What is Docking?

According to Halperin et al., docking is a term used for computational schemes that attempt to find the "best" matching between two molecules: a receptor and a ligand. The molecular docking problem is also defined as follows: Given the atomic coordinates of two molecules, predict their "correct" bound association.[1] The first widely used docking program was Kuntz's DOCK,[2] which described the binding site by intersecting spheres. The three components to docking are the representation of the system, the conformational space search, and the raking of potential solutions. Solving the docking problem involves an efficient search procedure and a good scoring function, and ideally, the best matching algorithms and scoring schemes should be combined.[1]

The basic description of the protein or ligand surface is the atomic representation of exposed residues, and the surface often is represented by its geometric features. For docking, points describing the molecular surfaces are needed to calculate a shape function, and the complexity of the algorithm depends on the number of points. A surface normal or *critical point*[1] associated with each point is also computed, and the docking strategy reduces to matching only pairs of critical points with the additional geometric information of their surface normals. In order to compute a candidate rigid transformation, it is necessary to detect a pair of critical points in both molecules that share the same internal distance, and if superimposed, have opposing surface normals.

Docking is computationally difficult because there are many ways of putting two molecules together. In a docking problem, the search for solutions can be addressed by either a scan of the entire solution space in a predefined systematic manner, or by a scan of only part of the solution space in a partially random and partially criteria-guided manner.[1] The search stage of molecular docking of ligands to proteins can be divided into two independent procedures, depending on whether the binding site is known,[3] although in all methods, the search part creates a population of solutions, with each assessed by some energy function.

A search algorithm may produce an immense number solutions,[4] so a scoring function must be used to discriminate between "correct" native solutions from the crystal complex and others within a reasonable computation time.[1] The solution to scoring appears to be a two-stage ranking in which traditional scoring is used to rapidly scan possible solutions and obtain initial "good" candidates, followed by more advanced methods to further discriminate the limited conformations.[5]

Novel Docking Methods

Molecular docking programs screen chemical databases for novel ligands that fit protein binding sites. When one compound fits the site well, close analogs typically do the same, so many of the compounds that are found in such screens resemble one another, thus reducing the variety and novelty of the compounds suggested. To increase the diversity of docking hit lists, Su et al. grouped the Available Chemicals Directory into families of related structures. All members of every family were docked and scored, but only the best scoring molecule of a high-ranking family was allowed in the hit list.[6] Compared with molecule-by-molecule docking, this novel method resulted in increased hit list diversity and more families of known ligands found.

Chen and Zhi introduced a ligand-protein inverse-docking approach for finding potential protein targets of a small molecule by the computer-automated docking search of a protein cavity database.[7] In a standard ligand-protein docking study, potential ligands are typically selected from the lowest energy docked structures. However, in an inverse-docking procedure potential protein targets are selected on the basis of an energy threshold, and docked structures with ligand-protein interaction energies lower than that threshold are considered putative targets.[7]

Docking Applications to Diabetes: Enzyme Inhibition

Hallmarks of type 2 diabetes and obesity include resistance to the hormones insulin and leptin. Drugs that can prevent this resistance should be effective in treating type 2 diabetes. Protein tyrosine phosphatase 1B (PTP1B) is thought to function as a negative regulator of insulin and leptin signal transduction. Defective or inappropriate regulation of PTPase activity leads to aberrant tyrosine phosphorylation, which contributes to the development of many human diseases.[8] Therefore, a challenge for treatment of diabetes will be to develop small-molecule inhibitors of this phosphatase.

The DOCK program is designed to identify novel compounds complementary to the ligand binding site of an enzyme or receptor of known 3D structure.[9] Sarmiento et al. used the DOCK algorithm to screen about 150,000 compounds in the Available Chemicals Directory for compounds displaying good geometric and electrostatic complementarity to the PTP1B active site. Using DOCK, it was possible to generate selective inhibitors targeted primarily to the PTP1B catalytic site and its immediate surroundings. The potent inhibition exhibited by the compounds identified illustrates the power of the DOCK approach for the discovery of a diverse array of low molecular weight compounds which have not been shown previously to be PTPase inhibitors.[10]

High-throughput screening (HTS) of compound libraries is used to discover novel leads for drug development. When a structure is available for the target, computer-based screening using molecular docking may also be considered. Doman et al. used HTS and molecular docking together to discover novel inhibitors for the PTP1B.[11] They screened a corporate library of approximately 400,000 compounds using high-throughput experimental techniques for compounds that inhibited PTP1B. Concurrently, molecular docking was used to screen approximately 235,000 commercially available compounds against the X-ray crystallographic structure of PTP1B, and high-scoring molecules were tested as inhibitors of the enzyme. It was found that structure-based docking enriched the hit rate by 1700-fold over random screening.[11] More inhibitors were discovered by docking than by HTS, even though 1000-fold more compounds were tested by HTS. However, the diversity of both hit lists and their dissimilarity to each other suggest that docking and HTS may be complementary techniques for future discovery.[11]

Problems and Future Considerations

The simpler problem in docking is referred to as “bound” docking, in which a “bound” structure is extracted from a structure of more than one molecule.[1] The goal is the more difficult predictive docking, or “unbound docking.” The unbound problem relates to computational schemes that attempt to reconstruct a complex using the unbound

structures of the receptor and the ligand.[1] Predictive docking is far more complex than bound docking, with the additional complexity derived from conformational changes that take place between the bound and unbound structures. No efficient method for reliable discrimination between correct solutions and false positives generated by predictive docking algorithms is currently available.[12] Therefore, a lack of a reliable method for quickly locating correct solutions, in particular if the binding site is unknown, is the major obstacle in using predictive docking for practical applications.[1]

The improvement of search speed is also a key in improving docking quality, which allows for the introduction of more sophisticated docking algorithms such as the search for the optimum binding mode in a cavity and more accurate modeling of protein flexibility.[7] Also, further development of Chen and Zhi's inverse-docking approach may bring interesting applications such as the determination of unknown and secondary therapeutic targets of drugs, drug leads, natural products, and synthetic chemicals and the identification of protein targets related to the side effects and toxicity of these molecules.

Current docking methods need improved scoring procedures to discriminate against false-positive predictions,[13] although scoring by empirical free energy functions and using refinement by flexible docking has improved the ranking. However, according to Vajda et al., while substantial progress has been made during the last few years, docking procedures are still too slow for application to large sets of proteins and their capacity to identify near-native structures among a large number of docked complexes is still low, so better docking tools will be needed for future analyses.

References

1.) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins.* **2002**, 47:409-443.

2.) Kuntz, I.; Blaney, J.; Oatley, S.; Langridge, R.; Ferrin, T. *J Mol Biol.* **1982,** 161:269-288.

3.) Balbes, L.M.; Mascarella, S.W.; Boyd, D.B. *Reviews in Computational Chemistry.* **1994**, p 337-378.

4.) Palma, P.N.; Krippahl, L.; Wampler, J.E.; Moura, J.J.G. *Proteins.* **2000,** 39:178-194.

5.) Hoffmann, D.; Kramer, B.; Washio, T.; Steinmetzer, T.; Rarey, M.; Lengauer, T. *J Med Chem.* **1999,** 42:4422-4433.

6.) Su, A.I.; Lorber, D.M.; Weston, G.S.; Baase, W.A.; Matthews, B.W.; Brian, K.S. *Proteins*. **2001,** 42:279-293.

7.) Chen, Y.Z.; Zhi, D.G. *Proteins.* **2001,** 43:217-226.

8.) Shen, K.; Keng, Y.-F.; Wu, L.; Guo, X.-L.; Lawrence, D.S. *J. Biol. Chem.* **2001,** 276: 47311-47319.

9.) Kuntz, I.D. *Science*. **1992,** 257, 1078-1082.

10.) Sarmiento, M.; Wu, L.; Keng, Y.-F.; Song, L.; Luo, Z.; Huang, Z.; Wu, G.-Z.; Yuan, A.K.; Zhang, Z.-Y. *J. Med. Chem.* **2000,** 43, 146-155.

11.) Doman, T.N.; McGovern, S.L.; Witherbee, B.J.; Kasten, T.P.; Kurumbail, R.; Stallings, W.C.; Connolly, D.T.; Shoichet, B.K. *J. Med. Chem.* **2002,** 45, 2213-2221.
12.) Norel, R.; Petrey, D.; Wolfson, H.; Nussinov, R. *Proteins.* **1999,** 35:403-419.
13.) Vajda, S.; Vakser, I.A.; Sternberg, M.J.E.; Janin, J. *Proteins.* **2002,** 47:444-446.