Academic Sciences International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 5, Suppl 1, 2013

Review Article

MOLECULAR MODELLING: A NEW SCAFFOLD FOR DRUG DESIGN

SONIKA REDHU[1], ANIL JINDAL[2]

[1]JCDM College of Pharmacy, Sirsa 125055, Haryana, India, [2]L. R. Institute of Pharmacy, Rajgarh Road, Solan 173212, India.
Email: sonika285@yahoo.co.in

Received: 30 Oct 2012, Revised and Accepted: 05 Dec 2012

ABSTRACT

Molecular modeling is becoming ever more important in the fields of protein engineering and drug design. X-ray diffraction and nuclear magnetic resonance techniques are improving rapidly, but nevertheless it is unlikely that the three-dimensional structures of all molecules of interest will be elucidated in the foreseeable future. Another line of work is drug design. Drug design is a creative act of the same magnitude as composing, sculpting or writing. The results can touch the lives of millions, but the creator is rarely one scientist and the rewards are distributed differently in the arts than in the sciences. The aim of this review is to give an outline of studies in the field of medicinal chemistry in which molecular modeling has helped in the discovery process of new drugs. The emphasis will be on lead generation and optimization.

Keywords: Molecular modeling, Drug design, Molecular software, Computational chemistry

INTRODUCTION

Molecular modeling encompasses all theoretical methods and computational techniques used to model or mimic the behavior of molecules. The techniques are used in the fields of computational chemistry, drug design, computational biology and materials science for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies. The simplest calculations can be performed by hand, but inevitably computers are required to perform molecular modeling of any reasonably sized system. The common feature of molecular modeling techniques is the atomistic level description of the molecular systems. Most molecular modeling studies involve three stages. In the first stage a model is selected to describe the intra- and inter- molecular interactions in the system. The two most common models that are used in molecular modeling are quantum mechanics and molecular mechanics. These models enable the energy of any arrangement of the atoms and molecules in the system to be calculated, and allow the modeler to determine how the energy of the system varies as the positions of the atoms and molecules change. The second stage of a molecular modeling study is the calculation itself, such as an energy minimization, a molecular dynamics or Monte Carlo simulation, or a conformational search. Finally, the calculation must be analyzed, not only to calculate properties but also to check that it has been performed properly. Computational chemistry/molecular modeling is the science (or art) of representing molecular structures numerically and simulating their behavior with the equations of quantum and classical physics. Computational chemistry programs allow scientists to generate and present molecular data including geometries (bond lengths, bond angles and torsion angles), energies (heat of formation, activation energy, etc.), electronic properties (moments, charges, ionization potential and electron affinity), spectroscopic properties (vibrational modes, chemical shifts) and bulk properties (volumes, surface areas, diffusion, viscosity, etc.). As with all models however, the chemist's intuition and training is necessary to interpret the results appropriately. One of the earliest and still one of the largest uses of computers is to solve complex problems in the natural sciences and engineering disciplines and more specifically to obtain solutions of mathematical models that describe chemical or physical phenomena (or processes). The techniques used to obtain such solutions are part of the general area called Scientific Computing, and the use of these techniques to obtain insights into scientific or engineering problems is called Computational Science. Computational Science is a rapidly-emerging transdisciplinary field at the intersection of the natural sciences, computer science, and mathematics because much scientific investigation now involves computing as well as theory and experiment:

Computational Science = mathematics+ computer science + field of application.

Computational Science typically unifies three distinct elements: (i) modeling, algorithms and simulations (e.g. numerical and nonnumerical, discrete and continuous); (ii) software developed to solve natural science, social science, engineering, and medical problems; (iii) computer and information science that develops and optimizes advanced hardware systems, software, networking and data management components.

DRUG DESIGN

Drug design is a synthesis of scientific knowledge, experience, intuition, and aesthetics. However, unlike the arts, this beauty has limited distribution; the general public is severely under informed about the creative process whereby molecules are designed and created. Indeed, like artists, scientists are hard-pressed to enunciate their intuitive insights. Three crucial components have converged to accentuate structure-based drug design as a product of the end of this century. Because of expanded *databases* (especially from crystallography and NMR), computer graphics *displays,* and linkage by facile *web tools* (Meyer & Funkhouser, 1998), these resources put precise molecular structures before our eyes and at our fingertips. The structural databases, which have been growing exponentially for three decades, now offer a number of interesting targets for structure-based drug design. While graphics hardware and prototypical software for drug design have been available since the 1960s (Meyer, 1980), only during the past decade have they matured to the extent that a synthetic chemist can now master molecular modeling without having to become a computer scientist; and not very many computer scientists have become synthetic chemists

MODELLING CHALLENGES

The current explosion in gene databases and “structural genomics” eventually will provide essential sequence information, but the lack of knowledge of the laws of protein folding generally prohibits the inference of structure from sequence. Targeted studies, which seek to understand the basic chemistry and physiology of a disease are more promising. For example, human immunodeficiency virus (HIV) proteinase inhibitor development surely must be one of the most spectacular successes in the brief history of structure-based drug design. A novel approach scans a pathological vector using the tools of molecular biology of the many relevant proteins produced a few can be isolated, crystallized and structurally elucidated. For example, *Pyrobaculum aerophilum* is reported to be a cofactor for the Rev and Rex Transactivator proteins of HIV-1 and T-cell leukemia virus I. It has been studied (Peat et al., 1998) at 1.75 Å resolutions and identified as a target for chemical interdiction. The structures of normal and pathological molecules can be compared and compounds designed to inhibit pathogenic enzymes or receptors selectively. So with the structure of even one target

protein and the knowledge of function of its receptor or active site it is now possible to use computer tools to build and dock a ligand or inhibitor ("new leads") prior to investing time and resources for synthesis and testing. Conversely, large-scale screening may detect "new leads" that then must be modeled in order to explore subsequent synthetic analogs. In either case, molecular modeling is essential for understanding and exploring the structure-function relationship. Attractive and repulsive forces can be summed and the fit quantified. Ideally, one seeks a correlated listing of experimental and computational values to give assurance that novel compounds can be evaluated before being synthesized.

COMPUTER GRAPHICS SOFTWARE

Software for molecular modeling is steadily improving both in software standards and the underlying science. Commercial interests play a part in this process. There are indications that this improvement will continue. The first molecular modeling facilities were relatively simple program which took a file of protein atomic positions from disk storage or tape displayed them and allowed a user to alter the positions with I/O devices. Periodically during the session the modeler would write the new coordinate file back onto the disk. If a new representation was used this involved a different program which could require the coordinate data in a different format or which used a different style of interaction. Today researchers realize the benefits of using a true computer system where all modules (replacing program) have the same internal data format and the modeler has one mode of interaction. There are four parts of a molecular modeling system:

- Molecular database
- Graphics system
- Application programs
- System skeleton

The molecular database holds the positions of the atoms (coordinates) and other information such as the protein sequence, amino acid properties, and atom properties. The *Winchester Graphics System* is based upon a relational database which lacks the efficiency of the UCSF implementation but is more flexible.

Graphics systems make the picture from the data. Having done this, the graphics system looks after the user interaction with the device. This involves polling the I/O devices, carrying out analog to digital conversion of the device signals, and loading the resulting transformation matrices into the special registers of the graphics hardware. These are applied and a new picture is calculated. This is carried out about 30 times a second to obtain smooth rotation of the molecules. There are two parts to the graphics subsystem. There is one part that is dependent upon the graphics hardware used and the second which depends upon the structure of the database and the system skeleton.

Application programs are very important in protein modeling. In particular, little modeling can be carried out effectively without using a molecular mechanics energy program. This program can be used for finding the minimum energy conformation, modeling the potential field around the protein, and simulating molecular dynamics. The need for energy minimization which can keep up with the graphics system has caused groups to look at new algorithms, the use of parallel processors, and specialized hardware. The system skeleton handles the passing of data between the subsystems and takes care of input/output.

MOLECULAR MODELING SOFTWARE

1. *Abalone* is designed for macromolecular simulations (proteins, DNA). It supports both explicit and implicit solvent models. In contrast to Ascalaph, tailored to the simulation of small molecules, Abalone is focused on molecular dynamics modeling of biopolymers. It supports such effective methods as the Replica Exchange and hybrid Monte Carlo.
2. *Ascalaph* is general purpose molecular modeling software that performs quantum mechanics calculations for initial molecular model development, molecular mechanics and dynamics simulations in the gas or in condensed phase. It can interact with external molecular modeling packages (MDynaMix, NWChem, CP2K, PC GAMESS/Firefly and Delphi).
3. *Yasara* is a molecular-graphics, modeling and simulation package for Linux and Windows. Yasara is powered by PVL (Portable Vector Language), a new development framework. PVL allows you to visualize even the largest proteins and enables true interactive real-time simulations with highly accurate force fields on standard PCs.
4. *RasMol* is a molecular graphics program developed at the University of Edinburgh. The software is intended for the visualization of proteins, nucleic acids and small molecules. The program has the ability to read in PDB as well as several other formats. Coloring schemes including atom type, temperature factor and hydrophobicity.
5. *MacroModel* is a computer program for molecular modelling of organic compounds and biopolymers. It features various force fields coupled with energy minimization algorithms for the prediction of geometry and relative conformational energies of molecules. MacroModel also has the ability to perform molecular dynamics simulations to model systems at finite temperatures using stochastic dynamics and mixed Monte Carlo algorithms.
6. *SYBYL-X* provides capabilities for crucial small molecular modeling and simulation, includng structure-activity relationship modeling, pharmacophore hypothesis generation, molecular alignment, conformational searching, homology modeling, sequence alignment, and other key tasks required to understand and model the static and dynamic 3D structural properties of proteins and other biological macromolecules.
7. *SOMA2* environment is a web server based system offering a framework for integrating molecular modelling applications including molecular data exchange. SOMA2 allows users to combine software available in the computing system into unique workflows which are automatically executed.
8. *Amber* is a suite of programs for molecular simulation and analysis of proteins, nucleic acids, lipids, carbohydrates. Amber" refers to two things: a set of molecular mechanical force fields for the simulation of biomolecules (which are in the public domain, and are used in a variety of simulation programs); and a package of molecular simulation programs which includes source code and demos.
9. *VMD* is designed for modeling, visualization, and analysis of biological systems such as proteins, nucleic acids, lipid bilayer assemblies etc. VMD provides a wide variety of methods for rendering and coloring a molecule: simple points and lines, CPK spheres and cylinders, licorice bonds, backbone tubes and ribbons, cartoon drawings, and others. VMD can be used to animate and analyze the trajectory of a molecular dynamics (MD) simulation.
10. *MOE* internal representation of organic chemical structures and flexible architecture provide a solid foundation for molecular modeling and computational chemistry.
11. The *PUPIL* (Program for User Packages Interfacing and Linking) system is designed to facilitate communication and control among the different software packages (Calculation Units, "CU"s) used in a multi-scale simulation. All simulation data were packed and unpacked appropriately to pass through the PUPIL application. All stubs were blocked properly (to synchronize with the information on which they are waiting), and released at the right moment (that is, when the framework had put the information into the stub data structures).
12. *SIMLYS* is a tool to aid in the analysis of molecular dynamics, Monte Carlo and other stimulations. Its purpose is twofold it is a system performing the actual analysis and it serves as a shell to integrate new analysis functions. SIMLYS allows one to analyse the results from various simulations, as for example from proteins or polymers, by using the trajectories. The program is separated into modules performing the input/output, building the interface to the user, preparing the coordinates and performing the calculations.

PROTEIN DATA

To access crystal structures rapidly, all files containing protein atom coordinates are stored on a local network Furthermore, the in-house protein structural data is preprocessed-- in the *.pdb* files we explicitly define ligands in active sites with HETATM records and all

other atoms with ATOM records. This enables the active sites of proteins to be identified immediately. Extra information about the structures, necessary for certain program macros, is coded into the CGI programs. This information includes tables of conserved active site residues (used for the superimposition of related proteins), the names of structures that do not contain waters, and the names of dimeric proteins.

MOLECULAR DOCKING

Molecular docking can suggest a favorable configuration for two molecules forming a complex system. Molecular docking has been applied to studies of protein–ligand interactions, and structural information from the theoretically modeled complex may help us clarify the mechanism of molecular recognition. Such models can even suggest modifications to lead structures to improve biological activity. We have developed different scoring functions for two stages of molecular docking. In the first step, the dot surface is generated using the program written by Connolly. The coordinates of the probe molecule and the target molecule surfaces are randomly rotated and translated. An initial solution is randomly generated containing six variables: three translational degrees of freedom, and three rotational degrees of freedom. The three rotational variables are described by three Euler angles. The position of the target molecule is fixed and the six variables define the orientation of the probe molecule. The initial solution is evaluated using surface complementarity. The evaluation score is composed of two parts: the matching score and penalty score of atomic overlapping.

$$Fitness = Score_{match} - const \times Score_{overlap}$$

Where $Score_{match}$ is the matching score and $Score_{overlap}$ is the penalty score. Const is a coefficient balancing the contributions of the two parts. Const is mainly determined by the dot density, an important parameter of MS program, which is defined as the average number of dots per square angstrom area of both probe and target molecules.

Molecular Docking Studies of Two Unbound Complexes

It has been shown that for bound complexes, surface complementarity usually is sufficient to obtain correct binding conformations. However, the ultimate goal of molecular docking is to predict protein–protein and protein–peptide interaction without requiring a complexed crystal structure. Compared with docking of these complexes with crystal structures, calculations for complexes without crystal structures are more difficult. During the formation of a complex, some molecules will undergo conformational changes, so the docking procedure must be sufficiently flexible to manage conformational changes, yet specific enough to identify the correct solution. In some cases, especially when the binding regions between proteins and/or peptides are unknown, complete conformational searches are not tractable. Even using rigid-body approximations, it is very difficult to determine the global minimum using conventional minimization algorithms**.** To test our hybrid minimization algorithm and docking procedure, two uncomplexed systems were studied. One study employed an uncomplexed trypsin inhibitor (4PTI in PDB) and an uncomplexed trypsin (3PTN in PDB). The other example used an uncomplexed serine proteinase (2PKA) and an uncomplexed bovine pancreatic trypsin inhibitor (2BPI). The PDB codes of these two cases are 2PTC and 2KAI. All crystallographic water molecules were eliminated from the structures. Some missing hydrogen atoms were added to the complexes using the molecular design software. The crystal structures of uncomplexed trypsin (3PTN) and an uncomplexed trypsin inhibitor (4PTI) have been solved separately in different systole forms. A comparison of their structures with the corresponding components of a complex has indicated that relatively large conformational changes have occurred, especially in the trypsin inhibitor.

THE MONTE-CARLO SIMULATION

The principle of a Monte-Carlo algorithm is the following: Firstly consider a conformation (number N) of energy E N. Then modify one (or more) torsion angle(s) by adding a random increment to the current torsion value. This process is repeated until reaching a conformation without van der Waals contacts. The random increment must be in user-defined limits. The energy associated to the new conformation N+1 is computed (E_{N+1}). Subsequently, we compare both energies E_N and E_{N+1}. If the new energy is lower than the previous one, the program accepts the new conformation N+1 as the starting conformation for the next perturbation. If $E_{N+1} > E_N$,a Boltzman distribution is used for a decision. A random number between 0 and 1 is generated. Then the following Boltzman distributed term is computed: $\exp[-(E_{N+1} - E_N/RT]$ where R is the gas constant and T the absolute temperature. If the exponent is lower than the random number, the conformation N+1 is taken as the new reference conformation. N will be re-used as the starting conformation for the next perturbation. The purpose of the Monte-Carlo method is to provide an efficient sampling of the conformational space. In order to increase the speed of the algorithm, it was decided to treat bond length and angles as fixed. As a consequence, only van der Waals and torsional terms need to be computed to evaluate energetic variations from the conformation N to N + 1. A simplified force field including only previous terms has also been implemented in the Monte-Carlo algorithm.

APPLICATIONS

- Molecular modeling has been used to predict the ability of the enzymes to discriminate between the enantiomers of a series of substrates, and has also been applied for guidance on redesigning CRL by site-directed mutagenesis in order to test the importance of a couple of amino acids selected on the basis of the computational results.
- Molecular modeling techniques have been applied to allow the identification of a possible binding model for a class of suramin-related anti-HIV-1 inhibitors (Suradistas) on the HIV-1 cellular receptor (CD4) surface, without detailed structural information about the binding site for the ligand on CD4.
- In a third application, finally, in the absence of any experimental structure (NMR or X-ray) of the macromolecular biological target (*Candida albicans* lanosterol 14α-demethylase) a model has been deduced from the ligands that bind to it by means of pseudo receptor modeling.
- Computer modeling has the potential to provide new insights into reaction pathways, to predict properties of catalysts that have not yet been synthesized, and to bring information for a given system from many different experimental techniques. e.g. kinetic studies of reaction rates, thermodynamic information on adsorption and spectroscopic data on molecular-level structure.

REFERENCES

1. http://en.wikipedia.org/wiki/Molecular_modelling
2. Andrew R,Leach. Molecular Modelling:Principles and Applications
3. Allen B. Richon: An Introduction to Molecular Modeling Molecular Solutions, Inc. 4411 Connecticut Avenue NW, STE 514 Washington, DC 20008-8677.
4. Rama Rao Nadendla. Molecular Modeling: A Powerful Tool for Drug Design and Molecular Docking
5. T Perun, C L Propst. Computer Aided Drug Design, Marcel Dekker Inc. New York.1989;2-4.
6. N Claude Cohen, Guide Book on Molecular Modelling in Drug Design Academic Press. 1995;56.
7. Boyd DB, Lipkowitz KB, J.Chem.Educ. 1982:29(4):269.
8. Marshall GR, Maylor CB. Comprehensive Medicinal Chemistry Pergamon.New York.1990:431.
9. Venkataraghavan R, J. Med. Chem.1982;29:.2149.
10. http: www. netsci.org/Science/ Compchem/feature 01.html.
11. Aithal KS, Singh UV, Satyanarayan K, Udupa N.Indian Drugs.1998;60(2):68.
12. Daniluk A.Visual modeling for scientific software architecture design. A practical approach Computer Physics Communications 2012;183:213–230
13. Denning PJ. Computer science: the discipline in: Encyclopedia of Computer Science. Nature Publishing Group, London, 2000.
14. Eykhoff P, System Identification: Parameter and State Estimation, Wiley. New York.1974.

15. Vriend G. WHAT IF: A molecular modeling and drug design program BIOSON Research Institute BIOSON Research Institute, Laboratory of Chemical Physics, Department of Chemistry, University of Groningen, Groningen. The Netherlands J. Mol. Graphics. 1990;8.
16. Bernstein FC. J. Mol. Biol. 1977;112:535-542
17. Cherfils J, Vaney MC, Morize I, Surcouf E, Colloc N, Momon JP. J. Mol. Graphics 1988;3:155-160
18. Vinter JG, Gardner M. Molecular Modeling and Drug Design. London: Macmillan; 1994.
19. Kollman P: Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. Accounts Chem Res 1996;29:461-469.
20. Jorgensen WL, Kaminski G: Performance of the AMBER94, MMFF94 and OPLS-AA force fields for modeling organic liquids. J Phys Chem 1996;100:1801-8013.
21. Case DA, Cheatham TE, Darden T, Gohlke, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods B. The Amber biomolecular simulation programs. J. Computat. Chem. 2005;26:1668-1688.
22. Ponder JW, Case DA. Force fields for protein simulations. Adv. Prot. Chem. 2003;66:27-85.
23. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. J Comput Chem. 2005 Dec;26(16):1668-88.
24. Taylor R, Mullier GW, Sexton GJ. Automation of conformational analysis and other molecular modeling calculations J. Mol. Graphics, 1992;10.
25. Cohen NC, Blaney JM, Humblet C, Gund P, Barry DC. J. Med. Chem. 1990;33:883-894
26. Odell BJ. Comput.Aid. Molec. Design. 1988;2:191-216.
27. SYBYL Users Manual,Tripos Associates, St. Louis, MO 63117, USA
28. Broughton HB. Molecular modeling Current Opinion in Chemical Biology. 1997;1:392-398
29. Hou T, Xu X. A new molecular simulation software package: Peking University Drug Design System (PKUDDS) for structure-based drug design. Journal of Molecular Graphics and Modelling 2001;19(5):455–465.
30. Wang JM, Hou TJ. Automated docking of peptides and proteins by genetic algorithm. Chemometr. Intell. Lab. 1999;45:281–286.
31. Hou TJ, Wang JM, Xu XJ. Automated docking of peptides and proteins by using a genetic algorithm combined with a tabu search. Protein Eng. 1999;12:639–647.
32. Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. Science 1983, 221, 709–713.
33. Ooms F. Molecular Modeling and Computer Aided Drug Design. Examples of their Applications in Medicinal Chemistry Current Medicinal Chemistry. 2000;7:41-158
34. Torras J, He Y, Cao C, Muralidharan K, Deumens E, Cheng HP, TrickeySB. PUPIL:A systematic approach to software integration in multi-scale simulations Computer Physics Communications 2007;177:265–279
35. Aqvist J, Warshel A, Chem. Rev. 2001;93:597–653.
36. Zhu T, Li J, Yip S, Bartlett RJ, Trickey SB, Leeuw DNH. Molecular Simulation 2003;29:671.
37. Kruger P, Luke M, Szameit A. SIMLYS - a software package for trajectory analysis of molecular dynamics simulations. Computer Physics Communications.1991;62:371-380.
38. Kinnunen T, Nyrönen T, Lehtovuori P. SOMA2 – Open source framework for molecular modelling workflows CSC – Finnish IT Center for Science
39. Lehtovuori P, Nyrönen T. J. Chem. Inf. Model. 2006;46(2):620.
40. Taylor RN, Smith R: The World Wide Web as a graphical user interface to program macros for molecular graphics, molecular modeling, and structure-base drug design. Journal of Molecular Graphics 1996;14:291-296.
41. Meyer FE, Swanson MS, Williams AJ. Molecular modelling and drug design. Pharmacology & Therapeutics 2000;85:113–121.