Molecular Docking and Drug Design

J.J. Irwin, D.M. Lorber, S.L. McGovern, B.Wei and B.K. Shoichet
Northwestern University Medical School, 303 E Chicago Ave, Chicago IL 60611 USA

ABSTRACT

Molecular docking screens large databases of small molecules by orienting and scoring them in the binding site of a protein. Top-ranked molecules may be tested for binding affinity *in vitro*, and may become lead compounds, the starting point for drug development and optimization. General background in molecular docking and an overview of some recent developments in this field are given in this section. Examples of where docking has been applied to experimental systems will be discussed.

Keywords: Drug discovery, drug design, structure-based drug design, pharmaceutical discovery.

1 OVERVIEW OF MOLECULAR DOCKING

Molecular docking screens large databases of small molecules by orienting and scoring them in the binding site of a protein. Top-ranked molecules may be tested for binding affinity *in vitro*, and may become lead compounds, the starting point for drug development and optimization. Figure 1 is a surface representation of the active site of AmpC beta lactamse, a bacterial resistance enzyme and a target for development of antibiotics and molecular docking.

Very recently, there has been tremendous growth in the number of protein structures determined to atomic resolution [1]. This trend can only be expected to accelerate under the impact of structural genomics [2]. The big problem in structural biology is no longer what does the protein look like, but what does the structure mean for functional analysis, and often, ligand design. Every week, dozens of new proteins of therapeutic interest appear in the Protein Databank. Many of these are amenable to molecular docking studies for lead discovery.

There are many problems with molecular docking as it is currently implemented. The conformational and orientational sampling of the putative drug in the protein site is crude. The model of the protein site itself is even cruder: often, the receptor is kept completely rigid. The scoring function used to rank good ligands above poor compounds is problematic too: the net energy of binding is a small difference of large values with large uncertainties, and the calculation of desolvating the ligand when it binds to the protein is complex to calculate.

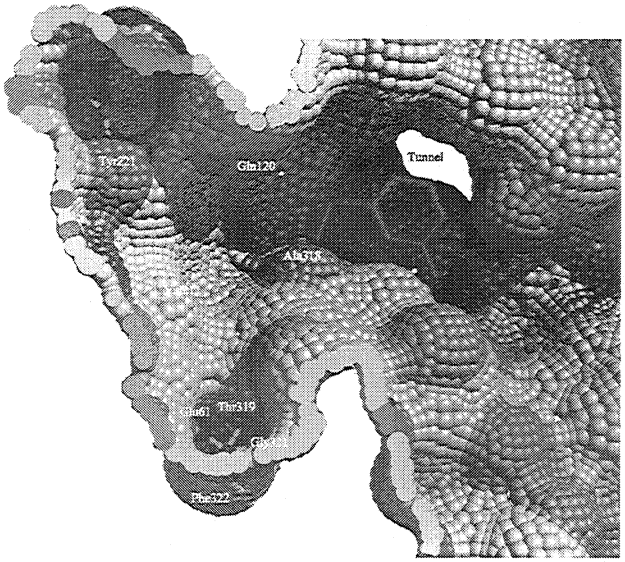
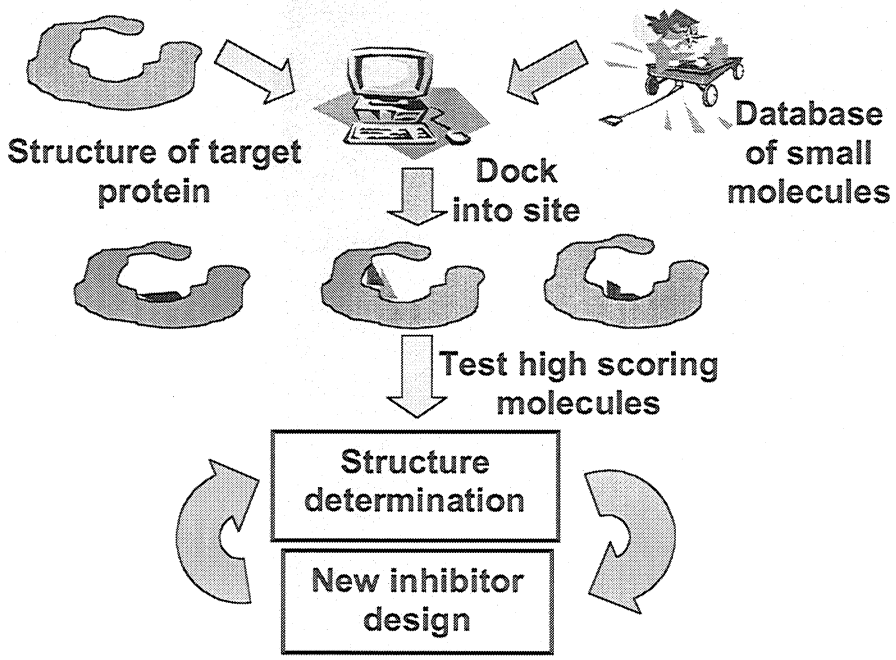


Figure 1 - The active site of AmpC beta lactamase, a target for molecular docking.

Despite these considerable problems, molecular docking has been used successfully numerous times to identify promising lead compounds that may be used to begin a process of lead optimization towards a drug candidate. Why does docking work even though there are so many problems with the calculation? Docking is a screening method: false positives are tolerated, and a very low ratio of hits is considered acceptable.

Figure 2 is a schematic outlining the docking procedure in this lab. A database of small molecules is oriented in the site, conformationally sampled to allow for molecular flexibility, and scored in the computer model of a protein binding site. The top scoring ligands from this procedure may then be purchased or synthesized, and subsequently tested *in vitro*. True binders may be chemically modified to improve affinity. Knowledge of true binders may be fed back into the calculation to more accurately parameterize the model.

Computational Nanoscience and Nanotechnology 2002, (www.cr.org), ISBN 0-9708275-6-3.



2 RECENT INNOVATIONS IN MOLECULAR DOCKING

Several recent innovations in the docking method from this laboratory have been successfully used to identify true binders. The **ensemble docking method** of Lorber and Shoichet [3] is an efficient technique for sampling the internal degrees of freedom of the small molecule. It does this by pre-calculating molecular flexibility, partitioning the molecule into a rigid fragment and flexible side chains. The **hierarchy method** of Lorber and Shoichet [4] builds upon the ensemble method by efficiently re-combining side chains and eliminating redundant or unnecessary calculations, resulting in orders of magnitude of improvement of speed and of conformational sampling. Combined with an array of computers, this enables huge databases to be screened rapidly, increasing the probability of identifying promising lead compounds. The **full ligand desolvation method** of Shoichet et.al [5] adjusts for the cost of bringing a drug molecule out of solution to bind to the protein, resulting in a truer estimate of the binding affinity, and a better chance of giving top ranking to true binders.

3 CONCLUSION

Molecular docking continues to holds great promise in the field of computer based drug design. Considering the simplicity of the earliest approaches the level of success they achieved is remarkable. In the next decade these algorithms will be significantly improved through efficient incorporation of both target site and inhibitor mobility. In addition, their accuracy will be improved through better desolvation accounting.

REFERENCES

1. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne: *Nucleic Acids Research*, **28** pp. 235-242 (2000)

2. H.M. Berman, T.N. Bhat, P.E. Bourne, Z. Feng, G. Gilliland, H. Weissig, J. Westbrook (2000): *Nature Structural Biology* 7 (11), pp. 957-959.

3. D.M. Lorber and B.K. Shoichet, Protein Science (1998) 7:938-950

4. D.M. Lorber and B.K. Shoichet, Manuscript in preparation.

5. B.K. Shoichet, A.R. Leach, I.D. Kuntz, *Proteins* (1999) 34:4-16 and references cited therein