DEVELOPMENT OF AN EFFICIENT AUTOMATED DOCKING METHOD

Miho YAMADA and Akiko ITAI*

Faculty of Pharmaceutical Sciences, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

An efficient automated method of searching for stable docking structures of protein-ligand complex has been developed. The method outputs several promising docking models in which protein and ligand interact favorably, covering all possible binding modes and ligand conformations. Our search method is excellent in terms of not only ease of use and speed of calculation, but also reliability and reproducibility of the docking results. It should become an indispensable tool for investigating biochemical reaction processes and designing new drug structures based on receptor structure.

KEYWORDS docking simulation; protein-ligand interaction; ligand conformation; computer-aided drug design

Docking simulation is one of the most important techniques for rational drug design in a case where the three-dimensional (3D) structure of the target macromolecule (receptor) is known. For the purpose of understanding intermolecular interactions of receptor and known bioactive compounds and designing new drug structures, it would be very helpful to predict the binding modes to the receptor and the ligand conformations in the complex, without performing experiments. The most important objective of docking simulation is to find the global-minimum energy structure and stability of the issued complexes of receptor and various ligand molecules.

The problems are how to cover all possible binding modes (relative positions and orientations of two molecules) and ligand conformations and how to choose the energetically most favorable one from among them. Needless to say, molecular mechanics calculation or even molecular dynamics calculation cannot generate the global-minimum energy structure without an appropriate starting structure. In the usual methods for docking simulation, various probably stable docking structures are located interactively using computer graphic displays, and then are optimized by energy minimization methods. But the global-minimum energy structure is difficult to find by such a manual method, unless either or both the binding mode and ligand conformation can be assumed properly based on experiments with analogous compounds. Since the problems of binding mode and ligand conformation are interlinked, an enormous number of docking structures resulting from the combination of them should be examined in order to find the global-minimum energy structure.

An efficient automated search method has long been desired for obtaining the correct docking models for any ligand or ligand candidate. The method should cover all possible binding modes and ligand conformations without any presumptions. It should be sufficiently rapid to search a vast number of possibilities within a practical time. Above all, the method should reliably generate the global-minimum energy structure with the highest ranking. Several research groups have approached this difficult problem,[1] but no one has yet proposed a satisfactory method. Some of the methods cannot be applied to highly flexible molecules, and some of them have a problem in accuracy or ranking of docking models. Recently, Leach and Kuntz have reported a search method in which an appropriate rigid group in the ligand is used as the anchor.[2] In their method, possible positions of the anchor group are searched for considering the shape complementarity and hydrogen bond (H-bond) formation, and then conformations in the flexible part are searched for each anchor position. Choice of an appropriate anchor is the key to success in their method.

Here we describe the development of an excellent search method which satisfies all the requirements described above. The computer program for realizing the method was named ADAM.

METHODS AND ALGORITHMS

Our docking procedure is outlined in Chart.1.

Prior to the docking, a 3D grid is generated inside the ligand-binding region of the protein.[3] At each grid point, van der Waals and electrostatic potentials are tabulated in order to estimate intermolecular interaction energy rapidly. Moreover, H-bonding flags, which show the possibility of H-bond formation with protein functional groups and H-bonding character (acceptor or donor), are set at each grid point. H-bonding dummy atoms are located at the center of grid points with common H-bonding flags inside the binding region.

In the first step of docking, possible H-bonding schemes are searched by covering all combination sets of correspondences between H-bonding dummy atoms in the protein and heteroatoms in the ligand, rather than covering all possible binding modes by systematic rotation and translation of the ligand. When p H-bonds are formed from m dummy atoms in the protein and n heteroatoms in the ligand, the number of combination sets $N(p)$ is given by the following equation. Each combination set consists of p pairs of H-bonding atoms in both molecules, and the combination sets in which the H-bonding characters in all pairs are not complementary are excluded.

$$N(p) = {}_mP_p \times {}_nC_p$$

As regards the ligand conformation, all rotatable bonds are rotated systematically. In each combination set, possible conformations in the H-bonding part, which involves H-bonding functional groups, are selected together with possible H-bonding schemes. In order to judge the possibility of forming H-bonds between the corresponding atoms in two molecules placed in different coordinate systems, distances among dummy atoms are compared with those among corresponding heteroatoms in each conformation for each combination set. If deviations between all corresponding distances and ligand intramolecular energy of the structure obtained by minimizing the root-mean-squares (rms) deviations of the distances are within given criteria, the H-bonding scheme is regarded as possible in that ligand conformation.

Then the obtained ligand conformers with possible H-bonding schemes are put into the grid space by a least-squares fitting[4] of positions of heteroatoms with those of the corresponding dummy atoms. The docked structures are pruned by total energy after energy minimization by the Simplex method,[5] optimizing position and orientation as well as bond rotation in the H-bonding part of the ligand.

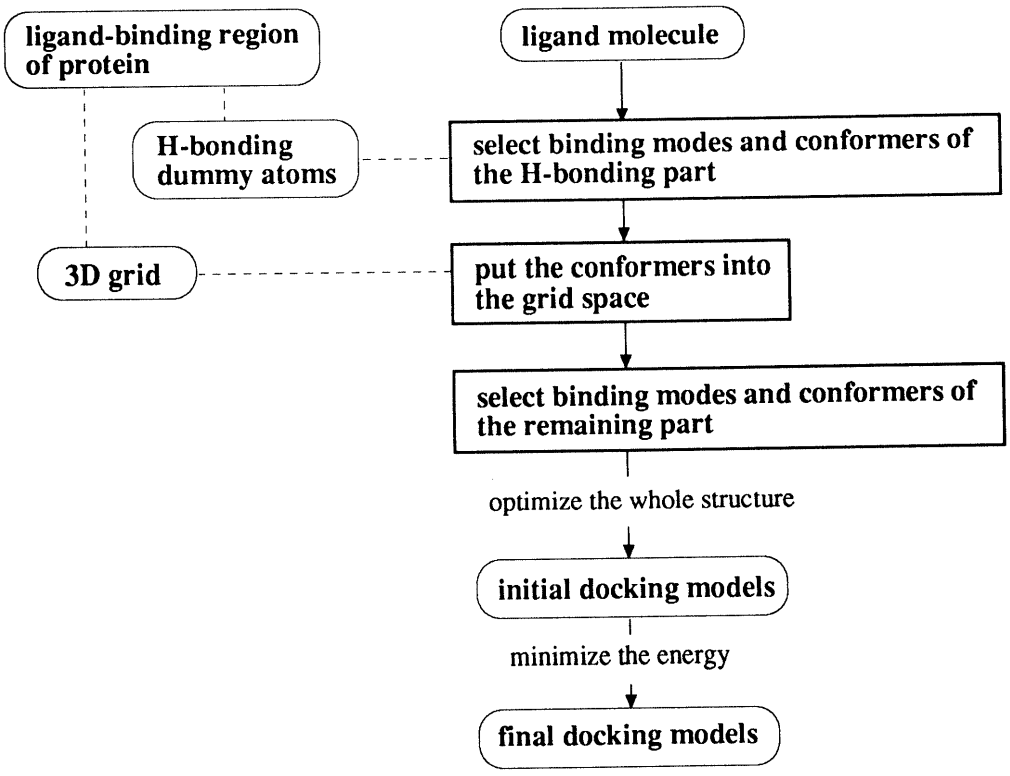

Chart 1. Flow Diagram of Our Docking Procedure

Conformations in the remaining part are generated for each likely structure of the H-bonding part. Low-energy models are subject to energy minimization by the Simplex method, taking into account the degrees of freedom for rotation, translation and conformation of the whole ligand structure, and are further pruned. From several to several dozen stable docking structures, which we call "initial docking models", are output here. The number of output models varies depending on the energetic criteria for pruning.

Finally, the "final docking models" are obtained by energy minimization of the initial docking models, using the AMBER program.[6] In this energy minimization, the positions of the protein atoms and the water molecules near the ligand-binding region are optimized together with those of the ligand atoms. The energy-minimized models are ranked by total energy.

The whole process of obtaining initial docking models is conducted continuously and automatically for any molecule by our program ADAM. However, in the case of a ligand molecule with many rotatable bonds or many H-bonding heteroatoms generating a vast number of possibilities, a straightforward search of all the possibilities is not a good policy. To finish the docking of such molecules in a short computational time, we have developed a new method, the Pre-Pruning (PP) procedure. The method is based on the assumption that H-bonding schemes impossible in a partial structure are also impossible in the whole structure. The partial structure does not need to be conformationally rigid. By indicating a partial structure in the ligand at the beginning of the docking, impossible H-bonding schemes and unused H-bonding atoms are preliminarily searched for the structure. The docking of the whole molecule proceeds automatically, dispensing with unnecessary searches, based on the information obtained by the preliminary search step.

The application of the program ADAM to an enzyme system and the evaluation of the method are reported in the following paper.[7]

Although our method is most efficient for systems where several H-bonds play important roles in specific molecular recognition, it is also useful in systems where only one or two H-bonds are working together with hydrophobic interaction.

The program has no limitation in manageable numbers of combination sets and rotated bonds, and any large molecule can be docked continuously. Nevertheless, the PP procedure is very efficient for reducing the required computational time for rather large molecules. Conformational rigidity is not required for the partial structure used for the PP procedure, unlike the anchor groups used in the method of Leach and Kuntz. Choice of the partial structure does not affect the docking results of the whole molecule, though it is favorable if the structure includes more than two H-bonding groups.

The automated search method is greatly superior to manual methods on graphic displays in terms of reliability, reproducibility, objectivity and coverage of possibilities, in addition to reducing time and labor requirements. This method should become an indispensable tool for rational drug design and investigation of the biochemical reaction process. The full details of the method will be published elsewhere.

REFERENCES

1) I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge and T. E. Ferrin, *J. Mol. Biol.*, **161**, 269 (1982); D. S. Goodsell and A. J. Olson, *Proteins*, **8**, 195 (1990); F. Jiang and S. -H. Kim, *J. Mol. Biol.*, **219**, 79 (1991); A. S. Smellie, G. M. Crippen and W. G. Richards, *J. Chem. Inf. Comput. Sci.*, **31**, 386 (1991); D. J. Bacon and J. Moult, *J. Mol. Biol.*, **225**, 849 (1992).
2) A. R. Leach and I. D. Kuntz, *J. Comp. Chem.*, **13**, 730 (1992).
3) N. Tomioka, A. Itai and Y. Iitaka, *J. Comput. -Aided Mol. Design*, **1**, 197 (1987).
4) W. Kabsch, *Acta Crystallogr., Sect. A*, **32**, 922 (1976); W. Kabsch, *ibid.*, **34**, 827 (1978).
5) J. A. Nelder and R. Mead, *Computer J.*, **7**, 308 (1965).
6) S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr. and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984); S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *J. Comp. Chem.*, **7**, 230 (1986).
7) M. Yamada and A. Itai, *Chem. Pharm. Bull.*, following paper in this issue.

(Received April 13, 1993)