Published in final edited form as:

Methods Mol Biol. 2012 ; 823: 359–366. doi:10.1007/978-1-60327-216-2_23.

Structure-Based Functional Design of Drugs: From Target to Lead Compound

Amy C. Anderson

Abstract

Proteomic and genomic discoveries have identified vast numbers of new drug targets for investigation. In the quest to discover drugs that modulate the function of these targets, identification of small-molecule drug leads is one of the earliest steps. Structure-based drug design has emerged as a valuable, inexpensive, and rapid computational resource that identifies lead compounds that are complementary to the structure of the target. Leads identified through this process are biologically evaluated and “hit compounds” with affinity and activity are further optimized. This chapter introduces the process of structure-based drug design, including preparation of the ligand database, preparation of the target structure, docking and scoring, and evaluation.

Keywords

Docking; Virtual screening; Structure-based drug design; *In silico* library; Flexibility

1. Introduction

Recently, there has been an explosion of genomic, proteomic, and structural information for new drug targets and consequently, hundreds of opportunities for new drug lead discovery. Structure-based drug design (SBDD) continues to play a critical role in this process. The advantages of SBDD are many: hundreds of thousands of ligands can be described and virtually screened as potential drug leads without the need for initial purchase or synthesis, the speed of the SBDD process is rapid relative to in vitro screening, and the cost of the process is relatively low.

The process of SBDD is iterative and fits nicely within the context of a larger drug discovery program (1–3). In this process, software is used to identify optimal binding modes of small-molecule ligands in the structure of a target (docking); these binding modes are then scored for their noncovalent interactions (see Note 1). The ranked list of ligands is then visually evaluated in the complex with the target and top-scoring molecules are often purchased or synthesized. Of these, some are considered “hits” and exhibit affinity for the target. Often, 50% inhibition constants (IC₅₀) for hits are in the range of 10–100 μM. These

© Springer Science+Business Media, LLC 2012

¹The choice of target structure is often critical to achieving accurate results in docking. Crystal structures are preferred over solution structures or homology models. Specifically, the best targets for docking are crystal structures determined with data at a resolution better than 2.5 Å, R-factors below 30%, low coordinate error (the Luzzati coordinate error should be in the range 0.2–0.3 Å), and good stereochemistry. The structure should also place at least 90% of the backbone ϕ and ψ angles in the most favored regions of the Ramachandran plot. Analyses of structure evaluations are available from the PDB. Solution structures determined with NMR data can also be used as docking targets, but care must be exercised to ensure accuracy of the docking results (16). Solution structures are presented as ensembles of structures, all of which satisfy the experimental data; individual ensemble members vary widely and may not be accurate receptors for docking. This problem may be alleviated by docking against each individual member of the ensemble and combining the scores (see below). Homology models generally represent poor targets for docking since they lack the accuracy necessary to predict the proper noncovalent interactions with ligands.

hits can be optimized toward a higher affinity interaction ($IC_{50} = 10\text{--}100\text{ nM}$) by undergoing additional cycles of SBDD using focused databases comprising analogs of the hit scaffold. The optimized hit is a lead that must then be developed for drug-like properties, such as bioavailability, stability, and efficacy.

There are several docking and scoring programs available. Each program has its own strengths and weaknesses (1–6) as well as its own procedures and nuances. The choice of program depends on priorities placed on requirements for flexibility of the target and ligand, virtual screening of whole molecules or de novo construction of a molecule from docked functional groups, and, lastly, purchase price. To describe the exact procedures for even a few of the available programs would be prohibitively lengthy in this chapter; therefore, a more general approach with overall considerations is presented.

2. Materials

2.1. In Silico Ligands

1. A ligand database can consist of compounds that are commercially available, a private collection of compounds, or a collection of functional groups (see Note 2). One example of a collection of commercially available compounds is the ZINC database ((7) and <http://zinc.docking.org/>).
2. Software for converting the two-dimensional representation of ligands in the database to three-dimensional representations. The programs, CONCORD (Tripos, Inc.) and CORINA, are common examples of software to perform this function.

2.2. Target Structure

The macromolecular structure can often be obtained from the protein database ((8) and <http://www.rcsb.org/>). Structures determined with X-ray diffraction data are most commonly used for drug design, although solution structures determined with NMR methods and homology models can also be effective.

2.3. Docking Software

Several programs are available and each has key features (Table 1). A few programs are available free of charge to academic users (DOCK and Autodock are two examples), and others are associated with a fee.

2.4. Software for Scoring Corrections (Optional)

Most docking software packages have associated scoring functions; however, additional scoring functions to assess the contribution of solvent and the free energy of the target:molecule interaction may increase the predictive accuracy of the docking process.

3. Methods

The process of SBDD begins with preparation of the *in silico* library of ligands and structure of the target. Using docking software, the ligands are then positioned in the target and scored and ranked for noncovalent interactions with the target.

²Ligands may adopt different conformations with different proteins, and therefore ligand flexibility has been incorporated in many docking algorithms (all programs listed in Table 1 allow ligand flexibility). Ligand conformations can either be precalculated and docked individually or may be “incrementally constructed” during the docking process; the choice between these options is dependent on the docking program. Incremental construction docks an anchor fragment first and then builds up the remainder of the ligand by exploring the potential interactions for various placements of the next fragment across a rotatable bond. If ligand flexibility is available as an option, it should be utilized.

3.1. Ligand Preparation

1. Ligands in the database are usually represented as “strings” that describe the two-dimensional connectivity of atoms. These strings are automatically converted to three-dimensional, minimized representations for docking with software available within the docking package or as a stand-alone utility.
2. The ligand library can be initially filtered to select compounds that are more likely to be bioavailable in later stages. Several criteria may apply, including molecular weight, number of rotatable bonds, and number of hydrogen bond donor and acceptor groups (9, 10).
3. Ligands are checked for proper geometry, including reasonable bond distances and angles. The conformation of the ligands can be minimized if necessary to achieve proper geometry.
4. Ligands with stereocenters are examined as independent enantiomers.
5. Ligands are appropriately protonated for the pH of the target solution.

3.2. Target Preparation

1. Hydrogens, usually absent from crystal structures determined with data at resolutions lower than 1 Å, are added to macromolecular structures.
2. Charges are calculated and assigned for individual residues.
3. The general docking site is defined. Ideally, the docking site is a pocket in which small-molecule ligands can interact. The ligand docking site can be the active site of an enzyme or an assembly site with another macromolecule. RNA secondary structural elements can also form good ligand docking sites since they have available functional groups arranged in a specific fashion (see Note 3).
4. The specific docking site is defined within the software using the procedures specific to the program (see Note 4). One approach is to define individual residues within the general docking site, and another approach uses a 3.5–6 Å radius around a preexisting ligand.
5. A decision to keep metals and cofactors that may be bound in the docking site is made at this point. Metals and cofactors may form an integral part of the binding interaction with the ligand and, if so, are considered part of the docking site. However, if displacement of the metal or cofactor is desired, they should be removed in order to make the functional groups to which they bind available to the ligand.
6. A decision to keep or remove ordered water molecules that exist in the docking site is made at this point. The water molecules can remain in the docking site if they are

³Ligands can induce targets, especially proteins, to undergo significant conformational changes. Incorporating protein flexibility can be critical to achieving accurate docking results (16–19). Some docking programs, such as Flo (20), SLIDE (21), and GOLD (22), directly incorporate protein flexibility and permit the user to identify key residues that are allowed to move during the docking process. Another approach to simulate protein flexibility is to create an ensemble of target structures that represent potential conformations (16, 18, 19, 23). The ensemble can be generated using molecular dynamics simulations, amino acid rotamer libraries, or multiple experimental structures. Ligands can either be docked to individual members of the ensemble and the scores combined or ligands can be docked to a single unified version that represents the individual ensemble members.

⁴Evaluation of docking programs is often based on “enrichment” or the identification of known hits from a large database containing decoys. High enrichment values are valuable in selecting initial hits from a diverse database. Nevertheless, accurately ranking a set of subtly different analogs with a conserved scaffold is important during lead optimization in order to properly guide new synthetic chemistry efforts. Advances in docking software progress toward the goal of accurate ranking, but the goal remains elusive. In order to judge whether novel compounds are properly ranked, the docking scores of several compounds with known affinity should be compared.

critical to ligand binding; they may be removed if the ligand is designed to replace them.

7. If the docking program allows target flexibility in order to accommodate conformational changes induced by the ligand, the number and identity of flexible residues and the degree of flexibility are defined.

3.3. Docking

1. There are individual procedures associated with different docking software packages. For instance, the format of the ligands in the ligand database may change, the ligand may be docked in entirety or in fragments, or the algorithms for ligand placement may differ. However, the general principles for docking software packages are similar: compounds are first positioned in the target site using an algorithm specific to the software and then evaluated for noncovalent interactions using a scoring algorithm available in the package. Most packages use an automatic format to place and score individual ligands from the database.
2. If the orientation of the ligands in the site is known a priori, this information can be used during and after the docking process. In some programs that allow an anchor fragment, ligand placement can be guided during the docking process. After the docking process, knowledge of binding orientation can be an important filter for evaluating results. As an example, once the orientation of a hit compound has been established by a crystal structure with the target, analogs of that lead can be assumed to dock in a similar fashion.

3.4. Optional Scoring Corrections

Docking scores are inherently approximations to the true binding constant, based primarily on the noncovalent interactions between the ligand and target (11). Postdocking software that attempts to narrow the difference between the docking score and true binding constant can be useful. Three postdocking corrections are presented here: estimating the contribution of solvent, consensus scoring, and calculating a free energy of binding.

1. Solvent plays an important role in ligand binding both in forming specific interactions with ligands and in affecting the dielectric constant and therefore the strength of electrostatic interactions. Some docking algorithms account for solvent effects, others do not and in this case, increased accuracy can be achieved by including a solvation correction to the score (12, 13). In order of increasing accuracy, (a) docking and scoring operations can be performed without modeling the effect of solvent, (b) simple approaches estimate a dielectric constant (either fixed or distance dependent), and (c) explicit solvation models are employed.
2. Consensus scoring, in which the top hits from the docking exercise are rescored with a variety of different scoring algorithms, has proven valuable in achieving greater accuracy (14). Hits appearing at the top of multiple lists are then selected for further investigation.
3. Free energy perturbation (FEP) calculations provide a rigorous measure of the changes in free energy between the unbound and bound complexes in solvent (15).

3.5. Interpretation of Results

1. The docking and scoring software provides a ranked list of ligands (see Note 4). Using computer graphics, these are visually evaluated in complex with the target to assess goodness of fit, formation of key hydrogen bonds or hydrophobic interactions, surface complementarity, and stability of the bound conformation of

the ligand compared to the free conformation. Several top-scoring ligands are then purchased or synthesized and evaluated in the laboratory using in vitro assays with the target.

2. Compounds that succeed as hits in the laboratory assays are then subjected to further rounds of *in silico* screening using more focused libraries of potential analogs. The structure of the complex comprising the target and hit molecules is often experimentally determined in order to validate the binding mode.

Acknowledgments

The author thanks Dr. Erin Bolstad for her critical review and comments. This work was supported by NIGMS (GM 067542).

References

1. Anderson AC. The process of structure-based drug design. *Chem Biol.* 2003; 10:787–97. [PubMed: 14522049]
2. McInnes C. Virtual screening strategies in drug discovery. *Curr Opin Chem Biol.* 2007; 11:494–502. [PubMed: 17936059]
3. 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):S7–26. [PubMed: 18037925]
4. Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP. Comparison of automated docking programs as virtual screening tools. *J Med Chem.* 2005; 48:962–76. [PubMed: 15715466]
5. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, et al. A critical assessment of docking programs and scoring functions. *J Med Chem.* 2006; 49:5912–31. [PubMed: 17004707]
6. Wang R, Lu Y, Wang S. Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem.* 2003; 46:2287–303. [PubMed: 12773034]
7. Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model.* 2005; 45:177–82. [PubMed: 15667143]
8. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28:235–42. [PubMed: 10592235]
9. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 2001; 46:3–26. [PubMed: 11259830]
10. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem.* 2002; 45:2615–23. [PubMed: 12036371]
11. 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; 12:12899–908. [PubMed: 20730182]
12. Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. *Proteins.* 1999; 34:4–16. [PubMed: 10336382]
13. Liu HY, Kuntz ID, Zou X. Pairwise GB/SA scoring function for structure-based drug design. *J Phys Chem B.* 2004; 108:5453–5462.
14. Hattotuwa CK, Davies MN, Flower DR. Receptor-ligand binding sites and virtual screening. *Curr Med Chem.* 2006; 13:1283–304. [PubMed: 16712470]
15. Plouffe ML, Jorgensen WL. Analysis of Binding Affinities for Celecoxib Analogues with COX-1 and COX-2 from Combined Docking and Monte Carlo Simulations and Insight into the COX-2/COX-1 Selectivity. *J Am Chem Soc.* 2000; 122:9455–9466.
16. Bolstad ES, Anderson AC. In pursuit of virtual lead optimization: the role of the receptor structure and ensembles in accurate docking. *Proteins.* 2008; 73:566–80. [PubMed: 18473360]

17. B-Rao C, Subramanian J, Sharma SD. Managing protein flexibility in docking and its applications. *Drug Discov Today*. 2009; 14:394–400. [PubMed: 19185058]
18. Paulsen JL, Anderson AC. Scoring ensembles of docked protein:ligand interactions for virtual lead optimization. *J Chem Inf Model*. 2009; 49:2813–9. [PubMed: 19950979]
19. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol*. 2008; 18:178–84. [PubMed: 18302984]
20. McMartin C, Bohacek RS. QXP: powerful, rapid computer algorithms for structure-based drug design. *J Comput Aided Mol Des*. 1997; 11:333–44. [PubMed: 9334900]
21. Schnecke V, Swanson CA, Getzoff ED, Tainer JA, Kuhn LA. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins*. 1998; 33:74–87. [PubMed: 9741846]
22. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*. 1997; 267:727–48. [PubMed: 9126849]
23. Knegtel RM, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. *J Mol Biol*. 1997; 266:424–40. [PubMed: 9047373]
24. Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, Brooijmans N, et al. Development and validation of a modular, extensible docking program: DOCK 5. *J Comput Aided Mol Des*. 2006; 20:601–19. [PubMed: 17149653]
25. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009; 30:2785–91. [PubMed: 19399780]
26. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 2004; 47:1739–49. [PubMed: 15027865]
27. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*. 1996; 261:470–89. [PubMed: 8780787]
28. Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem*. 2003; 46:499–511. [PubMed: 12570372]
29. Boehm, H-J. *Journal of Computer-Aided Molecular Design*. Springer; Netherlands: 1992. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. 6; p. 61-78.
30. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem*. 1985; 28:849–57. [PubMed: 3892003]

Table 1

Ten common docking programs and their characteristics

Program	Description	References
DOCK 6	Docks small molecules, includes solvent effects, uses incremental construction for ligands. Available to academic users without charge	(24)
Autodock	Uses an interaction grid to account for receptor conformations and simulated annealing to account for ligand conformations. Available free of charge	(25)
Glide	Performs complete conformational, orientational, and positional search for ligand	(26)
GOLD	Uses genetic algorithms. Allows partial protein flexibility	(22)
FlexX	Uses incremental construction for ligands	(27)
Flo	Allows protein flexibility and ligand flexibility	(20)
Surflex	Uses incremental construction with fragment assembly	(28)
SLIDE	Allows protein side chain and ligand flexibility	(21)
LUDI	Calculates interaction energy for small-molecule fragments	(29)
GRID	Calculates interaction energy of functional groups	(30)