

Molecular dynamics in predicting the stability of drug-receptor interactions

Devendra Prajapat^{1,2} and B. Jayaram^{1,2,3,*}

1– Supercomputing Facility for Bioinformatics & Computational Biology, 2 – Department of Chemistry, 3 – Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India – 110016.

devendra@scfbio-iitd.res.in, bjayaram@chemistry.iitd.ac.in

Abstract

With the growing number of small molecule libraries, Computer Aided Drug Discovery (CADD) now has a profound role in proposing novel molecules to cure ever-increasing diseases and disorders. Molecular simulations are a key step in this process, allowing a closer look into the molecular motions associated with the recognition of biomolecular targets by ligands. This article focuses on the fundamentals of molecular dynamics (MD) and several state-of-the-art concepts in vogue, such as free energy perturbation, umbrella sampling, steered molecule dynamics, MMBAPPL, and MMPBSA/GBSA methods in the context of drug discovery. The study emphasizes how molecular dynamics efficiently helps elucidate the mode of action of drug molecules, identify allosteric or cryptic binding pockets, and decipher mutational effects in the target proteins, not all of which are within the accessible domain of crystallographic experiments. We also present some case studies in which molecular dynamics and free energy simulations, combined with virtual screening and molecular docking, have successfully contributed to novel pharmacological therapeutics. With the growing computer power and development of enhanced sampling techniques, simulation-driven CADD has a prominent role in developing novel drugs and has a bright future.

Keywords: Molecular dynamics, Monte Carlo, Free energy perturbation, Steered MD, MMPBSA/GBSA/Bappl, Meta dynamics, Umbrella sampling, PMF, CADD and lead compounds.

8.1 Introduction

Significant health concerns exist worldwide due to the continuous emergence of disabling and fatal diseases. New therapeutic molecules must be discovered because of the persistent difficulty of dealing with the shortage of FDA-approved drugs and vaccines and growing drug resistance. As per the WHO report, more than 12000 diseases exist worldwide, and only 2724 FDA-approved drugs are available. Drug discovery is a lengthy process, and it takes a drug up to 14 years to come into the market and costs around \$2.6 billion. However, with the advent of CADD, time and cost have been reduced with improved efficiency. In CADD, screening and docking are rapid studies that help to filter out molecules with low affinity and shrink the sample space of drug-like molecules for a target protein/receptor. Molecular dynamics simulations provide the drug-receptor complex with nearly the same physical and chemical conditions as in an *in vitro* system, if not the true biological system. Thus, MD simulations and *post facto* binding free energy and structural analyses have now routinely become the end game in CADD research.

Indeed, understanding molecular motion, as presaged by Feynman, is essential for drug discovery. The earliest lock and key model of drug binding to the receptor [1], where it was considered that both the receptor and ligand remain rigid, gave way to adaptive models, which accounted for the motion of atoms [2-6]. One of several examples available in the literature that account for the importance of atomic motion is the mollusk acetylcholine binding protein (AChBP), a functional and structural surrogate of the human nicotinic acetyl

receptor (AChR) [7-9]. Figure (8.1) shows that nicotine-bound AChBP has a C loop partially closed around the nicotine, whereas a large AChR antagonist (snake α -neurotoxins) bound AChBP crystal structure has the same loop displaced by 10 Å, which opens the active site [10]. Bourne and co-workers [10] later proposed that the unbound AChR and AChBP are highly conformationally flexible and exist in many conformations. It is quite likely that any of these conformations is druggable and relevant from a pharmacological perspective.

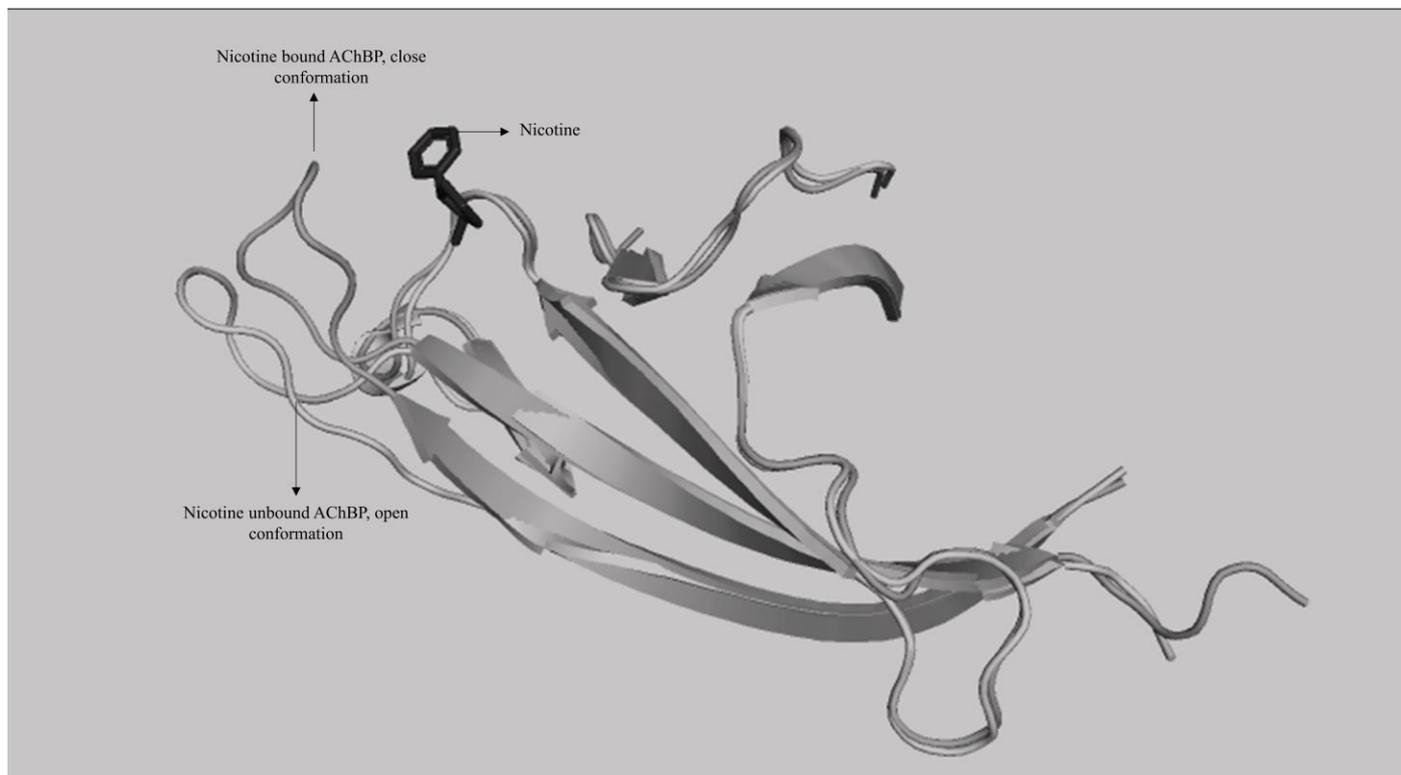


Figure 8.1: Conformational changes in AChBP protein associated with nicotine binding

Though crystallographic studies reveal that the conformational flexibility of a protein plays a crucial role in ligand binding events, the high cost associated with these techniques to generate such structures led the scientific community to seek computational methods that can deal with the associated atomistic motions. A quantum mechanical treatment of these motions is very complex and computationally intensive. Therefore, the community has utilized Newtonian physics to simulate atomic motions [11]. Molecular dynamics (MD) is a physical method for investigating the motions of atoms and molecules according to Newton's laws. A force field determines the forces between atoms and calculates the system's total energy. Following the integration of Newtonian physics into MD simulation, the trajectories that describe the positions and velocities of particles over time are generated in successive configurations of an evolving system. These MD trajectories can be used to derive a wide range of variables, including structural, dynamic and thermodynamic properties. Karplus, McCammon, Levitt and others established early on that classical MD simulations were fundamental to the study of biological systems. These early attempts at simulating complicated, spontaneous events, such as protein folding, were made possible by the use of molecular dynamics simulations.

Serious challenges have emerged with static structure-based drug design (SBDD) in the last few decades as it does not consider protein conformational changes, which are often observed during ligand binding. Recently, advanced molecular docking procedures allowed receptor flexibility while screening chemical libraries to address the issue of structural flexibility in SBDD. As a result of this protein flexibility, the utilization of several target conformations in SBDD has shown improved results [12]. Ensemble docking is an example of

this, where multiple protein conformations are targeted simultaneously. As a result of averaging and combining the data, better results are obtained [13]. While experimental methods may be used to provide a structural description of the receptor in ensemble docking, MD simulations offer a preferred alternative [14-17]. Pang and Kozikowski used a 40 ps trajectory of the acetylcholinesterase enzyme in 1994 to extract several conformations of the enzyme. They were able to accurately forecast the bound position of huperzine-A using rigid docking [18]. The target plasticity has since been discretely represented using snapshots taken from MD trajectories by several research groups [19]. McCammon and coworkers' relaxed complex approach is a famous example [20].

Although simulations lasting as long as a few micro to milliseconds are becoming feasible [21, 22], it is advisable to use a variety of trajectories to get sufficient data and a comprehensive sampling of the conformational space. Working on a single lead molecule to study long binding/unbinding kinetic studies to find drug likeliness is still a daunting task [23]. Conventional MD approaches cannot yet characterize gradual binding/unbinding events even when running on specialised hardware, which is one of the major challenges that limit the application of MD simulations in predicting the kinetic aspects of drug discovery [24]. The inadequate sampling, however, stimulated the development of several methods that are fundamental for the enhanced sampling methods, such as umbrella sampling [25], translational-path sampling [26], accelerated MD [27], free-energy perturbation [28, 29], replica exchange [30], steered MD [31, 32], milestoning [33], metadynamics [34], and their many possible combinations and variations. Recent studies have shown the effectiveness of these techniques for analyzing protein-ligand interactions and for calculating the free energy and kinetic parameters [35, 36]. In this article, we briefly touch upon some theoretical foundations of traditional MD, with an emphasis on how they relate to the development of new drugs. We then discuss some of the most cutting-edge and fruitful uses of classical MD for investigating protein-ligand interactions and for extracting key thermodynamic data, such as binding free energies. We then look at instances where MD is used in conjunction with enhanced sampling approaches to elucidate the conformational and energetic landscapes of the receptor-ligand systems. Again, we concentrate on the most commonly used methods in drug development, such as free-energy perturbation (FEP), umbrella sampling, MMPBSA/GBSA/Bappl etc..

8.2 Theoretical Background

The fundamental objective of an MD simulation is to study the time-dependent properties of microscopic systems, which is achieved by solving the differential equation [Equation (1)] given below:

$$f_j = m_j a_j(t) = -\frac{\partial V(x(t))}{\partial x_j(t)} \quad (1)$$

where f_j is the net force on the j^{th} atom, m_j and $a_j(t)$ are the mass and acceleration at time t , respectively. The vector $x(t)$ represents the instantaneous configuration of the system, which contains the coordinates of N interacting particles. Due to the high computational cost associated with solving the Hamiltonian, a classical mechanical description of the force is adapted, and an empirical potential energy function $V(x(t))$ is employed to calculate the forces. The simplest model arising from such a representation is referred to as the molecular mechanics force field (FF).

$$V = \sum_i^{bonds} \frac{K_r}{2} (r - r_{eq})^2 + \sum_i^{angles} \frac{K_\theta}{2} (\theta - \theta_{eq})^2 + \sum_i^{dihedrals} \left\{ \sum_k^M \frac{V_{ik}}{2} [1 + \right.$$

$$\cos(n_{ik}\omega_{ik} - \omega_{0,ik})\}} + \sum_{i,j}^{pairs} \epsilon_{i,j} \left[\left(\frac{r_{0,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0,ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j}^{pairs} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad (2)$$

The initial three terms in Equation (2) represent the intramolecular interactions of the particles, which are represented as summations over bond lengths (r), bond angles (θ) and dihedrals (ω), respectively. Angle bending and bond stretching share the same harmonic functional form with the equilibrium value r_{eq} and θ_{eq} , and force constants K_θ and K_r , respectively. The dihedrals have been defined using the cosine series of M terms, where n_{ik} describes the multiplicity for the k^{th} term of the series, ω_{ik} is the corresponding phase angle and V_{ik} is referred to as an energy barrier. 4^{th} and 5^{th} terms combinedly are referred to as nonbonded descriptors capturing van der Waals and electrostatic interactions between pairs of atoms, respectively. The van der Waals term here is also called "(12,6) Lennard-Jones potential", where the parameter ϵ_{ij} describes the energy well depth, and $r_{0,ij}$ is the sum of the van der Waals radii corresponding to minimum energy. Lastly, Coulombic potential defines the electrostatic interaction, where q_i and q_j are the partial charges on the atom i and j , respectively. ϵ_0 refers to the dielectric constant (permittivity) of free space and ϵ_r stands for relative permittivity.

Though quantum mechanical (QM) treatment of molecular systems produces more accurate results [37], it is inherently difficult to solve the time-dependent Schrodinger equation for a system having thousands of atoms, while classical force fields (FFs) considerably speed up the calculations. AMBER [38], CHARMM [39], GROMOS [40, 41] and OPLS [42] are a few widely used FFs in the scientific community for biological simulations.

By appropriately expanding the initial parametrization, parameters for nucleic acids, lipids, carbohydrates, and several ionic species have recently been added to the parent FFs in addition to amino acids. The structural diversity of tiny molecules, or ligands, has presented a significant barrier to the condensed-phase FFs. The user often has to provide certain parameters to get around this. Many efforts have been made to streamline and automate this laborious and error-prone process, including the creation of specialized parametrization toolkits and general force field sets for organic molecules (e.g., GAFF [43] for AMBER and CGenFF [44] for CHARMM).

A near-exact solution of Equation (1) is possible for a system consisting of only a few atoms with the help of numerical methods. The integrations in the equation can be divided into discrete time steps (tiny). While solving the integration of equations for each step, force is considered to be constant. However, as stated by Equation (1), forces depend on the positions of atoms that change over time. Thus, to take care of this, the time step is usually taken in the order of femtoseconds, which guarantees a reliable force over time. Verlet algorithm [45, 46] is one of the integrators used in most of the MD codes, where position ($t + \delta t$) is given in terms of $x(t)$, $v(t)$ and acceleration by Equation (3):

$$x_j(t + \delta t) = x_j(t) + v_j(t)\delta t + \frac{1}{2}a_j(t)\delta t^2 \quad (3)$$

Similarly, the $v(t + \delta t)$ is related to the $v(t)$ and acceleration (a) through Equation (4):

$$v_j(t + \delta t) = v_j(t) + \frac{1}{2}[a_j(t) + a_j(t + \delta t)]\delta t \quad (4)$$

Acceleration is obtained using the forces acting on the particle, which is calculated by taking the first derivative of potential energy with respect to the position. As can be seen in Equation (4), $v(t + \delta t)$ requires $a(t + \delta t)$, which means positions must be calculated before advancing the velocities. SHAKE [47] is one of

the commonly used algorithms where equations are solved considering only relevant degrees of freedom.

Classical MD involves solving Newton's equations of motion. Therefore, the total energy of the system must remain constant, that is:

$$H(p, x) = K(p) + V(x) \quad (5)$$

Where $H(p,x)$ stands for classical mechanical Hamiltonian that depends on the momentum and position of the particle. It is evident from these equations that as long as the integrator is working fine, the system behaves like a micro-canonical ensemble (constant NVE). By regulating the system's temperature and pressure while simulating, it is feasible to simulate real macroscopic phenomena more accurately. Constant temperature is achieved by rescaling the system's velocity, whereas volume rescaling helps control the system's pressure. Thermostats [48] and barostats [45, 46] are commonly used algorithms to control the temperature and pressure of the system, respectively. Note that one must not draw too many parallels between this and a temperature-controlled bath. Instead, by appropriately altering the equations of motion, the system's temperature is driven to reach, on average, the required macroscopic value.

Periodic boundary conditions (PBC) are often employed in MD to better realize bulk characteristics with finite-sized systems. The system is inserted into a unit cell with PBC reproduced in all directions to create an endless lattice. This way, all the atoms of periodic images are considered while evaluating energy terms for the atoms present in the original unit cell [45, 46]. In order to reduce the computational cost, a spherical cutoff technique with a radius of 10 Å may be utilized, which can be used to calculate the short-ranged van der Waals terms. Contrarily, the assessment of the Coulomb energy over the complete periodic lattice is necessary due to the inherently long-ranged character of electrostatic interactions which decay as r^{-1} . Ewald sum techniques are used to calculate entire electrostatics and handle the rapidly evolving electrostatic interactions at short ranges and gradually vanishing potential at far ranges differently. Particle Mesh Ewald (PME) algorithm [49] is a commonly employed method in biomolecular simulations for improved electrostatic description. For more details, the readers can refer to the articles and reviews mainly focused on PME [50]. With the help of all these algorithms, it is now feasible to simulate biomolecules in realistic conditions, enabling the scientific community to look closely at drug binding/unbinding and its kinetic and thermodynamic aspects. The drug-binding/unbinding events can quantitatively be discussed with the concept of free energy, and the following section discusses binding free energy estimations in detail.

8.3 Free Energy Calculations

Free energy in physical chemistry is arguably the most fundamental quantity [51-53]. The propensity of molecular systems to associate, dissociate, and hence react is described by their free energies. As a result, the ability to estimate free energy using molecular theories is an appealing goal. It is possible to arrive at a quantitative estimate of free energy by applying the principles of quantum mechanics and statistical mechanics. The statistical mechanical theory [54-56] of binding in aqueous media is given below:



P (receptor) and L(ligand) are the reactants and $P^{\#}L^{\#}$ is their non-covalent association. The superscript # represents the changes in the structure upon non-covalent association. At equilibrium:

$$\mu_{P.aq} + \mu_{L.aq} = \mu_{P\#L\#.aq} \quad (6)$$

$$\mu_{P.aq}^0 + RT \ln\left(\frac{\gamma_P C_P}{C_0}\right) + \mu_{L.aq}^0 + RT \ln\left(\frac{\gamma_L C_L}{C_0}\right) = \mu_{P\#L\#.aq}^0 + RT \ln\left(\frac{\gamma_{P\#L\#} C_{P\#L\#.aq}}{C_0}\right) \quad (7)$$

Where μ is partial molar Gibbs free energy (chemical potential), μ^0 represent its standard condition (1 bar), and γ and C stands for activity coefficient and concentration, respectively.

$$\Delta G_{aq}^0 = \mu_{P\#L\#}^0 - (\mu_{P.aq}^0 + \mu_{L.aq}^0) = -RT \ln\left[\frac{\gamma_{P\#L\#} C_{P\#L\#}^0}{(\gamma_L C_L)(\gamma_P C_P)}\right] = -RT \ln K_{eq.aq} \quad (8)$$

$$\Delta G_{aq}^0 = -RT \ln K_{eq.aq} \quad (9)$$

From statistical mechanics, we know:

$$U = \frac{\sum E_i \exp\left(\frac{-E_i}{k_B T}\right)}{\sum \exp\left(\frac{-E_i}{k_B T}\right)} = -\left[\frac{\partial \ln Q_{N,V,T}}{\partial \left(\frac{1}{k_B T}\right)}\right]_{N,V} = \left[\frac{k_B T^2}{Q} \frac{\partial Q}{\partial T}\right]_{N,V} \quad (10)$$

Where Q is the canonical ensemble (name given to an ensemble for constant temperature, number of particles and volume) and is given by Equation (11) below:

$$Q_{NVT} = \frac{1}{N!} \frac{1}{h^{3N}} \int \int dp^N dr^N \exp\left[-\frac{H(p^N, r^N)}{k_B T}\right] \quad (11)$$

Hamiltonian, H can be considered the same as the total energy, $E(p^N, r^N)$ and, therefore, can be replaced. The factor $N!$ comes from the indistinguishability of the particles and h^{3N} in the denominator ensures that the partition function resembles the quantum mechanical results of a particle in a box. From thermodynamics, we know:

$$U = \left[\frac{\partial \left(\frac{A}{T}\right)}{\partial \left(\frac{1}{T}\right)}\right]_{N,V} \quad (12)$$

Comparing Equations (10) and (12) gives Helmholtz free energy (A), analogous to Gibbs free energy expression [Equation (9)] in terms of the partition function $Q_{N,V,T}$:

$$A = -k_B T \ln Q_{N,V,T} + T \phi(V, N) = -k_B T \ln \left[\frac{Q_{P\#L\#}}{Q_P Q_L}\right]_{N,V,T} + T \phi(V, N) \quad (13)$$

ϕ is an unknown function of N and V . Since it's independent of T ; therefore, ϕ does not contribute to the derivative and is often considered zero. Therefore,

$$A = -k_B T \ln \left[\frac{Q_{P\#L\#}}{Q_P Q_L}\right]_{N,V,T} \quad (14)$$

The partition function Q can further be divided into individual degrees of freedom; translational, rotational and vibrational. Q_P and Q_L can be computed analytically, whereas molecular simulations are employed to deduce $Q_{P\#L\#}$. Mechanical properties of a system, such as internal energy, heat capacity, pressure etc., can be obtained using conventional molecular dynamics (MD) simulations, whereas statistical properties like free energy and entropy are difficult to calculate accurately. These latter properties are directly related to the

partition function, whereas the mechanical properties are obtained using the derivatives of the partition function. The subsequent section addresses the free energy problem associated with conventional MD. From Equations (10) and Equation (11), we can write:

$$U = \frac{k_B T^2}{Q} \int \int dp^N dr^N \frac{E(p^N, r^N)}{k_B T^2} \exp\left[\frac{-E(p^N, r^N)}{k_B T}\right] \quad (15)$$

$$= \int \int dp^N dr^N E(p^N, r^N) \frac{\exp\left[\frac{-E(p^N, r^N)}{k_B T}\right]}{Q}$$

The exponential term $\frac{\exp\left[\frac{-E(p^N, r^N)}{k_B T}\right]}{Q}$ is the probability, $[\rho(p^N, r^N)]$ of the state having energy $E(p^N, r^N)$. Thus the internal is given by:

$$U = \int \int dp^N dr^N E(p^N, r^N) \rho(p^N, r^N) \quad (16)$$

Equation (16) states that the states with high values of $E(p^N, r^N)$ have very low probability and contribute insignificantly to the overall integral. By default, the conventional molecular dynamics and Monte Carlo methods sample the low energy states (representative of the equilibrium state) of the phase space, which makes a significant contribution to the properties like internal energy, heat capacity and pressure (mechanical properties in general); therefore, these methods can make accurate estimates of equilibrium properties. Now, consider the problem associated with calculating the Helmholtz free energy. From Equations (11) and (14), we can write:

$$A = -K_B T \ln Q = K_B T \ln \frac{N! h^{3N}}{\int \int dp^N dr^N \exp\left[\frac{-E(p^N, r^N)}{k_B T}\right]} \quad (17)$$

$$\therefore \frac{1}{(8\pi^2 V)^N} \int \int dp^N dr^N \exp\left[\frac{-E(p^N, r^N)}{k_B T}\right] \exp\left[\frac{+E(p^N, r^N)}{k_B T}\right] = 1 \quad (18)$$

Inserting Equation (18) into the free energy Equation (17) and ignoring the constants:

$$A = k_B T \ln \frac{\int \int dp^N dr^N \exp\left[\frac{-E(p^N, r^N)}{k_B T}\right] \exp\left[\frac{+E(p^N, r^N)}{k_B T}\right]}{\int \int dp^N dr^N \exp\left[\frac{-E(p^N, r^N)}{k_B T}\right]} \quad (19)$$

Substituting the $\rho(p^N, r^N)$ for the probability density, we get:

$$A = k_B T \ln \int \int dp^N dr^N \exp\left[\frac{+E(p^N, r^N)}{k_B T}\right] \rho(p^N, r^N) \quad (20)$$

Equation (20) explains that high-energy states, despite their low probabilities, significantly contribute to free energy. Since the conventional molecular dynamics and Monte Carlo simulations do not sample high energy phase space with the same frequency as that of low energy, the results obtained for the free energy and other properties such as entropy as a consequence of inadequate sampling are inaccurate and poorly converged. In addition, integrating the free energy given by Equation (20) across $3N$ degrees of freedom is often not feasible.

8.4 Enhanced Sampling Methods for Free energy Calculations

8.4.1 Free Energy Perturbation

Given the constraints of the conventional MD associated with free energy determination, enhanced sampling techniques coupled with MD can be used to investigate the energy surface to retrieve the kinetic and thermodynamic properties. Free energy perturbation and umbrella sampling are some commonly used methods to calculate free energy changes. Steered MD and Meta-dynamics simulations are also often employed in drug discovery to monitor the ligand binding/unbinding events.

8.4.1.1 Focusing on the Calculation of Free Energy Differences

Researchers exploited the path independence of free energy, meaning that the difference in free energy does not rely on the route taken to break free from the limitations of Boltzmann sampling (conventional Monte Carlo/MD). If one is curious about the free energy difference ($\Delta A = A_Y - A_X$) between systems X and Y, represented by Hamiltonian H_X and H_Y , respectively, it can be calculated with the expression [Equation (21)]:

$$\Delta A = A_Y - A_X = -k_B T \ln \langle \exp\left(\frac{-\Delta H}{k_B T}\right) \rangle_X \quad (21)$$

Where $\Delta H = H_Y - H_X$ and $\langle \rangle_X$ refers to the ensemble average corresponding to the state X. Equation (21) is the basic concept of FEP calculations and is attributed to Zwanzig [57]. To implement the thermodynamic perturbation, one must define H_X and H_Y , and begin a simulation at state X to generate an ensemble average of $\exp\left[-\frac{\Delta H}{k_B T}\right]$ as the simulation proceeds. The idea is to calculate the energy for every instantaneous conformation corresponding to each configuration by mutating X to Y partially, i.e., X is assigned the potential energy parameters of Y partially. If the differences between X and Y systems are not negligible, i.e., if X and Y do not overlap in the phase space, then Equation (21) will not converge to give sensible free energy difference because the sample space of Y will not be sampled adequately while simulating X. This mainly arises when the energy difference ($H_Y - H_X$) between the states is much larger than the room temperature thermal energy, i.e., $|H_Y - H_X| \gg k_B T$. In such circumstances, one can, however, generalize the Hamiltonian using the coupling parameter λ as given below in Equation (22):

$$H(\lambda) = \lambda H_B + (1 - \lambda) H_A \quad (22)$$

Where λ ranges from 0, H_A to 1, H_B . Equation (21) then can be generalized:

$$\Delta A = A_Y - A_X = \int_{\lambda=0}^1 -k_B T \ln \langle \exp\left(\frac{-\Delta H^*}{k_B T}\right) \rangle_\lambda \quad (23)$$

Where $\Delta H^* = H_{\lambda+d\lambda} - H_\lambda$ i.e., one divides the free energy into windows, and several simulations are performed using molecular dynamics or Monte Carlo techniques, and a change in coupling parameter, $d\lambda$ is chosen (small enough) such that the free energy change converges in each simulation. The accuracy of H_X and H_Y determines the reliability of the free energy calculations obtained from Equation (21). As we all know, practically all methodologies assume that the Hamiltonian's kinetic energy term may be neglected. How realistic is this assumption? Let us consider the earliest application of free energy estimation of solvation of small chemical species: Jorgensen and Ravimohan's (JR) [28] analysis of the relative solvation free energy of methanol and ethane. Consider the free energy cycle shown in Figure (8.2).

8.4.1.1.1 The Relative Solvation-Free Energy of Methanol and Ethane

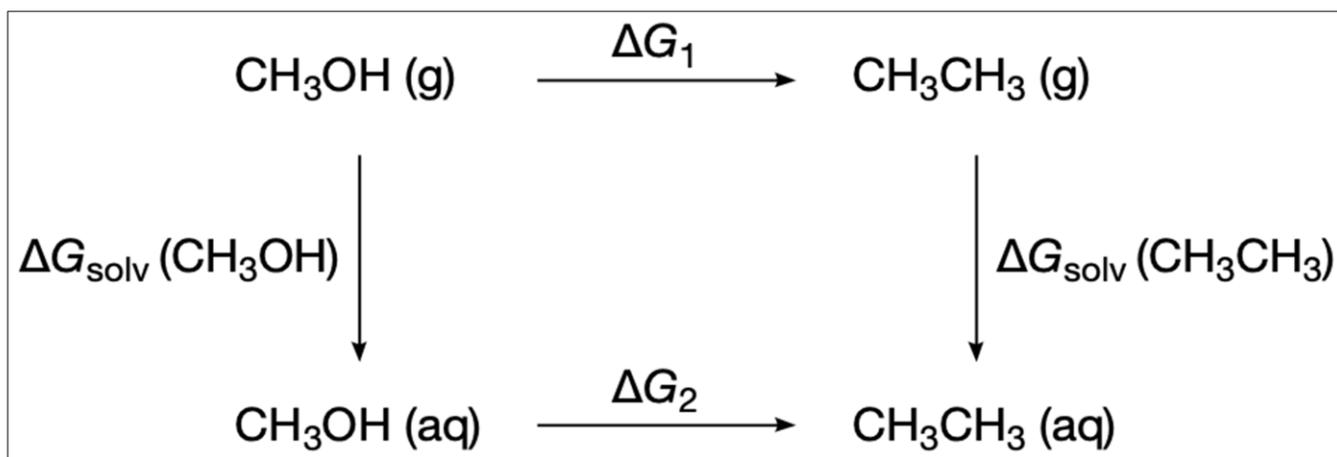


Figure 8.2: The free energy cycle of ethane and methanol solvation. The free energy difference of ethane and methanol solvation is given as $\Delta\Delta G = \Delta G_{\text{solv}}(\text{CH}_3\text{OH}) - \Delta G_{\text{solv}}(\text{CH}_3 - \text{CH}_3) = \Delta G_2 - \Delta G_1$, where ΔG_1 and ΔG_2 stands for the free energy change associated with the mutation of methanol into ethane in the gas and solution, respectively.

JR calculated ΔG_2 with the help of Equation (23) utilizing Monte Carlo simulations. JR's OPLS solute model includes a united atom CH_3 group, making CH_3OH a triatomic molecule and $\text{CH}_3 - \text{CH}_3$ a diatomic. They also presumed that changes in ΔG_2 due to kinetic energy variations would be equivalent when estimating ΔG_1 and ΔG_2 ; hence they were excluded from both calculations. Equation (24) below describes the potential energy functions used in the OPLS model by Weiner et al. [58, 59]:

$$V = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{\text{nonbonded } (i < j)} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] + \sum_{\text{H-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right] \quad (24)$$

Jorgensen and Tirado-Rives [42] adapted the first three terms, bond stretching, angle bending and torsional rotation using Equation (24) for bonding parameters, and employed non-bonding terms from their Monte Carlo calculations of requisite liquids for the molecular mechanics and dynamics. To employ free energy perturbation calculations using Equation (23), a variety of hybrid systems intermediate between CH_3OH and $\text{CH}_3 - \text{CH}_3$ were created. For instance, the C-O bond length in methanol is 1.43 Å, while the C-C bond length in ethane is 1.53 Å. So, an intermediate step ($\lambda = 0.5$) would have a 1.48 Å bond length between CH_3 group and the changing atom oxygen. Similarly, the charge on oxygen is -0.7, 0.265 on hydrogen (considering the methanol OPLS model) and zero on ethanol ($\lambda = 0$). Therefore, for ($\lambda = 0.5$), average charges corresponded to a 50% mutation of methanol to ethane. The van der Waals parameters were also similarly obtained for $\lambda = 0.5$.

Each term in the force field (Equation 24) for an intermediate can be written as a linear combination of two

states, methanol (X) and ethane (Y), using the coupling parameter λ .

$$k_r(\lambda) = \lambda k_r(Y) + (1 - \lambda)k_r(X) \quad (25)$$

$$r_0(\lambda) = \lambda r_0(Y) + (1 - \lambda)r_0(X) \quad (26)$$

$$k_\theta(\lambda) = \lambda k_\theta(Y) + (1 - \lambda)k_\theta(X) \quad (27)$$

$$\theta(\lambda) = \lambda \theta(Y) + (1 - \lambda)\theta(X) \quad (28)$$

$$V_n(\lambda) = \lambda V_n(Y) + (1 - \lambda)V_n(X) \quad (29)$$

$$q_i(\lambda) = \lambda q_i(Y) + (1 - \lambda)q_i(X) \quad (30)$$

$$q_j(\lambda) = \lambda q_j(Y) + (1 - \lambda)q_j(X) \quad (31)$$

$$\epsilon(\lambda) = \lambda \epsilon(Y) + (1 - \lambda)\epsilon(X) \quad (32)$$

JR simulated methanol, inserting it into a box of TIP4P water model and equilibrated the system using Monte Carlo method in the NPT ensemble. Then they mutated the system to $\lambda = 0.125$, (1/8 ethane and 7/8 methanol). In order to check for free energy convergence, the authors used a technique known as "double-wide sampling". The "double wide sampling" requires computing the free energy difference for both the intervals: $\lambda \rightarrow \lambda'$ and $\lambda' \rightarrow \lambda$. The free energy of transition from state λ to state λ' may be calculated from the ensemble average energy at state λ using Equation (23).

JR found that the free energy of contact with surrounding waters changes faster near methanol. So, they required more (contiguous) values of λ near methanol than ethane. By changing methanol to ethane to test ΔG , they discovered that the calculated ΔG , 6.75 kcal/mol, was in accordance with the observed ΔG value, 6.93 kcal/mol. Similarly, the next section discusses two more examples of how perturbation methods can be utilized to calculate a drug's binding affinity.

8.4.1.1.2 Calculation of a drug's binding affinity to the receptor/protein using the perturbation method

Consider the thermodynamic cycle shown in Figure (8.3) to calculate the binding energy (ΔG_{bind}) of a drug molecule [35, 60]. It is currently not possible to simulate a receptor-ligand system long enough to capture the entire binding event. The perturbation approach can still be used to determine a drug's binding affinity. The transformation is similar to the example discussed above. Since free energy is path independent, a system that moves from one state to the next in this cycle of free energy while staying in the same beginning state should not see any change in the overall amount of free energy change, i.e.

$$\Delta G_{bind} + \Delta G_{protein} - \Delta G - \Delta G_{water} = 0 \quad (33)$$

The ligand completely disappeared in both states (in solvent as well as in receptor-bound, as shown in the bottom half of Figure 8.3), implying that neither of these states can interact with solvent water or the receptor. Therefore, ΔG in Equation (33) equals zero. Thus,

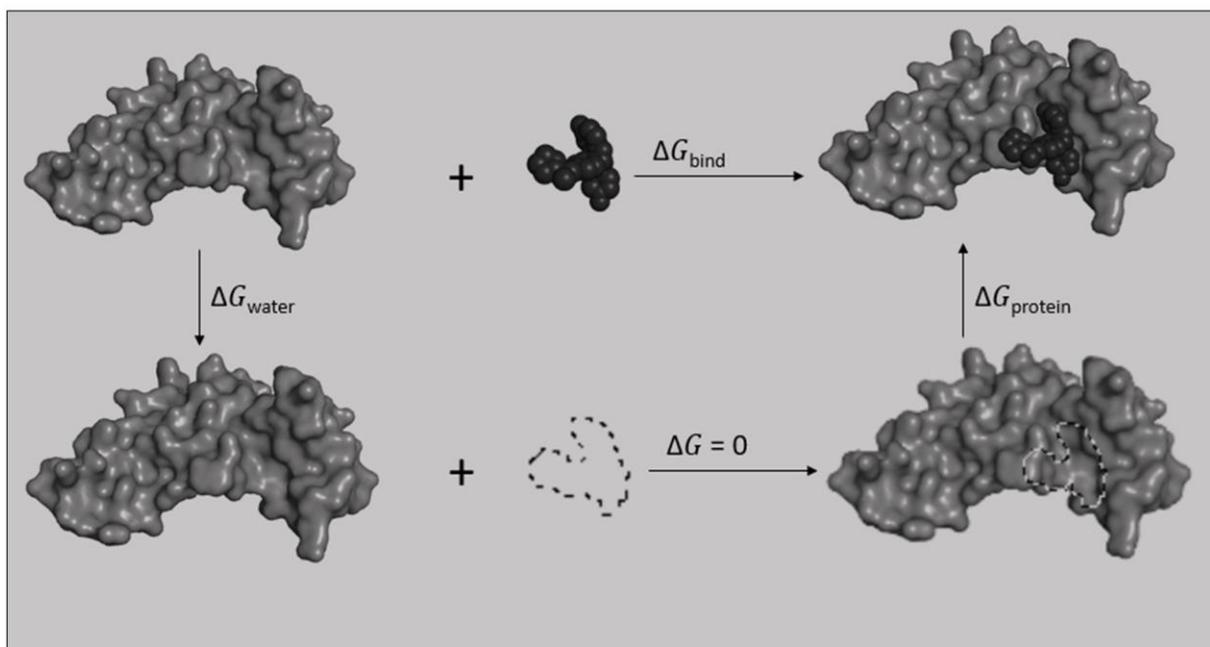


Figure 8.3: The thermodynamic cycle to estimate binding energy using FEP. ΔG_{bind} is free energy change when the drug binds to the protein/receptor. $\Delta G_{protein}$ refers to the change in free energy when a bound drug is gradually removed and ΔG_{water} is free energy change associated with desolvation of the drug molecule, whereas ΔG is free energy change when a ghost drug molecule binds to the protein. Since ghost drug molecules can not interact with a receptor, ΔG is always considered zero.

$$\Delta G_{bind} = \Delta G_{water} - \Delta G_{protein} \quad (34)$$

Equation (34) demonstrates that performing two simulations can yield the binding energy of a ligand/drug; one corresponding to the process where the receptor-bound ligand is annihilated gradually and another in which the solvated ligand/drug vanishes to estimate $\Delta G_{protein}$ and ΔG_{water} , respectively. The electrostatic and van der Waals forces generated by the ligand atoms are gradually reduced throughout a molecular dynamics simulation to prevent unwanted artefacts. Consequently, the ability of the ligand to bind with the protein or solvent eventually wears off.

A similar free energy estimation is employed in quantitative structure-activity relationship (QSAR) studies to calculate the relative ligand binding for drug optimization when one is interested in determining whether a functional group alteration can improve drug binding. In such cases, only a fragment/functional group is altered rather than annihilating the entire ligand. Figure (8.4) shows a thermodynamic cycle to compute relative free binding energy ($\Delta\Delta G_b = \Delta G_{bA} - \Delta G_{bB}$) between con-generic ligands A and B. The horizontal arrows in Figure (8.4) exhibit the physical process, and the vertical legs represent non-physical transition of ligand A into ligand B. One can estimate $\Delta\Delta G_b$ by performing two simulations [each corresponding to vertical leg in Figure (8.4)] same as discussed in the previous example.

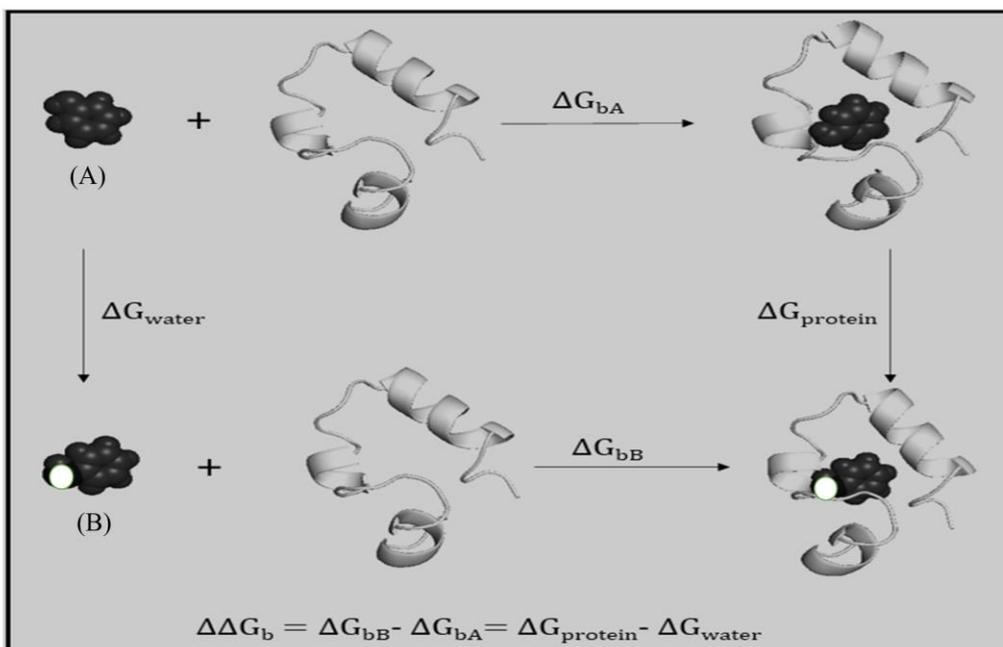


Figure 8.4: Thermodynamic cycle to estimate relative binding energies.

8.4.2 Umbrella Sampling

In the perturbation method discussed in the previous section, free energy was calculated by means of chemical mutation. It may be of interest to know how free energy varies in relation to a specific degree of freedom, for example, the change in free energy of binding of a ligand as a function of the distance between hydrogen bond donor and acceptor. The free energy surface as a parameter of a chosen variable is known as the potential of mean force (PMF). Unlike perturbation calculations (non-physical), it is possible to calculate PMF for a physical process, and umbrella sampling (US) is one of the methods used.

In order to adequately sample the unfavourable situations, umbrella sampling modifies the potential function in an attempt to solve the sampling issue. This technique can be implemented in both molecular dynamics and Monte Carlo simulations. It is possible to express the potential function change as perturbation:

$$V'(r^N) = V(r^N) + W(r^N) \quad (35)$$

Where $W(r^N)$ is a biased/weighing function, which is often quadratic and can be written as:

$$W(r^N) = k_w(r^N - r_0^N)^2 \quad (36)$$

Since the weighting function is largest for configurations far from the equilibrium state r_0^N , a simulation employing the modified energy function $V'(r^N)$ will be biased along the relevant reaction coordinates (also called collective variables (CVs)) away from the configuration r_0^N , which makes the overall distribution non-Boltzmann. Torrie and Valleau [25] introduced an approach to get the Boltzmann averages from the non-Boltzmann distribution, which is given by the Equation (37) below:

$$\langle M \rangle = \frac{\langle M(r^N) \exp[\frac{V(r^N)}{k_B T}] \rangle_W}{\langle \exp[\frac{V(r^N)}{k_B T}] \rangle_W} \quad (37)$$

The practical application of the US in pharmaceutical research is largely limited by its high processing expense, despite being one of the most precise techniques for determining free energy. For instance, to calculate the free energy of a ligand unbinding event from the protein, numerous overlapping windows must be created, properly equilibrated, and then sampled.

Woo and Roux published a useful umbrella sampling-based technique in 2006 for accurately calculating a peptide's binding affinity to the SH2 domain of human Lck kinase [61]. This method uses a one-dimensional PMF to describe the ligand (un)docking. The PMF can be estimated by utilising simulations with umbrella sampling to connect the protein and the ligand along a clearly defined axis called the CV. Constraining potentials (such as changes in ligand structure, orientation, or radial translations along the CV) are crucial because they allow for proper management of the transverse degrees of freedom that might alter as a result of ligand (un)binding. Integration of the PMF yields the equilibrium-binding constant, which in turn yields the absolute binding free energy. Later, this approach was effectively used for additional issues with pharmacological implications [62].

8.4.3 Steered MD

In this technique, the ligand is unbound from the target protein by applying an external potential that varies with time. Thus, a descriptor (or CV), often the receptor-ligand distance or a vector indicating the ligand escape route, may be used to speed up ligand undocking. The centre of the harmonic constraint is shifted to a limited velocity along the descriptor in order to induce a smooth escape of the ligand from the pocket [31, 32]. For this purpose, the ligand is attached to a spring whose force constant has been previously established. As a nonequilibrium technique, steered MD relies heavily on the values of the pulling velocity and the spring constant to accurately model the system under study. Although the basics are similar, Steered MD often involves far higher pulling velocities than those used in experimental AFM techniques. Notably, for each ligand under consideration, Steered MD enables the determination of the force applied and the external work performed on the system. This was initially shown by Grubmüller et al. [32] in their seminal study of the streptavidin-biotin complex. If the force constant is high, the rupture force grows linearly with the quantity of irreversible work. This gives us valuable information on the unbinding procedure, both qualitatively and semi-quantitatively. On the other hand, Jarzynski [63] established a crucial connection between reversible and irreversible work (i.e., the free-energy difference) in the year 1997. The so-called "Jarzynski equality" allowed the PMF for the studied process to be determined independently of the applied speed via a series of pulling simulations.

One interesting study used steered molecular dynamics (MD)-based rational drug design to compare how five structurally identical flavonoids bound to FabZ, a protein that might be targeted by antimalarial medicines [64]. Steered MD simulations were performed to determine the amount of force required to dislodge each of the tested ligands from the target protein, much similar to single-molecule tugging assays. In contrast to weakly bound ligands, which produced a flatter force profile, strongly bound inhibitors produced profiles with greater peak forces. Steered MD was used to design and forecast a novel flavonoid, and experimental results were used to confirm the validity of the method for computational drug development. Although the calculated observables (forces necessary to remove each inhibitor from the targeted enzyme) remained only qualitative, this archetypal investigation clearly showed that Steered MD could differentiate between active and inactive inhibitors [64]. Moreover, the unbinding force profile may be used to enhance the docking score function or

as postprocessing in virtual screening. It is possible to get a more thorough theoretical explanation of steered MD elsewhere. [65].

8.5 Rapid/Approximate Methods For Free Energy Calculation: Representation of Solvent

A majority of chemical transformations occur in the presence of a solvent. Therefore, it is essential to consider the solvent effects and their influence on the overall behaviour of chemical species. In cases where the solvent directly interacts with the solute, one needs to treat the solvent molecules explicitly. While in situations where the solvent does not directly interact with the solute but instead influences how the solute behaves, the solvent molecules may not require to be treated explicitly, but special treatment, such as mean-field theories, may be required. In the third case, the solvent acts as a bulk medium and affects the solute behavior via its dielectric properties; it is implicitly modeled. The introduction of continuum solvent models [66] perturbs the solute's behavior in the gas phase by the solvent. A variety of solvent models have been proposed both using classical as well as quantum mechanics [67]. We will only discuss the most widely used ones.

8.5.1 Thermodynamic Background

Free energy of solvation is defined as the energy required to move a molecule from a vacuum to a solvent and can be calculated from Equation (38):

$$\Delta G_{sol} = \Delta G_{ele} + \Delta G_{vdw} + \Delta G_{cav} \quad (38)$$

Where ΔG_{ele} stands for the free energy change associated with electrostatic interactions. In the case of polar solutes ΔG_{ele} significantly contributes to the overall free energy. Similarly, ΔG_{vdw} represents the change in the free energy due to van der Waals's interactions between solute and solvent. Lastly, ΔG_{cav} is free energy associated with work to create a solvent cavity. When the cavity is created, reorganization of solvent molecules takes place as work is done against the solvent pressure. Therefore, the ΔG_{cav} is positive. We will discuss all three aspects separately in detail in the following sections.

8.5.1.1 The Born and Onsager Models: Calculation of electrostatic contribution

The generalized Born model [68] views molecules as a collection of charged spheres that overlap and are embedded in a polarizable dielectric medium. The generalized Born theory is characterised by the following Equations (39-43):

$$G_{ele} = 332 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{q_i q_j}{r_{ij} \epsilon_r} - 166 \left(1 - \frac{1}{\epsilon_r}\right) \sum_{i=1}^n \frac{q_i^2}{a_i} \quad (39)$$

$$= 332 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{q_i q_j}{r_{ij}} + G_{pol} \quad (40)$$

$$G_{pol} = -166 \left(1 - \frac{1}{\epsilon_r}\right) \sum_{i=1}^{n-1} \sum_{j=1}^n \frac{q_i q_j}{f_{GB}} \quad (41)$$

$$f_{GB} = (r_{ij}^2 + a_{ij}^2 e^{-D})^{0.5}; a_{ij} = (a_i a_j)^{0.5}; D = \frac{r_{ij}^2}{4a_{ij}^2} \quad (42)$$

$$G_{pol} = -166\left(1 - \frac{1}{\epsilon_r}\right) \sum_{i=1}^n \sum_{j=1}^n \frac{q_i q_j}{f_{GB}} - 166\left(1 - \frac{1}{\epsilon_r}\right) \sum_{i=1}^n \frac{q_i^2}{a_i} \quad (43)$$

The total electrostatic free energy of a molecule is given by Equation (39), where the first component is the Coulomb interaction energy between any two charges q_i and q_j separated by a distance r_{ij} in a solvent of dielectric constant ϵ_r , and the second term represents the Born (self) solvation energy. a_i stands for Born radii. The free energy is represented as the sum of the polarisation free energy G_{pol} and the Coulombic interaction free energy in Equation (40). With careful selection of the effective distance parameter f_{GB} , as indicated by Equation (42), the GB polarisation energy captures all the electrostatic effects related to the solvent in a single term (Equation 41). As can be shown in Equation (43), the GB polarisation energy is a combination of solvent-shielding and self-energy components.

The Born model is only applicable for charged species, and this is where Onsager's model [69] becomes relevant, and there is a much wider range of molecules for which Onsager's dipole model holds true. An electric field (referred to as the reaction field) is generated within the cavity when the dipole of the solute causes a perturbation in the surrounding solvent environment and the solvent acts back on the dipole in the cavity. The electric field generated interacts with the solute and stabilizes the system. Expression of the electric field (ϕ_{RF}) generated is given by Equation (44):

$$\phi_{RF} = \frac{2(\epsilon_r - 1)}{(2\epsilon_r + 1)a^3} \mu \quad (44)$$

Here μ represents the dipole moment of the solute. ϵ_r and 'a' have their usual meaning, as mentioned before. For a polarizable dipole, work done (of the magnitude $\frac{\phi_{RF}\mu}{2}$), assembling the charge distribution within the cavity makes an additional contribution to electrostatic interaction. Therefore, the overall free energy expression becomes:

$$\Delta G_{ele} = -\frac{\phi_{RF}\mu}{2} \quad (45)$$

Quantum mechanical treatment of the electric field generated from the polarization of the solvent is possible with the help of perturbation of the Hamiltonian of an isolated molecule (H_0). The reaction field generated is denoted as a self-consistent reaction field (SCRf) and given as the Equation (46):

$$H_{tot} = H_0 + H_{RF} \quad (46)$$

Where H_{RF} stands for perturbed Hamiltonian and is given by:

$$H_{RF} = -\hat{\mu}^T \frac{2(\epsilon-1)}{(2\epsilon+1)a^3} \langle \Psi | \hat{\mu} | \Psi \rangle \quad (47)$$

Where $\hat{\mu}$ is the dipole moment operator matrix and $\hat{\mu}^T$ is the transpose of that matrix. Using the perturbed Hamiltonian's wave function Ψ and Equation (48), one can calculate the electrostatic part of the free energy:

$$\Delta G_{ele} = \langle \Psi | H_{tot} | \Psi \rangle - \langle \Psi_0 | H_0 | \Psi_0 \rangle + \frac{1}{2} \frac{2(\epsilon-1)}{(2\epsilon+1)a^3} \mu^2 \quad (48)$$

Ψ_0 is the wave function in the gas phase, and the third additional term refers to the work done to create the charge distribution of the solute molecule.

The SCRF approach has a disadvantage since it uses a spherical cavity, yet molecules are seldom perfectly round. Alternatively, better outcomes may be attained by contemplating an ellipsoidal cavity, which may be closer to the shape of several molecules. A different method involves deducing the molecule's volume from an appropriate electron density map. Since the solvent molecule cannot approach the solute molecules directly due to the interatomic repulsion, the radii acquired via these methods are sometimes corrected by adding an empirical constant. A more realistic approach to determining the cavity size is to calculate the atomic radii of the solute, which is widely used in polarizable continuum methods.

8.5.1.2 Methods Based on the Poisson-Boltzmann Equation

Finally, we review here approaches related to the Poisson/Poisson-Boltzmann equations for determining the electrostatic interactions and their contribution to the overall free energy. The Poisson equation treats the solvent as a continuum of high dielectric and has been shown to be very helpful for investigating the polar characteristics of biological molecules such as proteins, RNA and DNA. If ϵ_r be the dielectric of the medium (function of the position, varies with position/distance) and ρ be the charge density, the variation in the potential generated within the medium is given by:

$$\nabla[\epsilon_r \nabla \phi(r)] = -\frac{\rho(r)}{\epsilon_0} = -\frac{4\pi\rho(r)}{4\pi\epsilon_0} \quad (49)$$

The same can be written in the reduced electrostatic units as:

$$\nabla[\epsilon_r \nabla \phi(r)] = -4\pi\rho(r) \quad (50)$$

When the dielectric of the medium does not vary with distance/position, Equation (50) becomes:

$$\nabla^2 \phi(r) = -\frac{4\pi\rho(r)}{\epsilon_r} \quad (51)$$

When mobile ions are present or salt is added, the Poisson Equation must be changed to account for the small ion (counter and co-ion) distribution in the solution in response to the electric potential. Due to repulsive interactions with other ions and their inherent thermal motion, the ions are precluded from accumulating at the sites of severe electrostatic potential. The Boltzmann distribution is used to describe the spatial organization of ions as given below in Equation (52):

$$n(r) = N \exp(-V(r)/k_B T) \quad (52)$$

N represents the bulk number density, $V(r)$ is the energy change needed to move one ion from infinity to the position r , and the number density of ions at a position r is given by $n(r)$. The Poisson-Boltzmann Equation (53) emerges when these effects are added to the Poisson Equation (50):

$$\nabla[\epsilon_r \nabla \phi(r)] - \kappa'^2 \sinh[\phi(r)] = -4\pi\rho(r) \quad (53)$$

Where κ' is the Debye-Huckel inverse length and given by:

$$\kappa^2 = \frac{\kappa'^2}{\epsilon_r} = \frac{8\pi N_A e^2 I}{1000 \epsilon k_B T} \quad (54)$$

Where e and I stand for electronic charge and ionic strength of the solution, respectively. Expanding the sinh using the Taylor series:

$$\nabla[\epsilon_r \nabla \phi(r)] - \kappa'^2 \phi(r) \left[1 + \frac{\phi(r)^2}{6} + \frac{\phi(r)^4}{120} + \dots \right] = -4\pi\rho(r) \quad (55)$$

Approximating Equation (55) by only considering the first term in the expansion:

$$\nabla[\epsilon_r \nabla \phi(r)] - \kappa'^2 \phi(r) = -4\pi\rho(r) \quad (56)$$

In the absence of computers, only elementary geometries could be studied. Proteins, for instance, were represented as spherical or elliptical shapes (Tanford-Kirkwood theory), DNA as a cylindrical structure with a constant charge, and membranes as flat surfaces (Gouy-Chapman theory). The Poisson-Boltzmann Equation (56) can now be solved for arbitrary shaped solutes/biomolecules using numerical methods on modern computers.

8.5.1.3 Free energies via analytical solutions to Laplace and Poisson-Boltzmann equations for simple geometries with multi-layered solvent descriptions

For an ion in a solvent, Jayaram and Beveridge [70] used PB equations [Equation (58)] for ion atmosphere and Laplace [Equation (57)] for solvent effects to calculate the electrostatic free energy:

$$\nabla^2 \phi = 0 \quad (57)$$

$$(\nabla^2 - \kappa'^2)\phi = 0 \quad (58)$$

For an ion/solute in the spherical cavity, the general solution of Equation (57) can be written as:

$$\phi = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} \left(B_{nm} r^n + \frac{E_{mn}}{r^{n+1}} \right) P_n^m(\cos\theta) e^{-im\phi} \quad (59)$$

And for Equation (58), it is:

$$\phi = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} \left[\frac{C_{nm}}{r^{(n+1)}} e^{-\kappa r} X_n(\kappa r) \right] P_n^m(\cos\theta) e^{im\phi} \quad (60)$$

$$X_n(\kappa r) = \sum_{s=0}^n \left[\frac{2^s n! (2n-s)!}{s! (2n)! (n-s)!} \right] (\kappa r)^s \quad (61)$$

Where B_{nm} and C_{nm} are constants, $P_n^m \cos\theta$ are Legendre polynomials, and E_{mn} is associated with charge distribution in the cavity. Due to the charge distribution/polarization, there exist three different regions of varying dielectric. Region 1: medium inside the cavity of radius a with dielectric ϵ_i . Region 2: medium surrounding the cavity with dielectric ϵ_{local} in a radius of the thickness of $(b-a)$, and Region 3: the bulk medium having dielectric ϵ_0 . The potential due to the ϵ_i inside the cavity can be given as:

$$\phi_i = \frac{1}{\epsilon_i} \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} (B_{nm}r^n + F_{nm}r^n + \frac{E_{mn}}{r^{(n+1)}})P_n^m(\cos\theta)e^{im\phi} \quad (62)$$

Similarly, The potential due to the charge distribution around the cavity can be given as:

$$\phi_R = \frac{1}{\epsilon_i} \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} (B_{nm}r^n + F_{nm}r^n)P_n^m(\cos\theta)e^{im\phi} \quad (63)$$

Using appropriate boundary condition between the three regions, the electrostatic contribution to solvation free energy becomes:

$$\phi_R = \frac{1}{\epsilon_i} \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} (B_{nm}r^n + F_{nm}r^n)P_n^m(\cos\theta)e^{im\phi} \quad (64)$$

$$\begin{aligned} A &= \frac{1}{2} \sum_k q_k \phi_R(r_k) \\ &= \frac{1}{2\epsilon_i} \sum_{n=0}^{\infty} \left(\left[\frac{(n+1)(1-\epsilon'_a)}{(n+1)\epsilon'_a+n} \right] \frac{Q_n}{a^{2n+1}} + \left[\frac{(n+1)(1-\epsilon'_b)}{(n+1)\epsilon'_b+n} \right] \left[1 - \frac{n(1-\epsilon'_a)}{(n+1)\epsilon'_a+n} \right] \frac{Q_n}{b^{2n+1}} \right) \end{aligned} \quad (65)$$

$$\text{Where:} \quad Q_n = \sum_k \sum_l q_k q_l r_k r_l P_n(\cos\theta_{kl}) \quad (66)$$

$$\epsilon_{a'} = \epsilon_a / \left(1 + \frac{(n+1)(1-\epsilon_a)(1-\epsilon_b)}{[(n+1)\epsilon_b+n]} \frac{a^{(2n+1)}}{b^{(2n+1)}} \right) \quad (67)$$

The authors extended their study to non-spherical cavities to improve the results, for example, a coaxial cylindrical continuum to investigate electrostatic contribution for DNA [52] and later in 1994, generalised to interactions between two arbitrary charge distributions [71].

8.5.1.4 Non-electrostatic Contribution to the Solvation-Free Energy

Only the electrostatic component of the free energy of solvation has been considered in the account above. Even while this is significant, there are certainly other elements that also contribute to the total free energy of solvation. These additional contributions may be particularly important for solutes that are neither charged nor extremely polar. The contributions due to the van der Waals and cavity terms to the free energy are often combined and written as Equation (68) given below:

$$\Delta G_{cav} + \Delta G_{vdw} = \gamma A + b \quad (68)$$

Where γ and b are constants, and A stands for the total solvent accessible area. Free energy of transfer of alkanes from vacuum to water is often used to estimate γ and b . Lastly, ΔG_{cav} is free energy associated with work done to create a cavity in the solvent (as discussed before). The value of b in Equation (68) is often considered to be zero, which makes the surface area proportional to the cavity and van der Waals terms. As a result, the van der Waals and cavity terms should roughly correlate with the solvent-accessible surface area. The first solvation shell solvent molecules are those that are most impacted by this reorganization. The solute-solvent van der Waals interaction energy would mostly rely on how many solvent molecules are present in

the initial solvent shell, as van der Waals interactions are short-range. Therefore, to a first approximation, the number of solvent molecules in the initial solvation shell is proportional to the solute surface area that the solvent may contact. Thus, the computation of solvent accessible area along with γ provides the free energy contribution due to the van der Waals and cavity terms.

8.6 Applications of PBSA/GBSA methods to determine the binding energetics of a ligand to its biomolecular target (receptor)

The receptor-ligand binding energy for the process, $R + L \rightleftharpoons RL$, can be represented as follows:[72-76]:

$$\Delta G_{binding} = G_{complex} - G_{Receptor} - G_{ligand} \quad (69)$$

Where $\Delta G_{binding}$ stands for the free energy of ligand binding to the receptor. G_{ligand} and $G_{receptor}$ refer to the free energy of solvation of ligand and receptor, respectively. The expressions for the individual terms in Equation (69) can be given by [73, 77-79]:

$$G_y = \langle G_{solvation} \rangle - TS + \langle E_{MM} \rangle \quad (70)$$

Where y refers to the protein-ligand or protein or ligand, T and S represent the temperature and entropy, respectively. The term TS combinedly denotes the entropic contribution to the free energy. The average molecular mechanical potential energy (in vacuum) is given by $\langle E_{MM} \rangle$ and the $\langle G_{solvation} \rangle$ stands for the average free energy of solvation.

8.6.1 Molecular Mechanics Potential Energy ($\langle E_{MM} \rangle$)

It consists of energy terms (calculated using molecular mechanics (MM) force field [43, 80, 81]) for bonded and nonbonded interactions in vacuum. The energy expressions are given in Equation (71) below:

$$E_{MM} = E_{bonded} + E_{non-bonded} = E_{bonded} + (E_{vwd} + E_{ele}) \quad (71)$$

Interactions involving bond length, angle, and dihedrals are incorporated in bonded energy terms. Electrostatic and van der Waal's interactions are modelled using Coulombic and Lennard-Jones (also known as 12-6 potential) potentials, respectively. One thing here to note is that, in a single simulation approach, the conformation of the protein and ligand remains the same in both bound and unbound states. Therefore, ΔE_{MM} is always taken zero [82]. $\langle G_{solvation} \rangle$ is calculated as follows [74, 83, 84]:

$$G_{solvation} = G_{polar} + G_{non-polar}$$

G_{polar} is calculated by solving the Poisson-Boltzmann equation (as discussed in section 8.5.1.2 and 8.5.1.3), whereas $G_{non-polar}$ is evaluated by solving Equation (68).

8.7 Combined Quantum Mechanical/Molecular Mechanical methods

Utilizing both quantum and molecular mechanics is another method for simulating chemical processes in solutions. The system is handled quantum mechanically for the "reacting" components, and a force field is used to represent the remaining components. The expression of the energy for such a system can be described as Equation (72):

$$E_{TOT} = E_{MM} + E_{QM} + E_{QM/MM} \quad (72)$$

Where E_{MM} is the energy of the system's molecular mechanical components, and E_{QM} is the energy of those portions of the system treated by quantum mechanics. The quantum mechanical and molecular mechanical components of the system interact with each other via energy denoted by the term $E_{QM/MM}$. In some instances, nonbonded interactions between quantum mechanical and molecular mechanical atoms are the only cause of $E_{QM/MM}$ term. An example of the same could be ions in a solvent, where ions are treated quantum mechanically, whereas the solvent behaves classically. The Hamiltonian for the QM/MM region can be written as Equation (73) below:

$$H_{QM/MM} = -\sum_i \sum_M \frac{q_M}{r_{i,M}} + \sum_\alpha \sum_M \frac{Z_\alpha q_M}{R_{\alpha,M}} + \sum_\alpha \sum_M \left(\frac{A_{\alpha M}}{R_{\alpha,M}^{12}} - \frac{C_{\alpha,M}}{R_{\alpha,M}^6} \right) \quad (73)$$

In Equation (73), the subscripts i and ' α ' stand for a quantum mechanical electron and nucleus, respectively. A subscript M stands for a molecular mechanical nucleus, and q_M is its partial atomic charge. Thus, there exist electrostatic interactions between the molecular mechanical nuclei and the electrons of the quantum mechanical component and electrostatic interactions between molecular mechanical and quantum mechanical nuclei. Finally, quantum mechanical and molecular mechanical atoms interact via van der Waal's expression. The last two terms in Equation (73) are independent of electronic coordinates. Therefore, they do not change for a given nuclear configuration. Quantum mechanical calculations are required to solve the first term via one-electron integrals added to the one-electron matrix, H^{core} . The one-electron integral can be given as Equation (74):

$$\phi_\mu(1) \frac{1}{r_{1,M}} \phi(1) dv(1) \quad (74)$$

The cases where a bond shares both Molecular Mechanics and Quantum Mechanics regions, the $E_{QM/MM}$ must contain terms to describe these interactions. One of the ways is to introduce molecular mechanics like energy terms having bond stretching, angle bending and dihedral parameters for the atoms from both regions. Two general approaches have been adapted to deal with such a junction. In one, a hybrid sp^2 orbital with one electron is established along the QM-MM [85]. Whereas in other, hydrogen atoms are used as link atoms to ensure that the valency is maintained.

8.7.1 1st QM/MM Study: Theoretical Studies of Enzymic Reaction

The first application of QM/MM investigated the stability of carbonium ions, which are created when lysozyme breaks the glycosidic linkage [85]. This method considers the whole enzyme-substrate complex in addition to the solvent to analyse all the quantum mechanical and classical energy potential components that may affect the reaction pathway.

The partition of the total potential energy can be given as Equation (72). The Classical (MM) potential adapted in the study is given in Equation below (75):

$$V_{MM} = \sum_i K_b (b_i - b_0)^2 + \sum_i K_\theta (\theta_i - \theta_0)^2 + \sum_i K_\phi \cos\{n(\phi_i - \phi_0)\} \\ + \sum_{i>j} \epsilon_{i,j} \left\{ \left(\frac{r_{i,j}^0}{r_{i,j}} \right)^{12} - 2 \left(\frac{r_{i,j}^0}{r_{i,j}} \right)^6 \right\} + \sum_{i>j} \frac{Q_i Q_j}{r_{i,j}} \quad (75)$$

Where b_i , θ_i , ϕ_i , $r_{i,j}$ stands for bond length, bond angles, torsional angles and interatomic distances, respectively. The values of the internal parameters K_b , K_θ , K_ϕ , θ_0 , b_0 , ϕ and n are obtained from the spectroscopic data [86]. Crystal structures of hydrocarbons, amides, and amino acids were analyzed, and the atomic partial charges Q_i and nonbonded parameters $\epsilon_{i,j}$, $r_{i,j}^0$ were calculated to suit the experimental data.

The QCFF/PI approach was used to determine the quantum mechanical part of the investigation [87, 88]. In order to include all the valence atoms, the method is extended to a semi-empirical method QCFF/ALL. In this method, neighbouring orbitals on the same atom that are not necessarily orthogonal, overlap using hybrid atomic orbitals with Löwdin corrections. Hybrid orbitals also have the unique ability to express the covalent connection between a molecule's quantum and classical components in a single orbital. The approach yields the first derivatives of potential energy, which makes energy minimization easy to perform. Finally, the coupling component sharing both classical and quantum mechanical parts is given by Equation (76):

$$V_{QM/MM} = \sum_{i,j} \frac{Q_i Q_j}{r_{i,j}} + \sum_{i,j} \epsilon_{i,j} \left\{ \left(\frac{r_{i,j}^0}{r_{i,j}} \right)^{12} - 2 \left(\frac{r_{i,j}^0}{r_{i,j}} \right)^6 \right\} + V_{ind}^E + V_{ind}^W \quad (76)$$

Where i represents quantum, while j is in the classical region. The first and second terms, respectively, describe electrostatic and van der Waal's interactions. The third and fourth terms represent the electrostatic and van der Waal's interactions due to the induced dipoles and the dielectric. It is beyond the scope of this chapter to go into depth on the interactions brought on by polarization. Still, it is in the reader's best interest to check out the whole issue [85].

By including the polarizability of atoms in the computation, it, for the first time, became feasible to accurately replicate the energy balance seen in hydrogen transfer processes and account for electrostatic interactions. The approach primarily investigates the elements that influence the intermediate carbonium ion's stability. As a result, steric strain quickly relaxes in response to even modest changes in substrate and enzyme coordinates. It is probable that steric strain plays little or no function in the catalytic process. On the other hand, electrostatic interactions were discovered to be crucial to enzyme activity. It was discovered that the carbonium ion has a very high electrostatic stabilization (-40kcal/mol) compared to the vacuum. This demonstrates how the polarizability model can be used to explain the stabilization of ionized groups in proteins. Relative stabilization of the carbonium ion is often significantly less than 40 kcal/mol, as measured by comparing the enzyme's reaction rate and transition state stabilization to that in water. Now that it has been shown that the polarizability of the protein is primarily responsible for the considerable solvation effect, the question that has to be answered is to what extent the ionisation of neighbouring Asp52 leads to the overall stabilisation. Asp52 was shown to decrease the carbonium energy by around 9 kcal/mol compared to its ground state. The authors also investigated the substrate's D ring's transition state and relative ground state stability. The detailed discussion of the other analysis is out of the scope of this study. The reader can follow the complete article for more information [85]. With the introduction of QM/MM methods, it is now feasible to make quantum mechanical calculations of bond-making/breaking and ionic processes. The addition of a microscopic dielectric effect in QM/MM has made it possible to calculate the energy balance of real chemical and physiological processes, where the energy changes are of the order of 10 kcal/mol.

The above is a bird's eye view of the diverse methodologies adopted for molecular dynamics simulations.

Some questions do arise as to how molecular dynamics becomes necessary for drug discovery.

8.8 Why molecular dynamics?

We know that combinatorial and fragment-based drug discovery (FBDD) are the classical ways of drug discovery which takes 12-14 years for single drug development with a cost of \$2 billion or more [89]. The rise of molecular dynamics in FBDD [90] has made the process more efficient and has been shown to reduce cost and time. The latest example is the covid 19 pandemic, which created an urgency for a quick search for therapeutic strategies to combat the SARS-Cov-2. Pathak and co-workers [91] utilized the database of FDA-approved drugs to propose 16 candidates with low toxicity and higher efficacy within days. As discussed in other chapters, virtual screening and docking deal with only one protein conformation and do not consider the molecular flexibility, solvent medium and temperature of the system. Binding energies obtained from these methods do not always correlate well with experimental data. Therefore, virtual screening and docking are considered only as preliminary but essential studies in the process of drug discovery. In comparison, molecular dynamics takes care of the flexibility of a protein molecule upon ligand binding by considering all the conformers of a protein as well as ligand over the entire trajectory for binding energy calculations along with solvent and temperature effects. The following subsections will briefly discuss why molecular dynamics and free energy computations have proved invaluable, along with virtual screening and docking in drug discovery.

8.8.1 Validation of molecular dynamics on FDA-approved drugs on their respective targets

To start with, it would be best to have a look at how molecular dynamics, along with screening and docking, performs on the existing FDA-approved drugs. Ruchika et al. [92] considered some life-threatening diseases and their target proteins to validate their drug discovery pipeline against FDA-approved drugs. They utilized the *Sanjeevini* drug design software [93, 94] and its tools for screening, docking and scoring as the software has been benchmarked with many success stories [95]. Figure (8.5) below represents the flow chart of CADD used in *Sanjeevini* software.

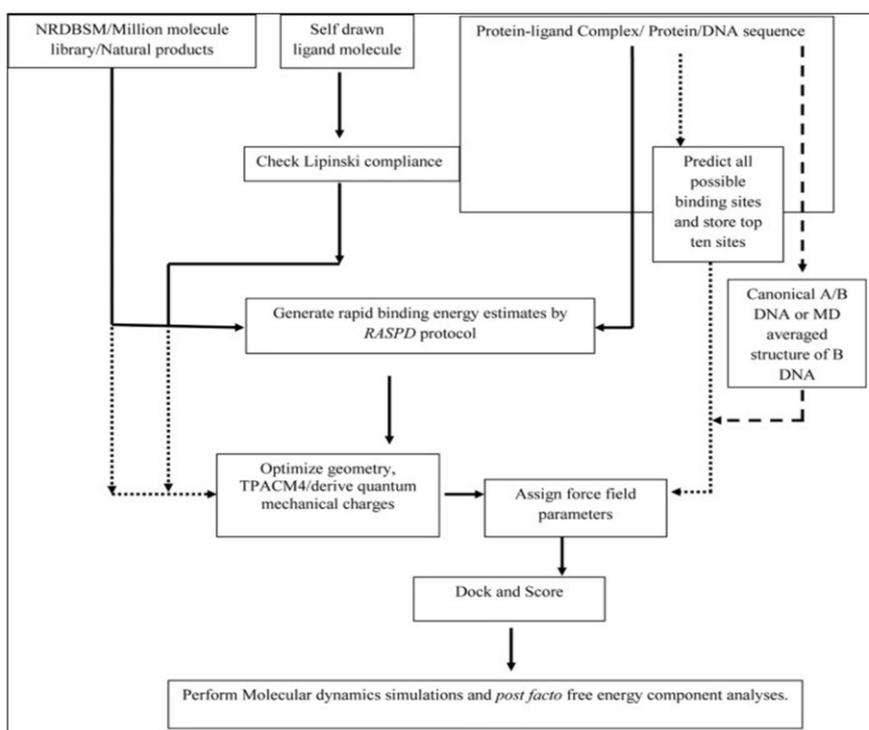


Figure 8.5: Workflow of CADD adopted in Sanjeevini software.

The authors considered 33 different proteins as targets for 17 life-threatening diseases and validated the methodology against their FDA-approved drugs (a total of 111 in number). In ~90% of cases, the known FDA-approved drugs were found to be the hit molecules identified. Therefore, molecular dynamics with screening and docking increases the reliability of CADD. Likewise, Tripathi et al. [96] conducted *in silico* studies against FDA-approved drugs and identified Novobiocin and Telmisartan as inhibitors of chikungunya nsP2 protease and validated them experimentally. Bhasker et al. [97] also identified Novobiocin as a Heat Shock Protein inhibitor. This is where *in silico* methods become very efficient as the FDA-approved drugs do not require much experimental testing and have very low toxicity.

8.8.2 Molecular dynamics to determine the mode of action

The mode of action of any drug has close associativity with the molecular motion and flexibility of a protein, and this is where dynamics is desirable over static structures. Artemisinin is a molecule that helps in preventing malarial symptoms [98]. However, the mechanism of action of Artemisinin is not yet fully known. One school of thought contends that the cleavage of the peroxide bridge in the presence of ferrous ion (Fe^{2+}) from heme (as the parasite is rich in heme iron) results in the production of highly reactive free radicals that quickly reorganise into more stable carbon-centred radicals [99, 100]. Many different parasite molecules may be chemically modified and inhibited by these free radicals created by artemisinin, leading to the parasite's demise [101]. Another hypothesis claims that Artemisinin directly interacts with the only SERCA-type Ca^{2+} -ATPase, PfATP6 (calcium pump), found in the malarial parasite [102]. In the previous studies, PfATP6 was modelled using 1IWO [103, 104] and 2DQS [105] as templates. Naik et al. [104] and Jung et al. [103] performed docking studies of Artemisinin with the models predicted and demonstrated the antimalarial activity using *in vitro* antimalarial studies. However, a correlation was not obtained with other antimalarial compounds [105]. The possible reason could be the docking of Artemisinin to a static closed conformation of PfATP6. To address the issue associated with molecular flexibility Ashutosh Shandilya et al. [106] proposed the mode of action of Artemisinin by involving the dynamics of different complexes with an open conformation of PfATP6 (modelled using 1SU4): PfATP6 enzyme, Artemisinin bound PfATP6 and Fe-Artemisinin adduct bound PfATP6. The authors also performed the same analysis for Thapsigargin (an endogenous molecule that binds to mammalian SERCA and is responsible for the open-to-close conformational transition to induce Ca pumping). The free energy of binding (kcal/mol) of Artemisinin, Fe-Artemisinin and Thapsigargin was found to be -6.5, -8.3 and -6.7 for PfATP6 (modelled using 1SU4), and -4.2, -5.1 and -9.1 for Mammalian SERCA, respectively.

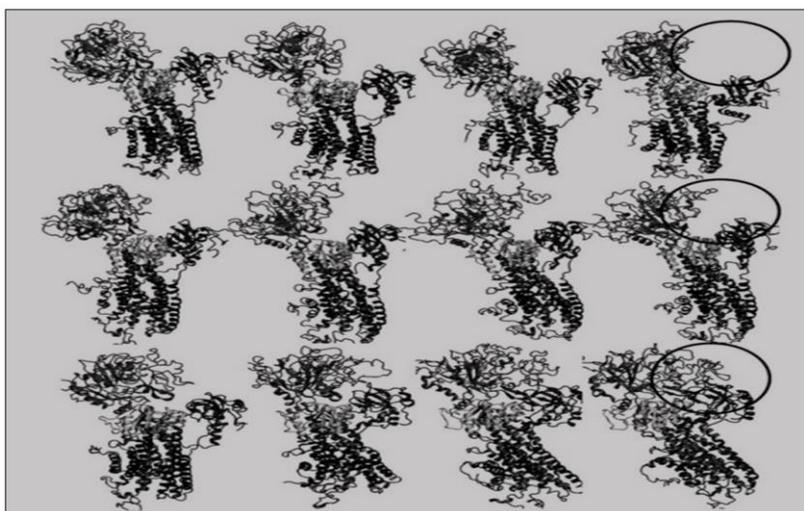


Figure 8.6: MD snapshots of the PfATP6 enzyme (top row), artemisinin-bound PfATP6 (second row), and Fe-artemisinin adduct (third row) bound PfATP6 were created from separate trajectories every 25 ns.

To investigate the high binding associated with the Fe-Artemisinin adduct, the authors performed the trajectory analysis. Figure (8.6) exhibits the snapshots generated from each of the complexes studied. It is evident from Figure (8.6) that Fe-Artemisinin adduct triggers the open-to-close conformational transition to arrest the Ca pumping, whereas the rest of the two do not undergo any significant conformational changes, suggesting the mode of action of Artemisinin. Authors performed several analyses such as distance and principal component analysis (details can be found in their article) and finally concluded that when Fe-Artemisinin binds to an open jaw-like conformation having an actuator, phosphorylation and nucleotide-binding domains, it leads to a close structure leading to the closure of all their domains causing an inability to reach the calcium-binding site, which leads to the parasite's Ca pump failing and clearance of the parasite from the infected host as a consequence.

Similarly, MD revealed the identification of the right conformation of thermolysin to be inhibited by β -phenylpropionyl-L-phenylalanine among many favourable conformers [107]. Unlike MD, electronic and structural analysis of lead-like molecules can also be used to devise the mode of action. Bharatam [108] studied the pharmacophore features of biguanide derivatives to determine their mode of action. Thus, electronic and structural analysis of lead-like molecules can also find application in CADD.

8.8.3 Molecular dynamics to study mutational effects

Even a point mutation in a protein often results in the loss of function of the protein. Molecular dynamics helps to study the structural changes causing the loss of function. One example is the mutations in angiogenin protein (ANG) that results in a fatal neurodegenerative disorder called amyotrophic lateral sclerosis (ALS). A catalytic triad His13, Lys40 and His114 is essential for the neuroprotective function of wild-type ANG. Aditya Padhi et al. [109] studied the molecular dynamics simulation of the various mutants such as G20S, P38R, R51H, P88H, R95L, A98V and P123L to understand the structural and functional changes upon mutation and found that conformational changes of His114 are responsible for the loss of ribonucleic activity in certain ANG mutants. For further validation, they extracted several snapshots from the MD trajectory at both native and altered His114 conformations and performed docking studies with a known inhibitor of ANG. The average binding free energy for the mutant having altered His114 conformation was found to be significantly lower than the wild ANG and mutant with native His114 conformation.

8.8.4 Molecular dynamics to identify allosteric sites

Allosteric sites may not directly involve a protein's functional activity, but they regulate the structural integrity of the active site. Molecular dynamics can be used to study the role of allosteric sites and their indirect roles. Testosterone and other sex hormones, fatty acid and drugs are carried in the blood by human serum albumin (HSA). Still, the binding site of the HSA and the dynamics of testosterone binding to it is not yet completely known. Jayaraj et al. [110], through molecular dynamics, revealed that testosterone's binding to fatty acid binding site (Phe403, Leu430, Val 433 and Leu387) on HSA is associated with the conformational changes at another active site (Ser480 and Phe204) showing that two distant sites are allosterically coupled. A recent study by Jasuja and workers [111] also demonstrated that estradiol binding to the sex hormone binding globulin induces allosteric sites on it.

Similarly, Ivetac and McCammon [112] generated different conformers of human adrenergic receptors using molecular dynamics and flooded the protein surface with small organic probes using FTMAP [113]. Allosteric

sites were defined as the locations on the protein surface where organic probes clustered together across different protein conformations.

8.8.5 Molecular dynamics to identify cryptic sites

The binding pockets to accommodate signalling molecules are often revealed by crystallographic studies, but the structures obtained from these experimental techniques conceal other therapeutic sites, known as cryptic sites. Schames et al. [114] explored the dynamics of HIV integrase, which was not amenable for SBDD and identified a new trench that was not evident from the crystallographic studies. Later, NMR and X-ray crystallographic studies confirmed that the existing drugs indeed bound to this trench. Subsequent research by Merck & co. [115] on the newly found cryptic site led to the emergence of raltegravir, a very effective new antiretroviral drug.

8.8.6 Molecular dynamics to design libraries of lead-like molecules from chemical templates

The development of various libraries having drug-like molecules has revolutionised drug discovery. An early example of this [116] is the development of new chemical candidate molecules, including Non-Steroidal Anti-inflammatory Drugs (NSAIDs) from the chemical templates against the cyclooxygenase-2 (COX-2). The authors used empirical filters with docking and simulation protocol combinations to extract the drug-like properties of each candidate molecule. Finally, the studies proved successful in separating drug-like compounds from non-drug-like ones and establishing the technique as a viable means of generating lead molecules for biomolecular targets. Jayaraj et al. [110] have recently developed a web server to identify lead molecules to multiple protein targets. Likewise, Shaikh et al. [117, 118] developed couple of new strategies for sketching lead-like molecules for proteins as well as for DNA. Similar to proteins, DNA can also be targeted, Das and Jayaram [119] studied a Brownian dynamics simulation to investigate the contribution of hydrodynamic and intermolecular forces in guiding the drug DAPI (4,6 diamidino-2-phenylindole) to its DNA binding site. The rate constant (68.9×10^6) obtained from the study was found to be nearer to the experimental value (63.3×10^6). Shaikh et al. [120] similarly studied 25 DNA-drug complexes to devise the structural and energetic parameters essential for a molecule to be drug-like through molecular dynamics and MMGBSA methods. In the following section, we will briefly discuss some more case studies in detail on how molecular dynamics and free energy simulations, combined with virtual screening and molecular docking, have successfully contributed to the development of novel drugs/lead molecules.

8.9 Some success stories of MD-based Computer-Aided Drug Discovery

As stated previously, the availability of large databases of targets for various diseases [121, 122] and purchasable small molecules has accelerated the drug discovery process. There are several examples where molecules from various databases, such as Cambridge DIVERSet, Zinc Database, Indian Medicinal Plants etc., have been shown to have bioactivity. Kiruthuka et al. [123] performed screening, docking, molecular dynamics simulations and free energy analysis using MMBAPPL, MMPBSA and MMGBSA for hepatitis B surface antigen (HBsAg) protein target using a million molecules from the ZINC database. The subsequent study of cytotoxicity and biological activity on the cell line identified a small molecule (ZINC20451377) having a high binding affinity to HBsAg with a KD value of 65.3nM. In fact, the molecule was also found to be effective against the HBV quadruple mutant, which is resistant to the existing drug tenofovir.

8.9.1 Case study 01: Assessing molecules from some Indian Medicinal Plants against cancer

Cancer is one the deadliest diseases and has the highest mortality among all diseases. We have utilized compounds from Indian Medicinal Plants: Neem, Tulsi and Turmeric, to find drug-like molecules against the disease. The workflow given below [Figure (8.7)] spells out the methodologies adopted.

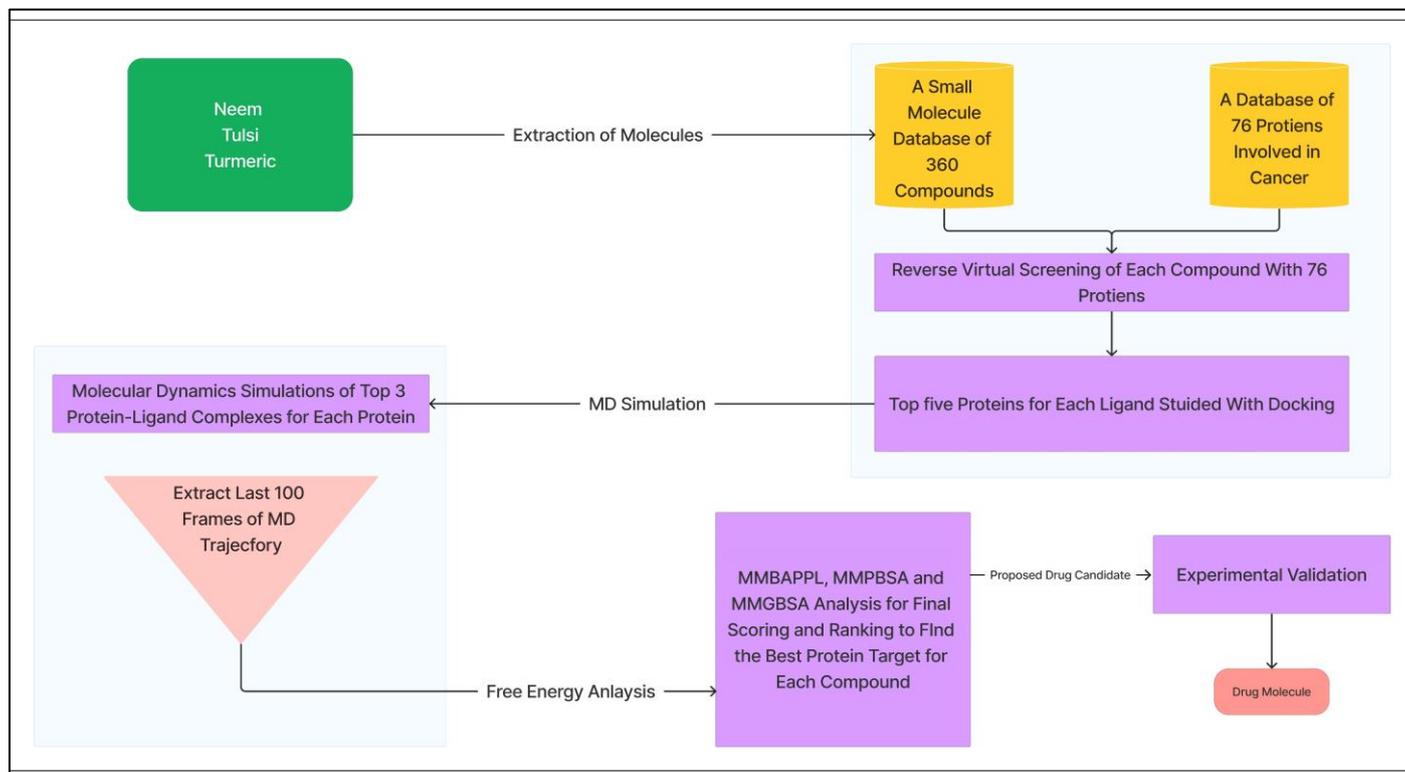


Figure 8.7: Workflow adopted to identify drug-like molecules using Indian medicinal plants against cancer.

We prepared a database of 360 molecules using Indian Medicinal Plants: Neem, Tulsi and Turmeric and performed reverse screening and docking using *in house* ParDock software [124] against a database of 76 proteins implicated in cancer. The top three protein-ligand complexes corresponding to each protein were studied with MD simulations and free energy analyses. The study provided potential lead molecules for some of the proteins, for example, Procurecumenol for estrogen receptor with a binding free energy of -32.1 kcal/mol (MMGBSA) and -10.1kcal/mol with MMPBSA method. The molecule is under experimental validation. This study is expected to revalidate the idea that understanding the magical cures that Indian Medicinal Plants offer at a molecular level will help in disseminating the knowledge and cures to the whole world.

8.9.2 Case Study 02: Anti-Cancer Drugs

Estrogen receptors (ER) play a crucial role in breast cancer [125]. Bhatnagar et al. [126] performed a computational study of biphenyl compounds [Figure (8.8)]. They performed the docking of four biphenyl compounds using *in-house* ParDock software.



Figure 8.8: Biphenyl compounds considered for the study.

The docking binding energy (kcal/mol) of compounds A, B, C and D was found to be -7.5, -8.0, -6.1 and -6.01, and -7.43, -7.67, -6.48 and -5.46 for ER α and ER β , respectively. All four compounds were further studied with molecular dynamics and free energy calculations (MM-PBSA) [127] using AMBER [128] to investigate the stability of all the complexes. The free energy of binding (Kcal/mol, for ER α) was found to be -23.63, -16.87, -22.52 and -5.09 for compounds A, B, C and D, respectively.

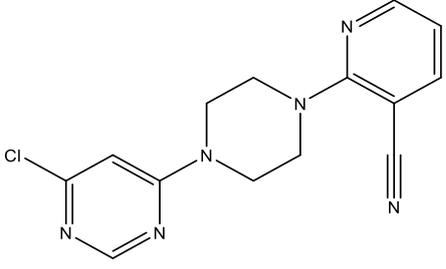
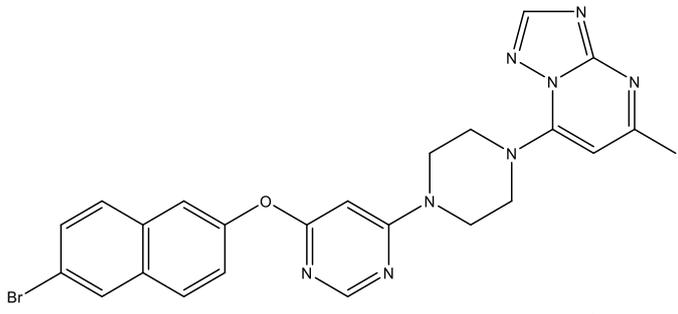
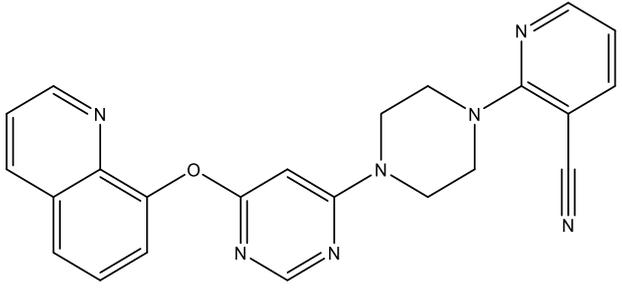
It is clear from the free energy analyses that compound (A) shows preferential binding to estrogen receptors. The compounds (A-D) were further evaluated against ER α (MCF-7) and ER β (MDA-MB-231) cell lines to study their biological activity. IC₅₀ values of A, B, C, D against ER α were found to be 62.5, 0.97, 31.25 and 1.96 μ g/ml, respectively and >125, 62.5, >125 and 62.5 μ g/ml for ER β , respectively. IC₅₀ values demonstrate that all the compounds (A-D) have preferential inhibitory activity against ER α over ER β . It is also evident from experiments that compound (A) is the most potent molecule among all four, showing a correlation with the results obtained from computational studies.

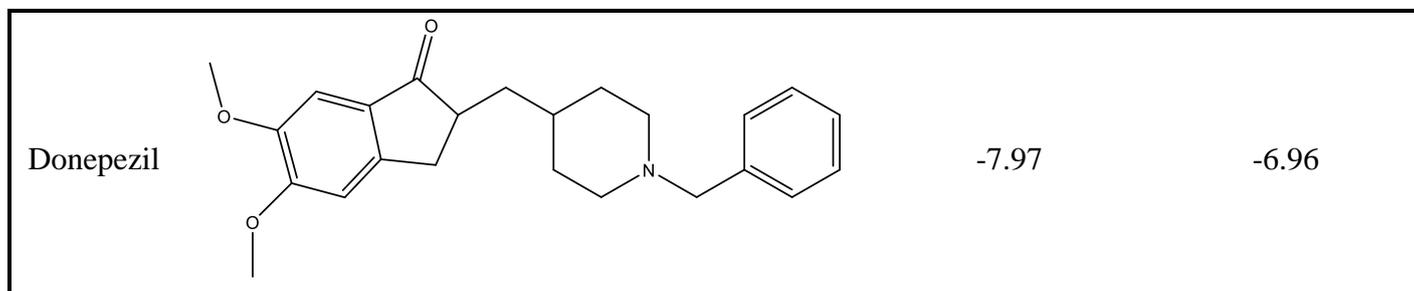
To further study the anti-proliferation activity of compound A, a cotreatment study was performed with known drugs tamoxifen and estradiol. Compound A showed enhanced cytotoxicity when administered with tamoxifen showing anti-proliferation activity. The reverse was observed when compound A was administered with estradiol. However, co-treatment of compound A with ER β (MDA-MB-231) cell line did not show cytotoxicity. Thus, the cotreatment findings show that compound A and Tamoxifen had a substantial synergistic effect on ER α + (MCF-7) cell lines but not on ER β + cells (MDA-MB-231). Therefore, compound A is selective for ER α , and its effects on cell proliferation and cell death directly result from its ability to target ER α . Additionally, authors conducted siRNA silencing studies to verify the ER α selectivity of compound A. After downregulating the ER α expression, silenced cells and parent MCF-7 cells were exposed to compound A and Tamoxifen, and it was observed that these compounds combinedly showed reduced cell death and demonstrated selectivity for ER α (correlating computational predictions).

8.9.3 Case-Study 03: Anti-Alzheimer's Drugs

Alzheimer's disease is often thought to be caused by β -amyloid build up, biometal dyshomeostasis, and dysfunctional cholinergic systems [129]. However, most drugs are designed against the dysfunctional cholinergic system to reduce acetylcholine (ACh) levels in the pre-synaptic region. Acetylcholinesterase (AChE) is the enzyme that catalyses the hydrolysis of ACh into acetyl and choline to stop signal transmission [130, 131]. The crystal structure of recombinant human AChE (rhAChE) and Torpedo californica AChE (TcAChE) demonstrated two ligand binding sites, an active catalytic site (CAS) at the core and another peripheral anionic site (PAS) at the mouth [132]. The catalytic triad in CAS, consisting of serine, histidine, and glutamate residues, plays a crucial role in ACh hydrolysis, whereas amino acids present in PAS help cationic ammonium substrates to move inward into the catalytic triad. It is also found that the PAS site in AChE serves as the nucleation centre for amyloid β accumulation and provides a promising target for the development of drugs that attempt to prevent the production of amyloid fibrils. Therefore, the scientific community has considered a bivalent approach targeting both PAS and CAS [133-137]. Kumar et al. [138] designed a couple of molecules [Table (8.1)] with triazolopyrimidine and quinoline moieties linked by piperazine and performed molecular docking and molecule dynamics studies. Molecules with favourable binding energy upon docking are reported along with the existing drug Donepezil in Table (8.1) below:

Table 8.1: Best binding molecules against both rhAChE and TcAChE.

Molecule	Structure	Binding Energy (kcal/mol) Against rhAChE	Binding Energy (kcal/mol) Against TcAChE
10e		-8.6	-8.4
11c		-8.8	-8.4
12b		-9.7	-9.3



Molecules 10e, 11c and 12b were further studied with molecular dynamics simulations and free energy calculation using the MMGBSA method. The free binding energy (kcal/mol) of compounds 10e, 11c and 12b were found to be -9.5, -9.8 and -11.9, and -8.8, -9.2 and -8.9 for rhAChE and TcAChE, respectively. The IC_{50} (μM) values when performed AChE inhibitory activity using Ellman's method were found to be 0.67, 0.161 and 0.036 for 10e, 11c and 12b, respectively.

The data clearly demonstrates that 12b is a potent molecule (binds 3-fold more tightly to the AChE than the existing drug Tacrine (IC_{50} value 0.13 μM) and similar to as Donepezil (IC_{50} value 0.038 μM) and can be used to block AChE signalling. Authors also studied the MTT assay and found that these molecules neither caused cell death nor resulted in a cell cycle blockage; consequently, they are not carcinogenic and may be safe for treating Alzheimer's disease.

8.9.4 CASE-STUDY 04: Antiviral Drugs

HIV-1 protease cleaves the translated viral gag-pol polyproteins into small discrete components responsible for the viral activity and infectivity. Therefore, HIV-1 protease is an attractive target to prevent the disease associated with the viral proteins. In one of the very early CADD/MD studies, Kalra et al. [55] performed a molecular dynamics study of HIV-1 protease complexed with two peptides, 4HVP and 8HVP. They employed various protocols while performing molecular dynamics to obtain a range of trajectories, estimated binding free energies and performed energy component analyses. The free energy value (kcal/mol) of various components G_{ele}^0 , ΔG_{vwd}^0 , ΔG_{cav}^0 , $\Delta G_{entropy}^0$, ΔG_{ions}^0 and ΔG_{adp}^0 were found to be 38.95, -16.73, -74.42, 37.47, -0.23, and 5.23, and 24.37, -13.85, -76.44, 39.32, 0.73 and 8.14 for 4HVP and 8HVP, respectively. ΔG_{tot}^0 obtained for 4HVP from simulation studies (-9.73 kcal/mol) showed almost similar results as that of experimental (-8.40 kcal/mol) data, whereas a slight deviation was observed for 8HVP between predicted (-17.73 kcal/mol) and experimental data (-12.30 kcal/mol).

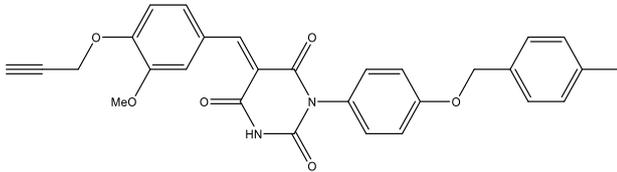
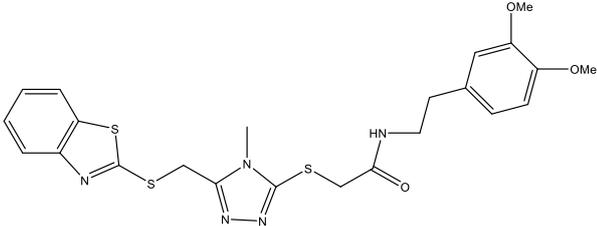
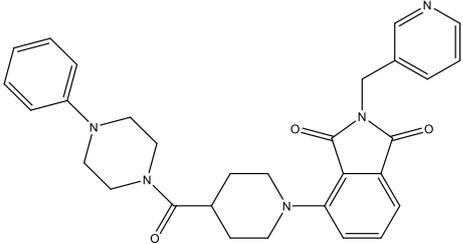
Finding the inhibitor's residues or functional groups that can be changed for good binding is important from the standpoint of drug design. The trajectories obtained from the MD simulations can be utilized to find the residues that interact well or poorly with an enzyme. The authors carried out residue-wise direct electrostatic and van der Waal's interactions of each amino acid of the inhibitor and compared that to the entire protein to derive possible modifications that can enhance free energy favourably. They replaced the Arginine side chain with ethyl in 4HVP and found the mutation favourable. However, when they mutated isoleucine with valine to see changes in the packing, free energy did not change much.

The authors studied both systems with various protocols using both implicit (utilizing different dielectric functions; Sigmoidal and 4r dielectric function) and explicit solvent models with varying minimization steps to find a robust method for accurate free energy estimation. They concluded that explicit water with periodic boundary conditions provides more accurate results. Also, the free energy obtained after minimization in continuum solvent models (-9.73 kcal/mol) showed a correlation with the experimental values (-8.40 kcal/mol). However, this does not include the cost of adaptation, and the resemblance to the experimental values could be case specific.

8.9.5 Case-Study 05: Antiviral Drugs

Developing novel antiviral agents is challenging, as an in-depth understanding of the viral life cycle is required. 3C protease plays a crucial role in the post-translation proteolysis of viral polyprotein in Hepatitis A virus and other picornaviruses and helps them to attack innate host immune system. The viral polymerase, required for replication initiation, is one of the structural proteins released when 3C protease auto-cleaves the translated polyprotein. [139]. Therefore, 3C protease is a good target for developing antiviral drugs. Banerjee et al. [140] carried out an *in silico* study to identify potent inhibitors of the 3C protease. They identified a binding pocket with a catalytic triad consisting of Cys172, His44, and Asp84 using the in-house active site prediction software, AADS [141]. Using a rapid screening protocol with RASPD [142] and RASPD+ [143] to determine the top 1000 hits for this binding site of HAV 3C protease, nearly one million chemical compounds from the ZINC database were screened against it. Then, using ParDock, atomic-level docking and scoring was applied to these molecules. Short molecular dynamics simulations of 250 compounds with the highest predicted binding energies were performed, and eventually, seven compounds that showed strong interactions were considered (Table (8.2) below) for further 100ns simulation and free energy analyses using MMPBSA/GBSA methods. Additionally, they computationally evaluated a range of isatin derivatives against the HAV 3C protease and discovered that two changes to the isatins' protective groups improved their interactions with the target.

Table 8.2: Compounds (1-9) identified against the Hepatitis A virus

Compound	Structure	Calculated Binding Energy (kcal/mol)	Experimental Inhibition Constant (K_i , μM)
1.		-16.25	3.0 ± 0.1
2.		-11.05	-8.6 ± 0.7
3.		-10.58	-2.5 ± 0.1

The dynamic investigation demonstrated that compound 6, on average, formed four hydrogen bonds with the active site residues of HAV 3C protease during the simulation. Table (8.3) below shows the MMPBSA, MMGBSA and MMBAPPL analysis obtained from the 100ns trajectories.

Table 8.3: MMPBSA, MMGBSA and MMBAPPL analysis of the compounds studied

Compounds	MMPBSA	MMGBSA	MMBAPPL	Experimental K_i values (μM)
1.	-38.60 \pm 0.70	52.20 \pm 0.53	-8.96 \pm 0.14	3
2.	-13.25 \pm 0.16	-41.80 \pm 0.47	-6.97 \pm 0.20	8.6
3.	-15.78 \pm 0.23	-42.56 \pm 0.42	-8.56 \pm 0.36	2.5
4.	-29.96 \pm 0.58	-50.22 \pm 0.12	-8.61 \pm 0.11	1.4
5.	-27.47 \pm 0.73	-46.78 \pm 0.54	-6.30 \pm 0.33	117.8
6.	-21.66 \pm 0.19	-57.20 \pm 0.35	-9.01 \pm 0.18	3.3
7.	-20.23 \pm 0.43	-55.18 \pm 0.33	8.92 \pm 0.13	1.2
8.	-21.32 \pm 0.11	-53.87 \pm 0.57	-8.19 \pm 0.14	2.1
9.	-18.95 \pm 0.14	-52.58 \pm 0.25	-8.56 \pm 0.14	1.6

When compared to the MMGBSA and MMPBSA scores, authors discovered that predictions of artificial intelligence and machine learning augmented algorithm MMBAPPL+ correlated well with experimental K_i values. The latest version of MMBAPPL is MMBAPPL+ [145], and it has shown improved performance. They conducted similar studies with other virus's 3C proteases from the Picornavirus family and demonstrated that designing and creating generic drugs that were effective against the 3C proteases was possible.

8.10 Conclusion

In the article, we discussed the many roles that molecular dynamics simulations play in drug discovery, including deciphering the mode of action of drug molecules, identification of allosteric sites and how mutations result in the loss of function of proteins. The study by Bhat et al. on FDA-approved drugs is one the biggest pieces of evidence that molecular dynamics based CADD has a promising future. Thus, with an increasing number of small molecule libraries and computing power, computer aided drug discovery provides an opportunity to come up with potential drug candidates within hours for a target. MMGBSA/PBSA methods for free energy estimation are quite quick and can be used for qualitative/semiquantitative ranking/scoring of molecules by performing only a single simulation. However, the statistical mechanical noise associated with MMPBSA/GBSA methods can sometimes lead to false rankings. A user can choose to go with the more accurate free energy perturbation methods as per the availability of resources and time. New trends such as artificial intelligence and machine learning (AI/ML) augmented algorithms like MMBAPPL+ are seen to perform well, broadening the scope for further improvements in ranking and scoring schemes. Lastly, it is hoped that simulations and AI techniques will address toxicity issues in the near future to ensure the prediction of safe lead molecules that work like magic bullets to cure diseases and disorders. Good days for MD/CADD

are here to stay.

Acknowledgements

We gratefully acknowledge funding to SCFBIO, IIT Delhi, from the Department of Biotechnology, Govt. of India and from Meity, GoI, Centre for Development of Advanced Computing (CDAC), India under the NSM project.

References

1. Fischer E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.*, 27: 2985-2993.
2. Teague, S. J. (2003). Implications of protein flexibility for drug discovery. *Nature reviews Drug discovery*, 2(7), 527-541.
3. Ma, B., Kumar, S., Tsai, C. J., & Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein engineering*, 12(9), 713-720.
4. Kumar, S., Ma, B., Tsai, C. J., Wolfson, H., & Nussinov, R. (1999). Folding funnels and conformational transitions via hinge-bending motions. *Cell biochemistry and biophysics*, 31(2), 141-164.
5. Tsai, C. J., Kumar, S., Ma, B., & Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Science*, 8(6), 1181-1190.
6. Ma, B., Shatsky, M., Wolfson, H. J., & Nussinov, R. (2002). Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein science*, 11(2), 184-197.
7. Bouzat, C., Gumilar, F., Spitzmaul, G., Wang, H. L., Rayes, D., Hansen, S. B., ... & Sine, S. M. (2004). Coupling of agonist binding to channel gating in an ACh-binding protein linked to an ion channel. *Nature*, 430(7002), 896-900.
8. Talley, T. T., Yalda, S., Ho, K. Y., Tor, Y., Soti, F. S., Kem, W. R., & Taylor, P. (2006). Spectroscopic analysis of benzylidene anabaseine complexes with acetylcholine binding proteins as models for ligand– nicotinic receptor interactions. *Biochemistry*, 45(29), 8894-8902.
9. Babakhani, A., Talley, T. T., Taylor, P., & McCammon, J. A. (2009). A virtual screening study of the acetylcholine binding protein using a relaxed–complex approach. *Computational biology and chemistry*, 33(2), 160-170.
10. Bourne, Y., Talley, T. T., Hansen, S. B., Taylor, P., & Marchot, P. (2005). Crystal structure of a Cbtx– AChBP complex reveals essential interactions between snake α -neurotoxins and nicotinic receptors. *The EMBO journal*, 24(8), 1512-1522.
11. McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded proteins. *nature*, 267.
12. Bottegoni, G., Kufareva, I., Totrov, M., & Abagyan, R. (2009). Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. *Journal of medicinal chemistry*, 52(2), 397-406.
13. Bottegoni, G. (2011). Protein-ligand docking. *Frontiers in Bioscience-Landmark*, 16(6), 2289-2306.
14. Rauh, D., Klebe, G., & Stubbs, M. T. (2004). Understanding protein–ligand interactions: the price of protein flexibility. *Journal of molecular biology*, 335(5), 1325-1341.
15. Damm, K. L., & Carlson, H. A. (2007). Exploring experimental sources of multiple protein conformations in structure-based drug design. *Journal of the American Chemical Society*, 129(26), 8225-8235.
16. Alonso, H., Bliznyuk, A. A., & Gready, J. E. (2006). Combining docking and molecular dynamic simulations in drug design. *Medicinal research reviews*, 26(5), 531-568.

17. Morra, G., Genoni, A., Neves, M. A., Merz Jr, K. M., & Colombo, G. (2010). Molecular recognition and drug-lead identification: what can molecular simulations tell us?. *Current medicinal chemistry*, 17(1), 25-41.
18. Pang, Y. P., & Kozikowski, A. P. (1994). Prediction of the binding sites of huperzine A in acetylcholinesterase by docking studies. *Journal of Computer-Aided Molecular Design*, 8(6), 669-681.
19. Ivetic, A., & Andrew McCammon, J. (2011). Molecular recognition in the case of flexible targets. *Current pharmaceutical design*, 17(17), 1663-1671.
20. Lin, J. H., Perryman, A. L., Schames, J. R., & McCammon, J. A. (2003). The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers: Original Research on Biomolecules*, 68(1), 47-62.
21. Dror, R. O., Jensen, M. Ø., Borhani, D. W., & Shaw, D. E. (2010). Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *Journal of General Physiology*, 135(6), 555-562.
22. Lane, T. J., Shukla, D., Beauchamp, K. A., & Pande, V. S. (2013). To milliseconds and beyond: challenges in the simulation of protein folding. *Current opinion in structural biology*, 23(1), 58-65.
23. Lu, H., & Tonge, P. J. (2010). Drug-target residence time: critical information for lead optimization. *Current opinion in chemical biology*, 14(4), 467-474.
24. Borhani, D. W., & Shaw, D. E. (2012). The future of molecular dynamics simulations in drug discovery. *Journal of computer-aided molecular design*, 26(1), 15-26.
25. Torrie, G. M., & Valleau, J. P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2), 187-199.
26. Bolhuis, P. G., Chandler, D., Dellago, C., & Geissler, P. L. (2002). Transition Path Sampling: Throwing Ropes. *Annu. Rev. Phys. Chem*, 53, 291-318.
27. Hamelberg, D., Mongan, J., & McCammon, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *The Journal of chemical physics*, 120(24), 11919-11929.
28. Jorgensen, W. L., & Ravimohan, C. (1985). Monte Carlo simulation of differences in free energies of hydration. *The Journal of chemical physics*, 83(6), 3050-3054.
29. Jorgensen, W. L., & Thomas, L. L. (2008). Perspective on free-energy perturbation calculations for chemical equilibria. *Journal of chemical theory and computation*, 4(6), 869-876.
30. Sugita, Y., & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1-2), 141-151.
31. Isralewitz, B., Gao, M., & Schulten, K. (2001). Steered molecular dynamics and mechanical functions of proteins. *Current opinion in structural biology*, 11(2), 224-230.
32. Grubmüller, H., Heymann, B., & Tavan, P. (1996). Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271(5251), 997-999.
33. Faradjian, A. K., & Elber, R. (2004). Computing time scales from reaction coordinates by milestoning. *The Journal of chemical physics*, 120(23), 10880-10889.
34. Laio, A., & Parrinello, M. (2002). Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20), 12562-12566.
35. Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC biology*, 9(1), 1-9.
36. Harvey, M. J., & De Fabritiis, G. (2012). High-throughput molecular dynamics: the powerful new tool for drug discovery. *Drug discovery today*, 17(19-20), 1059-1062.
37. De Vivo, M. (2011). Bridging quantum mechanics and structure-based drug design. *Frontiers in Bioscience-Landmark*, 16(5), 1619-1633.
38. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., ... & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19), 5179-5197.
39. MacKerell Jr, A. D. (1998). Bashford D, Bellott M, Dunbrack RL Jr. *Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S,*

- Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus MJ *Phys. Chem. B*, 102, 3586.
40. van Gunsteren, W. F., & Berendsen, H. J. (1987). Groningen molecular simulation (GROMOS) library manual. *Biomos, Groningen*, 24(682704), 13.
 41. van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P. K. H. C., Mark, A. E., ... & Tironi, I. G. (1996). Biomolecular simulation: the GROMOS96 manual and user guide. *Vdf Hochschulverlag AG an der ETH Zürich, Zürich*, 86, 1-1044.
 42. Jorgensen, W. L., & Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6), 1657-1666.
 43. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9), 1157-1174.
 44. Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., ... & Mackerell Jr, A. D. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry*, 31(4), 671-690.
 45. Frenkel, D., & Smit, B. (2001). *Understanding molecular simulation: from algorithms to applications* (Vol. 1). Elsevier.
 46. Allen, M. P., & Tildesley, D. J. (2017). *Computer simulation of liquids*. Oxford university press.
 47. Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3), 327-341.
 48. Hünenberger, P. H., Holm, C., & Kremer, K. (2005). Advanced computer simulation. *Adv. Polym. Sci*, 173, 105-149.
 49. Darden, T., York, D., & Pedersen, L. (1993). Jayaram, B., & Beveridge, D. L. (1990). A simple method to estimate free energy from a molecular simulation: renormalization on the unit interval. *Journal of physical chemistry*, 94(18), 7288-7293. *The Journal of chemical physics*, 98(12), 10089-10092.
 50. Sagui, C., & Darden, T. A. (1999). Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure*, 28(1), 155-179.
 51. Jayaram, B., & Beveridge, D. L. (1990). A simple method to estimate free energy from a molecular simulation: renormalization on the unit interval. *Journal of physical chemistry*, 94(18), 7288-7293.
 52. Jayaram, B., & Beveridge, D. L. (1990). Free energy of an arbitrary charge distribution imbedded in coaxial cylindrical dielectric continua: application to conformational preferences of DNA in aqueous solutions. *Journal of Physical Chemistry*, 94(11), 4666-4671.
 53. Jayaram, B., DiCapua, F. M., & Beveridge, D. L. (1991). A theoretical study of polyelectrolyte effects in protein-DNA interactions: Monte Carlo free energy simulations on the ion atmosphere contribution to the thermodynamics of lambda. repressor-operator complex formation. *Journal of the American Chemical Society*, 113(14), 5211-5215.
 54. Gilson, M. K., Given, J. A., Bush, B. L., & McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical journal*, 72(3), 1047-1069.
 55. Kalra, P., Reddy, T. V., & Jayaram, B. (2001). Free energy component analysis for drug design: A case study of HIV-1 protease-inhibitor binding. *Journal of medicinal chemistry*, 44(25), 4325-4338.
 56. Latha, N., & Jayaram, B. (2005). A binding affinity based computational pathway for active-site directed lead molecule design: some promises and perspectives. *Drug Design Reviews-Online (Discontinued)*, 2(2), 145-165.
 57. Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics*, 22(8), 1420-1426.
 58. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., ... & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106(3), 765-784.

59. Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *Journal of computational chemistry*, 7(2), 230-252.
60. Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC biology*, 9(1), 1-9.
61. Woo, H. J., & Roux, B. (2005). Calculation of absolute protein–ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences*, 102(19), 6825-6830.
62. Ge, X., & Roux, B. (2010). Absolute binding free energy calculations of sparsomycin analogs to the bacterial ribosome. *The Journal of Physical Chemistry B*, 114(29), 9525-9539.
63. Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14), 2690.
64. Colizzi, F., Perozzo, R., Scapozza, L., Recanatini, M., & Cavalli, A. (2010). Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *Journal of the American Chemical Society*, 132(21), 7361-7371.
65. Xiong, H., Crespo, A., Marti, M., Estrin, D., & Roitberg, A. E. (2006). Free energy calculations with non-equilibrium methods: applications of the Jarzynski relationship. *Theoretical Chemistry Accounts*, 116(1), 338-346.
66. Smith, P. E., & Pettitt, B. M. (1994). Modeling solvent in biomolecular systems. *The Journal of physical chemistry*, 98(39), 9700-9711.
67. Cramer, C. J., & Truhlar, D. G. (1999). Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chemical Reviews*, 99(8), 2161-2200.
68. Jayaram, B., Sprous, D., & Beveridge, D. L. (1998). Solvation free energy of biomacromolecules: Parameters for a modified generalized Born model consistent with the AMBER force field. *The Journal of Physical Chemistry B*, 102(47), 9571-9576.
69. Onsager, L. (1936). Electric moments of molecules in liquids. *Journal of the American Chemical Society*, 58(8), 1486-1493.
70. Jayaram, B., & Beveridge, D. L. (1988). Tanford–Kirkwood theory for concentric dielectric continua: Application to dimethylphosphate. *Biopolymers: Original Research on Biomolecules*, 27(4), 617-627.
71. Jayaram, B. (1994). Free Energy of Solvation, Interaction, and Binding of Arbitrary Charge Distributions Imbedded in a Dielectric Continuum. *The Journal of Physical Chemistry*, 98(22), 5773-5777.
72. Kumari, R., Kumar, R., Open Source Drug Discovery Consortium, & Lynn, A. (2014). g_mmpbsa□ A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of chemical information and modeling*, 54(7), 1951-1962.
73. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., ... & Cheatham, T. E. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research*, 33(12), 889-897.
74. Gilson, M. K., & Honig, B. (1988). Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Structure, Function, and Bioinformatics*, 4(1), 7-18.
75. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate– DNA helices. *Journal of the American Chemical Society*, 120(37), 9401-9409.
76. Jayaram, B., McConnell, K. J., Dixit, S. B., & Beveridge, D. L. (1999). Free energy analysis of protein–DNA binding: the EcoRI endonuclease–DNA complex. *Journal of Computational Physics*, 151(1), 333-357.
77. Gohlke, H., Kiel, C., & Case, D. A. (2003). Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes. *Journal of molecular biology*, 330(4), 891-913.
78. Wang, W., & Kollman, P. A. (2000). Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *Journal of molecular biology*, 303(4), 567-582.

79. Lee, M. R., Duan, Y., & Kollman, P. A. (2000). Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins: Structure, Function, and Bioinformatics*, 39(4), 309-316.
80. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3), 712-725.
81. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8), 1950-1958.
82. Homeyer, N., & Gohlke, H. (2012). Free energy calculations by the molecular mechanics Poisson–Boltzmann surface area method. *Molecular informatics*, 31(2), 114-122.
83. Still, W. C., Tempczyk, A., Hawley, R. C., & Hendrickson, T. (1990). Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society*, 112(16), 6127-6129.
84. Honig, B., & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science*, 268(5214), 1144-1149.
85. Warshel, A., & Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology*, 103(2), 227-249.
86. Levitt, M. (1974). Energy refinement of hen egg-white lysozyme. *Journal of molecular biology*, 82(3), 393-420.
87. Warshel, A., & Karplus, M. (1972). Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. *Journal of the American Chemical Society*, 94(16), 5612-5625.
88. Warshel, A. (1973). Quantum mechanical consistent force field (QCFF/PI) method: Calculations of energies, conformations and vibronic interactions of ground and excited states of conjugated molecules. *Israel journal of chemistry*, 11(5), 709-717.
89. Moore, T. J., Zhang, H., Anderson, G., & Alexander, G. C. (2018). Estimated costs of pivotal trials for novel therapeutic agents approved by the US Food and Drug Administration, 2015-2016. *JAMA internal medicine*, 178(11), 1451-1457.
90. Bissaro, M., Sturlese, M., & Moro, S. (2020). The rise of molecular simulations in fragment-based drug design (FBDD): an overview. *Drug Discovery Today*, 25(9), 1693-1701.
91. Pathak, A., Singh, B., Chaurasia, D. K., & Jayaram, B. (2021). Molecular Simulation–Driven Drug Repurposing for the Identification of Inhibitors Against Non-Structural Proteins of SARS-CoV-2. In *In Silico Modeling of Drugs Against Coronaviruses* (pp. 683-713). Humana, New York, NY.
92. Bhat, R., Kaushik, R., Singh, A., DasGupta, D., Jayaraj, A., Soni, A., ... & Jayaram, B. (2020). A comprehensive automated computer-aided discovery pipeline from genomes to hit molecules. *Chemical Engineering Science*, 222, 115711.
93. Jayaram, B., Latha, N., Jain, T., Sharma, P., Gandhimathi, A., & Pandey, V. S. (2006). “Sanjeevini: A Comprehensive Active-Site Directed Lead Design Software”. *Indian Journal of Chemistry-A*. 45A, 1834-1837.
94. Jayaram, B., Singh, T., Mukherjee, G., Mathur, A., Shekhar, S., & Shekhar, V. (2012, December). Sanjeevini: a freely accessible web-server for target directed lead molecule discovery. In *BMC Bioinformatics*, 13(Suppl 17):S7 DOI:10.1186/1471-2105-13-S17-S7.
95. Singh, A., Shekhar, S., & Jayaram, B. (2020). CADD: Some Success Stories from Sanjeevini and the Way Forward. In *Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design* (pp. 1-18). Springer Nature, Singapore Pte Ltd. S. K. Singh (ed.), https://doi.org/10.1007/978-981-15-8936-2_1.
96. Tripathi, P. K., Soni, A., Yadav, S. P. S., Kumar, A., Gaurav, N., Raghavendhar, S., ... & Patel, A. K. (2020). Evaluation of novobiocin and telmisartan for anti-CHIKV activity. *Virology*, 548, 250-260.

97. Chatterjee, B. K., Jayaraj, A., Kumar, V., Blagg, B., Davis, R. E., Jayaram, B., ... & Chaudhuri, T. K. (2019). Stimulation of heat shock protein 90 chaperone function through binding of a novobiocin analog KU-32. *Journal of Biological Chemistry*, 294(16), 6450-6467.
98. Klayman, D. L. (1985). Qinghaosu (artemisinin): an antimalarial drug from China. *Science*, 228(4703), 1049-1055.
99. Meshnick, S. R., Taylor, T. E., & Kamchonwongpaisan, S. (1996). Artemisinin and the antimalarial endoperoxides: from herbal remedy to targeted chemotherapy. *Microbiological reviews*, 60(2), 301-315.
100. Posner, G. H., Wang, D., Cumming, J. N., Oh, C. H., French, A. N., Bodley, A. L., & Shapiro, T. A. (1995). Further evidence supporting the importance of and the restrictions on a carbon-centered radical for high antimalarial activity of 1, 2, 4-trioxanes like artemisinin. *Journal of medicinal chemistry*, 38(13), 2273-2275.
101. Meshnick, S. R., Thomas, A., Ranz, A., Xu, C. M., & Pan, H. Z. (1991). Artemisinin (qinghaosu): the role of intracellular hemin in its mechanism of antimalarial action. *Molecular and biochemical parasitology*, 49(2), 181-189.
102. Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., ... & Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419(6906), 498-511.
103. Jung, M., Kim, H., Nam, K. Y., & No, K. T. (2005). Three-dimensional structure of *Plasmodium falciparum* Ca²⁺-ATPase (PfATP6) and docking of artemisinin derivatives to PfATP6. *Bioorganic & medicinal chemistry letters*, 15(12), 2994-2997.
104. Naik, P. K., Srivastava, M., Bajaj, P., Jain, S., Dubey, A., Ranjan, P., ... & Singh, H. (2011). The binding modes and binding affinities of artemisinin derivatives with *Plasmodium falciparum* Ca²⁺-ATPase (PfATP6). *Journal of molecular modeling*, 17(2), 333-357.
105. Garah, F. B. E., Stigliani, J. L., Coslédan, F., Meunier, B., & Robert, A. (2009). Docking studies of structurally diverse antimalarial drugs targeting PfATP6: no correlation between in silico binding affinity and in vitro antimalarial activity. *ChemMedChem: Chemistry Enabling Drug Discovery*, 4(9), 1469-1479.
106. Shandilya, A., Chacko, S., Jayaram, B., & Ghosh, I. (2013). A plausible mechanism for the antimalarial activity of artemisinin: a computational approach. *Scientific reports*, 3(1), 1-7.
107. Ghosh, I., & Rao, V. S. R. (1982). A conformational approach to the study of the dynamics of enzyme inhibition: studies on thermolysin. *International Journal of Biological Macromolecules*, 4(3), 130-136.
108. Bharatam, P. V., Patel, D. S., & Iqbal, P. (2005). Pharmacophoric features of biguanide derivatives: an electronic and structural analysis. *Journal of medicinal chemistry*, 48(24), 7615-7622.
109. Padhi, A. K., Vasaikar, S. V., Jayaram, B., & Gomes, J. (2013). Fast prediction of deleterious angiogenin mutations causing amyotrophic lateral sclerosis. *FEBS letters*, 587(12), 1762-1766.
110. Jayaraj, A., Bhat, R., Pathak, A., Singh, M., & Jayaram, B. (2018). Development of a Web-Server for Identification of Common Lead Molecules for Multiple Protein Targets. In *Multi-Target Drug Design Using Chem-Bioinformatic Approaches* (pp. 487-504). Humana Press, New York, NY.
111. Jasuja, R., Spencer, D., Jayaraj, A., Peng, L., Krishna, M., Lawney, B., ... & Bhasin, S. (2021). Estradiol induces allosteric coupling and partitioning of sex-hormone-binding globulin monomers among conformational states. *Isience*, 24(6), 102414.
112. Ivetac, A., & Andrew McCammon, J. (2010). Mapping the druggable allosteric space of G-protein coupled receptors: a fragment-based molecular dynamics approach. *Chemical biology & drug design*, 76(3), 201-217.
113. Brenke, R., Kozakov, D., Chuang, G. Y., Beglov, D., Hall, D., Landon, M. R., ... & Vajda, S. (2009). Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics*, 25(5), 621-627.

114. Schames, J. R., Henchman, R. H., Siegel, J. S., Sotriffer, C. A., Ni, H., & McCammon, J. A. (2004). Discovery of a novel binding trench in HIV integrase. *Journal of medicinal chemistry*, 47(8), 1879-1881.
115. Hazuda, D. J., Anthony, N. J., Gomez, R. P., Jolly, S. M., Wai, J. S., Zhuang, L., ... & Young, S. D. (2004). A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proceedings of the National Academy of Sciences*, 101(31), 11233-11238.
116. Latha, N., Jain, T., Sharma, P., & Jayaram, B. (2004). A free energy based computational pathway from chemical templates to lead compounds: a case study of COX-2 inhibitors. *Journal of Biomolecular Structure and Dynamics*, 21(6), 791-804.
117. Shaikh, S. A., Jain, T., Sandhu, G., Latha, N., & Jayaram, B. (2007). From drug target to leads-sketching a physicochemical pathway for lead molecule design in silico. *Current pharmaceutical design*, 13(34), 3454-3470.
118. Shaikh, S. A., Ahmed, S. R., & Jayaram, B. (2004). A molecular thermodynamic view of DNA–drug interactions: a case study of 25 minor-groove binders. *Archives of biochemistry and biophysics*, 429(1), 81-99.
119. Das, A., & Jayaram, B. (1998). Brownian dynamics simulations of DNA-ligand interactions: a theoretical study on the kinetics of DAPI-DNA complexation. *Journal of molecular liquids*, 77(1-3), 157-163.
120. Shaikh, S. A., & Jayaram, B. (2007). A Swift All-Atom Energy-Based Computational Protocol to Predict DNA–Ligand Binding Affinity and ΔT_m . *Journal of medicinal chemistry*, 50(9), 2240-2244.
121. Singh, A., Kaushik, R., Chaurasia, D. K., Singh, M., & Jayaram, B. (2020). PvP01-DB: computational structural and functional characterization of soluble proteome of PvP01 strain of Plasmodium vivax. *Database*, 2020. Volume 2020. doi.org/10.1093/database/baaa036
122. Singh, A., Kaushik, R., Kuntal, H., & Jayaram, B. (2018). PvaxDB: a comprehensive structural repository of Plasmodium vivax proteome. *Database*, 2018, doi/10.1093/database/bay021/4938395.
123. Kiruthika, S., Bhat, R., Dash, R., Rathore, A. S., Vivekanandan, P., & Jayaram, B. (2021). A novel piperazine derivative that targets hepatitis B surface antigen effectively inhibits tenofovir resistant hepatitis B virus. *Scientific reports*, 11(1), 1-13.
124. Gupta, A., Gandhimathi, A., Sharma, P., & Jayaram, B. (2007). ParDOCK: an all atom energy based Monte Carlo docking protocol for protein-ligand complexes. *Protein and peptide letters*, 14(7), 632-646.
125. Musgrove, E. A., & Sutherland, R. L. (2009). Biological determinants of endocrine resistance in breast cancer. *Nature Reviews Cancer*, 9(9), 631-643.
126. Bhatnagar, S., Soni, A., Kaushik, S., Rikhi, M., Santhoshkumar, T. R., & Jayaram, B. (2018). Nonsteroidal estrogen receptor isoform-selective biphenyls. *Chemical biology & drug design*, 91(2), 620-630.
127. Miller III, B. R., McGee Jr, T. D., Swails, J. M., Homeyer, N., Gohlke, H., & Roitberg, A. E. (2012). MMPBSA.py: an efficient program for end-state free energy calculations. *Journal of chemical theory and computation*, 8(9), 3314-3321.
128. Case, D. A., Aktulga, H. M., Belfon, K., Ben-Shalom, I. Y., Berryman, J. T., Brozell, S. R., Cerutti, D. S., Cheatham, J. T., Cisneros, G. A., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Giambasu, G., Gilson, M. K., Gohlke, H., Goetz, A. W., Harris, R., Izadi, S., Izmailov, S. A., Kasavajhala, K., Kaymak, M.C., King, E., Kovalenko, A., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, L., Luchko, T., Luo, R., Machado, M., Man, V., Manathunga, M., Merz, K. M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., O'Hearn, K. A., Onufriev, A., Pan, F., Pantano, S., Qi, R., Rahnamoun, A., Roe, DR., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shajan, A., Shen, J., Simmerling, CL., Skrynnikov, NR., Smith, J., Swails, J., Walker, RC., Wang, J., Wei, H., Wolf RM., Wu, X., Xiong, Y., Xue, Y., York, DM., Zhao, S., and Kollman, P. A., (2022). Amber 2022. University of California, San Francisco.

129. Mao, F., Huang, L., Luo, Z., Liu, A., Lu, C., Xie, Z., & Li, X. (2012). O-Hydroxyl-or o-amino benzylamine-tacrine hybrids: Multifunctional biometals chelators, antioxidants, and inhibitors of cholinesterase activity and amyloid- β aggregation. *Bioorganic & medicinal chemistry*, 20(19), 5884-5892.
130. CHATONNET Aand LOCKRIDgE, O. (1989). Comparison of butyrylcholinesterase and acetylcholinesterase. *Biochem J*, 260, 625-634.
131. Scarpini, E., & Cogiamanian, F. (2003). Alzheimer's disease: from molecular pathogenesis to innovative therapies. *Expert Review of Neurotherapeutics*, 3(5), 619-630.
132. Harel, M., Sonoda, L. K., Silman, I., Sussman, J. L., & Rosenberry, T. L. (2008). Crystal structure of thioflavin T bound to the peripheral site of Torpedo californica acetylcholinesterase reveals how thioflavin T acts as a sensitive fluorescent reporter of ligand binding to the acylation site. *Journal of the American Chemical Society*, 130(25), 7856-7861.
133. Barril, X., Kalko, S. G., Orozco, M., & Luque, F. (2002). Rational design of reversible acetylcholinesterase inhibitors. *Mini Reviews in Medicinal Chemistry*, 2(1), 27-36.
134. Mishra, C. B., Kumari, S., Manral, A., Prakash, A., Saini, V., Lynn, A. M., & Tiwari, M. (2017). Design, synthesis, in-silico and biological evaluation of novel donepezil derivatives as multi-target-directed ligands for the treatment of Alzheimer's disease. *European journal of medicinal chemistry*, 125, 736-750.
135. Li, Y., Qiang, X., Li, Y., Yang, X., Luo, L., Xiao, G., ... & Deng, Y. (2016). Pterostilbene-O-acetamidoalkylbenzylamines derivatives as novel dual inhibitors of cholinesterase with anti- β -amyloid aggregation and antioxidant properties for the treatment of Alzheimer's disease. *Bioorganic & Medicinal Chemistry Letters*, 26(8), 2035-2039.
136. Sun, Q., Peng, D. Y., Yang, S. G., Zhu, X. L., Yang, W. C., & Yang, G. F. (2014). Syntheses of coumarin-tacrine hybrids as dual-site acetylcholinesterase inhibitors and their activity against butylcholinesterase, A β aggregation, and β -secretase. *Bioorganic & Medicinal Chemistry*, 22(17), 4784-4791.
137. Shidore, M., Machhi, J., Shingala, K., Murumkar, P., Sharma, M. K., Agrawal, N., ... & Yadav, M. R. (2016). Benzylpiperidine-linked diarylthiazoles as potential anti-Alzheimer's agents: synthesis and biological evaluation. *Journal of medicinal chemistry*, 59(12), 5823-5846.
138. Kumar, J., Gill, A., Shaikh, M., Singh, A., Shandilya, A., Jameel, E., ... & Jayaram, B. (2018). Pyrimidine-Triazolopyrimidine and Pyrimidine-Pyridine Hybrids as Potential Acetylcholinesterase Inhibitors for Alzheimer's Disease. *ChemistrySelect*, 3(2), 736-747.
139. Blom, N., Hansen, J., Brunak, S., & Blaas, D. (1996). Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. *Protein Science*, 5(11), 2203-2216.
140. Banerjee, K., Bhat, R., Rao, V. B., Nain, A., Rallapalli, K. L., Gangopadhyay, S., ... & Jayaram, B. (2019). Toward development of generic inhibitors against the 3C proteases of picornaviruses. *The FEBS journal*, 286(4), 765-787.
141. Singh, T., Biswas, D., & Jayaram, B. (2011). AADS-An automated active site identification, docking, and scoring protocol for protein targets based on physicochemical descriptors. *Journal of chemical information and modeling*, 51(10), 2515-2527.
142. Mukherjee, G., & Jayaram, B. (2013). A rapid identification of hit molecules for target proteins via physico-chemical descriptors. *Physical Chemistry Chemical Physics*, 15(23), 9107-9116.
143. Holderbach, S., Adam, L., Jayaram, B., Wade, R. C., & Mukherjee, G. (2020). RASPD+: fast protein-ligand binding free energy prediction using simplified physicochemical features. *Frontiers in molecular biosciences*, 7, 601065.
144. Jain, T., & Jayaram, B. (2005). An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes. *FEBS letters*, 579(29), 6659-6666.
145. Soni, A., Bhat, R., & Jayaram, B. (2020). Improving the binding affinity estimations of protein-ligand complexes using machine-learning facilitated force field method. *Journal of Computer-Aided Molecular Design*, 34(8), 817-830.

