CHAPTER 6

INTRODUCTION TO MOLECULAR MODELING AND PHARMACOPHORE MODELING

6.1 Introduction

Molecular modeling, also referred to as computational chemistry, is a set of techniques for investigating chemical problems on a computer [6.1–6.6]. Molecular modeling consists of a range of computerized techniques based on theoretical chemistry and experimental data that are used either to analyze molecules and molecular systems or to predict molecular, chemical, and biochemical properties. This chapter is intended to provide an overview of some of the techniques and methods used in molecular modeling. Later illustrated is the use of these techniques in the field of rational drug design, followed by a brief discussion on pharmacophore modeling.

The practice of molecular modeling started in the late 1960s. Since then it has been a useful technique to understand the complexity of the molecular world through the scalability and accuracy of computation. The main objective of molecular modeling is to: (a) Extract results for a particular model; (b) Compare experimental results of the system; (c) Compare theoretical predictions for the model; (d) Help to understand and interpret experimental observations; (e) Correlate microscopic details at the atomic and molecular level with macroscopic properties; (f) Provide information not available from real experiments.

6.2 Tools for molecular modeling

Most molecular modeling studies involve three stages. The first stage is model parameterization, wherein the intra- and inter-molecular interaction of the molecule is described on the basis of interaction energy, taking into account molecular position and/or arrangements in 3D space. The second stage is the memory crunching calculation involving energy minimization, molecular dynamics (MD), Monte Carlo simulation (MC), or a conformational search. The final step is the result analysis. Above all the accuracy of all these calculations depends on the modeler’s decision to choose the correct method for

problem solving and in-depth knowledge of the “pros and cons” of the several molecular modeling tools. The most common tools used in the various stages of molecular modeling are, (a) *Ab initio* calculations, (b) Semiempirical methods, (c) Density functional theory, (d) Molecular mechanics, (e) Quantum mechanics/Molecular mechanics (QM/MM), (e) MD and MC simulations.

6.2.1 *Quantum mechanics*

Quantum mechanics (QM) is the correct mathematical description of the behavior of electrons. QM predicts exactly the properties of atoms and molecules. So far QM equations have only been solved exactly for one electron systems. QM deals with the motion of electrons under the influence of the electromagnetic force exerted by nuclear charges. The accurate understanding of the electronic behavior in the molecule helps in turn to understand molecular structure and the nature of chemical reactions. In practice, quantum chemistry is represented through *ab initio*, semiempirical methods and density functional theory. There are at least four important areas of drug design where quantum mechanical calculations can be applied: charge calculations, molecular electrostatic potential, parameter development for molecular mechanics and modeling chemical reactions and design of transition-state inhibitors [6.7].

6.2.1.1 *Ab initio*

The term *ab initio* is assigned to computations that are derived directly from theoretical principles with no inclusion of experimental data [6.8–6.11]. The most common type of *ab initio* calculation is called a Hartree–Fock calculation (abbreviated HF). The disadvantage of *ab initio* methods is that they are expensive. These methods often take enormous amounts of computer CPU time, memory, and disk space.

6.2.1.2 *Semiempirical methods*

Semiempirical methods are less strict compared to *ab initio* methods, because certain pieces of information are approximated or completely omitted [6.8, 6.11–6.14]. The core electrons are not included in the calculation and only a minimal basis set is used; further the two-electron integrals are not considered. Such parameter reduction introduces errors into the semiempirical methods. These errors are removed by further parameterization by estimating the omitted values. These values are obtained by fitting the results to

experimental data or *ab initio* calculations. The advantage of semiempirical calculations is that they are much faster than *ab initio* calculations. The disadvantage of semiempirical calculations is that the results can be erratic and fewer properties can be predicted reliably. If the molecule being computed is similar to molecules in the database used to parameterize the method, then the results may be very good. If the molecule being computed is significantly different from anything in the parameterization set, the answers may be very poor. Listed are some of the semiemperical methods practiced today, HÜCKEL, EXTENDED HÜCKEL, PPP, CNDO, MINDO, MNDO, INDO, ZINDO, SINDO1, PRDDO, AM1, PM3, PM3/TM, FENSKE–HALL, TNDO, SAM1, GAUSSIAN THEORY.

6.2.1.3 *Density functional theory*

Density functional theory (DFT) which was earlier confined to small molecules is now being introduced to macromolecules as well [6.7, 6.14, 6.15]. In DFT the energy of a molecule is determined from the electron density instead of a wave function. This formally reduces the many-body electronic problem to three coordinates only. This theory originated with a theorem by Hohenburg and Kohn to find out the ground-state electronic energy of a molecule. Later Kohn and Sham developed the practical application of this theory. Some of the more widely used functional are, Xα, HFS, VWN, BLYP, B3LYP, Becke3LYP, PW91, G96, P86, B96, B3P86, B3PW91.

6.2.2 *Molecular mechanics*

Molecular mechanics (MM) [6.16–6.18] is based on a mathematical model of a molecule as a collection of balls (corresponding to the atoms) held together by springs (corresponding to the bonds). The principle of Hooke's law is used to describe the ability of bonds to stretch, bend, and twist. The MM model ignores electrons. Therefore MM can be applied to molecules containing many atoms.

There are two types of interactions observed in atoms, bonded and non-bonded. The interactions are described on the basis of potential energy functions. In MM the minimum-energy geometry of a molecule is obtained through mechanical models. This is carried out by the use of a set of parameters that are empirically derived (usually termed as force-constant) and this collection of parameters is known as the force field [6.19, 6.20]. The term arises because the negative of the first derivative of the potential energy of a particle with respect to displacement along some direction is the force on the particle; a "force field"

E(*x*, *y*, *z* coordinates of atoms) can be differentiated to give the force on each atom. Many different force fields have been designed for various type of molecule. They have much in common, but they also have many differences. Some commonly used force fields are, AMBER, CHARMM, CFF, CHEAT, DREIDING, ECFF, EFF, GROMOS, MM1, MM2, MM3, MM4, MMFF, MOMEC, OPLS, Tripos, UFF, YETI.

6.2.3 *QM/MM*

Quantum mechanics/Molecular mechanics (QM/MM) is a hybrid method which encompasses both QM and MM [6.7, 6.20, 6.21]. This calculation is designed to give results that have very good speed when only one region needs to be modeled. The earliest QM/MM calculations were done simply by modeling different parts of the system with different techniques. For example, some crucial parts of the system could be modeled by using *ab initio* geometry-optimized calculations. The complete system could then be modeled using MM, by holding the geometry of the initial region fixed and optimizing the rest of the molecule. Sometimes the QM/MM technique uses an *ab initio* method to parameterize force field terms specific to a single system. Another variant of QM/MM is the energy subtraction method. In this method, calculations are done on various regions of the molecule with various levels of theory. Then the energies are added and subtracted to give suitable corrections. This results in computing energy for the correct number of atoms and bonds analogous to an isodesmic strategy. Bersuker and co-workers have proposed a technique called self-consistent method, whereby the atoms on the boundary between regions are included in both calculations. In this procedure, optimizations are done with each method, using the boundary atom charge from the other method, and this is repeated until the geometry is consistent between the levels of theory. The QM/MM methods are now routinely practiced for macromolecules.

6.3 Simulation tools

Simulation tools are very useful in elucidating biological intricacies. At present, simulations are carried out by statistical mechanics, MD or MC simulations. Statistical mechanics is the mathematical means to calculate the thermodynamic properties of bulk materials from a molecular description of the materials. Statistical mechanics provides a means for determining physical properties that are associated with the macroscopic sample of the bulk

liquid, solid, and so on. The gathering of information about possible energy levels, and about conformations of a macroscopic sample with reference to time and position is difficult through statistical mechanics. These difficulties are overcome by associating statistical mechanics with MD and MC simulations.

MD is a simulation of the time-dependent behavior of a molecular system, such as vibrational motion or Brownian motion [6.22–6.24]. It requires a way to compute the energy of the system, most often using a molecular mechanics calculation. This energy expression is used to compute the forces on the atoms for any given geometry. In MC simulation the location, orientation, and geometry of a molecule or collection of molecules are chosen according to a statistical distribution [6.25, 6.26]. MC methods are built around random sampling, which is simulated with a random-number-generating algorithm. For example, many possible conformations of a molecule could be examined by choosing the conformation angles randomly. If enough iterations are done and the results are weighted by a Boltzmann distribution, this gives a statistically valid result.

In summary, there exist several molecular modeling tools to understand the complexity and diversity of molecular system on the basis of theoretical calculations. Very often these tools are used in a variety of applications, such as material science research and rational drug design. The following section deals with some molecular modeling applications in rational drug design.

6.4 Rational drug design another face of molecular modeling

Today almost all pharmaceutical companies use molecular modeling as an integral part of their research and development programs [6.27]. According to a recent report, it costs a minimum about $600 million to $800 million and 12-15 years to bring a compound from the identification stage to the market [6.28]. Pharmaceutical companies are hard-pressed and so take a multitude of computer aided drug design (CADD) or rational drug design (RDD) approaches to shorten the time and reduce the cost of identifying new chemical entities (NCEs). The scalability and accuracy achieved by various molecular modeling tools mentioned above have further strengthened this process. Various avenue of RDD include quantitative structure activity relationship (QSAR), pharmacophore modeling, docking and scoring functions, *de novo* ligand design, molecular dynamics, virtual screening (VS),

ADMET predictions and chemoinformatics. In this context, pharmacophore modeling supplemented with virtual screening is an attractive research proposition in RDD.

6.4.1 *Quantitative structure activity relationship*

The functional groups present in a molecule determine the biological activities. For medicinal chemists, structure activity relationships (SAR) are the basis of synthesizing molecules with desired properties. Given a collection of compounds and their biological activities, it is possible to derive a statistical model which quantitatively establishes the structure–activity relationship. The importance of the functional groups present in molecules is described through chemical descriptors. QSAR is an attempt to correlate the structural features of a series of known compounds with their biological activities with the help of chemical descriptors [6.29, 6.30]. The QSAR model thus generated is used to predict the biological activity of novel inhibitors pertaining to the similar series of compounds.

6.4.2 *Docking and scoring functions*

The receptor–ligand complexation is the resultant of perfect complementarities of the steric and electrostatic factors between receptor and ligand. The task of predicting a favorable binding pose (lowest free energy binding mode) between receptor and ligand through computation by incorporating the essential ingredients of molecular recognition to molecular modeling techniques is the essence of docking. This is carried out through energy optimization process [6.31–6.34]. This is analogous to the docking of a ship in the harbor. In the present context the receptor is the harbor and ligand is the ship.

The scoring function used in docking is empirical and a rule based parameterized mathematical function is used to rank docking solutions based on the binding affinities of drug molecules to a receptor [6.35, 6.36]. The scoring functions are deemed to establish a linear correlation between the biological activity and the binding affinity of a diverse set of drugs for any particular target.

6.4.3 *De Novo ligand design*

Recent progress in structural biology has shown new light to many structural intricacies in molecular recognition [6.37]. By analyzing the active sites in terms of functional groups, the intuition of a complementary fragment very often comes to the mind of the medicinal chemist. This intuition of building complementary fragments *in silico* is the

basis of *de novo* ligand design. The fist step in *de novo* design is molecular fragment fitting in active sites followed by growth of these fragments to build a complete molecule [6.38]. This exercise leads to the design of a focused virtual compounds library after several iterations. However, the virtual compounds generated in this method are often not accepted due to difficulties in their synthetic feasibilities. Therefore the problem of synthetic feasibilities is assessed by sophisticated software before attempting compound synthesis.

6.4.4 *Virtual screening*

The method of screening large compound database *in silico* is termed as VS [6.39-6.44]. The VS method deploys a combination of known RDD techniques to generate first, a focused library of compounds and then selecting possible lead molecules *in silico* after subsequent property filtering. VS protocols include ligand based screens like 1D filters (e.g. molecular weight), 2D filters (similarity, substructure fingerprints) [6.43], 3D filters (3D-pharmacophore [6.40], 3D shape matching) and structure based screens like docking. The VS method has been implemented in many therapeutic targets [6.44].

6.4.5 *ADMET prediction*

The final lap of the successful journey of a drug molecule depends on its Absorption, Distribution, Metabolism, Elimination, and Toxicity (ADMET) profiles. In most cases, the success of drug discovery projects depends particularly on better ADMET profile of a drug. Therefore the majority of funds in drug discovery research is diverted for designing improved ADMET profile of a drug molecule. This is the reason that ADMET property evaluation has been incorporated into the drug design strategies today [6.45, 6.46]. The ADMET model estimates physicochemical properties of compounds like ionizability (*pKa*), and lipophilicity (*log P* or *log D*). These estimates provide a reasonable answer to ADMET profile of a given compound.

6.4.6 *Chemoinformatics*

Chemoinformatics is the application of informatics methods to solve chemical problems [6.47–6.50]. Over the years chemistry has generated large amount of digital information. The prime objective of chemoinformatics is to represent chemical object digitally, to manage chemical data, elucidation of structure, the design chemical reaction and estimate synthetic feasibility. Above all, the niche of chemoinformatics is extended to

molecular modeling and RDD methods. Hence chemoinformatics is widely used in drug research.

6.5 Pharmacophore modeling

Pharmacophore, as defined by an IUPAC working party and published in 1998, is "*the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response"* [6.51]. As a consequence the pharmacophore: (1) describes the essential, steric and electronic, function-determining points necessary for an optimal interaction with a relevant pharmacological target. (2) does not represent a real molecule or a real association of functional groups, but it is rather a purely abstract concept that accounts for the common molecular interaction capacities of a group of compounds towards their target structure. (3) is not just a specific functional group or "pieces of molecules" [6.52, 6.53].

6.5.1 *Historical perspectives*

The structure activity relationship is the basis of pharmacophore modeling. One of the earliest to recognize structure–activity relationships was Robert Boyle in 1685, who tried to explain the specific effects of drugs in terms of mechanical philosophy [6.54]. Later, at the turn of the 20th century, the German scientist Sigmund Fränkel discussed the pharmacological action and selectivity of drug molecule in cells and tissues on the basis of the presence of certain groups in the drug molecule [6.55]. The inadequacy in the understanding of molecular worlds in the beginning of 20th century laid several controversies on the pharmacological action and the physiological activity of drugs. However, the discovery of the pharmacological receptor concept in the later part of the century confirmed the structure–activity relationships in drug molecules. The initial hypothesis on receptor–drug interaction was put forth by Langley in 1878 [6.56], who introduced the term "receptive substance" [6.57]. However, the word "receptor" was introduced later, by Paul Ehrlich [6.58, 6.59]. During the first half of the 20th century, several observations highlighted the critical features associated with the concept of receptors [6.60]. The selective interaction of drug to their respective receptor was first demonstrated by Paul Ehrlich's through the discovery of salvarsan. This gave rise to the concept of a

chemotherapeutic “magic bullet” against specific infectious organisms. The work on enzyme and glucoside interaction by Emil Fischer later in 1894 further strengthened Ehrlich’s seminal discoveries of receptor hypothesis [6.61]. He proposed the famous “lock and key” hypothesis for enzyme substrate recognition. To extend this phenomenon in the present day context, substrate and receptor should have a greater degree of steric, electrostatic and molecular flexibility complementarities for better fit. With this background it is clear that the pharmacological action and the physiological activity of a drug molecule are hidden in the structure. Thus this provided food for thought to many to quantify and correlate structure–activity relationships in drug molecules through experiments and modeling.

6.5.1.1 *Pharmacophore models in the early years*

The pharmacophores were described in the literature before the advent of CADD. Due to the knowledge of the bond lengths and the van der Waals sizes in the 1940s, initial structure–activity relationships were established through the construction of simple two-dimensional model structures [6.62–6.65]. With the availability of X-ray analysis and conformational chemistry, the three-dimensional models became possible in the 1960s [6.66–6.68]. In the early 1970s, three-dimensional depictions of pharmacophore models were built for “Clonidine and Its Interaction with the *a*-Adrenergic Receptor” [6.69–6.72]. This knowledge was later incorporated to model building using computers.

6.5.1.2 *Pharmacophore modeling after the use of computers*

The advent of the computer enabled scientists to quantify structure–activity more efficiently. The initial step in this direction was carried out by Korolkovas, Gund [6.73, 6.74] and by Humbelt and Marshall [6.75]. Immediately this application found its ground in the field of medicinal chemistry to correlate the effect of structure on bioactivity. These events later incorporated the ingredients of chemical descriptor to enhance the accuracy of pharmacophore modeling. With the help of computer programs the pharmacophoric patterns in molecules were identified. In this regard the program developed at Princeton University in early 1970s [6.76, 6.77] is noteworthy. This program presented the topographic patterns of molecules based on distance measurement through Dreiding models. The distances were derived from a knowledge of small molecule crystal structures. The new era of

pharmacophore modeling began thereafter with the help of many sophisticated computers [6.78, 6.79].

6.5.2 *Pharmacophore modeling methods*

Broadly, pharmacophore methods can be classified into analog and structure based pharmacophore modeling [6.52]. The analog based pharmacophore modeling gathers information from already existing drugs/ligands that are active against target biological molecules (protein or DNA/RNA) of interest. Based on this information, a set of pharmacophore hypotheses is generated to compare biological activities between actives and in-actives and to search for lead molecules in chemical database. In this approach, the relative importance of different functionalities in drug receptor recognition is first determined; the hypothesis is next generated based on the correspondence between functional groups in congeneric series of drug molecules. Examples of analog based pharmacophore modeling package are DISCO, HipHop, GASP, HypoGen, PHASE. In the structure based approach the pharmacophore hypothesis is generated from the three-dimensional structure of the receptor. The method involves analyzing the active site to generate an interaction map of desirable features that a ligand should satisfy to adequately interact with the receptor. The features include various non-covalent interactions observed in receptor–ligand complexes. The generated hypotheses are later submitted as queries in chemical databases. The available software to carry out structure based pharmacophore modeling are structure based focusing in Cerius^2 and LigandScout, and pharmacophore modeling in MOE.

6.5.3 *Application of pharmacophore modeling in rational drug design*

Pharmacophore models can be used in different ways in drug design programs: (1) as a 3D query tool in virtual screening to identify potential new compounds from 3D databases; (2) to predict the activities of a set of new compounds yet to be synthesized; (3) to filter suitable docked poses implemented in the docking exercise; (4) to understand the possible mechanisms of action; (5) to generate a focused virtual library. The pharmacophore generation approach is quite powerful and finds many applications in drug discovery research.

6.6 As we move on to following chapters

The following two chapters deal with the pharmacophore modeling study on two therapeutically important targets. The studies are based on an amalgamation of techniques used in molecular modeling and drug design. These studies have a potential role in medicinal chemistry and drug design.