region selected by the user; this region is defined by selecting consensus clusters within the region. FTFlex will then select the residues for consideration within a 5 Å radius of the selected consensus clusters. Only Lys, Arg, Tyr, Phe, Trp, His, Met, Gln, Glu, Asn, and Asp are considered; all other residues are excluded. FTFlex then generates rotamers for each of these residues. Feasible rotamers are filtered based on the following criteria:

1. Only rotamers with a high population and low energy will be considered, as discussed previously.
2. Rotamers with an apo-like conformation as determined by an RMSD value will be excluded.

Each selected rotamer is then inserted back into the PDB one at a time and the structure is mapped again using FTMap. If there is an improvement in the mapping results, judged on the basis of an increase in the number of probe clusters within the binding site, then the rotamer is kept. If not, the rotamer is discarded from consideration. Since all rotamers are individually tested, the time required for this calculation will be highly dependent upon the input structure and the nature of the binding site. After all rotamers have been individually tested, the rotamers that yield the most improvement are substituted back into the PDB file to yield a modified structure, which is mapped and the results are returned to the user. A tutorial on how to use FTFlex can be found at: http://ftflex.bu.edu/tutorial.

SUPPLEMENTAL RESULTS AND ANALYSIS

Test set

Protein pairs to test the FTFlex algorithm were selected from the Astex non-native docking data set (Verdonk et al., 2008), which includes 65 targets, each represented by a number of structures. First, all targets that did not have an apo structure were discarded, thereby removing 36 targets. Next, targets were chosen that exhibited only very small backbone changes upon ligand binding, reducing the number of targets to 15. An apo and a bound structure were chosen for each of the 15 targets (Table S1). Both the apo and the reference bound structures were mapped after removing all water molecules, ligands, cofactors, and metals, with the exception of the heme cofactor in the structures 1p2y and 1mmv. The protein chains mapped, with the heme groups retained are shown in Table S1. The correlation coefficient between the % contact frequencies for apo and bound structures was calculated to score the similarity of the mapping results for bound and apo structures.

Table S1. Targets, PDB files of the apo and bound structures, and chains and heteroatoms used to test the FTFlex algorithm.

Target	PDB file/chain	
	Apo	Bound
TGT	1pud/A	1n2v/A
Carbonic anhydrase II	2nxr/A	1oq5/A
Cyclin-dependent kinase 2	1pw2/A	1ke5/A
Dipeptidyl peptidase IV	1pfq/A	1n1m/A
Acetylcholinesterase	1ea5/A	1gpk/A
Dihydrofolate reductase	1ai9/A	1ia1/A
Neuraminidase	2b8h/A	1l7f/A
nNOS	1zvi/A, heme	1mmv/A, heme
Chitinase B	1e15/A	1w1p/A
Carbonic Anhydrase XII	1jcz/B	1jd0/B
Neuraminidase	1nsb/A	1vcj/A
Protein Kinase 5	1ob3/A	1v0p/A
Dihydrofolate reductase	1pdb/A	1s3v/A
Cytochrome P450cam	1phc/A, heme	1p2y/A, heme
Beta-lactamase	2bls/A	1l2s/B

Analysis of side chain flexibility within the test set

The structural differences between the pairs of bound and apo structures in the test set are shown in Table S2. These RMSD values are for all residues, as selected by FTFlex, within 5 Å of the ligand binding site and meeting the hydrophobicity and cavity criteria. The first column lists the average RMSD for every residue across the set. To measure the extent of movement of only the residues that have conformational differences between the bound and apo structures, the residues that do not undergo any conformational change were filtered out and the average RMSD was recalculated (Table S2, second column). Some residues were never observed with a different conformation. For instance, Asn, Asp, Met, and Ser always had a similar conformation in both the bound and apo structures. As expected due to their size and larger range of motion, the RMSDs of Trp, Tyr, Lys and His are larger than those of Leu, and Val. It is interesting to note that in the case of Glu and Gln, which are both the same size, Gln has a larger RMSD. However, this result is largely an artifact of the nature of the Gln side chain. The RMSD

calculation does not take into account the spatial symmetry of the C_γ atom. That is, a Gln side chain with a 180° rotation at C_γ would be calculated as having a large RMSD even though the overall spacial conformation is identical.

Table S2. RMSD for each residue (bound structure vs initial apo structure) within the ligand binding site, number of rotamers predicted by FTFlex and RMSD of each rotamer to the initial apo conformation.

	Average RMSD	Average RMSD[a]	Average number of rotamers	Average number of rotamers[b]	RMSD (rotamer vs. apo)
Asn	0.398		4.67	1.86	1.74
Arg	0.896	1.298	10	2.75	2.75
Asp	0.364		3.63	1.63	1.41
Glu	0.517	0.808	3.5	2.22	1.61
Gln	1.717	2.458	6.5	1.67	2.13
His	1.287	2.677	1.71	1.4	2.54
Ile	0.796	1.847	5.62	1	0.65
Leu	0.448	1.040	2.5	1.33	0.89
Lys	2.630	3.379	14.5	3.75	2.1
Met	0.469		7	2	1.74
Phe	0.525	1.628	1.94	1.44	1.2
Ser	0.548		2.8	2.2	1.58
Thr	0.693	1.436	3.44	1.75	1.15
Trp	0.678	4.085	3.78	1.17	1.29
Tyr	0.648	3.254	3.64	1.65	1.89
Val	0.411	1.086	2.88	1.31	0.57

[a]For each of the bound-apo pairs, the RMSD of each side chain that moves was calculated and then averaged for each type of amino acid.
[b]The average number of low energy, high population rotamers.

The average number of rotamers generated by FTFlex was also determined for each residue (Table S2). Lys, Arg, and Met, are predicted to have an average of 14.5, 10, and 7 rotamers, respectively. Residues such as Gln, Asn and Ile are intermediate and all other residues are predicted to have an average of 4 rotamers or less. These results are in good agreement with data gathered for experimental ensembles, which invariably list Arg, Lys, Gln, and Met as the most flexible residues (Zhao et al., 2001; Najmanovich et al., 2000). Furthermore, in the case of Ile, Leu, and Val, only one low energy, high population rotamer is predicted by FTFlex and this rotamer is similar to the conformation of the residue in the apo state (Table S2, last column); thus, for our data set FTFlex does not predict any new conformations for Ile, Leu, and Val and these residues are considered to be nonmovable.

Impact of side chain flexibility on mapping results

Minimizing the number of amino acid residues considered by FTFlex not only enhances the speed of the calculation, but also reduces the probability of false positives; thus certain residues were excluded from the analysis performed by FTFlex. As explained above, Ile, Leu and Val were excluded. Rotamers for Ser and Thr were individually examined and did not yield a significant change in mapping results (see Table S3), thus these rotamers were also excluded from the FTFlex calculation. Thus, FTFlex only considers rotamers of the following residues: Asn, Arg, Asp, Glu, Gln, His, Lys, Met, Phe, Trp and Tyr.

Table S3. Rotamer energies and populations, correlation coefficients and probe cluster counts for Ser and Thr rotamers.

PDB	Residue	Rotamer Energy (kcal/mol)	Rotamer Population	RMSD to Apo (Å)	Correlation to bound	Probe Clusters
1pud					0.75	54
1pud	SER110	-1.22	44903	0.11	0.76	54
1pud	SER110	-0.91	67046	1.39	0.77	51
1ea5					0.84	75
1ea5	SER122	-7.26	67046	1.17	0.88	74
1ea5	SER122	-4.67	44903	2.71	0.88	75
1ea5	SER122	-0.34	30925	2.05	0.89	73
1ea5	SER200	-13	67046	2.38	0.89	76
1ea5	SER200	-6.05	39441	2.29	0.87	76
1ea5	SER200	-4.05	36387	0.4	0.89	76
1ea5	SER286	-4.17	39422	1.63	0.9	75
1ea5	SER286	-2.51	60526	0.22	0.87	75
1ea5	SER286	-2.07	42926	2.54	0.9	75
1pfq					0.46	26
1pfq	SER630	-8.15	10648	2.08	0.46	29
1pfq	SER630	-4.92	36387	0.84	0.39	27
1phc					0.87	29
1phc	THR101	-13.51	62069	0.46	0.84	29
1phc	THR101	-7.88	56641	1.59	0.83	24
1phc	THR185	-7.5	79811	0.36	0.84	28
1phc	THR252	-26.95	63921	2.06	0.82	29
1phc	THR252	-25.59	54787	0.45	0.84	29
1ai9					0.98	81
1ai9	THR133	-9.48	63923	0.26	0.98	77
1ai9	THR133	-4.03	54787	1.9	0.97	80
1ai9	THR147	-16.27	67527	1.42	0.98	79
1ai9	THR147	-9.28	51181	0.97	0.98	80
2nxr					0.35	61
2nxr	THR199	-5.6	51657	0.31	0.38	62
2nxr	THR200	-5.16	77796	0.88	0.32	64
2nxr	THR200	-3.34	42929	2.38	0.35	65
1pdb					0.95	83
1pdb	THR56	-18.17	54787	2.23	0.93	82
1pdb	THR56	-17.61	63923	0.76	0.94	79

To assess the agreement of the mapping results of each bound structure and the corresponding apo structure, a

correlation value was calculated (Table S4). The correlation value is a measure of the similarity of nonbonding interactions between probe clusters and binding site residues for two different structures.

Table S4. Correlations between the initial apo structure and bound structures (middle column), in addition to the modified and bound structures (last column). In the cases of the apo and bound structures, the mapping results were obtained using FTMap.

Target	Apo structures	Modified structures
TGT (1pud)	0.7540	0.9000
Carbonic anhydrase II (2nxr)	0.3533	0.7089
Cyclin-dependent kinase 2 (1pw2)	0.4803	0.6692
Dipeptidyl peptidase IV (1pfq)	0.4602	0.7506
Acetylcholinesterase (1ea5)	0.8433	0.8433
Dihydrofolate reductase (1ai9)	0.9751	0.9751
Neuraminidase (2b8h)	0.9427	0.9427
NNOS (1zvi)	0.8442	0.8533
Chitinase B (1e15)	0.7531	0.8281
Carbonic Anhydrase XII (1jcz)	0.8754	0.8388
Neuraminidase (1nsb)	0.8527	0.8603
Protein Kinase 5 (1ob3)	0.8229	0.8922
Dihydrofolate reductase (1pdb)	0.9495	0.9196
Cytochrome P450cam (1phc)	0.8654	0.8654
Beta-lactamase (2bls)	0.8879	0.9210

Table S4 shows two different correlation values for each of the fifteen targets. These two different correlation values were calculated using the mapping results of the initial apo structure, the bound structure, and the modified apo structure. In the case of the initial apo structure and the bound structure, the correlation values were calculated based on the mapping results obtained from FTMap. The correlation value is improved for eight targets: TGT, CAII, CDK2, dipeptidyl peptidase IV (DPP-IV), chitinase B (chit B), neuraminidase (PDB ID 1nsb), protein kinase 5 (PK5), and beta-lactamase.

An example is shown in Figure S1 for four different target proteins. The plots show the % contact frequency of the probe clusters to each of the ligand binding site residues for three structures: apo, modified apo and bound. The higher the number of nonbonded contacts between a probe cluster and a residue, the higher the % contact frequency; thus the % contact frequency serves as a fingerprint of the mapping results for a given structure. The more similar the % contact frequencies of two structures, the higher the correlation coefficient. In the following analysis, the success of FTFlex versus FTMap is compared. A correlation value was calculated using the mapping results obtained from FTMap for the initial apo structure versus the FTMap results of the bound structure. A second correlation value has been calculated for an FTFlex-generated modified apo structure versus the FTMap-generated mapping results of the bound structure. Thus the comparison of the correlation value between the initial apo-bound and modified apo-bound pairs is a direct assessment of the improvement in mapping obtained using FTFlex versus FTMap. For example, in the case of TGT, the agreement in the % contact frequency for Tyr106 varies significantly between the apo and bound structures (Figure S1a). FTFlex identifies Asn70, Tyr106, and Gln107 as moveable residues and generates a modified structure, which has a higher similarity to the % contact frequency profile of the bound structure. Two other examples are shown: carbonic anhydrase II (CAII) and cyclin-dependent kinase 2 (CDK2) in Figures S1b and S1c, respectively. In these cases, His64 of CAII and Lys33 and Lys89 of CDK2

were found to be moveable by FTFlex. The modified structure generated by FTFlex yielded mapping results in better agreement with those of the bound structure. In the case of dihydrofolate reductase (DHFR) (PDB ID 1ai9) (Figure S1d) there are minimal structural differences between the apo and bound structures. Hence, the mapping results of the bound and apo structures are already in good agreement and FTFlex predicts no new conformations

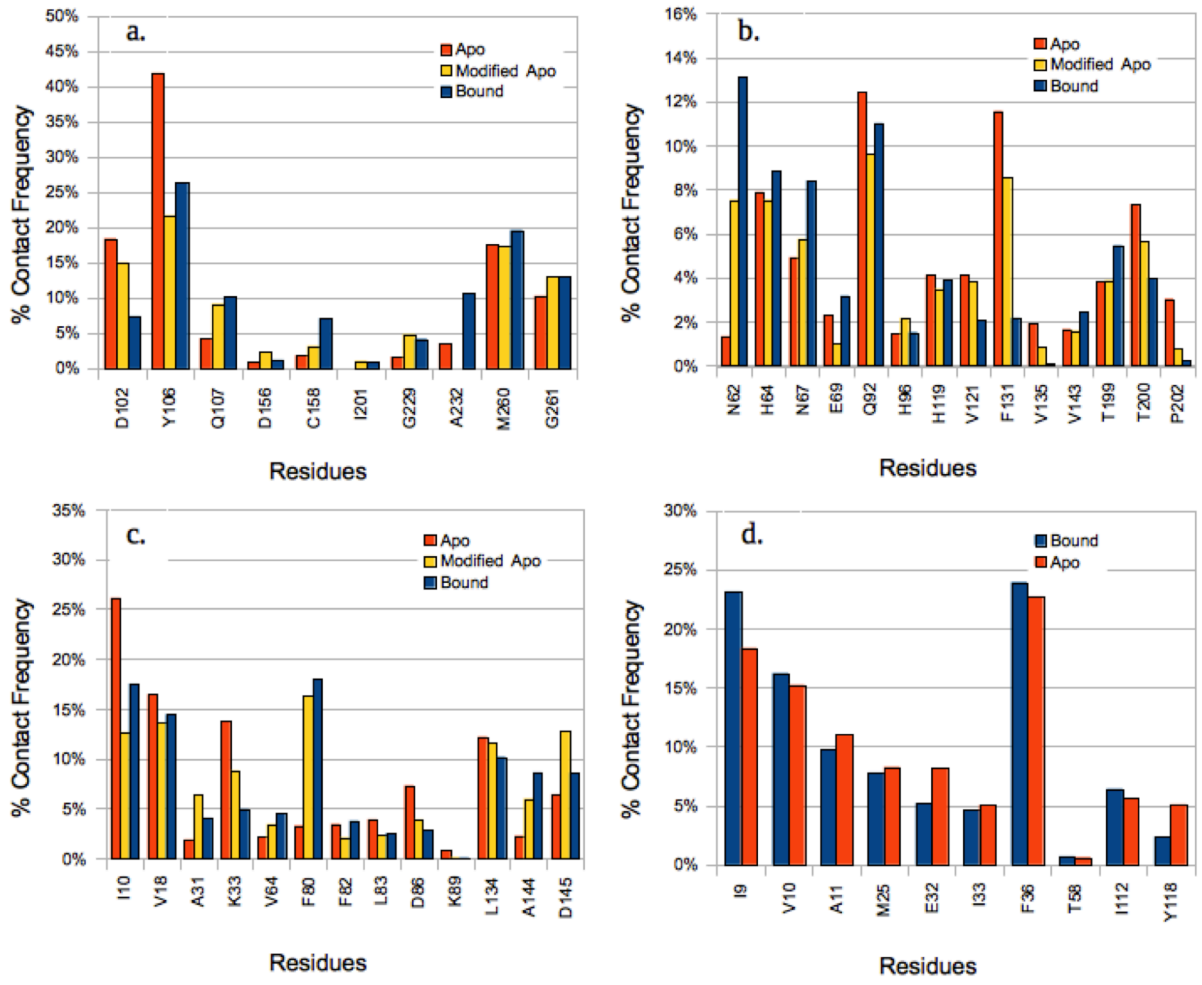

Figure S1. Distribution of nonbonded contacts between probe clusters and residues from the mapping results of apo, FTFlex-modified apo, and bound structures. The results are shown for four proteins: TGT (a), CAII (b), CDK2 (c), and DHFR (d).

The correlation values are also illustrated for TGT, CAII and CDK2 in Figures S1a, S1b and S1c, respectively. For these cases the improvement in the correlation value is quite significant, ranging from 0.15 to 0.35. For four cases, the correlation values are already quite high and FTFlex correctly predicts that no residues should be moved. These cases are DHFR (PDB ID 1ai9), neuraminidase (PDB ID 2b8h), NNOS, and cytochrome P459cam. In two cases, carbonc anhydrase XII and DHFR (PDB ID 1pdb), there is a slight worsening of the correlation value by approximately 0.04, but the value is still quite high. In the case of acetylcholinesterase (AChE), while the correlation score is good, it should be noted that the position of Phe330 is significantly different between the apo and bound structure (RMSD of 2.169 Å). While FTFlex predicts a Phe330 conformation that corresponds to the conformation in the bound structure, due to current selection rules this conformation is not selected. This case

warrants further exploration and perhaps a future implementation within FTFlex to allow users increased access to multiple conformations. The individual rotamer results, including RMSDs, and the final mapping results for the 15 targets are shown in Tables S5 and S6, respectively.

Table S5. FTFlex predicted rotamers, energies, populations, RMSDs, and mapping results for targets.

PDB	Rotamer[a]	Energy (kcal/mol)	Population	RMSD[b] (Å)	RMSD[c] (Å)	RMSD[d] (Å)	Correlation[e]	Probe Clusters
1pud	None						0.7540	54
	Asn70 (1)	-2.607	18873	0.420	0.510	0.811	0.7503	53
	Asn70 (2)*	-0.763	54937	2.235		2.513	0.7824	57
	Asp102	-5.643	14608	1.113	0.183	1.042	0.8055	54
	Tyr106 (1)	-9.939	43358	1.019	5.282	5.818	0.7637	52
	Tyr106 (2)*	-0.841	27839	5.176		0.684	0.8798	58
	Gln107 (1)	-3.692	57556	0.209	2.787	2.657	0.7510	59
	Gln107 (2)	-3.526	19358	3.005		2.472	0.7967	41
	Gln107 (3)*	-2.531	11004	3.192		2.428	0.8433	64
	Glu235 (1)	-8.998	13752	0.651	0.808	1.100	0.7699	53
	Glu235 (2)	-1.066	26261	1.899		2.444	0.7590	52
	Glu235 (3)	-0.462	103175	2.171		2.537	0.7807	52
	Final structure						0.9000	64
2nxr	None						0.3533	61
	His64*	-6.691	38212	4.115	3.305	2.169	0.7089	70
	Gln92 (1)	-3.009	15291	1.566	0.167	1.564	0.3255	61
	Gln92 (2)	-0.537	43247	3.430		3.393	0.4404	56
	Phe131	-4.480	32104	0.528	0.258	0.512	0.3851	60
	Trp209	-7.175	18977	0.294	0.188	0.467	0.3275	59
	Final structure						0.7089	70
1pw2	None						0.4803	28
	Lys33 (1)	-3.044	64053	0.375	3.657	3.922	0.4652	27
	Lys33 (2)*	-2.763	36408	2.916		1.517	0.6742	43
	Lys33 (3)	-2.454	26332	2.371		1.867	0.6607	43
	Phe80	-3.166	67631	0.161	0.493	0.630	0.4751	28
	Phe82	-4.000	49563	0.247	0.578	0.595	0.4792	27
	Gln85	-7.988	37000	3.367	2.220	2.393	0.4909	28
	Lys89 (1)	-11.832	36488	2.556	4.103	4.397	0.4614	27
	Lys89 (2)	-10.995	22279	3.623		4.818	0.4628	28
	Lys89 (3)*	-10.638	10683	5.423		5.944	0.5195	34
	Lys89 (4)	-9.919	12927	4.493		1.738	0.5066	34
	Asp145	03.415	83180	0.823	0.807	0.919	0.4706	24
	Final structure						0.6692	46
1pfq	None						0.4602	26
	Glu205 (1)	-5.024	32974	0.439	0.493	0.900	0.4396	24
	Glu205 (2)	-2.872	74099	2.448		2.739	0.4117	25
	Glu205 (3)	-2.801	50041	1.555		1.920	0.4541	25
	Glu206 (1)	-29.562	87494	2.301	0.627	2.078	0.3278	20

	Glu206 (2)	-12.537	69620	2.311		2.253	0.4276	23
	Phe357	-12.771	49854	0.721	0.206	0.632	0.3938	24
	Tyr547	-6.016	33972	0.567	1.224	1.321	0.4235	27
	Trp629*	-9.044	13565	3.854	4.085	0.413	0.7271	27
	Trp659	-14.818	13688	0.726	0.776	0.530	0.4163	25
	Arg669	-3.525	54738	0.331	1.210	1.277	0.4765	22
	Asn710 (1)	-3.119	22060	1.961	0.861	1.703	0.4302	24
	Asn710 (2)	-3.040	13997	1.259		1.846	0.4163	23
	Asn710 (3)	-1.207	28583	2.370		2.717	0.4243	23
	Asn710 (4)*	-0.471	33246	3.143		3.235	0.4111	29
	His740	-12.608	14316	4.744	2.048	4.754	0.3806	25
	Final structure						0.7506	27
1ea5	None						0.8433	75
	Asp72	-8.253	45977	0.991	0.437	1.307	0.8907	74
	Trp84	-4.818	11757	0.182	0.215	0.304	0.8834	74
	Asn85 (1)	-2.890	54249	0.392	0.123	0.431	0.9028	74
	Asn85 (2)	-0.607	19426	2.305		2.290	0.8871	73
	Tyr121	-3.499	24382	0.125	0.209	0.261	0.8523	73
	Tyr130	0.000	44384	0.366	0.218	0.307	0.8926	75
	Glu199	-2.078	49088	0.211	0.085	0.195	0.8862	75
	Trp279	-5.305	20962	0.718	0.156	0.730	0.9002	73
	Phe288	-3.530	64637	0.259	0.156	0.167	0.8893	74
	Phe290	-2.974	50411	0.147	0.216	0.193	0.8824	74
	Phe330	-5.965	85432	1.747	2.169	0.698	0.9595	71
	Phe331	-4.111	67776	0.617	0.228	0.732	0.8948	72
	Tyr334 (1)	-5.742	44289	0.217	0.217	0.353	0.8893	75
	Tyr334 (2)	-1.253	23943	3.349		3.356	0.8968	71
	His440	-16.298	33662	0.784	0.316	0.883	0.8770	74
	Tyr442	-5.199	48271	0.305	0.174	0.388	0.8855	74
	Final structure						0.8433	75
1ai9	None						0.9751	81
	Met25 (1)	-6.353	11460	1.538	0.645	1.782	0.9617	80
	Met25 (2)	-5.419	12678	2.303		2.310	0.9072	75
	Glu32 (1)	-8.680	50778	0.720	0.674	1.313	0.9623	79
	Glu32 (2)	-6.296	68465	1.709		1.797	0.9523	79
	Glu32 (3)	-5.810	37871	1.488		1.112	0.9788	78
	Phe36	-8.105	50135	0.171	0.461	0.359	0.9825	80
	Tyr118	-12.736	47502	0.251	0.206	0.259	0.9809	82
	Final structure						0.9751	81
2b8h	None						0.9427	31

1zvi	None						0.8442	36
	Glu592 (1)*	-3.295	26201	2.421	0.570	2.136	0.8533	43
	Glu592 (2)	-1.358	45830	1.726		1.812	0.8434	24
	Glu592 (3)	-1.207	17271	2.770		2.592	0.8899	36
	Arg596	-7.123	18065	6.098	0.493	6.385	0.8572	15
	Final Structure						0.8533	43
1e15	None						0.7531	83
	Tyr10	-12.313	23667	0.455	0.196	0.315	0.7482	82
	Phe12	-4.522	88623	0.400	0.485	0.463	0.7609	81
	Asn48	-3.934	64745	0.762	0.395	0.777	0.7331	83
	Phe51	-3.008	67629	0.182	0.256	0.250	0.8049	86
	Tyr98 (1)	-2.414	27968	0.611	0.489	0.727	0.7345	83
	Tyr98 (2)	-1.579	43044	6.800		6.477	0.6943	81
	Tyr99	-5.066	50900	0.387	0.513	0.628	0.7534	83
	Asp142 (1)*	-8.541	74674	1.711	0.209	1.651	0.7551	84
	Asp142 (2)	-6.279	54437	0.508		0.582	0.7503	84
	Glu144	-4.644	116250	0.680	0.350	0.434	0.8052	86
	Met212 (1)	-6.296	10425	1.445	0.293	1.533	0.6149	85
	Met212 (2)*	-4.235	19414	1.669		1.616	0.7573	87
	Tyr214 (1)	-6.543	29622	0.628	0.606	0.213	0.7697	84
	Asp215*	-19.576	40404	2.149	0.286	1.992	0.8297	84
	Arg294 (1)*	-22.238	50917	1.593	1.530	0.912	0.7746	85
	Arg294 (2)	-14.060	30002	2.589		2.970	0.7442	82
	Trp403	-4.972	17237	3.055	0.256	3.070	0.5652	79
	Gln407	-2.094	63080	0.302	0.120	0.281	0.7418	83
	Final Structure						0.8281	91
1jcz	None						0.8754	28
	His119	-5.823	48023	0.219	0.178	0.213	0.8817	34
	Trp209	-4.488	20730	0.217	0.144	0.324	0.8672	33
	Asn244 (1)	-3.864	10018	0.521	0.120	0.446	0.8481	31
	Asn244 (2)*	-0.571	20509	2.230		2.311	0.8388	35
	Final Structure						0.8388	35
1nsb	None						0.8527	28
	Arg373 (1)*	-6.607	19463	0.962	1.152	1.445	0.8603	31
	Arg373 (2)	-5.481	30952	3.112		2.890	0.8383	36
	Arg373 (3)	-4.242	16929	2.404		2.492	0.8066	35
	Arg373 (4)	-3.802	23177	1.564		1.259	0.8492	33
	Final Structure						0.8603	31
1ob3	None						0.8229	49
	Lys32 (1)*	-15.827	33044	1.614	2.378	1.371	0.8693	51

	Lys32 (2)	-13.602	51626	3.800		2.050	0.8103	40
	Lys32 (3)	-10.713	18284	3.470		2.043	0.8266	39
	Lys32 (4)	-10.196	28693	2.072		1.617	0.7974	47
	Phe79	-6.389	67631	0.278	0.341	0.290	0.8246	48
	Glu80 (1)	-10.458	93744	0.676	0.776	0.362	0.8229	46
	Glu80 (2)	-9.790	28080	1.077		1.503	0.8470	50
	Glu80 (3)*	-8.950	35290	2.922		3.050	0.8410	51
	His81 (1)	-9.333	26890	1.913	0.587	2.014	0.8329	50
	His81 (2)*	-4.576	14316	4.064		3.960	0.7905	46
	Lys88 (1)	-6.040	12452	1.015	3.951	3.844	0.8376	52
	Lys88 (2)	-5.197	34361	0.817		4.335	0.8049	53
	Lys88 (3)	-4.182	22279	2.407		4.079	0.8411	50
	Lys88 (4)*	-3.871	10683	4.171		4.738	0.8527	57
	Lys127 (1)	-6.175	18819	0.829	0.384	0.913	0.8419	50
	Lys127 (2)	-6.056	17326	1.151		1.410	0.8453	56
	Lys127 (3)*	-4.874	31034	1.779		1.889	0.8454	52
	Gln129 (1)	-6.687	24629	2.465	2.921	3.998	0.7823	54
	Gln129 (2)	-6.638	19506	2.217		1.909	0.8751	39
	Gln129 (3)	-6.378	15692	2.899		4.871	0.7154	41
	Gln129 (4)*	-6.301	16525	2.991		1.383	0.8671	53
	Asn130 (1)	-4.214	14966	2.443	0.432	2.441	0.8363	51
	Asn130 (2)*	-1.959	50788	1.674		1.666	0.8299	53
	Asn130 (3)	-0.338	13459	2.431		2.664	0.8596	53
	Asp143	-4.271	73273	0.545	0.382	0.707	0.8326	51
	Final Structure						0.8922	55
1pdb	None						0.9495	83
	Glu30 (1)	-21.961	36674	0.426	0.274	0.404	0.9571	82
	Glu30 (2)	-19.899	47956	2.122		2.104	0.9520	84
	Glu30 (3)	-12.878	26717	2.371		2.319	0.9166	85
	Glu30 (4)	-11.984	20544	2.531		2.484	0.9042	85
	Glu30 (5)*	-2.016	24105	3.157		3.154	0.9196	87
	Phe31	-14.277	49854	1.138	0.620	1.211	0.9562	82
	Tyr33	-19.387	31494	0.528	0.590	0.632	0.9562	82
	Phe34	-21.245	33391	0.437	0.264	0.314	0.9613	83
	Gln35 (1)	-15.933	36831	3.007	1.724	3.172	0.9439	83
	Gln35 (2)	-15.609	17377	0.427		1.861	0.9403	82
	Final Structure						0.9196	87
1phc	None						0.8654	29
	Phe87	-5.866	55232	0.261	1.087	0.940	0.8335	26
	Tyr96	-12.259	65972	0.822	0.956	0.777	0.8387	29

	Phe98	-7.995	72398	0.826	0.717	1.232	0.8494	24
	Asp297	-8.609	85637	1.132	0.340	1.185	0.8290	25
	Gln322	-7.733	11526	0.502	0.414	0.748	0.8467	26
	Final Structure						0.8654	29
2bls	None						0.8879	66
	Gln120 (1)	-17.946	21801	1.503	3.140	3.851	0.8534	66
	Gln120 (2)	-16.251	12191	2.484		4.119	0.8952	64
	Tyr150	-21.719	51526	0.841	0.238	0.798	0.6035	60
	Asn152 (1)	-31.551	29984	1.260	0.433	1.171	0.8801	62
	Asn152 (2)	-30.239	29605	1.845		2.109	0.9042	66
	Asn152 (3)*	-26.462	26113	3.005		2.753	0.9210	72
	Final Structure						0.9210	72

[a]Residues that have been selected as movable are listed. The individual rotamers that were selected based on our criteria are designated by the rotamer, ranked by energy, given in parentheses. The asterisk next to some rotamers indicates rotamers that were used to construct the modified apo conformation.
[b]The RMSD is for the atoms of the new rotamer and the corresponding residue in the apo structure.
[c]The RMSD is for the atoms of the original rotamer and the corresponding residue in the bound structure.
[d]The RMSD is for the atoms of the new rotamer and the corresponding residue in the bound structure.
[e]The mapping performance of the apo structure relative to that of the nonapo structure was scored based on the correlation of the nonbonded contacts of the clusters predicted by FTmap.

Table S6. Consensus site rankings and number of probe clusters for each bound, apo, and modified apo structure of the fifteen targets.

Target	Probe clusters (Total)	Hot spot ranking (number of probe clusters)						
TGT								
1n2v	44	1 (19)	4 (10)	6 (8)	7 (7)			
1pud	54	1 (32)	2 (14)	5 (8)				
1pud (modified)	64	1 (29)	2 (17)	3 (11)	7 (4)	9 (3)		
Carbonic anhydrase II								
1oq5	58	1 (19)	3 (14)	4 (14)	5 (9)	8 (2)		
2nxr	61	1 (28)	3 (13)	4 (10)	6 (6)	7 (3)	9 (1)	
2nxr (modified)	70	1 (21)	3 (16)	4 (13)	5 (10)	6 (9)	8 (1)	
Cyclin-dependent kinase 2								
1ke5	43	3 (14)	4 (13)	5 (11)	8 (3)	9 (2)		
1pw2	28	1 (17)	3 (11)					
1pw2 (modified)	46	1 (17)	2 (16)	4 (13)				
Dipeptidyl peptidase IV								
1n1m	23	2 (15)	6 (8)					
1pfq	26	3 (17)	5 (6)	9 (3)				
1pfq (modified)	27	2 (15)	8 (4)	9 (4)	14 (2)	15 (2)		
Acetylcholinesterase								
1gpk	61	2 (15)	3 (13)	4 (12)	6 (8)	7 (7)	8 (6)	
1ea5	75	1 (21)	2 (15)	3 (11)	4 (10)	7 (7)	8 (7)	9 (4)
Dihydrofolate reductase								
1ia1	81	1 (39)	2 (18)	3 (10)	4 (10)	5 (4)		
1ai9	81	1 (29)	2 (23)	3 (18)	4 (11)			
Neuraminidase								
1l7f	40	1 (24)	2 (16)					
2b8h	31	1 (19)	5 (8)	10 (4)				
nNOS								
1mmv	28	3 (15)	4 (8)	8 (5)				
1zvi	36	3 (18)	5 (9)	7 (5)	10 (4)			
1zvi (modified)	43	2 (18)	3 (12)	6 (7)	7 (6)			
Chitinase B (1e15)								
1w1p	89	1 (27)	2 (22)	3 (16)	4 (12)	5 (6)	6 (6)	
1e15	83	1 (23)	2 (21)	3 (16)	4 (11)	5 (7)	6 (5)	
1e15 (modified)	91	1 (33)	2 (24)	3 (13)	4 (11)	5 (5)	6 (3)	7 (2)
Carbonic Anhydrase XII (1jcz)								
1jd0	36	2 (21)	4(8)	5 (7)				
1jcz	28	3 (16)	4 (7)	7 (5)				

1jcz (modified)	35	2 (18)	3 (13)	5 (4)			
Neuraminidase							
1vcj	36	2 (21)	4 (8)	5 (7)			
1nsb	28	3 (16)	4 (7)	7 (5)			
1nsb (modified)	31	1 (26)	7 (5)				
Protein Kinase 5							
1v0p	50	1 (24)	3 (12)	4 (8)	6 (6)		
1ob3	49	1 (20)	3 (14)	4 (11)	9 (4)		
1ob3 (modified)	55	1 (23)	2 (15)	4 (10)	5 (7)		
Dihydrofolate reductase							
1s3v	75	1 (27)	2 (11)	4 (10)	5 (9)	6 (8)	8 (6)	9 (4)
1pdb	83	1 (30)	2 (17)	3 (15)	4 (11)	5 (10)		
1pdb (modified)	87	1 (31)	2 (18)	3 (13)	4 (8)	5 (7)	6 (7)	7 (3)
Cytochrome P450cam (1phc)							
1p2y	37	1 (22)	6 (7)	7 (7)	12 (1)		
1phc	29	1 (29)					
Beta-lactamase							
1l2s	69	1 (26)	2 (14)	3 (13)	4 (10)	5 (6)	
2bls	66	1 (18)	2 (12)	3 (12)	4 (10)	4 (9)	9 (5)
2bls (modified)	72	1 (16)	2 (16)	3 (14)	4 (10)	5 (8)	6 (8)

Of the residues that are considered by FTFlex, two trends were observed. First, in the case of Phe, Trp, Tyr and His, the best mapping results were obtained when the rotamer position aligned with that of the bound structure. For these residues the RMSD between the modified structure and the bound structure is much improved compared to the initial apo and bound structure, 1.81 versus 3.31 Å, respectively (Table S7). On the other hand, for Lys and Arg, there is no apparent improvement in the RMSD (Table S7). These results suggest that Phe, Trp, Tyr and His require specific conformations in the bound state. In the case of Lys and Arg, these residues simply need to be moved out of the binding site to yield improved mapping results. This result is consistent with the observation that Lys and Arg generally collapse into the active site upon crystal formation or are poorly defined in the X-ray crystallographic data (Teague, 2003; Davis et al., 2003).

Table S7. Average RMSDs (Å) of residues selected for use in the modified structure.

Averages	Apo vs bound	Apo vs modified	Modified vs. Bound
Phe, Trp, Tyr, His	3.31	4.3	1.81
Lys, Arg	2.2	2.38	2.18
Total	1.53	1.19	2.13

Examples of FTFlex applied to targets from the test set

In the following sections we briefly discuss the results obtained using FTFlex for four different targets from the data set.

tRNA-guanine transglycosylase.

tRNA-guanine transglycosylase (TGT) catalyzes the exchange of the queuine precursor, preQ1, with the guanine at position 34 in tRNAs specific for Asn, Asp, His, and Tyr (Romier et al., 1996). Mutation of the tgt gene in *Shigella flexneri* results in a significant loss of pathogenicity due to the reduced synthesis of the virF protein, which is a key regulatory protein for the expression of virulence (Durand et al., 1994). Hence, TGT is an appealing target for rational drug design against *Shigellae*, the causative agents of dysentary. The structure of *Zymomonas mobilis* TGT, which is identical to that of *S. flexneri* TGT with the exception of a Tyr106 → Phe 106 mutation, has served as the starting point for structure-based drug design. The X-ray crystallographic data of *Z. mobilis* reveal an irregular $(\beta/\alpha)_8$-barrel structure with a C-terminal zinc-binding domain. Soaking crystals of TGT in the presence of PreQ1 results in the binding of PreQ1 in the barrel pocket on the C-terminal face (Romier et al., 1996). PreQ1 intercalates between Tyr106 and Met260, with the side chain of Gln107 oriented towards the purine ring. PreQ1 is also closely positioned next to Asp102, which is the residue believed to act as a nucleophile in the proposed associative mechanism based on site-directed mutagenesis studies (Kittendorf et al., 2001).

Mapping results of the apo structures. The apo structure was taken from the X-ray crystallographic data in PDB file 1pud (1.85 Å resolution) (Romier et al., 1996). Mapping of this structure predicts three consensus sites within the active site (Figure S2, left). The highest ranked consensus site lies at the entrance of the active site in a region of primarily hydrophilic residues. The other two consensus sites (corresponding to the second and eighth ranked sites) are located below CS1 (consensus site 1) and are also relatively solvent-exposed (Figure S2); although they do not overlap with known inhibitors. All three consensus sites were chosen as the cluster around which to search for movable amino acid side chains.

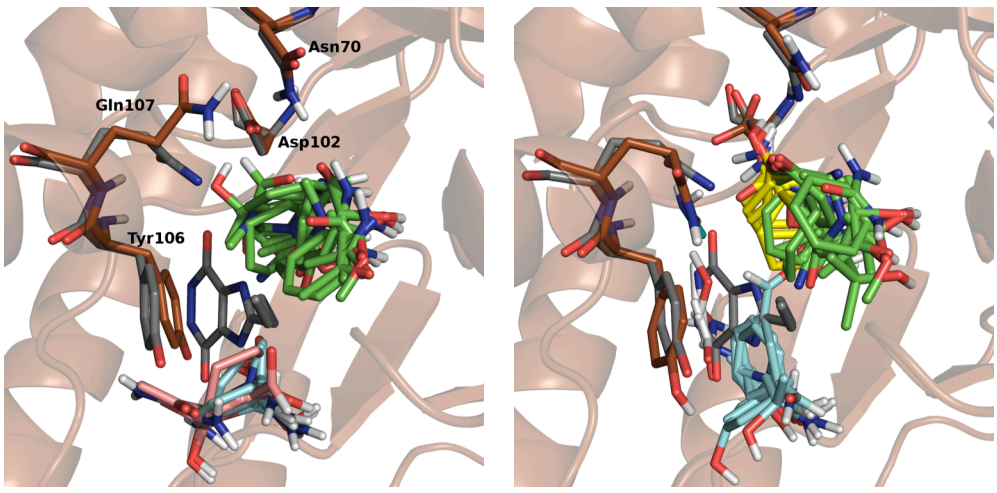

Figure S2. Mapping results (colored sticks) are shown for TGT (PDB file 1pud, copper) prior to (left) and after (right) FTFlex. The residues and ligand of the bound structure (PDB file 1n2v) are shown in grey. Residues of interest are shown as sticks and labeled (left).

FTFlex identified Asn70, Tyr106, and Gln107 as the movable residues within the active site of the apo structure (Table S5). The biggest conformational difference involves Tyr106. Between the apo and the modified apo structures Tyr106 has rotated approximately 90 degrees around Cβ orienting the phenol ring towards solvent and resulting in an RMSD of 5.176 Å compared to its original position. Mapping the modified apo structure yields an improvement in the total number of probe clusters from 54 to 64. The highest ranked consensus site is similarly positioned at the entrance of the active site. Inhibitors that take advantage of this binding site have been shown to have nanomolar inhibition (Meyer et al., 2002; Brenk et al., 2004; Stengl et al., 2007). Three additional consensus sites occupy a deeper binding pocket similar to the binding of PreQ1 and known inhibitors, with probe clusters in consensus site 2 (cyan clusters in Figure S2, right) participating in π-stacking with Tyr106.

Comparison of the mapping results of the apo and bound structures. The X-ray crystallographic data of the bound structure came from protein data bank (PDB) file 1n2v (2.1 Å resolution) (Brenk et al., 2003b). The inhibitor of this structure is 2-butyl-5,6-dihydro-1H-imidazo[4,5-D]pyridazine-4,7-dione and has 1a Ki of 83 ± 18 μM. In the bound structure, the largest conformational change is the opening up of the PreQ1 binding pocket upon inhibitor binding via the movement of Tyr106 (Table S5 and Figure S2) (Brenk et al., 2003a). The phenol side chain of Tyr106 is oriented towards solvent to facilitate π-stacking interactions between the aromatic ring of phenol and the aromatic heterocyclic ring of the inhibitor. The only other significant movement involves Gln107, which has rotated approximately 180 degrees around Cγ in the bound structure.

Mapping of the bound structure yields four consensus sites within the active site (Table S6 and Figure S2), in good agreement with those predicted for the modified apo structure. The movement of Tyr106 is key for opening up the buried ligand binding site. The RMSD of Tyr106 in the apo and bound structure improves from 5.282 Å to 0.684 Å for the modified apo structure. To further evaluate the success of FTFlex, correlation coefficients were calculated between the nonbonded contacts of the probe clusters of the bound and apo structures (Table S5 and Figure S1a). For the unaltered apo and bound structures, the correlation coefficient is 0.7540. However, this number significantly improves to 0.9000 when the mapping results of the modified apo structure are compared to those of the bound structure.

Carbonic anhydrase II.

Carbonic anhydrases (CAs) catalyze the Zn(II)-dependent reversible hydration of carbon dioxide into hydrogen carbonate and a proton (Supuran et al., 2003). There are fifteen isozymes known in vertebrates, all with different physiological roles. Some are localized in the cytosol or mitochondria, while others are membrane-bound with extra-cellular active site domains. Although there are some variations in the amino acid sequence, the active site for all CA isozymes are quite similar, with the catalytic Zn occupying a deep conical site (Fisher et al., 2007). The Zn^{2+} ion is coordinated by three His residues and a hydroxide ion. The catalysis of the hydration and dehydration of CA II occurs in two distinct steps and requires a proton transfer from solvent, facilitated by His64 (Nair and Christianson, 1991), which is located approximately 7.5 Å away from the catalytic Zn (Fisher et al., 2007). Thus His64 can adopt an inward (directed towards the active site) or outward (directed away from the active site)

conformation. Carbonic anhydrases (CAs) are widely targeted in the prevention of many diseases, including glaucoma, hypertension, diabetes (Scozzafava et al., 2004); some evidence suggests that CAs may also be a potential target for cancer (Cecchi et al., 2004). Inhibitors of CA predominantly are anchored to the zinc ion through sulfonamide groups (Abbate et al., 2002). Unfortunately, the lack of tissue-selective and isozyme-specific inhibition of CAs results in unwanted side effects; the design of isozyme-specific inhibitors remains a critical challenge.

Mapping results of the apo structures. The apo structure of CA II was taken from PDB file 2nxr (1.70 Å resolution) (Fisher et al., 2007) and mapped. The mapping results are shown in Figure S3 (left) and Tables S5 and S6. There are a total of 61 probe clusters found with the active site; 38 of those probe clusters correspond to the region of sulfonamide binding near the zinc ion (green and yellow sticks). An additional large binding site is also identified near Ile91, Gln92, Gln69, and Asn67 (purple sticks).

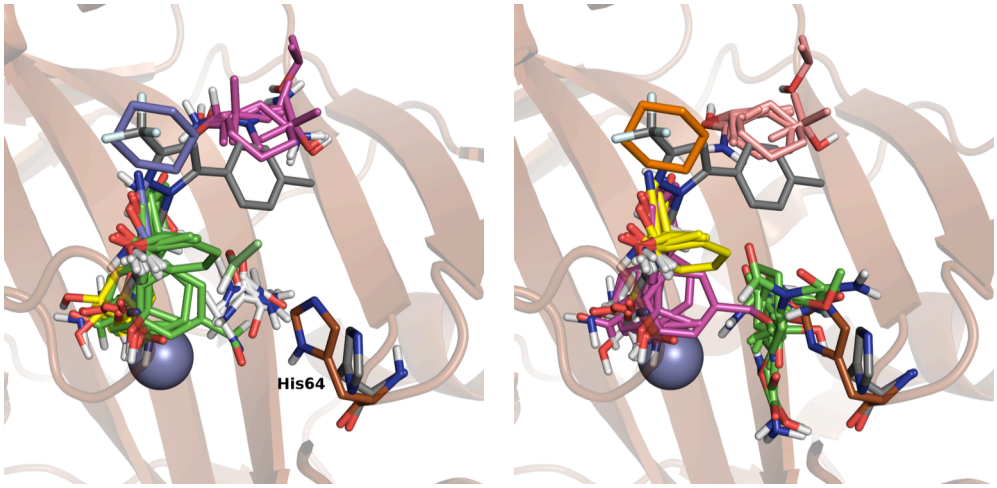

Figure S3. Mapping results (colored sticks) are shown for CA II (PDB file 2nxr, copper) prior to (left) and after (right) FTFlex. The residues and ligand of the bound structure (PDB file 1oq5) are shown in grey. Residues of interest are shown as sticks and labeled (left).

FTFlex identified His64 as the only flexible side chain. In the modified apo structure, His64 is oriented away from the active site, with an RMSD of 4.115 Å from its original position (Table S5). This movement creates a new hot spot as shown in Figure S3 (right, green sticks). This new hot spot is flanked by Tyr7, Asn62, His64, His94 and Thr200. Indeed, this new site is known to be a key binding site for recently discovered nanomolar inhibitors of CA II and CA IV (Vernier et al., 2010). The hot spots identified for the unmodified structure were also conserved in the mapping of the modified structure.

Comparison of the mapping results of the apo and bound structures. The bound structure of CA II was also mapped in the absence of its ligand. In this case the structure was taken from PDB file 1oq5 (1.50 Å resolution) (Weber et al., 2004). The inhibitor in this structure is celecoxib, which has a K_i of 21 nM and is also a cyclooxygenase-2 (COX-

2) specific inhibitor. The mapping results of the bound structure yield a total of 64 probe clusters within the active site. The majority of these probe clusters are found in the vicinity of the zinc ion (39 probe clusters). Two other hot spots are identified. One corresponds to the hot spot shown in Figure S3 (right) as pink sticks with 9 probe clusters. The other is the hot spot opened up by the movement of His64.

The correlation coefficient calculated for the bound and initial apo structure is 0.3533, but is a much improved value of 0.7089 for the modified and bound structures (Table S5 and Figure S1b). The RMSD of His64 has improved from 4.115 for the apo versus bound structure to 2.169 Å for the modified versus bound structure. Thus, the RMSD is still quite large, but the structure output by FTFlex is in significantly better agreement with the conformation and the mapping results of the bound structure.

Cyclin-dependent kinase 2.

The cyclin-dependent kinases (CDK) play a key role in cell cycle regulation. Cyclin-dependent kinase 2 (CDK2) in particular is involved in the regulation of the G1 phase and mitosis (Morgan, 1997). Activation of CDK2 requires the binding of a cyclin (cyclin E in the G1 phase and cyclin B during mitosis), which entails minimal activity, as well as phosphorylation of a threonine residue, which yields full activity (Pavletich, 1999). The binding of cyclin and a ligand to CDK2 results in large conformational changes to yield the fully aligned catalytic triad, namely Lys33, Glu51, and Asp145, within the active site (Wu et al., 2003). The catalytic triad lies in a cleft between the N-terminus β-sheet region and the C-terminus α-helical domain. This is the same pocket in which ATP binds, with the triphospodiester oriented towards the catalytic triad (Schulze-Gahmen et al., 1996). In the inactive state (non-cyclin bound), helix 1 is flipped, exposing Glu51 to the solvent; however, much of the ATP-binding pocket remains intact. This binding pocket has attracted interest in the use of protein kinase inhibitors as drugs against cancer and similar proliferative diseases due to the observation that the activity of kinases is deregulated in many tumor cells (Helal et al., 2009; Bramson et al., 2001).

Mapping results of the apo structures. The apo structure was taken from PDB file 1pw2 (1.95 Å resolution) (Wu et al., 2003). Mapping of this structure yields two consensus sites with a total probe cluster count of 28 (Figure S4, left and Tables S5 and S6). The highest ranked consensus site is shown as green and lies in the adenosine binding pocket. Above this site is a second consensus site which corresponds to the phosphodiester binding site of ATP. These two sites were chosen as the cluster around which to search for movable amino acid side chains.
FTFlex identified Lys33 and Lys89 as flexible binding site residues. The rotamer selected for Lys33 has moved out of the binding pocket; a 2.916 Å RMSD from the initial position of the residue. In the case of the selected rotamer of Lys89, the RMSD between the final selected rotamer and the initial position in the apo structure is quite large (5.423 Å), with the side chain rotating approximately 180 degrees around Cβ.

The modified apo structure was mapped and the results are shown in Table S6 and Figure S4. The highest ranked consensus site (green sticks) has now moved to occupy the position originally occupied by Lys33; this is also at the apex of the catalytic triad. Similarly, the third highest consensus site (purple sticks) occupies the site originally

occupied by Lys89. A consensus site is also found in the phosphodiester binding region as was observed for the original apo structure. Most notably, the improvement in the number of probe clusters (Table S5) is approximately additive in this case. By itself, movement of Lys33 yielded an additional 16 probe clusters while movement of Lys89 resulted in 6 additional probe clusters. Inserting both new rotamers into the final structure yields an additional 24 probe clusters.

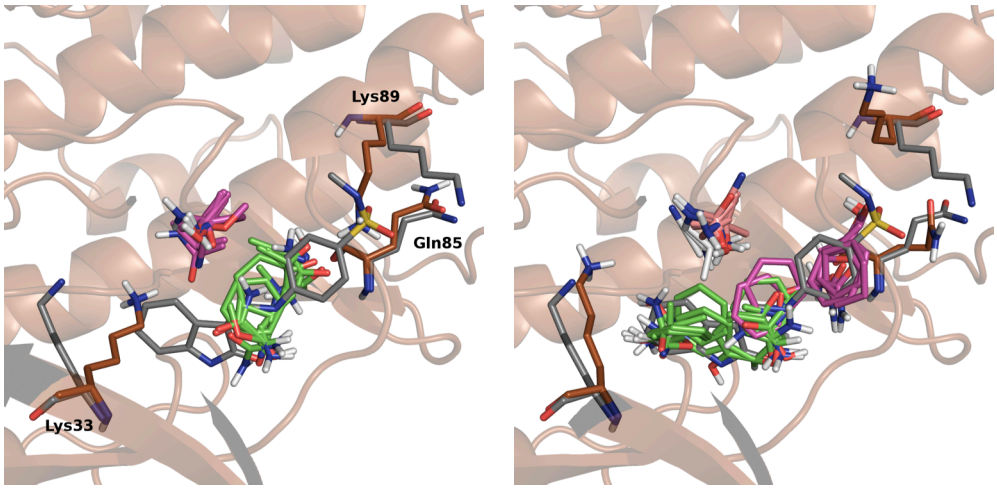

Figure S4. Mapping results (colored sticks) are shown for CDK2 (PDB file 1pw2, copper) prior to (left) and after (right) FTFlex. The residues and ligand of the bound structure (PDB file 1ke5) are shown in grey. Residues of interest are shown as sticks and labeled (left).

Comparison of the mapping results of the apo and bound structures. As mentioned above, while the CDK2 enzyme undergoes substantial conformational movement upon cyclin binding, the ATP binding site remains accessible in both states (Schulze-Gahmen et al., 1996). Hence, inhibitors have been targeted to lock CDK2 in the inactive state. The bound structure considered in this case is that of PDB file 1ke5 with inhibitor N-methyl-4-[2-oxo-1,2-dihydro-3H-indol- 3-ylidene) methyl] amino benzenesulfonamide (labeled and referred to as LS1 for the remained of this discussion); inhibitor LS1 exhibits an LC50 of 560 nM (Bramson et al., 2001). In this structure, the indolin-2-one ring sits in proximity to Lys33 and Asp145 of the catalytic triad, while the sulfonamide end binds closer to the surface of the cleft, with hydrogen bond contacts to Lys89 and the backbone amide of Asp86.

Mapping results of 1ke5 reveal 4 main hot spots within the binding cleft. The consensus sites determined for 1ke5 agree well with those determined for the final modified structure of 1pw2. The deep binding pocket near the catalytic triad (corresponds to the green sticks in Figure S4, right) and the phosphodiester (corresponds to the magenta sticks in Figure S4, right) binding sites are both identified. Notably, the mapping results for the final apo structure align very well with the ligand from the bound structure. This observation is supported by the improvement in the correlation from 0.4803 for the initial apo structure to 0.6603 for the final structure (Figure S1c).

As noted above, both Lys33 and Lys89 are important for opening up the binding site of the apo structure. In the case of Lys33, the RMSD between the initial apo and bound structures is 3.657 Å. The RMSD of Lys33 between bound and apo structure has decreased to 1.517 Å for the rotamer used in the final structure. In the case of Lys89, the RMSD between the bound and apo was even larger at 4.103 Å; however, the RMSD of the new rotamer predicted by FTFlex is 5.944 Å. In this case, selecting a rotamer that allowed for the side chain to be removed from the binding site gave sufficient results.

Dihydrofolate reductase

Dihydrofolate reductase (DHFR) catalyzes the reduction of dihydrofolic acid to tetrahydrofolic acid, using NADPH as the electron donor. DHFR is an important target for a number of organisms, including bacterial and fungal species (De Pauw and Meunier, 1999). In particular, life-threatening disease caused by organisms such as *Candida albicans*, is now relatively common among patients with compromised immune systems (Barnes, 2008). The structure of DHFR from *C. albicans* reveals a nine-stranded β-sheet core, flanked by α-helices on either side. Inhibitors are known to bind deep in the folic acid binding pocket, which is wider in the DHFR structure of *C. albicans* relative to the structure from humans (Whitlow et al., 2001). NADPH binds next to the folic acid binding pocket, with its nicotinamide ring positioned below the hydroxypteridin ring of folic acid.

Mapping results of the apo structures. The apo structure was taken X-ray crystallographic data from PDB file 1ai9 (resolution of 1.85 Å) (Whitlow et al., 2001). Mapping calculations revealed a total of 81 probe clusters within the folic acid/NADPH binding pocket (Figure S5). Of these probe clusters, the first and second highest count consensus site corresponded to the binding region of the hydropteridin ring of folate and the nicotinamide ring of NADPH. Indeed, inhibitors that took advantage of both the folate and NADPH binding pockets were found to more selective towards *C. Albicans* (Whitlow et al., 2001). The third and fourth consensus sites correspond to the binding sites of the ribose ring of NADPH and the phenyl ring of folate, respectively. FTFlex identified fourteen amino acids in the vicinity of the probe clusters. Of these amino acids, no rotamers were found to improve the mapping results. Hence, the initial mapping results were not improved by FTFlex.

Comparison of the mapping results of the apo and bound structures. The bound structure of DHFR from *C. Albicans* was taken from PDB file 1ia1 (1.70 Å resolution) (Whitlow et al., 2001). This structure includes both NADPH and the inhibitor, 5-phenylsulfanyl-2,4-quinazolinediamine, which has an IC_{50} of 0.034 μM. The largest differences between the binding site of the bound and apo structures are between Met25 and Glu32, with RMSDs of 0.645 and 0.674 Å, respectively. The mapping results of 1ia1 yield five consensus sites within the binding pocket for a total of 80 probe clusters, which is in good agreement to the number found for the apo structure. The first and third consensus sites overlap with the diaminoquinazoline ring system of the inhibitor, while the fourth consensus site corresponds to the phenyl ring. The consensus sites found for the bound structure are in good agreement with those determined for the apo structure, with a correlation of 0.9751 (Figure S1d). The high correlation indicates a high degree of similarity between the binding sites of the apo and bound structures, indicating that no conformational changes

are needed for the apo structure. Indeed, this result supports our findings for the apo structure, in which no movable amino acids were found by FTFlex.

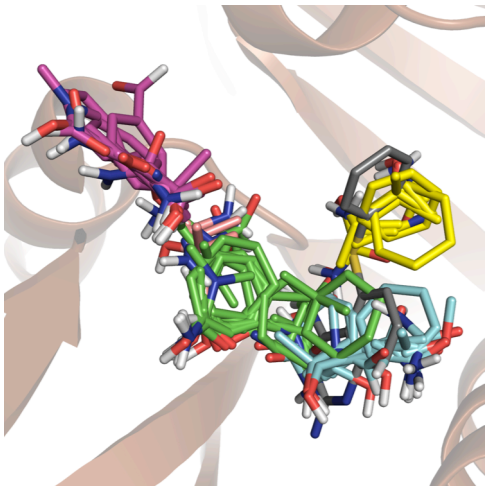

Figure S5. Mapping results (colored sticks) are shown for DHFR (PDB file 1ai9, copper). The ligand of the bound structure (PDB file 1ia1) is shown in grey.

SUPPLEMENTAL REFERENCES

Abbate, F. e*t al.* (2002) Nonaromatic sulfonamide group as an ideal anchor for potent human carbonic anhydrase inhibitors: Role of hydrogen-bonding networks in ligand binding and drug design. *J. Med. Chem.*, **45**, 3583-3587.

Barnes, R. A. (2008) Early diagnosis of fungal infection in immunocompromised patients. J. Antimicrob. Chemother., 61(suppl 1), i3-i6.

Beglov, D.*, et al.* (2012) Minimal ensembles of side chain conformers for modeling protein-protein interactions, *Proteins*, **80**, 591–601.

Bramson, H. N. e*t al.* (2001) Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): Design, synthesis, enzymatic activities, and x-ray crystallographic analysis. *J. Med. Chem.*, **44**, 4339-4358.

Brenk, R. e*t al.* (2003) Flexible adaptations in the structure of the tRNA-modifying enzyme tRNA-guanine transglycosylase and their implications for substrate selectivity, reaction mechanism and structure-based drug design. *ChemBioChem*, **4**, 1066-1077.

Brenk, R. e*t al.* (2003) Virtual screening for submicromolar leads of RNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem.*, **46**, 1133-1143.

Brenk, R. e*t al.* (2004) Crystallographic study of inhibitors of tRNA-guanine transglycosylase suggests a new structure-based pharmacophore for virtual screening. *J. Mol. Biol.*, **338**, 55-75.

Brenke, R. *et al*. (2009) Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics*, **25**, 621-627.

Brooks, B. R. *et al*. (1983) Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, **4**, 187-217.

Cecchi, A. e*t al.* (2004) Carbonic anhydrase inhibitors: synthesis and inhibition of cytosolic/tumor-associated carbonic anhydrase isozymes I, II, and IX with sulfonamides derived from 4- isothiocyanato-benzolamide. *Bioorg. Med. Chem. Lett*., **14**, 5775 - 5780.

Chuang, G.-Y. *et al*. (2008) Dars (decoys as the reference state) potentials for protein-protein docking. *Biophys. J*., **95**, 4217-4227.

Chuang, G.-Y. *et al*. (2010) Domain motion and inter-domain hot spots in a multi-domain enzyme. *Protein Sci.,* **19,** 1662-1672.

Davis, A. M. *et al*. (2003) Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem., Int. Ed.*, **42**, 2718-2736.

De Pauw, B. E. and Meunier, F. (1999) The challenge of invasive fungal infection. *Chemotherapy (Basel, Switz.)*, **45**, 1-14.

Durand, J. M. B. e*t al.* (1994) Transfer rna modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *shigella flexneri*: The expression of the virF gene. *Mol. Microbiol.*, **35**, 924-935.

Fisher, S. Z. e*t al.* (2007) Speeding up proton transfer in a fast enzyme: Kinetic and crystallographic studies on the effect of hydrophobic amino acid substitutions in the active site of human carbonic anhydrase II. *Biochemistry*, **46**, 3803-3813.

Hall, D. H. *et al.* (2011) Robust identification of binding hot spots using continuum electrostatics: Application to hen egg-white lysozyme. *J. Am. Chem. Soc.*, **133**, 20668-20671.

Helal, C. J. e*t al.* (2009) Potent and cellularly active 4-aminoimidazole inhibitors of cyclin-dependent kimase 5/p25 for the treatment of Alzheimer's disease. *Bioorg. Med. Chem.*, **19**, 5703-5707.

Kittendorf, J. D. e*t al.* (2001) tRNA-guanine transglycoslyase from *Escherichia coli*: Molecular mechanism and role of aspartate 89. *Biochemistry*, **40**, 14123-14133.

Kozakov, D. *et al*. (2011) Structural conservation of druggable hot spots in protein-protein interfaces. *Proc. Nat. Acad. Sci*., **108**, 13528-13533.

Landon, M. *et al.* (2008) Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. *Chem Biol Drug Des.,* **71,** 106-116.

Ma, B. *et al*. (2002) Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations. *Protein Sci.*, **11**, 184-197.

Meyer, E. A. *et al*. (2002) De novo design, synthesis, and in vitro evaluation of inhibitors for prokaryotic tRNA-guanine transglycosylase: A dramatic sulfur effect on binding affinity. *ChemBioChem*, **3**, 250-253.

Morgan, D. O. (1997) Cyclin-dependent kinases: Engines, clocks, and microprocessors. *Annu. Rev. Cell Dev. Biol.*, **13,** 261-291.

Nair, S. K. and Christianson, D. W. (1991) Unexpected ph-dependent conformation of His-64, the proton shuttle of carbonic anhydrase II. *J. Am. Chem. Soc.*, **113**, 9455-9458.

Najmanovich, R. *et al*. (2000) Side-chain flexibility in proteins upon ligand binding. *Proteins: Struct., Funct., and Genet.*, **39**, 261-268.

Pavletich, N. P. (1999) Mechanisms of cyclin-dependent kinase regulation: structures of CDKs, their cyclin activators, and CIP and INK4 inhibitors,. *J. Mol. Biol.*, **287**, 821 - 828.

Romier, C. *et al*. (1996) Crystal structure of tRNA-guanine transglycoslyase RNA modification by base exchange. *EMBO J*., **15**, 2850-2857.

Schulze-Gahmen, U. e*t al.* (1996) High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: Bound waters and natural ligand as guides for inhibitor design. *J. Med. Chem*., **39**, 4540-4546.

Stengl, B. *et al*. (2007) Crystal structures of tRNA-guanine transglycosylase (TGT) in complex with novel and potent inhibitors unravel pronounced induced-fit adaptations and suggest dimer formation upon substrate binding. *J. Mol. Biol.*, **370**, 492-511.

Teague, S. J. (2003) Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discovery*, **2**, 527-541.

Scozzafava, A. e*t al.* (2004) Modulation of carbonic anhydrase activity and its applications in therapy. *Expert Opin. Ther. Pat.*, **14**, 667-702.

Stengl, B. e*t al.* (2007) Crystal structures of tRNA-guanine transglycosylase (TGT) in complex with novel and potent inhibitors unravel pronounced induced-fit adaptations and suggest dimer formation upon substrate binding. *J. Mol. Biol*., **370**, 492-511.

Supuran, C. T. e*t al.* (2003) Carbonic anhydrase inhibitors. *Med. Res. Rev*., **23**, 146-189.

Verdonk, M. L.*et al*. (2008) Protein-ligand docking against non-native protein conformers. *J. Chem. Inf. Model*., **48**, 2214-2225.

Vernier, W. e*t al.* (2010) Thioether benzenesulfonamide inhibitors of carbonic anhydrases II and IV: Structure-based drug design, synthesis, and biological evaluation. *Bioorg. Med. Chem*., **18**, 3307 - 3319.

Weber, A. e*t al.* (2004) Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: New pharmacological opportunities due to related binding site recognition. *J. Med. Chem*., **47**, 550-557.

Whitlow, M. e*t al.* (2001) X-ray crystal structures of *Candida albicans* dihydrofolate reductase: High resolution ternary complexes in which the dihydronicotinamide moiety of NADPH is displaced by an inhibitor. *J. Med. Chem*., **44**, 2928-2932.

Zhao, S. *et al*. (2001) Analysis of a data set of paired uncomplexed protein structures: New metrics for side-chain flexibility and model evaluation. *Proteins: Struct., Funct., and Genet*., **43**, 271-279.