REVIEW ARTICLE

Accepted 14 April 1998
Received 10 March 1997
Indian J. Pharm Sci., 1998, 60(6) pp. 333-337

Importance of Docking studies in Drug Design

K.G. BOTHARA*, A. U. PATIL A. SAXENA
All India Shri Shivaji Memorial Society's College of Pharmacy
Kennedy Road, Near RTO, Pune-411 001

Docking Studies have become nearly indispensable for study of macromolecular structures and interactions. Mechanical model construction requires heroic patience and endurance to complete a structure which may contain several thousand atoms while computer graphics can build and display in seconds. Macromolecular modeling by Docking studies provides most detailed possible view of drug-receptor interaction and has created a new rational approach to drug design where the structure of drug is designed based on its fit to three dimensional structures of receptor site, rather than by analogy to other active structures or random leads.

LIGAND-macromolecule "Docking" is typically done interactively with molecular surface displays (e.g. "extra radius" surface) using to guide the fit, based on hydrophobic or electrostatic potential color coding. The binding site is usually treated as completely rigid initially, while the conformation of ligand is adjusted interactively[1]. High energy contacts can be highlighted with color-coded vectors. Docking studies may help to increase ligand specificity and if the drug produces undesirable side effects by binding to another macromolecule, it may be possible to diminish affinity for that competing site and thus to achieve a better therapeutic index.

EXPERIMENTAL TECHNIQUES

X-ray crystallography

In X-ray crystallography, primary observations are used to calculate[1] the electron density of a molecule in a crystal. It is the density that defines the atomic positions. X-ray crystallography requires relatively large amounts (mg) of pure sample.

Macromolecules are notoriously difficult to crystallize and the preparation of high quality crystals which diffract well is still in art. High resolution (< 2.5 A°) defined X-ray structures offer the potential of designing drugs to fit their receptor with high affinity and selectivity. Medium resolution structures (2.5-3.0 A°) may be unreliable regarding the exact placement of side chains, but still provide a good starting point for qualitative modeling since the overall shape of binding side is generally correct. The protein-folding pattern should be well determined with resolutions up to 4-5 A°. The X-ray crystal structure of a macromolecule is usually a good model of the biologically active conformation in the solution. Additional information such as the characteristic values of bond lengths and bond angles and amino acid sequence of the protein, is also used to give a final structure that accurately reflects the conformation of the macromolecule in the structure.

Neutron diffraction

In X-ray diffraction analysis of macromolecules, it is unlikely that the H-atoms will be located because of their X-ray scattering. This limitation has been overcome by neutron diffraction[2]. Such high resolution neutron diffraction investigations have not only revealed H-atom position but, because of different scattering lengths, have also distinguished nitrogen from carbon and oxygen.

*For correspondence

Amino acid sequencing

Amino acid sequencing[3] demonstrates the value of a correctly determined protein sequence when interpreting X-ray observation. Moreover, protein sequences may be used to infer the three-dimensional structure of a protein by comparison with related conformations already determined for similar macromolecules. Fortunately, protein sequences occasionally show high sequence homology with proteins whose X-ray structure is known, suggesting the possibility of modeling the unknown structure based on the crystal structure of the homologous protein. This has become popular approach and has been applied to modeling several proteins. Sequence homologies of 50% or more are probably required for this approach to be successful; the resulting models in several cases have been accurate to within 1-1.5 A° for the main chain atoms, although side chains are not as well located.

Docking by molecular shape

The availability of three-dimensional structural information on a potential therapeutic target does not guarantee identification of the site of action of the substrate or inhibitor, unless the structure of a relevant complex has been determined. In fact, conformational changes often occur during binding of ligands to enzymes that are not reflected in the three-dimensional structure of the enzyme alone.

A method to explore the interaction of flexible ligands with receptors of known geometry on the basis of molecular shape[3] is reported. Here, goal is to find series of sterically reasonable binding geometries for a given ligand in a receptor binding site whose structure is known. Each ligand is divided into a small set of large fragments that are docked separately into the binding site and then rejointed later in the calculation. This step is repeated if necessary, until all fragments are joined and the entire ligand is regenerated. The rejoined fragments are then energy minimized in the receptor site.

Definition of site[4]

A program DOCK, which explores the steric complementarity between ligands and receptors of known three-dimensional structure is explored. Using the molecular surface of a receptor, a volumetric representation of the chosen binding cavity is approximated using a set of spheres of various sizes that have been mathematically 'packed' within it. The set of distances between the centers of the spheres serve as a compact representation of the shape of the cavity. The relative distance paradigm allows comparison without the need for orientation of one shape with respect to the other. Potential ligands are characterized in a similar fashion by generating a set of spheres that mimic the shape of the ligand.

A CAVITY program allows the investigator to isolate a single cavity of interest by specifying a seed point. From this point, the algorithm systematically explored the entire volume of the cavity, following its borders and effectively filling every crevice within it, i.e. a three-dimensional cast of the internal volume was produced using technique of solid modeling.

Characterization of site[4]

Hydrogen bonding and other group binding sites

The popular program GRID allows a probe atom or group to explore the receptor site cavity on a lattice or grid while estimating the enthalpy of interaction. A 3-D contour map can be generated from the lattice of interaction energies, which gives a graphical representation of the optimal positions for the atom, or group, in question. Similar ideas are embodied in field mapping used in the Comparative Molecular Field Analysis (CoMFA) paradigm. Ideal positions for H-bond donors or acceptors can the mapped in this fashion as a preface to ligand design.

In yet another method, multiple copies of ligands can be distributed in the active site by simulation. After simulated annealing, the quenched populations of ligands concentrated in various orientations at points within the receptor where optimal binding would occur. By connecting the ligands with the most energetically favoured binding with fragments from a library of molecular fragments, novel compounds could be designed for possible synthesis.

Electrostatic and Hydrophobic fields

Researchers interested in studying the electrostatic interactions between binding molecular entities usually do so by color coding the molecular surface of each molecule by electrostatic potential. These surfaces are than docked and visually inspected to note regions of electrostatic complementarity and dispiraty. In a similar way an estimate

of the surface can be quantitated and indicated through color coding.

Design for ligands[4]

Visually assisted design

In the process of optimization of a lead, one needs to ascertain where modification is feasible. When the ligand-receptor distances have been calculated at all cavity-pocket interface lattice points, a user-defined color-coding scale is implimented to generate the display. This highlights those areas that are less well packed and available for ligand modification.

Three-dimensional databases

Several databases are well known, such as the Cambribge Structural Database (CSD), which contains nearly 90,000 structures of small molecules. The crystal coordinates of proteins and other large macromolecules are deposited into the Brookhaven Protein Databank. The conformations present in the crystallographic database reflect low energy conformers that should be readily attainable in solution and in the receptor complex.

Noncrystallographic databases have also been developed. One example is the three-dimensional database structures from Chemical Abstracts generated using CONCORD, which contain 70,000 entries. The use of such databases is most applicable when the binding of a particular ligand and its receptor is well understood in terms of functional group recognition and a crystal structure of the complex is known. One approach to ligand design is to develop novel chemical architectures (i.e., scaffolds) that position the pharmacophoric groups or their bioisosters in the correct three-dimensional arrangement.

Similar database searching methods have been incorporated into a number of current database searching systems. Program such as CAVEAT, ALADDIN (Abott), 3D, SEARCH (Lederly), MACCS-3D, CHEM-X, UNITY-3DB and others contain considerable functionality useful for such an approach. Additional constraints have been incorporated into 3D SEARCH and ALADDIN, including the consideration of retrieved ligand-receptor volume complementarity. Furthermore, CHEM-X performs a rule-based conformational search on each structure in the database to account for conformational flexibility.

De Novo design

It has become quite evident that much of a molecule acts simply as a scaffold to align the appropriate groups in the three-dimensional arrangement that is crucial for molecular recognition. By understanding the pattern for a particular receptor, one can transcend a given chemical series by replacing one scaffold with another of geometric equivalence. This offers a logical way to change dramatically the side-effect profile of the drug as well as its physical and metabolic attributes. Various software tools are already under development to assist the chemist in this design objective.

CAVEAT is a program developed to find cyclic scaffold by searching the CSD for the correct vectorial arrangement of appended groups. All these approaches attempt to help the chemist to discover novel compounds which will be recognized at a given receptor. ALADDIN have described a program for the design or recognition of compounds that meet geometric, steric or substructural criteria. The cavity-matching algorithm DOCK has been quite successful in finding noncongeneric molecules of the correct shape to interact with a receptor cavity.

Designing ligands to fit a specific macromolecule site

As far as direct drug design is concerned, full interactive control[5] over the position (translation and rotation along the x,y and z coordinate axes) and conformation (adjustment of torsion angles) in both the macromolecule and the ligand(s) should be simultaneously available. Good torsion angle adjustment facilities are essential, especially where most of the time is spent in interactive modeling.

Simultaneous control of six to eight torsion angles should be possible, where the torsions may be spread over several residues or even several molecules. It should be possible to adjust both ends of a rotatable bond without having to continually reassign the bond rotation.

The system should be capable of handling several molecules simultaneously to enable the comparison of different legends in the binding sites or different fits of same ligands. Molecular surface displays (dot surfaces) should associate a set of dots with each atom, so that the dots move together with the atom as the molecule is moved during bond rotation. The system should also indicate the option to calculate fast van der Waals surfaces, which are

useful for surfacing small molecules and small portions of a macromolecule. Additional useful features include the ability to read new molecules into the current modeling session at any time (not just the beginning of the session) and to save individual molecules at any time.

The other approaches are usually to design and built the ligand piece by piece in the binding site by combining formed three-dimensional fragments from a library. The library may contain several hundred different ring system, chains and functional groups, which should be conveniently selectable from within the modeling system. Small molecules can be built rapidly this way and resulting structures are usually accurate enough for initial fitting or 'docking' into the model site. Computer Graphics enable us to qualitatively visualize drug receptor interactions and molecular mechanics can provide rough estimates of the interaction energy, which allows to design molecules that are apparently complementary to a bind site. For close analogs these can be sufficient to both rationalize the relative activities of a series of analogs and design new analogs; in particularly fortunate cases the X-ray structure of a suitable lead compound complexed with the receptor is available and new analogs can be docked by analogy with the lead compound. The most significant advance in modeling drug receptor interactions is a recent application of the free energy perturbation theory methods and solvent accessible molecular surface calculation.

Free energy perturbation theory[6] is based on a thermodynamic perturbation method with molecular dynamics being used to change or 'mutate' one molecule into another. It has been applied to variety of small molecules to calculate relative changes in free energy of solvation within 1 kcal/mole of experimentally measured values. Perturbation calculations are carried out by specifying the parameters in the beginning and end states, and the length of the time for equilibration and data collection in each window. The program automatically carries out the transformation between the two states during which intermediate energy values for each window are accumulated, stored and reported.

The Solvent accessible molecular surface calculation[7] is presented for analytically calculating a smooth, three dimensional contour about a molecule. Solvent accessible area is defined as the traced out by the center of the probe sphere representing a solvent molecule as it is rolled over the surface of the molecule of interest.

Simply measuring a quantity of area is insufficient for the study of many aspects of the protein and nucleic acid function, such as substrate binding and catalysis, drug-nucleic acid interaction. The molecular surface envelope may be drawn on either color raster computer displays or real time vector computer graphic systems. Molecular areas and volumes may be computed analytically from this surface representation. Unlike most previous computer graphic representations of molecules, which imitate wire models or space filling plastic spheres, this surface shows only the atoms that are accessible to solvent.

Special graphics requirements for macro molecular applications[8]

A macromolecule-modeling system should be capable of simultaneously handling a large number of molecules with several thousand atoms and thousands of molecular surface points in depth-cued color and time-sliced stereo, where each molecule can be individually controlled in the three dimensions, while simultaneously monitoring inter/intramolecular distances and adjusting multiple contiguous or noncontiguous torsion angles-all in real time.

Raster graphic has recently become the dominant technology in interactive molecular modeling. Although raster displays have apparent advantages in providing 'realistic' color solid-shaded images, these images cannot be updated fast enough for real time modeling of macromolecules. Color-coded molecular surfaces can provide qualitative or quantitative displays of hydrophobic or hydrophilic regions, neutral or charged amino acid side chains, electrostatic potential or mobility. Color coding by hydrophobicity and electrostatic potential are particularly useful in drug design applications, where the goal is to design a molecule which is complementary in shape, hydrophobicity and charge to a binding site. Hydrophobic color coding was originally done simply by coloring surface points associated with carbon 'hydrophobic' (e.g. red) and all nitrogen and oxygen surface points 'hydrophilic' (e.g. blue); a more detailed approach includes 'neutral' or semihydrophilic' surface (e.g. yellow) for sulfur, α-carbons of amino acids.

CONCLUSION

The principal use of Docking studies by scientists has been in fitting the model to the electron density and in refining the structure. The methods described above will help scientists in succeeding step of interpreting the solved structures. The color Raster display of solved accessible surface made possible by the analytical algorithm is better able to communicate structures discoveries because of its higher resolution and greater visual realism.

REFERENCES

1. Goodford, P.J., **J. Med. Chem.,** 1984, 27, 557.
2. Schoenborn, B.P., **Nature,** 1969, 224, 143.
3. DesJarlias, R.L., Scheridan, R.P., Dixon, S.J., Kuntz, D.I., and Venkataraghavan, R., **J. Med. Chem.,** 1986, 29, 2149.
4. Wolff, M. E. In; Mandrede, E.W., Eds. Burger's Medicinal Chemistry and Drug Delivery, 5th Edn. Vol 1, A Wiley Interscience Publication. USA, 1995, 600.
5. Gosh, A.K. and Crippen G.M., **J. Med. Chem.,** 1982, 25, 892.
6. Bash, P.A., Singh, U.C., Brown, F.K., Langridge, R., and Kollman, P.A., **Science,** 1987, 235, 574.
7. Conally, M.L., **Science,** 1983, 221, 709.
8. Hansch, C., Sammes, P.G., and Taylor, J.B. In. Ramsden, C.A. Eds; Comprehensive Medicinal Chemistry-The Rational Design, Mechanistic Study and Therapeutic Applications of Chemical Compounds, 1st Edn. Vol 4, Pergman Press, New York, 1990, 460.