International Journal of Bioinformatics Research
ISSN: 0975-3087 & E-ISSN: 0975-9115, Volume 4, Issue 1, 2012, pp.-245-248.
Available online at http://www.bioinfo.in/contents.php?id=21

MOLECULAR DYNAMICS: BASIC STUDY

RICHA GUPTA*, ARCHANA TIWARI AND MAHAVIR YADAV

School of Biotechnology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, University of Technology of Madhya Pradesh, Bhopal, MP, India.
*Corresponding Author: Email- richa069@gmail.com

Received: January 13, 2012; Accepted: March 09, 2012

Abstract- Molecular mechanics (MM) these days tends to be concerned only with prediction of local minima on molecular potential energy surfaces. QSAR properties are often calculated in order to assist high-volume screening studies in pharmaceuticals applications. Should we want to study the motions of the molecule, all that would be needed would be to investigate the normal modes of vibration (which can be obtained from the hessian). MM does not take account of zero point vibrations and the calculations refer to a molecule at 0 K, when it is completely at rest. Workers in the modeling field often refer to MM as energy minimization. Molecular modeling is readily available for QSAR studies, pharmacophore studies. Its implementation can be for the design of vaccine, and drugs for therapeutic use..
Keywords- Molecular modeling, protein, force field, simulation, molecular dynamics, bioinformatics, conformation, homology modeling.

Citation: Richa Gupta, Archana Tiwari and Mahavir Yadav (2012) Molecular Dynamics: Basic Study. International Journal of Bioinformatics Research, ISSN: 0975-3087 & E-ISSN: 0975-9115, Volume 4, Issue 1, pp.-245-248.

Introduction

Molecular modeling is powerful methodology for analyzing the three dimensional structures of biological macromolecules. There are many ways in which molecular modeling methods have been used to address problems in structural biology. It is not widely appreciated that modeling methods are often an integral component of structure determination by NMR spectroscopy and X- ray crystallography. In the review we consider some of the numerous ways in which modeling can be used to interpret and rationalize experimental data and in constructing hypothesis that can be tested by experiment. Genome sequencing project are producing a vast wealth of data describing the protein coding regions of the genome under study. However, only a minority of the protein sequences thus identified will have a clear sequence of homology to a known protein. In such cases valuable three- dimensional models of protein coding sequence can be constructed by homology modeling methods.

How to compute molecular modeling potential

There are many levels of theory at which computational models of three- dimensional molecular structure can be constructed. The overall aim of modeling methods is to relate biological activity to structure. An important step towards this goal is to be able to compute the potential energy of the molecule as a function of the position of the constituent atoms. Quantum methods in which electronic structure is explicitly considered [1] are more rigorous, but because of the computational demands involved, they can usually only be applied to smaller molecules. The molecular mechanics potential energy function can be then written as

$$V=\frac{1}{2}\sum_{\text{bond}} K_b\,(b_{eqm}-b)^2+\frac{1}{2}\sum_{\text{angles}} K_\theta\,(\theta_{eqm}-\theta)^2+\frac{1}{2}\sum_{\text{torsions}} K_\theta[\,1+\cos n\varphi\,]+$$

$$\sum_{\text{non- bonded pairs i-j}} [\,A/\,r_{ij}{}^{12}-B/r_{ij}{}^{6}+q_i q_j/\,Dr_{ij}\,]$$

The equilibrium bond lengths (b_{eqrm}), bond angles (Θ_{eqrm}), partials

charge values (q), force constants (K) and Vander Waals parameters (A, B) must be carefully determined. This is normally achieved by fitting to experimental data and/ or higher level quantum calculations. Such an expression treats the molecular system in a highly simplified fashion, for example, the electrostatics of the system are assumed to be describable by uniform dielectric constant (D) with partial charges placed at atom centers. Dielectric constants that vary with distance are sometimes used in simulation as a very approximate scheme to mimic the screening of the electrostatic interactions by solvent. A more realistic treatment of solvent is to include explicit ions and solvent molecules in the molecular model and use a dielectric constant of one. However modeling using explicit solvent and ions greatly increases the number of atoms in a model, leading to an increase in the time required to complete calculation. Such calculation may also require knowledge of salt concentration, ionization state of particular residue etc. Other calculation schemes known as continuum electrostatic models [2] address the electrostatic effect of solvation, by treating the macromolecule as a region of low dielectric material surrounded by a continuous medium of high dielectric. The simple nature of the above molecular mechanisms expression means that it can be rapidly evaluated and applied to molecular systems with many thousands of atoms. The selection of given mathematical or functional form along with a chosen set of molecular mechanisms parameters is usually referred to as a force field. Some force fields also include additional terms designed to maintain non planarity of certain atom environments (eg N atoms in amine groups), or to explicitly model hydrogen bonding. Other force field allow for so called cross terms in which, for example, an energy term may arise for the product of bond length and bond angle deformation. In the field of protein modeling there are many examples of using sets of known protein structures to derive 'knowledge based' potential functions. These functions are fundamentally different from the molecular mechanics potential function described above. The knowledge based potentials utilize a representative database of solved protein structures to provide sampling types of interactions that occur in proteins. Interactions that occur more or less frequently than those expected on a statistical basis, are parameterized to represent favorable or unfavorable terms in the scoring function respectively. Many of these potential functions are designed for use with a simplified model of each amino acid in the protein.

Exploring conformation space

When a method for evaluation of molecular potential energy is available, it is natural to try to find an optimum molecular geometry by minimizing the energy of the system. A number of distinct algorithms are available for seeking the values of adjustable parameters that minimize a mathematical scoring function. Algorithms differ in the way in which they use the gradient of energy as well as in their search efficiency. In a biological macromolecule the potential energy surface is a complicated one, in which there are many local energy minima as well as a single overall global energy minimum. All the above energy minimization algorithms have a marked tendency to locate only a local energy minimum that is close to the starting conformation. For a biological macromolecule the number of conformations that must be searched rises exponentially with the size of the molecule, hence systematic searching is not a practical method for larger molecules.

Molecular dynamics (MD) is a conformational space search procedure in which the atoms of biological macromolecule are given an initial velocity, and are then allowed to evolve in time according to the laws of Newtonian mechanics [3]. Depending upon the simulated temperature of the system, the macromolecule can then overcome barriers in the potential energy surface in a way that is not possible with a minimization procedure. This scheme provides a picture of the molecular conformation that is a dynamic one, as opposed to the static picture provided by a minimization procedure. One useful combination of molecular dynamics and minimization schemes is a method known as simulated annealing. This method uses a molecular dynamics calculation in which the system temperature is raised to a high value, allowing for a widespread exploration of the available conformation space. Then system temperature is gradually decreased as further dynamics is performed. Finally a minimization phase may be used to select a minimum energy molecular conformation.

The annealing scheme is typically run several times in order to create an ensemble of representative molecular conformation. One of the most important applications of molecular modeling techniques in structural biology is the simulation of the docking of a ligand molecule to a receptor, such as protein. If the structure of the receptor is known then the application is essentially one of structure based drug design.

Common modeling force fields and packages

One of the most widely used packages is AMBER (Assisted Model Building with Energy Refinement). The AMBER force field (4- 6) was originally developed with the intent of enabling the simulation of protein and nucleic acid molecules. This package is capable of simulating a wide range of biological macromolecules. Details of AMBER force field and software package can be obtained from the website *The Amber Molecular Dynamics Package* [7]. The AMBER software allows simulations such as simple minimization, molecular dynamics and simulated annealing. There are other force field and molecular mechanics software packages that offer a broadly similar range of capabilities to AMBER. The CHARMM (Chemistry at HARvard Molecular Mechanics) force field/software [8] is another widely used and highly capable package. The program was designed for simulations of proteins, nucleic acid and lipids. Both AMBER and CHARMM software are available to academic users for a modest cost. AMBER and CHARMM force fields have also been implemented in a number of different software packages.

Commercially available modeling packages such as those from Molecular Simulation [9] or Tripos [10] include several of the force field mentioned above; they often provide an integral graphical environment for simulations that is reasonably easy for novice user to use.

Protein homology modeling

Why create protein homology models

Protein and nucleic acid sequencing are now well advanced and available in many laboratories. As a result sequence databases such as the protein information resource, SwissProt and TrEMBL have been growing rapidly in recent years. In contrast the determination of protein structure by NMR or X- ray crystallography has

tended to proceed much more slowly. Hence there are many important proteins where the sequence is available but the three dimensional structure is not yet known. One of the grand challenges of computational science is to be able to predict the overall fold of a protein purely from its sequence. This is commonly known as the protein folding problem. Homology modeling or comparative modeling methods, first reported by [11] are able to predict the 3-D structure of a protein sequence by using information derived from a homologous protein of known structure [12, 13]. The utility of homology modeling is evident when considering the vast numbers of open reading frames (ORFs), which are potential protein coding sequences, produced as a result of genome sequencing proteins. It has been estimated that the order of 20- 30% of these open reading frames can be assigned to a fold classification derived from structures in the PDB protein structural databank [14, 15].

If a three dimensional model of the protein of interest can be derived, it may be usable as the basis for a structure based drug design study. In addition such models can be useful aid to the rational design of experiments such as site directed mutagenesis or in understanding protein stability and function. In short it may be easier to rationalize the behavior of a protein if an experimental or model three dimensional structure is available, than it is to do so solely on the basis of sequence information alone.

Outline of homology modeling schemes

In order to construct a homology model for a query protein sequence, the query must first be aligned with one or more homologous reference proteins of known structure. Experience of homology modeling shows that when the sequence identity between two proteins falls to 30% or less, then alignment process becomes increasingly unreliable. The consequence of this will be regions of the protein model that are incorrectly folded in relation to the true structure. Fragment based homology modeling procedures use alignment between query sequences and known protein(s) to identify a number of structurally conserved regions (SCRs). Unlike fragment based, the restrain based homology modeling methods do not generally break the model building process into two distinct phase's i.e. building conserved regions then finding variable loop regions. Instead the alignment is used to derive geometrical restrains, such as limits on distances between pair of C^{α} atoms, ranges of backbone and side chain dihedral angles etc. in the method reported by [16] distance geometry structure generation procedure was tested on Kazal type trypsin inhibitors and used to predict the structure of the human pancreatic secretory trypsin inhibitor. In this approach the restrains are typically atom- atom distances derived from corresponding atoms in the known structure and compared to the distances in the model protein. Restrain based molecular dynamics procedures for the structure generation has been used in the MODELLER program [12]

Example of automated homology modeling

Increased automation of the homology model building process can make the benefits of modeling available to the wider audience of non- experts, although caution and expertise will always be required for a critical appreciation of the results. The Swiss- Model program suite [17, 18] provides one level of operation (known as first approach mode) that requires only a protein sequence as input. The method is also provided as a World Wide Web accessible server [19]. This development allows simplified access to homology modeling methods without the need to purchase specialized hardware and software.

Evaluating protein homology models

Difficult cases in homology modeling correspond to protein sequences that only possess distant homologues of known structure, where the level of sequence identity may be low. In such cases incorrect alignment can lead to regions of a model protein structures that have significant structural errors. Tools that can predict the quality of model protein structures and identify erroneous regions are valuable for model selection and helping to identify alignment errors. Simple checks on the geometry of the model protein bonds, bond angles and torsions etc. can be performed using a program such as PROCHECK [20]. Several other groups have designed to highlight residues that don't possess favorable environments or interactions [21-23]. When the model protein is compared with the known structure the utility of this scoring scheme can be tested. The knowledge based scoring functions have real value in assessing the reliability of various regions of a model protein structure. The accuracy of model protein structure is most strongly limited by the accuracy of protein loop regions.

Modeling protein- ligand complexes

The goals of protein- ligand docking

One of the most important and useful areas of application of molecular modeling is the approach of fitting together, or docking, a protein to a second molecule. Typically the latter is a small molecule ligand. This is of interest because it models the possible interactions between the proteins and the ligand in the formation of biologically important protein- ligand complex.

Protein- ligand docking

Bitomsky [24] have reported a comparison of the programs GRID, DOCK and AUTODOCK when applied to the task of docking heparin oligosaccharides to three proteins. The protein studied were acidic and basic fibroblast growth factors (FGFs) along with antithrombin; these were the only proteins for which heparin- protein complex crystal structures were available at the time of the report. All three programs were able to correctly identify the heparin-binding site on the protein.

Summary and Conclusions

From the range of application reported, it should be clear that molecular modeling is a very versatile technique and can be applied to many areas of macromolecular structural studies. Pure prediction in which no direct experimental data are used is still an area that must be approached cautiously; there are many difficulties and pitfalls that await the unwary. Successful prediction methods are likely to require the careful development of a model that is realistic and not yet computationally tractable. In this regard the growing trend for holding prediction contests is very valuable. It allows numerous computational methods to be applied on a common set of problems and for them to be evaluated in a common way. Prior to the emergence of such contests it was often difficult to evaluate the relative merits of methods from different laboratories, or variations in methods from within the same laboratory.

Simulation schemes can incorporate data from a number of experimental disciplines, including EM. As computer system become more powerful with time the utility of modeling and simulation methods can only increase. This will happen in two ways. Firstly, existing types of simulation will be able to run for longer time periods, thus allowing better sampling of conformational space and property statistics. Secondly, more realistic but expensive computation schemes will be accessible in a reasonable time. The growth of structure database such as PDB will make the application of methods such as protein homology modeling, threading and virtual docking applicable to a wider range of proteins. Molecular modeling methods have much to contribute to our understanding of structural biology.

References

[1] Foresman J.B., Frisch A. (1996).
[2] Jan Antosiewicz, Andrew McCammon J., Michael K. Gilson (1994) *Journal of Molecular Biology*, 238, 415-436.
[3] Van Gunsteren, Brenedsen H.J.C. (1997) *Molecular Physics*, 34, 1311-1327.
[4] Weiner S.J. (1984) *Journal of American Chemical Society*, 106, 765-784.
[5] Weiner S.J. (1986) *Journal of Computational Chemistry*, 7, 230-252.
[6] Cornell (1995) *Journal of American Chemical Society*, 117, 5179, 519.
[7] AMBER software package*, The Amber Molecular Dynamics Pacage.htm*.
[8] Brooks B.R. (1983) *Journal of Computational Chemistry*, 4, 187-217.
[9] Molecular Simulation, *http://www.msi.com, http:// www.accelrys.com.*
[10] Tripos, *http://www.tripos.com*.
[11] Browne W.J (1969) *Journal of Molecular Biology*, 42, 65-86.
[12] Sali (1993) *Journal of Molecular Biology*, 234, 779-815.
[13] Sanchez (1997) *Current Opinion in Structural Biology*, 7, 206-214.
[14] Gerstein (1997) T*he National Academy of Sciences*, 94, 11911-11916.
[15] Fischer (1999) *Current Opinion in Structural Biology*, 9, 208-211.
[16] Havel (1999) *Journal of Molecular Biology*, 217, 1-7.
[17] Peitsch (1996) *Biochemical Society Transactions*, 24, 274-279.
[18] Guex (1999) *Trends for Biochemical Sciences*, 24, 364- 367.
[19] SWISS- MODEL *SWISS-MODEL.htm*.
[20] Laskowski (1993) *Journal of Applied Crystallography*, 26, 283-291.
[21] Sippl (1993) *Proteins*, 17, 355- 362.
[22] Luthy (1992) *Nature*, 356, 83-85.
[23] Melo (1998) *Journal of Molecular Biology*, 277, 1141-1152.
[24] Bitomsky (1999) *Journal of American Chemical Society*, 121,3004-3013.