7-8-2008

Interpretation and Prediction of Structural and Energetic Factors Controlling ABC Transporters

xianfeng chen
Georgia State University

INTERPRETATION AND PREDICTION OF STRUCTURAL AND ENERGETIC FACTORS CONTROLLING ABC TRANSPORTERS

-- The implementation and improvement of molecular dynamics simulations

by

Xianfeng Chen

Under the Direction of Dr. Robert W. Harrison

Abstract

ATP Binding Cassette (ABC) transporters are trans-membrane proteins that exist in all phyla. Mutations in this family of proteins can cause inherited diseases like Cystic Fibrosis. ABC transporters consist of dimers of nucleotide binding domains (NBDs) and transmembrane domains (TMDs). NBDs regulate ABC transporters by binding to and hydrolyzing ATP. Although NBD-ATP interactions, NBD-TMD interactions and NBD-water interactions are known to be crucial to the function of these proteins, it is still not clear what structural and energetic factors are involved in the NBD-NTP interactions, how NBD and TMD interact with each other, how water is involved in the functions of ABC transporters and what are the structures and energetics of protein bound water. Molecular modeling and molecular dynamics (MD) simulations were conducted to interpret and predict the structural and energetic factors in control and action of two ABC

transporters, CvaB and SUR2B. Water is essential for ABC transporters to carry out their functions, to increase the accuracy of simulations. Therefore, water potentials in molecular modeling and dynamics simulations were improved based on the calculation of water structures from protein surface.

Previous study showed the NBDs of ABC transporter CvaB bind tighter to GTP than to ATP at lower temperature but not at high temperature. The MD simulations in this study suggested the velocity of water molecules initiates the temperature dependent functional change of proteins. Previous study found that Ser1387 in the NBD of SUR2B, an ABC transporter in vascular smooth muscles, is critical to Kir6.1/SUR2B channel. The molecular modeling and dynamics simulation conducted on SUR2B showed that Ser1387 is located at a region that contacts a TMD. Upon the phosphorylation, the interaction between the NBD and TMD was enhanced which led to an inter domain movement. Water is essential for ABC transporters to carry out their functions, to increase the accuracy of simulations, and, therefore, the structures and energetics of protein bound water were studied. The water radial distribution function for protein bound water was calculated from 105 atomic resolution protein crystal structures and was found to be sharper than that observed for bulk water.

INTERPRETATION AND PREDICTION OF STRUCTURAL AND ENERGETIC FACTORS CONTROLLING ABC TRANSPORTERS

-- The implementation and improvement of molecular dynamics simulations

by

Xianfeng Chen

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Appropriate Degree

in the College of Arts and Sciences

Georgia State University

2008

INTERPRETATION AND PREDICTION OF STRUCTURAL AND ENERGETIC FACTORS CONTROLLING ABC TRANSPORTERS

-- The implementation and improvement of molecular dynamics simulations

by

Xianfeng Chen

Committee Chair: Robert W. Harrison

Committee: Irene T, Weber
Phang C, Tai
Chun Jiang

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2008

Acknowledgements

As the particularity of this research, I would like to thank my Advisor Dr. Robert W. Harrison, a leading scientist in Computational Chemistry and Bioinformatics, and Dr. Irene T. Weber, a leading scientist in Crystallography and Bioinformatics, for providing me the opportunity to work on protein modeling and dynamics simulations. Their guidance, encouragement and sometimes even "pressure", led me to great achievements. I would also like to thank Dr. Phang C. Tai, a leading scientist in Microbial Physiology and Genetics and Chairman of the Biology Department, and Dr. Chun Jiang, a leading scientist in CBP/Cellular and Molecular Physiology, for their complementary scientific expertise and support throughout this research and for their invaluable advices and critical reading of this dissertation.

I would like to thank Drs. Phang C. Tai, Xiangxue Guo, Chun Jiang and Yun Shi for their cooperation with ABC transporters, providing me experimental data and verifying my predictions.

I also thank the colleagues in Drs Harrison and Weber's lab and in the Department for their technical support and friendship. Especially, Ms Yuanfang Wang taught me crystallography and Mr. Preston May provided me computer support.

Finally, I would like to give my deep appreciation to my family members, including my wife Yumei Zhou, parents and in-laws, my brother and sisters for their patience and understanding of my long-lab-staying research.

Table Of Contents

List Of Tables

List Of Figures

CHAPTER

1. Introductions

This dissertation research used two ABC transporters CvaB and SUR2B as examples, made models and ran molecular dynamics simulations on them to study the protein inter domain interactions, protein-ligand interactions, and how water and temperature factors are involved in these interactions. In addition, to increase the accuracy of simulations, water potentials in molecular modeling and dynamics simulation were improved based on the calculation of water structures from protein surface. Therefore, this introduction includes three parts: ABC transporters, molecular dynamics and water structures on protein surface.

1. ABC transporters

The ATP Binding Cassette (ABC) transporter is an important protein superfamily and widely exists in all phyla. More than 80 ABC transporters in *E. coli* (1) and 48 in humans (2) have been identified. Transporters are involved in many cellular processes including nutrient uptake, antigen processing, bacterial immunity, cell division, cholesterol and lipid trafficking, and osmotic homoeostasis (3). They export toxins, proteases, antibiotics, ions and so on, and import nutrients such as histidine, maltose, ribose, oligopeptide, vitamin B12, and ions (4). The mutation of transporter proteins will cause cell functional change or disability. 14 transporters are found related to human disease due to mutation, such as *CFTR* which causes Cystic Fibrosis characterized by defective exocrine activity of lung, pancreas, sweat ducts and intestine (5,6), *TAP1* which causes inherited immunodeficiency (7), and *MDR1* which is overexpressed in tumor cells and shows resistance to multiple drugs (8). Understanding the mechanism of this superfamily will definitely help in developing therapy to improve our health.

1.1 The ABCtransporter core structure and mechanism for transportation

ABC transporters are cross-membrane proteins, as reviewed(3). The members of ABC superfamily vary widely in their components, but share a conserved core structure of two nucleotide domains (NBDs) and two transmembrane domains (TMDs) (figure 1-1a). The NBD dimer binds to nucleotides and has ATPase activity which provides energy for the transportation. The NBD amino acid sequence is more conserved indicating the functions it carries out are common in all the ABC family. The TMDs form the channel for transporting the allocrites (the transported substrates). The TMD is less conserved among ABC family than the NBD because of the chemical diversity of allocrites. Transportation needs the cooperation between domains, the inter domain communications in an ABC transporter are mediated by inter domain/NTP/allocrite interactions. For example, the dimerization of ABC transporter is pre-required for NTP hydrolysis (9), NBD-NTP interactions are essential for dimerization (9), NBD-TMD interaction is the communication route between the two domains (10,11) which transfers the hydrolysis energy from NBDs to TMDs for transportation and passes the signals of TMD-alloscrites binding to NBDs to trigger NTP hydrolysis.

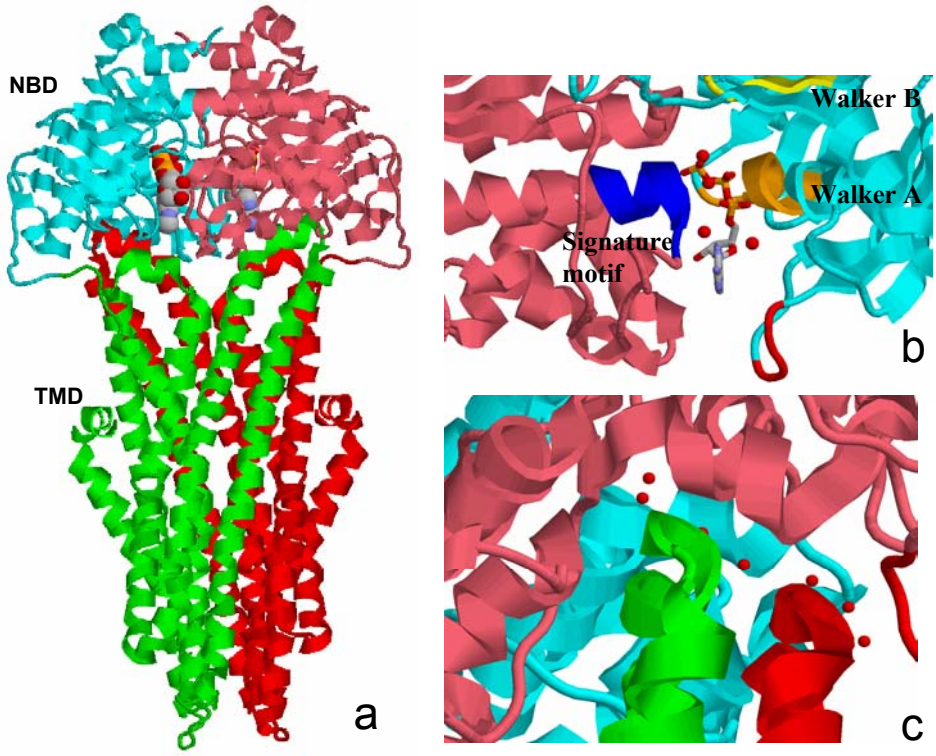


Figure 1-1 the structure of ABC transporter.

a) The core structure of ABC transporter (SAV1866) (12). Two NBDs are on the top (cyan and pink) with ADP bound (CPK, ball). The two TMDs are at the bottom (green and red). b) The nucleotide binding site (HlyB) (13) showing ATP (CPK, stick), Walker A (orange), Walker B (yellow), signature motif (blue), the corresponding double mutation site of CvaB (red) and water molecules (red dot). c) The NBD-TMD interaction region (SAV1866) showing NBDs (cyan and pink), TMDs (green and red) and water molecules (red dot).

1.2 Nucleotide binding site

Based on a several dimer crystal structures of ABC transporters such as HlyB (13), HisP (14), MJ0796 (15), MalK (16,17), GlcV (18) and SAV1866(12), the nucleotide binding site is at the junction of the two NBD subunits. The two binding sites are only formed when the protein is in a dimerized state. Each binding site is composed

of the highly conserved sequences Walker A, Walker B in one subunit and signature motif in the other subunit. The Walker A and B sites interact with the oxygen atoms of the β- and γ-phosphates of NTP and orientate the nucleotide phosphates in the binding site (19). The amino acid residues from the ABC Signature motif are also involved in the binding of nucleotide as shown in the dimer structures of ABC transporters (15,16,20,21) (figure 1-1b). Some single residues, such as Q-loop (22-24), D-loop (20,24-26) and H-loop (27-29) either directly interact with nucleotide or are mediated by a nucleotide-binding cofactor magnesium (Mg^{2+}) (30). The bound magnesium also mediates the interactions between the oxygen atoms from β - and γ-phosphates of the NTP and from residues of the Walker A and B sites (18). It neutralizes the negative charges of phosphates, stabilizes dimer structure by enhancing NBD-NTP interactions (9), and directly participates the nucleotide hydrolysis (19). The magnesium along with Walker A, Walker B, signature motif and the other conserved single residues in the substrate binding site, such as Q-loop (22-24), D-loop (20,24-26), and H-loop (27-29) are important to the functions of ABC transporters, removal of magnesium or mutations in those regions breaks down ATPase activity resulting in the loss of transport function (27-29,31).

1.3 NBD-nucleotide interaction and dimerization

Nucleotide interacts with Walker A, Walker B, on one subunit and signature motif and on the other subunit and facilitates the two subunits to form a dimer. For example, dimer will not be detected except in the presence of ATP in case of CvaB(NBD) (9) and HlyB(NBD) (29). Rad50 (20) forms monomer crystal structure without ATP and dimer with bound ATP. Furthermore, the tighter binding affinity between nucleotide and dimer, the more dimers are formed. For example, GTP possesses an additional amino group at

the 2nd carbon of guanine relative to ATP. UV cross-linking (9) between nucleotides and CvaB(NBD) dimer shows more intense band from bound GTP than ATP under the same condition. The model of CvaB(NBD) dimer structure indicates the additional amino group forms hydrogen bond with Y^{501}DSQ (9). The scanning data from Western blot (32) reveals that the oligomers of the CvaB(NBD) can be induced to dimerize by both GTP-Mg2+ and GDP-Mg2+, but GTP-Mg2+ is faster than GDP-Mg2+. The modeled structure of CvaB(NBD) (9) predicts the γ-phosphorate group in GTP forms additional interactions with residues in Walker A and Walker B regions. This evidence suggests that dimerization and NTP binding are intrinsically linked.

1.4 The conformations of ABC transporters

In the ABC transportation cycle, ABC transporters shift between different states which are indicated by different conformations from several crystal structures of dimers (12,14,15,17,33). For example, in SAV1866 (12) and AcrB (33), two exporters from bacteria, the transportation cycle includes two important states: an inward-facing conformation with the allocrite binding site accessible from the cell interior, and an outward-facing conformation with an extrusion pocket for allocrite exposed to the external medium. Changes between states or conformations are mediated by NBD-NBD and NBD-TMD interactions upon NTP binding/release, hydrolysis, phosphorylation and allocrite binding/release (figure 1-2).

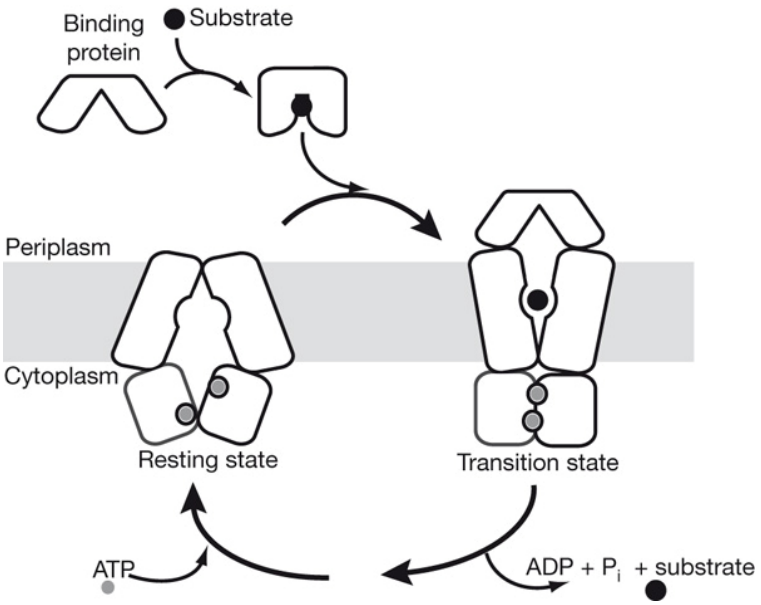


Figure 1-2 The conformations of ABC transporter

The resting state and transition state correspond to an inward conformation and an outward conformation respectively (figure is taken from Oldham et al (34))

1.5 Water and ABC transporters

Water is important in diffusion, solubility, and aggregation of protein and mediates protein-protein, protein-ligand interactions. Water molecules even directly participate in enzyme hydrolysis of substrates. In the case of ABC transporters, a single mutation can even change the solubility dramatically in HlyB (35) indicating water-protein interaction is important to keep protein in solution. In crystal structures, water molecules are found located between NBD and NBD (21), NBD and nucleotide (21) (figure 1-1b), NBD and TMD (12) (figure 1-1c), and fill in the empty space of allocrite binding pocket (3) suggesting that water molecules mediate interactions between them. The water molecules between Q-loop, bound magnesium and the oxygen of γ-phosphate even directly participate in the chemical reaction of NTP hydrolysis, providing energy for

transportation (36). Therefore, water molecules are highly involved in the critical functions of ABC transporters.

1.6 The introduction to CvaB

CvaB is an ABC transporter located on the cytoplasmic membrane of *E. coli* and plays pivotal roles in the export of antibacterial toxin colicin V (37). Colicin V (ColV), produced by *Enterobacteriaciae*, kills sensitive target cells by disrupting their cell membrane potential (38). ColV is secreted by a dedicated type I exporter system across both the cytoplamic and outer membranes directly into the environment outside the cells. At least three protein components participate in the ColV export system: two plasmid-borne proteins CvaA and CvaB in the cytoplasmic membrane and a host chromosomal gene product TolC in the outer membrane (37,39).

Like the other members in this family, such as the α-hemolysin exporter HlyB (13), the cystic fibrosis transmembrane conductance regulator (40), and the multidrug resistance P-glycoprotein (41), CvaB consists of dimerized transmembrane domains (TMDs) and dimerized nucleotide binding domains (NBDs) (9). The TMDs form the channel for colicin V secretion. The NBDs have ATPase activity which provides energy for the transportation. The sequence of NBD is much conserved in ABC transporters. Regions specific to this super family are also found in this protein such as Walker A, Walker B, Q-loop, D-loop, H-loop and signature motif. Thus, a study of the NBD domain of CvaB will help to understand the functions of this domain in the other family members.

Recently, UV cross-linking was conducted on the binding affinity between nucleotide and NBD domain of CvaB and its homolog HlyB (13) by Guo et al (9). They

found wild type CvaB(NBD) prefers binding of GTP over ATP at low temperature (273K) but not at high temperature (310K). As a control, HlyB(NBD) does not prefer GTP binding at both temperatures, indicating that the nucleotide binding of CvaB(NBD) is temperature sensitive. Further more, Guo et al found that the double mutation S503A/Q504L of CvaB(NBD) can eliminate this property, and acts like HlyB(NBD) (figure 1-1). The double mutation site is located in a loop region Y^{501}DSQ in CvaB(NBD) and the aromatic base of Y^{501} has a π-π stack interaction with the purine base of ATP in model. Even though Guo et al found that the double mutation site is close to the amino group at the 2nd carbon of guanine in GTP and may have effect on the protein-GTP binding, it is known how this effect changes with temperature.

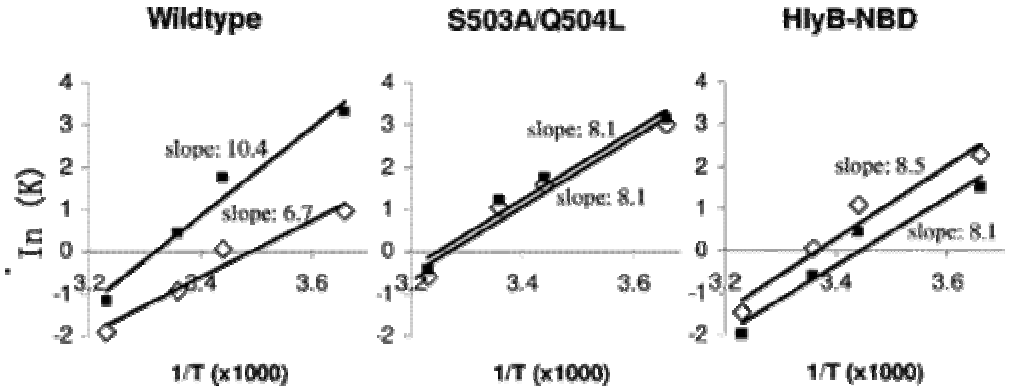


Figure 1-3 Temperature sensitivity of the binding affinity

Temperature sensitivity of the binding affinity for ATP (◆) and GTP (■). X-axis and Y-axis show temperature and binding affinity respectively. (Figure is taken from Guo et al 2006 (9)).

Vitkup et al (42) found that the fluctuations of the solvent atoms (mostly water), rather than protein atoms, mostly respond to temperature changes. Later on, Chen et al found that the water-protein interactions actually change the structures of proteins (43). Therefore, "the temperature of the solvent and thus its mobility is the dominant factor in determining the functionally important protein fluctuations", as stated by Martin Karplus

and J. Andrew McCammon (44). Then the question pops up: how do the velocities of water molecules trigger the temperature dependent functional changes of CvaB? The functional difference between wild type CvaB and the double mutant shows that the local effect of Y^{501}DSQ is critical to the temperature dependence of CvaB, and therefore provides a good example to study temperature effects on protein by only focusing on limited structural and energetic factors at a local region. This study will run MD simulation on CvaB(NBD) at 273K and 310K to answer the above question.

1.7 The introduction to SUR2B

SUR2B is a regulatory sulfonylurea receptor and a component of Kir6.1/SUR2B channel, the major isoform of vascular ATP-sensitive K^+ channels in vascular smooth muscles (45-48). ATP-sensitive K+ (KATP) channels play an important role in vascular tone regulation (49-51). The KATP channels are expressed in vascular smooth muscles (VSMs) and activated by several vasodilating hormones and neurotransmitters through the Gs-cAMP-PKA signaling system (49,52,53). The channel activation leads to hyperpolarization of VSM cells, decrease in voltage-dependent Ca2+ channel activity and relaxation of resistant arteries (54).

The channel does not open spontaneously at rest. Several groups of channel openers activate the channel, including pharmacological KATP channel openers (KCOs, e.g., pinacidil and nicorandil) (55), metabolites (MgADP, acidosis) (56,57), and hormonal vasodilators and neurotransmitters (calcitonin gene-related peptide, epoxyeicosatrienoic acids, β-adrenergic receptor agonists and vasoactive intestinal peptide) (53,58-60) . KCOs and Mg2+-nucleotides activate the KATP channels via binding to the SUR

subunits (55,61). The hormonal vasodilators activate the vascular KATP channel through direct phosphorylation of the channel protein by PKA (52,53).

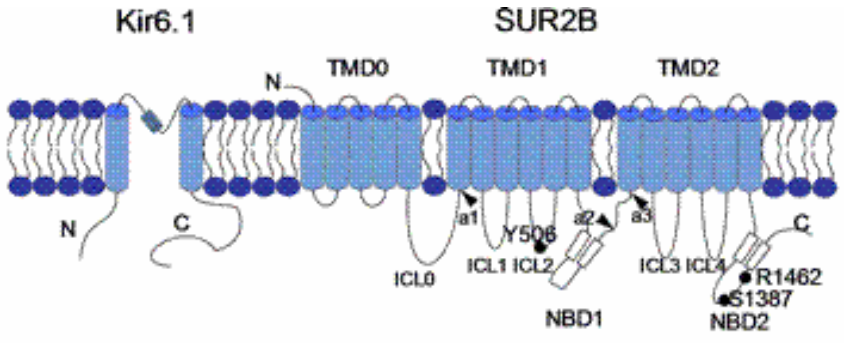


Figure 1-4 the structure of Kir6.1/SUR2B channel.

The TMDs of kir6.1 form the core structure of the channel. The interactions between Kir6.1 and SUR2B are mediated by TMDs of SUR2B.

SUR2B is a heterodimer like other ABC transporters. It consists of a nucleotide binding domain (NBD) and transmembrane domain (TMD) named NBD1, NBD2, TMD1 and TMD2. In addition, SURs have another transmembrane domain containing 5 helical segments termed TMD0 (figure 1-3). Several lines of evidence suggest that TMDs of SUR2B play a major role in interacting with Kir6 subunits (62). Such interaction has been observed in recombinant Kir6.2/SUR1 channels using electronic microscopy (63). The NBD domains form four intracellular linking regions with TMD domains (ICL1-4).

Previous study showed that the phosphorylation of Ser1387 in NBD of SUR2B is critical for Kir6.1/SUR2B channel opening (53), indicating the phosphorylation causes interdomain interaction and movement. However, it was not known where the Ser1387 was in the 3-D structure of NBD and how the phosphorylation of Ser1387 causes inter domain interaction and movement. In this study, SUR2B model will be made based on a bacterial ABC transporter (SAV1866) (12) to predict the 3-D position and the

interactions of the Ser1387. Then molecular dynamics will be applied to SUR2B before and after the phosphorylation of Ser1387 to study the interdomain movement.

2. Molecular Dynamics

MD simulation is a computer simulation wherein atoms in protein molecules are allowed to interact with each other following the laws of mathematics, physics, and chemistry. Algorithms from computer science and information theory are employed to give a view of the motion of the atoms during a period of time. A macro system consists of a pool of particles such as molecules, atoms, and ions, which is difficult to analyze the properties of the individual components from such a complicated system. MD simulation simplifies this problem and mimics the behaviors of a limited number of molecules (usually only one protein molecule with some small molecules such as water and ions) by using numerical methods and provides a "virtual experiment". MD simulations can be run to make a prediction and provide guidance to a laboratory experiment and/or interpret the experimental results and establish a new theory.

2.1 Free energy surface

The free energy surface of a protein molecule (figure 1-4) can be considered as a multi-dimensional broad and ragged landscape with one global minimum and many local minima. The path from one minimum to another is hindered by energy barriers. Normally, the native structure of a protein in a dynamic state is considered to stay in its global minimum or sometimes jumps to several local minima, each associated with a particular conformation. The distribution of the protein native structure among the conformations is determined by the free energies of the global and local minima. The less

free energy a global or local minimum is, the more distribution of the native structure the conformation shares.

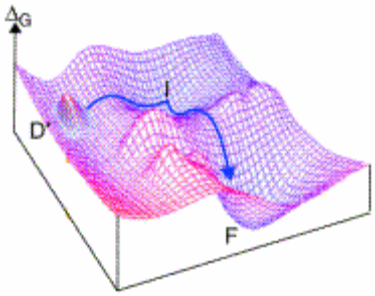


Figure 1-5 The free energy surface in a three-dimensional space.

D and ***F*** represent a local minimum and a global minimum respectively, and ***I***, amongst of the path from ***D*** to ***F***, represents a transition state. Taken from Gruebele 2002 (64).

2.2 Force fields

An MD simulation with classical method requires the definition of a force field (also known as a potential function) which is a description of the potential energy surface. The potential energy surface is represented by potentials in the force field that define how the particles such as atoms, ions and small molecules will interact with each other in a system such as a protein in aqueous environment. A force field includes functional forms of potentials. For example, Lennard-Jones potential is used for calculating van der Waals interactions (65). Tersoff potential is used for a pair wise interaction effected by its neighborhood atoms (66). Coulomb's law is used to calculate a pair wise ionic interaction. In addition to the functional form of the potentials, a force field defines a set of parameters for each type of atom. For example, a force field would include distinct parameters for an oxygen atom in a carbonyl functional group, in a hydroxyl group or in a water molecule. The typical parameter set includes values for atomic mass, van der Waals radius, and partial charge for individual atoms, and

equilibrium values of bond lengths, bond angles, and dihedral angles for pairs, triplets, and quadruplets of bonded atoms, and values corresponding to the effective spring constant for each potential. A force field is developed by empirical fitting from experimental data.

2.3 Energy minimization

An MD simulation of a protein starts from an initial structure such as a crystal, or NMR structure. Not all proteins have available crystal or NMR structures, such as CvaB and SUR2B, thus, a modeled initial structure must be built first based on a homologous protein with an available crystal or NMR structure and optimized by energy minimization.

Energy minimization or energy optimization is a common technique to calculate protein folding, normally, from a distorted structure. The basic idea is that a stable state of a protein molecule should correspond to a local or global minimum of the free energy surface (64). The energy minimization can be considered as a special type of MD which assumes a protein at very low temperature with kinetic energy close to zero and thus the free energy is very close to its potential energy. Then the mathematical procedure of optimization algorithms such as conjugate gradient or steepest descent is applied to the unfolded protein in such a way that atoms in the protein structure are moved to reduce the net forces or the gradients of potential energy to nearly zero, corresponding to a local or global minimum. This technical is also applied to the prediction of a protein structure from a homologous protein with a known crystal or NMR structure, known as homology modeling or comparative modeling. This kind of calculation generally starts from a raw structure by transferring the atom coordinates of a homologous protein to the target

protein after sequence alignment. Then energy minimization is applied to remove steric clashes and optimize the atomic geometry and internal energy of the raw structure.

Energy minimization has a drawback in the calculation of protein structure. In the landscape of the free energy surface of a protein, where a global minimum may exist along with many local minima, the native structure of the protein is considered to be in the global minimum like a crystal or NMR structure. However energy minimization only drives the free energy of protein structure downhill. If the free energy of the protein structure at the starting point is far from the global minimum, it may be trapped in a local minimum during energy minimization leading to incorrect structure prediction. This limitation affects the correctness of the homology model as well. Normally, homologous proteins share the similar primary, second, tertiary structures, and therefore the free energy surfaces that may have their global minimum close to each other. Thus one can predict a native protein structure by energy minimization based on a homologous template. The correctness of the predicted static structure is highly dependent on the similarity of the template. Because the less similar the two proteins are, the farther the global minima of the two proteins are, and the energy minimization will be more likely to be trapped into a local minimum that is far from the global minimum.

2.4 Molecular dynamics theories

Proteins are large molecules, so the MD simulations are computationally expensive. Many theories and algorithms are applied to MD simulations to speed up the calculations. In classical methods, the electrons and nucleus of an atom are treated separately based on Born-Oppenheimer approximation (67). The effect of the electrons of each atom is approximated as a single potential energy surface, usually representing the

ground state. The nucleus, much heavier than electrons and representing the weight of the atom, follows classical Newtonian dynamics. Thus, classical methods are widely used to study the molecular motion, vibration and thermodynamics of proteins.

The energy expression of classical methods can be written as an abstract form, consisting of the bonding and nonbonding potential energies, or as individual terms:

$V_{\text{total}} = \sum (V_{\text{bonding}} + V_{\text{nonbonding}})$

$V(\text{conformations}) = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{torsion}} + E_{\text{vdw}} + E_{\text{electrostatic}}$

There are also many algorithms applied to calculate MD trajectory. In case of Leap-frog, the position of an atom at time $t + \Delta t$ is calculated from

$$r(t + \Delta t) = r(t) + v\left(t + \frac{1}{2}\Delta t\right)\Delta t$$

The velocity of an atom is calculated from

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + a(t)\Delta t$$

Where, *a* is the acceleration calculated from the energy expression based on Newton's law.

2.5 Molecular dynamics and CvaB

Temperature is an important factor that affects the functions of proteins. Many protein-ligand systems have temperature dependent effects, either as a feature of the native protein or as the result of the temperature sensitive mutation. When temperature changes, the vibrations of atoms (bond stretching and angle bending) within a protein molecule change which can lead conformational changes and therefore promote chemical

reactions, hormone-receptor binding, and even trigger other complex processes such as signal transduction.

The study of temperature dependent functions of protein by MD simulation with classical methods is often restricted because many experimental data cannot be simulated. For example, temperature can change many properties of proteins such as polymer formation, aggregation, phase transition, radiation emission, light absorption, conformation of structure, molecular motion and even chemical reactions. In classical methods, radiation emission, light absorption and chemical reactions cannot be simulated. In addition, since MD simulation generally mimics a single protein molecule out of a pool of molecules in a macro system, polymer formation, aggregation and even phase transition cannot be simulated either.

Another restriction in the study of the temperature dependent function of protein is that a control system is often not available. Proteins are characterized by their primary, secondary and tertiary structures. Many studies relate the functional changes of protein to their structural changes originated by primary structure changes such as a mutation. Without a control system such as a point mutation, it is hard to focus on what structural factors are affected by temperature and how the structural factors relate to the functional changes of proteins.

The experiments done by Guo et al 2006 (9) identified a specific temperature dependent effect. The ABC transporter CvaB has a preference for GTP over ATP at low temperature but not at high temperature, and a double mutation at NTP binding site can remove this temperature dependent function. The significance of Guo et al's research to this study is to provide a series of experimental data in which the property changes of

proteins can be simulated properly. Temperature can be taken as a parameter, and the double mutation can be easily modeled by making a "virtual mutation", UV cross-linking bands can be treated as representing relative binding affinities which can be calculated as external binding energies of ATP or GTP, and the double mutation can be taken as a control system to focus on the structural factors causing the temperature sensitivity of CvaB(NBD).

This study ran MD simulations with classical methods on CvaB(NBD) and its mutants at different temperatures because conventional energy minimization calculations or single temperature molecular dynamics cannot explain Guo's experimental data, nor can these calculations explain how point mutations alter this temperature dependence. Two things were examined: 1) the consistency of the binding affinity of CvaB(NBD)-NTP between experimental data and prediction to answer the question whether classical MD simulations are accurate enough to study the temperature dependent function. Then based on this agreement, 2) what energetic and structural factors cause wild type CvaB(NBD) to bind GTP preferentially over ATP at low temperature.

2.6 Molecular dynamics and SUR2B

Previous study showed that the phosphorylation of Ser1387 in NBD of SUR2B is critical for Kir6.1/SUR2B channel opening (52), indicating the phosphorylation causes inter domain interaction and movement. However, the location of Ser1387 in the 3-D structure of NBD and how the Ser1387 causes inter domain interaction and movement upon phosphorylation were unknown.

Among ABC transporters whose crystal have been solved, Sav1866 (PDB entry 2HYD) (12), from the bacterium *S. aureus*, has the highest similarity with SUR2B with

31% identity in NBD and 21% in TMD. Thus, the location of the Ser1387 can be predicted by homology modeling. SUR2B presents presumably at least two state or conformations upon phosphorylation. The inter domain movement of SUR2B after phosphorylation is a conformational change wherein a multi-residue movement between NBDs and TMDs is accompanied by multi-bond breaking and forming. However, in the free energy surface landscape, the two conformations of SUR2B should have different global minimum with energy barriers between them. Thus if the SUR2B structures of the two conformations are built based on the single structure 2HYD, at least one of the energy minimizations has the danger of being trapped in a local minimum resulting in a reduced inter domain movement.

Unlike energy minimization resulting in a static structure, molecular dynamics movies a protein molecule through dynamic states where atoms in the protein are assigned with kinetic energy under a certain temperature. The assigned kinetic energy allows the protein to roam along the free energy surface and climb energy barriers between minima. The higher the temperature, the stronger the ability of the protein to climb the free energy barriers. Thus, even if the starting point of free energy is not the global minimum of the protein, the kinetic energy may bring protein cross the energy barriers to reach the global free energy minimum. In this dissertation, two SUR2B models will be made based on a bacterial ABC transporter (SAV1866) (12) before and after phosphorylation. Molecular dynamics will be run on the two modeled structures to study interdomain movement upon the phosphorylation of Ser1387.

3. Water structures

3.1 The general water structures

A water molecule is composed of an oxygen atom covalently bound to two hydrogen atoms forming a tiny V-shape with an angle of 109°. The complexity is that water molecules are polar and able to form hydrogen bonds with each other or the other ions and result in a complex network structure. For example, metal ions form complexes with water molecules (68), water molecules form four-water-molecule structures with the formation of bifurcated bonds (69) as well as two-water-molecule structures with one proton attached ($H_5^+O_2$) (70). They also form clusters such as ten-molecule tetrahedral clusters (71) and 280-Molecule expanded icosahedral water cluster consisting of 14-molecule tetrahedral units (72). In addition, these complex water structures are sensitive to temperature, gas pressure, ion concentration, pH value, and many other physical conditions.

3.2 The category of water

Water forms 70~95% of the mass of the cell and is essential to life. Many biological processes are driven by water-protein interactions (73,74). Water provides a platform for proteins to fold, solve, diffuse, hydrolyze, interact with the other molecules such as the bindings of enzymes to substrates and antibodies to antigens, and even the formation of micelles and bilayers. Generally, water around a protein molecule can be divided into three forms: 1) bulk water that surrounds the protein molecule without binding to it, 2) individually bound water that is directly bound to charged or polar atoms of protein by hydrogen bonds in cavities or inside of the protein, and 3) hydration water

that is attached to protein surface by directly interacting with the charged or polar atoms of protein or indirectly through the other water molecules forming water networks around the protein surface. Hydration water keeps the protein molecules separated from each other in solution (75). The difference between hydration water and individually bound water is defined by the local structural geometry. Individually bound water is found in a specific binding site similar to how a ligand would be bound, and usually conserved among similar proteins, while hydration water is not bound in a specific manner. Individually bound water and hydration water are also known as protein bound water.

3.3 Water and molecular dynamics

In addition to the general water-protein interactions, in the case of ABC transporters, water molecules directly participate in ATP hydrolysis and mediate interactions between NBD-substrate (21), NBD-NBD (21), NBD-TMD (12) and TMD-allocrite (3). It is important to include water molecules in molecular dynamic simulations (MD) to study protein function. Therefore, the type of water potential and the representation of water structure are critical in computer modeling of proteins. The current water potentials in MD programs are based on some water models such as TIP3P (76), TIP4P, TIP5P (77) and SPC (78) that are calculated from bulk water. Applying these water potentials to hydration surface in MD simulation can cause disagreement with experimental data. For example, if TIP3P is used with the AMBER (79) or CHARMM (80) force fields, the calculated carbohydrate-water interactions on the surface of disaccharides are less structured than observed (81). This disagreement may be caused by two of the following: 1) hydration water and bulk water differ from each other; 2) the MD can only include a limited number of water molecules around the protein surface due

to the expense of calculation. This may cause bias. Since the force field in classical method is obtained by empirical fitting, in either case, an elaboration of re-parameterization of water potentials will increase the correctness of MD calculations. In this dissertation, the water radial distribution will be calculated from a number of protein crystal structures with bound water molecules and based on which water potentials will be improved.

Specific aims of this dissertation

This dissertation has three aims. Two of the aims are the implementations of molecular modeling and dynamics for interpreting experimental data and predicting the structures and functions, two ABC transporters. One of the aims is the improvement of hydration water potentials applied in molecular modeling and dynamics by the analysis of water structures on protein surface.

SA1: Interpretation of the temperature sensitivity of ATP/GTP binding preferential of CvaB(NBD).

Many protein-ligand systems have temperature dependence, either as a feature of the native protein or as the result of the temperature sensitive mutation. One example is the nucleotide binding domain of the ABC transporter CvaB, which has a preference for GTP over ATP at low temperature but not at high temperature, but a double mutation at nucleotide binding site can eliminate this temperature effect. This cannot be explained by conventional energy minimization calculations or single temperature molecular dynamics. To study the effects of temperature on the function of a protein, MD simulations will be run for wild type CvaB(NBD)-NTP and mutants (as controls) at low and high temperature, respectively, to interpret how the structural factors at nucleotide binding site affect the temperature dependent function of CvaB(NBD).

SA2: Prediction of the structure and inter domain movement of SUR2B

SUR2B is a regulatory sulfonylurea receptor and a component of Kir6.1/SUR2B channel, the major isoform of vascular ATP-sensitive K^+ channels in vascular smooth muscles (45-47). Previous study showed that the phosphorylation of Ser1387 in NBD of SUR2B is critical for Kir6.1/SUR2B channel opening (52), indicating the

phosphorylation causes inter domain interaction and movement. To understand where the Ser1387 is in NBD and how the Ser1387 causes inter domain movement upon phosphorylation, molecular modeling and MD simulations will be made and run on SUR2B and its phosphorylated form to make predictions which will be verified by experiments afterward.

SA3: Calculation of water structures on protein surface

Water is important for a protein to diffuse, solve, folding and interact with other molecules, and it should be included in molecular modeling and dynamics. In classical methods, water potentials are set by empirical fitting. The experimental data of water structures on protein surface is useful to improve water potentials. To increase the accuracy of the simulations applied to proteins such as CvaB and SUR2B, water structures are calculated from 105 atomic resolution protein crystal structures.

2. Investigation of Temperature Dependent ATP/GTP Binding Preference of CvaB by Molecular Dynamics Simulations

Investigation of Temperature Dependent ATP/GTP Binding Preference of CvaB by Molecular Dynamics Simulations

*Xianfeng Chen[1], Phang C. Tai[1], Irene Weber[1] and Robert W. Harrison[1,2]**

[1]Department of Biology and [2]Department of Computer Science, Georgia State University, 33 Gilmer Street, Atlanta, Georgia 30302–4010, USA.

Email address: xchen5@gsu.edu, iweber@gsu.edu and rhrrison@cs.gsu.edu

* Correspondence to: Phone: 404-413-5724; Fax: 404-413-5717; E-mail: rharrison@cs.gsu.edu.

Abstract

While much progress has been made in understanding the physical basis of molecular recognition and this progress has translated into effective techniques for molecular and drug design, one area of protein-ligand interaction has remained stubbornly intractable. Many protein-ligand systems have temperature dependence, either as a feature of the native protein or as the result of the temperature sensitive mutation. One example system is the ABC transporter CvaB, which shows a preference for GTP over ATP at low temperature (273K) but not at high temperature (310K), and a double mutation at nucleotide binding site can eliminate this temperature dependent preference. Conventional energy minimization calculations or single temperature molecular dynamics cannot explain this. Nor can these calculations explain how point mutations alter this temperature dependence. In this paper, we showed that molecular dynamics performed at different temperatures and averaged over many independent trajectories can reproduce the temperature dependent preferences of not only the native enzyme, but also the mutant. The simulations implied that the velocities of water molecules and water-protein interactions play a pivotal role in temperature dependent functional changes. The mechanism discussed in this study may also be applied to the other temperature dependent functions of proteins.

1. Introduction

Temperature is an important factor in the functions of proteins. Studies show that some defective strains of bacteria or diseases of humans result from the gain or loss of temperature sensitivity of proteins (82-85). In a micro view, higher temperature increases the vibrations of atoms (bond stretching and angle bending) within a protein molecule resulting in flexibility and conformational changes (86-88). In a macro view, these micro changes can be detected as changes in chemical reactions such as hydrolysis of ATP (adenosine 5'-triphosphate) (9) and disulfide bond formation (89), protein folding (82,90,91), protein-protein interactions (85,92), protein-ligand binding (9,93), aggregation (94,95), radiation emission or absorption (96-98), and even crystallization (99). Martin Karplus and J. Andrew McCammon (44) propose that the temperature dependence of proteins originates from the velocity of solvent such as water molecules because the fluctuations of the solvent atoms rather than protein atoms mostly respond to temperature change (42), and the surface structure of a protein is sensitive to the velocity of surrounding water molecules (43). To understand why and how micro changes lead to macro changes, one needs to build a bridge between them by molecular dynamics (MD) simulations with classical methods.

Though important, the study of temperature dependent functions of protein by MD simulation has remained stubbornly intractable because it is often restricted for several reasons. First, the available experimental data for simulation are limited. For example, temperature can change many properties of proteins such as polymer formation, aggregation, phase transition, radiation emission and absorption, conformation of structure, molecular motion and even chemical reaction. In classical methods, radiation

emission, light absorption and chemical reaction can not be simulated because the potentials for each atom only represent the ground state. In addition, since MD simulation generally mimics a single protein molecule out of a pool of molecules in a macro system, polymer formation, aggregation and even phase transition are hard to simulate. Second, MD simulations on a protein need a starting structure such as a crystal or a NMR structure or a homologue structure, however, the number of available crystal or NMR structures is very limited. Third, a control system is often not available. Proteins are characterized by their primary, secondary and tertiary structures. The functional changes of protein result from their structural changes originating from their primary structure changes such as a mutation. Without a mutation as a control system; it is hard to focus on what structural factors are affected by temperature and how the structural factors relate to the temperature dependent functional change of protein. Fortunately, these restrictions are conquered by a recent experiment conducted on two ABC transporters CvaB and HlyB by Guo et al 2006 (9).

ABC (ATP binding cassette) transporters are trans-membrane proteins which belong to a big superfamily. Transporters export toxin, protease, antibiotics, ions and so on, and import nutrient such as histidine, maltose, ribose, oligopeptide, vitamin B12, and ions (4). CvaB and HlyB are ABC transporters located on the cytoplasmic membrane of *E. coli* and export ColV and Hly respectively (13,37). Like the other members in this family, they consist of dimerized transmembrane domains (TMDs) and dimerized nucleotide binding domains (NBDs) (9,13). The TMDs form a channel to transport substrates. NBDs bind to two ATPs between the interfaces of the two NBD subunits in each site (12,16,29). The conserved sequence Walker A, Walker B in one subunit and

signature motifs in the other subunit are found in each ATP binding site. ATP interacts with these motifs and facilitates the dimer formation of two subunits (9,29).

According to Guo et al 2006 (9), UV cross-linking was conducted to measure the binding affinity between ATP/GTP and NBD domain of CvaB and its homolog HlyB (13) in vitro. Wild type CvaB(NBD) prefers GTP to ATP at low temperature (273K) but not at high temperature (310K). As a control, HlyB(NBD) does not show the preference at any temperature indicating that the binding affinity between CvaB(NBD) and nucleotide is temperature sensitive. Furthermore, they found that a double mutation S503A/Q504L of CvaB(NBD), located at a loop region Y^{501}DSQ in nucleotide binding site, can eliminate this preference acting like HlyB(NBD) (figure 2-1).

The significance of Guo et al's research to this study is to provide a set of experimental data in which the property changes of proteins can be simulated by MD simulations. Temperature can be taken as a parameter, ATP, GTP and the double mutation can be easily set into models by making a "virtual mutation", UV cross-linking bands can be treated as the binding affinities or external binding energies of ATP/GTP, the double mutation can be taken as a control system for focusing on the structural factors which cause the temperature dependent nucleotide binding preference of CvaB(NBD), and HlyB can be taken as another control system for verifying the qualification of the double mutation as a control system.

This study will run molecular dynamics (MD) simulations on CvaB(NBD), its mutations and HlyB at low (273K) and high (310K) temperature respectively. Two aspects were examined: 1) the consistency of the binding affinity of CvaB(NBD)-ATP/GTP between experimental data and calculation to ensure if our classical methods

are qualified to simulate this system. Then, 2) the investigation of how residues 503 and 504 cause the changes of energetic and structural factors in a micro view and how these changes lead wild type CvaB(NBD) to prefer GTP over ATP at low temperature but not at high temperature in a macro view.

2. Methods

2.1. Modeling of the NBD dimer of CvaB

Among ABC transporters whose crystal structures have been solved, the HlyB(NBD) is most similar to the CvaB(NBD) (38% identity), both NBDs form a head to tail dimer (21,32). Thus, the dimerized NBD in the crystal structure of HylB mutant H662A (1XEF, resolution 1.8 Å) was used as a template for modeling. After Lalign alignment (http://www.ch.embnet.org/software/LALIGN_form.html) (figure 2-2), the coordinates of amino acids, ATP, ATP-binding Mg^{2+}, and water molecules in the template were transferred to the dimerized NBD model of CvaB, and the coordinates of ATP in the template were used to form GTP. Energy minimization was performed using 1000 steps of conjugate gradients optimization with the latest AMMP potential set (version tuna) (100,101) where the water potential was adjusted based on current hydration water experimental data (Chen et al., 2008). The model was viewed with PyMol (*DeLano, W.L. The PyMOL Molecular Graphics System (2002) on World Wide Web http://www.pymol.org*).

2.2. MD simulation of the NBD dimer of CvaB

6000 additional water molecules were added into the dimerized CvaB(NBD) models (wild type or mutants with bound ATP or GTP) and HlyB(NBD) to ensure an

aqueous environment. Additional ions were added to neutralize the total charge of the proteins. No screening dielectric term or bulk solvent correction was included. A constant dielectric of one was used. The amortized fast multipole algorithm in AMMP was used for the long-range terms in the non-bonded and electrostatic potentials so that no-cutoff radius was employed (102). Simulations were performed with a constant NVT ensemble corresponding to a pressure of about 1 atm with classical molecular mechanics MD program AMMP (100,101). The temperature was set at 273K and 310K, respectively. One thousand frames, each for one picosecond, were logged over 1-ns MD run. To make sure that the results of MD simulations were not random events, each MD simulation was repeated eleven times for a total of twelve times (runs) with different random starting velocities.

2.3. Estimation of leave-one-out errors

To determine if the number of repeats was enough to represent the distribution of binding energy, the convergence of each MD simulation was evaluated by calculating Leave-one-out errors from twelve MD runs. For each individual MD run, the average binding energy $\langle E \rangle$ is

$$\langle E \rangle = \frac{1}{2 \times 1000} \sum_{m=1}^{1000} \sum_{n=1}^{2} E_{mn}$$

Where, E stands for the external binding energy of ATP or GTP, m and n stand for the indexes of frames and nucleotide binding sites respectively. Thus, among 12 MD runs, if $\langle E_j \rangle$ is left out, the average binding energy of the rest MD runs is

$$\langle E_{(j)} \rangle = \frac{1}{12-1} \sum_{i=1, i \neq j}^{12} \langle E_i \rangle$$

Finally, the Leave-one-out error is evaluated as

$$STDEV\left\langle E_{(j)}\right\rangle \ (j=1 \text{ to } 12) \quad (1)$$

Where *STDEV* stand for the calculation of standard deviation, *i* and *j* are the indexes of MD runs respectively.

2.4. Possibility of hydrogen bond formation

The possibility of hydrogen bond formation was used to evaluate the local interactions between protein and bound nucleotide. Since there are total 12 runs for each combination of mutant, bound nucleotide, and temperature, 1000 frames for each run and two binding sites in each frame, the possibility of the hydrogen bond formation (POH_{ai}) between a non-hydrogen atom *a* of protein and a non-hydrogen atom *i* of bound nucleotide is defined as the total number of hydrogen bonds between the two atoms from 12 runs, 1000 frames and two binding sites divided by 24000. Thus, the total number of hydrogen bonds (including ionic bond) between atom *i* and residue *r* of a protein is defined as

$$POH_{ri} = \sum_{a=1}^{A} POH_{ai} \begin{Bmatrix} 1 : a \in acceptor, i \in donor \\ 1 : a \in donor, i \in acceptor \\ 0 \end{Bmatrix}$$

The total number of hydrogen bonds between bound nucleotide and a residue *r* is defined as

$$POH_{rn} = \sum_{n=1}^{N} POH_{ri}$$

Where, A and N stand for the total number of non-hydrogen atoms in a residue or nucleotide, respectively. During MD simulation, a big POH value stands for strong interaction(s) at a particular local region.

The above calculations were performed with PDBAnalyzer, a JAVA program developed in-house, and using Microsoft SQL Server 2000 as the database engine. The data for each pair of hydrogen bond interaction such as mutant name, temperature, ATP/GTP name, run ID, frame ID, binding site ID, residue atom ID, ATP/GTP atom ID, bond distance (hydrogen bond 2.5~3.2 Å, ionic bond 3.0~4.0 Å) were input to a database table as one record. Then, SQL scripts were used to extract data for POHs from sixteen million records. The definition of hydrogen donor or acceptor of each protein residue was obtained from *Atlas of Side-Chain and Main-Chain Hydrogen Bonding, by Ian McDonald and Janet M Thornton, WWW Edition December 1994, original edition April 1993 (http://www.biochem.ucl.ac.uk/bsm/atlas/).*

2.5. Binding preferences

In order to describe the effects of temperature on GTP/ATP binding preference, three concepts of preference were used: temperature dependent binding energy, temperature dependent formation of hydrogen bond from MD simulations, and difference of the slope from experimental data (figure 2-1). In the temperature dependent binding energy preference (E_{prefer}),

$$E_{prefer} = (E_{GTP@310}-E_{ATP@310})-(E_{GTP@273}-E_{ATP@273})$$

Where, each E stands for the averaged binding energy over 12 runs, 1000 frames and two binding sites with bound ATP/GTP and at 273/310K, respectively. Since stable

ATP/GTP-protein binding corresponds to a negative value of binding energy, the high value of E_{prefer} corresponds to a GTP binding preference at 273K.

To find which individual interaction between bound nucleotide and CvaB(NBD) mostly contributes to the temperature dependent nucleotide preference of wild type CvaB(NBD), the temperature sensitivity or the temperature dependent formation of hydrogen bond ($POH_{prefer.ai}$) of each individual interaction between an atom *i* in nucleotide and an atom *a* was calculated for wild type,

$$POH_{prefer.ai} = (POH_{ai.wild@310} - POH_{ai.wild@273}) - (POH_{ai.sqal@310} - POH_{ai.sqal@273})$$

Because the double mutant (sqal) can completely eliminate the nucleotide temperature dependence of wild type CvaB, it is an experimental control (9), the temperature sensitivity of an individual interaction derived from the double mutant should be subtracted to normalize that of wild type.

In the experimental data in figure 2-1, the binding preference is represented as

$$Slope_{prefer} = Slope_{GTP} - Slope_{ATP}$$

The bigger value responds to GTP preference at low temperature based on Arrhenius equation.

3. Results

3.1. Overall structures and RMSDs of CvaB(NBD)

The average structures (main chain) of CvaB(NBD) for each mutant, bound nucleotide and temperature are superimposed and shown in figure 2-3a. All of them share the similar structure. The bound nucleotides also share the similar orientation as shown in figure 2-3b. Besides the strong interactions with Walker A and signature motif, common in the ABC family, many of them have strong interactions with E^{628}, G^{629} and K^{537} as

well. As an example of nucleotide-CvaB(NBD) interactions, the residues having interactions with GTP in wild type are shown in figure 2-2.

The RMSDs of each Cα of CvaB(NBD) for each mutant, bound nucleotide and temperature were plotted in figure 2-4. They are similar to each other, but the values at high temperature are bigger than at low temperature with a range of 0.1~0.2 Å. The α-helix and β strands show lower RMSD values than the loop regions indicating they are more stable. The conserved sites in ABC family such as Walker A, B, signature motif, H and Q loops have lower RMSD values because they are located at α-helix and/or β strand or have interactions with bound nucleotide (figure 2-2, 2-3a, 2-3b & 2-4).

3.2. Temperature dependent nucleotide binding preference

Figure 2-5 is a bar graph of the E_{prefer} of CvaB, its mutants and HlyB and shows that in comparison to the double mutant and HlyB, wild type CvaB(NBD) prefers GTP at low temperature (273K). This is consistent with the experimental data of Guo et al 2006 (9). The calculated binding affinities of single mutants Q504L and S503A show a similar, but reduced, GTP preference at low temperature relative to wild type which agrees with the experimental data [*the ATP/GTP binding affinities of CvaB(NBD) of the two single mutants are similar to that of wild type at both low (9) and high temperature (personal communication with Dr. X. Guo)*].

The significance of the results in figure 2-5 is valuated by the estimation of leave-one-out error (Table 1). The maximum value in table 2-1 is 4.88 kcal/Mol which means that by adding one more MD run, the ATP/GTP binding energy is only expected to change by 4.88 kcal/Mol. 4.88 kcal/Mol is a small amount compared with the 65

kcal/Mol difference of the binding preference between wild type and double mutant, therefore the number of MD runs is enough, and the results in figure 2-5 are not random events.

3.3. Temperature dependent changes in individual interactions between nucleotide and CvaB

There is no doubt that the nucleotide binding preference of wild type CvaB(NBD) results from the differences between ATP and GTP and the differences between wild type and double mutant. In GTP, a carbonyl oxygen (O6) replaces a amino group (N6) in the sixth carbon of the purine resulting in a change from a hydrogen donor to an acceptor, and an amino group (N2) located at the 2nd carbon resulting in an additional hydrogen donor. The three atoms cause ATP/GTP to interact differently with protein. In the double mutant, S503A/Q504L changes two hydrophilic residues to hydrophobic ones. The two differences coordinate and lead to the nucleotide binding preference. However, the double mutation site has no strong interactions with any of the three atoms (POH_{ri}<0.05), or even the entire bound nucleotide (POH_{rn}<0.05) under any temperature, mutant or bound nucleotide. This indicates that there is an indirect coordination between double mutation site and the three non hydrogen atoms that differ between ATP and GTP.

The temperature sensitivity of each individual interaction between nucleotide and CvaB is shown in figure 2-6. Four individual interactions show significant temperature sensitivities ($POH_{prefer.ai}$>0.1). Y^{501}:OH-ATP:N6 and G^{629}:O-ATP:N6 show a preference at high temperature, D^{502}:OD1/2-GTP:N2 and E^{628}:OD1/2-ATP:N6 show a preference at low temperature. 501-502 (belonging to Y^{501}DSQ) and 628-629 (next to the N-terminus of signature motif) are located at two loop regions (figure 2-7). The interactions between

the loop 628-629 and ATP:N6 have no temperature sensitivity because the interactions of G^{629}:O-ATP:N6 and E^{628}:OD1/2-ATP:N6 cancel out. On the other hand, the interactions of Y^{501}DSQ show a preference for ATP at high temperature and GTP at low temperature. The interactions between Y^{501}DSQ and bound nucleotide have temperature sensitivity. Especially, D^{502}:OD1/2-GTP:N2 has strong low temperature preference ($POH_{prefer.ai}$= -0.32) which may dominate the temperature dependent nucleotide preference of CvaB(NBD).

3.4. Relationship between residues 503&504 and 502

Figure 2-7 shows the RMSD plots of the Cαs and the averaged side chains of residues 500~506 with GTP bound for different mutants and temperatures. The four panels show that there is an RMSD peak between 501~505 with 503 and 504 at the top. This is understandable because 503 and 504 form a turn structure located on a long loop region and the loop sticks out of the protein enlarging the surface area. In addition, 503&504 have very weak interactions with ATP/GTP or any other protein atoms (data not shown). Both RMSDs of the Cαs and the averaged side chains of 502, 503 and 504 in wild type are significantly bigger at 310K than 273K, while those from double mutant are the same. This indicates the 502, 503 and 504 in wild type are more stable at 273K than at 310K and the stability of residue 502 is consistent with the $POH_{prefer.GTP:N2\text{-}502:OD}$ in wild type in figures 2-6 & 2-9.

The correlation coefficient between the RMSDs of the Cαs of 502 and 503&504 is 0.82, and the correlation coefficient between the averaged side chain of 502, and 503&504 is 0.68 (figure 2-7), which indicates the strong relationship between the side chains of 502 and 503&504 is mediated by their neighboring main chain.

4. Discussion

4.1. ATP/GTP binding site

The bound nucleotides share the same orientation (figure 2-3b) and similar interactions with CvaB(NBD) even under different temperatures, mutants and nucleotides. The detailed interactions between bound GTP and CvaB(NBD) are shown in figure 2-10 as an example. Common in all ABC members, the strong interactions were observed between nucleotides and Walker A, signature motif and H-loop (POH_{rn}>0.1), and the Mg^{2+} forms ionic interactions between β- and γ-phosphates. Similar to HlyB (NBD) (PDB entry 1XEF), the purine base forms π-π interaction with Y^{501} in the Y^{501}DSQ loop, however, the interactions between nucleotides and Walker B, and Q-loop were not observed (POH_{rn}<0.01). More specifically, the phosphates have strong interactions with Walker A (S^{528}GAGKTT), K^{537}, partial signature motif (S^{631}GG), and H-loop (H^{684}), the sugar has strong interactions with partial signature motif (Q^{634}) and partial Walker A (G^{529} and T^{534}) and the purine base has strong interactions with K^{537} and partial signature motif (E^{628}GL). The significant difference between ATP and GTP is that the additional amino group at the 2^{nd} carbon of purine of GTP (GTP:N2) forms strong interactions with D^{502} in Y^{501}DSQ loop.

4.2. Explanation of temperature effect on ATP/GTP binding affinity

There are no strong direct interactions found between the 503&504 and ATP/GTP (POH_{rn}<0.05), however, they may interact through a second order indirect interaction. Here is the evidence: 1) there is a strong interaction between GTP:N2 and D^{502}, 2) there

is strong positive relationship between the RMSDs of the side chains and Cαs of the D^{502} and 503&504, 3) 503&504 are located at a loop region on protein surface with rare interactions with ATP/GTP or any other atoms of protein (figure 2-9), 4) high temperature is associated with a weaker D^{502}-GTP:N2 interaction in wild type, 5) high temperature is associated with a high RMSD value of the side chains of both D^{502} and 503&504. 6) the hydrophilic residues 503&504 (in wild type) show higher RMSD values at any temperature with both bound GTP (figure 2-7) or ATP (data not shown) than hydrophobic one (double mutant). Thus, when temperature increases, the velocities of water molecules increase, which increases the movement and vibration of the hydrophilic side chains of 503&504 on the wild type protein surface because of their interactions with water molecules. The flexible side chains of 503&504 cause flexibility in the side chain of D^{502} through their neighboring main chain located in the turn structure. On the other hand, when the S503 and Q504 are replaced in the mutant by alanine and leucine, which have weaker interaction with water molecules, the stability of the turn structure will be less affected by the increased velocity of water molecules at 310K. As a result, the distance between D^{502}:OD1/OD2 and GTP:N2 at 273K is closer than at 310K in wild type (figure 2-9), but the same in the double mutant (figure not shown).

The interactions between the turn structure and water molecules are also supported by previous studies (42,43). Chen et al found the β-switch region, a loop region on Glycoprotein Ib surface, can be changed to a hairpin structure by interaction with flowing water (43). In our simulations, formation and dissociation of the hydrogen bond between D^{502}:O and 505S:N were observed during MD simulations (data not shown). Vitkup et al (42) found that the fluctuations of the solvent atoms rather than protein

atoms mostly respond to temperature change, which was also observed in our MD simulations. Therefore, "the temperature of the solvent and thus its mobility is the dominant factor in determining the functionally important protein fluctuations" as reviewed by Martin Karplus and J. Andrew McCammon (44), consequently, the less structured loop region may be more easily affected by solvent molecules than a helix or β-sheet.

4.3. MD sampling

In protein dynamics, protein structures change (movement, vibration as well as conformational change) all the time. Proteins are big molecules with thousands of atoms and the distributions of protein structure or the free energy surface in dynamics occupy a multiple dimensional broad area. However, the area is always limited and constant under a certain temperature. Each individual MD run, with different initial velocity for instance, may cause the structural changes roaming from different directions which results in different paths in the broad area (103). Thus, the structural distributions from any two MD runs may not be consistent with each other as actually occurred in our samples. With the increase of the number of MD runs, a larger distribution or free energy surface area may be explored, thus the averaged structure distribution should show a convergence which evaluates if MD runs are enough to represent the TRUE structure distribution as shown in table 2-1.

4.4. Additional consistency of our MD simulations with experimental data

Besides the consistency of temperature binding preference shown in figure 2-5 and table 2-1, more evidence is found to agree with our MD simulations. Guo et al 2006

(9) made a series of mutations and measured the ATP/GTP binding preference, but none of them showed a significant change with wild type (figure 2-2, red letter). In MD simulations, the important sites for ATP/GTP binding preference are shown in figure 2-2 (green highlighted), and none of them overlap with the mutations lacking ATP/GTP preference (Guo et al 2006). In addition, the explanation that the increased binding affinity between D^{502} and GTP:N2 at 273K causes wild type to bind GTP tighter at 273K than at 310K can also explain why HlyB(NBD) does not have temperature dependent ATP/GTP binding preference. In HlyB(NBD), there is a lysine corresponding to D^{502} in CvaB and the main chain oxygen of the lysine did not form hydrogen bond with GTP:N2 either (POH_{rn}= 0.00).

5. Summary

The consistency of the temperature dependent binding affinity between experimental data and MD simulation indicates our classical method MD simulation is able to simulate temperature dependent functional change of protein. The effects of temperature on protein function occur everywhere, this study provides an exciting computational method to analyze this problem. The MD simulations also predicted a mechanism for how temperature affects protein functions: starting from the change of the velocity of water molecules causing the change of the vibration and movement of the side chain of a residue on the protein surface, which is transmitted through the main chain of neighboring residues in a loop region to alter the stability of the side chain of the target residue.

FIGURES & TABLES

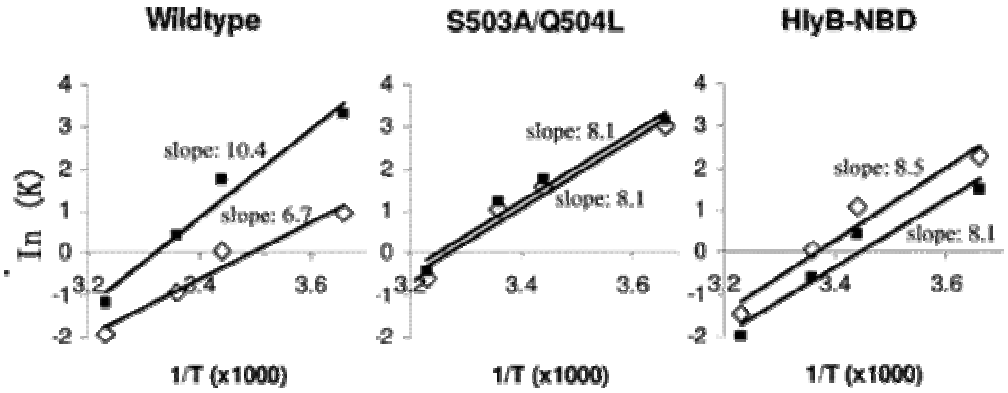


Figure 2-1 Temperature sensitivity of the binding affinity.

Temperature sensitivity of the binding affinity for ATP (◆) and GTP (■) of wild type, double mutation and HlyB(NBD). X-axis and Y-axis stand for temperature and binding affinity respectively. The relationship between binding affinity and temperature is evaluated by Arrhenius equation. The difference between the slopes for ATP and GTP of each protein represents the temperature dependent nucleotide binding preference. (Figure is taken from Guo et al 2006(9)).

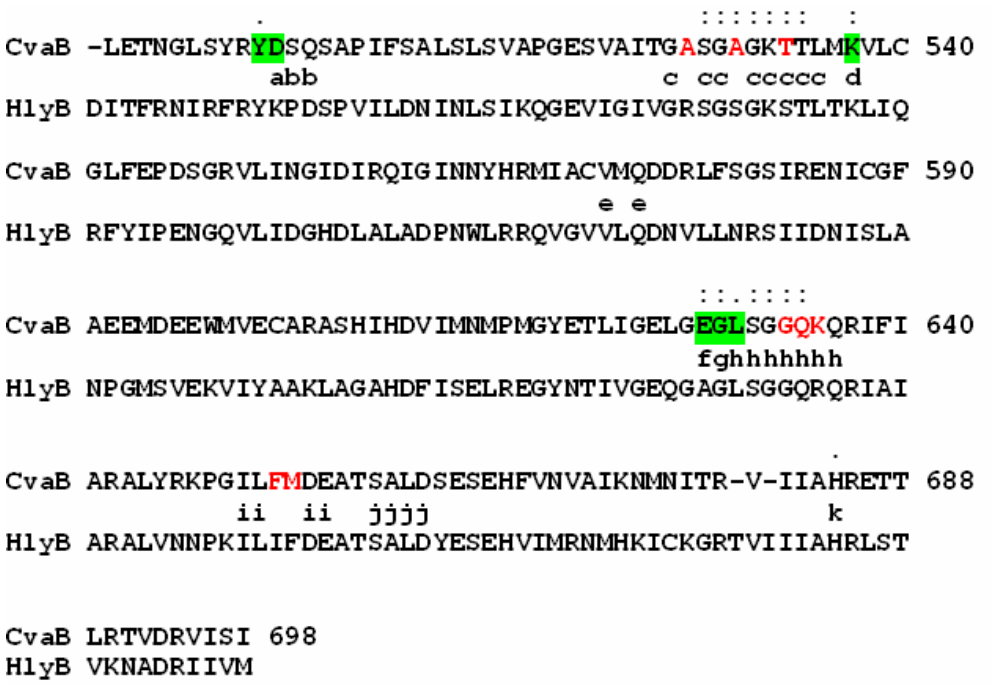


Figure 2-2 The pairwise alignment between CvaB(NBD) and HlyB(NBD).

a: D502, b: double mutation site, c: Walker A, d: K537, e: Q-loop, f: E628, g: G629, h: signature motif, i: Walker B, j: D-loop, k: H-loop.

The residues highlighted in green correspond to direct contacts in the modeled structures between ATP/GTP and protein. The red marked residues have been mutated by Guo et al (9) and showed no effects on the nucleotide binding preference.

'.' stands for the possible conservation of the interactions between nucleotide and a protein residue with POH_{rn}>0.1, ':' stands for that with POH_{rn}>0.5. Data is averaged from 12 runs x 1000 frames x 2 binding sites x 2 temperatures (273K and 310K) of wild type with bound GTP.

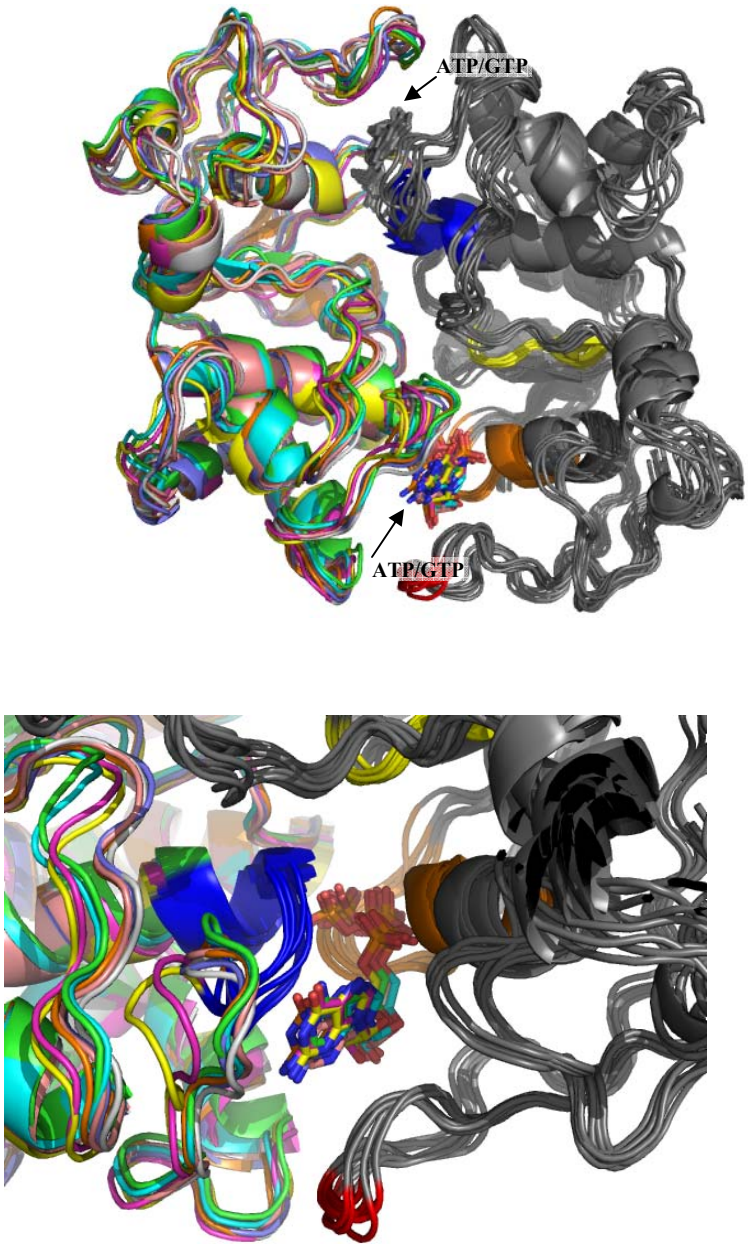


Figure 2-3 Superposition of CvaB(NBD) structures for each mutation, nucleotide and temperature.

Each structure is averaged from 12 runs x 1000 frames.

Figure 3a. Dimerized CvaB(NBD). Two subunits are shown as multi-color and grey. Walker A (orange), Walker B (yellow), signature motif (blue) and double mutation site (red) are shown in the grey subunit. Figure 3b. Superposition of CvaB(NBD) structures for each mutant, nucleotide and temperature. Walker A (orange), Walker B (yellow), signature motif (blue) and double mutation site (red) are shown.

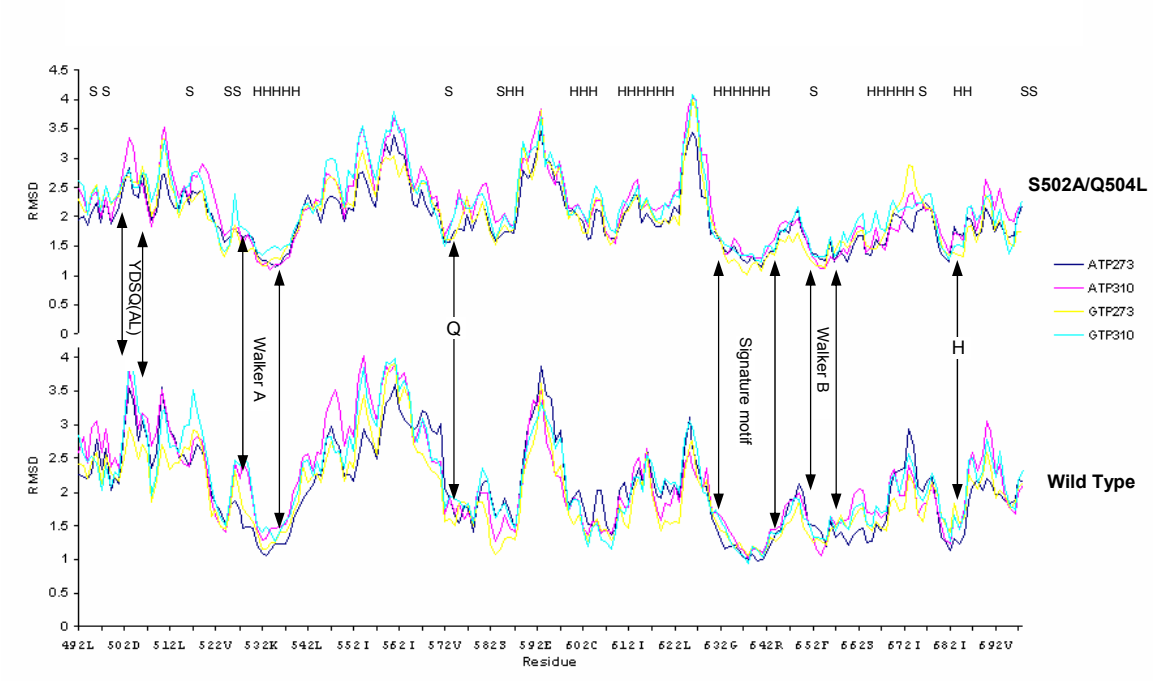


Figure 2-4 The RMSD plots of all Cαs over 12 runs x 1000 frames x 2 subunits. H: α-helix, S: β-strand.

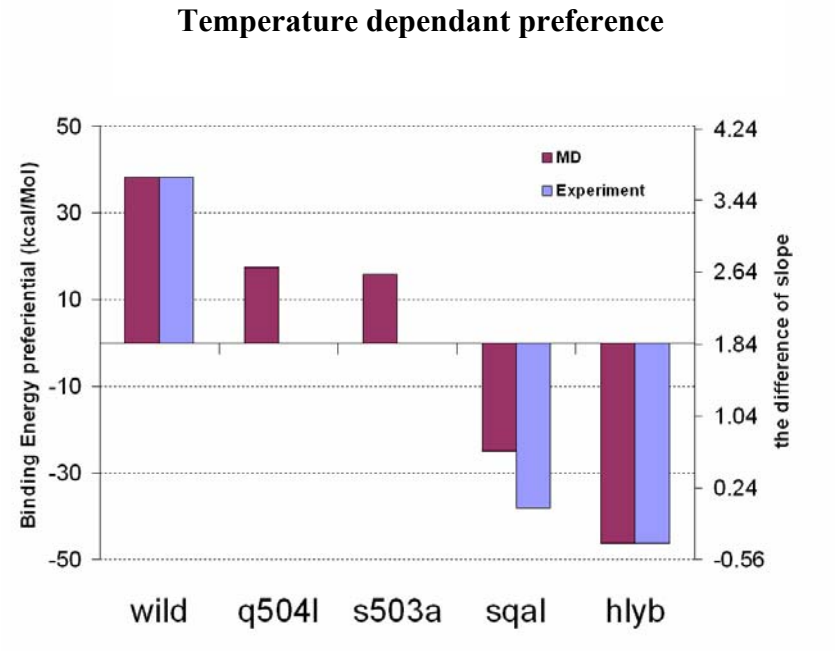


Figure 2-5 The differences of binding energies with GTP over ATP and 273K over 310K.

Data are averaged from 12 runs, 1000 frames and two binding sites. Higher value of temperature dependent preference (E_{prefer}) represents a GTP preference at low temperature.

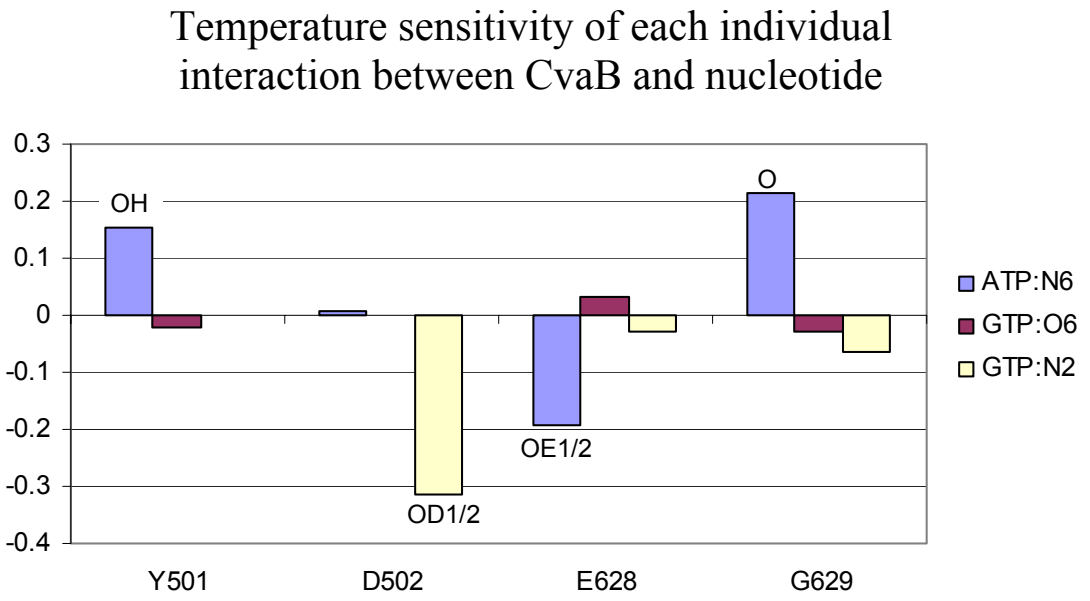


Figure 2-6 The temperature sensitivity of each individual interaction between CvaB and nucleotide.

The temperature sensitivity is normalized by $POH_{prefer.ai} = (POH_{ai.wild@310}-POH_{ai.wild@273})- (POH_{ai.sqal@310}-POH_{ai.sqal@273})$

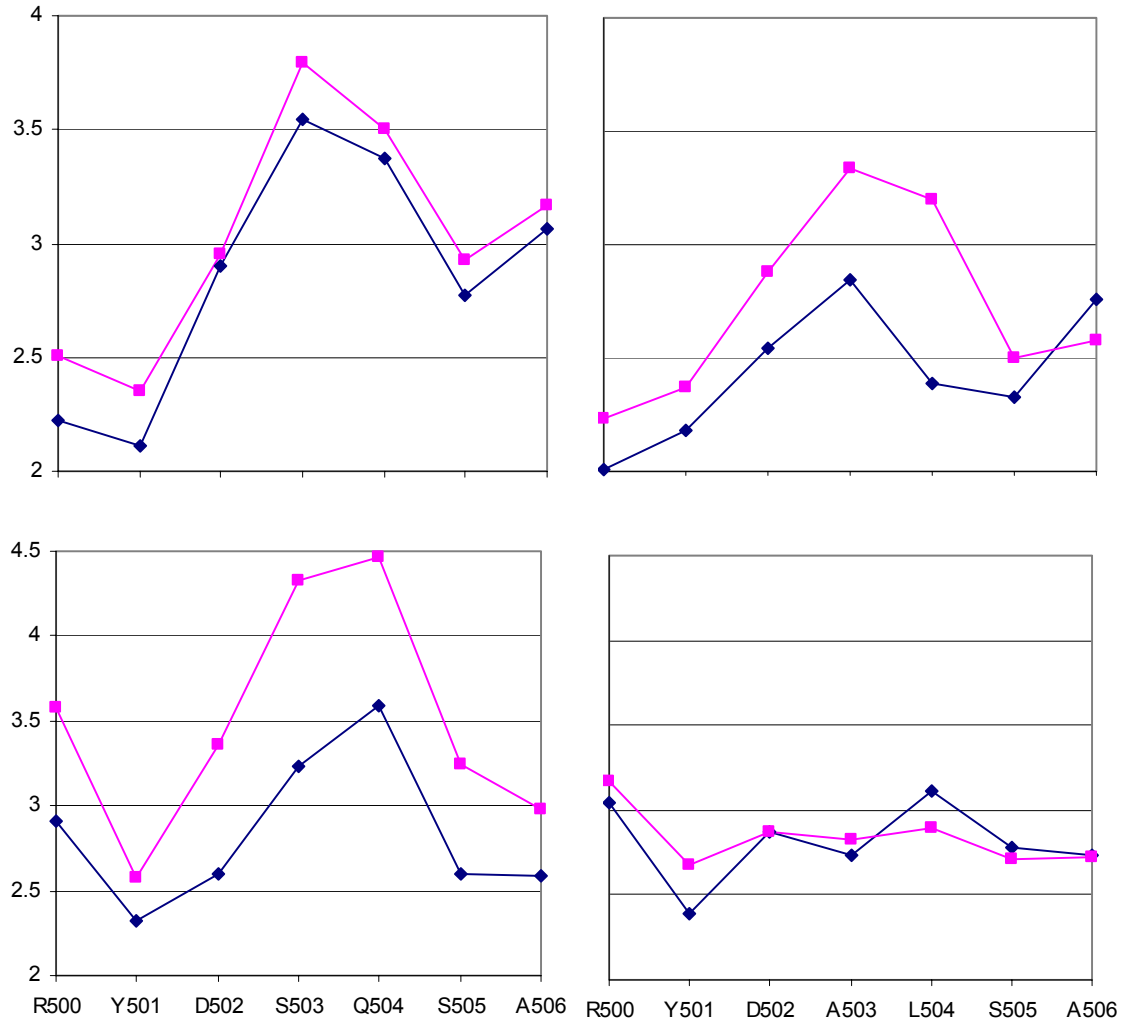


Figure 2-7 The RMSD plots of Cαs and averaged side chains.

The RMSD plots of Cα (upper) and averaged side chains (lower) of 500~506 with bound GTP for different mutation and temperature. (Data is calculated from twelve runs, 1000 frames for each run, and two binding sites for each frame).

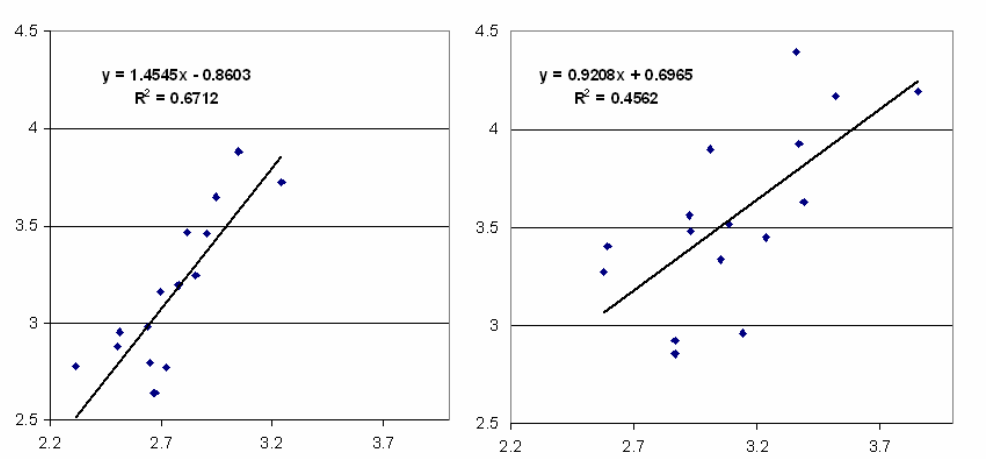


Figure 2-8 The relationship of the RMSDs of Cα (left) and averaged side chain (right) between 502 and the averaged 503&504.

The RMSD plots of Cα and averaged side chains. There are 16 data points in each panel for different mutant, ATP/GTP binding and temperature. (Data are calculated from twelve runs, 1000 frames for each run, and two binding sites for each frame).

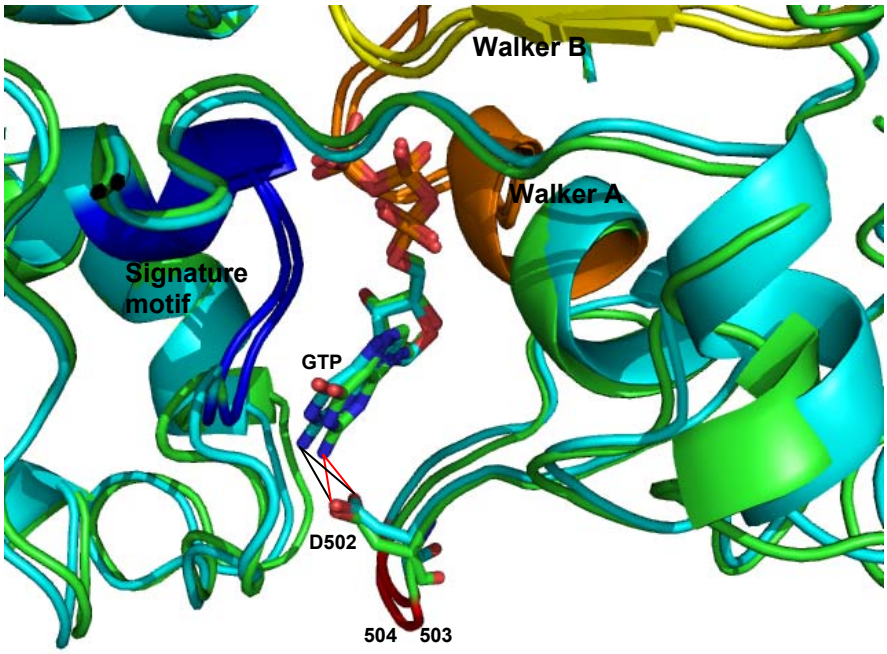


Figure 2-9 The GTP binding site of wild type.

The GTP binding site of wild type at 273K (green) and 310K (cyan). The side chain of D502, Walker A (blue), Signature motif (yellow), 503&504 (red) are shown. The distance between D502:OD1/OD2 and GTP:N2 at 273K (2.40/3.09Å) is shorter than at 310K (2.93/3.41Å). The structures are averaged from 12 runs, 1000 frames for each run, at 273K and 310K respectively. The picture was made by PyMol (www.pymol.org).

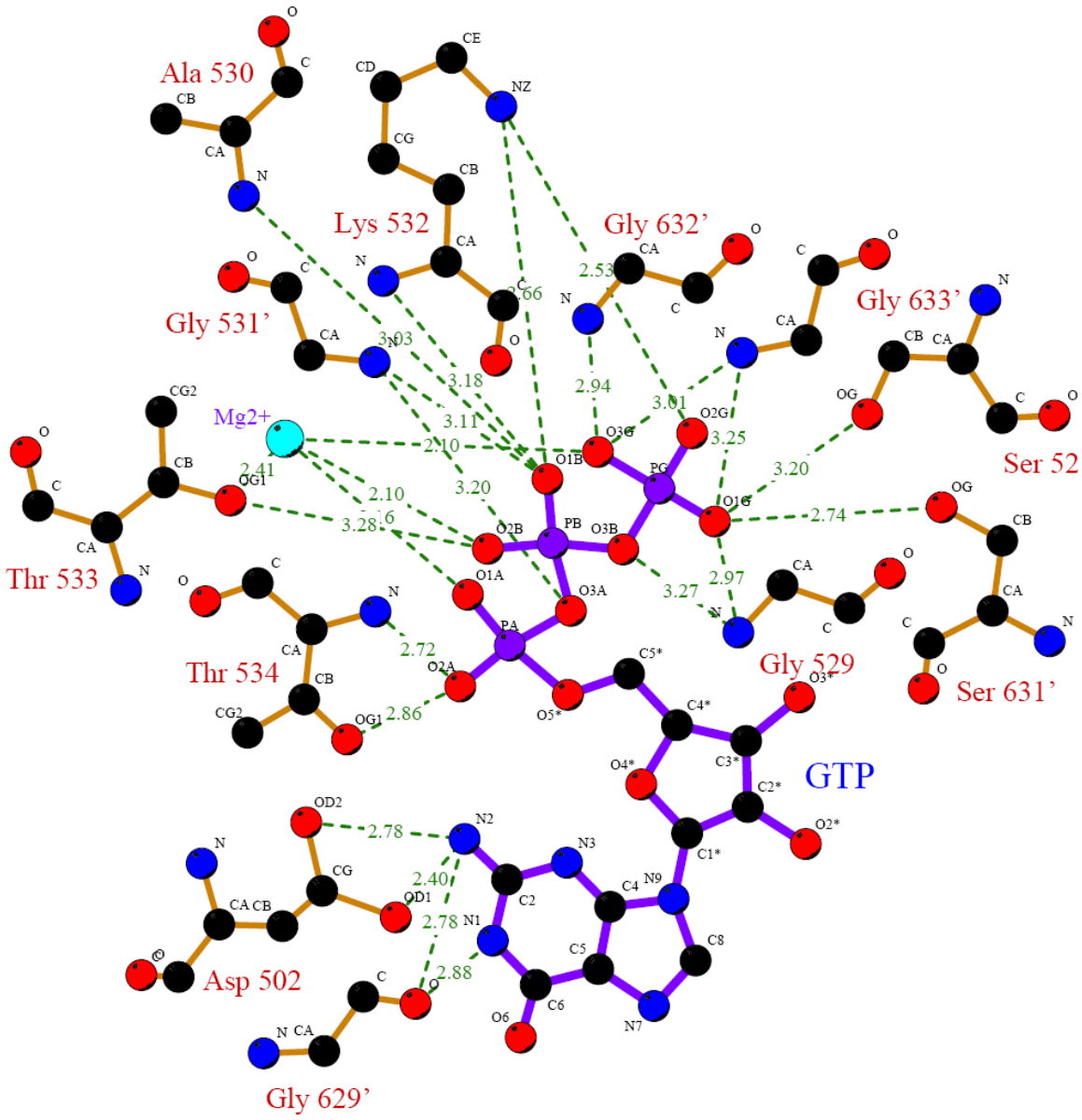


Figure 2-10 The nucleotide binding site of wild type CvaB with bound GTP at 273K.
Detailed individual hydrogen bond and ionic interactions with distances (Å) are shown. Structure was averaged from 12 runs x 1000 frames x 2 binding sites. The apostrophes after residue numbers indicate another subunit. Figure was made by Ligplot (104).

Table 2-1

	atp-273	atp-310	gtp-273	gtp-310	maximum
wild	1.88	3.34	4.26	3.53	4.26
q504l	3.63	2.58	3.10	2.20	3.63
s503a	2.83	2.13	2.92	3.73	3.73
sqal	2.59	3.27	4.88	3.99	4.88
hlyb	2.41	2.40	4.61	2.79	4.61

The Leave-one-out errors (kcal/Mol) out of twelve MD runs. (Data are calculated from equation 1).

3. PKA Phosphorylation Produces Interdomain Movement in SUR2B Leading to Activation of the Vascular KATP Channel

PKA Phosphorylation Produces Interdomain Movement in SUR2B Leading to Activation of the Vascular K_{ATP} Channel

Yun Shi*[1], Xianfeng Chen*[1, 2], Zhongying Wu[1], Weiwei Shi[1], Yang Yang[1], Ningren Cui[1],

Chun Jiang†[1] and Robert W. Harrison†[2]

[1] Department of Biology and [2] Department of Computer Science, Georgia State University, 33 Gilmer Street, Atlanta, Georgia 30302–4010, USA

Running head: ***Interdomain movement of SUR2B by phosphorylation***

Key words: K^+ channel, antagonist, second messenger, protein kinase A, vascular tones, phosphorylation

* These authors contributed equally to this work.
† Correspondence to:
Chun Jiang: Tel, 1-404-413-5404; Fax, 1-404-413-5301; Email, cjiang@gsu.edu

Robert W. Harrison: Tel, 1-404-413-5724; Fax, 1-404-413-5717; Email, rharrison@cs.gsu.edu

Abstract

Vascular ATP-sensitive K^+ channels are activated by multiple vasodilating hormones and neurotransmitters via PKA. A critical PKA phosphorylation site (Ser1387) is found in the second nucleotide-binding domain (NBD2) of SUR2B subunit. To understand how phosphorylation at Ser1387 leads to changes in channel activity, we modeled the SUR2B using a newly crystallized ABC protein SAV1866. The model showed that Ser1387 was located on the interface of NBD2 with TMD1 and physically interacted with Tyr506 in TMD1. A positively charged residue (Arg1462) in NBD2 was revealed in the close vicinity of Ser1387. Mutation of either these three residues abolished the PKA-dependent channel activation. Molecular dynamics simulations suggested that Ser1387, Tyr506 and Arg1462 formed a compact triad upon Ser1387 phosphorylation, leading to reshaping of the NBD2 interface and movements of NBD2 and TMD1. Restriction of the interdomain movements by engineering a disulfide bond between TMD1 and NBD2 prevented channel activation in a redox-dependent manner. Thus, a channel-gating mechanism is suggested via enhancing the NBD-TMD coupling efficiency following Ser1387 phosphorylation, which is shared by multiple vasodilators.

1. Introduction

ATP-sensitive K^+ (K_{ATP}) channels play an important role in vascular tone regulation (105-107). The K_{ATP} channels are expressed in vascular smooth muscles (VSMs) and activated by several vasodilating hormones and neurotransmitters through the G_s-cAMP-PKA signaling system (105,108,109). The channel activation leads to hyperpolarization of VSM cells, decrease in voltage-dependent Ca^{2+} channel activity and relaxation of resistant arteries (110).

K_{ATP} channels consist of 4 pore-forming Kir6x and 4 regulatory SURx subunits (111). The Kir6.1/SUR2B channel is the major isoform of K_{ATP} channels in VSMs (46,112-114). The channel does not open spontaneously at rest. Several groups of channel openers activate the channel, including pharmacological K_{ATP} channel openers (KCOs, e.g., pinacidil and nicorandil) (115), metabolites (MgADP, acidosis) (116,117), and hormonal vasodilators and neurotransmitters (calcitonin gene-related peptide, epoxyeicosatrienoic acids, β-adrenergic receptor agonists and vasoactive intestinal peptide) (58,59,109,118). KCOs and Mg^{2+}-nucleotides activate the K_{ATP} channels via binding to the SUR subunits (61,115). The hormonal vasodilators activate the vascular K_{ATP} channel through direct phosphorylation of the channel protein by PKA (108,109). Our resent study has shown that Ser1387 is a key phosphorylation site (109).

It is unclear how phosphorylation at the Ser1387 residue in SUR2B leads to channel activation. SURs belong to the ABC transporter protein family (119). All ABC proteins have an essential domain assembly, i.e. two transmembrane domains (TMD1 and TMD2) and two intracellular nucleotide-binding domains (NBD1 and NBD2). In addition, SURs have another transmembrane domain containing 5 helical segments termed TMD0

(figure. 3-1A). The cytosolic NBDs of several ABC proteins have been crystallized (15,18,21,120). They all show similar sandwich-like structures with nucleotides clamped between two NBDs. The NBD motifs, such as Walker A (116,121), Walker B (122-124), and signature sequences (125) are known to be important in nucleotide and/or KCO activation of K_{ATP} channels. How the channel is activated with nucleotide/KCO-binding, however, is still not fully understood. Several lines of evidence suggest that TMDs play a major role in interacting with Kir6 subunits (62). Such interaction has been observed in recombinant Kir6.2/SUR1 channels using electronic microscopy (126). Moreover, the TMD0 and ICL0 are involved in K_{ATP} gating by channel modulators and nucleotides (127,128). These studies suggest that NBDs couple with TMDs during SUR-mediated channel gating.

Several ABC proteins have been crystallized in full length with TMDs (12,129-131). These include the recently crystallized SAV1866 that is a homologue of mammalian ABC proteins (12) and shows the highest sequence similarity to SUR2B. Using the SAV1866 crystal structure, we modeled the core of SUR2B (SUR2B_core) containing TMD1, TMD2, NBD1 and NBD2 with particular attention to conformational changes after Ser1387 phosphorylation. Our combined studies of modeling, molecular dynamic simulations, functional assays and mutational analysis showed that Ser1387 was located on the interface between TMDs and NBDs where it formed a triad with Tyr506 and Arg1462 after phosphorylation. The triad appeared to strengthen the interaction of NBDs with TMDs, and enhance the necessary force coupling between the two protein domains for channel gating. The model also supports the accuracy of the SAV1866

crystal structure on the NBD-TMD interface and indicates that the SUR2B can be well represented by the SAV1866 structure.

2. Results

2.1 Modeling of the SUR2B_core

The dimerized SAV1866 was crystallized in the ADP-binding form (PDB: 2hyd). Based on this structure, we built a model of the SUR2B_core, in which sequence identities to SAV1866 are 21.5% for TM-NBD1 of SUR2B (46.5% similarities), and 21.3% for TM-NBD2 (47.8% similarities) (Online figure 3-1; figure 3-1A, B). NBD1 and NBD2 formed a tight hetero-dimer (figure 3-1C). There were two nucleotide-binding pockets on the interface between NBD1 and NBD2. A MgATP molecule was incorporated in the first nucleotide binding pocket and a MgADP in the second pocket. Different nucleotides were used because the first pocket is likely to be a high-affinity ATP-binding site, and the second is likely to bind to MgADP (61,132). In the SUR2B_core model, TMDs interact with NBDs via four short helices of intracellular linkers (ICLs, figure 3-1C). ICL1 and ICL3 interacted with the NBD surface across the nucleotide-binding regions. ICL2 from TMD1 inserted deeply into NBD2, while ICL4 from TMD2 inserted into NBD1 (figure 3-1C). NBD2 formed a spindle-like groove where a β sheet (β5) formed the bottom while Q-loop, C-loop and several α-helices (α2, α3 and α6) lined the wall. Ser1387 was located on β5 within the reach of ICL2 (<4Å, figure 3-1D). The side-chain of an aromatic residue (Tyr506) in ICL2 was inserted deeply into the NBD2 groove and was close to Ser1387 in ~4Å (figure 3-2A). Around Ser1387

there was a positively charged residue (Arg1462) in NBD2 at a distance that allows electrostatic interaction between them when Ser1387 is phosphorylated (figure 3-2A).

2.2 Activation of the Kir6.1/SUR2B channel by protein kinase A

In our previous studies (109,118), we found that β-adrenergic receptor agonists and vasoactive intestinal polypeptide activated vascular K_{ATP} channels in rat mesenteric artery, a response that could be reproduced in recombinant Kir6.1/SUR2B channels expressed in HEK293 cells. We thereby chose to use the recombinant Kir6.1/SUR2B in our current study because the expression system allowed us to manipulate the channel protein. Also because PKA is a common signal molecule by which multiple hormonal vasodilators activate the channel, we used forskolin (10μM) to investigate the PKA effects. Kir6.1 and SUR2B were expressed in HEK293 cells. Forskolin was applied to the cell when the baseline currents were stabilized in 4-6 min. The exposure to forskolin activated K^+ currents that were sensitive to both pinacidil and glibenclamide (figure 3-2B; Online figure 2). After currents were normalized between maximum channel inhibition by 10μM glibenclamide and maximum activation by 10μM pinacidil, the baseline currents averaged 6.0±1.9% (n=10) of the maximum channel activity. Forskolin (10μM) increased the currents to 38.5±3.0% (n=10, figure 3-2E)

2.3 Elimination of PKA activation by mutation at Arg1462 or Tyr506

In the SUR2B_core model, a positively charged residue Arg1462 was located in the close vicinity of Ser1387. The distance of the charge of Arg1462 (measured at the center of two N_h atoms) to Ser1387 side chain (O_g) was ~8Å (figure 3-2A). Such a distance may allow an electronic attraction when Ser1387 is phosphorylated. Mutation of Arg1462 to

alanine indeed eliminated the forskolin-induced channel activation (figure 3-2C). Based on our model prediction, a positive charge at this site is important. This prediction was verified by mutating the Arg1462 to lysine, and the R1462K mutant remained strongly activated by forskolin (68.5±5.8%, n=5). The Arg1462 is located in a consensus PKA site (**R**KSS), although neither Ser1464 nor Ser1465 is a functional PKA phosphorylation site (109). Thus, a positively charged residue critical for the PKA-dependent channel activation is identified in the SUR2B_core model, although it is far from Ser1387 in the primary sequence.

An aromatic residue, Tyr506, was found on the ICL2 segment, side chain of which dipped deeply into NBD2 and physically interacted with Ser1387. When the Tyr506 was mutated to alanine, the forskolin-induced channel activation was almost completely abolished (figure 3-2D). A phenylalanine residue is found at the corresponding site in SAV1866 and ICL4 of SUR2B (Online figure 1). Phenylalanine has a similar side chain as tyrosine but lacks a hydroxyl group. Interestingly, when Tyr506 was mutated to phenylalanine, the channel failed to be activated by forskolin (figure 3-2E), indicating that the hydroxyl group of Tyr506 residue is required for channel activation.

2.4 Formation of the Ser1387-Arg1462-Tyr506 triad following PKA phosphorylation

Our modeling suggests that PKA phosphorylation at Ser1387 may induce local conformational changes. In order to understand these changes, we used the molecular dynamics (MD) simulations to reveal changes in residues and the peptide backbone. Our SUR2B_core contained many hydrophobic transmembrane segments that would necessitate time-consuming simulations in the lipid bilayer environment. We therefore

constructed two additional models consisting of NBDs and ICL2 helix with and without Ser1387 phosphorylation. MD simulations were undertaken at 310K and 1 atm. After 1000 runs, ICL2 still stayed in the NBD2 groove, suggesting a strong interaction between ICL2 and NBD2. The side chain distances among the three critical residues were monitored (figure 3-3A, B). Without phosphorylation, the average distance between Tyr506 and Ser1387 was 3.8Å, a distance that allows hydrogen bond formation between these two residues. The Tyr506-Arg1462 and Ser1387-Arg1462 distances were around 6-7Å (figure 3-3A, C), suggesting that there is no direct contact between Arg1462 and the other two residues before Ser1387 phosphorylation. When the Ser387 was phosphorylated, distances of all three were less than 4Å (figure 3-3C), indicating formation of a compact triad (figure 3-3D, E). The hydroxyl group of Tyr506 appeared to be critical for stabilization of the triad, which was able to form hydrogen bonds with oxygen atoms of the phosphate group of p-Ser1387 and N_h atoms in Arg1462, consistent with our observation that forskolin failed to activate the Y506F mutant (figure 3-2E).

2.5 The movement of ICL2

Following phosphorylation, the Tyr506 C_α atoms moved by 9.2Å in the *mode* structures. The movement led to a horizontal shift of ICL2 from a position close to the center of NBD2 toward the edge of NBD2 along the NBD2 groove. Since the MD data were acquired after removing most TMDs, the restriction of ICL2 by TMDs was not evaluated. Therefore, this ~10Å movement might be overestimated. However, the direction appeared reasonable. The Arg1462 was located at the end of α6 helix, close to the edge of NBD2, while Ser1387 was located on the bottom of the NBD2 groove. With the triad formation, ICL2 was pulled toward NBD2 with lateral sliding along the groove.

Meanwhile, the α2 helix in NBD2 made a lateral move and left space for ICL2 (figure 3-4C).

2.6 Reshaping of the NBD2 groove by phosphorylation

Studying the conformations of NBD2 of pre- vs. post-phosphorylation, we found that NBD2 underwent major conformational changes on the interface with ICL2. The groove for ICL2 binding ICL2consisted of residues from β5, α2, α3, α6, Q-loop and C-loop. The Arg1462 was the last residue of the long α6 helix, located close to the edge of NBD2. The α6 helix was expected to be stable because it was buried in the structure. Indeed, C_αs in the corresponding segment in NBD1 did not significantly change their positions (1.8±0.2Å, figure 3-5D) in comparison to overall movements of C_αs in NBDs (3.4±0.1Å). However, the phospho-Ser1387 (p-Ser1387) attracted the positively charged Arg1462, moving the residue by 4.5Å toward ICL2 (measured at C_α, figure 3-5C, D). This movement had major effects on the C-terminal half (tail) of α6 (residues 1455-1462) but not the N-terminal half (head) (residues 1448-1454). The average movement of C_αs in the tail was 4.2±0.3Å, much greater than the head (2.2±0.3Å). Consequently, the α6 helix was slightly bent around Phe1454 and turned toward ICL2 with an angle ~10^o. The movement of the α6 helix appeared to cause the α3 helix to move toward ICL2 about 4.2±0.7Å. Like the head of α6 helix, the Q-loop was quite stable (2.4±0.3Å). The C_αs in β5 sheet moved toward ICL2 with distance of 3.0±0.4Å. The β5 movement was greater than the overall movements of β-sheets in NBD2 (2.3±0.2Å, P=0.11) and the corresponding segment in NBD1 (2.4±0.2Å, P=0.07). The movement of these segments toward ICL2 suggested a stronger interaction between ICL2 and NBD2 following phosphorylation. On the model we also observed a large movement (6.9±0.3Å) of the α2

helix. The α2 helix appeared flexible in the SUR2B_core model; it moved toward the edge of NBD2 and yielded space for ICL2.

To make sure that these movements were not random events, the MD simulations were repeated twice for a total of three times with different starting velocities. We observed the same triad formation and protein domain movements in every runs. Therefore, even though there may be technical limitations in the MD simulations, we were able to conclude that ICL2 interacted with NBD2 rather loosely before phosphorylation, and the interaction of the two domains was strengthened with the Tyr506-pSer1387-Arg1462 triad formation after phosphorylation.

2.7 Restriction of the interdomain movement by introduced disulfide bond

Mutations were engineered in the channel to test the hypothesis that interdomain motion plays a critical role in PKA-dependent channel activation. Forming a disulfide bond between the two domains can restrict the motion thus limiting the channel activation if the motion is critical. Examining all residues in the NBD2 groove and ICL2, we found that a cysteine (Cys1408) on the wall of the NBD2 groove was close to Ala507 in ICL2 (figure 3-6A). The distance between the C_αs of the two residues was ~8Å. When the Ala507 was mutated to cysteine, the channel activation by forskolin was completely eliminated (figure 3-6B). Since the Ala507 is located next to Tyr506, the mutation itself instead of the disulfide bond formation could have affected the channel activation. To address this possibility, we included 3mM reduced glutathione in the pipette solution and found that the Kir6.1/SUR2B_A507C channel activation by forskolin was nicely restored (29.7±3.0%, N=5, figure 3-6C). To show whether the introduced Cys507 indeed formed a disulfide bond with Cys1408 but not with another unidentified cysteine residue, we

mutated the Cys1408 to serine, a residue close to cysteine in side chain properties but is unable to form disulfide bonds. The mutant channel (Kir6.1/SUR2B_A507C_C1408S) was activated by forskolin to the same extend as wild type channel (figure 3-6D, E). Thus, the formation of an artificial disulfide bond between residues 507 and 1408 locked the channel in the pre-phosphorylation state and disrupted channel activation by forskolin, supporting the hypothesis that a relative movement between ICL2 and NBD2 is necessary for channel activation by PKA. We also tried to use disulfide bond to lock the channel in its open state without success.

3. Discussion

Modeling SUR2B_core with SAV1866. The regulatory subunits SURs (*ABCC8* and *ABCC9*) of K_{ATP} channels belong to the ABC transporter superfamily. Four ABC proteins have been crystallized in full length with TMDs and NBDs. The NBDs are similar to each other in their overall structures and assemblies. The TMDs however, are quite different. There are 10 TMHs in each TMD in HI1470/1 (131) and BtuCD (130), while there are 6 in each TMD in SAV1866 (12) and ModBC (129). The TMD in ModBC is much shorter than that in SUR2B, and no evident sequence homology can be found between them. In contrast, the TMD topology of SAV1866 is the same as that of SUR2B, suggesting the validity of modeling the SUR2B_core based on the SAV1866 structural template.

In our SUR2B_core model, NBD1 and NBD2 dimerize to form two nucleotide-binding pockets on their interface, consistent with the models of SUR1-NBDs in previous

studies (133,134). A significant expansion of our model is to include TMDs and to show the interaction between TMDs and NBDs. SUR2B_TMD1 interacts with NBDs via two short segments ICL1 and ICL2. ICL1 interacts with both NBDs at their borders around the first nucleotide-binding region, while ICL2 is inserted deeply into NBD2. Therefore, the TMD1 mainly interacts with the NBD2. Similar cross-interaction is found in TMD2-NBD1. Our disulfide bond experiment and functional analysis indicate that Cys1408 in NBD2 is located close to Ala507 in the TMD1. Evidence supporting such a TMD-NBD cross-interaction also comes from a recent biochemical study (135) on human multidrug transporter (*ABCB1*), a close homologue of SURs.

3.1 Importance of defining TMD-NBD interaction in SURs.

The TMD-NBD interaction in SUR2B is not only important for PKA activation, but may also provide a structural basis to explain SUR-mediated channel gating by other ligands. For instance, it is well known that NBDs are important for Mg^{2+}-nucleotide activation on K_{ATP} channels (116,122,123,125). Although KCOs activate K_{ATP} channels via binding to segments in TMDs (115), NBDs are also necessary for KCO effects (116,121,122), suggesting an allosteric modulation between these two distinct groups of channel activators. It is unclear how signals in NBDs are coupled to TMDs. Indeed, cis-interaction (TMD1 interacts with NBD1, TMD2 with NBD2) has been assumed to occur in SUR1 (127) and SUR2 (121) based on the mis-assembled MSBA structure (136). Therefore, experimental demonstration of the SAV1866-based NBD-TMD cross-interaction in SUR2B contributes significantly to the understanding of the SUR-mediated gating of K_{ATP} channels.

3.2 Mechanism of channel activation by PKA phosphorylation.

The SUR2B_core model improves our understanding of the mechanism underlying PKA activation of vascular K_{ATP} channels. Ser1387 is a critical phosphorylation site for PKA-dependent activation of Kir6.1/SUR2B channel (109). Data from the current study further support its role in channel activation. Phosphorylation of Ser1387 leads to a change in the NBD2 conformation and strengthens the interaction of NBD2 with ICL2 through the formation of an interacting triad of residues in different protein domains. Every residue of the triad is critical, as shown in our mutational analysis. The interaction of p-Ser1387 and Arg1462 is apparently caused by the electrostatic attraction. The interactions of Tyr506 with the other two seem to rely on hydrogen bonds. Our simulation study indicates that the triad is stable and may act as a primary force for changing conformations of NBD2 and IC2 on their interface. ICL2 appears to move toward NBD2 with phosphorylation, while NBD2 is reshaped and moves toward ICL2, especially the α6-tail and the α3 helix and probably the β5 sheet as well. Thus the two domains interact with each other more tightly post phosphorylation, allowing NBDs to couple with TMDs more effectively. Such a strong NBD-TMD coupling is likely to ensure the transfer of mechanical forces and movements necessary for channel gating between these protein domains. Since such a coupling is weakened without Ser1387 phosphorylation, the mechanical forces and movements produced in NBDs may not be sufficient for channel gating. By enhancing the NBD-TMD interaction, vasodilators seem to be able to use the existing forces to augment the channel activity without producing additional mechanical work. It is noteworthy that the NBD-TMD force transfer depends

on elaborate interactions of the triad, as restriction of these domains with disulfide bond diminishes channel activation.

3.3 Possible cooperation of PKA and nucleotide activators.

The SAV1866 is crystallized in ADP binding forms (PDB: 2hyd), which is represented as a dimer in NBDs. The NBD dimerization has been seen in almost all the ABC proteins including NBDs of SUR1 (133,134). In our SUR2B_core model, the two NBDs form a hetero-dimer with two nucleotide-binding pockets on the interface. Such dimerization of NBD1 and NBD2 in SURs suggests a channel conformation in the 'active' state. Nevertheless, the channel activity is rather low (less than 10% of maximal activity by pinacidil according to our observation in previous and current studies), probably because the interaction between TMDs and NBDs is quite loose at the baseline. PKA phosphorylation enhances the interaction of TMD1 (via ICL2) with NBD2 and thus drastically activates the channel. It is possible that the NBD dimerization by nucleotide binding is necessary for PKA activation. A study (137) on the Kir6.2/SUR1 channel indeed indicates that ADP concentrations determine whether the channel is activated or inhibited by PKA. Although we have not observed inhibition on Kir6.1/SUR2B, we did see a reduction in the PKA effects with decreasing ADP concentrations (unpublished data). It is of interest to know whether PKA acts on the channel without nucleotides. Since ATP is necessary for PKA phosphorylation, ATP/ADP cannot be completely removed from the system. Therefore, the information of the SUR conformation without nucleotide is needed tp fully understand the cooperative effects between PKA and nucleotide activators on the channel.

In conclusion, our SAV1866-based modeling provides novel information of SUR2B structures, and suggests a mechanism for the Kir6.1/SUR2B channel activation by PKA phosphorylation, information that is necessary for the understanding of how vasodilators activate K_{ATP} channels and relax vascular tension. The discovery of the Ser1387-Tyr506-Arg1462 triad and its critical location for the interaction between TMDs and NBDs contributes significantly to our understanding of SUR-mediated K_{ATP} channel gating. Furthermore, our combined studies of SUR2B_core model with mutational analysis support the accuracy of the SAV1866 crystal structure on the NBD-TMD interface and indicate that the SUR2B can be well represented by the SAV1866 structure.

4. Materials and Methods

Modeling of the SUR2B_core was based on the crystal structure of SAV1866 (PDB: 2HYD). The sequences of SAV1866 were aligned with both halves of the SUR2B_core (TM-NBD1 and TM-NBD2) using ClustalW. SAV1866 was 21.5% identical to TM-NBD1 of SUR2B (46.5% similarities) and 21.3% to TM-NBD2 (47.8% similarities). The NBDs were highly conserved: 28.7% identical in NBD1 and 33.5% identical in NBD2. The SAV1866_TMD were 16.0% and 12.8% identical to SUR2B_TMD1 and TMD2, respectively (~35% when amino acids with similar side chain were considered). Although the homology in amino acid sequences in TMDs is low, they can be aligned using the conserved topology within each TMD, i.e., 6 transmembrane helices (determined by hydrophobicity of the residues) in each domain with similar lengths. The sizes and positions of extracellular and intracellular linker regions were also comparable. There were only a few short gaps in the alignment (4 in TMD1 and 2 in TMD2, Online figure 1). The atomic coordinates of amino acids,

nucleotides and water molecules in the template (SAV1866) were transferred to the SUR2B_core model. The ADP molecule in first nucleotide binding pocket was replaced with ATP. The coordinates of ATP-binding Mg^{2+} and ADP-binding Mg^{2+} were obtained using the crystal structures of ABC transporters HlyB (1XEF) and TAP1 (1JJ7), respectively. The linker regions between TMD1 and NBD1 (residues 618-665), and between TMD2 and NBD2 (residues 1290-1309) were modeled as surface loops in the structure. The linker between NBD1 and TMD2 (residues 914-975) was omitted from the model due to lack of a template. Energy minimization was performed using 1000 steps of conjugate gradients optimization with the latest AMMP potential set (version tuna) (101). The model was viewed with PyMOL (http://pymol.sourceforge.net/).

Molecular dynamics (MD) simulations were carried out to study the conformational change in the NBD-ICL2 interface. The SUR2B_core model contained many hydrophobic transmembrane segments that that would require a very large molecular dynamics simulation including the lipid bilayer. Therefore, we constructed two models containing NBDs and an extended segment of ICL2 (residues 500-512) with and without a phosphate group linked to atom O_g of Ser1387. Seven thousand and five hundred H_2O were added to the model molecules to ensure an aqueous environment. Additional ions were added to neutralize the total charge of the protein. No screening dielectric term or bulk solvent correction was included. A constant dielectric of one was used. The amortized fast multipole algorithm in AMMP was used for the long-range terms in the non-bonded and electrostatic potentials so that no-cutoff radius was employed (102). Simulations were performed with a constant NVT ensemble corresponding to a pressure about 1 atm with classical molecular mechanics MD program AMMP (100,101). The

temperature was set at 310K. One thousand frames, one for each picosecond, were logged over 1-ns MD run. The *mode* structures which represent the most frequent position of each atom (101) and the RMSDs of C_αs of each residue were calculated over 1000 frames.

Rat Kir6.1 (GenBank #D42145) and mouse SUR2B (GenBank #D86038) were used in the present study. The cDNAs were cloned in a eukaryotic expression vector pcNDA3.1. Site-specific mutations were generated using a site-directed mutagenesis kit based on the Pfu DNA polymerase (Stratagene, La Jolla, CA). The cDNAs were expressed in Human embryonic kidney cells (HEK293) as previously described (109). Whole-cell patch clamp experiments were performed as detailed in previous work (Online figure 2) (109).

Data were presented as means ± s.e. Differences in means were tested with the ANOVA or Student t test and were accepted as significant if $P \leq 0.05$.

Acknowledgments

The authors are grateful to Dr. S. Seino and Y. Kurachi for their gifts of the Kir6.1 and SUR2B cDNAs, respectively. This work was supported by the NIH (HL067890) and Georgia State University Research Program Enhancement Fund. We thank Dr. I. Weber for her critical reading of the manuscript.

Figures

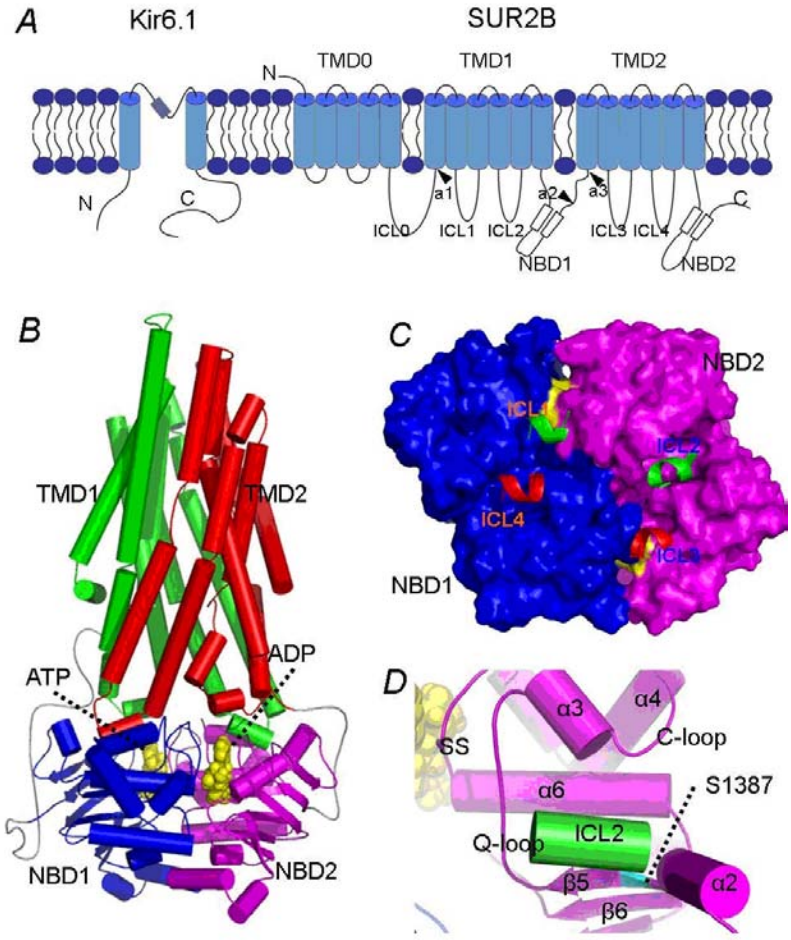


Figure 3-1 SUR2B_core model.

A. A skeletal representation of K_{ATP} channel. Kir6.1 (left) has two transmembrane helices (TMHs) with both N-, C- termini located in cytosol. SUR2B has 17 TMHs, forming three transmembrane domains (TMD0, TMD1 and TMD2) in a 5-6-6 pattern. The N-terminus is located extracellularly and the C-terminus intracellularly. The nucleotide binding domain-1 (NBD1) is located between TMD1 and TMD2, and NBD2 lies in the C-terminal. The intracellular linkers (ICLs) between TMHs are indicated. The SUR2B_core includes sequences from TMD1 to C-terminus. The region between residue 286 (arrow a1) and 914 (a2) was used as TM-NBD1. The residues between 976 (a3) and 1546 (C-terminus) were used as TM-NBD2. TM-NBD1 and TM-NBD2 were aligned to SAV1866 and used to generate the SUR2B_core model. A few residues between a2 to a3 were excluded in the SUR2B_core model as it does not have

homology with SAV1866. **B.** The overall structure of SUR2B_core. TMD1 (residues 286-617, green) was linked with NBD1 (residues 666-913, blue) with 38 residues (loop in gray). TMD2 (residues 976-1289, red) was linked with NBD2 (residues 1310-1546, purple) with 20 residues. The linker regions between TMD1 and NBD1, TMD2 and NBD2 are longer than the corresponding regions in SAV1866, and thus they were looped beside the NBD domains (gray). The first nucleotide binding pocket (formed by Walker A in NBD1 and signature sequence in NBD2) was modeled with ATP (yellow). The second nucleotide binding pocket (formed by Walker A in NBD2 and signature sequence in NBD1) was modeled with ADP. **C.** Interaction of the intracellular linkers with NBDs. ICL1 and ICL3 interacted with both NBD1 and NBD2 across the nucleotide-binding regions (yellow). ICL2 was inserted into a groove formed by the NBD2 surface, and ICL4 interacted into NBD1. **D.** Details of the interaction between ICL2 and NBD2. NBD2 formed a big groove to host ICL2, in which β5 formed the bottom, Q-loop, α2, α3, α6, and a loop between α3 and α4 (named the C-loop because a cysteine residue, Cys1408 was located in this segment) formed the walls. The phosphorylation site Ser1387 was located in the β5 sheet and is shown in cyan.

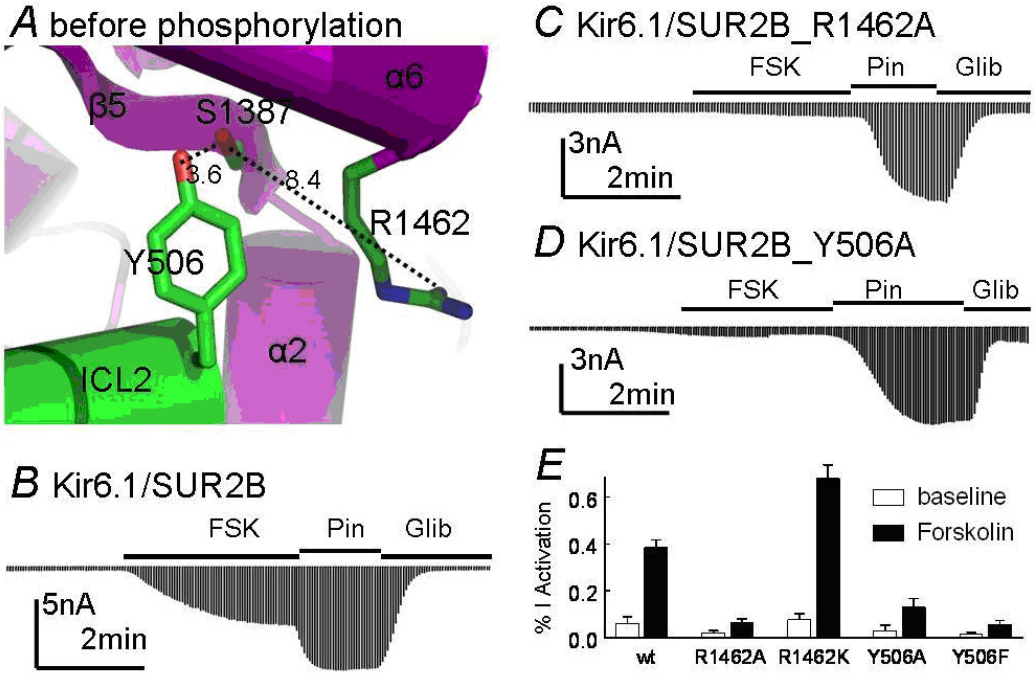


Figure 3-2 Ser1387 and surrounding sites in PKA phosphorylation.

A. Ser1387 was close to Tyr506 and Arg1462. The atom O_g in Ser1387 side-chain was 3.6Å from atom O_h of Tyr506, and ~8Å from the positive charges of Arg1462 (the middle of the two N_h atoms). **B.** Forskolin (FSK, 10μM) activated whole cell currents of Kir6.1/SUR2B, which were further activated by pinacidil and inhibited by glibenclamide (See online figure 2 for details). **C.** The channel with SUR2B_R1462A mutation failed to be activated by forskolin. **D.** SUR2B_Y506A mutant greatly diminished channel activation by forskolin. **E.** Summary of forskolin effects on wt and mutant Kir6.1/SUR2B channels. Note the Y506F mutant was not activated, and the R1462K mutant was activated by forskolin.

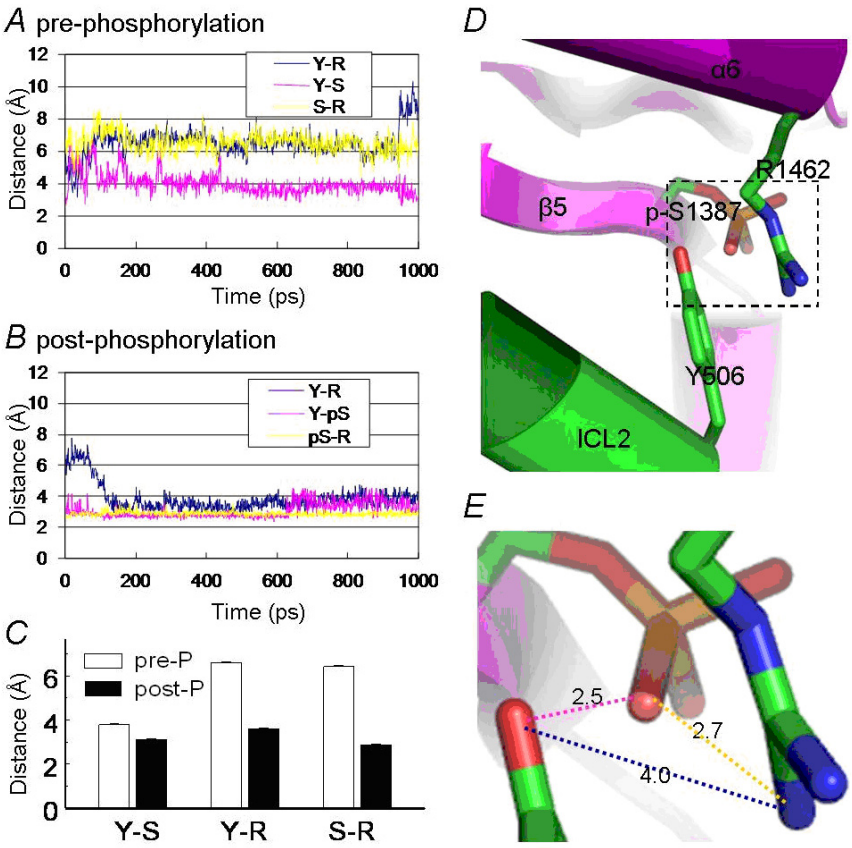


Figure 3-3 The interaction between Tyr506, Ser1387 and Arg1462.

Interaction of Tyr506, Ser1387 and Arg1462 following phosphorylation. Molecular dynamics simulations were performed on ICL2 with a short flanking sequence (residues 500 to 512) and NBDs. A phosphate group was linked to atom O_g of Ser1387 representing phosphorylation. **A.** The distances between each pair of residues were monitored before phosphorylation. Y-S was measured between Tyr506-O_h and Ser1387-O_g. Y-R was measures between Tyr506-O_h and Arg1462-N_{h1} or N_{h2} depending on which one atom was

closer to Tyr-O_h. So was S-R. **B.** The triad distances post phosphorylation. S represented one of the three O_p atoms of phosphorylated-Ser1387. Six distances were measured between p-Ser1387 and Arg1462, and three between Tyr506 and p-Ser1387, with the closest distances shown. **C.** Summary of the distances among the three residues pre- and post-phosphorylation. The last 800 time frames were counted after stabilization with the initial ~200 runs (see A and B). All three distances were shorter in the post-phosphorylated form. Major changes were seen in distances between Ser1387 and Arg1462, and Tyr506 and Arg1462. **D.** *Mode* structure of the triad post phosphorylation. **E.** Close review of the triad shown in D. The numbers in the figure are the distances (in Å) between the interested residues. They are slightly and insignificantly different from the average distances shown in C.

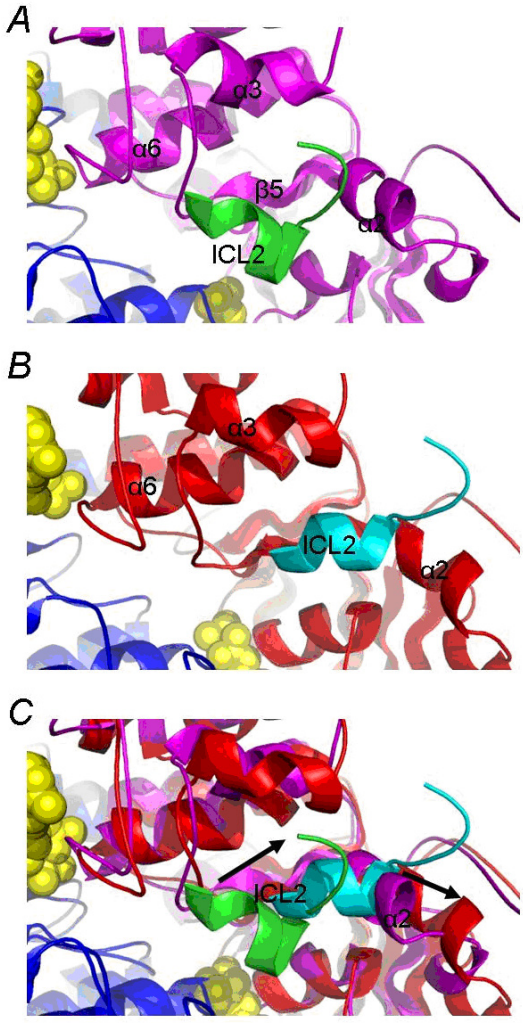


Figure 3-4 The movement of ICL2 following phosphorylation.

All pictures represented the *mode* structures of the MD simulations. **A-C.** Top views of ICL2 and NBD2 in pre-phosphorylation, post-phosphorylation and overlay, respectively. The pre-phosphorylation structure was shown in the same set of colors as figure 3-1. In post-phosphorylation structure, NBD2 was colored red and ICL2 was colored cyan. Compared to the *mode* structure of the pre-phosphorylated form, ICL2 helix moved horizontally away from the center of NBDs by 8.4±0.4Å (measured at C_αs of residues 500-506). Accordingly, the α2 helix in NBD2 moved to yield space for ICL2. The arrows in **C** indicated the direction of movement of ICL2 and α2 helix.

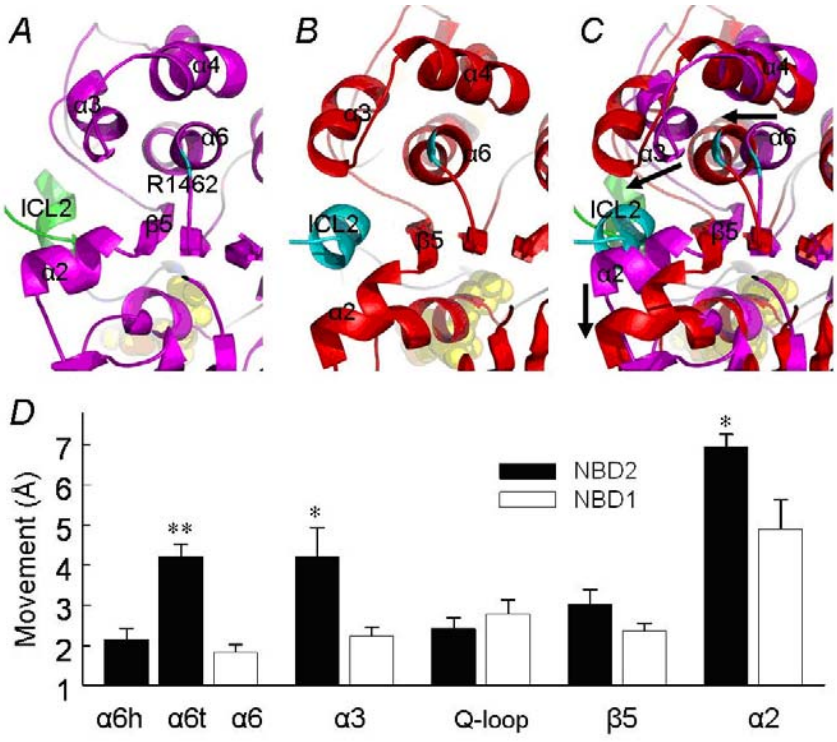


Figure 3-5 Conformational changes in NBD2 interface.

A-C. The side views of the NBD2 interface in *mode* structures of pre-phosphorylation, post-phosphorylation and overlay, respectively. The structures were displayed with the same set of colors in figure 3-4. The backbone of the Arg1462 moved by 4.5Å toward ICL2, leading to a movement of the α6-tail (residue 1455-1462). Major position changes were also seen in α2 and α3 helices. The arrows in **C** indicated the movement direction of α6, α3 and α2. **D.** Analysis of critical segments in the NBD2 groove. Major changes were seen on NBD2-α6-tail, α3 and α2, compared to the corresponding segments in NBD1 (by measuring position changes at C_αs. **, $P<0.001$; *, $P<0.05$). The β5 in NBD2 changed more than

overall β sheets (2.3±0.2Å) in NBD2 and the corresponding segment in NBD1. The α2 helices in both NBDs changed by large distances suggesting they are flexible, while α2 helix in NBD2 moved laterally, making space for ICL2. In contrast, NBD2-α6-head (α6h) and Q-loop did not show significant position changes.

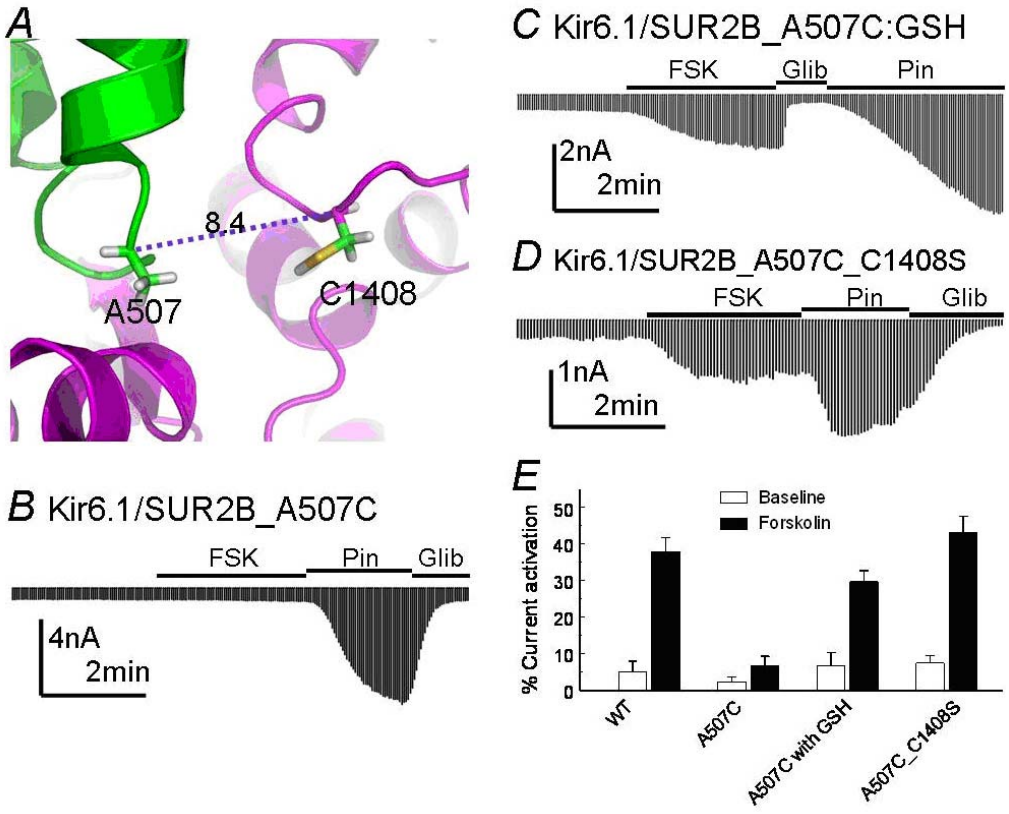


Figure 3-6 The blockade of the PKA effects with disulfide bond.

Blockade of the PKA effects with disulfide bond between ICL2 and NBD2. **A.** A close-up of residues A507 and C1408. The distance between C_αs of the two residues was ~8Å, allowing the two residues to form disulfide bond. **B.** When Ala507 was mutated to cysteine, channel activation by forskolin was abolished. **C.** When 3 mM of reduced glutathione were included in the pipette solution, forskolin activation of the mutant channel was restored. **D.** When the Cys1408 was mutated to serine, a similar residue but cannot form a disulfide bond, the channel behaved like wt channel. **E.** Summary of the forskolin effects on wt and mutants.

Online Figure 1
Shi et al.

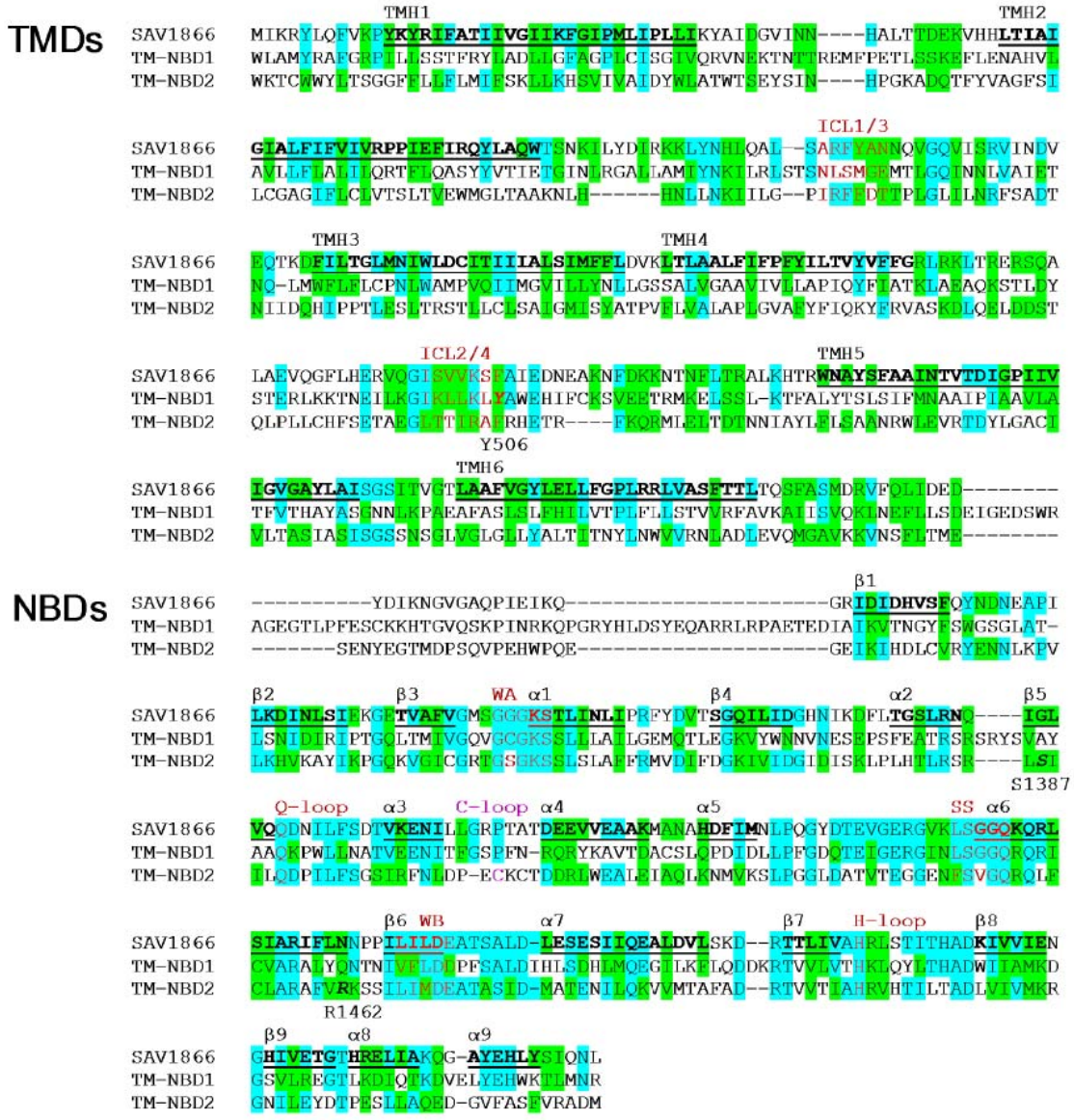


Figure 3-7 Alignment of SUR2B and SAV1866 sequences (Online figure 1).

TM-NBD1 and TM-NBD2 relative to residues 286-914 and 976-1546 of SUR2B. Residues in SUR2B that are identical and similar to SAV1866 were shaded in cyan and green, respectively. Six transmembrane helices (TMH1-6) in SAV1866 were shown in bold and underlined, which correspond to TMH6-11 and TMH12-17 in SUR2B_TMD1 and TMD2, respectively. Sequences between TMH2 and 3, and between

TMH4 and 5 are intracellular linkers. Short helices that are critical for TMD-NBD interactions were shown in red. ICL1, ICL2 are located in SAV1866 and SUR2B_TMD1 while ICL3, ICL4 in SUR2B_TMD2. Although the homology in TMDs between SAV1866 and SUR2B is low, the TMHs' lengths are the same. The lengths of intracellular and extracellular linkers are comparable. Therefore, when they are aligned, there are only 4 short gaps in TMD1 and 2 in TMD2. The similar topology allowed us to model SUR2B TMDs with the SAV1866 crystal structure as the template. The homology in NBDs is higher. Critical sequence motifs such as Walker A (WA), Walker B (WB), signature sequence (SS), Q-loop (between β5 and α3) and H-loop in NBDs are conserved (in red). The secondary structures in SAV1866_NBD were shown in bold and underlined. The loop between α3 and α4 helices was named C-loop as Cys1408 was found in it. Tyr506, Ser1387 and Arg1462R in SUR2B shown in bold and italic were critical for PKA activation.

Online Figure 2
Shi et al.

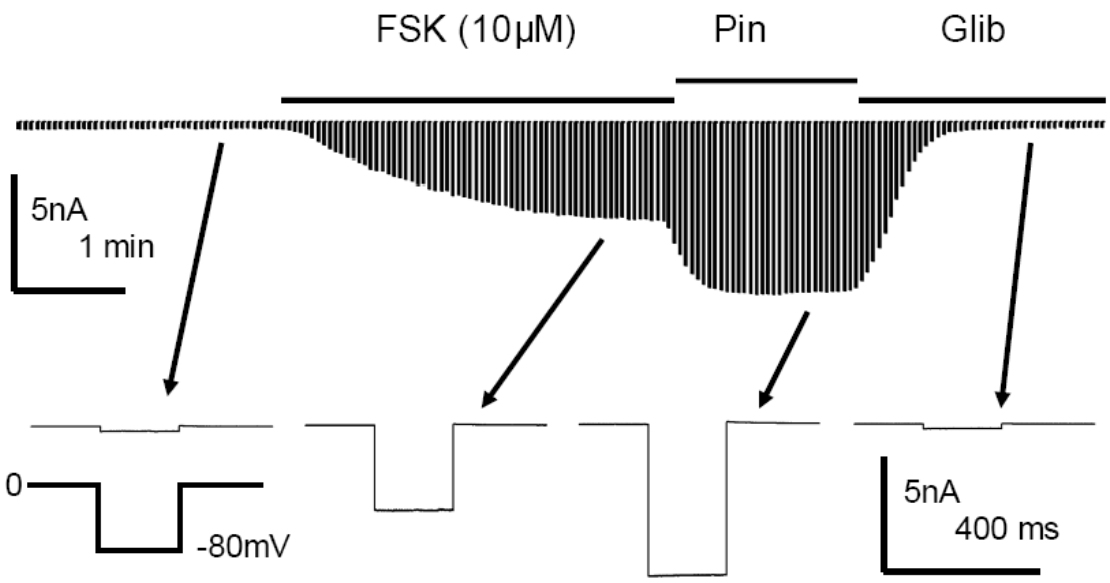


Figure 3-8 Activation of the Kir6.1/SUR2B channel by forskolin in HEK293 cells (online figure 2). Whole-cell currents were recorded from a cell transfected with Kir6.1/SUR2B. Application of forskolin (10μM) increased the whole-cell currents rapidly and reached a plateau in about 2 min. K_{ATP} channel opener pinacidil (Pin, 10 μM) strongly augmented the currents that was potently inhibited by glibenclamide subsequently (Glib, 10 μM). Note that arrows point to where each bottom trace was taken from.

4. Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from atomic resolution crystal structures

Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from atomic resolution crystal structures

*Xianfeng Chen[1], Irene Weber[1] and Robert W. Harrison[1,2]**

[1]Department of Biology and [2]Department of Computer Science, Georgia State University, 33 Gilmer Street, Atlanta, Georgia 30302–4010, USA.

Email address: xchen5@gsu.edu, iweber@gsu.edu and rhrrison@cs.gsu.edu

* Correspondence to: Phone: 404-413-5724; Fax: 404-413-5717; E-mail: rharrison@cs.gsu.edu.

Abstract

Water plays a critical role in the structure and function of proteins, although the experimental properties of water around protein structures are not well understood. The water can be classified by the separation from the protein surface into bulk water and hydration water. Hydration water interacts closely with the protein and contributes to protein folding, stability and dynamics, as well as interacting with the bulk water. Water potential functions are often parameterized to fit bulk water properties because of the limited experimental data for hydration water. Therefore, the structural and energetic properties of the hydration water were assessed for 105 atomic resolution (≤1.0 Å) protein crystal structures with a high level of hydration water by calculating the experimental water-protein radial distribution function or surface distribution function (SDF) and water radial distribution function (RDF). Two maxima are observed in SDF: the first maximum at a radius of 2.75 Å reflects first shell and hydrogen bond interactions between protein and water, the second maximum at 3.65 Å reflects second shell and van der Waals interactions between water and non-polar atoms of protein forming clathrate-hydrate-like structures. Thus, the two shells do not overlap. The RDF showed the features of liquid water rather than solid ice. The first and second maxima of RDF at 2.75 and 4.5 Å, respectively, are the same as for bulk water, but the peaks are sharper indicating hydration water is more stable than bulk water. Both distribution functions are inversely correlated with the distribution of B factor (atomic thermal factors) for the waters, suggesting that the maxima reflect stable positions. Therefore, the average water structure near the protein surface has experimentally observable differences from bulk water. This analysis will help improve the accuracy for models of water on the protein surface by

providing rigorous data for the effects of the apparent chemical potential of the water near a protein surface.

1. Introduction

Water is essential to the structure and function of proteins. Generally, water around proteins can be divided into three categories (138): 1) bulk water that surrounds the protein molecule at a separation of greater than van der Waals contact, 2) individually bound water that forms hydrogen bonds with charged or polar protein atoms in cavities inside the protein(139), and 3) hydration water at the protein surface with direct interactions with the protein (140,141). The three categories have different functions. Bulk water is free to move, it assists in protein diffusion relative to the other interacting molecules by random movement in solution. Hydration water forms water networks around the protein surface to keep protein in solution. Individually bound water has multiple contacts that stabilize the protein structure. For example, individually bound water stabilizes cavities in thrombin and trypsin (142,143), and mediates contacts between proteins and their ligands (144). The presence of individually bound water depends on the local structure of the protein, and is often conserved among related proteins. However, a water molecule can share the properties of both individually bound water and hydration water if it has multiple contacts with protein and contributes to a hydration water network. Both individually bound water and hydration water are also known as protein-bound water and form strong interactions with proteins. Bound water is visible in protein crystal structures, but bulk water is not observed because there is no stable location for the water molecules.

Hydration water is a major form of protein-bound water that contributes to many properties of proteins such as protein folding, solubility, drug docking, and oligomer formation as reviewed in Raschke 2006 (145). Hydration water directly interacts with the

protein surface and affects the local protein structure. Protein enzymes without a hydration layer will lose their catalytic abilities (146). Direct observation of hydration water properties in proteins is difficult, except for visualization in atomic resolution (~1.0Å) crystal structures of proteins (147). Elegant calorimetric experiments on hydrated collagen show that there is a specific binding energy or chemical potential associated with hydration (148). Terahertz spectroscopy has shown that the effects of the hydration environment can extend to 10Å from the surface of a protein (149). Neutron scattering of dipeptide solutions showed small differences between bulk and hydration water (150), but the different effects may be observed with proteins due to their larger size. X-ray scattering has been used with myglobin to study hydration water (151).

Although hydration water is very important for protein structure and function, there are no specific parameters for the potential of hydration water in simulations such as molecular dynamics simulations (MD). The water potential function is defined by a set of parameters in simulations that represent the potential energy of water. The water potentials currently used in MD programs are based on water models such as TIP3P (152), TIP4P (153), TIP5P (76) and SPC (78) that are parameterized to reproduce the experimental data of bulk water. Recently, neutron scattering data on dipeptides has been used as a basis for generating water potentials (150). Theoretical calculations using these potentials have been decomposed into polar, apolar and charged interactions (154) in order to fit compressibility data. These potentials derived from bulk water may not accurately represent the properties of hydration water. For example, when TIP3P was used with CSFF(155) and GLYCAM-2000a(156), the new versions of CHARMM (157) and AMBER (158) force fields respectively, the modeled carbohydrate-water interactions

on the surface of disaccharides were less structured than in experimental data (81). Even though the compressibility data were well reproduced in (Dardarlat 2006) (154) the radial distribution curves show less detail than is seen experimentally. Hence, the differences between the properties of hydration water and bulk water need to be defined in order to improve the water potentials and especially to reproduce the critical interactions between water and proteins or other soluble macromolecules.

The structural and energetic factors of hydration water can be studied using the water protein radial distribution function or surface distribution function (SDF) and water radial distribution function (RDF). The SDF describes the density of water as a function of the distance from the protein surface (the closest non-hydrogen atom). The RDF describes the density of water as a function of the distance from a particular water molecule. The RDF is determined by averaging over the many different local structures formed in liquid water (70-72). Local structures include neutral (uncharged) water structures, and two-water-molecule structures with one proton attached ($H_5^+O_2$) (70). Water molecules also form higher order clusters, such as ten-molecule tetrahedral clusters (71) and 280-molecule expanded icosahedral water clusters comprising 14-molecule tetrahedral units (72).

Previous studies on the hydration water were highly dependent on computational simulations due to the difficulty of direct measurements. Direct measurements of visible hydration water became possible when the atomic resolution crystal structures of proteins became available in significant numbers. Moreover, an average over a number of crystal structures will reduce the deviation caused by the differences in crystallization conditions and other sources of experimental variation.

This study examined 105 crystal structures from the protein data bank (http://www.wwpdb.org) (159) solved with X-ray data at resolutions better than 1.0 Å. The SDFs and RDFs were calculated and averaged. Two density maxima representing two shells are observed in SDF, the water molecules in the first shell are much denser than those in the second shell. The maxima of RDF are higher than those of bulk water and the minima are lower. In both SDF and RDF, the higher density of water corresponds to the lower B-factor value of water, confirming that water molecules are stable in each maximum.

2. Methods and Algorithms

2.1 The water-protein distance

The ideal model of a protein is a sphere surrounded by first and second water shells (figure 4-1a). The positions and densities of the two shells are determined by the interactions between water and the protein. In practice, the protein-water distance is defined as the distance between the water oxygen and the closest non-hydrogen atom in the protein molecule.

2.2 The water-protein radial distribution

In protein crystal structures, the average density of water around protein (ρ) within any water-protein distance h is

$$\rho = \frac{n}{\frac{4}{3}\pi\left(R_p + h\right)^3 - \frac{4}{3}\pi R_p^3}$$

Where, n is the number of water molecules with water-protein distance less than h, R_p is the radius of the protein molecule p.

The bulk water is invisible in protein crystal structures because it is not bound to protein. The density of "visible" water (mostly hydration water) around the protein decreases sharply with distance from the protein surface. The SDF is the water density as a function of the water-protein distance (h) and termed g_{wp}

$$g_{wp}(h) = \frac{d\rho}{dh} \qquad (1)$$

2.3 Water radial distribution

The RDF in the protein crystal structure must be adjusted due to the presence of protein molecules, unlike the RDF in bulk water where only water-water interactions are considered. In the presence of protein atoms, the volume over which the water can be observed is restricted and thus the manner in which the density is normalized must be modified. The raw counts of water molecules within each radial shell are converted into an expected value for the RDF based on the restriction of the shell due to interactions between water and protein molecules. This process is shown in figure 4-1b where the thin orange shell represents a volume shell at a distance r_w from the water w. The thin orange shell illustrates the unimpeded differential volume element (UDV). The water density at any point within the area surrounded by the orange shell is not only related to the distance to water w, but also related to the distance to protein surface. Steric interactions between the protein and water prevent or impede water from binding in all of the positions in the UDV, thus the area surrounded by the orange shell defines the available differential volume element (ADV). Thus, two water densities can be calculated within ADV: the

water density (ρ_{ww}) which is related to the distance from the water *w* and caused by water-water the interaction and the water density (ρ_{wp}) which is related to the distance from the protein surface and caused by the water-protein interaction.

$$\rho_{ww} = \frac{n_{ww}}{v_w}$$

$$\rho_{wp} = \frac{n_{wp}}{v_w} = \frac{\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\int_{r=0}^{r_w} r^2 g_{wp}(h_{wr\phi\theta})\sin\phi d\phi d\theta dr}{v_w}$$

Where, n_{ww} is the actual total number of water molecules within ADV, v_w is the volume of ADV, r_w is the distance from UDV to water *w*, n_{wp} is the number of water molecules within ADV evaluated by SDF (g_{wp}), $h_{wr\phi\theta}$ is the distance from each point within ADV to the protein surface.

Thus, the raw RDF (g_{raw}) and the water density distribution for normalization (g_{nor}) can be calculated as

$$g_{raw}(r_w) = \frac{d(\rho_{ww})}{dr_w}$$

$$g_{nor}(r_w) = \frac{d(\rho_{wp})}{dr_w}$$

$d(\rho_{ww})$ is the water density in the UDV calculated from the actual number of waters (*w1*, *w2* and w3 indicate the three water molecules in figure 4-1b) divided by the volume of the UDV. $d(\rho_{wp})$ is the average of the water density of each small cell in the UDV (*e1*, *e2*, *e3*….etc), which is determined by its water-protein distance ($h_{wr\phi\theta}$). dr_w is the thickness of the UDV.

Thus, the averaged raw RDF ($\langle g_{raw} \rangle$) and the averaged water density for normalization ($\langle g_{nor} \rangle$) over all sampled crystal structures can be expressed as

$$\langle g_{raw}(r_{wp}) \rangle = \frac{1}{P} \sum_{p=1}^{P} \left(\frac{1}{W_p} \sum_{w=1}^{W_p} g_{raw}(r_{wp}) \right) \qquad (2)$$

$$\langle g_{nor}(r_{wp}) \rangle = \frac{1}{P} \sum_{p=1}^{P} \left(\frac{1}{W_p} \sum_{w=1}^{W_p} g_{nor}(r_{wp}) \right) \qquad (3)$$

Where P is the total number of protein structures in our sample, W_p is the total number of water molecules in protein p, r_{wp} is the distance from the UDV to water w in protein p.

Therefore, in crystal structures the normalized RDF (g_{nmlzed}) can be expressed as

$$g_{nmlzed}(r_{wp}) = \frac{\langle g_{raw}(r_{wp}) \rangle}{\langle g_{nor}(r_{wp}) \rangle} \qquad (4)$$

2.4 The number of water molecules

Every water molecule has a refined occupancy in the PDB (Protein Data Bank) file. The majority of water has an occupancy value of 1.0 indicating the presence of this water in all protein molecules forming the crystal. Some waters have occupancies of less than 1.0 indicating the water is not present in all protein molecules in the crystal. In this study, the sum of the occupancy of each water molecule was used instead of the actual number of waters.

2.5 The sharpness of a RDF

A numerical measure based on the width at half height was defined in order to describe the sharpness of the features of the RDF. The peaks in the RDF are asymmetric, especially the first peak, which makes the standard symmetric half height measure

difficult to apply. Because of this asymmetry, the measure was defined as the amplitude between a maximum and its next minimum (HD) divided by the width of the maximum at the half value of HD (figure 4-1c). This corresponds to the width at the smaller of the two peak heights.

2.6 The water- protein radial distribution of B-factors

The surface (or water-protein radial) distribution function of B-factors (SDFB) is the averaged B-factor of water as a function of water-protein distance (h) which can be calculated from

$$B(h) = \frac{d}{dh}\left(\frac{1}{W}\sum_{w=1}^{W}\frac{b_w}{o_w}\right)$$

Where W is the total number of water molecules in the sphere with distance ($<h$) from protein p, b_w is B-factor of each water oxygen in the sphere, o_w is the occupancy of each water molecule in the sphere. Thus, the averaged SDFB over all sampled crystal structures can be expressed as

$$\left\langle B(h_p)\right\rangle = \frac{1}{P}\sum_{p=1}^{P}\left(\frac{1}{W_p}\sum_{w=1}^{W_p}B(h_p)\right) \qquad (5)$$

Where P is the total number of protein structures in our samples, W_p is the total number of water molecules in protein p, h_p is the distance from the surface of the sphere to the closest atom in protein p.

2.7 The water radial distribution of B-factor

The water radial distribution of B-factor is the averaged B-factor of water as a function of water-water distance. At any distance from water w (r_w), the radial distribution function of water B-factor (RDFB) can be calculated as

$$B(r_w) = \frac{d}{dr_w}\left(\frac{1}{W}\sum_{w=1}^{W}\frac{b_w}{o_w}\right)$$

Where W is the number of water and partial water molecules in the ADV with radial ($<r_w$) from water w (figure 4-1b), b_w is B-factor of each water in the ADV, o_w is the occupancy of each water in the ADV. Thus, the averaged RDFB over all sample crystal structures can be expressed as

$$\left\langle B(r_{wp})\right\rangle = \frac{1}{P}\sum_{p=1}^{P}\left(\frac{1}{W_p}\sum_{w=1}^{W_p}B(r_{wp})\right) \quad (6)$$

Where P is the total number of protein structures, W_p is the total number of water molecules in protein p, r_{wp} is the distance from the UDV to water w in protein p.

The above calculations are performed with PDBAnalyzer, a JAVA program developed in-house, and using Microsoft SQL Server 2000 as the database engine. The data for each water molecule such as occupancy, B factor, PDB file name, atom number, distance to the closest atom in protein, and the polarity of this atom in protein were input to a database table as one record. Then, SQL scripts were used to extract data for SDF from more than ten thousands records. Similarly, the data for each pair of water molecules with separation less than 10.5 Å were input into a database table as one record, along with both occupancies, B factors, PDB file names, atom numbers and distances to protein surface. The SQL scripts were used to extract data for RDF from over one million records. Hydrogen atoms can rarely be seen due to the presence of a single electron, even in sub-atomic resolution crystal structures of proteins. Therefore, the position of the water molecule was defined by the position of the oxygen atom, and the RDF is the O-O distribution. The water-protein distance is also calculated as the distance from the O to the non hydrogen atom of the protein.

3. Results

One hundred and five protein crystal structures refined with water molecules and diffraction data at resolutions ≤ 1.0 Å were analyzed (the structures are listed in the Appendix). These structures contain a total of 33376 water molecules, including 9026 partially occupied waters. The variable crystallization conditions such as temperature, ion concentration and pH value, and differences in crystallographic methodology are assumed to have little or no effect on the calculated radial distributions.

3.1 Water-protein radial distribution

The calculated SDF showed two maxima, representing two water shells around proteins, at radial distances of 2.75 Å and 3.65 Å (figure 4-2a). The first maximum of 0.19 atoms/$Å^3$ is twice the average density of bulk water at 267K temperature (160). The second maximum of 0.024 atoms/$Å^3$ is much lower. The SDF (blue) is the sum of the water-polar atom radial distribution (SDFpol, red) and the water-non polar atom radial distribution (SDFnon, green), since the protein surface consists of polar atoms (or charged atoms, mostly oxygen and nitrogen) and non-polar atoms (mostly carbon). SDF overlaps with SDFpol almost perfectly between 2.25~3.25 Å, indicating that the first shell consists of water molecules with hydrogen bonds to the polar (or charged) atoms of the protein. The second maximum between 3.2-4.2 Å has a larger contribution from SDFnon, indicating the importance of van der Waals interactions between water molecules and protein. Thus, the first shell is formed around polar atoms and the second shell is formed around the non-polar atoms of the protein, and the two shells are distinct

(figure 4-2b). The first shell is much denser than the second shell, because polar and charged atoms predominate on the solvent accessible surface of proteins.

The water-water interactions for each water molecule were calculated from all samples. More than 98% of water molecules in the second shell, compared to 82% of all waters, have hydrogen bond interactions with at least one additional water molecule (with average 1.7), indicating further stabilization of these waters by formation of networks. According to Petrenko and Whitworth 1999 (161), water molecules interact to form small cage-like structures called clathrate hydrates around methane, ethane, $(CH_2)_2O$ and other hydrophobic molecules. Similar cages around methyl groups are observed in a protein crystal structure (162). Thus, the majority of water molecules in the second shell of SDF form clathrate-hydrate-like structures.

3.2 The Water Radial Distribution Function

The equation (2) used to calculate the actual RDF was developed for calculation of RDF in a system of bulk water where the expected density of water is constant. The observed density of water as a function of water-water distance in proteins is not constant (figure 4-3a. dashed pink) which causes the uncorrected or observed RDF to decrease rather than remaining constant with increasing distance from the protein surface (figure 4-3a. blue). To eliminate this bias, the normalized RDF (g_{nmlzed}) (figure 4-3b. red) is used as calculated from equation (4). The normalized RDF is compared to the RDF of bulk water from ALS x-ray scattering and neutron diffraction at 298K (163,164) (figure 4-3b). The normalized RDF in crystal structures fluctuates more sharply than in bulk water, although the positions of the maxima and minima are the same. The amplitude of each maximum or minimum in the normalized RDF is larger than in the RDF of pure bulk

water (figure 4-3b) and the peaks are sharper. This suggests that steric hindrance and specific interactions between the protein surface and the water lead to a more ordered water structure in the vicinity of the protein surface. Unlike bulk water where three peaks are seen in the RDF (164), these surface effects extend to a (weak) fourth peak at about 8.5 Å distance (figure 4-3b, table 4-1) and evidence is seen of ordering at 10 Å.

3.3 The affect of individually bound water on the normalized RDF

Individually bound water is a small component of the water observed in protein structures. Some individually bound water molecules are deeply buried inside the protein with no surface accessibility. For example, there is a water molecule deeply embedded in inverting glycosidase (water 513 in PDB entry 1KWF), which is 9.3 Å away from the closest water. Such water molecules can bias the analysis of hydration water because they do not belong to the hydration water network. Methodologically, individually bound water was defined as water with no other water molecules within 3.2 Å (the maximum distance to form a hydrogen bond) in this study. Using this definition, individually bound water formed 18% of our sample. The normalized RDF of "pure" hydration water, after removing the individually bound water molecules, also is sharper than that for bulk water and even more than the normalized RDF (Table 1). Similar results were obtained by changing the water-water distance that defines individually bound water to 3.0 Å or 3.4 Å. However, a discontinuity in the RDF is introduced at the cutoff which is not physically realistic. Therefore, the RDF of "pure" hydration water in the experimental data is sharper than bulk water indicating hydration water is less mobile and more ordered than bulk water.

3.4 The relationship between water density and its B-factor

Atoms in molecular structures move due to thermal vibrations. In crystal structures, the atomic B factor or thermal factor is used to describe this motion, as well as the effects of static disorder (165). The distribution of atomic B factors can be used to study protein flexibility (166), thermal stability (167,168) and protein dynamics (101). The relationship between water density and water thermal factor could shed light on the dynamics of hydration water.

The radial distribution of averaged B factor of water molecules as a function of the water-protein distance (SDFB) is shown in figure. 4-4a. The maxima and minima are the inverse of those in the SDF (figure. 4-4b). Similarly, the radial distribution of averaged B factor of water molecules as a function of the water-water distance (RDFB) show inversed positions of maxima and minima relative to the RDF. In conclusion, water density and its B-factor are negatively correlated, confirming that the water molecules in the shells of high density have relatively low mobility.

4. Discussion

4.1 What is the small maximum in the normalized RDF?

The small maximum in the vicinity of 2.3 Å (figure 4-3a, 4-3b.) has not been reported previously. One possibility is just sodium misassigned as water, because sodium is often abundant in the crystallization solutions and can easily resemble a “water” molecule during the refinement of the crystal structure. Harding(169) analyzed the radial distribution of sodium-water in protein crystal structures and found the peak to be

between 2.3~2.5 Å. However, the small maximum is very sharp within the narrow range of 2.28~2.3 Å, and distinctly different from the maximum due to the assigned sodium-water distribution. The water molecules with separation of 2.28~2.3 Å were checked manually, and none resembled sodium ions with short interactions with several water molecules or polar atoms of protein as described previously (169-171). Therefore, the small maximum is unlikely to be caused by sodium mistakenly assigned as water.

It is not clear if the small maximum near 2.3 arises from any particular water structure, although some water structures such as the RDFs of ice (Ih) (164) produces a maxima at 2.5 Å and $H_5^+O_2$ (70) has a OO distance of 2.52Å. On the other hand, 90% of the crystal structures in this sample were refined with SHELX (http://shelx.uni-ac.gwdg.de/SHELX/). SHELX picks the maximum of electron density with a distance cut-off of 2.3 Å by default. In addition, most of the waters in the small maximum were observed to have poorly defined electron density and high B factors. Therefore, the small maximum is possibly caused by the distance cut-off of SHELX, or over-interpretation of the electron density maps. However, the possibility that the small maximum represents protonated water structure cannot be excluded.

4.2 How closely does the hydration water in protein crystals represent water in solvated proteins?

Water and protein structures are affected by many factors such as temperature and ion concentration. Protein crystals are grown from solutions with various buffers, salts and/or organic precipitating agents and additives. Many protein crystals are grown at room temperature and atomic resolution X-ray data recorded at very low temperature (liquid

nitrogen). However, these factors were ignored in the calculations of the normalized RDF in protein crystal structures, which raises several questions as discussed below.

Crystals are often grown in high ionic strength solutions which could affect the RDF. The high ion concentration in the hydration water of protein crystals, based on previous studies (172,173), does not increase the sharpness of RDF maxima. The RDF of 10 M lithium chloride solution has no change relative to the RDF of pure water (173). However, with an 8M NaCl solution the sharpness of the RDF maxima decreased about 30% from that of pure water (172). Salts and buffers tend to reduce the sharpness if they have any affect. The RDF in this study shows an increase in sharpness for RDF maxima (figure 4-3b red), which strongly suggests that the effects of the salts are not causing the difference between hydration water and bulk water.

On possible explanation for the increase in the order in the RDF is the formation of ice. The ice specific maxima (from RDF of ice (Ih) (164)) are not observed in the normalized RDF of water in protein crystals, indicating the hydration water on the protein surface is still liquid-like even at very low temperature. Even though the normalized RDF in protein crystal structures is sharper than in bulk liquid water at 298K or at 275K (174), it is unclear how much the low data collection temperature affects the normalized RDF. To answer this question it would be necessary to obtain atomic resolution protein structures with data collected at room temperature, which is impossible for most protein crystals with current technology.

Due to the difficulties of experimental measurement, the current studies on hydration water on the protein surface highly depend on computational simulations. Many simulations use water models such as TIP3P (76), TIP4P, TIP5P (76) and SPC (78)

derived from bulk water. This study, with experimental atomic resolution data from 105 protein crystal structures, showed a significant difference in the water structure and water potentials between hydration water and bulk water in protein crystals. The results agree with previous studies on a much smaller number of proteins (147). The RDF of liquid water in vycor, a kind of porous glass with a hydrophilic surface similar to proteins, has higher first and second maximum and lower first minimum than that of bulk water at 300K (175). The RDF of water around the myoglobin molecule showed that both the depth of the first minimum and the height of the second maximum are larger than those for pure bulk water at 298K (151). Neutron scattering on dipeptide solutions shows smaller but distinct effects (McLain 2008)(150), as expected due to the smaller size of the solute. The SDFpol and SDFnon in this study are similar to those seen in computer simulations of trypsin, ribonuclease A, hen egg white lysozyme and α-lactalbumin (154).

5. Conclusions

There are two layers of hydration water in protein crystals as evidenced by the two maxima observed in SDF calculated for 105 atomic resolution structures. This study expands on the analysis of three protein structures by neutron and X-ray diffraction (147) by having a much larger and more complete sample that allows for better normalization. Like (Svergun et al 1998) (147) the first significant maximum is sharper than is seen in bulk water, but there is ordered structure at a larger distance and a correlation between RDF and thermal factors. The first maximum at 2.75 Å represents the center of the inner layer of hydration water formed by hydrogen bonds between water and polar atoms of the protein. The second maximum at 3.65 Å represents the outer layer of hydration water

formed by water-water hydrogen bonds and van der Waals interactions between water and non-polar atoms of the protein. The structure of water in the outer layer resembles that of a clathrate-hydrate. The RDF displays the characteristics of liquid water rather than solid ice. The first and second maxima of the RDF are at 2.75 and 4.5 Å, respectively, the same distances as those of bulk water. However, the maxima/minima of the RDF of hydration water are higher/lower than those of bulk water, indicating that hydration water is denser and narrowly defined, likely due to stronger interactions induced by the protein. The observed experimental differences between hydration water and bulk water will serve as a solid foundation for theoretical calculations to analyze the apparent chemical potential (or zeta-potential) of water near protein surfaces.

Acknowledgements

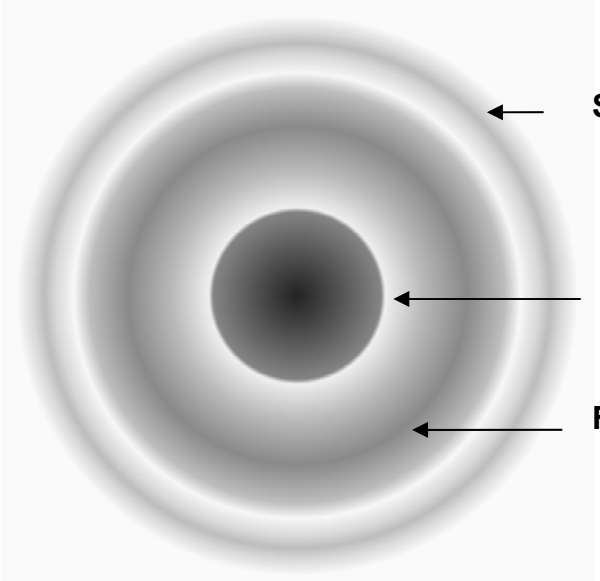
We thank Dr. Long Wang for discussing the equations. X.C. was supported in part by the Georgia State University Research Program Enhancement award. I.T.W. and R.H.W. are Georgia Cancer Coalition Distinguished Cancer Scholars. This research was supported in part by the Molecular Basis of Disease Program of Georgia State University, the Georgia Research Alliance, the Georgia Cancer Coalition, and the National Institutes of Health awards GM065762 and GM062920.

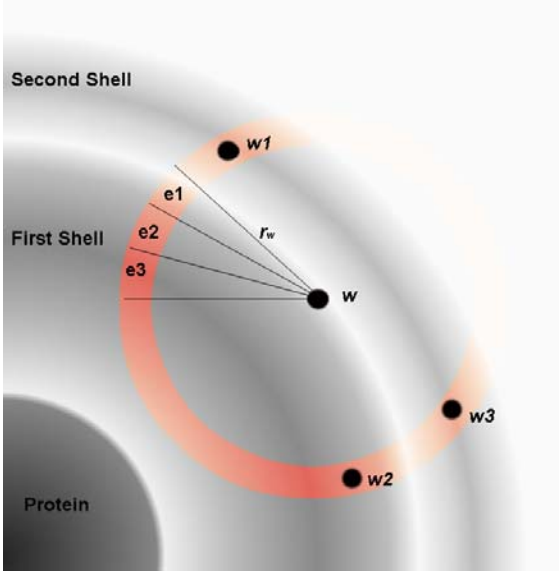
Supporting Information

105 protein crystal structures:

1a6m,1aho,1b0y,1brf,1bxo,1c75,1ceq,1ea7,1eb6,1etl,1etm,1etn,1f94,1f9y,1g66,1gkm,1gwe,1hje,1i1w,1ic6,1iee,1ir0,1iua,1k2a,1k4p,1k6u,1kwf,1m40,1mc2,1mnz,1mxt,1n1p,1n4u,1n4v,1n4w,1n55,1n9b,1ob4,1ob7,1ok0,1ot6,1ot9,1p9g,1pq5,1pwm,1rtq,1s5n,1ssx,1sy3,1tt8,1ucs,1unq,1us0,1v0l,1vb0,1vyr,1w0n,1x8p,2fdn,4lzt,7a3h,8rxn,1byi,1c7k,1cex,1exr,1g4i,1g6x,1gci,1gqv,1hj8,1hj9,1iqz,1j0p,1jfb,1k4i,1k5c,1kcc,1kth,1l9l,1lug,1m1q,1mj5,1muw,1nls,1nwz,1od3,1pjx,1pq7,1q6z,1r6j,1rb9,1s5m,1sy2,1tqg,1ufy,1ug6,1vl9,1x8q,2bf9,2erl,2pvb,3lzt,3pyp,8a3h

FIGURES


Second shell
Protein molecule
First shell


Second Shell
First Shell
Protein
w1
e1
e2
e3
r_w
w
w2
w3

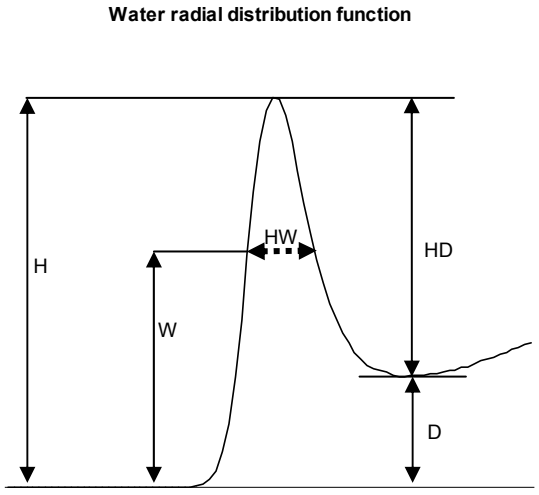
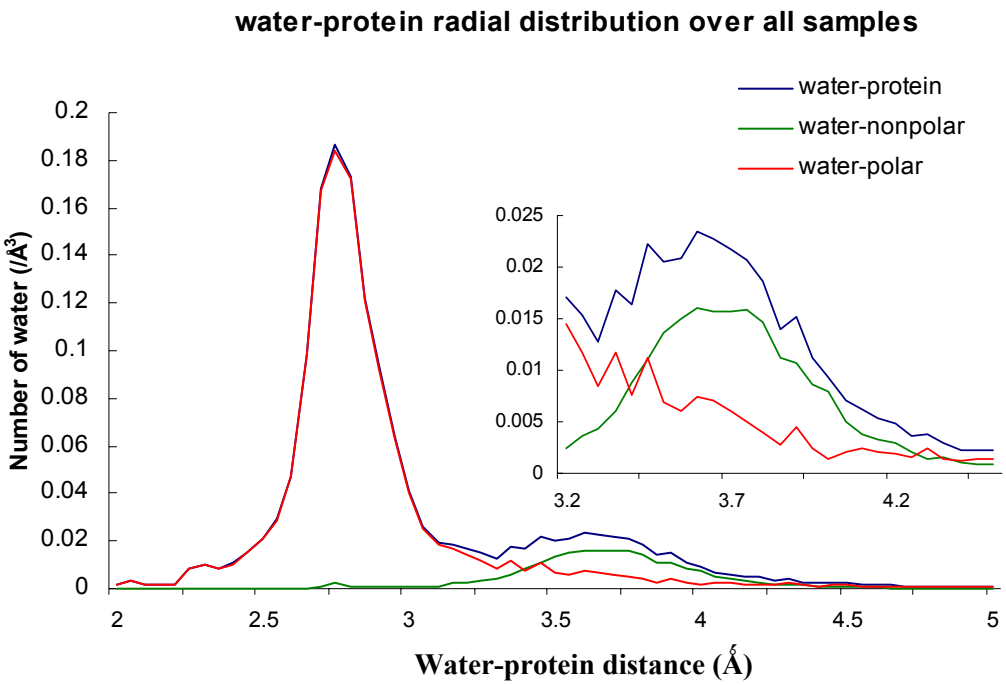


Figure 4-1 the image of water distribution on protein surface.

Figure 4-1a. The ideal model of water distributed around a protein molecule.

Figure 4-1b. Representation of terms. The RDF is the density of water as a function of distance from a particular water molecule *w*. The actual density of water in the orange shell (UDV) is the number of waters (*w1*, *w2* and *w3* represent three of the water molecules) divided by the volume of the UDV. The small cells (*e1*, *e2*, *e3* etc) are expected to have different water density because they have different distances from the protein surface. The observed density of water in the UDV is the average of the expected density of each small cell.

Figure 4-1c. The first maximum of an RDF. The sharpness of a maximum of an RDF is defined as HD/HW, where HD=H-D, HW (dashed line) is the width of the peak at height W and W= HD/2 + D.



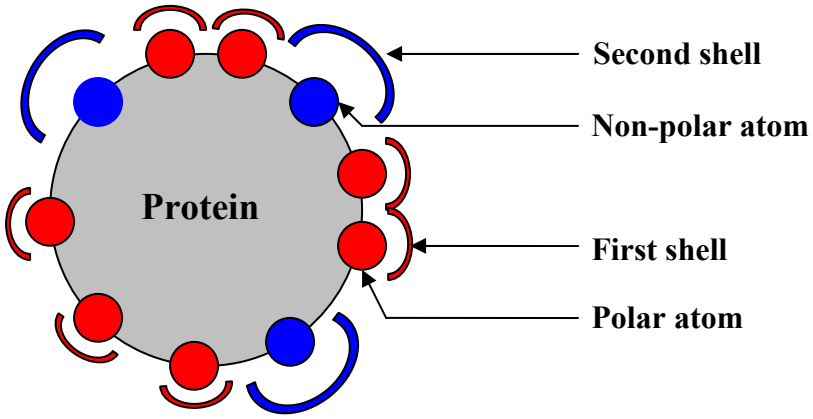
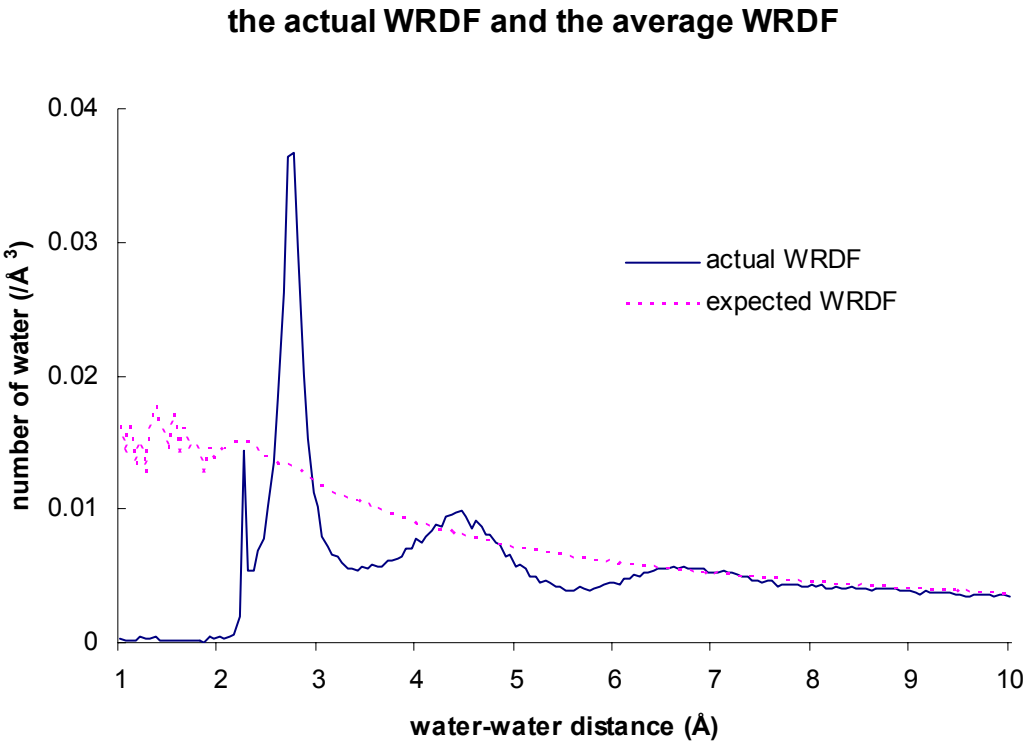


Figure 4-2 surface distribution of water.

Figure 4-2a. The SDF (blue), water-polar atom radial distribution function (SDFpol, red), and water-non polar atom radial distribution function (SDFnon, green). The number of waters is averaged over 105 crystal structures. The plots are calculated from equation (1). The inset shows the enlarged second shell.

Figure 4-2b. The model of the first and second water shells on the protein surface. Red and blue beads indicate the polar (or charged) and non-polar atoms on protein surface, respectively. The red and blue arcs represent the first and second shell, respectively. The two shells do not overlap.


the actual WRDF and the average WRDF
0.04
0.03
0.02
0.01
0
number of water (/Å 3)
actual WRDF
expected WRDF
1
2
3
4
5
6
7
8
9
10
water-water distance (Å)

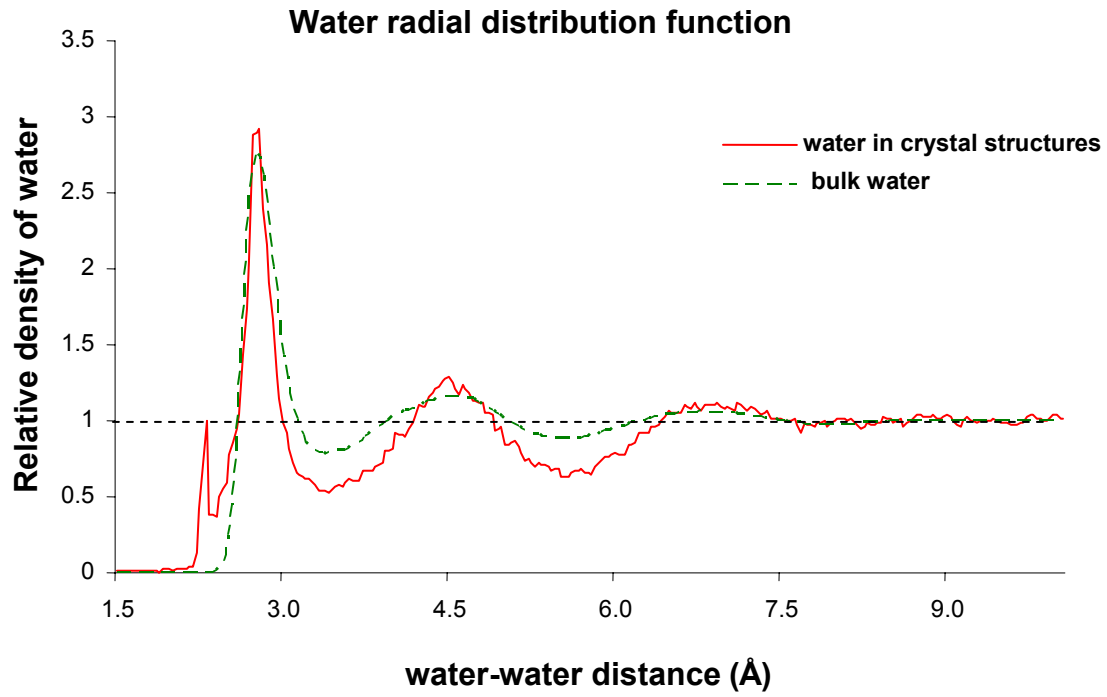
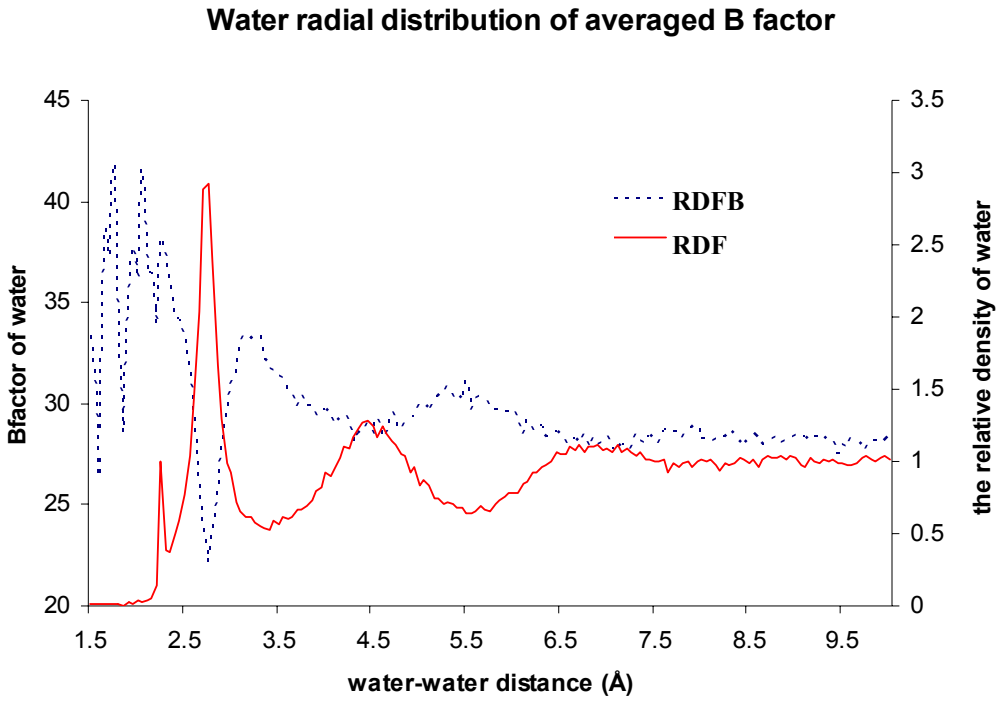

Water radial distribution function
3.5
3
2.5
2
1.5
1
0.5
0
Relative density of water
water in crystal structures
bulk water
1.5
3.0
4.5
6.0
7.5
9.0
water-water distance (Å)

Figure 4-3 Water Radial Distribution.

Figure 4-3a. The raw RDF (blue) from equation (2) and the water density distribution for normalization (pink) from equation (3) among 105 crystal structures.

Figure 4-3b. The normalized RDF (red) of water in crystal structures and the RDF of pure bulk water at 298K (green dashed line) (164).



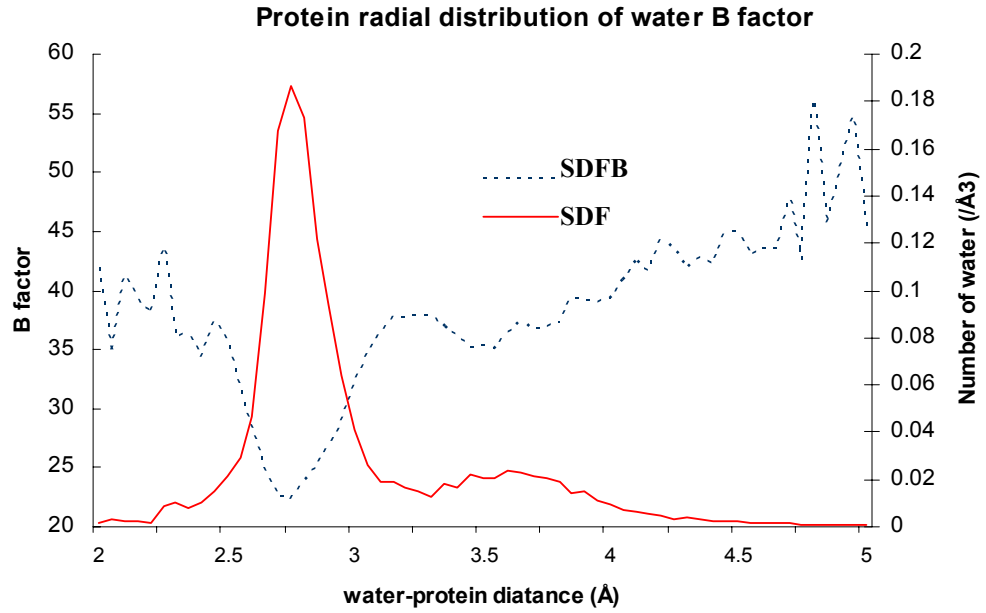


Figure 4-4 water distribution and water B factor distribution.

Figure4-4a. The normalized RDF (red line) and RDFB (dotted blue line) calculated from equation (5).

Figure4-4b. The SDF (red line) and SDFB (dotted blue line) calculated from equation (6).

Table 4-1 Comparison of Maxima and Minima in RDF calculated for protein structures and pure bulk water

	Proteins[1]			Hydration Water in Proteins[2]			Pure bulk water[3]		
	Position	Height	Sharpness	Position	Height	Sharpness	Position	Height	Sharpness
First maximum	2.76	2.92	10.08	2.76	4.04	16.95	2.73	2.75	6.35
First minimum	3.39	0.50		3.33	0.48		3.33	0.78	
Second maximum	4.47	1.29	0.84	4.47	1.48	0.99	4.53	1.16	0.27
Second minimum	5.52	0.68		5.52	0.68		5.58	0.88	
Third maximum	6.90	1.15	N/A	6.81	1.15	N/A	6.87	1.06	N/A
Forth maximum	8.85	1.05	N/A	8.82	1.05	N/A	N/A	1.0	N/A

The position (Å), height (relative density) and the sharpness of each maximum or minimum are noted.

(1) The normalized RDF from figure 4-3b (red line)
(2) The normalized RDF of "pure" hydration water
(3) The RDF of pure bulk water at 298K(164) from figure 4-3b (dashed green line)

REFERENCES

1. Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science (New York, N.Y* **277**, 1453-1474
2. Dean, M. (2002) The Human ATP-Binding Cassette (ABC) Transporter Superfamily. in *http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=mono_001.chapter.137 (*website, N. ed.
3. Jones, P. M., and George, A. M. (2004) *Cell Mol Life Sci* **61**, 682-699
4. Holland, I. B., and Blight, M. A. (1999) *Journal of molecular biology* **293**, 381-399
5. Quinton, P. M. (1999) *Physiological reviews* **79**, S3-S22
6. Andersen, D. (1938) *Amer J Dis Child*, 344-399
7. de la Salle, H., Zimmer, J., Fricker, D., Angenieux, C., Cazenave, J. P., Okubo, M., Maeda, H., Plebani, A., Tongio, M. M., Dormoy, A., and Hanau, D. (1999) *The Journal of clinical investigation* **103**, R9-R13
8. Allen, J. D., Brinkhuis, R. F., Wijnholds, J., and Schinkel, A. H. (1999) *Cancer research* **59**, 4237-4241
9. Guo, X., Chen, X., Weber, I. T., Harrison, R. W., and Tai, P. C. (2006) *Biochemistry* **45**, 14473-14480
10. Qu, Q., and Sharom, F. J. (2001) *Biochemistry* **40**, 1413-1422
11. Liu, P. Q., Liu, C. E., and Ames, G. F. (1999) *The Journal of biological chemistry* **274**, 18310-18318
12. Dawson, R. J., and Locher, K. P. (2006) *Nature* **443**, 180-185
13. Schmitt, L., Benabdelhak, H., Blight, M. A., Holland, I. B., and Stubbs, M. T. (2003) *Journal of molecular biology* **330**, 333-342
14. Hung, L. W., Wang, I. X., Nikaido, K., Liu, P. Q., Ames, G. F., and Kim, S. H. (1998) *Nature* **396**, 703-707
15. Smith, P. C., Karpowich, N., Millen, L., Moody, J. E., Rosen, J., Thomas, P. J., and Hunt, J. F. (2002) *Mol Cell* **10**, 139-149
16. Chen, J., Lu, G., Lin, J., Davidson, A. L., and Quiocho, F. A. (2003) *Molecular cell* **12**, 651-661
17. Diederichs, K., Diez, J., Greller, G., Muller, C., Breed, J., Schnell, C., Vonrhein, C., Boos, W., and Welte, W. (2000) *EMBO J* **19**, 5951-5961
18. Verdon, G., Albers, S. V., Dijkstra, B. W., Driessen, A. J., and Thunnissen, A. M. (2003) *Journal of molecular biology* **330**, 343-358
19. Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982) *EMBO J* **1**, 945-951
20. Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P., and Tainer, J. A. (2000) *Cell* **101**, 789-800
21. Zaitseva, J., Jenewein, S., Jumpertz, T., Holland, I. B., and Schmitt, L. (2005) *EMBO J* **24**, 1901-1910

22. Hunke, S., Mourez, M., Jehanno, M., Dassa, E., and Schneider, E. (2000) *The Journal of biological chemistry* **275**, 15526-15534
23. Hunke, S., Landmesser, H., and Schneider, E. (2000) *Journal of bacteriology* **182**, 1432-1436
24. Urbatsch, I. L., Gimi, K., Wilke-Mounts, S., and Senior, A. E. (2000) *Biochemistry* **39**, 11921-11927
25. Jones, P. M., and George, A. M. (1999) *FEMS microbiology letters* **179**, 187-202
26. Lamers, M. H., Winterwerp, H. H., and Sixma, T. K. (2003) *The EMBO journal* **22**, 746-756
27. Nikaido, K., and Ames, G. F. (1999) *The Journal of biological chemistry* **274**, 26727-26735
28. Davidson, A. L., and Sharma, S. (1997) *Journal of bacteriology* **179**, 5458-5464
29. Zaitseva, J., Jenewein, S., Wiedenmann, A., Benabdelhak, H., Holland, I. B., and Schmitt, L. (2005) *Biochemistry* **44**, 9680-9690
30. Higgins, C. F., and Linton, K. J. (2004) *Nat Struct Mol Biol* **11**, 918-926
31. Yuan, Y. R., Blecker, S., Martsinkevich, O., Millen, L., Thomas, P. J., and Hunt, J. F. (2001) *The Journal of biological chemistry* **276**, 32313-32321
32. Guo, X., Harrison, R. W., and Tai, P. C. (2006) *Journal of bacteriology* **188**, 2383-2391
33. Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T., and Yamaguchi, A. (2006) *Nature* **443**, 173-179
34. Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L., and Chen, J. (2007) *Nature* **450**, 515-521
35. Benabdelhak, H., Schmitt, L., Horn, C., Jumel, K., Blight, M. A., and Holland, I. B. (2005) *The Biochemical journal* **386**, 489-495
36. Akola, J., and Jones, R. (2003) *J. Phys. Chem. B* **107**
37. Gilson, L., Mahanty, H. K., and Kolter, R. (1990) *EMBO J* **9**, 3875-3884
38. Zhang, L. H., Fath, M. J., Mahanty, H. K., Tai, P. C., and Kolter, R. (1995) *Genetics* **141**, 25-32
39. Fath, M. J., Skvirsky, R. C., and Kolter, R. (1991) *J Bacteriol* **173**, 7549-7556
40. Vergani, P., Lockless, S. W., Nairn, A. C., and Gadsby, D. C. (2005) *Nature* **433**, 876-880
41. Altenberg, G. A. (2004) *Curr Med Chem Anticancer Agents* **4**, 53-62
42. Vitkup, D., Ringe, D., Petsko, G. A., and Karplus, M. (2000) *Nat Struct Biol* **7**, 34-38
43. Chen, Z., Lou, J., Zhu, C., and Schulten, K. (2008) *Biophys J*
44. Karplus, M., and McCammon, J. A. (2002) *Nat Struct Biol* **9**, 646-652
45. Cui, Y., Tran, S., Tinker, A., and Clapp, L. H. (2002) *American journal of respiratory cell and molecular biology* **26**, 135-143
46. Li, L., Wu, J., and Jiang, C. (2003) *The Journal of membrane biology* **196**, 61-69
47. Yamada, M., Isomoto, S., Matsumoto, S., Kondo, C., Shindo, T., Horio, Y., and Kurachi, Y. (1997) *The Journal of physiology* **499 (Pt 3)**, 715-720
48. Cole, W. C., and Clement-Chomienne, O. (2003) *J Cardiovasc Electrophysiol* **14**, 94-103
49. Quayle, J. M., Nelson, M. T., and Standen, N. B. (1997) *Physiological reviews* **77**, 1165-1232

50. Seino, S., and Miki, T. (2003) *Prog Biophys Mol Biol* **81**, 133-176
51. Jackson, W. F. (2005) *Microcirculation* **12**, 113-127
52. Quinn, K. V., Giblin, J. P., and Tinker, A. (2004) *Circulation research* **94**, 1359-1366
53. Shi, Y., Wu, Z., Cui, N., Shi, W., Yang, Y., Zhang, X., Rojas, A., Ha, B. T., and Jiang, C. (2007) *Am J Physiol Regul Integr Comp Physiol* **293**, R1205-1214
54. Nelson, M. T., and Quayle, J. M. (1995) *Am J Physiol* **268**, C799-822
55. Moreau, C., Prost, A. L., Derand, R., and Vivaudou, M. (2005) *J Mol Cell Cardiol* **38**, 951-963
56. Gribble, F. M., Tucker, S. J., and Ashcroft, F. M. (1997) *EMBO J* **16**, 1145-1152
57. Wang, X., Wu, J., Li, L., Chen, F., Wang, R., and Jiang, C. (2003) *Circ Res* **92**, 1225-1232
58. Wellman, G. C., Quayle, J. M., and Standen, N. B. (1998) *J Physiol* **507 (Pt 1)**, 117-129
59. Lu, T., Hoshi, T., Weintraub, N. L., Spector, A. A., and Lee, H. C. (2001) *J Physiol* **537**, 811-827
60. Yang, Y., Shi, Y., Guo, S., Zhang, S., Cui, N., Shi, W., Zhu, D., and Jiang, C. (2008) *Biochim Biophys Acta* **1778**, 88-96
61. Matsuo, M., Kimura, Y., and Ueda, K. (2005) *J Mol Cell Cardiol* **38**, 907-916
62. Schwappach, B., Zerangue, N., Jan, Y. N., and Jan, L. Y. (2000) *Neuron* **26**, 155-167
63. Mikhailov, M. V., Campbell, J. D., de Wet, H., Shimomura, K., Zadek, B., Collins, R. F., Sansom, M. S., Ford, R. C., and Ashcroft, F. M. (2005) *EMBO J* **24**, 4166-4175
64. Gruebele, M. (2002) *Curr Opin Struct Biol* **12**, 161-168
65. Lennard-Jones, J. (1931) *Proceedings of the Physical Society* **43**, 21
66. Tersoff, J. (1988) *Phys Rev B Condens Matter* **37**, 6991-7000
67. Born, M., and OppenheiMer, J. (1927) *Ann. Physik* **84**
68. Ogasawara, H., Brena, B., Nordlund, D., Nyberg, M., Pelmenschikov, A., Pettersson, L. G., and Nilsson, A. (2002) *Phys Rev Lett* **89**, 276102
69. Solovei, A. B., and Lobyshev, V. I. (2006) *Russian Journal of Physical Chemistry* **80**, 1578-1583
70. Huang, X., Braams, B. J., and Bowman, J. M. (2006) *The journal of physical chemistry* **110**, 445-451
71. Hajdu, F. (1977) *Acta Chim. (Budapest)* **93**, 24
72. Müller, A., Bögge, H., and Diemann, E. (2003) *Inorg. Chem. Commun* **6**, 52
73. Israelachvili, J., and Wennerstrom, H. (1996) *Nature* **379**, 219-225
74. Breslow, R. (1991) *Accounts of Chemical Research* **24**, 159-164
75. Rowe, A. J. (2001) *Biophys Chem* **93**, 93-101
76. Mahoney, M. W., and Jorgensen, W. L. (2000) *J. Chem. Phys* **112**
77. Mahoney, M. W., and Jorgensen, W. L. (2000). *Chem. Phys* **112**
78. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. v., and Hermans, J. (1981) *Intermolecular Forces* Reidel, Dordrecht
79. Cornell, W., Cieplak, P., Bayly, C., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc* **117**

80. Brooks BR, B. R., Olafson BD, States DJ, Swaminathan S, Karplus M (1983) *J Comp Chem* **4**
81. Corzana, F., Motawia, M. S., Du Penhoat, C. H., Perez, S., Tschampel, S. M., Woods, R. J., and Engelsen, S. B. (2004) *J Comput Chem* **25**, 573-586
82. Brown, C. R., Hong-Brown, L. Q., and Welch, W. J. (1997) *The Journal of clinical investigation* **99**, 1432-1444
83. Hoffman-Goetz, L., and Kluger, M. J. (1979) *Am J Clin Nutr* **32**, 1423-1427
84. Chiu, W. L., Szajner, P., Moss, B., and Chang, W. (2005) *J Virol* **79**, 8046-8056
85. Vogeser, M., and Briegel, J. (2007) *Clin Biochem* **40**, 724-727
86. Leeson, D. T., and Wiersma, D. A. (1995) *Nat Struct Biol* **2**, 848-851
87. Schlichter, J., Friedrich, J., Herenyi, L., and Fidy, J. (2001) *Biophys J* **80**, 2011-2017
88. Trabbic-Carlson, K., Meyer, D. E., Liu, L., Piervincenzi, R., Nath, N., LaBean, T., and Chilkoti, A. (2004) *Protein Eng Des Sel* **17**, 57-66
89. Loo, T. W., Bartlett, M. C., and Clarke, D. M. (2002) *J Biol Chem* **277**, 41303-41306
90. Hagen, S. J., Carswell, C. W., and Sjolander, E. M. (2001) *Journal of molecular biology* **305**, 1161-1171
91. Huang, D. M., and Chandler, D. (2000) *Proc Natl Acad Sci U S A* **97**, 8324-8327
92. Kodama, H., Kodama, Y., Shinozawa, S., Kanemaru, R., Todaka, K., and Mitsuyama, Y. (1999) *Eur J Pharm Biopharm* **47**, 295-298
93. Qu, Q., Chu, J. W., and Sharom, F. J. (2003) *Biochemistry* **42**, 1345-1353
94. Meersman, F., and Heremans, K. (2003) *Biochemistry* **42**, 14234-14241
95. Wu, D. H., Laidman, D. L., and Smith, C. J. (1993) *Journal of Experimental Botany* **44**, 457-461
96. Copty, A., Sakran, F., Popov, O., Ziblat, R., Danieli, T., Golosovsky, M., and Davidov, D. (2005) *Synthetic Metals* **155**, 422-425
97. Pang, x.-f., and Xu, C.-t. (2001) *International Journal of Infrared and Millimeter Waves* **22**
98. Suskind, R. G. (1967) *J Cell Biol* **34**, 721-734
99. Each, O. C. C. (1999) *CURRENT SCIENCE* **77**, 855
100. Harrison, R. W., and Weber, I. T. (1994) *Protein engineering* **7**, 1353-1363
101. Chen, X., Weber, I. T., and Harrison, R. W. (2004) *J Mol Model* **10**, 373-381
102. Harrison, R. W. (2003) Amortized fast multipole algorithm for molecular modeling. in *Proceedings of the International Conference on Computer Science and its Applications*, Gatton TM (eds) Dey PP, Amin MN
103. Caves, L. S., Evanseck, J. D., and Karplus, M. (1998) *Protein Sci* **7**, 649-666
104. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein engineering* **8**, 127-134
105. Quayle, J. M., Nelson, M. T., and Standen, N. B. (1997) *Physiol Rev* **77**, 1165-1232
106. Seino, S., and Miki, T. (2003) *Prog Biophys Mol Biol* **81**, 133-176
107. Jackson, W. F. (2005) *Microcirculation* **12**, 113-127
108. Quinn, K. V., Giblin, J. P., and Tinker, A. (2004) *Circ Res* **94**, 1359-1366
109. Shi, Y., Wu, Z., Cui, N., Shi, W., Yang, Y., Zhang, X., Rojas, A., Ha, B. T., and Jiang, C. (2007) *Am J Physiol Regul Integr Comp Physiol* **293**, R1205-1214

110. Nelson, M. T., and Quayle, J. M. (1995) *Am J Physiol* **268**, C799-822
111. Shyng, S., and Nichols, C. G. (1997) *J Gen Physiol* **110**, 655-664
112. Yamada, M., Isomoto, S., Matsumoto, S., Kondo, C., Shindo, T., Horio, Y., and Kurachi, Y. (1997) *J Physiol* **499 (Pt 3)**, 715-720
113. Cui, Y., Tran, S., Tinker, A., and Clapp, L. H. (2002) *Am J Respir Cell Mol Biol* **26**, 135-143
114. Cole, W. C., and Clement-Chomienne, O. (2003) *J Cardiovasc Electrophysiol* **14**, 94-103
115. Moreau, C., Prost, A.-L., Derand, R., and Vivaudou, M. (2005) *J Mol Cell Cardiol* **38**, 951-963
116. Gribble, F. M., Tucker, S. J., and Ashcroft, F. M. (1997) *Embo J* **16**, 1145-1152
117. Wang, X., Wu, J., Li, L., Chen, F., Wang, R., and Jiang, C. (2003) *Circ Res* **92**, 1225-1232
118. Yang, Y., Shi, Y., Guo, S., Zhang, S., Cui, N., Shi, W., Zhu, D., and Jiang, C. (2007) *Biochim Biophys Acta*
119. Higgins, C. F. (1992) *Annu Rev Cell Biol* **8**, 67-113
120. Gaudet, R., and Wiley, D. C. (2001) *Embo J* **20**, 4964-4972
121. Yamada, M., and Kurachi, Y. (2004) *Mol Pharmacol* **65**, 1198-1207
122. Shyng, S., Ferrigni, T., and Nichols, C. G. (1997) *J Gen Physiol* **110**, 643-654
123. Yamada, M., Ishii, M., Hibino, H., and Kurachi, Y. (2004) *Mol Pharmacol* **66**, 807-816
124. Huopio, H., Reimann, F., Ashfield, R., Komulainen, J., Lenko, H. L., Rahier, J., Vauhkonen, I., Kere, J., Laakso, M., Ashcroft, F., and Otonkoski, T. (2000) *J Clin Invest* **106**, 897-906
125. Matsuo, M., Dabrowski, M., Ueda, K., and Ashcroft, F. M. (2002) *Embo J* **21**, 4250-4258
126. Mikhailov, M. V., Campbell, J. D., de Wet, H., Shimomura, K., Zadek, B., Collins, R. F., Sansom, M. S. P., Ford, R. C., and Ashcroft, F. M. (2005) *Embo J* **24**, 4166-4175
127. Babenko, A. P., and Bryan, J. (2003) *J Biol Chem* **278**, 41577-41580
128. Chan, K. W., Zhang, H., and Logothetis, D. E. (2003) *Embo J* **22**, 3833-3843
129. Hollenstein, K., Frei, D. C., and Locher, K. P. (2007) *Nature* **446**, 213-216
130. Locher, K. P., Lee, A. T., and Rees, D. C. (2002) *Science* **296**, 1091-1098
131. Pinkett, H. W., Lee, A. T., Lum, P., Locher, K. P., and Rees, D. C. (2007) *Science* **315**, 373-377
132. Bienengraeber, M., Alekseev, A. E., Abraham, M. R., Carrasco, A. J., Moreau, C., Vivaudou, M., Dzeja, P. P., and Terzic, A. (2000) *FASEB J* **14**, 1943-1952
133. Campbell, J. D., Proks, P., Lippiat, J. D., Sansom, M. S. P., and Ashcroft, F. M. (2004) *Diabetes* **53 Suppl 3**, S123-127
134. Campbell, J. D., Sansom, M. S. P., and Ashcroft, F. M. (2003) *EMBO Rep* **4**, 1038-1042
135. Zolnerciks, J. K., Wooding, C., and Linton, K. J. (2007) *FASEB J*
136. Chang, G. (2003) *J Mol Biol* **330**, 419-430
137. Light, P. E., Manning Fox, J. E., Riedel, M. J., and Wheeler, M. B. (2002) *Mol Endocrinol* **16**, 2135-2144

138. Purkiss, A. (2001) *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **359**, 1515-1527
139. Roberts, B. C., and Mancera, R. L. (2008) *J Chem Inf Model*
140. Nakasako, M. (1999) *J Mol Biol* **289**, 547-564
141. Higo, J., and Nakasako, M. (2002) *J Comput Chem* **23**, 1323-1336
142. Rashin, A. A., Iofin, M., and Honig, B. (1986) *Biochemistry* **25**, 3619-3625
143. Meyer, E. (1992) *Protein Sci* **1**, 1543-1562
144. Raymer, M. L., Sanschagrin, P. C., Punch, W. F., Venkataraman, S., Goodman, E. D., and Kuhn, L. A. (1997) *J Mol Biol* **265**, 445-464
145. Raschke, T. M. (2006) *Current opinion in structural biology* **16**, 152-159
146. Rupley, J. A., and Careri, G. (1991) *Adv Protein Chem* **41**, 37-172
147. Svergun, D. I., Richard, S., Koch, M. H., Sayers, Z., Kuprin, S., and Zaccai, G. (1998) *Proceedings of the National Academy of Sciences of the United States of America* **95**, 2267-2272
148. Cameron, I. L., Short, N. J., and Fullerton, G. D. (2007) *Cell biology international* **31**, 531-539
149. Ebbinghaus, S., Kim, S. J., Heyden, M., Yu, X., Heugen, U., Gruebele, M., Leitner, D. M., and Havenith, M. (2007) *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20749-20752
150. McLain, S. E., Soper, A. K., and Watts, A. (2008) *Eur Biophys J*
151. Seki, Y., Tomizawa, T., Khechinashvili, N. N., and Soda, K. (2002) *Biophys Chem* **95**, 235-252
152. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and Klein, M. L. (1983) *J. Chem. Phys* **79**, 10
153. Jorgensen, W. L., and Madura, J. D. (1985) *Mol. Phys* **56**, 11
154. Dadarlat, V. M., and Post, C. B. (2006) *Biophysical journal* **91**, 4544-4554
155. Kuttel, M., Brady, J. W., and Naidoo, K. J. (2002) *Journal of computational chemistry* **23**, 1236-1243
156. Basma, M., Sundara, S., Calgan, D., Vernali, T., and Woods, R. J. (2001) *Journal of computational chemistry* **22**, 1125-1137
157. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. Wiley
158. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. American Chemical Society
159. Berman, H., Henrick, K., and Nakamura, H. (2003) *Nature structural biology* **10**, 980
160. Botti, A., Bruni, F., Isopo, A., Ricci, M. A., and Soper, A. K. (2002) Experimental determination of the site–site radial distribution functions of supercooled ultrapure bulk water. American Institute of Physics
161. Petrenko, V. F., and Whitworth, R. W. (1999) *Physics of Ice*, Oxford University Press, Oxford
162. Head-Gordon, T. (1995) *Proceedings of the National Academy of Sciences of the United States of America* **92**, 8308-8312

163. Sorenson, J. M., Hura, G., Glaeser, R. M., and Head-Gordon, T. (2000) What can x-ray scattering tell us about the radial distribution functions of water?
164. Soper, A. K., Rossky, P. J., and Eds. (2000) *Chem. Phys.* **258** 121-137
165. Harrison, R. W., Kourinov, I. V., and Andrews, L. C. (1994) *Protein Eng* **7**, 359-369
166. Vihinen, M., Torkkila, E., and Riikonen, P. (1994) *Proteins* **19**, 141-149
167. Vihinen, M. (1987) *Protein engineering* **1**, 477-480
168. Parthasarathy, S., and Murthy, M. R. (2000) *Protein engineering* **13**, 9-13
169. Harding, M. M. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 872-874
170. Nayal, M., and Di Cera, E. (1996) *J Mol Biol* **256**, 228-234
171. Kovalevsky, A. Y., Tie, Y., Liu, F., Boross, P. I., Wang, Y. F., Leshchenko, S., Ghosh, A. K., Harrison, R. W., and Weber, I. T. (2006) *Journal of medicinal chemistry* **49**, 1379-1387
172. Sherman, D. M., and Collings, M. D. (2002) Ion association in concentrated NaCl brines from ambient to supercritical conditions: results from classical molecular dynamics simulations.
173. Tromp, R. H., Neilson, G. W., and Soper, A. K. (1992) Water structure in concentrated lithium chloride solutions. AIP
174. Hura, G., Russo, D., Glaeser, R. M., Head-Gordon, T., Krack, M., Parrinello, M., and Matter, S. (2003) Water structure as a function of temperature from X-ray scattering experiments and ab initio molecular dynamics.
175. Puibasset, J., and Pellenq, R. J. (2003) *Eur Phys J E Soft Matter* **12 Suppl 1**, 67-70