VIBRATIONAL BAND ASSIGNMENT AND

ELECTROSTATIC PROPERTIES OF

BIOMOLECULES BASED ON *AB*

*INITIO* COMPUTATIONAL
STUDIES


By

BEINING NIE

Bachelor of Science

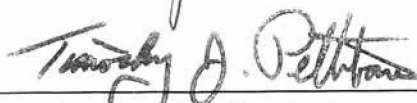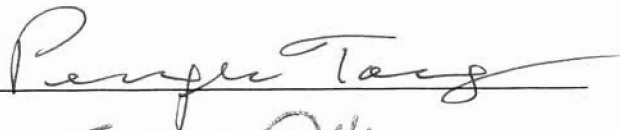Beijing University of Aeronautics & Astronautics

Beijing, P. R. China

1997


Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 2002

VIBRATIONAL BAND ASSIGNMENT AND

ELECTROSTATIC PROPERTIES OF

BIOMOLECULES BASED ON *AB*

*INITIO* COMPUTATIONAL

STUDIES

Thesis Approved:

_____

Thesis Adviser

_____

_____

_____

Dean of the Graduate

# ACKNOWLEDGEMENTS

I especially appreciate the advice, support and encouragement given by Dr. Aihua Xie during past two years. Her guidance, patience and kindness accompany thoroughly my courses study, research, and my thesis. I would like to thank Dr. Penger Tong and Dr. Bruce Ackerson for serving on my committee and reviewing this thesis.

I wish to extend special thanks for the friendship, help and support of professors, visiting faculty, graduate students, department secretaries and support staff, especially including Dr. Wouter D. Hoff, Lorand Kelemen, Brandy J. White, Ye Xiong, Kai Wang, Zeynep Gunay, Yong Yang, Susan Cantrell, Cindi Raymond and Warren Grider.

Finally, I would like to thank my parents, my sister's and my brother's family, and my husband for their understanding, love and encouragement.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# Nomenclature

IR — infrared

DFT — density functional theory

DMSO — dimethyl sulfoxide

RMSD — root mean square derivation

a.u. — atomic unit

pCA-ME-*trans* — protonated *p*-coumaric acid methyl ester in *trans* form

FTIR — Fourier transform infrared

H-bond — hydrogen bond

PYP — photoactive yellow protein

bR — bacteriorhodopsin

SR-I — sensory rhodopsin I

R — rhodopsin

D — aspartic acid

E — glutamic acid

pCA — *p*-coumaric acid

MVA — 4-Methylvaleric acid

SCRF — Self-Consistent Reaction Field

PCM — Polarized continuum Model

# CHAPTER I
# INTRODUCTION

## 1.1 Background

Quantum mechanical studies of biomolecules have been rapidly developed recently by enhanced computational power and improvement of computational methods (2, 3, 4, 5, 6). Our research interests are mainly focused on biomolecules including individual amino acids, the basic building block of proteins and p-coumaric acid, the chromophore of photoactive yellow protein. These molecules contain 20-40 atoms. It is now feasible to perform numerical *ab initio* calculations on molecules with 20 - 40 atoms and more than one hundred of electrons on a 1GHz personal computer. This allows us to study important properties of biomolecules from the first principles of quantum mechanics, including geometry optimization, hydrogen bonding interactions, vibrational force constant and electrostatic properties. One of our main goals of *ab initio* studies of biomolecules is vibrational band assignment for time-resolved FTIR spectroscopy to probe structural dynamics of proteins.

Vibrational band assignment is a complex, time consuming, and costly process. The general procedure is outlined as below. Vibrational frequencies are calculated based on a set of force constants, masses, and normal mode analysis (7, 8, 9). Traditionally, the typical values of force constant were obtained from vibrational studies for a specific chemical bond of simple, small molecules. Such typical values of various chemical bonds serve as initial values in a trial and error process to find the actual force constants. For each set of force constants,

vibrational frequencies are computed based on normal mode analysis. Such frequencies are compared with the experimental data. If the frequencies do not match with experimental data in a specific spectral region, this set of force constants must be modified. This process will be continued until the computed vibrational frequencies match with experimental vibrational bands. Since vibrational data are often collected in a limited spectral region, this set of force constants is still not guaranteed to be correct. The next test of the accuracy of force constants is by the use of isotopic labeling, namely replace a specific atom(s) with its isotope ($^1H \rightarrow {}^2D$, $^{12}C \rightarrow {}^{13}C$, $^{14}N \rightarrow {}^{15}N$, $^{16}O \rightarrow {}^{18}O$). Force constants of a molecule are independent of atomic weight of a particular atom while vibrational frequencies do shift upon isotopic labeling. So, a set of correct force constants should give correct vibrational frequencies of the isotopically labeled molecule as well. With traditional method for vibrational band assignment, it is necessary to use isotopically labeled compound, which can be chemically challenging and costly. We will use *ab initio* method for vibrational band assignment. A great advantage is that no trial and error is needed and the frequencies will be calculated straightforward based on structures.

There are a number of *ab initio* methods to use for *ab initio* studies. It is important to evaluate the accuracy of specific *ab initio* method to specific need. To evaluate the accuracy of the performed computations, we emphasize the importance to compare computational data with experimental data. The computational data may provide us valuable information of biomolecules such as band assignment of vibrational spectroscopy for structural and kinetic studies of

2

proteins. Researchers have been assigning infrared vibrational modes of biomolecules to structural motions for decades (10, 11, 12). For instance, to identify a band in the infrared (IR) spectrum by experimental approach, one has to prepare isotopically labeled molecules, observe the peak shift, make empirical calculations of vibrational frequencies with force constants, and conclude that what kind of vibration generates this peak. At the same time, it is very convenient and timesaving to assign these bands for specific biomolecules by *ab initio* computational studies. Since no analytical solution is available for systems with two or more electrons, approximation is necessary. There are two basic quantum theories for *ab initio* calculation, Hartree-Fock theory and density functional theory (13, 14, 15) (see chapter II for further information). However, computational results make sense only when evaluated them via comparing to experimental data. In particular, calculations of vibrational frequencies offer strong control over reliable evaluation of computational methods. In addition, without *ab initio* quantum mechanical calculations, vibrational band assignment is a trial and error process and requires isotopic labeling, which can be chemically challenging, costly, and time consuming.

## 1.2 Objectives

Quantum mechanical computation is a valuable tool for us to examine the properties of biomolecules while experimental approaches have limitations. Our research goals are:

Goal I. Band assignments of biomolecules: a free chromophore in solvent, protonated *p*-coumaric acid methyl ester in *trans* form, model

3

compound of the chromophore of photoactive yellow protein, a solvent molecule (dimethyl sulfoxide) and one amino acids (L-alanine at low pH).

Goal II. Conformational effect, hydrogen bonding effect and solvent effect on the vibrational frequencies; electrostatic properties (charge distribution and electrostatic fields) of biomolecules.

Goal III. Band assignment of the chromophore in the ground state of functional intermediates of photoactive yellow protein.

In order to achieve optimal computational speed and good computational accuracy, we test four methods, HF/STO-3G, HF/3-21G, HF/6-31G(d), and B3LYP/6-31G(d) by comparing vibrational frequency calculations to the experimental data. Then we choose the appropriate computational methods to achieve these goals.

# CHAPTER II

# BAND ASSIGNMENT OF BIOMOLECULES USING *AB INITIO* METHODS

## 2.1 *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:

$$H\psi(\vec{r}) = E\psi(\vec{r}) \tag{2.1}$$

where H is the Hamiltonian operator, E is the eigenenergy of the particle, $\psi$ is the eigenwavefunction (16). 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 (13):

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

$$\tag{2.1}$$

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:

$$H = \sum_{i=1}^{N} \frac{p_i^2}{2m} + \frac{1}{4\pi\varepsilon_0} \frac{1}{2} \sum_{i=1,i\neq j}^{N} \sum_{j=1}^{N} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} - \frac{1}{4\pi\varepsilon_0} \sum_{n=1}^{K} \sum_{i=1}^{N} \frac{Z_n e^2}{|\vec{r}_i - \vec{R}_n|} \qquad (2.2)$$

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 (2.2) 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 (2.2).

Hartree-Fock theory or self-consistent field method was first proposed by Douglas Hartree to find the approximate wave functions of a multielectron atom and then modified by Vladimir Fock (17). 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 (2.2) is taken into account by integrating the repulsion term. This gives the average effect of the repulsion, but not the explicit repulsion interaction (14). 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 (17, 14). Gaussian functions have the general form:

$$g(\alpha, \vec{r}) = c x^n y^m z^l e^{-\alpha r^2} \qquad (2.3)$$

where $\vec{r}$ is composed of x, y, z. The quantities n, m, l are integers. $\alpha$ is a constant determining the size of the function (15).

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 wave function (18). Following on the work of Kohn and Sham, the approximate functionals employed by current DFT methods partition the electronic energy into four terms:

$$E = E^T + E^V + E^J + E^{XC} \qquad (2.4)$$

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 (15). 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 (14). 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 (2, 3, 15).

We used the commercial software Gaussian98 to perform the *ab initio* calculations. Gaussian98 can be used to perform geometry optimization, calculations of vibrational frequency, electrostatic field, charge distribution, and system energy. There are many semi-empirical and *ab initio* methods available in Gaussian98. Semi-empirical methods in Gaussian98 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 *ab initio* methods, both Hartree-Fock and DFT are employed in our studies.

## 2.2 Geometry optimization

The initial chemical structure of the molecule is created first using the software ChemDraw, next imported into another software Chem3D to generate the internal coordinates, i.e., bond length, bond angle, and dihedral angle, and then to optimize the geometry using semi-empirical method — PM3, based on

energy minimization. The resulting coordinates are imported into Gaussian98 and geometry is optimized using a specific *ab initio* method by minimizing the energy iteratively. Based on the finally optimized geometry, we may perform the following computations: force constants for vibrational frequency calculations, the ground state energies for fine structures, electrostatic potential maps, etc. Hence, geometry optimization is a crucial step to perform further calculations.

To illustrate applications of various *ab initio* methods, we show the results of *ab initio* studies of a small molecule — dimethyl sulfoxide (DMSO) as an example. Table I shows the computed ground state energies and computational time for DMSO calculated with the HF/STO-3G, HF/3-21G, HF/6-31G(d), and B3LYP/6-31G(d) methods (The first three methods are based on Hartree-Fock theory and B3LYP is based on density function theory. Among these four methods, HF/STO-3G uses the smallest number of basis functions, HF/3-21G employs larger number of basis functions than HF/STO-3G, and HF/6-31G(d) and B3LYP/6-31G(d) employ larger number of basis functions than HF/3-21G. The number of the basis functions is dependent of the size of the system). The energy difference arises from different methods used. We calculate root-mean-square derivation (RMSD) of the Cartesian coordinates of optimized structure for HF/STO-3G, HF/3-21G, and HF/6-31G(d) from the Cartesian coordinates for B3LYP/6-31G(d). HF/STO-3G method yields very small RMSD of 0.124Å compared to the resolution of crystal structure (usually one to two angstroms) by X-ray diffraction experimental method and it is the least time consuming. Therefore, HF/STO-3G method that gives high enough accuracy can be used as

9

an optimal method for geometry optimization calculations. As we will see later, however, this method is not accurate enough for vibrational frequency calculations.

**Table I. The ground state energy of DMSO**

| Method | Energy (a.u.) | RMSD (Å) | CPU time |
|---|---|---|---|
| HF/STO-3G | -545.19 | 0.124 | 1m10s |
| HF/3-21G | -548.68 | 0.062 | 1m19s |
| HF/6-31G(d) | -551.54 | 0.024 | 3m19s |
| B3LYP/6-31G(d) | -553.19 | 0.000 | 18m6s |

1a.u. = 2619.6kJ/mol (See Appendix I), m = minute, s = second.

## 2.3 Searching for an optimal method for vibrational frequency calculations

We test the above four *ab initio* methods to calculate vibrational frequencies of protonated *p*-coumaric acid methyl ester in *trans* form (pCA-ME-*trans*) in DMSO solvent using 1GHz Personal Computers in order to find the
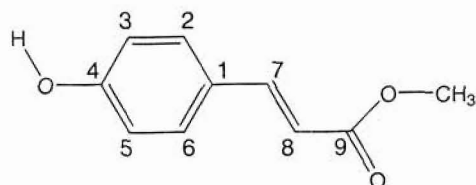


Figure 2.1 The chemical structure of the lowest energy conformation of pCA-ME-*trans*.

optimal method for vibrational frequency calculations. The chemical structure of pCA-ME-*trans* is shown in Figure 2.1. Here *trans* denotes the protons are on opposite sides of C7=C8 bond. We choose

pCA-ME-*trans* to study because it has structural importance as a model compound of photoactive yellow protein chromophore. There are five double bonds in this molecule, three in the phenolic ring, one in carbonyl group (C=O), and one in the central C=C bond. The frequencies of C=C and C=O are generally in the region of 1800 cm$^{-1}$ - 1450 cm$^{-1}$ (10). The experimental Fourier transform infrared (FTIR) spectrum of pCA-ME-*trans* shows 5 prominent IR bands in this

region, with peak positions located at 1703 cm$^{-1}$, 1632 cm$^{-1}$, 1605cm$^{-1}$, 1588cm$^{-1}$, and 1515 cm$^{-1}$ (Figure 2.2).

The *ab initio* calculations of vibrational frequencies are performed in three major steps. First, the chemical structure of the molecule is created in ChemDraw, and then imported into Chem3D for structural optimization using the semi-empirical method — PM3, as mentioned before. Second, the structure is imported to Gaussian98 where it is optimized in vacuum and then in a solvent using an *ab initio* method in Gaussian98 software. Finally, the force constants are computed based on the optimized structure and then the frequency calculations are performed. The output structures of pCA-ME-*trans* from Gaussian98 and vibrational modes are viewed using GaussianView software.

Before comparing to experimental FTIR data, computed frequencies have to be scaled to reduce known systematic errors. For instance, neglecting of electron correlation based on Hartree-Fock theory results in overestimated frequencies by about 10% - 12% (15). Since there is no default scale factor for HF/STO-3G method, we use 0.8214 as a scale factor by optimizing the positions of calculated frequencies to best match the experimental data. As for HF/3-21G, HF/6-31G(d), and B3LYP/6-31G(d) methods, the scale factors are 0.9085, 0.8929, and 0.9613, respectively, recommended by the book "Exploring Chemistry with Electronic Structure Methods" (15). These scale factors are allowed to be changed in order to optimally match the experimental data (15). Therefore, we use the scale factors 0.8979 for HF/3-21G method, 0.8844 for HF/6-31G(d) method, and 0.9961 for B3LYP/6-31G(d) method to optimally match

the experimental data as shown in Figure 2.2. The intensities were normalized to the third peak in all cases and the experimental data used are fitted results.

By comparing the computed frequencies to the experimental FTIR data, we found that B3LYP/6-31G(d) method shows excellent agreement whereas the other three methods — HF/STO-3G, HF/3-21G, HF/6-31G(d) give less good accuracy (see Figure 2.2). Therefore, we use B3LYP/6-31G(d) method for vibrational frequency calculations henceforth unless otherwise stated.
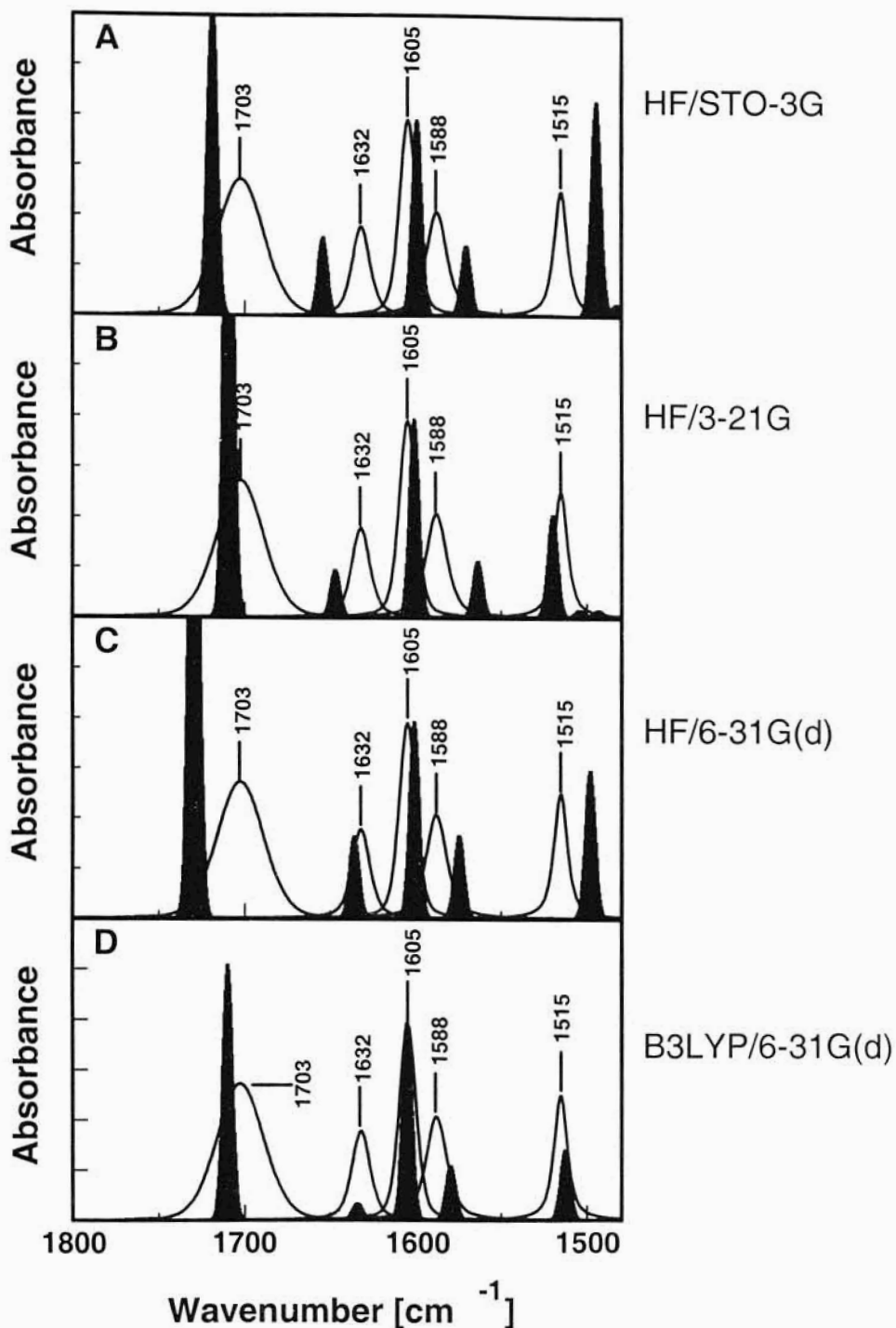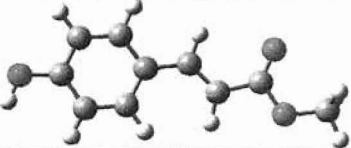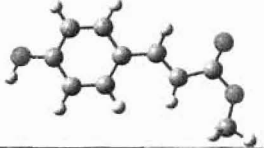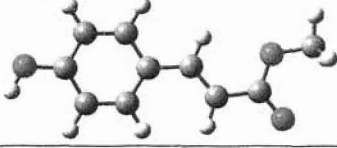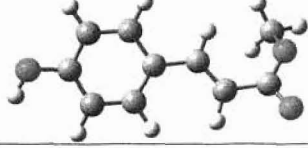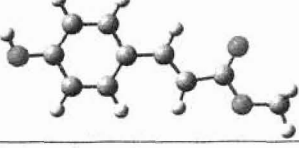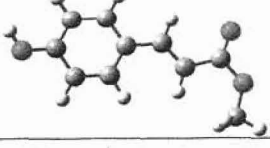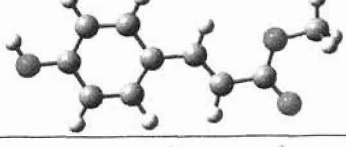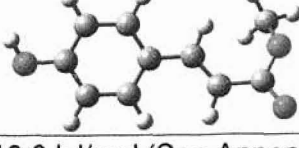
Figure 2.2 Comparisons of four *ab initio* methods for vibrational frequency calculations of pCA-ME-*trans* in DMSO to the experimental IR absorption spectrum. The vibrational bands from Gaussian98 calculations are depicted in shaded bands with *ab initio* methods of HF/STO-3G (A), HF/3-21G (B), HF/6-31G(d) (C) and B3LYP/6-31G(d) (D). The fitted IR absorption spectrum of pCA-ME-*trans* from FTIR measurement is shown in (A-D). The scale factor is 0.8214 for HF/STO-3G, 0.8979 for HF/3-21G, 0.8844 for HF/6-31G(d), and 0.9661 for B3LYP/6-31G(d).

## 2.4 Importance of conformations

Groups bonded only by a single bond can undergo rotation of motions about that bond with respect to each other. The temporary molecular geometries that result from rotations of groups about single bonds are called conformations of the molecule (19). Rotational freedom around three single bonds (C4 – OH, C8 – C9, and C9 – O) of pCA-ME-*trans* (Figure 2.1) gives rise to eight ($2^3 = 8$) conformations (see Table II). Each conformation has its own conformational energy, respectively. The relative populations of these conformations are determined by their ground state energies based on Bolzmann distribution. Therefore, when free molecules are in solvent, the conformations that have the lowest energies will be dominating rather than all conformations being present. So, it is important to find the conformations with low energies in the cases where molecules have freedom to rotate.

We calculated and compared conformational energies of these eight conformations of pCA-ME-*trans*, as shown in Table II. pCA-ME-*trans* has two predominant conformations. According to Boltzmann distribution, these two conformations occupy 47.4% and 41.8%, another two occupy 5.8% and 5.0%, contributing to 99.98% of the total population at room temperature of 298K, and the rest four conformations contribute trivial. The energy values of these conformations are calculated in DMSO using B3LYP/6-31G(d) method. In addition, the population is computed based on Boltzmann distribution Equation (2.5),

## Table II. Calculated energy of *p*CA-ME-*trans*
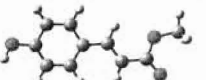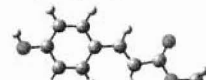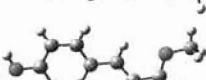
| | Molecule | Energy (a.u.) | Energy (kJ/mol) | Population |
|---|---|---|---|---|
| 1 |  | -612.74980 | 0.52* | 4.998% |
| 2 |  | -612.74372 | 16.45 | 0.008% |
| 3 |  | -612.75182 | -4.77 | 41.831% |
| 4 |  | -612.73579 | 37.22 | 0.000% |
| 5 |  | -612.74994 | 0.16 | 5.776% |
| 6 |  | -612.74389 | 16.01 | 0.010% |
| 7 |  | -612.75194 | -5.08 | 47.377% |
| 8 |  | -612.73580 | 37.20 | 0.000% |

1a.u. =2619.6 kJ/mol (See Appendix I), *— with respect to -612.750 a.u.

The calculated frequencies of the five double bond vibrational modes of four populated conformations of pCA-ME-*trans* are shown in Table III. The overall frequencies represent the calculations averaged over four conformations according to their populations. The frequency of Band III (1604 cm$^{-1}$) changes by

16

only 0.3 cm$^{-1}$ or less among the four conformations. Therefore, the Band III is regarded not sensitive to pCA-ME-*trans* conformations. The frequencies of Band II, IV and V change by 2.6, 2.2 and 2.1 cm$^{-1}$, respectively, these three bands are rather weakly sensitive to pCA-ME-*trans* conformations. In contrast, the frequency of Band I changes by 6.6 cm$^{-1}$, and shows the largest sensitivity to pCA-ME-*trans* conformations. The scale factor used here is 0.9659 and the vibrational frequencies were calculated in DMSO using B3LYP/6-31G(d) method.

**Table III. Vibrational frequencies of pCA-ME-*trans* in DMSO**

| pCA-ME-*trans* conformations | Relative Population | Band I (cm$^{-1}$) | Band II (cm$^{-1}$) | Band III (cm$^{-1}$) | Band IV (cm$^{-1}$) | Band V (cm$^{-1}$) |
|---|---|---|---|---|---|---|
|  | 5.0% [a] | 1716.3 | 1634.9 | 1604.9 | 1577.6 | 1511.3 |
|  | 41.8% [b] | 1709.7 | 1635.1 | 1605.2 | 1579.8 | 1513.4 |
|  | 5.8% [c] | 1715.5 | 1636.4 | 1605.2 | 1577.9 | 1511.2 |
|  | 47.4% [d] | 1709.8 | 1633.9 | 1604.8 | 1579.4 | 1512.6 |
| Overall $\nu_{cal}$ | N/A | 1710.4 | 1634.6 | 1605.0 | 1579.4 | 1512.8 |
| Experiment $\nu_{exp}$ | N/A | 1703.0 | 1632.0 | 1604.7 | 1587.6 | 1515.0 |
| $\Delta\nu = \nu_{cal} - \nu_{exp}$ | N/A | 7.1 | 2.6 | 0.3 | -8.2 | -2.2 |
| $|\Delta\nu|/\nu_{exp}$ | N/A | 0.42% | 0.16% | 0.02% | 0.52% | 0.15% |

[a] is the 1st conformation, [b] is the 3rd conformation, [c] is the 5th conformation and [d] is the 7th conformation in Table II. The scale factor for calculated frequencies is 0.9659. The first four rows are the calculated vibrational frequencies after scaling.

Since pCA-ME-*trans* in solution occupies different conformations according to their populations, we may simulate the infrared absorption spectrum using following equation (2.6)

$$\varepsilon_{calc} = \sum_{i=1}^{4} P_i(v)\varepsilon_i(v) \tag{2.6}$$

and compare the averaged overall frequencies to the experimental data. The largest $\Delta v$ is 8.2 cm$^{-1}$. This variance is only 0.52% of the measured frequency. These results demonstrate that the force constants from Gaussian98 calculations using B3LYP/6-31G(d) are highly accurate.

The comparison between calculated frequencies of four conformations of pCA-ME-*trans* and the corresponding experimental frequencies is shown in Figure 2.3. Here we use a scale factor 0.9659 for the overestimates of B3LYP/6-31G(d) method. In this way, we include the contributions of all populated conformations and it shows excellent agreement to the experimental data. So,
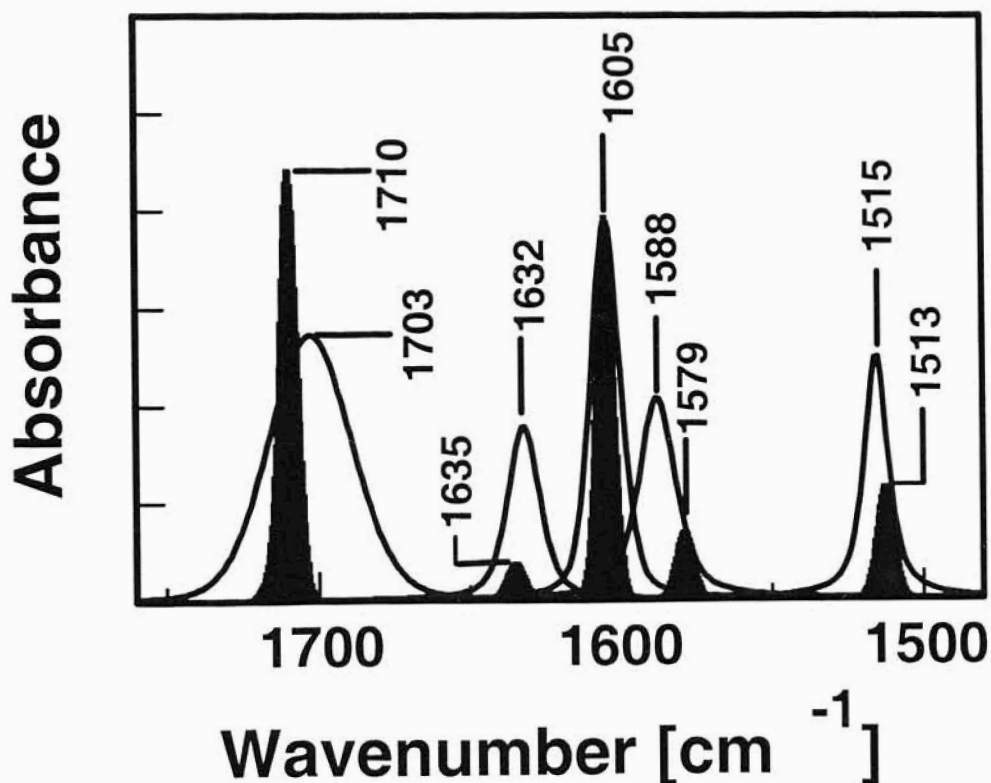


Figure 2.3 IR absorbance spectra: fitted experimental data of pCA-ME-*trans* in DMSO and computed data of the mixture of four calculated conformation according to their populations (shaded). The scale factor is 0.9659.

## 2.6 Band assignment and vibrational modes of pCA-ME-*trans*

We know from Figure 2.3 that the calculation and the experimental data are in excellent agreement and that we use GaussianView software to visualize the vibrational modes for each calculated vibrational frequency. Then we may assign the bands of the first conformation of pCA-ME-*trans* by connecting the vibrational modes to the experimental data. The first conformation is in the same configuration as the chromophore in photoactive yellow protein. Figure 2.4 shows the vibrational modes of five bands in pCA-ME-*trans*. The vibrational mode of



1712 cm$^{-1}$ (C=O stretching)

1634 cm$^{-1}$ (C7=C8 stretching)

1604 cm$^{-1}$ (*s* ring vibration)

1577 cm$^{-1}$ (*as* ring vibration)

1511 cm$^{-1}$ (*as* ring vibration)
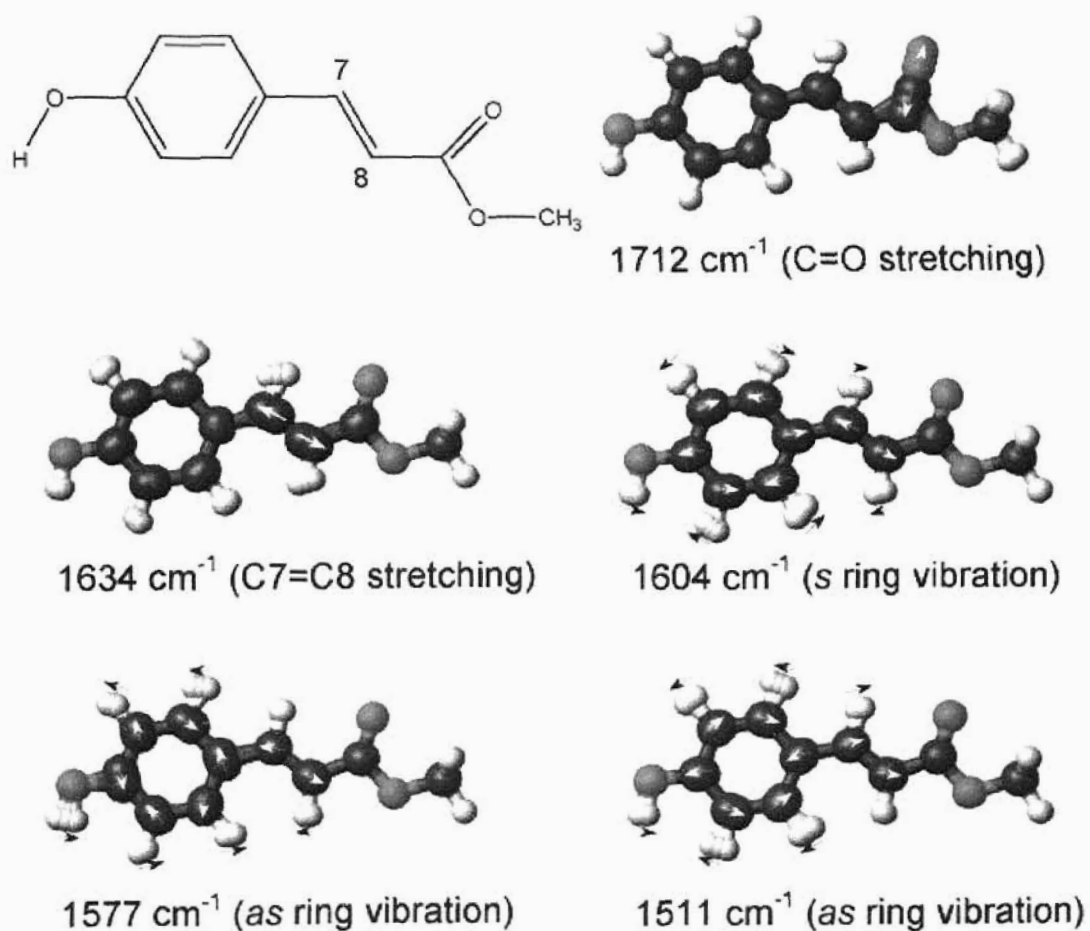
Figure 2.4 Five vibrational modes of pCA-ME-*trans*. Arrows indicate the motions of the atoms.

Band I at 1712 cm$^{-1}$ is assigned to C=O stretching, Band II at 1634 cm$^{-1}$ to C7=C8 stretching, Band III at 1604 cm$^{-1}$ to symmetric ring vibration, Band IV at 1577 cm$^{-1}$ and Band V at 1511 cm$^{-1}$ to asymmetric ring vibrations. These movements are the major components of the vibration, which may also involve displacement of other atoms. The vibrational modes of other vibrational frequencies are beyond our discussion.

## 2.7 Band assignment of other biomolecules
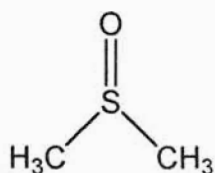
We have also performed band assignment for two other molecules, DMSO and L-alanine monomer at pH*=0.8 in D$_2$O.

Figure 2.5 The chemical structure of DMSO.

DMSO is a simple molecule we are very interested in and is a polar solvent. The structure of DMSO is shown in Figure 2.5. We calculated the vibrational frequencies of DMSO in DMSO solvent using B3LYP/6-31G(d) method and the scale factor is 0.9550. Figure 2.6 shows the IR absorbance spectrum and the second derivatives spectrum of calculated (blue) and experimental data (black) of DMSO. The reason we take the second derivatives is that there are wide,
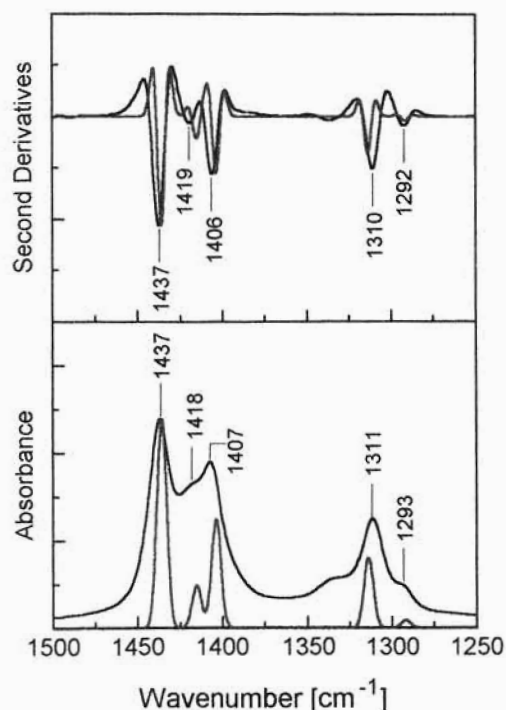


Figure 2.6 The absorbance spectra (bottom) and the second derivatives spectra (top) of DMSO. The experimental data is shown in black line and the calculation in blue. The scale factor is 0.9550.

overlapping bands in the infrared spectrum and their positions are very clear in the second derivatives spectrum. Based on *ab initio* calculations, the bands at 1437 $cm^{-1}$, 1418 $cm^{-1}$, 1407 $cm^{-1}$, 1311 $cm^{-1}$, 1293 $cm^{-1}$ are assigned to various types of symmetric and asymmetric bending motions of C-H on both methyl groups. Other motions are not mentioned here.
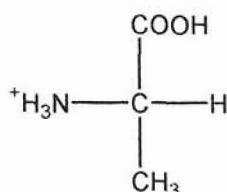
Figure 2.7 The chemical structure of L-alanine at low pH.

Proteins are composed of 20 different amino acids (1). Monitoring structural changes of amino acids at active sites is crucial to probe structural functions of proteins (20, 21), therefore, it is also very important to do band assignment of amino acids. Here we take alanine as an example.

The chemical structure of alanine at low pH is shown in Figure 2.7. It has a carboxyl group (COOH), an amino group ($NH_3^+$), and a sidechain group $CH_3$. It is the second smallest of the 20 amino acids. We calculated the vibrational frequencies of L-alanine in $D_2O$ solvent using B3LYP/6-31G(d) method and the scale

Figure 2.8 The absorbance spectra (bottom) and the second derivative spectra (top) of alanine at low pH in $D_2O$. The black line represents the experimental data while the blue line is for the calculation. The scale factor is 0.9613.

factor is 0.9613. Figure 2.8 shows the IR absorbance spectrum and the second derivatives spectrum of calculated (blue) and experimental data (black) of L-alanine at low pH in $D_2O$. The band at 1729 cm$^{-1}$ is assigned to the carbonyl stretching of COOD group, 1464 cm$^{-1}$ to C-H bending of CH$_3$ group, 1405 cm$^{-1}$ to C-H bending of CH$_3$ group coupled with OD bending of COOD group, 1377 cm$^{-1}$ to C-H bending of the center and 1322 cm$^{-1}$ and 1269 cm$^{-1}$ to C-H bending of the center coupled with OD bending of COOD group.

We demonstrated band assignment of L-alanine at low pH as an example. We can perform vibrational band assignment in the same way for other amino acids.

In conclusion, we choose optimal methods in terms of different criteria to achieve our goals: HF/STO-3G method for geometry optimization calculations, B3LYP/6-31G(d) method for vibrational frequency calculations, and HF/6-31G(d) method for electrostatic potential calculations as we will see in Chapter III.

# CHAPTER III

# ELECTROSTATIC PROPERTIES OF CHARGED

# AND IONIZABLE GROUPS

Embedding a charged group inside a protein in a nonpolar (low-dielectric) environment is energetically unfavorable. Therefore, most charged groups in a protein are solvent-exposed. The buried charges Glu46 is fully conserved in the PYP family (22) and Glu134 is conserved in rhodopsin family (23), indicating their structural and functional importance (21). In this chapter, we report our *ab initio* computational studies on the electrostatic properties of charged and ionizable groups, namely hydrogen bonding interactions, charge distribution, and electrostatic field.

## 3.1 Introduction of hydrogen bonds

Hydrogen bond (H-bond) is important for protein structure and functions. H-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 3.1 shows various types of interactions that involve H-bonds. The distance between two electronegative atoms in H-bonding ranges from 2.7Å to 3.2Å (17, 24). H-bonds can be found in biological systems for structural stability. For instance, $\alpha$ Helix structure is considerably stabilized by many internal H-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. $\beta$ Sheet structure is formed by many hydrogen bonding

interactions between adjacent polypeptide chains (1). H-bonds are also formed between amino acid side chains.
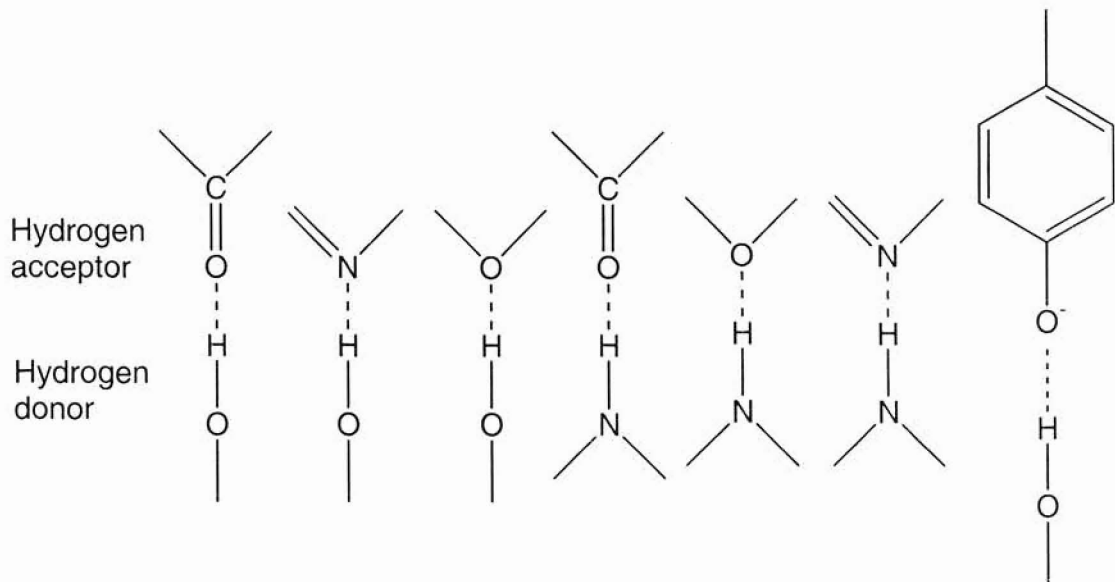


Figure 3.1 H-bonds in biological system. Dashed lines represent H-bonds. The hydrogen acceptor is usually oxygen or nitrogen (1).

H-bonds are weaker than covalent bonds. The H-bond dissociation energy is about 15 - 30 kJ/mol (1). We take a water molecule as an example to see how H-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 of another.

This attraction and partial sharing of electrons lead to the formation of H-bond. H-bonds are individually weak, but collectively they have a very significant influence on the three-dimensional structure of proteins (1). In addition, hydrogen bonding interactions are important for proton transfer in proteins.

## 3.2 Hydrogen bonds effect on vibrational frequency of C=O stretching
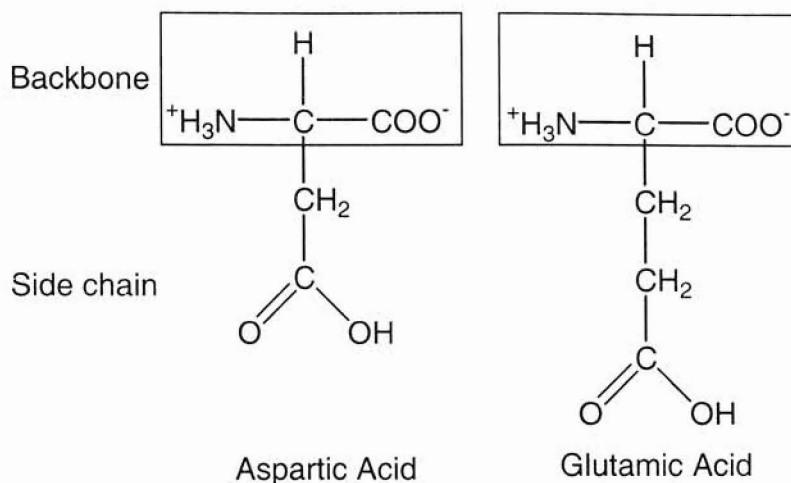


Figure 3.2 The chemical structures aspartic acid and glutamic acid. The groups for forming backbones are indicated in the rectangular square, and the rests are side chains.

Protonation/deprotonation of aspartic acid (Asp) side chain or glutamic acid (Glu) side chain (see Figure 3.2) are extensively studied for their important roles in proton transfer for energy transduction and signal transduction in four photoreceptor proteins — bacteriorhodopsin (bR) (25), sensory rhodopsin I (SR-I) (26), rhodopsin (R) (27), photoactive yellow protein (PYP) (20). Proton transfer in these photoreceptor proteins is facilitated by hydrogen bonding interactions, which involve the carboxyl group Asp and Glu as a proton donor or as a proton acceptor (28). The vibrational frequency of C=O stretching mode of protonated carboxyl group (COOH) is in the range of 1700 cm$^{-1}$ – 1765 cm$^{-1}$, depending on the hydrogen bonding interactions with neighboring group. Vibrational mode of

26
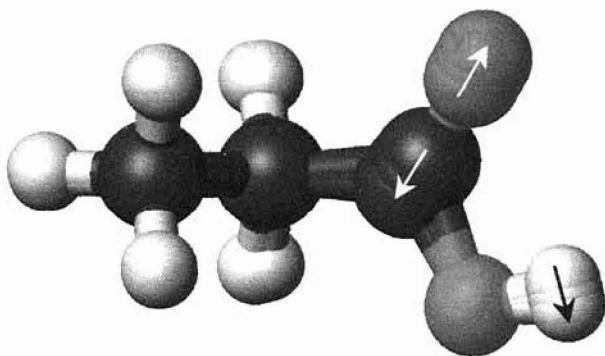
Figure 3.3 Vibrational mode of carboxylic C=O stretching. The arrows indicate the vibrational direction.

C=O stretching is depicted in Figure 3.3. Table I shows the vibrational frequencies of protonated Asp(D) or Glu(E) side chains in different proteins based on FTIR data and the number of H-bonds formed between the corresponding Asp/Glu and other groups. In the case of PYP, the vibrational frequency at 1737 cm$^{-1}$ in the ground state is assigned to the protonated Glu46 (E46). Deuteration of the carboxylic group COOH→COOD, leads this peak down-shifted to 1726 cm$^{-1}$ (20). D85, D96, D115, D204, D212 are ordered according to bR sequence. M is an important intermediate state of the bR photocycle for light driven transmembrane proton pump.

**Table I. FTIR band position of C=O stretching vs. number of H-bonds**

| Protein | COOH group | Type of H-bonds (Code) | IR Freq. $\lambda$ (cm$^{-1}$) | No. of H-bonds |
|---------|------------|------------------------|--------------------------------|----------------|
| M412 | D85 | A000 | 1760[25] | 0 |
| R | D83 | A000 | 1767[32] | 0 |
| bR568 | D115 | B100 | 1742[30] | 1 |
| bR568 | D96 | B100 | 1740[30] | 1 |
| M412 | D212 | D100 | 1738[25] | 1 |
| PYP | E46 | B010 | 1737[29] | 1 |
| M412 | D115 | B100 | 1734[30] | 1 |
| R | E122 | B100 | 1734[32] | 1 |
| bR568 | E204 | C110 | 1700[31] | 2 |

E represents glutamic acid and D represents aspartic acid. M412 denotes that peak absorption of bR's M intermediate state is at 412nm and bR568 denotes that peak absorption of bR ground state is at 568nm. See Table II for hydrogen bonding types for code. Refer to Appendix II for H-bonds formation for each residue.

For a COOH group, the vibrational frequency of C=O stretching mode is sensitive to hydrogen bonding interactions of its carbonyl oxygen with a H-bond donor(s) and/or of its hydroxyl group with a H-bond acceptor(s). Table I shows that without hydrogen bonding, C=O stretching frequency is around 1760 cm$^{-1}$, with one H-bond, the C=O vibrational frequency is characteristically down-shifted to 1734 -1742 cm$^{-1}$, and with two H-bonds, it is further down-shifted to 1700 cm$^{-1}$.

In quantum mechanics, vibrations are treated as harmonic oscillators, where their 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 (33):

$$-\frac{\hbar^2}{2m}\frac{d^2\varphi}{dx^2} + \frac{C}{2}x^2\varphi = E\varphi \qquad (3.1)$$

where $\hbar$ is plank constant, m is the mass of the atom, $C$ is the force constant, and E is the eigenenergy. By solving this equation, we found the eigenenergy has discrete values:
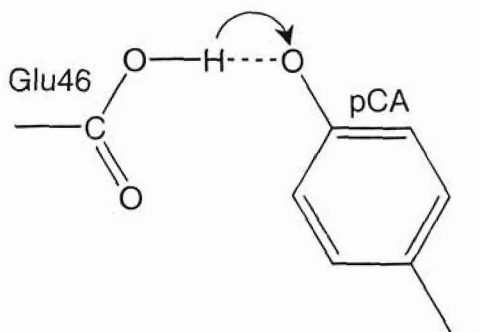


Figure 3.4 The pCA chromophore of PYP forms an H-bond with Glu side chain.

$$E_n = (n + \frac{1}{2})h\nu , \text{ n} = 0,1,2,3,... \quad (3.2)$$

The vibrational frequency is proportional to the square root of force constant $\nu = \frac{1}{2\pi}\sqrt{\frac{C}{m}}$, where $\nu$ is the frequency. Triple bonds have larger force constants and exhibit higher vibrational frequencies than double and single bonds, and double bonds have larger force constants and display higher vibrational frequencies than single bonds. When a H-bond is formed in proteins between a COOH group and another group (see Figure 3.4), the electrostatic interaction between the H-bond and the C=O bond will weaken the force constant of C=O (34, 35). As a result, the frequency of C=O stretching will be lower.
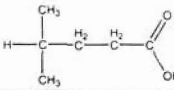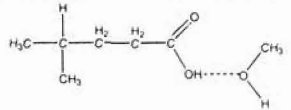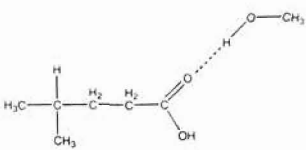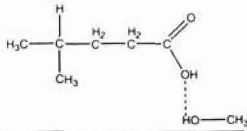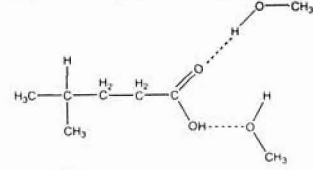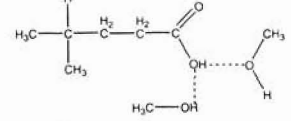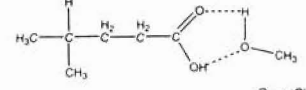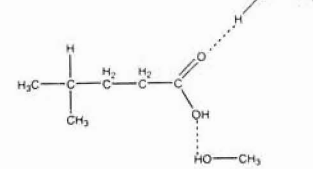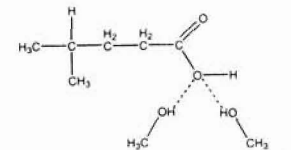
We chose 4-Methylvaleric acid (MVA) to model the side chain of Glu and examine the H-bond effect on energy and vibrational frequency of the carboxyl group. The chemical structure of MVA is shown in the first column and the second row of Table II. We use B3LYP/6-31G(d) method to calculate the vibrational frequencies of this molecule in vacuum. Then we add methanol molecules to form H-bonds and calculate the vibrational frequencies and conformational energies. Each oxygen atom has two lone pairs of electrons; thus, it may form two H-bonds with methanol molecules, as illustrated in Table II.

Dashed lines indicate the H-bonds. Without hydrogen bonding interaction, the vibrational frequency of carboxylic C=O stretching is at 1776 cm$^{-1}$ (A000). Upon formation of one H-bond between the carbonyl oxygen (B100) or carboxyl hydrogen (B010) and hydroxyl group of methanol, this frequency is down-shifted to 1745 cm$^{-1}$. However, when the carboxyl oxygen (B001) forming H-bond with methanol, it is up-shifted to 1787 cm$^{-1}$. According to the energies of B010, B100 and B001, B001 occupy the highest energy state indicating this formation is the least likely. A, B, C and D denotes that the number of H-bond is zero, one, two, and three respectively, whereas three numbers denote the positions of H-bonds. The frequency is down-shifted to 1712 - 1720 cm$^{-1}$ for C110 and C110* types, to 1756 cm$^{-1}$ for C011 and C101 types because of the presence of 001 type of H-bond, and it is up-shifted to 1799 cm$^{-1}$. As for three H-bonds, this frequency is down-shifted to 1693 cm$^{-1}$ without 001 type of H-bond while it is down-shifted to 1725 cm$^{-1}$ and 1734 cm$^{-1}$ with 001 type of H-bond. Overall, these calculations are reasonably matching the corresponding experimental frequencies as shown in Table III.

**Table III.** Vibrational frequency of C=O stretching of protonated carboxylic group
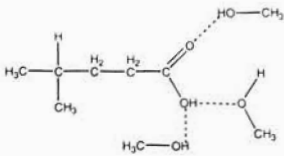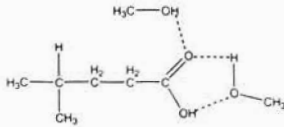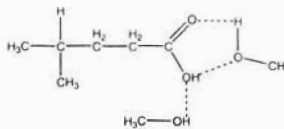
|  | FTIR Freq. (cm$^{-1}$) | Calculated Freq. in vacuum (cm$^{-1}$) |
|---|---|---|
| with 0 H-bond | 1760 -1767 | 1776 |
| with 1 H-bond | 1734 - 1742 | 1745 - 1747 |
| with 2 H-bond | 1700 - 1715 | 1713 |

# Table II. Frequency shifts vs. number of H-bonds

| Model | Type of H-bond (Code) | No. of H bond | Scaled Freq. $\nu_p(cm^{-1})$ | $\Delta\nu$ $(cm^{-1})$ | E (a.u.) | $\Delta E$ (kJ/mol) |
|---|---|---|---|---|---|---|
| *(structure)* | A000 | 0 | 1775.5 | 0 | -386.337 | 0 |
| *(structure)* | B010 | 1 | 1744.5 | 31.0 | -386.355 | -47.15 |
| *(structure)* | B100 | 1 | 1746.7 | 28.8 | -386.350 | -34.05 |
| *(structure)* | B001 | 1 | 1786.8 | -11.3 | -386.347 | -26.20 |
| *(structure)* | C110 | 2 | 1712.8 | 62.7 | -386.368 | -81.21 |
| *(structure)* | C011 | 2 | 1755.7 | 19.8 | -386.366 | -75.97 |
| *(structure)* | C110* | 2 | 1720.1 | 55.4 | -386.361 | -62.87 |
| *(structure)* | C101 | 2 | 1756.5 | 19.0 | -386.358 | -55.01 |
| *(structure)* | C002 | 2 | 1799.7 | -24.2 | -386.354 | -44.53 |

The energy of methanol alone is -115.714 a.u. The energies are the subtraction of the energies of methanols involved from the total energy. 1a.u. = 2619.6 kJ/mol (See Appendix I). The scale factor is 0.9613.

## Table II. Frequency shifts vs. number of H-bonds (continued)

| Model | Type of H-bond (Code) | No. of H bond | Scaled Freq. $v_p$(cm$^{-1}$) | $\Delta v$ (cm$^{-1}$) | E (a.u.) | $\Delta E$ (kJ/mol) |
|---|---|---|---|---|---|---|
|  | D111 | 3 | 1725.3 | 50.2 | -386.378 | -107.40 |
|  | D210* | 3 | 1692.9 | 82.6 | -386.373 | -94.31 |
|  | D111* | 3 | 1733.6 | 41.9 | -386.371 | -89.07 |

The Energy of methanol alone is -115.714 a.u. The energies are the subtraction of the energies of methanols involved from the total energy. 1a.u. = 2619.6 kJ/mol (See Appendix I). The scale factor is 0.9613.

## 3.3 Solvent effect on vibrational frequency of protonated carboxylic group

The *ab initio* calculations mentioned above are performed in vacuum while the IR experimental data of protonated carboxylic group are obtained in dielectric medium. Therefore, we have to consider solvent effect when we perform *ab initio* calculations. There is a family of models for systems in non-aqueous solution, called Self-Consistent Reaction Field (SCRF) methods. These methods all model the solvent as a continuum of uniform dielectric constant $\varepsilon$: the reaction field. The solute is placed into a cavity within the solvent. We select Tomasi's Polarized continuum Model (PCM) which defines the cavity as the union of a series of interlocking atomic spheres to calculate the model system in solvents (15, 5, 6).

Solvents can be classified into three types based on their ability to form H-bonds: nonpolar aprotic solvent that can not form H-bond with the model system; polar aprotic solvent that can be H-bond acceptor, protic solvent that can serve H-bond donor and acceptor (19), provided that the model system has hydrogen acceptor or hydrogen donor ready to form H-bonds. The vibrational frequency of carboxylic C=O stretching is dependent of dielectric medium, which is apparent in Table I. If this correlation between vibrational frequency and dielectric medium is known, it can be used to monitor environmental changes of the carboxylic group inside proteins from FTIR spectra. This will help us to obtain a better understanding of the structural basis of protein functions. Therefore, we carry out a number of *ab initio* calculations using B3LYP/6-31G(d) theory to compute the vibrational frequencies of MVA in all possible solvents provided in the program Gaussian98. We found that the vibrational frequency of carboxylic C=O

stretching is linearly proportional to the inverse dielectric constant of the aprotic solvent (see Figure 3.5). When protic solvent is present, the vibrational frequency is significantly lower due to strong H-bonds formation (small box in Figure 3.5). The dashed line is a least-squares fit of the vibrational frequencies of COOH in aprotic solvents and in vacuum while the solid line is a least-squares fit in aprotic solvents only. The standard deviation between fitted result and data points is 2.58 cm$^{-1}$ for aprotic solvents and vacuum, and is 1.93 cm$^{-1}$ for aprotic solvents only.
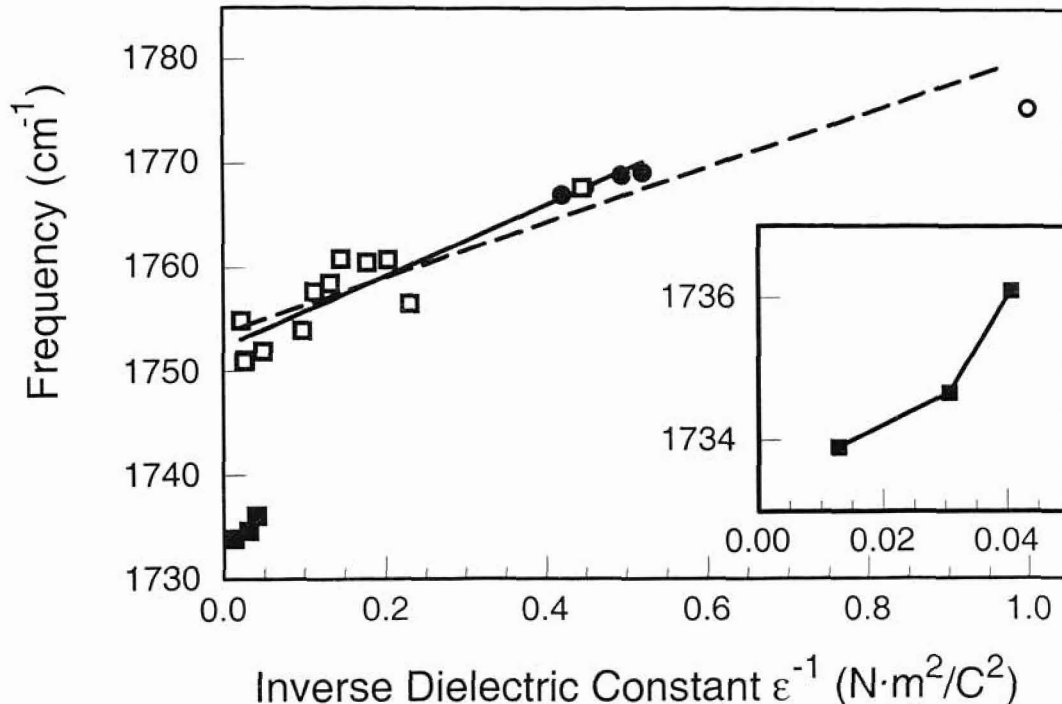


Figure 3.5 Dependence of calculated vibrational frequency of carboxylic C=O stretching on inverse dielectric constant of solvent. Hollow circle point represents the frequency in vacuum, solid circles in nonpolar aprotic solvent, hollow squares in polar aprotic solvent, and solid squares in protic solvent. The solid line is the fitting result of those data points for aprotic solvents and the dashed line is that for aprotic solvents and for vacuum. The small box is the enlarged area of three points for protic solvents.

Table IV shows detailed information of Figure 3.5 including the dielectric constants, structural formulas of the solvents and corresponding calculated vibrational frequencies of carboxylic C=O stretching.

**Table IV. Dependence of vibrational frequency on solvent based on *ab initio* DFT calculations**

| Solvent Name | Dielectric Constant $\varepsilon$ | $1/\varepsilon$ | Solvent Formula | Scaled Vibrational Freq. (cm$^{-1}$) [a] |
|---|---|---|---|---|
| Vacuum | 1 | 1 | N/A | 1775.5 |
| Heptane | 1.9 | 0.52 | $C_7H_{16}$ | 1769.2 |
| CycloHexane | 2.0 | 0.49 | $C_6H_{12}$ | 1768.9 |
| CarbonTetrachloride [c] | 2.2 | 0.45 | $CCl_4$ | 1767.7 |
| Benzene | 2.2 | 0.45 | $C_6H_6$ | 1767.7 |
| Toluene | 2.4 | 0.42 | $C_7H_8$ | 1767.1 |
| DiEthylEther [c] | 4.3 | 0.23 | $C_4H_{10}O$ | 1755.4 |
| Chloroform [c] | 4.9 | 0.20 | $CHCl_3$ | 1760.9 |
| ChloroBenzene [c] | 5.6 | 0.18 | $C_6H_5Cl$ | 1760.5 |
| Aniline [c] | 6.9 | 0.15 | $C_6H_5NH_2$ | 1760.9 |
| TetraHydroFuran [c] | 7.6 | 0.13 | $C_4H_8O$ | 1758.5 |
| DiChloroMethane [c] | 8.9 | 0.11 | $CH_2Cl_2$ | 1757.7 |
| DiChloroEthane [c] | 10.4 | 0.10 | $C_2H_4Cl_2$ | 1754.0 |
| Acetone [c] | 20.7 | 0.05 | $C_3H_6O$ | 1752.0 |
| Ethanol [b] | 24.6 | 0.04 | $C_2H_6O$ | 1736.1 |
| Methanol [b] | 32.6 | 0.03 | $CH_4O$ | 1734.7 |
| Acetonitrile [c] | 36.6 | 0.03 | $CH_3CN$ | 1751.1 |
| NitroMethane [c] | 38.2 | 0.03 | $CH_3NO_2$ | 1751.0 |
| DMSO [c] | 46.7 | 0.02 | $C_2H_6SO$ | 1754.9 |
| Water [b] | 78.4 | 0.01 | $H_2O$ | 1733.9 |

Values of dielectric constant are given for 1 atm and 20° C (36, 37). [a] scale factor is 0.9613. The vibrational frequency is for C=O stretching of carboxylic acid group. [b] protic solvents. [c] polar aprotic solvents. Other solvents are nonpolar aprotic solvents.

## 3.4 Electrostatic field and charge distribution studies of biomolecules

It is difficult to measure the electrostatic field and charge distributions experimentally. Calculated charge distribution is an important parameter in papers to perform molecular dynamics simulation (38, 39, 40) and to make further conclusion (41, 42). Based on our studies, however, these charge

distribution calculations are not reliable while we should rely much more on the electrostatic field. *Ab initio* methods can perform a Mulliken population analysis, which partitions the shared electrons to the atoms in the molecule even though the net charge of the molecule is zero or localized on a particular atom. *Ab initio* methods can also compute the electrostatic potentials and fields generated by the molecule itself. Since there are no experimental data to compare to, are these calculations reliable? We calculate the charge distribution and electrostatic
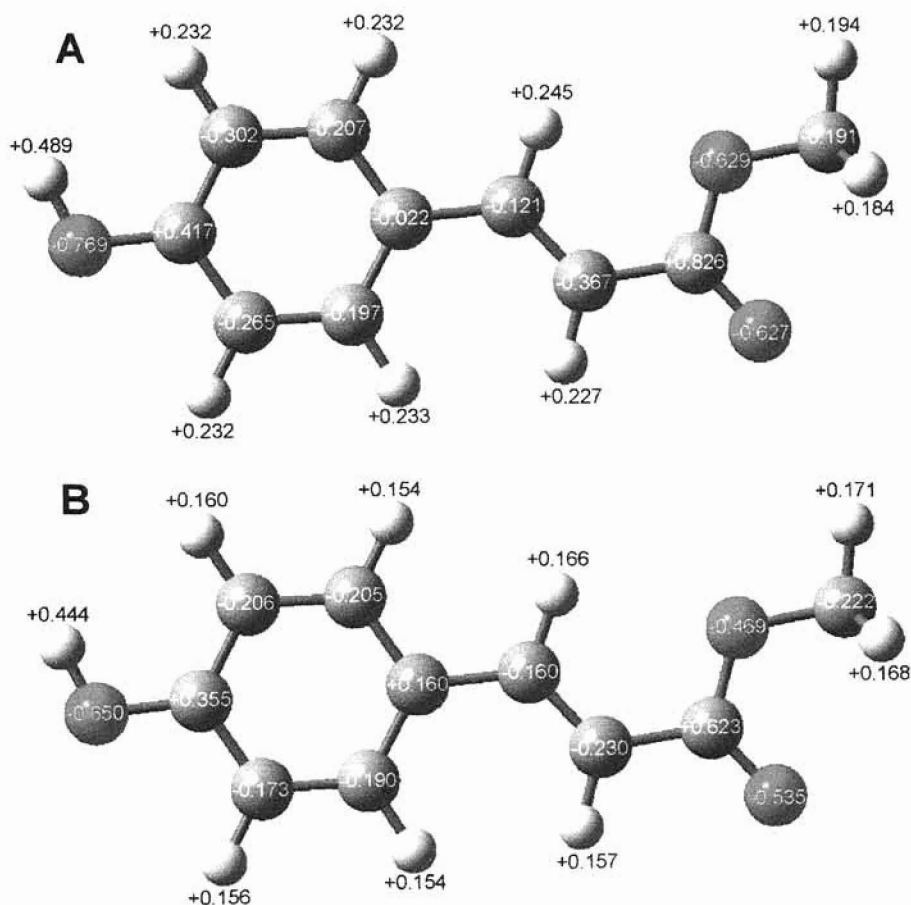


Figure 3.6 Charge distributions over pCA-ME-*trans* molecules. (A) HF/6-31G(d) method; (B) B3LYP/6-31G(d) method. The unit of the charge is e. The carbon atoms are depicted in gray, the oxygen atoms in red, and the hydrogen atoms in white. The average difference in charge is 0.077e. The largest variance in charge is 0.198e.

36

potentials of pCA-ME-*trans* molecule with all four methods HF/STO-3G, HF/3-21G, HF/6-31G(d), and B3LYP/6-31G(d). The charge distributions of HF/6-31G(d) and B3LYP/6-31G(d) as examples are shown in Figure 3.6. Their contour plots of electrostatic potentials of two methods are shown in Figure 3.7.

In Figure 3.6, atoms in red are oxygen atoms carrying partial negative charges for their strong electronegativities (the ability to attract electrons), atoms
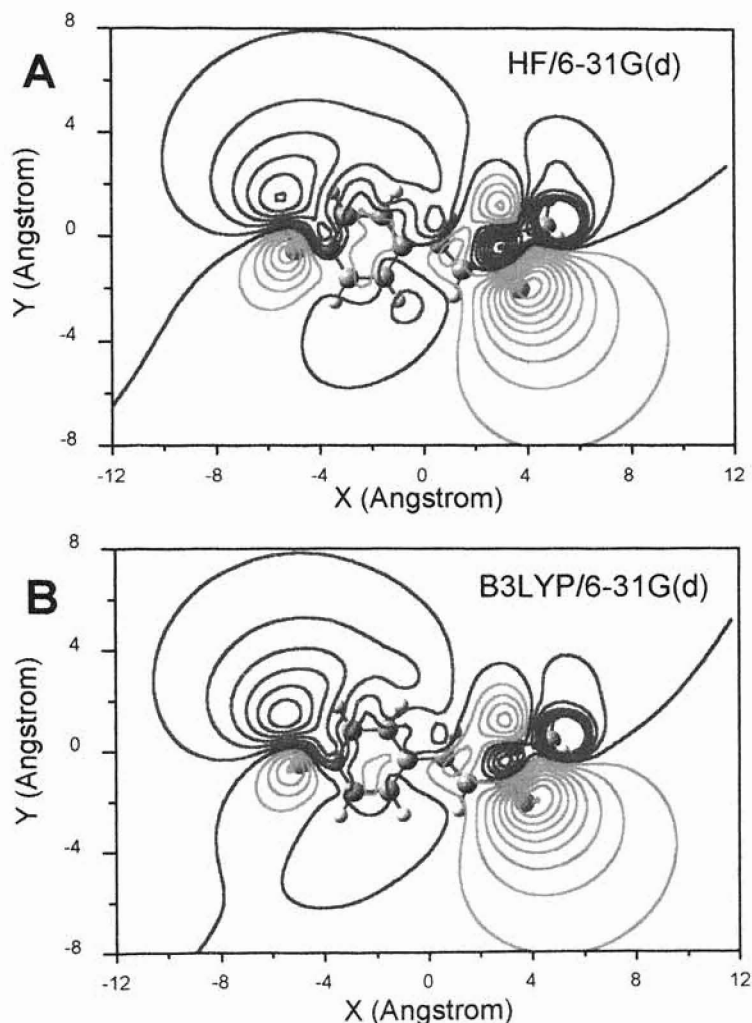


Figure 3.7 Contour plots of equipotential lines (red: negative; blue: positive; and green: zero potential) at Z=1.6 Å calculated using (A) HF/6-31G(d) and (B) B3LYP/6-31G(d) methods. The increment between two neighboring equipotential lines is 0.2 V.

in gray are carbon atoms carrying partial negative or positive charges, and atoms in white are hydrogen atoms carrying partial positive charges. The net charge of this pCA-ME-*trans* molecule is zero. The charge distributions resulted by these two methods are very different in that two methods divide the electrons shared between atoms in different ways. So, calculated charge distribution is not dependable in characterizing the electrostatic properties. Since the atoms carrying electric charges generate electrostatic field around them, the electrostatic potential reflect the electrostatic field through $\vec{E} = -\vec{\nabla} V$, where E is the electrostatic field, V is the electrostatic potential, and $\vec{\nabla}$ is the vector differential operator. Therefore, the calculated electrostatic potentials indirectly reflect the charge distribution.

In Figure 3.7, red circles are indicating negative equipotential lines, green circles are zero equipotentials, and blue circles are positive equipotential lines. The position of the molecule is determined by the equipotential of nucleus (see Appendix III). In contrast with the computed charge distributions, which give very different values for the four methods, it is evident that the two contour plots for different methods are almost the same. Therefore, we conclude that HF/6-31G(d) method is the optimal method for electrostatic potential calculations and electrostatic field is more unambiguous in characterizing electrostatic properties of a molecule than point charge distribution.

# CHAPTER IV

# BAND ASSIGNMENT OF THE CHROMOPHORE
# OF PHOTOACTIVE YELLOW PROEIN

## 4.1 Photoactive yellow protein

Photoactive yellow protein (PYP) is a small water-soluble protein. It was first isolated from the bacterium *Halorhodospira halophila* in 1985 (43, 44). PYP is the putative photoreceptor for the negative phototaxis response to blue light (45). It bears a unique thioester-linked p-coumaric acid (pCA) chromophore for light detection (46, 47). The protein in the receptor state is characterized by a strong absorption band centered at 446 nm, resulting in its yellow color (43).

Upon absorption of a blue photon, PYP undergoes a cyclic process involving structural transitions at the chromophore, photoactive site residues, and over the global protein conformation. The photocycle of PYP is shown in Figure 4.1. Following chromophore *trans* to *cis* around the C7=C8 bond (Figure 4.2) and C9=S bond photoisomerization, a red-
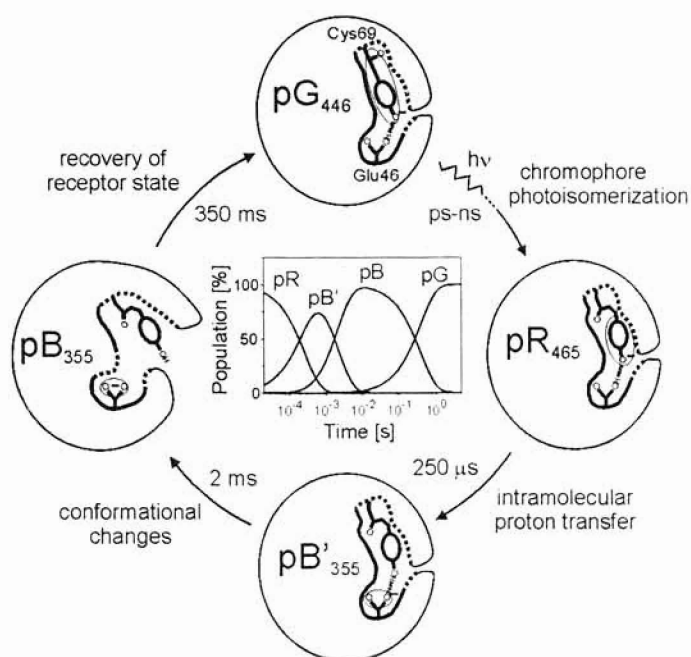


Figure 4.1 The photocycle of PYP. The pG446 state is the receptive state, and the pB355 state is the signaling state. The pB'355 state is the most recently discovered intermediate, and is of central importance for understanding PYP receptor activation.

shifted intermediate state pR with $\lambda_{max}$ at 465 nm is formed in nanoseconds. Subsequently, a proton transfer event occurs in which the pCA accepts a proton from Glu46 (20), leading to a blue-shifted intermediate state pB' with $\lambda_{max}$ at 355 nm. Subsequent large conformational changes result in a new blue-shifted pB state with $\lambda_{max}$ at 355 nm. Finally, pB returns to the initial state pG.

Monitoring structural changes occurring during the photocycle of PYP from FTIR difference spectra is crucial to probe structural functions of PYP. The infrared signals in the pR - pG, and pB' - pG difference spectra are dominated by the chromophore and key residues at active site of PYP. Therefore, we will obtain important information of vibrational modes for these IR signals by assigning these bands through *ab initio* calculations. The reason why we use model compounds is that the chromophore is connected covalently to the entire protein via thioester linkage, thus it is necessary to separate the chromophore from the protein appropriately in order to perform *ab initio* calculations. The chemical structures of the chromophore and the model compounds are shown in Figure 4.2. We use the model
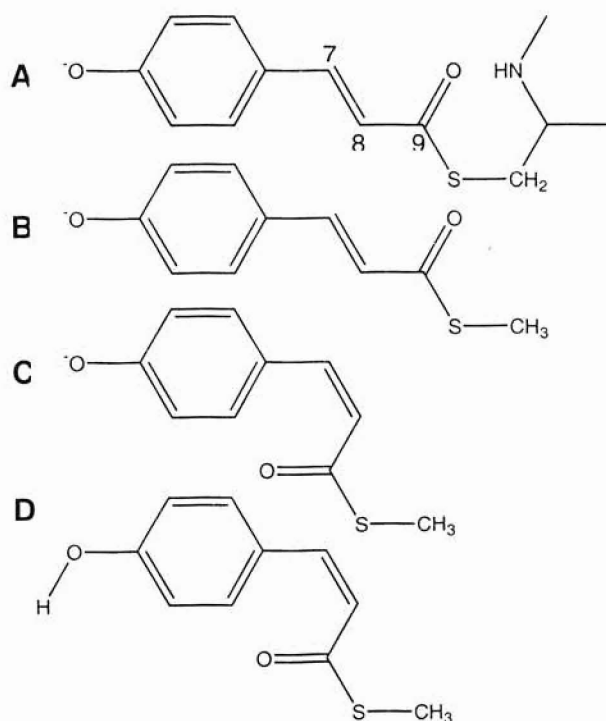


Figure 4.2 The Chemical structures of the chromophore in the ground state of PYP (A), the model compounds for pG (B), pR (C), and pB' (D).

compound Figure 4.2 (B) as the deprotonated *trans* chromophore in pG state Figure 4.2 (A), Figure 4.2 (C) as the deprotonated *cis* chromophore in pR state, and Figure 4.2 (D) as the protonated *cis* chromophore in pB' state.

## 4.2 Band assignment of the chromophore

In order to assign the bands of the FTIR difference spectra pR – pG and pB' – pG, we first calculate the vibrational frequencies of the model compound of pG, pR, and pB' in ethanol using B3LYP/6-31G(d) method, then subtract the scaled frequencies of pG from those of pR and pB'. The calculated frequencies of pG, pR and pB' are shown in Table I.

**Table I. Calculated vibrational frequencies of pG, pR and pB' in ethanol**

| Model Compound | Band I ($cm^{-1}$) | Band II ($cm^{-1}$) | Band III ($cm^{-1}$) | Band IV ($cm^{-1}$) | Band V ($cm^{-1}$) |
|---|---|---|---|---|---|
| pG | 1641 | 1606 | 1557 | 1499 | 1487 |
| pR | 1629 | 1599 | 1557 | 1494 | 1489 |
| pB' | 1674 | 1628 | 1591 | 1559 | 1523 |

The scale factor is 0.9623 for pG and pR, and 0.9779 for pB'.

The calculated vibrational frequencies of pR and pG before subtraction are shown in Figure 4.3 (A) and after subtraction are shown in Figure 4.3 (B). The experimental difference spectrum of pR - pG revealing the structural changes of protein and the chromophore is also shown in Figure 4.3 (B). The negative band at 1726 $cm^{-1}$ arises from the stretching mode of deuterated carboxylic sidechain of Glu46 (20). This band is not present in computational data. The second negative band at 1643 $cm^{-1}$ is assigned to the stretching mode of carbonyl group (C=O) coupled with C7=C8 stretching and with a little ring vibration on the pG chromophore because the calculated frequency is exactly overlapping the experimental frequency. The third negative band at 1610 $cm^{-1}$ is

Figure 4.3 The difference spectrum of pR – pG. (A) The calculated frequencies of pR (black) and pG (blue) before subtraction; (B) the difference spectra of the experimental data is shown in black and the computational data in blue. The scale factor for pR and pG is 0.9623.

assigned to the stretching mode of C7=C8 double bond next to the phenolic ring coupled with symmetric ring vibration. The negative band at 1552 cm$^{-1}$ is contributed by asymmetric ring vibration. And the band at 1501 cm$^{-1}$ is assigned C7=C8 stretching mode coupled with asymmetric ring vibration. In the same way, we assign the positive band at 1622 cm$^{-1}$ to the stretching mode of carbonyl group coupled with symmetric ring vibration on the pR chromophore, the positive band at 1597 cm$^{-1}$ arises from the stretching mode of C7=C8 double bond coupled with carbonyl stretching and with symmetric ring vibration, and the positive band at 1485 cm$^{-1}$ is assigned to the asymmetric ring vibration of the pR chromophore. The intensities of the negative band of pG chromophore at 1487 cm$^{-1}$ and the positive bands of pR chromophore at 1557 cm$^{-1}$ and 1489 cm$^{-1}$ are cancelled out because the positive and negative intensities are close to each other. Therefore, these bands are not shown in the

difference spectrum. The remaining bands are considered as contributions of amino acids of PYP.

The band assignment of pB' chromophore bands is performed the same way as those of the pR and pG chromophore. The calculated vibrational frequencies of pB' and pG before subtraction are shown in Figure 4.4 (A) and after subtraction are shown in Figure 4.4 (B). The experimental difference spectrum of pB' − pG is also shown in Figure 4.4 (B). The positive band at 1674 cm$^{-1}$ is assigned the stretching mode of carbonyl group coupled with C7=C8 stretching and a little ring vibration on pB' chromophore. The positive band at 1622 cm$^{-1}$ arises from the stretching mode of C=C double bond coupled with symmetric ring vibration. The positive band at 1594 cm$^{-1}$ is assigned to asymmetric ring vibration. The band at 1563 cm$^{-1}$ are assigned to the asymmetric ring vibrations coupled with C7=C8 stretching. And the positive
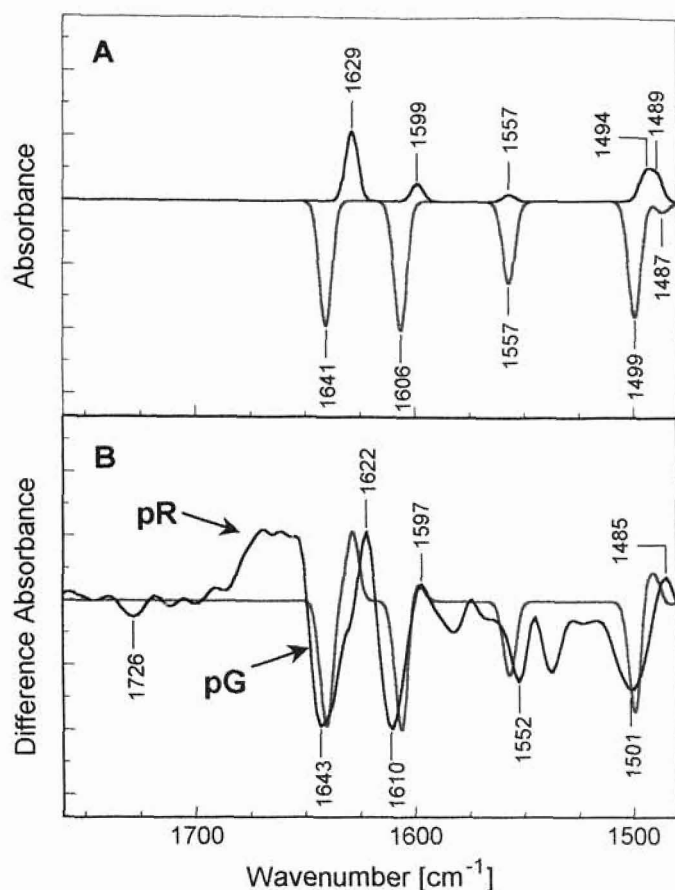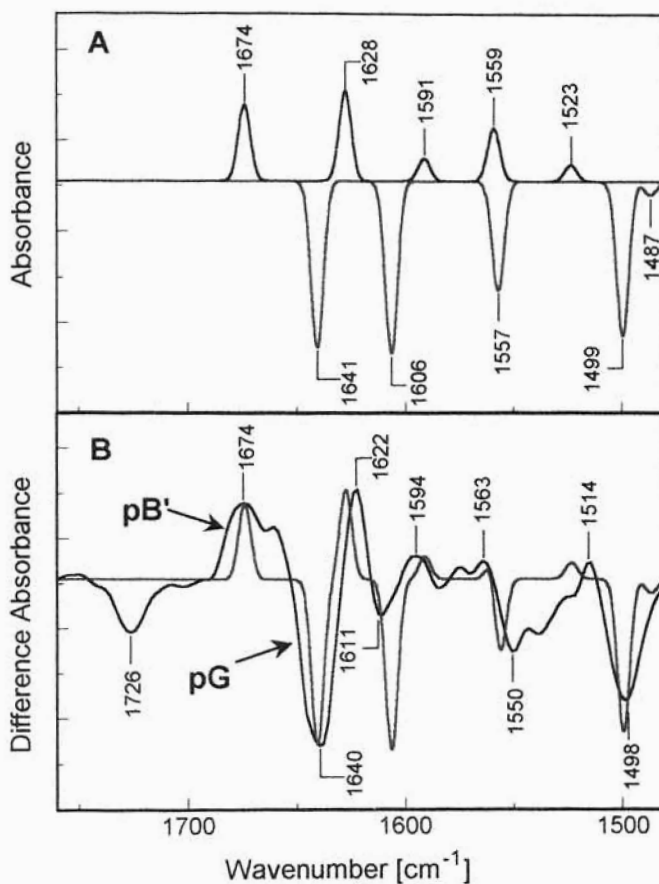


Figure 4.4 The difference spectrum of pB' − pG. (A) The calculated frequencies of pB' (black) and pG (blue) before subtraction; (B) the difference spectra of the experimental data is shown in black and the computational data in blue. The scale factor for pB' is 0.9779 and for pG is 0.9623.

band at 1514 cm$^{-1}$ is assigned to the asymmetric ring vibration of the pB' chromophore.

Among each five bands arising from the chromophore, three of them are considered as spectral markers for they can provide us important information of structural changes. The first spectral marker is the negative band at 1643 cm$^{-1}$, it is down-shifted to 1622 cm$^{-1}$ for pR chromophore and up-shifted to 1674 cm$^{-1}$ for pB' chromophore. The vibrational modes of this spectral marker for pG, pR and pB' chromophore are shown in Figure 4.5. It is characterized by carbonyl group (C=O) stretching mode and isomerization of C7=C8 double bond, but it coupled with different motions in pG, pR, and pB' states. The second spectral marker is the negative band at 1610 cm$^{-1}$, it is down-shifted to 1597 cm$^{-1}$ for pR chromophore and up-shifted to 1622 cm$^{-1}$ for pB' chromophore. It is noticeable that the intensity of this band in pB' - pG FTIR difference spectrum decreased due to the cancellation of a positive band at the same position arising from other amino acids of PYP. The vibrational modes of this spectral marker for pG, pR and pB' chromophore are shown in Figure 4.6. It is characterized by C7=C8 stretching



Figure 4.5 The vibrational modes of the first spectral marker. (A) 1643 cm$^{-1}$ for pG chromophore; (B) 1622 cm$^{-1}$ for pR chromophore; (C) 1674 cm$^{-1}$ for pB' chromophore.

coupled with symmetric ring vibration. The third spectral marker is the negative
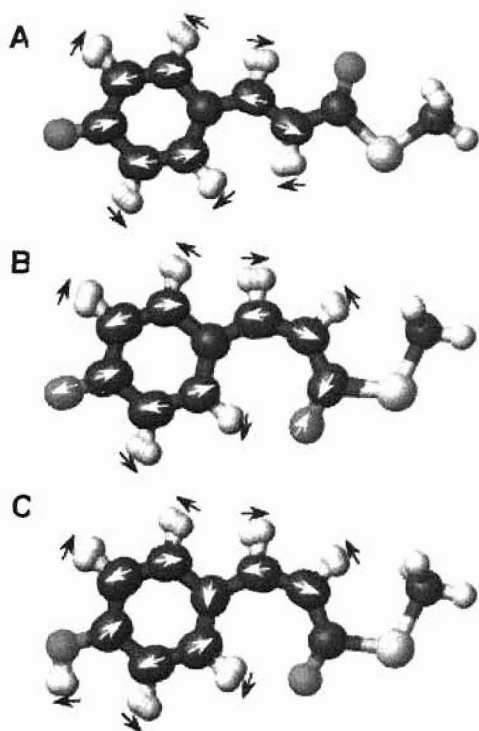


Figure 4.6 The vibrational modes of the second spectral marker. (A) 1610 cm$^{-1}$ for pG chromophore; (B) 1597 cm$^{-1}$ for pR chromophore; (C) 1622 cm$^{-1}$ for pB' chromophore.
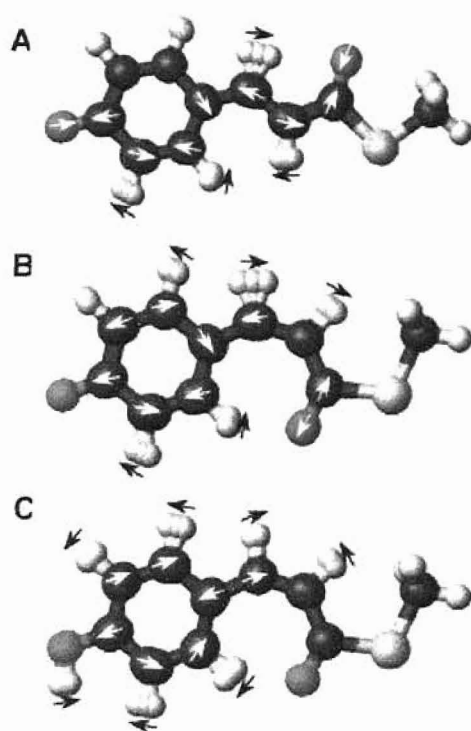
Figure 4.7 The vibrational modes of the third spectral marker. (A) 1501 cm$^{-1}$ for pG chromophore; (B) 1485 cm$^{-1}$ for pR chromophore; (C) 1514 cm$^{-1}$ for pB' chromophore.

band around 1501 cm$^{-1}$, it is down-shifted to 1485 cm$^{-1}$ for pR chromophore and

up-shifted to 1515 cm$^{-1}$ for pB' chromophore. The vibrational modes of this

spectral marker for pG, pR and pB' are shown in Figure 4.7. It is characterized by

protonation of the chromophore and asymmetric ring vibration.

Thus, the band assignment of the chromophore bands in pG, pR, and pB'

states for spectral region 1760 cm$^{-1}$ - 1480 cm$^{-1}$ is complete.

# CHAPTER V
# SUMMARY AND REMARKS

## 5.1 Summary

Time-resolved infrared absorption spectroscopy is an important technique for structural and kinetic studies of protein functional mechanism. Vibrational band assignment is a crucial step in order to obtain structural information from infrared absorption spectra. Traditionally, vibrational band assignment is a challenging task, requires repetitive, trial and error empirical calculations and synthesis and use of isotopically labeled molecules. We employed a new approach for vibrational band assignments based on *ab initio* computational studies. We tested and evaluated the accuracy of vibrational band assignment of four *ab initio* methods, HF/STO-3G, HF/3-21G, and HF/6-31G(d) based on Hartree-Fock theory, and B3LYP/6-31G(d) based on density function theory, by comparing the computational results with experimental data. We found that B3LYP/6-31G(d) is the optimal method. This method is therefore employed for vibrational band assignment of a free chromophore, *p*-coumaric acid methyl ester, and obtained excellent results. When assigning vibrational bands of biomolecules, geometry optimization, conformational effect, and solvent effect have to be taken into account to best match the experimental data.

Electrostatic interactions play a major role in the structure and function of proteins. It is difficult to perform direct experimental studies on electrostatic interactions in proteins. We employ *ab initio* computational method to study hydrogen bonding interactions and their effects on vibrational frequencies of participating groups. This allows the use of vibrational spectral markers to probe

hydrogen bonding interactions in proteins that is a major contributor to protein structural stabilization. In addition, we tested two methods that characterize electrostatic properties of a molecule, (i) point charge distribution, and (ii) electrostatic potential map. We found that despite the fact that many groups have reported point charge distributions in biomolecules, this method is inaccurate and misleading. A more reliable method is electrostatic potential map that provides clear view of electrostatic field generated by the molecule and insight how this electrostatic field will influence the proton affinity of ionizable groups and the structure of polar groups in the area.

In addition, we made vibrational frequency calculations of three chromophore structures in the ground state and functional intermediates of photoactive yellow protein, a photoreceptor protein for bacterial vision. Without use of any isotopic labelings, the results are in good agreement with experimental data. We also identified three spectral markers characterizing the structural changes from the chromophore.

## 5.2 Remarks

Computational biophysics based on *ab initio* computational methods is a rapidly growing field, empowered by the increasing computational powers and new developments in *ab initio* computational methods. The results reported in this thesis demonstrate that *ab initio* computational studies of biomolecules can provide valuable structural, spectral, and energetic information for biophysical studies of proteins. We are in the process of further expanding the scope of our

computational studies of biomolecules in combination with experimental studies of the functional mechanism of proteins.

*Hydrogen bonding interactions in proteins:* Hydrogen bonding interaction is crucial in protein structural stabilization and in many functional processes of proteins, including signal transduction, energy transduction, and enzymatic catalytic processes. A typical protein folding energy, the energy difference between the native structure and the unfolded structure of a protein, is approximately 40–50 kJ/mole, equivalent to the sum of only three hydrogen bonding interactions (~15 kJ/mol each). By combining time-resolved infrared absorption spectroscopy with *ab initio* computational studies, it is possible to study functionally important structural changes in proteins involving the formation and disruption of hydrogen bonding interactions. This approach will be further employed in studies of the activation mechanism of photoactive yellow protein.

*Proton affinity of ionizable groups in Proteins:* Proton transfer is a fundamental process in biological energy transduction, and is essential for many receptor proteins in biological signal transduction. Proton transfer takes place as a result of induced changes in the proton affinities of participating ionizable groups during the functional process of a protein. In order to understand the structural basis of proton transfer, it is insightful to carry out *ab initio* computational studies on proton affinities of ionizable groups in proteins.

Since protein environment of an ionizable group in not homogeneous, reliable calculations should include the specific environment of the ionizable group. We are in the process of performing such *ab initio* calculations by

combining with molecular dynamics simulation. That is, we select an ionizable group and its surrounding groups at the active site of the protein to be treated using *ab initio* computational method, and the rest of proteins to be studied using classical molecular dynamics simulation. The structural optimization is performed via an alternative iterative process, with *ab initio* calculations, followed with molecular dynamics simulations. Upon completion of structural optimization, we then calculate the energies of the protein with and without a proton residing on the chosen ionizable group. The energy difference between the two protonation states is a direct measure of proton affinity of the ionizable group. Since a protein is a dynamic system, experiences constant structural fluctuations, a number of protein conformational frames will be selected and employed to perform proton affinity calculations. This ongoing project is in collaboration with Dr. Benjamin H. McMahon at the Theoretical Biophysics Division in Los Alamos National Laboratory.

# REFERENCE

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

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

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

4.  Becke, A.D., *Density-functional thermochemistry. IV. A new dynamical correlation Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing*. J. Chem. Phys., 1996. **104**: p. 1040.

5.  Cammi, R., et al., *On the calculation of infrared intensities in solution within the polarizable continuum model*. J. Phys. Chem., 2000. **104**: p. 9874 - 9879.

6.  Cappelli, C., et al., *Nonequilibrium formulation of infrared frequencies and intensities in solution: analytical evaluation within the polarizable continuum model*. J. Chem. Phys., 2000. **113**(24): p. 11270 - 11279.

7.  Mathies, R.A. and X. Li, *On modeling the vibrational spectra of 14-s-cis retinal conformers in bacteriorhodopsin*. Biophys. Chem., 1995. **56**: p. 47-55.

8.  Hu, X., H. Frei, and T.G. Spiro, *Nanosecond step-scan FTIR spectroscopy of hemoglobin: ligand recombination and protein conformational changes*. Biochemistry, 1996. **35**: p. 13001-13005.

9.  Eads, D.D., et al., *Selective enhancement of resonance Raman spectra of separate bacteriopheophytins in Rb. sphaeroides reaction centers*. Biopolymers, 2000. **57**(2): p. 64-76.

10. Nakanishi, K. and P.M. Solomon, *Infrared absorption spectroscopy*. 1977, San Francisco: Holden-Day.

11. Lin, S.L., et al., *Deprotonation of tyrosines in bacteriorhodopsin as studied by Fourier transform infrared spectroscopy with deuterium and nitrate labeling*. Biochemistry, 1987. **26**: p. 8327-8331.

12. Liu, X., et al., *Site-direct isotope labeling and FTIR spectroscopy: assignment of tyrosine bands in the bR- M difference spectrum of bacteriorhodopsin.* Biophysical Chemistry, 1995. **56**: p. 63-70.

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

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

15. Foresman, J.B. and Æ. Frisch, *Exploring chemistry with electronic structure methods.* 2nd ed. 1996, Pittsburgh: Gaussian Inc.

16. Shankar, R., *Principles of quantum mechanics.* 2nd ed. 1994, New York: Plenum Press.

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

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

19. Solomons, G. and C. Fryhle, *Organic chemistry.* Seven ed. 1998, New York: John Wiley & Sons, Inc.

20. 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.

21. Xie, A., et al., *Formation of a new buried charge drives a large-amplitude protein quake in photoreceptor activation.* Biochemistry, 2001. **40**: p. 1510-1517.

22. Taylor, B.L. and I.B. Zhulin, *PAS domains: Internal sensors of oxygen, redox potential, and light.* Microbiol. Mol. Biol. Rev., 1999. **63**(2): p. 479-506.

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

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

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

26.    Zhang, X. and J.L. Spudich, *His166 is critical for active-site proton transfer and phototaxis signaling by sensory rhodopsin I.* Biophys. J., 1997. **73**: p. 1516-1523.

27.    Szundi, I., et al., *Proton transfer reactions linked to rhodopsin activation.* Biochemistry, 1998. **37**: p. 14237-14244.

28.    Creighton, T.E., *Protein structures and molecular properties.* Second Edition ed. 1997, New York: W. H. Freeman and Company.

29.    Xie, A., A.F.G. van der Meer, and R.H. Austin, *Excited-state lifetimes of far-infrared collective modes in proteins.* Phys. Rev. Lett., 2002. **88**(1): p. 018102-1 - 4.

30.    Sasaki, J., et al., *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, 1994. **33**(11): p. 3178-3184.

31.    Rammelsberg, R., et al., *Bacteriorhodopsin's intramolecular proton-release pathway consists of a hydrogen-bonded network.* Biochemistry, 1998. **37**: p. 5001-5009.

32.    Fahmy, K., et al., *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, 1993. **90**: p. 10206-10210.

33.    Eisberg, R. and R. Resnick, *Quantum physics of atoms, molecules, solids, nuclei, and particles.* 2nd ed. 1985: John Wiley & Sons, Inc.

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

35.    Dioumaev, A.K., *Evaluation of intrinsic chemical kinetics and transient product spectra from time-resolved spectroscopic data.* Biophys. Chem., 1997. **67**(1-3): p. 1-25.

36.    Griffiths, D.J., *Introduction to electrodynamics.* Third ed. 1999: Prentice-Hall, Inc.

37.     Frisch, Æ. and M.J. Frisch, *Gaussian 98 user's reference*. 2nd ed. 1999, Pittsburgh: Gaussian, Inc.

38.     Groenhof, G., et al., *Signal transduction in the photoactive yellow protein. I. photon absorption and the isomerization of the chromophore*. Proteins: structure, function, and genetics, 2002. **48**: p. 202-211.

39.     Groenhof, G., et al., *Signal transduction in the photoactive yellow protein. II. proton transfer initiates conformational changes*. Proteins: structure, function, and genetics, 2002. **48**: p. 212-219.

40.     Yamato, T., N. Niimura, and N. Go, *Molecular dynamics study of femtosecond events in photoactive yellow protein after photoexcitation of the chromophore*. Proteins, 1998. **32**(3): p. 268-275.

41.     He, A., et al., *Theoretical studies on excited states of a phenolate anion in the environment of photoactive yellow protein*. J. Phys. Chem., 2000. **104**: p. 2939-2952.

42.     Molina, V. and M. Merchan, *On the absorbance changes in the photocycle of the photoactive yellow protein: a quantum-chemical analysis*. Proc. Natl. Acad. Sci. USA, 2001. **98**(8): p. 4299-4304.

43.     Meyer, T.E., *Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium Ectothiorhodospira halophila*. Biochim. Biophys. Acta, 1985. **806**(1): p. 175-183.

44.     Imhoff, J.F. and J. Suling, *The phylogenetic relationship among Ectothiorhodospiraceae: A reevaluation of their taxonomy on the basis of 16S rDNA analyses*. Arch. Microbiol., 1996. **165**: p. 106-113.

45.     Sprenger, W.W., et al., *The eubacterium Ectothiorhodospira halophila is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein*. J. Bacteriol., 1993. **175**(10): p. 3096-3104.

46.     Hoff, W.D., et al., *Thiol ester-linked p-coumaric acid as a new photoactive prosthetic group in a protein with rhodopsin-like photochemistry*. Biochemistry, 1994. **33**(47): p. 13959-13962.

47.     Baca, M., et al., Complete chemical structure of photoactive yellow protein: novel thioester-linked 4-hydroxycinnamyl chromophore and photocycle chemistry. Biochemistry, 1994. 33(48): p. 14369-14377.

## Appendix I

### Atomic units

The fundamental equations of quantum chemistry are usually expressed in units designed to simplify their form by eliminating fundamental constants. The atomic unit of length is the *Bohr radius*:

$$a_0 = \frac{\varepsilon_0 h^2}{\pi m_e e^2} = 0.529\text{Å} \tag{1}$$

where $\varepsilon_0$ is permittivity of free space, $\varepsilon_0 = 8.85 \times 10^{-12} C^2/Nm^2$, h is planck's constant, $h = 6.63 \times 10^{-34} Js$, $m_e$ is electron mass, $m_e = 9.109 \times 10^{-31} kg$, e is unit charge, $1\ e = 1.6 \times 10^{-19} C$. Coordinates can be transformed to Bohrs by dividing them by $a_0$. Energies are measured in hartrees (i.e. a.u.), defined as the Coulomb interaction between two charges of e separated by one Bohr distance (15):

$$1\ \text{hartree} = \frac{e^2}{4\pi\varepsilon_0 a_0} \tag{2}$$

Masses are also specified in terms of electron mass units (i.e. define $m_e=1$). Therefore, we may exchange the units as follows,

$$1\ \text{a.u.} = \frac{e^2}{4\pi\varepsilon_0 a_0} J = \frac{e}{4\pi\varepsilon_0 a_0} eV = \frac{e^2}{4\pi\varepsilon_0 a_0} \cdot \frac{N_A}{1000} kJ/mol$$

where $N_A$ is Avogadro's number. After carrying out the calculations, we have the following relations,

$$1\ \text{a. u.} = 27.2\ eV = 2619.6\ kJ/mol$$

## Appendix II

### H-bond formation in proteins

| Protein | COOH group | IR Freq. $\lambda$ (cm$^{-1}$) | No. of H-bonds | H-bond partner | Distance (Å) |
|---------|------------|--------------------------------|----------------|----------------|--------------|
| M412    | D85        | 1760[25]                       | 0              | 0              | N/A          |
| R       | D83        | 1767[32]                       | 0              | 0              | N/A          |
| bR568   | D115       | 1742[30]                       | 1              | T90            | 2.582        |
| bR568   | D96        | 1740[30]                       | 1              | T46            | 2.402        |
| M412    | D212       | 1738[25]                       | 1              | Y57            | 2.273        |
| PYP     | E46        | 1737[29]                       | 1              | HC69           | 2.687        |
| M412    | D115       | 1734[30]                       | 1              | T90            | 3.100        |
| R       | E122       | 1734[32]                       | 1              | H211           | 2.906        |
| bR568   | E204       | 1700[31]                       | 2              | E194/ S193     | 3.025/ 2.601 |

E: glutamic acid, D: aspartic acid, T: threonine, Y: tyrosine, HC: the chromophore, H: histidine, and S: serine. M412 denotes that peak absorption of bR's M intermediate state is at 412nm and bR568 denotes that peak absorption of bR ground state is at 568nm. R denotes rhodopsin and PYP denotes photoactive yellow protein. See Table II in Chapter III for hydrogen bonding types for code.

**Appendix III**

Overlapping procedure

1.  Plot the contour plot of equipotential lines at z=0 Å indicating the electric fields generated by the nuclei of pCA-ME-*trans*. (Figure 1).

2.  Overlap the chemical structure of pCA-ME-*trans* on the contour plot according to the nuclei positions of the atoms.
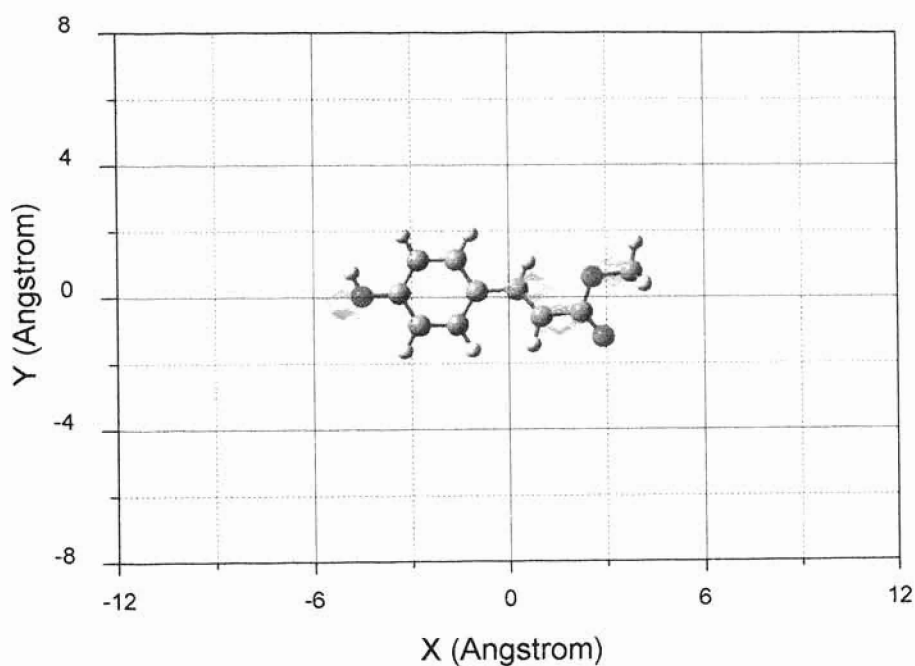


Figure 1 The contour plot of equipotential lines of pCA-ME-*trans* at z=0 Å is shown in blue. The chemical structure is overlapping the contour plot

VITA 2

Beining Nie

Candidate for the Degree of

Master of Science

Thesis: VIBRATIONAL BAND ASSIGNMENT AND ELECTROSTATIC
PROPERTIES OF BIOMOLECULES BASED ON *AB INITIO*
COMPUTATIONAL STUDIES

Major Field: Physics

Biographical:

Personal Data: Born in Kashi, Xinjiang, China, on October 19, 1974.

Education: Graduated from Jinzhou High School, Jinzhou, Liaoning, 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 Graduate
College of Beijing University of Aeronautics and Astronautics,
Beijing, China, in December 1999. Completed the requirements of
the Master of Science degree with a major in physics at Oklahoma
State University in December 2002.

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; employed as research assistant in the Department of
Physics, Oklahoma State University, Spring 2001 to present;
employed as a graduate research assistant (GRA) in the theoretical
biophysics division, Los Alamos National Laboratory during the
summer 2002.

Professional Membership: Member of Biophysical Society since 2002.