CHAPTER I
INTRODUCTION

1.1 NAMD and molecular dynamics simulation

Molecular modeling and simulation have become a powerful tool for better understanding of fundamental biology and improving the rational design of drugs. A number of molecular dynamics (MD) simulation programs for studying biological molecules have been well-established and commonly-used worldwide either commercial license or free-license such as NAMD[1], CHARMM[2], AMBER[3]. Some of these softwares are made for beginners as well as developers to be easy to understand and modify, offering a flexible structure for user-interface manipulation via external script commands. One of the most popular software when taking about the free-license program is NAMD which was established by the Theoretical Biophysics Group University of Illinois and Beckman Institute, Urbana. NAMD was specially designed for the large scale of biomolecular system, over 100,000 atoms that is compatible with the high-performance computing system, cluster or parallel computing platform, even an individual laptop or desktop computer. It also works together with the molecular graphics program VMD from the same group of developer. The NAMD trajectory file will be easily visualized and analyzed on VMD (For more details and download programs see http://www.ks.uiuc.edu/Research/namd/). The reference paper of NAMD was published by James C. Phillip and co-workers in 2005[1] that describes algorithms in NAMD, key features, and some application of NAMD. Up to July 2013, this work was cited over 3,000 citations and still growing up.

The advantages of NAMD are:

1. Useable for simulations of large biomolecular systems
2. Free of charge
3. Scalable to 200,000 cores for parallel simulation of large system
4. Compatible with AMBER, CHARMM and X-PLOR force field
5. Able to modify source code
6. Cooperative with VMD
7. Able to run on various platforms
8. Able to work with external Tcl or python scripts

Like many other MD simulation programs, the motions of atoms in NAMD are based on the Newtonian equations of motion.

$$m_i\ddot{\vec{r}}_i = -\frac{\partial}{\partial \vec{r}_i}U_{total}(\vec{r}_1,\ \vec{r}_2,\ \ldots,\ \vec{r}_N),\ \ i = 1, 2 \ldots N \qquad (1.1)$$

Where m_i, $\vec{r}_i$ and U_{total} are mass of atom *i*, the position of atom *i* and the total potential energy of the system, respectively. In biological system, the total potential energy is commonly calculated using a summation of the molecular mechanic potential terms.

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{vdW} + U_{Coulomb} \qquad (1.2)$$

The first three terms are bonded terms and the others are non-bonded terms, U_{bond} describes the stretching of covalent bonds which represent the vibration motion between two bonding atoms. (Figure 1.1 A). The bond stretching energy term is expressed as a simple harmonic equation:

$$U_{bond} = \sum_{bond\ i} k_i^{bond}(r_i - r_{0i})^2 \qquad (1.3)$$

Where r_i is the distance of bond *i* and r_{0i} is the equilibrium distance of bond *i*, and k_i^{bond} is the bond constant of bond *i*.

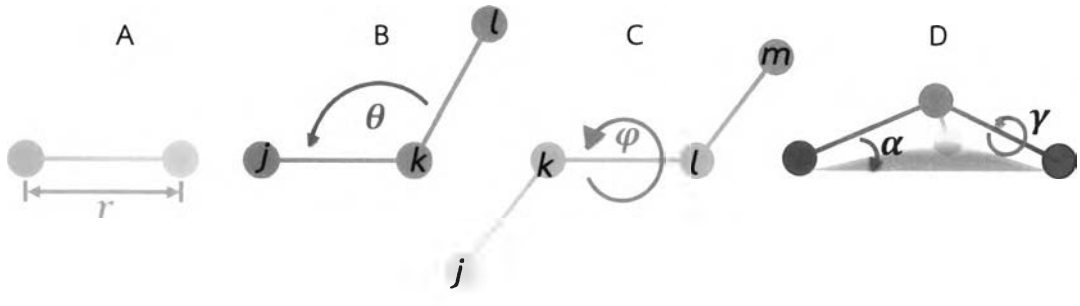

Figure 1.1 Atomic connectivity defined for bonds and angles in the bonded potential energy A. The covalence bonding and the distance *r* between two atoms. B. The angle of three adjacent atoms and θ is the angle defined in U_{angle}. C. The dihedral angle, φ and D. The improper dihedral angle. φ is controlled by α and γ.

U_{angle} describes the angular vibration of atoms *j*, *k* ,*l* that are covalently connected together via two adjacent covalent bonds, (Figure 1.1 B)

$$U_{angle} = \sum_{angle\ i} k_i^{angle}(\theta_i - \theta_{0i})^2 \qquad (1.4)$$

Where θ_i is the angle in radians between two vectors $\overrightarrow{r_{jk}} = \overrightarrow{r_k} - \overrightarrow{r_j}$ and $\overrightarrow{r_{kl}} = \overrightarrow{r_l} - \overrightarrow{r_k}$, the θ_{0i} is the equilibrium angle and of course k_i^{angle} is the angle constant.

The dihedral angle potential $U_{dihedral}$, is the 4-body term representing the angle between two planes, each of which contains three adjacent covalent bond, atom *j*, *k*, *l* and *m* (Figure 1.1 C).

$$U_{dihedral} = \sum_{dihedral\ i} k_i^{dihedral}[1 + cos(n_i\varphi_i - \phi_i)] \quad (1.5)$$

Where φ_i is the angle (in radians) between two planes: the (*j*, *k*, *l*) plane and the (*k*, *l*, *m*) plane. The ϕ_i will be the equilibrium angle when a positive integer, n_i is equal to zero and $k_i^{dihedral}$ is the force constant for dihedral angles. An improper dihedral (Figure 1.1D) the φ is depending on α and γ.

The last two non-bonded terms are the van der Waals term and the Coulomb term. The van der Waals potential, U_{vdW} , represents the interaction of non-directly-bonded pair of atoms. This term can be approximated by using the Lennard-Jones 6-12 potential.

$$U_{vdW} = (-E_{min})\left[\left(\frac{R_{min}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min}}{r_{ij}}\right)^{6}\right] \quad (1.6)$$

Where r_{ij} is the distance between i^{th}- and j^{th}-atom, E_{min} is U_{vdW} at the minimum distance, R_{min}. U_{vdW} represents the attraction or repulsion of paired atoms depending on distances and atom types.

The electrostatic potential or Coulomb term accounts for the repulsive of the same charge and attractive of the difference charge of atoms. Commonly the Coulombic energy term is expressed as

$$U_{Coulomb} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (1.7)$$

Where q_i and q_j are atomic charges of an atom *i* and *j* respectively, the ε_0 is the dielectric constant and r_{ij} gives the distance between atom *i* and *j*.

From the potential terms above, the value of parameters or constants can be obtained from both experimental and empirical method, and represented in terms of the molecular mechanic force field. The force field is the most important part for calculating the potential energy of the system in classical MD simulation.

NAMD can work with the most popular force field such as CHARMM force field or AMBER force field.

1.2 Site-directed spin labeling and electron paramagnetic resonance

Site-directed spin labeling (SDSL) and electron paramagnetic resonance (EPR) spectroscopy provide a powerful tool for studying of the structure and conformational dynamics of membrane proteins. SDSL introduces the spin probe into a protein via a chemical modification of thiol (-SH) group in the cysteine residue with the nitroxide-based unpaired electron and thus gives rise to an EPR signal. (Figure 1.2) Methanethiosulfonate spin label containing the nitroxide moiety is the most commonly used as the spin labeling agent[4].

Generally, three structural parameters can be obtained from SDSL/EPR technique. These are i) probe mobility (ΔH_0^1) which relates to nitroxide dynamics from line-shape analysis, ii) solvent accessibility (Π) of the labeled site through power saturation paramagnetic relaxation experiments and iii) distances from dipolar coupling between nitroxide spins via spectral broadening.

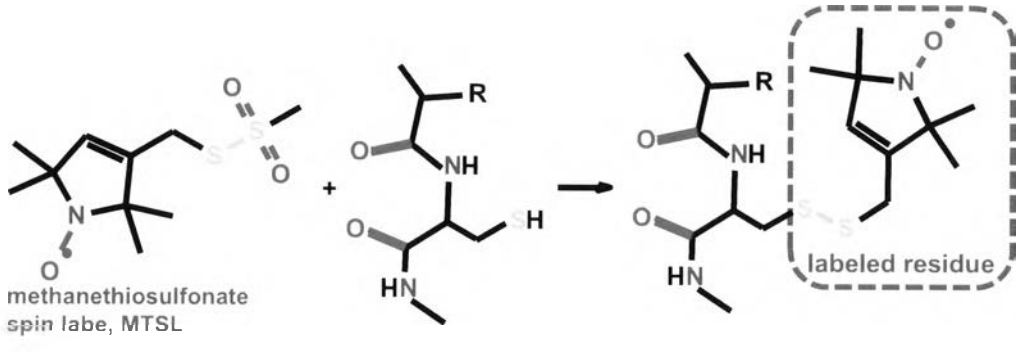

Figure 1.2 The chemical modification of cysteine residue with the spin labeling, methanethiosulfonate spin label (MTSL).

The study in this thesis involves with the method called PaDSAR (Pseudoatom-Driven Solvent Accessibility Refinement). This computational method aims to refine membrane protein structures by making use of two types of solvent accessibility data: molecular oxygen accessibility (ΠO_2) and Nickel-Ethylenediaminediacetate accessibility (ΠNiEDDA). These accessibility data are obtained from EPR measurement of the spin labeled protein sample in the presence of either O_2 gas or the NiEDDA solution. They are paramagnetic and have different solubility properties. O_2 dissolves in hydrophobic bilayer whereas NiEDDA dissolves in aqueous. The value of ΠO_2 and ΠNiEDDA of a spin labeled residue indicates how difficult the spin probe can interact with O_2 or NiEDDA (Figure 1.3).

Based on their solubility properties of O_2 and NiEDDA complex, the solvent accessibility can be used to indicate whether the spin probe is likely in hydrophobic membrane or in aqueous environment, or being buried in the protein core. Typically, a single residue position is subjected to spin labeling one at a time. For a single labeled residue, ΠO_2 and ΠNiEDDA give only information on the local environment of the spin label sidechain which may not be useful for structural characterization.

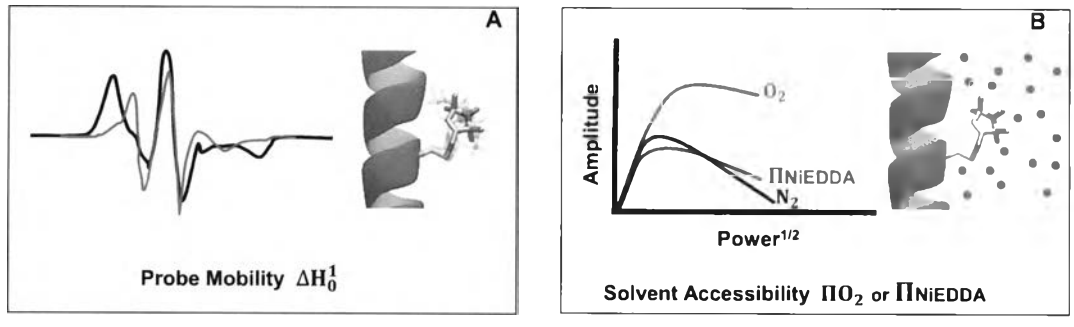

Figure 1.3 The experimental EPR data obtained from SDSL/EPR technique. A. Mobility data B. Solvent accessibility.

However, a more useful information for this experiment is to obtain a set of ΠO_2 and ΠNiEDDA data as a function of amino acid sequence. Therefore, one needs to have a set of several single-point mutants with different cysteine-substituted position. This is because the solvent accessibility profiles of a protein segment provide valuable information on protein secondary structure and tertiary contact. For instance, ΠO_2 and ΠNiEDDA patterns shown in Figure 1.4 is associated with the character of α-helix and its orientation within the lipid bilayer[5].

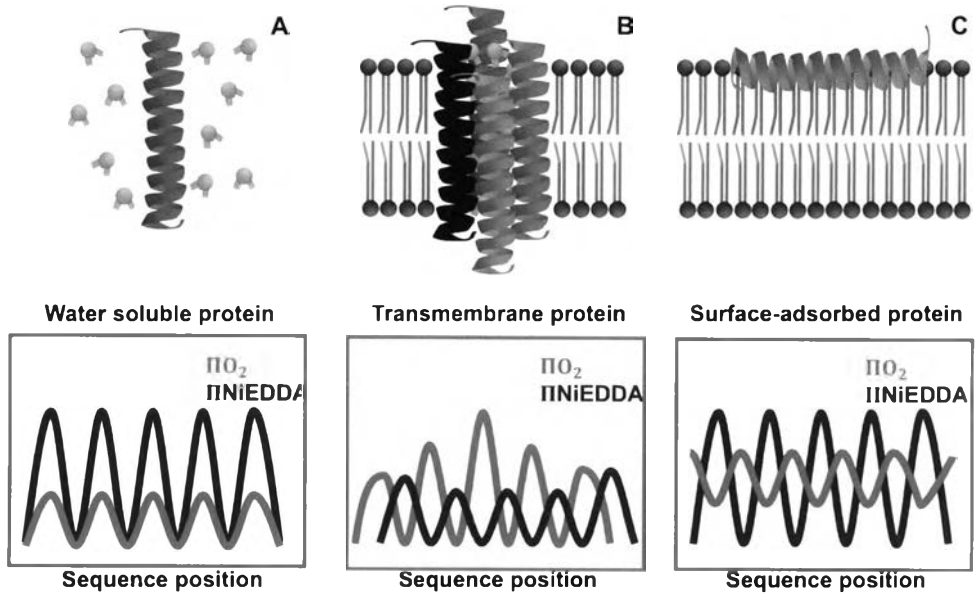

Figure 1.4 Models of three type helical proteins and their ΠO_2, non-polar solvent, ΠNiEDDA, polar solvent, as a function of residue position A. A soluble protein, represents the residues that expose to the solvent B. A transmembrane protein and C. A surface-adsorbed helix.

1.3 Membrane proteins and potassium channels

Proteins that interact with the lipid bilayer are called membrane proteins. It can be classified into three major groups: i) integral, ii) peripheral and iii) lipid-anchored membrane proteins[6]. Integral membrane proteins contain at least one portion or segment passing though the lipid bilayer. Peripheral membrane proteins do not penetrate into the biological membrane, but non-covalently attach on the surface of the lipid membrane. Lipid-anchored proteins bind to the membrane through a specific covalent bond. (Figure 1.5)

Integral membrane proteins are transmembrane proteins that play an important role in transportation of matters though the cells. They contain both hydrophobic and hydrophilic portions. Generally, the hydrophobic portions consist of a series of hydrophobic amino acid residues such as valine, leucine or tryptophan, which can favorably interact with the acyl chain of phospholipid bilayers. The hydrophilic portions that are exposed outside the lipid membrane can interact with ions or polar groups or molecules such as water and phosphate groups of the lipid. Some transmembrane proteins such as ion channels form hydrophilic pore within the membrane to allow the passage of ions across the membrane.

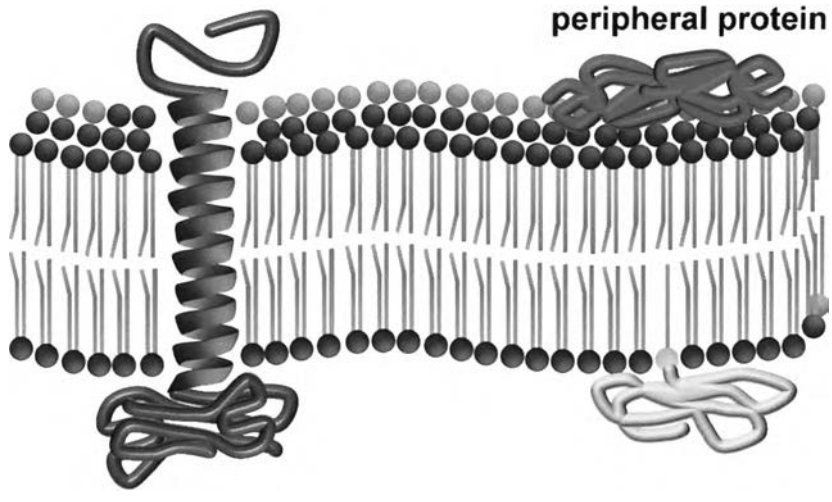

Figure 1.5 Three classes of membrane proteins.

Voltage-gated potassium channels, Kv, are an integral transmembrane protein. They serve the selective transportation of potassium ions, K^+, across the cell to control the electric membrane potential of an excitable cell. The homotetrameric structure of Kv composed of four identical subunits and each subunit consist of six helical transmembrane (TM) proteins S1 to S6 (Figure 1.6). S1 to S4 are voltage-sensor domains and S4 contains positive amino acid residues while S5 and S6 are the ion conduction pore domains and a loop between S5 and S6 is the selectivity filter. Both voltage-sensor domains and conduction pore domains are structural and functional independent[7].

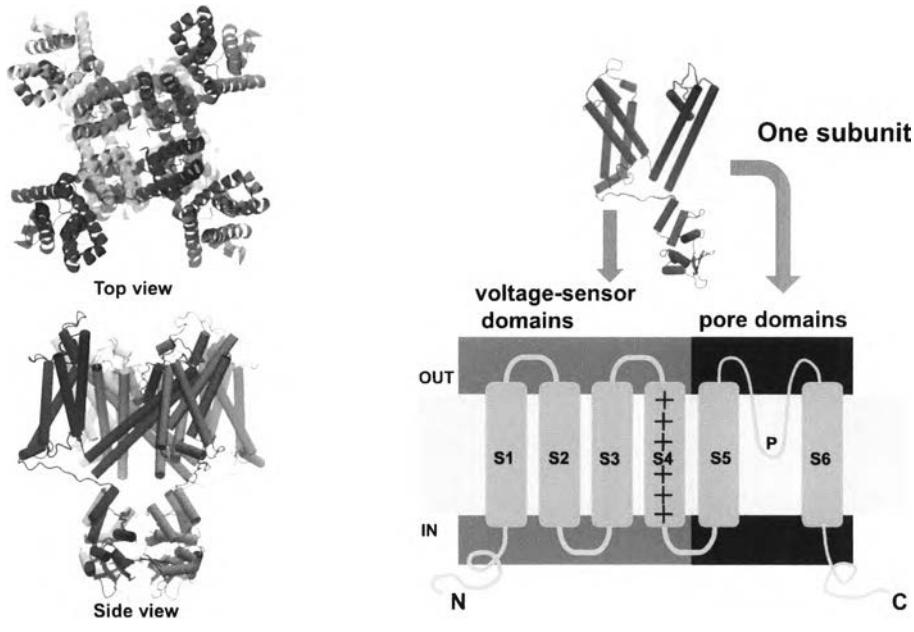

Figure 1.6 The illustration of potassium channel

KcsA is a prokaryotic potassium channel from soil bacteria *Streptomyces lividans*. The high resolution of x-ray crystal structure of KcsA was carried out by MacKinnon, R. and their co-workers[8] with pdb code 1K4C. It contains only two transmembrane helixes (TM1 and TM2) forming pore domains. TM1 and TM2 represent outer and inner helixes, respectively (Figure 1.7)

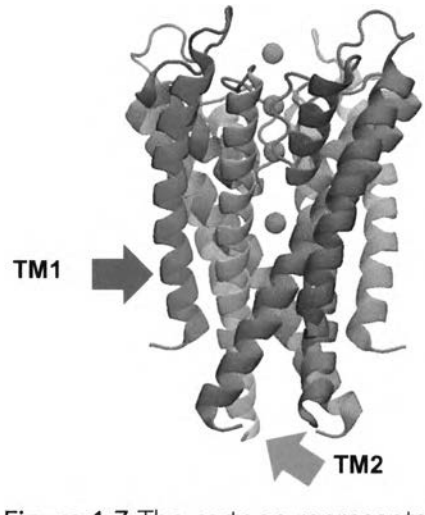

Figure 1.7 The cartoon represents the KcsA.

1.4 Pseudoatom-driven solvent accessibility refinement

Knowledge of structure basis of various states of ion channels at the molecular level in the living cell is essential to understand their functional mechanism. There are two major methods to obtain the three dimensional (3D) high-resolution structure of proteins: x-ray diffraction (XRD) and nuclear magnetic resonance (NMR). However, it is difficult and not straightforward to obtain multiple-state conformations from high-resolution structure techniques especially membrane protein systems. Besides well-known problems of protein expression, crystallization and stability, membrane proteins that are typically carried out in detergent micelles can distort their conformation to non-native state.

As described previously, SDSL/EPR serves as an alternative approach to explore the structure, dynamics and conformational changes for a great number of membrane proteins. The current available techniques of SDSL/EPR offer three sources of structure parameters: mobility (ΔH_0^1), spin accessibility (Π) from paramagnetic agents such as molecular oxygen (ΠO_2) and nickel-ethylenediaminediacetate (ΠNiEDDA) and spin-spin distances (Figure 1.8A). Nevertheless, the low-to-moderate resolution of such structural data has made it insufficient to obtain 3D structure from the direct information. Thus, they have heavily been used on qualitative interpretations, mostly due to the intrinsic limitations of the technique.

There are a number of efforts exploiting SDSL/EPR data for protein structure determination. One of the strategies that have been used to model 3D structures and conformational changes for a number of membrane proteins is called PaDSAR (Pseudoatom-driven solvent accessibility refinement)[9]. PaDSAR is a method for incorporating solvent accessibility data from EPR in the structural refinement of membrane proteins through restrained molecular dynamics (MD) simulation.

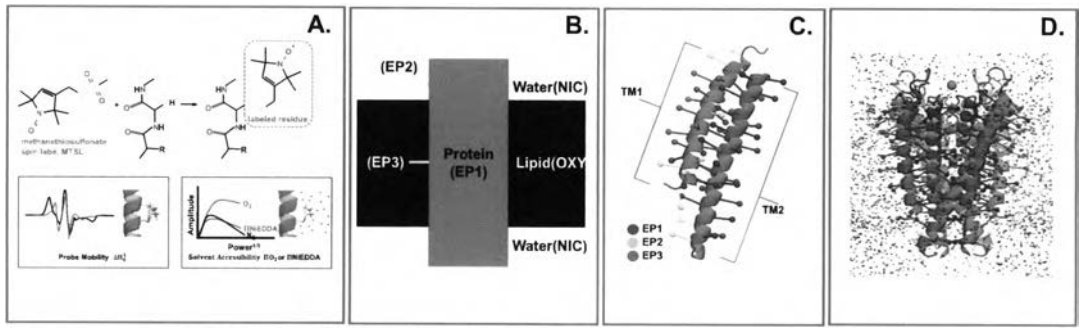

Figure 1.8 A. Electron paramagnetic resonance experimental schemes B. Scenario of a membrane protein having different surrounding environment C. Attachment of pseudoatom EP-type into transmembrane segments of KcsA channel D. A snapshot (a tetramer structure of KcsA channel and the environment pseudoatoms (OXY=red, NIC=blue)).

1.5 Rationale and Objectives

According to the introduction of PaDSAR. CHARMM is one of the most popular MD simulation programs but it is a commercial program and limit user to modify. For wildly use and application of PaDSAR in long term, it is good to implement PaDSAR concept to another simulation program that free to use. NAMD is the best alternative answer. It is not only free of charge but it is also suitable for large scale and parallel simulation and especially allow user to modify program source code. CHARMM and NAMD are very differ in details such as program command, program algorithm, even program feature etc. But both programs provide set of the basic features that equivalence each other.

The main objective of this study is to implement PaDSAR approach to a macromolecular dynamics software, NAMD.