Published in final edited form as:

Nat Protoc. 2015 May ; 10(5): 733–755. doi:10.1038/nprot.2015.043.

The FTMap family of web servers for determining and characterizing ligand binding hot spots of proteins

Dima Kozakov^{a,*}, Laurie E. Grove^b, David R. Hall^c, Tanggis Bohnuud^a, Scott Mottarella^d, Lingqi Luo^d, Bing Xia^a, Dmitri Beglov^a, and Sandor Vajda^{a,*}

Dima Kozakov: vajda@bu.edu; Laurie E. Grove: grovel@wit.edu; David R. Hall: david@acpharis.com; Tanggis Bohnuud: tanggis@bu.edu; Scott Mottarella: semottarella@gmail.com; Lingqi Luo: marsluo@bu.edu; Bing Xia: sixpi@bu.edu; Dmitri Beglov: dbeglov@bu.edu; Sandor Vajda: Vajda@bu.edu

^aDepartment of Biomedical Engineering, Boston University, Boston, MA 02215, USA

^bDepartment of Sciences, Wentworth Institute of Technology, Boston, MA 02115, USA

^cAcpharis Inc., Holliston, MA 01746, USA

^dProgram in Bioinformatics, Boston University, Boston, MA 02215, USA

Abstract

FTMap is a computational mapping server that identifies binding hot spots of macromolecules, i.e., regions of the surface with major contributions to the ligand binding free energy. To use FTMap, users submit a protein, DNA, or RNA structure in PDB format. FTMap samples billions of positions of small organic molecules used as probes and scores the probe poses using a detailed energy expression. Regions that bind clusters of multiple probe types identify the binding hot spots, in good agreement with experimental data. FTMap serves as basis for other servers, namely FTSite to predict ligand binding sites, FTFlex to account for side chain flexibility, FTMap/param to parameterize additional probes, and FTDyn to map ensembles of protein structures. Applications include determining druggability of proteins, identifying ligand moieties that are most important for binding, finding the most bound-like conformation in ensembles of unliganded protein structures, and providing input for fragment based drug design. FTMap is more accurate than classical mapping methods such as GRID and MCSS, and is much faster than the more recent approaches to protein mapping based on mixed molecular dynamics. Using 16 probe molecules, the FTMap server finds the hot spots of an average size protein in less than an hour. Since FTFlex performs mapping for all low energy conformers of side chains in the binding site, its completion time is proportionately longer.

*Corresponding authors: Sandor Vajda, vajda@bu.edu, 1-617-353-4757, Dima Kozakov, midas@bu.edu, 1-617-353-4742.

Author contributions statements

D.K., D.R.H., T.B., B.X., and S.M. developed the servers; D.K., L.E.G., D.R.H., T.B., D.B, L.L., and B.X. performed experiments; L.E.G., D.R.H., and S.V. prepared the manuscript with editing by T.B., S.V., and L.E.G.

Competing financial interests

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/nprot/index.html>.

Keywords

ligand-protein interaction; ligand binding site; drug discovery; druggability; fragment based drug design

INTRODUCTION

The interactions of macromolecules (proteins, DNA, and RNA) with other macromolecules and small ligands are at the core of many biological fields. The nature of these interactions is important for understanding fundamental biological processes, as well as applications in drug discovery. It has been established that the binding sites of macromolecules include smaller regions called hot spots that are major contributors to the binding free energy, and hence are crucial to binding any ligand at that particular site^{1–3}. This concept was originally introduced in the context of mutating interface residues to alanine in protein-protein or protein-peptide interfaces^{4–7}. On the basis of this method, a residue is considered a hot spot if its mutation to alanine gives rise to a substantial drop in binding affinity. An alternative experimental method to determine binding hot spots, more directly related to the binding of small ligands, is based on screening libraries of fragment-sized organic molecules for binding to the target protein⁸. A fundamental property of hot spots is their capability to bind a variety of small organic probe molecules^{3,8–10}. Since the binding of the small compounds is very weak, the interactions are most frequently detected by X-ray crystallography^{11–13} or nuclear magnetic resonance (NMR)⁸. In the multiple solvent crystal structures (MSCS) method, X-ray crystallography is used to determine the structure of the target protein soaked in aqueous solutions of 6–8 organic solvents used as probes. By superimposing the structures, regions that bind multiple different probes can be detected^{11,12}. While individual probes may bind at a number of locations, their clusters indicate binding hot spots. Similarly, in the structure-activity relationship (SAR) by NMR method, proteins are immersed in a series of organic solvents and perturbations in residue chemical shifts are used to identify residues that participate in small molecule binding⁸. It was shown that the small “probe” ligands cluster at hot spots, and the hit rate predicts the importance of the site^{8,11}. The NMR based screening correctly identified known druglike molecule binding sites in 94% of cases within a set of 23 target proteins, and the method has been extended to a much larger test set⁸. While the existence of binding hot spots has been experimentally verified beyond doubt, there is no generally accepted explanation for their origin. Based on simulations, our hypothesis is that hot spots are distinguishable from other regions of the protein due to their concave topology combined with a mosaic-like pattern of hydrophobic and polar functionality^{9,14,15}.

The main advantage of studying hot spots is that they are less sensitive to conformational changes than binding sites are, and can be identified in almost any structure of a protein, including ones without a bound ligand^{14–17}. The knowledge of hot spots is very valuable for a variety of applications. First, hot spots identify the most important regions of binding sites that should be considered when exploring macromolecule-ligand interactions. Second, the strength of hot spots determines druggability of a site, defined as the ability of a site to bind drug-size compounds with at least low micromolar affinity^{9,18–21}. Third, an important

application is the identification of binding sites²². Fourth, since hot spots are the energetically important regions of binding sites, the ligand moieties interacting with hot spots are the ones that are essential for binding²³. Fifth, determining hot spots provides information on the importance of residues in protein-protein interfaces. In particular, it was shown that over 90% of side chains at such interfaces that are identified as hot spots by alanine scanning protrude into hot spots of the partner protein²⁴. Finally, probably the most important use of hot spot determination is providing input for fragment-based ligand discovery (FBLD). FBLD is a combinatorial approach in which individual fragments binding to regions of the target site are selected from a fragment library, and then combined to form potential lead compounds^{25–27}. Hot spots identify the important subsites, determine their druggability^{18,19,28}, help to select an appropriate fragment library, and once fragment hits are identified, can be used in optimally extending such fragment hits into higher affinity ligands²⁹.

Experimental techniques to determine binding hot spots are time consuming and can be limited by the physical constraints of the protein-solvent system. Here we describe a protocol using the FTMap family of web servers for determining and characterizing binding hot spots using computational approaches that can replace these experimental methods, and provide the benefit of ease of use (Table 1). Each algorithm has been developed for a specific application and has been implemented as a separate server. The basic algorithm and server is FTMap, a close computational analogue of the X-ray crystallography or NMR based screening experiments¹⁴. FTMap provides direct information on binding hot spots and their druggability, and can be used for extending fragment hits into larger ligands. The second server is FTSite, aimed at the identification of ligand binding sites based on the structure of ligand-free proteins²². The third server is FTFlex, which performs repeated mapping calculations while exploring low energy conformers of side chains in the vicinity of hot spots³⁰, primarily for opening pockets in protein-protein interfaces that have the potential to bind small ligands²⁸. The fourth server is FTMap/param³¹, which can be used to determine whether small molecules selected by the user bind in the hot spot regions predicted by FTMap. Finally, the FTDyn server has been developed to map ensembles of conformationally diverse structures, obtained by Nuclear Magnetic Resonance (NMR) experiments³² or by molecular dynamics (MD) simulations^{21,33}. While FTSite, FTFlex, FTMap/param, and FTDyn all are built on the FTMap algorithm, there are slight differences in the details of implementation, and the methods serve very different applications (Table 1). In addition, FTMap already had a sizeable user base by the time the other servers were developed. In view of these factors and in order to retain the simplicity of use, we decided to implement each algorithm as a separate server rather than creating a single server with a complex interface and potentially confusing presentation of results.

The FTMap algorithm and server

FTMap has been developed as a close computational analogue of the X-ray crystallography or NMR based screening experiments¹⁴. The method distributes small organic probe molecules of varying size, shape, and polarity on a macromolecule surface, finds the most favorable positions for each probe type, then clusters the probes and ranks the clusters on the basis of their average energy (Fig. 1a). FTMap¹⁹ uses 16 organic molecules as probes

(ethanol, isopropanol, isobutanol, acetone, acetaldehyde, dimethyl ether, cyclohexane, ethane, acetonitrile, urea, methylamine, phenol, benzaldehyde, benzene, acetamide, and N,N dimethylformamide, see Fig. 1b). Regions that bind several different probe clusters are called consensus sites (CSs), and the site containing the largest number of probe clusters is considered the main hot spot; all other consensus sites are secondary hot spots. Since the hot spots are identified by consensus sites, and in turn the consensus sites are defined by consensus clusters, we will use these terms somewhat interchangeably when computational mapping is discussed. For each consensus site, each probe cluster contained is shown as a single structure representing the cluster center (Fig. 1c). Because the existence of hot spots is a fundamental property of a protein structure, unbound structures can be used as the input and thus no information related to ligand binding is required. Even in cases where there exist conformational differences between unbound and bound structures, the structural features of the hot spot are robust enough to be identified in the unbound structure^{28,34}.

The only input required for FTMap is a protein, DNA, or RNA structure, typically obtained by X-ray crystallography or NMR techniques. The program removes all bound ligands and water molecules that the structure file may contain (Fig. 2)^{19,35}. For each probe, the algorithm generates bound positions using rigid body docking. This step uses the Fast Fourier Transform (FFT) correlation approach to sample billions of probe positions on translational and rotational grids, consisting of 0.8 Å translations and of 500 rotations at each location. The goal of using the FFT algorithm is to speed up the calculation. In fact, while the same systematic sampling could be performed in real space, for an average size protein it would take at least 1000-fold longer, which would not be feasible even on very large computing systems. Probe positions are scored using a detailed energy expression that includes attractive and repulsive van der Waals terms, electrostatic interaction energy based on Poisson-Boltzmann calculations, a cavity term, and a structure-based pairwise interaction potential. The 2000 lowest energy poses for each probe are retained and energy minimized using the CHARMM potential³⁶ with the Analytic Continuum Electrostatic (ACE) model³⁷ to account for electrostatics and solvation. The minimized probe conformations are clustered with a 4Å radius, starting with the lowest energy structure. The clusters are ranked on the basis of their Boltzmann averaged energies, and the six clusters with the lowest average energies are retained for each probe. The clusters of different probes are then clustered into consensus clusters. In this clustering step, a center of mass is defined for each probe cluster and the cluster with the highest number of cluster centers within 4Å is selected as the highest ranked consensus cluster. The position of this consensus cluster is defined as the top consensus site, denoted as CS000. The clusters in CS000 are removed from the pool, and the clustering is repeated until all probe clusters have been assigned to a CS. The CSs are ranked based on their number of probe clusters. An example is shown in Fig. 1c, which indicates only a single molecule from each probe cluster. We note that seeking consensus clusters is in good agreement with the observation that only clusters of different probes identify binding hot spots, whereas individual probes may bind in a number of crevices. Focusing on clusters also gives substantial robustness to the method, as the positions of consensus clusters are much less dependent on small structural changes than the docked positions of individual probes. Due to this property, the FTMap results are essentially free of the many

false positive probe binding positions predicted by classical methods such as GRID³⁸ or MCSS³⁹ (see also the paper by Mattos and Ringe⁴⁰ for a discussion of this problem).

The FTMap server is available at <http://ftmap.bu.edu> and has over 1300 registered users. The only mandatory step required by the user is the selection of a structure to map either by providing a PDB (Protein Data Bank) ID⁴¹, in which case FTMap will fetch the structure for the user, or by uploading a structure in PDB format. The user may also identify the chains to map; however, if this field is left blank, all chains will be used. There are three optional inputs. The first is a protein mask. A masking file is a listing of the protein atoms to ignore during the mapping calculation. This file should be in PDB format. This option can be useful if the user wishes to ignore known binding sites for cofactors, proteins or other biomolecules. Masking can also be used in cases in which the user knows the binding site already and wishes to simply focus on the identification of hot spots within this region, ignoring the rest of the protein. Second, the user may select protein-protein interaction (PPI) mode to indicate that the goal of the mapping is to identify hot spots on protein-protein interfaces. Third, selecting the nucleic acid mode indicates that the PDB file may include nucleic acids, possibly in addition to amino acids, and the nucleic acids will be considered as part of the macromolecule to map. We note that FTMap has been initially developed to map only proteins¹⁹, but the server has been recently extended to DNA and RNA molecules⁴².

The FTSite algorithm and server

The goal of FTSite is the identification of ligand binding sites based on the structure of ligand-free proteins. The method is based on the observation that the binding site of a macromolecule generally includes a strong “main” hot spot and some other hot spots that are close enough to be reached by a ligand binding at the main hot spot. The program uses the consensus sites (CSs) detected via the FTMap server to identify and rank binding sites (<http://ftsitesite.bu.edu>)²². One difference from the FTMap algorithm is that FTSite ranks the consensus clusters by the number of nonbonded contacts between the protein and all probes in the consensus cluster, rather than by the number of probe clusters, since the former approach provides slightly better binding site predictions²². A residue of the protein and a probe are considered to be in contact if any atom of the residue is less than 4Å from any atom of the probe. To identify the ligand binding sites of a protein, FTSite selects CS000 and expands it by adding any neighboring CS if the center of any of its probe is closer than 3.5Å to the center of any probe in CS000. The protein residues that are within 4Å of the expanded CS constitute the top prediction of the binding site, defined as Site 1, while other CSs identify lower ranked predictions. Since the goal of FTSite is to rank all binding sites, no part of the protein can be masked. The server facilitates application to multiple protein structures by allowing the upload of a zip file that contains up to 15 PDB files. The user may provide an e-mail address, in which case upon completion of the job an email will be sent with link to the results. There are no additional input parameters.

The FTFlex algorithm and server

FTFlex performs repeated mapping calculations while exploring low energy conformers of side chains in the vicinity of user-selected hot spots³⁰. The primary goals of this approach are opening pockets in protein-protein interfaces that are potentially capable of binding

small molecular inhibitors, and to determine the druggability of such sites. FTFlex is a two-stage algorithm (Fig. 3). Stage 1 is the identification of the hot spots using the FTMap algorithm. After Stage 1 is completed, the program stops, and the user can select the ligand binding region, or other region of interest, by specifying the consensus sites that define this region. After this selection the server proceeds to Stage 2, which itself includes multiple computational steps. First, FTFlex examines the residues within 5 Å of the selected hot spots, and selects the ones that satisfy hydrophobicity and cavity size restrictions. A hydrophobicity potential is used to calculate the hydrophobicity value for each residue^{19,43}, and only the residues above a hydrophobicity threshold are retained. A cavity measure, developed for use in FTMap¹⁹, is also calculated for each residue, and only residues with a cavity measure of at least 60% of the maximum value are retained. Second, possible rotamers are determined for each of the selected residues, but the user may specify side chains that will be kept fixed. To determine the potential rotamers, FTFlex performs energy minimization starting from each rotamer in a rotamer library⁴⁴, to reach local minima. 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 field³⁶ with the Analytical Continuum Electrostatics (ACE) solvation model³⁷. The local minima are then clustered to define an “environment-dependent” rotamer library⁴⁴. For each cluster the average Boltzmann energy is calculated. Side chains that are predicted to have multiple conformers with low energy and/or high population are considered to be movable. In the next step of FTFlex, each selected rotamer is inserted back into the initial PDB structure one at a time, and the resulting structure is mapped again using FTMap. Rotamers with apo-like conformations are excluded in this step, since they were already considered in the initial mapping. A rotamer is retained if its use increases the number of probe clusters within the binding site, indicating a pocket with improved binding properties, otherwise the rotamer is discarded from consideration. After all rotamers have been tested, the rotamers that yield the most improvement are substituted back into the initial PDB file to yield a modified structure, which is mapped and the results are returned to the user. Since mapping may be performed for a large number of side chain rotamers, running FTFlex may take substantially longer than mapping the single initial structure by FTMap, and the time required for the calculation is highly dependent upon the input structure and the nature of the binding site.

The FTMap/param algorithm and server

FTMap/param, available at (<http://ftmap.bu.edu/param/>)³¹, enables the user to define small molecules that are not included in the standard set of 16 probes, to use the mapping program to generate low energy clusters of these additional molecules, and to check the locations of these clusters relative to the hot spots based on the original probes³¹. FTMap/param includes the identification of hot spots using FTMap. In addition to the standard FTMap input, however, the user may provide SMILES string specifications and formal charges for up to 10 compounds. The program determines the parameters that are required to run FTMap on these additional molecules, and generates a number of low energy conformations for the compounds that have rotatable bonds. Each conformer is considered as a new probe, and used in a separate mapping run. Based on our experience, the standard set of 16 probes is sufficient for finding the hot spots and hence only these are used when forming the

consensus clusters and thus determining the consensus sites²². However, the structures representing the low energy clusters of the user-selected additional molecules show whether or not the compound is likely to bind at the hot spot. If a compound binds, these clusters may provide information on the preferred position of specific functional groups, thereby helping the design of larger ligands. Compounds that do not cluster in the hot spots most likely do not bind to the protein at all³¹.

We have two main reasons for developing a separate FTMap/param server rather than adding an extra option to FTMap. First, the parameters of the 16 molecules used in the standard probe set have been manually improved to be more compatible with the parameterization of the CHARMM potential used for the refinement of probe positions³⁶. This was possible because most of these probes are amino acid side chain analogs, and CHARMM has been primarily developed for protein modeling, with limited resources for parameterizing arbitrary small molecules. Second, the parameterization and conformational search steps of the algorithm are still being improved, partially based on future user experience with various probe molecules. Currently the server generates conformers for the additional probe molecules using the program Confab⁴⁵. The parameterization involves well established computational chemistry programs including ANTECHAMBER⁴⁶ based on the general AMBER force field (GAFF)⁴⁷, and General Atomic and Molecular and Electronic Structure Systems (GAMESS)⁴⁸. The charge model Austin Model 1 bond charge correction (AM1-BCC)⁴⁹ is used to calculate atomic charges. However, since no approach can generate optimal parameters for every class of compounds, most likely the parameterization server will require further development, although these changes will not affect the user interface and hence this Protocol.

The FTDyn algorithm and server

FTMap has also been extended to work with ensembles of conformationally diverse structures, based on Nuclear Magnetic Resonance (NMR) experiments³² or on molecular dynamics (MD) simulations^{21,33}. We added FTDyn to the FTMap family recently, but the program has already been used for mapping NMR derived ensembles of protein structures³². The advantage of FTDyn is that it uses a faster version of the FTMap algorithm without local minimization, and applies this simplified mapping to each structure in the ensemble. This may reduce the accuracy of ranking the hot spots, but makes it computationally feasible to map even very large ensembles such as snapshots from MD simulations. In addition to the usual mapping results for the individual structures, FTDyn determines the average number of nonbonded and hydrogen bond interactions between the probes and each residue for the entire ensemble. Selecting the residues interacting with many probes helps to determine the most likely binding site residues³². Once the binding site residues are selected, for each structure we can calculate the binding site hit rate (HR), defined as the sum of probe-protein interactions for all binding site residues divided by the total number of probe-protein interactions. It was shown that, at least for protein-protein interaction targets that also bind peptides or small ligands in the interface region, the hit rate calculated for ligand-free structures is a good predictor of their similarity to a peptide or ligand bound conformation³². Since FTDyn does not refine the probe positions by minimizing the CHARMM energy, the results may be somewhat less accurate than using FTMap. However, results appear to have

adequate quality for selecting the most bound-like conformations of the ensemble, and individual structures can be further investigated using FTMap.

To describe an alternative approach to the analysis of FTMap results for ensembles, we note the recent release of the FTProd plugin (<https://amarolab.ucsd.edu/ftprod/>) for VMD, a molecular visualization program⁵⁰. FTProd combines FTMap results obtained for multiple experimental structures of the same target, clusters the results, and displays them for analysis within VMD. In contrast to our servers, FTProd is a plugin, which should be downloaded and installed. While FTProd is a useful tool extending the capabilities of FTMap, it is not part of the FTMap family, and hence we refer potential users to the instructions for downloading, installing, and using the program to the FTProd online tutorial at <https://amarolab.ucsd.edu/ftprod/tutorial/01>.

Comparison to existing methods

As mentioned, hot spot residues were originally defined as the ones in protein-protein interfaces that give rise to a substantial drop in binding affinity when mutated to alanine. A variety of computational methods have been developed for the prediction of such residues^{7,51–54}. However, in this work we focus on hot spots that bind small molecules, and hence do not review methods of hot spot residue prediction. Since FTFlex simply accounts for receptor flexibility and FTMap/param defines new probes for FTMap, we consider the basic functions of the programs, i.e., the identification of binding sites and the determination and ranking of binding hot spots. A number of methods exist for binding site detection. These methods largely fall into one of three categories: (1) geometry-based methods, (2) knowledge-based methods, and (3) energy-based methods⁵⁵. In geometry-based methods, cavity size is measured and the ligand binding site is identified as the largest pocket. This category includes methods such as: POCKET⁵⁶, LIGSITE^{57,58}, PASS⁵⁹, and CASTp⁶⁰. However, these methods sometimes fail to detect shallow cavities or pockets that may be partially closed, and the results generally do not correlate with ligand binding energetics. That is, it is impossible to assess which sites contribute most to the ligand binding free energy. Knowledge based methods can be used to identify binding sites in proteins by comparison to structures with high sequence similarity. These methods include 3DLigandSite⁶¹ and FINDSITE⁶². For targets with highly conserved binding sites, this type of method can be quite successful⁶³; however, for other targets, this method has success rates closer to 70%. In the energy-based method, the propensity of the site to interact favorably with probe molecules is measured, thus making this method favored in drug design applications. FTMap falls under this category, as does Q-SiteFinder⁶⁴, SiteMap⁶⁵, and SITEHOUND⁶⁶. Q-SiteFinder and SITEHOUND both use a methyl probe, whereas SiteMap uses a water molecule as the probe. Both GRID³⁸ and Multiple Copy Simultaneous Search (MCSS)³⁹ use multiple functional group types to probe the surface and identify regions that are capable of binding multiple probe types, in a similar manner as the experimental MSCS method¹¹.

Both GRID³⁸ and Multiple Copy Simultaneous Search (MCSS)³⁹ also rank the regions where probes bind, provide some information similar to the results FTMap, and have been used for structure based drug discovery. Both programs have strengths and shortcomings.

GRID performs efficient global sampling, finds the potential binding pockets, and discriminates between polar and non-nonpolar regions³⁸. The main shortcoming of the method is that the very small probes bind in many different pockets, resulting in a large number of false positive local minima^{11,67}. The original implementation had no solvation term in the scoring function, which may have contributed to the false positives. MCSS is using CHARMM³⁶, which is a detailed and well established potential. MCSS performs simultaneous Monte Carlo minimization of multiple probes distributed on the protein surface. The probes do not interact with each other, but interact with the protein. Since all probes and the protein are fully flexible in the simulation, the conformation of the protein can be influenced by the entire ensemble of probes, which is a potential advantage³⁹. However, it was reported that MCSS also yield a large number of false positive energy minima^{11,68}, possibly due to the lack of solvation, which was added later⁶⁹. Another problem is that both GRID and MCSS generally predict different locations for different types of probes, contradicting to the results of experiments such as the X-ray crystallography based Multiple Solvent Crystal Structures (MSCS), showing that the different small organic molecules overlap at a few locations. More recently, mixed molecular dynamics (MD) has emerged as an alternative approach to protein mapping. The method performs MD simulations of the target in an aqueous solution of probe molecules^{70–73}. While the use of explicit molecules potentially improves the accuracy of simulations, they also reduce the diffusivity of the probes, and hence it is frequently questionable whether equilibrium distribution can be achieved on reasonable time scales³⁴. In addition, due to the need for long simulations, current mixed MD methods rely only on a few probe types, which is likely to limit the reliability of hot spot prediction.

Relative to the above methods, FTMap has several advantages. First, it uses 16 probe molecules with a variety of sizes, shapes, and functional groups; this variety of probes is known to provide the robustness required to accurately identify binding sites and eliminate false positives, such as sites in narrow cavities^{19,34}. Second, FTMap is able to provide adequate accuracy with maximum efficiency¹⁹. A key to FTMap's high accuracy is the use of a detailed energy expression to sample probe positions on the protein surface. To achieve maximum efficiency using this energy expression, the fast Fourier transform (FFT) correlation approach is used¹⁹. FTMap also accounts for solvation using a continuum electrostatic model within the CHARMM implementation³⁷. Lastly, FTMap is available as a free, easy-to-use server, in which the only required input is a protein, DNA, or RNA structure in PDB format. This enables the exploration of binding properties of macromolecules and answering biological questions with limited efforts.

Limitations

Currently FTMap maps structures that include only naturally occurring amino acid residues and nucleotides. Most HETATM records are removed prior to mapping. While a few cofactors can be explicitly added by the user using the chains field, this list is limited (See Table S1 in Supplementary Information). FTMap itself does not include protein flexibility; however, the FTFlex server (at <http://ftflex.bu.edu>) accounts for side chain flexibility³⁰. Another limitation is that, due to memory limits on our computational resources, the analysis of proteins over 1100 residues frequently fails.

Availability

All servers are available to any user, but their jobs will be publicly accessible. User with education or government email addresses can also set up FTMap accounts. The advantage of such accounts is that the results are available only to the user, and the job does not show up on the website. While results can be viewed online, most analyses require the use of protein visualization software; PyMOL is the recommended software and is used within this Protocol to demonstrate data analysis.

MATERIALS

EQUIPMENT

- A computer with internet access and a web browser.
- Atomic resolution structures of biomolecules under investigation, in PDB format. The PDB ID can be used to directly fetch in structure or structure may be uploaded from computer.
- Access to PyMOL or similar structure viewing software is recommended but not required. PyMOL can be downloaded at <http://www.PyMOLwiki.org/index.php/Category:Installation>.

Running FTMap

- 1| Locate the server at <http://ftmap.bu.edu>. FTMap can be used with or without a user account. To create an account, register on the FTMap server website using an educational or governmental email address. A password will be sent to the email address and can be later changed. If you already have a username and password, fill in the boxes and click Login. Proceed to Map, which is the server home screen. From this page you will be able to submit a new job. If you prefer not to use an account, click the option below Login to use the server without an ID.

CRITICAL STEP Users who run FTMap without an account will have their results publicly accessible.

- 2| (Optional) Provide a job name for this submission. If you choose to leave this blank, a unique ID will be created for this field.
- 3| Input the coordinates of the target structure using PDB format. Only atoms of amino acid residues and nucleotides will be retained. All HETATM records, including waters, ligands, and cofactors, will be automatically removed. Some HETATMs can be selectively added back as detailed in Step 5. There are two options for inputting a structure: use Option A to directly import coordinates from the PDB, or option B to upload a structure directly.

A. Import coordinates directly from PDB

- i. Import coordinates directly from the PDB by typing the four digit PDB ID into the field PDB ID.

B. Upload structure directly

- i. Upload a structure directly from your computer by clicking on the Upload PDB option under the PDB ID field. Select Browse to upload a file containing a structure in PDB format.

CRITICAL STEP At this point only structures containing standard amino acid residues and nucleotides can be mapped. Input of structures with ATOM records of nonstandard amino or nucleic acids will result in an error.

CRITICAL STEP If you are using an NMR structure, you must upload the PDB of the model number you are interested in, otherwise only the first model will be submitted for mapping. To create this PDB file, download the PDB file of the NMR structure from the PDB. In either PyMOL or a text editor, select only the ATOMs lines of the model of interest. Save these lines as a new PDB file. Upload this file to the FTMap server.

CRITICAL STEP The user should indicate if the file includes any nucleic acid (see Step 9).

? TROUBLESHOOTING

- 4| In the Chains field, enter the protein chains that you wish to include in the mapping calculation. List chains using their chain id and separate multiple chains with a whitespace. If no chains are specified, then all chains will be mapped.

? TROUBLESHOOTING

- 5| (Optional) Include certain heteroatoms in the PDB file by prefacing the HETATM code, found in the PDB file, with an H and adding the chain id to the end of the HETATM code. For instance, if you want to include a Zn atom in chain A, you would input HZNA into the Chains field, using whitespace to separate multiple entries. There are parameters for common metals, such as Zn, Mg, and Fe, as well as heme. A complete list of available HETATMS can be found in Supplementary Table S1.

? TROUBLESHOOTING

- 6| (Optional) Click the Advanced Options label on the Map page to see more options.
- 7| (Optional) Upload a protein mask file to tell FTMap to ignore a certain region of the protein surface. The mask file contains coordinates for atoms in the defined region and FTMap will prevent probes from going into that particular region of the protein. Masking can be useful in cases where there is a ligand binding site separate from a coenzyme binding site and one wants FTMap to ignore the coenzyme binding site. The mask file can be created using PyMOL. First, load the protein into PyMOL. Select the molecule in the binding pocket that you

would like to ignore by right clicking on the ligand or cofactor. Expand the selection to nearby atoms by right clicking on the selection and choosing actions > around > atoms within 8Å. This selection can then be saved using File > Save Molecule.

- 8| (Optional) If you are looking for binding hot spots in a protein-protein interface, select the PPI mode option.
- 9| (Optional) Select the Nucleic Acids option if the PDB file includes nucleic acids. This option uses a specific set of parameters.
- 10| Click on the Map button to begin the calculation. You can immediately check the status of the job on the Queue page. Jobs are run in order of their submission. Your calculation will be listed with Id number, job name, user name, and a status update. Click on the ID of your job to see a detailed Status Page. The Status Page shows the job ID number, job name, job status, submission time stamp, and PDB ID. The page also shows pictorial representations of the uploaded and processed inputs, and Probe Status, i.e., status for each small molecule probe. See Box 1 for information pertaining to status abbreviations.
- 11| An email will be sent when the job has completed or if an error occurred (see Box 2 for a listing of possible errors and their meanings). The email will contain a link to the mapping results. Click the link, or alternatively, locate the results under the Results tab on the server. All user results will be listed in order of Id number. If there was an error, an error message will be provided in an email to the user as well as in the Results tab.

CRITICAL STEP The mapping result files will only be stored on the server for 2 months. After this time, the results will be deleted.

? TROUBLESHOOTING

- 12| View results by clicking on the Id number of the job under the Results tab. The Result page (see Fig. 4) starts with the job name. The page also offers direct visualization that shows the protein and all consensus sites. Click on the image to interact with the resulting structure using JSmol.

CRITICAL STEP Allow several seconds for JSmol to load for interactive visualization.

- 13| To manipulate the structure in JSmol, use the left mouse button and drag to rotate your structure. Use your mouse wheel to zoom in and out. Use the checkboxes along the bottom to select/deselect any specific consensus cluster. Use the color options to change the structure coloring. Execute JSmol scripts from the text field. Information on using JSmol provided on several websites, e.g., http://wiki.jmol.org/index.php/Jmol_JavaScript_Object#JSmol.
- 14| The result page also includes two bar graphs that show the percent of nonbonded and hydrogen bond interactions, respectively, between the probes and the protein for each residue along the protein sequence (Fig. 4). Download these

results as tab separated files with exact residue contact counts by clicking the link located directly beneath each respective graph.

- 15] On the result page, click on Probe Summary to download a text file summary of clusters and their probe compositions. Finally, you can also download the PDB file containing the protein and representatives of probe clusters in the consensus sites, and the PyMOL session (.pse file; see Fig. 5a) containing all results to be opened in PyMOL. (See Box 3 for a description of the contents within each file.) Examples of types of analyses that can be performed using these results (e.g., focusing on hot spots around the ligand as shown in Fig. 5b) are detailed in the Anticipated Results section.

? TROUBLESHOOTING

BOX 1

STATUS UPDATES FOR FTMAP RUNS

The progress of the FTMap run can be monitored in the “Queue” tab. Note that since the timings for each step are highly dependent on the input structure, they are not provided here.

Processing pdb files - downloading PDB file from the www.pdb.org web site, processing chain information, extracting the chains the user specified

Pre-docking minimization - running CHARMM to add missing atoms and polar hydrogens, minimizing the added atoms in the presence of the protein.

Calculating PB potential - using CHARMM to calculate the Poisson-Boltzmann potential around the protein

Copying to supercomputer - copying the PDB file and Poisson-Boltzmann potential to the cluster where FTMap will run

Held on supercomputer - files are on cluster, but job is not yet submitted

In queue on supercomputer - jobs for all probes have been submitted on the cluster, but they have not started running

Running on supercomputer - jobs have begun running on the cluster. By clicking on the job id, one can access additional details of the run as related to probe docking and clustering. These statuses are as follows:

Queued - waiting in the cluster queue to run FFT

Running FFT - probe is running FFT on the cluster

Clustering - post-FFT clustering of probes

Queued for minimization - waiting in the cluster queue to minimize probes

Running Minimization - probe is being minimized on the cluster

Finished - probe has finished minimization on the cluster

Finished on supercomputer - all probes have run FFT and minimization on the cluster

Copying to local computer - results are being copied from the cluster back to the FTMap server

Clustering and minimization - individual probes are being clustered, then consensus sites are generated by clustering across probes

Calculating interactions - calculate nonbonded and hydrogen bonded interactions between probes and the protein using HBPlus

Finished - everything is complete

BOX 2

ERROR MESSAGES AND THEIR MEANINGS

When the calculation encounters an error, the job will be terminated and the user will receive an email with an error reason. The error will also be listed next to the corresponding job id in the “Results” tab. The error codes and explanations are listed below.

xxxx not found in PDB - xxxx is the four letter PDB ID. This error occurs when the computer is unable to download the entered PDB ID from the website, www.pdb.org. Usually this error occurs when the PDB ID does not exist, but it can also occur when the PDB website is down.

Unknown residue xxx in receptor. Please remove. - xxx is the three letter amino acid code. This error occurs when a residue in an atom record is not recognizable by FTMap and thus FTMap does not have parameters for it. Check to make sure that the proper three letter code is used and the amino acid is one of the 20 naturally occurring amino acids.

Processing failed on receptor – This error occurs during the initial steps when CHARMM is being used to add missing hydrogens and minimize the structure. Usually, this occurs when the protein structure has sterically clashing atoms or the structure generally does not make physical sense in terms of bonds.

PB Potential not created – CHARMM could not create a Poisson-Boltzmann potential. If this occurs, it is usually because the protein is too big to create a potential.

Not enough lines in output for xxx - FFT failed to run for a probe.

Minimization error for xxx - Minimization failed to run for a probe.

Failed during clustering – This error usually means that probes could not be clustered because FTMap could not find low energy positions for one or more probes.

Protein too large for FTMap – This error may occur for proteins longer than 1100 residues.

BOX 3

DESCRIPTIONS OF RESULT FILES

Five different result files are available to users at the conclusion of their FTMap calculation. The results can be used in a number of applications as described in the **ANTICIPATED RESULTS** section. The output files are:

1. A **PDB** file that contains the input structure ATOM lines followed by the coordinates for the individual crossclusters, which are defined as the consensus clusters generated using the individual probe clusters. This file can be opened as a text file or can be used in any structure-viewing program, such as PyMol. The crossclusters are ranked in order of number of probe clusters. Their name format is: “crosscluster.xxx.yyy.pdb” where the number “xxx” provides the ranking of the cluster, starting at 000 for the highest ranked cluster. The number “yyy” provides the total number of probe clusters. Note that only representative probes poses are provided in the output PDB file.
2. A **PyMol Session** file that is pre-formatted to open in PyMol. The structure is appropriately scaled and shown in cartoon mode. Crossclusters are shown as sticks and individually colored.
3. A **Nonbonded Interactions** file that lists all noncovalent contacts between probes and amino acids, calculated using HBPlus. This file contains 4 columns. The first column lists the amino acid number, the second column lists the chain id, the third column lists the amino acid 3-letter code, and the fourth column lists the number of contacts between all docked probe molecules and the amino acid.

The residues with the highest number of nonbonded interactions define the main and secondary hot spots. The nonbonded interaction file can also be used to generate a mapping fingerprint by calculating a percentage contact frequency for each amino acid. This mapping fingerprint is then the profile consisting of the amino acids with the highest percentage contact frequencies. The percentage contact frequency for an individual amino acid (aa_i) for amino acids $i = 1 \dots n$ is defined as:

$$\% \text{ contact frequency for } aa_i = \frac{\text{Number of nonbonded contacts for } aa_i}{\text{Sum of contacts for all } aa} \times 100\%$$

4. An **Hbond Interactions** file that lists all hydrogen bonding contacts between probes and amino acids, calculated using HBPlus. The column format is the same as for the Nonbonded Interactions file.

5. A **Probe summary** file that lists the individual probe clusters contained within each crosscluster.

Running FTSite

- 16| Locate the server at <http://ftsites.bu.edu>, and (optionally) repeat Step 2. The use of FTSite does not require registration.

CRITICAL STEP All FTSite results are publicly accessible.

- 17| Input a target structure using either options A or B from Step 3, or you can upload a Zip file containing up to 15 PDB files. Such a file should contain only the chains of interest, and the individual PDB files should not be in folders inside the zip file. Use options A, B or C respectively to prepare the zip file for the three major operating systems (Windows, Mac OS X and Linux) as follows:

A. Windows

- i. Select the files, then right click and select “Send to” > “Compressed (zipped) folder”. You will then be able to choose a name for your file.

B. Mac OS X

- i. Select the files, then right click and select “Compress N Items”. This will create a zip file called Archive.zip with your files.

C. Linux

- i. On the command line, navigate to the folder with the pdb files. Then run `zip filename.zip *.pdb` to create a file called filename.zip with your pdb files.

CRITICAL STEP At this point only structures containing the 20 standard amino acids can be used in FTSite. Input of structures with nonstandard amino acids in ATOM records will result in an error.

- 18| Specify the chain to be mapped as in Step 4.
- 19| (Optional) You may provide an e-mail address. If this option is used, e-mail will be sent to the address with a link to the results when the FTSite run is completed, or if an error occurred (see Step 11 for details).
- 20| Click on the “Find My Binding Site” button to begin the calculation. Unless there is an error, you will immediately see the message “Success, job submitted” and can check the status of the job on the Queue page. Your calculation will be listed with job name and a status update. See Box 1 for a listing of status updates.

? TROUBLESHOOTING

- 21 After the FTSite run is completed, view the results (Fig. 6) using option A via a web interface, or select option B to use PyMOL:

A. Web interface

- i) View the results via a web interface. The interface contains a Java Applet. Mesh representations of the top three sites and sticks of the residues around the sites can be turned on and off along with various representations of the protein. Residues near each site are also listed. An example of results is shown in the Anticipated Results section.
- i) Click on the “Download PyMOL Session” label to download a file containing the PyMOL session which shows the protein, the detected sites, and the residues that surround the site.

Running FTFlex

- 22| Locate the server at <http://ftflex.bu.edu>. FTFlex can be used without a user account. (Optionally) repeat Step 2, and input target structures/specify chains to be mapped as described in Steps 3 and 4.
- 23| (Optional) As for FTSite (Step 19), you may provide an e-mail address.
- CRITICAL STEP** If no e-mail is provided, the user must keep track of the link provided post-submission (see Step 24) to follow the job. Due to the two-stage nature of FTFlex, providing an e-mail address is highly recommended.
- 24| Click on the “(Flexibly) Map My Protein” button to begin the initial mapping. Successful submission is indicated by the prompt “Success: Job submitted”, and a link is provided for keeping track of the job. The job tracking page will show the status of the initial mapping along with how many jobs are in the FTMap queue.
- 25| From the results of the initial mapping (Fig. 7), choose the consensus sites (CSs) around which low energy side chain conformers will be explored in repeated mapping calculations. Examine the CSs on the output page from the initial mapping by FTFlex in JSmol, and select by clicking on the appropriate CS labels. Alternatively, the mapping results can be downloaded as a PDB file and viewed in a molecular viewer such as PyMOL, but the CSs for side chain analysis still must be selected in JSmol on the output page from the initial mapping.

CRITICAL STEP Results are downloaded as a PDB file rather than a PyMOL session provided by the other servers. In the file the consensus sites are defined as separate chains, separated by HEADER records, which allows some viewers (e.g. PyMOL) to show them as separate objects. Opening the file shows the protein in line representation, including the side chains. However, the structure is clipped to focus on a slab around the consensus sites. The slab width can be increased in PyMOL using the Zoom or Clip menus. However, after the analysis

of the results using PyMOL you still will have to return the result page from the initial mapping by FTFlex for selecting the hot spots for side chain analysis (see Step 27).

- 26| (Optional) In addition to selecting CSs for side chain analysis, you can select an arbitrary number of side chains whose rotamers will be kept fixed in the further steps of the algorithm from the output page of the initial mapping. The format of the selection is “chain_id - residue_number chain_id - residue_number...”. For example, if you want FTFlex to ignore Trp127 on chain B, you should enter B-127.
- 27| After choosing the CSs, initiate the analysis of side chain rotamers by clicking on the “Submit” button. FTFlex will individually generate and test rotamers for all side chains within the 5Å radius of the selected CSs, and will map all resulting structures. The status page will show all side chains and their rotamers being tested along with their status and position in the mapping queue. After mapping with all the rotamers individually, FTFlex determines whether a change in a side chain rotamer would increase the number of probe clusters. If the answer is positive, a new protein structure is generated with all selected side chains changed to their appropriate rotamers simultaneously, and the resulting structure is considered in a final mapping. In this structure the number of probe clusters is maximized within the binding site defined by the selected CSs, indicating a maximally opened binding pocket. The status page will show that FTFlex is performing the final mapping, the actual steps performed by the algorithm, and the position of the job in the mapping queue.
- 28| The status page indicates when the final mapping is finished, and an e-mail is sent to the user if an e-mail address has been provided in Step 23. This e-mail will contain the mapping results (see Box 3) and information regarding which residues, if any, were moved to obtain the final mapping results. Alternatively, download this information from the results page along with the mapping results in form of a PyMOL session and counts of hydrogen bonded and nonbonded contacts between the probes and the protein in the final conformation with the selected side chain conformers. These latter results have the same format as the results from the FTMap server.

Running FTMap/param

- 29| Locate the server at <http://ftmap.bu.edu/param/>. Repeat Steps 1–9 to set up FTMap/param to perform a standard mapping calculation.
- 30| (Optional) Following the FTMap input, click on the “Advanced Options” label “Upload Small Molecules”. This will open a window where you can specify up to 10 small molecules that will be used as additional probes. Each molecule is described in a separate line by its formal charge and its SMILES string.

CRITICAL STEP Isomeric SMILES must be used to specify the stereochemistry.

CRITICAL STEP In order to limit the computer time required for the mapping, at this time the server accepts SMILES strings only for small molecules containing three or fewer rotatable bonds; typically this translates up to 100 conformers per molecule.

- 31| Repeat Steps 10 and 11. Your uploaded small molecules will be run and listed alongside the sixteen original probes in the default probe set. They will have been assigned generated names where the first two digits represent the order they were uploaded and the second two digits represent the particular conformation.
- 32| View results as described in Step 12, and manipulate the JSmol file as detailed in Steps 13–15. The results from FTMap/param are very similar to the results from FTMap. The difference from FTMap is that, in addition to clustering the sixteen standard probes, FTMap/param maps the protein with the user-defined molecules, finds low energy clusters for these additional compounds, and returns the representative of any cluster that overlaps with one of the consensus clusters. Accordingly, the PDB file and the PyMOL session (Fig. 8) includes such low energy conformations labeled as Molecule_1, Molecule_2, etc. Clicking on any of these labels shows the different conformers of the molecule, labeled as mol1-conf1, mol1-conf2, etc. An example is shown in the Anticipated Results section. Activating JSmol initially shows the protein, all consensus clusters, and the conformations of the user-defined molecules at the consensus sites. As in FTMap, at the bottom of the JSmol panel checkboxes are provided to select/deselect any specific consensus cluster. The difference from FTMap is that after deselecting all consensus clusters, the panel still shows low energy conformers of the user-defined molecules. As a matter of fact, these conformers cannot be removed. The graphs as in Step 14 are based on the 16 standard probes only, and clicking on Probe Summary as detailed in Step 15 will download a text file summary of clusters and their probe compositions. The user-selected molecules are not included.

CRITICAL STEP Allow several seconds for JSmol to load for interactive visualization.

CRITICAL STEP The JSmol panel simultaneously shows all conformers for each user-selected molecule if it is in a consensus cluster. If the number of rotamers is large, the JSmol figure becomes too crowded, emphasizing the advantages of looking at the .pse file in PyMOL.

Running FTDyn

- 33| Locate the server at <http://ftdyn.bu.edu>. The use of FTDyn does not require registration.
- 34| Input the coordinates for an ensemble of structures, using Option A to import coordinates directly from PDB, or option B to upload coordinates from your computer. The coordinates of the target structure ensemble must be specified in the PDB model record format as specified at <http://deposit.rcsb.org/adit/docs/>

[pdb_atom_format.html#MODEL](#). A single coordinate entry in this format contains multiple structures with models numbered sequentially beginning with 1. Each MODEL file must have a corresponding ENDMDL record. Only atoms of amino acids will be retained. All HETATM records, including waters, ligands, and cofactors, will be automatically removed.

A. Import coordinates directly from PDB

- i. Import coordinates directly from the PDB by typing the 4 letter PDB ID of a PDB file stored in model record format, which is the case for NMR structures in the PDB.

B. Upload coordinates directly from your computer

- i. Upload coordinates arranged in PDB model record format directly from your computer by clicking on the Upload PDB File option under the PDB ID field.

CRITICAL STEP At this point only structures containing the 20 standard amino acids can be mapped. Input of structures with ATOM records of nonstandard amino acids will result in an error. All HETATM records are stripped out, removing any ligands in the file.

35| Specify the chain to be mapped as in Step 4.

36| (Optional) The user may provide an e-mail address. If this option is used, e-mail will be sent to the address with a link to the results when the FTDyn run is completed, or if an error occurs.

CRITICAL STEP If no e-mail address is provided, your FTDyn results will be publicly accessible.

37| Click on the “Map Ensemble” button to begin the calculation. You will see a success message once the structures are submitted to the computing queue, and can check the status of the job on the Queue page. This will take up to a couple minutes depending on the number and size of uploaded structures. Your calculation will be listed with job name and a status update. See Box 1 for a listing of status updates.

38| The structures in the ensemble are mapped one-by-one. Once the mapping of a structure is completed, the interactions can be plotted on the job results page by clicking the View link under the Fingerprint tag. The page offers direct visualization that shows the protein color-coded according to the number of nonbonded interactions with the probes. Click on the Load button under the Map tag to activate JSmol for the visualization of the selected structure. The structure can be manipulated as described in Step 13. By default, the residues are colored on a spectrum from blue to red, indicating the frequency of probe interactions from zero to maximum. You can also color the protein by direction or secondary structures using the radio buttons. A command line box is provided so you can execute arbitrary JSmol command to control the view.

CRITICAL STEP Allow several seconds for JSmol to load for interactive visualization.

- 39| The result page includes two bar graphs for the selected individual structure that show the percent of nonbonded and hydrogen bond interactions, respectively, between the probes and the protein for each residue along the protein sequence. Download these results as tab separated files by clicking the links located in the Downloads column of the result table.
- 40| When mapping of all structures is completed, the result page shows the first structure of the ensemble, but the color coding is based on the number of probe-protein contacts averaged over the entire ensemble (Fig. 9a). Residues are colored from blue to red based on contact frequency. View the structure in JSmol as for the individual structures. Similarly, the two bar graphs on the page show the percent of nonbonded and hydrogen bond interactions, respectively, averaged over all structures (Fig. 9b).
- 41| Download one or all of the following results files from the results page: (1) .pdb: a PDB file including all submitted models divided by MODEL/ENDMDL records, with the frequency of interactions written in columns 61–66. (2) .pse: PyMOL session containing all structures with residues colored from blue to red based on contact frequency (Fig. 10a). (3) .tar.xz: an archived directory containing all the individual contact files, pdb files, and combined files. For each structure of the ensemble, the file includes the nonbonded and hydrogen bond interactions with each residue, and the PDB file with the number of interactions recorded in columns 61–66. Further files are a .pdb file with all models, a .pse file with all models to view in PyMOL, files containing the numbers of nonbonded and hydrogen bond interactions, averaged over all structures, and a .pdb file of the first structure, with the ensemble-average number of nonbonded interactions recorded in columns 61–66. The file can be extracted in bash with `$ tar xvfJ 424b991f-9546-4fef-a662-2f4f00116925.tar.xz` or in Windows with 7-Zip.

TIMING

Steps 1–9, input the coordinates of the target structure and select options for FTMap: ~ 1 min

Steps 10–11, running FTMap: ~ 1 hour to more than a day; depends on the size of the protein and the number of jobs in the queue. On average, results are returned in less than 4 hours.

Steps 12–15, analysis of FTMap results for determining the binding hot spots: ~ 15–30 min

Steps 16–19, input the coordinates of the target structures and select options for for FTSite,: ~ 1–10 min; this time includes creating a zip file if multiple protein structures are analyzed.

Steps 20, running FTMap: ~ 1 hour to more than a day for each protein structure; running time depends on the size of the protein and the number of jobs in the queue. On average, results are returned in less than 4 hours for each structure

Step 21, analysis of FTSite results for identifying the binding sites, including the residues that surround the sites: ~ 15–30 min

Steps 22, input the coordinates of the target structure and select options for the initial mapping by FTFlex: ~ 1 min

Steps 23–24, running the initial phase of FTFlex: ~ 1 hour to more than a day; depends on the size of the protein and the number of jobs in the queue. On average, results are returned in less than 2 hours

Steps 25–26, choosing the consensus sites (CSs) around which low energy side chain conformers will be explored: ~5 min

Step 27, mapping structures with the low energy conformers of the side chains in the selected hot spots: Since mapping may take one hour for more or each low energy rotamer, the length of this step will depend on the number of the low energy side chain rotamers around the selected hot spots, and hence the time that this step takes is highly variable. The timing also depends on the number of jobs in the mapping queue. On average, using FTFlex requires 12–24 hours.

Step 28, analysis of FTFlex results, including the list of residues that were moved to obtain the final hot spots: ~ 15–30 min

Step 29, input the coordinates of the target structure and select options for FTMap/param: ~ 1 min

Step 30, input formal charges and isomeric SMILES strings to specify up to 10 small molecules to be used as additional probes: ~ 5 min

Step 31, running FTMap/param: Mapping with the standard set of 16 probes takes about one hour or more, depending on the size of the protein and the number of jobs in the queue. However, the total run time may be substantially longer if you submit several additional molecules, some having many conformers.

Step 32, analysis of FTMap/param results for determining the binding hot spots and the most likely binding modes of additional probes: ~ 15–30 min

Steps 33–36, input the coordinates of the target structure and select options for FTDyn: ~ 10 min

Steps 38, running FTDyn: ~ 10 min for each protein structure; the total run time is proportional to the number and size of structures in the ensemble. The running time also depends on the size of the protein and the number of jobs in the queue.

Steps 39–41, analysis of FTDyn results: ~ 1 hour

TROUBLESHOOTING

Troubleshooting advice can be found in Table 2.

ANTICIPATED RESULTS

Identification of hot spots using FTMap

Fig. 4 shows an example of the results page at the successful completion of a FTMap run. In this case, results are shown for the mapping of renin using a structure of the apo protein (PDB ID 2ren)⁷⁴. The target protein is shown as a green cartoon, and the small molecules representing each probe cluster at the consensus sites are shown in various colors as sticks. Clicking on the image will activate JSmol, and the image can be manipulated using the JSmol tools. The CSs are defined by the positions of the consensus clusters, which are shown as stick, with each consensus cluster represented by a different color. The protein can be rotated by pressing the left mouse button and moving the mouse, and its size can be increased or decreased by rotating the mouse wheel. The consensus sites, each with the number of probe clusters they contain, are listed below the JSmol panel, and can be individually turned on and off. The consensus clusters are colored in the following order, with the highest ranking color first: cyan, magenta, yellow, salmon, white, periwinkle, orange, pale green, teal, fuchsia, and so on. Each CS is shown as the collection of probes representing the centers of the probe clusters in the consensus cluster. The two graphs below the JSmol panel show the percent of nonbonded and hydrogen bond interactions, respectively, between the probes and the protein for each residue along the protein sequence.

In addition to viewing the results in JSmol, you can download the PDB file containing the protein and representatives of probe clusters in the consensus sites, as well as a PyMOL session (.pse file) containing all results to be opened in PyMOL (Fig. 5a). In the session the naming of consensus clusters indicates the number of probe clusters at each consensus site. The CSs are labeled in the format crosscluster.xxx.yyy where xxx refers to the ranking of the CS, starting at 000, and yyy denotes the number of probe clusters in the CS. The ranking indicates the importance of the CS, with the highest ranking sites having the highest number of probe clusters. Mapping the apo structure of renin (PDB ID 2ren)⁷⁴ yields the consensus sites CS000 (cyan, 25 probe clusters), CS001 (magenta, 18 probe clusters), CS002 (yellow, 16 probe clusters) and CS003 (salmon, 10 probe clusters), CS004 (white, 8 probe clusters), CS005 (blue, 8 probe clusters), CS006 (orange, 7 probe clusters), and CS007 (green, 3 probe clusters). Downloading and working with the .pse file is very useful, because you can add further structures, e.g., ones with bound ligands, align the proteins, select particular side chains, and generally use the versatility of PyMOL. For example, Fig. 5b shows Aliskiren, a novel non-peptidic renin inhibitor from the inhibitor-bound renin structure 2v0z superimposed on the hot spots of the apo renin⁷⁵. This result demonstrates that the hot spots, based on ligand-free protein structures, can very well trace out the shape of high affinity ligands. Indeed, Aliskiren overlaps with all consensus sites apart from CS001 (not shown in Fig. 5b). We note that CS001 overlaps with the transition state inhibitor, CGP 38'560 (PDB ID 1rne)⁷⁶, not shown here. Of course, such high level of agreement assumes that the conformational changes of the protein upon ligand binding are moderate, which is the case for renin¹⁸. However, we have demonstrated that the hot spots are relatively robust, and in most cases can be determined by mapping unbound structures²⁸.

Agreement between computational and experimental mapping results when using FTMap

The hot spots identified by FTMap generally are in good agreement with those obtained using experimental mapping techniques such as Multiple Solvent Crystal Structures (MSCS) based on X-ray crystallography¹⁹. Fig. 11a shows MCSS results for ribonuclease A (RNase A)⁷⁷. The method identified four consensus sites using a screening library of 8 aqueous organic solvents⁷⁷. These sites are: two in the B1 pocket, one in the P1 pocket, and one in the B2 pocket. The two sites in the B1 pocket are very close together, and hence Fig. 11a shows only three distinct hot spots, one in each of the three pockets. The X-ray structure of RNase A (PDB ID 2e3w)²⁴ was also computationally mapped using FTMap. FTMap identifies the same three sites as MSCS (B1, P1, and B2), but also detects an additional site (Fig. 11b)²⁴. The consensus sites CS000 (cyan) binds 26 probe clusters in the B1 pocket, CS001 (magenta) includes 20 clusters in the P1 pocket, CS002 (yellow) with 15 clusters is at the site that was not found by MSCS, and CS003 (salmon) binds 13 clusters in the B2 pocket. Note that as all our figures, Fig. 11b shows only a single representative pose from each probe cluster rather than all bound probe positions. It is likely that the CS002 site was not seen in the MSCS study because it is part of a crystal contact interface in the crystal form used in the MSCS experiments⁷⁷. However, this site accommodates a hot spot residue of the ribonuclease inhibitor (RNI), identified by alanine scanning mutagenesis in the RNase A/RNI complex, indicating that the site is an important contributor to the binding free energy, and hence is indeed a hot spot²⁴.

Selecting the important regions of the binding site using FTMap

FTMap can also be used to identify specific hot spot residues that make the most interactions with probe molecules in a manner similar to that used in NMR experiments. This information can be found in the nonbonded interaction file provided at the successful completion of an FTMap run. This file lists the number of contacts between probe molecules and atoms in a residue for all residues. By simply ranking the residues by number of contacts, the user can quickly obtain those residues with the highest number of contacts. Fig. 4 shows such nonbonded and hydrogen bond contacts directly from the result page of FTMap. The nonbonded interaction graph shows the residues that have the highest number of contacts with the probes, and very likely are the most important for ligand binding. The percent contact frequency for an individual residue is calculated as the number of nonbonded interactions for an individual residue divided by the number of interactions for all residues.

Determining druggability and providing information for FBDD using FTMap

Perhaps the most important application of FTMap is determining druggability, defined as the ability of a target protein to bind ligands with high affinity. The first requirement for a site to be druggable is the consensus cluster strength (S), defined as the number of probe clusters within the consensus cluster. A consensus cluster of $S > 16$ will be druggable by some sort of ligand²⁸. Consensus clusters of $S < 13$ occupy sites that are not druggable due to very weak hot spots. Between 16 and 13 we have some uncertainty, predicting proteins that are borderline druggable, i.e., are likely to have ligands with low micromolar affinity, but no further improvement in binding would be possible. The second requirement for druggability

is the existence of at least one additional (possibly weaker) hot spot within 8 Å from the strong hot spot, where the distance is measured between the centers of mass for two consensus clusters²⁸. The distance between two selected atoms can be determined in PyMOL by choosing Wizard > Measurement > Distances, and by clicking on the two atoms of interest. Proteins that have a strong hot spot ($S > 16$) and other hot spots that are further from this site than 8 Å still may be druggable by molecules beyond the traditional druglike chemical space, i.e., macrocyclic compounds or peptides⁷⁸.

The FTMap results can also provide information to help the discovery of ligands via a Fragment Based Drug Discovery (FBDD) approach²⁹. FBDD uses a screening library of low molecular weight compounds to identify compounds that bind to the site of interest. However, the binding affinity is generally very low and can be greatly improved by adding further functional groups or by linking two fragments that occupy neighboring sites. Since FTMap identifies regions that are most likely to contribute to ligand binding free energy, it can be used to detect regions that will be important for FBDD. We have demonstrated that the core fragment generally binds in CS000 with the highest probe cluster count²⁹. Surrounding secondary CSs designate the hot spots that can be used to expand the initial fragment into higher affinity ligands. In a recent study²⁹, we mapped ten different targets in which structural data were available for structures with just a core fragment bound as well as structures with a ligand expanded around the core site. Mapping the apo structures revealed that for nine out of ten of these targets, the highest ranking CS coincided with the core fragment. In eight of these cases, secondary CSs overlapped with known fragment extensions. In the case where the highest ranked CS did not overlap with the fragment, the second highest CS overlapped with the fragment and the first highest CS coincided with an extension site. An example is shown in Fig. 12 for thrombin. The mapping of an apo structure of thrombin (PDB ID 1ths)⁷⁹ reveals 4 CSs: CS000 with 25 probe clusters, CS002 with 13 probe clusters, and CS004 and CS005, both with 7 probe clusters. CS000 and CS004 overlap with a known fragment (Fragment 1 in Fig. 12, from PDB ID 2c90)⁸⁰ with an IC_{50} value of 330 μ M. CS002 and CS005 overlap with a second fragment (Fragment 2 in Fig. 12, from PDB ID 2c93)⁸¹ with an IC_{50} value of 12 μ M. Linking these two fragments provided a starting point for optimization of a high affinity ligand, of which several ligands with nanomolar affinity were found⁸⁰. A ligand with an IC_{50} value of 4 nM is shown in the right panel of Fig. 12, overlaid with the mapping results²⁹.

Identification of ligand binding sites using FTSite

As an extension of FTMap, the server FTSite (<http://ftsites.bu.edu>) uses additional clustering of consensus clusters to yield ligand binding sites. FTSite is based on the observation that ligand binding sites also bind small organic molecules of various shapes and polarity, and thus contain at least one hot spot²². Further analysis has shown that nearby weaker hot spots also belong to the same binding site. Accordingly, FTSite first selects the consensus cluster with the highest number of contacts. This cluster is then expanded by adding any neighboring consensus cluster if the center of any of its probe is closer than 3.5 Å to the center of any probe in the consensus cluster. The protein residues that are within 4 Å of the expanded consensus cluster constitute the top prediction of the binding site. The first consensus cluster is then removed, and the procedure is repeated using the next consensus

cluster with the highest number of contacts to identify lower ranked predictions of the ligand binding site. As an example, Fig. 6 shows the screen capture of FTSite results for human lymphocyte kinase (Lck, PDB ID 3lck)⁸². A bound inhibitor from the structure with PDB ID 1qpe⁸³ is superimposed for reference, shown in white ball-and-stick representation. The figure on the left shows the top prediction of the ligand binding site, named Site 1, using mesh representation for the cluster of probe molecules found at this site. The output page also lists the residues that are within 5 Å of the binding site found. The plot reveals that the binding site identified covers only about half of the inhibitor. While site 2 extends the binding pocket in a direction that does not interact with the this particular inhibitor, adding Site 3 to Site 1 covers the entire region of inhibitor binding (see figure on the right). The residues interacting with Site 3 are also listed. Unlike many other binding site prediction methods, the FTSite algorithm is purely energy based and does not rely on any evolutionary or statistical information. Nevertheless, applications of FTSite to established test sets have shown that we achieve near experimental accuracy, as the top ranked prediction of the binding site was correct for over 94% of apo proteins²². Based on the available information, it appears that FTSite is the best server for predicting ligand binding sites based on the x-ray structure of ligand-free proteins²².

Accounting for side chain flexibility using FTFlex

The function of FTFlex is accounting for binding site flexibility around user-selected hot spots in order to improve the accuracy of mapping. The method is particularly useful when the goal is opening potentially druggable pockets in protein-protein interfaces. Stage 1 of FTFlex maps the given structure using FTMap and presents the results for the selection of hot spots for further analysis. Fig. 7 shows the transition from Stage 1 to Stage 2 of FTFlex applied to mapping chain A of the apo structure of the cyclin-dependent kinase 2 (CDK2, PDB ID 1pw2)⁸⁴. Based on preliminary information on kinase binding sites, consensus sites 002 (13 probe clusters) and 004 (10 probe clusters) were selected by clicking on the appropriate buttons for further analysis of binding site hot spots. The outcome of Stage 1 is also shown in Fig. 13a, where an inhibitor bound to a CDK2 from the structure with PDB ID 1ke5⁸⁵ (shown as grey sticks) is superimposed on the mapping results. In addition to the bound inhibitor, the composite figure includes two side chains, Lys33 and Lys89 that substantially change conformation upon inhibitor binding. In the apo structure the two side chains, shown as brown sticks, protrude into the ligand binding site and substantially reduce its volume, preventing the formation of larger probe clusters. In the inhibitor-bound structure the two side chains, shown as grey sticks, move out of the pocket, and thus do not interfere with ligand binding. After the selection of hot spots for further analysis, clicking on the “Submit” button (shown in Fig. 7), initiates Stage 2 of FTFlex. The results for CDK2 in Fig. 13b show that the algorithm moves both Lys33 and Lys89 out from the ligand binding site, resulting in conformers shown as brown sticks. The Lys33 side chain is close to the conformer in the inhibitor-bound structure 1ke5. Although this is not the case for Lys89, the side chain moves farther from the pocket, and thus does not interfere with probe binding. As a result, the ligand binding site now includes the two most populated consensus sites, CS000 (green, 18 probe clusters) and CS001 (magenta, 14 probe clusters). Notice that based on the FTMap results (i.e., Stage 1 of FTFlex) we would consider CDK2 only borderline

druggable, but after adjusting the side chains by FTFlex the mapping results indicate that the protein is druggable.

Adding user-defined probes in FTMap/param

The goal of FTMap/param is to examine the binding mode of user-selected small molecules relative to the hot spots defined by the clustering of the standard 16 probes. As an example, we have mapped the unbound structure of thrombin (PDB ID: 1hxf, see⁸⁶) by adding the small compound 1-(3-chlorophenyl)methanamine (HETATM ID: C2A), which was used in an FBDD campaign and has been co-crystallized with thrombin (PDB ID: 2c8z)⁸¹. FTMap/param identifies the important hot spots based on the consensus clusters of the 16 standard probes, and reports the lowest energy cluster representatives of C2A within 4 Å of each of the consensus cluster, using the geometric center distances in the calculations. Fig. 8 shows the PyMOL session created by FTMap/param. The molecules shown are thrombin (green cartoon), the strongest consensus site represented by the probes at clusters centers (cyan lines), and the highest probability pose of C2A (yellow sticks) generated by the mapping and located close to a consensus site. It is interesting to note that this pose is almost identical to the bound pose of C2A, co-crystallized with thrombin (PDB ID: 2c8z)⁸¹. Although C2A is a weak binder with an IC₅₀ of only 300 µM, FTMap/param was capable of detecting the interaction based on the mapping of the apo protein. The pose identified by computational mapping of C2A is particularly interesting, because the chlorophenyl group occupies the S₁ site of thrombin fully and the NH₂ group protrudes from the pocket, indicating the possibility of expanding the molecule. Indeed, C2A was subsequently joined with a 12 µM ligand to generate a 220 nM inhibitor⁸⁷. Therefore, in this case, computational mapping recapitulated the important protein-ligand interactions, and this type of information can be very useful for screening candidate molecules in fragment based drug design.

Mapping ensembles of structures using FTDyn

FTDyn is a server designed for hot spots analysis on structural ensemble of proteins using the FTMap without local energy minimization, which reduces somewhat the accuracy of locating the hot spots in the individual structures, but enables the server to map even very large ensembles³². The primary mapping results are provided in the form of frequencies of probe-residue interactions. Fig. 9 shows the result page applying FTDyn to an ensemble of 24 structures of the human MDM2 (PDB ID 1z1m)⁸⁸, determined by Nuclear Magnetic Resonance (NMR). The page shows the first structure of the ensemble in cartoon representation with residues color-coded from blue to red based on the frequency of nonbonded contacts averaged over the entire ensemble (Fig. 9a). Clicking on the picture of the molecule activates JSmol. The result page also presents the ensemble-averaged frequency of nonbonded and hydrogen bond interactions in form of bar graphs (Fig. 9b). Below the graphs the table can be used to view or download mapping results for all structures, or for any of the structures of the ensemble individually. The screenshot of the PyMOL session, downloaded from the result page (Fig. 10a) show substantial variation among the structures within the NMR derived MDM2 ensemble. However, the color-coding from blue to red based on contact frequency reveals that largely the same residues interact with the probes in all structures. The bar graph in Fig. 10b compares the distribution of nonbonded probe-protein contacts for residues of MDM2, based on the mapping of model 9

in the ensemble (blue bars), to the distribution of nonbonded ligand-protein contacts observed in the complex of MDM2 with piperidinone, a small molecular inhibitor of the MDM2-p53 interaction (PDB ID 2lzg)⁸⁹, shown as red bars. Model 9 was selected because it had the maximum hit rate, defined as the total number of probe-protein interactions in the binding site, among all models. The comparison demonstrates that the averaging of probe-protein interactions identifies and ranks by importance the binding site residues fairly well, in spite of the substantial structural variation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This investigation was supported by grants GM064700 from the National Institute of General Medical Sciences

References

1. DeLano WL, Ultsch MH, de Vos AM, Wells JA. Convergent solutions to binding at a protein-protein interface. *Science*. 2000; 287:1279–1283. [PubMed: 10678837]
2. Thanos CD, DeLano WL, Wells JA. Hot-spot mimicry of a cytokine receptor by a small molecule. *Proc Natl Acad Sci U S A*. 2006; 103:15422–15427. [PubMed: 17032757]
3. DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol*. 2002; 12:14–20. [PubMed: 11839484]
4. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. *Science*. 1995; 267:383–386. [PubMed: 7529940]
5. Keskin O, Ma BY, Nussinov R. Hot regions in protein-protein interactions: The organization and contribution of structurally conserved hot spot residues. *J Mol Biol*. 2005; 345:1281–1294. [PubMed: 15644221]
6. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J Mol Biol*. 1998; 280:1–9. [PubMed: 9653027]
7. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci U S A*. 2002; 99:14116–14121. [PubMed: 12381794]
8. Hajduk PJ, Huth JR, Fesik SW. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem*. 2005; 48:2518–2525. [PubMed: 15801841]
9. Vajda S, Guarnieri F. Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr Opin Drug Discov Devel*. 2006; 9:354–362.
10. Seco J, Luque FJ, Barril X. Binding Site Detection and Druggability Index from First Principles. *J Med Chem*. 2009; 52:2363–2371. [PubMed: 19296650]
11. Mattos C, Ringe D. Locating and characterizing binding sites on proteins. *Nat Biotechnol*. 1996; 14:595–599. [PubMed: 9630949]
12. Allen KN, et al. An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem*. 1996; 100:2605–2611.
13. Ciulli A, Williams G, Smith AG, Blundell TL, Abell C. Probing hot spots at protein-ligand binding sites: A fragment-based approach using biophysical methods. *J Med Chem*. 2006; 49:4992–5000. [PubMed: 16884311]
14. Dennis S, Kortvelyesi T, Vajda S. Computational mapping identifies the binding sites of organic solvents on proteins. *Proc Natl Acad Sci U S A*. 2002; 99:4290–4295. [PubMed: 11904374]
15. Silberstein M, et al. Identification of substrate binding sites in enzymes by computational solvent mapping. *J Mol Biol*. 2003; 332:1095–1113. [PubMed: 14499612]
16. Landon MR, et al. Detection of ligand binding hot spots on protein surfaces via fragment-based methods: application to DJ-1 and glucocerebrosidase. *J Comput Aid Mol Des*. 2009; 23:491–500.

17. Kuttner YY, Engel S. Protein Hot Spots: The Islands of Stability. *J Mol Biol.* 2012; 415:419–428. [PubMed: 22100447]
18. Landon MR, Lancia DR Jr, Yu J, Thiel SC, Vajda S. Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J Med Chem.* 2007; 50:1231–1240. [PubMed: 17305325]
19. Brenke R, et al. Fragment-based identification of druggable ‘hot spots’ of proteins using Fourier domain correlation techniques. *Bioinformatics.* 2009; 25:621–627. [PubMed: 19176554]
20. Chuang GY, et al. Binding hot spots and amantadine orientation in the influenza A virus M2 proton channel. *Biophys J.* 2009; 97:2846–2853. [PubMed: 19917240]
21. Landon MR, et al. 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.* 2008; 71:106–116. [PubMed: 18205727]
22. Ngan CH, et al. FTSite: high accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics.* 2012; 28:286–287. [PubMed: 22113084]
23. Villar EA, et al. How proteins bind macrocycles. *Nat Chem Biol.* 2014; 10:723–731. [PubMed: 25038790]
24. Zerbe BS, Hall DR, Vajda S, Whitty A, Kozakov D. Relationship between Hot Spot Residues and Ligand Binding Hot Spots in Protein-Protein Interfaces. *J Chem Inf Model.* 2012; 52:2236–2244. [PubMed: 22770357]
25. Rees DC, Congreve M, Murray CW, Carr R. Fragment-based lead discovery. *Nat Rev Drug Discov.* 2004; 3:660–672. [PubMed: 15286733]
26. Erlanson DA, McDowell RS, O’Brien T. Fragment-based drug discovery. *J Med Chem.* 2004; 47:3463–3482. [PubMed: 15214773]
27. Hartshorn MJ, et al. Fragment-based lead discovery using X-ray crystallography. *J Med Chem.* 2005; 48:403–413. [PubMed: 15658854]
28. Kozakov D, et al. Structural conservation of druggable hot spots in protein-protein interfaces. *Proc Natl Acad Sci U S A.* 2011; 108:13528–13533. [PubMed: 21808046]
29. Hall DR, Ngan CH, Zerbe BS, Kozakov D, Vajda S. Hot spot analysis for driving the development of hits into leads in fragment-based drug discovery. *J Chem Inf Model.* 2012; 52:199–209. [PubMed: 22145575]
30. Grove LE, Hall DR, Beglov D, Vajda S, Kozakov D. FTFlex: accounting for binding site flexibility to improve fragment-based identification of druggable hot spots. *Bioinformatics.* 2013; 29:1218–1219. [PubMed: 23476022]
31. Ngan CH, et al. FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res.* 2012; 40:W271–275. [PubMed: 22589414]
32. Bohnuud T, Kozakov D, Vajda S. Evidence of conformational selection driving the formation of ligand binding sites in protein-protein interfaces. *PLoS Comput Biol.* 2014; 10:e1003872. [PubMed: 25275445]
33. Ivetac A, McCammon JA. Mapping the druggable allosteric space of G-protein coupled receptors: a fragment-based molecular dynamics approach. *Chem Biol Drug Des.* 2010; 76:201–217. [PubMed: 20626410]
34. Hall DH, et al. Robust identification of binding hot spots using continuum electrostatics: application to hen egg-white lysozyme. *J Am Chem Soc.* 2011; 133:20668–20671. [PubMed: 22092261]
35. Hall DR, Kozakov D, Vajda S. Analysis of protein binding sites by computational solvent mapping. *Methods Mol Biol.* 2012; 819:13–27. [PubMed: 22183527]
36. Brooks BR, et al. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem.* 1983; 4:187–217.
37. Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. *J Phys Chem.* 1996; 100:1578–1599.
38. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem.* 1985; 28:849–857. [PubMed: 3892003]

39. Miranker A, Karplus M. Functionality Maps of Binding-Sites - a Multiple Copy Simultaneous Search Method. *Proteins-Structure Function and Genetics*. 1991; 11:29–34.
40. Mattos C, et al. Multiple solvent crystal structures: probing binding sites, plasticity and hydration. *J Mol Biol*. 2006; 357:1471–1482. [PubMed: 16488429]
41. Berman HM, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28:235–242. [PubMed: 10592235]
42. Bohnuud T, et al. Computational mapping reveals dramatic effect of Hoogsteen breathing on duplex DNA reactivity with formaldehyde. *Nucleic Acids Res*. 2012; 40:7644–7652. [PubMed: 22705795]
43. Chuang GY, Kozakov D, Brenke R, Comeau SR, Vajda S. DARS (Decoys As the Reference State) potentials for protein-protein docking. *Biophys J*. 2008; 95:4217–4227. [PubMed: 18676649]
44. Beglov D, et al. Minimal ensembles of side chain conformers for modeling protein-protein interactions. *Proteins*. 2012; 80:591–601. [PubMed: 22105850]
45. O'Boyle NM, Vandermeersch T, Flynn CJ, Maguire AR, Hutchison GR. Confab - Systematic generation of diverse low-energy conformers. *J Cheminform*. 2011; 3:8. [PubMed: 21410983]
46. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model*. 2006; 25:247–260. [PubMed: 16458552]
47. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*. 2004; 25:1157–1174. [PubMed: 15116359]
48. Alexeev Y, Mazanetz MP, Ichihara O, Fedorov DG. GAMESS as a free quantum-mechanical platform for drug research. *Curr Top Med Chem*. 2012; 12:2013–2033. [PubMed: 23110536]
49. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem*. 2002; 23:1623–1641. [PubMed: 12395429]
50. Votapka L, Amaro RE. Multistructural hot spot characterization with FTProd. *Bioinformatics*. 2013; 29:393–394. [PubMed: 23202744]
51. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J Mol Biol*. 2002; 320:369–387. [PubMed: 12079393]
52. Zhu X, Mitchell JC. KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins*. 2011; 79:2671–2683. [PubMed: 21735484]
53. Tuncbag N, Gursoy A, Keskin O. Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics*. 2009; 25:1513–1520. [PubMed: 19357097]
54. Deng L, et al. PredHS: a web server for predicting protein-protein interaction hot spots by using structural neighborhood properties. *Nucleic Acids Res*. 2014; 42:W290–295. [PubMed: 24852252]
55. Laurie AT, Jackson RM. Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. *Curr Protein Pept Sci*. 2006; 7:395–406. [PubMed: 17073692]
56. Levitt DG, Banaszak LJ. Pocket - a Computer-Graphics Method for Identifying and Displaying Protein Cavities and Their Surrounding Amino-Acids. *J Mol Graph*. 1992; 10:229–234. [PubMed: 1476996]
57. Huang B, Schroeder M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct Biol*. 2006; 6:19. [PubMed: 16995956]
58. Hendlich M, Rippmann F, Barnickel G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J Mol Graph Model*. 1997; 15:359–363. 389. [PubMed: 9704298]
59. Brady GP Jr, Stouten PF. Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aid Mol Des*. 2000; 14:383–401.
60. Binkowski TA, Naghibzadeh S, Liang J. CASTp: Computed atlas of surface topography of proteins. *Nucleic Acids Res*. 2003; 31:3352–3355. [PubMed: 12824325]

61. Wass MN, Kelley LA, Sternberg MJ. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res.* 2010; 38:W469–473. [PubMed: 20513649]
62. Brylinski M, Skolnick J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc Natl Acad Sci U S A.* 2008; 105:129–134. [PubMed: 18165317]
63. Chou KC, Cai YD. A novel approach to predict active sites of enzyme molecules. *Proteins.* 2004; 55:77–82. [PubMed: 14997541]
64. Laurie ATR, Jackson RM. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics.* 2005; 21:1908–1916. [PubMed: 15701681]
65. Halgren TA. Identifying and Characterizing Binding Sites and Assessing Druggability. *J Chem Inf Model.* 2009; 49:377–389. [PubMed: 19434839]
66. Hernandez M, Ghersi D, Sanchez R. SITEHOUND-web: a server for ligand binding site identification in protein structures. *Nucleic Acids Res.* 2009; 37:W413–W416. [PubMed: 19398430]
67. English AC, Done SH, Caves LS, Groom CR, Hubbard RE. Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins.* 1999; 37:628–640. [PubMed: 10651278]
68. English AC, Groom CR, Hubbard RE. Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Eng.* 2001; 14:47–59. [PubMed: 11287678]
69. Haider MK, Bertrand HO, Hubbard RE. Predicting Fragment Binding Poses Using a Combined MCSS MM-GBSA Approach. *J Chem Inf Model.* 2011; 51:1092–1105. [PubMed: 21528911]
70. Lexa KW, Carlson HA. Improving Protocols for Protein Mapping through Proper Comparison to Crystallography Data. *J Chem Inf Model.* 2013; 53:391–402. [PubMed: 23327200]
71. Bakan A, Nevins N, Lakdawala AS, Bahar I. Druggability Assessment of Allosteric Proteins by Dynamics Simulations in the Presence of Probe Molecules. *J Chem Theory Comput.* 2012; 8:2435–2447. [PubMed: 22798729]
72. Raman EP, Yu W, Lakkaraju SK, MacKerell AD Jr. Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach. *J Chem Inf Model.* 2013; 53:3384–3398. [PubMed: 24245913]
73. Yu W, Lakkaraju SK, Raman EP, Mackerell AD Jr. Site-Identification by Ligand Competitive Saturation (SILCS) assisted pharmacophore modeling. *J Comput Aided Mol Des.* 2014
74. Sielecki AR, et al. Structure of Recombinant Human Renin, a Target for Cardiovascular-Active Drugs, at 2.5 Å Resolution. *Science.* 1989; 243:1346–1351. [PubMed: 2493678]
75. Rahuel J, et al. Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. *Chem Biol.* 2000; 7:493–504. [PubMed: 10903938]
76. Rahuel J, Priestle JP, Grutter MG. The crystal structures of recombinant glycosylated human renin alone and in complex with a transition state analog inhibitor. *J Struct Biol.* 1991; 107:227–236. [PubMed: 1807356]
77. Dechene M, Wink G, Smith M, Swartz P, Mattos C. Multiple solvent crystal structures of ribonuclease A: an assessment of the method. *Proteins.* 2009; 76:861–881. [PubMed: 19291738]
78. Villar EA, et al. How proteins bind macrocycles. *Nat Chem Biol.* 2014; 10:723–731. [PubMed: 25038790]
79. Qiu XY, Yin ML, Padmanabhan KP, Krstenansky JL, Tulinsky A. Structures of Thrombin Complexes with a Designed and a Natural Exosite Peptide Inhibitor. *J Biol Chem.* 1993; 268:20318–20326. [PubMed: 8376390]
80. Howard N, et al. Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. *J Med Chem.* 2006; 49:1346–1355. [PubMed: 16480269]
81. Howard N, et al. Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. *J Med Chem.* 2006; 49:1346–1355. [PubMed: 16480269]
82. Yamaguchi H, Hendrickson WA. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature.* 1996; 384:484–489. [PubMed: 8945479]
83. Zhu XT, et al. Structural analysis of the lymphocyte-specific kinase Lck in complex with non-selective and Src family selective kinase inhibitors. *Struct Fold Des.* 1999; 7:651–661.

84. Wu SY, et al. Discovery of a novel family of CDK inhibitors with the program LIDAEUS: Structural basis for ligand-induced disordering of the activation loop. *Structure*. 2003; 11:399–410. [PubMed: 12679018]
85. Bramson HN, et al. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): Design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J Med Chem*. 2001; 44:4339–4358. [PubMed: 11728181]
86. Zhang E, Tulinsky A. The molecular environment of the Na⁺ binding site of thrombin. *Biophys Chem*. 1997; 63:185–200. [PubMed: 9108691]
87. Howard N, et al. Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. *J Med Chem*. 2006; 49:1346–1355. [PubMed: 16480269]
88. Uhrinova S, et al. Structure of free MDM2 N-terminal domain reveals conformational adjustments that accompany p53-binding. *J Mol Biol*. 2005; 350:587–598. [PubMed: 15953616]
89. Michelsen K, et al. Ordering of the N-Terminus of Human MDM2 by Small Molecule Inhibitors. *J Am Chem Soc*. 2012; 134:17059–17067. [PubMed: 22991965]

FIGURE 1.

Principles and tools of the FTMap algorithm. **(a)** Visualization of the FTMap process. In Step 1, probe molecules are docked. They are then energy minimized to find the most favorable positions (Step 2) and each probe type is clustered to yield probe clusters. These probe clusters are then clustered on the basis of their average free energy in Step 3 to yield consensus clusters, which defines the consensus sites (CSs). The CS with the largest number of probe clusters is the main hot spot; all other CSs are secondary hot spots. **(b)** The 16 small organic probe molecules that are used by FTMap. These probes vary in size, shape, and polarity. **(c)** An example of a consensus cluster in a protein pocket. The probes

molecules depicted in this case are representatives of their probe clusters. The figure shows the binding pose of the lowest energy structure from each of the 16 probe clusters that define the hot spot. The color codes represent different atoms, carbon cyan, oxygen red, nitrogen blue, and hydrogen white.

FIGURE 2.

Flowchart of the FTMap algorithm. Users upload a PDB to the server manually or using pdb.org. FTMap checks the structure, removes bound ligands and water molecules, and adds any missing atoms, including polar hydrogens. After calculation of the Poisson-Boltzmann potential, positions for the first probe molecule are sampled using rigid body docking. This step uses the FFT correlation approach and a detailed energy expression, including terms for the van der Waals energy, the electrostatic interaction energy, a cavity term describing the hydrophobic contributions of the cavity and a knowledge-based pairwise potential. The 2000 best probe poses are retained and minimized using the CHARMM potential, which includes

solvation. The probes are then clustered starting with the lowest energy structure and using a 3 Å clustering radius. Clusters with less than 10 members are excluded from consideration. The clusters are ranked on the basis of their Boltzmann averaged energies and the six lowest energy clusters are retained. This step is repeated for all remaining probe molecules. When all 16 probe molecule types have been sampled and clustered, the probe clusters themselves are clustered. The probe cluster with the maximum number of neighbors within 4 Å is selected as the top consensus site (CS000) and all probe clusters within 4 Å of this CS are included. This procedure is repeated until all probe clusters have been assigned to a CS. CSs are ranked on the basis of the number of their probe clusters.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 3.

Flowchart of the FTFlex algorithm. Stage 1 is the identification of the hot spots using the FTMap algorithm. After Stage 1 is completed, the program stops, and the user can select the ligand binding region, or other region of interest, by specifying the consensus sites that define this region. After this selection the server proceeds to Stage 2, considering the residues that are within 5Å of the selected hot spots and satisfy hydrophobicity and cavity size restrictions. In Stage 2 possible rotamers are determined for each of the selected residue, inserted back into the initial PDB structure, and the resulting structures are mapped again using FTMap. After all rotamers have been tested, the rotamers that yield the most improvement are substituted back into the initial PDB file to yield a modified structure, which is mapped and the results are returned to the user.

FIGURE 4.

Screen image of FTMap results for PDB ID 2ren (apo structure of renin). Five different output files are available for download; descriptions can be found in Box 3. The user can view the mapping results using the PyMol plugin, or by downloading either the PDB file containing the protein and probe coordinate or a PyMol session. The target protein is shown as a green cartoon, and the probes representing each cluster at the consensus sites are shown in various colors as sticks. Clicking on the image will activate JSmol, and the image can be manipulated (e.g., rotated and translated) using the JSmol tools. JSmol also provides checkboxes along the bottom of the picture of the protein to select/deselect any of the consensus clusters. The two bar graphs at the bottom of the page show the percent of nonbonded and hydrogen bond interactions, respectively, between the probes and the protein for each residue along the protein sequence. You can download these results as tab separated

files with exact residue contact counts by clicking the link located directly beneath each respective graph.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 5.

Viewing FTMap results for PDB ID 2ren (apo structure of renin) using PyMOL. **(a)** Opening the PyMOL session provided by the server. In addition to the protein and the representative probes at the consensus sites, PyMOL lists consensus clusters on the right hand side of the screen in order of probe cluster ranking. The format of the consensus clusters is “crossclusters.xxx.yyy” where xxx is the ranking of the consensus site, starting at 000, and yyy is the number of probe clusters. **(b)** Demonstrating the value of looking at the FTMap results using PyMOL: The inhibitor Aliskiren (magenta sticks) from the inhibitor-bound renin structure 2v0z is superimposed onto the hot spots calculated for the renin apo structure, which shows that the entire inhibitor binding site is covered by hot spots (For better viewing the protein is not shown).

FIGURE 6.

Screen capture of FTSite results for human lymphocyte kinase (Lck, PDB ID 3lck). A bound inhibitor from the structure with PDB ID 1qpe is superimposed for reference, shown in white ball-and-stick representation. The figure on the left shows the top prediction of the ligand binding site, named Site 1, using mesh representation for the cluster of probe molecules found at this site. The output page also lists the residues that are within 5 Å of the binding site found. The plot reveals that the binding site identified covers only about half of the inhibitor. While site 2 extends the binding pocket in a direction that does not interact with this particular inhibitor, adding Site 3 to Site 1 covers the entire region of inhibitor binding (see figure on the right). The residues interacting with Site 3 are also listed.

FIGURE 7.

Transition from Stage 1 to Stage 2 in FTFlex: selection of consensus clusters around which low energy side chain conformers will be explored in repeated mapping calculations. The figure shows the output page from Stage 1 of FTFlex applied to Chain A of the apo structure of the cyclin-dependent kinase 2 (CDK2, PDB ID 1pw2). Consensus sites 002 and 004 were selected by clicking on the appropriate buttons. Notice the small “Submit” button for starting Stage 2 of FTFlex.

FIGURE 8.

Screenshot of the PyMOL session from the application of FTMap/param to the unbound structure of thrombin (PDB ID: 1hxf) with the user-defined probe molecule 1-(3-chlorophenyl)methanamine, a thrombin inhibitor. We show the lowest energy pose for conformer 1 of this additional probe (yellow sticks), overlapping with the largest consensus site (cyan, 30 probe clusters). The 1-(3-chlorophenyl)methanamine molecule bound to thrombin (magenta) from the structure with PDB ID 2c8z is superimposed for reference.

FIGURE 9.

Screenshots of the result page when applying FTDyn to map an ensemble of 24 structures of the MDM2 protein (PDB ID 1z1m), determined by Nuclear Magnetic Resonance (NMR). **(a)** The page shows the first structure of the ensemble in cartoon representation, but the color coding is based on the number of probe-protein contacts averaged over the entire ensemble. Residues are colored from blue to red based on contact frequency. Clicking on the picture of the molecule activates JSmol, and thus the structure can be rotated and translated. **(b)** The two bar graphs on the page show the percent of nonbonded and hydrogen bond interactions, respectively, averaged over all structures. Below the graphs the table can be used to view or download mapping results for all structures together, or any of the structures of the ensemble individually. Clicking on the Load button under the Map tag will activate JSmol for the visualization of the selected structure. The PDB file of the mapping results and the lists of nonbonded and hydrogen bond interactions can also be downloaded for each structure.

FIGURE 10.

Mapping of the 24 MDM2 structures obtained by NMR using FTDyn. **(a)** Screenshot of the PyMOL session shows the ensemble, with residues color-coded from blue to red based on contact frequency. **(b)** Comparing the distribution of nonbonded probe-protein contacts for residues of MDM2, based on the mapping of model 9 in the ensemble (blue bars), to the distribution of nonbonded ligand-protein contacts observed in the complex of MDM2 with piperidinone, a small molecular inhibitor of the MDM2-p53 interaction (PDB ID 2lzg), shown as red bars. Horizontal axes list residues of MDM2 from Glu25 to Tyr104 (unstructured regions were removed before mapping analysis). The vertical axis shows the fraction of atom-atom interactions that each protein residue makes with probe or ligand atoms.

FIGURE 11.

Comparison of FTMap results to experimental data. **(a)** Probe binding to Ribonuclease A (RNase A), observed by the Multiple Solvent Crystal Structures (MSCS) methods using X-ray crystallography. In all cases, the CSs are shown as colored sticks and RNase A is shown as a tan surface. The sites identified using MSCS are circled and labeled as B1, B2, and P1. **(b)** Consensus sites for the mapping of the unbound structure of RNase using PDB ID 2e3w. The CSs are as follows: CS000 (20 probe clusters, cyan), CS001 (16 probe clusters, magenta), CS002 (13 probe clusters, yellow), CS003 (13 probe clusters, salmon), and CS004 (11 probe clusters, white). Although CS002 is not seen in the MSCS experiments, the existence of a hot spot at that location has been demonstrated by alanine scanning data (see Anticipated results).

FIGURE 12.

Consensus sites identified using FTMap for thrombin (PDB ID 1ths) are shown in line representation, overlapping with fragments and a high affinity ligand shown as sticks. Mapping identifies 4 consensus sites (left). Of these sites, CS000 (cyan) has 25 probe clusters, CS002 (yellow) has 13 probe clusters, CS004 (white) has 7 probe clusters, and CS005 (periwinkle) has 7 probe clusters. These consensus sites overlap with two known fragments (middle), Fragment 1 (fushia sticks, PDB ID 2c90) and Fragment 2 (pale green sticks, PDB ID 2c93). Joining the two fragments to take full advantage of all hot spots yields an inhibitor (PDB ID 2c8w) with an IC_{50} value of 4 nM, shown in teal sticks overlaid with the consensus sites.

FIGURE 13.

Flexible mapping the apo structure of the cyclin-dependent kinase 2 (CDK2, PDB ID 1pw2) using FTFlex. **(a)** Results from Stage 1, i.e., without adjusting the side chains in the X-ray structure. The figure shows CS002 (green, 13 probe clusters) and CS004 (magenta, 10 probe clusters) in the ligand binding site. An inhibitor bound to CDK2 (from the structure with PDB ID 1ke5) is superimposed for reference (grey sticks). We also show two “moving” side chains, Lys33 and Lys89, that with their conformers in the apo structure protrude into the ligand binding site and substantially reduce its volume (brown sticks). In the inhibitor-bound structure (PDB ID 1ke5) the two side chains move out of the pocket, and thus do not interfere with ligand binding (also shown as grey sticks). **(b)** Results from Stage 2 of FTFlex. Both Lys33 and Lys89, shown as brown sticks, have largely moved out from the ligand binding site. The Lys33 side chain is close to the conformer in the inhibitor-bound structure 1ke5. Although this is not the case for Lys89, the side chain moves farther from the pocket, and thus does not interfere with probe binding. As a result, the ligand binding site now includes the two most populated consensus sites, CS000 (green, 18 probe clusters) and CS001 (magenta, 14 probe clusters).

TABLE 1

The FTMap family of servers.

Server (URL)	Function	Required Input
http://ftmap.bu.edu	Identification of binding hot spots; determining druggability; and providing information for fragment based drug discovery.	PDB ID of a protein, DNA, or RNA structure; or structure file in PDB format.
http://ftsites.bu.edu	Identification of likely ligand binding sites; ranking of sites in terms of probe-protein interactions; and listing of binding site residues.	PDB ID of a protein, structure; or protein structure file in PDB format; or a Zip file containing up to 15 protein PDB files.
http://ftflex.bu.edu	Identification of binding hot spots of proteins and determining druggability, while accounting for side chain flexibility around selected hot spots; opening pockets in protein-protein interfaces.	Stage 1: PDB ID of a protein, structure; or protein structure file in PDB format. Stage 2: Selection of hot spots for side chain adjustment and re-mapping.
http://ftmap.bu.edu/param/	Same as FTMap, plus determining the low energy binding poses of up to 15 user selected probe molecules in hot spot regions.	PDB ID of a protein, DNA, or RNA structure; or structure file in PDB format; plus formal charges and SMILES strings to define additional probes.
http://ftdyn.bu.edu	Mapping potentially large ensembles of protein structures; determining probe-protein interactions for each structure, and averages over the ensemble; Identification of the structure most similar to a ligand-bound conformation.	Ensemble of protein structures specified in PDB model record format.

TABLE 2

Troubleshooting table.

Step	Problem and solution
3	<p>Problem: I uploaded a structure from my computer, but as soon as I submitted the job failed with the message: "Processing failed on receptor." What is wrong?</p> <p>Solution: Make sure that your file is in PDB format and has the required spacing between columns. Documentation with respect to the PDB file format can be found at: http://www.wwpdb.org/docs.html. Also, make sure that the PDB file does not contain any non-protein residues within the ATOM record. If it does, remove these using a visualizing tool (e.g. PyMol) or text editor.</p> <p>Problem: I want to map all models for an NMR ensemble, how do I do this?</p> <p>Solution: Using FTMap each model needs to be mapped individually. Individual models can be mapped by generating a separate PDB file for each model in the set using PyMol or a text editor. However, a less accurate mapping of the entire ensemble can be performed using FTDyn. Note that models that yield the highest number of probe clusters within a binding site most likely have the highest relevance to the bound structure.</p> <p>Problem: The target I am mapping has multiple experimental structures. Some have ligands bound and some are of the unbound state. How do I select the best structure to map?</p> <p>Solution: Each of the structures can be mapped individually and the final results can be compared. Typically the unbound structure yields the best results for the analysis of druggability. Mapping both unbound and ligand-bound structures and comparing the results provide information on the changes in binding properties due to ligand binding.</p>
4	<p>Problem: How do I know which chains to include?</p> <p>Solution: Many structures have multiple chains with the same sequence, but small structural differences. In this case, each chain should be independently mapped. The differences in mapping results provide some information on the effect of conformational variation. Generally, the structure with the largest consensus site population can be used for the analysis of druggability.</p> <p>Problem: What if the chain I am submitting has multiple domains? Should I split them up or map the entire structure?</p> <p>Solution: FTMap works best by mapping only single domains. In this case, you can use one of three different methods for determining the cutoffs for the domains: SCOP, CATH, or PFAM. Access to all three of these methods can be found on the PDB website for individual structures. A PDB format file including only the selected domain should be prepared and submitted to FTMap. The separate file can be prepared by selecting the appropriate residues in PyMOL.</p>
5	<p>Problem: I entered a HETATM record to include in the calculation, but it was not recognized. What do I do?</p> <p>Solution: At this point there is no option for considering heteroatoms (HETATM entries) that are not included in Table S1 of the Supplementary Information, which shows the list of parameterized HETATM entries. If your HETATM record is a coenzyme, you might consider creating a mask file so that FTMap will not consider that region of the protein surface. See Step 7 for information on how to create a mask file.</p>
11	<p>Problem: My job keeps crashing even though my input file looks fine.</p> <p>Solution: Big proteins run out of memory, potentially causing the calculation to crash during the pre-minimization, PB potential, or CHARMM minimization steps. Try to make your protein smaller by using a single chain for a multi-chain protein, or split a single chain into domains.</p>
15	<p>Problem: I cannot load the mapping results into my molecular viewer.</p> <p>Solution: Your viewer likely does not support multiple structures in one PDB file. Each crosscluster (cluster of probe clusters defining a consensus site) is listed individually, thus the viewer must be capable of simultaneously loading multiple structures. Try using PyMol, available at www.pymol.org.</p> <p>Problem: When clicking on the PyMol Session tab, a text file is downloaded rather than a .pse file that can be opened in PyMOL</p> <p>Solution: Edit the file name by removing the .txt extension</p> <p>Problem: The PyMol session tab opens with the red background.</p> <p>Solution: The older free version of PyMOL is not fully compatible with the newer version used by FTMap for creating the .pse file. In PyMOL, open the Display menu and change the background to black to best viewing.</p>
20	<p>Problem: I uploaded a structure from my computer, but as soon as I submitted the job failed with the message: "Processing failed on receptor." What is wrong?</p> <p>Solution: see Step 2. Also, check the content of the zip file if you intend to upload one.</p>