Chapter 6
Molecular Modeling

**Dr. Preethi Badrinarayan, Chinmayee Choudhury
and Prof. G. Narahari Sastry**

Abstract Recent advances in theoretical methods based on quantum mechanics and classical mechanics with visualization tools have played a very important role in chemistry and biology. Further, computers have a profound influence on the way we do science in the last few decades. The current chapter provides a preliminary exposure to a range of molecular modeling approaches applicable to small to medium sized molecules to proteins. Recent advances in the theoretical and computational methodologies which are aimed to treat large molecules are described. Emphasis is given on methods of computer aided drug design. These are preceded by a simple introduction to quantum mechanics, classical mechanics and molecular dynamics. A major area in modelling biomolecules is a proper quantitative treatment of non-bonded interactions. The importance of understanding various non-bonded interactions is highlighted. The role of these non-bonded interactions which determines the biological structure and functions is described. Thus, the current chapter provides a brief overview of computational methods applied to biomolecules.

Keywords Quantum mechanics · Molecular dynamics · QM/MM · Coarse grained simulations · Computer aided molecular design · Dokcing · Quantitative structure activity relationships · Virtual screening · Multi-scale modeling

List of Abbreviations

ANN	Partial least square, artificial neural network
CADD	Computer aided drug design
DFT	Density function theory
FBDD	Fragment based drug design
FEP	Free energy perturbation
F-value	Fisher statistic
FTA	Fragment tailoring approach

Prof. G. N. Sastry (✉) · Dr. P. Badrinarayan · C. Choudhury
Centre for Molecular Modeling, CSIR-Indian Institute of Chemical Technology, Tarnaka,
Hyderabad 500 607, India
e-mail: gnsastry@gmail.com

GFA	Genetic function approximation
HTS	High-throughput screening
Ki	Inhibitory constant
LBVS	Ligand-based virtual screening
MCSCF	Multi-configurational self-consistent field
MD	Molecular dynamics
MLR	Multi-linear regression
MM	Molecular mechanics
PCR	Principal component regression
PDE	Phosphodiesterase
QM	Quantum mechanics
QM/MM	Quantum mechanics/molecular mechanics
QSAR	Quantity structure activity relationship
SBVS	Structure-based virtual screening
SCF	Self-consistent field
TI	Thermodynamic integration
VS	Virtual screening

6.1 Introduction

Models are essential to comprehend the incomprehensible reality. While biology and chemistry are empirical sciences, which are largely based on observations and observing, it has become necessary to develop the understanding at atomistic level. Therefore, modeling molecules has become an indispensable tool to not only complement the structural and spectroscopic attempts to characterise molecules, but also to provide the basic tool for imaging molecular action. Thus, tools for molecular visualization on one hand and computational approaches based on rigorous theory have occupied central stage in the computer age. When dealing with small molecules, the objective has been to obtain highly reliable properties often comparable with those of experiments. Quantum mechanical treatment of atoms and molecules thus provide the most fundamental dimensions to treat small molecules (Cramer 2004; Leach 2001; Hinchliffe 2003). In natural sciences there are different kinds of theories. While theory has been a strong and integral part of physics, the science of chemistry has grown with experimentation and biology with observations. In addition to theory and experiment, the entry of computations has provided another dimension to pursue science. While the *ab initio* computational approaches have the ability to model and obtain every possible experimental property, their applicability becomes very limited as the size of the system increases. Consequently, for macro-molecules, the application of quantum mechanical approaches is severely limited. When the time dependency in solving the Schrödinger wave equation is considered, its application is restricted still further. Also the time scales that can be probed using *ab initio* molecular dynamics (MD) are very small. Thus application of quantum chemical methods based on *ab initio* theory is practically limited to systems with very limited length and
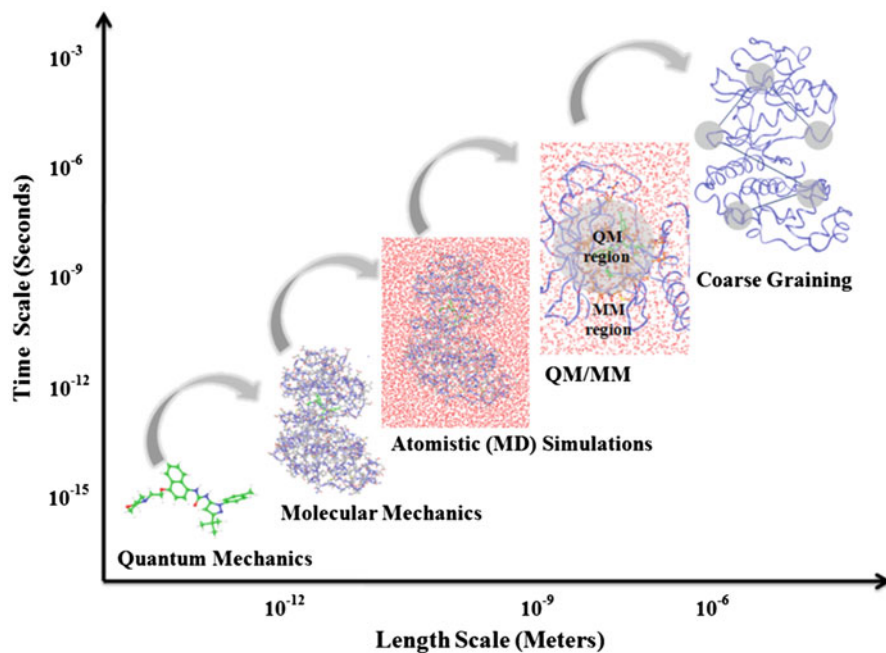
Fig. 6.1 Hierarchical order of molecular modeling approaches at different time and length scales. The figure depicts typical systems and methods which can be useful for varying time and length scales

time scales. Therefore, one needs to resort to methods based on classical mechanics to employ computational methods with larger length and time scales (Fig. 6.1).

In addition to the quantitative theories, many qualitative theories have emerged out of a large body of data obtained from experiments and observations. These methods based on informatics have been extremely successful in analysing the massive data of protein and nucleic acid sequences and thus ushering the area of bioinformatics. Most of the data in the area of bioinformatics is experimental. However, in the area of small molecules, computations have played a very important role and produced a large body of data of high reliability obtained by employing computational tools. Therefore, molecular modeling and computation have been complementary to the experimental efforts as far as the small molecules are concerned.

In this chapter, we introduce various computational methods which may be applied to molecular systems with varying length and time scales. Further, we discuss the basic principles of structure and analogue based approaches briefly. Finally, a cursory outlook on the application of these methods on biomolecules and materials has been given.

6.2 Quantum Mechanics

Computational chemistry is aimed at theoretically determining the properties of the molecules based on quantum chemical or classical mechanical equations of motion. Quantum chemical approaches are needed to accurately model the systems at atomistic scale and more importantly for obtaining electronic structure information. According to quantum mechanics (QM), all possible information on a molecular system can be obtained from a wavefunction, ψ which is obtained by solving the Schrödinger wave equation. However, the Schrödinger wave equation can be solved only for one electron systems thus rendering it unsolvable for many electron systems. The Schrödinger equation is the fundamental equation in QM and provides the basis for providing a complete description of the electronic structure of a molecule. Due to the difficulty associated with solving the Schrödinger equation for many electronic systems a large number of approximations were provided. There are excellent treatises available in the literature on computational QM, which deals with electronic structure calculations based on either *ab initio* molecular orbital theory or density functional theory (Cramer 2004; Jensen 2007; Levine 2013). Thus, we refrain from providing any further details on this section for two reasons. The first being the limitation of the space and the second and most important one is the limitation of the applicability of quantum chemical approaches to large biomolecules in general. However, in the following sections we discuss about the hybrid methods and multiscale modelling approaches, which employ and also largely based on the principle of quantum mechanics.

Herein we provide a cursory look at the various approximations that are being employed to solve the Schrödinger wave equation for medium sized molecules. The primary approximation for most of the electronic structure calculations is the Born-Oppenheimer approximation which essentially decouples the nuclear and electronic part of the kinetic energy operator. Following this variation theorem has become extremely effective for setting a lower bound for the energy. Such a condition has played a very important role in getting a better wavefunction through iterative procedures. The second important approximation is perturbation theory truncated to second, third or higher orders. However, the most effective theory based on electronic structure methods is *ab initio* self-consistent field (SCF) theory, where the fundamental level of reference wavefunction for the single determinant wavefunction is obtained by using the Hartree-Fock method. Electron correlation, which in principle may be divided into static and dynamic, is one of the most important parameter which needs to be included for the accurate description of the wavefunction. Thus, methods which go beyond Hartree-Fock level were warranted to obtain reliable properties of the molecular systems. The most popular variants of these methods, where a single Slater determinant can reasonably describe the system, are based on Moller-Plesset perturbation theory, configuration interaction and couple cluster methods. However, for open shell systems, the non-dynamic electron correlation becomes important, and one needs to have more than one Slater determinant for the reference wavefunctions. In such conditions the multi-configurational self-consistent field (MCSCF) procedures become imminent.

Methods based on these approximations have become very popular and have contributed greatly to the understanding of molecular structure, function and property relationships. The most rigorous method among these is based on the *ab initio* molecular orbital theory. One of the main bottlenecks in the application of the rigorous *ab intio* calculations to large molecules is computational capacity. In order to overcome that, several economical semiempirical SCF methods have emerged. However, the recent advances in the density functional theory have become very effective in dealing with the medium to large biomolecules.

6.3   Molecular Mechanics

Molecular mechanics (MM) applies Newtonian mechanics on a molecule or a molecular system to model its detailed structure and physical properties by calculating the energy of a molecule in terms of the bonded and non bonded interactions. MM is useful to study a broad range of molecular systems starting from small molecules to large biological systems or material assemblies with many thousands to millions of atoms (Field et al. 2007). The atomistic MM methods treat molecules as balls joined by springs wherein each atom is a single particle with an assigned radius (typically the van der Waals radius), polarizability, constant net charge (generally derived from quantum calculations and experiment). The bonded interactions are treated as springs with an equilibrium distance equal to the experimental or calculated bond length. All these bonded and non bonded terms all together are represented as a functional abstraction or force field to calculate the potential energy of a molecular system in a given conformation.

6.3.1   Energy of a Molecule

The steric energy of a molecule is the energy due to the geometry of a molecule. By nature, a molecule always tends to be in its lowest energy conformation to attain stability. As stated earlier, MM assumes the steric energy of a molecule to arise from a few, specific interactions within a molecule. These interactions include the stretching or compressing of bonds beyond their equilibrium lengths and angles, torsional effects of twisting about single bonds, the van der Waals attractions or repulsions of atoms that come close together, and the electrostatic interactions between partial charges in a molecule due to polar bonds (Hirschfelder 1954). To quantify the contribution of each, these interactions can be modeled by a potential function that gives the energy of the interaction as a function of distance, angle, or charge. The total steric energy of a molecule can be written as a sum of the energies of the interactions:

$$E = E_{bonded} + E_{non-bonded} \tag{6.1}$$

$$E_{bonded} = E_{stretch} + E_{bend} + E_{stretch-bend} + E_{dihedral} + E_{improper} \tag{6.2}$$

Table 6.1 Bonded and non-bonded energy components

S. No.	Energy component	Energy function	Constants and variables
	Bonded		
1	Stretch	$E_{str}(r) = \sum\limits_{bonds} \frac{1}{2}k_r(r - r_0)^2$	k_r = force constant for the bond, r = actual bond length between the two atoms defining the bond and r_0 = equilibrium distance for the bond
2	Bend	$E_{bend}(\Theta) = \sum\limits_{angles} \frac{1}{2}k_\Theta(\Theta - \Theta_0)^2$	Θ = bond angle, Θ_0 = equilibrium angle, k_Θ = force constant
3	Strech-bend	$E_{str-bend}\sum\limits_{bonds,angles} = \frac{1}{2}k_{sb}(r - r_0)(\Theta - -\Theta_0)$	k_{sb}= stretch-bend force constant
4	Dihedral	$E_{dihedral} = \sum\limits_{dihedrals} \frac{1}{2}V_n[1 + \cos(n\varphi - \delta)]$	n = periodicity of the angle, δ = phase of the angle, V_n = force constant
5	Improper	$E_{improper}(\omega) = \sum\limits_{improper} \frac{1}{2}k_\omega(\omega - \omega_0)^2$	k_ω and ω_0 = force constant
	Non-bonded		
6	Electrostatic	$E_{elec} = \frac{q_i q_j}{r_{ij}}$	q_i and q_j = partial atomic charges for atoms i and j separated by a distance r_{ij}
7	van der Waals	$E_{LJ} = \sum \frac{A_{ij}}{r_{ij}{}^{12}} - \frac{B_{ij}}{r_{ij}{}^6}$	Aij and Bij are positive constants whose values depend on the depth of the Lennard-Jones well Ïµij

$$E_{non-bonded} = E_{electrostatic} + E_{van\ der\ Waals} \qquad (6.3)$$

The bond stretching, bending, torsion and improper interactions are called bonded interactions because the atoms involved must be directly bonded or bonded to a common atom. The van der Waals and electrostatic interactions are between non-bonded atoms (Table 6.1).

6.3.2 The Force Fields

Force field refer to a mathematical function with a set of parameters (obtained experimentally as well as theoretically from computer intensive quantum calculations)
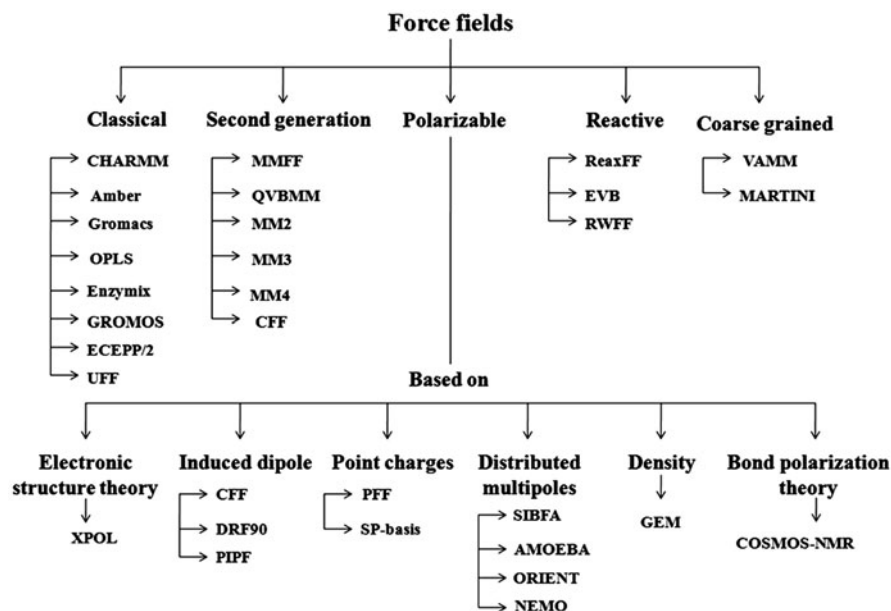
Force fields

Fig. 6.2 Examples of various types of fields

to represent the potential energy of a molecular system. There are various types of force fields depending upon the level of accuracy (Fig. 6.2). For example, the "coarse grained" force fields which are used to simulate large proteins provide a crude representation to save computational time, while the "all atom" force fields, although computationally expensive, can accurately treat even the terminal hydrogen atoms (Ponder 2003). The basic functional form of a typical force field is given by Eqs. 1–3 which has already been discussed. In this section we discuss the various classes of force fields.

Apart from a representative function for the potential energy, each force field has a set of parameters for each bonded and non-bonded terms. Also, each force field has a particular atom typing. For example, the parameters for an oxygen atom in a carbonyl group and in a hydroxyl group are given distinct parameters. The typical parameter set includes values for atomic mass, van der Waals radii, and partial charge for various atom types, and equilibrium values of bond lengths, bond angles, dihedral angles, impropers and also the spring constants associated with them. The parameters for given atom types are generally derived from observations on small organic molecules that are more tractable for experimental studies and quantum calculations and extrapolated for larger molecules like proteins and DNA.

6.4 Molecular Dynamics

Biological processes are complex and involve a repertoire of atomic interactions. Although experiments help to deduce the molecular level understanding of the biological processes, the atomic interactions need to be modelled computationally. Thus MD simulations are used to estimate the microscopic properties and dynamic motions of assemblies of a biomolecular structure. MD simulations provide an access to the thermally-accessible states and help to correlate them with the functions of biomolecular systems (Frenkel and Smit 2002). It is thus a method, which integrates the Newtonian equations of motion for 'N'particles of a system over a period of time resulting in a trajectory which is used for the calculation of the micro and macroscopic properties. The calculation of the MD trajectories is based on the principles of statistical mechanics (Allen and Tildesley 1987). MD simulations calculate the microscopic properties of the system such as position and velocities of each individual atom of the system. However, the properties that are of higher practical value are the macroscopic properties such as number of particles (N), volume (V), energy (E), temperature (T), pressure (P), chemical potential of particles μ (Rapaport 2004). These bulk properties are used to gauge the thermodynamic modulation with time.

The positions and momenta of all the particles of a system define a microscopic state. The positions and momenta of all the particles in the system are adjudicated as coordinates of a 6N dimensional space also called as phase space. Thus at any given time, the system corresponds to a point of the multidimensional space. The evolution of a system with time therefore corresponds to a trajectory in the phase space and can be determined by solving the equations of motion based on the potential energy (PE). There are different ways to distribute the total energy among the N particles of the system. An ensemble (EN) constitutes a collection of systems with similar macroscopic properties wherein each system corresponds to a point in the phase space. There are different types of ensembles based on the set of constant macroscopic properties such as canonical (NVT), the grand canonical (μVT), microcanonical (NVE) and the isothermal-isobaric (NPT) ensemble.

The partition function as given in Table 6.2 defines the microscopic state of a system explicitly. However due to the existence of large number of microscopic states in a biomolecular system and their sampling according to Boltzmann distribution in canonical ensemble, direct calculation of Z_{NVT} is not feasible. A MD simulation is initialized with the assignment of initial positions and velocities of all particles in the system. The initial velocities are assigned to particles in such a way that the total momentum is ensured to be zero, whereby the Maxwellian velocity distribution law is obeyed (Zeigler et al. 2000).

$$v^2{}_\alpha = \frac{\kappa_B T}{m} \tag{6.4}$$

With the initial velocities assigned, the next step is the calculation of potential energy of the system using Eqs. (1–3). This is followed by deducing the force acting on each particle of the system by differentiating the calculated energy with respect to

Table 6.2 Types of statistical ensembles used in MD simulations

S. No.	Ensemble	Features	Partition function	Remarks
1	Microcanonical	Constant number of particles (N), volume (V) and energy (E). Entropy of the system increases continuously	$\Omega(E)$	$\Omega(E)$ is the number of micro-states corresponding to the system's energy E
2	Canonical	Number of particles (N), volume (V) and temperature (T) are constant	$Z_{NVT} = \sum_{j=1}^{jmax} e^{-\beta PE(x)}$	$\beta = \frac{1}{K_B\,T}$, $PE(x) =$ energy of the xth microstate of the system
3	Grand canonical	Chemical potential or fugacity (μ), volume (V) and temperature (T) are constant	$Z_{\mu VT} = \sum_i e^{(N_i \mu - PE(x))}/T K_B$	$\beta = \frac{1}{K_B\,T}$, $PE(x) =$ energy of the xth microstate of the system

Table 6.3 Mathematical functions used to generate velocity and position at each time step (Δt) with Verlet and Verlet-like integrators. (Frenkel and Smit 2002; Fermann and Valeev 1997)

S. No.	Integrators	Equation to update new position (x)	Equation to update new velocity (v)
1	Verlet	$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{f(t)}{m}(\Delta t)^2$	$v(t) = \frac{x(t+\Delta t)-x(t-\Delta t)}{2x(t)} + \mathcal{O}(\Delta t)^2$
2	Velocity Verlet	$x(t+\Delta t) = x(t)+v(t)\Delta t+\frac{f(t)}{m}(\Delta t)^2$	$v(t + \Delta t) = v(t) + \frac{f(t+\Delta t)+f(t)}{2m}\Delta t$
3	Leap-Frog	$x(t + \Delta t) = x(t) - \Delta t v\left(t + \frac{\Delta t}{2}\right)$	$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \Delta t\,\frac{f(t)}{m}$
4	Velocity corrected Verlet	$x(t + \Delta t) = x(t) + v(t)\Delta t$ $\quad + \frac{4f(t)-f(t-\Delta t)}{6m}\Delta t^2$	$v(t + \Delta t) = v(t)+$ $\quad \frac{2f(t+\Delta t)+5f(t)-f(t-\Delta t)}{6m}\Delta t$

the atomic positions (Fig. 6.3). After calculating the force on each particle, Newton's laws of motions are integrated to generate new positions and velocities for specified time-steps (Table 6.3).

The behaviour of a system is determined by its thermodynamic properties such as free energy, entropy and enthalpy. In a biomolecular system, the formation of a protein-ligand complex involves a change in free energy (Reddy and Erion 2001). The free energy determines the equilibrium properties of a system and is usually considered as the Gibbs or Helmholtz free energy. The configurational Helmholtz free energy (*A*) and can be represented as follows for a canonical ensemble:

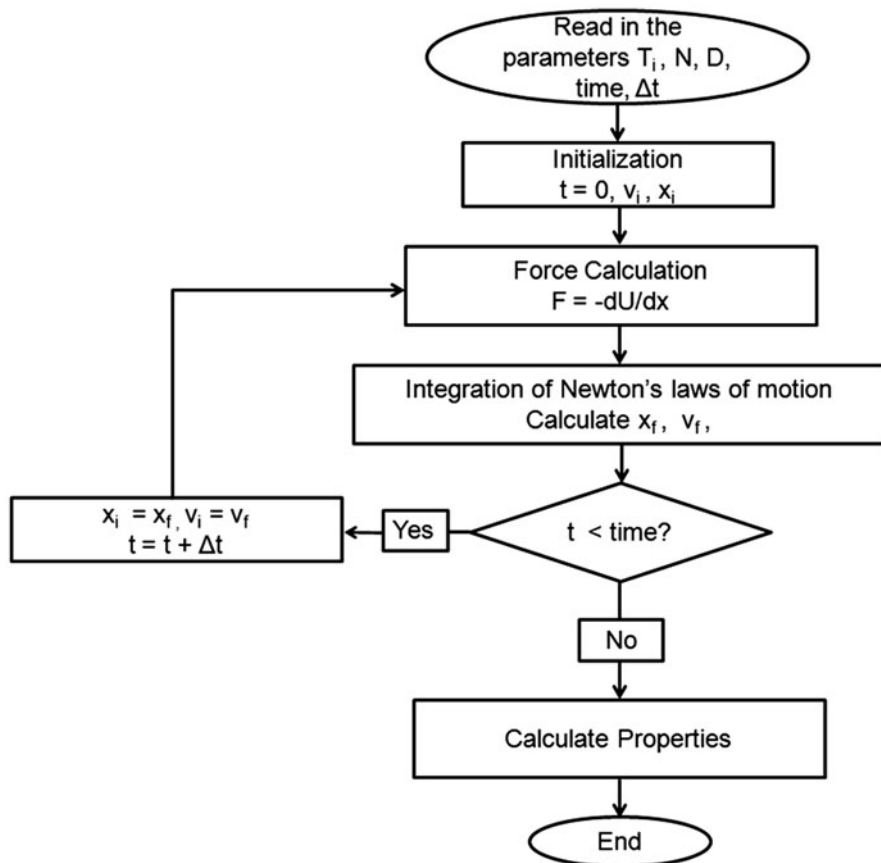$$A = -\beta^{-1} ln Z_{NVT} \tag{6.5}$$

Fig. 6.3 Flow-chart depicting the general steps of a typical MD simulation

Calculation of absolute free energy is difficult due to inadequate sampling. There is therefore a need for different methods to calculate free energy. The free energy simulation techniques aim at computing ratios of partition functions using various techniques. The most common methods include free energy perturbation, thermodynamic integration, umbrella sampling and potential of mean force. The free energy calculation methods thus calculate the ratio between the two partition functions Z_{a1} and Z_{a2} to obtain the difference in free energy (ΔA). The free energy is calculated between two states. Thus the free energy differences between states '$a1$' and '$a2$' with partition functions Z_{a1} and Z_{a2} respectively can be calculated as:

$$\Delta A = -\beta^{-1} ln \frac{Z_{a2}}{Z_{a1}} \tag{6.6}$$

The ΔA thus obtained, is used to calculate the binding affinity of a protein-ligand complex. The solvation-free energy or binding-free energy is usually calculated through alchemical transformations through addition or removal of ligand related energy terms from the total Hamiltonian. The calculation of free energy by two most widely used methods namely free energy perturbation (FEP) and thermodynamic integration (TI) have been described below.

Free energy is a state function therefore the difference in free energy in the two states can be represented as:

$$\Delta A = -\beta^{-1} \ln \langle \exp[-\beta(V_{a2} - V_{a1})] \rangle_{a1} \tag{6.7}$$

In the TI approach, the difference in free energy between two states is calculated by integrating over enthalpy changes between the transition states. These states can be described with a parameter λ wherein λ_0 and λ_1 represent states 0 and 1 respectively (Wang and McCammon 2012). Thus the difference in free energy between the two states 0 and 1 is (i.e. from λ_0 to λ_1) obtained by as

$$\Delta A = \int_{\lambda_0}^{\lambda_1} \frac{\partial V}{\partial \lambda}_\lambda \tag{6.8}$$

6.5   Computer Aided Molecular Design

Computational approaches have become an integral and indispensable part of both academia and industry. Deciphering of the human genome is one of the first definitive accomplishments towards the molecular level understanding of biology. This has provided a quantitative understanding of the structural and functional aspects of biology unraveling a multitude of disease targets for drug discovery (Hopkins and Groom 2002). New drugs are constantly required for improving the treatment of existing and the newly identified diseases, in addition to the production of safer drugs by the reduction or removal of adverse side effects. Consequently huge investments are being channeled from pharmaceutical industries in research and development activities. The interdisciplinary nature of drug discovery warrants a fruitful collaboration among chemists, biologists, pharmacologists, physicians, computational and informatics scientists etc. New lead design is now more a strategic than a serendipity driven process. Thus, in the last couple of decades, *in silico* approaches have become an integral part of essentially all rational drug discovery programs. The rational approaches in drug discovery are traditionally classified as structure and analogue based (Fig. 6.4).

Medicinal chemistry driven approaches for several decades have relayed on the analogue based approaches wherein finding the quantity activity relationship was the key. However the recent advances in structure and molecular biology have provided more fundamental insights at molecular level. These approaches have been applied to obtain insights on the binding characteristics of the drug with the target. A drug
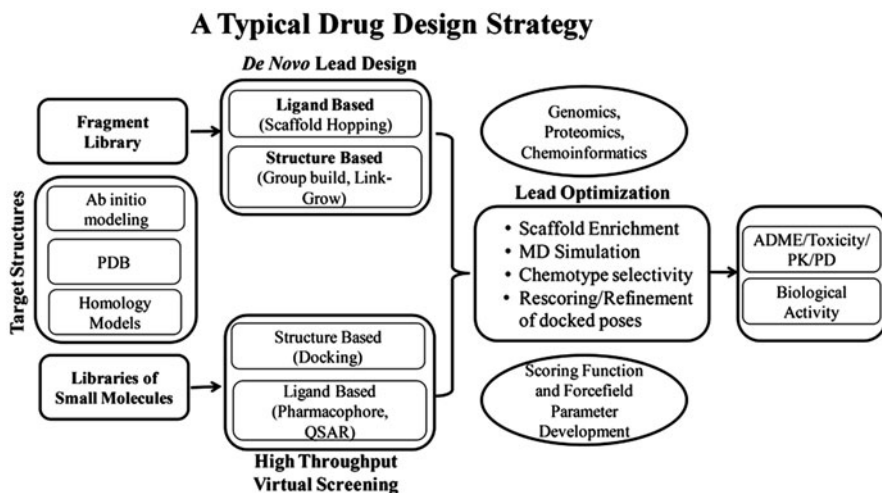
A Typical Drug Design Strategy

De Novo **Lead Design**

Fig. 6.4 A general work-flow of computer aided molecular design

contains various sub-units which contribute to the druglikeliness parameters such as the ADMET, pK/pD, blood brain barrier, drug metabolism, human intestinal absorption and permeability (Lipinski 1997). The long term use of a drug is however restricted by a multitude of inter-related factors such as development of resistance due to mutations, drug-drug interactions and most importantly target specificity. Rational design of a target specific drug is capable of overriding the other restricting factors. Specificity is of prime importance in the design of leads (Badrinarayan and Sastry 2013). There are different types of specificities such as target specificity, chemotype specificity and sub-type specificity. The different kinds of enzyme specificity that can be obtained by targeting different binding sites in a protein like the active site, allosteric sites. An inhibitor scaffold constitutes several fragments or chemotypes which individually can contribute to selectivity which can be used in the design of inhibitor for the active site to obtain high efficacy. This can possibly be achieved through molecule design or fragment based drug design (FBDD) using selectivity rendering fragments (Ringe and Reynolds 2010). The difference in the shape and constituency of the additional binding site called allosteric site can be exploited to design selective inhibitors for targets which share the same active site. Most proteins or enzymes exist in several isoforms which share high structural similarity among themselves. In such cases, small binding pockets or sub-pockets can be detected which are non-conserved and these can be used to obtain selectivity among the various subtypes.

Quantum chemical calculations based on *ab initio* have been proved to be highly reliable. However, the computational time rises exponentially as the number of electrons in the system increases thereby precluding their practical application to molecules with more than few dozens of atoms. Although it is possible to treat small

ligand sized molecules quantum mechanically, it is expensive to apply them for larger bimolecular systems like proteins and nucleic acids. Thus MM has become a definitive choice of application for biomolecular targets. Computer aided drug design (CADD) focuses on understanding three essential factors for the design of drugs namely the features which render a macromolecule druggable, the properties which distinguish a drug from a small molecule and the interactions which facilitate an optimal fit of a drug-like molecule into a druggable target. A disease target is one which plays a pivotal role in the cause and expression of a disease phenotype and can be modulated by a drug. The therapeutic targets are thus both disease modifying and druggable. At present the currently approved drugs interact with only 2 % of human proteins hence there still exists a repertoire of undiscovered targets (Hopkins and Groom 2002). CADD uses amalgamation of structure and analogue based approaches. The structure based approaches include homology modeling, docking, virtual screening (VS), MD, MM-PBSA/GBSA, free energy calculations (FEP, TI) while the analogue based approaches include quantity structure activity relationship (QSAR), pharmacophore mapping, toxicity prediction and chemoinformatics methods.

6.5.1   Structure Based Drug Design

Advances in sophisticated large-scale automation were expected to generate an unprecedented number of novel leads resulting in a substantial increase in novel drug entities to be launched in market every year. This could not materialize as the discovered hits failed to optimize into actual leads. Thus, the initial euphoria associated with these approaches has subsided owing to the significantly high costs and disappointingly low hit rates involved in high-throughput screening (HTS). This calls for the rational application of drug design approaches such as virtual screening or docking to obtain lead compounds, which can be optimized further as drugs.

6.5.1.1   Docking

Docking is carried out using an automated computer algorithm that determines the binding of a compound to the active site of a protein (Stahl and Rarey 2001). This includes determining the orientation of the compound, its conformational geometry, and the scoring. There are two key components of a docking program namely the search algorithm and the scoring algorithm (Table 6.4). The search algorithm positions molecules in a multitude of locations, orientations, and conformation within the active site (Young 2009). The identified orientations are sampled further through downhill minimization to obtain bioactive conformations. The choice of the search algorithm determines the thoroughness of the program in checking the possible positions of the molecule and time taken.

Table 6.4 Scoring functions based on different algorithms implemented in some of the popular docking softwares. (Friesner et al. 2004; Morris et al. 1998; Rarey et al. 1996; Jones et al. 1997)

Software	Search algorithm	Scoring function
GOLD	Genetic Algorithm	$GOLD_{Fitness\ Score} = HB_{ext} + 1.3750 * vdW_{ext} + HB_{int} + 1.0000 * Lig_{int};$ $\$Li\{\{g\}\_\{int\}\}=vd\{\{W\}\_\{int\}\}+\{torsion\}\$$
GLIDE	Systematic Conformational Search	SP : $\Delta G_{bind} = C_{lipo-lipo} \sum f(r_{lr}) + C_{hbond-neut-neut} \sum g(\Delta r)h(\Delta\alpha) + C_{hbond-neut-charged} \sum g(\Delta r)h(\Delta\alpha)$ $+ C_{hbond-charged-charged} \sum g(\Delta r)h(\Delta\alpha) + C_{max-metal-ion} \sum f(r_{lr}) + C_{rotb} H_{rotb}$ $+ C_{polar-phob} V_{polar-phob} + C_{coul} E_{coul} + C_{vdW} E_{vdW} + solvation\ terms$
		XP : $XP_{Glide\ Score} = E_{coul} + E_{vdW} + E_{bind} + E_{penalty}$
FlexX	Incremental Construction	$\Delta G = \Delta G_0 + \Delta G_{rot} N_{rot} + \Delta G_{hb} \sum_{neu\ Hbonds} f(\Delta R, \Delta\alpha) + \Delta G_{lipo} \sum_{lipo\ int} f^*(\Delta R, \Delta\alpha) + \Delta G_{ion} \sum_{ionic\ int} f(\Delta R, \Delta\alpha)$ $+ \Delta G_{aro} \sum_{aro\ int} f(\Delta R, \Delta\alpha)$
AutoDock	Lamarckian Genetic Algorithm	$\Delta G = \Delta G_{vdW} \sum_{ij} \left(\dfrac{A_{ij}}{r_{ij}^{12}} - \dfrac{B_{ij}}{r_{ij}^6} \right) + \Delta G_{hbond} \sum_{ij} E(t) \left(\dfrac{C_{ij}}{r_{ij}^{12}} - \dfrac{D_{ij}}{r_{ij}^{10}} \right)$ $+ \Delta G_{elec} \sum_{ij} \dfrac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + \Delta G_{tor} N_{tor} + \Delta G_{sol} \sum_{ij} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)}$

Orientations with closely placed atoms are scored and others discarded. The energetically favorable modes of binding of a ligand are stored as different poses. These poses representing the protein-ligand interactions are scored in terms of binding enthalpies or Gibbs free energies, or a qualitative numerical measure or from a potential of mean force equation. This is concomitant to the inhibitory constant (K_i) calculated experimentally. Most of the scoring functions correlate well with the K_i values whereas others provide a qualitative ranking of the compounds tested. Some programs retain all the poses generated whereas some provide a scrutinized list based on the scores. Evaluation of closely placed atom pairs using full force field equations consumes a lot of computer time (Table 6.5).

This is overcome by using a grid based algorithm wherein potential fields are created which can be numerically evaluated over a grid generated over the active site. At a given point, the value of the potential on the grid is equivalent to the energy required for placing a unit charge at that point. Thus different types of scoring functions such as knowledge based, empirical, force field and many more have been developed to gauge the strength of interactions between the receptor and the small molecules. Most of the components of these scoring functions predict the non-covalent interactions and they are hardly any accounting for the covalent interactions (Cross et al. 2009). Considering the lacuna in individual scoring functions, consensus scoring is in vogue. Although all the small molecules undergo a conformational change during docking, the protein is held rigid in a fixed geometry in majority of the cases. Some programs facilitate alteration in the conformation of the active site as in flexible docking. This takes a longer time therefore options such as side chain repositioning and scaling are introduced which results in an induced fit approach so as to mimic the physiological conditions as far as possible.

In addition to search algorithms, scoring functions and flexibility, solvation is a major issue in defining the accuracy of results obtained as it has a direct impact on the binding energies of dissimilar molecules. An algorithm deficient of a solvation term results in identification of charged ligands which are large in size. The free energy of interaction relative to the free energies in solution of two molecules determines the binding affinity of the ligand for a particular receptor. FEP techniques accurately calculate the relative free energy however these are time consuming and are biased for similar set of molecules and are impractical for application for screening huge datasets of relatively diverse molecules. Therefore, energy correction for the solvent surrounding the protein needs to be included rather than considering only those occupying the active site (Gohlke and Klebe 2002). This has made this method as a protocol of choice in both academia and industry prediction of binding mode of a ligand during lead optimization as well in the identification of the potent lead itself through virtual screening.

6.5.1.2 Virtual Screening

Drug here is treated as a chemical substance that is used to prevent or cure diseases. In ancient times, a wide range of natural products obtained from animal, vegetable and

Table 6.5 Force field parameters used in docking algorithms. (Leach [2001]; Ponder [2003]; Friesner et al. [2004]; Morris et al. [1998]; Rarey et al. [1996])

S. No.	Force Field	Energetics
1	CVFF	$E_{MMFF} = \sum_{i}^{3N} \frac{P_i^2}{2m_i} + \sum_{ij}^{nbonds} V_b(r_{ij}) + \sum_{ijk}^{nangles} V_a(\Theta_{ijk}) + \sum_{ijkl}^{ntorsions} V_t(\varphi_{ijk}) + \sum_{ij}^{npairs} V_{np}(r_{ij})$
2	Tripos	$E = \sum E_{str} + \sum E_{bend} + \sum E_{oop} + \sum E_{tors} + \sum E_{vdW} + \sum E_{elec} + \sum E_{multifit}$
3	MMFF94	$E_{MMFF} = \sum EB_{ij} + \sum EA_{ijk} + \sum EAB_{ijk} + \sum EOOP_{ijk;l} + \sum ET_{ijkl} + \sum EvdW_{ij} + \sum EQ_{ij}$
4	OPLS	$E(r^N) = E_{bond} + E_{angle} + E_{dih} + E_{db}$
5	CHARMM	$V = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\Theta(\Theta - \Theta_0)^2 + \sum_{dihedrals} k_\phi[1 + cos(n\phi - \delta)] + \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_u(u - u_0)^2$ $+ \sum_{nonbonded} \epsilon \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon r_{ij}}$
6	AMBER	$V(r^n) = \sum_{bonds} \frac{1}{2} k_b(l - l_0)^2 + \sum_{angles} \frac{1}{2} k_a(\Theta - \Theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n[1 + cos(n\omega - \gamma)]$ $+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \epsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}} \right\}$

mineral sources were used for medicinal purposes but with the increase in knowledge the focus is centered on the use of pharmaceutically active compounds as starting point for the development of drugs. However, the increase in chemical space has made the identification of lead a tedious and erudite process. Therefore computational based screening methods such as VS are used for the identification of these pharmaceutically active compounds from a core set of molecules (Badrinarayan and Sastry 2011; Reddy et al. 2007a). The approaches used can be classified as ligand-based and structure-based methods. The availability of the physicochemical information dictates what strategy is more likely to be applied. Existence of structural data of the target protein calls for the application of structure-based virtual screening (SBVS) strategies. In the absence of structural information, ligand-based virtual screening (LBVS) protocols are usually applied. Virtual screening is carried out in concord with a number of different tools such as informatics (chemo and bio), docking, QSAR, pharmacophore mapping, machine learning tools (MLT), fingerprints, quantum mechanics/molecular mechanics (QM/MM), QM etc.

LBVS functions on the similarity principle which considers that the molecules which are structurally similar have similar biological activities. In the absence of structural information, LBVS use similarity, QSAR and pharmacophore based methods to correlate the physiochemical properties of known ligands with their structural characteristics and generate a query. LBVS methods look out for desired patterns in molecules such as fragments, pharmacophore, and core scaffold through graph theory like approaches or use molecular descriptors. Molecular fingerprints are however emerging as the most sought after options due to the ease of handling and speed. Fingerprints can be formulated with both 2D and 3D features (Sastry et al. 2010). The fingerprints are defined in the form of vectors which constitute bit strings of ones and zeros or position vectors indicating the presence or absence of a particular feature. A large part of the LBVS work being done is driven by informatics wherein the similarity indices are used to scale the nature if identity between the query and the database molecules.

With the increase in the repertoire of crystal structure data and the efficiency of docking in deducing the binding modes of ligands, SBVS is still employed fervently. Given a 3D structure, the SBVS approaches employ docking to generate the binding modes of the database compounds which are then shortlisted based on their scores. VS is a multi-step protocol and with the advent of multi-drug resistant strains and cross-target reactions, each step of the screening process is embedded with filters for druglikeness, Lipinski, target-selectivity, toxicity-ADME etc. in order to garner the best of the lot at every step and curb the percentage of false positives (Klebe 2006). These filters vary in complexity and dimension (1D-3D). Time and precision are the two endearing factors of VS. The time required depends on the type of query used and the complexity of the databases (molecule or fragment) being screened. The query used for screening can be simple 2D co-ordinates in the form of SMILES, fingerprints, bit vectors or a complex 3D representation of the active site, ligand template, pharmacophore or surface maps. The precision of the endeavor relies greatly on the stringency of filters used. The query can be complex 3D active site,

molecule, pharmacophore, surface-volume or simple 2D co-ordinates, and feature-trees. SMILES, correlation vectors, bit strings and fingerprints are the simplest. The discrepancies in scoring functions are being encountered through incorporation of parameters extracted from MM-PBSA or QSAR calculations (Stahl and Bohm 1998). Although such procedures increase the accuracy of the hits obtained they however increase the overall time to get the desired outcome. The numbers of molecules that can be screened with virtual screening are several orders of magnitude higher than that of HTS. This number crunching ability of virtual screening and its inexpensive execution makes it endearing. However the crux of the process lies in the development of a screening strategy with efficient filters to obtain target specific leads.

We have developed a three step filtering strategy to identify target specific allosteric fragments for the inflammatory target p38 MAP kinase (Badrinarayan and Sastry 2012; Badrinarayan and Sastry 2010). The study entails the design of two target specific virtual screening filters based on docking score components and substructure interaction fingerprints. The components of the scoring function of two well-known docking protocols were evaluated to gauge their individual contribution in identification of lead. Eight thresholds were identified for the active and inactive conformations of kinase and were used in the identification of lead for the inactive conformation of the target kinase. The fragments or chemotypes demonstrating specific interactions with the study target were garnered from the known set of inhibitors. These interacting chemotypes were converted into substructure interaction fingerprints. The filters were used to screen a database of 10 million compounds and extract the interacting chemotypes from the identified leading hits. The extracted allosteric fragments itself constitute a new library of target specific allosteric fragments and are a good starting point for many lead design endeavors. Such protocols can easily be extended for different druggable targets to ensure the retrieval of target specific hits positively.

6.5.1.3   Fragment Based Methods

A drug or an inhibitor molecule constitutes a number of sub-parts called fragments whose presence either enhances their efficacy or renders them synthetically feasible. The action of drugs emancipate either from their physicochemical properties or from their chemical structure. The former are non-specific, act in large doses by forming a monomolecular layer over the entire cellular surface of an organism as in case of general anaesthetics, hypnotics such as aliphatic alcohols, antiseptics and antifungals (Lemke and Williams 2008). Those which are structure driven are specific and act in small doses on specific protein molecules which are usually located in the cell membrane to trigger a series of physiological and biochemical response. The specific recognition of receptors is driven by a fragment called chemotype which specifically endears the small molecule to that particular receptor (Badrinarayan and Sastry 2010). Identification of specific low molecular weight fragments in a molecule sets a stage for the stepwise design of new leads incorporating the identified chemotype. FBDD samples the chemical space to a greater extent than virtual screening of

Table 6.6 Postulates for the design of drugs, leads, scaffolds and fragments

S. No.	Properties	Drug-like (RO5)	Lead-like	Scaffold-like	Fragment-like (RO3)
1	Hydrogen bond donors (sum of -OH and -NH)	≤ 5	≤ 5	≤ 3	≤ 3
2	Hydrogen bond acceptors (sum of O and N)	≤ 10	≤ 8	≤ 8	≤ 3
3	Molecular weight (daltons)	< 500	< 450	< 350	< 300
4	Lipophilicity (clog P)	≤ 5.0	-3.0 to 4.5	≤ 2.2	≤ 3.0
5	Number of rotatable bonds (NROT)	–	≤ 9	≤ 6	≤ 3
6	Polar surface area (PSA) $\mathrm{\AA}^2$	–	–	–	60

molecules (Hajduk and Greer 2007). The fragments adhering to a set of properties as specified by the 'rule of three' are used embellished, linked and then grown into new leads. Identification of the right fragment through virtual screening or optimizing a prioritized one is computationally complicated since the existing scoring functions have been formulated for molecules and the cut offs prescribed do not suit the rule of three (Table 6.6).

However, fragments are more rigid with lesser numbers of degrees of freedom as compared to small molecule and are therefore can be easily docked. Majority of the initial work in FBDD has been carried out for the kinase targets. The fragment libraries are designed using reduced topological graph to compare the modes of the feature tree as in LoFT (Fischer et al. 2010) or use a set of bond rules to model ring substitution and cleavage sulphur groups as in BRICS (Degen et al. 2008). Certain protocols such as BROOD identify the fragment similar to the query template and design leads through bioisosteric replacements (Chen and Wang 2003). SeeDs on the other hand use pharmacophore fingerprints to screen for fragments (Baurin et al. 2004). The design of lead from fragments is usually carried out by linking fragments that bind to different parts of the target active site through a linker. We have developed a new fragment based lead design called 'Fragment Tailoring Approach (FTA)' based on similar principle wherein the existing set of kinase inhibitors binding to its highly conserved ATP site is reengineered into a target specific inhibitor by linking it with a chemotype which binds to its non-conserved allosteric site in the inactive conformation. The newly designed leads thus acquire efficacy from the ATP site fragment and specificity from the allosteric fragment (Badrinarayan and Sastry 2010). Self binding fragments as in click-chemistry on the other hand need no linker for connecting to each other and form a lead. The leads can also be derived by embellishing an individual fragment with function groups complementing the target active site. This has led to the development of different FBDD protocols

such as BREED, LUDI, RECAP, ADAPT, LEA3D, LigBuilder based on genetic algorithm and SKELGEN, SMoG, SPROUT based on Monte Carlo simulations. FBDD has thus popularized the concept of 'prioritized sub-structures'. FBDD has resulted in the successful design of leads for several important diseases such as BACE-1, Phosphodiesterase (PDE) 4, Bcl-XL, Urokinase, Thrombin and Aurora kinase (Loving et al. 2010) to name a few.

6.5.2 Analogue Based Drug Design

Analogue based approaches for rational drug design have also emerged in parallel to the structure based approaches. These approaches complement the structure based approaches where the structure of the target is unknown, but the active inhibitors for the target are known. The main concept of analogue based drug design is based on a belief that chemical structure and biological activity of the analogues of a drug are often similar to the lead drug. In the last few decades, computational methods have significantly contributed to model new analogues for an existing drug as well as to predict the activities of new analogues. These predictive models rapidly screen large databases to identify new hit and lead molecules with improved biological activity profile and greater potency, thus opening up the way to new types of structures for drug research. QSAR modeling, pharmacophore modeling is some of the most important methods in analogue based drug design.

6.5.2.1 QSAR

The QSAR modeling is one of the analogue based computational tools, which establishes a quantitative correlation between biological activity/toxicity/property of a molecule and its structural features. In QSAR study, the variations of biological activity/toxicity/property within a series of compounds are correlated with changes in a group of computed features of the molecules referred to as descriptors.

QSAR method to predict a certain property of a molecule from its structure as a mathematical expression in the form of

$$y = m_1 x_1 + m_2 x_2 + \ldots + C \tag{6.9}$$

Where, y is the predicted property (the dependent variable) and $x_1, x_2, \ldots$ are the known molecular properties called descriptors. QSAR uses descriptors that are a single number describing some aspect of the molecule, such as molecular weight, number of atoms, topological indices etc. The coefficients $m_1, m_2 \ldots$ in the QSAR equation are weights of the descriptors obtained by using various curve fitting methods.

The activities and properties being modeled by QSAR/QSPR are known as dependent variables (y) of the QSAR model. A dependent variable can be a biological
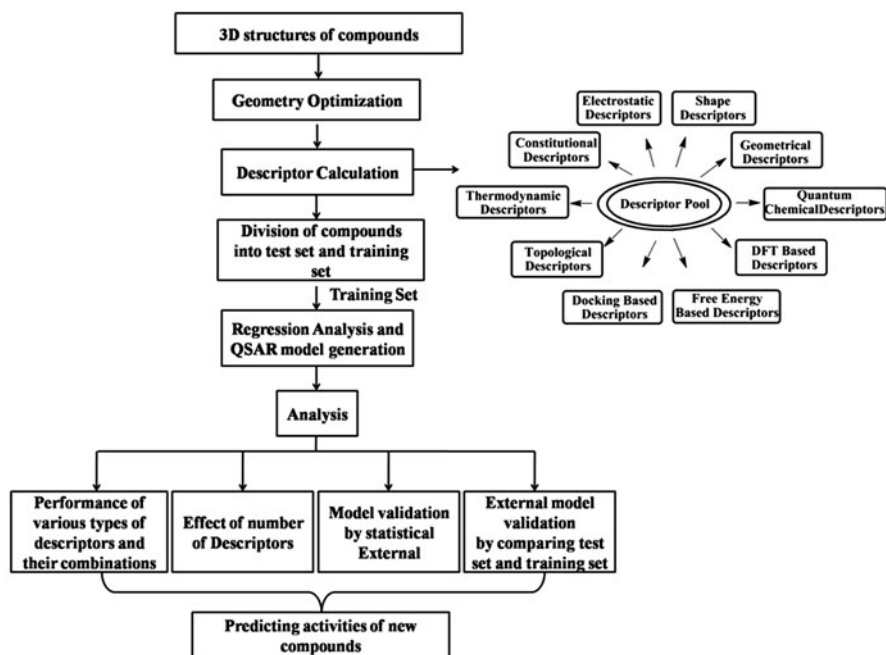
Fig. 6.5 Steps of QSAR modeling

property such as receptor binding, inhibition constant, permeability, pharmacokinetics, biodegradation, carcinogenicity, drug metabolism and clearance, mutagenicity, toxicity etc. or a chemical property such as boiling point, chromatographic retention time, dielectric constant, diffusion coefficient, dissociation constant, melting point, reactivity, solubility, stability, thermodynamic properties, viscosity etc. (Young 2009).

QSAR modeling typically describes molecular structures in terms of the descriptors and then correlates these molecular descriptors with observed activities using various statistical methods. The first step of QSAR modeling is preparation of a dataset of molecules with their activities, which follow a uniform distribution and calculation of descriptors. Molecular descriptors are chemical information that is encoded within the molecular structures and are collectively responsible for a particular activity of the molecule (Todeschini and Consonni 2000). The descriptors serve as the independent variables of a QSAR model. Various categories of descriptors employed in QSAR (Katritzky et al. 1994; Karelson et al. 1996). Constitutional descriptors are simple descriptors that represent only the molecular composition of the compound independent of the geometry and electronic structure (Fig. 6.5).

Examples are number of atoms, number of bonds, molecular weight etc. Topological descriptors/topological indices describe the atomic connectivity in the molecule. Examples are Wiener index, Randic and Kier & Hall indices, Kier flexibility index,

Information content index and its derivatives etc. (Katritzky et al. 1994). Geometrical descriptors are dependent upon 3D-coordinates of the atoms in the given molecule. For example, moments of inertia, shadow indices, molecular volume, molecular surface area, gravitation indexes etc. Electrostatic descriptors are calculated based on the charge distribution of the molecule. Examples are topological electronic index and charged partial surface area descriptors. Quantum-chemical descriptors are calculated from quantum chemical data at various levels of theory. For example Extreme (maximum and minimum) values of the atomic nucleophilic (N_A), electrophilic (E_A) and one-electron (R_A) Fukui reactivity indices, ε_{LUMO} and ε_{HOMO} etc. (Karelson et al. 1996). Hydrophobicity descriptors such as log P, aqueous solubility and chromatographic parameters are also very useful for QSAR studies (Helguera et al. 2008). However, development of simple and new descriptors is still a topic of high interest (Badrinarayan et al. 2011; Srivani et al. 2007). Among the new descriptors the density function theory (DFT) based ones are extensively studied. In many studies DFT based descriptors show good performance in predicting the biological activities (Parr 1983; Singh et al. 2004; Wadehra and Gosh 2005; Srivastava and Sastry 2012). Employment of docking scores as QSAR descriptors is one of the new approaches. The free energies of binding calculated by MMPBSA/GBSA methods are also tested in several studies and they show excellent correlation with the bioactivities (Srivastava et al. 2012). Once descriptors are computed, it is very crucial to choose the descriptors that should be included in the QSAR model. Preprocessing of the dataset should also be performed carefully as anomalies, errors, missing/incomplete data may lead to severe erroneous/misleading predictions. The data should also be normalized or standardized where there is a large range of variability in the dataset. Inter-correlated descriptors should be removed from the dataset before the model construction. (Nantasenamat et al. 2009).

Various techniques based on the multi-linear regression (MLR) analysis are employed in order to achieve the QSAR equation. This equation essentially correlates the variation of activities of the molecules as a function of the variations of the molecular structures present in the molecular data set (Kubinyi 1993). MLR analysis is usually used to correlate a given bioactivity with molecular descriptors. Different statistical methods come into play for building a QSAR model. Depending on the type of dataset and other parameters, however, it is possible to generate nonlinear equations that contain exponents of best fit, logarithms of descriptors, etc. MLR, principal component regression (PCR), partial least square, artificial neural network (ANN), genetic function approximation (GFA), factor analysis, discriminant analysis, cluster analysis are a few of the statistical methods that can be employed in the QSAR modeling (Dehmer et al. 2012).

For the linear QSAR equations the correlation coefficient r^2 gives a quantitative measure of how well the descriptor describes the activity (Wold 1991). r^2 is calculated as follows

$$r^2 = \frac{1 - \sum (y_{obs} - y_{calc})^2}{\sum (y_{obs} - y_{mean})^2} \tag{6.10}$$

where, y_{calc}, y_{obs} *and* y_{mean} are predicted, actual, and mean values of the target property respectively. Thus, the descriptors with the highest correlation coefficient can be selected. The predictive power of a QSAR model can be verified through statistical measures such as the correlation coefficient between actual and predicted values. Various statistical parameters such as cross validated correlation coefficient, Fisher statistic (F-value) values etc.

Crossvalidated r^2, lso called as q^2 signifies how best the model predicts. It is calculated by omitting each compound once from the training set, then predicting its activity using the model constructed from the remaining compounds. The model thus built with the remaining molecules is used to predict the response of the deleted compound/compounds. This cycle is repeated till all the molecules of the dataset have been deleted once. The cross-validated squared correlation coefficient q^2 is calculated as follows

$$1 - \frac{\sqrt{\sum (y_{obs} - y_{calc})^2}}{\sum (y_{obs} - y_{mean})^2} \qquad (6.11)$$

where, y_{calc}, y_{obs} *and* y_{mean} are predicted, actual, and mean values of the target property respectively. F-value is also an important measure of the statistical significance of the regression model, which is given by the following equation (Wold 1991).

$$F = \frac{r^2}{1 - r^2} \qquad (6.12)$$

where r^2 is the correlation coefficient. Also as an external validation, some of the compounds with known results are left out of the training set to be used as a test of the predictive ability of the QSAR model.

QSAR is a valuable tool for predicting molecular properties that cannot be computed any other way. It is very useful or the prediction of a wide range of biological properties, essential to identify potential leads (Nantasenamat et al. 2009). Although it may not be a reliable tool to predict drug activity, pharmacokinetic properties, such as blood–brain barrier permeability and passive intestinal absorption etc. can be fairly predicted by QSAR method. Hence, QSAR models are of immense help to predict the properties of new and untested compounds possessing analogous molecular structures as compounds used in the development of the models.

6.5.2.2   Pharmacophore Modeling

Pharmacophore modeling has gained immense importance as an analogue based approach in past few days because of its simplicity. According to IUPAC definition (Wermuth et al. 1998), "*A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supra-molecular interactions with a specific biological target structure and to trigger (or to block) its biological response.*" However, different researchers define it through their own view glasses
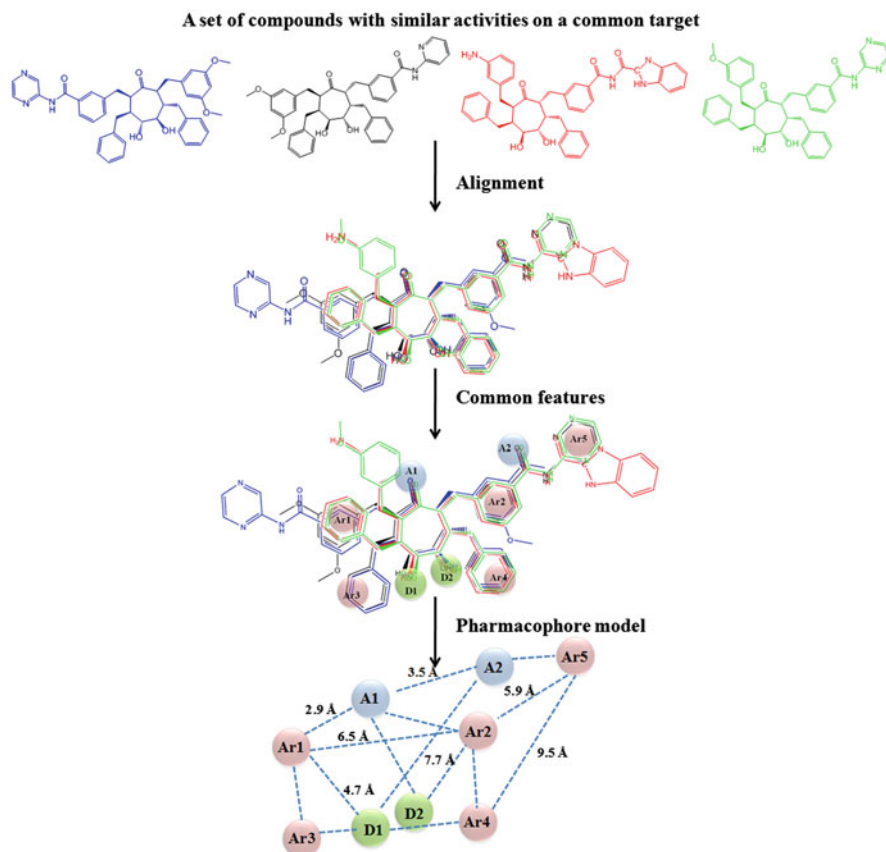
Fig. 6.6 Steps of pharmacophore modeling

depending on the suitability. A pharmacophore can be considered as the maximal set of common features extracted from a group of molecules exhibiting a similar pharmacological profile on a common target protein (Guner 2000).

A pharmacophore does not represent any real molecule, but represents the common interaction pattern of a group of compounds with their target (Wermuth 2006). The chemical signatures identified in a molecule which are actually responsible for making a certain type of non-covalent interaction with the receptor are called as pharmacophore features (Fig. 6.6).

A few examples of such functional features are hydrogen bond donors, hydrogen bond acceptors, aromatic rings (may be ring atoms, ring center, or normal to the ring), hydrophobic centers (also called neutral centers), positive charge centers, negative charge centers, acidic groups, basic groups, bulky groups engaged in steric interactions, planar atoms, CO_2 centroid (i.e., ester or carboxylic acid), metal (also

called a metal ligator) and excluded volumes—forbidden regions, where the protein is and the ligand cannot have functional groups (Dror et al. 2006). Pharmacophore models provide a reasonable qualitative prediction of binding by modeling the spatial arrangement of a small number of atoms or functional groups (Yang 2010; Ekins et al. 2001). A detailed quantitative prediction of active molecules based upon the binding pattern requires sophisticated computational techniques as well as lots of computer time. Pharmacophore models are of immense use in analogue based virtual screening. Its usefulness covers three major domains. The generation of a relevant pharmacophore model, consistent with structure property relationship in a series of molecules helps in design of optimal ligands. Scaffold hopping may be an important implication of pharmacophore modeling, which consists in the design of functional analogues by searching within large virtual compound libraries of structures with similar activity profiles, but based on a different scaffold. New active compounds can also be designed by combining the key pharmacophore features of two different pharmacophore models (Langer and Hoffmann 2006; Wolber et al. 2008).

6.6 Modeling Large Molecular Systems

Large-scale biomolecular simulations are significantly important to study the functionality of large biomolecular systems. MD simulations have substantially contributed to the advancement of knowledge in biology, chemistry and material science. Although the MD simulations are being conducted for systems with millions of atoms and for millisecond timescale, the atomistic MD simulations fare to be too long and large when studying a biological phenomenon. The functioning of the living cell is a complex process, characterised by multiple interactions between macromolecules that act across multiple levels of structural and functional organisation—from molecular reactions to target-drug binding to protein-protein interactions. Since biological systems are multiscale in nature, there should be efficient model building and biological knowledge integration and prior data at all biological scales. Hence to explore such multiscale systems quantitatively, one has to integrate several different simulation techniques at different time and length scales. This calls for a paradigm shift in the simulation techniques wherein the atomistic treatment of the large biomolecular system as in MD simulations is replaced by the partitioning of the system. Such approaches either partition the large systems based on the level of precision required for its various components as in case of QM/MM wherein the protein is divided into the active site region which is treated quantum mechanically and the non-active site region which is treated with molecular mechanics. The other approaches include the multiscale approach which constitutes a framework comprising of different levels accuracy and couples them to enable a hierarchical handshake which leads to effective transfer of information across the different scales. These approaches divergent from the basic atomistic MD are useful in predicting the structure activity relationships and provide a fundamental mechanistic understanding of biological process. This facilitates efforts in predictive modeling and molecule design efforts.

6.6.1 QM/MM

The QM methods are too complex to be applied on the large biomolecular systems whereas MM methods fail to model the enzyme mediated reaction mechanisms. Therefore considering the individual shortcomings of each of these methods, a hybrid method such QM/MM employing their individual strengths is warranted (Ayton et al. 2007). The QM/MM partitions the biomolecular system into two regions. The active site comprises the smaller region and is treated quantum mechanically while the rest of the system is treated with the classical molecular mechanics force fields. There are two schemes to calculate the total energy of the system namely the additive and subtractive (Sherwood et al. 2008; Senn and thiel 2009; Sherwood et al. 2003). The subtractive scheme consists of four components namely the total energy of the system $E_{QM/MM}$ (system), E_{MM} (system) the MM energy of the entire system, E_{QM} (QM) the QM energy of the QM region and E_{MM} (QM) the MM energy of the QM region. The equation used to calculate the energy of the system though the subtractive scheme is represented as follows:

$$E_{QM/MM}\left(system\right) = E_{MM}\left(system\right) + E_{QM}(QM) - E_{MM}\left(QM\right) \qquad (6.13)$$

The scheme encounters shortcomings due to the treatment of interactions between the QM and MM region only at MM level which is inaccurate. The scheme requires MM parameters for the QM region. Parameters are not usually available for those systems which are present in excited electronic states or contain transition metals. The additive scheme has therefore gained popularly. In this scheme, the total energy of the system $E_{QM/MM}$ (system) comprises of only three components viz., E_{MM} (MM) the MM energy of the MM region only, E_{QM} (QM) the QM energy of the QM region and the E_{QM-MM} (QM, MM) a term which interfaces between the QM/MM through the inclusion of bonded and non-bonded interactions. The bonded interactions account for bond stretching, bending and torsion while the non-bonded account for the van der Waals and electrostatic interactions.

$$E_{QM/MM}(system) = E_{MM}(MM) + E_{QM}\left(QM\right) - E_{QM-MM}(QM, MM)$$
$$(6.14)$$

The key to such QM/MM methods is the coupling between the electric field from the surrounding and the QM Hamiltonian in the active-site region. This requires careful treatment of the boundary between the QM and MM regions, either by using hybrid orbitals for the connection or a linked atom approach. The calculation of free energies from QM/MM simulations can be performed by averaging over the system's configurations via perturbations from a reference surface; however, such sampling for accurate free energy evaluations as well as calculations of pKa values remain challenging and form an active area of research.

Fig. 6.7 Different types of coarse grained models: the low resolution model constituting the major functional domains as beads and used to study molecular level interactions, the high resolution mesoscale model clustering groups of amino acid residues of the protein as a bead. The third is the ultra high resolution model considering each amino acid residue as a bead and used to study atomistic level interactions
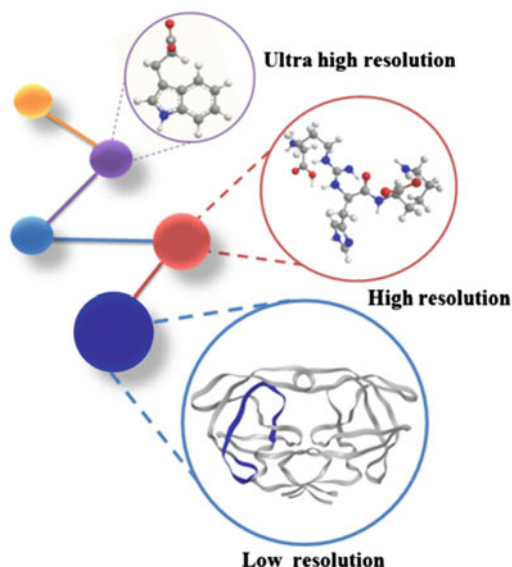
6.6.2 Coarse Graining

MD simulations, scaling over longer timescales, work well with biomolecular systems such as proteins, lipids and nucleic acids however they fall short in investigating complex phenomenon such as protein-protein assembly, vesicle diffusion, membrane deformation, DNA super coiling, DNA packaging in bacteriophage, folding of RNA in ribosome etc. Therefore, approaches such as coarse grained simulations are used wherein the single complex system is divided in to a couple of systems by grouping several atoms (Saunders and Voth 2013). Coarse graining clusters groups of atoms into beads or sites. Based on the accuracy desired either one amino acid is defined as a bead or a group of amino acids form a bead (Fig. 6.7).

These beads which are a kind of quasi-particles interact with each other. The combination of these interactions and the reduce degrees of freedom help to span the spatiotemporal scales. The accuracy and utility of a coarse grained model is largely dependent on the force field parameterization which implicitly account for the enthalpic and entropic contributions of free energy. The key steps in coarse graining include development of primary models based on experimental results followed by large scale simulation and identification of interactions influencing the energetic of the model system. Coarse graining retains the primary physical features of the system thus distills the atomistic scale information into simplistic but low resolution models. The final step therefore is to link with the molecular scale through all atom MD or Monte Carlo simulations based on the previous coarse grained simulation results so as to bridge the atomistic and mesoscopic scales. The precision in such cases can be

obtained with multiple iterations of the entire protocol. Coarse grain simulations provide information on complex phenomenon such as biomolecular self-assembly at the mesoscopic scale and with iterative information transfer guides leverage of the atomistic details of the studied phenomenon through MD and Monte Carlo simulations. Thus the property or responses which are inaccessible at the atomistic or continuum levels of theory can be effectively simulated through coarse grained approach.

There are two main approaches to coarse graining called the inversion approach and the multiscale approach. The inversion approach employs thermodynamics, structural and experimental properties to developed coarse grained models. There a gamut of inverse coarse grained methods such as the Monte Carlo inverse Newton method (Lyubartsev and Laaksonen 1995), direct Boltzmann inversion approach (Tschop et al. 1998), iterative Boltzmann inversion method by Muller-Plathe (Muller-Plathe 2002). Most of these methods use reduced statistical distributions such as radial distribution instead of calculating the many body coarse grained potential mean force functions and detect the most appropriate coarse grained potential by inversing the data. The multiscale approach builds a hierarchical ladder to bridge atomic interactions to the mesoscale coarse grained model. In this the basic functions depicting the many body coarse grained potential mean force is mapped by atomistic scale forces. One of the earliest contributions has been made by Levitt & Warshel who identified the essential components contributing to the problem of protein folding and constructed a coarse grained model (Levitt and Warshel 1975). Gholke et al. have developed a three step multiscale coarse grained approach to model the conformational changes in proteins (Kruger et al. 2012). A high resolution coarse grained model with multiple beads per amino acid residue in protein is can effectively delineate the atomistic level interactions. However, to decipher the molecular scale motions occurring at the cell level necessitates simulation of large protein assemblies using the multiscale modeling approaches.

6.6.3   Multiscale Modeling

Biological systems are made up of several individual components or strata organised in a hierarchical manner (Schnell et al. 2007). The transfer of information among them leads to the functioning of the system as a whole. There are two different ways to scale the biological systems namely the spatial and temporal scales. The spatial scales hierarchically classify the biological processes based on the organization of biological systems. These scales are called 'levels of organization' and range from quantum, molecular, cellular, tissue, organ, organism to its ecosystem (Southern et al. 2008). Associated with the spatial levels of organization are the temporal scales of biological processes which range from microsecond for molecular interactions to years for an average lifespan of human being (Walker and Southgate 2009). According to this theory, a cell is made up of millions of molecules, while a tissue is made up of billions of cells and the number game thus augments. These key components of the biological system have intra- and inter-connections (Twycross 2010).

The diversity and connectivity among these scales increase the complexity of the biological systems. It is therefore necessary to model the individual components at multiple scales and integrate them to understand the impact of intra- and inter-scale interactions on the system and its surrounding ecosystem (Noble 2002). The multiscale modeling is an integrated and iterative approach which couples information obtained from various scales.

Mathematical representation of a complex system is termed as a 'model'. Models representing complex systems span a wide range of time and length scales and such models are termed as 'multiscale models'. The use of such multiscale models in addition with experimental data to understand the functioning of a biological system is termed as 'systems biology' while the engineering of these multiscale models to construct an artificial biological system to study its functioning comes under the preview of 'synthetic biology'. The modeling of biological systems is associated with different levels of complexity and therefore it requires a laddered approach instituting simulations at various time and length scales using methods offering varied degrees of precision and speed. Multiscale approaches thus encompass the combined use of computations and mathematics to obtain a simulated representation of a physiological system at different scales of time and biomolecular organization. This is an ingrained concept in the areas of engineering, aerodynamics and fluids associated with physics and material science however it is still comparatively raw to chemical and biological science.

The multiscale approach in modeling biological systems integrates the well-established disciplines like quantum chemistry, classical MD, systems biology, pathway modeling, and bioinformatics. They lie at the crossroads of frontier research areas in physics, biology, chemistry, and medicine. The multiscale models integrate (QM), molecular mechanics (MM), hybrid QM/MM, MD (MD), coarse-grained (CG), linear scaling and heuristic approaches. Multiscale modeling of biological systems is thus a measure to understand various scales of life at different resolutions. Each scale offers different features and therefore it is up to the discretion of the modeller to choose the appropriate strategy for the maximum abstraction of data and to bridge the gap between the various scales. There are two main approaches in multiscale modeling namely the 'top-down approach' and 'bottom-up approach' (Qu et al. 2011). The top-down approach treats the system as an individual entity and studies the macroscopic properties of the system. Hodgkin et al. created an action potential model of the giant axion using this approach (Hodgkin and Huxley 1952). To do so the individual ion-channels were overlooked and the voltage dependence of whole currents was modelled based on experimental data. This simplified the process to a great extent but they fail to account for the impact of individual components which participate in the expression of studied phenomenon. The bottom-up approach on the contrary simulates each individual component of a system and models their interactions to understand the nature of the system as a whole. This approach is useful in studying the behaviour of the interactive elements of a system and is therefore used in the study of cell-transport, protein folding and working of ion-channels (Kamerlin and Warshel 2011). The main aim of multiscale modeling is not only to model a particular system at different scales but also to conserve the data accurately during

its transit from a lower scale to a higher scale or vice-versa. It has been employed to understand the functioning of important biological processes such as protein folding, membrane remodeling, drug metabolism and nucleic acid packaging.

The main objectives of a multiscale protocol are the identification of the individual processes constituting a complex system, the scales for modeling them and development of a link to couple these individual processes. In a multiscale strategy, the system is first decomposed into several sub-units. The temporal and spatial scales are allocated to model each of these sub-units. For example if the diffusion process is modelled then the temporal scale would define the rate of diffusion (Dada and Mendes 2011). The coupling of the micro-, meso- and macro-level processes leads to the development of multiscale model. Establishing coupling between scales is an intricate process (Martins et al. 2010). Solutions like multiscale Simulation Library and Environment (MUSCLE), Model Coupling Toolkit, XML based multiscale model management in systems biology have been developed to ensure smooth coupling of the scaled models.

6.7 Non-covalent Interactions

The drug once taken, travels through the body and elicits a pharmacological response. The site of drug action is the receptor while the pharmacodynamics is controlled by the different forces of interaction which bind a drug to a specific receptor (Holtje et al. 2008). The drugs and receptors exist as an ensemble of conformers in solvent. Thus to form a solvated complex with the receptor, the drug molecule needs to displace the solvent molecules occupying the binding site of the receptor. This is possible only when the interactions between the drugs and receptor are stronger than their individual interactions with the solvent molecules (Bissantz et al. 2010). The complex formation is entropically unfavorable and induces a loss in the conformational, rotational and translational degrees of freedom of both the drug and the receptor. The entropic loss is therefore expected to be compensated by favorable enthalpic contacts i.e. interactions. The bonds are spontaneously formed between atoms with a decrease in free energy (ΔG) i.e. when ΔG is negative. The activity of a small molecule (drug) is initiated by its atomic level interaction with the macromolecule (receptor or target). This association is stabilized by a plethora of intermolecular drug-receptor interactions which are either covalent or non-covalent in nature. The interaction of a drug with the binding site of a receptor depends on the complementarity of fit between the two molecules as stated in the Lock and Key Hypothesis by Emil Fischer (Silverman 2004). The interactions comply with the law of mass action. The binding of a drug to its receptor is therefore usually orchestrated though a gamut of non-covalent interactions rather than the covalent ones.

One of the most important and exhaustively studied drug-receptor interactions is the H-bond. Strong H-bonds like N-H...O, N-H...N and O-H...O are formed by the Glu, Leu and His residues which interact with the donor atom of inhibitor whereas the Leu and Gly residues interact with the acceptor atom (Sarkhel and

Desiraju 2004). The linked heterocyclic systems of the inhibitors are stabilized by the weak H-bonds such as C-H...N, C-H...O. Of the 20 amino acids comprising the protein, Gly and Glu play a substantial role as H-bond donor and acceptor. The propensity and strength to form H-bond varies with different functional groups. Thus the constitution of protein's active site has a substantial influence on the desolvation effects and SAR (Foloppe et al. 2005). The ammonium groups found in drug-receptor complexes are usually not permanently charged quaternary ions. This leaves at least one proton on the nitrogen atom which can be used in binding. The strength of the H-bond shows a dramatic increase when augmented with additional H-bonds (Neela et al. 2010). The hydrophobic nature of the active site can be attributed to a large extent to the side-chains of several aromatic residues which open into it. The aromatic rings of Phe, Trp, Tyr, and His form cation-π interactions with the cationic side-chains of Lys and Arg (Reddy and Sastry 2005). The cation-π interactions in biological systems stem from the interaction of nitrogen, phosphorous, oxygen and sulphur based onium ions (Mahadevi and Sastry 2013). The shape and electronic properties of the aryl rings of the aromatic amino acids give rise to large polarizabilities and a considerable quadrupole moment. The π-motifs engage in a T-shaped edge-to-face and the parallel-displaced stacking arrangement and interact with the heterocyclic rings of inhibitor. In proteins, the π-systems cluster into networks of various sizes. A database study by our group has shown that the CH-π and π-π stacking interactions formed by the side-chains of the aromatic residues provide stability to the protein hydrophobic pockets (Reddy et al. 2007b). The correlation between π-π stacking and hydrogen bonding is a very well studied example, owing to its relevance in nucleic acids (Vijay et al. 2008). The π-motifs form networks and their influence manifest strongly on the nature of inhibitor binding (Chourasia et al. 2011). The binding and stabilization is also contributed by the alkyl-aryl interactions as well. The aromatic π-motif forms one of the strongest non-covalent interactions on interacting with a metal ion. Such an interaction is a key player in enzyme regulation, stabilization, and functioning of nucleic acids. A subtle competition is also seen to exist between the π and σ- (in plane) approach of metal ion with the aromatic motifs (Reddy et al. 2006). The non-covalent interactions either complement or compete with each other in a cooperative or non-cooperative manner (Mahadevi and Sastry 2013; Vijay and Sastry 2010). The cooperativity of non-covalent interactions is an interesting phenomenon which is known to influence stability, conformational transitions and allosteric interactions in addition to inhibitor binding. The array of non-covalent interactions and their role in bio-macromolecules contribute to in the drug-receptor interactions, stabilization, and functional reorganization necessitating their consideration in design of leads (Fig. 6.8).

The non-covalent interactions engage in reversible binding are therefore preferred in CNS drugs, depressants etc. where the pharmacological effect needs to be terminated after some time. The role and relevance of non-covalent interactions in biological systems and the ability of computational methods to model them makes them a topic of high contemporary interest.
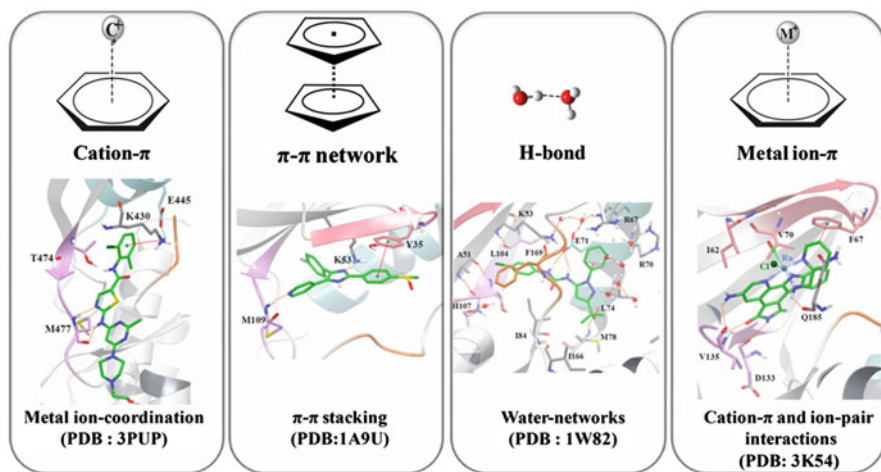
Fig. 6.8 Non-covalent interactions at the interface of chemistry and biology

6.8 Outlook

Molecular modeling has occupied the central space in basic, applied and industrial research. At the interface of chemistry, biology and material science, computational modeling has played a pivotal role in understanding the structure function relationships at atomistic levels. Although reliable and rigorous approaches have strong limitations in their applicability as the size of the system increases, several practical alternatives have been steadily emerged. This chapter provides a brief overview of the computational methods which can be applied to small and large molecules particularly bio-molecules.

References

Allen MP, Tildesley DJ (1987) Computer simulations of liquids. Clarendon Press, Oxford

Ayton GS, Noid WG, Voth GA (2007) Multiscale modeling of biomolecular systems: in serial and in parallel. Curr Opin Struct Biol 17:192–198

Badrinarayan P, Sastry GN (2010) Sequence, structure, and active site analyses of p38 MAP kinase: exploiting DFG-out conformation as a strategy to design new type II leads. J Chem Inf Model 51:115–129

Badrinarayan P, Sastry GN (2011) Virtual high-throughput screening in new lead identification. Comb Chem High T Scr 14:840–860

Badrinarayan P, Sastry GN (2012) Virtual screening filters for the design of type II p38 MAP kinase inhibitors: a fragment based library generation approach. J Mol Graph Modell 34:89–100

Badrinarayan P, Sastry GN (2013) Rational approaches towards lead optimization of kinase inhibitors: the issue of specificity. Curr Pharm Des 19:4714–4738

Badrinarayan P, Srivani P, Sastry GN (2011) Design of 1-arylsulfamido-2-alkylpiperazine derivatives as secreted PLA2 inhibitors. J Mol Model 17:817–831

Baurin N, Aboul-Ela F, Barril X, Davis B, Drysdale M, Dymock B, Finch H, Fromont C, Richardson C, Simmonite H, Hubbard RE (2004) Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. J Chem Inf Comput Sci 44:2157–2166

Bissantz C, Kuhn B, Stahl MA (2010) Medicinal chemist's guide to molecular interactions. J Med Chem 53:5061–5084

Chen X, Wang W (2003) The use of bioisosteric groups in lead optimization. Annu Rep Med Chem 38:333–346

Chourasia M, Sastry GM, Sastry GN (2011) Aromatic-aromatic database, A2ID: an analysis of aromatic π-networks in proteins. Int J Biol Macromol 48:540–552

Cramer CJ (2004) Essentials of computational chemistry: theories and models, 2nd edn. Wiley, Chichester

Cross JB, Thompson DC, Rai BK, Baber JC, Fan KY, Hu Y, Humblet C (2009) Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. J Chem Inf Model 49:1455–1474

Dada JO, Mendes P (2011) Multi-scale modelling and simulation in systems biology. Integr Biol 3:86–96

Degen J, Wegscheid-Gerlach C, Zaliani A, Rarey M (2008) On the art of compiling and using 'drug-like' chemical fragment spaces. Chem Med Chem 10:1503–1507

Dehmer M, Varmuza K, Bonchev D (eds) (2012) Statistical modeling of descriptors in QSAR and 881 QSPR. Wiley-Blackwell, Weinheim

Dror O et al (2006) Predicting molecular interactions *in silico*. I. An updated guide to pharmacophore identification and its applications to drug design. Front Med Chem 3:551–584

Ekins S, De Groot MJ, Jones JP (2001) Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome P450 Active Sites. Drug Metab Dispos 29:936–944

Fermann JT, Valeev EF (1997) Fundamentals of molecular integrals evaluation. Tech. rep

Field MJ, Moleculaire LD, Grenoble (2007) Practical introduction to the simulation of molecular systems, 2nd edn. Cambridge University Press, Cambridge

Fischer JR, Lessel U, Rarey MJ (2010) LoFT: similarity-driven multiobjective focused library design. Chem Inf Model 50:1–21

Foloppe N, Fisher LM, Howes R, Kierstan P, Potter A, Robertson AG, Surgenor AE (2005) Structure-based design of novel Chk1 inhibitors: insights into hydrogen bonding and protein-ligand affinity. J Med Chem 48:4332–4345

Frenkel D, Smit B (2002) Understanding molecular simulations: from algorithms to applications, 2nd edn. vol 1, Computational Science Series Academic Press, San Diego

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47:1739–1749

Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands. Angew Chem Int Ed Engl 41:2644–2676

Guner OF (ed) (2000) Pharmacophore perception, development, and use in drug design. International University Line, La Jolla

Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Reviews Drug Discov 6:211–219

Helguera AM, Combes RD, Gonzalez MP, Cordeiro MN (2008) Applications of 2D descriptors in drug design: a DRAGON tale. Curr Top Med Chem 8:1628–1655

Hinchliffe A (2003) Molecular modelling for beginners, vol xviii. Wiley, Chichester

Hirschfelder JO, Curtiss L, Bird RB (1954) Molecular theory of gases and liquids. Wiley, New York

Hodgkin AL, Huxley, AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 117:500–544

Holtje HD, Sippl W, Rognan D, Folkers G (2008) Molecular modeling: Basic principles and explanations, 3rd edn. Wiley-VCH, Weinheim

Hopkins AL, Groom CR (2002) The druggable genome. Nat Rev Drug Discov 1:727–730

Jensen F (2007) Introduction to computational chemistry, 2nd edn. Wiley, UK

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and Validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748

Kamerlin SC, Warshel A (2011) Multiscale modeling of biological functions. Phys Chem Chem Phys 13:10401–10411

Karelson M, Lobanov VS, Katritzky AR (1996) Quantum-chemical descriptors in QSAR/QSPR Stud. Chem Rev 96:1027–1043

Katritzky AR, Lobanov VS, Karelson M (1994) CODESSA 2.0 comprehensive descriptors for structural and statistical analysis. University of Florida, U.S.A.

Klebe G (2006) Virtual ligand screening: strategies, perspectives and limitations. Drug Discov Today 11:580–594

Kruger DM, Ahmed A, Gohlke H (2012) NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins. Nucleic Acids Res 40:W310–316

Kubinyi H (1993) QSAR, Hansch analysis and related approaches. In: Timmerman H, Mannhold R (eds) Methods and principles in medicinal chemistry. Wiley-VCH, Weinheim

Langer T, Hoffmann RD (eds) (2006) Pharmacophore and pharmacophore searches, vol 32. Wiley-VCH, Weinheim

Leach AR (2001) Molecular modelling—principles and applications, 2nd edn. Prentice Hall, Essex

Lemke TL, Williams D, Roche VF, Zito SW (eds) (2008) Foye's principles of medicinal chemistry, 6th edn. Lippincott Williams & Wilkins, Philadelphia

Levine IN (2013) Quantum chemistry, 7th edn. Prentice Hall, UK

Levitt M, Warshel A (1975) Computer simulation of protein folding. Nature 253:694–698

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25

Loving K, Alberts I, Sherman W (2010) Computational Approaches for Fragment-Based and *De Novo* Design. Curr Top Med Chem 10:14–32

Lyubartsev AP, Laaksonen A (1995) Calculation of effective interaction potentials from radial distribution functions: a reverse Monte Carlo approach. Phys Rev E 52:3730–3737

Mahadevi AS, Sastry GN (2013) Cation-π interaction: its role and relevance in chemistry, biology, and material science. Chem Rev 113:2100–2138

Martins ML, Ferreira SC Jr, Vilela MJ (2010) Multiscale models for biological systems. Curr Opin Colloid Interface Sci 15:18–23

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662

Muller-Plathe F (2002) Coarse-graining in polymer simulation: from the atomistic to the mesoscopic scale and back. Chemphyschem 3:754–769

Nantasenamat C, Ayudhya CIN, Naenna T, Prachayasittikul V (2009) A practical overview of quantitative structure-activity relationship. EXCLI J 8:74–88

Neela YI, Mahadevi AS, Sastry GN (2010) Hydrogen bonding in water clusters and their ionized counterparts. J Phys Chem B 114:17162–17171

Noble D (2002) Modeling the heart from genes to cells to the whole organ. Science 295:1678–1682

Parr RG (1983) Density functional theory. Annu Rev Phys Chem 34:631–656

Ponder JW, Case DA (2003) Force fields for protein simulation. Adv Prot Chem 66:27–85

Qu Z, Garfinkel A, Weiss JN, Nivala M (2011) Multi-scale modeling in biology: how to bridge the gaps between scales? Prog Biophys Mol Biol 107:21–31

Rapaport DC (2004) The art of MD simulation, 2nd edn. Cambridge University Press, New York

Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261:470–489

Reddy MR, Erion MD (eds) (2001) Free energy calculations in rational drug design. Kluwer/Plenum Press, New York

Reddy AS, Sastry GN (2005) Cation [M = H$^+$, Li$^+$, Na$^+$, K$^+$, Ca$^+$, Mg^{2+}, NH$_4^+$, and NMe$_4^+$] interactions with the aromatic motifs of naturally occurring amino acids: a theoretical study. J Phys Chem A 109:8893–8903

Reddy AS, Vijay D, Sastry GM, Sastry GN (2006) From subtle to substantial: role of metal ions on pi-pi interactions. J Phys Chem B 110:247924-81

Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN (2007a) Virtual screening in drug discovery—a computational perspective. Curr Prot Peptide Sci 8:329–351

Reddy AS, Sastry GM, Sastry GN (2007b) Cation-aromatic database. Proteins: Struc Func Bioinfo 67:1179–1184

Ringe D Jr, Reynolds CH, Merz KM (eds) (2010) Drug design: structure- and ligand-based approaches. Cambridge University Press, UK

Sarkhel S, Desiraju GR (2004) N–H...O, O–H...O, and C–H...O hydrogen bonds in protein-ligand complexes: strong and weak interactions in molecular recognition. Proteins 54:247–259

Sastry M, Lowrie JF, Dixon SL, Sherman W (2010) Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. J Chem Inf Model 50:771–784

Saunders MG, Voth GA (2013) 1. Coarse-graining methods for computational biology. Annu Rev Biophys 42:73–93

Schnell S, Grima R, Maini PK (2007) Multiscale modeling in biology. Am Sci 95:134–142

Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. Angew Chem Int Ed Engl 48:1198–229

Sherwood P, de Vries AH, Guest MF et al (2003) QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis. J Mol Struct Theochem 632:1–28

Sherwood P, Brooks BR, Sansom MS (2008) Multiscale methods for macromolecular simulations. Curr Opin Struct Biol 18:630–640

Silverman RB (2004) The organic chemistry of drug design and drug action, 2nd edn. Elsevier Academic Press, San Diego

Singh PP, Srivastava HK, Pasha FA (2004) DFT-based QSAR study of testosterone and its derivatives. Bioorg Med Chem 12:171–177

Southern J, Francis JP, Whiteley J, Stokeley D, Kobashi H, Nobes R, Kadooka Y, Gavaghan D (2008) Multi-scale computational modelling in biology and physiology. Prog Biophys Mol Biol 96:60–89

Srivani P, Srinivas E, Raghu R, Sastry GN (2007) Molecular modeling studies of pyridopurinone derivatives—potential phosphodiesterase 5 inhibitors. J Mol Graph Model 26:378–390

Srivastava HK, Sastry GN (2012) MD investigation on a series of HIV protease inhibitors: assessing the performance of MM-PBSA and MM-GBSA approaches. J Chem Inf Model 52:3088–3098

Srivastava HK, Choudhury C, Sastry GN (2012) The efficacy of conceptual DFT descriptors and docking scores on the QSAR models of HIV protease inhibitors. Med Chem 8:811–825

Stahl M, Bohm HJ (1998) Development of filter functions for protein-ligand docking. J Mol Graph Model 16:121–132

Stahl M, Rarey M (2001) Detailed analysis of scoring functions for virtual screening. J Med Chem 44:1035–1042

Todeschini R, Consonni V (2000) Handbook of molecular descriptors. In: Mannhold R, Kubinyi H, Timmermann H (eds) Methods and principles in medicinal chemistry. Wiley-VCH, Weinheim

Tschop W, Kremer K, Batoulis J, Burger T, Hahn O (1998) Simulation of polymer melts. I. Coarse graining procedure for polycarbonates. Acta Polym 49:61–74

Twycross J, Band LR, Bennett MJ, King JR, Krasnogor N (2010) Stochastic and deterministic multiscale models for systems biology: an auxin-transport case study. BMC Syst Biol 4:34–45

Vijay D, Sastry GN (2010) The cooperativity of cation-π and π-π interactions. Chem Phys Lett 485:235–242

Vijay D, Zipse H, Sastry GN (2008) On the cooperativity of cation-π and hydrogen bonding interactions. J Phys Chem B 112:8863–8867

Wadehra A, Ghosh SK (2005) A density functional theory-based chemical potential equalization approach to molecular polarizability. J Chem Sci 117:401–409

Walker DC, Southgate J (2009) The virtual cell-a candidate co-ordinator for 'middle-out modelling of biological systems. Briefings Bioinf 10:450–461

Wang Yi, McCammon JA (2012) Introduction to MD: theory and applications. In: Dokholyan NV (ed) Biomolecular modeling computational modeling of biological systems. From molecules to pathways. Springer, USA, pp 3–30

Wermuth CG (2006) Pharmacophores: historical perspective and viewpoint from a medicinal chemist. In: Langer T, Hoffmann RD (eds) Wiley-VCH Verlag GmbH & Co. KGaA

Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA (1998) Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure Appl Chem 70:1129–1143

Wolber G, Seidel T, Bendix F, Langer T (2008) Molecule-pharmacophore superpositioning and pattern matching in computational drug design. Drug Discov Today 13:23–29

Wold S. (1991) Validation of QSARs. Quant Struct Act Relat 10:191–193

Yang SY (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discov Today 15:444–450

Young DC (2009) Computational drug design: a guide for computational and medicinal chemists. Wiley, Hoboken

Zeigler B, Praehofer H, Kim T (eds) (2000) Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems, 2nd edn. Academic Press, New York