Published in final edited form as:

Methods Mol Biol. 2015 ; 1289: 75–87. doi:10.1007/978-1-4939-2486-8_7.

Site Identification by Ligand Competitive Saturation (SILCS) Simulations for Fragment-Based Drug Design

Christina E. Faller¹, E. Prabhu Raman², Alexander D. MacKerell Jr.², and Olgun Guvench^{1,*}

¹Department of Pharmaceutical Sciences, University of New England College of Pharmacy, 716 Stevens Avenue, Portland, ME 04103

²Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, 20 Penn St., HSF II-633, Baltimore, MD 21201

Abstract

Fragment-based drug design (FBDD) involves screening low molecular weight molecules (“fragments”) that correspond to functional groups found in larger drug-like molecules to determine their binding to target proteins or nucleic acids. Based on the principle of thermodynamic additivity, two fragments that bind non-overlapping nearby sites on the target can be combined to yield a new molecule whose binding free energy is the sum of those of the fragments. Experimental FBDD approaches, like NMR and X-ray crystallography, have proven very useful but can be expensive in terms of time, materials, and labor. Accordingly, a variety of computational FBDD approaches have been developed that provide different levels of detail and accuracy.

The Site Identification by Ligand Competitive Saturation (SILCS) method of computational FBDD uses all-atom explicit-solvent molecular dynamics (MD) simulations to identify fragment binding. The target is “soaked” in an aqueous solution with multiple fragments having different identities. The resulting computational competition assay reveals what small molecule types are most likely to bind which regions of the target. From SILCS simulations, 3D probability maps of fragment binding called “FragMaps” can be produced. Based on the probabilities relative to bulk, SILCS FragMaps can be used to determine “Grid Free Energies (GFEs),” which provide per-atom contributions to fragment binding affinities. For essentially no additional computational overhead relative to the production of the FragMaps, GFEs can be used to compute Ligand Grid Free Energies (LGFEs) for arbitrarily complex molecules, and these LGFEs can be used to rank-order the molecules in accordance with binding affinities.

Keywords

Fragment Based Drug Design (FBDD); molecular dynamics (MD); Site Identification by Ligand Competitive Saturation (SILCS); binding free energy; FragMap; Grid Free Energy (GFE); Ligand Grid Free Energy (LGFE)

*Corresponding author: T 207-221-4171; F 207-523-1926; oguvench@une.edu.

Conflict of Interest

OG and ADM are founders of SilcsBio LLC and are currently Manager and Chief Scientific Officer, respectively.

1. General FBDD Methods and the SILCS Approach

Fragment-based drug design (FBDD) seeks to identify low molecular weight molecules (“fragments”) that bind to target proteins or nucleic acids of interest. The identities of these fragments are chosen based on their similarity to functional groups commonly occurring in drug-like molecules. After determining which of these fragments bind, as well as their binding poses, fragments binding to adjacent sites on the target can be linked to create a molecule with a higher binding affinity (1). This approach derives from the principle of thermodynamic additivity, which states that if two components are independent in their contributions to the change in free energy, then the sum of their respective contributions gives the total change in free energy, i.e. $G_{\text{total}} = G_{\text{fragment1}} + G_{\text{fragment2}}$ (2). A variety of experimental techniques, including X-ray crystallography, NMR, surface plasmon resonance, isothermal titration calorimetry, and mass spectrometry, have proven to be very useful in determining binding affinities and binding poses of fragments to target proteins (3–6). However, FBDD can be an expensive endeavor using experimental approaches, as they are associated with high costs in materials and time, especially for high-throughput screening, as well as in labor.

Computational approaches to FBDD aim to minimize the various costs associated with experimental approaches. Many *in silico* methods utilize simplified representations of the target and of the solvent in order to reduce the computational burden by reducing the number of degrees of freedom in the system. Examples of common simplifications include a rigid target model and representing the solvent as a continuum (7–11). The rigid target model approach is often referred to as docking and has difficulty identifying ligands that require even minor changes in target conformation for binding (12–15). More recent work has sought to improve sampling in this regard by using several different rigid target conformations for docking calculations (16). Because of its computational speed, FBDD docking can allow high-throughput screening of large libraries of fragments that approach the theoretical limit of fragment diversity, which is 10^7 unique fragments (17). FBDD docking, in addition to having the capacity to test all possible fragments, benefits from the fact that fragments have few internal degrees of freedom, which greatly simplifies the conformational search problem in docking (18–20). However, development of sufficiently accurate scoring functions for ranking different docked molecules continues to be a challenge (21–24).

The opposite end of the spectrum from rigid target docking is the application of all-atom explicit-solvent molecular dynamics (MD) simulations, in which the solvent is explicitly modeled in atomic detail, and the ligands and target protein or nucleic acid are all fully flexible. In these MD simulations binding free energies can be determined and, in conjunction with the optimized empirical force fields presently available for biomolecules and small molecules (25–33), near-quantitative binding free energy agreement can be reached relative to wet-lab experiments (34–43). Unfortunately, while this level of detail provides accuracy, computational efficiency is lost due to the need to sample ligand, target, and solvent degrees of freedom sufficiently to obtain converged results. Therefore, using this type of thorough MD simulation to do high-throughput analysis of fragment binding is simply not possible for the foreseeable future.

While it is computationally inefficient to do MD simulations on individual small molecules from a large set, a possible advantageous approach is to employ a competitive method that simultaneously screens a simplified set of fragment molecules that represent various functional groups. The SILCS (Site Identification by Ligand Competitive Saturation) method (44) does exactly this by using atomic level of detail MD simulations of a target in an aqueous solution containing selected fragment molecules so as to determine regions of high probability binding for different fragment types.

2. SILCS Methodological Details

SILCS (44) uses nanosecond-length all-atom explicit-solvent MD simulations of the target in an aqueous solution containing a variety of fragments. Explicitly modeling water molecules allows for atomic-level solvation effects to be included. Multiple simulations are run for each system, the trajectories are combined, and 3D probability maps of each fragment type around the target are calculated. The 3D probability maps are then normalized relative to fragment probabilities in bulk solution, thereby incorporating fragment desolvation free energies into the final maps, which are referred to as “FragMaps.” As explicit water is included in the MD simulations, the free energy penalty for desolvation of the target to allow fragments binding is taken into account in the final FragMaps in addition to all other components of binding free energy including target-ligand interactions, target deformation energy, and entropic contributions.

2.1. Fragment selection

Selecting fragments to include in the aqueous solution is an important step in the SILCS methodology: the fragments should be small enough to allow adequate concentrations to facilitate conformational sampling, and should minimally represent hydrogen bond donors, hydrogen bond acceptors, aliphatic groups, and aromatic groups. In the original conception of SILCS, the water portion of the solution contributed both hydrogen bond donors and hydrogen bond acceptors. The other fragments, therefore, needed to include aliphatic and aromatic moieties, meaning two more fragments were required to provide these functional moieties to complement those provided by water. Low molecular weight fragments are particularly desirable due to their high diffusion rates, which lead to improved convergence of simulations. By these standards, the original SILCS fragments were water, benzene, and propane. However, while water is a convenient choice, there are potentially many other hydrogen bond donor and/or hydrogen bond acceptor-containing fragment choices that are more representative of moieties in drug-like molecules. As such, the original SILCS fragment set is now referred to as “Tier 1,” and a new “Tier 2” fragment set consisting of propane, benzene, methanol, formamide, acetaldehyde, methylammonium, and acetate has been validated (45). Notable in Tier 2 SILCS is the use of both neutral and charged donors and acceptors, allowing for regions of the target that bind these to be differentiated.

Low molecular weight fragments have an additional benefit, which stems from the competitive nature of the SILCS *in silico* assay: there is an upper limit to the ligand binding affinity per heavy atom (46), commonly referred to as “ligand efficiency” (47), and this limit is 0.4–0.5 kcal·mol⁻¹ per heavy atom (17). As a consequence, smaller fragments (fewer heavy atoms) translate to weaker binding and higher turnover of fragments on target binding

sites, which improves sampling. Characterization of such weakly-binding fragments through NMR and X-ray crystallography experiments can be challenging, which limits the number of fragment types that are recognized in experimental FBDD efforts. In contrast, SILCS does not have this limitation. Of course, fragments other than those mentioned above can be used to further broaden the range of chemical space represented by FragMaps, but again it is emphasized that larger fragments may slow convergence because of slower diffusion and greater binding affinity.

2.2. Preventing fragment aggregation

The fact that SILCS uses fragment concentrations approaching 1 M in an aqueous solution brings about the problem of aggregation. The aggregation of hydrophobic molecules occurs because they prefer not to be solvated, but to associate with other hydrophobic molecules. The resulting phase separation has the serious consequence that the effective concentration of the hydrophobic molecules is substantially reduced. In Tier 2 SILCS, the presence of ions can lead to ion-pair formation in solution, again reducing the effective fragment concentration. In both instances, the chemical potential of the fragments in bulk solution is reduced, thereby reducing their sampling of the target surface.

SILCS overcomes this barrier to sampling by leveraging the fact that it is a computational method: in SILCS, a repulsive potential between fragments is used to prevent fragment association (44). This repulsive potential – unique to SILCS – only alters fragment-fragment interactions in the system, thereby maintaining an “ideal” solution of fragments in water while leaving all other interactions in the system unperturbed (Figure 1).

2.3. Balancing target flexibility and denaturation

The value of including target flexibility can be visualized by comparing atomic resolution structures of apo- and ligand-bound proteins; in many cases, binding of a ligand is coupled to a conformational change, including in therapeutically-relevant targets such as kinases and proteases (48). Common to SILCS-like approaches (49–53) is the risk of fragment-induced target denaturation, especially in cases of inherently less-stable target proteins, such as those with no disulfide bonds. A range of options has been considered with regard to treatment of target flexibility in SILCS to optimally balance inclusion of target flexibility while minimizing the risk of fragment-induced target denaturation: a fully flexible protein (no positional restraints), weak C α positional restraints, and weak positional restraints on non-hydrogen atoms near the protein core (54). The weak positional restraints allow for relatively large motions of the restrained atoms while limiting the motions enough to avoid denaturation. What is clear is that a lack of restraints (i.e. “full flexibility”) can allow target denaturation regardless of whether the fragments are hydrophobic or hydrophilic.

Should full target flexibility be required, a protocol has been developed to identify denaturing SILCS trajectories for exclusion from subsequent analysis. It employs a combined metric consisting of the root-mean square deviation (RMSD) and the radius of gyration (R_{gyr}). The average RMSD (with the starting structure as reference coordinates) and the average R_{gyr} are computed for each SILCS trajectory. Likewise, they are computed for non-SILCS standard MD control trajectories of the target in the absence of fragments. Each

SILCS trajectory average RMSD, average R_{gyr} pair is then compared to the cluster of these values from the control trajectories, and the SILCS trajectory is excluded from further analysis if its values lie outside the cluster for the control simulation (54). While large RMSD values alone may be good indicators of denaturation, intermediate values can require visualization of snapshots from the trajectory to confirm the RMSD results. In cases of structural rearrangement, like loop rearrangements or sliding of helices relative to one another, intermediate RMSD can be misleading, as the increase in RMSD is not due to denaturation, but due to functionally-relevant conformational change. The use of R_{gyr} has a long history as a reaction coordinate in computational studies of protein folding (55–57) and is a metric of the overall spatial extent of the protein that increases as a protein unfolds. In the context of SILCS, R_{gyr} is especially capable of identifying situations where fragments tunnel into and disrupt the protein hydrophobic core, which may lead to ambiguous intermediate changes in RMSD.

2.4. FragMaps

FragMaps are 3D probability distributions of the fragment atom types in the context of the target. They serve to identify which functionalities (e.g. hydrogen bond donors, hydrogen bond acceptors, aromatic groups, aliphatic groups, etc.) associate most strongly with different areas of the target. FragMaps can be conveniently visualized as isocontour surfaces in the context of the target using freely-available molecular graphics software like VMD (58) (Figure 2) or PMV/ADT (59, 60), or any of the widely-used commercial molecular visualization software packages, since FragMaps can be stored in the same formats as those used for electron densities (61) or docking grids (60).

In the initial SILCS implementation using Tier 1 fragments, to be included in a FragMap a fragment atom must meet a distance criterion relative to the target protein: for example, for inclusion in the hydrogen bond acceptor FragMap, water molecule oxygen atoms must be within 2.5 Å of the protein (44). This criterion was particularly relevant in order to distinguish whether the water was acting in a hydrogen bond donor or acceptor capacity, as a target-bound water molecule may be serving in either role or simultaneously in both roles. With the move to the more varied and specific Tier 2 fragments, this is less of an issue, and inclusion of all fragments during FragMap generation, regardless of distance from the target, can be useful to capture longer-range interactions such as water-mediated interactions of polar molecules with the target. In either of these two approaches, fragment atom locations are binned to create a 3D histogram (FragMap) having $1 \text{ Å} \times 1 \text{ Å} \times 1 \text{ Å}$ voxels.

2.5. Determining convergence

A practical means to evaluate SILCS simulation convergence is to run ten independent simulations, and create two sets of FragMaps by averaging over two sets of five independent FragMaps. Data in the second set are combined and subtracted from the combined data in the first set to generate a difference map. If the simulations are converged, differences should be due to random error and therefore have a tight distribution centered around zero (44). Alternatively, the overlap coefficient of the two maps can be calculated to gauge the extent of convergence (62).

2.6. Grid Free Energies (GFEs)

To obtain quantitative free-energy information, FragMaps are normalized relative to fragment occupancies in bulk solvent and converted to “Grid Free Energies (GFEs)” via inverse-Boltzmann weighting of the normalized FragMap occupancies (63). In order to normalize results, simulations with conditions similar to those of the target + fragments + water system are run with only the SILCS solution (i.e. water + fragments). As with the target-containing simulations, the solution-only simulations are run in the isothermal-isobaric (*NPT*) ensemble to allow for system size relaxation to account for the volume occupied by fragments in the solution. After relaxation, the bulk occupancy for a particular fragment type is computed by simply dividing the number of atoms of that fragment type by the average volume of the system computed from the *NPT* simulations. The GFE for a fragment atom type *f* in a particular voxel centered at *x*, *y*, *z* is then,

$$GFE_{x,y,z}^f = \min \left\{ -RT \log_e \frac{occupancy_{x,y,z}^f}{\langle bulk\ occupancy \rangle}, GFE_{\max} \right\}, \quad (\text{Eqn. 1})$$

where $GFE_{\max}$ can be set as a maximum unfavorable value. A $GFE_{\max}$ of 0 has been used previously (63), which removes any unfavorable contributions arising from voxels having an occupancy lower than bulk.

2.7. Ligand Grid Free Energy scores (LGFEs)

Ligand Grid Free Energy scores (LGFEs) can provide an estimate of binding free energy of a particular target-ligand conformation for an arbitrarily-complex ligand molecule. In order to calculate LGFE scores, ligand molecule atoms are classified into FragMap types based on their chemical similarity to the various fragment atoms used to compute the FragMaps. To this end, an assignment convention has been developed that translates force-field atom types into the FragMap classes. Additionally, certain ligand atoms may be excluded from the LGFE calculation. For example, aromatic hydrogen atoms are implicitly accounted for in the FragMap for the parent benzene carbon atom. The LGFE for a ligand molecule is computed as a sum of the GFEs of its classified atoms:

$$LGFE = \sum_{\text{FragMaps } f} \sum_{\text{atoms } i_f} GFE_{x,y,z}^f(i_f) \quad (\text{Eqn. 2})$$

where the outer summation is over the FragMap types denoted by *f* and the inner summation is over the atoms denoted by *i_f* that are classified into each FragMap type. In addition to single conformations, LGFE scores can be computed for an ensemble of conformations and thermodynamically averaged. Such ensembles of conformations may be obtained, for example, from relatively short (e.g. 1–2 ns) Langevin simulations of the target-ligand complex in the gas phase or with a continuum solvent model. When this is performed it is suggested that the simulations be repeated multiple times with different target conformations.

Figure 3 demonstrates the utility of LGFE scores in structure based drug design. The crystallographic conformations of three ligands that bind to the protein α-thrombin with progressively increasing affinities are shown, along with LGFE scores and experimentally-

determined binding free energies. The overlap of the optimized ligands with FragMaps reflected in the increasingly favorable LGFE scores captures the experimental trend of increasing binding affinities (63, 64).

3. SILCS Workflow

3.1. System construction

1. Determine size of system: a cube with edge lengths x is typical, where x is 16 Å longer than the longest dimension of the target.
2. Generate a box of water molecules of edge length x at the experimental density.
3. Choose a fragment palette, e.g.: Tier 1 SILCS = propane and benzene; Tier 2 SILCS = propane, benzene, methanol, formamide, acetaldehyde, methylammonium, and acetate.
4. Compute fragment placement grid overlapping with water box, such that a ~1M for Tier 1 or ~0.25M for Tier 2 solution in each of the fragments will result.
5. At each placement grid point, randomly select and place a fragment from the palette to generate a fragment solution box.
6. Center the target in fragment solution box from above.
7. Delete water molecules and fragments overlapping with target.

3.2. System simulation

1. Nonbonded conditions: 8 Å real-space cutoff; Particle-mesh Ewald for long-range electrostatics; Switching function between 5 and 8 Å for Lennard-Jones; Virtual particles are added to center of each fragment to serve as interaction sites for inter-fragment repulsion (repulsive potential between the virtual particle pairs is modeled using Lennard-Jones potential combined with above switching function, where Lennard-Jones parameters are $\epsilon = -0.01$ kcal/mol, $R_{\min} = 12.0$ Å).
2. Positional restraints for protein target sidechain flexibility: Harmonic restraints on C α atom positions of the form $k(\mathbf{r})^2$, where $\mathbf{r}$ is the displacement in Å from the crystallographic position and $k = 0.1$ kcal* mol^{-1} *Å $^{-2}$. In addition to full sidechain flexibility, these restraints are sufficiently weak to allow modest backbone flexibility.
3. *Or* positional restraints for loop flexibility: Harmonic restraints on non-hydrogen atoms within x Å of the target center of mass, where x is sufficiently small (e.g. half of the radius of gyration of the target) so as not to include residues near the surface of the target. Restraint functional form and force constants are same as for “sidechain flexibility” above.
4. *Or* full flexibility: No positional restraints. In this case, run control simulations of just the target without fragments for use in post-run determination of denaturing trajectories using average RMSD, average R_{gyr} pair metric.
5. Simulate with isothermal-isobaric (*NPT*) molecular dynamics (MD)

6. Typical simulation length is 50 ns.
7. It is recommended that ten independent simulations be run: Same solution box with ten different random seeds to initiate MD, or, preferably, ten different solution boxes.

3.3. FragMap generation

1. For a given fragment atom type f , bin all atomic positions from all SILCS trajectory snapshots to create a 3D histogram spanning the size of the system and having $1 \text{ \AA} \times 1 \text{ \AA} \times 1 \text{ \AA}$ voxels.
2. If, as recommended, multiple independent simulations were run, generate two FragMaps for each fragment atom type f , using half the simulations for each FragMap, and compute a difference map or overlap coefficient to estimate convergence.
3. n.b.: If “full flexibility” was used, snapshots from denaturing trajectories must be excluded and each snapshot needs to be aligned to a single reference orientation of the target prior to binning of fragment atom positions.

3.4. GFE generation

1. See Eqn. 1.

3.5. LGFE computation

1. Ligand conformation generation: Can be a docking pose, multiple poses from an MD simulation of the target-ligand complex, etc.
2. Ligand atom classification: Each ligand atom is mapped to a fragment atom type f based on chemical similarity.
3. LGFE is computed using Eqn. 2.

Acknowledgements

NIH AI080968, CA107331, and R15GM099022; NSF XSEDE TG-MCB120007; Samuel Waxman Cancer Research Foundation; The University of Maryland Computer-Aided Drug Design Center; and University of New England start-up funds

References

1. Erlanson DA, McDowell RS, O'Brien T. Fragment-based drug discovery. *J Med Chem.* 2004; 47:3463–3482. [PubMed: 15214773]
2. Dill KA. Additivity principles in biochemistry. *J Biol Chem.* 1997; 272:701–704. [PubMed: 8995351]
3. Hennig M, Ruf A, Huber W. Combining biophysical screening and X-ray crystallography for fragment-based drug discovery. *Top Curr Chem.* 2012; 317:115–143. [PubMed: 21837555]
4. Erlanson DA. Introduction to fragment-based drug discovery. *Top Curr Chem.* 2012; 317:1–32. [PubMed: 21695633]
5. Orita M, Warizaya M, Amano Y, Ohno K, Niimi T. Advances in fragment-based drug discovery platforms. *Expert Opin Drug Discov.* 2009; 4:1125–1144. [PubMed: 23480433]

6. Sancineto L, Massari S, Iraci N, Tabarrini O. From small to powerful: the fragments universe and its "chem-appeal". *Curr Med Chem*. 2013; 20:1355–1381. [PubMed: 23410157]
7. Miranker A, Karplus M. Functionality maps of binding-sites - a multiple copy simultaneous search method. *Proteins: Struct, Funct, Genet*. 1991; 11:29–34. [PubMed: 1961699]
8. Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins: Struct, Funct, Genet*. 1999; 37:88–105. [PubMed: 10451553]
9. 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]
10. Clark M, Guarnieri F, Shkurko I, Wiseman J. Grand canonical Monte Carlo simulation of ligand-protein binding. *J Chem Inf Model*. 2006; 46:231–242. [PubMed: 16426059]
11. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, Briggs JM, McCammon JA. Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem*. 2000; 43:2100–2114. [PubMed: 10841789]
12. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. *Med Res Rev*. 2006; 26:531–568. [PubMed: 16758486]
13. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct*. 2003; 32:335–373. [PubMed: 12574069]
14. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: Current status and future challenges. *Proteins: Struct, Funct, Bioinf*. 2006; 65:15–26.
15. Zhong S, Chen X, Zhu X, Dziegielewska B, Bachman KE, Ellenberger T, Ballin JD, Wilson GM, Tomkinson AE, MacKerell AD. Identification and validation of human DNA ligase inhibitors using computer-aided drug design. *J Med Chem*. 2008; 51:4553–4562. [PubMed: 18630893]
16. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol*. 2008; 18:178–184. [PubMed: 18302984]
17. Congreve M, Chessari G, Tisi D, Woodhead AJ. Recent developments in fragment-based drug discovery. *J Med Chem*. 2008; 51:3661–3680. [PubMed: 18457385]
18. Majeux N, Scarsi M, Caflisch A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins: Struct, Funct, Genet*. 2001; 42:256–268. [PubMed: 11119650]
19. Gozalbes R, Carbajo RJ, Pineda-Lucena A. Contributions of computational chemistry and biophysical techniques to fragment-based drug discovery. *Curr Med Chem*. 2010; 17:1769–1794. [PubMed: 20345344]
20. Rabal O, Urbano-Cuadrado M, Oyarzabal J. Computational medicinal chemistry in fragment-based drug discovery: what, how and when. *Future Med Chem*. 2011; 3:95–134. [PubMed: 21428828]
21. Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br J Pharmacol*. 2008; 153(Suppl 1):S7–S26. [PubMed: 18037925]
22. Guvench O, MacKerell AD Jr. Computational evaluation of protein-small molecule binding. *Curr Opin Struct Biol*. 2009; 19:56–61. [PubMed: 19162472]
23. Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys*. 2010; 12:12899–12908. [PubMed: 20730182]
24. Huang SY, Zou X. Advances and challenges in protein-ligand docking. *Int J Mol Sci*. 2010; 11:3016–3034. [PubMed: 21152288]
25. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem*. 2004; 25:1656–1676. [PubMed: 15264259]
26. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc*. 1995; 117:5179–5197.
27. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct, Funct, Bioinf*. 2006; 65:712–725.

28. Wang JM, 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]
29. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc*. 1996; 118:11225–11236.
30. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B*. 2001; 105:6474–6487.
31. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*. 2010; 31:671–690. [PubMed: 19575467]
32. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, Mackerell AD Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput*. 2012; 8:3257–3273. [PubMed: 23341755]
33. Guvench O, MacKerell AD Jr. Comparison of protein force fields for molecular dynamics simulations. *Methods Mol. Biol*. 2008; 443:63–88. [PubMed: 18446282]
34. Boresch S, Tettinger F, Leitgeb M, Karplus M. Absolute binding free energies: A quantitative approach for their calculation. *J Phys Chem B*. 2003; 107:9535–9551.
35. Jayachandran G, Shirts MR, Park S, Pande VS. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J Chem Phys*. 2006; 125:084901. [PubMed: 16965051]
36. Jiao D, Golubkov PA, Darden TA, Ren P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc Natl Acad Sci U S A*. 2008; 105:6290–6295. [PubMed: 18427113]
37. Lee MS, Olson MA. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys J*. 2006; 90:864–877. [PubMed: 16284269]
38. Lee MS, Olson MA. Calculation of absolute ligand binding free energy to a ribosome-targeting protein as a function of solvent model. *J Phys Chem B*. 2008; 112:13411–13417. [PubMed: 18821791]
39. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting absolute ligand binding free energies to a simple model site. *J Mol Biol*. 2007; 371:1118–1134. [PubMed: 17599350]
40. Shirts MR, Mobley DL, Chodera JD, Pande VS. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J Phys Chem B*. 2007; 111:13052–13063. [PubMed: 17949030]
41. Wang J, Deng Y, Roux B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys J*. 2006; 91:2798–2814. [PubMed: 16844742]
42. Woo HJ, Roux B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc Natl Acad Sci U S A*. 2005; 102:6825–6830. [PubMed: 15867154]
43. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct*. 2007; 36:21–42. [PubMed: 17201676]
44. Guvench O, MacKerell AD Jr. Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comp Bio*. 2009; 5:e1000435.
45. 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]
46. Kuntz ID, Chen K, Sharp KA, Kollman PA. The maximal affinity of ligands. *Proc Natl Acad Sci U S A*. 1999; 96:9997–10002. [PubMed: 10468550]
47. Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today*. 2004; 9:430–431. [PubMed: 15109945]

48. Spyrakis F, BidonChanal A, Barril X, Luque FJ. Protein flexibility and ligand recognition: challenges for molecular modeling. *Curr Top Med Chem*. 2011; 11:192–210. [PubMed: 20939788]
49. Seco J, Luque FJ, Barril X. Binding site detection and druggability index from first principles. *J Med Chem*. 2009; 52:2363–2371. [PubMed: 19296650]
50. Yang C-Y, Wang S. Hydrophobic Binding Hot Spots of Bcl-xL Protein-Protein Interfaces by Cosolvent Molecular Dynamics Simulation. *ACS Med Chem Lett*. 2011; 2:280–284. [PubMed: 24900309]
51. Lexa KW, Carlson HA. Full Protein Flexibility Is Essential for Proper Hot-Spot Mapping. *J Am Chem Soc*. 2011; 133:200–202. [PubMed: 21158470]
52. Tan YS, Sledz P, Lang S, Stubbs CJ, Spring DR, Abell C, Best RB. Using ligand-mapping simulations to design a ligand selectively targeting a cryptic surface pocket of polo-like kinase 1. *Angew Chem Int Ed Engl*. 2012; 51:10078–10081. [PubMed: 22961729]
53. 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]
54. Foster TJ, Mackerell AD Jr, Guvench O. Balancing target flexibility and target denaturation in computational fragment-based inhibitor discovery. *J Comput Chem*. 2012; 33:1880–1891. [PubMed: 22641475]
55. Boczko EM, Brooks CL III. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science*. 1995; 269:393–396. [PubMed: 7618103]
56. Sheinerman FB, Brooks CL III. Calculations on folding of segment B1 of streptococcal protein G. *J Mol Biol*. 1998; 278:439–456. [PubMed: 9571063]
57. Shea JE, Brooks CL III. From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu Rev Phys Chem*. 2001; 52:499–535. [PubMed: 11326073]
58. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996; 14:33–38. [PubMed: 8744570]
59. Sanner MF. Python: a programming language for software integration and development. *J Mol Graph Model*. 1999; 17:57–61. [PubMed: 10660911]
60. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009; 30:2785–2791. [PubMed: 19399780]
61. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*. 2011; 67:235–242. [PubMed: 21460441]
62. Bernard D, Coop A, MacKerell AD Jr. Quantitative conformationally sampled pharmacophore for delta opioid ligands: reevaluation of hydrophobic moieties essential for biological activity. *J Med Chem*. 2007; 50:1799–1809. [PubMed: 17367120]
63. Raman EP, Yu W, Guvench O, MacKerell AD Jr. Reproducing Crystal Binding Modes of Ligand Functional Groups Using Site-Identification by Ligand Competitive Saturation (SILCS) Simulations. *J Chem Inf Model*. 2011; 51:877–896. [PubMed: 21456594]
64. Baum B, Muley L, Heine A, Smolinski M, Hangauer D, Klebe G. Think twice: understanding the high potency of bis(phenyl)methane inhibitors of thrombin. *J Mol Biol*. 2009; 391:552–564. [PubMed: 19520086]

Figure 1. Fragment aggregation after 20 ns of SILCS MD simulation. a) No inter-fragment repulsive potential. b) With SILCS inter-fragment repulsive potential. The protein target is displayed as ribbons, water oxygen atoms are in red, and benzene and propane carbon atoms are in blue.

Figure 2.

Tier 1 FragMaps. a) Target protein molecular surface in white, propane FragMap in green mesh, and benzene FragMap in purple mesh. b) Same as (a) but with clipping and depth-cueing to expose additional FragMap density (white boxes) beneath the molecular surface of the apo-target crystal structure; this density corresponds to experimentally-known ligand-binding sites (54).

Figure 3.

Crystallographic complexes of α -thrombin with three ligands of progressively higher affinity, along with benzene and propane FragMaps. The benzene and propane FragMaps are in purple and green, respectively. The ligand grid free energy (LGFE) for each ligand is displayed on the right-bottom side of each panel and the experimentally-measured binding affinity difference is at the interface of each pair of panels. Protein-ligand structures are from PDB IDs (a) 2ZGX, (b) 2ZDA, and (c) 2ZO3.