PROBING HYDROGEN BONDING INTERACTIONS

AND PROTON TRANSFER IN PROTEINS

By

BEINING NIE

Bachelor of Science
Beijing University of Aeronautics & Astronautics
Beijing, China
1997

Master of Science
Oklahoma State University
Stillwater, OK, USA
2002

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2006

PROBING HYDROGEN BONDING INTERACTIONS

AND PROTON TRANSFER IN PROTEINS

Thesis Approved:

Dr. Aihua Xie
_____
Thesis Adviser

Dr. Robert H. Hauenstein
_____

Dr. Benjamin H. McMahon
_____

Dr. Wouter D. Hoff
_____

Dr. Gordon Emslie
_____
Dean of the Graduate College

ACKNOWLEDGEMENTS

The completion of this dissertation marks the end of my education. In the past few years, there were numerous supports, encouragements, and love without which I would not be able to come to this far. At this moment, I would like to take this opportunity to express my sincere gratitude to those who have supported and loved me all along.

To begin, I thank my parents for their endless love throughout my life. Their caring and full support always extends to the deepest of my heart. I would not be able to achieve this milestone of my life without expressing my sincere gratitude to my parents.

Especially, I would like to express my deep appreciation to my family. My husband, Jianning Wang, has accompanied and encouraged me all along. He had supported me in many aspects. In addition, I thank my two baby girls Amy and Alice for bring incomparable joys to my life. My family's deep love was the driving force for the write-up of this dissertation.

I would like to thank Dr. Robert Hauenstein, Dr. Wouter Hoff, and Dr. Ben McMahon (alphabetically ordered) for serving on my committee and reviewing this thesis. I had worked with Dr. Hauenstein on a research project for more than a year. His eagerness of learning biophysics, sound knowledge of physics, and professional spirits are remarkable. I know Wouter since I came to the United

States. I am very amazed by his kindness, broad knowledge of biochemistry, and his spirits of living healthy. I had worked with Ben for two summers at Los Alamos National Lab on molecular dynamics simulations of proteins in explicit and inexplicit solvents. He taught me from very basic Linux commands to complicated simulation techniques. His talents and full understanding of biophysics deeply impressed me. These three scientists are very nice, intelligent, and professional. I am greatly honored to have them on my committee.

I wish to extend special thanks for the friendship, help and support of many faculty members, lab mates, graduate students, department secretaries and support staff, especially including Dr. Paul Westhaus, Dr. Jacques Perk, Dr. Bret Flanders, Lorand Kelemen, Qinan Bao, Jarmila Guijarro, Anu Thubagere, Ye Xiong, Kai Wang, Zhenyue Zhu, Susan Cantrell, Cindi Raymond, Stephanie Hall, and Warren Grider.

Last and foremost, I would like to thank Dr. Aihua Xie for being my advisor. She is very thoughtful, intelligent, and professional. She always cared about me, gave me guidance and advice, and encouraged and supported me in countless circumstances. Without her, I would not have completed my Ph.D. degree with *Outstanding Computational Research Assistant Award* and I would not have achieved so much in research. I appreciate all the opportunities and assistance she had provided me over the past few years. She exemplified a successful and accomplished female professor and demonstrated her independence and capability in science. Her numerous guidance and advice

have deeply enhanced my chances for success at OSU and my subsequent career and life. Thank you, Dr. Xie!

TABLE OF CONTENTS

LIST of TABLES

LIST of FIGURES

CHAPTER I

INTRODUCTION

1.1 Background

Proton transfer

Proton transfer is an essential process in life and plays a crucial role in biological processes, such as enzyme catalysis, generating proton gradients in membrane-bound proteins, ATP synthesis, certain forms of cell motility, accumulation of neurotransmitters, and acid/base equilibrium. There are many examples of such proton transfer processes. In bacteriorhodopsin, a well-studied light-driven proton pump protein for solar energy transduction, protonated Schiff base of the retinal chromophore and Asp85 form an ion pair, which is the center of hydrogen bonding network comprise of polar amino acid side chains and water molecules surrounded by seven helical strands of the membrane protein. After absorbing light, all-*trans* to 13-*cis* photoisomerization of the retinal chromophore takes place, followed by the primary proton transfer from the Schiff base to Asp85 that triggers sequential proton transfer reactions for the proton pump from the cytoplasmic side to extracellular side (Heberle, 2000; Kandori, 2004; Lanyi, 2004). In photoactive yellow protein (PYP), negatively charged chromophore is stabilized at the active site via hydrogen bonding interactions with Glu46, Tyr42, and Cys69 (Borgstahl et al., 1995). After light absorption, *trans* to *cis*

1

Photoisomerization of the chromophore occur, followed by proton transfer from Glu46 to the chromophore that the resulting negative charge on Glu46 triggers and drives large conformational changes (Xie et al., 1996; Xie et al., 2001). In the structures of the photosynthetic reaction center (Adelroth and Brzezinski, 2004; Paddock et al., 2003), the *bc*1 complex (Crofts et al., 2006; Hunte et al., 2003), and Cytochrome *c* Oxidase (Adelroth and Brzezinski, 2004; Namslauer and Brzezinski, 2004), pathways for proton transfer have been identified that within the proteins, protons move rapidly through hydrogen-bonded chains of water and amino acid residues, presumably via a hopping mechanism (Nagle and Tristram-Nagle, 1983). A mechanism was proposed that there is a sequential, isoenergetic exchange of hydrogen bonds and covalent bonds, which allows a proton to be added to one end of a chain of waters while another proton is released from the other end.

Hydrogen bonding interactions

Hydrogen bonds are formed between a hydrogen acceptor (an electronegative atom, usually oxygen or nitrogen with a lone pair of electrons) and a hydrogen atom covalently bonded to another electronegative atom (a hydrogen donor). Figure 1.1 shows various types of interactions that involve hydrogen bonds. The length between two electronegative atoms in hydrogen bonding interactions ranges from 2.7 Å to 3.2 Å (Bergethon, 1998; Vinogradov and Linnell, 1971). This length may be shorter for strong hydrogen bonds or longer for weak hydrogen bonds. The angle of X–H $\cdots$ X is in a range of 180° ± 20°. Hydrogen bonds are found in biological systems for structural stability. For

instance, α Helix structure is considerably stabilized by many internal hydrogen bonds each of which connects the hydrogen atom on the nitrogen atom of the peptide bond of one amino acid and the carbonyl oxygen atom of the fourth amino acid's peptide bond. β Sheet structure is formed by many hydrogen bonding interactions between adjacent polypeptide chains (Nelson and Cox, 2000). Hydrogen bonds are also formed between amino acid side chains.
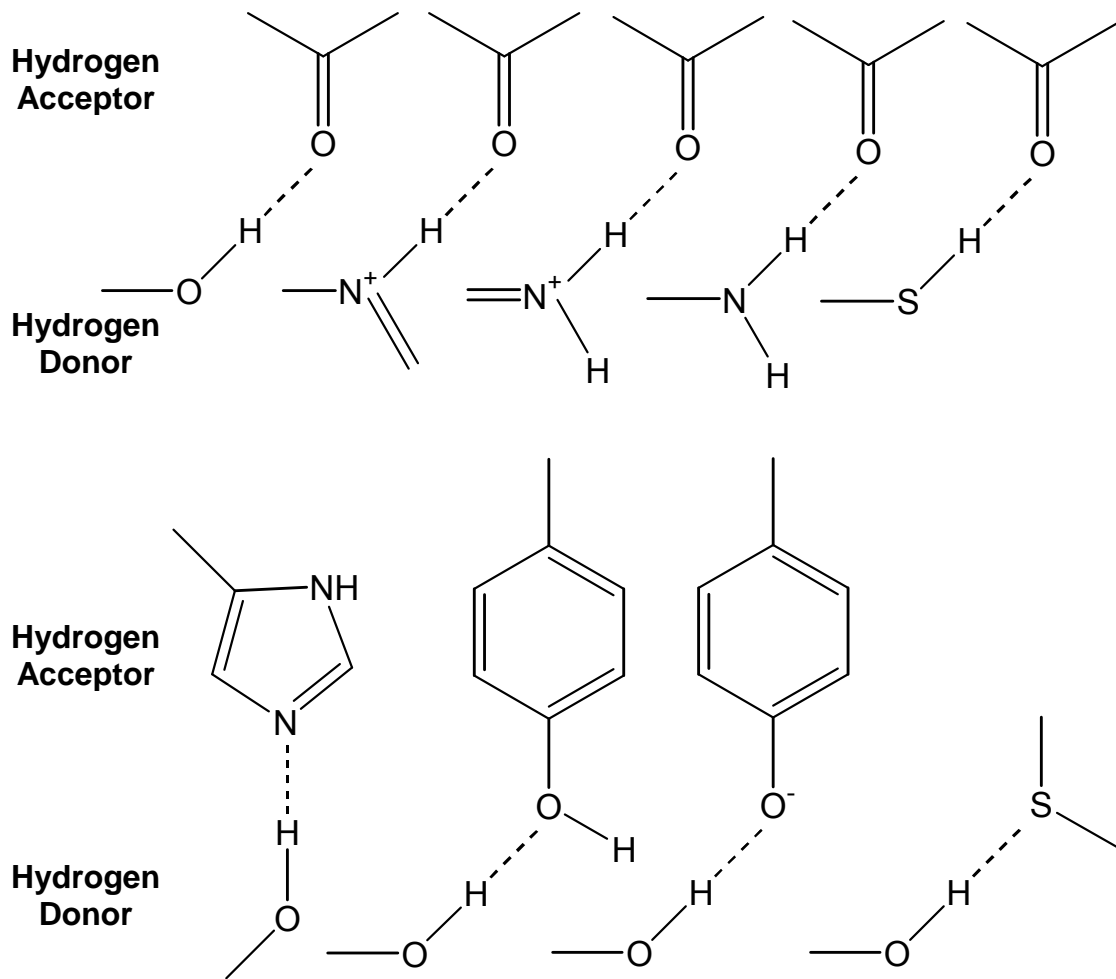
Figure 1.1 Hydrogen bonds in biological system. Dashed lines represent Hydrogen bonds. The hydrogen acceptor is usually oxygen or nitrogen atom (Nelson and Cox, 2000).

Hydrogen bonds are weaker than covalent bonds. The Hydrogen bond dissociation energy is about 10 – 40 kJ/mol (Nelson and Cox, 2000). We take a water molecule as an example to see how Hydrogen bond is formed. The oxygen atom of a water molecule forms two covalent bonds with two hydrogen atoms by sharing two pairs of electrons, respectively. This results in two pairs of unshared electrons. Since oxygen is more electronegative than hydrogen, the oxygen nucleus attracts electrons more strongly than does the hydrogen nucleus. The sharing of electrons between H and O is therefore unequal. The electrons are more often in the vicinity of the oxygen atom than of the hydrogen. The result of this unequal sharing of electrons is that the oxygen atom bears a partial negative charge and each hydrogen atom bears a partial positive charge. As a result, there is an electrostatic attraction between the oxygen atom of one water molecule and the hydrogen atom of the other. This attraction and partial sharing of electrons lead to the formation of Hydrogen bond. Hydrogen bonds are individually weak, but collectively they have a very significant influence on the three-dimensional structure of proteins (Nelson and Cox, 2000). In general, hydrogen bonding interaction(s) is central to all biological processes, from stabilizing of secondary structures such as the α-helix or β-sheet and driving proton transfer in proteins to the reversible bond formation that allows the action of enzymes.

Ab initio methods

The term *ab initio* is Latin for "from the beginning." This name is given to computations that are derived directly from theoretical principles with no inclusion

of experimental data. *Ab Initio* methods are based on the fundamental laws of quantum mechanics and employ a variety of mathematical transformation and approximation techniques such as using a combination of gaussian functions to solve the fundamental equations or finding an approximate solution to a differential equation. Two different quantum theories are applied in *ab initio* approaches to the many-electron systems: the Hartree-Fock theory and the density functional theory.

The time-independent Schrödinger equation describing the wavefunction of a particle has the general form:

$$H\psi(\vec{r}) = E\psi(\vec{r})$$ (1.1)

where H is the Hamiltonian operator, E is the eigenenergy of the particle, ψ is the eigenwavefunction (Shankar, 1994). The Hamiltonian is made up of kinetic and potential energy. The Hamiltonian of a system consisting of N electrons and K nuclei with charges Z_n reads (Thijssen, 1999):

$$H = \sum_{i=1}^{N}\frac{p_i^2}{2m} + \sum_{n=1}^{K}\frac{P_n^2}{2M_n} + \frac{1}{4\pi\varepsilon_0}\frac{1}{2}\sum_{i=1,i\neq j}^{N}\sum_{j=1}^{N}\frac{e^2}{\left|\vec{r}_i - \vec{r}_j\right|} - \frac{1}{4\pi\varepsilon_0}\sum_{n=1}^{K}\sum_{i=1}^{N}\frac{Z_n e^2}{\left|\vec{r}_i - \vec{R}_n\right|} + \frac{1}{4\pi\varepsilon_0}\frac{1}{2}\sum_{\substack{n,n'=1;\\n\neq n'}}^{K}\frac{Z_n Z_{n'} e^2}{\left|\vec{R}_n - \vec{R}_{n'}\right|}$$
(1.2)

where index *i* refers to the electrons and *n* to the nuclei, m is the electron mass, and M_n are the masses of the different nuclei. The first two terms represent the kinetic energies of the electrons and nuclei respectively; the third term represents the Coulomb repulsion between the electrons and the fourth term represents the Coulomb attraction between electrons and nuclei, and the last term represents the Coulomb repulsion between the nuclei. It is impossible to solve this equation analytically for any system that contains two or more electrons. Therefore,

adequate approximations must be made. The first approximation is Born-Oppenheimer approximation, which separates electron and nuclear motion based on the idea that the nuclei move much more slowly than the electrons since the nuclei are much heavier than the electrons. Then the Hamiltonian for electrons has the following form:

$$H = \sum_{i=1}^{N} \frac{p_i^2}{2m} + \frac{1}{4\pi\varepsilon_0} \frac{1}{2} \sum_{i=1,i\neq j}^{N} \sum_{j=1}^{N} \frac{e^2}{\left|\vec{r}_i - \vec{r}_j\right|} - \frac{1}{4\pi\varepsilon_0} \sum_{n=1}^{K} \sum_{i=1}^{N} \frac{Z_n e^2}{\left|\vec{r}_i - \vec{R}_n\right|} \qquad (1.3)$$

where the first term is the kinetic energy of the electrons, the second term is the Coulomb repulsion between the electrons, and the third term is the Coulomb attraction between electrons and nuclei. The electrostatic energy of the nuclei should be added to the energy of the electrons to arrive at the total energy. Although the Hamiltonian (1.3) is much more simplified, it remains intractable because of the second term containing the interactions between the electrons. Hartree-Fock theory is a basic method for approximating the eigenfunctions of the Hamiltonian (1.3).

Hartree-Fock theory or self-consistent field method was first proposed by Douglas Hartree to find the approximate wave functions of a multi-electron atom and then modified by Vladimir Fock (Bergethon, 1998). There are two approximations in Hartree-Fock theory. The first approximation is the center field approximation, which means that the Coulombic electron-electron repulsion in equation (1.3) is taken into account by integrating the repulsion term. This gives the average effect of the repulsion, but not the explicit repulsion interaction (Young, 2001). The second approximation involves expressing the wavefunction

as linear combinations of a pre-defined set of one-electron functions know as basis functions. The functions used most often are linear combinations of Gaussian-type orbitals (Bergethon, 1998; Young, 2001). Gaussian functions have the general form:

$$g(\alpha, \vec{r}) = c x^n y^m z^l e^{-\alpha r^2} \qquad (1.4)$$

where $\vec{r}$ is composed of x, y, z. The quantities n, m, and l are integers. α is a constant determining the size of the function (Foresman and Frisch, 1996).

Density functional theory (DFT) has become very popular in recent years. It originated from a theorem by Hoenburg and Kohn stating that the energy of a molecule can be determined from the electron density instead of a wavefunction (Hohenberg and Kohn, 1964). Following on the work of Kohn and Sham, the approximate functionals employed by current DFT methods partition the electronic energy into four terms:

$$E = E^T + E^V + E^J + E^{XC} \qquad (1.5)$$

where E^T is the kinetic energy term of the electrons, E^V includes the Coulomb repulsion between pairs of nuclei and attraction between nuclear and electron, E^J is the Coulomb repulsion between the electrons, and E^{XC} is the exchange-correlation term including the exchange energy arising from the antisymmetry of the quantum mechanical wavefunction and dynamic correlation in the motions of the individual electrons. All terms except the nuclear-nuclear repulsion are functions of the electron density (Foresman and Frisch, 1996). The advantage of using electron density is that the integrals for Coulomb repulsion need be done only over the electron density and at least some electron correlation can be

included in the calculation (Young, 2001). The B3LYP method that used for this thesis is a hybrid calculation including Hartree-Fock exchange and DFT exchange-correlation functionals. B3LYP indicates that E^{XC} uses Becke's three parameters and LYP correlation correction (Becke, 1993; Foresman and Frisch, 1996; Lee et al., 1988).

The commercial software Gaussian03 was utilized to perform the *ab initio* calculations including geometry optimization, vibrational frequency, electrostatic field, charge distribution, and system energy. There are many semi-empirical and *ab initio* methods available in Gaussian03. Semi-empirical methods in Gaussian03 can be applied to systems where parameters have been developed for all of their component atoms and are not as accurate as appropriate *ab initio* methods. Only B3LYP method was employed in our computational studies for its accuracy of calculating vibrational frequencies (Nie, 2002). For proton affinities studies, we had used Semi-empirical methods and HF methods because of the large size of the model system and limitation of computer power.

Vibrational spectral marker for probing hydrogen-bonding status

Intramolecular vibration between two covalently bonded atoms is treated as harmonic oscillator in quantum mechanics, where its force constants can be obtained from *ab initio* calculations. We know that the time-independent Schrödinger equation in the energy basis for harmonic oscillators is as follows (Eisberg and Resnick, 1985):

$$-\frac{\hbar^2}{2m}\frac{d^2\varphi}{dx^2} + \frac{C}{2}x^2\varphi = E\varphi \qquad (1.6)$$
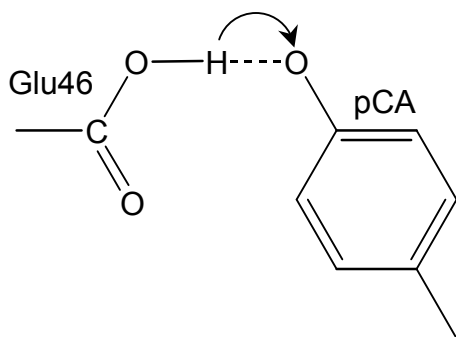
Figure 1.2 The pCA chromophore of PYP forms a hydrogen bond with Glu46 side chain.

where $\hbar$ is Planck's constant, m is the mass of the atom, φ is the wavefunction, x is the displacement of the atom, C is the force constant, and E is the eigenenergy. By solving this equation, we found that the eigenenergy has discrete values:

$$E_n = (n+\frac{1}{2})h\nu \text{ , n = 0,1,2,3,...} \qquad (1.7)$$

The vibrational frequency ν is proportional to the square root of force constant $\nu = \frac{1}{2\pi}\sqrt{\frac{C}{m}}$, indicating that stronger bonds have a larger force constant and vibrate at higher frequencies than weaker bonds and bonds between atoms of higher masses vibrate at lower frequencies than bonds between lighter atoms. In general, triple bonds are stronger than double or single bonds and exhibit higher vibrational frequencies, and double bonds are stronger than single bonds and display higher vibrational frequencies than that of single bonds. When a hydrogen bond is formed between a COOH group and a phenolate group (see Figure 1.2), the proton on the Glu46 was pulled toward the phenolate oxygen of pCA due to the hydrogen bonding interaction; thus weakens the O–H bond of Glu46 and its force constant becomes smaller than that of O–H without any hydrogen bond. As a result, the frequency of O–H stretching will be lower. In Chapter II, our computational studies show that the C=O stretching frequency is also down–shifted. This is probably because the stronger structural resonance of

the COOH group when the small hydrogen atom was pulled toward the phenolate oxygen atom via hydrogen bonding interaction.

Some of the vibrational frequencies are known to be sensitive to hydrogen bonding interaction(s), e.g. O−H and C=O stretching frequency of the COOH group. However, it lacks quantitative studies to establish a strong and clear correlation between these vibrational frequencies and the nature of hydrogen-bonding interactions of the amino acid side−chain groups of interest. Therefore, it is necessary for us to perform extensive computational studies that supported by experimental evidence to first search for such vibrational frequencies that are sensitive to hydrogen bonding interaction in a unique infrared spectroscopic region (not overlapping with other infrared bands); then establish vibrational spectral marker(s) for probing the number and type of hydrogen bonding interactions of key residues buried in proteins; eventually utilize the vibrational spectral marker(s) to get insights into the hydrogen-bonding status and local environment of key residues during functional processes of proteins that their crystal structures are not yet available.

Photoactive yellow protein: An excellent model system

PAS domain proteins encompass hundreds of receptor and regulatory proteins found from all three branches of life, including a human potassium channel for cardiac function. Photoactive yellow protein (PYP) is a prototype PAS domain photoreceptor (Figure 1.3 A). It was identified as a blue light photoreceptor by spectroscopic techniques and responsible for the negative phototaxis response of *Ectothiorhodospira halophila*, a halophilic purple

phototrophic bacterium (Meyer, 1985; Meyer et al., 1987). Unlike other photoreceptor proteins such as rhodopsin or bacteriorhodopsin, PYP binds a unique anionic *p*-coumaric (pCA, or 4-hydroxycinnamic acid) chromophore (Baca et al., 1994; Hoff, W. D. et al., 1994; van Beeumen et al., 1993), which is covalently attached to Cys69 via thioester linkage (Figure 1.3 B). PYP is present in a range of purple photosynthetic bacteria (Jiang et al., 1999; Koh et al., 1996; Kort et al., 1996), and form a recently discovered family of photoreceptor proteins.
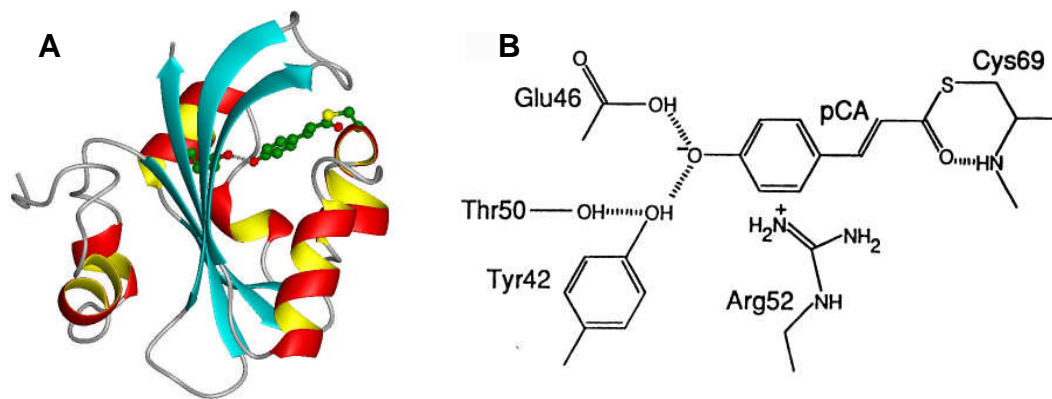
Figure 1.3: (A) The structure of PYP. PYP is folded into a PAS domain structure with a central beta sheet flanked on both sides by a total of 5 alpha helices. The pCA and Glu46 are shown as ball-and-sticks. (B) The photoactive site of PYP. The pCA chromophore interacts with key amino acid side chains via hydrogen bonding and electrostatic interactions.

PYP is a small water soluble protein (van Beeumen et al., 1993), approximately one third the size of rhodopsin. Its 125 amino acid residues are folded into a PAS domain structure (Figure 1.3 A) with a central beta sheet flanked on both sides by a total of 5 alpha helices (Borgstahl et al., 1995; Dux et al., 1998). Upon absorbing a blue photon, the pCA chromophore of PYP undergoes *trans* to *cis* photo-isomerization, initiates the photocycle of PYP

(Figure 1.4). In the initial receptor state, pG_{446} (446 nm is the peak absorption wavelength of PYP in pG receptor state), the pCA chromophore is buried, with its negatively charged phenolic oxygen stabilized by two hydrogen bonds with protonated Glu46 and Tyr42 (Figure 1.3 B) and a counter ion from Arg52, exhibiting the pK_a values of both the phenolic group of the pCA chromophore and the carboxylic group of Glu46 are anomalously shifted in opposite directions by 6 to 8 pH units. The neutral charge state of Glu46 is embraced by a highly hydrophobic binding pocket. In pR_{465}, chromophore photoisomerization dramatically increases the proton affinities of the phenolic oxygen of the pCA chromophore so that proton transfer from Glu46 to the pCA chromophore takes place in reversible pR_{465} to pB'_{355} states. The proton transfer process is localized at the active site of the protein without dramatic protein conformational changes. The resulting negative charge on Glu46 buried in hydrophobic core of the protein is energetically unstable. It therefore triggers and drives a large-amplitude protein quake, leading to the formation of putative signaling state pB_{355} (Hoff, W. D. et al., 1999; Sprenger et al., 1993; Xie et al., 2001). The conformational change pB_{355} is thought to initiate a signal transduction cascade that ultimately reverses the flagella motor to produce negative phototaxis (Hoff, W. D, 1995). Then it recovers back to the receptor state pG_{446} and the photocycle is complete. This molecular mechanism for PYP activation is proposed by our research group (Xie et al., 1996; Xie et al., 2001). The pR_{465} and pB_{355} states are also named I_1 and I_2 in the literature (Ujj et al., 1998).
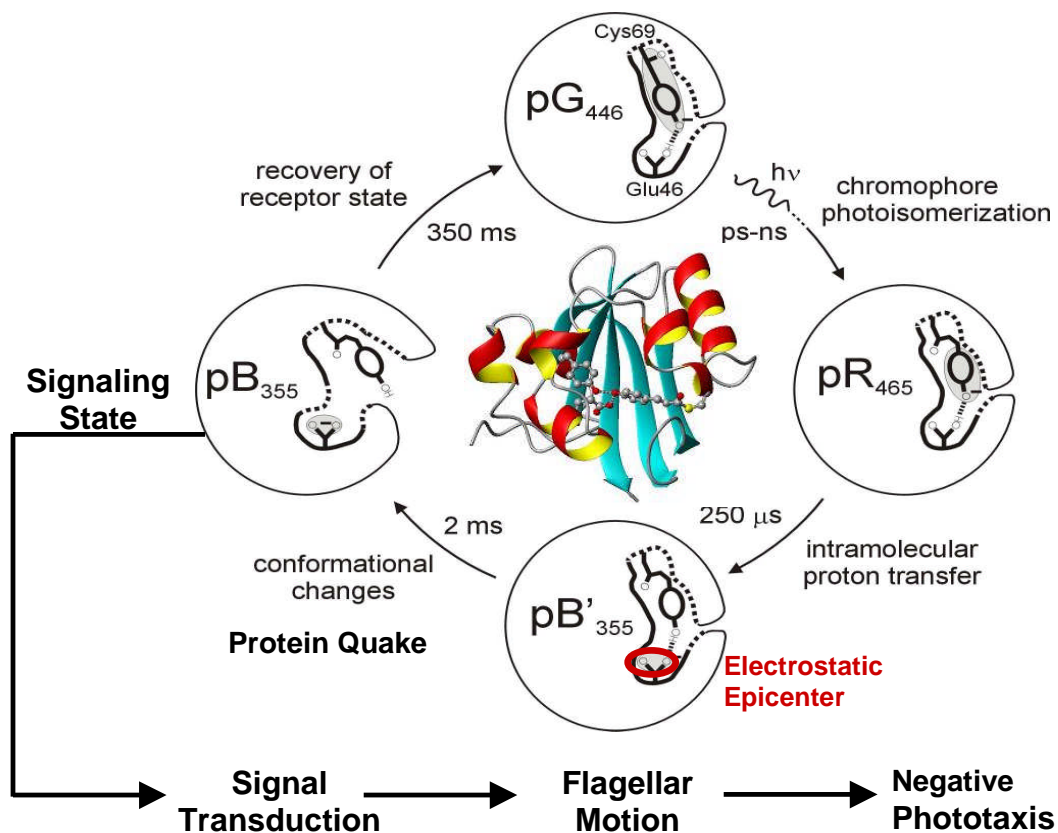
Figure 1.4: PYP photocycle and a molecular mechanism for PYP activation. Three intermediates (pR_{465}, pB'_{355}, and pB_{355}) of the PYP photocycle are depicted. The 3D structure of PYP (2PHY.pdb) is depicted in the inset. The electrostatic properties of the binding pocket for the chromophore and Glu46 are denoted using thick dashed lines for high-dielectric environments, thick solid lines for low-dielectric environments, and thin solid lines for unspecified dielectric environments. The charge delocalizations are qualitatively represented by shaded ovals.

Molecular dynamics simulation

The molecular dynamics (MD) method was first introduced by Alder and Wainwright in the late 1950's (Alder and Wainwright, 1957; Alder and Wainwright, 1959) to study the interactions of hard spheres. Many important insights concerning the behavior of simple liquids emerged from their studies. The next major advance was in 1964, when Rahman carried out the first simulation using a realistic potential for liquid argon (Rahman, 1964). The first MD simulation of a

realistic system was done by Rahman and Stillinger in their simulation of liquid water in 1974 (Stilling and Rahman, 1974). The first protein simulations appeared in 1977 with the simulation of the bovine pancreatic trypsin inhibitor (BPTI) (McCammon et al., 1977). Nowadays MD simulations of solvated proteins, protein-DNA complexes as well as lipid systems can often be found in the literatures addressing a variety of issues including the thermodynamics of ligand binding, protein stability, conformational changes, and protein folding. The number of simulation techniques has greatly expanded; there exist now many specialized techniques for particular problems, including combined quantum mechanical and mechanical mechanics (QM/MM). Molecular dynamics simulation techniques are widely used in experimental procedures such as X-ray crystallography and NMR structure determination.

In PYP field, protein dynamics using MD simulations of solvated PYP at early intermediate state had been studied before (Antes et al., 2002; Groenhof, G. et al., 2004; Groenhof, G et al., 2002a; Groenhof, G. et al., 2002b; Thompson et al., 2003; van Aalten et al., 2002; Yamato et al., 1998). However, whether the crystal structures of PYP at early intermediate state that had been used for the simulations were correct was still under debate. It is essential for us to investigate the structure nature of PYP at early intermediate state using MD simulations that will allow us to follow the kinetics of hydrogen bonding network and get insight into structural features that are experimentally difficult to determine. With experimental evidence of broken or intact hydrogen bonding interactions, we will then determine plausible structure of PYP at the early

intermediate state. This will provide a specific guide for researchers on how to use time-resolved infrared spectroscopy to identify which photoproduct is formed after photoisomerization.

Due to the limitation of computer power, MD simulations of solvated PYP on microsecond and millisecond timescale are not possible. Special strategy has been developed to explore protein conformations on this timescale, namely, Go model simulation. It involves striping all the water molecules and hydrogen atoms from MD simulations of protein in explicit solvents and defining constraints such as Leonard-Johns potential, coulomb repulsion and attraction potential, hydrogen bonding interactions, hydrophobic contacts, etc. Although we had performed Go model simulations of PYP, the data analysis is beyond scope of this thesis. In this thesis, we mainly focused on MD simulations of solvated PYP on nanosecond timescale studying protein structures at early intermediate state to understand the mechanism of proton transfer.

1.2 Goals

Our ultimate goal is to understand the mechanism of proton transfer process in proteins using PYP as a model system.

Goal I. Establish a vibrational spectral marker for probing hydrogen-bonding status of Glu46 for proton transfer at the active site of PYP (*ab initio* calculations)

Goal II. Establish a vibrational spectral marker for probing hydrogen-bonding status of Tyr42 for proton transfer at the active site of PYP (*ab initio* calculations).

Goal III. Study structural basis of PYP at intermediate states (Molecular dynamics simulation) and the impact of hydrogen bond interactions on proton affinity (*ab initio* calculations) toward understanding proton transfer mechanism in proteins.

CHAPTER II

A VIBRATIONAL SPECTRAL MARKER FOR PROBING THE HYDROGEN

BONDING STATUS OF PROTONATED ASP AND GLU RESIDUES

2.1 Introduction

Buried carboxylic groups from the side chains of aspartic (Asp) and glutamic (Glu) acids are often found in the active sites of proteins. Despite the low pK_a values (typically around 4) of exposed carboxylic groups, many buried carboxylic groups exhibit anomalously shifted-pK_a values (Creighton, 1997), some as high as 11 (Meyer et al., 2003; Szaraz et al., 1994). Due to large increases in pK_a values, these ionizable groups are protonated at neutral pH conditions. The pK_a values of these groups may change dramatically during the functional processes of proteins, resulting in deprotonation and proton transfer (Song et al., 2003; Xie et al., 2001). Hydrogen bonding interactions play key roles in regulating the pK_a values of ionizable groups and in driving proton transfers.

There are many examples of Asp and Glu residues that are buried in the active sites of proteins. These residues play important roles in protein functions. For photoactive yellow protein (PYP), a blue light bacterial photoreceptor, only one out of nineteen Asp and Glu residues is buried (Borgstahl et al., 1995; Xie et al., 1996). This buried residue, Glu46, is located at the active site of PYP, and exhibits an abnormal pK_a of 11 or above in the initial, receptor state (Meyer et al.,

2003). Therefore, Glu46 is protonated at neutral pH in the receptor state (pG), however becomes deprotonated and ionized during the formation of the putative signaling state (pB) upon light activation (Brudler et al., 2001; Imamoto et al., 1997; Xie et al., 1996; Xie et al., 2001). Experimental evidence (Xie et al., 1996; Xie et al., 2001) strongly suggests that Glu46 is essential for driving large protein conformational changes, so called "protein quake", during PYP receptor activation. For bacteriorhodopsin, a well studied light-driven proton pump for solar energy transduction, four out of its eighteen Asp and Glu residues are buried (Engelhard et al., 1985; Krebs and Khorana, 1993; Lanyi and Schobert, 2002; Pebay-Peyroula et al., 1997; Subramaniam and Henderson, 2000). They are Asp85, Asp96, Asp115, and Asp212 residues that are found in either the chromophore binding pocket or the proton transfer pathway. Asp85 and Asp212 are essential for charge stabilization of the protonated Schiff base of the retinal chromophore (Marti et al., 1992; Otto et al., 1990). Asp85 is the proton acceptor for Schiff base deprotonation during the proton pumping photocycle (Subramaniam et al., 1992), whereas the deprotonation of Asp96 is directly coupled to the reprotonation of the Schiff base. Mutation of Asp96 to Asn residue results in 100 fold slow down of the Schiff based re-protonation process (the M2 to N transition) (Brown et al., 1998). For rhodopsin, an animal visual photoreceptor protein, Glu134 is highly conserved and plays a central role in modulating proton uptake and receptor activation of rhodopsin (Arnis et al., 1994). Mutation of Glu134 results in constitutive activation of rhodopsin (Acharya and Karnik, 1996), exemplifying its vital role in the structures and functions of

rhodopsins. For heme-copper oxidases, redox-driven proton pumps, the conserved Glu242 residue from mitochondrial or bacterial membrane of cytochrome bo$_3$ plays a key role in the proton translocation mechanism (Lubben et al., 1999; Riistama et al., 1997). These examples show that there are three types of buried carboxylic groups, some remain ionized, some stay neutral, and some change their protonation states during the functional processes of proteins. In this paper, we focus on developing a vibrational spectral marker for probing the hydrogen bonding status of buried, protonated COOH groups, including those that change their protonation states during the functional processes.

Hydrogen bonding is a fundamental element in protein structure and function. The hydrogen bond dissociation energy in proteins is in the order of 10 – 40 kJ/mol (2-10 kcal/mol) (Creighton, 1997; Hoppe et al., 1983) (Perrin and Nielson, 1997). (Much stronger hydrogen bonds with up to 160 kJ/mol have been found in some chemical systems, particularly those with small, charged groups (Hibbert and Emsley, 1990; Perrin and Nielson, 1997; Steiner, 2002).) In comparison, a typical value for protein folding energy is approximately 40 kJ/mol (Honig and Yang, 1995), only about one to four times of the hydrogen bond dissociation energy. Therefore, breaking one or more hydrogen bonds may impair the stability of a protein.

A protonated carboxylic group may form up to four hydrogen bonds with neighboring hydrogen bond donors and acceptors: the carbonyl oxygen atom of a COOH group may form a maximum of two hydrogen bonds, the hydroxyl oxygen and the hydroxyl hydrogen can each form one hydrogen bond. It is

therefore energetically preferable for buried carboxylic groups to form as many hydrogen bonds as possible. It is challenging to monitor the hydrogen bond number of buried, protonated Asp and/or Glu residues for functional intermediate states of proteins and for proteins that their X-ray crystal structures are not available.

Time-resolved infrared difference spectroscopy is a powerful method to probe and characterize the structural dynamics of proteins, particularly involving proton transfer (Brudler et al., 2001; Friedrich et al., 2002; Gerwert, 1999; Heberle, 2000; Xie et al., 2001). Most infrared studies on amino acid residues in proteins are carried out on Asp and Glu residues due to not only their functional importance, but also the presence of an excellent vibrational spectral marker. The C=O stretching frequency of a protonated carboxylic group is typically in the region of 1700 to 1775 cm^{-1} (see Table 2.8), while the asymmetric and symmetric stretching frequencies of a COO^- group are characteristically around 1555 and 1410 cm^{-1} respectively (Barth, 2000). This large difference in the vibrational frequencies of protonated and ionized carboxylic groups allow conclusive identification of their protonation states. This method has been successfully applied to study the protonation states of functionally important carboxylic groups in both transient and steady states of proteins (Bousche et al., 1992; Braiman et al., 1988; Brown et al., 1995; Fahmy et al., 2000; Jager et al., 1994; Rothschild, 1992; Xie et al., 2001; Zscherp et al., 1999) or reaction products (Nabedryk et al., 2001).

The distribution of the C=O stretching frequency is thought to be related to the dielectric medium or hydrogen bonding interactions of the concerned carboxylic groups in proteins (Maeda et al., 1992; Puustinen et al., 1997; Sasaki et al., 1994; Xie et al., 1996). However, it lacks quantitative studies to establish a strong and clear correlation between the C=O stretching frequency and the nature of hydrogen bonding interactions of protonated carboxylic groups. In this paper, we report that the C=O stretching frequency is a sensitive infrared structural sensor (vibrational spectral marker) for detecting and monitoring the number of hydrogen bonds of buried COOH groups. This vibrational spectral marker is established based on ab initio computational studies and is supported by experimental data. When the C=O stretching frequency is combined with the use of the O$-$H stretching frequency, our computational studies further demonstrate that it is possible to distinguish one type of hydrogen bond from another, such as one hydrogen bond with carboxyl oxygen from that with hydroxyl hydrogen. The applications of vibrational spectral markers are expected to enhance the power of time-resolved infrared difference spectroscopy for structural characterization of functionally important intermediate states of proteins, and consequently for understanding the functional mechanism of proteins.

2.2 Computational method

A butyric acid molecule, shown in Figure 2.1 A, was employed to model the side chains of Asp and Glu residues in proteins. Methanol molecules were selected to serve as the hydrogen-bond partner for the butyric acid, based on its

structural simplicity and the fact that its hydroxyl group (OH) can serve as both hydrogen bond donor and acceptor.

Both energy and vibrational frequency calculations were performed using ab initio methods based on density function theory (Gaussian03 (Frisch et al., 2003)). Each input structure of butyric acid, with or without hydrogen bonds (Figure 2.1), for Gaussian03 calculations was first generated using ChemDraw, then preliminarily optimized using Chem3D based on the semi-empirical PM3 method (CambridgeSoft Co.). In molecular structure specification section of Gaussian calculation (Frisch et al., 2003), the hydrogen-bond between hydrogen atom (hydrogen-bond donor) and oxygen atom (hydrogen-bond acceptor) in the butyric acid-methanol complex was defined for geometry optimization involving hydrogen-bond interactions.
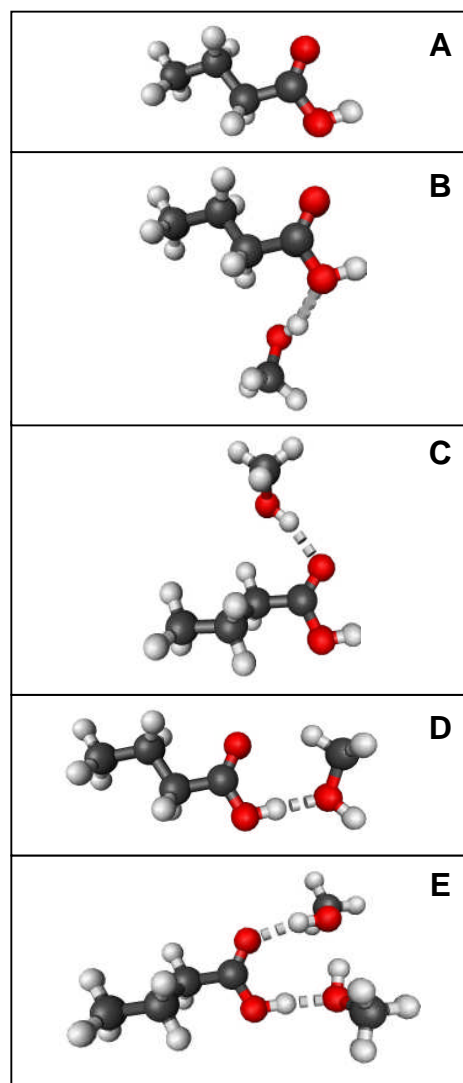
Figure 2.1 The structure of a butyric acid molecule (A) and its hydrogen bonding interactions with 0, 1, and 2 methanol molecules. The notation, H**O**−C=O (B), represents that the hydroxyl oxygen atom (in bold and underlined) forms one hydrogen bond with a methanol molecule. Similarly, HO−C=**O** (C) and **H**O−C=O (D) indicate that the carbonyl oxygen atom and the hydroxyl hydrogen atom form a hydrogen bond with a methanol molecule, respectively. The notation, **H**O−C=**O** (E), denotes that both the hydroxyl hydrogen and the carbonyl oxygen atoms form a hydrogen bond with a methanol molecule.

Calculations of Energy

To calculate the hydrogen bond dissociation energy of butyric acid with methanol, we calculated the energies of an isolated butyric acid, an isolated

methanol, and a butyric acid-methanol hydrogen-bonding complex. The hydrogen bond dissociation energy is defined as:

$$\Delta E = E_{BAC} + E_{ME} - E_{BAC,ME} \qquad (2.1)$$

where BAC stands for butyric acid while ME for methanol.

To calculate the energy, the input structure was first optimized using B3LYP/6-31G(d), and then the energy was calculated using B3LYP/6-311+G(2d,p). This method is shown to be one of the most accurate methods for energy calculation (Foresman and Frisch, 1996). The resulting output energy is in the atomic unit (au), which was converted into kJ/mol with 1 au = 2619.6 kJ/mol.

To calculate the energy profile as a function of hydrogen-bond length between butyric acid and its hydrogen-bond partner, methanol, the hydrogen bond length was fixed at a number of values from 2.4 Å to 4.0 Å during geometry optimization. In most cases when the hydrogen bond-length was fixed, the direct geometry optimization using B3LYP/6-31G(d) did not converge. Such failure is often due to inaccurate initial force constants that are estimated at the first step, and then refined at each subsequent optimization step. By first calculating the initial force constants using HF/3-21G or HF/6-31G, we were able to optimize the butyric acid-methanol structure using B3LYP/6-31G(d) (see Table 2.7). Once the geometry was optimized for each fixed hydrogen bond-length, the energy was then calculated using B3LYP/6-311+G(2d,p).

Energy calculation in dielectric medium

Self-Consistent Reaction Field (SCRF) methods were employed in Gaussian03 for structural optimization and energy calculations in the presence of a dielectric medium. Using self-consistent reaction fields, the solvent is modeled as a continuum of uniform dielectric constant ε, while the solute is placed into a cavity within the solvent. We select Polarized Continuum Model (PCM) which defines the cavity as the union of a series of interlocking atomic spheres to calculate energies of an isolated butyric acid, an isolated methanol or a model compound of amino acid, and a butyric acid and hydrogen-bond partner complex in various solvents (Cammi et al., 2000; Cappelli et al., 2000; Foresman and Frisch, 1996). The values of the dielectric constants for various solvents (at 293.2 K or 20°C and atmospheric pressure) were quoted from the CRC Handbook of Chemistry and Physics (78th edition) (Lide, 1997-1998).

Calculations of Vibrational Frequencies

Vibrational frequency calculations were carried out in three steps. First, each structure was optimized using B3LYP/6-31G(d) method. Then all the force constants for vibrational motions were calculated, also using B3LYP/6-31G(d) method. Finally the frequencies for all the vibrational modes were computed based on normal mode analysis. Calculated vibrational frequencies are systematically higher than the experimental values. A scaling factor is recommended for each type of computational method to compensate the over-estimated force constants (Foresman and Frisch, 1996). However, as long as we use the same computational method, this scaling factor is the same for all the calculations of different molecules. For the B3LYP/6-31G(d) computational

method, the recommend scaling factor is 0.9613 (Foresman and Frisch, 1996), representing approximately 3.9% reduction from the calculated vibrational frequencies.

Except for those studies in specified dielectric media (Table 2.2), all the calculations were made in vacuum. This is due to the fact that the computational theories and methods for treating dielectric media are complicated and have not been well developed. In contrast, the computational methods for use in vacuum are far more accurate than those in dielectric media. Therefore, all the frequency calculations in this paper were made in vacuum. This issue is further addressed in the discussion.

Identify the number and the type of hydrogen-bond interactions from protein crystal structures

The pdb files for high resolution crystal structures of proteins were obtained from the protein data bank. To determine the number and the type of hydrogen bonding interactions for each buried carboxylic group, we first distinguished carbonyl oxygen from hydroxyl oxygen by measuring and comparing the two C−O bond lengths. The shorter bond is regarded as the carbonyl group, while the longer one as the hydroxyl group. Then we examined and identified all the plausible hydrogen-bond donors and/or acceptors within the hydrogen bonding distance of a carboxylic group. Next, we checked the distance, bond angle and dihedral angles between the potential hydrogen bonding partners. A hydrogen bond is expected to meet all three criteria: less than 3.2 Å for the hydrogen-bond length, 109±15º for the C−O…X angle, and 160 to 180° or

0 to 20° for the C−C−O…X dihedral angle. X stands for a heavy atom that is hydrogen bonded to a COOH group. Finally we examined the compatibility of hydrogen bonding interaction. For example, it is impossible for a hydrogen-bond acceptor to form a hydrogen bond with another hydrogen-bond acceptor.

2.3 Results and discussion

Figure 2.1 A depicts the structure of a butyric acid molecule (C_3H_7COOH), a model compound for the side chains of protonated Asp and Glu residues in proteins. Figure 2.1 B−D illustrates three different types of single hydrogen bonding interaction between a butyric acid and its partner, the hydroxyl group of a methanol. Figure 2.1 E displays a complex of a butyric acid and two methanol molecules. The hydrogen bonding properties of protonated carboxylic groups were studied regarding their hydrogen-bond dissociation energy and hydrogen-bond length using density function theory (DFT) based ab initio computational methods (Gaussian03). This study provides insights into the roles of Asp and Glu residues in protein structure and function. We propose a vibrational spectral marker for probing the hydrogen-bond numbers of buried COOH groups. Such a vibrational spectral marker is particularly valuable in order to characterize the structures and dynamics of short-lived intermediate states of proteins while in action using time-resolved infrared difference spectroscopic techniques. Our computational results are presented and discussed below.

Hydrogen-bond strength of protonated carboxylic (COOH) groups

The strength of a hydrogen bond is characterized by its hydrogen-bond dissociation energy (defined in Eq. 2.1). This is the energy required to completely

separate a pair of hydrogen-bond donor and acceptor. The stronger a hydrogen bond is, the more energy is required to break the hydrogen bond. We first calculated and examined the hydrogen-bond dissociation energies of a protonated carboxylic group (butyric acid) for four different types of hydrogen bonding interactions (Figures 2.1 B–E). These results are summarized in Table 2.1. The notation, H**O**–C=O, represents that the hydroxyl oxygen atom is hydrogen bonded to a methanol molecule (see Figure 2.1 B). Similarly, HO–C=**O** (Figure 2.1 C) and **H**O–C=O (Figure 2.1 D) indicate that the carbonyl oxygen atom and the hydroxyl hydrogen atom are hydrogen bonded to a methanol molecule, respectively. The notation, **H**O–C=**O** (Figure 2.1 E), denotes that both the hydroxyl hydrogen and the carbonyl oxygen atoms are hydrogen bonded. It is interesting to note that the calculated hydrogen-bond strength strongly depends on the specific type of hydrogen bonding interactions: 11.4 kJ/mol for H**O**–C=O, 18.5 kJ/mol for HO–C=**O**, and 31.3 kJ/mol for **H**O–C=O.

This result sheds light on the preferred hydrogen bonding interaction of a protonated carboxylic group in the vicinity of a hydrogen-bond donor in proteins. Due to the small size of a hydrogen atom, the hydrogen bonding interactions of HO–C=**O** and H**O**–C=O are sterically similar. However, they differ in hydrogen-bond strength. The HO–C=**O** hydrogen bonding interaction is more stable than that of H**O**–C=O by 7.2 kJ/mol. This energy difference is significant compared with the thermal energy at room temperature (2.5 kJ/mol). Therefore, it is more likely for a buried COOH group to form a hydrogen bond in form of HO–C=**O** instead of H**O**–C=O. In the case that the hydrogen bond partner of a buried,

protonated carboxylic group has a hydroxyl group which can serve as either a hydrogen-bond donor or an acceptor, the <u>H</u>O−C=O interaction is energetically more favorable than both HO−C=<u>O</u> and H<u>O</u>−C=O interactions.

Table 2.1 also shows the optimized hydrogen-bond length for each type of hydrogen bonding interactions. The hydrogen-bond length between heavy atoms increases progressively from 2.75 Å for the strongest hydrogen bond of <u>H</u>O−C=O, to 2.87 Å for the medium hydrogen bond of HO−C=<u>O</u>, to 2.94 Å for the weakest hydrogen bond of H<u>O</u>−C=O. In the presence of two hydrogen bonding interactions, <u>H</u>O−C=<u>O</u> (Figure 2.1 E), the hydrogen-bond length involving the carbonyl oxygen is reduced from 2.87 Å (for HO−C=<u>O</u>) to 2.75 Å, while the hydrogen-bond length involving the hydroxyl hydrogen is reduced from 2.75 Å (for <u>H</u>O−C=O) to 2.65 Å. These results reveal that for a given pair of hydrogen bond partners, the stronger a hydrogen bond is, the shorter the hydrogen-bond length. Such correlation between the hydrogen-bond length and hydrogen-bond dissociation energy is consistent with previous reports (Hoppe et al., 1983; Steiner, 2002).

TABLE 2.1 Calculated hydrogen-bonding properties of a COOH group

No. of H-bonds	Type of H-bond	H-bond dissociation energy (kJ/mol)	H-bond length* (Å)	C=O bond length (Å)	C–O bond length (Å)	Difference in bond length$^{\#}$ (Å)
0	HO–C=O	0.0	N/A	1.211	1.359	0.148
1	H**O**–C=O	11.1	2.94	1.208	1.374	0.166
1	HO–C=**O**	18.3	2.87	1.220	1.351	0.131
1	**H**O–C=O	31.3	2.75	1.218	1.341	0.123
2	**H**O–C=**O**	63.7	2.65 / 2.75	1.228	1.325	0.097

*The hydrogen-bond length is defined here as the distance between two heavy atoms (oxygen) from a pair of hydrogen-bond donor and acceptor. This is due to the fact that the positions of hydrogen atoms are often not resolved in protein crystal structures.

$^{\#}$The difference in the bond length is defined as $\delta l = l_{C-O} - l_{C=O}$ where l_{C-O} and $l_{C=O}$ are the bond lengths of C–O and C=O respectively.

As shown in Table 2.1, the covalent bond length of C=O is slightly shorter than the C–O bond length by 0.10 to 0.17 Å. Such differences are often observable in high resolution protein crystal structures (eg. 1.50 Å or better). Note that the precision in bond length from protein X-ray crystallography is generally much better than the spatial resolution. Since the positions of hydrogen atoms are not resolved in most protein crystal structures, this small, but detectable difference between the two CO covalent bond lengths provides a practical and valuable guide to distinguish the carbonyl oxygen from the hydroxyl oxygen of a protonated carboxylic group in proteins. This approach was employed in identifying hydrogen bonding interactions in protein crystals presented later in the paper.
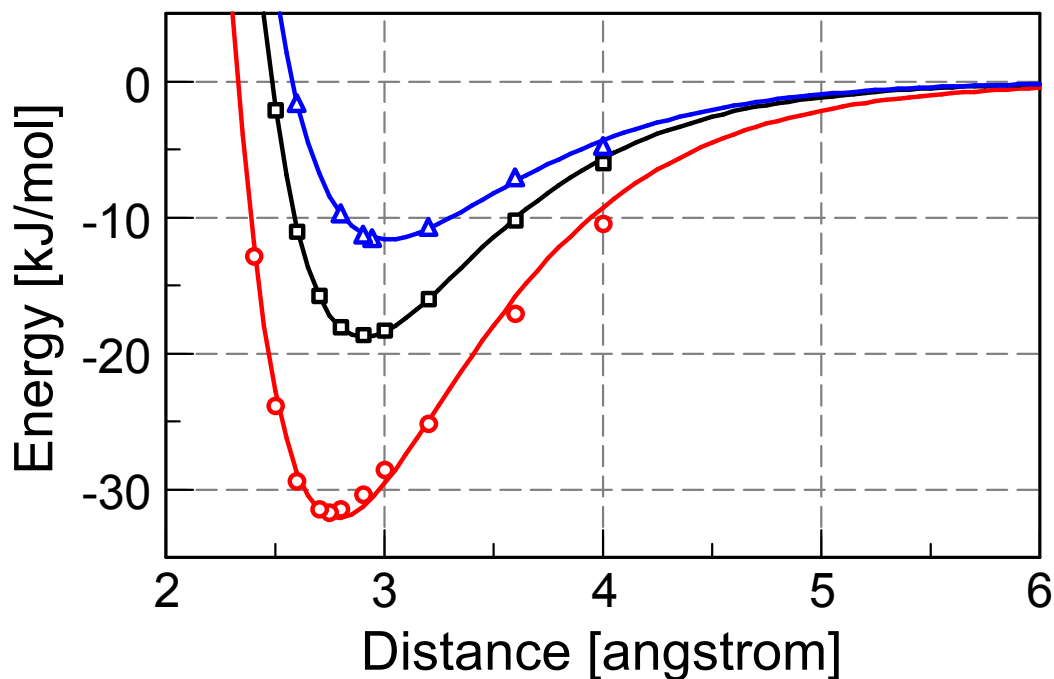
Figure 2.2 Hydrogen-bond dissociation energy of a protonated carboxylic group (butyric acid) interacting with a methanol molecule. The hydrogen-bond dissociation energies were calculated at selected bond lengths using B3LYP/6-311+G(2d,p) (see method for details) for H**O**–C=O (blue triangles), HO–C=**O** (black squares), and **H**O–C=O (red circles). The curves are the results of non-linear least square fitting of the computational data to Morse potentials. The strongest hydrogen bonding interaction occurs at 2.94 Å with –11.6 kJ/mol for H**O**–C=O, 2.87 Å with –18.8 kJ/mol for HO–C=**O**, and 2.75 Å with –32.1 kJ/mol for **H**O–C=O.

Figure 2.2 shows the profiles of the calculated hydrogen-bond dissociation energy versus hydrogen-bond length. Such curves are not characterized by parabolic functions, indicating that hydrogen bonding interactions are strongly anharmonic. Instead, the data fit well to Morse potentials, $E = E_e * \{[1 - e^{-a(r-r_e)}]^2 - 1\}$, where r_e is the optimal hydrogen-bond length and E_e is the hydrogen-bond dissociation energy. Due to thermal motions that allow energy fluctuation by 2.5 kJ/mol, the hydrogen-bond length is allowed

to vary from 2.78 to 3.41 Å by 0.63 Å for H**O**−C=O, from 2.72 to 3.18 Å by 0.46 Å

for HO−C=**O**, and from 2.63 to 3.00 Å by 0.37 Å for **H**O−C=O around their

optimal hydrogen-bond lengths. That is, the stronger the hydrogen bonding

interaction, the smaller the structural fluctuations regarding the hydrogen-bond

length.

Influence of local dielectric medium on hydrogen-bond strength

The hydrogen-bond strength of a buried COOH group in proteins is

sensitive to the local dielectric environment. To examine the influence of a local

dielectric medium on the hydrogen-bond strength, we calculated the hydrogen-

bond dissociation energy in a variety of dielectric media. The results are shown in

Table 2.2. In benzene (C_6H_6), a non-polar solvent with a dielectric constant of

2.3, the hydrogen-bond dissociation energy is 4.3 kJ/mol for H**O**−C=O, 10 kJ/mol

for HO−C=**O** and 21 kJ/mol for **H**O−C=O. Therefore, in benzene-like dielectric

environment, the hydrogen bonding interactions of **H**O−C=O are highly stable,

whereas that of HO−C=**O** is much stronger than H**O**−C=O. Note that the

hydrogen-bond dissociation energies of three single hydrogen bonding

interactions are reduced by different extents from vacuum to benzene: 34% for

HO−C=O, 44% for HO−C=**O**, and 61% for H**O**−C=O. If the electrostatic

interactions were dominant, each hydrogen-bond dissociation energy would be

reduced by a factor of $(\varepsilon-1)/\varepsilon$ (55 % for benzene), where ε is the dielectric

constant of the solvent. The smaller reductions in the calculated hydrogen-bond

dissociation energy than 55% for HO−C=**O** and **H**O−C=O in benzene indicate

that the other contributions are significant, such as chemical interactions between hydrogen bonding partners and with solvent.

In ether (C_2H_5–O–C_2H_5), a weakly polar solvent with a dielectric constant of 4.3, the hydrogen-bond dissociation energies for both H**O**–C=O and HO–C=**O** are rather small, no longer provide effective stabilization to the protein structure. However, the hydrogen bond for **H**O–C=O remains strong, with 14.6 kJ/mol, much larger than the thermal energy, 2.5 kJ/mol. Therefore, only the hydrogen bonding interaction of **H**O–C=O is significant in ether-like dielectric medium. Similar conclusions can be made for chlorobenzene (Cl-C_6H_5), a polar solvent with a dielectric constant of 5.7. In water, a highly polar solvent with a dielectric constant of 80.1, all three types of hydrogen bonding interactions are weak. In fact, the carboxylic group of Asp or Glu residue prefers to be ionized in water above pH 4. The fact that the same carboxylic group favors to be protonated below pH 4 is due to the entropic contribution prevailing over enthalpic contribution. In conclusion, our computational results show that a low dielectric interior of a protein makes the hydrogen bonding interactions between buried groups significantly stronger than those in a high dielectric interior. Therefore, a low dielectric medium of a protein interior is favorable for the structural stability of a protein.

TABLE 2.2 Calculated hydrogen-bond dissociation energy of a COOH group in dielectric media

Hydrogen bond		Hydrogen-bond dissociation energy (kJ/mol)				
Number	Type	Vacuum $\varepsilon = 1.0$	Benzene $\varepsilon = 2.3$	Ether $\varepsilon = 4.3$	Chlorobenzene $\varepsilon = 5.6$	Water $\varepsilon = 80.1$
0	HO–C=O	0	0	0	0	0
1	H**O**–C=O	11.1	4.3	−1.2	0.1	−5.6
1	HO–C=**O**	18.3	10.2	3.1	4.6	1.5
1	**H**O–C=O	31.3	20.7	14.6	12.8	6.7
2	**H**O–C=**O**	63.7	38.8	20.4	24.6	4.7

There is no simple answer to what is the effective, average dielectric constant inside a protein. Proteins have a non-uniform dielectric medium. The presence of external and internal water molecules makes it difficult to perform direct experimental studies. Dielectric constants of dry protein powders have been directly measured to be 2 to 4 (Dwyer et al., 2000). Indirect studies of hydrated proteins suggest that dielectric constants of protein interiors may be as high as 12 (Garcia-Moreno et al., 1997). Examination of protein structures and flexibility may provide reasonable insight into the effective dielectric constants of proteins. Dielectric constants are determined by two major factors: reorientation of permanent dipoles of polar groups and induced dipoles of non-polar and polar groups (Griffiths, 1999). Dielectric constants of non-polar solvents arise from induced dipoles. They are typically small (with 1.9 for heptane, 2.0 for cyclohexane, 2.2 for carbon tetrachloride, 2.3 for benzene, and 2.4 for toluene). Dielectric constants of polar solvents are dominated by re-orientation of their permanent dipoles. They are generally much larger than those of non-polar solvents (with 33.0 for methanol, 47.2 for DMSO, and 80.1 for water). Inside proteins, most buried polar groups are carbonyl and amide groups in the protein backbone. They are exceptionally difficult to be re-orientated due to their limited degrees of freedom. There are a small number of buried polar groups from the side chains of polar and charged amino acid residues in proteins. These polar groups have some degrees of freedom for reorientation, but generally much more constrained than the corresponding free molecules in solvents due to covalent linkages and steric hindrance. Therefore, buried polar groups, except for

internal water molecules, have limited contributions from re-orientations of their permanent dipoles to the dielectric constants of protein interiors. Based on these considerations, the dielectric constant for a protein interior on average is expected to be slightly larger than those of non-polar solvents (eg. ~2.5 to 3.5). In the active sites of proteins where many polar groups and/or internal water molecules are found, the localized effective dielectric constant may be higher than the average dielectric constant of protein interior. However, contributions of these polar and charged groups can be treated explicitly in terms of hydrogen bonding interactions. The effective dielectric constant of the remaining part of a protein is small.

A vibrational spectral marker for probing the hydrogen-bond number of a protonated carboxylic group

Certain vibrational modes carry specific structural information. Therefore, they can be employed to probe the structures of proteins. The vibrational band of such a mode is regarded as a vibrational spectral marker, or infrared structural sensor. For example, the C=O stretching frequency of a carboxylic group is an excellent vibrational spectral marker for detecting the protonation state of this group. When a carboxylic group is protonated, its C=O stretching frequency is typically in the spectral region of 1700 – 1770 cm^{-1} (Barth, 2000). Once it is ionized upon deprotonation, this frequency is shifted to approximately 1555 and 1400 cm^{-1} for its asymmetric and symmetric O–C–O stretching modes respectively (Barth, 2000). This vibrational spectral marker has been extensively used to probe the protonation status of the side chains of Asp and/or Glu

residues in bacteriorhodopsin (Dioumaev, 2001; Rothschild, 1992), rhodopsin (Fahmy et al., 2000; Jager et al., 1994), photoactive yellow protein (Brudler et al., 2001; Imamoto et al., 1997; Xie et al., 1996; Xie et al., 2001), and other proteins (Bergo et al., 2003; Friedrich et al., 2002; Lubben et al., 1999; Nabedryk et al., 2000; Nabedryk et al., 2001). Another good example is the Amide I band of the protein backbone C=O stretching. The Amide I frequency is in the range of 1610 – 1690 cm^{-1}, and has been extensively employed to determine the secondary structures of proteins (Barber-Armstrong et al., 2004; Hering et al., 2002; Susi et al., 1967), namely alpha helices, beta sheets, turns, and random coils. Here we present a vibrational spectral marker, the C=O stretching frequency of a protonated carboxylic group (COOH), that detects the number of hydrogen bonding interactions.

Table 2.3 shows the calculated frequencies of the C=O and O–H stretching modes of a protonated carboxylic group with different hydrogen bonding interactions. With no hydrogen bond, the C=O bond is strong. Consequently the C=O stretching frequency is high, at 1776 cm^{-1}. When one hydrogen bond is formed with either the carbonyl oxygen (HO–C=<u>O</u>) or the hydroxyl hydrogen (<u>H</u>O–C=O), the C=O stretching frequency is down shifted by approximately 30 to 35 cm^{-1}, to 1746 and 1741 cm^{-1} respectively. Since the hydrogen bonding interaction with the hydroxyl oxygen (H<u>O</u>–C=O) is energetically not stable in the dielectric environment of proteins (see discussion above), it is not expected to observe the characteristic up-shift of the C=O frequency for this type of hydrogen bonding interaction. Therefore, our

calculations reveal that the characteristic vibrational frequency of a protonated carboxylic group with one hydrogen bond interaction with methanol molecule in vacuum is in the region of 1741 to 1746 cm^{-1}. When two hydrogen bonds are formed with a COOH group, with both its carbonyl oxygen and hydroxyl hydrogen (HO−C=O), the C=O stretching frequency is further down-shifted by 35 to 40 cm^{-1} to 1705 cm^{-1}. This frequency is well separated from the C=O stretching frequency for one hydrogen bonding interaction. Therefore, the C=O stretching frequency of a protonated carboxylic group may serve as a sensitive vibrational spectral marker for probing the nature of its hydrogen bonding interactions.
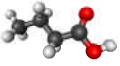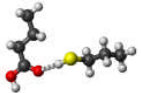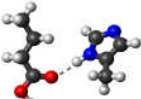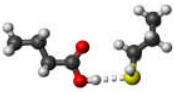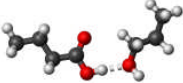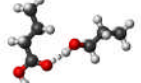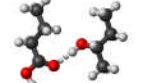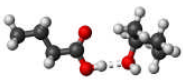
TABLE 2.3 Calculated C=O and O–H stretching frequencies of a protonated carboxylic group
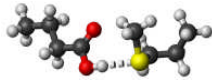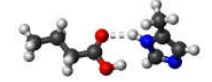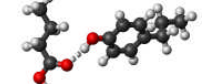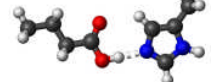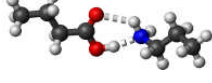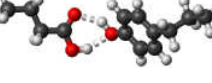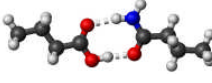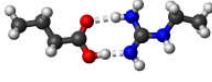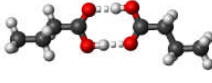
No. of H- bonds	Type of H-bond	C=O Freq (cm^{-1})	ΔFreq of C=O (cm^{-1})*	O–H Freq (cm^{-1})	ΔFreq of O–H (cm^{-1})*
0	HO–C=O	1776.4	0	3542.6	0
1	H**O**–C=O	1787.8	+11.4	3538.2	−4.4
1	HO–C=**O**	1740.6	−35.8	3537.7	−4.9
1	**H**O–C=O	1744.1	−32.3	3213.9	−328.7
2	**H**O–C=**O**	1705.5	−70.9	2860.2	−682.4

*The shifts in the C=O and the O–H stretching frequencies are due to hydrogen bonding interactions. A methanol molecule is employed as the hydrogen-bond partner to the COOH group.

To further explore, evaluate, and determine the qualification of the C=O stretching frequency as a vibrational spectral marker, we performed a range of calculations of a protonated carboxylic group interacting with different polar and charged side chain groups of amino acids via hydrogen bonding interactions. These side chain groups include 10 polar side chain groups (Ser, Thr, Cys, Met, Asn/Gln, Tyr, Asp/Glu, His, Lys, and Arg), 2 negatively charged side chain groups (Tyr and Asp/Glu) and 3 positively charged side chain groups (Arg, His, and Lys). Tables 2.4 and 2.5 show the computational results as well as the structures of molecules that are employed to model the different side chain groups of polar and charged amino acids. For the convenience of discussion, we will refer these model compounds as their corresponding amino acids.

TABLE 2.4 Calculated hydrogen-bonding properties of butyric acid interacting with polar amino acid side chains

Structure[*]	Amino Acid[*]	No. of H-bond[†]	H-bond Length[†] (Å)	H-bond Angle[†] (deg)	H-bond Energy[‡] (kJ/mol)	C=O Frequency[‡] (cm^{-1})
	N/A	0	N/A	N/A	N/A	1776.4
	Cys – 1	~0	3.60	175.8	5.9	1759.1
	His – 1a	1	2.99	174.7	20.3	1748.8
	Cys – 2	1	3.37	175.7	24.8	1746.9
	Ser – 2	1	2.75	177.7	31.9	1744.2
	Ser – 1	1	2.89	171.3	17.6	1743.7
	Thr – 1	1	2.90	172.8	17.4	1742.6
	Thr – 2	1	2.75	176.4	32.2	1739.4

42

	Met	1	3.30	174.7	21.2	1738.2
	His − 1b	1	2.93	151.8	21.2	1736.3
	Tyr − 1	1	2.84	173.6	24.0	1736.2
	His − 2	1	2.77	178.5	45.3	1733.4
	Lys	~1.5	2.73 2.99	166.3 126.5	44.8	1726.2
	Tyr	~1.5	2.77 2.72	154.0 145.0	37.5	1717.8
	Asn/Gln	2	2.68 2.88	173.4 168.0	61.2	1709.9[§]
	Arg	2	2.65 2.90	175.3 173.2	70.4	1704.8
	Asp/Glu	2	2.69 2.69	178.9 179.4	63.7	1702.9[§]
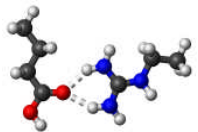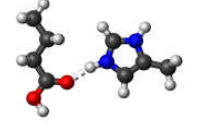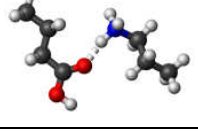
[*]The structures presented here are optimized using B3LYP/6-31G(d) method. Color code: red for oxygen atoms, blue for nitrogen and sulfur atoms, dark gray for carbon atoms, and white for hydrogen atoms. The dashed lines represent hydrogen bonds. His−1 indicates that the carbonyl oxygen of a COOH group forms a hydrogen bond, while His−2 designates that the hydroxyl hydrogen of a COOH group forms a hydrogen bond. The same notation is used for other amino acids.

[†]The hydrogen-bond length is measured between the heavy atoms of a pair of hydrogen-bond donor and acceptor. When two hydrogen bonds are formed, the upper value is for hydroxyl oxygen and the bottom value is for carbonyl oxygen.

[‡]The vibrational frequencies were calculated using B3LYP/6-31G(d) method on optimized structures. The energy was computed using B3LYP/6-311+G(2d,p) method.

[§]The C=O stretching frequency was initially coupled to that of its hydrogen-bond partner in forms of asymmetric and symmetric C=O stretching of two carbonyl groups. To remove this vibrational coupling, we isotopically labeled the hydrogen bond partner ($^{18}O={}^{13}C-{}^{15}N^2H_2$ for Asn/Gln, and $^{18}O={}^{13}C-{}^{18}O^2H$ for Asp/Glu), then calculate the C=O stretching frequency of the COOH group (butyric acid).

TABLE 2.5 The calculated hydrogen-bonding properties of butyric acid with charged amino acid side chain

Structure[*]	Amino Acid[*]	No. of H-bond[§]	H-bond Length[§] (Å)	H-bond Angle[§] (deg)	H-bond Energy[#] (kJ/mol)	C=O Frequency[#] (cm^{-1})
	Arg	~1.5	2.86 2.89	148.2 146.8	74.2	1707.4
	His	1	2.72	165.5	70.5	1707.1
	Lys	1	2.72	166.6	84.8	1689.2

The color codes, the methods for frequency and energy calculations, and the definition of hydrogen-bond lengths are the same as those in Table 2.4.

Tables 2.4 shows the calculated hydrogen bonding properties of a COOH group interacting with neutral, polar amino acids. A buried COOH group often forms hydrogen bond(s) with neutral polar group(s), since it is energetically unfavorable to bury charged groups in proteins. For a buried, isolated COOH group without any hydrogen bonding interactions, the calculated C=O stretching frequency is high, at 1776 cm^{-1}. When a COOH group forms one normal hydrogen bond with a polar side chain group, all nine computational results consistently show that the C=O stretching frequency is largely red-shifted from 1776 to 1733 − 1749 cm^{-1}. The strength of these hydrogen bonds is in the range between 17 and 45 kJ/mol. A significant part of this hydrogen-bond energy distribution is due to the fact that O=C−O**H** is stronger than **O**=C−OH by 14 to 24 kJ/mol. When a COOH group forms two hydrogen bonds with polar side chain groups (Asp/Gln, Arg, or Asp/Glu, see Table 2.4), the C=O stretching frequency is even further red-shifted to 1703 − 1710 cm^{-1}. The average hydrogen-bond strength is approximately 30 − 35 kJ/mol per hydrogen bond. These results support the notion that the C=O stretching frequency of a COOH group is a sensitive vibrational spectral marker for probing the number of hydrogen bonds of this COOH group. Figure 2.3 shows the distribution of calculated C=O stretching frequencies of COOH groups with zero, one and two hydrogen bonding interactions. The data shows that these distributions are well separated from one another, indicating that it is possible to identify the number of hydrogen bonds of a COOH group based on its C=O stretching frequency.

The results discussed above are based on well-formed hydrogen bond(s) with neutral, polar side chain groups. That is, the hydrogen-bond angles are optimal (170 − 180°, not deformed), and the hydrogen-bond strength is normal (17 − 45 kJ/mol, not extreme). Before we discuss experimental data on C=O stretching frequencies, we will first address three special cases. The first special case is the hydrogen bonding interaction, Cys−1, shown in Table 2.4. The hydrogen-bond strength for this interaction is exceptionally weak, only 5.9 kJ/mol. This hydrogen-bond dissociation energy is less than one fourth of the average calculated hydrogen-bond dissociation energy for single hydrogen bonding interactions (~26 kJ/mol) (Table 2.4). This is due to the fact that sulfur atom is large so that the hydrogen bond is long and weak, and that the hydrogen bonding in form of HO−C=**O** is weaker interaction than that of **H**O−C=O. Therefore, this hydrogen bonding interaction is closer to no hydrogen bond than to one hydrogen bond. Therefore, we count this interaction as no hydrogen bond. The C=O stretching frequency for Cys−1 is 1759 cm^{-1}. We thus classify the C=O frequency distribution of a COOH group with no hydrogen bond is as from 1759 to 1776 cm^{-1}.

The second special case is dealing with two deformed hydrogen bonds. When a COOH group forms two hydrogen bonds with either a neutral Lys or a neutral Tyr group as illustrated in Table 2.4, these two hydrogen bonds are deformed due to geometrical constrains. Their hydrogen-bond angles are 126°/166° for Lys and 145°/155° for Tyr, largely deviated from the optimal hydrogen-bond angle of 170° − 180°. Such deformations lead to significantly

reduced hydrogen-bond dissociation energy, ~49 kJ/mol. In comparison, the hydrogen-bond dissociation energy is on average 65 kJ/mol for two well-formed hydrogen bond interactions (Table 2.4), equivalent to 32.5 kJ/mol per hydrogen bond. The hydrogen-bond dissociation energy for two deformed hydrogen bonds, 49 kJ/mol, is 1.5 times of 32.5 kJ/mol. Therefore, for the calibration of this vibrational spectral marker, we classify such two deformed hydrogen bonds (Lys and Tyr) as forming 1.5 hydrogen bonds. The C=O stretching frequency for these deformed hydrogen bonding interactions is 1718 cm^{-1} for Lys and 1726 cm^{-1} for Tyr, lying between those C=O stretching frequencies for one and two hydrogen bonding interactions.

The third special case is that a COOH group forms hydrogen bond(s) with a charged side chain group. Most charged groups in proteins are solvent exposed. However, buried charged groups have been found in the active sites of proteins. Therefore, we examined the hydrogen bonding interactions of a buried COOH group with positively charged side chains (Arg+, Lys+, and His+) and with negatively charged side chains (COO− of Asp/Glu or a negatively charged phenolic oxygen of Tyr−).

In the case that a COOH group forms a hydrogen bond with a positively charged side chain of His or Lys (His+ or Lys+ in Table 2.5), the strength of such a hydrogen bond is very strong, 71 − 85 kJ/mol. This is approximately three times as strong as the average hydrogen-bond strength of a COOH group interacting with one polar hydrogen bonding partner, ~ 26 kJ/mol per hydrogen bond. When a COOH group is hydrogen bonded to a positively charged side chain of Arg, two

deformed hydrogen bonds are formed with the carbonyl oxygen. The hydrogen-bond dissociation energy for these two deformed bonds is 74.2 kJ/mol (Table 2.5), similar to that of single hydrogen bonding interactions. Since positively charged side chains do not have hydrogen-bond acceptors, no hydrogen bond is formed with the hydroxyl hydrogen of a COOH group. The C=O stretching frequency is 1707 cm^{-1} for hydrogen bonding interactions with Arg+ and His+, and further red-shifted to 1689 cm^{-1} for the hydrogen bonding interaction with Lys+. Therefore, a low C=O stretching frequency (1703 to 1710 cm^{-1}) is an indication of either two hydrogen bonds with neutral, polar side chain groups or hydrogen bonding interaction with a positively charged group. Alternatively speaking, a low C=O stretching frequency (1703 to 1710 cm^{-1}) indicates strong hydrogen bond interactions with 61 to 74 kJ/mol with either neutral or positively charged side chain groups.

In the case that the hydrogen-bond partner is negatively charged (Asp/Glu– or Tyr–), all our calculations lead to proton movements. Starting from an initial structure for a hydrogen bonding interaction between a COOH and a COO– group, we found that the labile proton is moved and equally shared by two carboxylic groups after structural optimization. Similarly, starting from an initial structure for a hydrogen bonding interaction between a COOH group and a negative charged phenolic group of Tyr–, the labile proton moves from the COOH group to the phenolic group after structural optimization. These results suggest that an isolated hydrogen bonding interaction between a COOH group and a negatively charged side chain groups is energetically unstable; additional

interactions must be present in order to stabilize any hydrogen bonding interaction between a COOH group and a negatively charged side chain group in proteins.

Experimental evidence for a vibrational spectral marker

In order to establish the C=O stretching frequency as a vibrational spectral marker for probing the numbers of hydrogen bonds of buried COOH groups, it is imperative to examine the related experimental evidence. We have carefully searched for proteins that their steady state structures have been studied using both high-resolution X-ray crystallography and infrared spectroscopy. The results from literature studies and crystal structure analysis are summarized in Table 2.6. Most of buried COOH groups form one hydrogen bond. The corresponding C=O stretching frequency is in the range of 1734 to 1748 cm^{-1}. This frequency distribution agrees extremely well with our computational results, 1733 to 1749 cm^{-1}, as shown in Table 2.4. Only one case is found for a buried COOH group to form no hydrogen bond, one case with two hydrogen bonds. Their C=O stretching frequencies are 1767 and 1708 cm^{-1} respectively. These data agree well with our computational results of 1759 to 1776 cm^{-1} for forming no hydrogen bond and 1703 to 1710 cm^{-1} for forming two hydrogen bonds (Table 2.4).

TABLE 2.6 Correlations between the C=O stretching frequency and the hydrogen-bond number of buried COOH groups in proteins

No. of H-bonds	IR Freq. λ (cm^{-1})	Protein*	COOH Group of	Type of H-bonds	H-bond Length (Å)	pdb code	Resolution
0	1767[1]	Rhodopsin	D83	-	NA	1L9H[2]	2.60 Å
1	1748[3]	CcO (B. H., oxidized)	E242	**H**O-C=O	3.14/3.22	1V54[4]	1.80 Å
	1746[3]	C*c*O (R. S., oxidized)	E286	**H**O-C=O	2.44/2.63	1M56[5]	2.30 Å
	1743[3]	CcO (B.H., reduced)	E242	**H**O-C=O	3.25	1V55[4]	1.90 Å
	1743[6]	Halorhodopsin	D141	**H**O-C=O	2.55	1E12[7]	1.80 Å
	1742[8]	Bacteriorhodopsin	D96	**H**O-C=O	2.79	1MOM[9]	1.43 Å
	1737[10]	PYP	E46	**H**O-C=O	2.58	1NWZ[11]	0.82 Å
	1735[12]	Rhodopsin	E134	HO-C=**O**	2.53/2.51	1L9H[2]	2.60 Å
	1735[13]	Reaction center	E104	**H**O-C=O	2.78	1M3X[14]	2.55 Å
	1734[8]	Bacteriorhodopsin	D115	HO-C=**O**	2.88	1MOM[9]	1.43 Å
	1734[1]	Rhodopsin	E122	**H**O-C=O	2.94/2.68	1L9H[2]	2.60 Å
2	1708[15]	C*c*O (T. T., oxidized)	Heme a$_3$ Propionate	**H**O-C=O HO-C=**O**	2.51 2.67	1EHK[16]	2.4 Å

[*]Abbreviations for proteins are: C*c*O for Cytochrome *c* oxidase in bovine heart (B. H.), Rhodobacter sphaeroides (R. S.), and Thermus thermophilus (T. T.); and PYP for photoactive yellow protein. References are quoted in numbers: [1] (Fahmy et al., 1993); [2] (Okada et al., 2002); [3] (Schmidt et al., 2004); [4] (Tsukihara et al., 2003); [5] (Svensson-Ek et al., 2002); [6] (Hutson et al., 2001); [7] (Kolbe et al., 2000); [8] (Sasaki et al., 1994); [9] (Lanyi and Schobert, 2002); [10] (Xie et al., 1996); [11] (Getzoff et al., 2003); [12] (DeLange et al., 1999); [13] (Breton et al., 1997); [14] (Camara-Artigas et al., 2002); [15] (Koutsoupakis et al., 2004); [16] (Soulimane et al., 2000).

[§]These are the two hydrogen-bond lengths corresponding to two structures of the same protein in the same state from one pdb data.

Comparisons of experimental studies using X-ray crystallography and infrared spectroscopy on intermediate states are not included in Table 2.6 due to the following reasons. First, there are only few proteins have been studied using both time-resolved or cold-trapped X-ray crystallography and time-resolved infrared spectroscopy techniques. Second, the experimental conditions for X-ray crystallography and infrared spectroscopy are often very different in temperature, with X-ray crystallography performed at much lower temperatures. If internal water molecules are involved in hydrogen bonding interaction, the internal water molecules may relocate in proteins as temperature changes. Therefore, it is difficult to make direct, reliable comparison. Third, the structural dynamics of proteins in crystals can be very different from those in solution state. Time-resolved infrared studies of photoactive yellow protein (PYP) in crystalline and solution states made in our lab show that the crystallization of PYP altered the proton transfer pathway and suppressed large protein conformational changes (Xie et al., 2001). The active site Glu46 is ionized in the solution upon chromophore protonation, but remains protonated in the crystalline state.

More time-resolved or cold-trapped infrared spectroscopic studies on the intermediate states of proteins have been made than those using X-ray crystallographic studies. Figure 2.3 A shows the histogram of the C=O stretching frequencies of all 31 buried COOH groups in the steady states of 25 different proteins that we found from the literature (see Table 2.8 for detailed information). The distributions of these C=O stretching frequencies form four clusters: three of them, shown as horizontal unfilled bars, overlap remarkably well with the

distributions of calculated C=O stretching frequencies with zero, one, and two hydrogen bonds (horizontal black bars). Experimental C=O stretching frequencies around 1725 cm^{-1} overlap well with the C=O frequency distribution (1718 to 1726 cm^{-1}) of two deformed hydrogen bonding interactions from our computational data (Table 2.4). Such good general agreements as well as the data in Table 2.6 support that the assignment of the C=O frequency as the vibrational spectral marker for probing the hydrogen-bond numbers of buried COOH groups.

Figure 2.3 B shows the histogram of the C=O stretching frequency of 31 buried COOH groups in the intermediate states of proteins (unfilled columns) from our extensive literature search. It is striking that the C=O stretching frequencies of these intermediate states are continuously distributed from 1706 to 1765 cm^{-1}. There are no clear gaps in the C=O stretching frequency distribution as observed from steady states of proteins. One plausible interpretation for this observation is that in the transient states more deformed hydrogen bonds (in forms of hydrogen-bond angle and/or hydrogen-bond length) are formed so that the C=O stretching frequencies are higher than those of corresponding normal hydrogen bonding interactions.

All vibrational frequency calculations presented in this paper were made in vacuum. This is due to the fact that the theories and methods for treating dielectric media are complex and have not been well developed. We have calculated the C=O stretching frequencies of COOH groups in 19 different dielectric media using Gaussian98 (Nie, 2002) and Gaussian03 (unpublished

results). We found that the C=O stretching frequencies calculated using Gaussian98 and Gaussian03 are very different, neither of them agree with experimental data. However, we found that if we treat the solvent explicitly in form of hydrogen bonding interactions between a COOH group and a solvent molecule, the C=O stretching frequency calculated in vacuum shows an excellent agreement with experimental data. For instance, the calculated C=O stretching frequency of a COOH group (of a butyric acid) that is hydrogen bonded to a chloride atom of a dichloromethane molecule (between the C**Cl** of dichloromethane and the O**H** of butyric acid) is 1748.9 cm^{-1}. In comparison, the experimentally measured
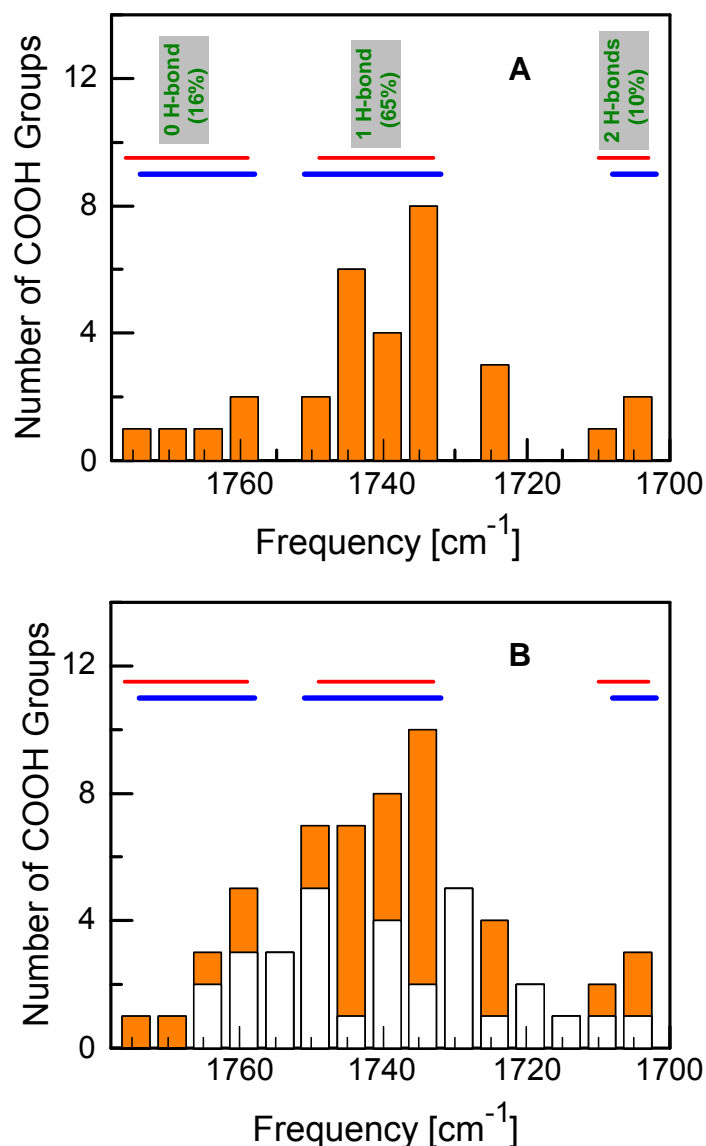
Figure 2.3 The distributions of calculated and experimental C=O stretching frequencies of a protonated carboxylic group (COOH). (A) The horizontal red bars show the distributions of calculated C=O stretching frequencies (see Table 2.4) with no hydrogen bond interaction (1759 – 1776 cm^{-1}), one hydrogen bond interaction (1733 – 1749 cm^{-1}), and two hydrogen bond interactions (1703 – 1710 cm^{-1}). The histogram illustrates the distribution of experimental C=O stretching frequencies of 31 buried COOH groups in the steady state structures of proteins. The horizontal blue bars depict the clustered distributions of experimental C=O stretching frequencies that correlates well with computational data. (B) The histograms for experimental C=O stretching frequencies of 31 buried COOH groups in intermediate states (unfilled columns), 31 buried COOH group in the steady states of proteins (filled orange columns), and the total distributions of both steady states and intermediate states (unfilled plus filled columns). The red and blue horizontal bars are the same as those in (A). The percentage of buried hydrogen bonds were calculated based on experimental data in the frequency regions that mirror the distribution of calculated C=O stretching frequencies with known hydrogen-bond information.

C=O stretching frequency in dichloromethane (ε=8.93) is 1750 cm^{-1} (Dioumaev and Braiman, 1995). Similarly our calculated C=O stretching frequency of a COOH group hydrogen bonded to the nitrogen atom of an acetonitrile molecule (between the C**N** of an acetonitrile and the O**H** of butyric acid) in vacuum is 1743.5 cm^{-1}. This value agrees extremely well with the experimentally measured C=O stretching frequency, 1743 cm^{-1}, of COOH model compounds in the acetonitrile solvent (ε=36.6) (Lascombe et al., 1962). These comparisons and good agreements between calculated and measured C=O stretching frequencies show that the C=O stretching frequency calculations in vacuum is accurate and reliable when solute-solvent interactions are treated explicitly in form of hydrogen bonding interactions. Furthermore, these agreements support the notion that the calculated C=O stretching frequencies of COOH groups in vacuum when hydrogen bonding interactions are treated explicitly can be applied directly to buried COOH groups in proteins, since the average, effective dielectric constant of protein interior is smaller than those of dichloromethane and acetonitrile.

It should be stressed that the vibrational spectral marker that is proposed in this paper is for use to probe buried COOH groups in proteins. It is not proper to generalize this spectral marker for a COOH group interacting with any non-biological molecules. Due to extreme chemical properties of some non-biological molecules, the C=O stretching frequency of a COOH group may be shifted out of our specified regions.

A linear correlation between the C=O stretching frequency and the hydrogen bonding strength
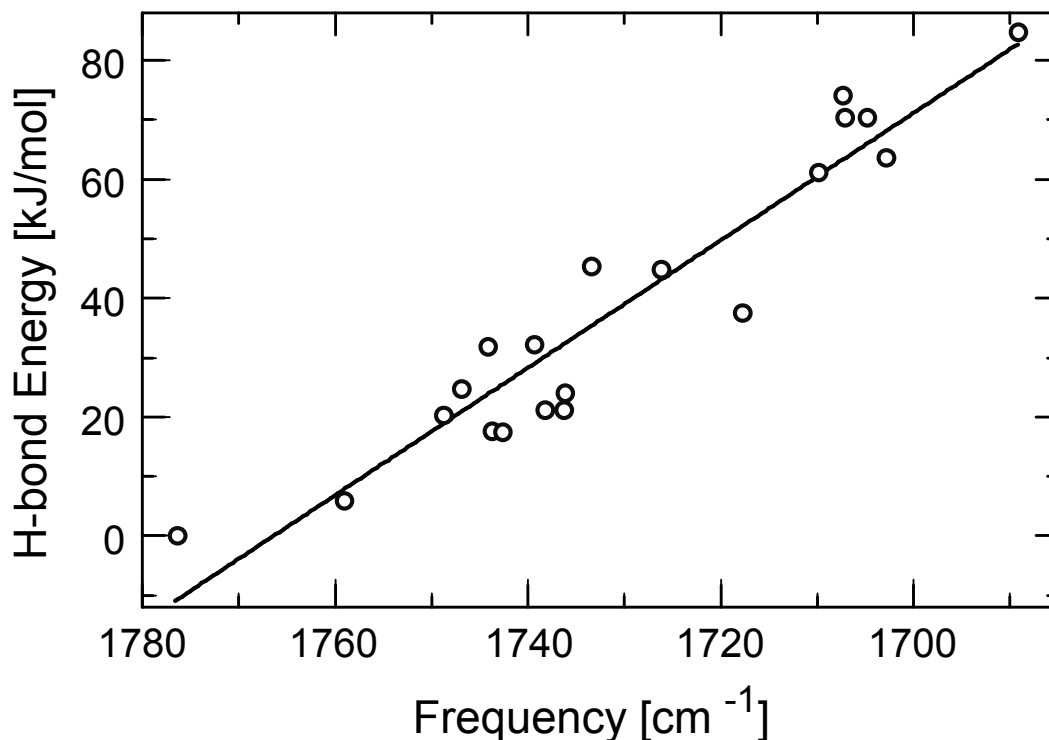
Figure 2.4 A correlation between the calculated hydrogen-bond dissociation energies and the calculated C=O stretching frequencies of a COOH group with neutral and positively charged side chain groups of amino acids (data shown in Tables 2.4 and 2.5). The calculated data are shown in open circles, while a linear straight line is the result of non-linear least square fitting. The root mean square deviation (RMSD) of the fitting is 7.49 kJ/mol, resulting in $E_{Hbond} = -1.07 \times (Freq - 1760) + 6.85$ with energy in unit of kJ/mol.

The data in Tables 2.4 and 2.5 show a general trend that the stronger the hydrogen bonding interactions, the larger the red-shift in the C=O stretching frequency. These data are shown in Figure 2.4. A non-linear least square fit of the data yields the following linear relation between the hydrogen-bond dissociation energy, E_{HB}, and the C=O stretching frequency, v_{CO}:

$$E_{HB} = 1.07 \times (1760 - v_{CO}) + 6.85 \qquad\qquad (2.2)$$

where the unit of vibrational frequency is cm^{-1}, and the unit of energy is kJ/mol. The root mean square deviation (RMSD) of the fitting is 7.5 kJ/mol. This deviation is much smaller than the distribution of the hydrogen-bond dissociation energies that is in the range between 0 and 85 kJ/mol (see Tables 2.4 – 2.5). Therefore, our computational data demonstrate that the C=O stretching frequency is a vibrational spectral marker for probing the approximate hydrogen-bond strength of a protonated carboxylic group in proteins.

It should be pointed out that a large part of the root mean square deviation from above fitting is due to the differences in the hydrogen-bond strength between the HO−C=**O** and **H**O−C=O forms of hydrogen bonding interaction. The calculated data shown in Tables 2.3 and 2.4 demonstrate that the hydrogen-bond dissociation energy for HO−C=**O** is consistently lower than that for **H**O−C=O by 15 to 24 kJ/mol between the same pair of hydrogen-bond partners, such as Thr−1 vs Thr−2, or His−1 vs His−2. Therefore, if we distinguish HO−C=**O** from **H**O−C=O, the root mean square deviation from the non-linear least square fitting is expected to be much smaller than 7.5 kJ/mol.

Vibrational spectral markers for 2D infrared spectroscopy
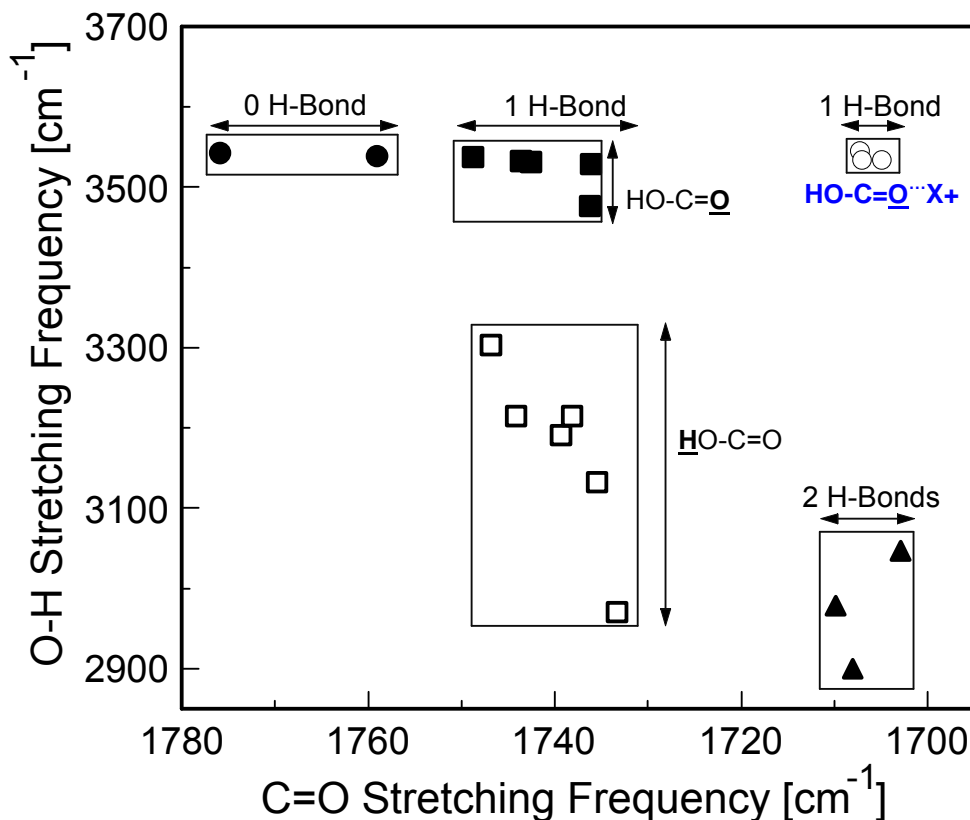
Figure 2.5 A 2D frequency distribution of the O−H and C=O stretching modes for probing the specific type of hydrogen bonding interactions in proteins: HO−C=O (filled circle), HO−C=**O** with neutral groups (filled square), **H**O−C=O (open square), HO−C=**O** with positively charged groups (open circle), and **H**O−C=**O** (filled triangle). The data are shown in Table 2.4.

As shown in Table 2.3, the O-H stretching frequency is also sensitive to hydrogen bonding interactions of a protonated carboxylic group. However this correlation between the O-H stretching frequency and specific hydrogen bonding interactions is different from that of the C=O stretching frequency. When both O-H and C=O stretching frequencies are used in a two-dimensional (2D) plot (see Figure 2.5), each data point represents one type of hydrogen bonding interaction. All five data points are separated by either C=O stretching or O-H stretching or

both in this 2D plot. That means, with such a 2D plot, it is possible to detect not only the number of hydrogen bonding interactions, but also the specific type of hydrogen bonding interactions.

The O-H stretching frequency is under-studied. Partly this is due to the fact that strong O-H absorption from solvent absorption overlaps with the O-H stretching modes from the COOH side chain groups. Some good experiments have been made and reported in this frequency region, in particular the studies on the structural dynamics of buried structural water molecules (Maeda, 2001; Maeda et al., 2003). In principle, this frequency range can provide valuable structural information for amino acid side chains as well. The use of hydrated protein films as samples greatly reduces solvent contribution to the infrared absorption in this region. Isotopic labeling helps to identify specific molecular groups under study.

The power of a vibrational spectral marker for structure-function studies of photoactive yellow protein

The ultimate goal of establishing vibrational spectral markers is to apply them to probe and characterize protein structures, in particular those of transient states during protein functions. Fourier transform difference infrared spectroscopy is a powerful method for monitoring structural developments of proteins while in action for studies of protein structure-function relationships. Vibrational spectral markers are essential to translate infrared spectral information into protein structural characterization.

For instance, the first infrared spectroscopic study of photoactive yellow protein (Xie et al., 1996) shows that the C=O stretching frequency of Glu46 at the active site is shifted from 1740 to 1732 cm^{-1} upon chromophore photo-isomerization at 100 K. The crystal structure of PYP in the receptor state (pG) reveals that Glu46 forms one hydrogen bond with the negatively charged phenolic oxygen of the chromophore prior to photo-isomerization. However, there was no X-ray structure for the cold-trapped photoproduct of PYP at that time. The infrared data (Xie et al., 1996) unambiguously shows that the Glu46 remains hydrogen bonded after photo-isomerization. This structural information was crucial to propose, test, and confirm a new chromophore structure for the photo-isomerization product of PYP. This significantly different photo-isomerization mechanism has direct implications to the subsequent structural dynamics of PYP during receptor activation. This example demonstrates the power of infrared spectroscopic techniques when they are combined with vibrational spectral markers for the study of protein structure-function relationships.

Furthermore, the use of this vibrational spectral marker, the C=O stretching frequency of Glu46 allow us to make two predictions: (i) the existence of a pre-signaling state (pB'), and (ii) the ionized Glu46 in pB' as an electrostatic epicenter for protein quake as a mechanism for PYP receptor activation. Later, by using four vibrational spectral markers (Xie et al., 2001), including the C=O stretching band of Glu46 combined with time-resolved step-scan FTIR spectroscopy, we unambiguously proved that indeed there exists a pre-signaling state pB' with predicted structural characteristics. At the same time, the study

provided strong supporting evidence that the COO$^-$ group of Glu46 in the pB'

state serves as an electrostatic epicenter for protein quake and receptor

activation. In conclusion, vibrational spectral markers paved the way for the

infrared difference spectroscopy to reach its utmost power in extracting and

revealing structural information of not only the steady state of proteins, but more

importantly of the short-lived functional intermediates of proteins.

2.4 Conclusions

Our computational studies on the hydrogen bonding properties of a buried

COOH group show that the hydrogen bonding interaction with the hydroxyl

hydrogen atom as the hydrogen-bond donor, <u>H</u>O–C=O, is the strongest among

three different types of one hydrogen-bond interactions (Table 2.1). This

hydrogen bonding interaction remains strong in a range of dielectric media (Table

2.2) with dielectric constants of 1 (vacuum) to 5.6 (chlorobenzene). The hydrogen

bonding interaction with the carbonyl oxygen atom as the hydrogen bond donor,

HO–C=<u>O</u>, is significantly stronger than that of H<u>O</u>–C=O. Therefore, this

interaction, HO–C=<u>O</u>, dominates the hydrogen bonding interaction in presence of

a hydrogen-bond donor in the vicinity. This hydrogen-bond dissociation energy is

large in the dielectric medium of benzene with a dielectric constant of 2.2, but is

small in the dielectric medium of ether with a dielectric constant of 4.3 or higher.

We conclude that low dielectric protein interiors make the hydrogen bonding

interactions stronger and more stable than high dielectric interiors.

Our density function theory based ab initio computational studies reveal a

strong correlation between the C=O stretching frequency and the hydrogen-bond

number of COOH groups that are hydrogen bonded to polar groups of amino acid side chains (Table 2.4). This correlation is further supported by available experimental evidence from 12 buried COOH groups in the steady states of 8 different proteins (Table 2.6): infrared vibrational frequencies of buried Asp and Glu residues in proteins and their hydrogen bonding interactions based on the protein crystal structures. We provided both computational and experimental evidence that support the establishment of the C=O stretching frequency of buried COOH groups as a vibrational spectral marker for probing the hydrogen-bond number (Figure 2.3): approximately 1759 to 1776 cm^{-1} for zero, 1733 to 1749 cm^{-1} for one, and 1703 to 1710 cm^{-1} for two hydrogen bonding interactions. This correlation is valid when the hydrogen-bond partners are polar side chain groups of amino acids, but not valid when the hydrogen-bond partners are positively charged (Table 2.5). In this case, it is better to apply an approximate linear correlation between the C=O stretching frequency and the net hydrogen-bond strength (Figure 2.4): $E_{HB} = -1.07 \times (v_{CO} - 1760) + 6.85$. This correlation applies to both polar and positively charged side chains as hydrogen-bond partners over a broad distribution of the C=O stretching frequency (1689 to 1776 cm^{-1}). Furthermore, we proposed that a two-dimensional infrared spectroscopy, C=O stretching vs. O−H stretching, may identify the specific type of hydrogen bonding interaction (Figure 2.5). This vibrational spectral marker for hydrogen bonding interaction is expected to enhance the power of time-resolved Fourier transform infrared spectroscopy in characterizing the structures of functionally

important intermediate states of proteins and in elucidating the functional mechanism of proteins.

2.5 Additional materials

Solving difficulty in geometry optimization in energy calculations with fixed distance

Table 2.7 shows detailed Gaussian03 energy calculation for Figure 2.2. All the energy calculations are based on geometry optimized structure with the distance between two heavy atoms of the hydrogen bond fixed at a number of values from 2.4 – 4.0 Å. In most of the calculations, we had difficulty in optimizing geometry using B3LYP/6-31G(d) method directly (Method C in the table). This is due to the initial force constants estimated by the optimization procedure differs substantially from the actual values. By default, a geometry optimization starts with an initial guess for the second derivative matrix derived from a simple valence force field. The approximate matrix is improved at each step of the optimization using the computed first derivatives (Foresman and Frisch, 1996). When this initial estimation is poor, we need to compute the force constants at the initial point and then optimize the structure using a less expensive method. In the next step, the force constants were read from optimized structure from previous step and the structures are further optimized using B3LYP/6-31G(d) method (see Method A and B). In a few cases, we calculated initial force constants and optimized structures using HF/3-21G, then read force constants and optimize again using larger basis sets, and at last read force constants and

optimize using more expensive method B3LYP (see Method D). However, there

are still three energy calculations that did not converge with this strategy.

TABLE 2.7   Computational methods for geometry optimization and energy calculation

Distance (Å)	HO–C=**O** (2.87 Å)[#]		**H**O–C=O (2.75 Å)[#]		H**O**–C=O (2.94 Å)[#]	
	Method	Energy[§] (kJ/mol)	Method	Energy[§] (kJ/mol)	Method	Energy[§] (kJ/mol)
2.30	N/A		[¶]A	7.4	N/A	N/A
2.40	[‖]B	13.8	B	−12.6		
2.50	B	−1.7	[*]C	−23.5	[**]D (failed)	N/A
2.60	B	−10.6	A	−29.0	D	-1.3
2.70	B	−15.5	A	−31.1	D (failed)	N/A
2.75	N/A	N/A	C	−31.3	N/A	N/A
2.80	B	−17.7	B	−31.1	B	−9.3
2.87	C	−18.3	N/A	N/A	N/A	N/A
2.90	B	−18.3	B	−30.0	B	−10.8
2.94	N/A	N/A	N/A	N/A	C	−11.1
3.00	B	−17.9	B	−28.2	D (failed)	N/A
3.20	C	−15.6	B	−24.8	A	−10.4
3.60	C	−9.9	B	−16.8	B	−6.8
4.00	C	−5.8	B	−10.2	B	−4.4

[#]Optimized hydrogen-bond length between two heavy atoms in hydrogen bond interactions

[§]Hydrogen bond dissociation energy

[¶]A: HF/6-31G(d) Opt=(CalCFC, Z-matrix)

Additional step 1: B3LYP/6-31G(d) Opt=(ReadFC, Z-matrix)

<superscript>||</superscript>B: HF/3-21G Opt=(CalCFC, Z-matrix)

Additional step 1: B3LYP/6-31G(d) Opt=(ReadFC, Z-matrix)

<superscript>*</superscript>C: B3LYP/6-31G(d) Opt=Z-matrix

<superscript>**</superscript>D: HF/3-21G Opt=(CalCFC, Z-matrix)

      Additional step 1: HF/6-31G(d) Opt=(ReadFC, Z-matrix)

      Additional step 2: B3LYP/6-31G(d) Opt=(ReadFC, Z-matrix)

Experimental infrared frequency of C=O stretching mode in proteins

We had found from literatures that there are only a few proteins that have been studied using both X-ray crystallography and infrared spectroscopy. We summarized these experimental infrared C=O stretching frequencies of buried COOH group in proteins and crystal structure analyses in Table 2.6. These experimental data agreed well with our computational results of C=O stretching frequency shift with number of hydrogen bonds.

There are much more time-resolved or cold-trapped infrared spectroscopic studies on the intermediate states of proteins of which the crystal structures were not available or at low resolution. We had searched a large number of literatures that contains experimental infrared C=O stretching frequency of buried in proteins in its steady/intermediate states. The results are summarized in Table 2.8 and also used to make histogram in Figure 2.3. With established vibrational spectral marker (C=O stretching frequency) for detecting the hydrogen-bonding status of protonated Asp and Glu, it serves general purposes for researchers that even though the protein structures may not be available, one may be able to tell by simply looking at the infrared frequency of COOH in $1775 - 1700 \text{ cm}^{-1}$ region

whether there is hydrogen-bond interaction(s) involving protonated Asp and Glu, the number of H-bond(s), and the strength of the H-bond(s).

TABLE 2.8 Experimental C=O stretching frequency of buried and protonated carboxylic group in proteins

No.	IR Freq. λ (cm^{-1})	Protein[*]	State[*]	COOH Group of	Reference
1	1774	Caa$_3$ (A. A)	Oxidized	Unknown	(Hellwig et al., 2003)
2	1767	R	Ground	D83	(Fahmy et al., 1993)
3	1765	psR-II	M	D75	(Furutani et al., 2002)
4	1763	SRI	Ground	D76	(Rath et al., 1996)
5	1762	bR	M	D85	(Dioumaev, 2001)
6	1761	Cbd	Oxidized	Unknown	(Zhang et al., 2002)
7	1761	HsSRII	SRII$_{540}$	D73	(Bergo et al., 2000)
8	1758	NR (NOP-1)	M	Unknown	(Bergo et al., 2002)
9	1758	Caa$_3$ (A. A)	Reduced	Unknown	(Hellwig et al., 2003)
10	1757	psR-II	O	Unknown	(Furutani et al., 2004b)
11	1754	bR	N	D85	(Braiman et al., 1996)
12	1754	bR	O	D85	(Braiman et al., 1996)
13	1754	NR (NOP-1)	N	Unknown	(Bergo et al., 2002)
14	1751	SRI	S373	D76	(Rath et al., 1996)
15	1751	Cbd	Reduced	Unknown	(Zhang et al., 2002)
16	1750	R	MII	D83	(Fahmy et al., 1993)
17	1749	HsSRII	SRII$_{540}$	Unknown	(Bergo et al., 2000)
18	1748	CcO (B. H)	Oxidized	E242	(Schmidt et al., 2004)
19	1748	bR	L	D96	(Dioumaev, 2001)
20	1747	PSII (Spinach)	S2	Unknown	(Hillier and Babcock, 2001)
21	1746	CcO (R. S.)	Oxidized	E286	(Schmidt et al., 2004)
22	1746	CcO (P. D.)	Oxidized	E278	(Hellwig et al., 1998)
23	1746	Caa$_3$ (A. A)	Reduced	Unknown	(Hellwig et al., 2003)
24	1743	CcO (B. H.)	Reduced	E242	(Schmidt et al., 2004)

25	1743	sHR	Ground	D141	(Hutson et al., 2001)
26	1742	bR	Ground	D96	(Dioumaev, 2001)
27	1742	bR	M	D115	(Dioumaev, 2001)
28	1742	pHR	Ground	D156	(Hackmann et al., 2001)
29	1740	Thioredoxin (E. Coli)	Reduced	D26	(Rollin-Genetet et al., 2004)
30	1739	bR	N	D115	(Dioumaev, 2001)
31	1739	HsSRII	Ground	Unknown	(Bergo et al., 2000)
32	1739	NR (NOP-1)	Ground	Unknown	(Bergo et al., 2002)
33	1738	PSII (T. E.)	S1	Unknown	(Noguchi and Sugiura, 2003)
34	1737	PYP	Ground	E46	(Xie et al., 1996)
35	1737	PSII (Spinach)	S1	Unknown	(Hillier and Babcock, 2001)
36	1736	bR	M	D96	(Dioumaev, 2001)
37	1736	Cbd	Oxidized	Unknown	(Zhang et al., 2002)
38	1735	R	Ground	E134	(DeLange et al., 1999)
39	1735	RC (R. S.)	QA	E104	(Breton et al., 1997)
40	1735	CcO (R. S.)	Reduced	E286	(Schmidt et al., 2004)
41	1734	CcO (P. D.)	Reduced	E278	(Hellwig et al., 1998)
42	1734	NR (NOP-1)	K	D161	(Furutani et al., 2004a)
43	1734	bR	Ground	D115	(Dioumaev, 2001)
44	1734	R	Ground	E122	(Fahmy et al., 1993)
45	1732	Caa$_3$ (A. A)	Oxidized	Unknown	(Hellwig et al., 2003)
46	1730	HsSRII	SRII$_{540}$	Unknown	(Bergo et al., 2000)
47	1730	Cbo$_3$ (E. Coli)	Light	E286	(Puustinen et al., 1997)
48	1728	bR	L	D115	(Dioumaev, 2001)
49	1728	RC (R. S.)	QA$^-$	E104	(Breton et al., 1997)
50	1728	MDP	Active	Unknown	(Bartl and Zundel, 1997)
51	1726	Cbo$_3$ (E. Coli)	Dark	E286	(Puustinen et al., 1997)
52	1725	Cbd	Reduced	Unknown	(Zhang et al., 2002)
53	1723	PSII (T. E.)	S1	Unknown	(Noguchi and Sugiura, 2003)
54	1722	psR-II	Ground	Unknown	(Furutani et al., 2004b)
55	1720	NR (NOP-1)	M	Unknown	(Bergo et al., 2002)

56	1718	PSII (Spinach)	S2	Unknown	(Hillier and Babcock, 2001)
57	1712	R	MII	E113	(Fahmy et al., 1994)
58	1708	PSII (Spinach)	S1	Unknown	(Hillier and Babcock, 2001)
59	1708	CcO (T. T.)	Oxidized	Heme a$_3$ Propionate	(Wolpert et al., 2004)
60	1706	CcO (R. S.)	Oxidized	Unknown	(Schmidt et al., 2004)
61	1704	PSII (T. E.)	S1	Unknown	(Noguchi and Sugiura, 2003)
62	1702	CcO (B. H.)	Oxidized	Unknown	(Schmidt et al., 2004)

[*]Caa$_3$ (A. A.) is abbreviation of cytochrome aa$_3$ from Acidianus ambivalens; R is of bovine rhodopsin; psR-II is of pharaonis phoborhodopsin; SRI is of sensory rhodopsin I; bR is of bacteriorhodopsin; Cbd is of cytochrome bd; HsSRII is of sensory rhodopsin II from Halobacterium salinarum; NR is of Neurospora rhodopsin (also called NOP-1); CcO is of cytochrome c oxidase from bovine heart (B. H.), Paracoccus denitrificans (P. D.), Rhodobacter sphaeroides (R. S.), and Thermus thermophilus (T. T.); PSII is of photosystem II from Spinach and from Thermosynechococcus elongatus; pHR is of halorhodopsin from Natronobacterium pharaonis; sHR is of halorhodopsin from Halobacterium salinarum; PYP is photoactive yellow protein; RC is of reaction center from Rhodobacter sphaeroides; MDP is of maltodextrinphosphorylase; and Cbo$_3$ is of cytochrome bo$_3$.

[*]Indication of each protein in ground state, reduced/oxidized state, or intermediate state. K, L, M, N, O denote intermediate states in pharaonis phoborhodopsin or bacteriorhodopsin or neurospora rhodopsin. SRII$_{540}$ is the intermediate state of sensory rhodopsin II with absorption maximum at 540 nm. S373 is intermediate state of sensory rhodopsin I with absorption maximum at 373 nm. MII is the intermediate state of bovine rhodopsin. S1 and S2 states are the intermediate states after one and two laser flashes on

photosystem II. QA⁻/QA is the light–induced/dark state of quinone acceptor in reaction center. Light/dark is the light–induced/dark state of fully reduced cytochrome bo$_3$. Active state of maltodextrinphosphorylase (MDP) is the state when substrate phosphate glucose-1-methylenephosphonate is added to MDP.
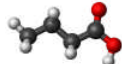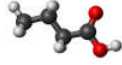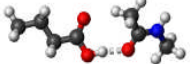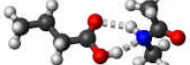
Hydrogen-bonding properties of COOH group interacting with backbone and water molecules

In addition to the hydrogen bonding interactions between Asp/Glu and other amino acid side chains, there are other possible hydrogen bonding interactions in proteins: protein backbone and water molecule(s).

Tables 2.9 shows the calculated hydrogen bonding properties of a COOH group interacting with a backbone model compound. When the hydroxyl hydrogen (<u>H</u>O–C=O) forming one hydrogen bond with carbonyl oxygen of the backbone molecule, the C=O frequency is 1735.5 cm⁻¹. It is in consistency with the frequency range of ~1733 – 1749 cm⁻¹ for one hydrogen-bonding interaction of our proposed vibrational spectral marker. The hydrogen bond dissociation energy is 42.6 kJ/mol, 11.3 kJ/mol higher than 31.3 kJ/mol for <u>H</u>O–C=O in Table 2.1. This is due to carbonyl oxygen forming a weak hydrogen bond with C –H of methyl group of the backbone molecule. When the carboxylic group forming two deformed hydrogen bonds with N – H of the backbone molecule, the C=O frequency is 1726.5 cm⁻¹. It is between C=O stretching frequencies for one and two hydrogen bonding interactions. This deformation yields reduced hydrogen bond dissociation energy of 22.8 kJ/mol. In comparison to the average hydrogen-bond strength of approximately 30 – 35 kJ/mol per hydrogen bond, such

deformed geometry is much less stable that might lead to breaking one hydrogen bond to form a more stable geometry of <u>H</u>O−C=O. Noticeably when the hydroxyl hydrogen is in a different orientation (first row of Table 2.9), the C=O stretching frequency is 1809.2 cm^{-1}. It is much higher than the frequency of carboxylic group both in computational studies of normal orientation (second row in Table 2.9, 1776.4 cm^{-1}) and in experimental studies ((Hellwig et al., 2003), 1774 cm^{-1}). The energy is 21.5 kJ/mol higher indicating that its population is much smaller than that of normal orientation according to Boltzmann distribution. Therefore, the conformation in the second row of Table 2.9 is dominantly present.

TABLE 2.9 Calculated hydrogen-bonding properties of a COOH group interacting with backbone

Structure	H-Bond Partner	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy (kJ/mol)	C=O Frequency (cm^{-1})
	N/A	0	N/A	N/A	-21.5	1809.2
	N/A	0	N/A	N/A	0.0	1776.4
	bb-CO	1	2.71	170.2	42.6	1735.5
	bb-NH	~1.5	2.92 2.98	141.2 146.8	22.8	1726.5

The color codes, the methods for frequency and energy calculations, and the definition of hydrogen-bond lengths are the same as those in Table 2.4.
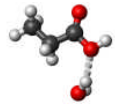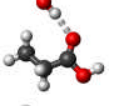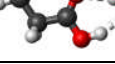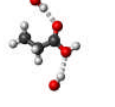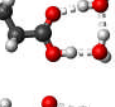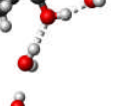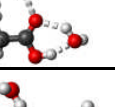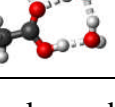
Tables 2.10 shows the calculated hydrogen bonding properties of a COOH group interacting with 1 − 3 water molecules. With zero hydrogen bond,

the vibrational frequency of COOH group is 1778 cm^{-1}, slightly higher than that of butyric acid with zero hydrogen bond 1776 cm^{-1}. This is due to the difference in the main chain. When the carbonyl oxygen (HO–C=$\underline{\text{O}}$) forms one hydrogen bond with a water molecule, the C=O frequency is 1748 cm^{-1}. It is also consistent with the frequency range of ~1733 – 1749 cm^{-1} for one hydrogen-bonding interaction of our proposed vibrational spectral marker. The hydrogen-bond dissociation energy is 21.5 kJ/mole, close to 18.3 kJ/mol for HO–C=$\underline{\text{O}}$ in Table 2.1. Whereas the hydroxyl oxygen H$\underline{\text{O}}$–C=O forms one hydrogen bond, the vibrational frequency of COOH is up–shifted to 1787 cm^{-1}, exactly the same as that of butyric acid hydrogen-bonded with a methanol group (H$\underline{\text{O}}$–C=O) in Table 2.3. It was discussed earlier in the chapter that H$\underline{\text{O}}$–C=O and HO–C=$\underline{\text{O}}$ are sterically similar and the strength of HO–C=$\underline{\text{O}}$ is stronger than H$\underline{\text{O}}$–C=O by ~7 kJ/mole, therefore, it is more likely for a buried COOH group to form a hydrogen bond in the form of HO–C=$\underline{\text{O}}$ rather than H$\underline{\text{O}}$–C=O. In the case of hydroxyl hydrogen forming one hydrogen bond with a water molecule, the initial structure for Gaussian03 calculation was made that there was only one hydrogen bond. The geometry optimized structure, however, shows that the COOH group forms two deformed hydrogen bonds with the water molecule in that the water molecule is small and it may serve as both proton donor and acceptor ($\underline{\text{H}}$O–C=$\underline{\text{O}}$-a). The vibrational frequency is 1724 cm^{-1} and the hydrogen-bond dissociation energy is 40.5 kJ/mole. It is in the category of 1.5 hydrogen bonds (Table 2.4).

When the COOH group forms two hydrogen bonds with two water molecules ($\underline{\text{H}}$O–C=$\underline{\text{O}}$-b), the C=O frequency is 1704 cm^{-1}, in consistency with

1703 – 1710 cm^{-1} for two hydrogen-bonding interactions of our proposed vibrational spectral marker. And the hydrogen-bond dissociation energy, 67.9 kJ/mole, is also in the range for two hydrogen bonds (Table 2.4). In the case of H**O**–C=**O**, the hydrogen-bond dissociation energy is significantly reduced in comparison with two hydrogen bonds (**H**O–C=**O**-b), indicating it is much less stable. When three hydrogen bonds (**HO**–C=**O**, **H**O–C=**O**–c, and **H**O–C=**O**–d) are formed, the C=O frequency is down–shifted upon combined frequency shifts: **HO**–C=**O** is the combination of H**O**–C=O and **H**O–C=**O**-a, **H**O–C=**O**–c is the combination of HO–C=**O** and **H**O–C=**O**-a, and **H**O–C=**O**–d is the combination of HO–C=**O** and **H**O–C=**O**-b. The hydrogen-bond dissociation energy is also under the same combination effects. Although the fact is that it is energetically favorable for a buried COOH group to form as many hydrogen bonds as possible, our experimental evidence (Table 2.6) shows that most of the buried COOH groups form one hydrogen bond and only one case with two hydrogen bonds. In the case of two hydrogen bonds, it is in the same configuration as **H**O–C=**O**-b. Therefore, the H**O**–C=**O** type and three hydrogen bonding interactions are not likely.

TABLE 2.10 Calculated hydrogen-bonding properties of a COOH group interacting with water molecule(s)

Structure	No. of H-bond	Type of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy (kJ/mol)	C=O Frequency (cm^{-1})
	0	HO–C=O	N/A	N/A	N/A	1777.9
	1	H**O**–C=O	2.98	155.6	14.7	1787.2
	1	HO–C=**O**	2.92	169.7	21.5	1747.7
	~1.5	**H**O–C=**O**–a	2.70* 2.75	157.7 141.3	40.5	1724.3
	2	H**O**–C=**O**	2.99 2.94	154.4 169.8	31.0	1757.3
	2	**H**O–C=**O**–b	2.66 2.76	175.7 166.6	67.9†	1704.1§
	~2.5	**HO**–C=**O**	2.96 2.67 2.75	158.3 158.7 138.7	56.3	1735.7
	~2.5	**H**O–C=**O**–c	2.69 2.77 2.92	159.4 137.6 170.2	61.8	1695.1§
	3	**H**O–C=**O**–d	2.64 2.78 2.89	175.2 165.7 179.3	84.9†	1672.7§

The color codes, the methods for frequency and energy calculations, and the definition of hydrogen-bond lengths are the same as those in Table 2.4.

*The hydrogen-bond length is measured between the heavy atoms of a pair of hydrogen-bond donor and acceptor. When three hydrogen bonds are formed, the upper value is for hydroxyl oxygen and the middle value and the bottom value are for carbonyl oxygen.

†The H – Bond energy was obtained by subtracting the total H – Bond energy from the H – Bond energy of two water molecules.
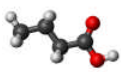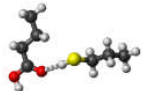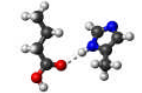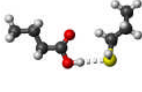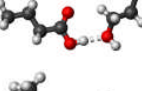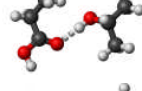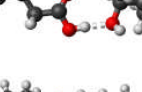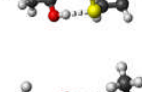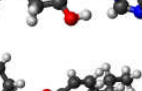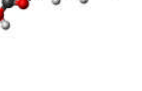
§The C=O stretching frequency was initially coupled to O–H bending frequency of the water molecules. To remove this vibrational coupling, we isotopically labeled the water molecules ($^{18}O–^{2}H_2$), then calculate the C=O stretching frequency of the COOH group.
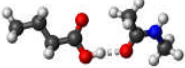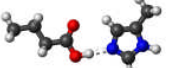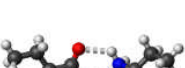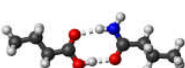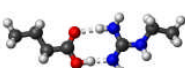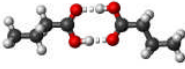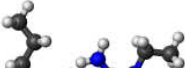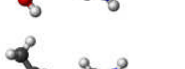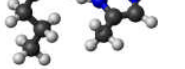
Isotopic shifts of the vibrational spectral marker

Infrared measurements of biological molecules are often impeded by the intense absorption of water near the 1650 and 3300 cm^{-1} bands. The amide I mode of protein falls very near or right under the 1650 cm^{-1} band of water and it has much smaller absorption than that of water. For this reason most FTIR measurements of biological samples are made in D_2O. FTIR difference spectroscopy, combined with isotopic labeling (H $\rightarrow$ D, $^{16}O \rightarrow ^{18}O$, $^{12}C \rightarrow ^{13}C$, and $^{14}N \rightarrow ^{15}N$) and site-directed mutagenesis of individual residues in the protein, makes it possible to assign peaks to specific single residue. To provide a quantitative view of isotopic shift of vibrational spectral marker, we calculated vibrational frequency of C=O stretching upon isotopic labeling of COOH $\rightarrow$ COOD, COOH $\rightarrow$ $C^{18}O^{18}OH$, and COOH $\rightarrow$ $^{13}COOH$. The results are shown in Table 2.11 and Table 2.12. Upon COOH $\rightarrow$ COOD deuteration, the frequency down–shift of C=O stretching is in a range of 7 – 9 cm^{-1} except for those deformed hydrogen bonding interactions (12 – 24 cm^{-1}) and hydrogen bonding interactions with positively charged Arg (15 cm^{-1}) or Lys (11 cm^{-1}). Upon $C^{18}O^{18}OH$ isotopic labeling, the C=O stretching frequency of most hydrogen

bonding interactions is down–shifted by 28 – 33 cm^{-1}. Upon ^{13}COOH isotopic labeling, the C=O stretching frequency of all but one (Arg+) hydrogen bonding interactions is further down–shifted by 37 – 44 cm^{-1}. These frequency shifts are insensitive to hydrogen-bonding interactions.

TABLE 2.11 Isotopic labeling effect on vibrational spectral marker of COOH

Structure	H-Bond Partner*	ν_{COOH} (cm^{-1})	ν_{COOD} (cm^{-1})	$\Delta\nu$ (cm^{-1})	$\nu_{C^{18}O^{18}OH}$ (cm^{-1})	$\Delta\nu$ (cm^{-1})	ν^{13}_{COOH} (cm^{-1})	$\Delta\nu$ (cm^{-1})
	N/A	1776.4	1768.2	8.2	1743.1	33.3	1732.1	44.3
	CYS – 1	1759.1	1750.4	8.7	1727.2	31.9	1715.5	43.6
	His – 1a	1748.8	1740.7	8.1	1717.5	31.3	1705.3	43.5
	CYS – 2	1746.9	1739.7	7.2	1715.3	31.6	1703.5	43.4
	SER – 2	1744.2	1735.1	9.1	1713.2	31.0	1701.2	43.0
	SER – 1	1743.7	1735.1	8.6	1712.7	31.0	1700.3	43.4
	THR – 1	1742.6	1734.1	8.5	1711.7	30.9	1699.3	43.3
	THR – 2	1739.4	1730.2	9.2	1708.6	30.8	1696.6	42.8
	MET	1738.2	1730.6	7.6	1707.1	31.1	1695.2	43.0
	HIS – 1b	1736.3	1727.6	8.7	1705.7	30.6	1693.1	43.2
	TYR – 1	1736.2	1727.4	8.8	1705.7	30.5	1693.1	43.1
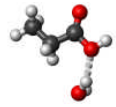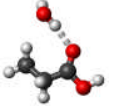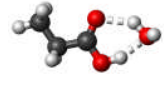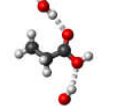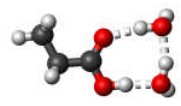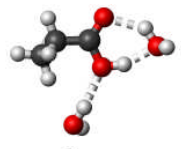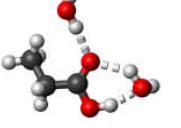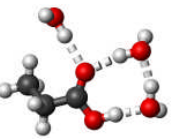
	bb–CO	1735.5	1728.4	7.1	1707.1	28.4	1691.5§	44.0
	HIS – 2	1733.4	1723.6	9.8	1702.9	30.5	1690.8	42.6
	bb–NH	1726.5	1714.7	11.8	1696.9	29.6	1684.1	42.4
	LYS	1726.2	1706.0	20.2	1698.5	27.7	1686.1	40.1
	TYR	1717.8	1699.8	18.0	1690.3	27.5	1676.7	41.1
	ASN/ GLN	1709.9§	1702.0§	7.9	1681.9§	28.0	1669.6§	40.3
	ARG	1693.3§	1670.5§	22.8	1667.5§	25.8	1655.5§	37.8
	Asp/Glu	1702.9§	1693.7§	9.2	1677.3§	25.6	1665.6§	37.3
	ARG+	1707.4	1692.4	15.0	1683.7	23.7	1676.1	31.3
	HIS+	1707.1	1698.1	9.0	1677.7	29.4	1664.2	42.9
	HIS+	1705.0	1695.9	9.1	1675.7	29.3	1662.2	42.8
	LYS+	1689.2	1678.6	10.6	1657.7§	31.5	1647.1	42.1

The color codes, the methods for frequency and energy calculations, and the definition of hydrogen-bond lengths are the same as those in Table 2.4.

§The C=O stretching frequency was initially coupled to C=O stretching or N−H bending of its hydrogen-bond partners. To remove this vibrational coupling, we isotopically

labeled the hydrogen bond partner ($^{18}O=^{13}C$ for backbone, $^{18}O=^{13}C-^{15}N^2H_2$ for Asn/Gln, $^{18}O=^{13}C-^{18}O^2H$ for Asp/Glu, $^{15}N^2H-^{13}C-^{15}N^2H_2$ for neutral Arg, and $^{15}N^2H_3$ for positively charged Lys+), then calculate the C=O stretching frequency of the COOH group.

TABLE 2.12 Isotopic labeling effect on vibrational spectral marker of COOH

Structure	No. of H-Bond	v_{COOH} (cm^{-1})	v_{COOD} (cm^{-1})	Δv (cm^{-1})	$v_{C^{18}O^{18}OH}$ (cm^{-1})	Δv (cm^{-1})	v^{13}_{COOH} (cm^{-1})	Δv (cm^{-1})
	0	1777.9	1770.2	7.7	1745.5	32.4	1734.2	43.7
	1	1787.2	1779.2	8.0	1754.3	32.9	1743.3	43.9
	1	1747.7	1739.3	8.4	1716.7	31.0	1704.4	43.3
	~1.5	1724.3	1701.5	22.8	1698.9	25.4	1686.7	37.6
	2	1757.3	1748.3	9.0	1726.0	31.3	1713.9	43.4
	2	1704.1§	1695.0	9.1	1675.2§	28.9	1661.7§	42.4
	~2.5	1735.7	1712.0	23.7	1709.5	26.2	1696.9	38.8
	~2.5	1695.1§	1674.5	20.6	1669.4§	25.7	1655.1§	40.0
	3	1672.7§	1662.2	10.5	1645.6§	27.1	1630.6	42.1

§The C=O stretching frequency was initially coupled to O–H bending frequency of the water molecules. To remove this vibrational coupling, we isotopically labeled the water molecules (^{18}O–$^{2}H_2$), then calculate the C=O stretching frequency of the COOH group.

Solvent effects on the vibrational frequency of a buried COOH group

The vibrational frequency of a buried COOH group in proteins is sensitive to the local dielectric environment. To investigate the influence of of a local dielectric medium on the C=O stretching frequency, we calculated the vibrational frequencies of butyric acid in 19 different dielectric media using Gaussian98 (Frisch et al., 1998) and later Gaussian03. The computational results of C=O stretching frequency in various solvents are listed in Table 2.13. We found that the C=O stretching frequencies calculated by two versions of Gaussian package are very different. To determine which set of vibrational frequencies is more accurate in comparison with experimental data, we performed FTIR absorption measurements of butyric acids in various solvents. The experimental results are shown in Table 2.14. Due to low solubility of butyric acid in nonpolar solvents, we listed the C=O stretching frequencies of butyric acid in those solvents that its IR absorptions are strong. We found in the table that neither of Gaussian98 and Gaussin03 frequencies agrees with the experimental data. Besides, the IR band at around 1710 cm^{-1} is consistently present in experimental data. Our calculation in Table 2.4 shows that the C=O stretching frequency is 1703–1710 cm^{-1} for two hydrogen bonds. Therefore, the band at 1710 cm^{-1} is probably because of the hydrogen-bonding interactions between two butyric acid molecules. The chemical structure of such hydrogen-bonding interactions is shown in Figure 2.6.

TABLE 2.13 C=O stretching frequency of butyric acid in various solvents based on *ab initio* DFT calculations

Solvent Name	Dielectric Constant ε	$1/\varepsilon$	Solvent Formula	Scaled IR Freq. G98 $(cm^{-1})^*$	Scaled IR Freq. G03 $(cm^{-1})^*$
Vacuum	1	1	N/A	1775.5	1776.4
Heptane	1.9	0.52	C_7H_{16}	1769.2	1755.1
CycloHexane	2.0	0.49	C_6H_{12}	1768.9	1754.0
CarbonTetrachloride[†]	2.2	0.45	CCl_4	1767.7	1751.3
Benzene	2.2	0.45	C_6H_6	1767.7	1751.3
Toluene	2.4	0.42	C_7H_8	1767.1	1749.5
DiEthylEther[†]	4.3	0.23	$C_4H_{10}O$	1755.4	1736.6
Chloroform[†]	4.9	0.20	$CHCl_3$	1760.9	1734.0
ChloroBenzene[†]	5.6	0.18	C_6H_5Cl	1760.5	1732.4
Aniline[‡]	6.9	0.15	$C_6H_5NH_2$	1760.9	1730.1
TetraHydroFuran[†]	7.6	0.13	C_4H_8O	1758.5	1727.6
DiChloroMethane[†]	8.9	0.11	CH_2Cl_2	1757.7	1726.7
DiChloroEthane[†]	10.4	0.10	$C_2H_4Cl_2$	1754.0	1724.9
Acetone[†]	20.7	0.05	C_3H_6O	1752.0	1720.4
Ethanol[§]	24.6	0.04	C_2H_6O	1736.1	1719.5
Methanol[§]	32.6	0.03	CH_4O	1734.7	1718.0
Acetonitrile[†]	36.6	0.03	CH_3CN	1751.1	1717.7
NitroMethane[†]	38.2	0.03	CH_3NO_2	1751.0	1717.6
DMSO[†]	46.7	0.02	C_2H_6SO	1754.9	1717.0
Water[§]	78.4	0.01	H_2O	1733.9	1716.0

Values of dielectric constant are given for 1 atm and 20° C (Frisch et al., 1998; Griffiths, 1999).

*The scale factor is 0.9613.

[†]Solvents that can only be H-bond acceptors.

[‡]Solvents that can only be H-bond donors.

[§]Solvents that can both H-bond donor and acceptor.

Other solvents are nonpolar solvents that can not form any H-bond.

TABLE 2.14 FTIR absorption of butyric acid in various solvents

Solvent Name	Dielectric Constant ε	Solvent Formula	Experimental IR Freq. (cm^{-1})*	Calculated Freq. G98 $(cm^{-1})^{\dagger}$	Calculated Freq. G03 $(cm^{-1})^{\dagger}$
Benzene	2.2	C_6H_6	1709	1767.7	1751.3
DiEthylEther	4.3	$C_4H_{10}O$	1711	1755.4	1736.6
ChloroBenzene	5.6	C_6H_5Cl	1750, 1711	1760.5	1732.4
Aniline	6.9	$C_6H_5NH_2$	1711	1760.9	1730.1
TetraHydroFuran	7.6	C_4H_8O	1733, 1711	1758.5	1727.6
DiChloroEthane	10.4	$C_2H_4Cl_2$	1749, 1711	1754.0	1724.9
Acetone	20.7	C_3H_6O	1735, 1709	1752.0	1720.4
Ethanol	24.6	C_2H_6O	1736, 1709	1736.1	1719.5
Methanol	32.6	CH_4O	1736, 1709	1734.7	1718.0
DMSO	46.7	C_2H_6SO	1744, 1714	1754.9	1717.0
Water	78.4	H_2O	1713, 1703	1733.9	1716.0

*The experimental data is obtained by taking second derivatives of IR absorption spectra

of butyric acid in solvents.

†The calculated frequencies are taken from Table 2.13.
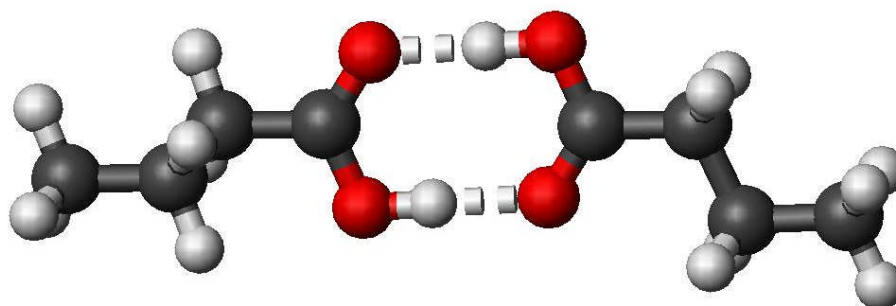

Figure 2.6 The chemical structure of a butyric acid molecule interacting with a butyric acid molecule via two hydrogen bonds.

CHAPTER III

A VIBRATIONAL SPECTRAL MARKER FOR PROBING THE

HYDROGEN-BONDING STATUS OF TYROSINE RESIDUES

3.1 Introduction

Tyrosine residues are often found at the active sites of proteins. The pKa value of tyrosine side−chain group is 10.5 (Nelson and Cox, 2000) so that these phenol groups are mostly protonated in proteins at steady states. The pKa values of these groups may change dramatically at the active sites of proteins in functional intermediate states, resulting in temporary or permanent deprotonation of these groups and therefore proton transfer between tyrosine residues and other key amino acids. Hydrogen-bonding interactions play key roles in regulating the pKa values of ionizable groups and driving proton transfer.

There are many examples of Tyrosine residues that are found at the active sites of proteins. They play important roles in protein functions. For example, in photoactive yellow protein (PYP), a blue-light bacterial photoreceptor, Tyr42 is found to stabilize a negatively charged pCA chromophore via hydrogen-bonding interaction in the receptor state (pG) (Borgstahl et al., 1995; Getzoff et al., 2003). For Bacteriorhodopsin, a well-studied light-driven proton pump for solar energy transduction, only Tyr185 out of 11 tyrosine is structural active (Braiman et al., 1988; Liu et al., 1995; Sonar et al., 1994). This residue is responsible for stabilizing Asp212 via a strong hydrogen bond in bR_{550}

(Rothschild et al., 1990) and M_{412} intermediate state (Ames et al., 1992). For rhodopsin, an animal visual photoreceptor protein, the pattern of a tyrosine/tyrosinate changes, similar but reversed to that of bacteriorhodopsin, was identified upon photoactivation of rhodopsin. The potential candidate was assigned to Tyr268 and it may be functional important paired with Pro267 so that they forms a hinge for α–helix movement (DeLange et al., 1998). For pharaonis phoborhodopsin (ppR; also pharaonis sensory rhodopsin II, psRII), Tyr199 is conserved completely in photorhodopsins among a variety of archaea. The dissociation constant of double mutants of D75N /Y199F or D75N/Y199V with its truncated transducer pHtrII was estimated ~10 fold larger than that of single mutant of D75N, suggesting the significant contribution of Tyr199 to binding (Sudo et al., 2002). In Cytochrome *bo,* a heme-copper terminal ubiquinol oxidase of *Escherichia coli* under highly aerated growth conditions, Tyr288 is located at the end of the K-channel and delivers one (or two) chemical protons to the heme-CuB binuclear center (Gennis, 1998). For a human B2 bradykinin receptor, a prominent member of the G protein–coupled receptor (GPCR), Tyr305 plays an important role in keeping the receptor in an inactive uncoupled state. Mutation of Tyr305 results in constitutive phosphorylation and ligand-independent receptor internalization (Kalatskaya et al., 2004). These examples show that buried tyrosine side–chain group may change their protonation states during the functional processes of proteins. In this chapter, we focus on developing a vibrational spectral marker for probing the hydrogen-bonding status of buried

neutral tyrosine side–chain group, including those that change their protonation states during the functional processes.

Hydrogen bonding is a fundamental element in protein structure and function. The hydrogen bond dissociation energy in proteins is in the order of 10 - 40 kJ/mol (2-10 kcal/mol) (Creighton, 1997; Hoppe et al., 1983) (Perrin and Nielson, 1997). (Much stronger hydrogen bonds with up to 160 kJ/mol have been found in some chemical systems, particularly those with small, charged groups (Hibbert and Emsley, 1990; Perrin and Nielson, 1997; Steiner, 2002).) In comparison, a typical value for protein folding energy is approximately 40 kJ/mol (Honig and Yang, 1995), only about one to four times of the hydrogen bond dissociation energy. Therefore, breaking one or more hydrogen bonds may impair the stability of a protein.

A phenol group of tyrosine may form up to two hydrogen bonds with neighboring hydrogen bond donors and acceptors, as shown in Figure 3.1: the phenolic oxygen and the hydroxyl hydrogen of a phenol group may each form one hydrogen bond. It is therefore energetically preferable for tyrosine to form as many hydrogen bonds as possible. It is challenging to monitor the hydrogen bond states of tyrosine residues for functional intermediate states of proteins and for proteins that their
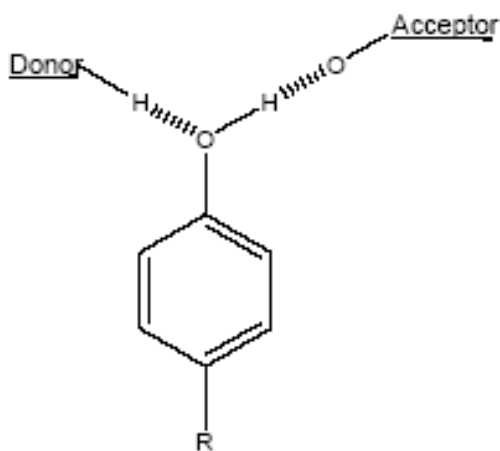
Figure 3.1 The schematic display of phenol group of tyrosine forming hydrogen bond(s) as hydrogen bond donor and acceptor with small model molecules.

X-ray crystal structures are not available. In this article, we reported that the C–O stretching frequency is a sensitive infrared structural marker for detecting and monitoring hydrogen-bonding status of tyrosine side–chain group. When the C–O stretching frequency is inadequate to distinguish zero and two hydrogen bonds, two dimensions of C–O stretching frequency combined with O–H bending or stretching frequency may be used to make such distinction. This vibrational spectral marker is based on *ab initio* computational studies and supported by experimental data. The applications of vibrational spectral markers are expected to enhance the power of time-resolved infrared difference spectroscopy for structural characterization of functionally important intermediate states of proteins, and consequently for understanding the functional mechanism of proteins.

3.2 Methods

A 4-propyl-phenol molecule was employed to model the side–chain of neutral Tyr residue in proteins, as shown in Figure 3.2 A. Both energy and vibrational frequency calculations were performed in vacuum using the method previously reported (Nie et al., 2005). The reason that solvent effects were not included in our calculations were discussed in the reference (Nie et al., 2005). Water molecules and molecules that model protein backbone and side–chain groups of polar and charged amino acids were utilized to serve as hydrogen-bond donor and/or acceptor for Gaussian03 (Frisch et al., 2003) calculations of hydrogen-bonding properties of tyrosine interacting with these molecules.
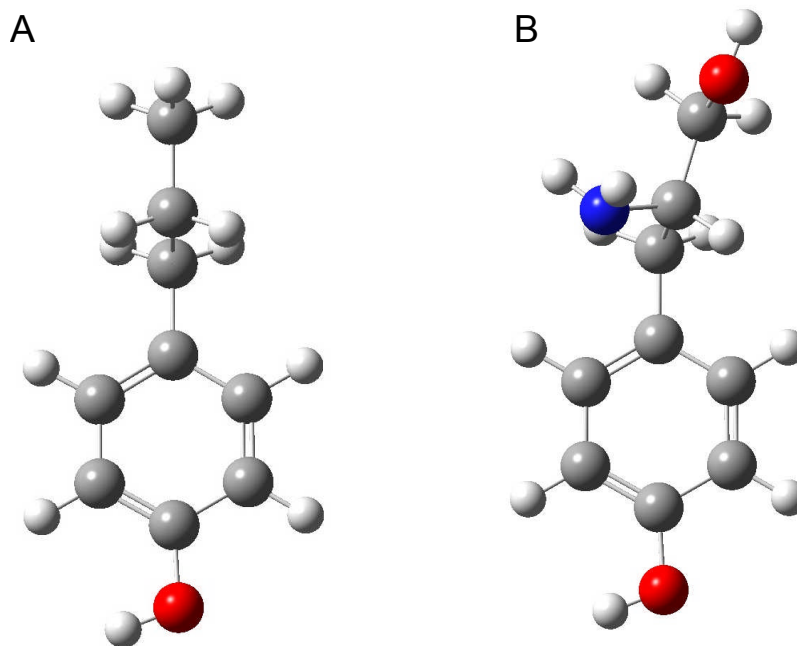
Figure 3.2 The chemical structures of 4-propyl-phenol (A) and L-tyrosinol (B). Color code: gray for carbon atoms, red for oxygen atoms, blue for nitrogen atoms, and white for hydrogen atoms. The bond order was also shown.

The notation, **O**H–X, represents that the phenolic oxygen is hydrogen bonded to a proton–donor molecule. Similarly, O**H**–X indicates that the hydroxyl hydrogen is hydrogen bonded to a proton–acceptor molecule. The notation, **OH**–X denotes that both the phenolic oxygen and hydrogen are hydrogen bonded.

Identify the number and the type of hydrogen-bond interactions from protein crystal structures

The pdb files for high resolution crystal structures of proteins were obtained from the protein data bank. To determine the number and the type of hydrogen bonding interactions for each buried phenol group, we first examined

and identified all the plausible hydrogen-bond donors and/or acceptors within the hydrogen bonding distance of a tyrosine side−chain group. Next, we checked the distance, bond angle and dihedral angles between the potential hydrogen bonding partners. A hydrogen bond is expected to meet all three criteria: less than 3.2 Å for the hydrogen-bond length, $120 \pm 20°$ for the CZ−O…X angle, and 150 to 180° or 0 to 30° for the CE−CZ−O…X dihedral angle. X stands for a heavy atom that is hydrogen bonded to a phenol group. Finally we examined the compatibility of hydrogen bonding interaction. For example, it is impossible for a hydrogen-bond acceptor to form a hydrogen bond with another hydrogen-bond acceptor.

Sample preparation

The chemical structure of L-Tyrosinol is shown in Figure 3.2 B. L−Tyrosinol hydrochloride (Sigma, CAS No. 87745-27-5) samples for FTIR experiments were prepared with 0.8 M concentration in H_2O and D_2O. The pH was adjusted to 0.5 by adding a small amount of hydrochloride or deuterium chloride to ensure complete protonation of phenol group. Each sample of 4 μl was sandwiched between two BaF2 plates 15 mm in diameter and 2 mm thick using a 12 μm spacer.

Rapid-scan FTIR spectroscopy

A Bruker IFS 66v FTIR spectrometer was utilized for rapid-scan measurements. Three samples of empty, tyrosinol in H_2O, and tyrosinol in D_2O were loaded to a custom-made sample exchanger driven by step motor driver (Si3540). Then the sample chamber was purged with nitrogen gas to get rid of

water vapor. The optics chambers of the spectrometer were evacuated to eliminate water vapor along optical path. The step motor driver is externally triggered by the spectrometer in rapid-scan mode to synchronize the data collection and changing sample. In this way, the quality of infrared spectra was largely improved due to little environmental change between the samples and the background (empty). The signal to noise ratio was large after 256 averages. The scanning rate was chosen to be 40 kHz and the spectral resolution to be 4 cm^{-1}.

Generation of an IR spectrum based on calculated vibrational frequencies

After each structure was optimized, there are 3 $\times$ N − 6 calculated vibrational frequencies corresponding to 3 $\times$ N − 6 vibrational motions, where N is the number of atoms in each structure. A Fortran90 code was written to generate a 2D data set (wavenumber vs. intensity of vibrational frequency) using Gaussian function:

$$Y(v) = I(v) * \exp(-\frac{(x - x_v)^2}{d^2})$$
(3.1)

where I(v) is the intensity of the vibrational frequency X$_v$ in Gaussian03 output and FWHM (full width at half maximum) is $2\sqrt{\ln 2}d$ where d was set to be 4 cm^{-1}.

3.3 Results and discussion

Band assignment of tyrosine side–chain group

Figure 3.3 depicts the comparison of calculated vibrational frequencies of isolated tyrosine (a), tyrosine forming one hydrogen bond with a water molecule as proton acceptor (b), tyrosine forming one hydrogen bond with a water molecule as proton donor (c), and tyrosine forming two hydrogen bonds with two
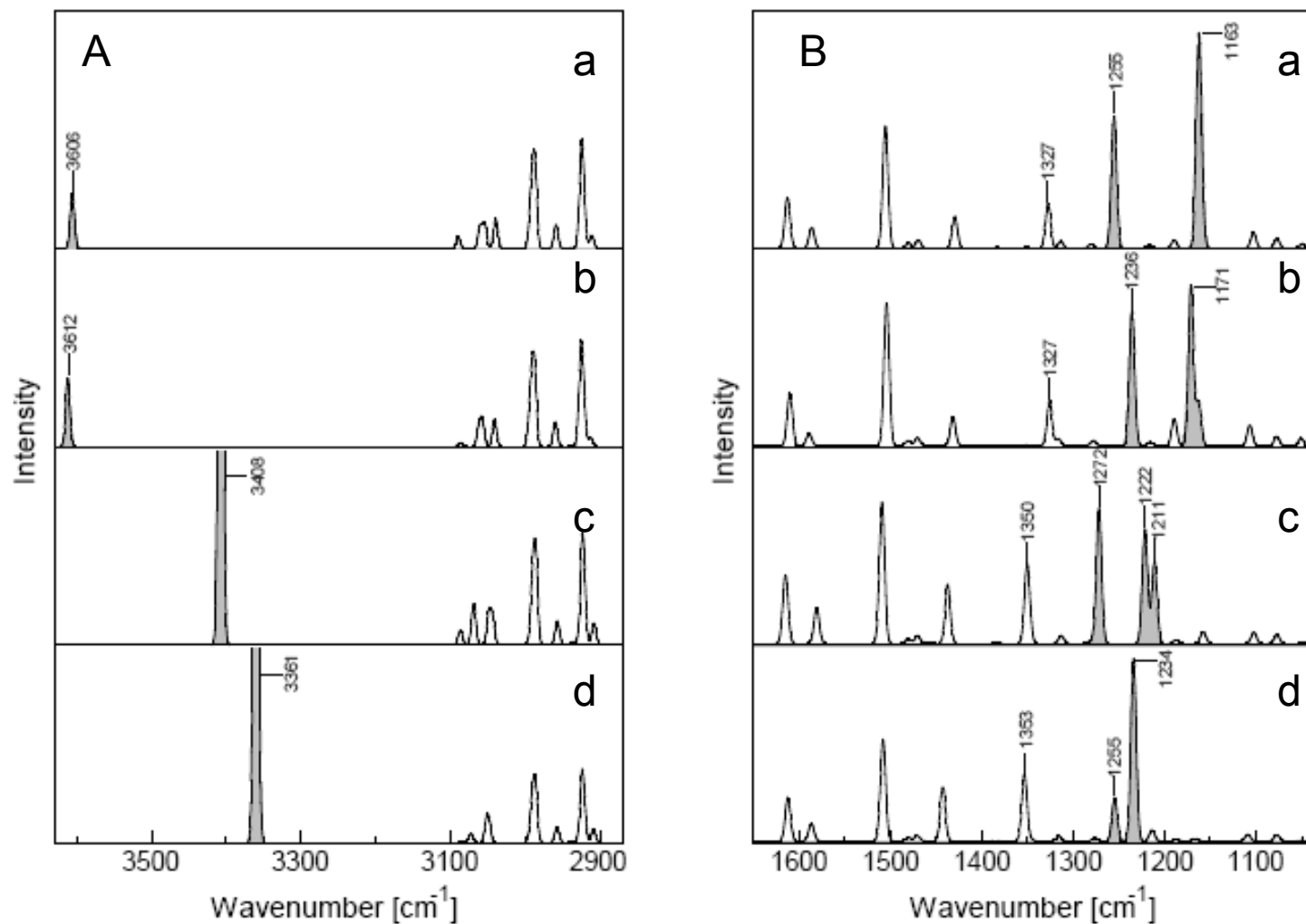
Figure 3.3 The calculated vibrational frequencies in the region of 2870 – 3630 cm^{-1} (A) and 1030 – 1650 cm^{-1} (B) of isolated 4-propyl-phenol (a); one hydrogen bond **O**H-X (b); one hydrogen bond O**H**-X (c); and two hydrogen bonds **OH**-X with water molecule(s) (d). The absorption of water has been removed.

water molecule as proton donor and acceptor (d). The band assignment of tyrosine side−chain is shown in Table 3.1. The vibrational frequencies of tyrosinol in H_2O and D_2O obtained using FTIR spectroscopy (see Figure 3.4) are also listed in the table for comparison.
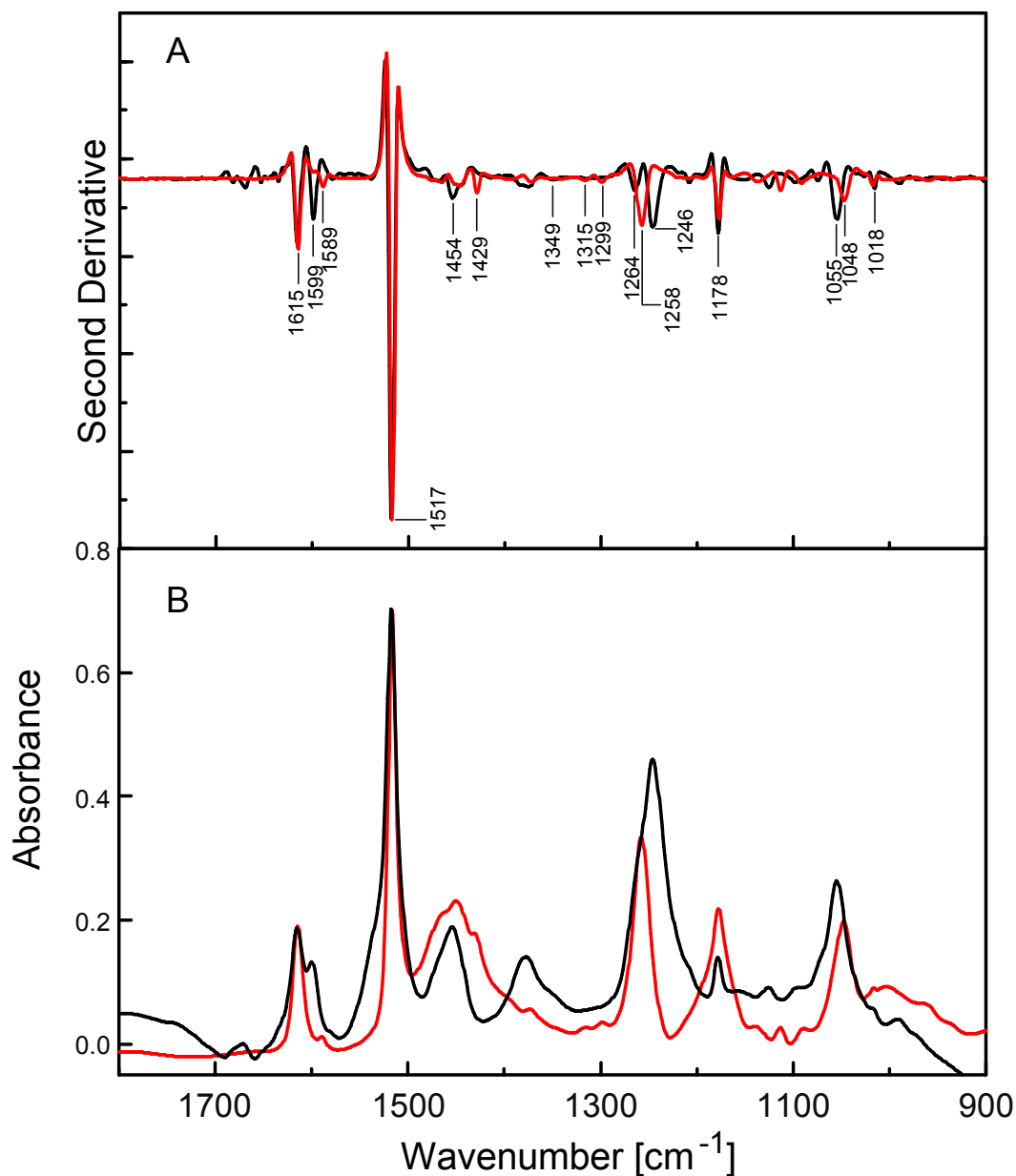

Figure 3.4 The second derivative (A) and absorbance spectrum (B) of L-tyrosinol in H_2O at pH =0.5 (black) and in D_2O at pH*=0.2 (red).

TABLE 3.1 Vibrational frequencies of tyrosine side–chain group

Vibrational mode	Tyr (cm^{-1})	Tyr–H_2O ($\underline{O}H–X$)	Tyr–H_2O ($O\underline{H}–X$)	Tyr–H_2O ($\underline{OH}–X$)	Exp. In H_2O	Tyr–D_2O ($\underline{OH}–X$)	Exp. In D_2O
Ring	1612	1610	1615	1612	**1615**	1612	**1615**
Ring	1586	1589	1581	1587	**1599**	1575	**1589**
Ring	1505	1504	1509	1508	**1517**	1506	**1515**
Ring & O–H Bending	1429	1432	1437	1443	**1454**	1420	**1429**
Ring & O–H Bending	1327	1326	1350	1353	**1349**	1302	**1299**
Ring	1314	1316	1313	1316	**N/A***	1316	**1315**
C–O stretching	1255	1235	1272	1255	**1264**	1247	**1258**
C_{ring}–C_{chain} stretching	1190	1189	1187	1187	**1178**	1188	**1178**
O–H Bending	1163	1171	1222 1211	1234	**1246**	978	**N/A***
Ring C–H Bending	1160	1162	1158	1167	**1055**	1167	**1048**
Ring C–H Bending	1103	1107	1107	1109	**1018**	1110	**1018**

*The experimental frequency was not visible.

Searching for bands that are sensitive to hydrogen-bonding interactions

In the fingerprint region, there are two bands of $\underline{O}H–X$ (Figure 3.3 b) that their frequencies largely shift due to hydrogen-bonding interactions: 1255 cm^{-1} down–shifted to 1235 cm^{-1} and 1162 cm^{-1} up–shifted to 1170 cm^{-1}. Using GaussView 2.1 for Windows (Gaussian, Inc.), we identified that 1255 cm^{-1} was attributed to C–O stretching and 1162 cm^{-1} to O–H bending modes. When the

hydroxyl hydrogen of tyrosine forms a hydrogen bond with the water molecule as proton donor (O**H**–X in Figure 3.3 c), the frequency of C–O stretching up-shifted from 1255 cm^{-1} to 1271 cm^{-1} and the frequency of O–H bending also up-shifted from 1162 cm^{-1} to 1221/1211 cm^{-1}. The 1221/1211 cm^{-1} doublet is due to coupled O–H bending and C–H bending on the main chain. Noticeably there is one more frequency shifted from 1327 cm^{-1} to 1350 cm^{-1} in Figure 3.3 c that is attributed to C–C stretching on the ring coupled O–H bending. This frequency is insensitive to the hydrogen-bonding interaction in Figure 3.3 b. The vibrational modes of 1327 cm^{-1}, 1255 cm^{-1}, and 1162 cm^{-1} are shown in Figure 3.4.

1327 cm^{-1} (Ring & O–H bending)   1255 cm^{-1} (C–OH stretching)   1162 cm^{-1} (O–H bending)
1299 cm^{-1} (Ring & O–D bending)   1249 cm^{-1} (C–OD stretching)   913 cm^{-1} (O–D bending)
1326 cm^{-1} (Ring & ^{18}O–H bending)   1245 cm^{-1} (C–^{18}OH stretching) 1156 cm^{-1} (^{18}O–H bending)
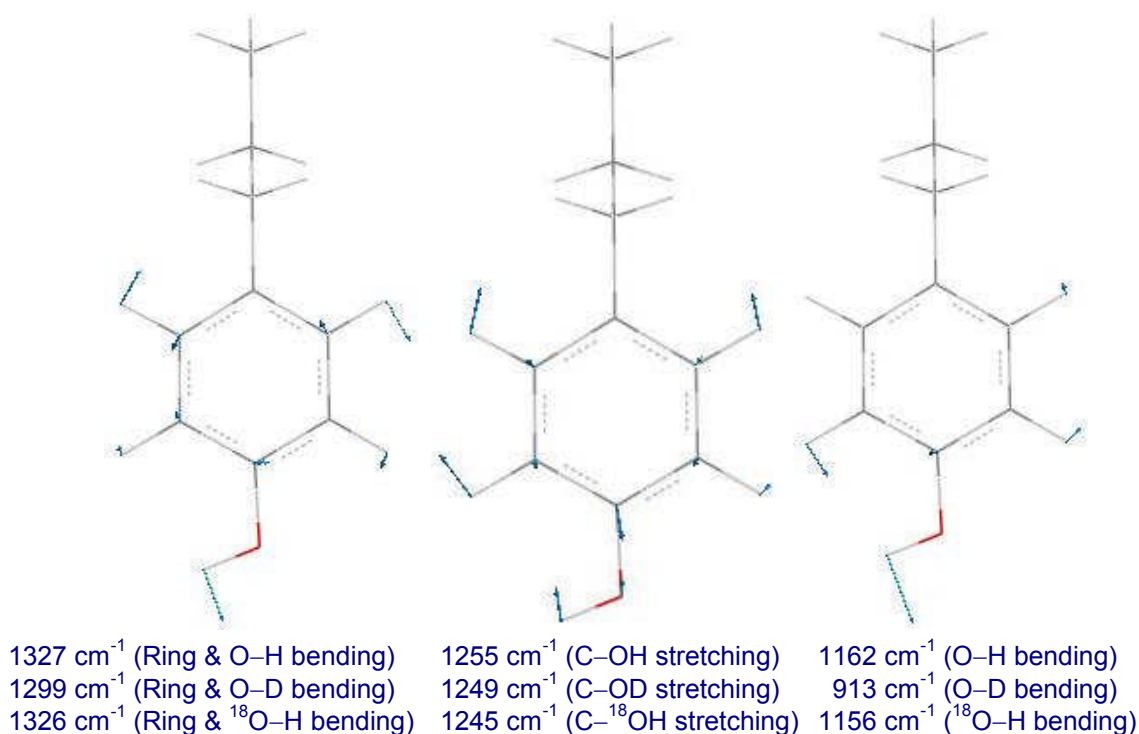
Figure 3.5 Normal vibrations of 4-propyl-phenol with zero hydrogen bond. Color code for atoms: gray for carbon atoms, red for oxygen atoms, and white for hydrogen atoms. The displacement vectors are shown in green arrow. These vibrational modes are obtained directly from GaussView.

In principle, the frequency 1327 cm^{-1} of C–C stretching on the ring coupled O–H bending may be considered as a vibrational spectral marker of the type of hydrogen-bonding interaction of tyrosine: it shifts less than 2 cm^{-1} for **O**H–X and it up-shifted to 1350 – 1357 cm^{-1} for O**H**–X (see supplemental material Table 3.8). However, this frequency is not a good marker for probing hydrogen-bonding status of tyrosine side–chain for three reasons: (1) the frequency shift of 2 cm^{-1} is so small between isolated tyrosine and **O**H–X that it is difficult to distinguish zero hydrogen bond and one hydrogen-bonding interaction of **O**H–X; (2) the frequency shift upon deuteration is insensitive to hydrogen-bonding interaction (see supplemental material Table 3.8); (3) The intensity of this frequency in experimental data is small in both absorption and second derivative spectra (data not shown). Therefore, we focused on the C–O stretching and O–H bending frequencies that are sensitive to the hydrogen-bonding interaction in the fingerprint region.

A vibrational spectral marker for probing the hydrogen-bonding status of a tyrosine side–chain group

To explore, evaluate, and determine the qualification of the C–O stretching and O–H bending frequency as vibrational spectral markers for detecting the hydrogen-bonding status of tyrosine, we performed a range of calculations of tyrosine side–chain group hydrogen-bonding interacting with water, protein backbone, and polar and charged side–chain groups of amino acids. These side–chain groups include 10 polar side–chain groups (Ser, Thr, Cys, Met,

Asn/Gln, Tyr, Asp/Glu, His, Lys, and Arg), 2 negatively charged side–chain groups (Tyr and Asp/Glu) and 3 positively charged side–chain groups (Arg, His, and Lys). Tables 3.2 and 3.3 show the computational results as well as the structures of molecules that are employed to model the different side–chain groups of polar and charged amino acids. For the convenience of discussion, we will refer these model compounds as their corresponding amino acids.

Table 3.2 shows the calculated hydrogen-bonding properties of tyrosine side–chain group interacting with water, backbone, and neutral polar amino acids. For an isolated phenol group without any hydrogen-bonding interactions, the calculated C–O stretching frequency is 1255 cm^{-1}. When the phenolic oxygen of tyrosine forms one normal hydrogen bond with a polar group (**O**H–X), all six computational results (from Arg to His) show that the C–O stretching frequency is red–shifted from 1255 cm^{-1} to 1235–1239 cm^{-1}. The strength of these hydrogen bonds is in the range of 8.7–16.7 kJ/mol. In the case of hydrogen-bonding interaction, Asp/Glu–1, this frequency is further red–shifted to 1230 cm^{-1} and the hydrogen-bonding strength is stronger (23.2 kJ/mol). This is due to an additional weak hydrogen bond between carbonyl oxygen of Asp/Glu and C–H on the ring of tyrosine (Desiraju and Steiner, 1999). When the hydroxyl hydrogen of tyrosine forms one normal hydrogen bond with a polar group (O**H**–X), nine computational results (from His to Asp/Glu) show that C–O stretching frequency is blue–shifted from 1255 cm^{-1} to 1264–1277 cm^{-1}. The strength of these hydrogen bonds is in the range of 23.0–38.4 kJ/mol, much stronger than that of **O**H–X (8.7–16.7

95

kJ/mol), indicating O**H**–X type of hydrogen-bonding interaction is more stable than **O**H–X type of hydrogen-bonding interaction. When a phenol group forms two hydrogen bonds, the C–O stretching frequency is $1252 - 1255$ cm^{-1}, fairly close to the C–O frequency of isolated phenol group. It is expected to be very close to the C–O stretching frequency of a phenol group with zero hydrogen bond because of combined hydrogen bonding effects on the C–O stretching frequency: average red shift of 18 cm^{-1} for **O**H–X and blue shift of 16 cm^{-1} for O**H**–X. The average hydrogen-bond strength is 22.9 kJ/mol per hydrogen bond for two hydrogen-bonding interactions.

The results discussed above are based on well-formed hydrogen bond(s) with neutral polar groups. Noticeably when two tyrosine side–chain groups form one hydrogen bond, the hydrogen-bond strength is stronger than that of **O**H–X and weaker than O**H**–X due to the coupling of two phenol groups. The C–O stretching frequency therefore is further red–shifted to 1231 cm^{-1} and less blue–shifted to 1261 cm^{-1}.

There are four special cases in Table 3.2 and 3.3 that need to be addressed. The first special case is the hydrogen-bonding interactions with Cys and Met, shown in Table 3.2. The hydrogen-bond strength for Cys–A is 5.1 kJ/mol, much weaker than other **O**H–X type of hydrogen-bonding interactions (average of 12.7 kJ/mol). The hydrogen-bond strength for Cys–B and Met–B is $14-17$ kJ/mol, also much weaker than other O**H**–X type of hydrogen-bonding interactions (average of 30.7 kJ/mol). The hydrogen-bond dissociation energy of these hydrogen-bonding interactions is approximately half of the average of

calculated hydrogen-bond dissociation energy for $\underline{\text{O}}$H–X and O$\underline{\text{H}}$–X. This is due to the fact that sulfur atom is large so that the hydrogen bond is long and weak, and that the hydrogen-bond strength of $\underline{\text{O}}$H–X is weaker interaction than that of O$\underline{\text{H}}$–X. Therefore, this hydrogen-bonding interaction is classified as half hydrogen bond or weak hydrogen bond. The C–O stretching frequencies for Cys–A, Cys–B, and Met–B are 1243 cm^{-1}, 1261 cm^{-1}, and 1259 cm^{-1}, respectively. These frequency shifts are less but in consistency with the trend of C–O stretching frequency shift due to hydrogen-bonding interactions.

The second special case is the hydrogen-bonding interactions with neutral Arg and Lys, shown in Table 3.2. The pKa value of Arg and Lys side–chain is 12.5 and 10.5 (Nelson and Cox, 2000). Therefore they are normally charged in proteins. The case that Arg or Lys is neutral is fairly rare. Therefore, we do not include hydrogen-bonding interactions with neutral Arg and Lys into our classification of vibrational spectral marker for probing hydrogen-bonding status of tyrosine side–chain group. We will address the special case that the hydrogen-bond partner is positively charged Arg+, Lys+, and His+ below.

The third special case is dealing with two deformed hydrogen bonds. Asn/Gln and protonated Asp/Glu may take part in hydrogen-bonding interactions both as proton donor and proton acceptor. When a phenol group forms two hydrogen bonds with these groups as illustrated in Table 3.2, these two hydrogen bonds are deformed due to geometrical constraints. Their hydrogen bond angles are 158°/137° for Asn/Gln and 145°/154° for Asp/Glu, largely deviated from the optimal hydrogen-bond angle of 170° – 180°. Such deformations lead to reduced

hydrogen-bond dissociation energy, ~35.3 kJ/mol. In comparison with 22.9 kJ/mol per hydrogen bond for two well-formed hydrogen-bonding interactions (2 H_2O in Table 3.2), the hydrogen-bond dissociation energy for two deformed hydrogen bonds, 35.3 kJ/mol, is approximately 1.5 times of 22.9 kJ/mol. Therefore, we classify such two deformed hydrogen bonds (Asn/Gln and Asp/Glu) as forming 1.5 hydrogen bonds.

The fourth special case is that a phenol group forms hydrogen bond(s) with a charged side–chain group. Most charged groups in proteins are solvent exposed. However, buried charged groups have been found in the active sites of proteins. Therefore, we examined the hydrogen-bonding interactions of tyrosine side–chain with positively charged side–chains (Arg+, Lys+, and His+) and with negatively charged side–chains (COO− of Asp/Glu or a negatively charged phenolic oxygen of Tyr).

In the case that a phenol group forms a hydrogen bond with a positively charged side–chain of His or Lys (His+ or Lys+ in Table 3.3), the strength of such a hydrogen bond is very strong, 60 − 71 kJ/mol. This is approximately three times as strong as the average hydrogen-bond strength of a phenol group interacting with one polar neutral hydrogen-bonding partner, ~22.9 kJ/mol per hydrogen bond. When a phenol group is hydrogen bonded to a positively charged side–chain of Arg, two deformed hydrogen bonds are formed with the phenolic oxygen. The hydrogen-bond dissociation energy for these two deformed bonds is 59 kJ/mol (Table 3.3), similar to that of single hydrogen-bonding interactions with His+ or Lys+. Since positively charged side–chains do not have hydrogen-bond

acceptors, no hydrogen bond is formed with the hydroxyl hydrogen of a phenol group. The C–O stretching frequency is largely red–shifted from 1255 cm^{-1} to 1194–1204 cm^{-1}. Therefore, a low C–O stretching frequency (1194 to 1204 cm^{-1}) indicates strong hydrogen-bonding interactions with positively charged side–chain groups.
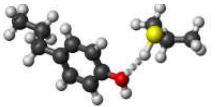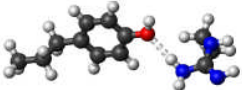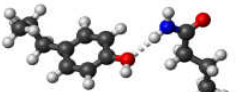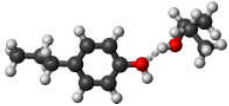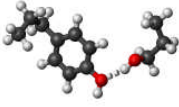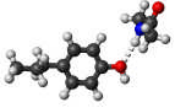
In the case that the hydrogen-bond partner is negatively charged (Asp/Glu– or Tyr–), our calculations lead to proton movements. Starting from an initial structure for a hydrogen-bonding interaction between a phenol group and a COO$^-$ group or a Tyr–O$^-$ group, we found that the proton on tyrosine is shared between the two groups after structural optimization. These results suggest that the hydrogen-bonding interaction between a phenol group and a negatively charged side–chain group is energetically unstable; additional interactions must be present in order to stabilize any hydrogen-bonding interaction between a phenol group and a negatively charged side–chain group in proteins.
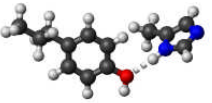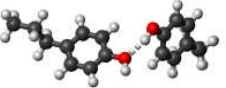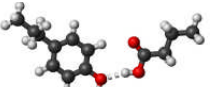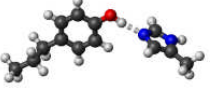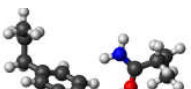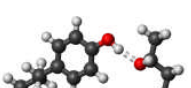
Overall, the hydrogen-bond strength of tyrosine side-chain group is stronger for O**H**–X than that of for **O**H–X with neutral polar groups. The hydrogen-bond strength of **O**H–X with positively charged side–chain groups is stronger than that of **OH**–X. Similar patt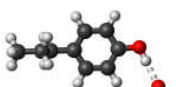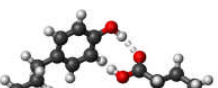ern was observed for the vibrational spectral marker of protonated carboxylic group (Nie et al., 2005). In comparison with the same kinds of hydrogen bond(s), all hydrogen-bond strengths of tyrosine side-chain group are weaker than that of protonated carboxylic group. For instance, the hydrogen-bond strength of a phenol group interacting with serine,
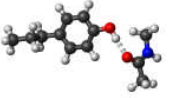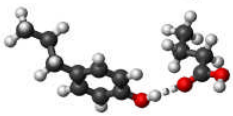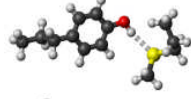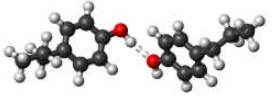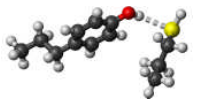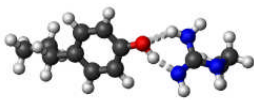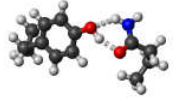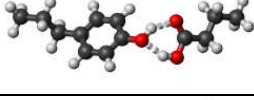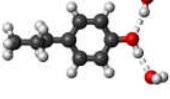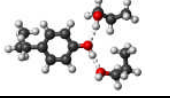
OH–X (Ser–A), is 11.2 kJ/mol, whereas the hydrogen-bond strength of a COOH group interacting with serine, HO–C=**O** (Ser–1), is 17.6 kJ/mol (Nie et al., 2005).

These computational results support that the C–O stretching frequency is a sensitive vibrational spectral marker for probing the hydrogen-bonding status of tyrosine side–chain group: it is 1255 cm^{-1} for zero hydrogen bond, 1230–1236 cm^{-1} for one hydrogen bond **O**H–X and 1264–1277 cm^{-1} for one hydrogen bond O**H**–X with neutral polar group, and 1194–1204 cm^{-1} for one hydrogen bond **O**H–X with positively charged side–chain group. The case that a phenol group forms two deformed hydrogen bonds in proteins is fairly rare because of structural constraints. In fact, most hydrogen-bonding interactions found in crystal structures of proteins are single hydrogen bond (Nie et al., 2005). In the case of single hydrogen-bonding interactions with Cys (1259 cm^{-1}) or Met (1261 cm^{-1}) or coupled Tyr–Tyr (1261 cm^{-1}), and two hydrogen-bonding interactions (1252–1255 cm^{-1}), it is difficult to use the C–O stretching frequency to distinguish them from zero hydrogen bond (1255 cm^{-1}) and one hydrogen bond O**H**–X (1264–1277 cm^{-1}). Therefore, additional vibrational information will be needed.

TABLE 3.2 Calculated hydrogen-bonding properties of a phenol group interacting with neutral groups

Structure	Amino Acid*	No. of H-bond[†]	H-bond Length[†] (Å)	H-bond Angle[†] (deg)	H-bond Energy[‡] (kJ/mol)	ν_{C-O} (cm^{-1})	$\nu_{O-H \text{ Bending}}$ (cm^{-1})	$\nu_{O-H \text{ Stretching}}$ (cm^{-1})
	TYR	0	N/A	N/A	N/A	1255.5	1162.9	3606.1
	CYS – A	~0.5	3.65	150.8	5.1	1243.2	1162.6[§]	3607.7
	ARG – A	1	3.11	174.4	11.9	1239.1	1164.4[§]	3609.0
	ASN/GLN – A	1	3.06	172.1	10.7	1236.4	1164.1[§]	3609.6
	THR – A	1	2.91	157.7	8.7	1236.1	1163.0	3601.3
	SER – A	1	2.91	162.8	11.2	1236.0	1166.2[§]	3612.9
	BB – A	1	3.09	171.4	13.1	1235.9	1164.6[§]	3608.8
	H$_2$O – A	1	2.93	158.8	14.8	1235.5	1166.8[§]	3612.4

	HIS – A	1	3.01	173.8	16.7	1235.0	1165.9§	3609.7
	TYR – A	1	2.87	167.2	18.5	1230.8	1166.0§	3614.1
	ASP/GLU – A	1	2.83	167.2	23.2	1229.8	1155.0	3614.7
	HIS – B	1	2.85	173.8	38.4	1276.7§	1230.1	3244.5
	LYS – B	1	2.81	168.2	34.9	1276.1§	1244.6	3105.7
	ASN/GLN – B	1	2.77	165.4	26.9	1273.1	1227.3	3312.0
	H₂O – B	1	2.82	179.0	23.0	1271.8	1221.5	3408.4
	ASP/GLU – B	1	2.81	158.4	27.7	1268.9§	1221.3 1210.6	3427.8 3383.8
	SER – B	1	2.79	172.7	24.1	1267.6	1225.7	3370.7
	THR – B	1	2.78	167.7	26.8	1266.2	1228.9	3364.4

BB – B	1	2.77	171.0	33.6	1265.5	1227.6	3367.8
ASP/GLU – B	1	2.84	173.6	24.0	1263.5[§]	1222.1[§]	3447.5
MET – B	~0.5	3.42	167.8	14.4	1261.4	1197.5	3440.7
TYR – B	1	2.87	167.2	18.5	1260.9	1201.0	3497.6
CYS – B	~0.5	3.35	149.1	17.1	1258.7	1194.6	3452.0
ARG	1.5	2.72 / 3.03	157.4 / 144.8	48.4	1265.8	1249.6	3028.2[§]
ASN/GLN	1.5	2.72 / 2.94	157.5 / 136.6	33.0	1257.5[§]	1235.9	3221.9
ASP/GLU	1.5	2.72 / 2.77	145.0 / 154.0	37.5	1245.9	1221.3	3338.9 / 3252.9
2 H$_2$O	2	2.77 / 2.89	173.1 / 164.7	45.8	1254.7	1233.8	3361.0
2 SER	2	2.76 / 2.87	173.1 / 166.0	39.3	1251.6	1237.1	3332.2
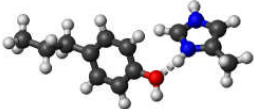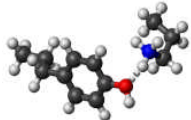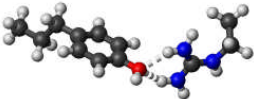
[*] The structures presented here are optimized using B3LYP/6-31G(d) method. Color codes for atoms: black for carbon atoms, red for oxygen atoms, blue for nitrogen atoms, yellow for sulfur atoms, and white for hydrogen atoms. The dashed lines represent hydrogen bonds. His−A indicates that the phenolic oxygen of a phenol group forms a hydrogen bond while His−B designates that the hydroxyl hydrogen of a phenol group forms a hydrogen bond. The same notation is used for other amino acids. BB is the abbreviation of backbone.

[†] The hydrogen-bond length is measured between the heavy atoms of a pair of hydrogen-bond donor and acceptor. When two hydrogen bonds are formed, the upper value is for hydroxyl hydrogen and the bottom value is for phenolic oxygen.

[‡] The vibrational frequencies were calculated using B3LYP/6-31G(d) method on optimized structures. The energy was computed using B3LYP/6-311+G(2d,p) method.

[§] The vibrational frequency was the average of two coupled vibrational modes that their frequencies are next to each other.

TABLE 3.3   The hydrogen bonding interactions of neutral Tyrosine with charged amino acid side chains

Structure	Amino Acid*	No. of H-bond[†]	H-bond Length[†] (Å)	H-bond Angle[†] (deg)	H-bond Energy[‡] (kJ/mol)	$\nu_{C\text{-}O}$ (cm^{-1})	$\nu_{O\text{-}H\ Bending}$ (cm^{-1})	$\nu_{O\text{-}H\ Stretching}$ (cm^{-1})
	HIS+	1	2.78	167.3	59.4	1204.4	1163.0	3595.5
	HIS+	1	2.75	175.1	60.6	1200.7	1176.5	3596.1
	LYS+	1	2.74	164.5	71.0	1198.0	1159.7[§]	3593.8
	ARG+	2	2.90 2.92	149.3 147.6	58.7	1194.1	1177.8	3585.6

The color codes, the methods for frequency and energy calculations, and the definition of hydrogen-bond lengths are the same as those in Table 3.2.

Two-dimensional infrared spectroscopy

As shown in Table 3.2 and 3.3, the O–H stretching and bending frequencies are also sensitive to hydrogen-bonding interactions. However, the correlation between the O–H stretching and bending frequencies and specific hydrogen-bonding interactions is different from that of the C–O stretching frequency. For example, O–H stretching and bending frequencies can not be used to distinguish zero hydrogen bond from one hydrogen-bonding interaction of **O**H–X with neutral polar group or with positively charged side–chain group. When both O–H frequency (either stretching or bending) and C–O stretching frequency are used in a two-dimensional (2D) plot (see Figure 3.5), six categories are well separated by either C–O stretching or O–H frequency or both: zero hydrogen bond (■), one hydrogen bond **O**H–X with neutral polar groups (□), one hydrogen bond **O**H–X with positively charged groups (●), one hydrogen bond O**H**–X with neutral polar groups (○), one hydrogen bond O**H**–X with Cys or Met or Tyr (▲), and two hydrogen bonds **OH**–X with neutral polar groups (✱). Due to the fact that strong O–H absorption from solvents overlaps with the O–H stretching modes from the tyrosine side–chain groups into consideration, the O–H stretching frequency is understudied. In principle, this frequency range can provide valuable structural information for amino acid side–chains. The use of protein samples in hydrated film or crystal form may greatly reduce solvent absorption in this region. Under these conditions, 2D plot of O–H stretching and C–O stretching frequencies can be helpful. In other circumstances, 2D plot of O–H bending and C–O stretching frequencies is more suitable to provide
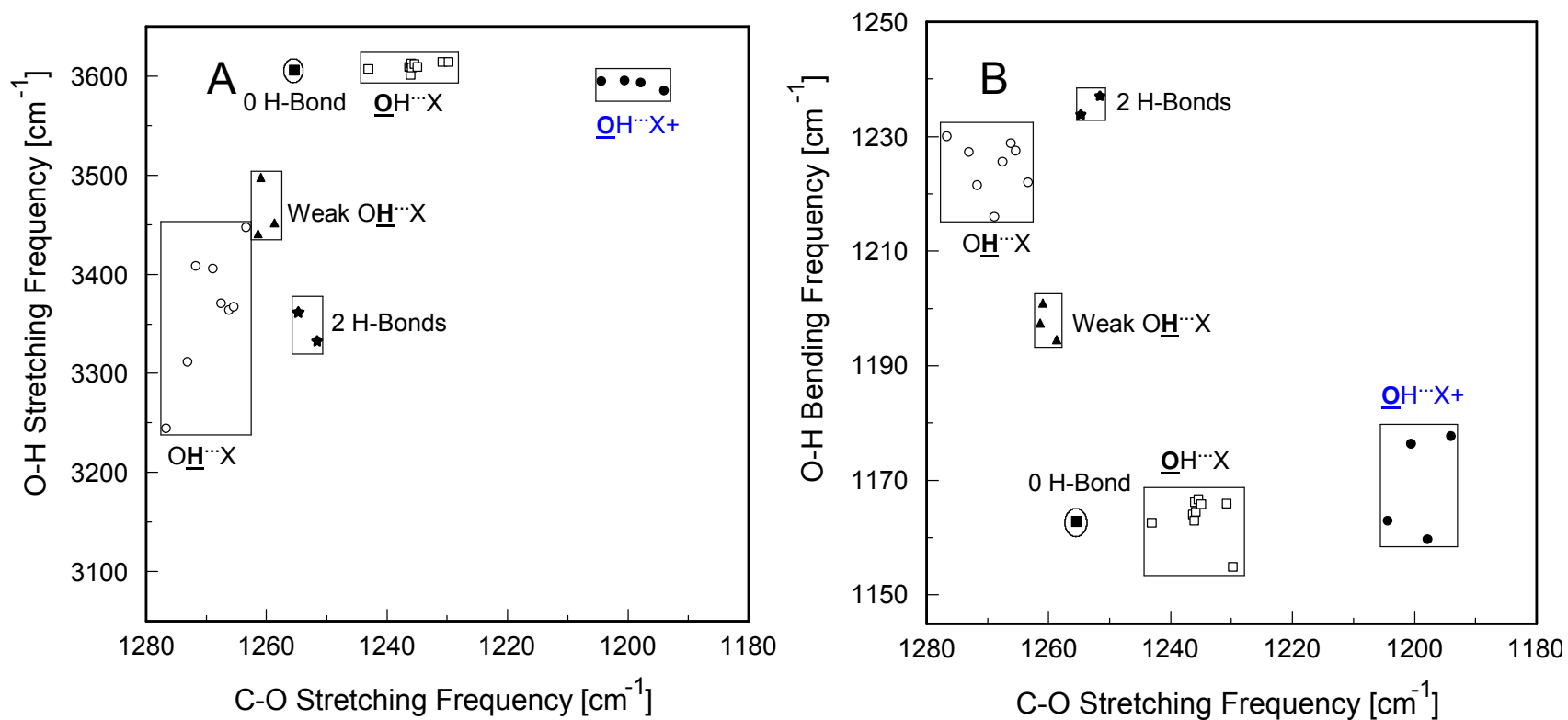
Figure 3.6 Two–dimensional correlations of the O–H and C–O stretching modes (A) and the O–H bending and C–O stretching modes (B) for probing the specific type of hydrogen–bonding interactions in proteins: zero hydrogen bond (■), one hydrogen bond **O**H–X with neutral polar groups (□), one hydrogen bond **O**H–X with positively charged groups (●), one hydrogen bond O**H**–X with neutral polar groups (○), one hydrogen bond O**H**–X with Cys or Met or Tyr (▲), and two hydrogen bonds **OH**–X with neutral polar groups (✱).

information on hydrogen-bonding status of tyrosine side–chain groups in proteins.

Isotopic shifts of vibrational spectral marker of tyrosine

Isotopic labeling directly affects the position of IR bands. Therefore, it is often used to test the isotopic–labeling induced frequency shifts so that labeled specific residues can be identified from IR spectra of proteins containing rich information of many amino acids. The common isotope substitutions are $H \rightarrow D$, $^{12}C \rightarrow {}^{13}C$, $^{14}N \rightarrow {}^{15}N$, and $^{16}O \rightarrow {}^{18}O$. For tyrosine side–chain group, the most commonly used isotopic labeling is deuteration of hydroxyl hydrogen of the phenol group and deuteration of four hydrogen atoms on the phenol ring. Therefore, we have calculated vibrational frequencies of tyrosine side–chain group (4-propyl-phenol) with these two isotope substitutions.
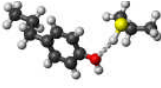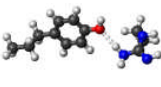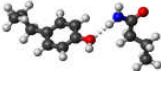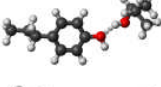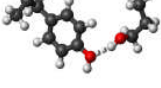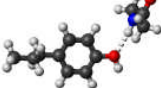
Table 3.4 shows the vibrational frequency shifts (shaded columns) of C–O stretching, O–H bending, and O–H stretching modes due to $OH \rightarrow OD$ and ring–$H_4 \rightarrow$ ring–D_4. In the case of $OH \rightarrow OD$ deuteration, the frequency down–shift of C–O stretching vibration of the phenol group is in a range of $3 - 9$ cm^{-1} except for the frequency shift of the phenol group forming two deformed hydrogen bonds with Asp/Glu, 15 cm^{-1}. This frequency shift is insensitive to hydrogen-bonding interactions. The experimental C–O stretching frequency of tyrosinol in H_2O, at 1264 cm^{-1}, is down–shifted by 6 cm^{-1} to 1258 cm^{-1} in D_2O. In practice, the phenol group of tyrosinol most likely forms two hydrogen bonds with solvents. Our computational results of tyrosine side–chain group with two hydrogen bonds show that the C–O stretching frequency at 1252–1255 cm^{-1} is down–shifted by 8
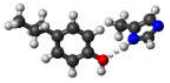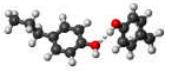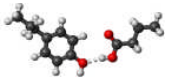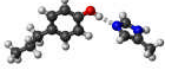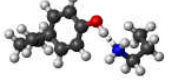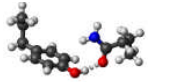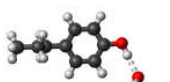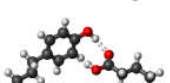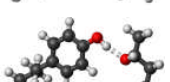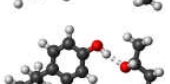
cm^{-1} to 1244–1247 cm^{-1} upon deuteration. It is in good agreement with experimental data. This frequency shift is considerably small in comparison with the frequency shifts of O–H bending and O–H stretching modes. Most computational results demonstrated that the frequency of O–H bending is down–shifted by 250 cm^{-1} to below 1000 cm^{-1} in which region the BaF$_2$ infrared windows are necessary for FTIR measurements in order to observe such big shift. However, it is not visible in our experiment of tyrosinol in D$_2$O for unclear reasons. For O–H stretching mode, its frequency is largely down–shifted by around 980 cm^{-1} for zero and **O**H–X with neutral polar groups and positively charged groups, 840 – 950 cm^{-1} for O**H**–X with neutral polar groups and two hydrogen bonds **OH**–X. This frequency shift can not be distinguished from experimental data of tyrosinol because of overlapping of water absorptions in this region.
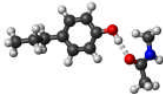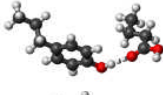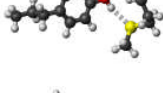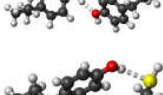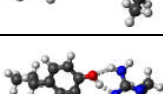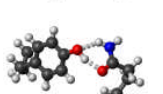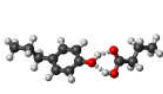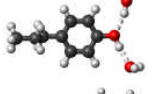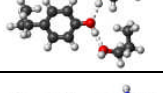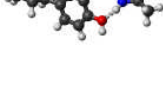
In the case of ring–H$_4$ → ring–D$_4$ deuteration, the C–O stretching frequency is largely downshifted compared to the OH → OD deuteration induced frequency shift. It is down–shifted by 38 cm^{-1} for zero hydrogen bond, 27 – 34 cm^{-1} for one hydrogen bond **O**H–X with neutral polar groups, 40 – 47 cm^{-1} for one hydrogen bond O**H**–X with neutral polar groups, 39 cm^{-1} for two hydrogen bonds **OH**–X, and 19 – 23 cm^{-1} for **O**H–X with positively charged groups. Therefore, this frequency shift is sensitive to hydrogen-bonding interactions. This is an additional indicator of hydrogen-bonding status of tyrosine side–chain group when experimental data of both conditions is available. Upon ring–H$_4$ → ring–D$_4$ deuteration, all our computational results show that the O–H bending motion is
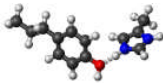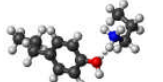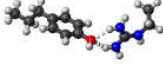
coupled with C–H bending motion on the main chain. The O–H bending motion is uncoupled in unlabeled tyrosine calculations. Therefore, the isotopic labeling induced frequency shift may not be compared. The frequency of O–H stretching is not shifted due to ring–H_4 $\rightarrow$ ring–D_4 deuteration. This means that the deuteration on the ring has no effect on O–H stretching frequency because the isotopes are distant from OH group. In comparison to the experimental C–O stretching frequency ~1226 cm^{-1} of ring–D_4 labeled tyrosine (Dollinger et al., 1986; Hellwig et al., 2002; Hienerwadel et al., 1997), the calculated C–O stretching frequency of ring–D_4 labeled tyrosine side–chain with two hydrogen bonds is 1216 cm^{-1}. The difference of 10 cm^{-1} is less than 1% of the C–O stretching frequency. Therefore, our computational results are in good agreement with experimental data. Since the amide I absorption of protein strongly overlaps absorption of water at around 1650 cm^{-1} and it has much smaller absorption than that of water, most IR spectra were measured for protein samples in D_2O to avoid this overlap. The 2D plot of O–D stretching and C–O stretching modes and the 2D plot of O–D bending and C–O stretching modes is presented in Figure 3.6 for reference.

TABLE 3.4 Isotopic labeling effects on vibrational spectral marker of Tyr

Structure	H-Bond Partner	$\nu_{C\text{-}O}$ (cm^{-1})					ν_{OH} bending (cm^{-1})					ν_{OH} stretching (cm^{-1})				
		OH	OD	$\Delta\nu$	ring–^{2}H$_4$	$\Delta\nu$	OH	OD	$\Delta\nu$	ring–^{2}H$_4$[#]	$\Delta\nu$[‖]	OH	OD	$\Delta\nu$	ring–^{2}H$_4$	$\Delta\nu$
	N/A	1255.5	1248.9	6.6	1217.2	38.3	1162.9	913.2	249.7	1230.4 1178.5	N/A	3606.1	2625.1	981.0	3606.1	0.0
	CYS – A	1243.2	1237.9	5.3	1209.2	34.0	1162.6[§]	910.8[§]	251.8	1227.9 1177.7	N/A	3607.7	2626.4	981.3	3607.7	0.0
	ARG – A	1239.1	1234.4	4.7	1207.8	31.3	1164.4[§]	910.0[§]	254.4	1226.5 1179.0	N/A	3609.0	2627.4	981.6	3609.0	0.0
	ASN/ GLN – A	1236.4	1231.9	4.5	1202.7	33.7	1164.1[§]	914.4	249.7	1222.3 1173.0	N/A	3609.6	2627.8	981.8	3609.6	0.0
	THR – A	1236.1	1230.4	5.7	1206.2	29.9	1163.0	911.0	252.0	1225.8 1177.7	N/A	3601.3	2621.7	979.6	3601.3	0.0
	SER – A	1236.0	1231.7[§]	4.3	1203.1	32.9	1166.2[§]	917.2	249.0	1222.2 1174.0	N/A	3612.9	2630.3	982.6	3612.9	0.0
	BB – A	1235.9	1231.2	4.7	1205.9	30.1	1164.6[§]	915.3	249.3	1227.1 1180.5	N/A	3608.8	2627.2	981.6	3608.8	0.0

H$_2$O – A	1235.5	1230.8	4.7	1205.9	29.6	1166.8§	916.5	250.3	1227.2 1180.4	N/A	3612.4	2629.9	982.5	3612.4	0.0
HIS – A	1235.0	1230.9	4.1	1204.8	30.2	1165.9§	917.8	248.1	1227.1 1181.9	N/A	3609.7	2627.8	981.9	3609.7	0.0
TYR – A	1230.8	1226.6	4.2	1201.3§	29.5	1166.0§	918.1	247.9	1227.7 1180.5	N/A	3614.1	2631.2	982.9	3614.1	0.0
ASP/GLU – A	1229.8	1227.3	2.5	1202.8	27.0	1155.0	919.1	235.9	1227.1¶ 1190.6 1163.1	N/A	3614.7	2631.5	983.2	3614.7	0.0
HIS – B	1276.7§	1270.2	6.5	1230.2	46.5	1230.1	979.2	250.9	1260.3 1208.3	N/A	3244.5	2366.0	878.5	3244.5	0.0
LYS – B	1276.1§	1267.1§	9.0	1229.0	47.1	1244.6	1032.4 1005.9	212.2 238.7	1268.2§ 1210.8	N/A	3105.7	2266.9	838.8	3105.1	0.6
ASN/GLN – B	1273.1	1268.1	5.0	1226.6	46.5	1227.3	980.9	246.4	1261.8 1195.1	N/A	3312.0	2415.0	897.0	3311.9	0.1
H$_2$O – B	1271.8	1266.0	5.8	1227.1	44.7	1221.5	966.2	255.3	1254.7 1204.9	N/A	3408.4	2483.1	925.3	3408.4	0.0
ASP/GLU – B	1268.9§	1263.3	5.6	1224.9	44.0	1221.3 1210.6	971.4	249.9 239.2	1252.5 1203.9	N/A	3427.8 3383.8	2492.4	935.4 891.4	3427.8 3383.8	0.0
SER – B	1267.6	1260.4	7.2	1221.0	46.6	1225.7	976.6	249.1	1262.9 1195.8	N/A	3370.7	2456.0	914.7	3370.6	0.1
THR – B	1266.2	1259.1	7.1	1223.0	43.2	1228.9	980.6	248.3	1260.0 1207.0	N/A	3364.4	2451.6	912.8	3364.4	0.0

BB – B	1265.5	1258.2§	7.3	1222.5	43.0	1227.6	975.7§	251.9	1258.9§ / 1206.9	N/A	3367.8	2454.6	913.2	3367.8	0.0
ASP/ GLU – B	1263.5§	1255.4	8.1	1221.8§	41.7	1222.1§	965.0	257.1	1254.0 / 1205.2	N/A	3447.5	2511.6	935.9	3447.4	0.1
MET – B	1261.4	1253.8	7.6	1219.7	41.7	1197.5	938.5	259.0	1242.6§ / 1196.8	N/A	3440.7	2506.9	933.8	3440.7	0.0
TYR – B	1260.9	1254.2	6.7	1219.5	41.4	1201.0	947.1	253.9	1245.0 / 1199.8	N/A	3497.6	2547.3	950.3	3497.6	0.0
CYS – B	1258.7	1251.6	7.1	1218.8	39.9	1194.6	940.2	254.4	1242.0 / 1195.5	N/A	3452.0	2514.7	937.3	3452.0	0.0
ARG	1265.8	1257.0	8.8	1223.4	42.4	1249.6	1001.6	248.0	1271.6 / 1212.2	N/A	2999.5§	2194.1	805.4	2999.5§	0.0
ASN/ GLN	1257.5§	1248.3	9.2	1215.1	42.4	1235.9	994.0§	241.9	1270.8 / 1197.8	N/A	3221.9	2352.6	869.3	3221.9	0.0
ASP/ GLU	1245.9	1230.8	15.1	1201.5	44.4	1221.3	975.4	245.9	1258.4 / 1211.3	N/A	3338.9 / 3252.9	2421.7	917.2 / 831.2	3338.9 / 3252.9	0.0
2 H$_2$O	1254.7	1246.9	7.8	1215.9	38.8	1233.8	977.7	256.1	1259.8 / 1207.8	N/A	3361.0	2449.0	912.0	3361.0	0.0
2 SER	1251.6	1243.8	7.8	1212.3	39.3	1237.1	984.3	252.8	1267.7 / 1197.1	N/A	3332.2	2428.9	903.3	3332.2	0.0
HIS+	1204.4	1199.9	4.5	1181.1	23.3	1163.0	939.1 / 897.1	223.9 / 265.9	1226.1¶ / 1191.2 / 1170.8	N/A	3595.5	2617.4	978.1	3595.5	0.0

	HIS+	1200.7	1196.4	4.3	1178.6	22.1	1176.5	900.3[§]	276.2	1229.1 1191.8[§]	N/A	3596.1	2617.6	978.5	3596.0	0.1
	LYS+	1198.0	1191.6	6.4	1187.4 1168.8	10.6 29.2	1159.7[§]	903.8[§]	255.9	1222.8[¶] 1187.4 1168.8	N/A	3593.8	2616.2	977.6	3593.8	0.0
	ARG+	1194.1	1190.0	4.1	1175.6	18.5	1177.8	923.5	254.3	1229.4 1191.6	N/A	3585.6	2609.8	975.8	3585.6	0.0

[#]The vibrational frequencies represent coupled vibrations of O–H bending and C–H bending on the main chain.

[§]The vibrational frequency was the average of two coupled vibrational modes that their frequencies are next to each other.

[¶]When the vibrations of O–H bending and C–H bending are coupled, the bottom two frequencies are due to further coupling of O–H bending motion and motions of hydrogen-bond partner.

[∥]The vibrational motion of O–H bending without isotopic labeling is uncoupled and it is coupled when four hydrogen atoms on the tyrosine ring are deuterated. Therefore, the frequencies are not comparable.

The color codes and the methods for frequency calculations are the same as those in Table 3.2.
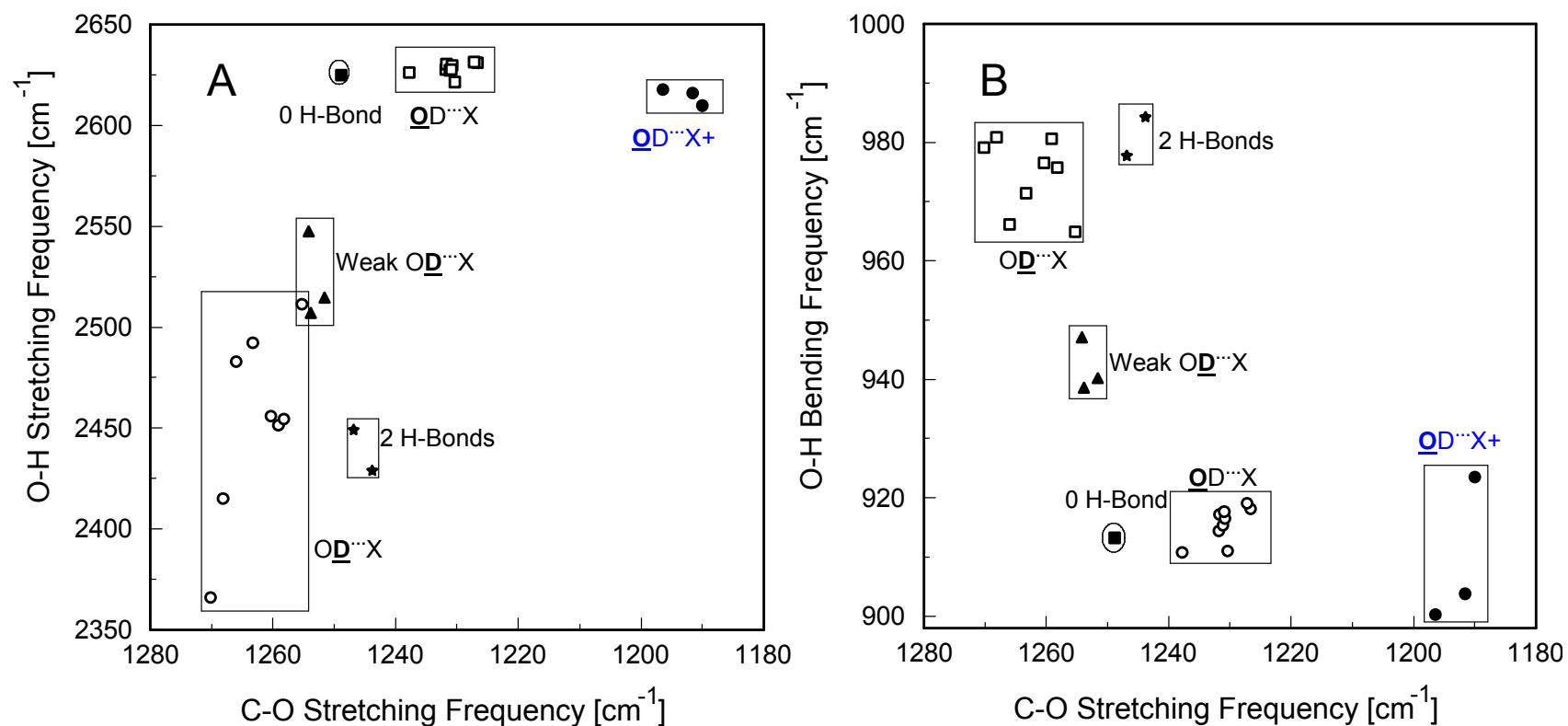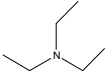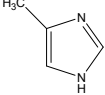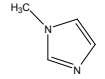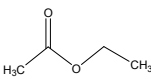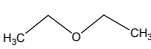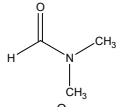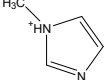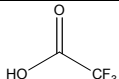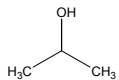
Figure 3.7 Two–dimensional correlations of the O–D and C–O stretching modes (A) and the O–D bending and C–O stretching modes (B) for probing the specific type of hydrogen–bonding interactions in proteins: zero hydrogen bond (■), one hydrogen bond **O**D–X with neutral polar groups (□), one hydrogen bond **O**D–X with positively charged groups (●), one hydrogen bond O**D**–X with neutral polar groups (○), one hydrogen bond O**D**–X with Cys or Met or Tyr (▲), and two hydrogen bonds **OD**–X with neutral polar groups (✱).

Experimental evidence for a vibrational spectral marker

To establish the C–O stretching frequency as a vibrational spectral marker for probing the hydrogen-bonding status of tyrosine side–chain groups, experimental evidence will be required. A model compound of tyrosine side–chain group, p-cresol in various solvents including aprotic and apolar solvent, proton acceptor solvents, and proton donor solvents, had been studied using FTIR and Raman spectroscopy (Hienerwadel et al., 1997; Takeuchi et al., 1989). The results from both literatures are summarized in Table 3.5. The vibrational modes of C–O stretching and O–H bending appear at 1255 cm^{-1} and ~1176 cm^{-1} for *p*-cresol in CCl$_4$, i.e., zero hydrogen-bonding interaction. In proton-acceptor solvents, the frequency of C–O stretching mode of *p*-cresol is up–shifted from 1255 cm^{-1} to 1263 – 1272 cm^{-1} and the frequency of O–H bending mode of *p*-cresol is up–shifted from 1176 cm^{-1} to 1199 – 1251 cm^{-1}. In proton-donor solvents (strong acid), the frequency of C–O stretching mode of *p*-cresol is down–shifted from 1255 cm^{-1} to 1235 – 1240 cm^{-1}. Furthermore, the Raman spectroscopy of L-tyrosine hydrochloride and glycyl-L-tyrosine hydrochloride showed that the C–O stretching frequency was 1232 – 1236 cm^{-1} and crystal structure analysis of the two chemicals identified the hydrogen bonding interactions to be <u>O</u>H–X (Takeuchi et al., 1989). In the solvents that serve as both proton donor and acceptor, the frequencies of C–O stretching and O–H bending modes of *p*-cresol are in the range of 1260 – 1240 cm^{-1}. These experimental data agrees well with our computational data (Table 3.6).

TABLE 3.5 Vibrational modes of C–O stretching and O–H bending of *p*-cresol in different solvents

Solvent*	Solvent structure	No. of H-Bond	Exp. ν_{C-O} (cm^{-1})	Exp. ν_{O-H} (cm^{-1})	Ref.[#]
CCl$_4$	CCl$_4$	0	1255	1177 1175	1–2
Triethylamine			1272	N/A	1
4-MeImH			1271	1251	2
1-MeIm			1269	1251	2
Pyridine			1268	1246	2
Ethyl acetate		1	1267	N/A	1
Diethyl ether			1267	N/A	1
N,N'-DMF			1267	1230	2
1,4-Dioxane			1265	1224	2
1-MeImH$^+$			1263	1229	2
Trifluoroacetic acid			1240	N/A	1
TCL-acetic	N/A	1	1235	N/A	2
Water	H$_2$O		1260	1240	1–2
2-propanol		2	1261	1244	2

*Abbreviations: 4-MeImH is 4-methylimidazole; 1-MeIm is 4-methylimidazole; N,N'-DMF is N,N'-dimethylformamide.

[#]References are quoted in numbers: (1) (Takeuchi et al., 1989) and (2) (Hienerwadel et al., 1997).

TABLE 3.6 Comparison of experimental and computational C–O stretching and O–H bending frequencies

No. of H-Bond	Type of H-bonds	Exp. ν_{C-O} (cm^{-1})*	Cal. ν_{C-O} (cm^{-1})#	Exp. ν_{O-H} (cm^{-1}) *	Cal. ν_{O-H} (cm^{-1}) #
0	OH–X	1255	1255	1175–1177	1163
1	**O**H–X	1263–1272	1259–1277	1229–1251	1195–1245
1	O**H**–X	1235–1240	1230–1243	N/A	1155–1163
2	**O**H–X	1260–1261	1252–1255	1240–1244	1234

*The experimental frequencies of *p*-cresol are taken from Table 3.5.

#The computational frequencies of 4-propyl-phenol are taken from Table 3.2.

Experimental evidence of was also found in proteins. We have carefully searched for proteins for which their structures have been studied using both high-resolution X-ray crystallography and infrared/resonance Raman (RR) spectroscopy. The results are summarized in Table 3.7. All tyrosine groups in Table 3.7 forms one hydrogen bond. There is more O**H**–X kind of hydrogen-bonding interactions than **O**H–X. This is due to the fact that O**H**–X is stronger than **O**H–X by 18 kJ/mol (the difference of averaged hydrogen-bond dissociation energy. The corresponding C–O stretching frequency is in the range of 1265 – 1279 cm^{-1} for O**H**–X and 1240 – 1249 cm^{-1} for **O**H–X. This frequency distribution qualitatively agrees well with our computational results. The case of tyrosine with zero hydrogen bond was not found. This is probably because buried phenol group forms hydrogen bond(s) easily. There are cases that tyrosine with two hydrogen bonds in crystal structures, but the infrared/RR data supporting the

band assignment of such tyrosine residue was not available. Noticeably, the absorptions of ionized tyrosine (tyrosinate, 1277 cm^{-1}), retinal chromophore (C-C stretching, 1236 cm^{-1}), tryptophan (1238 cm^{-1}), and amide III (1236 cm^{-1}) were often found in this region. We must carefully differentiate the effect of tyrosine hydrogen bonding from all the possible interactions of other groups in proteins.

TABLE 3.7   Correlations between the C–O stretching frequency and the hydrogen-bonding status of tyrosine side–chain groups in proteins

No. of H-bonds	IR/Raman Freq. C–O (cm^{-1})	IR/Raman Freq. O–H (cm^{-1})	Protein*	Phenol Group of	H-bond Partner	Type of H-bonds	H-bond Length (Å)	pdb code	Resolution
	1279[1]	1255[1]	PSII	Tyr$_Z$	His190	O**H**–X	2.78	2AXT[2]	3.00 Å
	1275[1]	1250[1]	PSII	Tyr$_D$	His189	O**H**–X	2.59	2AXT[2]	3.00 Å
	1274[3]	N/A	CcO (P.D., oxidized)	Tyr35	SER134	O**H**–X	3.09	1QLE	3.00 Å
1	1270/1268[3]	N/A	CcO (P.D., oxidized)	Tyr280	HEA2	O**H**–X	2.37	1QLE	3.00 Å
	1265[5]	N/A	SRII-HtrII	Tyr199	Asn74	O**H**–X	2.74	1H2S[6]	1.90 Å
	1249[7]	N/A	AppA	Tyr21	Gln63	**O**H–X	2.52/2.55/2.59#	1YRX[8]	2.30 Å
	1240[9]	N/A	bR (K)	Tyr185	Asp212	**O**H–X	2.75	1M0K[10]	1.43 Å

*Abbreviation for proteins are: PSII for photosystem II from thermosynechococcus elongates; CcO for cytochrome c oxidase in Paracoccus Denitrificans in fully oxidized state, SRII-HtrII for sensory rhodopsin II–HtrII receptor–transducer complex of *Natronomonas pharaoni*; AppA for flavin adenine dinucleotide (FAD)–containing photoreceptor protein AppA (in which the FAD is bound to a novel so-called BLUF domain) from the purple nonsulfur bacterium Rhodobacter sphaeroides, and bR (K) for K

intermediate state of Bacteriorhodopsin. References are quoted in numbers: [1] (Hienerwadel et al., 1997); [2] (Loll et al., 2005); [3] (Hellwig et al., 2002); [4] (Harrenga and Michel, 1999); [5] (Bergo et al., 2005); [6] (Gordeliy et al., 2002); [7] (Laan et al., 2003); [8] (Anderson et al., 2005); [9] (Rothschild et al., 1986); [10] (Schobert et al., 2002).

#These are the hydrogen-bond lengths corresponding to three chains of the same protein in the same state from on PDB data.

3.4 Conclusion

Our density function theory based *ab initio* computational studies on the hydrogen-bonding properties of a tyrosine side–chain group show that the hydrogen-bonding interactions of O**H**–X of a phenol group as the hydrogen-bond donor is stronger than that of **O**H–X. We identified that there is a strong correlation between the frequency of C–O stretching frequency and the hydrogen-bonding status of tyrosine side–chain that is hydrogen bonded to polar or positively charged amino acid side chains (Table 3.2 and 3.3). This correlation is further supported by available experimental evidence (Table 3.5 and 3.6). We provided both computational and experimental evidence that support the establishment of the C–O stretching frequency of tyrosine side chain as a vibrational spectral marker for probing the hydrogen-bonding status: 1255 cm^{-1} for zero, 1230 to 1243 cm^{-1} for phenolic oxygen **O**H–X and 1265 to 1277 cm^{-1} for hydroxyl hydrogen O**H**–X forming a strong hydrogen bond with neutral polar amino acid side–chain groups, and 1194 to 1204 cm^{-1} for hydroxyl hydrogen

<u>O</u>H–X forming a strong hydrogen bond with positively charges amino acid side chain groups. A two-dimensional infrared spectroscopy, C–O stretching vs. O–H stretching (Figure 3.6, valid only if protein samples are prepared in hydrated film or crystal state) or C–O stretching vs. O–H bending (Figure 3.6), may be helpful to identify two hydrogen bonding interactions or weak O<u>H</u>–X interaction from zero hydrogen bond when the marker of C–O stretching frequency is inadequate. Upon OH $\rightarrow$ OD deuteration (Table 3.4), the frequencies of C–O stretching and O–H stretching and bending modes are red–shifted to various extents: 10 cm^{-1} or less for C–O stretching, 250 cm^{-1} for O–H bending, and more than 800 cm^{-1} for O–H stretching frequency. These frequency shifts are not good indicator of hydrogen-bonding status of tyrosine side–chain group. Upon ring–H_4 $\rightarrow$ ring–D_4 deuteration, the frequency shift of C–O stretching mode is sensitive to specific type of hydrogen-bonding interactions and that of O–H stretching frequency is insensitive to hydrogen- bonding interactions (Table 3.4).
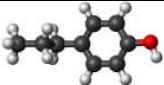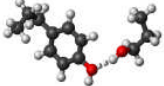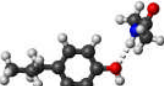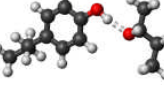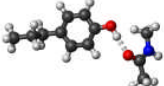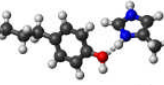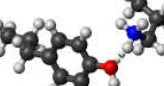
This vibrational spectral marker for hydrogen-bonding interaction is expected to enhance the power of time-resolved Fourier transform infrared spectroscopy in characterizing the structures of functionally important intermediate states of proteins and in elucidating the functional mechanism of proteins.

3.5 Supplemental material

Table 3.8 shows the isotopic labeling effects on vibrational frequency of ring coupled with O–H bending motion. For unlabeled phenol group, this band is up–shifted from 1326 cm^{-1} to >1350 cm^{-1} region for O<u>H</u>–X with neutral polar

group, it is down–shifted to 1322 cm^{-1} for $\underline{O}$H–X with positively charged groups, and it does not shift for $\underline{O}$H–X with neutral polar groups. This vibrational frequency shows sensitivity to hydrogen-bonding interactions. Upon OH → OD deuteration, however, it is down–shifted to 1296 – 1306 cm^{-1} for all hydrogen-bonding interactions. Hence it is difficult to distinguish four types of hydrogen-bonding interactions from 10 cm^{-1} frequency differences. In addition, the intensity of this band from our experimental data is small. Therefore, the vibrational frequency of ring coupled with O–H bending mode is not a good candidate of vibrational spectral marker for probing hydrogen-bonding status of tyrosine side–chain groups.

TABLE 3.8 Isotopic shift of vibrational frequency of ring plus O–H Bending

Structure	H-Bond Partner	$\nu_{\text{ring \& OH bending}}$				
		OH (cm^{-1})	OD (cm^{-1})	$\Delta\nu$ (cm^{-1})	ring–D4 (cm^{-1})	$\Delta\nu$ (cm^{-1})
	N/A	1327.7	1298.8	28.9	1314.1	13.6
	SER – 1	1326.1	1306.6	19.5	1311.2	14.9
	H$_2$O – 1	1325.9	1302.0	23.9	1312.3	13.6
	BB – NH	1326.1	1300.6	25.5	1311.5	14.6
	H$_2$O – 2	1350.4	1296.2	54.2	1321.7	28.7
	SER – 2	1357.3	1297.7	59.6	1321.8	35.5
	BB – CO	1356.0	1297.8	58.2	1321.9	34.1
	HIS+	1322.6	1298.5	24.1	1306.0	16.6
	LYS+	1319.6	1298.1	21.5	1305.9	13.7

CHAPTER IV

PROTON TRANSFER IN PYP

4.1 Introduction

Proton transfer is unambiguously the most fundamental and common process and it plays key roles in biological systems. Due to the facts that photoactive yellow protein (PYP) has rhodopsin-like photocycle (rhodopsin is a protein in the membrane of the rod photoreceptor cell in the retina of the eye) and it is considered a structural prototype of the PAS domain superfamily because of its small size (125 amino acids), PYP is an excellent model system to study proton transfer mechanism in proteins.

Photoactive yellow protein (PYP), a small water-soluble protein that was extracted from bacteria *Ectothiorodospira halophila*, is known responsible for the negative phototaxis of blue light (Hoff, W. D, 1995). Its unique *p*-coumaric acid (pCA) chromophore covalently linked to Cys69 via thioester linkage carries a negative charge in pG_{446} state (446 nm is the peak absorption wavelength of PYP in pG receptor state). It is energetically stabilized by hydrogen bonding network, namely, H-bonds between the chromophore and Glu46, Tyr42, and backbone of Cys69. Upon absorption of a blue photon, a reversible photocycle is initiated (Figure 4.1). In a few nanoseconds, a red-shifted intermediate state, named pR_{465} is formed (Hoff, W. D. et al., 1994; Meyer et al., 1989). In pR_{465},

chromophore photoisomerization dramatically increases the proton affinities of the phenolic oxygen of the pCA chromophore so that proton transfer from Glu46
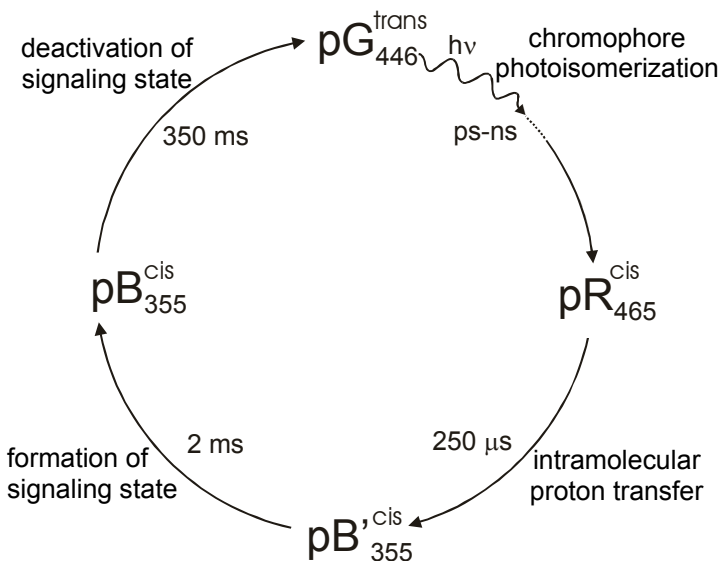
Figure 4.1 The photocycle model of PYP.

to the pCA chromophore has taken place in 250 μs based on modeling of the photocycle kinetics (Bao, 2004), thus a new intermediate state pB'_{355} is formed (Xie et al., 1996; Xie et al., 2001). The transition from pB'_{355} to pB_{355} signaling state takes place on 2 ms

timescale. Then the photocycle is completed by full recovery from pB_{355} to pG446 in a few seconds (Hoff, W. D. et al., 1994; Meyer et al., 1989). Our FTIR difference spectroscopy showed that the protein structure has no dramatic changes in photoisomerization (pR) and proton transfer (pB') process until in signaling state (pB) (Xie et al., 2001). A receptor activation mechanism has been proposed that a new buried charge E46 after proton transfer is the electrostatic epicenter that triggers and drives large conformational changes of the protein, so called "protein quake" (Xie et al., 1996; Xie et al., 2001).

Crystal structures of PYP at early intermediate state

The light-induced chromophore isomerization in pR is the crucial step that allows the conversion of light energy into a structural change in photosensitive

126

proteins for many light-mediated biological processes, such as vision, photosynthesis, and so on. It is therefore important to have accurate and correct chromophore structure in pR for understanding proton transfer mechanism in PYP.

The crystal structure of PYP at early intermediate state has been reported before. The first structure was determined by X-ray crystallography at 287 K (Perman et al., 1998). Later in the same year, the structure in cryotrapped intermediate state at 149 K were reported (Genick et al., 1998). Shortly after, picosecond and femtosecond spectroscopic studies reveal that there are two short-lived intermediate states (I_o and $I_o^{\ddagger}$) occurring within a few ps to 1 ns timescale of the photocycle (Devanathan et al., 1999; Ujj et al., 1998). Then the crystal structure was resolved by cryotrapping the photoactivated protein at 85 K (Kort et al., 2004) and at 110 K (Anderson et al., 2004b). The two structures at low temperature resemble the crystal structures of PYP in $I_o^{\ddagger}$ and pR intermediate states at 288 K reported using time-resolved Laue crystallography (Ihee et al., 2005). The pCA chromophore surrounded by the binding pocket in these crystal structures of PYP at early intermediate state is shown in Figure 4.2. Whether they are correct is still under debate. There were theoretical modeling of PYP at early intermediate state using molecular dynamics (MD) simulation in the past: some amino acids had been observed that have strong interactions with chromophore upon photoexcitation in femtosecond simulations (Yamato et al., 1998); concerted large chromophore – linked fluctuations were identified in pG ground state and in a form that has isomerized chromophore (van Aalten et al.,

1998); the *trans – cis* isomerization was modeled either by temporarily removing the double bond potential (van Aalten et al., 1998) or by simulating the relaxation of the chromophore from excited state (Groenhof, G. et al., 2004; Groenhof, G et al., 2002a; Groenhof, G. et al., 2002b). However, the pR chromophore structures used or obtained in the MD simulations were not consistent (Antes et al., 2002; Groenhof, G. et al., 2004; Groenhof, G. et al., 2002b). Here we performed MD simulations of PYP with three *cis* chromophore configurations, with the support of experimental evidence, we identified the structure nature of pCA chromophore of PYP at early intermediate state. This provides a specific guide on how to use time-resolved infrared spectroscopy to identify which photoproduct is formed after photoisomerization.
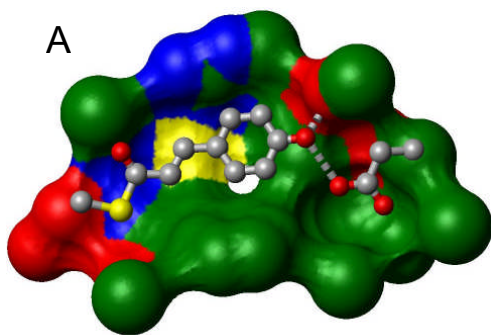
**2PHY (1.4 Å), Room Temperature
(Borgstahl et al., 1995)**
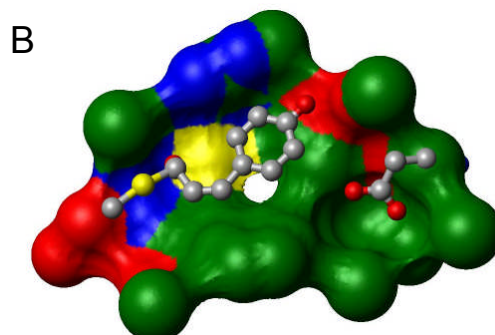
**2PYR (1.9 Å), 1 ns, 287
K(Perman et al., 1998)**
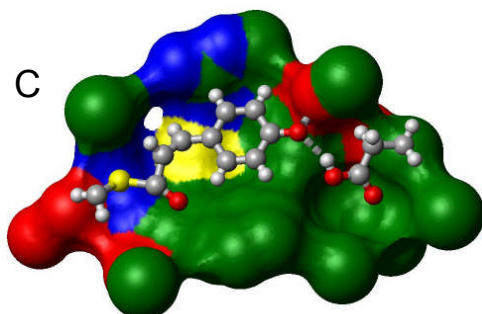
**3PYP (0.85 Å), 149 K
(Genick et al., 1998)**
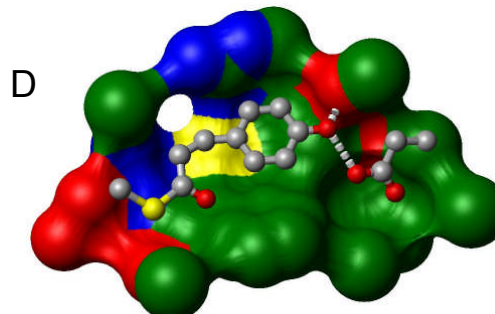
**1UWP (1.2 Å), 85 K
(Kort et al., 2004)**
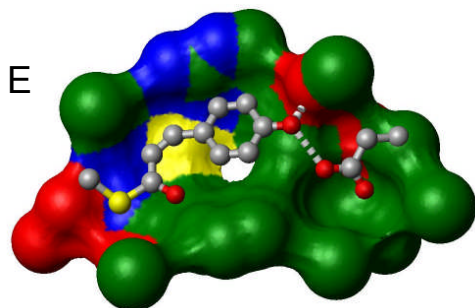
**1OT9 (1.05 Å), 110 K
(Anderson et al., 2004b)**
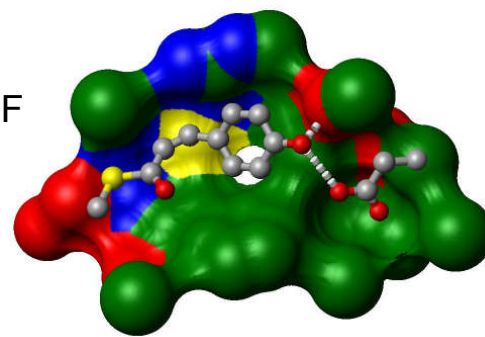
**1OT9 (1.05 Å), 110 K
(Anderson et al., 2004b)**
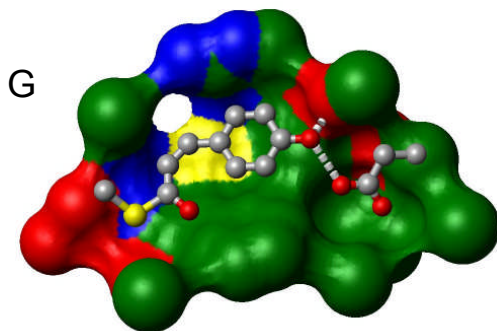
**1TS8 (1.6 Å), 20 ns, 288 K
(Ihee et al., 2005)**
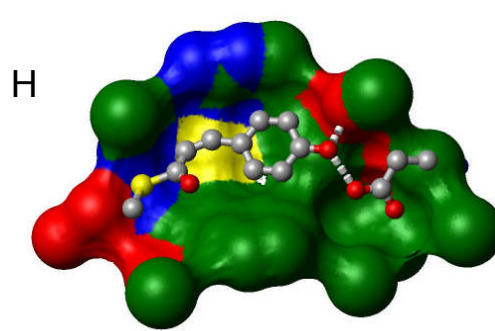
**1TS7 (1.6 Å), 180 μs, 288 K
(Ihee et al., 2005)**

Figure 4.2 The pCA chromophore and the COOH group of Glu46 in the binding pocket of PYP crystal structures in pG (2PHY, A) and in early intermediate states (2PYR, B; 3PYP, C; 1UWP, D; 1OT9, E and F; 1TS8, G; and 1TS7, H). The binding pocket is formed from the atoms on the first shell, and is depicted as a surface generated using a probe radius of 1.4 Å. These atoms are from Tyr42 (OH, CZ), Thr50 (OG1, CB, CG2), Arg52 (NH1, NH2, CZ, NE, CD), Tyr94 (CE2), Thr95 (O), Phe96 (C, CA, CB, CD1, CD2, CE1, CE2, CG, CZ), Asp97 (N, OD1, OD2), Tyr98 (N, CB), Met100 (SD), Gly29 (CA), Glu46 (CA, O), Ile49 (CB, CD1, CG2), and Val122 (CG1). The surface is presented with carbon atoms in green, oxygen atoms in red, nitrogen atoms in blue, and sulfur atoms in yellow. The chromophore and the side chain group of Glu46 are illustrated in ball and stick with gray for carbon atoms, red for oxygen atoms, yellow for sulfur atoms, and white for hydrogen atoms.

Proton affinities regulating proton transfer in proteins

Proteins are able to perform an enormous variety of functions, while using only a limited number of underlying processes. One of these is proton transfer, found in a wide range of biological energy transformation, signal transduction, and enzymatic reactions. The process of intra-protein proton transfer involves a proton donor and a proton acceptor. The proton affinity, defined as the negative of the enthalpy change at standard condition (i.e. temperature and pressure), is frequently used to describe the ability of a molecule to accept a proton. When the proton affinities of a proton donor are stronger than that of its proton acceptor, it is energetically more stable for the proton to remain on the donor. In order for proton transfer to taken place, such proton affinities of the proton donor and the proton acceptor must be reversed. However, it is difficult to determine the proton affinities experimentally whereas *ab initio* calculations provide reliable proton affinities.

The pKa of the pCA chromophore in free solution is 10 (Hoff, W. D, 1995). It is downshifted to 2.8 when the chromophore is buried in the protein (Hoff, W. D, 1995; Meyer et al., 1987). This indicates that in normal condition of pH 7, the

pCA chromophore is ionized in protein while it is neutral in solution. The pKa shift is 7.2, that is, 41.7 kJ/mol in terms of energy using following calculations:

$$\Delta G = -RT\ln\frac{K_a}{K_a(pH7)} = 2.303RT \times \Delta pK_a = 5.7kJ/mol \times \Delta pK_a = 41kJ/mol$$

where ΔG is the difference in Gibbs free energy, R is the gas constant, T is the temperature in Kelvin, and K_a is equilibrium constant. By calculating the energy difference before and after proton transfer using *ab initio* methods, we have quantitatively studied the effects of hydrogen bonding interactions, dielectric media, and structural configuration on the relative proton affinities. We have identified a plausible structural element that may be crucial to increase the proton affinity of the chromophore and trigger the intra-protein proton transfer during PYP photoreceptor activation. In addition, we have calculated energy barrier between the proton donor Glu46 and proton acceptor pCA chromophore along the proton transfer pathway, it provided us deep insights into proton transfer mechanism.

4.2 Computational methods

MD simulations of pG

The starting structure for MD simulation of pG was taken from PYP crystal structure at 1.4 ´ resolution, 2PHY (Borgstahl et al., 1995), in the Brookhaven Protein Databank (Bernstein et al., 1977). All titratable groups and the chromophore in the protein were considered to carry charges, with the exception of protonated Glu46. Therefore, eight sodium ions and two chloride ions were introduced to compensate net six negative charges of the protein by replacing the water molecules with the highest or the lowest electrostatic potential. Then

hydrogen atoms are added at proper locations of amino acids according to the built-in database of AMBER 6.0 molecular dynamics software package (Case et al., 1999). In particular, hydrogen atoms on the carboxylic oxygen of Glu46 and on the hydroxyl oxygen of Tyr42 were made pointing toward the phenolic oxygen of pG chromophore. Then the protein with 10 ions were hydrated by immersing it in a rectangular periodic box of water, with dimension of 55 Å × 65 Å × 61 Å, the volume of which was 218,075 $Å^3$. A periodic boundary condition was imposed on the box. In total, the system contained 4747 TIP3P water molecules (Jorgensen et al., 1983), including 92 crystal water molecules. The water/protein ratio is comparable to that of 10 mM PYP in water, a typical protein concentration for our FTIR experimental studies. Then the entire system containing 16,202 atoms was energy minimized for 200 steps by using a steepest descent algorithm. The energy reaches the minimum after 50 steps. Subsequently, the system was simulated for 200 ps with fixed protein coordinates to allow relaxation of water molecules and ions around the protein. At last, the entire system including protein, ions and water molecules was equilibrated for 5 ns. During equilibration the protein structure was stable and pCA chromophore remained a fairly planar conformation. The tertiary structure of PYP in a box of water is shown in Figure 4.3.
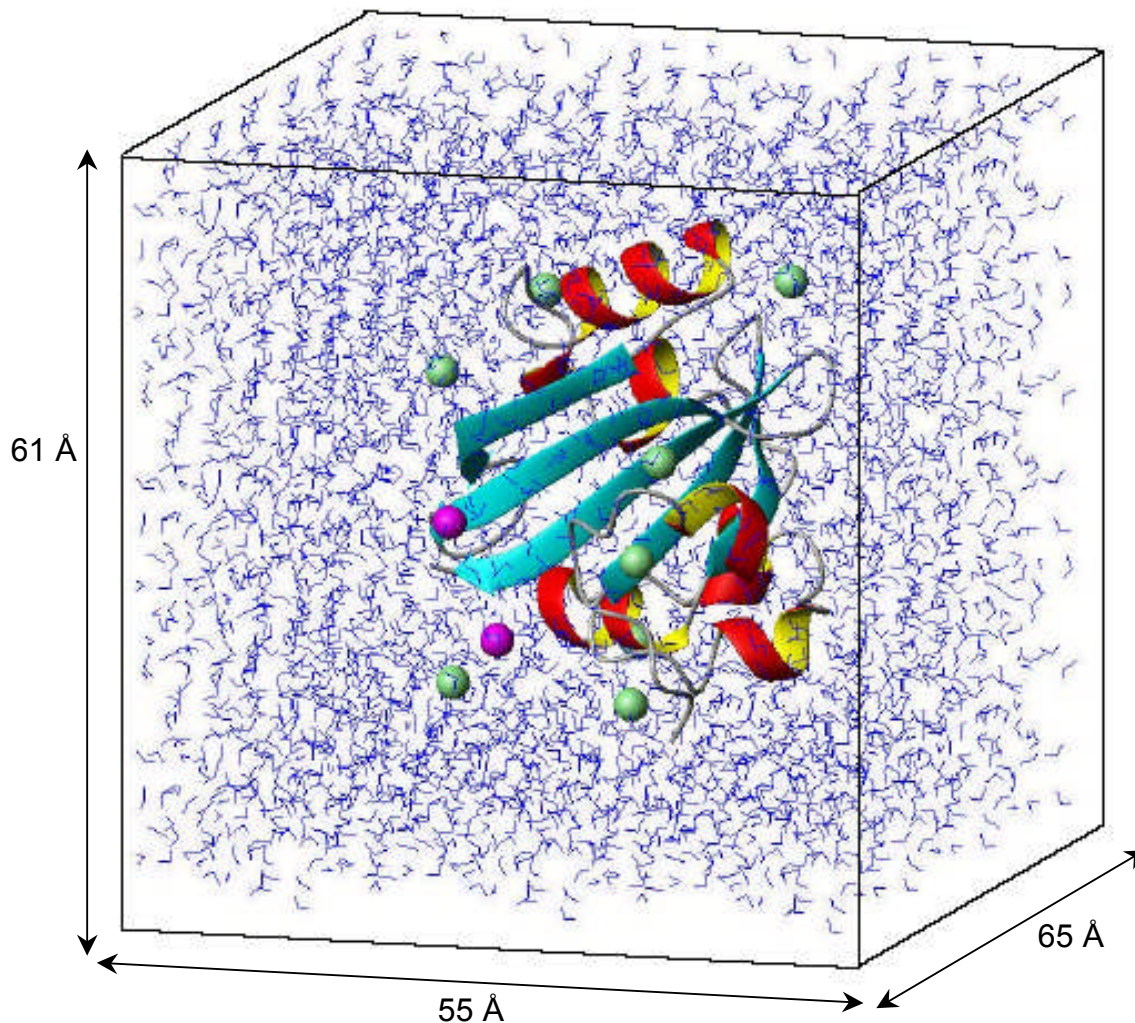
Figure 4.3 The PYP in a rectangular box of water. The 4747 water molecules are shown in blue lines. There are 8 sodium ions (light green) and 2 chloride ions (magenta) surrounding the protein in the box.

MD simulations of pR

The chromophore was manually isomerized from *trans* to *cis* in three orientations as in (1) the proposed model in the paper (Xie et al., 1996) (pR$_1$); (2) the pR crystal structure at 149 K (pR$_2$); (3) the model in the paper (Groenhof, G. et al., 2004) (pR$_3$). The chemical structures of pR$_1$, pR$_2$, and pR$_3$ chromophore are shown in Figure 4.4.
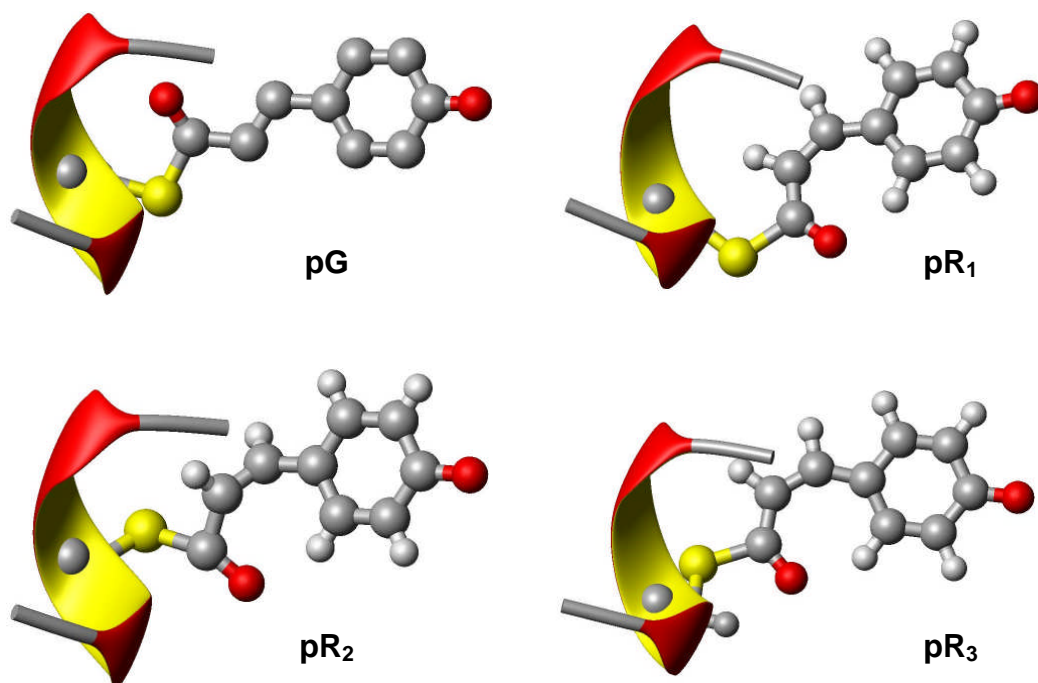
Figure 4.4 The chromophore structures of PYP in pG crystal (2PHY.pdb) and for MD simulations of pR_1, pR_2, and pR_3. Color code: gray for carbon atoms, red for oxygen atoms, yellow for sulfur atoms, and white for hydrogen atoms. The α-helix that contains the Cys69 is shown as reference.

To generate the starting structures for MD simulations of pR, eight frames were selected from 3 ns to 5 ns of pG simulations, every 250 ps a frame. The *trans* chromophore in these frames was replaced by a *cis* chromophore that is covalently linked to Cys69. Then 1000 steps energy minimization of each system including protein, water molecules and ions were carried out. Eventually 500 fs simulation every 1 fs a step, 2 fs a time window, followed by 100 ps simulation every 2 fs a step, 1 ps a time window, and continued by 5 ns simulations every 2 fs step, 1 ps a time window were performed. In total, we have 40 ns MD simulations of pR_1, pR_2, and pR_3, respectively.

Mimic chromophore isomerization

The *cis* chromophore was modeled so that the dihedral (torsion) angles of C=O and C–H on the central double bond was twisted from 0° to ±180°, every 30° a step (12 steps), to mimic chromophore co-isomerization. Then this chromophore replaced the *trans* chromophore in the last frame of 5 ns simulations of pG. Next, 50 steps energy minimization including protein, water molecules and ions were carried out to remove atomic clash. Eventually twelve simulations for 10 ps, every 2 fs a step, 2 fs a time window were performed.

Molecular dynamics simulations of pB'

The last frames were extracted from 5ns MD simulation of pR_1 and pR_3 and the velocities of atoms from the last frames of pR simulations were preserved for MD simulation of pB', in which the proton on the Glu46 was manually moved to the chromophore. Eventually 5 ns simulations of pB', every 2 fs step, 1 ps a time window were carried out.

All the simulations were performed with AMBER 6.0 program package (Case et al., 1999) at constant pressure (1 atm) and constant temperature (300 K). The AMBER6 all-atom force field (Cornell et al., 1995) was applied for PYP. The parameters for the chromophore are described below. Application of the SHAKE algorithm (van Gunsteren and Berendsen, 1977) to constrain bond lengths allowed a time step of 2 fs. A cut-off of 10 Å was used for long-range electrostatic interactions. The required time for 1 ns was approximately 1 week on a single processor with speed of Pentium IV 1.7 GHz of our Linux computer cluster.

The chromophore force fields

The force constants of bond lengths, angles, dihedral angles for the chromophore were taken from the reference paper if available (Cornell et al., 1995). Otherwise they were made analogous to other molecules in the paper. The equilibrium bond lengths, angles, dihedral angles as well as charge distribution (Mulliken) of pG, pR and pB' chromophore were taken from *ab initio* calculations of geometry optimized chromophore model compound using B3LYP/6-31G(d) method with Gaussian03 software package (Frisch, M. J. et al., 2003b). In *ab initio* calculations, the pG chromophore was in three hydrogen bonding interactions with $CH_3(CH_2)_2COOH$, C_6H_5OH, and CF_3NHCF_3, analogous to Glu46, Tyr42 and backbone of Cys69. The strong polarity of CF_3NHCF_3 made the hydrogen bond length close enough to that in crystal structure (2PHY). The pR and pB' chromophore model compounds were in two hydrogen bonding interactions with analogy to Glu46 and Y42. The chromophore model compounds were capped at the sulfur atom with methyl group.

Proton affinities calculation

The structure of the chromophore model compound with one, two, and three hydrogen bonding interactions with proton on E46 model compound and with proton on the chromophore was optimized using semi–empirical method PM3 (Gaussian03). Then the energy was calculated using B3LYP/6-31G(d) in vacuum and in dielectric media. For the calculation of structural configuration on the proton affinities, the structure of the chromophore and E46 with fixed dihedral
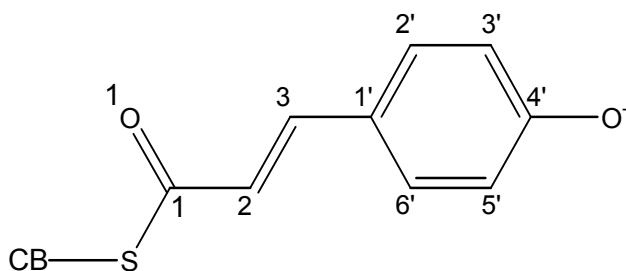
Figure 4.5 The chemical structure of pCA chromophore.

angle of C3=C2-C1=O1 (see Figure 4.5 for atom labels) was optimized using PM3 and the energy was then calculated using B3LYP/6-311+G(2d,p) method.

Proton barrier calculation

The structure of *trans* chromophore model compound in one hydrogen bonding interaction with Glu46 model compound was optimized using semi–empirical method PM3 (Gaussian03) with two fixed distances between the phenolic oxygen of pCA chromophore and hydroxyl oxygen at 2.7 Å and 3.0 Å and between the proton and the phenolic oxygen along the proton transfer pathway stepwise, i.e., from 0.8 Å to 1.8 Å, 0.1 Å a step. Then the energy was calculated for each step using HF/3-21G method (Frisch, M. J. et al., 2003b).

4.3 Results and discussion

Structural deformation of the chromophore in pR crystal structures at 287 K and 149 K

The chromophore photoisomerization takes place in less than 1 picosecond (Larsen et al., 2004). Based on FTIR spectroscopic data, Xie and coworkers concluded that the photoisomerization involves co–isomerization of the central C2=C3 double bond and C1-S (See Figure 4.5 for atom labels) single bond so that the phenolic ring and Cys69 are anchored, only limited structural changes occur in the pCA and its binding pocket (Xie et al., 1996; Xie et al.,

2001). The chromophore structures in both reported pR crystal structures at 287 K and at 149 K are highly deformed, with abnormal dihedral angle of either the carbonyl group or the central double bond.

The pR chromophore at 287 K displays abnormal dihedral angle of O1-C1-C2=C3, 83.6º (See Figure 4.2 B), while averaged value of this angle is -4.5º in the pR_1 simulation, as indicated in Table 4.1. The dihedral angles of CB-S-C1-C2 and of CB-S-C1=O1 also show abnormalities compared to those in MD simulation of pR_1. We performed *ab initio* energy calculation of a chromophore model compound, pCA methyl thioester, in relaxed *cis* conformation and in deformed conformation with fixed dihedral angle of O1-C1-C2=C3 of 83.6°. It is shown in Table 4.2 that the energy of the deformed chromophore structure in pR at 287 K is 39.8 kJ/mol higher than the relaxed structure, not to mention how much energy it would cost if both abnormal dihedral angles were fixed. This energy (39.8 kJ/mol) is comparable to the protein folding energy, and poses serious structural instability to pR. In addition, the phenolic ring is flipped away from Glu46 (See Figure 4.2 B) while it is supposed to stay intact with Glu46 and Tyr42 during photoisomerization. It is not likely for the binding pocket to open such big cavity to allow conformational change as flipped ring in very short time. Therefore, we conclude that this pCA chromophore structure is highly deformed and structural incorrect.

Table 4.1. Important dihedral angles of the chromophore in pG and pR

Dihedral Angle	pG		pR		pR	
	Crystal	MD pG	Crystal$_{287K}$	MD pR$_1$*	Crystal$_{149K}$	MD pR$_2$*
O1=C1-C2=C3	-22.7	-1.3	83.6	-4.5	38.2	8.8
C1-C2=C3-C1'	173.6	166.1	0.0	-6.4	-72.2	8.7
C2=C3-C1'=C2'	-176.9	-175.7	0.2	-3.6	36.5	13.7
CB-S-C1-C2	174.1	177.8	-74.9	-8.4	-126.7	179.3
CB-S-C1=O1	11.5	-0.2	51.8	174.0	87.4	-2.5
H2-C2=C3-H3	N/A	174.7	N/A	-1.7	-72.2	5.8
H2-C2=C3-C1'	N/A	-9.5	N/A	176.7	107.8	-173.2
H3-C3=C2-C1	N/A	-9.6	N/A	175.3	107.8	-172.3

The extent of chromophore deformation in pR is measured in terms of dihedral angles. *The dihedral angles for MD simulations are averaged over 1 ns.

The difference between the pR chromophore in crystal structure at 149 K and the chromophore by co-isomerization published in the paper (Xie et al., 1996) lies in the position of the sulfur atom of Cys69 apart from abnormal dihedral angles of the chromophore. These dihedral angles are compared with MD simulation of pR$_2$, in which the chromophore is relaxed to its planar structure, as shown in Table 4.1. The very important dihedral angle of central double bond C1-C2=C3-C1', -72.2° in crystal structure (See Figure 4.2 C), costs 55.8 kJ/mol energy increases than the relaxed pR chromophore conformation by *ab initio* calculation, while it is 8.7° in pR$_2$ simulation with its energy increase only by 1.0 kJ/mol. It does form H-bonds with Glu46 and Tyr42. However highly twisted central double bond is clearly observed as well as other abnormal angles.

Table 4.2: Minimum energy increase due to chromophore deformation

	Dihedral Angle		Energy Increase
pCA_{calc}	NA	NA	0
pCA_pR_{287K}	O1-C1-C2=C3	83.6	+39.8 kJ/mol
pCA_pR_{149K}	C1-C2=C3-C1'	-72.2	+55.8 kJ/mol
$pCA\_MD\ pR_2$	C1-C2=C3-C1'	8.7	+1.0 kJ/mol

The structure of the chromophore model compound, pCA methyl thioester in relaxed conformation and in fixed dihedral angle was optimized using HF/3-21G method. Then energy was calculated using B3LYP/6-311+G(2d,p) method (Gaussian03).

On the other hand, the PYP in crystal form at cryotrappped photoactive state reported by (Anderson et al., 2004b) has showed that two photoproducts were present at early intermediate state. One photoproduct is consistent with the proposed co−isomerization model in the paper (Xie et al., 1996). The other is consistent with the structure simulating relaxation of chromophore from excited state (Groenhof, G. et al., 2004). The two photoproducts have been assigned to be $I_0^{\ddagger}$ and pR intermediate state at room temperature (Ihee et al., 2005).

Hydrogen bonding network at the photoactive site

We have performed the MD simulations of pR with three chromophore configurations that resemble the proposed co-isomerization model (pR_1), the structure at 149 K (pR_2), and the structure similar to pR_1 with sulfur position flipped 180° (pR_3) according to MD simulations of chromophore isomerization (Groenhof, G. et al., 2004). In comparison with the crystal structures of PYP at early intermediate state, we found that the chromophore structure of pR_1 and pR_3

were consistent with the structures of two photoproducts at low temperature (Anderson et al., 2004b) and in $I_o^{\ddagger}$ and pR intermediate states at room temperature (Ihee et al., 2005).

Normally hydrogen bond length is in a range of 2.7 Å – 3.2 Å (Bergethon, 1998; Vinogradov and Linnell, 1971) and the hydrogen bonding angle of X–H ⋯ X is in the range of 160° – 180°. It is considered that the hydrogen bond is broken if the distance between two electronegative atoms is beyond 3.2 Å or the angle of X–H ⋯ X is less than 160°. Since we manually isomerized the chromophore for MD simulations of pR_1, pR_2, and pR_3 simulations, the hydrogen bonds between phenolic oxygen of the chromophore and hydroxyl oxygen of Glu46 and between phenolic oxygen of the chromophore and phenolic oxygen of Tyr42 broke because of contraction of *cis* chromophore. The two hydrogen bonds quickly reformed in less than 1 ps (data not shown). From out *ab initio* calculations, the length of the pCA chromophore in relaxed cis form is 0.8 Å shorter than that in trans form.

Hydrogen bond between pCA chromophore and Glu46

In 99.2% of 5ns pG simulations, the hydrogen bond between the pCA chromophore and Glu46 remained intact. The free energy in terms of the population of hydrogen bond length based on Bolzmann distribution is shown in Figure 4.6. It clearly indicates that the hydrogen bond within 2.6 Å – 2.7 Å that has the lowest energy is the most populated. And the hydrogen bond beyond 3.0 Å is much less populated.

Figure 4.6 The energy distribution of hydrogen bond length between phenolic oxygen of the pCA chromophore and hydroxyl oxygen of Glu46 taken from the MD simulations of PYP in pG (5 ns). The data points are in solid circles and the curve is polynomials least−squares fitting.

The kinetic traces of the distance between the pCA chromophore and Glu46 in pR simulations are illustrated in Figure 4.7.

Distance of pCA – E46
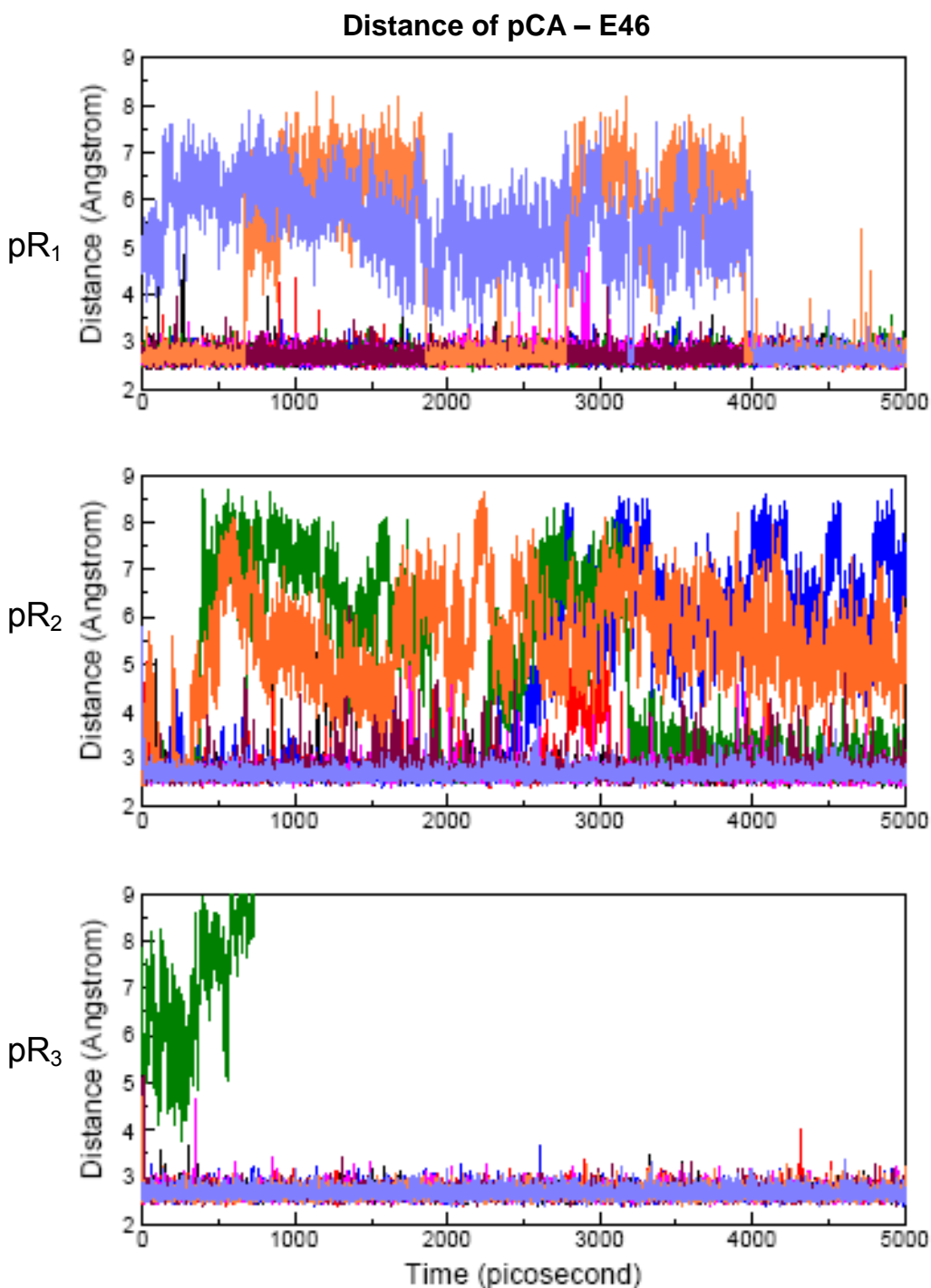
Figure 4.7 The kinetic traces of the distance between phenolic oxygen of the pCA chromophore and hydroxyl oxygen of Glu46 in eight 5 ns MD simulations of PYP in pR_1, pR_2, and pR_3.

For pR$_2$, the distance showed big fluctuation in majority of the eight 5 ns simulations and the hydrogen bond remains broken in two of the simulations. For pR$_3$, the distance was stably within the range of hydrogen bond length in seven of the eight 5 ns simulations except that in one simulation, the distance between the pCA chromophore and Glu46 was 6.8 Å at the first frame and it rapidly increased to 15 Å in later time frames. This was probably due to the bad starting structure.

From the structure point of view, the chromophore structures in pR$_1$ and pR$_3$ simulations were stable in *cis* form. Whereas the chromophore structure in pR$_2$ simulation yields to a *cis* form with flipping ring and isomerization of C–H on the central double bond closer to the ring. As we have discussed earlier in this chapter, the ring flip was not favored based on our FTIR experimental data shown that there were limited structural changes upon chromophore photoisomerization (Xie et al., 2001). In addition, ring flip causes the broken of the hydrogen bond between pCA chromophore and Glu46, as clearly shown in the kinetics traces of pR$_2$ simulations in Figure 4.7. It is inconsistent with our experimental data that unambiguously showed the hydrogen bond between the pCA chromophore and Glu46 remained intact after chromophore photoisomerization (Xie et al., 1996). Therefore, we concluded that the chromophore structure in pR$_2$ simulations was incorrect. We did not performed MD simulations of pB' based on pR$_2$ simulations. From now on, our discussion will be mainly focused on pR$_1$ and pR$_3$ related simulations. The data of pR$_2$ simulations are presented as reference.

The histograms of the distance between pCA chromophore and Glu46 in pG, pR_1, pR_2, pR_3, pB'_pR_1, and pB'_pR_3 simulations is shown in Figure 4.8. Statistically the distance was in the range of hydrogen bond length for 99.2% of 5 ns pG simulations; 86.7% of pR_1, 71.2% of pR_2, and 87.8% of pR_3 simulations along 40 ns; and 97.3% of 5 ns simulations of pB'_pR_1 and 93.6% of pB'_pR_3. We consider two criteria for the formation of hydrogen bonding interaction: (1) the distance is within hydrogen bond length of 3.2 Å and (2) the angle of X–H $\cdots$ X is

within the range of hydrogen bond angle of 160° – 180°. Hence a 2D correlation of distance and angle will be needed in order to see the population of
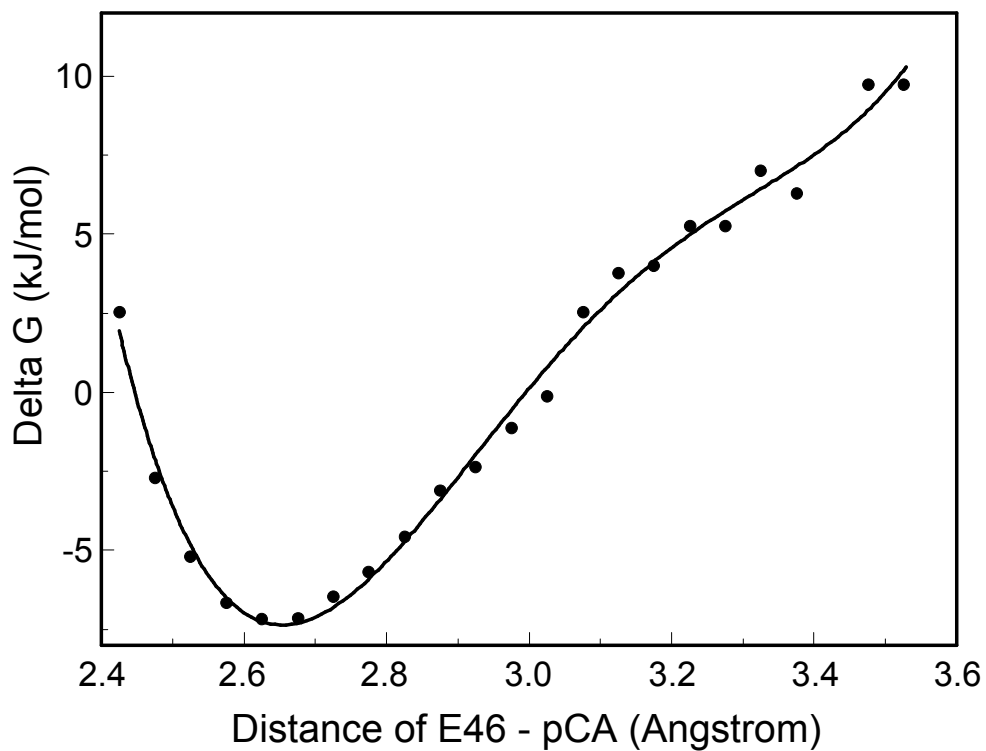

Figure 4.8 The histogram of the hydrogen bond length between phenolic oxygen of the pCA chromophore and hydroxyl oxygen of Glu46 taken from the MD simulations of PYP in pG (5 ns), pR_1, pR_2, and pR_3 (40 ns), and pB' simulations continued from pR_1 and from pR_3 (5 ns).

the hydrogen bond. Such correlation of pR_1 and pR_3 simulations is shown in Figure 4.9. The distance of pR_1 and pR_3 in the figure was presented in the same range for comparison. The data points that are beyond 9 Å in pR_3 simulations were not shown here. The condensed data points on the top left corner in both pR_1 and pR_3 denote the presence of hydrogen bond between the pCA chromophore and Glu46, indicating that this hydrogen bond was intact for most of the pR_1 and pR_3 simulations.

Figure 4.9 A 2D correlation of the distance between phenolic oxygen O4' of the pCA chromophore and hydroxyl oxygen OE2 of Glu46 and the angle of OE2–H ⋯ O4' in pR_1 (left) and pR_3 (right) simulations.

Hydrogen bond between pCA chromophore and Tyr42

The kinetic traces of the distance between the phenolic oxygen pCA chromophore and the phenolic oxygen of Tyr42 in pR simulations are illustrated in Figure 4.10. For pR_1, the distance was in the range of hydrogen bond length in

Distance of pCA – Y42

Figure 4.10 The kinetic traces of the distance between phenolic oxygen of the pCA chromophore and phenolic oxygen of Tyr42 in eight 5 ns MD simulations of PYP in pR_1, pR_2, and pR_3.

all eight simulations. For pR_3, the distance was in the range of hydrogen bond length in seven simulations but showed relatively larger fluctuation in comparison with that for pR_1. In the same simulation that the hydrogen bond between the pCA chromophore and Glu46 was broken, the distance between the pCA chromophore and Tyr42 was 4 Å at the first frame. It fell to the range of hydrogen bond length in 6 ps for 576 ps and then rapidly increased to 9 Å or larger. It never came back to the hydrogen bond length. This was indeed due to the bad starting structure that led to open up of the protein core.

The histograms of the distance between pCA chromophore and Tyr42 in pG, pR_1, pR_2, pR_3, pB'_pR_1, and pB'_pR_3 simulations is shown in Figure 4.11.
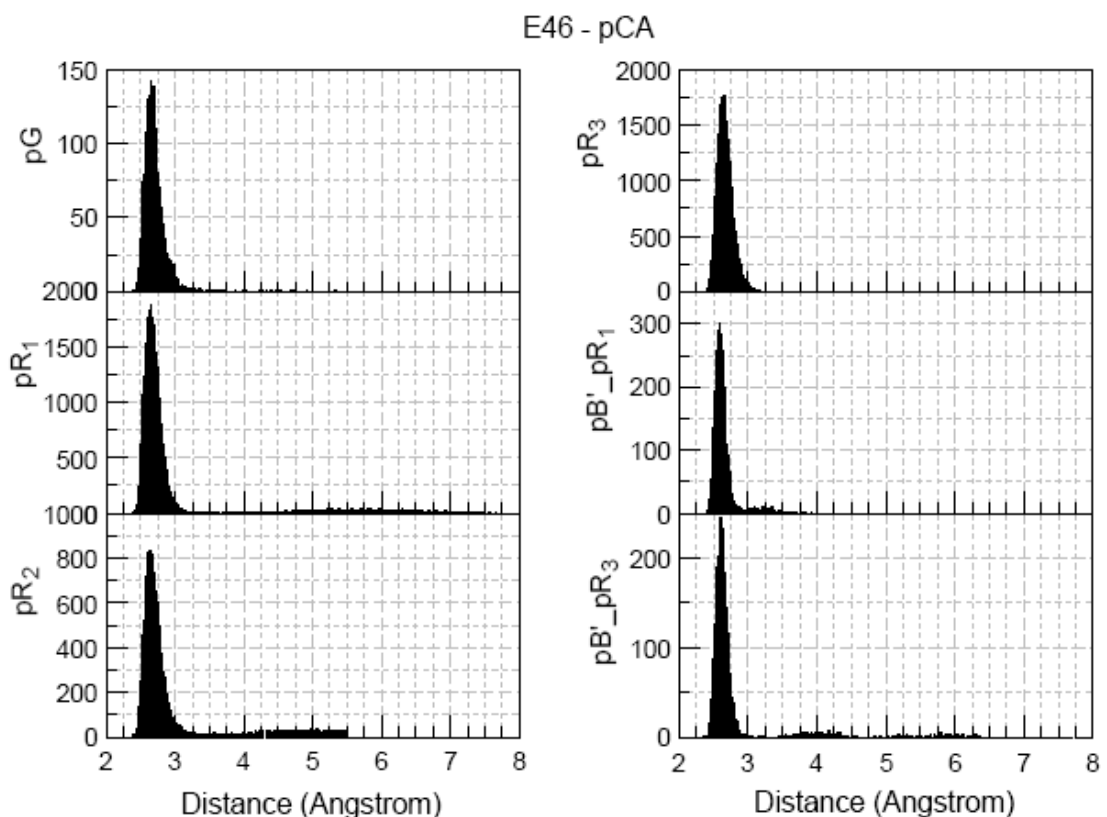

Figure 4.11 The histogram of the hydrogen bond length between phenolic oxygen of the pCA chromophore and hydroxyl oxygen of Tyr42 taken from the MD simulations of PYP in pG (5 ns), pR_1, pR_2, and pR_3 (40 ns), and pB' simulations continued from pR_1 and from pR_3 (5 ns).

Statistically the distance between phenolic oxygen of the chromophore and phenolic oxygen of Tyr42 was in the range of hydrogen bond length for 99.6% of pG simulations; 99.0% of pR_1 simulations and 85.4% of pR_3 simulations along 40 ns; 94.4% of pB'_pR_1 and 74.4% of pB'_pR_3 simulations of 5 ns.

Similarly, a 2D correlation of the distance and angles between the pCA chromophore and Tyr42 is shown in Figure 4.12. It clearly indicates that the hydrogen bond stayed intact for most of pR_1 and pR_3 simulations. Furthermore, the percentage of the hydrogen bond formed in pR_1 simulations is larger than that in pR_3, indicating the hydrogen bond is more stable in pR_1 than in pR_3 simulations.
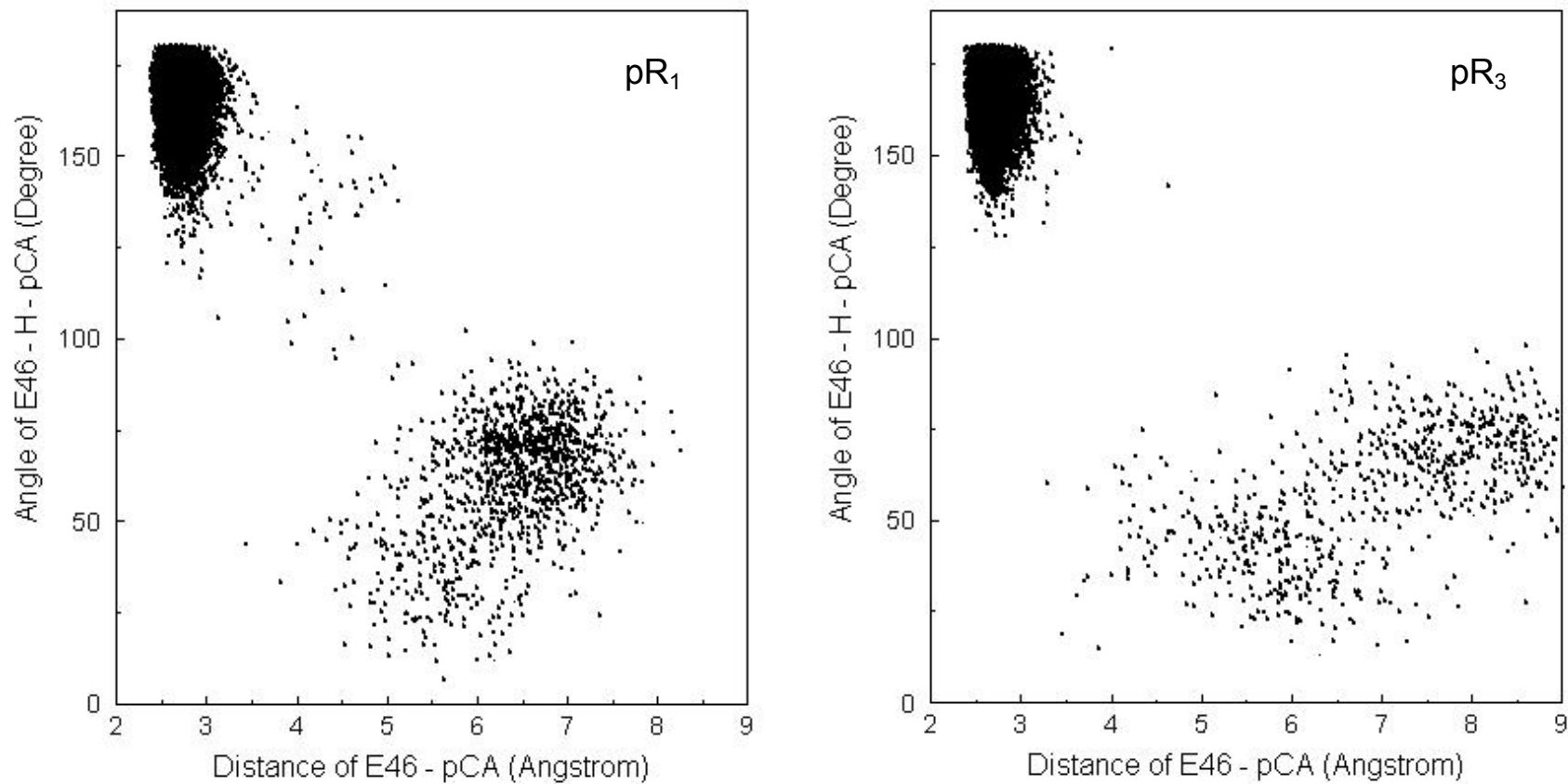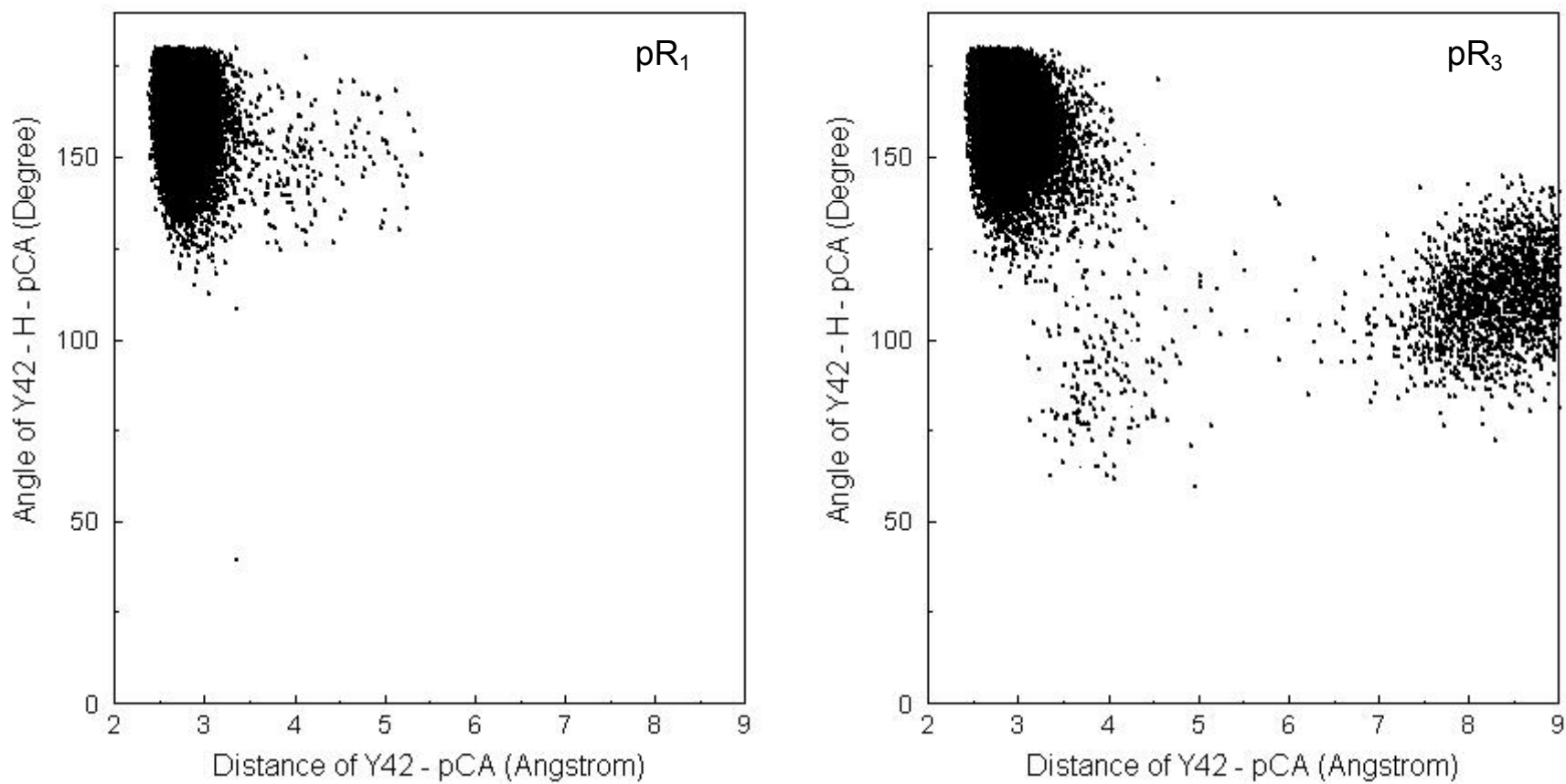
Figure 4.12 A 2D correlation of the distance between phenolic oxygen O4' of the pCA chromophore and phenolic oxygen of Tyr42 and the angle of O–H $\cdots$ O4' in pR_1 (left) and pR_3 (right) simulations.

Hydrogen bond between pCA chromophore and Cys69

In pG, this hydrogen bond is present to stabilize the pCA chromophore (Borgstahl et al., 1995). It is not clear whether it is formed at early intermediate
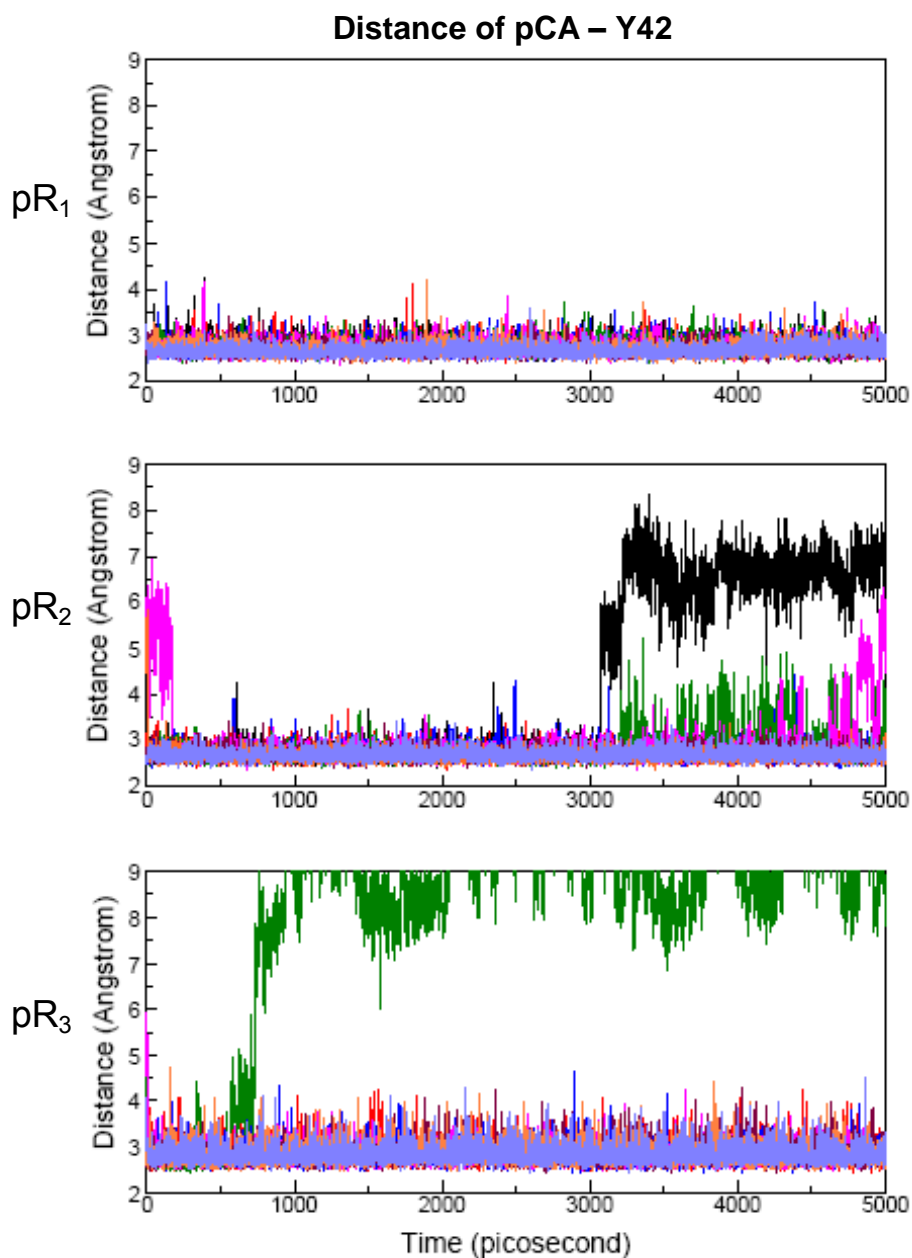
Distance of pCA – C69
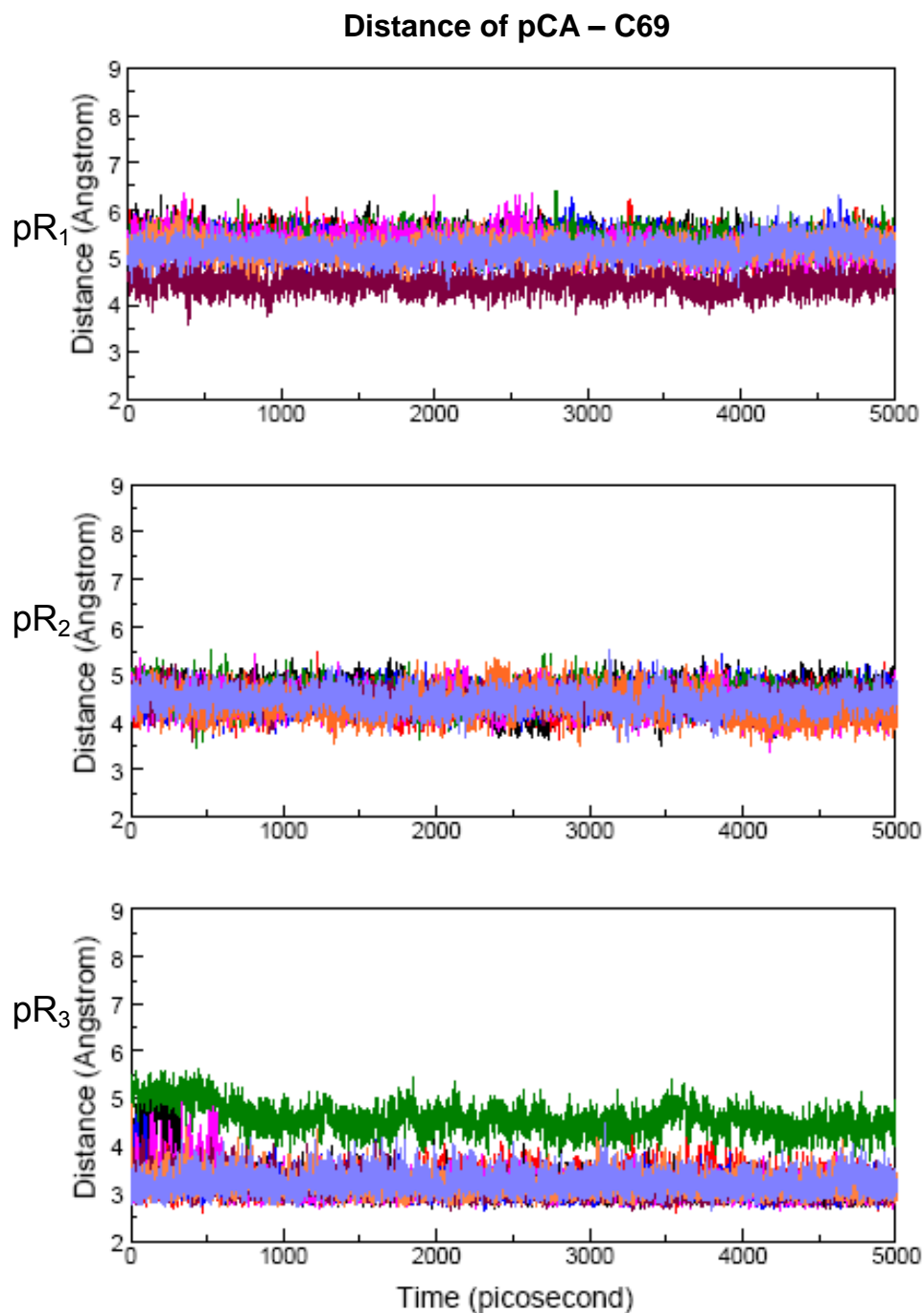
Figure 4.13 The kinetic traces of the distance between carbonyl oxygen of the pCA chromophore and backbone nitrogen of Tyr42 in eight 5 ns MD simulations of PYP in pR_1, pR_2, and pR_3.

state. The kinetic traces of the distance between the carbonyl oxygen pCA chromophore and backbone nitrogen of Cys69 in pR simulations are illustrated in Figure 4.13. For pR_1, the distance was clearly out of the range of hydrogen bond length in all eight simulations. For pR_3, the distance was fluctuating in and out of the range of hydrogen bond length in seven simulations. In the same simulation that has bad starting structure, the distance between the pCA chromophore and C69 was also out of the range of hydrogen bond length.

The histograms of the distance between pCA chromophore and C69 in pG, pR_1, pR_2, pR_3, pB'_pR_1, and pB'_pR_3 simulations is shown in Figure 4.14.
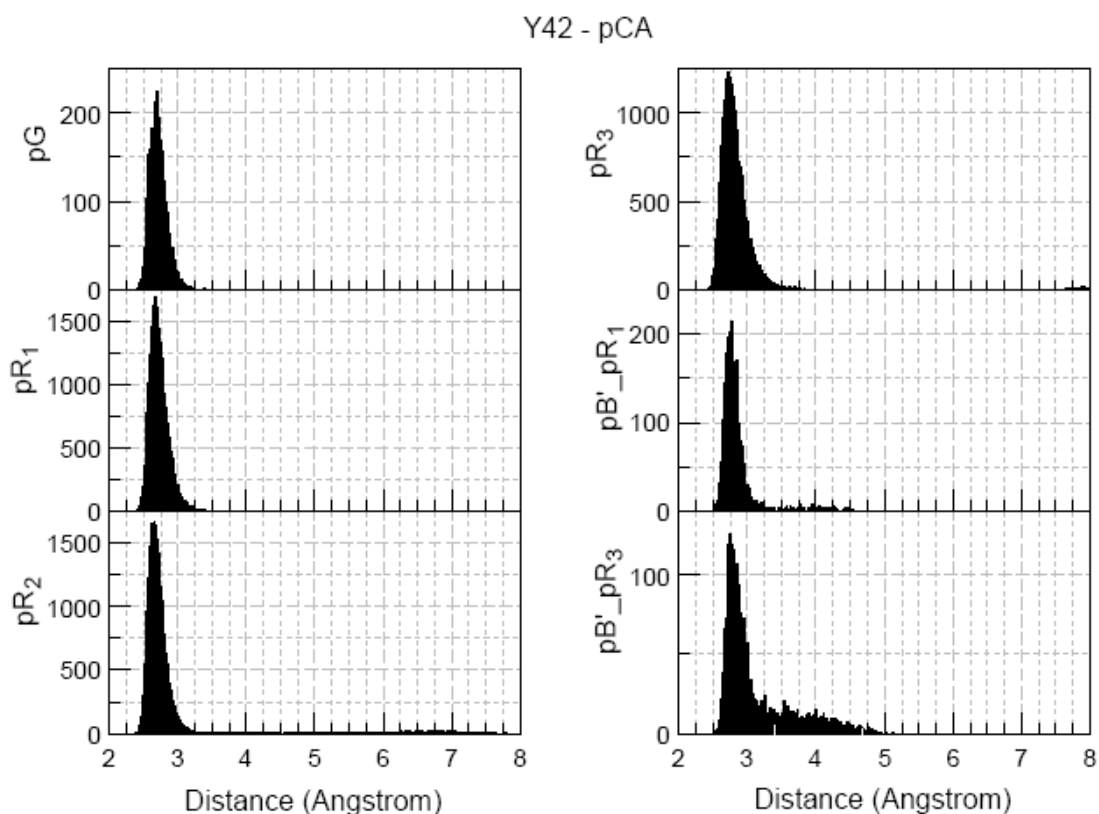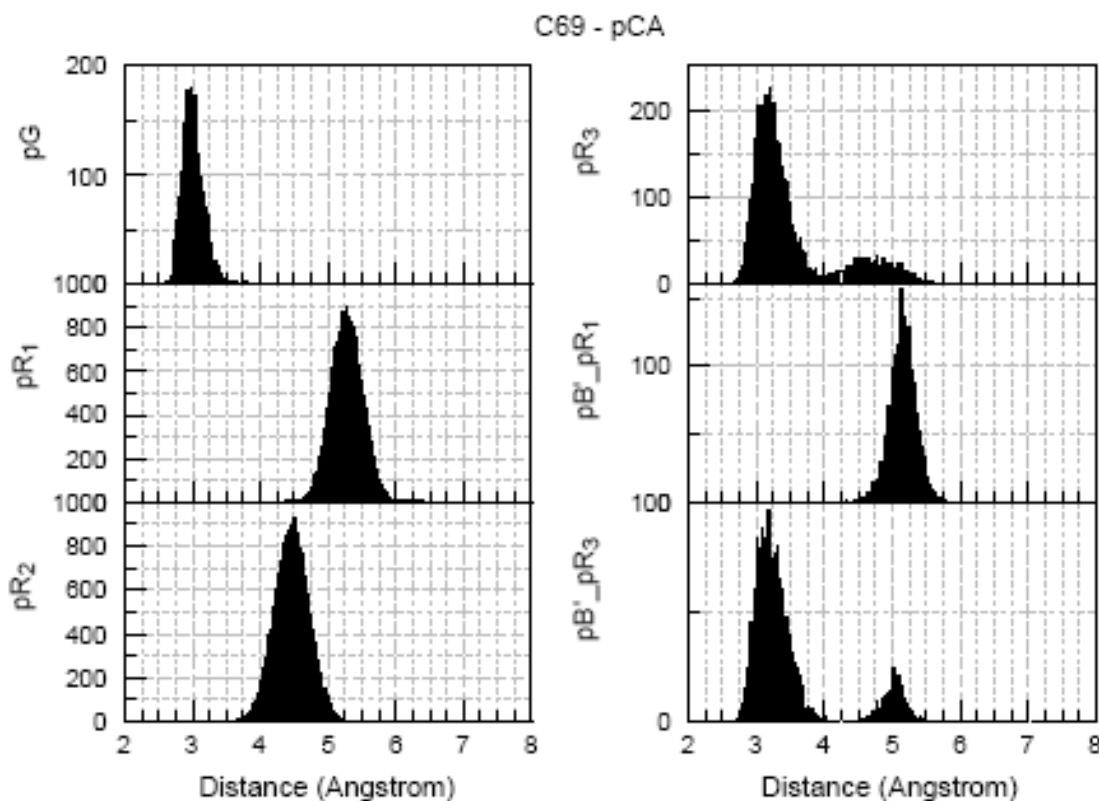

Figure 4.14 The histogram of the hydrogen bond length between carbonyl oxygen of the pCA chromophore and backbone nitrogen of Cys69 taken from the MD simulations of PYP in pG (5 ns), pR_1, pR_2, and pR_3 (40 ns), and pB' simulations continued from pR_1 and from pR_3 (5 ns).

Statistically the distance between carbonyl oxygen of the chromophore and backbone nitrogen of Cys69 was in the range of hydrogen bond length for in 89.1% of pG simulations; 0% of pR_1 simulations and 53.6% of pR_3 simulations along 40 ns; 0% of pB'_pR_1 and 45% of pB'_pR_3 simulations of 5 ns. The distance in about half of the pR_3 simulations was in the range of hydrogen bond length because the sulfur atom flipped in addition to co-isomerization of the central C2=C3 double bond and C1-S single bond (See Figure 4.5 for atom labels and Figure 4.2 F and H for structures) so that the carbonyl oxygen of the pCA chromophore was closer to the backbone nitrogen again.

Similarly, a 2D correlation of the distance and angles between the pCA chromophore and Cys69 is shown in Figure 4.15. It is fairly clear that the hydrogen bond was broken in pR_1 simulations. In addition, it indicates that even though the distance between the pCA chromophore and Cys69 is in the range of hydrogen bond length in half of pR_3 simulations, most of the angle was not linear enough to form hydrogen bond or it may form weak hydrogen bond. Indeed, only a small percent of hydrogen bond was formed in 40 ns pR_3 simulations. A Fortran90 code will be written to get exact number of frames in which the hydrogen bond is intact.

Overall, the hydrogen bond network in pR_1 and pR_3 is similar with some difference in the hydrogen bond between the pCA chromophore and Cys69. It should be emphasized that one can not rely only on the distance between two heavy atoms to draw the conclusion that there is formation of a hydrogen bond.
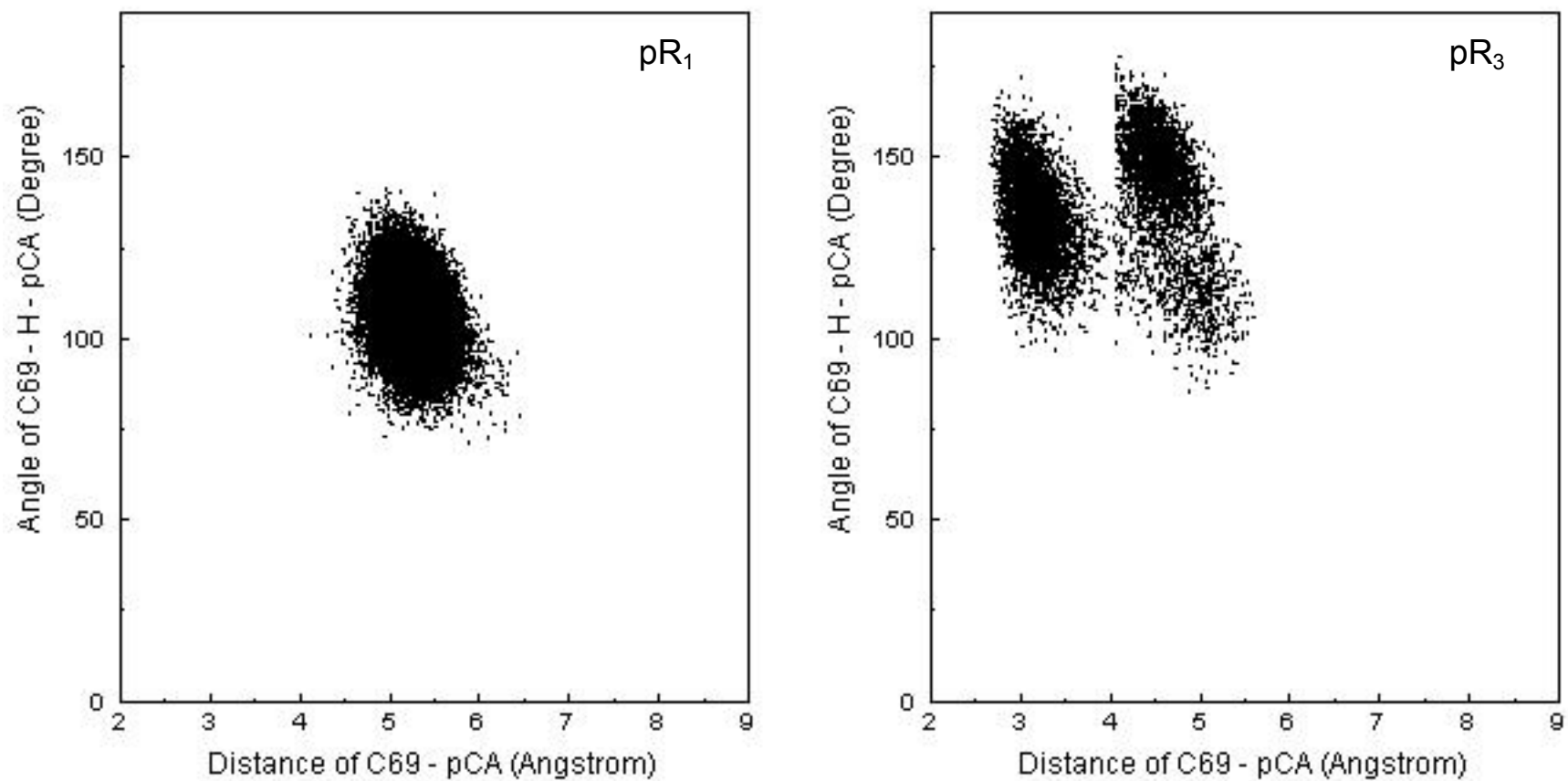
Figure 4.15 A 2D correlation of the distance between carbonyl oxygen O1 of the pCA chromophore and backbone nitrogen of Cys69 and the angle of N–H $\cdots$ O1 in pR_1 (left) and pR_3 (right) simulations.

In addition, we observed that there was a hydrogen bond formed between the hydroxyl oxygen of Tyr94 and the sulfur of the pCA chromophore for most of pR_1 simulations and pB' simulations continued from pR_1 meanwhile it was the backbone nitrogen of Asp97 that formed a hydrogen bond with the sulfur for most of pR_3 simulations and pB' simulations continued from pR_3, as shown in Figure 4.16 and 4.17.
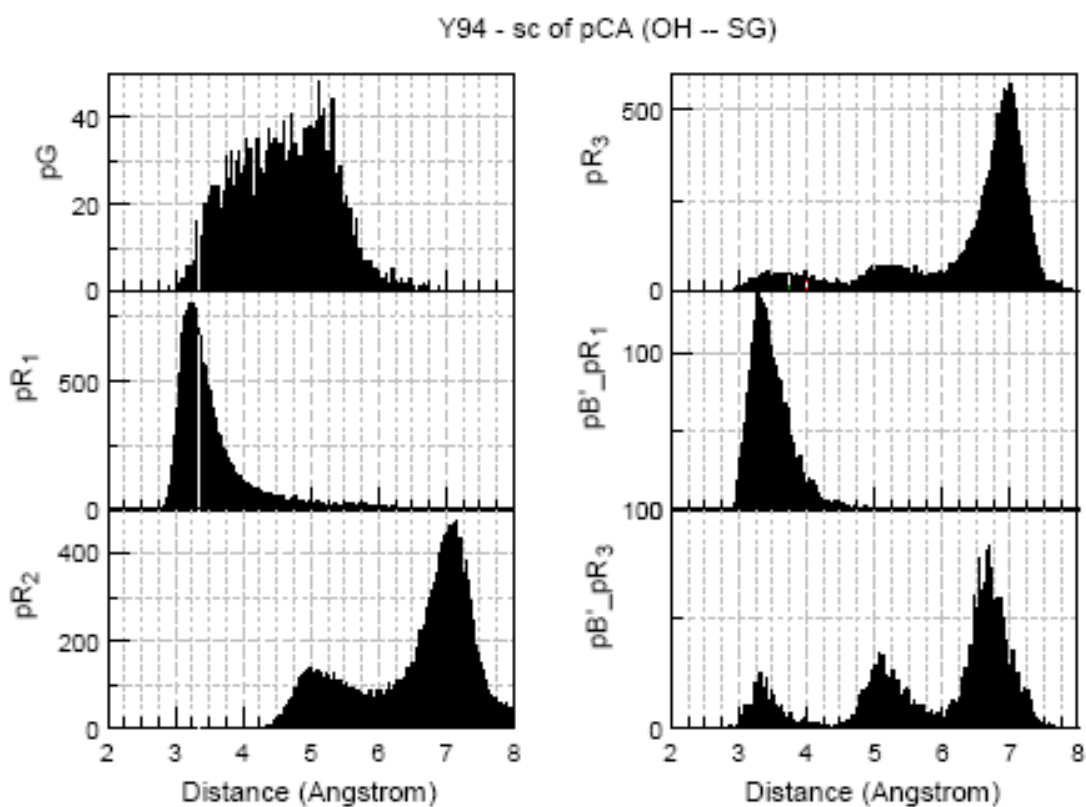
Figure 4.16 The histogram of the distance between the sulfur of the pCA chromophore and phenolic oxygen of Tyr94 taken from the MD simulations of PYP in pG (5 ns), pR_1, pR_2, and pR3 (40 ns), and pB' simulations continued from pR_1 and from pR_3 (5 ns).
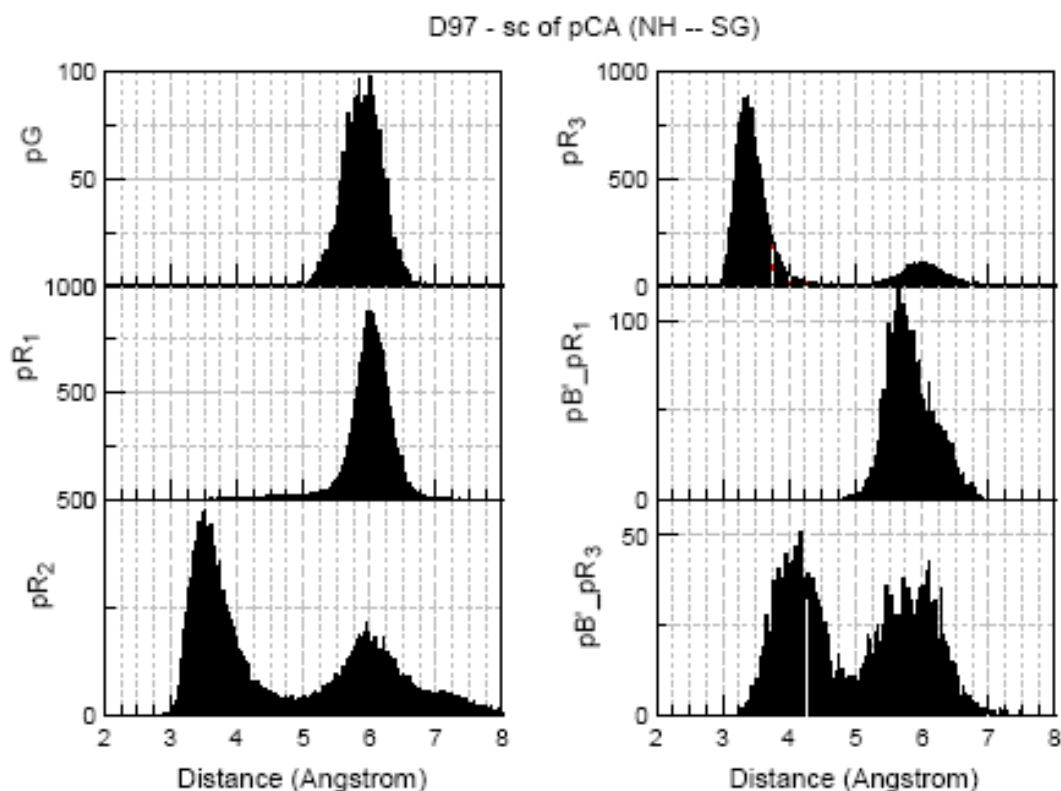
Figure 4.17 The histogram of the distance between the sulfur of the pCA chromophore and backbone nitrogen of Asp97 taken from the MD simulations of PYP in pG (5 ns), pR$_1$, pR$_2$, and pR3 (40 ns), and pB' simulations continued from pR$_1$ and from pR$_3$ (5 ns).

Mimic the chromophore isomerization

Starting from *trans* pCA chromophore, we modeled 12 chromophore structures in which C1=O1 and C2–H2 (see Figure 4.5 for atom labels) were in opposite direction but in the same plane that is in an angle from the chromophore ring plane, 30° a structure, from −180° to 180°. Then we have performed 10 ps of MD simulations for each structure. The kinetic traces of characteristic dihedral angles of the 12 simulations are shown in Figure 4.18. When the angle of the plane was 0°, -30°, -60°, -90°, that is, counterclockwise if we look from backbone to the direction of the ring, the chromophore relaxed to *trans* configuration.
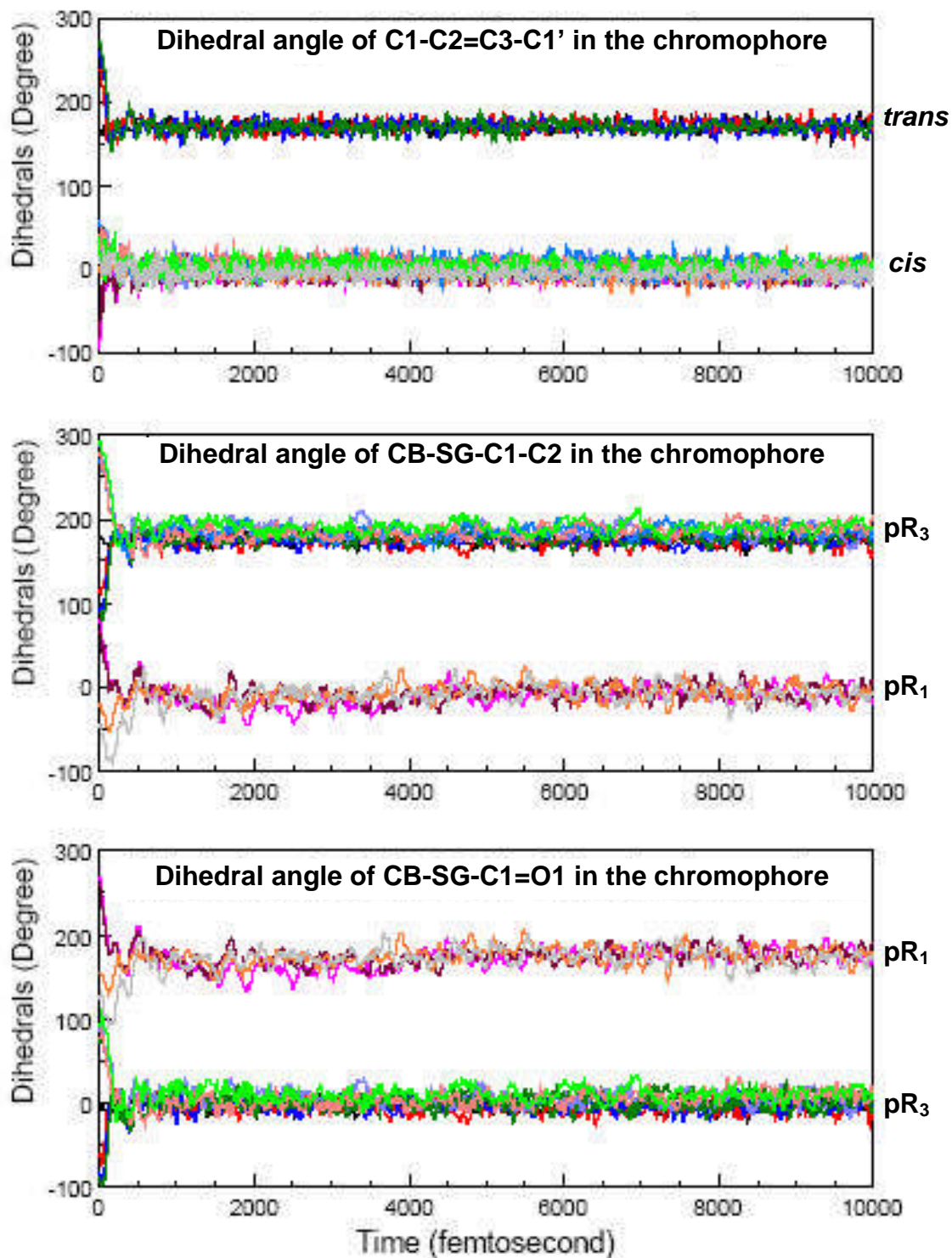
Figure 4.18 The kinetics traces of dihedral angles of the chromophore taken from 10 ps MD simulations of isomerization. The C1=O1 and C2–H2 (see Figure 4.5 for atom labels) were made so that they were in the same plane and in an angle to the chromophore plane during relaxation: -180°(orange), -150°(brown), -120°(magenta), -90°(green), -60°(blue), -30°(red), 0°(black), +30°(light purple), +60°(royal blue), +90° (light coral), +120°(light green), and +150° (gray).

When the angle of the plane was -120°, -150°, -180°, +150°, counterclockwise if we look from backbone to the direction of the ring, the chromophore relaxed to *cis* configuration of pR$_1$. When the angle of the plane was +30°, +60°, +90°, +120°, clockwise if we look from backbone to the direction of the ring, the chromophore relaxed to *cis* configuration of pR$_3$. Therefore, there are indeed two possible *cis* configuration dependent on in which direction the isomerization goes, namely, pR$_1$ and pR$_3$.

Proton affinities calculations

We have performed *ab initio* calculations of relative proton affinities of *trans* and *cis* chromophore model compound with the presence of 1, 2 and 3 hydrogen bonds before and after proton transfer in vacuum and cyclohexane with dielectric constant of ε=2.023 (Frisch, Æ. et al., 2003a). The structure was first optimized using PM3 method and then the energy was calculated using B3LYP/6-311+G(2d,p) method (Frisch, M. J. et al., 2003b). The optimized structures are shown in Figure 4.18 and the results are shown in Table 4.3.

In vacuum, the energy of *trans* pCA - Glu46 with proton on Glu46 was 21.8 KJ/mol lower than that of with proton on the pCA, i.e., 3.8 pKa shift (ΔpKa). In fact, all the calculations in Table 4.3 showed that the proton on Glu46 was energetically more stable. However, the energy of the *trans* pCA - Glu46 with proton on Glu46 was only 4.7 kJ/mol (ΔpKa=0.8) lower than that of with proton on *trans* pCA in dielectric medium of 2.023. The energy of *trans* pCA - Glu46 with the proton on Glu46 became higher in ether with dielectric constant of 4.335
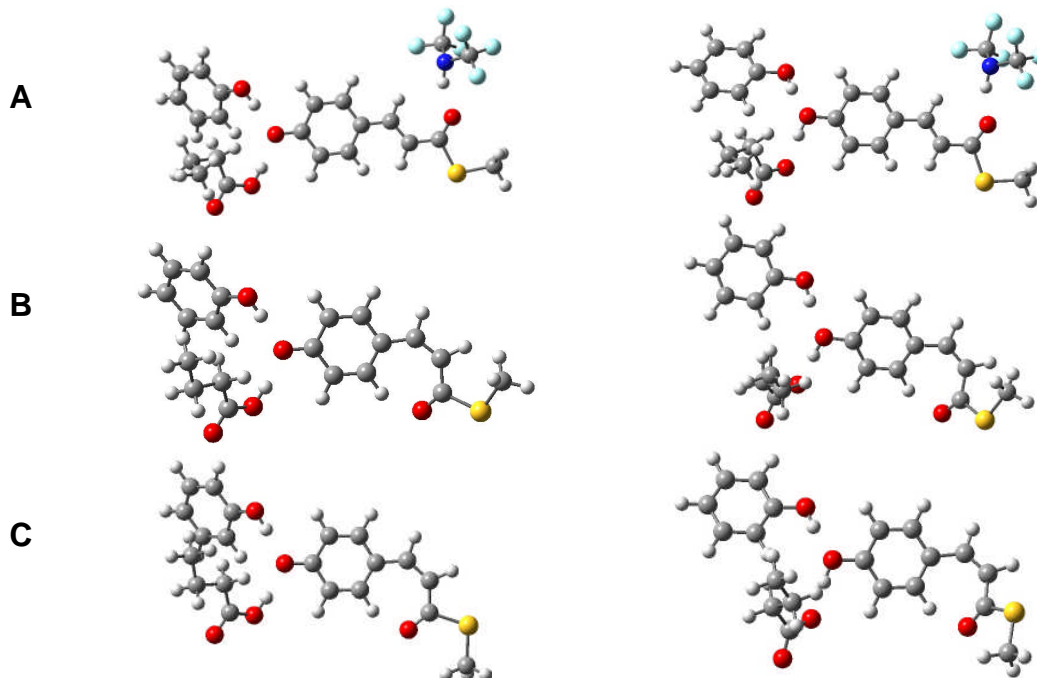
Figure 4.19 The optimized structures of *trans*-pCA in three hydrogen bonds with Glu46, Tyr42, Cys69 model compounds (A), cis-pCA$_{pR1}$ (B) and cis-pCA$_{pR3}$ (C) in two hydrogen bonds with Glu46 and Tyr42 model compounds using PM3 method. The proton in the left panel is on the Glu46 and in the right panel is on the pCA. For calculation of one hydrogen-bonding interaction, only Glu46 and the pCA were present. For calculation of two hydrogen-bonding interactions, Glu46 and Tyr42 or Glu46 and Cys69 as hydrogen bond partners together with the pCA were present.

(Table 4.4), indicating that the proton favored staying on the pCA in dielectric medium of 4.335 or higher. This implies that changes of protein environment at the active site have large effects on relative proton affinities.

When both Tyr42 and Glu46 formed hydrogen bonds with the pCA, the energy difference (ΔE) between the proton on Glu46 and proton on pCA increased to 57.8 kJ/mol in vacuum. The hydrogen bond of pCA - Tyr42 alone contributed 36 kJ/mol to relative proton affinities. This is 6.3 ΔpKa, comparable to the pKa shift of pCA chromophore 7.2 between in free solution and in protein. Similarly in dielectric medium of 2.023, the *trans* pCA - Tyr42 contributed 34.8 kJ/mol (ΔpKa=6.1) to the relative proton affinities. When both Cys69 and Glu46

were present as hydrogen bond partners, the ΔE was 27.9 kJ/mol in vacuum, 6

kJ/mol of which was attributed to the hydrogen bond of *trans* pCA - Cys69,

indicating the hydrogen bond between pCA and Cys69 is very weak. It was so

weak that it did not make any contribution in the case of pCA in three hydrogen

bonds. This was also the case in dielectric medium of 2.023.

Table 4.3. Effects of hydrogen bonding interactions on relative proton affinity

Model*	H-bond partners	No. of H-bond	Vacuum (ε=1)		Cyclohexane (ε=2)	
			$\Delta E^{\dagger}$ (kJ/mol)	$\Delta pKa^{\#}$	$\Delta E^{\dagger}$ (kJ/mol)	$\Delta pKa^{\#}$
trans-pCA	E46	1	21.8	3.8	4.7	0.8
trans-pCA	Y42, E46	2	57.8	10.1	39.5	6.9
trans-pCA	C69, E46	2	27.9	4.9	4.0	0.7
trans-pCA	C69, Y42, E46	3	56.0	9.8	39.2	6.9
cis-pCA$_{pR1}$	E46	1	36.9	6.5	18.4	3.2
cis-pCA$_{pR1}$	Y42, E46	2	71.1	12.5	51.2	9.0
cis-pCA$_{pR3}$	Y42, E46	2	76.3	13.4	56.4	9.9

*The structure was optimized using PM3 method in vacuum and energy was calculated

using B3LYP/6-31G(d) method.

$^{\dagger} \Delta E = E(pCA^{(0)}, E46^{(-)}) - E(pCA^{(-)}, E46^{(0)})$ in atomic unit.

1 Hartree = 2619.6 kJ/mol, 1eV=97 kJ/mol

$^{\#}$The ΔpK_a was calculated using the following equation.

$\Delta G \approx \Delta E = -RT \ln K_a / K_a(pH7) = 2.303RT \times \Delta pK_a = 5.7 kJ/mol \times \Delta pK_a$

As for *cis* pCA - Glu46 in vacuum, the ΔE was 36.9 kJ/mol, 15.1 kJ/mol higher than that of *trans* pCA - Glu46. This was due to the structural difference between *cis* and *trans* that changed the relative proton affinities. Again, the hydrogen bond of *cis* pCA - Tyr42 contributed 34.2 kJ/mol to the relative proton affinities in the configuration of pR_1 and 39.4 kJ/mol in the configuration of pR_3. It seems that Glu46 and Tyr42 made comparable contributions to proton affinities in vacuum. This is not the case in dielectric medium. It is unambiguous that in dielectric medium of 2.023, the hydrogen bond of *cis* pCA - Glu46 contributed 18.4 kJ/mol to the proton affinities, about half of 34.8 kJ/mol that was contributed by hydrogen bond of *cis* pCA - Tyr42 in pR_1 and 38.0 kJ/mol in pR_3 configuration. It indicates that Tyr42 may be the key residue that regulates proton affinities and therein triggers the intra-protein proton transfer during protein photoactivation based on the fact that the hydrogen bonding interaction between Tyr42 and pCA contributes most to the relative proton affinities.

We have also calculated effect of structural geometry on relative proton affinities of *trans* pCA with Glu46 before and after proton transfer in vacuum, cyclohexane (ε=2.023), and in ether (ε=4.335). The structure with the dihedral angle of C3=C2–C1=O1 was fixed to be −11.471°, the same as in crystal structure of PYP in pG at 0.82 Å resolution (Getzoff et al., 2003), was first optimized using PM3 method and then the energy was calculated using B3LYP/6-311+G(2d,p) method. The results are summarized in Table 4.4. When the dihedral angle was fixed, the hydrogen bond of *trans* pCA – Glu46 contributed 7.6 kJ/mol (ΔpKa=1.3) in vacuum to relative proton affinities, less

than half of 17.6 kJ/mol when the dihedral angle was free to move. The ΔE became negative, -8.4 kJ/mol ($\Delta pKa=-1.5$) in dielectric medium of 2.023 and more negative, -19.9 kJ/mol ($\Delta pKa=-3.5$) in dielectric medium of 4.335. This implies that the chromophore geometry has also large effects on the relative proton affinities.

Table 4.4. Effects of chromophore geometry on proton affinities

Model*	H-bond partners	No. of H-bond	Vacuum ($\varepsilon=1$)		CycloHexane ($\varepsilon=2.023$)		Ether ($\varepsilon=4.335$)	
			$\Delta E^{\dagger}$ (kJ/mol)	$\Delta pK_a^{\#}$	$\Delta E^{\dagger}$ (kJ/mol)	$\Delta pK_a^{\#}$	$\Delta E^{\dagger}$ (kJ/mol)	$\Delta pK_a^{\#}$
trans-pCA	E46	1	17.6	3.1	6.0	1.0	-2.5	-0.4
trans-pCA	Y42, E46	2	45.1	7.9	26.2	4.6	12.9	2.3
‡trans-pCA	E46	1	7.6	1.3	-8.4	-1.5	-19.9	-3.5

*The structure was optimized using PM3 method in vacuum and energy was calculated using B3LYP/6-311+G(2d,p) method.

$^{\dagger}\Delta E = E(pCA^{(0)}, E46^{(-)}) - E(pCA^{(-)}, E46^{(0)})$ in atomic unit.

1 Hartree = 2619.6 kJ/mol, 1eV=97 kJ/mol

$^{\#}$The ΔpK_a was calculated using the following equation.

$\Delta G \approx \Delta E = -RT \ln K_a / K_a (pH7) = 2.303RT \times \Delta pK_a = 5.7kJ / mol \times \Delta pK_a$

‡The dihedral angle of C3=C2-C1=O1 (See Figure 4.5 for atom labels) is fixed to be $-11.471°$, obtained from 1NWZ.pdb (Getzoff et al., 2003).

Proton transfer barrier

We had been trying to get proton transfer barrier using *ab initio* method for a long time. The proton in the optimized structure was shared between the

phenolic oxygen of the pCA and the hydroxyl oxygen of Glu46 so that there was

no proton transfer barrier that was found. Recently we obtained proton transfer

barrier by fixing the distance between the two oxygen atoms and fixing the

distance between the proton and the phenolic oxygen of the pCA stepwise,

optimizing structure at each step using PM3 method, and then calculating energy

using HF/3-21G method. The results are shown in Figure 4.20.
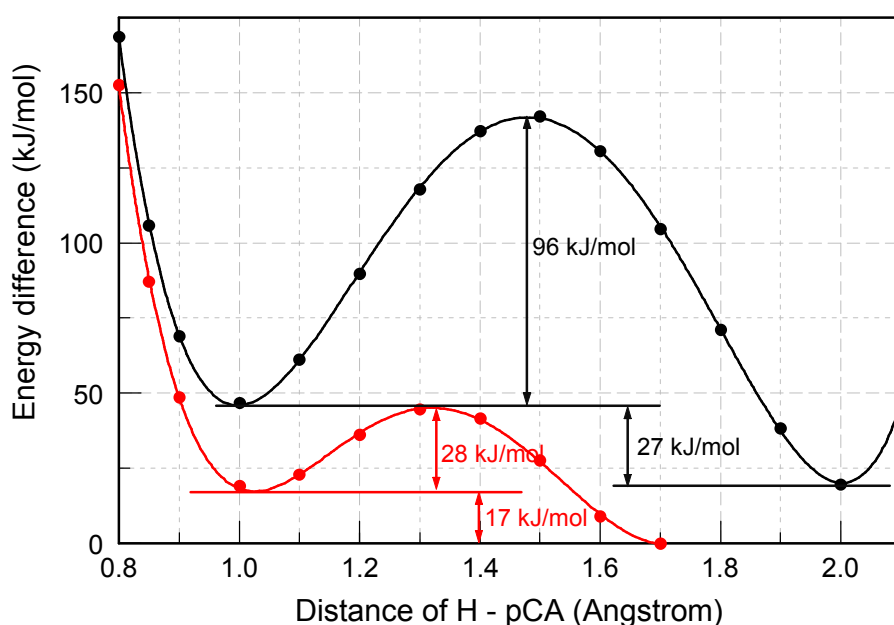
Figure 4.20 Calculated energy barrier for proton transfer between Glu46 and the pCA chromophore in *trans* form. The distance between the phenolic oxygen of the pCA and hydroxyl oxygen of Glu46 is fixed at 2.7 Å (red) and 3.0 Å (black). In addition, the distance between the proton and phenolic oxygen of the pCA chromophore is fixed from 0.8 to 1.7 Å (2.1 Å), 0.1 Å a step. The calculated data points are presented in solid dots and the curve is least squares fitting. The geometry was optimized in each step using PM3 and the energy was computed using HF/3-21G method. For constrained geometry with short distance, we had lowered the convergence criteria to get optimization.

It is apparent from the figure that the proton on Glu46 is energetically

more stable in both curves. When the hydrogen bond length is 3.0 Å (black

curve), the energy barrier for proton transfer from the pCA to Glu46 is 96 kJ/mol.

This barrier is 28 kJ/mol, much smaller when the hydrogen bond length is 2.7 Å (red curve), indicating that the shorter the hydrogen bond length between the pCA and Glu46, the easier for proton transfer to taken place. It is found in PYP that after chromophore photoisomerization, the hydrogen bond between the pCA and Glu46 became stronger and thereby the hydrogen bond length was shorter (Anderson et al., 2004a; Xie et al., 1996). Based on our computational results, it is unambiguous that the proton transfer barrier is much lower for stronger hydrogen bond, which promotes intra-protein proton transfer.

4.4 Conclusion

We have analyzed the two crystal structures of PYP at early intermediate state and found that the chromophore structures were highly deformed. That is the reason that we performed MD simulations in explicit solvent of three possible photoproducts of PYP to identify plausible structure(s) that was supported by experimental evidence. We found that the hydrogen bond network in pR_1 and pR_3 is similar with some difference in the hydrogen bond between the pCA chromophore and Cys69 whereas the results of pR_2 simulations were against the experimental evidence. We have mimicked the chromophore isomerization by twisting the C1=O1 and C2–H2 in 12 positions and let each structure relax for 10 ps MD simulations. We found that 4 structures went back to *trans* form, 4 structure relaxed to *cis* form of pR_1, and 4 structure relaxed to *cis* form of pR_3. Therefore, there are two plausible photoproducts that will depend on the direction of chromophore isomerization. Our simulation data of three plausible

photoproducts provide a specific guide on how to use time-resolved infrared spectroscopy to identify which photoproduct is formed after photoisomerization.

In addition, we studied the effects of hydrogen bonding interactions, dielectric medium, and chromophore geometry on the relative proton affinity that is an important indicator for proton transfer. We found that the hydrogen bond of pCA – Tyr42 made major contributions to the relative proton affinities that is crucial to trigger the intra-protein proton transfer during PYP photoreceptor activation. Both the dielectric medium and chromophore geometry have also large effects on the relative proton affinities. Our computational results of proton transfer barrier have showed that the stronger hydrogen bond had much smaller proton transfer barrier that may promote proton transfer process from the hydrogen bond donor to the hydrogen bond acceptor.

CHAPTER V

CONCULSION AND REMARKS

5.1 Conclusion

Vibrational spectral markers

Time-resolved infrared difference spectroscopy is a powerful technique to probe structural changes in functional processes of proteins and interactions of individual groups of the proteins with local environment. Some of the vibrational frequencies of the specific groups may shift upon change of local environment. The changes of local environment include water penetration, changes of hydrogen−bonding interactions including formation or breaking of hydrogen bonds and changes of hydrogen-bond strength, and changes of hydrophobic interactions. At the active sites of proteins where proton transfer occurs, the hydrogen−bonding interactions are predominant. It is therefore necessary to establish vibrational spectral markers to probing hydrogen-bonding status of key residues at the active site of proteins.

PYP was chosen as a model system to study proton transfer in proteins. It binds a unique pCA chromophore that is stabilized by the side−chain groups of Glu46 and Tyr42 and the backbone of Cys69 via hydrogen−bonding interactions. Due to the crucial roles of Glu46 and Tyr42 in proton transfer in PYP, we have performed extensive computational studies based on density functional theory on

the vibrational frequencies of protonated Asp/Glu and Tyr side–chain groups interacting with polar neutral side–chain groups (Ser, Thr, Cys, Met, Asn/Gln, Tyr, Asp/Gln, His, Lys, and Arg), charged side–chain groups (Arg+, Lys+, His+, COO– of Asp/Glu, and negatively charged Tyr–), water molecules, and backbone. Supported by experimental evidence, we have successfully established vibrational spectral markers for detecting hydrogen-bonding status of protonated Asp/Glu (Chapter II) and Tyr (Chapter III).

(1) The C=O stretching frequency of buried Asp/Glu is an excellent vibrational spectral marker for probing hydrogen-bond number: ~1759–1776 cm^{-1} for zero, ~1733–1749 cm^{-1} for one, ~1703–1710 cm^{-1} for two hydrogen–bonding interactions with polar neutral side–chain groups. We proposed an approximate linear correlation between the C=O stretching frequency and the net hydrogen-bond strength that applies to hydrogen–bonding interactions with polar neutral and positively side–chain charged groups. In addition, we proposed a two-dimensional infrared spectroscopy, i.e., C=O stretching and O–H stretching frequencies, may be helpful to identify the specific type of hydrogen–bonding interactions.

(2) The C–O stretching frequency of Tyr side–chain (phenol) group is also an excellent vibrational spectral marker for probing hydrogen-bonding status: 1255 cm^{-1} for zero, 1230–1243 cm^{-1} for hydroxyl oxygen of the phenol group and 1265–1277 cm^{-1} for hydroxyl hydrogen of the phenol group forming one hydrogen bond with polar neutral groups, and 1194–1204 cm^{-1} for hydroxyl

hydrogen forming one strong hydrogen bond with positively charged side–chain groups. A two-dimensional infrared spectroscopy, C–O stretching vs. O–H stretching or C–O stretching vs. O–H bending, may be helpful to identify the type of hydrogen–bonding interaction when the marker of C–O stretching frequency is inadequate.

When protonated Asp/Glu or Tyr interact with negatively charges side–chain groups (COO– of Asp/Glu, and negatively charged Tyr–), all our calculations yield the proton movement. The proton is shared between the two oxygen atoms of two side–chain groups that were involved in hydrogen–bonding interactions. Special treatment will be required to study such interactions, that is, force the proton attached to one of the oxygen atoms by fixing their covalent bond of neutral Asp/Glu or Tyr.

In addition to establishing vibrational spectral markers, we calculated the vibrational frequency of C=O stretching of a COOH group upon common isotopic labeling of COOH $\rightarrow$ COOD, COOH $\rightarrow$ C^{18}O^{18}OH, and COOH $\rightarrow$ ^{13}COOH in experimental studies. We also calculated vibrational frequencies of C–O stretching, O–H bending, and O–H stretching upon deuteration of hydroxyl hydrogen of the phenol group and ring–H$_4$ $\rightarrow$ ring–D$_4$ isotopic labeling. The frequency shifts of the vibrational spectral markers upon these isotopic substitutions have been discussed in the thesis. It has great advantage of obtaining vibrational frequencies and quantitatively studying the frequency shift upon isotopic labeling using *ab initio* method whereas traditional isotopic labeling is chemically challenging and expensive.

Proton transfer in PYP

We have analyzed the two crystal structures of PYP at early intermediate state and found that the chromophore structures were highly deformed. That is the reason that we performed MD simulations in explicit solvent of three possible photoproducts of PYP to identify plausible structure(s) that was supported by experimental evidence. We found that the hydrogen bond network in pR_1 and pR_3 is similar with some difference in the hydrogen bond between the pCA chromophore and Cys69 whereas the results of pR_2 simulations were against the experimental evidence. We have mimicked the chromophore isomerization by twisting the C1=O1 and C2–H2 in 12 positions and let each structure relax for 10 ps MD simulations. We found that 4 structures went back to *trans* form, 4 structure relaxed to *cis* form of pR_1, and 4 structure relaxed to *cis* form of pR_3. Therefore, there are two plausible photoproducts that will depend on the direction of chromophore isomerization. Our simulation data of three plausible photoproducts provide a specific guide on how to use time-resolved infrared spectroscopy to identify which photoproduct is formed after photoisomerization.

In addition, we studied the effects of hydrogen bonding interactions, dielectric medium, and chromophore geometry on the relative proton affinity that is an important indicator for proton transfer. We found that the hydrogen bond of pCA – Tyr42 made major contributions to the relative proton affinities that is crucial to trigger the intra-protein proton transfer during PYP photoreceptor activation. Both the dielectric medium and chromophore geometry have also large effects on the relative proton affinities. Our computational results of proton

transfer barrier have showed that the stronger hydrogen bond had much smaller proton transfer barrier that may promote proton transfer process from the hydrogen bond donor to the hydrogen bond acceptor.

5.2 Future outlook

Many interesting topics remain for further investigation. First of all, there are other polar side−chain groups at the active sites other than protonated Asp/Glu and Tyr. We have calculated the hydrogen−bonding properties of Asn/Gln, neutral Arg and Arg+, neutral His and His+, negatively charged Asp/Glu (Asp/Glu−), negatively charged Tyr (Tyr−), neutral Lys and Lys+, and Ser/Thr. The results are shown in the Appendix. Some of the vibrational frequencies of these polar groups are sensitive to hydrogen−bonding interactions. With data analysis data and experimental support from the literatures, a series of vibrational spectral markers may be established. Thus these vibrational spectral markers are powerful tools to get insights into the hydrogen-bonding status and local environments of important amino acids during functionally important intermediate states of proteins.

Second, hydrogen-bonding interactions involving the pCA chromophore in PYP and retinal chromophore in rhodopsin and bacteriorhodopsin always draw attentions due to their structural and functional importance. It would be of great interests to establish vibrational spectral marker(s) to probe hydrogen-bonding status of the protein chromophore(s). However, it is more challenging because the chromophore(s) has larger size and carries charges. Our computational

studies have shown that special strategies will be required for calculations of negatively charged groups hydrogen−bonding interacting with other polar groups.

Third, proton transfer barrier was not found for the pCA chromophore hydrogen bonding interacting with Glu46, Tyr42, and Cys69. There may be other potential residues that are important for proton transfer. We would continue looking for proton transfer barrier with larger model system containing Thr50 or Arg52 or both. In addition, we would extract the active site of the protein from MD simulations of pG, pR and pB' and calculate proton transfer barrier using *ab initio* method. This is so-called quantum mechanics/molecular mechanics (QM/MM).

Four, centered on proton transfer in PYP, we have calculated the pKa values of all titratable amino acids of PYP in pG receptor state using MCCE (Multi-conformation continuum electrostatics) program. The protein coordinates are taken from protein crystal structure (PDB code: 2PHY). We are in the process of obtaining reaction field energy of the negatively charged pCA chromophore since it is not standard amino acids therefore not included in the program package. Reaction field energy stands for the energy required to move a charge from the solution to the inside of protein. Once we get the pKa values of Glu46 and pCA chromophore close enough to the experimental values, we will apply the same strategies to calculate the selected frames of molecular dynamics simulations of PYP in pR and pB' in explicit solvents. The pKa calculations of key residues at the active site of the protein are expected to elucidate what crucial structural elements regulate relative proton affinities of proton donor and proton acceptor and therefore drive proton transfer in the functional intermediate states

of PYP. This ongoing project is primarily done by Dr. Robert Hauenstein, associate professor at the Department of Physics in Oklahoma State University and supported by Dr. Marilyn Gunner's research group at the Department of Physics in the City University of New York.

Last but not the least, we have performed molecular dynamics Go-model simulations of PYP that did not include the protons and water molecules. We will investigate these simulations to explore (1) protein structural conformations at microsecond to millisecond timescale, (2) protein folding, (3) protein stability, and (4) interactions at the active site of the protein. This ongoing project is in collaboration with Dr. Benjamin H. McMahon, staff scientist at the Theoretical Biophysics Division in Los Alamos National Laboratory.

REFERENCE

Acharya, S., and S. S. Karnik. 1996. Modulation of GDP release from transducin by the conserved Glu134-Arg135 sequence in rhodopsin. *J. Biol. Chem.* 271:25406-25411.

Adelroth, P., and P. Brzezinski. 2004. Surface-mediated proton-transfer reactions in membrane-bound proteins. *Biochim. Biophys. Acta.* 1655:102-115.

Alder, B. J., and T. E. Wainwright. 1957. Phase Transition for a Hard Sphere System. *Journal of Chemical Physics* 27:1208-1209.

Alder, B. J., and T. E. Wainwright. 1959. Studies in Molecular Dynamics .1. General Method. *Journal of Chemical Physics* 31:459-466.

Ames, J. B., M. Ros, J. Raap, J. Lugtenburg, and R. A. Mathies. 1992. Time-resolved ultraviolet resonance Raman studies of protein structure: application to bacteriorhodopsin. *Biochemistry* 31:5328-5334.

Anderson, S., S. Crosson, and K. Moffat. 2004a. Short hydrogen bonds in photoactive yellow protein. *Acta Crystallogr D Biol Crystallogr.* 60:1008-1016.

Anderson, S., V. Srajer, and K. Moffat. 2004b. Structural heterogeneity of cryotrapped intermediates in the bacterial blue light photoreceptor, photoactive yellow protein. *Photochem. Photobiol.* 80:7-14.

Antes, I., W. Thiel, and W. F. van Gunsteren. 2002. Molecular dynamics simulations of photoactive yellow protein (PYP) in three states of its photocycle: a comparison with X-ray and NMR data and analysis of the effects of Glu46 deprotonation and mutation. *Eur. Biophys. J.* 31:504-520.

Arnis, S., K. Fahmy, K. P. Hofmann, and T. P. Sakmar. 1994. A conserved carboxylic acid group mediates light-dependent proton uptake and signaling by rhodopsin. *J. Biol. Chem.* 269:23879-23881.

Baca, M., G. E. Borgstahl, M. Boissinot, P. M. Burke, D. R. Williams, K. A. Slater, and E. D. Getzoff. 1994. Complete chemical structure of photoactive yellow protein: novel thioester-linked 4-hydroxycinnamyl chromophore and photocycle chemistry. *Biochemistry* 33:14369-14377.

Bao, Q. 2004. The photocycle of photoactive yellow protein [Master thesis]. Oklahoma State University.

Barber-Armstrong, W., T. Donaldson, H. Wijesooriya, R. A. Silva, and S. M. Decatur. 2004. Empirical relationships between isotope-edited IR spectra and helix geometry in model peptides. *J. Am. Chem. Soc.* 126:2339-2345.

Barth, A. 2000. The infrared absorption of amino acid side chains. *Prog. Biophys. Mol. Bio.* 74:141-173.

Bartl, F., and G. Zundel. 1997. Molecular recognition and proton transfer processes in maltodextrinphosphorylase-An FTIR study. *J. Mol. Struct.* 404:1-12.

Becke, A. D. 1993. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98:5648-5652.

Bergethon, P. R. 1998. The physical basis of biochemistry. New York: Springer-Verlag New York, Inc.

Bergo, V., E. N. Spudich, K. L. Scott, J. L. Spudich, and K. J. Rothschild. 2000. FTIR analysis of the SII$_{540}$ intermediate of sensory rhodopsin II: Asp73 is the Schiff base proton acceptor. *Biochemistry* 39:2823-2830.

Bergo, V., E. N. Spudich, J. Spudich, and K. J. Rothschild. 2003. Conformational changes detected in a sensory rhodopsin II-Transducer complex. *J. Biol. Chem.* 278:36556-36562.

Bergo, V., E. N. Spudich, J. L. Spudich, and K. J. Rothschild. 2002. A Fourier transform infrared study of Neurospora rhodopsin: similarities with archaeal rhodopsins. *Photochem. Photobiol.* 76:341-349.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. A computer based archival file for macromolecular structures. *J. Mol. Biol.* 112:535-542.

Borgstahl, G. E., D. R. Williams, and E. D. Getzoff. 1995. 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry* 34:6278-6287.

Bousche, O., S. Sonar, M. P. Krebs, H. G. Khorana, and K. J. Rothschild. 1992. Time-resolved Fourier transform infrared spectroscopy of the bacteriorhodopsin mutant Tyr-185-->Phe: Asp-96 reprotonates during O formation; Asp-85 and Asp-212 deprotonate during O decay. *Photochem. Photobiol.* 56:1085-1095.

Braiman, M. S., A. K. Dioumaev, and J. R. Lewis. 1996. A large photolysis-induced pKa increase of the chromophore counterion in bacteriorhodopsin: implications for ion transport mechanisms of retinal proteins. *Biophys. J.* 70:939-947.

Braiman, M. S., T. Mogi, T. Marti, L. J. Stern, H. G. Khorana, and K. J. Rothschild. 1988a. Vibrational spectroscopy of bacteriorhodopsin mutants: Light-Driven proton transport invovles protonation charges of Aspartic acid residues 85, 96, and 212. *Biochemistry* 27:8516-8520.

Braiman, M. S., T. Mogi, L. J. Stern, N. R. Hackett, B. H. Chao, H. G. Khorana, and K. J. Rothschild. 1988b. Vibrational spectroscopy of bacteriorhodopsin mutants: I. Tyrosine-185 protonates and deprotonates during the photocycle. *Proteins* 3:219-229.

Breton, J., E. Nabedryk, J. P. Allen, and J. C. Williams. 1997. Electrostatic influence of QA reduction on the IR vibrational mode of the 10a-ester C==O of HA demonstrated by mutations at residues Glu L104 and Trp L100 in reaction centers from Rhodobacter sphaeroides. *Biochemistry* 36:4515-4525.

Brown, L. S., A. K. Dioumaev, R. Needleman, and J. K. Lanyi. 1998. Local-access model for proton transfer in bacteriorhodopsin. *Biochemistry* 37:3982-3993.

Brown, L. S., J. Sasaki, H. Kandori, A. Maeda, R. Needleman, and J. K. Lanyi. 1995. Glutamic acid 204 is the terminal proton release group at the extracellular surface of bacteriorhodopsin. *J. Biol. Chem.* 270:27122-27126.

Brudler, R., R. Rammelsberg, T. T. Woo, E. D. Getzoff, and K. Gerwert. 2001. Strucutre of the I1 early intermediate of photoactive yellow protein by FTIR spectroscopy. *Nat. Struct. Biol.* 8:265-270.

Camara-Artigas, A., D. Brune, and J. P. Allen. 2002. Interactions between lipids and bacterial reaction centers determined by protein crystallography. *Proc. Natl. Acad. Sci. USA* 99:11055-11060.

Cammi, R., C. Cappelli, S. Corni, and J. Tomasi. 2000. On the calculation of infrared intensities in solution within the polarizable continuum model. *J. Phys. Chem.* 104:9874-9879.

Cappelli, C., S. Corni, R. Cammi, B. Mennucci, and J. Tomasi. 2000. Nonequilibrium formulation of infrared frequencies and intensities in solution: analytical evaluation within the polarizable continuum model. *J. Chem. Phys.* 113:11270-11279.

Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, W. S. I. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng and others. 1999. AMBER 6. University of California, San Francisco.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, J. Merz, K. M., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179-5197.

Creighton, T. E. 1997. Proteins: Structures and Molecular Properties. New York: W. H. Freeman and Company.

Crofts, A. R., S. Lhee, S. B. Crofts, J. Cheng, and S. Rose. 2006. Proton pumping in the bc(1) complex: A new gating mechanism that prevents short circuits. *Biochim. Biophys. Acta.* Epub.

DeLange, F., P. H. M. Bovee-Geurts, A. M. A. Pistorius, K. J. Rothschild, and W. J. DeGrip. 1999. Probing intramolecular orientations in rhodopsin and metarhodopsin II by polarized infrared difference spectroscopy. *Biochemistry* 38:13200-13209.

DeLange, F., C. H. Klaassen, S. E. Wallace-Williams, P. H. Bovee-Geurts, X. M. Liu, W. J. DeGrip, and K. J. Rothschild. 1998. Tyrosine structural changes detected during the photoactivation of rhodopsin. *J. Biol. Chem.* 273:23735-23739.

Desiraju, G. R., and T. Steiner. 1999. The Weak Hydrogen Bond in Structural Chemistry and Biology. Oxford: Oxford University Press.

Devanathan, S., A. Pacheco, L. Ujj, M. Cusanovich, G. Tollin, S. Lin, and N. Woodbury. 1999. Femtosecond spectroscopic observations of initial intermediates in the photocycle of the photoactive yellow protein from Ectothiorhodospira halophila. *Biophys. J.* 77:1017-1023.

Dioumaev, A. K. 2001. Infrared methods for monitoring the protonation state of carboxylic amino acids in the photocycle of bacteriorhodopsin. *Biochemistry (Moscow)* 66:1269-1276.

Dioumaev, A. K., and M. S. Braiman. 1995. Modeling vibrational spectra of amino acid side chains in proteins: the carbonyl stretch frequency of buried carboxylic residues. *J. Am. Chem. Soc.* 117:10572-10574.

Dollinger, G., L. Eisenstein, S. L. Lin, K. Nakanishi, and J. Termini. 1986. Fourier transform infrared difference spectroscopy of bacteriorhodopsin and Its photoproducts regenerated with deuterated tyrosine. *Biochemistry* 23:6524-6533.

Dux, P., G. Rubinstenn, G. W. Vuister, R. Boelens, F. A. Mulder, K. Hard, W. D. Hoff, A. R. Kroon, W. Crielaard, K. J. Hellingwerf and others. 1998. Solution structure and backbone dynamics of the photoactive yellow protein. *Biochemistry* 37:12689-12699.

Dwyer, J., A. G. Gittis, D. Karp, E. Lattman, D. S. Spencer, W. Stites, and B. Garcia-Moreno. 2000. High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.* 79:1610-1620.

Eisberg, R., and R. Resnick. 1985. Quantum physics of atoms, molecules, solids, nuclei, and particles: John Wiley & Sons, Inc.

Engelhard, M., K. Gerwert, B. Hess, W. Kreutz, and F. Siebert. 1985. Light-driven protonation changes of internal aspartic acids of bacteriorhodopsin: an investigation by static and time-resolved infrared difference spectroscopy using [4-13C] aspartic acid labeled purple membrane. *Biochemistry* 24:400-407.

Fahmy, K., F. Jager, M. Beck, T. A. Zvyaga, T. P. Sakmar, and F. Siebert. 1993. Protonation states of membrane-embedded carboxylic acid groups in rhodopsin and metarhodopsin II: A Fourier-transform infrared spectroscopy study of site-directed mutants. *Proc. Natl. Acad. Sci. USA* 90:10206-10210.

Fahmy, K., T. P. Sakmar, and F. Siebert. 2000. Transducin-dependent protonation of glutamic acid 134 in rhodopsin. *Biochemistry* 39:10607-10612.

Fahmy, K., F. Siebert, and T. P. Sakmar. 1994. A mutant rhodopsin photoproduct with a protonated Schiff base displays an active-state conformation: a Fourier-transform infrared spectroscopy study. *Biochemistry* 33:13700-13705.

Foresman, J. B., and Æ. Frisch. 1996. Exploring chemistry with electronic structure methods. Pittsburgh: Gaussian Inc.

Friedrich, T., S. Geibel, R. Kalmbach, I. Chizhov, K. Ataka, J. Heberle, M. Engelhard, and E. Bamberg. 2002. Proteorhodopsin is a light-driven proton pump with variable vectoriality. *J. Mol. Biol.* 321:821-838.

Frisch, Æ., M. J. Frisch, and G. W. Trucks. 2003a. Gaussian 03 user's reference. Carnegie: Gaussian, Inc.

Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, J. Vreven, T., K. N. Kudin, J. C. Burant and others. 2003b. Gaussian 03, Revision A.1. Pittsburgh, PA: Gaussian, Inc.

Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, J. R. E. Stratmann, J.

C. Burant and others. 1998. Gaussian 98, Revision A.9. Pittsburgh, PA: Gaussian, Inc.

Furutani, Y., A. G. J. Bezerra, S. Waschuk, M. Sumii, L. S. Brown, and H. Kandori. 2004a. FTIR spectroscopy of the K photointermediate of Neurospora rhodopsin: structural changes of the retinal, protein, and water molecules after photoisomerization. *Biochemistry* 43:9636-9646.

Furutani, Y., M. Iwamoto, K. Shimono, N. Kamo, and H. Kandori. 2002. FTIR spectroscopy of the M photointermediate in pharaonis rhoborhodopsin. *Biophys. J.* 83:3482-3489.

Furutani, Y., M. Iwamoto, K. Shimono, A. Wada, M. Ito, N. Kamo, and H. Kandori. 2004b. FTIR spectroscopy of the O photointermediate in pharaonis phoborhodopsin. *Biochemistry* 43:5204-5212.

Garcia-Moreno, B. E., J. J. Dwyer, A. G. Gittis, E. E. Lattman, D. S. Spencer, and W. E. Stites. 1997. Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* 64:211-224.

Genick, U. K., S. M. Soltis, P. Kuhn, I. L. Canestrelli, and E. D. Getzoff. 1998. Structure at 0.85 A resolution of an early protein photocycle intermediate. *Nature* 392:206-209.

Gennis, R. B. 1998. How does cytochrome oxidase pump protons? *PNAS* 95:12747-12749.

Gerwert, K. 1999. Molecular reaction mechanisms of proteins monitored by time-resolved FTIR-spectroscopy. *Biol. Chem.* 380:931-935.

Getzoff, E. D., K. N. Gutwin, and U. K. Genick. 2003. Anticipatory active-site motions and chromophore distortion prime photoreceptor PYP for light activation. *Nat. Struct. Biol.* 10:663-668.

Griffiths, D. J. 1999. Introduction to electrodynamics: Prentice-Hall, Inc.

Groenhof, G., M. Bouxin-Cademartory, B. Hess, S. P. De Visser, H. J. Berendsen, M. Olivucci, A. E. Mark, and M. A. Robb. 2004. Photoactivation of the photoactive yellow protein: why photon absorption triggers a trans-to-cis Isomerization of the chromophore in the protein. *J. Am. Chem. Soc.* 126:4228-4233.

Groenhof, G., M. F. Lensink, H. J. C. Berendsen, and A. E. Mark. 2002a. Signal transduction in the photoactive yellow protein. II. proton transfer initiates conformational changes. *Proteins: Struct. Funct. Genet.* 48:212-219.

Groenhof, G., M. F. Lensink, H. J. C. Berendsen, J. G. Snijders, and A. E. Mark. 2002b. Signal transduction in the photoactive yellow protein. I. photon absorption and the isomerization of the chromophore. *Proteins: Struct. Funct. Genet.* 48:202-211.

Hackmann, C., J. Guijarro, I. Chizhov, M. Engelhard, C. Rodig, and F. Siebert. 2001. Static and time-resolved step-scan Fourier transform infrared investigations of the photoreaction of halorhodopsin from Natronobacterium pharaonis: consequences for models of the anion translocation mechanism. *Biophys. J.* 81:394-406.

Heberle, J. 2000. Proton transfer reactions across bacteriorhodopsin and along the membrane. *Biochim. Biophys. Acta.* 1458:135-147.

Hellwig, P., J. Behr, C. Ostermeier, O. M. Richter, U. Pfitzner, A. Odenwald, B. Ludwig, H. Michel, and W. Mantele. 1998. Involvement of glutamic acid 278 in the redox reaction of the cytochrome c oxidase from Paracoccus denitrificans investigated by FTIR spectroscopy. *Biochemistry* 37:7390-7399.

Hellwig, P., C. M. Gomes, and M. Teixeira. 2003. FTIR spectroscopic characterization of the cytochrome aa3 from Acidianus ambivalens: evidence for the involvement of acidic residues in redox coupled proton translocation. *Biochemistry* 42:6179-6184.

Hellwig, P., U. Pfitzner, J. Behr, B. Rost, R. P. Pesavento, W. V. Donk, R. B. Gennis, H. Michel, B. Ludwig, and W. Mantele. 2002. Vibrational modes of tyrosines in cytochrome c oxidase from Paracoccus denitrificans: FTIR and electrochemical studies on Tyr-D4-labeled and on Tyr280His and Tyr35Phe mutant enzymes. *Biochemistry* 41:9116-9125.

Hering, J. A., P. R. Innocent, and P. I. Haris. 2002. Automatic amide I frequency selection for rapid quantification of protein secondary structure from Fourier transform infrared spectra of proteins. *Proteomics* 2:839-849.

Hibbert, F., and J. Emsley. 1990. Hydrogen bonding and chemical reactivity. *Adv. Phys. Org. Chem.* 26:255-379.

Hienerwadel, R., A. Boussac, J. Breton, B. A. Diner, and C. Berthomieu. 1997. Fourier transform infrared difference spectroscopy of photosystem II tyrosine D using site-directed mutagenesis and specific isotope labeling. *Biochemistry* 36:14712-14723.

Hillier, W., and G. Babcock. 2001. S-state dependent Fourier transform infrared difference spectra for the Photosystem II oxygen evolving complex. *Biochemistry* 40:1503-1509.

Hoff, W. D. 1995. Photoactive yellow protein. A new family of eubacterial blue-light photoreceptors [Ph.D.]. University of Amsterdam.

Hoff, W. D., P. Dux, K. Hard, B. Devreese, I. M. Nugteren-Roodzant, W. Crielaard, R. Boelens, R. Kaptein, J. van Beeumen, and K. J. Hellingwerf. 1994a. Thiol ester-linked *p*-coumaric acid as a new photoactive prosthetic group in a protein with rhodopsin-like photochemistry. *Biochemistry* 33:13959-13962.

Hoff, W. D., I. H. van Stokkum, H. J. van Ramesdonk, M. E. van Brederode, A. M. Brouwer, J. C. Fitch, T. E. Meyer, R. van Grondelle, and K. J. Hellingwerf. 1994b. Measurement and global analysis of the absorbance changes in the photocycle of the photoactive yellow protein from *Ectothiorhodospira halophila. Biophys. J.* 67:1691-1705.

Hoff, W. D., A. Xie, I. H. Van Stokkum, X. J. Tang, J. Gural, A. R. Kroon, and K. J. Hellingwerf. 1999. Global conformational changes upon receptor stimulation in photoactive yellow protein. *Biochemistry* 38:1009-1017.

Hohenberg, P., and W. Kohn. 1964. Inhomogeneous electron gas. *Phys. Rev.* 136:B864-B871.

Honig, B., and A. S. Yang. 1995. Free energy balance in protein folding. *Adv. Protein Chem.* 46:27-58.

Hoppe, W., W. Lohmann, H. Markl, and H. Ziegler. 1983. Biophysics. New York: Springer-Verlag.

Hunte, C., H. Palsdottir, and B. L. Trumpower. 2003. Protonmotive pathways and mechanisms in the cytochrome bc1 complex. *FEBS Lett* 545:39-46.

Hutson, M. S., S. V. Shilov, R. Krebs, and M. S. Braiman. 2001. Halide dependence of the halorhodopsin photocycle as measured by time-resolved infrared spectra. *Biophys. J.* 80:1452-1465.

Ihee, H., S. Rajagopal, V. Srajer, R. Pahl, S. Anderson, M. Schmidt, F. Schotte, P. A. Anfinrud, M. Wulff, and K. Moffat. 2005. Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds. *PNAS* 102:7145-7150.

Imamoto, Y., K. Mihara, O. Hisatomi, M. Kataoka, F. Tokunaga, N. Bojkova, and K. Yoshihara. 1997. Evidence for proton transfer from Glu-46 to the chromophore during the photocycle of photoactive yellow protein. *J. Biol. Chem.* 272:12905-12908.

Jager, F., K. Fahmy, T. P. Sakmar, and F. Siebert. 1994. Identification of Glutamic Acid 113 as the Schiff Base Proton Acceptor in the Metarhodopsin I1 Photointermediate of Rhodopsin. *Biochemistry* 33:10878-10882.

Jiang, Z., L. R. Swem, B. G. Rushing, S. Devanathan, G. Tollin, and C. E. Bauer. 1999. Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. *Science* 285:406-409.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926-935.

Kalatskaya, I., S. Schussler, A. Blaukat, W. Muller-Esterl, M. Jochum, D. Proud, and A. Faussner. 2004. Mutation of tyrosine in the conserved NPXXY sequence leads to constitutive phosphorylation and internalization, but not signaling, of the human B2 bradykinin receptor. *J. Biol. Chem.* 279:31268-31276.

Kandori, H. 2004. Hydration switch model for the proton transfer in the Schiff base region of bacteriorhodopsin. *Biochim. Biophys. Acta.* 1658:72-79.

Koh, M., G. Van Driessche, B. Samyn, W. D. Hoff, T. E. Meyer, M. A. Cusanovich, and J. J. Van Beeumen. 1996. Sequence evidence for strong conservation of the photoactive yellow proteins from the halophilic phototrophic bacteria Chromatium salexigens and Rhodospirillum salexigens. *Biochemistry* 35:2526-2534.

Kolbe, M., H. Besir, L. O. Essen, and D. Oesterhelt. 2000. Structure of the light-driven chloride pump halorhodopsin at 1.8 A resolution. *Science* 288:1390-1396.

Kort, R., K. J. Hellingwerf, and R. B. Ravelli. 2004. Initial events in the photocycle of photoactive yellow protein. *J. Biol. Chem.* 279:26417-26424.

Kort, R., W. D. Hoff, M. Van West, A. R. Kroon, S. M. Hoffer, K. H. Vlieg, W. Crielaand, J. J. Van Beeumen, and K. J. Hellingwerf. 1996. The xanthopsins: a new family of eubacterial blue-light photoreceptors. *EMBO J.* 15:3209-3218.

Koutsoupakis, C., T. Soulimane, and C. Varotsis. 2004. Probing the Q-proton pathway of ba3-cytochrome c oxidase by time-resolved Fourier transform infrared spectroscopy. *Biophys. J.* 86:2438-2444.

Krebs, M. P., and H. G. Khorana. 1993. Mechanism of light-dependent proton translocation by bacteriorhodopsin. *J. Bacteriol.* 175:1555-1560.

Lanyi, J. K. 2004. Bacteriorhodopsin. *Annu. Rev. Physiol.* 66.

Lanyi, J. K., and B. Schobert. 2002. Crystallographic structure of the retinal and the protein after deprotonation of the Schiff base: the switch in the bacteriorhodopsin photocycle. *J. Mol. Biol.* 321:727-737.

Larsen, D. S., I. H. van Stokkum, M. Vengris, M. A. van Der Horst, F. L. de Weerd, K. J. Hellingwerf, and R. van Grondelle. 2004. Incoherent manipulation of the photoactive yellow protein photocycle with dispersed pump-dump-probe spectroscopy. *Biophys. J.* 87:1858-1872.

Lascombe, J., M. Haurie, and J. M. L. 1962. Influence des solvants sur la vibration de valence $v_{C=O}$ de quelques acides monocarboxyliques: etude par spectroscopie infrarouge. *J. Chim. Phys* 59:1233-1246.

Lee, C., W. Yang, and R. G. Parr. 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 37:785-789.

Lide, D. R. 1997-1998. CRC handbook of chemistry and physics. New York: CRC Press.

Liu, X., S. Sonar, C. Lee, M. Coleman, U. L. RajBhandary, and K. J. Rothschild. 1995. Site-direct isotope labeling and FTIR spectroscopy: assignment of tyrosine bands in the bR- M difference spectrum of bacteriorhodopsin. *Biophys. Chem.* 56:63-70.

Lubben, M., A. Prutsch, B. Mamat, and K. Gerwert. 1999. Electron transfer induces side-chain conformational changes of glutamate-286 from cytochrome bo3. *Biochemistry* 38:2048-2056.

Maeda, A. 2001. Internal water molecules as mobile polar groups for light-induced proton translocation in bacteriorhodopsin and rhodopsin as studied by difference FTIR spectroscopy. *Biochemistry (Moscow)* 66:1256-1268.

Maeda, A., J. Sasaki, Y. Shichida, T. Yoshizawa, M. Chang, B. Ni, R. Needleman, and J. K. Lanyi. 1992. Structures of aspartic acid-96 in the L and N intermediates of bacteriorhodopsin: analysis by Fourier transform infrared spectroscopy. *Biochemistry* 31:4684-4690.

Maeda, A., F. L. Tomson, R. B. Gennis, S. P. Balashov, and T. G. Ebrey. 2003. Water molecule rearrangements around Leu93 and Trp182 in the formation of the L intermediate in bacteriorhodopsin's photocycle. *Biochemistry* 42:2535-2541.

Marti, T., H. Otto, S. J. Rosselet, M. P. Heyn, and H. G. Khorana. 1992. Anion binding to the Schiff base of the bacteriorhodopsin mutants Asp-85---- Asn/Asp-212----Asn and Arg-82----Gln/Asp-85----Asn/Asp-212----Asn. *J. Biol. Chem.* 267:16922-16927.

McCammon, J. A., B. R. Gelin, and M. Karplus. 1977. Dynamics of Folded Proteins. *Nature* 267:585-590.

Meyer, T. E. 1985. Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium Ectothiorhodospira halophila. *Biochim. Biophys. Acta.* 806:175-183.

Meyer, T. E., S. Devanathan, T. T. Woo, E. D. Getzoff, G. Tollin, and M. A. Cusanovich. 2003. Site-specific mutations provide new insights into the origin of pH effects and alternative spectral forms in the photoactive yellow protein from halorhodospira halophilia. *Biochemistry* 42:3319-3325.

Meyer, T. E., G. Tollin, J. H. Hazzard, and M. A. Cusanovich. 1989. Photoactive yellow protein from the purple phototrophic bacterium, Ectothiorhodospira halophila. Quantum yield of photobleaching and effects of temperature, alcohols, glycerol, and sucrose on kinetics of photobleaching and recovery. *Biophys. J.* 56:559-564.

Meyer, T. E., E. Yakali, M. A. Cusanovich, and G. Tollin. 1987. Properties of a water-soluble, yellow protein isolated from a halophilic phototrophic bacterium that has photochemical activity analogous to sensory rhodopsin. *Biochemistry* 26:418-423.

Nabedryk, E., J. Breton, H. M. Joshi, and D. K. Hanson. 2000. Fourier transform infrared evidence of proton uptake by glutamate L212 upon reduction of the secondary quinone QB in the photosynthetic reaction center from Rhodobacter capsulatus. *Biochemistry* 39:14654-14663.

Nabedryk, E., J. Breton, M. Y. Okamura, and M. L. Paddock. 2001. Simultaneous replacement of Asp-L210 and Asp-M17 with Asn increases proton uptake by Glu-L212 upon first electron transfer to QB in reaction centers from Rhodobacter sphaeroides. *Biochemistry* 40:13826-13832.

Nagle, J. F., and S. Tristram-Nagle. 1983. Hydrogen bonded chain mechanisms for proton conduction and proton pumping. *J. Membrane Biol.* 74:1-14.

Namslauer, A., and P. Brzezinski. 2004. Structural elements involved in electron-coupled proton transfer in cytochrome c oxidase. *FEBS Lett* 567:103-110.

Nelson, D. L., and M. M. Cox. 2000. Lehninger principle of biochemistry. New York: Worth Publishers.

Nie, B. 2002. Vibrational band assignment and electrostatic properties of biomolecules based on *ab initio* computational studies [Master thesis]. Oklahoma State University.

Nie, B., J. Stutzman, and A. Xie. 2005. A vibrational spectral marker for probing the hydrogen-bonding status of protonated Asp and Glu residues. *Biophys. J.* 88:2833-2847.

Noguchi, T., and M. Sugiura. 2003. Analysis of flash-induced FTIR difference spectra of the S-state cycle in the photosynthetic water-oxidizing complex by uniform 15N and 13C isotope labeling. *Biochemistry* 42:6035-6042.

Okada, T., Y. Fujiyoshi, M. Silow, J. Navarro, E. M. Landau, and Y. Shichida. 2002. Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography. *Proc. Natl. Acad. Sci. USA* 99:5982-5987.

Otto, H., T. Marti, M. Holz, T. Mogi, L. J. Stern, F. Engel, H. G. Khorana, and M. P. Heyn. 1990. Substitution of amino acids Asp-85, Asp-212, and Arg-82

in bacteriorhodopsin affects the proton release phase of the pump and the pK of the Schiff base. *Proc. Natl. Acad. Sci. USA* 87:1018-1022.

Paddock, M. L., G. Feher, and M. Y. Okamura. 2003. Proton transfer pathways and mechanism in bacterial reaction centers. *FEBS Lett* 555:45-50.

Pebay-Peyroula, E., G. Rummel, J. P. Rosenbusch, and E. M. Landau. 1997. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science* 277:1676-1681.

Perman, B., V. Srajer, Z. Ren, T. Teng, C. Pradervand, T. Ursby, D. Bourgeois, F. Schotte, M. Wulff, R. Kort and others. 1998. Energy transduction on the nanosecond time scale: early structural events in a xanthopsin photocycle. *Science* 279:1946-1950.

Perrin, C. L., and J. B. Nielson. 1997. "Strong" hydrogen bonds in chemistry and biology. *Annu. Rev. Phys. Chem.* 48:511-544.

Puustinen, A., J. A. Bailey, R. B. Dyer, S. L. Mecklenburg, M. Wikstrom, and W. H. Woodruff. 1997. Fourier transform infrared evidence for connectivity between CuB and glutamic acid 286 in cytochrome bo3 from Escherichia coli. *Biochemistry* 36:13195-13200.

Rahman, A. 1964. Correlations in Motion of Atoms in Liquid Argon. *Physical Review a-General Physics* 136:A405-&.

Rath, P., E. Spudich, D. D. Neal, J. L. Spudich, and K. J. Rothschild. 1996. Asp76 is the Schiff base counterion and proton acceptor in the proton-translocating form of sensory rhodopsin I. *Biochemistry* 35:6690-6696.

Riistama, S., G. Hummer, A. Puustinen, R. B. Dyer, W. H. Woodruff, and M. Wikström. 1997. Bound water in the proton translocation mechanism of the haem-copper oxidases. *FEBS Lett.* 414:275-280.

Rollin-Genetet, F., C. Berthomieu, A. H. Davin, and E. Quemeneur. 2004. Escherichia coli thioredoxin inhibition by cadmium: two mutually exclusive binding sites involving Cys32 and Asp26. *Eur. J. Biochem.* 271:1299-1309.

Rothschild, K. J. 1992. FTIR difference spectroscopy of bacteriorhodopsin: toward a molecular model. *J. Bioenerg. Biomembr.* 24:147-167.

Rothschild, K. J., M. S. Braiman, Y. W. He, T. Marti, and H. G. Khorana. 1990. Vibrational spectroscopy of bacteriorhodopsin mutants. Evidence for the interaction of aspartic acid 212 with tyrosine 185 and possible role in the proton pump mechanism. *J. Biol. Chem.* 265:16985-16991.

Sasaki, J., J. K. Lanyi, R. Needleman, T. Yoshizawa, and A. Maeda. 1994. Complete identification of C=O stretching vibrational bands of protonated aspartic acid residues in the difference infrared spectra of M and N intermediates versus bacteriorhodopsin. *Biochemistry* 33:3178-3184.

Schmidt, B., W. Hillier, J. McCracken, and S. Ferguson-Miller. 2004. The use of stable isotopes and spectroscopy to investigate the energy transducing function of cytochrome c oxidase. *Biochim. Biophys. Acta.* 1655:248-255.

Shankar, R. 1994. Principles of quantum mechanics. New York: Plenum Press.

Sonar, S., C. P. Lee, M. Coleman, N. Patel, X. Liu, T. Marti, H. G. Khorana, U. L. RajBhandary, and K. J. Rothschild. 1994. Site-directed isotope labelling and FTIR spectroscopy of bacteriorhodopsin. *Nat. Struct. Biol.* 1:512-517.

Song, Y., J. Mao, and M. R. Gunner. 2003. Calculation of proton transfers in Bacteriorhodopsin bR and M intermediates. *Biochemistry* 42:9875-9888.

Soulimane, T., G. Buse, G. P. Bourenkov, H. D. Bartunik, R. Huber, and M. E. Than. 2000. Structure and mechanism of the aberrant ba(3)-cytochrome c oxidase from thermus thermophilus. *EMBO J.* 19:1766-1776.

Sprenger, W. W., W. D. Hoff, J. P. Armitage, and K. J. Hellingwerf. 1993. The eubacterium *Ectothiorhodospira halophila* is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein. *J. Bacteriol.* 175:3096-3104.

Steiner, T. 2002. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed. Engl.* 41:49-76.

Stilling, F. H., and A. Rahman. 1974. Improved Simulation of Liquid Water by Molecular-Dynamics. *Journal of Chemical Physics* 60:1545-1557.

Subramaniam, S., D. A. Greenhalgh, and H. G. Khorana. 1992. Aspartic acid 85 in bacteriorhodopsin functions both as proton acceptor and negative counterion to the Schiff base. *J. Biol. Chem.* 267:25730-25733.

Subramaniam, S., and R. Henderson. 2000. Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature* 406:653-657.

Sudo, Y., M. Iwamoto, K. Shimono, and N. Kamo. 2002. Tyr-199 and charged residues of pharaonis Phoborhodopsin are important for the interaction with its transducer. *Biophys. J.* 83:427-432.

Susi, H., S. N. Timasheff, and L. Stevens. 1967. Infrared spectra and protein conformations in aqueous solutions. I. The amide I band in H2O and D2O solutions. *J. Biol. Chem.* 242:5460-5466.

Svensson-Ek, M., J. Abramson, G. Larsson, S. Tornroth, P. Brezezinski, and S. Iwata. 2002. The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome c oxidases from Rhodobacter sphaeroides. *J. Mol. Biol.* 321:329-339.

Szaraz, S., D. Oesterhelt, and P. Ormos. 1994. pH-induced structural changes in bacteriorhodopsin studied by Fourier transform infrared spectroscopy. *Biophys. J.* 67:1706-1712.

Takeuchi, H., N. Watanabe, Y. Satoh, and I. Harada. 1989. Effects of hydrogen bonding on the tyrosine Raman bands in the 1300-1150 cm^{-1} region. *J. Raman Spectrosc.* 20:233-237.

Thijssen, J. M. 1999. Computational physics. Cambridge: Cambridge University Press.

Thompson, M. J., D. Bashford, L. Noodleman, and E. D. Getzoff. 2003. Photoisomerization and proton transfer in photactive yellow protein. *J. Am. Chem. Soc.* 125:8186-8194.

Tsukihara, T., K. Shimokata, Y. Katayama, H. Shimada, K. Muramoto, H. Aoyama, M. Mochizuki, K. Shinzawa-Itoh, E. Yamashita, M. Yao and others. 2003. The low-spin heme of cytochrome c oxidase as the driving element of the proton-pumping process. *Proc. Natl. Acad. Sci. USA* 100:15304-15309.

Ujj, L., S. Devanathan, T. E. Meyer, M. A. Cusanovich, G. Tollin, and G. H. Atkinson. 1998. New photocycle intermediates in the photoactive yellow
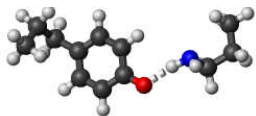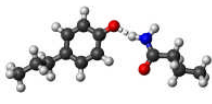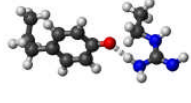
protein from *Ectothiorhodospira halophila*: picosecond transient absorption spectroscopy. *Biophys. J.* 75:406-412.

van Aalten, D. M. F., A. Haker, J. Hendriks, K. J. Hellingwerf, L. Joshua-Tor, and W. Crielaard. 2002. Engineering photocycle dynamics. Crystal structures and kinetics of three photoactive yellow protein hinge-bending mutants. *J. Biol. Chem.* 277:6463-6468.

van Aalten, D. M. F., W. D. Hoff, J. B. Findlay, W. Crielaard, and K. J. Hellingwerf. 1998. Concerted motions in the photoactive yellow protein. *Protein Eng.* 11:873-879.

van Beeumen, J. J., B. V. Devreese, S. M. Van Bun, W. D. Hoff, K. J. Hellingwerf, T. E. Meyer, D. E. McRee, and M. A. Cusanovich. 1993. Primary structure of a photoactive yellow protein from the phototrophic bacterium Ectothiorhodospira halophila, with evidence for the mass and the binding site of the chromophore. *Protein Sci.* 2:1114-1125.

van Gunsteren, W. F., and H. J. C. Berendsen. 1977. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Molecular Physics* 34:1311-1327.

Vinogradov, S. N., and R. H. Linnell. 1971. Hydrogen bonding. New York: Van Nostrand Reinhold Company.

Wolpert, M., O. Maneg, B. Ludwig, and P. Hellwig. 2004. Characterization of the CuA center in the cytochrome c oxidase from Thermus thermophilus for the spectral range 1800-500 cm-1 with a combined electrochemical and Fourier transform infrared spectroscopic setup. *Biopolymers* 74:73-76.

Xie, A., W. D. Hoff, A. R. Kroon, and K. J. Hellingwerf. 1996. Glu46 donates a proton to the 4-hydroxycinnamate anion chromophore during the photocycle of photoactive yellow protein. *Biochemistry* 35:14671-14678.

Xie, A., L. Kelemen, J. Hendriks, B. J. White, K. J. Hellingwerf, and W. D. Hoff. 2001. Formation of a new buried charge drives a large-amplitude protein quake in photoreceptor activation. *Biochemistry* 40:1510-1517.

Yamato, T., N. Niimura, and N. Go. 1998. Molecular dynamics study of femtosecond events in photoactive yellow protein after photoexcitation of the chromophore. *Proteins* 32:268-275.

Young, D. 2001. Computational chemistry: a practical guide for applying techniques to real world problems. New York: John Wiley & Sons, Inc.

Zhang, J., W. Oettmeier, R. B. Gennis, and P. Hellwig. 2002. FTIR spectroscopic evidence for the involvement of an acidic residue in quinone binding in cytochrome bd from Escherichia coli. *Biochemistry* 41:4612-4617.

Zscherp, C., R. Schlesinger, J. Tittor, D. Oesterhelt, and J. Heberle. 1999. In situ determination of transient pK changes of internal amino acids of bacteriorhodopsin by using time-resolved attenuated total reflection Fourier-transform infrared spectroscopy. *Proc. Natl. Acad. Sci. USA* 96:5498-5503.

Appendix A

The hydrogen bonding properties of other polar amino acids

In addition to protonated Asp/Glu and Tyr, we also calculated the hydrogen bonding properties of important amino acid side chains, such as ionized Asp/Glu, ionized Tyr, neutral and charged Arg, neutral and charged His, Asn/Gln, and Ser/Thr. We are currently in the process of identifying vibrational spectral markers of these amino acid side chains. Their calculated hydrogen bonding properties are listed in the Table 1 to 12.

TABLE 1. The hydrogen bonding interactions of ionized Tyrosine with neutral polar groups

Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	ν_{C-O} (cm^{-1})	ν_{ring} (cm^{-1})
	TYR	0	N/A	N/A	N/A	1383.8	1543.4
	LYS	1	2.88	169.6	34.3	1373.1	1524.7
	SER	1	2.61	172.3	69.7	1360.0	1520.7
	H$_2$O	1	2.69	169.3	67.1	1367.2	1520.1
	ASN/GLN	1	2.68	162.7	63.9	1362.9	1515.5
	ASP/GLU	1	2.52	171.8	88.2	1362.3	1516.9
	ARG	1	2.81	174.7	69.7	1361.6	1515.6
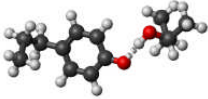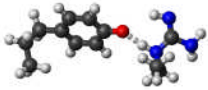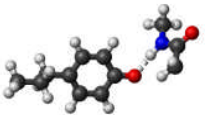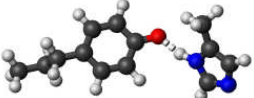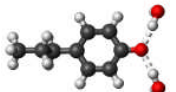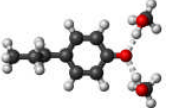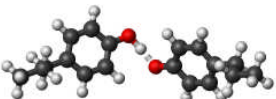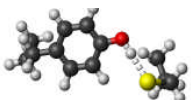
	THR	1	2.64	171.7	66.6	1358.9	1514.0
	ARG	1	2.76	164.2	69.6	1358.8	1514.6
	BB	1	2.73	173.1	90.2	1358.0	1513.4
	HIS	1	2.65	164.2	99.8	1354.1	1512.7 1510.6
	H2O	2	2.72 2.72	167.7 167.7	125.69	1342.9	1508.3
	MeOH	2	2.67 2.67	170.9 170.9	130.4	1335.0	1512.2
	TYR	1	2.49	177.4	108.8	1327.4 1294.0	1516.4
	CYS	1	3.07	177.3	69.4	1289.3	1506.6

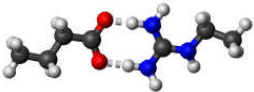**TABLE 4. The hydrogen bonding interactions of charged Asp/Glu with charged amino acid side chains**

Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	Anti-sym & sym v_{COO^-} (cm^{-1})	COOH Stretching (cm^{-1})
	HIS	1	2.77	178.5	523.6	N/A	1736.5
	HIS	1	2.78	171.0	524.8	N/A	1736.3
	LYS	1	2.78	165.0	541.4	N/A	1785.2
	ARG	1	2.59 2.66	178.8 178.0	495.2	1607.2 1392.7	N/A

TABLE 5. The hydrogen bonding interactions of Histidine with neutral polar groups

Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	N-H Stretching Freq. (cm^{-1})
	HIS	0	N/A	N/A	N/A	3501.2
	CYS	1	3.52	174.4	10.8	3502.5
	LYS	1	3.25	179.7	11.0	3501.8
	CYS	1	3.57	158.2	14.0	3391.8
	MET	1	3.58	171.7	16.1	3373.4
	TYR	1	3.01	173.8	16.7	3404.3

TABLE 5. The hydrogen bonding interactions of Histidine with neutral polar groups
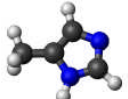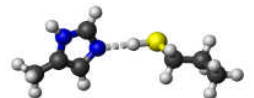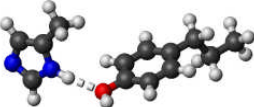
Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	N-H Stretching Freq. (cm^{-1})
	HIS	0	N/A	N/A	N/A	3501.2
	CYS	1	3.52	174.4	10.8	3502.5
	LYS	1	3.25	179.7	11.0	3501.8
	CYS	1	3.57	158.2	14.0	3391.8
	MET	1	3.58	171.7	16.1	3373.4
	TYR	1	3.01	173.8	16.7	3404.3

TABLE 4. The hydrogen bonding interactions of charged Asp/Glu with charged amino acid side chains
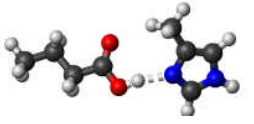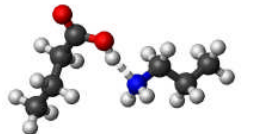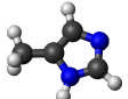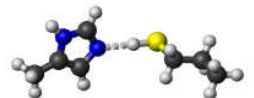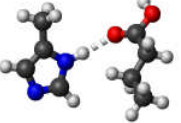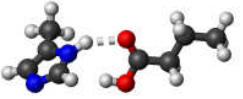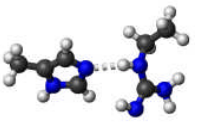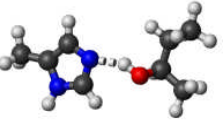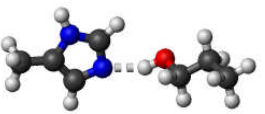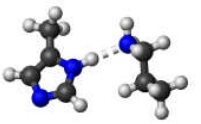
Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	Anti-sym & sym ν_{COO^-} (cm^{-1})	COOH Stretching (cm^{-1})
	HIS	1	2.77	178.5	523.6	N/A	1736.5
	HIS	1	2.78	171.0	524.8	N/A	1736.3
	LYS	1	2.78	165.0	541.4	N/A	1785.2
	ARG	1	2.59 2.66	178.8 178.0	495.2	1607.2 1392.7	N/A

TABLE 5. The hydrogen bonding interactions of Histidine with neutral polar groups

Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	N-H Stretching Freq. (cm^{-1})
	HIS	0	N/A	N/A	N/A	3501.2
	CYS	1	3.52	174.4	10.8	3502.5
	LYS	1	3.25	179.7	11.0	3501.8
	CYS	1	3.57	158.2	14.0	3391.8
	MET	1	3.58	171.7	16.1	3373.4
	TYR	1	3.01	173.8	16.7	3404.3

	ASP/GLU	1	2.99	174.7	20.3	3398.1
	ASP/GLU	1	2.93	151.8	21.2	3343.1
	ARG	1	3.06	155.0	21.3	3503.5
	SER	1	2.95	177.8	22.2	3327.5
	THR	1	2.93	178.0	22.4	3348.8
	THR	1	2.85	152.0	23.1	3504.1
	SER	1	2.83	151.6	25.2	3504.3
	LYS	1	2.99	177.3	27.3	3139.3

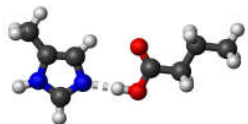|						
	ARG	1	3.06	173.8	27.5	3505.2
	ASN/GLN	1	2.92	176.2	28.5	3305.8
	HIS	1	3.00	176.8	34.1	3503.5 3225.9
	ASN/GLN	1	2.99	166.7	34.9	3505.0
	ARG	1	2.91	165.4	37.6	3099.3
	TYR	1	2.85	173.6	38.4	3504.2
	ASP/GLU	1	2.77	178.5	45.3	3504.9
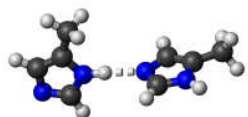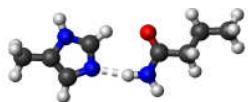
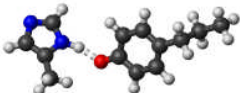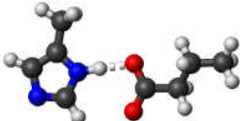**TABLE 6. The hydrogen bonding interactions of Histidine with charged amino acid side chains**
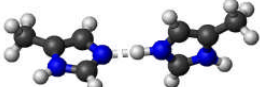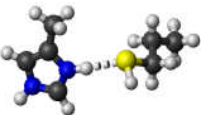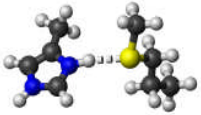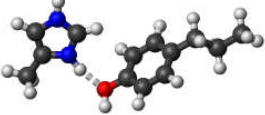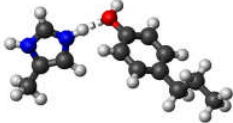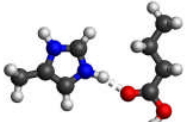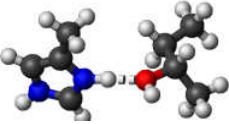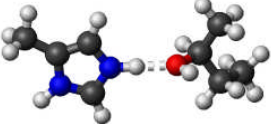
Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	O-H Stretching Freq. (cm^{-1})
	TYR	1	2.65	164.2	99.8	2558.3
	ARG	1	2.99 2.96	151.0 153.4	104.1	3492.2
	ASP/GLU	1	2.72	178.9	109.3	2682.9
	HIS	1	2.72	178.1	112.1	3494.7
	HIS	1	2.70	178.8	114.3	3495.6
	LYS	1	2.69	177.4	130.7	3492.8

TABLE 7. The hydrogen bonding interactions of charged Histidine with neutral polar groups

Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)
	HIS	0	N/A	N/A	N/A
	CYS	1	3.32	172.1	50.4
	CYS	1	3.30	169.9	52.4
	MET	1	3.27	166.5	58.0
	TYR	1	2.78	167.3	59.4
	MET	1	3.26	169.2	59.7

	TYR	1	2.75	175.1	60.6
	ASP/GLU	1	2.74	169.3	69.7
	ASP/GLU	1	2.72	165.5	70.5
	THR	1	2.72	176.4	70.7
	SER	1	2.71	169.9	72.3
	THR	1	2.70	180.0	73.7
	SER	1	2.70	176.5	74.2
	LYS	1	2.76	175.1	78.6
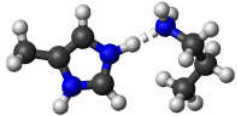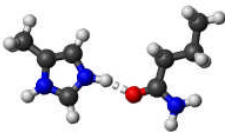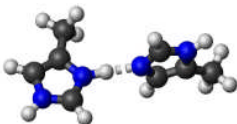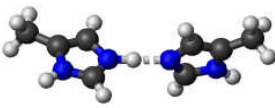
	LYS	1	2.74	173.7	81.7
	ASN/GLN	1	2.66	161.8	95.5
	ASN/GLN	1	2.64	161.8	98.4
	HIS	1	2.72	178.1	112.1
	HIS	1	2.70	178.8	114.3
	ARG	1	3.01 2.96	151.4 154.3	151.1
	ARG	1	3.00 2.96	151.3 153.9	154.9

TABLE 8. The hydrogen bonding interactions of charged Histidine with charged amino acid side chains
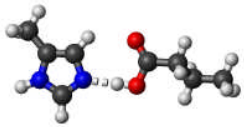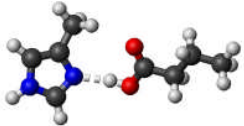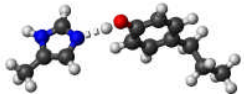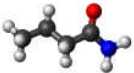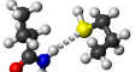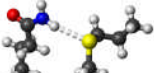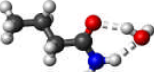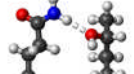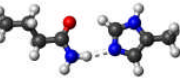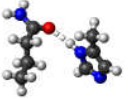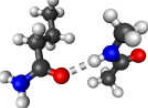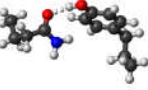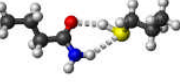
Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)
	ASP/GLU	1	2.77	178.5	523.6
	ASP/GLU	1	2.78	171.0	524.8
	TYR	1	2.85	173.5	527.8
	TYR	1	2.82	167.2	528.2

TABLE 9. The hydrogen bonding interactions of Asn/Gln with neutral polar groups

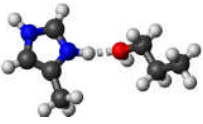| Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	$V_{C=O}$ (cm^{-1})	sym V_{NH2} (cm^{-1})	anti-sym V_{NH2} (cm^{-1})	ND_2 $V_{C=O}$ (cm^{-1})	$^{15}NH_2$ $V_{C=O}$ (cm^{-1})	$^{15}ND_2$ $V_{C=O}$ (cm^{-1})
	Asn/Gln	0	N/A	N/A	N/A	1734.6	3449.0	3571.3	1727.1	1733.5	1726.6
	CYS	1	3.66	167.8	7.8	1729.2	3407.8	3538.8	1721.5	1727.9	1720.9
	TYR	1	3.06	172.1	10.7	1728.1	3413.6	3541.5	1720.1	1726.8	1719.4
	MET	1	3.65	167.2	11.4	1727.3	3398.9	3538.0	1719.5	1725.9	1718.7
	H2O	1	3.00	173.2	16.6	1727.0	3371.0	3530.1	1718.9	1725.4	1718.1
	THR	1	3.01	175.5	18.4	1726.0	3384.7	3533.7	1717.9	1724.5	1717.2
	BB-CO	1	2.97	172.3	24.8	1725.5	3333.1	3513.8	1717.4	1724.2	1716.7

	SER	1	3.00	176.3	18.9	1725.1	3365.2	3526.1	1717.0	1723.8	1716.3
	LYS	1	3.05	175.7	22.5	1721.9	3231.3	3504.8	1713.4	1720.2	1712.6
	HIS	1	2.99	166.7	34.9	1706.6	3209.7	3530.0	1690.0	1703.4	1688.8
	HIS	1	2.92	176.2	28.5	1702.5	3454.9	3579.2	1693.1	1700.7	1692.4
	BB-NH	1	3.00	174.6	23.0	1702.4	3454.2	3578.5	1694.0	1700.9	1693.2
	THR	1	2.82	169.5	20.2	1698.0	3456.0	3580.4	1688.5	1696.3	1687.8
	H2O	1	2.78	152.6	37.5	1694.3	3281.3	3542.9	1670.5	1691.3	1669.5
	TYR	1	2.77	165.4	26.9	1685.1	3422.7	3559.2	1672.7	1682.7	1671.8
	CYS	1.5	3.42/3.57	154.5 146.7	16.4	1703.6	3387.2	3548.9	1692.8	1701.8	1692.0
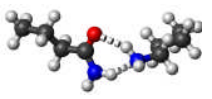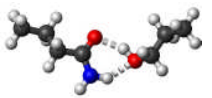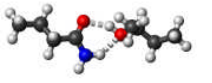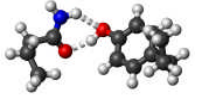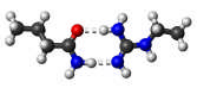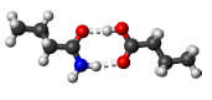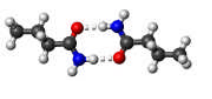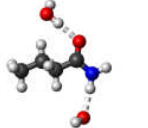
	LYS	1.5	3.05/2.94	145.1 148.9	28.8	1701.2	3198.0	3533.8	1681.3	1698.3	1680.3
	SER	1.5	2.78/2.87	153.0 141.0	37.1	1691.9	3297.4	3547.0	1670.5	1688.8	1669.5
	THR	1.5	2.79/2.87	153.1 141.1	32.5	1690.4	3306.2	3547.2	1669.1	1687.3	1668.0
	TYR	1.5	2.72/2.94	157.5 136.6	33.0	1684.2	3381.1	3552.7	1664.3	1681.1	1663.2
	ARG	2	2.93/2.92	176.2 174.3	53.0	1705.0	3033.8	3520.7	1666.0	1694.4	1661.2
	ASP /GLU	2	2.68/2.88	173.4 168.0	61.9	1717.5 1679.6	3246.9	3536.9	1716.1 1649.0	1717.2 1674.9	1715.8 1647.6
	ASN GLN	2	2.89/2.89	176.2 176.3	54.1	1705.4 1688.4	3213.0 3168.0	3530.4 3530.0	1701.7 1667.2	1704.0 1685.4	1701.3 1666.1
	H2O	2	2.97 2.82	112.9 169.9	46.9	1697.2	3357.1	3580.4	1685.8	1693.9	1684.8

TABLE 10. The hydrogen bonding interactions of Asn/Gln with charged amino acid side chains
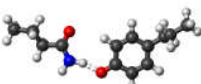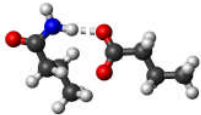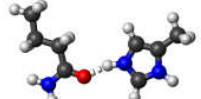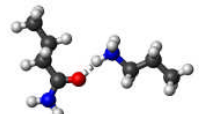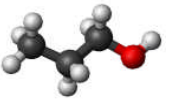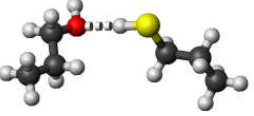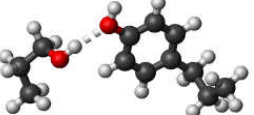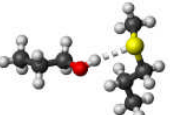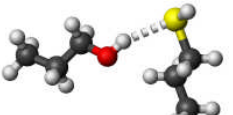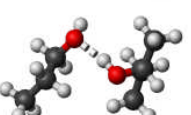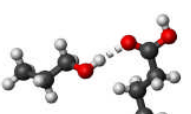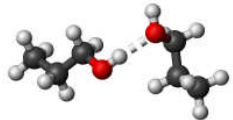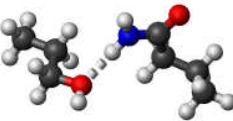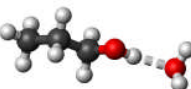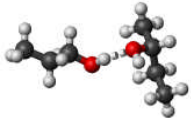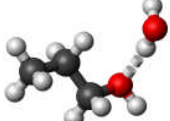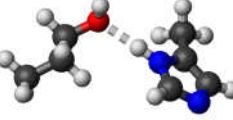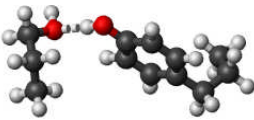
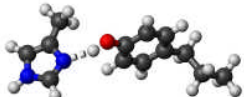| Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	$V_{C=O}$ (cm^{-1})	sym V_{NH2} (cm^{-1})	anti-sym V_{NH2} (cm^{-1})	ND_2 $V_{C=O}$ (cm^{-1})	$^{15}NH_2$ $V_{C=O}$ (cm^{-1})	$^{15}ND_2$ $V_{C=O}$ (cm^{-1})
	TYR4	1	2.68	162.7	63.9	1707.4	N/A	3489.5	1693.9	1702.4	1691.6
	ASP /GLU	1	2.80	173.9	98.1	1697.3	N/A	3489.6	1685.9	1694.7	1684.9
	ARG	1	2.77	176.9	76.2	1686.0	3455.6	3572.5	1682.9	1685.0	1682.7
	HIS	1	2.64	161.8	98.4	1668.0	3449.0	3568.1	1647.2	1660.5	1645.6
	HIS	1	2.67	161.8	95.5	1667.5	3448.4	3567.6	1652.1	1663.6	1650.6
	LYS	1	2.62	162.1	116.0	1660.3	3445.8	3564.3	1653.0	1656.1	1652.9

TABLE 11. The hydrogen bonding interactions of Serine with neutral polar groups

Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	O-H Stretching Freq. (cm^{-1})
	SER	0	N/A	N/A	N/A	3607.1
	CYS	1	3.49	179.7	8.2	3605.0
	TYR	1	2.91	162.8	11.2	3554.3
	MET	1	3.46	169.4	12.2	3501.4
	CYS	1	3.43	155.2	14.0	3509.3
	THR	1	2.88	167.2	14.8	3608.7
	ASP/GLU	1	2.89	171.3	17.6	3511.2

	SER	1	2.85	166.1	18.1	3610.0 3469.3
	ASN/GLN	1	3.00	176.3	18.9	3610.0
	H2O	1	2.85	165.1	19.5	3473.2
	THR	1	2.84	165.0	20.2	3453.3
	H2O	1	2.86	168.4	21.8	3607.0
	HIS	1	2.95	177.8	22.2	3609.6
	TYR	1	2.79	172.7	24.1	3610.5
	BB	1	3.08	179.9	25.0	3618.8
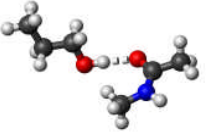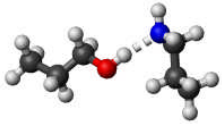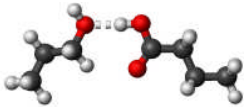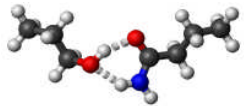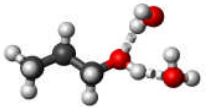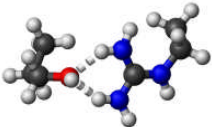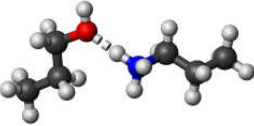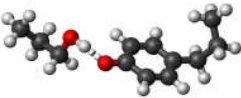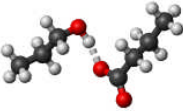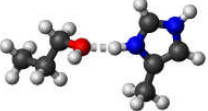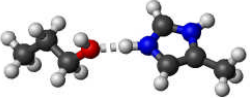
	BB	1	2.81	166.2	25.1	3454.0
	HIS	1	2.83	151.6	25.2	3401.2
	LYS	1	2.87	169.2	26.2	3279.3
	ASP/GLU	1	2.75	177.7	31.9	3610.9
	ASN/GLN	1.5	2.78/2.87	153.0/141.0	37.1	3350.7
	ARG	1.5	2.76/2.95	153.7/148.7	40.5	3181.2
	H2O	2	2.75/2.74	151.5/153.6	63.5	3337.8 3245.2

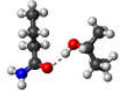**TABLE 12. The hydrogen bonding interactions of Serine with charged amino acid side chains**

Structure	Amino Acid	No. of H-bond	H-bond Length (Å)	H-bond Angle (deg)	H-bond Energy* (kJ/mol)	O-H Stretching Freq. (cm^{-1})
	ARG	1.5	2.88 2.87	148.2 148.6	62.3	3589.7
	LYS	1	2.71	170.3	79.4	3612.4
	TYR	1	2.61	172.3	69.7	2716.3
	ASP/GLU	1	2.64	168.5	70.7	2917.3
	HIS	1	2.71	169.9	72.3	3604.0
	HIS	1	2.70	176.5	74.2	3609.1

Appendix B

Protocol for Gaussian03 calculations

1. Draw the chemical structure starting from carbon atoms in Chemdraw software.
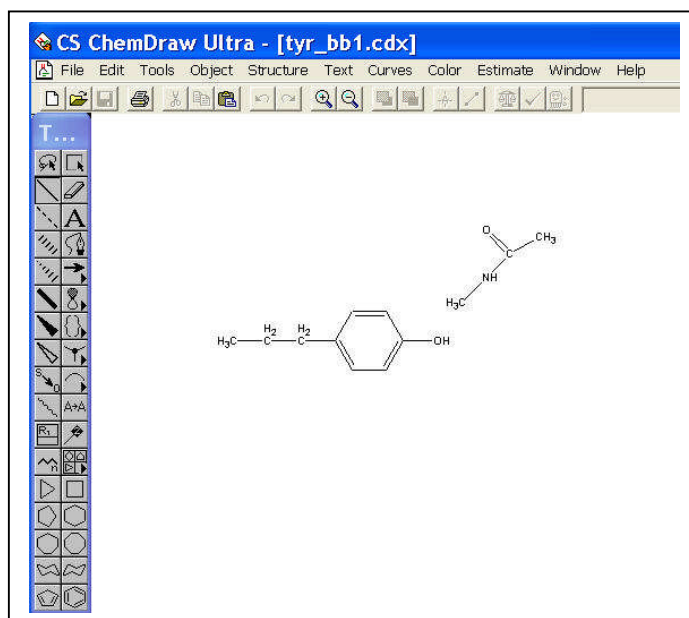
2. Optimize the structure in Chem3D with PM3 method with Minimum RMS gradient of 0.01. Save the optimized structure in MOPAC format.
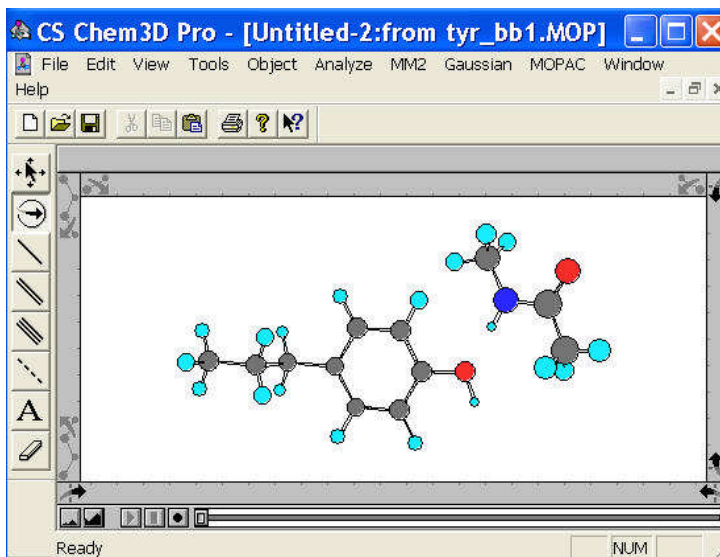
3. Change the extension of MOP to INP file. Select utilities and then NewZMat in Gaussian98 to convert the INP file to a Gaussian input file.
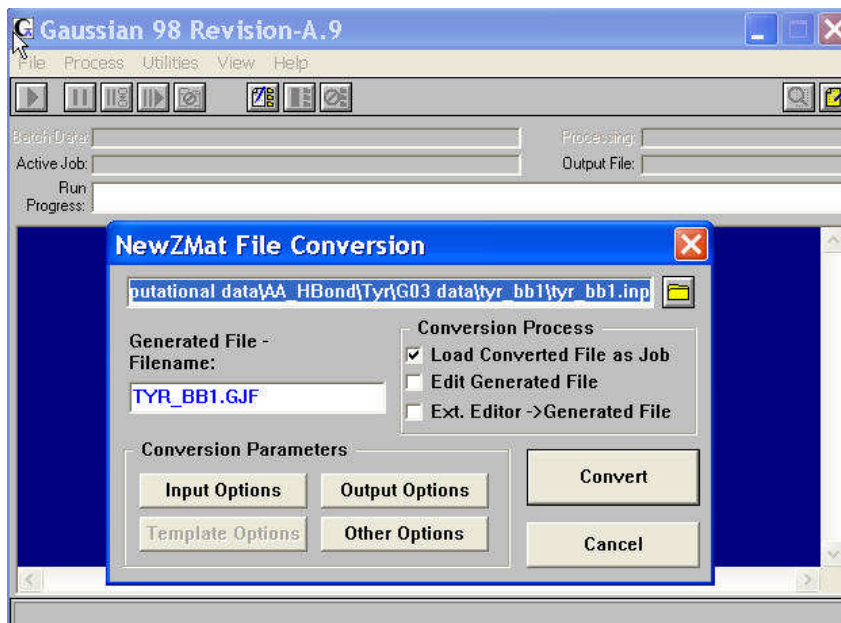
4. Optimize the geometry and calculate the frequency using B3LYP/6-31G(d) method. Then the energy is calculated using B3LYP/6-311+G(2d, p) method.
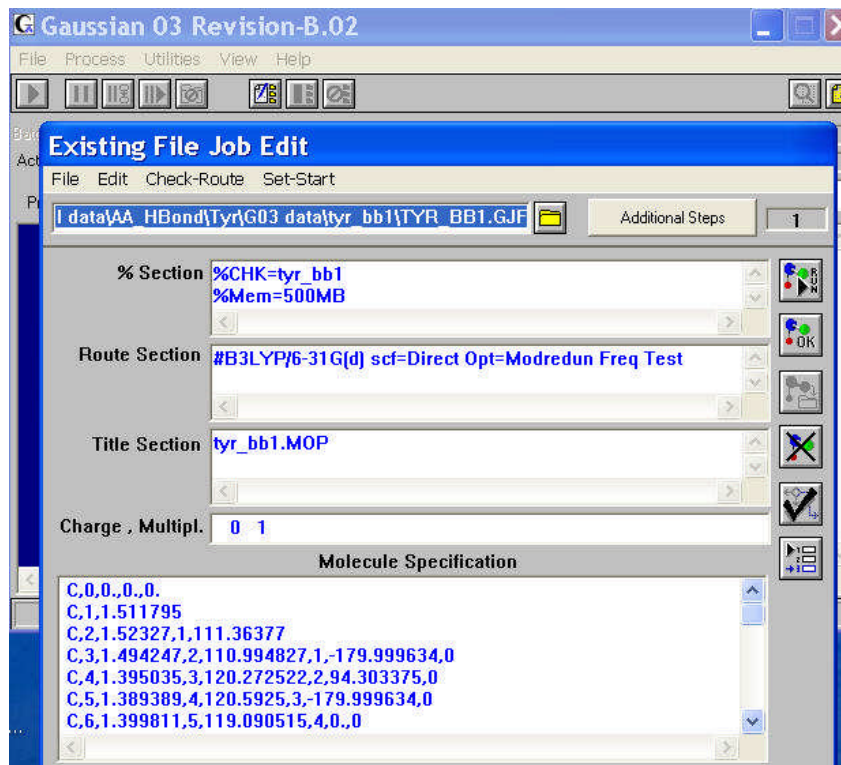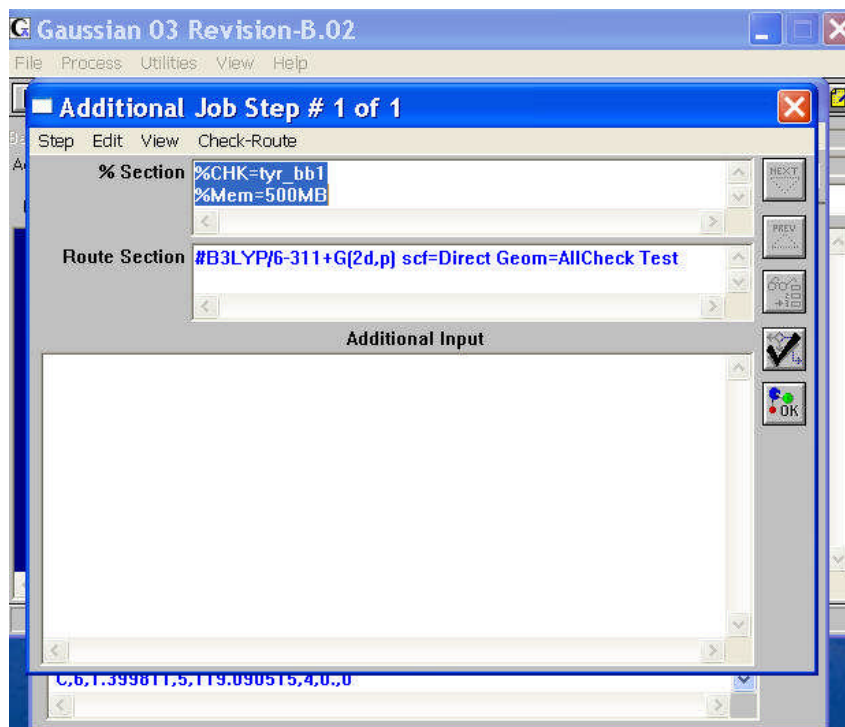
5. View the vibration for each frequency using GaussView.
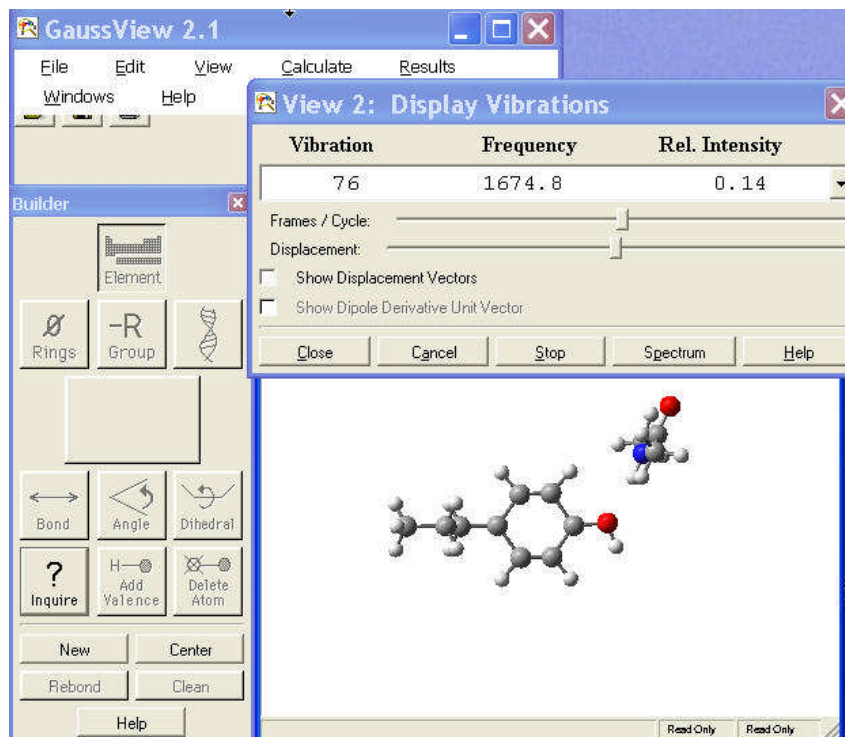
Technical Details

For difficult optimization, we had several strategies:

1. By default, a geometry optimization starts with an initial guess for the force constants. When the initial guess is poor, the geometry optimization will take long time or even fail. By calculating the force constants at the initial point using the same method and basis set, i.e., Opt=CalcFC, it often helps the convergence of the calculation.

2. If the first strategy does not work, we may optimize the structure using a lower level method and force constants will be stored in the checkpoint file. Then we read in the initial force constants from the checkpoint file and start optimization with higher level method. The keyword is Opt=ReadFC.

3. For some constrained geometry, we use keyword of Opt=Loose to lower the convergence criteria of optimization. On the other hand, Opt=Tight with higher criteria will be used for molecular system with very small force constants (low frequency vibrational modes).

4. There are definite numbers of iterations depending on the size of the molecular system. In some cases, the geometry was almost optimized but the iteration stopped by the limit of number of steps. In such condition, we need to increase the numbers of iteration by using keyword Opt=(MaxCycle=N), where N is the maximum allowed number of iterations.

Appendix C

Protocol for MD simulations of PYP

<u>MD simulations of pG</u>

1. Generating prmtop and prmcrd files for pG MD simulation using Amber 6.0:

    Three input files: (1) pG crystal structure (2phy.pdb), (2) parameter file that has force constants for the chromophore (pyp_090903.frcmod), and (3) parameter files containing internal coordinates of chromophore, Glu46 and Tyr42 (pCA_E46_Y42_pG.in). Glu46 and Tyr42 were made so that the protons on the carboxylic oxygen of Glu46 and phenolic oxygen of Tyr42 were pointing towards the phenolic oxygen of pCA chromophore while other atoms in Glu46 and Tyr42 were unchanged. The force constants for the chromophore were taken from reference paper [1] when available and were made analogous to the closest formation type when not available. Charge distributions of pG were taken from *ab initio* calculation of pG model compound with three H-bonds partners ($CH_3(CH_2)_2COOH$, C_6H_5OH, and CF_3NHCF_3) in vacuum. Three H-bonds distances were 2.59 Å, 2.66 Å, and 2.79 Å.

(a) Eight sodium and two chloride ions were introduced to the protein where the electrostatic potential was the highest or the lowest. These ions replaced the closest crystal water molecules when steric overlaps occur. There are twelve Aspartic acids and seven Glutamic acids in PYP. The pG chromophore is negatively charged whereas Glu46 is protonated so that the total number of negative charged group is nineteen. There are two Arginines, eleven Lysines, and two Histidines. Two Histidines are neutral at pH=7 according to the titration curve of PYP [2] resulting in thirteen positive charged groups in total. (The C-terminus and N-terminus also carry charges but the net charge of them is zero). Therefore, the net charge at pH=7 is –6 e. Please refer to Technical Details for reasons we added 10 ions instead of 32 ions to the protein.

(b) Then a rectangular periodic box with 6 Å constraints from the edge of the box to the nearest protein molecules in each direction was added. The box was in the dimension of 55Å × 65 Å × 61 Å containing 3884 water molecules. The volume of the box was approximately 218,075 Å. (There are approximately 4740 water molecules per protein if the protein concentration is 10 mM).

2. The system was energy minimized to relax possible strain in the molecules for 200 steps by using a steepest descent algorithm (a few minutes).

3. Then the system was simulated for 200 ps with fixed protein to optimize water molecules and ions. The "belly" command was used to fix the protein. The cutoff length of 10 Å is used for non-bonded interactions. (It took 3 days with one CPU at 1.7 GHz speed).

4. Next, the entire system was equilibrated at 0.1 K and constant pressure 1 atm for 20 ps to prevent breaking of H-bonds. The cutoff length of 10 Å is used for non-bonded interactions.

5. Last, the entire system was equilibrated at 300 K and constant pressure 1 atm for 5 ns [3]. The initial temperature is 0.1 K. (It took one week to get 1 ns MD simulations with one CPU at 1.7 GHz speed).

6. This equilibrated pG structure (3 – 5 ns) with all the water molecules was used to generate initial structures for MD simulation of pR_1, pR_2 and pR_3.

Technical Details

One or more of H-bonds between the chromophore and its partners were broken for 870 ps in pG simulation with 32 ions while H-bonds reformed within 10 ps in pG simulation with 10 ions. In addition, the simulation of the system containing 32 ions takes longer time than that of 10 ions because of larger size of the system. Therefore, we chose the system with 10 ions including 8 sodium ions and 2 chloride ions.

Control parameters

1. The control for energy minimization is named "min.in":

```
&cntrl
   imin=1, maxcyc=200, ncyc=200, ntx=1,
   ntt=1, temp0=300.0, tautp=0.2, irest=0,
   ntb=2, ntc=2, ntf=2, iwrap=1, ntp=1, pres0=1.0,
   ntpr=25, ntwx=25, ntwe=25
 &end
```

2. The control for equilibration of water molecules and ions is named "mdbelly.in":

```
&cntrl
   imin=0, ntx=1,
   ntt=1, temp0=300.0, tautp=0.2, irest=0,
   ntb=2, ntf=2, ntc=2, iwrap=1,
   ntp=1, pres0=1.0, ibelly=1,
   ntr=0, ntrx=1, cut=10.0, tempi=250.0,
   dt=0.002, ig=77751, nstlim=100000,
   ntpr=500, ntwx=500, ntwe=500
 &end
 belly
```

RES 126 4019
        END
        END

3. The control for the entire system including water, ions, and protein frozen at 0.1 K is named "mdfreeze.in":

    &cntrl
        imin=0, ntx=1,
        ntt=1, temp0=0.1, tautp=0.2, irest=0,
        ntb=2, ntf=2, ntc=2, iwrap=1,
        ntp=1, pres0=1.0,
        cut=10.0, tempi=0.1,
        dt=0.002, ig=77751, nstlim=10000,
        ntpr=500, ntwx=500, ntwe=500
     &end

4. The control for the entire system including water, ions, and protein is named "md.in":

    &cntrl
        imin=0, ntx=7,
        ntt=1, temp0=300.0, tautp=0.2, irest=1,
        ntb=2, ntf=2, ntc=2, iwrap=1,
        ntp=1, pres0=1.0,
        cut=10.0, tempi=0.1,
        dt=0.002, ig=77751, nstlim=2500000,
        ntpr=500, ntwx=500, ntwe=500
     &end

MD simulations of pR

1. Generating pR_1, pR_2, and pR_3 initial structures for pR MD simulation with Amber 6.0:

(a) Three input files: (1) eight different frames were selected from 3 ns to 5 ns of pG simulation, every 250 ps a frame. The pG chromophore from these frames was removed and saved as pR_3ns.pdb, and so forth. Each frame would be used for one simulation of pR; (2) parameter file that contains force constants for the chromophore (pyp_100103.frcmod), and (3) parameter file containing internal coordinates of chromophore: pCA_pR1.in, pCA_pR2.in, and pCA_pR3.in. (See Technical Details for more information about three pR chromophore structures). In the pdb files, the residue names TYR was replaced by TYY and GLH by GLE in order to make the protons pointing towards the phenolic oxygen of the chromophore. The atoms of CYP69 were deleted except for backbone atoms (until CB) so that pR chromophore was added. Charge distributions, bond lengths and angles of pR were taken from *ab initio* calculation of pR model compound with two H-bonds partners ($CH_3(CH_2)_2COOH$ and C_6H_5OH) in vacuum. Two H-bonds distances are 2.59 Å and 2.67 Å.

(b) Then a rectangular periodic box with 10^{-6} Å constraint from the edge of the box to the nearest water molecules of the box containing protein, 10 ions and 3884 water molecules was added in each direction. The box is in the

dimension of 56 Å × 57 Å × 61 Å containing water molecules. The volume of the box is approximately 194,712 Å^3. There were 44 − 73 water molecules added to the system including the protein, ions, and 3884 water molecules.

2. The entire system of each pR including water, ions, and proteins was energy minimized for 200 steps by using a steepest descent algorithm.

3. Next, the velocities of protein, 10 ions and 3884 water molecules from each frame of pG simulation were added to the end of the restart file generated from energy minimization step. Then the velocities of the additional water molecules were set to zero. Keep the number of the atoms at the beginning and the box size at the end of the restart file. This step was done by a Fortran90 code.

4. Eight MD simulations for each pR structure at constant temperature 300 K and constant pressure 1 atm were performed for 2000 fs, every 1 fs a step, and coordinates were printed out every 2 fs.

5. At last, eight MD simulations for each pR from the last frame of 2000 fs MD simulations were continued for 5ns, every 2 fs a step, and coordinates are print out every 1 ps.

Technical Details

The chromophore structure of pR$_1$ was modeled according to proposed model in the reference paper [4]. The structure of pR$_2$ chromophore was taken from the crystal structure of PYP at early intermediate state trapped at 149 K [5]. The chromophore of pR$_3$ was modeled as in the reference paper [6].

Control parameters

1. The control for energy minimization is named "min.in":

```
&cntrl
    imin=1, maxcyc=200, ncyc=200, ntx=1,
    ntt=1, temp0=300.0, tautp=0.2, irest=0,
    ntb=2, ntc=2, ntf=2, iwrap=1, ntp=1, pres0=1.0,
    ntpr=25, ntwx=25, ntwe=25
 &end
```

2. The control for MD simulation of the entire system including water, ions, and protein for 5 ns is named "md.in":

```
&cntrl
    imin=0, ntx=1,
    ntt=1, temp0=300.0, tautp=0.2, irest=0,
    ntb=2, ntf=2, ntc=2, iwrap=1,
    ntp=1, pres0=1.0,
    cut=10.0, tempi=300.0,
    dt=0.002, ig=77751, nstlim=250000,
    ntpr=500, ntwx=500, ntwe=500, ntwv=500
 &end
```

Appendix D

The protocol for bacterial growth and protein purification
(By Haeshin Lee, Yuhong Zeng, Beining Nie, and Qinan Bao)

Step One: Overproduction (pQE80 E. Coli, for 4 L culture)

1. Prepare the growth medium:

Ingredients for various volumes of the growth medium

Growth Medium	5 mL	30 mL	1 L	2 L	4 L
LB Broth	0.125 g	0.75 g	25 g	50 g	100 g
Ampicillin (50 mg/ml)	10 μl	60 μl	2 ml	4 ml	8 ml

Solution form of Ampicillin (50 mg/ml) is prepared by mixing 0.686 g Ampicillin with 13.72 g de-ionized water.

Add 100 g LB Broth to 4 L distilled water in a large beaker, stir well to get it dissolved. Dissolve 0.75 g LB Broth in 30 ml distilled water in a 225 ml flask. Adjust the pH of the growth medium to 7.0 with concentrated HCl or NaOH and also adjust the pH of the small volume culture.

Distribute 1 L of the growth medium to each of the short 2800 ml flasks. Cover opening of 2800 flasks and 225 ml flask with aluminum foil.

Tape a piece of autoclave paper to the aluminum foil cover and autoclave it for 1 hour. Check the autoclave paper (should show black strips after proper autoclave) to make sure that the autoclave process is completed. Also autoclave the wood applicators, beakers, column, reservoir, packing adaptors, I.E.X. and S.E.C. buffer, and centrifuge tubes.

Let the growth medium cool down to room temperature.

2. Grow the bacteria:

Add 60 μl Ampicillin (antibiotic) to 30 ml growth medium. Select six healthy, strong single colonies from a LB agar plate using sterilized wood applicators (one for each colony) to inoculate 30 ml growth medium. The antibiotic, Ampicillin, help to suppress the growth of other unwanted bacteria.

Grow bacteria in 225 ml flask at 37 °C and 210 rpm or higher overnight (~14 hours) in a shaker. (Do not cover the top with aluminum foil tightly.)

On the next day, add 2 ml Ampicillin (50 mg/ml, see table) to each flask containing 1 L LB broth and transfer 5 ml bacteria to the flasks, shaking at 37 °C with rpm at 210-220.

Monitor the exponential growth every hour until absorption reaches 0.8 at 600 nm. This step usually takes about 4 hours.

Add 1mM IPTG (1ml/L) to induce cells when absorption is 0.8 at 600 nm and monitor growth every hour until beginning of decrease (~1.2 OD at 600 nm), with continuous shaking at 37 °C (~5 hours).

3. Harvest:

Harvest cells through centrifugation at 3200 rpm for 15 minutes. After centrifugation, decant as much of the spent medium as possible.

Measure the weight of all the cells (~14 g for 4 L). Store the cells at -20°C.

4. Cell lysis:

Let the cell thaw at room temperature and resuspend the cell pellet in 50 mM sodium phosphate buffer (pH 7.2), 80ml per 4 L culture.

Distribute the solution into five or six tubes. Put these tubes in an ice bucket.

Sonicate the cells in the tubes by immersing the tip (VWR Scientific, Branson Sonifier 450) 0.25 – 0.5" in the solution, each tube for 1.5 minutes, and then change to next tube. Each tube should be sonicated for 10 times. (Parameters on the sonifier: 30 % duty cycle, output control is 5.)

5. pH adjustment

Combine the cell extract in a beaker. Increase pH to 10 (for culture large than 4L, at least 1ml NaOH is needed) with 10 M NaOH and incubate for 5 minutes at this pH. (Remember the volume of 10 M NaOH you put because a lot of precipitates produced in the next step make it difficult to measure the pH.)

Readjust pH with 10 M HCl to 7.

Spin the lysate for 15 minutes at 15,000 rpm (use a Ti 45 Beckman rotor) and continue with supernatant and check pH again (~5.9).

The supernatant was brought to ammonium sulfate 50% saturation (23.47 g per 75 ml protein solution).

Spin for 15 minutes at 15,000 rpm and continue with supernatant (use a Ti 45 Beckman rotor).

Step Two: Reconstitution from apoPYP to holoPYP

1. Preparing trans-p-coumaric acid to p-coumaric anhydride

Dissolve 1.64 g *p*CA (free acid) in 25 ml water-free DMF. Optimally, the DMF is dried over a column but it would be OK as long as it is kept sealed without contacting air.

Dissolve 2.06 g of DCC in 25 ml water-free DMF.

Combine these two solutions and stir overnight at 4 °C. p-coumaric anhydride in DMF solution is prepared, however, not 100% will change to p-coumaric anhydride. It is normal to have some precipitation.

Remove precipitates by low-speed centrifugation around 2,000-3,000 rpm for 5 minutes.

Store the supernatant at –70 °C.

2. Reconstitution of apoPYP

Add 100 µl p-coumaric anhydride at a time to apoPYP and check absorption at 446 nm until increases are no longer observed comparing to the baseline (~ 3 ml p-coumaric anhydride will be added for 4 L culture). Dilute the protein solution by ten to twenty times with 50 mM sodium phosphate buffer when measuring the UV/visible absorption.

Dialyze against 4 L ion-exchange buffer (20 mM Tris) at 4 °C and stir it with a big stirring bar. Change the buffer at least two times (totally 3 times). Use a semi-permeable membrane with molecular weight cutoff of 6,000 -

8,000. Be sure to leave at least the same free volume (leave 2 times free volume for safety) as the solution volume for the protein solution is so concentrated that a lot of water molecular will enter the dialyze bag. Collect the protein solution to a beaker (about 300 ml).

Step Three: Purification of PYP

1. Running ion-exchange column (I.E.X.)

Concentrate the PYP solution to around 100 ml by CentriPrep filters (up to 15 ml) of 10,000 MWCO at 3,000 rpm for 30 minutes at 4 °C.

Load the filtered PYP solution to I.E.X. column (resin: Q-sepharose).

After all protein binds to the column, begin to wash off the proteins. The washing buffer is NaCl with gradient concentration.

a) Prepare the ion-exchange buffer (10X, 2 L):

200mM Tris, pH 7.6 at room temperature

b) Prepare the washing buffer:

Prepare NaCl solution in 1X ion-exchange buffer, 7.31 g NaCl + 250 ml 1X I.E.X. buffer

For elution of the protein, use a gradient mixer. In the inner cylinder, add 1X I.E.X. buffer 250 ml. In the outer cylinder, add prepared NaCl solution 250 ml. Stir the inner solution for faster mixing rate.

Get the Fraction Collector ready, ~40 – 50 tubes. If the tubes are not of the same height, then rank the tubes from high to low.

Since the concentration of NaCl is increasing, the protein will be washed off faster and faster. During the washing procedure, make sure that the

solution in the inner and outer layers of the gradient mixer are at the same level.

Collect the protein fractions, about 2 minutes per tube.

Measure the purity index for each fraction and select fractions showing 0.54 or better purity index (Abs_{280}/Abs_{446}). Use the I.E.X. buffer as the background.

2. Running size-exclusion column (S.E.C.)

a) Prepare 10X gel filtration buffer 2 L for S.E.C:

1 M NaCl 116.88 g (MW 58.44)

200 mM Tris 48.456 g (MW 121.14)

10 mM EDTA 7.445 g (MW 372.24)

Adjust pH to 7.4 at room temperature or pH 7 at 4 °C

Concentrate PYP with purity index better than 0.54 to a volume around 5-10 ml. Concentrate to a volume as small as possible.

Load the concentrated PYP solution to the S.E.C.

Collect PYP fractions, about 2-3 ml per tube or about 5 minutes for each tube.

Measure the UV/visible spectrum of the protein in each tube with gel filtration buffer as the background. Collect fractions showing a purity index (Abs_{280}/Abs_{446}) of 0.49 or better.

Step Four: Final Concentration and final calculation:

Concentrate the PYP solution from S.E.C. by CentriPrep filters (up to 15 ml) of 10,000 MWCO at 3,000 rpm for 2 hours at 4 °C. Decant the waste solution once during centrifugation.

Calculate the total amount of PYP, OD, concentration and purity index of the sample and record it with a spectrum.

Appendix E

Sample Preparation

General procedure. Approximately 4 mg of PYP provides enough 0.7 IR OD (~7.5 mM for wild type) for more than $15 \times$ (2.5 μL~3.0 μL) samples. This is sufficient for averaging without exposing too much to the air before use. The PYP is dissolved into 400 μL of buffer (50 mM $KDPO_4$ in D_2O, pH=7 (pH*=7.4)) and put into a 10 kDa molecular weight cutoff filter. This is centrifuged for 10 min at 10,000 rpm (g force equal to 8169) and then 30 to 45 min at 12,000 rpm (g force equal to 11,750) in an Eppendorf 5415C centrifuge to concentrate to ~50 μL. This is followed by two more washes of 400 μL of buffer.

Sample holder. Two 15 x 2 mm BaF_2 (6000 cm^{-1} – 890 cm^{-1}) or CaF_2 (6000 cm^{-1} – 1000 cm^{-1}) windows and 12 μm spacer, diameter: d_{in}=9 mm and d_{out}=13 mm (Control the sample thickness. Improvement will be made in the future).

Appendix F

Experimental Protocol for rapid-scan FTIR of wt-PYP

Checklist:

- Warm up the laser according to laser startup procedure
- Internal MCT Detector is cooled down
- Vacuum on (EVAC. IF/ VENT SM on the FTIR)
- Temperature controller is on
- IR beam is not blocked (Check signal in OPUS-NT)
- IR signal (no sample) at the centerburst should be in the range of ± 0.75 and ADC count is around -12000 with Sample signal gain of 1
- Laser energy before sample (0.066 mJ/mm^2, 7.46 mJ on 6 mm area)
- Purge the LN2 in the sample chamber (with higher flow rate 140 (3.62 L/min, glass ball) for ~5 minute and then reduce it to ~80 (2.21 L/min, glass ball).
- Measurement parameters in OPUS are OK

Checklist at the end of day:
- Vent optics and turn off the vacuum pump
- Turn off purging nitrogen gas
- Slow down the scanner velocity to 10 kHz step by step
- Turn down the laser according to laser shutdown procedure
- Turn off the temperature controller
- Turn off the delay generators and power meter
- Remove the samples from sample chamber and clean the windows (for BaF2 windows, need to clean the windows immediately after each day's measurement)

- Back up the experimental data on CD (two copies) at the end of experiments

Measurements:
Experimental Parameters for Cary UV/Vis spectrum
Range: 210 – 700 nm
Double Beam Mode
SBW: 2.0 nm
Baseline Correction: ON
Scan Rate: 600 nm/min
Data Interval: 1.0 nm
Mode: Absorption

Experimental Parameters for standard measurement (saved at
D:\OPUS\MEAS\XPM\norm_FTIR_2cm-1.xpm).
Optic settings
　　Detector: MCT internal
　　Aperture: 6mm
　　Scanner Velocity: 10.0 kHz
　　Sample Signal Gain: 1 (with no sample ADC count is 20200)

Advanced settings
　　Spectral resolution: 2 cm^{-1}
　　Save data from 4000 cm^{-1} to 850 cm^{-1}

Acquisition settings
　　High Folding Limit: 7899.94 cm^{-1}
　　Low Folding Limit: 0.00 cm^{-1}
　　Wanted High Frequency: 6000 cm^{-1}
　　Wanted Low Frequency: 800 cm^{-1}
　　Acquisition mode: Double-sided forward backward

FT settings
　　Phase resolution: 4 cm^{-1}
　　Phase Correction Mode: Mertz
　　Apodization Function: Blackman-Harris 3-Term
　　Zero filling factor: 4

Experimental Parameters for rapid-scan FTIR (saved at
D:\OPUS\MEAS\XPM\wtPYP_rpsc_011906.xpm).
Method: RPD_4SP.TRS
Split Forward/Backward Inteferogram
Split Doublesided Inteferogram

Optic settings
　　Detector: MCT internal

Aperture: 6mm
Scanner Velocity: 200.0 kHz
Sample Signal Gain: 2

Advanced settings
Spectral resolution: 4.5 cm^{-1}
Save data from 4000 cm^{-1} to 850 cm^{-1}

Acquisition settings
High Folding Limit: 5266.62 cm^{-1}
Low Folding Limit: 0 cm^{-1}
Wanted High Frequency: 5250 cm^{-1}
Wanted Low Frequency: 800 cm^{-1}
Acquisition mode: Double sided forward backward

FT settings
Phase resolution: 6 cm^{-1}
Phase Correction Mode: Mertz
Apodization Function: Blackman-Harris 3-Term
Zero filling factor: 4

Delay Generators
All the output signals are TTL (inverted normal) signals. FTIR TKDA signal triggers the delay generator slave 1.

The TKDA signal goes high at the beginning of forward and at the beginning of backward scan. With scanner velocity of 200 KHz and spectral resolution of 4.5 cm^{-1}, we have TKDA signal of 49.4 ms time period (double of that gives 98.7 ms).

Slave 1   A = T + 13 ms
            B = A + 30 μs
            D= T + 95 ms
Slave 2   A = T + 13 ms + 278 μs
            B = A + 20 μs
            C = T + 0 ms
            D = T + 4.022 s
Add the delay time for the 2nd, 3rd, 4th measurements of 15 ms, 63 ms, 65 ms to **A**s on both delay generators.

TKDA cable of FTIR (trigger out) is connected to DG Slave 1's "To" (trigger in)
Flash trigger cable is connected to DG Slave 1's normal **AB** output
Q-switch trigger cable is connected to DG Slave 2's normal **AB** output
Rapid-scan's Trigger In cable is connected to DG Slave's normal **CD** output.

Here 278 μs is the Q-Switch delay, corresponding to 35% laser energy. The Q-Switch delay is in the range of 192 μs (100%) – 305 μs (5%). The 278μs delay

time is to give an example of how delay generators are setup.

Experimental procedure:
Measure single beam spectra of the sample holder with (1) no sample and (2) the buffer only. This will be used as background and it needs to be collected only once before starting the few-days step-scan measurements.

Daily procedure

1. Load the protein sample in a small sample holder.

2. Measure the VIS absorption spectrum of the sample using Cary 300 UV/Vis spectrometer. Insert samples into FTIR sample chamber (Bruker IFS66v).

3. Start purging with higher flow rate 140 (3.62 L/min, glass ball) for ~5 minute and then reduce it to ~80 (2.21 L/min, glass ball).

4. Set the desired temperature. Wait an extra 10 min for equilibration after the sample reached the desired temperature.

5. Measure IR absorption spectrum. Check IR absorption, Amide I OD. Take special care for the water vapor bands.

6. Change trigger mode in Opotek control software from internal to external

7. Turn on Flashlamp and Q-switch with internal triggering using the remote control box and block the laser beam

8. Start Rapid-scan measurement

9. Switch triggering mode of flashlamp from internal to external and turn on the flashlamp

10. Wait for 80 fires of flashlamp

11. Switch triggering mode of Q-switch from internal to external and turn on the Q-switch

12. When the laser beam is visible, remove the block of the laser

13. Measure one rapid scan for testing; 50 averages, test band positions, decay time.

14. Measure real rapid scan on the protein sample.

15. After rapid-scan is complete each time (when Fourier transformation starts, the frequency of TKDA has changed), change back the triggering mode of flashlamp and Q-switch to internal mode and turn on flashlamp and Q-switch

16. Set the delays and repeat the rapid-scan measurement three times.

17. Repeat collecting several sets of measurements.

18. Measure IR absorption spectrum again. The absorption of 5% protein bleaching is acceptable.

19. Measure VIS absorption spectrum of the protein sample.

Data analysis using macro programs:

1. Load all the rapid-scan measurement files in OPUS window.

2. Select "Run Macro" in the OPUS-NT "Macro" pull down menu.

3. Open the main macro program for rapid-scan data analysis.

4. Select all the single beam spectra of 0 ms delays and drag them to the blank space labeled "Rapidscan(s) D0" in the popup window.

5. Specify the data blocks for "First block to extract" and "Last block to extract".

6. Select "Calc mOD" and "Average".

7. Click "OK".

8. Select the single beam spectra of 15 ms delays and drag them to the blank space labeled "Rapidscan(s) D1" in the popup window and click "OK".

9. Select all the single beam spectra of 50 ms delays and drag them to the blank space labeled "Rapidscan(s) D2" in the popup window and click "OK".

10. Select all the single beam spectra of 0ms delays and drag them to the blank space labeled "Rapidscan(s) D3" in the popup window and click "OK".

Appendix G

Experimental Protocol for step-scan FTIR of wt-PYP

Checklist:

- Warm up the laser according to laser startup procedure
- MCT Detector is cooled down
- Vacuum on (EVAC. IF/ VENT SM on the FTIR)
- Table is floating (pressure 23.6 psi, 6 mm higher)
- Temperature controller is on
- IR filter is placed in the sample compartment. Cutoff: 1900 cm^{-1} (5% transmission).
- IR beam is not blocked (Check signal in OPUS-NT)
- IR signal (no sample) at the centerburst should be in the range of ± 0.75 and ADC count is around 22000 with Sample signal gain of 4 (AC mode)
- Laser energy before sample (0.066 mJ/mm^2, 7.46 mJ on 6 mm area)
- Test the control of sample exchanger
- Purge the LN2 in the sample chamber (with higher flow rate 140 (3.62 L/min, glass ball) for ~5 minute and then reduce it to ~80 (2.21 L/min, glass ball).
- Measurement parameters in OPUS are OK

Checklist at the end of day:

- Vent optics and turn off the vacuum pump
- Turn off purging nitrogen gas
- Turn off the high pressure nitrogen to stop floating table
- Turn down the laser according to laser shutdown procedure
- Turn off the temperature controller
- Turn off the delay generators and power meter

- Remove the samples from sample chamber and clean the windows (for BaF2 windows, need to clean the windows immediately after each day's measurement)

- Back up the experimental data on CD (two copies) at the end of experiments

Measurements:
Experimental Parameters for Cary UV/Vis spectrum
Range: 210 – 700 nm
Double Beam Mode
SBW: 2.0 nm
Baseline Correction: ON
Scan Rate: 600 nm/min
Data Interval: 1.0 nm
Mode: Absorption

Experimental Parameters for standard measurement (saved at D:\OPUS\MEAS\XPM\norm_FTIR_2cm-1.xpm).
Optic settings
    Detector: MCT Photovoltaic
    Aperture: 6mm
    Scanner Velocity: 20.0 kHz
    Sample Signal Gain: 4 (with no sample ADC count is 22000, AC mode)

Advanced settings
    Spectral resolution: 2 cm^{-1}
    Save data from 4000 cm^{-1} to 850 cm^{-1}

Acquisition settings
    High Folding Limit: 7899.94 cm^{-1}
    Low Folding Limit: 0.00 cm^{-1}
    Wanted High Frequency: 6000 cm^{-1}
    Wanted Low Frequency: 800 cm^{-1}
    Acquisition mode: Double-sided forward backward

FT settings
    Phase resolution: 4 cm^{-1}
    Phase Correction Mode: Mertz
    Apodization Function: Blackman-Harris 3-Term
    Zero filling factor: 4

Experimental Parameters for step-scan FTIR (saved at D:\OPUS\MEAS\XPM\wtPYP_stsc_010506.xpm).
Recorder settings
    Device: Internal ADC (±10 V)

Time resolution: 5 μs (200 KHz)
No. of time slices 3256 (15 ms)
Timebase: Linear Timescale

Input range: ±10 V
Repetition/Coadd. Count: 6 or 9 per mirror position (2 or 3 for each sample)
Trigger mode: External Positive Edge
Experiment recovery time: 0 ms
Stabilization delay after mirror movements: 20 ms

Optic settings
   Detector: MCT Photovoltaic
   Aperture: 6mm
   Scanner Velocity: 20.0 kHz
   Sample Signal Gain: 32 (with 11 mM wt-PYP sample ADC count is 20800, DC mode)

Advanced settings
   Spectral resolution: 4 cm^{-1}
   Save data from 1850 cm^{-1} to 930 cm^{-1}

Acquisition settings
   High Folding Limit: 1858 cm^{-1}
   Low Folding Limit: 929 cm^{-1}
   Wanted High Frequency: 1850 cm^{-1}
   Wanted Low Frequency: 930 cm^{-1}
   Acquisition mode: Single sided

FT settings
   Phase resolution: 16 cm^{-1}
   Phase Correction Mode: Mertz
   Apodization Function: Blackman-Harris 3-Term
   Zero filling factor: 4

DC output of MCT detector is connected to Detector X Input Analog Input.

The number of mirror positions is determined by spectral resolution and phase resolution (See the table below)

Phase resolution	4 cm^{-1}	6 cm^{-1}
32	**470**	**330**
16	**522**	**382**
8	**627**	**487**

Number of mirror positions at different phase and spectral resolutions with folding of 929-1858cm^{-1}

No. of spectra averaged until the noise is $\leq 10^{-5}$ OD ($\leq 10^{-4}$ OD?)

Calculation of measurement time

6 co-addition				9 co-addition			
Sample recovery time	DG Slave D	Mirror position	Meas. time	Sample recovery time	DG Slave D	Mirror position	Meas. time
2 s	0.667 s		34.8 m	2 s	0.667 s		52.5 m
4 s	1.333 s	522	1.16 h	4 s	1.333 s	522	1.74 h
6 s	2 s		1.74 h	6 s	2 s		2.61 h
8 s	2.667 s		2.32 h	8 s	2.667 s		3.48 h
2 s	0.667 s		25.5 m	2 s	0.667 s		38.2 m
4 s	1.333 s	382	50.9 m	4 s	1.333 s	382	1.27 h
6 s	2 s		1.27 h	6 s	2 s		1.91 h
8 s	2.667 s		1.70 h	8 s	2.667 s		2.55 h

Explanation (wt-PYP for example)

6 coadd/mirror positions means 2 flashes/sample.
6 coadd. $\times$ 522 mirror position $\times$ 4 s $\times$ 1/3 = 4176 s, ~70 min for one spectra series of 6 averages (plus ~5 min for FT). This is 1044 flashes/sample. One sample can be used for 4 step-scan measurements: ~4000 flashes (the absorption of 5% protein bleaching is acceptable). 100 averages: ~20 hrs. In the literature 4 – 8 cm^{-1} resolution is used.

Delay Generators
Master (Rate = 9.87 Hz, approximately 100ms)
      A = T + 2 ms
      B = A + 30 μs
Slave A = T + 2 ms + 278 μs
      B = A + 20 μs
      C = T + 0.998 ms
      D = T + 1.33 s

Flashlamp trigger cable is connected to DG Master's normal **AB** output
Q-switch trigger cable is connected to DG Slave's normal **AB** output
Step motor drive (for multiple sample changer) trigger cable is connected to DG Slave's **B** output
Step-scan's Experimental TTL trigger input is connected to DG Slave's normal **CD** output.

The 0.998ms in DG Slave's **C** is determined by: 2 ms-1.28 ms + 278 μs (Q-Switch delay, 278 μs is corresponding to 35% laser energy. The Q-Switch delay is in the range of 192 μs (100%) – 305 μs (5%). The 278 μs delay time is to give an example of how delay generators are setup. In case of 5 μs time resolved step-scan in 1.28 ms you can collect 256 background spectra.

Experimental procedure:
Measure single beam spectra of the sample holder with (1) no sample and (2) the buffer only. This will be used as background and it needs to be collected only once before starting the few-days step-scan measurements.

Daily procedure

1. Load the three samples in the small sample holders.

2. Measure the three samples' VIS absorption spectrum using Cary 300 UV/Vis spectrometer. The variation in the absorption of 3 samples should be within 0.05 OD (less than 10%) from each other.

3. Insert samples into FTIR sample chamber (Bruker IFS66v).

4. Start purging with higher flow rate 140 (3.62 L/min, glass ball) for ~5 minute and then reduce it to ~80 (2.21 L/min, glass ball).

5. Offset the DC output signal so that the baseline of interferogram is close to 0, make sure the interferogram is properly amplified (within ± 0.75, ADC counts around 22000 with no sample and sample signal gain of 4, AC mode).

6. Set the desired temperature. Wait an extra 10 min for equilibration after the sample reached the desired temperature.

7. Measure IR absorption spectrum. Check IR absorption, Amide I OD. Take special care for the water vapor bands.

8. Measure rapid scan on one sample for testing; 50 averages, test band positions, decay time.

9. Measure step scan spectra on the three samples.

10. Measure IR absorption spectrum again.

11. Measure VIS absorption spectrum of the three samples.

Data analysis using macro programs:
1. Load all the step-scan measurement files in OPUS window.

2. Select "Run Macro" in the OPUS-NT "Macro" pull down menu.

3. Open the main macro program for step-scan data analysis.

4. Select all the single beam spectra to be averaged and drag them to the blank space labeled "Stepscan(s)" in the popup window.

5. Specify the data blocks for "Last preflash block", "First block to extract" and "Last block to extract".

6. Select "Calc mOD" and "Average".

7. Click "OK".

8. Specify the data blocks for averaging in the following window according to "step-scan and rapid-scan data logarithmic averaging" scheme.

Appendix H

Operation protocol of Brilliant laser (10 Hz)

Daily startup procedure:

- Turn the control key on the front panel of the laser box to the "I" position.
- Wait about 15 minutes until the water temperature equilibrates for flashlamp operation.
- Turn on the computer. Wait until Windows has started.
- Turn on the motor driver.
- Start the OPO control software "OPOTEK". Select "Yes" to home the stepping motor.
- Press "Start Flashlamp" button to start flashlamp. Wait for 20 minutes to warm up the laser.
- Tune the OPO to a desired wavelength by selecting the target wavelength.
- Select the desired percent of energy. Check the repetition rate.
- Open the beam shutter of the laser. Wait 8 seconds before Q-switch control electronics enables the laser operation.
- Press the "Fire laser" button to start laser emission.
- Wait for 20 minutes to warm up the harmonic generators before measuring the energy and phase matching adjustment.

Shutdown procedure:

- Stop the laser emission by pressing the "stop" buttons of the Q-Switch and flashlamp on the remote control box or by pressing the "Stop laser" button on the OPO control panel on the computer and then press "Return" and "Quit" to exit the software.
- Turn off the power of the motor drive.
- Shut down the computer
- Close the beam shutter of the laser
- Turn the control key on the front panel of the laser box to the "O" position.

Appendix I

Calculation of laser energy at the sample

Basic knowledge of laser energy needed for wt-PYP sample is essential for doing FTIR rapid-scan and step-scan experiments. The proteins could be bleached with strong laser energy and they could be wasted with weak laser energy. To control the intensity of the laser, we calculated how much energy is needed for wt-PYP sample.

With a dimension of l = 12 μm thickness and d = 6 mm diameter, we get the volume of the protein in the light path

$$V = \pi \times (d/2)^2 \times l = 0.34 \text{ mm}^3 = 0.34 \text{ } \mu\text{L} \tag{1}$$

Then the number of protein molecules with c = 11 mM concentration is

$$N_{protein} = V \times c \times N_A = 0.34 \text{ } \mu\text{L} \times 11 \text{ mM} \times 6.02 \times 10^{23} = 2.25 \times 10^{15} \tag{2}$$

where N_A is the Avogadro's constant. With quantum efficiency of 0.5 (i.e. the number of photocycle initiated by absorption of one photon), the number of photon that excites $N_{protein}$ number of protein molecules is

$$n_{photon} = N_{protein} / 0.5 = 4.5 \times 10^{15} \tag{3}$$

The laser energy absorbed by the protein is

$$(I_o - I) \times \text{Area} = n_{photon} \times h \text{ } \nu \tag{4}$$

where I_o and I are the laser intensity before and after the sample, Area is the area of protein sample in the light path, h is Plank's constant, and ν is the laser frequency. The total laser energy before the sample is

$$E = I_o \times \text{Area} = N_{photon} \times h\nu \qquad (5)$$

Combining (4) and (5), the total number of photons before the sample is

$$N_{photon} = n_{photon} / (1 - I/I_o) \qquad (6)$$

According to the definition of absorbance and the visible absorption of 11 mM wt-PYP is measured to be 0.1264, we got the ratio of laser intensity before and after the sample

$$I / I_o = 10^{-A} = 10^{-0.1264} = 0.7475 \qquad (7)$$

So, combining (6) and (7), we have

$$N_{photon} = n_{photon} / (1 - I/I_o) = 4.5 \times 10^{15} / (1 - 0.7475) = 1.78 \times 10^{16} \qquad (8)$$

The laser energy of one photon at 475 nm is

$$h\nu_{475} = h \times c / \lambda_{475} = (6.63 \times 10^{-34} \text{ J·s} \times 3 \times 10^8 \text{ m/s}) / 475 \times 10^{-9} \text{ m}$$
$$= 4.19 \times 10^{-19} \text{ J} = 2.62 \text{ eV} \qquad (9)$$

Therefore, the total laser energy before the sample (5) is

$$E = N_{photon} \times h\nu = 1.78 \times 10^{16} \times 4.19 \times 10^{-19} \text{ J} = 7.46 \text{ mJ} \qquad (10)$$

With 6 mm diameter sample, the laser intensity before the sample is

$$I_o = E / \text{Area} = 7.46 \text{ mJ} / \pi \times (3 \text{ mm})^2 = 0.26 \text{ mJ/mm}^2 \qquad (11)$$

Appendix J

TKDA Signal

The TKDA signal is a square wave signal generated by the moving mirror. The signal goes high at the beginning of forward and at the beginning of backward scan of the moving mirror. In our FTIR rapid-scan measurements, we have utilized this signal to trigger the delay generators that synchronize the laser, Q-switch, and the start of the FTIR measurement. There are three parameters determining the time interval of the TKDA signal: scanner velocity (moving mirror), the spectral resolution, and acquisition mode. For example, with 240 kHz scanner velocity, 3.5 cm^{-1} spectral resolution, and double-sided forward backward acquisition mode, the period of the TKDA signal of 50.4 ms is obtained from the oscilloscope. We can calculate the actual scanning time and turning around time of the moving mirror (scanner) as follows:

The scanning time = the traveling distance / the optical velocity

The traveling distance of the scanner = 2* 0.9 / 3.5 cm^{-1} = 0.514 cm

The physical velocity of the scanner = 633 nm x 240 kHz / 2 = 7.596 cm/s, where 633 nm is the wavelength of the He-Ne laser. The optical velocity is twice the physical velocity, i.e., 15.192 cm/s.

The scanning time = 0.514 cm / 15.192 cm/s = 0.0338 s = 33.8 ms

The turning around time = 50.4 ms – 33.8 ms = 16.6 ms

With the same scanner velocity and spectral resolution, and single-sided forward backward acquisition mode, the period of the TKDA signal is 37 ms from the oscilloscope. We calculate the actual scanning time and turning around time again:

The traveling distance of the scanner = 0.9 / 3.5 cm^{-1} + 0.9 / 16 cm^{-1} = 0.313 cm, where 16 cm^{-1} is the phase resolution.

The optical velocity of the scanner is still 15.192 cm/s.

The scanning time = 0.313 cm / 15.192 cm/s = 0.0206 s = 20.6 ms

The turning around time = 37.0 ms – 20.6 ms = 16.4 ms, this is consistent with the calculated time for double-sided forward backward acquisition mode because of the same scanner velocity.

Appendix K

Calculating Crystal Contact in Proteins

Crystal contacts were obtained in three steps: (1) load the PDB file of a protein in its unit cell using SwissPdbViewer program (also called Deep View), (2) build crystallographic symmetry according to its space group, e.g., $P6_3$ space group for Photoactive yellow protein, (3) calculate the crystal contacts between the protein and its neighboring proteins in the unit cell within given distance. More details about how to calculate the crystal contacts using SwissPdbViewer can be found at the following website: http://molvis.sdsc.edu/protexpl/xxtl.htm. The procedure above is to get the number of crystal contacts in proteins. Noncovalent Bond Finder in Advanced Explorer of Protein Explorer program was utilized to see which atoms are in contact and what kind of interaction it is. Since it is very easy to miss some contacts when distance increases and it is also time-consuming to get quantitative values of each contact in Protein Explorer, I wrote a Fortran90 code to calculate the properties of each atom in contact. These properties are the actual contact distances, angles, dihedral angles, and to which atoms it forms angles. An example of Fortran90 output calculating the crystal contacts of PYP (2phy.pdb) within 3.2 Å is attached to this appendix.

Proteins		Crystal information				Number of crystal contacts		
& Inter-mediates	MW (Da)	*pdb ID, resolution (Å)*	*unit cell size a, b, c (Å)*	unit cell angle α, β, γ	*Space Group*	≤ 2.8 Å	≤ 3.0 Å	≤ 3.5 Å
PYP, pG		2PHY[7] 1.40	66.90, 66.90, 40.80	90, 90, 120	$P6_3$	2	17	36
PYP, pR	13.9k	2PYR[8] 1.90	66.90, 66.90, 40.80	90, 90, 120	$P6_3$	2	17	36
PYP, pB		2PYP[9] 1.90	66.90, 66.90, 40.80	90, 90, 120	$P6_3$	4	17	33
bR		1M0M[10] 1.43	61.07, 61.07, 110.10	90, 90, 120	$P6_3$	2	2	12
M1		1M0M[10] 1.43	61.07, 61.07, 110.10	90, 90, 120	$P6_3$	2	2	12
bR	28.2k	1CWQ[11] 2.25	61.08, 61.08, 110.40	90, 90, 120	$P6_3$	5	6	23
M2		1CWQ[11] 2.25	61.08, 61.08, 110.40	90, 90, 120	$P6_3$	2	6	33
Rhodopsin	39.0k	1U19[12] 2.20	96.68, 96.68, 150.20	90, 90, 90	$P4_1$	2	2	7
Deoxy Mb		1BZP[13] 1.15	35.15, 31.04, 64.59	90, 105.58, 90	$P2_11$	11	15	36
MbCO	17.2k	1BZR[13] 1.15	35.15, 31.04, 64.59	90, 105.58, 90	$P2_11$	4	14	34
Deoxy Hb		1A3N[N/A] 1.80	62.65, 82.43, 53.53	90, 99.61, 90	$P2_11$	2	16	50
HbCO	64.5k	1M9P[14] 2.10	56.0, 58.70, 175.20	90, 90, 90	$P2_12_12_1$	0	6	23

Crystal contacts of PYP within 3.2 Å (The unit of distance is angstrom and the unit of angles and dihedrals is degree.)

```
 1  ATOM      8  CE   MET     1      21.421 -13.325  -1.738
    ATOM   1104  OH   TYR    76      22.550 -11.032   -.245

Distance = 2.960

 2  ATOM     40  N    PHE     6      20.420  -4.021   -.689
    ATOM   1084  OD1  ASP    71      18.704  -6.479   -.715

    ASP 71  OD1      PHE  6  N       PHE  6  CA       PHE  6  C

Distance = 2.998    Angle = 117.625   Dihedral =  118.535

 3  ATOM    102  ND2  ASN    13      34.435   3.142   1.552
    ATOM   1038  O    LEU    40      36.247   1.156   2.504
```

```
       LEU 40  O        ASN 13  ND2      ASN 13  CG       ASN 13  CB

Distance = 2.852   Angle = 126.066   Dihedral =  -12.721

  4  ATOM   131  NZ  LYS    17      33.626  -6.831   -.006
     ATOM  1034  OD2 ASP    34      36.343  -8.000    .424

       ASP 34  OD2      LYS 17  NZ       LYS 17  CE       LYS 17  CD

Distance = 2.989   Angle =  94.864   Dihedral =  -62.861

  5  ATOM   254  OD2 ASP    34      30.557   8.000 -19.976
     ATOM  1169  NZ  LYS    17      33.274   6.831 -20.406

       ASP 34  OD2      LYS 17  NZ       LYS 17  CE       LYS 17  CD

Distance = 2.989   Angle =  94.864   Dihedral =  -62.861

  6  ATOM   290  O   LEU    40      30.653  -1.156 -17.896
     ATOM  1160  ND2 ASN    13      32.465  -3.142 -18.848

       LEU 40  O        ASN 13  ND2      ASN 13  CG       ASN 13  CB

Distance = 2.852   Angle = 126.066   Dihedral =  -12.721

  7  ATOM   373  O   GLY    51      16.544  -8.676 -22.396
     ATOM  1214  CA  THR    90      17.211 -11.096 -24.325

Distance = 3.166

  8  ATOM   373  O   GLY    51      16.544  -8.676 -22.396
     ATOM  1218  OG1 THR    90      17.382  -8.795 -25.080

       GLY 51  O        THR 90  OG1      THR 90  CB       THR 90  CA

Distance = 2.814   Angle = 106.656   Dihedral =   28.386

  9  ATOM   373  O   GLY    51      16.544  -8.676 -22.396
     ATOM  1220  N   MET    91      14.940 -10.337 -24.022

       GLY 51  O        MET 91  N        MET 91  CA       MET 91  C

Distance = 2.824   Angle = 129.369   Dihedral =   49.773

 10  ATOM   392  OD2 ASP    53      22.323 -11.316 -24.340
     ATOM  1205  N   ASN    89      20.641 -13.905 -24.514

       ASP 53  OD2      ASN 89  N        ASN 89  CA       ASN 89  C

Distance = 3.092   Angle = 119.118   Dihedral =   36.163

 11  ATOM   417  NE2 GLN    56      23.115  -9.506 -28.028
     ATOM  1190  O   LYS    78      22.074 -10.277 -30.668

       LYS 78  O        GLN 56  NE2      GLN 56  CD       GLN 56  CG
```

```
Distance = 2.941    Angle = 115.603    Dihedral = -164.909

12  ATOM    508  O    PRO    68      12.360   8.072 -21.656
    ATOM   1271  NZ   LYS   123      10.354  10.135 -21.433

    PRO 68  O        LYS123  NZ      LYS123  CE       LYS123  CD

Distance = 2.886    Angle = 106.823    Dihedral =  -53.349

13  ATOM    528  O    ASP    71      12.130  11.820 -20.233
    ATOM   1271  NZ   LYS   123      10.354  10.135 -21.433

    ASP 71  O        LYS123  NZ      LYS123  CE       LYS123  CD

Distance = 2.726    Angle = 122.987    Dihedral =   49.649

14  ATOM    531  OD1  ASP    71      14.963  12.959 -21.115
    ATOM   1252  N    PHE     6      13.692  15.674 -21.089

    ASP 71  OD1      PHE  6  N       PHE  6  CA       PHE  6  C

Distance = 2.998    Angle = 117.625    Dihedral =  118.525

15  ATOM    553  OE1  GLU    74       8.179  11.932 -10.010
    ATOM   1145  OE1  GLN    99       6.969   9.851  -8.557

Distance between carboxylate carbon and oxygen in protein  D(OD1/OE1)
= 1.251   D(OD2/OE2) = 1.254
Two carbonyl oxygen, Distance = 2.812

16  ATOM    554  OE2  GLU    74       7.246  11.298 -11.908
    ATOM   1007  NH1  ARG   124       4.239  12.194 -11.669

    GLU 74  OE2      ARG124  NH1     ARG124  CZ       ARG124  NE

Distance = 3.147    Angle = 103.403    Dihedral =  147.807

17  ATOM    577  OH   TYR    76      20.829  14.013 -20.645
    ATOM   1244  CE   MET     1      22.250  11.889 -22.138

Distance = 2.960

18  ATOM    585  O    LYS    78      19.937  13.978 -10.268
    ATOM   1137  NE2  GLN    56      19.790  15.265  -7.628

    LYS 78  O        GLN 56  NE2     GLN 56  CD       GLN 56  CG

Distance = 2.941    Angle = 115.593    Dihedral = -164.899

19  ATOM    662  N    ASN    89      22.362  10.923  -4.114
    ATOM   1128  OD2  ASP    53      20.961  13.674  -3.940

    ASP 53  OD2      ASN 89  N       ASN 89  CA       ASN 89  C

Distance = 3.092    Angle = 119.149    Dihedral =   36.159

20  ATOM    671  CA   THR    90      18.215   9.357  -3.925
```

```
        ATOM     1120  O    GLY     51       15.786    9.990   -1.996

Distance = 3.166

21  ATOM      675  OG1  THR     90       16.308   10.656   -4.680
    ATOM     1120  O    GLY     51       15.786    9.990   -1.996

        GLY 51  O         THR 90  OG1      THR 90  CB       THR 90  CA

Distance = 2.814    Angle = 106.652   Dihedral =    28.392

22  ATOM      677  N    MET     91       16.422    7.770   -3.622
    ATOM     1120  O    GLY     51       15.786    9.990   -1.996

        GLY 51  O         MET 91  N        MET 91  CA       MET 91  C

Distance = 2.824    Angle = 129.395   Dihedral =    49.779

23  ATOM      696  N    GLU     93       10.404    5.165   -6.342
    ATOM     1141  O    GLN     99        9.065    6.422   -3.862

        GLN 99  O         GLU 93  N        GLU 93  CA       GLU 93  C

Distance = 3.086    Angle = 122.866   Dihedral =   117.913

24  ATOM      754  OH   TYR     98       12.554    2.396  -28.313
    ATOM     1277  CG   ARG    124       10.739    4.843  -29.201

Distance = 3.173

25  ATOM      754  OH   TYR     98       12.554    2.396  -28.313
    ATOM     1278  CD   ARG    124       11.470    4.448  -30.482

Distance = 3.177

26  ATOM      758  O    GLN     99       10.094   -4.640  -24.262
    ATOM     1228  N    GLU     93        9.675   -6.428  -26.742

        GLN 99  O         GLU 93  N        GLU 93  CA       GLU 93  C

Distance = 3.086    Angle = 122.849   Dihedral =   117.941

27  ATOM      762  OE1  GLN     99       12.016   -1.110  -28.957
    ATOM     1185  OE1  GLU     74       14.423   -1.117  -30.410

Distance between carboxylate carbon and oxygen in neighboring protein
D(OD1/OE1) = 1.251   D(OD2/OE2) = 1.255
Two carbonyl oxygen, Distance = 2.812

28  ATOM      957  NZ   LYS    123       13.954   -3.900   -1.033
    ATOM     1074  O    PRO     68       13.171   -6.668   -1.256

        PRO 68  O         LYS123  NZ       LYS123  CE       LYS123  CD

Distance = 2.885    Angle = 106.844   Dihedral =   -53.340

29  ATOM      957  NZ   LYS    123       13.954   -3.900   -1.033
```

245

```
       ATOM    1081  O    ASP    71      16.301  -4.595    .167

       ASP 71  O         LYS123  NZ      LYS123  CE      LYS123  CD

   Distance = 2.726    Angle = 122.991    Dihedral =   49.680

30  ATOM     963  CG   ARG    124      9.564  -6.879  -8.801
    ATOM    1116  OH   TYR     98      8.352  -9.674  -7.913

   Distance = 3.173

31  ATOM     964  CD   ARG    124      9.587  -7.709 -10.082
    ATOM    1116  OH   TYR     98      8.352  -9.674  -7.913

   Distance = 3.177

32  ATOM     967  NH1  ARG    124      8.441  -9.768 -11.669
    ATOM    1017  OE2  GLU     74      6.161 -11.924 -11.908

       GLU 74  OE2       ARG124  NH1     ARG124  CZ       ARG124  NE

   Distance = 3.147    Angle = 103.386    Dihedral =  147.829
```

REFERENCE

1.   Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules.* J. Am. Chem. Soc., 1995. **117**: p. 5179-5197.

2.   Bao, Q., *The photocycle of photoactive yellow protein.* 2004, Oklahoma State University. p. 99.

3.   Saam, J., et al., *Molecular dynamics investigation of primary photoinduced events in the activation of rhodopsin.* Biophys. J., 2002. **83**: p. 3097-3112.

4.   Xie, A., et al., *Glu46 donates a proton to the 4-hydroxycinnamate anion chromophore during the photocycle of photoactive yellow protein.* Biochemistry, 1996. **35**(47): p. 14671-14678.

5.   Genick, U.K., et al., *Structure at 0.85 A resolution of an early protein photocycle intermediate.* Nature, 1998. **392**(6672): p. 206-209.

6.   Groenhof, G., et al., *Photoactivation of the photoactive yellow protein: why photon absorption triggers a trans-to-cis Isomerization of the chromophore in the protein.* J. Am. Chem. Soc., 2004. **126**(13): p. 4228-4233.

7.   Borgstahl, G.E., D.R. Williams, and E.D. Getzoff, *1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore.* Biochemistry, 1995. **34**(19): p. 6278-6287.

8.   Perman, B., et al., *Energy transduction on the nanosecond time scale: early structural events in a xanthopsin photocycle.* Science, 1998. **279**(5358): p. 1946-1950.

9.   Genick, U.K., et al., *Structure of a protein photocycle intermediate by millisecond time-resolved crystallography.* Science, 1997. **275**(5305): p. 1471-1475.

10. Lanyi, J.K. and B. Schobert, *Crystallographic structure of the retinal and the protein after deprotonation of the Schiff base: the switch in the bacteriorhodopsin photocycle.* J. Mol. Biol., 2002. **321**(4): p. 727-737.
11. Sass, H.J., et al., *Structural alterations for proton translocation in the M state of wild-type bacteriorhodopsin.* Nature, 2000. **406**(6796): p. 649-653.
12. Okada, T., et al., *The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure.* J. Mol. Biol., 2004. **342**(2): p. 571-583.
13. Kachalova, G.S., A.N. Popov, and H.D. Bartunik, *A steric mechanism for inhibition of CO binding to heme proteins.* Science, 1999. **284**(5413): p. 473-476.
14. Patskovska, L.N., et al., *COHbC and COHbS crystallize in the R2 quaternary state at neutral pH in the presence of PEG 4000.* Acta Cryst. D Biol Crystallogr., 2005. **61**: p. 566-573.

VITA

Beining Nie

Candidate for the Ph.D. Degree

Thesis: PROTON TRANSFER IN PHOTOACTIVE YELLOW PROTEIN

Major Field: Physics

Biographical:

Personal Data: Born in Kashi, Xinjiang, China, on October 19, 1974.

Education: Graduated from Jinzhou High School, Jinzhou, Liaoning Prov., China in July 1993; received Bachelor of Science degree in Thermal Physics Engineering from Beijing University of Aeronautics and Astronautics, Beijing, China, in July 1997; withdrew from the Graduate College of Beijing University of Aeronautics and Astronautics, Beijing, China, in December 1999 and entered Ph.D. program in Physics at Oklahoma State University in Jan 2000; received Master of Science degree in Physics from Oklahoma State University, Stillwater, OK in December 2002. Completed the requirements of the Ph. D. degree with a major in physics at Oklahoma State University in July 2006.

Experience: Worked as sales manager at Beijing Xin Ke Yu Computer Company for half year in 1998; employed as teaching assistant in the Department of Physics, Oklahoma State University, Spring and Fall 2000, Fall 2005 to present; employed as research assistant in the Department of Physics, Oklahoma State University, Spring 2001 to summer 2005; employed as a graduate research assistant (GRA) in the theoretical biophysics division, Los Alamos National Laboratory during the summer 2002 and 2004.

Professional Membership: Member of Biophysical Society since 2001.

Name: Beining Nie   Date of Degree: July, 2006

Institution: Oklahoma State University   Location: Stillwater, Oklahoma

Title of Study: PROBING HYDROGEN BONDING INTERACTIONS AND
                        PROTON TRANSFER IN PROTEINS

Pages in Study: 247   Candidate for the Degree of Doctor of Philosophy

Major Field: Physics

Scope and Method of Study: Hydrogen bonding is a fundamental element in protein structure and function. Breaking a single hydrogen bond may impair the stability of a protein. It is therefore important to probe dynamic changes in hydrogen bonding interactions during protein folding and function. Time-resolved Fourier transform infrared spectroscopy is highly sensitive to hydrogen bonding interactions. However, it lacks quantitative correlation between the vibrational frequencies and the number, type, and strength of hydrogen bonding interactions of ionizable and polar residues. We employ quantum physics theory based *ab initio* calculations to study the effects of hydrogen bonding interactions on vibrational frequencies of Asp, Glu, and Tyr residues and to develop vibrational spectral markers for probing hydrogen bonding interactions using infrared spectroscopy. In addition, proton transfer process plays a crucial role in a wide range of energy transduction, signal transduction, and enzymatic reactions. We study the structural basis for proton transfer using photoactive yellow protein as an excellent model system. Molecular dynamics simulation is employed to investigate the structures of early intermediate states. Quantum theory based *ab initio* calculations are used to study the impact of hydrogen bond interactions on proton affinity and proton transfer.

Findings and Conclusions: Our extensive density function theory based calculations provide rich structural, spectral, and energetic information on hydrogen bonding properties of protonated side chain groups of Asp/Glu and Tyr. We developed vibrational spectral markers and 2D FTIR spectroscopy for structural characterization on the number and the type of hydrogen bonding interactions of the COOH group of Asp/Glu and neutral phenolic group of Tyr. These developments greatly enhance the power of time-resolved FTIR spectroscopy as a major experimental tool for structural characterization of functionally important intermediate states of proteins. Our MD simulations on the structures of photoactive yellow protein (PYP) and its photoproducts provide deep insight into the much-debated structural nature of chromophore photo-isomerization. Our simulation data of three plausible photoproducts provide a specific guide on how to use time-resolved infrared spectroscopy to identify which photoproduct is formed after photoisomerization. Our data also indicate that hydrogen bond with Tyr42 may be crucial to increase the proton affinity of the chromophore and trigger the intra-protein proton transfer during PYP photoreceptor activation.

ADVISER'S APPROVAL: _____Aihua Xie_____