Oxidation of Methionine Residues in Protein Pharmaceuticals

by

Jhih-Wei Chu

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

[September 2004]
~~June 2004~~

Author ....
Department of Chemical Engineering
June 15, 2004

Certified by ....
Bernhardt L. Trout
Associate Professor of Chemical Engineering
Thesis Supervisor

Certified by ....
Daniel I.C. Wang
Institute Professor of Massachusetts Institute of Technology
Thesis Supervisor

Accepted by ....
Daniel Blankschtein
Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Oxidation of Methionine Residues in Protein Pharmaceuticals

by

Jhih-Wei Chu

Submitted to the Department of Chemical Engineering
on June 15, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Oxidation of the amino acid methionine by peroxides in aqueous formulations of proteins is a critical issue in the development of therapeutic products. It must be controlled so that therapeutic proteins can maintain their activity. In addition, oxidized therapeutics are undesirable due to their possible immunogenetic effects. An understanding of the mechanism and the factors that influence the reactivity of different methionine sites toward oxidation is therefore important. In this thesis, computational methods are applied and developed to address these problems.

First, a mechanism by which peroxides oxidize the sulfur atom of methionine is developed. The rate-limiting step was found to be the breaking of the O-O bond of H_2O_2 and the formation of the S-O bond during which significant charge separation is developed. The charge separation can be stabilized via specific interactions such as hydrogen bonding with surrounding water molecules. This "water-mediated" mechanism of oxidation is consistent with experimental data such as those on activation energies of oxidation and pH dependence of the rates of oxidation. Based on the "water-mediated" mechanism, a structural property, average 2-shell water coordination number (2SWCN), has been shown to correlate well to the rates of oxidation of different methionine groups in Granulocyte Colony-Stimulating Factor (G-CSF) and in a Human Parathyroid hormone fragment (hPTH(1–34)). Including the dynamics of protein and water molecules in an explicit manner was found to be important for such correlation. Via combined quantum mechanical and molecular mechanical free energy simulations, the activation free energies of the oxidation of methionine residues in G-CSF are found to be equivalent to the values for the oxidation of free methionine. Therefore, the environments surrounding different methionine sites in G-CSF mainly provide spatial restriction to the access to the solvent but do not affect oxidation in a specific manner, consistent with the good correlation between 2SWCN's and the rates of oxidation.

A comprehensive picture of oxidation is thus developed. It allows an accurate prediction of protein oxidation, and provides a rationale for developing strategies to control oxidation, such as modulating protein conformation via adding excipients.

This knowledge could aid in developing in a more rational manner solvent formulations that protect therapeutic proteins against oxidation.

Thesis Supervisor: Bernhardt L. Trout
Title: Associate Professor of Chemical Engineering

Thesis Supervisor: Daniel I.C. Wang
Title: Institute Professor of Massachusetts Institute of Technology

Acknowledgments

First and foremost, I would like to thank my thesis advisors, Professor Bernhardt L. Trout and Professor Daniel I.C. Wang for their help, guidance, and friendship. With this unique combination, I have learned both the beauty and power of computational theories and the philosophy and arts of practicing biotechnologies. I have benefited from them not only for my doctoral studies, but also from the diversified research areas and opportunities they have provided. Their creativity and critical thinking in tackling engineering problems are invaluable lessons to me, and it has been a great pleasure to work with them.

I would like to thank each of my thesis committee members, Dr. Bernard R. Brooks, Professor Charles Cooney, and Professor Bruce Tidor for their help in guiding this thesis. I thank very much Dr. Bernie Brooks at the NIH for welcoming me to his laboratory and his hands-on training on computational methodologies. Bernie has also provided tremendous guidance, friendship, and advice. I am grateful to Professor Bruce Tidor at the Biological Engineering division for taking interest in my work and continually offering wonderful guidance and valuable advice. I also thank Professor Charles Cooney for his input and advice to my research and for providing the opportunity of taking his excellent course on downstream processing.

It has been great to study in an exciting place like MIT and to know and interact with many fantastic people. I want to thank Dr. Jin Yin for his friendship and for providing experimental data on oxidation. My colleagues, Jung-Sheng Wu, Claudiu Giurumescu, Cynthia Lo, Brian Baynes, Xi Lin, and Yuen-Chung Cheng, deserve special thanks for their friendship and their help in getting my project going in many ways. It is also necessary to thank every member of the Trout group, both past and the current, for their kindness and friendship over the years.

Finally, I would like to acknowledge my parents and sisters for their constant support and love. Without them, I could not have gone through these years. I also want to thank my fabulous girl friend, Ms. Ming-Yi Tsai for her love and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

With the advent of recombinant DNA technology, protein pharmaceuticals, including functional regulators and supplements, enzyme activators and inhibitors, antibodies, and various vaccines, have become indispensable in combating human diseases. In developing a protein as a therapeutic product, one of the most difficult steps is to deal with physical and chemical instabilities such as aggregation, deamidation and oxidation (Cleland & Langer, 1994; Shahrokh, 1997; Wang, 1999; Meyer *et al.* , 2002). If physical and/or chemical degradation pathways modify a protein molecule during storage, the biological activity and immunogenicity of the molecule can be changed as well. Therefore, chemical and physical instabilities need to be prevented in any formulation of therapeutic proteins (Cleland *et al.* , 1993). The objective of protein stabilization is thus to design a formulation to prevent physical and chemical instabilities so that an acceptable shelf life (>2 years) can be achieved for a therapeutic protein product. During storage, various degradation pathways can proceed at different rates depending on the physical-chemical properties of the formulation (temperature, pH, ionic strength, additives, etc.) and the protein molecule (primary sequence, tertiary structures, unfolding melting temperatures). The complexity of protein systems makes the design of formulation a difficult challenge, and is addressed currently empirically. A detailed understanding of individual degradation pathway would be required for an accurate prediction of protein stability and is thus highly desired for the formulation design of protein pharmaceuticals (Meyer *et al.* , 2002).

Oxidation is one of the major chemical instabilities of protein pharmaceuticals, and the sidechains of methionine (Met), cysteine (Cys), histidine (His), tryptophan (Trp) and tyrosine (Tyr) residues are potential sites for oxidative modification (Li *et al.* , 1995). The thioether group of Met makes it the one of the amino acids most liable to oxidation in an aqueous solution (Wang, 1999; Li *et al.* , 1995). Potential reactive oxygen species involves singlet oxygen, $^1O_2$, superoxide radical, $\cdot O_2^-$, peroxides, $ROOH$, and hydroxyl radicals, $\cdot OH$. Active oxygen species may be introduced into the formulation from impurities in the additives or be produced by light-catalyzed or metal-catalyzed reactions (Stadtman & Oliver, 1991; Stadtman, 1992; Nguyan, 1994; Li *et al.* , 1995). Oxidation of methionine residues in protein molecules can be roughly divided into two categories: non-site specific and site-specific reactions. Non-site-specific reactions are those in which oxidation is affected by oxidants such as peroxides or upon exposure to light (Matheson *et al.* , 1979; Vogt, 1995; Levine *et al.* , 1996; Yao *et al.* , 1996; Gao *et al.* , 1998; Patten *et al.* , 1999; Yin *et al.* , 2000). Site-specific oxidation reactions are those in which reactive oxygen species are generated at specific metal-binding sites catalyzed by transition metal ions (Levine, 1983; Mark & Chevion, 1985; Garland *et al.* , 1986; Stadman, 1990; Eckardstein *et al.* , 1991; Stadtman, 1992; Schöneich *et al.* , 1993; Stadtman, 1993; Schöneich *et al.* , 1997).

Of the residues susceptable to oxidation, methionine is the most important, since the majority of the cases of protein oxidation identified thus far are related to modification of methionine (Cleland *et al.* , 1993). The covalent addition of an oxygen atom to the sulfur atom of methionine changes the chemical properties of proteins and generally leads to the loss of biological function. Examples of the proteins that are destabilized by oxidation of methionine include Human α-1 Proteinase Inhibitor (Matheson *et al.* , 1979), Calmodulin (Levine *et al.* , 1996; Gao *et al.* , 1998), Human parathoid hormone (Nabichi *et al.* , 1998), Antithrombin (Patten *et al.* , 1999), Glutamine synthetase (Levine, 1983), α1-antitrypsin (Griffiths & Cooney, 2002), and

Granulocyte Colony-Stimulating Factor (Lu *et al.* , 1999). Early studies suggested that the major reactive oxygen species responsible for the oxidation of methionine residues in proteins in aqueous solutions are hydroxyl radicals and hydrogen peroxide (Li *et al.* , 1995). Production of hydroxyl radicals is usually catalyzed by metal ions that bind to specific sites of a protein, i.e., the "site-specific" oxidation (Stadtman & Oliver, 1991; Stadtman, 1992; Li *et al.* , 1995). Oxidation then tends to ocur at or near the specific catalytic sites. Oxidation by hydrogen peroxide, on the other hand, can occur anywhere, and is thus "non-site-specific" oxidation (Cleland *et al.* , 1993; Li *et al.* , 1995).

In this thesis, the oxidation by peroxides of the amino acid methionine in proteins in aqueous formulations, i.e., the non-site-specific oxidation, is addressed via molecular simulations. A specific focus is placed on elucidating the fundamental mechanism of oxidation and the relationship of rates of oxidation to protein structures. The detailed knowdege and understanding derived from this study should be useful in developing better ways to control protein oxidation by peroxides.

1.2 The mechanism of "non-site-specific" oxidation of methionine residues by peroxides

Peroxide molecules can be introduced into a formulation of therapeutic proteins by a trace amount of impurities in the added excipients such as surfactants (Cleland *et al.* , 1993; Wang, 1999). As a result, oxidation of proteins in a forced condition of excess H_2O_2 is also one of the most commonly used models in order to understand protein oxidation in solutions (Levine *et al.* , 1996; Nabichi *et al.* , 1998; Patten *et al.* , 1999; Lu *et al.* , 1999; Griffiths & Cooney, 2002). Moreover, in the light-catalyzed and metal-catalyzed oxidation pathways, peroxide molecules are often present as one of the intermediates (Stadtman & Oliver, 1991; Cleland *et al.* , 1993; Stadtman, 1993; Nguyan, 1994; Li *et al.* , 1995), and may also contribute to the oxidative modification of the protein substrate. These facts make the mechanism by which

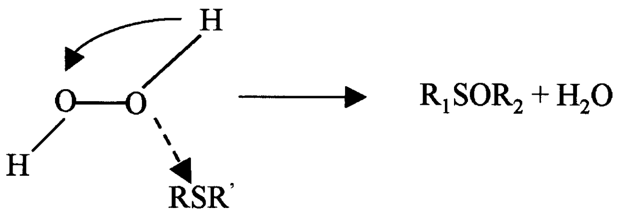


Figure 1-1: A mechanism for the oxidation of an organic sulfide by hydrogen peroxide.

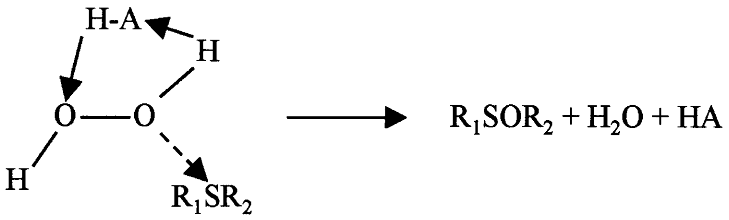


Figure 1-2: The acid-catabolized mechanism for the oxidation of an organic sulfide by hydrogen peroxide.

peroxide molecules oxidize methionine residue an extremely important subject to study in order to understand the oxidation of protein molecules.

Before 1968, the oxygen transfer from hydroperoxides to nucleophilic substrates in solution was generally accepted to be an S_N2 type displacement reaction as shown in Figure 1-1 (Edwards, 1960; Ranky & Nelson, 1961). In this mechanism, the transfer of oxygen is associated with a hydrogen shift to the distal oxygen. In 1968, Dankeff *et al.* (Dankleff *et al.* , 1968), proposed the mechanism shown in Figure 1-2, based on their data of solvent effects on the oxidation of thioxane by hydrogen peroxide. HA in Figure 1-2 is a general acid, for example, a solvent molecule, and was thought to serve as an intermediate agent which reduced the charge separation of the activation complex via *proton transfer*. Specifically, they found that the rates of oxidation in different solvents correlate better to solvent acidity (pKa) than to solvent polarity (dielectric constant). Moreover, transferring the reaction to an aprotic solvent lead to an increase in the reaction order of hydrogen from 1 to 2, strong evidence that the second hydrogen peroxide played the role of the generalized acid. The schematic picture in Figure 1-2 has been generally accepted to be the mechanism of oxidation of organic sulfides via peroxides (Bach *et al.* , 1991; Nguyan *et al.* , 1993; Nguyan, 1994; Bach *et al.* , 1994; Li *et al.* , 1995; Okajima, 2001; Ottolina & Carrea, 2001).

Starting in 1991, in order to evaluate the plausibility of the acid-catalyzed mechanism shown in Figure 1-2, and other hypothesized reaction mechanisms, *ab initio* studies on peroxide oxidation have been performed for both the hydrogen transfer reaction of hydrogen peroxide (Bach *et al.* , 1991; Bach *et al.* , 1994; Okajima, 2001) and for the oxidation of amines and sulfides with hydrogen peroxide (Bach *et al.* , 1991; Bach *et al.* , 1994; Ottolina & Carrea, 2001). The purpose of those studies was to develop a more detailed understanding of the oxidation mechanism. Important questions that were discussed include: what are the roles of the general acid? What are the essential reaction coordinates? What is the order by which the transfer of oxygen and hydrogen proceeds? And what are the predicted activation energies and how do they compare with experimental values? The last question was the most important in that proposed reaction mechanisms were accepted or eliminated depending on how their energy barrier compared to experimental values. For example, the direct 1,2 hydrogen transfer of H_2O_2 in a vacuum was found via *ab initio* computations to have a very high energy barrier (56 kcal/mol) (Pople *et al.* , 1983; Bach *et al.* , 1991), but typical observed activation barriers for peroxide oxidation of amines and organic sulfides in aqueous solutions are in the range of 10–20 kcal/mol (Dankleff *et al.* , 1968; Curci *et al.* , 1970). On this basis, the direct 1,2 hydrogen transfer of H_2O_2 was assumed not to be the correct mechanism. In general, however, *ab initio* calculations of the mechanism in Figure 1-2 yielded activation barriers in the range of 28–50 kcal/mol (Bach *et al.* , 1991; Bach *et al.* , 1994; Okajima, 2001; Ottolina & Carrea, 2001), all much higher than the range of experimental data (10∼20 kcal/mol). Therefore, Bach *at el.* concluded that general acid alone cannot catalyze the oxidation reaction (Bach

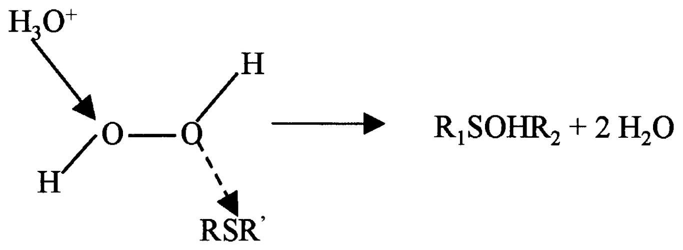


Figure 1-3: The acid-catabolized mechanism for the oxidation of an organic sulfide by hydrogen peroxide with the presence of a prontonated water molecule.

et al. , 1994). On the other hand, when both a protonated solvent and a general acid are present as shown in Figure 1-3, the calculated energy barriers lie between 5–15 kcal/mol, within the experiment range.

The mechanism shown in Figure 1-3 was, thus, concluded to be the reaction mechanism by (Bach *et al.* , 1994). Note that based on this mechanism (Figure 1-3), the oxidation rates should decrease with the increase of pH. However, subsequent to the proposal of this mechanism, it was found that the dependence on pH of the rate of oxidation of DMS (pH=2 to 10) (Amels *et al.* , 1997) and of methionine residues in human parathyroid hormone (pH=2 to 8) (Nabichi *et al.* , 1998) was negligible in aqueous solution. In general, experimental observations do not indicate a pH dependence of peroxide oxidation between pH=2 and pH=10. However, an increase in oxidation rate is observed when pH is very low (<2), indicating that the mechanism proposed by Bach et al. (Figure 1-3) could be valid at low pH values but is unlikely to occur at moderate pH values.

As a result, the existing discrepency regarding the mechanism of methionine oxidation by peroxides has prevented rational design of protein formulation. For example, if the mechanisms shown in Figure 1-2 and Figure 1-3 are valid, a protein molecule should be kept under basic condition, pH > 7, to prevent oxidation, but almost all protein formulations are designed between pH 4.5 and 7.5, including week acidic pH, to maintain stability (Powell, 1996). Therefore, it is necessary to develop detailed mechanism of oxidation in order to explain the experimental data and to develop rationales for the design of formulation against oxidation.

1.3 Oxidation of methionine and protein conformation

Methionine residues in a protein molecules are usually oxidized at different rates of oxidation (Yao *et al.* , 1996; Lu *et al.* , 1999; Nabichi *et al.* , 1998; Griffiths & Cooney, 2002). Differences in the oxidation rates of different methionine residues has been suggested to be due to the different structural environment of individual methionine residues (Cleland & Langer, 1994; Lu *et al.* , 1999; Griffiths & Cooney, 2002; Meyer *et al.* , 2002), however, the detailed mechanism of such correlation had not been estabilished.

Such difference in rates could be due to differences in interactions among residues in a protein molecule and differences in the interaction between the protein molecule and the solvent. It is also possible that the difference in rates is because the mechanism of oxidation is not the same at different methionine sites in a protein molecule. A better understanding of these interactions and mechanisms would lead to a better understanding of oxidation and how to control it. Moreover, having a useful tool to estimate the oxidation rates would be helpful. This property could be taken from the X-ray structure or from molecular dynamics simulations, but due to the thermodynamic fluctuations at a finite temperature, and it is unlikely that static X-ray data will be sufficient. Solvent accessible area (Lee & Richards, 1971) (SAA) is traditionally used to characterize the degree of interaction between solvent molecules and sites in a protein. Since the access of the sulfur atom to oxidants in the solution is required for the oxidation reaction, assuming that the oxidation rate is proportional to SAA is the simplest model for the relashtionship between protein structure and the oxidation of methionine residues. However, it has been shown that SAA's of the sulfur atom of a methionine residue from the X-ray structure can not be used to describe the kinetic data of oxidation (Griffiths & Cooney, 2002; Chu *et al.* , 2004a). Therefore, further work is in need to understand the correlation of structural properties to the oxidation rates of methionine residues.

1.4 Goal and objectives

The goal of this thesis is to investigate the oxidation of methionine by peroxides in protein pharmaceuticals in an aqueous formulation. The focus is on developing a detailed understanding of oxidation at the molecular level in order to explain experimental data on oxidation and to develop rationales for the design of formulations to control oxidation. Molecular simulations are employed for this purpose. This is the first time that this kind of approach has yet been applied to the field of formulation design in biotechnology. Specifically, the objectves of this thesis are to:

1. Develop a mechanism for methionine oxidation by peroxides that is consistent with all experimental data.

2. Quantify the effects of the solution and the protein environment on the rates of oxidation of methionine.

3. Develop a correlation between structural properties of protein molecules and the rates of oxidation of methionine in protein pharmaceuitical, in order to be able to predict the tendency of a particular residue to be oxidized.

4. Develop computational methodologies to achieve the above objectives.

Chapter 2

Computational Methods

In this chapter, the principles of molecular simulations that are employed in this thesis are introduced. The purpose is not to reproduce the existing information in literature and textbooks but to summarize the methodologies. Computational methods that are developed during the course of the thesis work, on the other hand, are described in greater detail.

2.1 Simulation of molecular systems

In computer simulations of molecular systems, microscopic models (10–100,000 atoms) are first built in order to compute the behavior of systems at the laboratory scale (order of 10^{23} atoms). If the system of interest is in a condense phase, periodic boundary conditions are usually applied (Allen & Tildesley, 1987; Frenkel & Smit, 1996). The intramolecular and intermolecular interactions in the microscopic model, i.e., the potential energy function, V, also need to be defined. Such interactions can be described at different levels of theories depending on the properties of interest. Two major categories for defining energy functions of chemical and biological systems are classical force fields (Brooks *et al.* , 1983; Pearlman *et al.* , 1995; Jorgensen, 1998) and *ab initio* quantum mechanical methods based on the Bohn-Oppenheimer approximation to separate nuclear and electronic coordinates (Szabo & Ostlund, 1989; Parr & Yang, 1989). The classical force fields of molecular mechanics (MM), are based on predefined functional forms and are fit from experimental data and/or from *ab initio* electronic structure calculations. A system size up to 100,000 atoms can be handled by modern computers using MM force fields. *Ab initio* methods, including methods based on density functional theory, solve approximately the Schrödinger equation without adjustable parameters (QM), but are limited to a system size of ~10–200 atoms by currently available computational power. The MM methods are efficient for describing phenomena that depend only on interactions with well-defined functional forms (for example, vibration of chemical bonds, electrostatic, and van der Waals interactions) but are not suitable for describing processes such as the breaking and formation of chemical bonds, for which the QM methods are appropriate. In order to keep the efficiency of MM methods in treating interactions among molecules and also be able to describe chemical reactions at the same time, combined quantum mechanical and molecular mechanical (QM/MM) methods have been developed (Gao & Truhlar, 2002; Das *et al.* , 2002; Gao *et al.* , 2003). In all cases, the starting point is to define the potential energy as a function of the coordinates of atoms or the degrees of freedom in the system:

$$V = f(\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N) \ , \tag{2.1}$$

where N is the number of atoms in the molecular model and $\vec{r}_i$ is the position vector of the atom i.

After the potential energy function is defined, the Hamiltonian and the equations of motion describing the dynamics of the system can then be defined:

$$H = \sum_{i=1}^{n} \frac{1}{2} \frac{\vec{p}_i \cdot \vec{p}_i}{m_i} + V(\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N) \ , \tag{2.2}$$

$$\begin{aligned} \dot{\vec{r}}_i &= \frac{\partial H}{\partial \vec{p}_i} \\ \dot{\vec{p}}_i &= -\frac{\partial H}{\partial \vec{r}_i} \ . \end{aligned} \tag{2.3}$$

In Eq. (2.2) and (2.3), $\vec{p}_i$ is the momentum vector, and m_i is the mass of the atom i. With the potential energy function and the equations of motion, equilibrium properties of the system can be obtained by sampling the configuration space of the system either by molecular dynamics (MD) or Monte Carlo (MC) simulations according the principles of statistical mechanics. Dynamical properties such as response functions are more readily available from MD simulations via linear response theory (McQuarrie, 1976).

2.2 Transition state theory

In order to link the phenomenological rate law of a transition process observed experimentally to microscopic mechanics, transition state theory (TST) has been developed (Eyring, 1935; Evans & Polanyi, 1935). TST is based on two assumptions. The first is the time-scale separation assumption that there exists a dividing surface, the transition state (TS), which can be described by a reaction coordinate (RC), ξ, for a process and the time scale for the activation of ξ is much slower than the time scale of other degrees of freedom such that they can be assumed to reach equilibrium even at the TS. The second assumption is the no-recrossing assumption that as long as the system has passed the TS, the system does not cross back. Via TST with the correction of the no-recrossing assumption, the rate constant, k, of a transition process can be represented as:

$$k = \nu \frac{k_B T}{h} \exp\left(-\Delta G^{\ddagger}\right) , \tag{2.4}$$

where ν is the transmission coefficient correcting for the dynamic recrossing effects, k_B is the Boltzmann constant, T is the temperature, h is the Planck's constant, and $\Delta G^{\ddagger}$ is the activation free energy for the system to reach the transition state. The TST can be further generalized to include quantum mechanical effects (Gao & Truhlar, 2002; Gao *et al.* , 2003), but this is not within the scope of this thesis.

For reactions in the gas phase, the transition states usually correspond to saddle points on the potential energy surface (PES). A saddle point is located at the maximum along one degree of freedom, but minimum along all others. By calculating the Hessian, or the second derivatives of V, at the saddle point, the activation free energy can be obtained based on a quasi-harmonic approximation (McQuarrie, 1976). However, for reactions occurring in condensed phases, especially in the liquid phase, finding a correct reaction coordinate and locating the transition state can be very difficult, since many degrees of freedom are involved and coupled. The coupling to the surrounding environment of a reaction coordinate can make the transmission coefficient an important factor in calculating the rate constant according to Eq.(2.2) (Gao & Truhlar, 2002; Nam *et al.* , 2004; Garcia-Viloca *et al.* , 2004). Moreover, the quasi-harmonic approximation of the PES is no longer correct for liquid-state systems, and the calculation of Hessian becomes intractable when many degrees of freedom are involved. Therefore, one needs to be careful when dealing with reactions in the liquid phase, and new computational methodologies are needed. In the next section, a newly developed method for finding reaction paths without calculating the Hessian is presented.

2.3 Finding reaction pathways in complicated systems

In this section, the development of a superlinear minimization scheme for the Nudged Elastic Band (NEB) method, which determines a minimum energy path (MEP) of a reaction via connecting intermediate "replicas" between the reactant and the product is presented. The minimization scheme is based on a quasi-Newton method: the adopted basis Newton-Raphson (ABNR) minimization scheme. In each step of ABNR minimization, the Newton-Raphson procedure is performed in a subspace with a user-defined dimension. The tangent directions of the path at a new ABNR step are determined self-consistently in the subspace. The acceleration of the proposed scheme over the quenched molecular dynamics minimization, the current practice for minimizing a path using NEB, is demonstrated in three nontrivial test cases: isomerization of an alanine dipeptide, α-helix to π-helix transition of an alanine decapeptide, and oxidation of dimethyl sulfide. New features are also added such that the distances between replicas can be defined in the root of mean squared (RMS) best-fit space with flexible weighting options. This offers a way to incorporate the effects of a mobile solvent in the process with a finite number of replicas. MEPs obtained from various initial structures can be used to investigate different proposed of reaction mechanisms, and the speedup of minimizing a path enhances the applicability of the NEB method. The combination of NEB force projection procedures, the flexible distance definition in the RMS best fit space with arbitrary weighting options, and the superlinear minimization scheme provides a framework to aid in the study of transition processes of biological molecules such as proteins.

2.3.1 Minimum energy paths and the nudged elastic band method

An useful approach to understand chemical and physical processes occurring in complicated systems is to find a minimum energy path (MEP) connecting two local minima on the potential energy surface of a system. MEP's also correspond to the most probable overdamped trajectories connecting the two stable or metastable states (Olender & Elber, 1997). The position of the maximum energy along the path is a first order saddle point and may be used to estimate the rate constant of the transition process using the transition state theory (TST) (Steinfeld, 1999). The MEP also defines a reaction coordinate along which the free energy profile of the process could be obtained for certain systems. Developing computational methods for finding MEPs has been an active research area (Elber & Karplus, 1987; Czerminski & Elber, 1990; Ulisky & Elber, 1990; Choi & Elber, 1991; Olender & Elber, 1997; Jónsson *et al.* , 1998; Henkelman & Jónsson, 2000; Henkelman *et al.* , 2000; Weinan *et al.* , 2002; Maragakis *et al.* , 2002). A very popular method for finding MEPs is the nudged elastic band (NEB) method (Jónsson *et al.* , 1998). The NEB method is based on the chain-of-states representation of a path (Elber & Karplus, 1987), but utilizes force projections along the path direction and the off-path directions (directions perpendicular to the path) in order to remove artifacts, such as corner cutting and sliding down, caused by the elastic-band restraints. If a path is fully optimized using NEB force projections, a number of equally spaced replicas connecting the reactant and the product with zero forces in the off-path directions, i.e., a minimum energy path, are obtained. Further enhancements of NEB include an improved tangent direction definition (Henkelman & Jónsson, 2000) and a more accurate estimation of the saddle point (Henkelman *et al.* , 2000). Therefore, the NEB method provides a framework to search for a MEP and the saddle point, and has been applied to a number of surface reactions and diffusion processes on surfaces (Eichler & Hafner, 1999; Ciobica *et al.* , 2000; Ciobica & van Santen, 2002; Fordell *et al.* , 2002; Trushin *et al.* , 2002; Xu & Mavrikakis, 2002).

NEB calculations, however, are extremely costly because of the difficulty in defining an objective function. Due to the projection of forces, it is difficult to define an accurate Lagrangian for the discrete systems in order to generate the NEB forces, and the implementation of more powerful (i.e. partial second-order) minimization schemes such as conjugate gradient and quasi-Newton methods may be problematic. A minimization scheme based on quenched Newtonian molecular dynamics is currently an accepted practice (Jónsson *et al.* , 1998). The MD minimization scheme is similar to a direct steepest descent method but kinetic energies are added and adjusted to increase the efficiency of minimization. Although the MD minimization works properly in most cases, convergence is very slow. Thus, a method with a super-linear convergence rate is highly desired. Minimization schemes such as conjugate gradient and quasi-Newtonian methods have better convergence behavior and require fewer energy calculations to locate a minimum (Press *et al.* , 1992; Payne *et al.* , 1992). For a chemical processes, performing as few energy calculations as possible is especially desirable since expensive quantum mechanical calculations may be involved.

In the next section, a super-linear minimization scheme for NEB is developed based on the adopted basis Newton-Raphson (ABNR) method implemented in the CHARMM program (Brooks *et al.* , 1983). First, the mathematical formulas of NEB are briefly reviewed based on the framework of the replica/path method in CHARMM. The current MD minimization method and the difficulties in implementing the conjugate gradient method because of the force projections are also discussed. The details of the quasi-Newton method, ABNR, are first described in terms of minimizing a scalar function. Then the procedures for optimizing a path using NEB are elaborated. The superlinear convergence rate of the scheme is tested by three examples: isomerization of an alanine dipeptide, α-helix to π-helix transition of an alanine decapeptide, and oxidation of dimethyl-sulfide (DMS) by hydrogen peroxide with three water molecules.

2.3.2 Development of a superlinear minimization scheme for NEB

The elastic band formulation

The objective function of the chain-of-states representation of a path employed in this study is:

$$S(\vec{r}_0, \ldots, \vec{r}_N) = \sum_{i=0}^{n} V(\vec{r}_i) + \sum_{i=0}^{n-1} \frac{1}{2} k_s \left(\Delta l_i^{RMS} - \Delta l^{RMS}\right)^2, \tag{2.5}$$

where N is the number of replicas connecting a reactant, $i = 0$, to a product, $i = N$. A spring (elastic band) is put in between each pair of neighboring replicas. The distance between replica $i-1$ and i, is denoted as Δl_i , and Δl is the average distance between replicas. In the calculation of Δl_i, it is desirable to take out contributions from rigid body translation and rotation of the replicas. One way is to fix an atom in each replica, constrain another atom moving along a line, and constrain a third atom to move only in a plane (Jónsson *et al.* , 1998). A more flexible way is to define the distance in the RMS best-fit space. In the best-fit procedure, the weighted centers of two replicas are first overlapped. Then a 3×3 rotational matrix, U, is obtained by minimizing the root-mean-square of the weighted distance between the two replicas:

$$\Delta l_i^{RMS} = \left| \sum_{j=1}^{Natoms} w_j \left(\vec{r}_{i,j} - U \cdot \vec{r}_{i+1,j}\right)^2 / \sum_{j=1}^{Natoms} w_j \right|. \tag{2.6}$$

The optimal rotation matrix is solved analytically using a Lagrangian. A weighting factor, w_j, can be assigned to each atom in the system (often equal to the mass of each atom). This additional flexibility is very useful for biological systems such as enzymatic reactions. Atoms at the active site can be assigned with larger weights to increase the resolution of important processes. Water molecules away from the active site can be assigned a zero weighing factor and thus be treated as a perturbation. This is important because the translation and rotation of mobile water molecules can cause large contributions to the RMS distance between replicas without having a significant effect on the MEP and thus decrease the resolution of the MEP.

In applying the best-fit procedure to the optimization of a path, derivatives of the rotational matrix, U, do not contribute to the gradient, since U is a stationary point of the constraint equation determined using a Lagrangian (Kabsch, 1976). However, there is a special case that may need to be treated in the gradient of the RMS best-fit procedure. This occurs when the selected atoms fit the mirror image of the reference coordinates better than the reference set itself (i.e. the determinant of the R matrix in Eq.(2.6) in (Kabsch, 1976) is negative) and when the eigenvalues of the $R^T R$ matrix become degenerate (i.e. two equally good RMS best-fit solutions). In this case, the eigenvalues and eigenvectors of $R^T R$ give the solution of the U matrix. This is a rather rare event that we have seen only in model systems involving few atoms at high temperatures, in which case there can be sudden and large changes of the forces of the system during a minimization or a molecular dynamic simulation. In order to cure this event, we monitor the sign and the magnitudes of the eigenvalues of the $R^T R$ matrix during a simulation, and use a smooth quadratic switching function to turn off the contributions of the nearly degenerate eigenvectors to the solution of U when the determinant of R is negative. Note that the use of the switching function is not an exact solution to the constraint problem, but has been adopted as an expedient measure for a rather rare event.

The nudged elastic band method

To perform the force projections based on the NEB method (Jónsson *et al.* , 1998), a tangent vector along the path has to be defined for each replica. For example, the tangent direction at replica i can be defined as:

$$\vec{\tau}_i = \mathrm{NORM}\left(\vec{w}_j \cdot \left(\vec{r}_{i+1,j}^{\,RMS\rightarrow j} - \vec{r}_{i-1,j}^{\,RMS\rightarrow j}\right)\right). \tag{2.7}$$

The "$RMS \rightarrow i$" superscript indicates that the neighboring replica is best fitted to replica i according to Eq.(2.6). "NORM" is the normalization operator. An alternate definition based on the relative energies of neighboring replicas has been proposed to have better stability, leading to a smoother path (Henkelman & Jónsson, 2000), and our results are consistent with this. The forces on each replica in the framework of NEB are:

$$\vec{F}_i = \vec{F}_i^{\perp} + \vec{F}_i^{\|}, \tag{2.8}$$

$$\vec{F}_i^{\perp} = -\vec{\nabla} V \cdot \left(1 - \vec{\tau}_i \vec{\tau}_i\right), \tag{2.9}$$

$$\vec{F}_i^{\|} = -\vec{\nabla}\left(\sum_{i=1}^{N-1} \frac{1}{2} k_s \left(\Delta l_j^{RMS} - \Delta l^{RMS}\right)\right), \tag{2.10}$$

where $\vec{F}_i^{\perp}$ is the force component in the off-path direction (perpendicular to the tangent direction), and comes only from the potential energy, not from the elastic bands. $\vec{F}_i^{\|}$ is the component in the tangent direction and comes only from the elastic bands. In Eq.(2.9), the spring forces do not interfere with the true forces in the off-path directions. This prevents the corner-cutting problem. The true forces also do not compete with the spring forces along the tangent direction in Eq.(2.10). This cures the sliding-down problem. A stationary point of the NEB forces therefore locates a MEP.

Difficulties in developing a super-linear minimization scheme for the NEB method

Due to the projection of force components in certain directions, it is not possible to define a Lagrangian to generate forces in Eq.(2.8)–Eq.(2.10), since $\vec{\tau}_i$ also depends on the coordinates. Note that the mixed partial second derivatives are not equal. Although in the continuum limit with an infinite number of replicas, a Lagrangian can be defined to approach the NEB forces asymptotically (Jónsson *et al.* , 1998), the definition for a discrete path to generate NEB forces is still not clear. On the other hand, the scalar work, $\sum_i \left|F_i^{\perp}\right| |\Delta l_i|$, that was derived from the Brownian dynamics (Olender & Elber, 1997), can be used as the cost function for minimizing a path. Using this quantity, the system could be optimized using gradient-based minimization methods. However, the gradient of the objective function requires the calculation of Hessian for each replica, and this becomes intractable for large systems.

Direct line minimization of NEB forces along a steepest descent direction would be another approach, but this would require updating the tangent direction along the line search and performing the force projection "on the fly" during the line minimization. The problem with this approach, however, is that the NEB forces cannot be used to construct a set of conjugate directions if the target function (Lagrangian) is not well-defined (Press *et al.* , 1992). As a result, implementation of a conjugate gradient minimization scheme for NEB is, at best, not straightforward. Currently, in order to avoid the problems described above, minimization using quenched molecular dynamics integration schemes such as the velocity Verlet algorithm is commonly employed to optimize a path using NEB. The quenched MD minimization method is reduced to a steepest descent scheme if the velocity is set to zero at each time step (i.e. fully quenched), but the velocity damping can be adjusted during the minimization to increase the convergence rate.

The adopted basis Newton Raphson (ABNR) minimization scheme

Although the MD minimization scheme works for most cases, the convergence rate is considerably slower than it would be for super-linear methods such as Newton-Raphson or conjugate gradient. Therefore, it is highly desirable to have such a minimization scheme for NEB. In this section, we describe an efficient quasi-Newton method, the Adopted Basis Newton Raphson method (ABNR). This method was developed originally by David J. States, and is used widely for geometry optimization of biological molecules, and has been available in the CHARMM program (Brooks *et al.* , 1983) since 1982. Here, we describe the ABNR in some detail, and then followed by the devised procedures for optimizing a path using NEB.

We start from the Newton-Raphson scheme for minimizing a multi-variable function with M variables:

$$\vec{\nabla} V\left(\vec{r}_{k+1}\right) \approx \vec{\nabla} V\left(\vec{r}_{k}\right) + \underline{\underline{A}}_{k} \cdot \left(\vec{r}_{k+1} - \vec{r}_{k}\right) = 0, \tag{2.11}$$

$$\underline{\underline{A}}_{k} = \vec{\nabla}\vec{\nabla}\left(\vec{r}_{k+1}\right), \tag{2.12}$$

where the subscript k denotes the iteration step. In the ABNR minimization, the Newton-Raphson (NR) scheme is only applied in a subspace of dimension M_s and is combined with a steepest-descent (SD) step:

$$\vec{r}_{k+1} = \vec{r}_{k} + \Delta\vec{r}_{k}^{\,NR} + \Delta\vec{r}_{k}^{\,SD}. \tag{2.13}$$

$$\Delta\vec{r}_{k}^{\,SD} = \frac{\vec{F}_{k}}{\left|\vec{F}_{k}\right|}\Delta s_{k}, \text{ and } \vec{F}_{k} = -\vec{\nabla} V\left(\vec{r}_{k} + \Delta\vec{r}_{k}^{\,NR}\right). \tag{2.14}$$

The step size is adjusted during the minimization according to the change of the target function. In the beginning stage of an ABNR minimization, only the SD procedure is conducted, i.e., $\Delta\vec{r}_k^{NR} = 0$ in Eq.(2.13). After several SD steps, position vectors of previous steps can be used as basis vectors:

$$\Delta\vec{r}_k^{NR} = \sum_{l=1}^{M_s} \Delta\vec{r}_{lk} c_{lk} + \sum_{l=1}^{M_s} (\vec{r}_{k-l} - \vec{r}_k)\, c_{lk}. \tag{2.15}$$

After a Taylor expansion at $\vec{r}_k$ with $\Delta\vec{r}_k^{NR}$ as the displacement vector, Eq.(2.11) gives:

$$\sum_{l'=1}^{M_s}\sum_{l=1}^{M_s} \Delta\vec{r}_{l'k} \cdot \left(\vec{\nabla}V(\vec{r}_{k-l}) - \nabla V(\vec{r}_k)\right) c_{lk} = -\sum_{l'=1}^{M_s} \Delta\vec{r}_{l'k} \cdot \nabla V(\vec{r}_k). \tag{2.16}$$

Eq.(2.16) is a set of M_s equations and can be solved by diagonalizing the $M_s \times M_s$ symmetric matrix. The solution vector, c_{kl}, is then used to determine the Newton-Raphson step, $\Delta\vec{r}_k^{NR}$ according to Eq.(2.15). The gradient vector at $\vec{r}_k + \Delta\vec{r}_k^{NR}$, is estimated as:

$$\vec{\nabla}V(\vec{r}_k + \Delta\vec{r}_k^{NR}) \approx \vec{\nabla}V(\vec{r}_k) + \sum_{l=1}^{M_s} \left(\vec{\nabla}V(\vec{r}_{k+1}) - \vec{\nabla}V(\vec{r}_k)\right) c_{lk}. \tag{2.17}$$

The SD displacement, $\Delta\vec{r}_k^{SR}$, is then determined via Eq.(2.14). The SD procedure adds an additional dimension in the new step and effectively updates the estimation of the local Hessian. Therefore, the Hessian-update steps in quasi-Newton minimization methods (Fletcher, 1981) is automatically incorporated. If any eigenvalue of Eq.(2.17) is negative, then the potential energy surface is concave downward in some direction and a NR procedure may lead to an increase of the objective function. In this case, the NR step is given up and only the SD procedure is performed. This approach also avoids the system from relaxing to stationary points other than local minimum, as may occur in the DIIS (direct inversion of the iterative subspace) (Farkas & Schlegel, 1999; Farkas & Schlegel, 2002) method. After c_{kl} are determined, energy and forces are calculated at the new position. This process is repeated until the convergence criteria are satisfied. In each step of ABNR, the Newton-Raphson procedure is performed in a space of dimension M_s. The Taylor expansion of Eq.(2.15) is always performed at the lowest energy configuration in the subspace. If the function is quadratic, then the minimum in the subspace is located, provided that the basis vectors are linearly independent. A new dimension is added by the SD step, the minimum of which will be found in the next NR procedure. For a system of N harmonic oscillators, $N + 1$ ABNR steps can locate a minimum. A subspace of dimension 5 suffices for satisfactory performance in most cases.

In ABNR minimization, the step size is adjusted based on the sign of the dot product:

$$\vec{\nabla} V\big(\vec{r}_k + \Delta\vec{r}_k^{NR}\big) \cdot \vec{\nabla} V\big(\vec{r}_k + \Delta\vec{r}_k^{NR} + \Delta\vec{r}_k^{SD}\big). \tag{2.18}$$

$\vec{\nabla} V\big(\vec{r}_k + \Delta\vec{r}_k^{NR}\big)$ is estimated by Eq.(2.17), and $\vec{\nabla} V\big(\vec{r}_k + \Delta\vec{r}_k^{NR} + \Delta\vec{r}_k^{SD}\big)$ is the gradient calculated at the new position. If the dot product is positive, the SD step on top of the NR procedure is moving toward the minimum in the new dimension, and the step size is increased. If the dot product is negative, the SD step overshoots the minimum, and the step size is decreased. In general, ABNR requires roughly the same number of iterations as conjugate gradient methods but is more efficient by factor of 3, because no line search is required.

Extending ABNR to minimize a path using NEB

In this section, we present a set of procedures to rapidly optimize a path using NEB, based on the framework of ABNR. If there are N replicas and each replica has M degrees of freedom, the overall dimension of the replicated system is $N \times M$. Following Eq.(2.14), the steepest-descent step is defined as:

$$\Delta\vec{R}_k^{SD} = \frac{\vec{F}_k^{\perp}}{\left|\vec{F}_k^{\perp}\right|}\Delta s_k + \frac{\vec{F}_k^{\parallel}}{\left|\vec{F}_k^{\parallel}\right|}\Delta s_k. \tag{2.19}$$

Similarly, the Newton-Raphson equations are:

$$\vec{F}_i^{\perp} = -\vec{\nabla}V(\vec{R}) \cdot (1 - \vec{\tau}\vec{\tau}), \tag{2.20}$$

$$\vec{F}_i^{\parallel} = -\vec{\nabla}V_s(\vec{R}) \cdot (\vec{\tau}\vec{\tau}). \tag{2.21}$$

The Newton-Raphson displacement at the k-th step is also expanded in a subspace of dimension M_s, constructed from previous steps:

$$\Delta\vec{R}_k^{NR} = \sum_{l=1}^{M_s} \Delta\vec{R}_{lk} c_{lk} + \sum_{l=1}^{M_s} \left(\vec{R}_{k-l} - \vec{R}_k\right) c_{lk}. \tag{2.22}$$

The expansion coefficients are obtained by solving:

$$\vec{\nabla}V\Big(\vec{R}_k + \Delta\vec{R}_k^{NR}\Big) \cdot \Big(1 - \vec{\tau}_{(\vec{R}_k+\Delta\vec{R}_k^{NR})}\vec{\tau}_{(\vec{R}_k+\Delta\vec{R}_k^{NR})}\Big) = 0, \tag{2.23}$$

$$\vec{\nabla}V_s\Big(\vec{R}_k + \Delta\vec{R}_k^{NR}\Big) \cdot \Big(\vec{\tau}_{(\vec{R}_k+\Delta\vec{R}_k^{NR})}\vec{\tau}_{(\vec{R}_k+\Delta\vec{R}_k^{NR})}\Big) = 0. \tag{2.24}$$

Eq.(2.20)–Eq.(2.21) are non-linear because the tangent direction also depend on the coefficient c_{lk}; as a result, defining a target function to generate the projected forces is apparently intractable. However, can still be approximated using Eq.(2.17), and Eq.(2.20)–Eq.(2.21) can then be solved self-consistently in the subspace of the expansion. Because of the equal-spacing constraints, two sets of reduced-space ($2 \times M_s$) equations need to be satisfied by c_{lk}. This results in an underdetermined set of equations, and the singular-value-decomposition (SVD) (Press *et al.* , 1992) scheme is used to solve for c_{lk}. To our knowledge, this is the first time that a self-consistent solution to find the tangent directions for optimizing a path with an NEB force projection is proposed.

The iteration of self-consistent solution (SCS) starts with $c_{lk} = 0$ and stops when a self-consistent solution of $c_{lk} = 0$ is obtained. The projected forces at $\left(\vec{R}_k + \Delta\vec{R}_k^{NR}\right)$ can then be calculated, and the SD displacement is added according to Eq.(2.19). This completes an ABNR minimization step of a path using the NEB method. The SCS solution usually converges in 3–10 iterations, using zeros as an initial guess. When the SCS does not converge, the NR procedure is given up and only the SD procedure is performed. Solutions which have a singular value below a lower limit are rejected in Eq.(2.20)–Eq.(2.21) in order to preserve linear dependence (Press *et al.* , 1992). Although these restrictions are important, optimal choices do not appear to very widely with system size or composition. Thus, hard coded rules may be acceptable.

In optimizing a path using ABNR, the NR expansion is still performed over the configuration with the lowest value of $S(\vec{R})$ in the subspace. $S(\vec{R})$ is also used to adjust the step size of the SD step. This choice assures that the system goes downhill at the beginning of the minimization where the system is far from the minimum and has a large magnitude of force vectors. Around the neighborhood of a "minimum-energy" path, choosing either the minimum gradient or the minimum function value has been observed to yield identical pathways. The dot products, $\vec{\nabla}V^{\perp}(\vec{R}_{k+1}) \cdot \left(\vec{R}_{k+1} - \vec{R}_k\right)$ and $\vec{\nabla}V_s^{\parallel}(\vec{R}_{k+1}) \cdot \left(\vec{R}_{k+1} - \vec{R}_k\right)$ become negative if an ABNR step overshoots the minimum, and the step size is decreased if either one is negative. The minimization procedure is then repeated until the magnitudes of the NEB forces in both tangent and off-path directions are less than the convergence criterion.

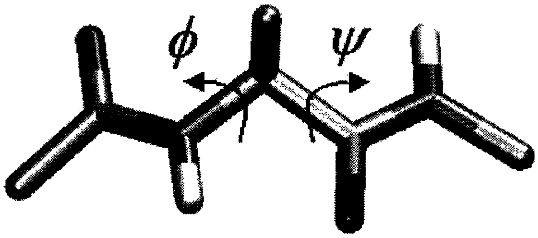


Figure 2-1: Stick representation of alanine dipeptide. The united atom model is used in the calculation. The backbone dihedral angles are labeled ϕ and ψ.

2.3.3 Performance of the superlinear minimization scheme

In order to demonstrate the convergence behavior of the ABNR minimization scheme presented in the previous section, three test cases are studied. The performance is compared with the molecular dynamics minimization method based on the velocity Verlet formulation and with a velocity updating scheme (Jónsson *et al.* , 1998). All calculations are performed using the CHARMM program (Brooks *et al.* , 1983). This method is available in CHARMM31.

Isomerization of an alanine dipeptide

The first test case is the isomerization of the alanine dipeptide(N-acetylalanyl-N-methyl-amide); a stick representation of the molecule is illustrated in Figure 2-1.

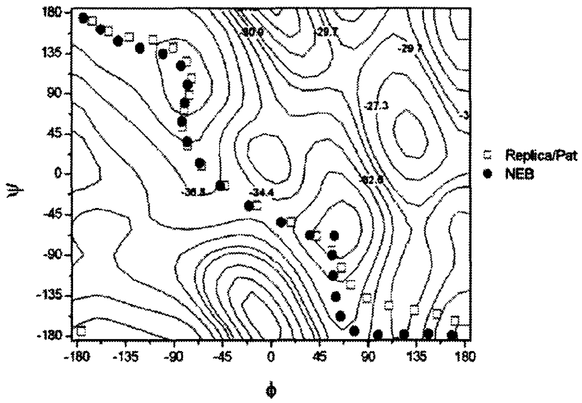


Figure 2-2: Reaction paths of the isomerization of alanine dipeptide obtained by the Replica/Path method and the nudged elastic method (NEB). The path is represented on a reduced potential energy surface of backbone ϕ and ψ angles.

The backbone dihedral angles ϕ and ψ can be used to describe the isomeric configurations of polypeptides (Mathews & van Holde, 1990). Isomerization of an alanine dipeptide has been of interest as a model system to study the dynamics and conformational changes of proteins (Elber & Karplus, 1987; Bolhuis *et al.* , 2000; Andricioaei *et al.* , 2003; Lavrich *et al.* , 2003). For simplicity, we used the united atom model, so that the molecule contains 12 atoms. The CHARMM19 force field is used for the calculation. A cyclic path composed of 25 replicas is generated by evenly rotate each backbone dihedral angle, ϕ and ψ, by 360 degrees. A reduced potential energy surface is constructed by restraining minimizations with a grid size of 2 degrees for both ϕ and ψ. The cyclic path is then minimized using NEB with the MD minimization method and the ABNR scheme.

In Figure 2-2, the MEP optimized using NEB and the path optimized without "nudging", and with an additional angle term to prevent the pathway from falling down into itself, is plotted on the reduced PES. The figure shows that a smooth path can be obtained by using the original Replica/Path method, but corner cutting may occur especially over the highly curved regions, such as the C_{ax} regime. After applying the NEB method, an MEP is clearly obtained. The convergence rate of the ABNR scheme and the MD method are compared in Figure 2-3. A spring constant

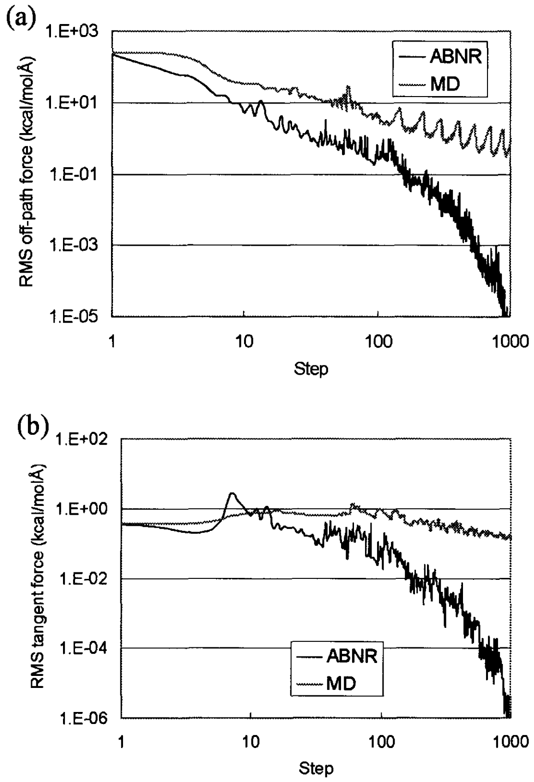


Figure 2-3: The convergence of the ABNR and the MD minimization methods for optimizing the reaction path of the isomerization of alanine dipeptide using the NEB method. The time step of the MD method is 0.001 ps. The RMS forces are plotted as functions of minimization steps on a log/log scale. (a) the off-path direction (b) the tangent direction.

of 1000 kcal/mol/Å^2 was used. The initial step size for ABNR was 0.02 Å and a time step of 0.001 ps was used for the MD minimization. The step size for the MD method is enlarged to 0.005 ps after 10 steps. In Figure 2-3, the RMS off-path and tangent forces are plotted with respect to the minimization step, indicating that the convergence rate using ABNR is much faster. A super-linear convergence rate for NEB is demonstrated. During the minimization of a path using the NEB method, fluctuations in the RMS forces are observed in both the ABNR and the MD methods, as Figure 2-3 shows. This is due to the fact that the tangent direction is also a dynamic variable and a displacement along the NEB force at one step may not decrease the magnitude of the force at the new step, because the tangent vector is changed as well. The change of tangent directions at the new step is taken into account in the NR step of ABNR minimization via self-consistent solutions, but not in the SD step. The phenomena of fluctuating forces justifies the use of $S(\vec{R})$ to increase or decrease the step-size in the ABNR scheme, since the step-size would decrease rapidly if the RMS force is used instead and would result in a worse convergence rate. A rapid decrease of step-size will occur after the system is very close the MEP, since then, no further decrease of $S(\vec{R})$ can be achieved. The rapid decrease of step-size is therefore a signal of convergence to the MEP. The MEP obtained via NEB does not depend on the value of the spring constant, k_s. However, the convergence rate does have some dependence on the value of k_s in the ABNR scheme. In Figure 2-4, the rates of convergence of the RMS off-path force using different spring constants are compared. For k_s=10 and 100 kcal/mol/Å^2, the convergence rate is about identical. As k_s reaches 10000 kcal/mol/Å^2, the convergence rate becomes slow. A large value of k_s results in high-frequency motions along the tangent directions. As a result, the frequency of positive $\vec{\nabla}V^{\perp}(\vec{R}_{k+1}) \cdot \left(\vec{R}_{k+1} - \vec{R}_k\right)$ or $\vec{\nabla}V_s^{\parallel}(\vec{R}_{k+1}) \cdot \left(\vec{R}_{k+1} - \vec{R}_k\right)$ may become much higher at a larger value of k_s, leading to a rapid decrease of step size and a slower convergence rate. Indeed, when k_s=10000, the total number of positive values is 864 in 1000 steps, and is about 400 times more than that of other cases. Moreover, a large k_s also leads to more frequent failure in converging the SCS. Therefore, the selection of the k_s value should be made so as not to introduce new high frequency motion

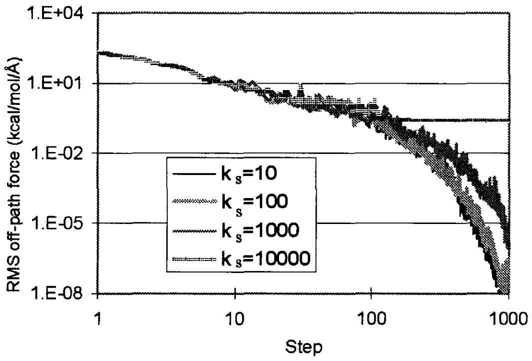


Figure 2-4: The convergence of the ABNR method for optimizing the reaction path of the isomerization of alanine dipeptide using the NEB method with different spring constants in units of kcal/mole/Å^2. The RMS off-path forces are plotted as functions of minimization steps on a log/log scale.

beyond those normally observed in the macromolecular system. There appears to be no significant penalty for choosing a rather small k_s value in this example.

α-helix to π-helix transformation of an alanine decapeptide

In this example, the transformation of an alanine decapeptide from a α-helical structure to a π-helix is studied as a model for processes that involves large-scale conformational changes. Currently, there is no tractable way of systematically finding all pathways, but the NEB/Replica-path method could be used to find different MEPs in order to gain an understanding of different pathways and as inputs into dynamic methods (Bolhuis *et al.* , 1998; Bolhuis *et al.* , 2002; Passerone & Parrinello, 2001).

The alanine decapeptide is terminated with methyl groups, and the CHARMM22 all-atom force field (MacKerell *et al.* , 1998) was used to describe the potential energy of the molecule (102 atoms). The optimized structures of the reactant (α-helix) and the product (π-helix) are shown in Figure 2-5. Calculations were performed in vacuum without energy cutoff. The minimized energy of the reactant (α-helix) is 67.8 kcal/mol and that of the product (π-helix) is 75.3 kcal/mol. The transformation of the alanine decapeptide from an α-helix to a π-helix involves the breaking and forming of hydrogen bonds, a general feature of conformational transformations of proteins.

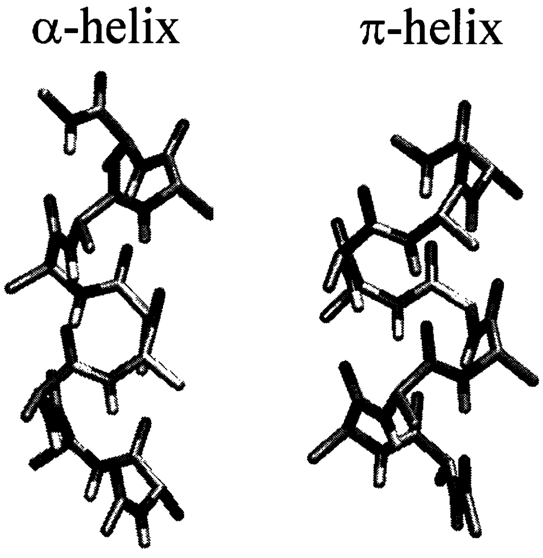


Figure 2-5: Stick representation of the α-helical and π-helical structures of alanine decapeptide. An all-atom model is used in the calculation. The non-polar hydrogen atoms are not shown.

A total of 51 replicas, including the reactant (α-helix) and the product (π-helix), are used to represent the transformation process. The initial structures were generated by linear interpolation between the reactant and the product. The initial path was first minimized with 50 steepest descent steps followed by 200 ABNR minimizations using the Replica/Path method with a spring constant of 100000 kcal/mol/Å^2 without NEB force projection. The resulting path wass taken as the starting point for NEB minimization using both ABNR and MD methods.

Figure 2-6 shows the RMS off-path force of the path as a function of minimization step. The preliminary Replica/Path minimizations result in an RMS off-path force of 0.24 kcal/mol/Å. A total of 1200 ABNR minimization steps were performed and reduced the final magnitude of the RMS force to 0.02 kcal/mol/Å. After about 880 ABNR steps, the step-size started to decrease rapidly. The rapid decrease of step size is a signal of convergence to the MEP, as mentioned earlier. For this system, the gradient did not continue to decrease with the number of steps; however, the direction of the gradient alternated as the step length went to zero, indicating a cusp

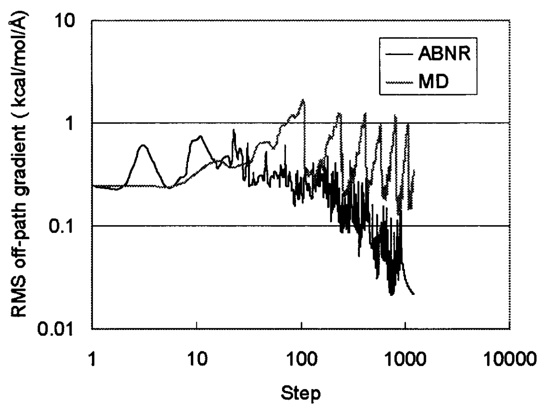


Figure 2-6: The convergence of the ABNR and the MD minimization methods in optimizing the reaction path of the α-helix to π-helix transition of alanine decapeptide using the NEB method. The off-path RMS forces are plotted as functions of minimization steps on a log/log scale.

like solution that has no zero gradient point. With the same number of MD minimization steps, however, the RMS off-path forces still fluctuated around the original magnitude of 0.2 kcal/mol/Åwith large amplitudes (1.0 kcal/mol/Å), but the tangent RMS forces remain relatively small (<0.01 kcal/mol/Å, results not shown). This indicates that the MD method is not efficient in relaxing along the off-path directions for this system. The total energies of the replicated system, Eq.(2.5), as a function of minimization steps are shown in Figure 2-7, which clearly indicates that the total energy function is dropping more rapidly with the ABNR minimization and that it plateaus after about 500 steps. The energy profiles obtained from ABNR and MD minimizations are shown in Figure 2-8. The MD method results in a path far away from the MEP after 1200 minimization steps. The MEP shows a single transition state at replica 29 and a local minimum at replica 38. Based on the fact that initial configurations were linear interpolations between the α helix and the π helix, we call this pathway a concerted transformation. The resulting MEP, however, indicates a sequential behavior, such that residues 4–9 start to transform earlier than residues 1–3. Residues 4–9 also reoriented significantly during the transformation with large deviation ($\pm$ 40 degree) of the backbone ϕ and ψ angles from the helical range (-40 –

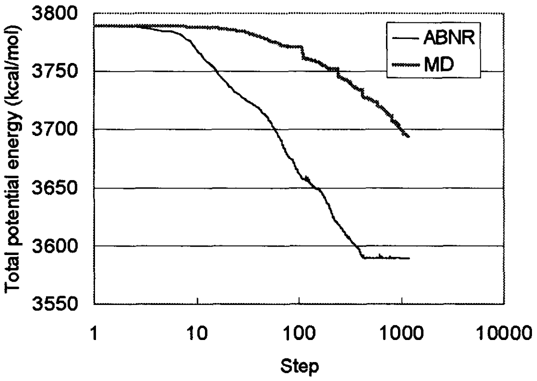


Figure 2-7: The convergence of the ABNR and the MD minimization methods for optimizing the reaction path of the α-helix to π-helix transition of alanine decapeptide using the NEB method. The total energies of the replicated system are plotted as functions of minimization steps on a semi-log scale.

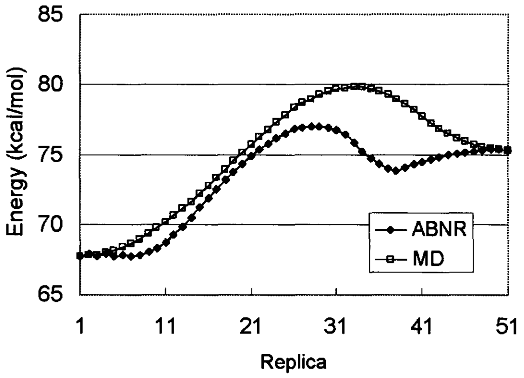


Figure 2-8: The energy profiles of the paths obtained by the ABNR and the MD methods of the α-helix to π-helix transition of alanine decapeptide using the linearly interpolated initial structure. The x-axis is the number of replicas and the y-axis is the energy in units of kcal/mol.

-70 degree). In the above calculations, the reaction was defined via all non-hydrogen atoms and the polar hydrogen atoms, and mass weighting was employed. The spring constant of the elastic bands was 500 kcal/mol/Å. A weighting factor of two was given to backbone atoms and a factor of one was used for the rest. Mass-weighting and a larger weighting factor for the backbone atoms were both found to be important in studying this process. Without mass-weighting or by using an equal weighting factor for all atoms resulted in slower convergence to the MEP. Therefore, it can be concluded that backbone atoms are more important than side chain atoms in defining the reaction coordinate of the α- to π-helical transformation of the alanine decapeptide and that the capability of defining distances between replicas in the RMS best-space with arbitrary weighting options is very helpful in studying processes involving large biological molecules such as proteins. Increasing the weighting factor of the backbone atoms to 5 resulted in the same pathway and similar convergence performance. In addition to the concerted pathway, other hypotheses of the transformation can be made and examined using the proposed minimization scheme. Another example is the zipup mechanism, i.e., the a-helix unwinds from the N-terminal to form the p-helix. The initial structure of the zipup mechanism was constructed by interpolating internal coordinates between the α-helix the π-helix from the N-terminal to the C-terminal. The MEP was then obtained using the same procedures described earlier. The resulting MEP indeed shows the zipup character, and the resulting energy profile is shown in Figure 2-9. Three transition states were located in the MEP using the zipup initial configuration. The activation barrier is 9.54 kcal/mol, slightly higher than that of the concerted case (9.23 kcal/mol). Similarly, by assigning the internal coordinates, one can also generate a zipdown initial configuration in which the α-helix unwinds from the C-terminal to form the π-helix. The resulting energy profile is also shown in Figure 2-9. The MEP using the zipdown initial structure does not contain a transition state with an energy much larger than thermal energy at room temperature. Therefore, out of the mechanisms investigated, the zipdown mechanism is the pathway that has the lowest activation energy. Analyzing the MEP's indicates that the MEP with larger backbone ϕ and ψ variations, estimated by RMS fluctuations of

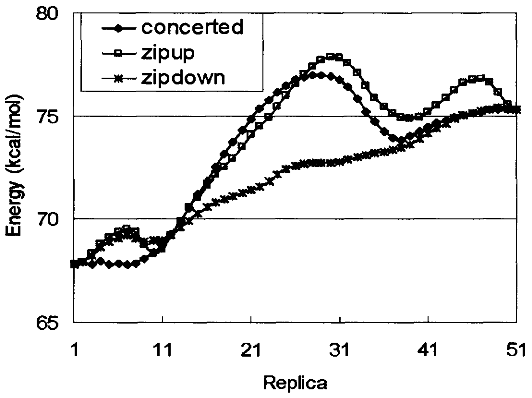
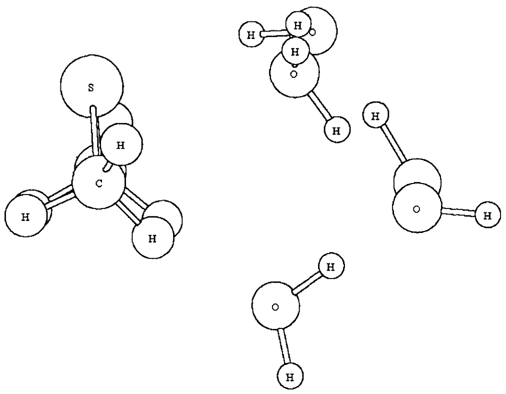


Figure 2-9: The energy profiles of the MEPs of the α-helix to π-helix transition of alanine decapeptide obtained with different initial paths.

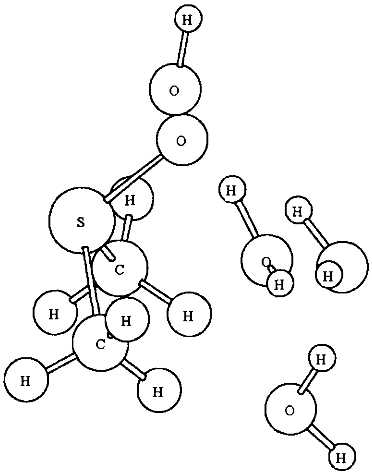
these angles along each path, has higher activation barrier for the transformation, i.e., zipup > concerted > zipdown. Simplistic descriptions such as concerted, zipup, and zipdown are useful to describe qualitatively a process such as in this example, and the molecular details involved can be found using the Replica/Path NEB method. As demonstrated in this example, the conformational transformation of a large molecule such as a protein can be very complicated. NEB force projections can be combined with the RMS best-fit distance definitions and the superlinear minimization scheme proposed to provide a way to systematically study such processes. Effects of solvent can be examined by using an implicit description of solvation or by placing the protein molecule in a simulation cell composed of explicit water molecules. Explicit water molecules need not all be replicated and can be set to respond to the average structure of the replicated protein system. Water molecules close to the protein can be replicated and assigned a non-zero weighting factor, if it is necessary to include them in the definition of the reaction coordinate. The reversible work of the reaction can then be obtained via integrating the potential of mean force along the MEP, leading to the activation free energy. Finally, the MEP's can also be used as initial guesses (non-dynamical) in transition path sampling simulations38.

Oxidation of dimethyl sulfide

In this example, a process involving the breaking and forming of chemical bonds is examined. Dimethyl sulfide can be oxidized to form dimethyl sulfoxide (DMSO) in aqueous solutions by hydrogen peroxide (Bach *et al.* , 1991; Bach *et al.* , 1994). The ground state structures of the reactant and product clusters are shown in Figure 2-10. Three water molecules are included in the system, in addition to DMS and H_2O_2. The reactant cluster involves the unoxidized DMS, and the product cluster is composed of DMSO and four water molecules. The reaction involves reorientation of water molecules, transfer of the oxygen atom of H_2O_2 to the sulfur atom of DMS, and proton transfer to form the leaving water molecule. All atoms were treated quantum mechanically using the B3LYP functional and 6-31G basis functions. The quantum mechanical total energy was calculated using the GAMESS-UK (Guest *et al.* , 1996) program and the CHARMM/GAMESS-UK interface was used to perform the calculations. 20 replicas including the reactant and the product were used to represent the path. The initial path was generated by linearly interpolating between the reactant and the product cluster. The weighting factor is 5 for the sulfur atom and the oxygen atoms of H_2O_2, and 1 for the rest of the atoms. Mass weighting is not applied. 50 steps of steepest descent minimization were first performed with regular Replica/Path method, in which the spring constant used was 10000 kcal/mol/Å. The path was then optimized with the NEB force projection with the ABNR scheme and the MD method. Figure 2-11 compares the convergence rate of the ABNR and MD methods. A total of 600 steps were performed for both methods. The force constant for the elastic bands is 5000 kcal/mol/Å for both methods. The time step for the MD method is 0.001 ps. The ABNR scheme we developed also has a much faster convergence rate in this example. With the ABNR scheme the final RMS off-path force is 0.01 kcal/mol/Å after 600 steps. Note that free minimization of a single QM structure using ABNR results in a similar terminal RMS gradient. Since the total-energy calculations become more expensive when a quantum mechanical Hamiltonian is involved, the enhancement of efficiency in employing the ABNR scheme is very im-



Reactant



Product

Figure 2-10: A ball and stick representation of the reactant cluster and the product cluster of the dimethyl sulfide oxidation reaction. The reactant cluster includes dimethyl sulfide, hydrogen peroxide, and three water molecules. The product cluster includes dimethyl sulfoxide and four water molecules.

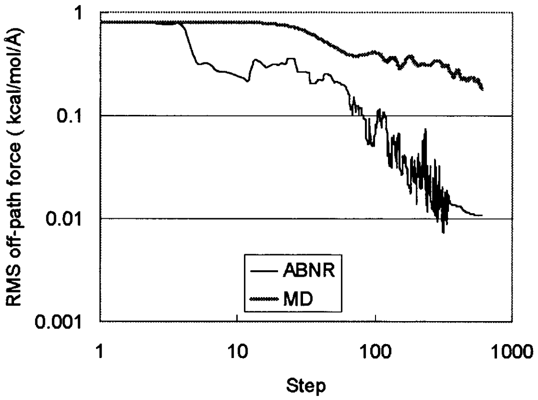


Figure 2-11: The convergence of the ABNR and the MD minimization method for optimizing the reaction path of the oxidation of DMS via H_2O_2 with three water molecules, using the NEB method. The off-path RMS forces are plotted as functions of minimization steps on a log/log scale.

portant. This MEP suggests a hydrogen transfer limiting mechanism, i.e., the major cost of the energy is the proton transfer of hydrogen peroxide to form water oxide, and the estimated activation energy is 21.9 kcal/mol. Other the hand, a different MEP can be found by starting from a different initial structure. The alternative initial structure is generated via linear interpolation with additional intermediates obtained via restraint minimizations, forcing the O-O bond length of H_2O_2 at 1.5, 1.8, and 2.0 Å, combined with the reactant and the product cluster. The MEP is obtained using the same minimization procedure and is denoted as $\mathrm{MEP}_r es$, and the previous MEP is denoted as MEP_{lint}. $\mathrm{MEP}_r es$ indicates that the breaking of the O-O bond in H_2O_2 and the formation of the S-O bond is the governing step. The transfer of hydrogen occurs after the transition state is passed. This example demonstrates that the faster convergence rate of the ABNR scheme can provide a more efficient use of the NEB method in studying chemical reactions. In addition to reactions of protein molecules, the information of MEP is also very useful for studying surface reactions and surface diffusion processes. The superlinear minimization method we proposed can also greatly enhance the efficiency of applying the NEB method in those fields.

2.3.4 Summary

In this section, a superlinear minimization scheme, useful for finding minimum energy pathways using the NEB method is presented. The procedure does not require a Lagrangian, the first derivative of which would generate the NEB forces, and the proposed scheme is therefore based on a quasi Newton-Raphson method, ABNR, instead of the conjugated gradient method. Moreover, a self-consistent solution of the tangent vectors along the path is included in predicting the new step during the minimization. The speedup of the new procedure compared with the currently used quenched MD method has been demonstrated in three non-trivial examples, isomerization of an alanine dipeptide, α-helix to π-helix transition of an alanine decapeptide, and the oxidation of dimethyl sulfide. The acceleration of convergence makes NEB a more powerful and reliable method to study transition processes, especially when the process needs to be described via a large number of degrees of freedom or when a quantum mechanical Hamiltonian is involved. The capability of defining distances between replicas in the discrete representation of a path in the RMS-fit space with a user-assigned weighting distribution is also found to be very helpful for studying flexible macromolecules, such as proteins. It also provides a way to incorporate the effects of mobile solvent molecules without decreasing the resolution of the process of interest in a discrete representation of the path. The combination of NEB force projection procedures, the flexible distance definition in the RMS best fit space with arbitrary weighting options, and the proposed superlinear minimization scheme thus provides a framework to study transition processes in complicated systems.

2.4 Free energy simulations of transition processes

The NEB (Nudged Elastic Band) method with the superlinear minimization scheme introduced in section 2.3 is very efficient for finding the MEPs (Minimum Energy Paths) and the activation energies at 0 K for reactions in large systems. However, the oxidation of methionine residues that is of interest in this thesis occurs in aqueous formulations at a finite temperature, and the free energy is the meaningful quantity to characterize the oxidation process and to compare with experimental data. In order to study transition processes with barriers much larger than k_BT using molecular simulations, it is necessary to sample the configuration space that is not accessible via standard MD or MC techniques. Therefore, specially designed computational methods are required to facilitate the calculation of free energies.

The free energy difference between two states characterized by a thermodynamic variable, λ (for example, volume and temperature), is obtained as:

$$W(\lambda_2) - W(\lambda_1) = \int_{\lambda_1}^{\lambda_1} d\lambda' \left\langle \frac{\partial H}{\partial \lambda'} \right\rangle_{\lambda'}, \tag{2.25}$$

where H is the Hamiltonian defined as in Eq.(2.2), and $\langle \ldots \rangle_{\lambda'}$ is the statistical average over the equilibrium ensemble corresponding to the parameter value λ'. Eq.(2.25) can also be used to compute the activation free energy along a reaction path defined by varying a reaction coordinate, ξ, and the free energy for a given value of $\xi = \xi'$ is:

$$W(\xi') = -k_BT \ln P_\xi(\xi'), \tag{2.26}$$

with the probability density, $P_\xi(\xi')$ given by:

$$P_\xi(\xi') = \langle \delta(\xi - \xi') \rangle . \tag{2.27}$$

Analogous to Eq.(2.25), the free energy difference between two values of ξ is:

$$W(\xi_2) - W(\xi_1) = \int_{\xi_1}^{\xi_1} d\xi' \left\langle \frac{\partial H}{\partial \xi} \right\rangle_{\xi'}^{\text{cond.}} . \tag{2.28}$$

The superscript of the integrand in Eq.(2.28) indicates that the distribution of the ensemble is conditional with the constraint: $\xi = \xi'$. After the reaction coordinate of a reaction is defined, the free energy profile can be obtained via techniques such as constrained MD and umbrella sampling according to Eq.(2.28) (Allen & Tildesley, 1987; Frenkel & Smit, 1996). These computational methods add biased potentials or constraints to the Hamiltonian, Eq.(2.2), in order to sample configuration space around the reaction coordinate in a sequence of simulations.

In this thesis, constrained MD procedures are used to compute the free energy barriers of methionine oxidation in different environments The SHAKE algorithm (Allen & Tildesley, 1987; Frenkel & Smit, 1996) is implemented in the CHARMM program (Brooks *et al.* , 1983) for constraining a reaction coordinate in MD simulations. Since in the SHAKE algorithm, both $\xi = \xi'$ and $\dot{\xi} = 0$, the ensemble distribution in the MD simulation is thus different from the desired distribution in Eq.(2.28). Fortunately, the correction of the additional $\dot{\xi} = 0$ constraint has been developed (Sprik & Ciccotti, 1998):

$$\frac{dW}{d\xi'} = \frac{\langle Z^{-1/2}[-\lambda + k_B T G]\rangle_{\xi'}}{\langle Z^{-1/2}\rangle_{\xi'}}, \tag{2.29}$$

where λ is the force in order to satisfy the $\xi = \xi'$ and $\dot{\xi} = 0$ constraints from the SHAKE procedure during an MD simulation, and Z and G are geometrical quantities which depend on the reaction coordinate, ξ:

$$Z = \sum_i \frac{1}{m_i} \left(\frac{\partial \xi}{\partial \vec{r}_i} \right)^2, \tag{2.30}$$

$$G = \frac{1}{Z^2} \sum_{i,j} \frac{1}{m_i m_j} \frac{\partial \xi}{\partial \vec{r}_i} \cdot \frac{\partial^2 \xi}{\partial \vec{r}_i \partial \vec{r}_j} \cdot \frac{\partial \xi}{\partial \vec{r}_j}, \tag{2.31}$$

where m_i is the mass of atom i and $\vec{r}_i$ is the position vector of atom i. Therefore, the activation free energy can be obtained via constrained MD simulations according to Eq.(2.28)–Eq.(2.31).

2.5 Dynamic effects of solvent molecules and other degrees of freedom

Although the activation free energy of transition processes in condensed phases can be computed as discussed in section 2.4, the determination of a suitable reaction coordinate can be very difficult, especially if the activation process is complicated, for example, when solvent degrees of freedom are involved (Bolhuis *et al.* , 2000;

Geissler *et al.* , 2001). In a liquid phase, the coupling of solvent molecules to a transition process not only affects the underlying free energy surface (Chandler, 1978; Garcia-Viloca *et al.* , 2004), but may also change the governing degree of freedom (Bolhuis *et al.* , 2000; Geissler *et al.* , 2001). Transition path sampling (TPS) (Bolhuis *et al.* , 1998; Bolhuis *et al.* , 2002) is a computational method that samples reaction pathways without an a priori definition of the reaction coordinate, but it is extremely computationally intensive to compute the free energy of a reaction via TPS.

Thus, in order to compute the free energy barrier of the oxidation of methionine as well as to understand the effects of solvent molecules, an approach that is a compromise between TPS and methods described earlier in the chapter is adapted in this thesis. Constrained MD simulations are still performed along a chosen reaction coordinate, but they are based on minimum energy paths (MEP) obtained by the method developed in section 2.3. The committor probabilities (as described below) of the configurations at the maximum of the free energy profile are then calculated to determine wether or not the transition state ensemble was found (Du *et al.* , 1998). These results not only broaden the current knowledge of the mechanism of oxidation of methionine residues in solution but also add to our understanding of transition path sampling methods by increasing our experience with their applications.

The committor probability, $p_B(r, t_s)$, or, p_B, is the probability that a trajectory, $\overline{\chi}(t_s)$, initiated from configuration r and of length t_s ends in state B. Here, the vector overbar is omitted on r for simplicity. Mathematically, p_B can be represented as (Bryant *et al.* , 2000; Bolhuis *et al.* , 2002):

$$p_B = p_B(r, t_s) = \sum_{\overline{\chi}(t_s)} \rho[\overline{\chi}(t_s)]\, \delta(\overline{\chi}(t_0) - r)\, h_B(\overline{\chi}(t_s)) \,/\, p(r), \tag{2.32}$$

where $\rho[\overline{\chi}(t_s)]$ is the probability density of the trajectory $\overline{\chi}(t_s)$, and h_B is the step function (1 for a state belong to B and 0 otherwise). States A and B can be defined by appropriate order parameters. A configuration r with a committor probability,

$$p_B = p_A = \frac{1}{2}, \tag{2.33}$$

is located on the dividing surface separating states A and B along the trajectory $\overline{\chi}(t_s)$, i.e., the transition state. The transition state ensemble is thus an ensemble of structures satisfying Eq.(2.33). This definition of transition states depends on properties of dynamic pathways but not on the free energy profile from a predefined reaction coordinate. Therefore, one of the uses of Eq.(2.33) is to test the dynamic effects of other degrees of freedom on a chosen reaction coordinate. For example, the committor probabilities of configurations at the maximum of a free energy profile of a reaction coordinate, $\xi = \xi'$, can be calculated by assigning different initial velocities according to a Boltzmann distribution. A committor probability distribution centered at 0.5 indicates that ξ is not coupled to the activation of other degrees of freedom and is thus a representative reaction coordinate of the transition process (Klosek *et al.* , 1991; Du *et al.* , 1998; Bolhuis *et al.* , 1998; Bryant *et al.* , 2000; Bolhuis *et al.* , 2002).

Chapter 3

The Mechanism of the Oxidation of Organic Sulfides by H_2O_2

As mentioned in section 1.2, a mechanistic understanding of the peroxide induced oxidation of organic sulfides in an aqueous solution has not yet been developed. In this chapter, this question is addressed by analyzing proposed oxidation mechanisms using high level *ab initio* simulations with the objective of finding a mechanism that is consistent with all experimental observations in moderate pH ranges. Both the specific effects of solvent molecules and the effects of the solvent as a polarizable dielectric continuum are carefully included. It turns out that including water molecules explicitly leads to a new reaction mechanism, which is consistent with all experimental data. Specifically, two reactions, hydrogen transfer of hydrogen peroxide to form water oxide and the oxidation of dimethyl sulfide (DMS) by hydrogen peroxide to form dimethyl sulfoxide, were studied as models of these processes in general. Solvent effects are included both via explicitly including water molecules and via the polarizable continuum model. The former was found to have a much more significant effect than the latter. When explicit water molecules are included, a mechanism different from those proposed in the literature was found. Specific interactions including hydrogen bonding with 2–3 water molecules can provide enough stabilization for the charge separation of the activation complex. The energy barrier of the oxidation of DMS by hydrogen peroxide was estimated to be 12.7 kcal/mol, within the experimental range of the oxidation of analogous compounds (10–20 kcal/mol). The major reaction coordinate of the reaction is the breaking of the O-O bond of H_2O_2 and the formation of the S-O bond, and the transfer of hydrogen to the distal oxygen of hydrogen peroxide occurs after the system has passed the transition state. Reaction barriers of the hydrogen transfer of H_2O_2 are an average of 10 kcal/mol or higher than the reaction barriers of the oxidation of DMS. Therefore, a two step oxidation mechanism in which the transfer of a hydrogen atom occurs first to form water oxide and the transfer of oxygen to a substrate occurs afterwards is unlikely to be correct. Our proposed oxidation mechanism does not suggest a pH dependence of the oxidation rates within a moderate range around neutral pH (i.e. under conditions in which hydronium and hydroxide ions do not participate directly in the reaction), and it agrees with experimental observations over moderate pH values. Also, without including a protonated solvent molecule, it has activation energies that correspond to measured activation energies.

3.1 Details of calculations

Ab initio calculations were performed using the GAUSSIAN 98 package (Frisch *et al.* , 1998). B3LYP (Becke's three-parameter functional (Berke, 1993)) with the 6-31++G(d,p) basis set was employed for geometry optimizations, transition-state searches and frequency calculations throughout this study; a scaling factor of 0.980630 was used for zero-point-energies (ZPEs) corrections. All transition states reported in this chapter have only one imaginary vibrational mode, and were found without imposing constraints during the process of geometry optimization. We have tested the accuracy of the B3LYP functional and the 6-31++G(d,p) basis set against higher levels of theories (MP2, MP4 (Frisch *et al.* , 1990a; Frisch *et al.* , 1990b) and CCSD, CCSD(T) (Purvis & Bartlett, 1982),(Scuseria & Schaefer, 1989)) and larger basis sets (6-31++G(3df,3dp), cc-pVDZ and aug-cc-pVDZ) for the 1,2 hydrogen transfer reaction of hydrogen peroxide (Bach *et al.* , 1991), and the results indicate that the B3LYP/6-31++G(d,p) theory is sufficient to predict accurate geometries and relative energetics. In order to treat electron correlation in a reliable manner, the MP4 method is used for single energy corrections. The SAC (scaling-all-correlation) (Fast *et al.* , 1999; Lynch & Truhlar, 2003) factor for the MP2//B3LYP6-31++G(d,p) calculations is found to be 1.1052 to scale to the MP4 calculations. Unless specifically mentioned, the energetics reported in this study are based on MP4//B3LYP/6-31++G(d,p) theory. The polarized continuum model (Miertus *et al.* , 1981; Miertus & Tomasi, 1982), PCM, was used to describe the long-range electrostatic polarization of the surrounding environment. Water is the only solvent considered (ε=80), and the united atom topological model (Barone *et al.* , 1997) was used to build up the solute cavity. The effects of solvent polarization were examined at the B3LYP/6-31G++(d,p) level.

3.2 Results and discussions

In this chapter, two reactions, the hydrogen transfer of H_2O_2 to form water oxide and the oxidation dimethyl sulfide (DMS) by H_2O_2, were studied using *ab initio* calculations in order to identify the mechanism by which organic sulfides are oxidized by H_2O_2 in solution. The hydrogen transfer of H_2O_2 corresponds to the rate-limiting step of a 2-step mechanism (Bach *et al.* , 1991; Bach *et al.* , 1994), the first step being the formation of water oxide and the second step being the transfer of oxygen from H_2OO to the nucleophile. The activation barrier of the second step was estimated to be less than 5 kcal/mol for the oxidation of DMS (Bach *et al.* , 1991). A correct mechanism must satisfy all experimental observations described earlier, i.e., the activation energies estimated by the mechanism should be between 10–20 kcal/mol and the mechanism should not suggest a pH dependence of oxidation rates between pH=2–10 (Amels *et al.* , 1997).

3.2.1 Hydrogen transfer of hydrogen peroxide to form water oxide

Water oxide, H_2OO, is a local energy minimum on the potential energy surface of hydrogen peroxide and is 46–50 kcal/mol higher in energy than H_2O_2. The reported activation energies of the hydrogen transfer reaction of H_2O_2 in earlier *ab initio* studies were in the range of 29–50 kcal/mol (Bach *et al.* , 1991; Okajima, 2001), depending on the number of water molecules involved. The lower end of reported energy barriers is still much higher than the measured activation barriers (10–20 kcal/mol) for the oxidation of organic sulfides. In this section, the potential role of the general acid in stabilizing the transition complex of the hydrogen transfer reaction of H_2O_2 is analyzed in detail. We would like to find out that whether or not the rate-limiting step in the oxidation of organic sulfides is the formation of water oxide. Various reaction configurations as well as the effects of electrostatic polarization on activation barriers of the hydrogen transfer reaction of H_2O_2 are examined. Figure 3-1 shows optimized geometries of the reactant cluster **1** and the transition state **TS-1** of the hydrogen transfer reaction with a single water molecule. The MP4/6-31++G(d,p)//B3LYP6-31++G(d,p) energy difference is 39.1 kcal/mol. This value is similar to the result in a previous study with an MP2/6-31G* calculation (42.7 kcal/mol) (Bach *et al.* , 1991). For the transition state **TS-1**, the imaginary vibrational mode corresponds to the 1,4 hydrogen transfer, and the water molecule serves as a catalyst by accepting the H_3 atom from O_1 and donating the H_7 atom to O_2. In order to test the sensitivity of the calculated energy barrier with respect to the number of basis sets, post-SCF calculations on the B3LYP/6-31++G(d,p) geometries were performed with a number of basis sets at the MP2 level; results are shown in Figure 3-2. The energy barriers are not very sensitive to the number of basis functions, although a slight decrease in ΔE with an increase in the number of basis functions was observed between basis sets with <100 basis functions and those with >100 basis functions. Differences are within 2.5 kcal/mol in all cases. Compared to the direct 1,2 hydrogen transfer of H_2O_2 ($\Delta E^{\ddagger}$=54.6 kcal/mol), the 1,4 hydrogen transfer pathway in the presence

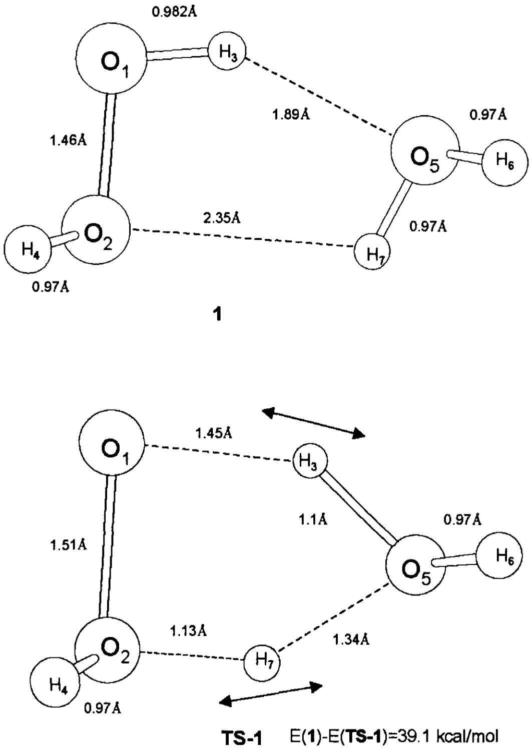


Figure 3-1: The hydrogen transfer reaction of hydrogen peroxide with a single water molecule. Reactant cluster **1** and transition state **TS-1** cluster were fully optimized at the B3LYP/6-31++G(d,p) level; energies were calculated at the MP4//B3LYP/6-31++G(d,p) level.

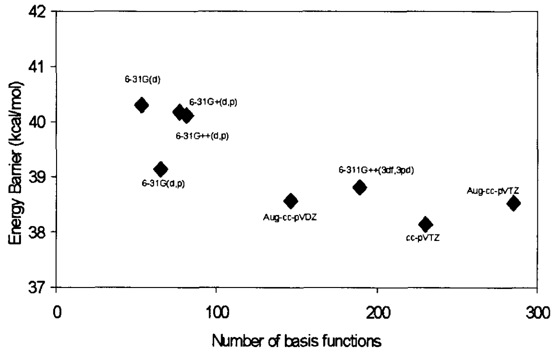
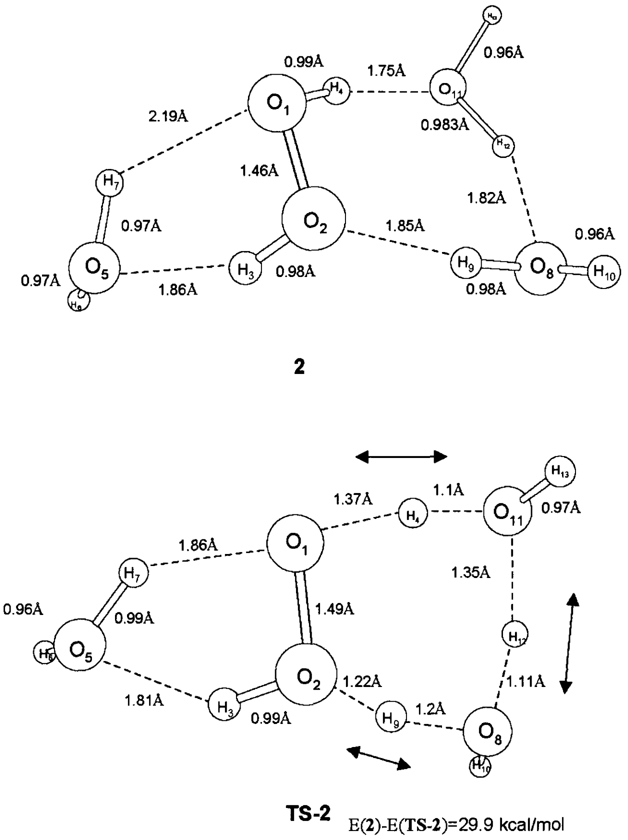


Figure 3-2: Sensitivity of the energy barrier between **1** and **TS-1** to the number of basis functions. Post-SCF calculations were performed at the MP2 level.

of a single water molecule has an activation energy that is lower by ~15 kcal/mol, indicating that solvent molecules can efficiently reduce the activation barrier of this reaction. We have also explored various reaction configurations involving 2 and 3 water molecules, and our results show that the second and the third water molecule do have significant effects on reducing the activation barrier. For example, the configurations with 2 water molecules have activation barriers that are 7.5–8.0 kcal/mol lower than the one water model, and the configurations with 3 water molecules have activation barriers that are 3–6 kcal/mole lower than the two water configurations. However, the fourth water molecule does not result in significant difference of activation barrier compared to the three water cases (difference <1 kcal/mol). Compared to the results of earlier theoretical studies of the hydrogen transfer reaction of H_2O_2 (the lowest value reported: 29.5 kcal/mol (Bach *et al.* , 1991; Okajima, 2001)), the activation energy (with ZPE) of reactant cluster **2** and **TS-2** shown in Figure 3-3 is lower (26.2 kcal/mol). The major cause of this reduction is the specific inclusion of water molecules with the reactant (Figure 3-3). Two water molecules are involved in the hydrogen-transfer process in **TS-2**, and this type of 1,6 of hydrogen transfer has been reported (Bach *et al.* , 1991; Okajima, 2001). The third water molecule in **TS-2** is not involved in the hydrogen transfer process, but it is hydrogen bonded to H_2O_2 and the presence of the third water molecule has reduced the barrier by ~5 kcal/mol. This type of configuration has not been explored in earlier studies. This result indicates that specific interactions with solvent molecules can have significant stabilization effects on the reaction. In addition, the effects of the long-range electrostatic polarization of the solvent (ϵ_r=0 in gas phase and ϵ_r=80 in aqueous solution) has been analyzed by incorporating Tomasi's polarizable continuum model (PCM) (Miertus *et al.* , 1981; Miertus & Tomasi, 1982) in locating the geometries of the reactant clusters and transition states and in total energy calculations. Including polarization in this way does not led to a significant difference (<1.5 kcal/mol) in reaction activation energies compared to those in vacuum. This result indicates that solvent molecules stabilize the hydrogen transfer reaction of H_2O_2 via specific interactions rather than dielectric polarization; it is also consistent with the experimental



2

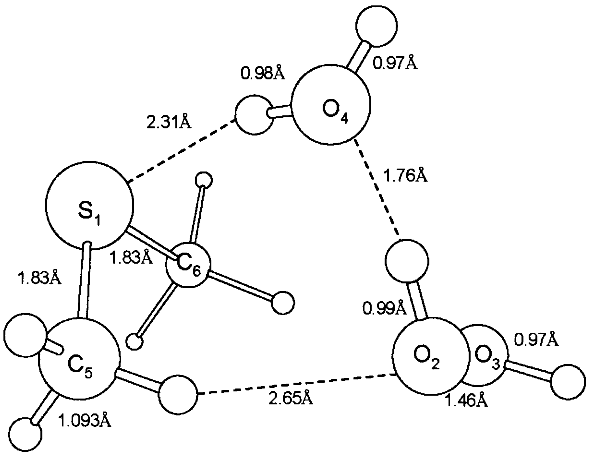
TS-2 E(2)-E(TS-2)=29.9 kcal/mol

Figure 3-3: The hydrogen transfer reaction of hydrogen peroxide with three water molecules, two at one side and the third one at the other side of H_2O_2. Two water molecules are involved in the concerted proton transfer at the transition state. Reactant cluster **2** and transition state **TS-2** were fully optimized at the B3LYP/6-31++G(d,p) level; energetics were calculated at the MP4//B3LYP/6-31++G(d,p) level.

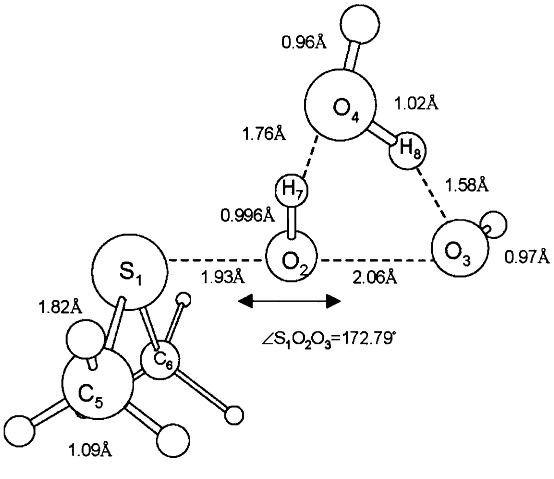
observation that the reaction rate of peroxide oxidation is insensitive to solvent dielectric constants (Dankleff *et al.* , 1968). Although the minimum activation energy obtained (26.2 kcal/mol) is lower than that obtained in previous studies, it is still too high compared to experimentally determined activation energies of the oxidation of organic sulfides by hydrogen peroxide (10–20 kcal/mol).

3.2.2 Oxidation of dimethyl sulfide(DMS)

Dimethyl sulfide (CH_3-S-CH_3) is an organic sulfide and can be oxidized by H_2O_2 to form dimethyl sulfoxide (DMSO) in aqueous solution (Amels *et al.* , 1997). DMS oxidation in aqueous solution is also of biological interest, because it represents part of the methionine side chain, and oxidation of the sulfur atom in methionine residues by peroxides is one of the major degradation pathways of therapeutic proteins. The starting point of this section is the fact that of the mechanisms previously proposed in the literature, as presented in Chapter 1 and as studied in the previous section, none of them are consistent with experimental data, including both the pH dependence of the rates of oxidation and the estimated activation energies. Aside from the mechanism proposed by Bach *et al.* (Bach *et al.* , 1994) (Figure 1-3), the activation energies estimated in previous theoretical studies are too high compared to experimental estimations. The problem with the mechanism of Bach *et al.* (Figure 1-3) is that it implies a marked pH dependence of rates of oxidation of organic sulfides via H_2O_2, but a subsequent experimental study reported that the rates of oxidation of DMS are independent of pH over the range of pH=2–10 in aqueous solutions (Amels *et al.* , 1997). Note that Figure 1-3 may still be the route of oxidation at low pH values, since an increase in the rates of oxidation of DMS was observed for pH<2 (Amels *et al.* , 1997). We wish to determine the mechanism of the oxidation process in a moderate pH range (2–10), including understanding the governing factors that bring about this reaction. Not only will this add to our understanding of a class of important chemical reactions, but also, it will enhance our ability to develop ideas to hinder this process, when desired. Figure 3-4 shows geometries optimized at the B3LYP/6-31++G(d,p) level for the reactant cluster **3** and the transition state **TS-3** for DMS oxidation by



3



TS-3

E(**3**)-E(**TS-3**)=24.4 kcal/mol

Figure 3-4: Oxidation of dimethyl sulfide with a single water molecule. Reactant cluster **3** and transition state **TS-3** cluster were fully optimized at the B3LYP/6-31++G(d,p) level; energies were calculated at the MP4//B3LYP/6-31++G(d,p) level.

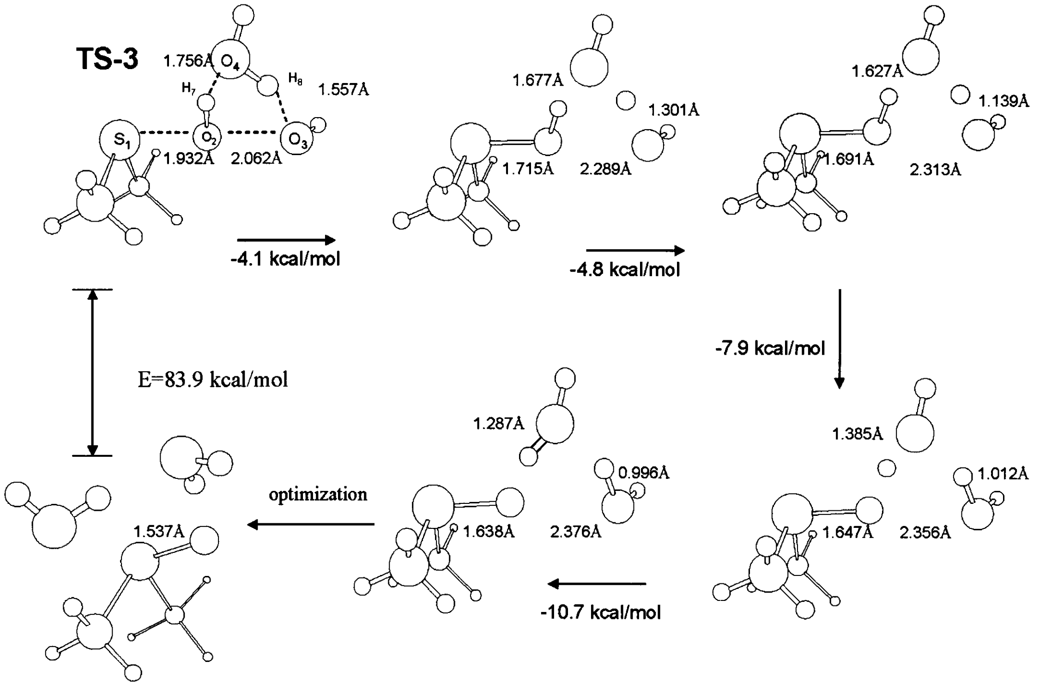


Figure 3-5: Oxidation of dimethyl sulfide with a single water molecule. Reactant cluster **3** and transition state **TS-3** cluster were fully optimized at the B3LYP/6-31++G(d,p) level; energies were calculated at the MP4//B3LYP/6-31++G(d,p) level.

H_2O_2 in the presence of a single water molecule. At **TS-3**, the reaction mode is mainly the transfer of O_2 to S_1 and does not exhibit significant hydrogen transfer character. This indicates that the proper reaction coordinate involves the S_1-O_2 distance and the O_2-O_3 distance, and that our picture is different from the mechanism found earlier for DMS oxidation where the transfer of hydrogen to form water oxide occurs before the transfer of oxygen (Bach *et al.* , 1994). A reaction path following analysis (Gonzalez & Sclegel, 1989; Gonzalez & Schlegel, 1990) shows that the transfer of hydrogen indeed occurs after the transfer of oxygen. Selected frames and corresponding energies are shown in Figure 3-5. The energy difference between the reactant cluster and the transition state in this model is 24.4 kcal/mol at the MP4//B3LYP/6-31++G(d,p) level. Although this value is still too high compared with experimental observations,

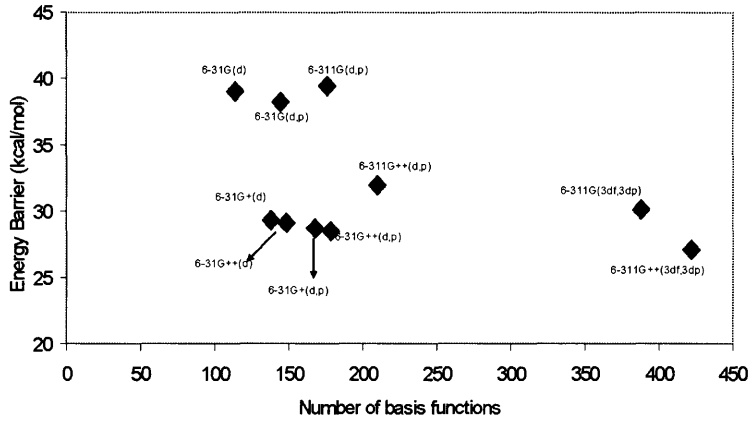
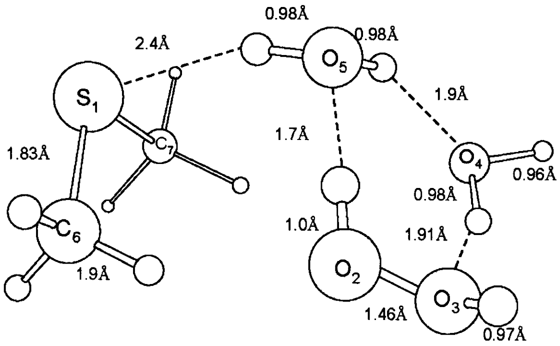


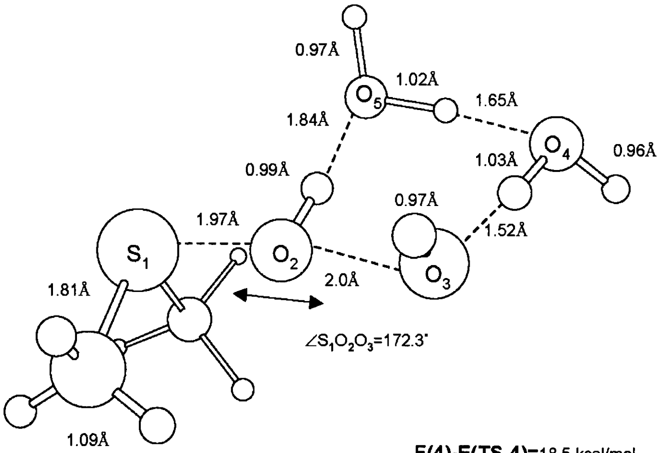
Figure 3-6: Sensitivity of the energy barrier between **3** and **TS-3**. Post-SCF calculations were performed at the MP2 level of theory for all data points.

it is lower than the activation energy of a similar reaction: oxidation of DMS with two H_2O_2 molecules calculated at the MP4//MP2/6-31G(d) level, in which the transfer of hydrogen occurs first to form water oxide, and the reported energy barrier of this process was 40.8 kcal/mol (Bach *et al.* , 1994). A sensitivity analysis of the reaction energy barriers with respect to the number basis functions was performed and is shown in Figure 3-6; all data points were calculated at the MP2 level of theory. By including diffuse functions with 6-31(d) and 6-31(d,p) basis functions, the calculated energy difference is decreased by ~10 kcal/mol. Among the cases investigated, only the 6-311G(3df,3dp) basis set can predict a reasonable energy barrier without diffuse functions (Frisch *et al.* , 1998). In conclusion, including diffuse functions can greatly enhance the performance of the 6-31G(d) and 6-31G(d,p) basis sets and are necessary to predict correct energy barriers of DMS oxidation by H_2O_2 with a water molecule, and the 6-31++G(d,p) basis set is sufficiently accurate compared to larger basis sets. The effects of the electrostatic polarization of the surrounding solvent were also analyzed with the PCM model for DMS oxidation with a single water molecule. The difference in activation energy between the unsolvated phase and the solvated phase is insignificant (0.2 kcal/mol). The insensitivity of activation energy to dielectric polarization of solvent for this reaction is similar to the hydrogen transfer reaction of

H_2O_2 as described earlier, and again, is also consistent with the experimental observation that the reaction rate of peroxide oxidation is insensitive to solvent dielectric constants (Dankleff *et al.* , 1968). We analyzed the charge distribution along the reaction path of **TS-3** using the Merz-Singh-Kollman (MSK) scheme (Frisch *et al.* , 1998) in order to understand the charge separation in the S_N2 oxidation of DMS by H_2O_2. The sulfur atom of DMS, S_1, has an atomic charge of -0.22 in the reactant cluster **3**, Figure 3-4. Both of the two oxygen atoms of H_2O_2 have similar atomic charges, -0.43 for O_2 and -0.48 for O_3, indicating the symmetric nature of hydrogen peroxide. As the hydrogen peroxide molecule approaches DMS (S_1-O_2=2.8 Å, S_1-O_2=4.2 Å in reactant cluster **3**), the atomic charges of O_2 and O_3 became different (O_2=-0.24, O_3=-0.45). During this part of reaction, the energy is increased by 9 kcal/mol. Note that the O_2-O_3 bond length remains near the equilibrium value of 1.46 Å at this point. In this part of reaction, the hydrogen peroxide and the water molecule rearrange such that the S_1, O_2 and O_3 atoms lie on a line, and the O_3-H_8 and O_4-H_7 hydrogen bonds are formed (reference to Figure 4). This arrangement results in a more significant change of the atomic charge on the transferring oxygen (-0.41 to -0.24) than on other atoms. The O_2-O_3 bond then starts to break. The negative charge on the distal oxygen, O_3, begins to increase, and the sulfur atom, S_1, begins to lose its negative charge. For the transferring oxygen, O_2, however, the MSK charge remains around the value of -0.24. At the transition state, **TS-3**, the charges are: S_1 (0.15), O_2 (-0.24), and O_3 (-0.83) (reactant, S_1 (-0.22), O_2 (-0.43), and O_3 (-0.48)). The atomic charge on the sulfur atom goes from negative in the reactant to positive in the transition state. This trend continues after passing the transition state. Although the values of the point charges depend on the method used for population analysis, the nature of charge separation and transfer is qualitatively the same if other population analysis methods were used, such as the Mulliken population analysis (Szabo & Ostlund, 1989) and the natural orbital analysis. The analysis of charge separation during the oxidation process also provides insight into how solvent molecules may stabilize the reaction. For example, hydrogen bonding to the more negative distal oxygen may stabilize the transition state structure more than the reactant cluster so that the activation energy can be decreased. Based on the experience of the H_2O_2 hydrogen transfer reaction, we anticipate that the number of water molecules in the theoretical model can dramatically change the reaction energy barrier. Three configurations were studied for the model containing two water molecules. For configurations 1 (**4** and **TS-4**) and 2 (**5** and **TS-5**), two water molecules are at the same side of H_2O_2, as shown in Figure 3-7 and Figure 3-8, respectively. The difference is that the two water molecules are at the trans side with respect to the methyl groups of DMS in the first configuration, (Figure 3-7, **4** and **TS-4**), and at the cis side in the second, (Figure 3-8, **5** and **TS-5**). For the third configuration, each water molecule is at one side of the hydrogen peroxide, (Figure 3-9 and **TS-6**). Geometries were optimized in the unsolvated systems only. Recall that both theoretical calculations (DMS oxidation with one water molecule) and experimental observations suggest that the oxidation rates of organic sulfides are not strong functions of the dielectric constant of the media (Dankleff *et al.* , 1968). We conclude that calculations performed in the unsolvated systems can provide adequate information for the reaction kinetics if enough water molecules are included explicitly. Reactant cluster **4** (Figure 3-7) has the lowest energy among the three reactant clusters. It is 0.1 kcal/mol lower than that in **5** (Figure 3-8), and 1.7 kcal/mol lower than that in **6** (Figure 3-9). The predicted energy barriers for the three configurations with two water molecules are 18.5 kcal/mol (**4** and **TS-4**), 18.9 kcal/mol (**5** and **TS-5**) and 14.0 kcal/mol (**6** and **TS-6**), respectively. At **TS-4** and **TS-5**, the oxygen atom of the second water molecule, O_5, forms a hydrogen bond with the hydrogen atom bonded to O_2, and the hydrogen atom of the other water is hydrogen bonded to the distal oxygen, O_3. In the one water case (**TS-4**, Figure 3-7), the two hydrogen bonds are formed by the single water molecule. The arrangement of the second water molecule results in a decrease of 5–7 kcal/mol in the energy barrier. For **TS-6**, the hydrogen atom of the second water molecule forms a hydrogen bond to O_3, and the other water retains a similar configuration as that in **TS-4**. Therefore, the additional hydrogen bond to the more negative distal oxygen in **TS-6** versus **TS-4** and **TS-5** is responsible for the lower activation energy of configuration 3, again, as a result of the stabilization of the charge separation in the transition state. The



4



TS-4 E(4)-E(TS-4)=18.5 kcal/mol

Figure 3-7: Oxidation of dimethyl sulfide with two water molecules, configuration 1. The two water molecules are at the same side of H_2O_2 and at the trans side of the methyl groups of DMS molecule at the transition state. Reactant cluster **4** and transition state **TS-4** were fully optimized at the B3LYP/6-31++G(d,p) level; energies were calculated at the MP4//B3LYP/6-31++G(d,p) level.

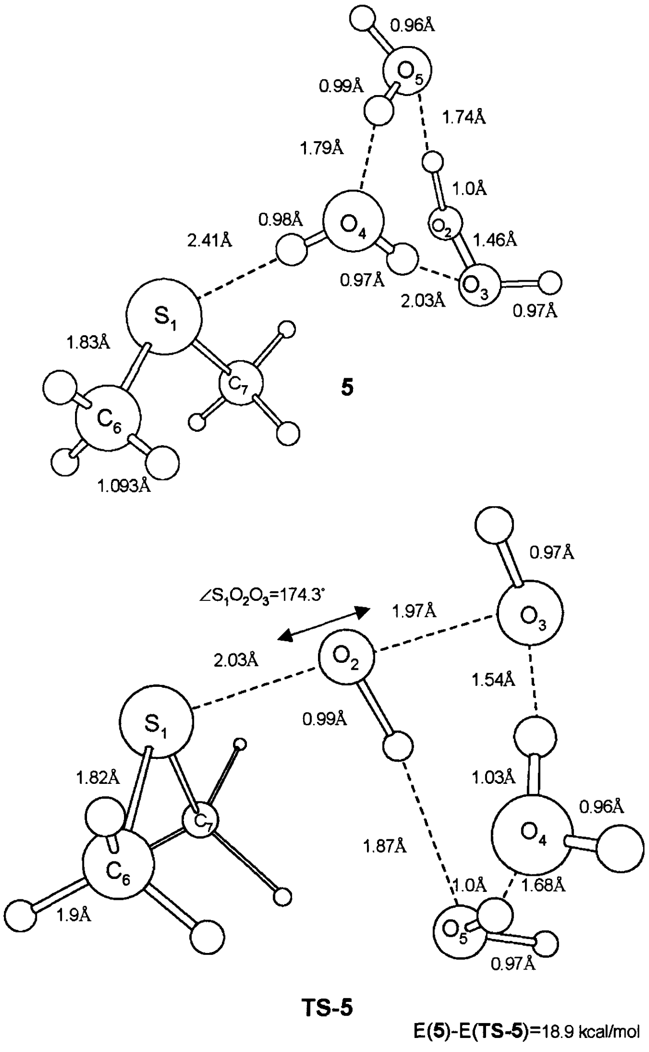
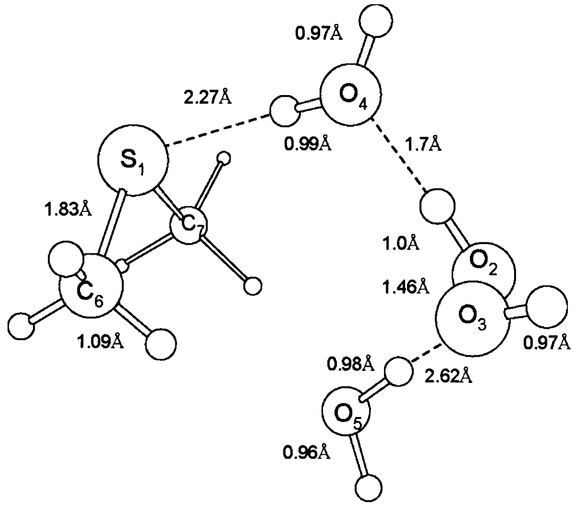
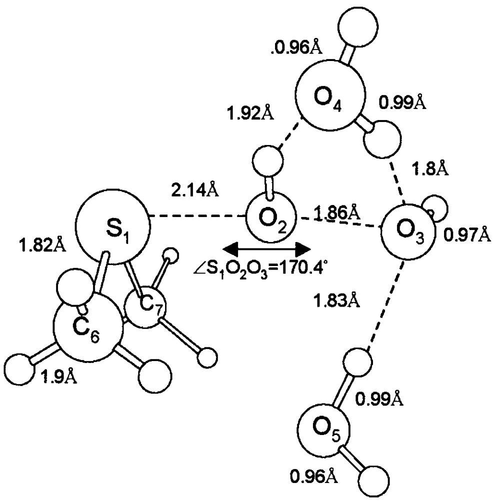


Figure 3-8: Oxidation of dimethyl sulfide with two water molecules, configuration 2. The two water molecules are at the same side of H_2O_2 and at the cis side of the methyl groups of DMS molecule at the transition state. Reactant cluster **5** and transition state **TS-5** were fully optimized at the B3LYP/6-31++G(d,p) level; energies were calculated at the MP4//B3LYP/6-31++G(d,p) level.



6



TS-6 E(**6**)-E(**TS-6**)=14.0 kcal/mol

Figure 3-9: Oxidation of dimethyl sulfide with two water molecules, configuration 3. Each water molecule is of one side of H_2O_2. Reactant cluster **6** and transition state **TS-6** were fully optimized at the B3LYP/6-31++G(d,p) level; energies were calculated at the MP4//B3LYP/6-31++G(d,p) level.

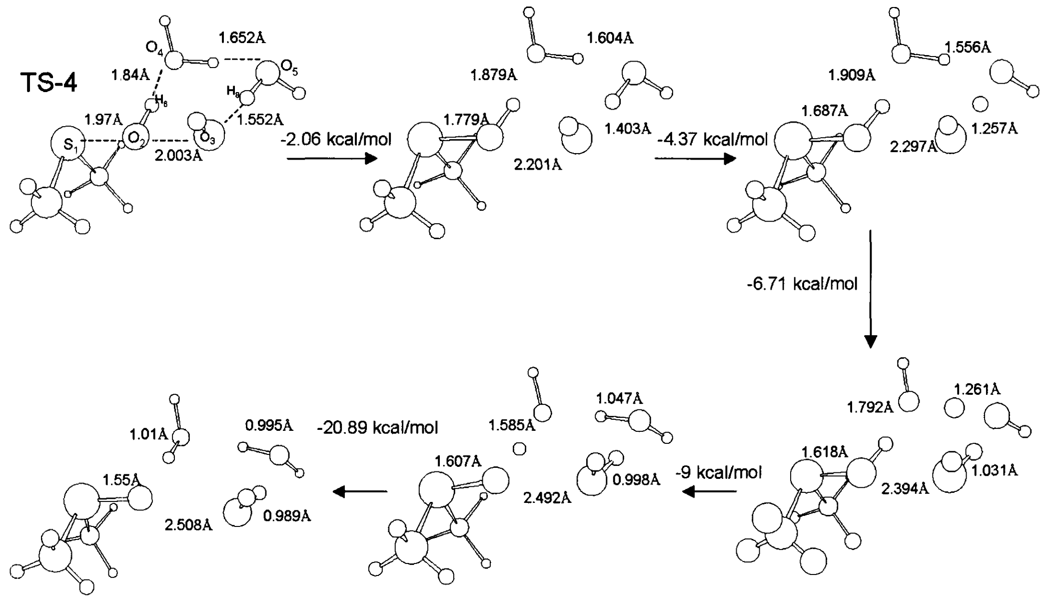


Figure 3-10: Reaction path following analysis of DMS oxidation with two water molecules for **TS-4**. Bond lengths in each snapshot correspond to the dotted lines connecting the indicated atoms of the upper-left frame.

imaginary vibrational modes for all transition states demonstrate that the characteristic mode at the transition state is oxygen transfer to the sulfur atom, but not the transfer of hydrogen. A reaction path following analysis for transition state **TS-4** is shown in Figure 3-10. First, the O_2-O_3 bond breaks, and the H_8 transfers from O_5 to O_3. Second, H_7 transfers to O_5 from O_4. Finally, H_6 transfers to O_4 from O_2 forming the leaving water molecules. In this reaction, both water molecules are involved in the hydrogen transfer process after passing the transition state. For **TS-5**, a reaction path following analysis is shown in Figure 3-11. In this case, the transfer of O_2 to S_1 also occurs before the transfer of hydrogen. The hydrogen atom H_4, goes directly to the distal oxygen, O_3 from O_2. Water molecules only stabilize the system via hydrogen bonding. For **TS-6**, the hydrogen transfer process is similar to that in the one water model, Figure 3-5. The hydrogen atom of the second water molecule forms a hydrogen bond with O_3, but is not involved in the transfer of hydrogen. Compared to the one water case, this arrangement has lowered the activation energy by ~10 kcal/mol.

For the three-water model, two water configurations were considered. In the first

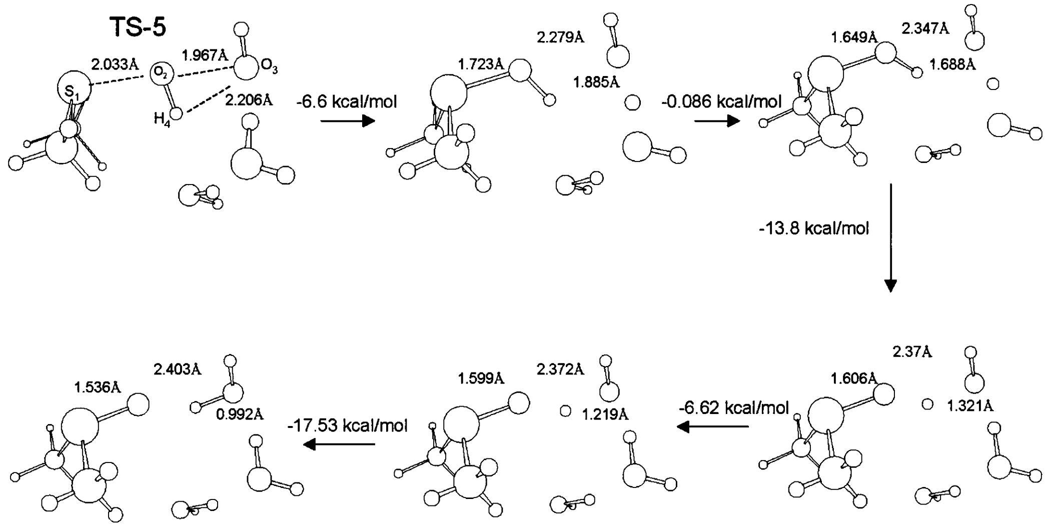
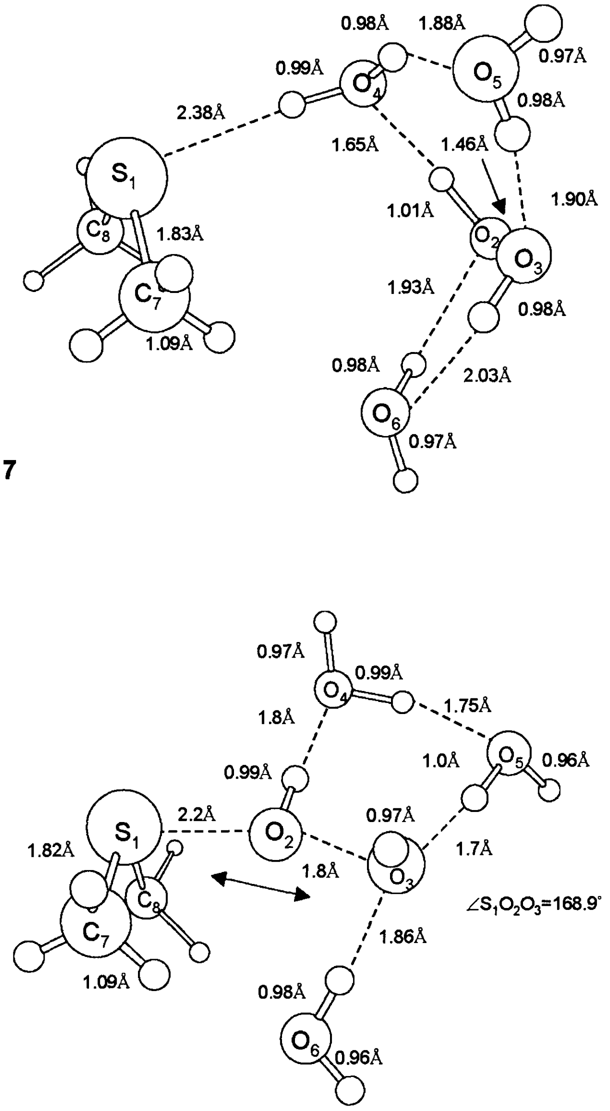


Figure 3-11: Reaction path following analysis of DMS oxidation with two water molecules for **TS-5**. Bond lengths in each snapshot correspond to the dotted lines connecting the indicated atoms of the upper-left frame.

configuration, two water molecules are at one side of H_2O_2, but at the trans side of the methyl groups (**7** and **TS-7**, Figure 12). This configuration corresponds to configuration 1 in the two water model, and the third water molecule forms a hydrogen bond with the distal oxygen, O_3 at the transition state. In configuration 2, two water molecules are at the cis side of the methyl groups and the other one is at the trans side (**8** and **TS-8** in Figure 13). This configuration corresponds to configuration 3 in the two water model, and in this case, the third water molecule forms a hydrogen bond with the transferring oxygen, O_2 at the transition state. Since the distal oxygen is more negative than the transferring oxygen at the transition state, it is expected that the third water molecule in configuration 1 will have more stabilization effects. This is indeed the case. The energy barriers are 12.6 kcal/mol between **7** and **TS-7** and 13.7 kcal/mol between **8** and **TS-8**. Compared to configuration 3 in the two water model, the third water molecule in configuration 2 yields only 0.3 kcal/mol in stabilization energy. However, compared with configuration 1 in the two-water model, the third water molecule of configuration 1 has a significant stabilization effects, reducing the activation energy by 5.9 kcal/mol. These results indicate that forming a hydrogen

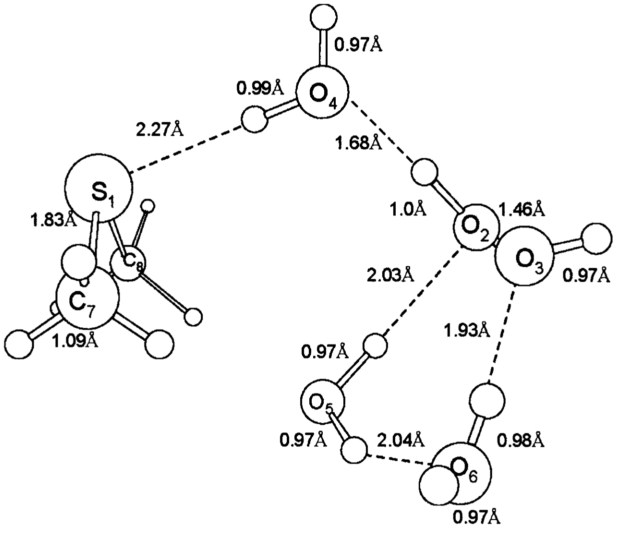


7

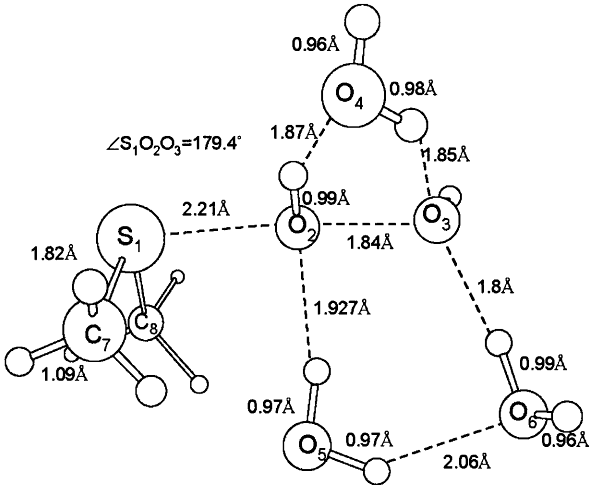
TS-7

E(**7**)-E(**TS-7**)=12.7 kcal/mol

Figure 3-12: Oxidation of dimethyl sulfide with three water molecules, configuration 1. Reactant cluster **7** and transition state **TS-7** were fully optimized at the B3LYP/6-31++G(d,p) level; energies were calculated at the MP4//B3LYP/6-31++G(d,p) level.



8



TS-8

E(**8**)-E(**TS-8**)=13.7 kcal/mol

Figure 3-13: Oxidation of dimethyl sulfide with three water molecules, configuration 2. Reactant cluster **8** and transition state **TS-8** were fully optimized at the B3LYP/6-31++G(d,p) level; energies were calculated at the MP4//B3LYP/6-31++G(d,p) level.

bond with the more negative distal oxygen at the transition state has more stabilization effect than doing this with the less negative O_2 atom. For both configurations, the calculated activation energies are within the experimentally observed range of organic sulfides (10∼20 kcal/mol). When compared with the activation energy (13.5 kcal/mol) for thioxane oxidation by hydrogen peroxide in aqueous solution, the above result is also quite close (Dankleff *et al.* , 1968). For each model of DMS oxidation, the activation barrier is about 10 kcal/mol or more lower than that of the corresponding hydrogen transfer reaction of hydrogen peroxide to water oxide. Therefore, it is unlikely that oxidation occurs in two steps in which the hydrogen transfers first to form water oxide. Instead, oxidation occurs via stretching of the O-O bond of H_2O_2 and reduction in the S-O distance. Solvent molecules stabilize the charge separation at the transition state via specific interactions, including hydrogen bonding, and may also be involved in the transfer of hydrogen after the system has passed the transition state. Hydrogen bonding to the more negative distal oxygen at the transition state has a significant stabilization effect as shown in the two-water and three-water models. Reaction path following analyses of transition states indicate that one or two water molecules can be involved in the hydrogen transfer process, similar to the hydrogen transfer reaction of hydrogen peroxide in which both 1,4 and 1,6 concerted proton transfers have been observed. In addition, a direct 1,2 hydrogen transfer to the distal oxygen occurring after the system has passed the transition state was also observed (Figure 3-11). Since the transfer of hydrogen occurs after the transition state is passed, it is not the determining factor of the activation barriers. Instead, 2–3 explicit water molecules can provide enough stabilization energy for the oxidation of organic sulfides by peroxides. Water molecules stabilize the charge separation at the activation complex via specific interactions including hydrogen bonding. This mechanism is also in agreement with the experimentally observed pH-independent rates over moderate pH ranges, since an additional proton is not involved in this model. Our results imply that unprotonated water molecules can provide enough stabilization for the oxidation of organic sulfides by hydrogen peroxide. At extremely low pH conditions, of course, high concentrations of protonated solvent molecules may result in different oxidation pathways (Bach *et al.* , 1994; Amels *et al.* , 1997).

3.3 Conclusions

The following conclusions can be drawn from the above theoretical analyses:

1. For moderate pH values, perhaps between 2 and 12, the reaction coordinate leading to oxidation of dimethyl sulfide by H_2O_2 is the separation (breaking) of the O-O bond together with the formation of the S-O bond. Thus, under these conditions, water molecules stabilize the transition complex via specific interactions including formation of hydrogen bonds with H_2O_2 but not proton transfer as previously assumed in the literature (Bach *et al.* , 1991; Nguyan *et al.* , 1993; Nguyan, 1994; Bach *et al.* , 1994; Li *et al.* , 1995; Okajima, 2001; Ottolina & Carrea, 2001).

2. Hydrogen transfer does occur during oxidation of DMS, but it is not the determining factor of the activation barrier. Hydrogen transfer can occur via multiple different pathways depending on the local solvent configuration (Figure 3-5, Figure 3-10 and Figure 3-11).

3. During DMS oxidation, an uneven charge distribution between the oxygen atoms of H_2O_2 is developed. The distal oxygen is more negative than the transferring oxygen. Water molecules stabilize the charge separation by local polarization and via formation of hydrogen bonds with H_2O_2.

4. For hydrogen transfer reactions involving H_2O_2 and for DMS oxidation with a single water molecule, a polarizable solvent results in small effects on activation energies (<1 kcal/mol in most cases). This is also consistent with the experimental observation that the rates of oxidation of organic sulfides are not a strong function of solvent dielectric constants (Dankleff *et al.* , 1968).

5. The energy barrier of DMS oxidation with 3 water molecules is estimated to

be 12.7 kcal/mol at the MP4//B3LYP/6-31++G(d,p) level, within the experimental range of barriers of oxidation of organic sulfides (Dankleff *et al.* , 1968).

6. The activation energies of the hydrogen transfer reaction of hydrogen peroxide are an average of 10 kcal/mol or more higher than those for the oxidation of DMS by H_2O_2. This, combined with conclusions 1–5, indicates that a two-step mechanism in which the hydrogen transfer of H_2O_2 occurs first to form water oxide followed by the transfer of oxygen to the nucleophile from water oxide is not the mechanism by which organic sulfides are oxidized by hydrogen peroxide.

7. The proposed oxidation mechanism does not suggest a pH dependence of oxidation rates within a moderate range around neutral pH (i.e. under conditions in which hydronium and hydroxide ions do not participate directly in the reaction). In this respect, it agrees with experimental observations over moderate pH values (Amels *et al.* , 1997). It also predicts activation energies corresponding to measured activation energies without including a protonated solvent molecule.

Chapter 4

Oxidation of Methionine Residues in G-CSF

In order to understand the connection between the conformation of a protein molecule and the oxidation of its methionine residues, the structural properties of G-CSF (Granulocyte Colony-Stimulating Factor) are investigated via molecular dynamics simulations and are correlated with the rates of oxidation of methionine residues at different pH values. Each of the four methionine groups in G-CSF have significant and different rates of oxidation as a function of pH. Even Met1 in the unstructured N-terminal region, has a rate of oxidation as low as half that of free methionine. The structural properties of G-CSF as a function of pH are evaluated in terms of properties such as hydrogen bonding, deviations from X-ray structure, helical/helical packing, and the atomic covariance fluctuation matrix of α-carbons. The results show that dynamics (structural fluctuations) are essential in explaining oxidation and that a static picture, such as that resulting from X-ray data, fails in this regard. Moreover, the simulation results also indicate that the solvent-accessible-area (SAA), traditionally used to measure solvent accessibility of a protein site, of the sulfur atom of methionine residues does not correlate well with the rate of oxidation. Instead, a structural property, average two-shell water coordination number (WCN), is identified to correlate well with measured oxidation rates.

4.1 Motivation

As discussed in Chapter 1, the covalent addition of an oxygen atom to the sulfur atom of methionine changes the chemical properties of proteins and generally leads to the loss of biological function. Early studies suggested that the major reactive oxygen species causing the oxidation of methionine in proteins in aqueous solutions are hydroxyl radicals and hydrogen peroxide (Li *et al.* , 1995). Oxidation by hydroxyl radicals is catalyzed by metal ions that bind to specific sites of a protein. This type of mechanism is referred to as "site-specific" oxidation(Stadtman & Oliver, 1991; Stadtman, 1992; Nguyan, 1994; Li *et al.* , 1995). Oxidation by hydrogen peroxide, on the other hand, is generally accepted to be an S_N2 reaction. It is referred to as "non-site-specific" oxidation (Wang, 1999; Li *et al.* , 1995). The charge separation developed during the breaking of the O-O bond in H_2O_2 contributes to most of the energetic penalty of the oxidation reaction (Edwards, 1960; Ranky & Nelson, 1961; Dankleff *et al.* , 1968; Curci *et al.* , 1970). This observation led to the proposal in a 1994 theoretical study that a strong proton donor, such as a hydronium ion, would be necessary to affect the oxidation (Bach *et al.* , 1994). Subsequent to the 1994 study, it was observed experimentally that over a large range of pH(2–10) there is no pH dependence on the rate of oxidation (Amels *et al.* , 1997). This observation called into question the validity of the 1994 mechanism over an intermediate range of pH. The studies presented in Chapter 3 were thus performed, and the results led to the proposal of a new mechanism of charge stablization resulting from specific interactions, such as hydrogen bonding, with surrounding water molecules. These interactions stabilize the charge separation and result in the observed activation energy of the reaction (10–15 kcal/mol). The new mechanism can thus be called a "water-mediated" mechanism and is consistent with all available experimental information about the oxidation mechanism, such as activation energies and the pH dependence of oxidation rates as shown in Chapter 3. It also leads to the conclusion that exposure to solvent molecules plays an essential role in oxidation reactions. In several studies, it was shown that amino acids that are located in buried positions (determined qualitatively from static X-ray data) are less easily oxidized by exogenously added peroxide oxidants (Pearlman & Bewley, 1993; Cleland *et al.* , 1993; Lu *et al.* , 1999; Griffiths & Cooney, 2002) compared to those at the surface. This implies that the access of solvent to methionine sites affects oxidation. While this observation is consistent with our proposed mechanism, it is qualitative at best, because it is based on static data. Clearly, proteins have essential dynamic behavior that is not captured by static X-ray structure. This study addresses the effects of protein dynamics and thermal fluctuations on the rates of oxidation of methionine sites in therapeutic proteins. In this chapter, the focus is put on G-CSF, a protein pharmaceutical useful in treating patients suffering from neutropenia and in mobilizing peripheral blood progenitor cells for harvesting and transplantation after myeloablative chemoradio therapy (Lu *et al.* , 1999). G-CSF has four methionine groups, Met1, Met122, Met127 and Met138. Each of the methionine groups has been shown to be oxidized by hydrogen peroxide with distinct oxidation rates in the order: Met1 > Met138 > Met127 > Met122. Furthermore, the bioactivity measured by a cell mitogenesis bioassay was shown to decrease to 3% after all methionine groups were oxidized (Lu *et al.* , 1999).

A measure typically used to quantify the degree to which the solvent can interact with different parts of the protein is solvent accessible area (SAA) (Lee & Richards, 1971). However, using the X-ray structure by Aritomi *et al.* (Aritomi *et al.* , 1999; Aritomi *et al.* , 2000), PDB entry 1CD9, and a probe radius of 1.6 Å, the computed SAA's of Met122, Met127, and Met138 are zero. Since all three of these Met groups have appreciable rates of oxidation, it is unlikely that this structural quantity either relates to oxidation or provides a good measure of the degree to which solvent can interact with sites in the protein. It seems to be important at least to incorporate protein and solvent dynamics in any model that relates structural properties to rates of oxidation. In this study, molecular dynamics simulations of G-CSF were performed to understand the relationship between oxidation and the structures of G-CSF. The simulations incorporated an explicit treatment of the solvent molecules and realistic dynamical motions of each atom in the protein. The goal was to better understand the oxidation of methionine by relating structural properties to rates of oxidation.

It was hoped that a structural parameter could be identified that correlates well to rates of oxidation. In addition, the effects of pH on the rates of oxidation and on structural properties are also analyzed.

4.2 Details of calculations

Molecular dynamics simulations of G-CSF were performed using the CHARMM software (Brooks *et al.* , 1983) and the CHARMM22 all-atom potential (MacKerell *et al.* , 1998). The initial structure of G-CSF was obtained from the X-ray structure of Aritomi *et al.* (Aritomi *et al.* , 1999; Aritomi *et al.* , 1999), determined at pH 7.5, PDB entry 1CD9. The coordinates of residues missing in the PDB data (residues 1–6) were generated using the internal coordinate table; the coordinates of hydrogen atoms were then generated using the HBUILD function of CHARMM (Brooks *et al.* , 1983). 100 steepest descent steps followed by 500 conjugate gradient steps were applied to the initial geometry with heavy atoms fixed at their PDB coordinates. The resulting geometry was then further minimized with 300 conjugate gradient steps with heavy atoms restrained by their PDB coordinates. The restraint force constant was 30 kcal/mol/Å and was decreased by 10 kcal/mol/Å at every 100 minimization steps. The resulting geometry had a 0.12 Å all-atom RMSD (Root of Mean Squared Deviation) from the original X-ray coordinates and was used as the reference structure for analysis. G-CSF has five histidine residues, and the protonation states were determined by examining the relative distances to surrounding hydrogen bond donors and acceptors. The reference structure was then solvated in a pre-equilibrated (295 K, 1 atm) truncated octahedral water cell with a lattice length of 68 Å, containing 8280 water molecules. Counter ions were placed using the SOLVATE program (Grubmuller, 1996). Water molecules with oxygen atoms within 2.4 Å of the heavy atoms of G-CSF and counter ions were removed. Periodic boundary conditions were applied and the long-range Coulombic interactions were calculated using the particle-mesh Ewald summation method (Darden *et al.* , 1993; Essmann *et al.* , 1995a). The cutoff radius for electrostatic interactions was 15 Å. Covalent bonds were fixed during the simulations using the SHAKE (Allen & Tildesley, 1987) algorithm, thus allowing a 0.002 ps integration time step. 50 steepest descent steps followed by 200 conjugate gradient minimization steps were performed first with protein atoms fixed. Next, the system was heated up to 295K by gradually scaling the velocities (3 K/0.1 ps) for 10 ps. During the heating period, the backbone atoms of G-CSF were restrained to the reference structure by a force constant of 10 kcal/mol/Å. After the heating period, a 30 ps equilibration simulation at 295 K was performed using the Nosé-Hoover thermostat (Nosé, 1991). During the equilibration period, the restraint force constant for the protein backbone atoms was gradually decreased by 5 kcal/mol/Å every 10 ps. For the last 10 ps of equilibration, no restraint force was applied. Then the molecular dynamics simulation was performed in the canonical ensemble at 295 K for 2.96 ns, during which averaged properties were computed. Trajectories were saved every 0.1 ps for analysis, and the statistical uncertainties of averaged properties were estimated based on standard procedures (Allen & Tildesley, 1987; Frenkel & Smit, 1996). Simulations were performed separately for three regimes of pH. The first is called "high-pH", in which only one of the nitrogen atoms of histidine residues was protonated. The carboxyl oxygen of aspartic acid and glutamic acid residues were not protonated. The second is called "medium-pH" in which both nitrogen atoms of the histidine residues were protonated but not the carboxyl oxygen of aspartic acid and glutamic acid residues. The last is called "low-pH", in which all acidic residues were protonated. In order to compare the results of molecular dynamics simulations with the measured rates of oxidation of methionine residues, a mapping between protonation states and pH values is needed. G-CSF has five histidine groups. From the X-ray structure, the protonation states of His80 and His157 may affect the fluorescence intensity of the nearby Trp59(His157) and Trp119(His80) residues (Chen & Barkley, 1998; Berms, 2002). The pH dependence of the fluorescence intensities of G-CSF suggests that the pKa's of His80 and His157 are around 5.8 (Narhi *et al.* , 1991). The rest of histidine residues (His44, His53, and His171) are more exposed to the solvent, and we assume that the pKa's are close to the value of the free residue, 6.1. G-CSF also contains 4 Asp and 9 Glu residues over the protein surface, and we also assume that their pKa's do not deviate significantly from the values of the free residues. Based on the above information, we assume that the "high-pH" protonation configurations correspond roughly to pH 8.5, the "medium-pH" configurations correspond roughly to pH 5.5, and the "low-pH" configurations correspond roughly to pH 2. In order to evaluate the effects of different protonated configurations of His80 and His157, three different protonation configurations of His80 and His157 are examined. They are all considered to be in the "high-pH" regime. The first configuration is called "high-pH-1". In this configuration, the ε-nitrogen of His80 and the d-nitrogen of His157 are protonated. The second configuration, in which the d-nitrogen of His80 and the d-nitrogen of His157 are protonated, is called "high-pH-2". The last configuration, in which the ε-nitrogen of His80 and the ε-nitrogen of His157 are protonated, is called "high-pH-3".

4.3 Summery of experimental data on the oxidation of methionine residues in G-CSF at different pH values

In this section, the experimental data of the oxidation rates of methionine groups in G-CSF are summarized as the basis to compare with the computational results. The experimental details are reported in literature (Chu *et al.* , 2004a). In Figure 4-1, the concentration of unoxidized methionine residues in G-CSF at pH=4.5 are normalized to their initial concentration and plotted versus time on a semi-log scale. The very high linearity ($R^2 > 0.98$) in Figure 4-1 under an oxidation condition of excess H_2O_2 indicates that the reaction is pseudo first order for each methionine residue. The first-order nature of the reaction also indicates that the oxidation of one methionine residue does not affect the rates of oxidation of other methionine residues in the experimental time scale, i.e. a significant change of the protein conformation did not occur. For Met122, on the other hand, there indeed exists a different oxidation rate constant after the rest of methionine residues are oxidized, as reported earlier

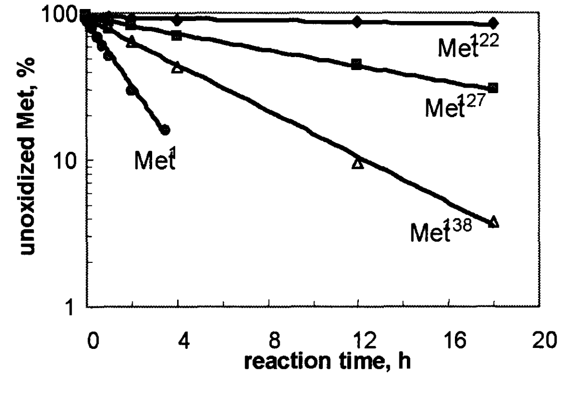


Figure 4-1: Time course of the percentage of unoxidized methionine residue on a semi-log scale. Oxidation reactions were carried out with 0.1 mg/ml G-CSF at a pH of 4.5 and 25 °C. H_2O_2 concentrations were 30 mM.

(Lu *et al.* , 1999). In this study, the focus is on the stability of native G-CSF, i.e., initial rates of oxidation, and thus, change in the conformation of G-CSF as a result of oxidation is not a concern. The reaction order of H_2O_2 was also examined by performing the oxidation experiments at different H_2O_2 concentrations, and the results indicate that the reaction is also of first order with respect to H_2O_2. The pseudo first order rate constants of each methionine residue as well as the results in Lu *et al.* (Lu *et al.* , 1999) are shown in Table 4.1. The buffers used were phosphate at pH 2.0, citrate at pH 3.0, 5.5 and 6.5, acetate at pH 4.0, and tris-HCl at pH 8.5. The pH dependence of oxidation rates is shown in Figure 4-2. The oxidation rates of Met1 show a minimum at pH 4.5 and are 20 % higher at pH 2 and 77 % higher at pH 8.5. Met138 is oxidized the second fastest and shows a moderate pH dependence of oxidation rates with a minimum at pH=4.5. The oxidation rate of Met138 is 33 % higher at pH 2 and 13 % higher at pH=8.5 than at pH 4.5. For Met127, the pH dependence of oxidation rates is small (the difference between the maximum and minimum rate constant is $<$ 15%). Met122 oxidizes at the lowest rate among the methionine residues of G-CSF and shows a weak pH dependence of oxidation rates between pH 2 $\sim$ pH 5.5 but increases by 120 % from pH 4.5 to pH 8.5. The experimental results indicate that each methionine residue in G-CSF has a different rate of oxidation and that the rates of oxidation of each methionine residue

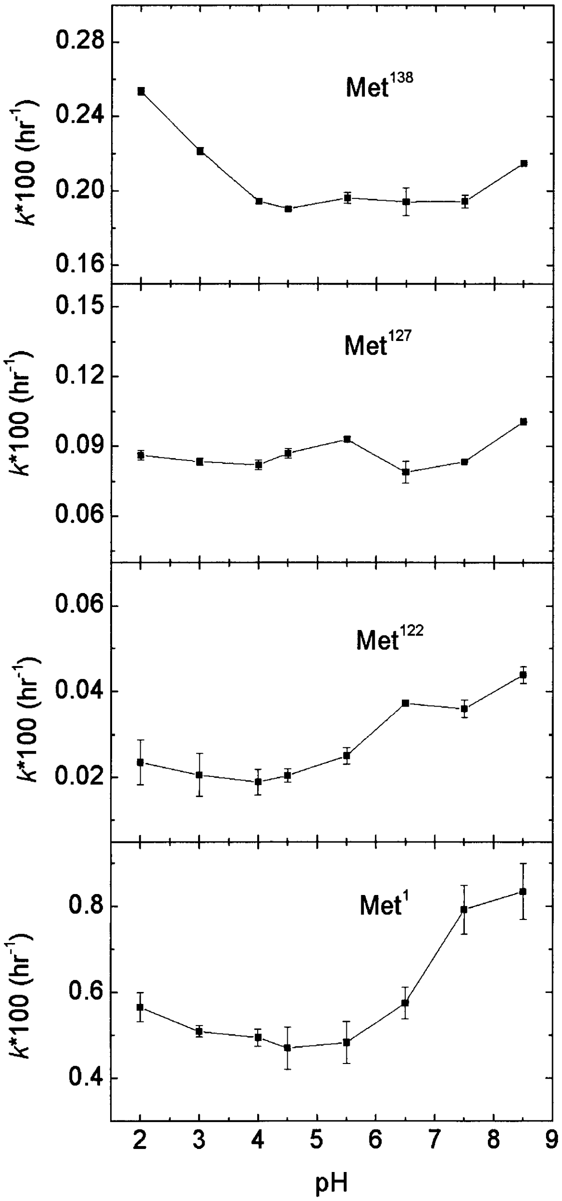


Figure 4-2: The rates of oxidation in G-CSF at different pH values. The pseudo first-order rate constants in unit of (hr^{-1}) are plotted as a function of pH for each methionine residue. Oxidation reactions were carried out with 0.1 mg/ml G-CSF at 25 °C. H_2O_2 concentrations were 30 mM.

Table 4.1: Pseudo first-order oxidation rate constants of methionine groups in G-CSF at different pH values. Oxidation reactions were carried out with 30 mM H_2O_2 and 0.1 mg/ml G-CSF at 25 °C. The data recorded is (rate constant, $hr^{-1}*100$).

pH	Met1	Met122	Met127	Met138
2	0.566	0.024	0.086	0.253
3	0.509	0.021	0.083	0.221
4	0.495	0.019	0.082	0.194
4.5	0.470	0.020	0.087	0.190
4.5[a]	0.486	0.002	0.180	0.263
5.5	0.483	0.025	0.093	0.196
6.5	0.575	0.037	0.079	0.194
7.5	0.792	0.036	0.083	0.194
8.5	0.834	0.044	0.101	0.215

---

[a] Results from (Lu *et al.* , 1999), converted to 30 mM H_2O_2.

are different functions of pH. Note that there is no pH dependence of the rates of oxidation of free methionine between pH 2 and 8.5. Also note that a previous study on the oxidation of dimethyl sulfide (DMS), which represents part of the methionine side chain, reports no pH dependence between pH 2 and 10 (Amels *et al.* , 1997). Furthermore, the proposed "water-mediated" mechanism discussed in Chapter 3 does not lead to a pH dependence of the oxidation of organic sulfides by hydrogen peroxide over intermediate ranges of pH. All of these results suggest that the pH dependence of the rates of oxidation of a given methionine is due to the changes in the local environment in response to solution pH. This hypothesis is examined in detail below using the results of molecular dynamics simulations.

4.4 Results and discussions

4.4.1 Identification of the governing factors of the oxidation of methionine via molecular dynamics simulations

The structures and conformation of G-CSF at different pH values were investigated via molecular dynamics simulations of G-CSF with different protonation configurations in

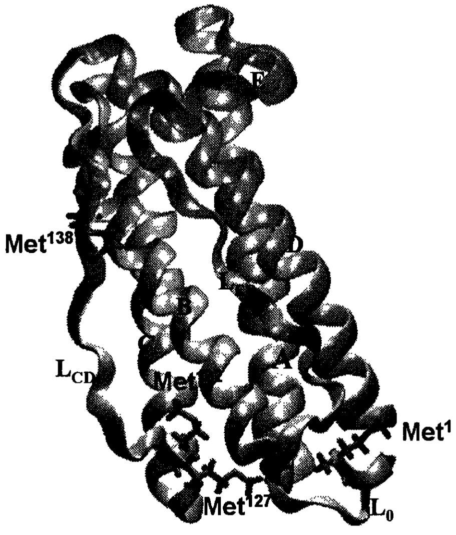


Figure 4-3: The ribbon representation of the X-ray structure of G-CSF (PDB code 1CD9). The main bundle helices A (residues 12–40), B (residues 73–93), C (residues 100–125) and D (residues 144–173) are labeled near their N termini. The short 3_{10} (residues 45–48) and α (residues 51–55) helices are also indicated.

the unoxidized form. Since all methionine groups show first order kinetics in the early phase of oxidation, using the unoxidized form of G-CSF to relate G-CSF conformation to its initial rates of oxidation is valid. In this section, we first describe the structural properties of G-CSF at different pH values and then study the correlation between the structural properties of G-CSF and the initial rates of oxidation of its methionine residues.

Labeling of regions in G-CSF

First, different regions of the protein are labeled in corder to describe their conformational properties in the following sections. G-CSF belongs to the four-helix-bundle structural superfamily of growth factors (Bazan, 1990). The ribbon representation of the X-ray structure of G-CSF in Figure 4-3 (prepared using the VMD package (Humphrey *et al.* , 1996), and the secondary structure was identified using the DSSP progeam. (Kabsch & Sander, 1983)) shows that G-CSF has four main helices, helix A (residues 12–40), B (73–93), C (100–125) and D (143–173), and a short helical

Table 4.2: Distribution of charged residues and the net charges of sub-domains of G-CSF.

Domain	$\mathrm{L}_0{}^a$	A^a	L^a_{AE}	E^a	L^a_{EB}	B^a	L^a_{BC}	C^a	L^a_{CD}	D^a	Total^b
	(1–11)	(12–40)	(41–44)	(45–55)	(56–72)	(73–93)	(94–99)	(100–125)	(126–143)	(144–175)	
Arg		23								147 148 167 170	5
Lys		17 24 35	41								4
His			44	53		80				157 171	5
Asp		28						105 110 113			4
Glu		20 34		46 47			94 99	123 124		163	9
pH^c						net charge					
low	0	4	2	1	0	1	0	0	0	6	14
medium	0	1	2	-1	0	1	-2	-5	0	5	1
high	0	1	1	-2	0	0	-2	-5	0	3	-4

---

[a] The numbers shown in the column are the residue numbers of charged residues.
[b] The total number of charged residues in G-CSF.
[c] The protonated state of each pH category is described in the text.

region, helix E (45–55). The four main helices are arranged in an up-up-down-down fashion by the long loops L_{EB} (residues 56–72) connecting helices E and B, and L_{CD} (126–172), connecting helices C and D. The residues 1–11 of G-CSF, denoted by L_0, do not form a well-defined secondary structure. The short segment between helices A and E (41–44) is denoted by L_{AE} and the segment between helices B and C (94–99) is denoted as L_{BC}. The net charges of G-CSF and its sub-domains at various pH ranges are summarized in Table 4.2. Basic residues are mostly located in helices A and D, and acidic residues are mostly located in helix C. Note that helices A and C also contain the receptor binding domains of G-CSF (Reidhaar-Olson *et al.* , 1996; Aritomi *et al.* , 1999). Loops L_0, L_{BE}, and L_{CD} do not have charged residues.

Overall structure and fluctuations

The time course of the root mean square deviations (RMSD) of the backbone heavy atoms of G-CSF from the X-ray reference structure is shown in Figure 4-4. Backbone RMSDs for the helical domains (A–D), for L_0-excluded residues (denoted as L_0-ex in

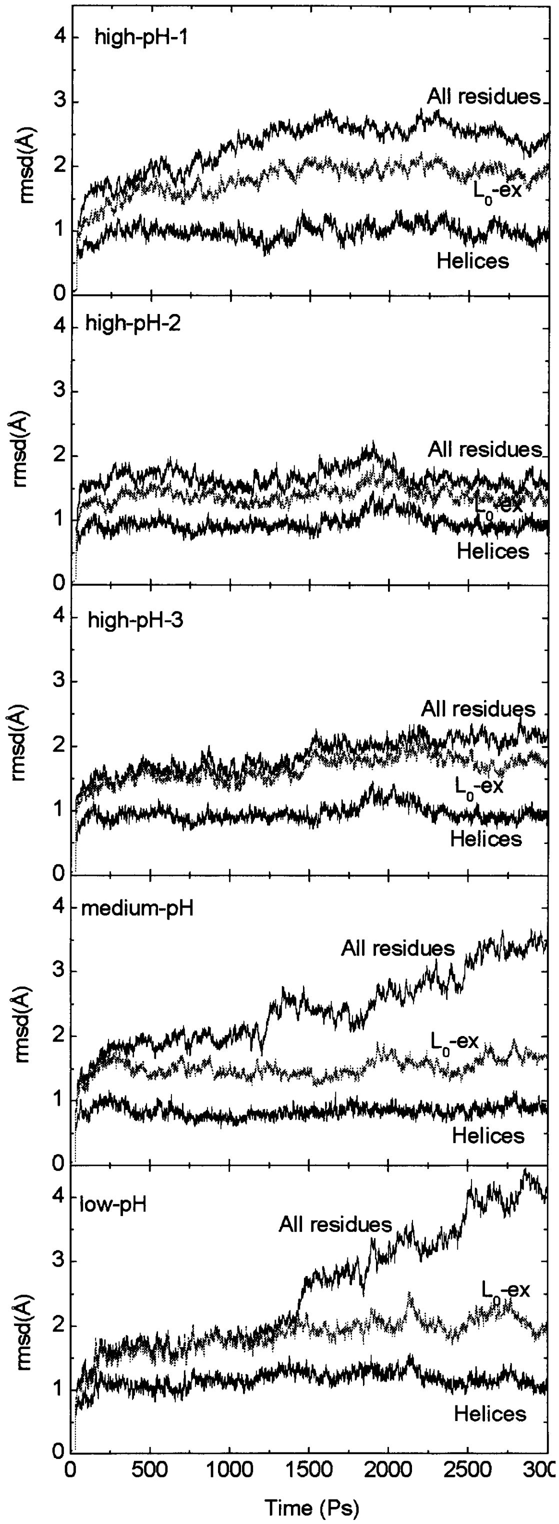


Figure 4-4: Time course of the Root-Mean-Square Deviations (RMSD) of backbone atoms from the X-ray structure (PDB code 1CD9) of G-CSF. RMSDs for the helical domains (A–D), for all residues except the L_0 region (L_0-ex), and for all residues are plotted. The data points are shown for every 1 ps.

Figure 4-4) and for all residues are presented. The helical regions have relatively small RMSDs from the X-ray structure for all protonation configurations throughout the simulations. The largest average helical-region RMSD (last 1.5 ns) occurs in the low-pH simulation (1.2 Å). When we include residues other than those in the L_0 region, the RMSDs from the X-ray structure become larger but stay constant after 1.5 ns in the simulation for all protonation configurations. Including the L_0 region further increases the RMSD values, indicating that the loop regions of G-CSF have deviated more from the reference structure than the helical domains have. The small backbone RMSDs for helical domains in the simulation of all protonation configurations suggest that the secondary as well as the up-up-down-down arrangement of the helical domains of G-CSF are stable for a wide pH range, even at low pH values where many proteins may exhibit the pH-induced partially unfolded "A" state (Fink *et al.* , 1993; Fink *et al.* , 1994). The conformational stability of G-CSF at low pH values has also been reported using fluorescence and circular dichroism spectra (Kolvenbach *et al.* , 1997). In the high-pH regime, the high-pH-2 simulation has the smallest average backbone all-residue (1.71 Å) and L_0-ex (1.44 Å) RMSDs among all three cases (High-pH-1: 1.96 Å L_0-ex RMSD, 2.58 Å all-residue RMSD, High-pH-3: 2.0 Å L_0-ex RMSD, 1.8 Å all-residue RMSD). These differences highlight the fact that changes in the protonation of the nitrogen atoms in histidine residues may result in distinct protein conformational changes. The medium-pH simulation, on the other hand, has a quite small L_0-ex RMSD (1.56 Å), but the all-residue RMSD is significantly larger (2.86 Å). This indicates that the position of the L_0 region has deviated significantly from its initial structure, but the rest of the protein units maintain positions similar to those in the X-ray structure. Finally, the low-pH simulation has the largest average L_0-ex RMSD (2.03 Å) and all-residue RMSD (3.4 Å).

Root-Mean-Squared atomic fluctuations (RMSF) of backbone heavy atoms, averaged within a residue are shown in Figure 4-5. In high-pH-1, high-pH-2 and medium-pH simulations, the RMSF's from molecular dynamic simulations and from the X-ray (pH 7.5) experimental B-factor (Hill *et al.* , 1993) agree very well. In general, the loop regions exhibit higher mobility than the helical regions. The RMSF in the high-pH-3

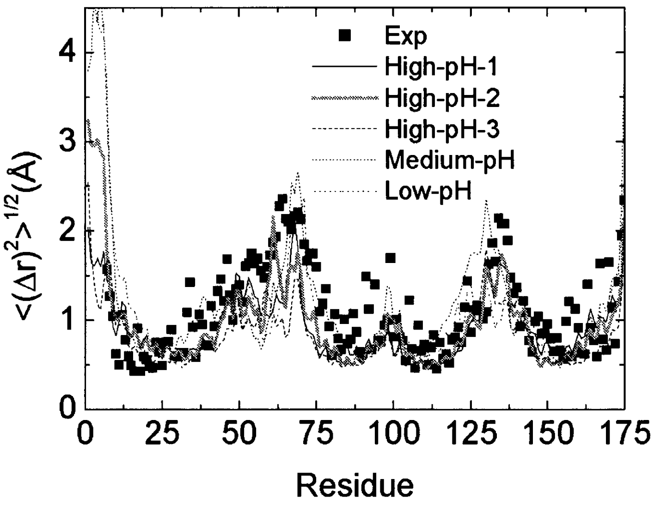


Figure 4-5: The root-mean-squared atomic fluctuations (RMSFs) averaged for each residue for the last 1.5 ns of each simulation with a 0.1 ps interval between data points. Gray lines: calculated from crystalline B-factors; black lines: calculated from the molecular dynamic simulations. Helical regions are marked on the X-axis of the top panel.

and low-pH simulations also agree well with the experimental data, but the high-pH-3 simulation has smaller RMSF's in the L_{EB} region, while the low-pH simulation has smaller RMSF's in the L_{CD} region.

The backbone flexibility of G-CSF can be evaluated using the squared order parameter (Lipari & Szabo, 1982):

$$S^2 = \frac{3}{2}\left[\langle x^2\rangle^2 + \langle y^2\rangle^2 + \langle x^2\rangle^2 + 2\langle xy\rangle^2 + 2\langle yx\rangle^2 + 2\langle xz\rangle^2\right], \tag{4.1}$$

where x, y and z are components of the unit vector along the backbone in the N-H direction. The squared order parameter of each residue is shown in Figure 4-6 for all protonation configurations and compared to that from the NMR experimental data at pH 3.5 (Zink *et al.* , 1994). Similar to the RMSD and RMSF results, the loop regions demonstrate higher mobility than the helical domains. Although both the NMR and simulation order parameters in the L_{EB} and L_{CD} regions show smaller

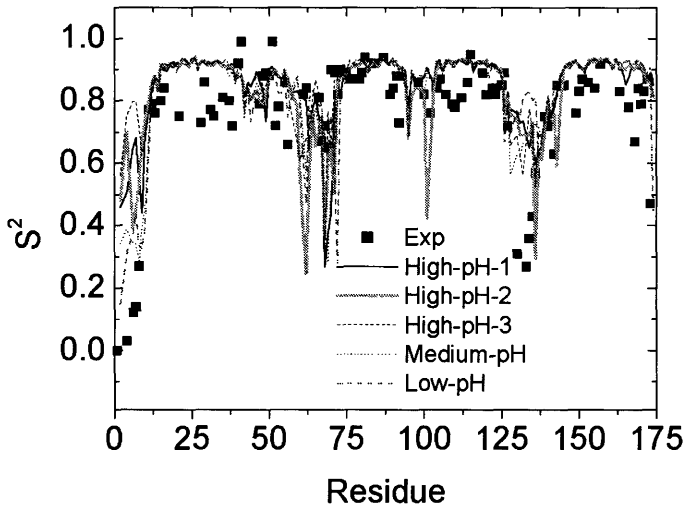


Figure 4-6: The squared order parameter of the backbone N-H vector of each residue calculated according to Eq. 4.1 for the last 1.5 ns of each simulation with a 0.1 ps interval between data points. Gray dots: NMR experiments at pH 3.5; black lines: calculated from the molecular dynamic simulations. Helical regions are marked on the X-axis of the top panel.

values compared to those of the helices, there exist deviations in magnitude. In any case, the simulation results averaged over the last 1.5 ns reproduce the general trend of the NMR data.

Conformational properties at different pH values

Titratable and charged residues of G-CSF are mainly contained in helical domains; the two long loops, L_{EB} and L_{CD}, however, do not contain these groups (Table 4.2). The change of charge distribution upon changing the protonation configuration affects the interactions among different regions in G-CSF and thus its structure. The change of conformation of G-CSF in the loop regions is shown in Figure 4-7. Structural changes can be also captured by properties such as hydrogen bonding patterns, and helical-helical distances and angles (Loof *et al.* , 1992). Hydrogen bonds (HB) not involved in forming the helical structures (see Figure 4-8) are analyzed and the results are shown in Table 4.3. Analyses show that the protonation/deprotonation of acidic

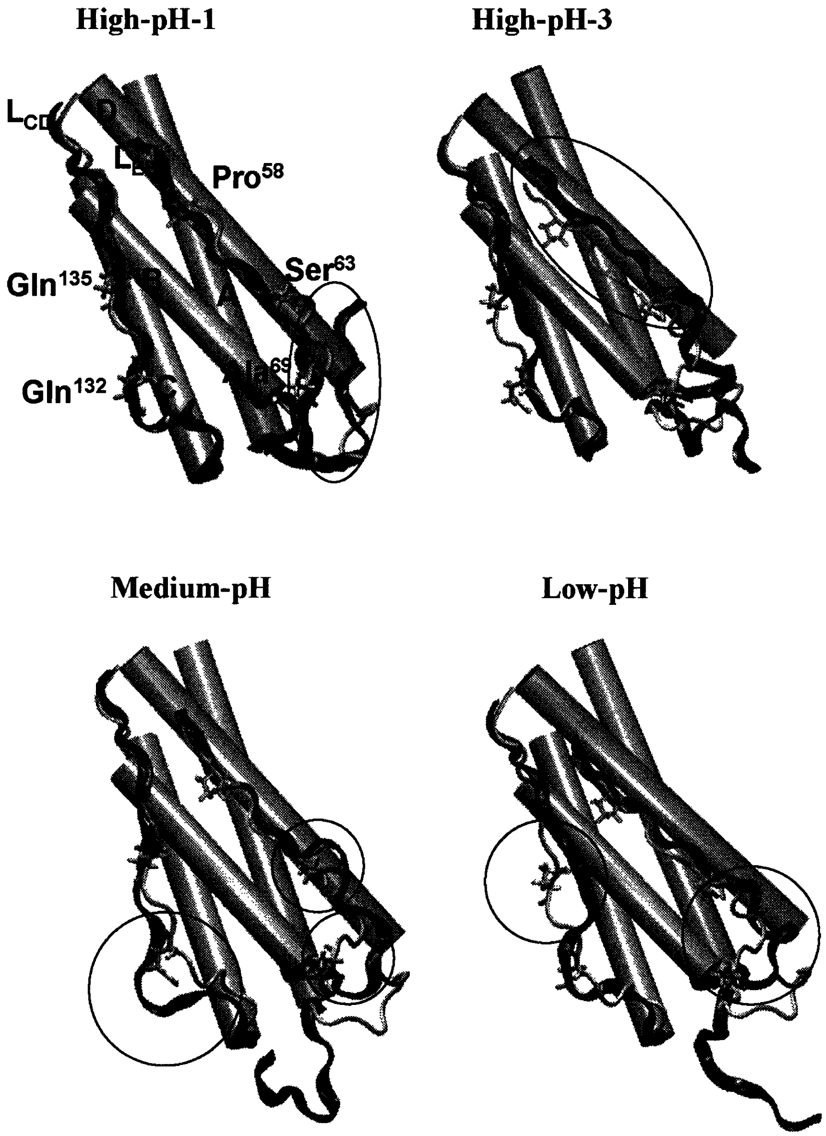


Figure 4-7: Representations of characteristic loop conformations of G-CSF at different protonation states. The X-ray structure (in gray) is also plotted for comparison.

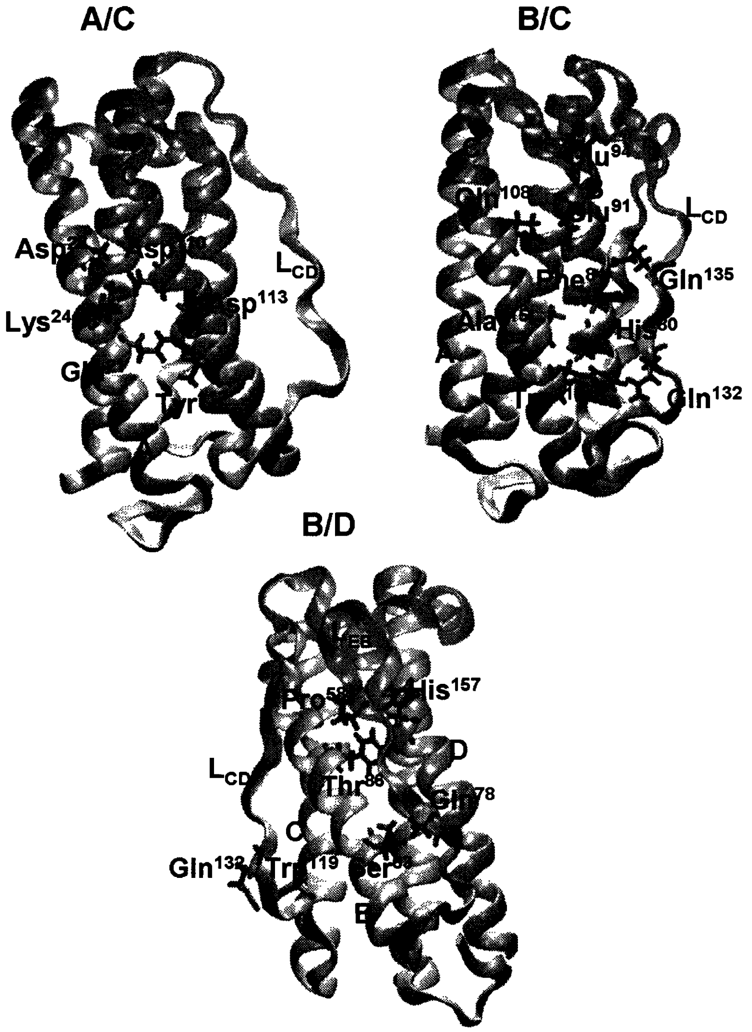


Figure 4-8: Representations of the persistent non-helical hydrogen bonds of G-CSF.

Table 4.3: Strengths of several persistent (see text) non-helical hydrogen bonds of G-CSF. The strength is the percentage of the dynamic trajectory satisfying the 2.4 Ådonor/acceptor distance and 150 degree D–H$\cdots$A angle criteria. The strength of hydrogen bonds is estimated by computing the percentage satisfying the above criteria in the last 1.5 ns of the simulations. Superscript "SC" denotes side-chain atoms and"BB" denotes backbone atoms.

			Strength		
Hydrogen bonded residues	high-pH-1	high-pH-2	high-pH-3	medium-pH	low-pH
			A/C		
Gln21(O^{SC})–Thr117(H^{SC})	56	0	73	19	56
Gln21(H^{SC})–Asp113(O^{SC})	11	41	48	24	0
Lys24(H^{SC})–Asp110(O^{SC})	94	96	89	96	0
Asp28(O^{SC})- Asp110(H^{SC})	0	0	0	0	21
			B/C		
Glu94(O^{SC})–Gln108(H^{SC})	0	0	0	0	45
Glu91(H,O^{SC})–Gln108(H,O^{SC})	34	10	19	12	0
His80(H^{SC})–Ala115(O^{BB})	0	11	0	56	0
			B/L_{CD}		
His80(H^{SC})–Gln132(O^{SC})	15	0	0 0	0	
Phe84(O^{BB})–Gln135(H^{SC})	68	77	59 69	39	
			C/L_{CD}		
Trp119(H^{BB})–Gln132 (O^{BB})	20	13	1.5	8	17
			L_{EB}/D		
Pro58(O^{BB})–His157(H^{SC})	4.5	54	0 91	87	
			L_{EB}/B		
Pro58(O^{BB})–Thr86(H^{BB})	0	9	20	0	0
Ser63(H,O^{BB})–Gln78(H,O^{BB})	76	70	0	0	5

groups significantly affects the helical/helical packing between helices A and C and between B and C since acidic residues are more populated in these regions of G-CSF. The change of helical packing also leads to the change of loop conformation such as shown in Figure 4-7. The protonation/deprotonation of histidine groups, on the other hand, mainly affects the conformation in the loop regions, especially His80 and His157, which form hydrogen bonds with other residues. G-CSF conformation is sensitive to the protonation states of His80 and His157; the results can be seen by difference in the loop conformation and hydrogen bond strengths as shown in Figure 4-7 and Table 4.3 for "high-pH-1", "high-pH-2", and "high-pH-3" configurations in which the net charge of G-CSF remains the same but the protonation state of His80 and His157 are different. These analyzes show that although the overall tertiary structure of G-CSF is stable across different pH values, there exists difference in conformation, even for the exposed L_0 region, at different pH values.

The protonation states of titratable residues, or the pH values, also change the correlation of the motion of different residues, as measures by the covariance fluctuation matrix, $C_{i,j}$:

$$C_{i,j} = \frac{\langle \Delta x_i \Delta x_j \rangle}{\sqrt{\langle \Delta x_i \rangle \langle \Delta x_j \rangle}}, \tag{4.2}$$

where Δx_i is the deviation of the i-th coordinate from its mean structure. The covariance fluctuation matrix can also be used in quasiharmonic mode determination. The results of the covariance fluctuation matrix between α-carbons are summarized in Figure 4-9. It can be seen in Figure 4-9 that there exist different types of long range correlation in different protonation states. These various types of correlations can be related to the conformational properties of each protonation configuration.

In summary, the molecular dynamics simulations show that altering the protonation states of the titratable residues of G-CSF, located mainly in the helical regions, can lead to distinct conformations for each protonation state. Despite the fact that the up-up-down-down arrangement of the helical domains of G-CSF remains stable, different conformational features, such as hydrogen bonding patterns, helix-helix packing, and pair-wise correlations, can be identified in each protonation configura-

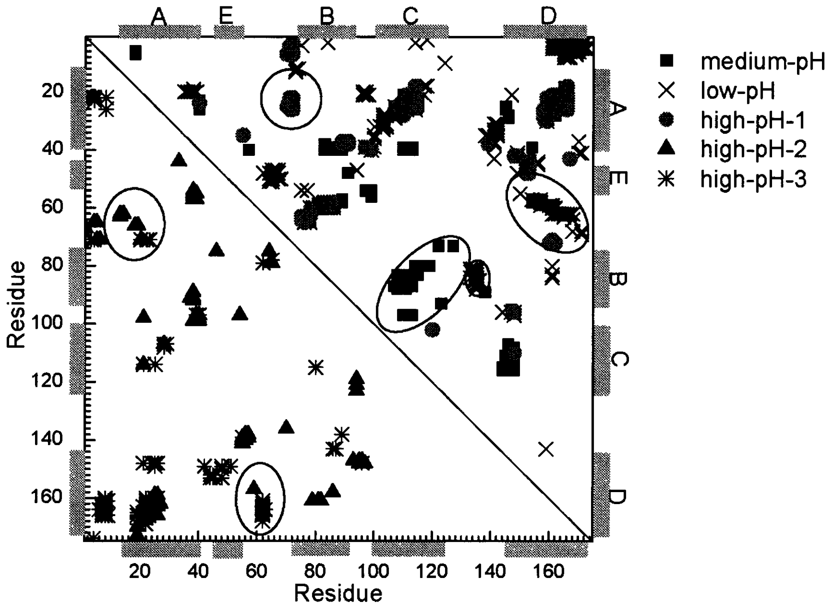


Figure 4-9: The atomic covariance fluctuation matrix of a-carbons calculated according to Eq. (4.2). The covariance matrices of the fluctuation of a carbons of G-CSF are calculated over 300 ps trajectory segments of the last 1.5 ns simulation, i.e., the reference structure is the average over 300 ps in each segment. Only persistent pairwise correlations, i.e. > 0.2 in all of the 5 segments, are considered. Short-range pairwise correlations between residues within ±10 in the primary sequence are also ignored.

tion even in the most exposed L_0 region. Since our proposed oxidation mechanism suggests that interaction of methionine sites with solvent molecules is governing, the oxidation rate must be a function of the local conformation of solvent and protein surrounding each methionine residue. Here we propose that the pH dependence of the conformation of G-CSF is most likely responsible for the observed pH dependence of the rates of oxidation (Figure 4-2). In the next section, we focus on evaluating why different methionine residues exhibit different rates of oxidation.

4.4.2 Characterization of solvent configuration and correlation with rates of oxidation

Differences in the oxidation rates of different methionine residues has been proposed to be due to the different structural environments of individual methionine residues (Lu *et al.* , 1999; Griffiths & Cooney, 2002; Meyer *et al.* , 2002). This difference could be due to differences in interactions among residues in a protein and differences in the interaction between the protein and the solvent. A better understanding these interactions would lead to a better understanding of oxidation and how to control it. Moreover, having a useful tool to estimate the relative oxidation rates would be helpful. This property could be taken from the X-ray structure or from molecular dynamics simulations, but as noted above, properly treating thermodynamic fluctuations is important, and it is unlikely that static X-ray data will be sufficient. We address the correlation of structural properties to the oxidation rates of methionine residues in this section. Solvent accessible area (SAA) (Lee & Richards, 1971) is traditionally used to characterize the degree of interaction between solvent molecules and sites in a protein. Since the access of the sulfur atom to oxidants in the solution is required for the oxidation reaction, assuming that the oxidation rate is proportional to SAA is the simplest model. The SAA (probe radius 1.6 Å) of the sulfur atom of methionine determined using the X-ray structure is only non-zero for Met1 and is zero for Met122, Met127, and Met138. Clearly, SAA's computed from X-ray structure are not good structural properties for describing oxidation. Comparison of average SAA's of methionine residues computed from molecular dynamics simulation results (last 1.5 ns) to the rates of oxidation at each protonation state is shown in Figure 4-10. All methionine residues have non-zero average SAA's, suggesting that each methionine residue is liable to oxidation. This result is qualitatively consistent with the experimental results shown earlier. This result also further emphasizes that using the SAA's from the X-ray structure of methionine residues is not sufficient, even qualitatively, for making reasonable estimates of relative rates of oxidation of G-CSF, and that considering the dynamic nature of protein motion is essential. Nevertheless, the SAA's captured from simulation do not show good quantitative agreement with the oxidation rates of methionine residues. Therefore, we conclude that SAA is not a good measure of the susceptibility of a methionine site to oxidation. Another measure of the degree to which the solvent interacts with sites on the protein is the water coordination number of the sulfur atom (WCN), i.e., the number of water molecules within a sphere of a predefined cutoff radius from the atom. *Ab initio* calculations discussed in Chapter 3 indicate that water molecules near the reaction site (the sulfur atom) stabilize the charge separation occurring during the breaking of the O-O bond of hydrogen peroxide in the oxidation process. The results discussed below use a cutoff radius of 5.5 Å to compute the WCN. This cutoff radius corresponds to two solvation shells around the sulfur atom (from the pair correlation function of the sulfur atom to the oxygen atoms of water). The choice of including two solvation shells was made based on the proposed water-mediated mechanism of oxidation in Chapter 3. Results are shown in Figure 4-10. Adjusting the cut-off radius does not lead to a better correlation between the of rate constants and WCN's, also leading to the conclusion that two solvation shells participate in the oxidation processes. Here, the WCN's are normalized to the WCN's of Met1 at each pH. In the high-pH-1 simulation, the correlation between the WCN's and relative oxidation rates of each methionine residues is very good, providing strong evidence that the rates of oxidation can be characterized by the local conformational and structural environment of each methionine residue. In the high-pH-2 simulation, the agreement is also good, in general, but the WCN is a bit larger than the oxidation rate of Met127. In the high-pH-3 simulation, however,

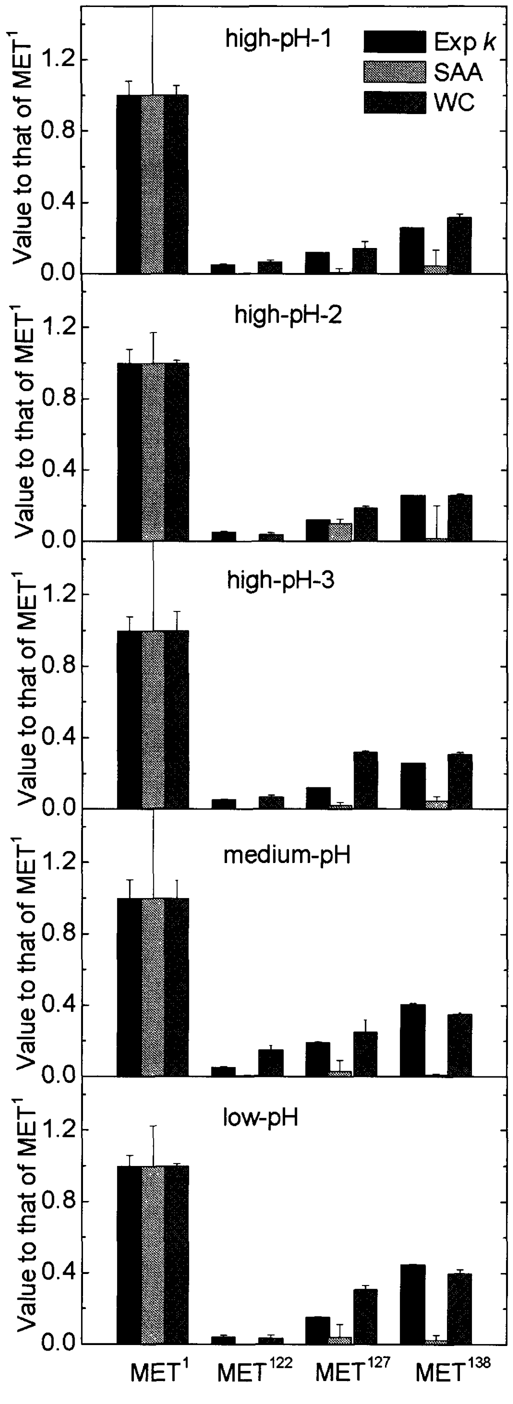


Figure 4-10: Comparison of structural properties to the rates of oxidation of methionine residues in G-CSF. Solvent accessible area (SAA) of methionine sulfur atoms (1.6 Åprobe radius), water coordination number (WCN) of methionine sulfur atoms with a cutoff radius of 5.5 Å, and the measured rates of oxidation of methionine residues (Exp) are normalized to the value of Met1 at each pH value. The average values over the last 1.5 ns are shown.

the WCN's of Met127 and Met138 are in reverse order to the rates constants. We interpret these results to mean that the protonation configuration adapted in the case of high-pH-1 is more representative of the experimental system than those of high-pH-2 and high-pH-3 at pH 8.5. In the medium-pH simulation, relative oxidation rates and WCN also correlate well at pH 5.5, but the normalized WCN of Met122 is larger than the normalized rate of oxidation. In the low-pH simulation, the WCN's of Met122 and Met138 correlate well to the experimental data, while the normalized WCN of Met127 is larger than the normalized rate of oxidation. At all pH values (taking the high-pH-1 simulation as representative of the high-pH case), the order of the normalized rates of oxidation found via experiment, Met1>Met138>Met127>Met122, are consistent with the order of the normalized WCN's for the methionine residues of G-CSF. The success in the correlation of the WCN's of sulfur atoms to the relative rates of oxidation and the failure of using SAA's also indicates that taking into account solvent effects in an explicit manner is essential. Some additional understanding of the relative rates of oxidation can be gained from viewing the results in Figure 4-10 in light of the protein structure in Figure 4-3. Met1 is the most exposed methionine residue, and therefore has the highest methionine oxidation rate. Met138 and Met127 are located in the L_{CD} region (Figure 4-3). Met138 is close to helix B and near the strong Trp119-Gln132 HB region of L_{CD} (Figure 4-8 and Figure 4-8). Met127 is close to the C-terminal of helix C (residues 100–125), and has a slower oxidation rate than Met138 at all pH values (2–8.5) and also smaller WCN's. Finally, Met122 is buried in helix C and has the lowest oxidation rate. A more general comparison between the WCN and the rates of oxidation of methionine residues can be made by normalizing the WCN'S and oxidation rate against the WCN's and oxidation rates of the free methionine residue (free Met). This is done in Figure 4-11. Note that the WCN's of the free methionine are determined in separate simulations, and the rate of oxidation of the free methionine acid does not show a pH dependence. As shown in Figure 4-11, the oxidation rate of Met1 at pH 5.5 is 49% of that of free Met, and similarly, the WCN of Met1 at that pH is 45% of that versus that of the free Met. This indicates that the sulfur atom of Met1 is not as exposed as that of free Met.

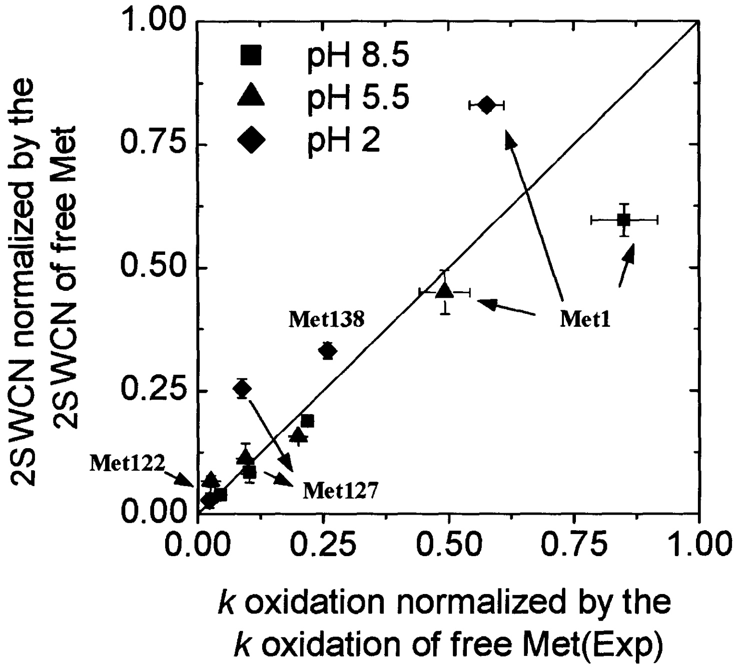


Figure 4-11: Correlation between the water coordination numbers (WCN) of the methionine sulfur atoms (cutoff radius 5.5 Å) and the rates of oxidation of methionine residues of G-CSF at different pH values. The rates of oxidation of methionine residues and methionine sulfur WCN's are normalized to the values of the free methionine amino acid.

Both the oxidation rates and WCN's of Met1 at different pH values are 49 85% of the corresponding values of free Met. The differences, particularly the minimum at pH 5.5, indicate that the local conformation around Met1 is also pH dependent. Similarly, both the oxidation rates and WCN's of Met138 depend on pH (minimum at pH 5.5) and agree well with each other quantitatively. For Met127, the pH dependence of oxidation rates is not significant, and the WCN's correlate well with oxidation rates at pH 5.5 and pH 8.5. The WCN is, however, larger than the normalized rate of oxidation at pH 2. For Met122, the pH dependence of oxidation rates is also not significant, and the WCN's correlates well with the oxidation rates. The above results demonstrate that the rates of oxidation of different Met groups correlate well to the WCN's of different Met groups. However, differences in the rates of oxidation of a particular Met residue as a function of pH can be explained only in part by differences in WCN, mainly because the effects of pH on conformation are subtle. Therefore, WCN can be used as a quantitative measure of the relative susceptibility to oxidation of methionine residues by H_2O_2 in a protein molecule. In this model, the effects of surrounding side-chain residues on the oxidation reaction are not considered. The fundamental mechanism of oxidation described earlier in Chapter 3 does not rule out the possibility that surrounding residues of a methionine sidechain with hydrogen bond donors and acceptors can play a similar role as solvent molecules, stabilize the charge separation when the O-O bond of H_2O_2 breaks. Moreover, it has been shown that carboxylic acids can form reactive intermediate with H_2O_2 and catalyzed the oxidation of various compounds (Shul'pin & Lindsay-Smith, 1998; Shul'pin *et al.* , 2001). Along the same line, Glu and Asp residues may also have catalytic effects on the oxidation of surrounding methionine residues. However, for methionine residues in G-CSF, the closest Asp or Glu residues are 10 Å away from the Met residues during the MD simulations. Therefore, the Asp and Glu residues in G-CSF likely do not affect the oxidation reaction; as a result, WCN's provide an accurate description on the relative rates of oxidation. The effects of residues with hydrogen bond donors and acceptors, if present, on the rates of oxidation of methionine residues are currently under investigation in our laboratory via *ab initio* calculations. Moreover, correlating a property such as WCN to rates of oxidation also assumes that each water molecule surrounding the sulfur atom, on average, stabilizes the oxidation reaction equally. The water molecules within the protein are sometimes bound (Sue *et al.* , 2001) and therefore may not be as available as the non-bound water molecules to stabilize the oxidation reaction. Such bound water molecules around methionine sidechains have not been observed in the MD simulations as estimated by monitoring the duration of each water molecule surrounding the sulfur atoms. This is probably because G-CSF is a protein with four compactly packed helices and does not have water binding sites near methionine residues. Moreover, we did not find a significant difference in the life time of a water molecule within 4 Å of the sulfur atom for different methionine groups. A more detailed analysis on the effects of protein environment of the oxidation reaction by QM/MM simulations is discussed in Chapter 6.

4.5 Conclusions

In this chapter, the measured the rates of oxidation by H_2O_2 of the Met groups as a function of pH are compared with the detailed information on the local environment of the methionine residues of G-CSF obtained via molecular dynamics simulations. It was found that the water coordination number (WCN) of each sulfur atom correlates well to the rate of oxidation. Thus, WCN can be used as a semi-quantitative measure of oxidation. Note that solvent accessible area (SAA), typically used in the analysis of proteins, does not correlate well with the rates of oxidation. Furthermore, static structural data from X-ray measurements cannot provide much insight into the oxidation process; dynamic information, such as that resulting from molecular dynamics (MD) simulations, is needed. The MD simulations show that the conformation of the protein is significantly different at different pH's. These differences can be characterized via properties such as hydrogen bonding, helix packing, and the covariance fluctuation matrix. Experiments show differences in the rates of oxidation of a Met site with pH. These results, plus the fact that there is no pH dependence of the oxidation rate of free methionine amino acids, suggest that the observed pH-dependence of methionine residues in a protein molecule is very likely to be due to the change of protein conformation as a function of pH.

Chapter 5

Oxidation of Methionine Residues in hPTH(1–34)

The relationship between conformational properties of 1–34 human parathyroid hormone (hPTH(1–34)) and the oxidation of its methionine residues, Met8 and Met18, by hydrogen peroxide is analyzed by performing molecular dynamics (MD) simulations with an explicit representation of water molecules. Between pH=4 to pH=8, both Met8 and Met18 have nearly pH independent rates of oxidation, and Met18 is oxidized at a rate 90 %–100 % of that of free methionine amino acid (freeMet) and 10–20 % faster than Met8. We found that average two-shell water coordination numbers (2SWCN's) calculated from MD simulations correlate well to the rates of oxidation of Met8 and Met18. The use of 2SWCN's is based on the mechanism that is proposed in Chapter 3, the water-mediated mechanism, in which water molecules stabilize the transition state via specific interactions, but the transfer of protons (acid-catalyzed mechanism) does not play a role. Only at very low pH values, pH=1 for the oxidation of freeMet, does the acid-catalyzed oxidation mechanism become important. For the oxidation of Met8 and Met18 in hPTH(1–34), the acid-catalyzed mechanism becomes significant at a higher pH value, pH=2, probability due to the proximity of nearby acidic residues to Met8 (Glu4) and Met18 (Glu22). In this chapter, we have demonstrated that the chemistry of oxidation and the structure of polypeptides can be correlated with a detailed understanding of the reaction mechanism, appropriate sampling of configurational space, and a suitable choice of a structural property, the water coordination number.

5.1 Motivation

The mechanism by which hydrogen peroxide, or hydroperoxides in general, oxidizes organic sulfides, such as methionine, has been elucidated in Chapter 3 by using *ab initio* calculations and molecular simulations. A summary of the conclusions is that specific interactions, such as hydrogen bonding, with surrounding water molecules stabilize the charge separation developed during the S_N2 oxidation of H_2O_2, i.e., during the breaking of the O-O bond and the formation of the S-O bond. This "water-mediated" mechanism is consistent with available experimental data such as activation energies (10–20 kcal/mol), and is significantly different from the acid-catalyzed mechanism that has been prevalent in the literature (Edwards, 1960; Ranky & Nelson, 1961; Dankleff *et al.* , 1968; Curci *et al.* , 1970; Bach *et al.* , 1991; Bach *et al.* , 1994; Li *et al.* , 1995). The water-mediated mechanism is also consistent with the fact that within a medium range of pH (pH=2–10), the rates of oxidation are generally pH independent (Amels *et al.* , 1997; Nabichi *et al.* , 1998).

The water-mediated mechanism also leads to the conclusion that access of solvent molecules to methionine sites governs oxidation, and the molecular property that correlates well to the rates of oxidation is found in Chapter 4 to be the ensemble average of the two-shell water coordination number (2SWCN) of the sulfur atoms of methionine residues. This property, incorporates both the dynamics of the protein and the explicit molecular simulation of water. For G-CSF, a protein with four compactly packed helices and well-defined tertiary structure, the average 2SWCN's were shown to correlate well with the rates of oxidation of different methionine residues as shown in Chapter 4. Application of this approach to flexible polypeptides without well-defined tertiary structure is one of the objectives in this chapter as an aid in gaining further understanding of the structural property/oxidation relationship of polypeptides.

In this chapter, we investigate the oxidation of methionine residues of a 34 residue peptide, 1–34 human parathyroid hormone (hPTH(1–34)) by hydrogen peroxide. We measure the circular dichroism (CD) spectra of hPTH(1–34) as a function of pH, and also perform molecular dynamics simulations of this molecule in order to find the correspondence between the conformational properties of hPTH(1–34) and the rates of oxidation of it methionine groups.

Human parathyroid hormone (hPTH) is a 84-amino-acid peptide with influence on cell proliferation of skeletal tissues and on bone metabolism (Sömjen *et al.* , 1991). It is generally accepted that the N-terminal 34 amino-acid fragment of hPTH is sufficient for biological activity, (Potts *et al.* , 1982) and in fact, it playings an important role in osteoporosis (Barden & Kemp, 1993) and hypoparathyroidism (Winer *et al.* , 1996) treatments. Compared to G-CSF(175 residues, compactly packed), studied in Chapter 4, hPTH(1–34) manifests significant structural variation in an aqueous solution as determined by NMR spectroscopy (Barden & Cuthbertson, 1993; Barden & Kemp, 1993; Klaus *et al.* , 1991; Strickland *et al.* , 1993; Pellegrini *et al.* , 1998; Marx *et al.* , 2000). HPTH(1–34) has two methionine residues, Met8 and Met18, each of which can be oxidized by H_2O_2 at a different rate (Nabichi *et al.* , 1998). Oxidation of methionine residues has also been shown to lead to the loss of biological activity of hPTH(1–34) (Nabichi *et al.* , 1998). These properties of hPTH(1–34) make it an important system to study in order to better understand the structural/oxidation relationship of polypeptides.

5.2 Details of experiments and calculations

5.2.1 CD spectroscopy

CD spectra were recorded at 20 °C in an 0.1 cm cell from 250 to 200 nm at 0.2 nm/sec (4 scans) on a Aviv 62 DS spectrometer. A protein concentration of 100 μM is prepared. CD spectra of hPTH(1–34) were collected at 5 pH values, controlled by different buffers, 1 (100 mM HCl), 2 (50 mM phosphate), 4 (50 mM citrate), 6(40 mM citrate), and 8(50 mM Tris). Ionic strengths were controlled at 200mM using NaCl. Estimations of the helical fraction of the peptide were calculated using the value of $[\theta]_{obs}$ at 222 nm, $(([\theta]^{222}_{obs} - 3000)/ - 39000 * 100)$ (Pellegrini *et al.* , 1998).

5.2.2 Molecular dynamics simulations

Molecular dynamics (MD) simulations of hPTH(1-34) and analyses thereof were performed using the CHARMM program (Brooks *et al.* , 1983) with the CHARMM 22 all-atom potential (MacKerell *et al.* , 1998). The initial configurations of hPTH(1–34) were obtained from the NMR structures of Marx *et al.* (Marx *et al.* , 2000), determined at pH 6.2, PDB entry 1ZWA, containing 10 different structures. Each structure was first solvated in a pre-equilibrated rectangular water box with at least 3 solvation shells from any atom of hPTH(1–34) to the box boundary. The largest box needed to satisfy this criteria contains 5326 water molecules and was used for the simulation of all structures. After inserting the molecule into the the water box, water molecules with oxygen atoms within 2.6 Å of the atoms of hPTH(1–34) were deleted. The total system then contained 15711 atoms and was subject to minimization before the dynamics simulation.

During the minimization calculations and MD simulations, periodic boundary conditions were applied to water molecules. Long range electrostatic interactions were calculated via the particle mesh Ewald summation method, (Darden *et al.* , 1993; Essmann *et al.* , 1995b) with a cut-off of 15 Å for real space interactions and for van Der Waals interactions. Bond lengths associated with hydrogen atoms were fixed at their equilibrium values using the SHAKE algorithm (Allen & Tildesley, 1987), thus allowing a time step of 0.002 ps for the integration of the equations of motion. The solvated structures were first minimized by 200 steepest descent steps followed by ABNR (Adopted-Basis Newton-Raphson) (Brooks *et al.* , 1983) minimizations, until the RMS (root of mean squared) force was less then 0.2 kcal/mol/Å. During the minimization steps, the backbone atoms of hPTH(1–34) were restrained to their NMR positions with a spring constant of 30 kcal/mol/Å. The minimized system was then brought to 300 K by assigning velocities according to Botzmann distributions with an increasing rate of 3 K/0.1 ps for 10 ps. After heating, the system was coupled to a Langevin piston (Zhang *et al.* , 1995) to control the system pressure at 1 atm and a Nosé-Hoover thermostat (Nosé, 1991) to control the temperature at 300 K. The simulations were then continued for 2.04 ns. In order to sample the conformation of hPTH(1-34) more efficiently, separate 2.05 ns MD simulations started from each of the 10 reported NMR structures were performed, instead of one long simulation started from one of the NMR structures (Caves *et al.* , 1998). Simulation results with only one of the sidechain nitrogen is protonated are denoted as "high-pH" (net charge= 1), when both nitrogen atoms of histidine residues are protonated, the results are donoted as "medium-pH" (net charge= 4). The same procedures were applied to all simulations performed in this study. If not otherwise mentioned, statistics reported in this study are taken from the last 1 ns of each simulation. Statistical uncertainties of each simulation were estimated by using the standard asymptotic block-averaging method (Allen & Tildesley, 1987; Frenkel & Smit, 1996).

The water coordination number is defined as the number of water molecules within a cutoff region around a specified atom, for example, the sulfur atom of a methionine residue. The choice of cutoff radius to describe the rates oxidation is 5.5 Å (2SWCN). The rationale behind this option is that water molecules surrounding hydrogen peroxide participate the reaction via specific interactions such as hydrogen bonding. Such interactions are needed to stabilize the charge separation during the S_N2 oxidation by H_2O_2. In order to capture the effects of water molecules surrounding the reactants (the sulfur atom and H_2O_2), two solvation shells are used and indeed have better performance in correlating the rates of oxidation than different cutoff radii as shown in Chapter 4.

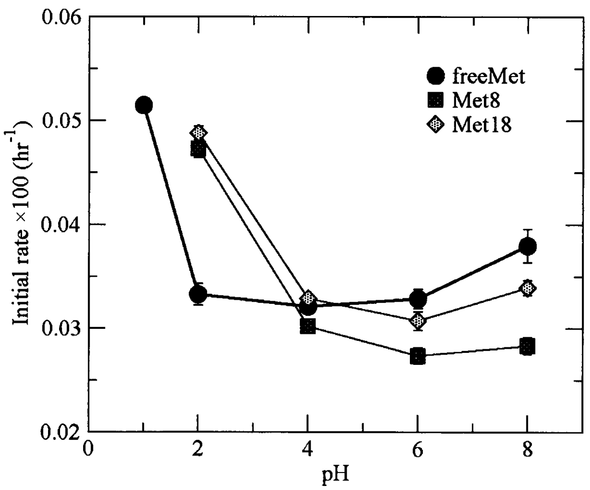


Figure 5-1: Pseudo first order rate constants for the oxidation of the methionine amino acid (freeMet) and methionine residues in hPTH at different pH values.

5.3 Summary of experimental data on the oxidation of free methionine and methionine residues in hPTH(1–34) at different pH values

In this section, the experimental data of the oxidation rates of free methionine (freeMet) and methionine groups in hPTH(1–34) are summarized as the basis to compare with the computational results. The experimental details are reported elsewhere (Chu *et al.* , 2004b). The measured initial rates of oxidation of free methionine (freeMet), Met8, and Met18 as a function of pH are shown in Figure 5-1. The rates of oxidation of freeMet do not depend on pH significantly between pH=2 to 8, another validation of the water-mediated mechanism that is proposed for oxidation in Chapter 3. The rate of oxidation of freeMet, however, jumps significantly when the pH is lowered to 1. This is attributes to the manifestation of the acid-catalyzed mechanism proposed by Bach *et al.* (Bach *et al.* , 1994), which becomes important as the concentration of hydronium ions become large.

For the oxidation of methionine residues in hPTH(1–34) above pH=2, Met18 is oxidized at a rate of 90 %–100 % of that of freeMet, and Met8 is oxidized 10–20 % slower than Met18. The rates of oxidation of both Met8 and Met18 are not very

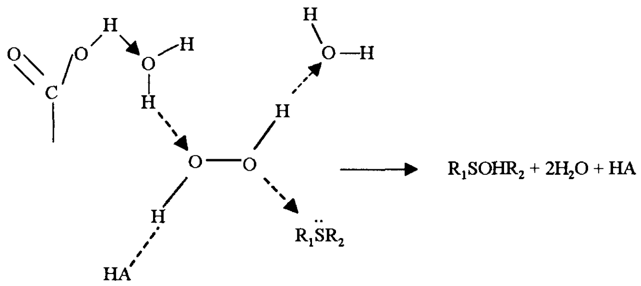


Figure 5-2: Proposed the acid-catalyzed mechanism in the presence of a protonated acidic residue at low pH values. In the acid-catalyzed mechanism, a protonated solvent molecule is needed, and at low pH values, a surrounding acidic residue can provide the additional proton. A carbonyl oxygen at a distance of 2–3 solvation shells from the sulfur atom suffices for this purpose

sensitive to pH for pH > 2, as is the case for freeMet. The small difference in the rates of oxidation between Met8 and Met18 indicates that the local environments around these two residues are only slightly different. These differences will be examined later. At pH=2, however, both Met8 and Met18 are oxidized faster than freeMet, and the rates of oxidation of each is close that of freeMet at pH=1. We interpret this result to mean that the acid-catalyzed mechanism becomes significant in hPTH(1–34) at a slightly higher pH value than in freeMet. A possible reason for this observation is that an acid amino residue, if is in the vicinity of a methionine residue, can also donate its proton for the acid-catalyzed oxidation at low pH. A schematic representation of the acid catalyzed oxidation mechanism near an acidic group is shown in Figure 5-2. HPTH(1–34) has a total of 4 acidic residues, including Glu4, which is near Met8 in the N-terminal helix, and Glu19 and Glu22, which are near Met18 in the C-terminal helix. (See Figure 5-3.) Therefore, the effective concentration of protons around Met8 and Met18 of hPTH(1–34) could be higher than that around freeMet at the same pH value, and this increased concentration results in a more significant effect of the acid-catalyzed mechanism at pH=2. Note that there are no acidic residues in the proximity of Met1 in G-CSF, the protein studied in Chapter 4, and this residue has a rate of oxidation less than that of freeMet at pH=2. For most practical situations, however, the acid-catalyzed mechanism will not be important, since the range of pH values that

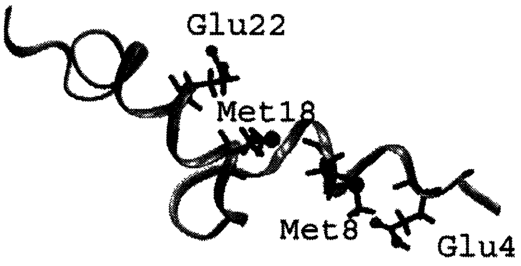


Figure 5-3: Ribbon representation of 1ZWA-2, the second of the selected NMR structures (PDB entry: 1ZWA). (Glu4 and Met8) and (Met18 and GLu22) are shown in black. Note that the distances between Glu4 and Met8 and between Met18 and Glu22 are around 5.5 Å.

is usually used in the formulation of protein pharmaceuticals is pH=4.5–7.5 (Powell, 1996), where water-mediated mechanism is most relevant.

5.4 Results and discussions

5.4.1 Structural properties of hPTH(1–34)

Now we start to address the relationship between the rates of oxidation of the methionine residues of hPTH(1–34) and the structural properties of the peptide in the range of pH values where the water-mediated oxidation mechanism is dominant (pH=4–8). In addition to CD spectra, the configuration space of hPTH(1–34) in solution are also explored by performing MD simulations according to the protocols described earlier. We first present the results of CD spectra and MD simulations. The correspondence between the structural properties of hPTH(1–34) and the rates of oxidation of Met8 and Met18 is then examined by using the 2-shell water coordination number (2SWCN), which has been shown in Chapter 4 to correlate well to the rates of oxidation of methionine residues in G-CSF.

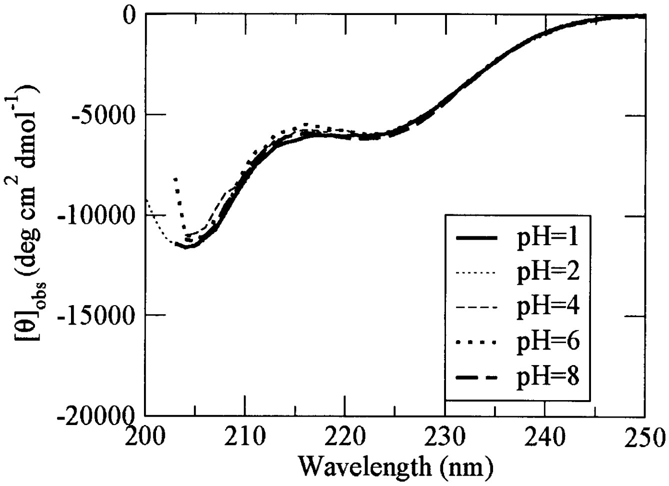


Figure 5-4: The measured ellipticity of hPTH(1–34) as a function of pH.

CD spectra

HPTH(1–34) has many chargeable residues, including three His residues (His9, His14, and His32), four acidic residues (Glu4, Glu19, Glu22, and Asp30), and five basic residues (Lys13, Arg20, Arg25, Lys26, and Lys27). Therefore, the intramolecular electrostatic interactions are expected to play important roles in the structure of hPTH(1–34). Indeed, it has been shown that increasing the ionic strength results in an slight increase of the helicity of hPTH(1–34) (from 24 % to 28 %) (Pellegrini *et al.* , 1998).

The measured ellipticity of hPTH(1–34) at different pH values in shown in Figure 5-4, and the estimated helical fraction of hPTH(1–34) as a function of pH is shown in Figure 5-5. It can be seen that the helical content of hPTH(1–34) is around 23 %, independent of pH from pH=1 to pH=8, and this is close to the value reported in an earlier study, 24%, prepared at pH=6.3 (Pellegrini *et al.* , 1998). This data suggests that the secondary structure of hPTH(1–34) does not change even in acidic conditions, implying that the conformational properties are not sensitive to pH. This result is also consistent with the result that the oxidation rates are not sensitive to pH in the range of pH=4–8, according to the water-mediated oxidation mechanism. It also adds further confirmation of our theory that the increases in the rates are due to the

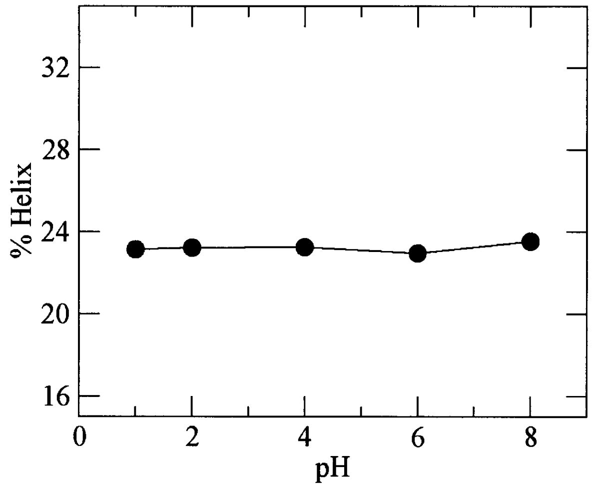


Figure 5-5: The helix fraction of hPTH(1–34) as a function of pH estimated via the measured ellipticity from CD spectroscopy, % helix= $(([\theta]^{222}_{obs} - 3000)/ - 39000 * 100)$ (Pellegrini *et al.*, 1998)

acid-catalyzed mechanism coming into play. Note that the helical content of G-CSF, which has a well-defined tertiary structure, is sensitive to pH (Narhi *et al.* , 1991), and thus the rate of oxidation of methionine groups in G-CSF is a function of pH, even in the intermediate ranges of pH, when solely the water-mediated mechanism is important.

MD simulations of hTPH(1–34) and its conformational properties

While some studies indicate that hPTH(1–34) has a short N-terminal helix from Gln6 to His9, and a longer C-terminal helix from Met18 to Gln29 connected by a flexible hinge around Gly12, the secondary structures of hPTH(1–34), are not strictly regular, but are flexible and consist of extended or irregular structures (Pellegrini *et al.* , 1998; Marx *et al.* , 2000). For example, the chemical shifts of the residues in the N-terminal helix are only marginally lower than those of a random coil (Pellegrini *et al.* , 1998; Marx *et al.* , 2000), and the α-helical H$\alpha(i)$-HN$(i+3)$ NOEs are of low intensity for i=6,7 in the N-terminal helix and for i= 24,26 in the C-terminal helix (Pellegrini *et al.* , 1998). Indeed, the number of residues with helical NOE signals is around 17 (Marx *et al.* , 2000), but the CD spectra only suggest the presence of 8 helixal residues.

Moreover, distance geometry calculations also indicate that a single structure cannot fulfill all of the experimental restraints, suggesting the existence of both structured and unstructured peptide structures in an aqueous solution (Pellegrini *et al.* , 1998). HPTH(1–34) also has very few long-range NOE signals, a further indication of its flexibility, and the major long-range interactions identified by NMR spectroscopy are between the δ-protons of Leu15 and the aromatic protons of Trp23, both hydrophobic. Based on the looseness of the structure on this molecule, we choose to study the 10 most representative structures determining via matching NOE constraints measured at pH=6.3 (Marx *et al.* , 2000), PDB entry: 1ZWA. Each structure is denoted as 1ZWA-i, i=1–10 in this work.

The 10 structures can be roughly clustered into four groups based on the overall folding as determined by geometrical quantities, such as radius of gyration. Ribbon representations of these structures are shown in Figure 5-6. Since each of the 10 structures are very distinct from one another, we report both the properties of each structure and the average properties of all 10 structures. In the following sections, the "Avg-" symbol will be placed in front of a property if it is averaged over the properties of the 10 structures.

The Root of Mean Squared Deviations (RMSD) of α-carbons in each MD trajectory are computed with respect to the NMR structure. The average of all 10 of these RMSDs is 4.8±0.9 Å for high-pH and 5.2±0.76 Å for medium-pH. The fairly large α-carbon RMSDs indicate that hPTH(1–34) indeed is flexible and deviates significantly from the starting NMR structures during the period of the 2 ns simulations. Note that the cross RMSDs (RMSDs of the trajectory of 1ZWA-j with respect to the starting structure of 1ZWA-i) indicate that each simulation is more similar to its own starting structure than to other NMR structures, and transitions from one structure to another were not observed.

The NMR structures of hPTH(1–34) were determined by Marx *et al.* via 207 NOEs using constrained MD simulations and simulated annealing techniques (Marx *et al.* , 2000). We computed the percentage of time that NOEs are satisfied during the last 1 ns of a trajectory, and present the results in Table 5.1. The criteria for satisfying

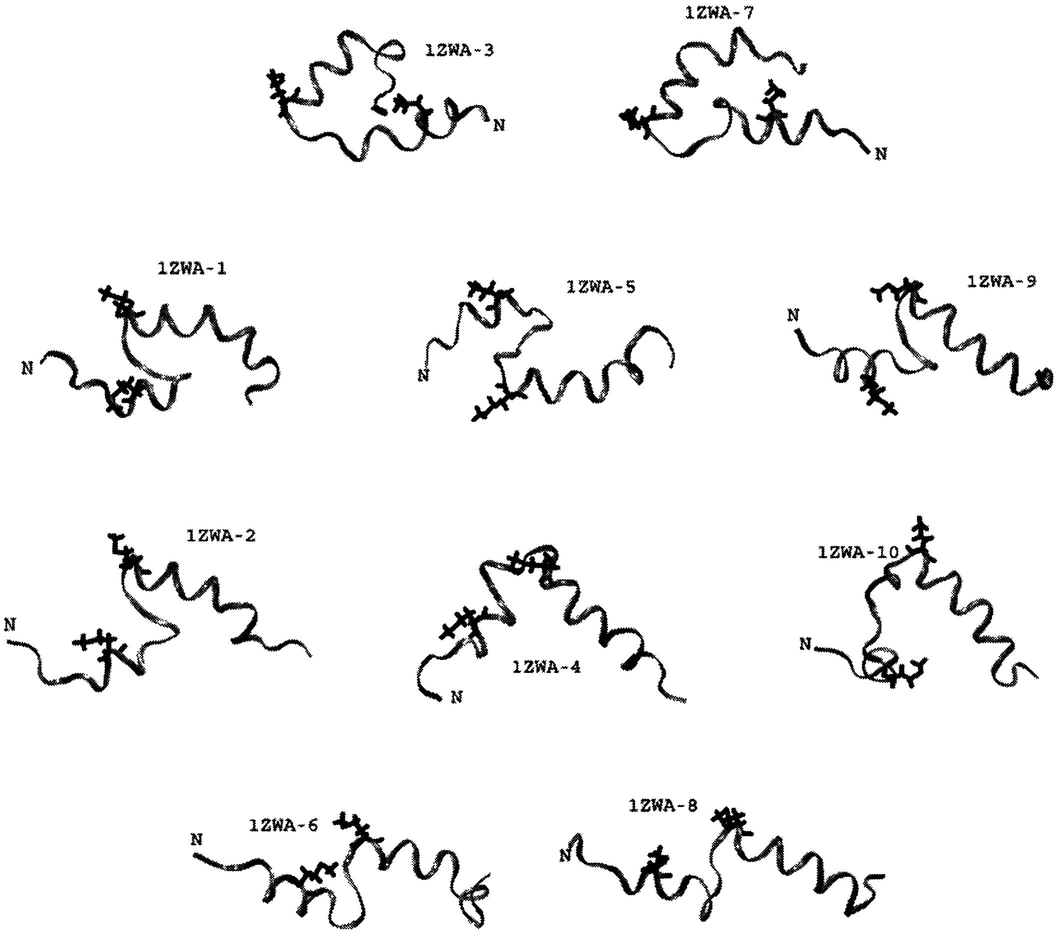


Figure 5-6: Ribbon representations of NMR structures of hPTH (1–34), (PDB entry: 1ZWA). Each structure is labelled as 1ZWA-*i*, *i*=1–10. The 10 NMR structures are roughly clustered into 4 groups, *i*= 3 and 7, *i*= 1, 5, and 9, *i*=2, 4, and 10, and *i*= 6 and 8. The groups are listed in an ascending order according to the value of the radius of gyration of their α-carbons. Met8 and Met18 are shown in black, and the location of the N-terminus is labelled by the capital letter, N.

Table 5.1: The percentage of time that the various NOE constraints are satisfied in the MD trajectories.

simulation	short-range[a]		medium-range[b]		helical residues[c]		Leu15-Trp23[d]	
	high[e]	medium[f]	high	medium	high	medium	high	medium
1ZWA-1	71	69	52	37	13.21	12.81	87	93
1ZWA-2	68	69	39	35	10.54	9.52	28	75
1ZWA-3	69	69	27	37	8.66	8.49	88	100
1ZWA-4	70	71	55	42	14.81	13.86	100	0
1ZWA-5	74	74	44	52	13.63	11.93	81	2
1ZWA-6	67	67	53	37	19.63	11.22	89	95
1ZWA-7	70	68	12	25	3.29	12.74	0	25
1ZWA-8	70	68	52	48	14.76	11.94	100	9
1ZWA-9	62	62	34	21	9.69	10.22	0	5
1ZWA-10	70	68	53	42	12.85	12.62	0	57
Average-1[g]	69	69	42	38	12.11	11.53	57	46

---

[a] The percentage of i+1 and i+2 NOEs (127 in total) that are satisfied by a trajectory.

[b] The percentage of i+3, i+4, and i+5, NOEs (75 in total) that are satisifed by a trajectory.

[c] The average number of helical residues in a trajectory identified via the DSSP program(Kabsch & Sander, 1983).

[d] The percentage of a trajectory with long range interactions between Leu15 and Trp23 (3 NOEs).

[e] Only one sidechain nitrogen of histidine is protonated.

[f] Two sidechain nitrogens of histidine are protonated.

[g] Average over 10 simulations.

the constraints depend on the relative intensities (Marx *et al.* , 2000), and 2.0–4.0 Å criteria is used for strong intensity NOEs, 2.0–5.0 Å is used for medium intensity NOEs and 2.0–5.5 Å for weak intensity NOEs in the analyzes. The NOEs can be divided into 3 categories: "short-range" includes i+1 and i+2 NOEs, a total of 127, "medium-range" includes i+3, i+4, and i+5 NOEs, a total of 75, and "long-range" is for $> i + 5$ and is observed in hPTH(1–34) only between Leu15 and Trp23.

As shown in Table 5.1, over 60 % of the short-range NOEs are satisfied in each trajectory and the average is 69 % for both high-pH and medium-pH simulations. For medium-range NOEs, however, the percentages vary significantly among different NMR structures, and the average % of NOEs is around 40 in high and medium-pH simulations. Since the medium-range NOEs are closely related to helical secondary structures, this result of MD simulations indicates that the helical structures of hPTH(1–34) have significant variation in an aqueous solution, consistent with the results of earlier NMR studies (Pellegrini *et al.* , 1998; Marx *et al.* , 2000). Although only 40 % of medium-range NOEs are satisfied, the number of helical residues identified using the DSSP program (Kabsch & Sander, 1983) still has an average value of around 12, as shown in Table 5.1. Note also that in the reported NMR structures (Marx *et al.* , 2000), the average number of helical residues is 16, but CD spectra only suggest the existence of 8 helical residues. These results indicate that not all NOEs related to the helical structures are satisfied in an aqueous environment. This is similar to the conclusion of earlier distance geometry calculations (Pellegrini *et al.* , 1998) on hPTH(1–34).

The Leu15-Trp23 interactions have been observed both at pH=6 and pH=4 from NMR spectroscopy(Barden & Kemp, 1993; Pellegrini *et al.* , 1998; Marx *et al.* , 2000). In the MD simulations, an average of 57 % of the 10 trajectories exhibit the Leu15-Trp23 interaction in the high-pH simulations and 46 % in the medium-pH simulations. Similar to the medium-range NOEs, the Leu15-Trp23 interaction has very different values over the 10 simulations due to the structural diversity of hPTH(1–34). Although the Leu15-Trp23 NOEs are not satisfied in a significant percent of the trajectories, in most of these trajectories, the amount by which the distances exceeds the NOE constraint is below 2.0 Å, and only 1 out of the 10 trajectories has an average violation > 6 Å (1ZWA-9 in high-pH, and 1ZWA-4 in medium-pH). Therefore, we conclude that there exists long-range interactions between Leu15 and Trp23 in the MD simulations. Note also that there does not exist a clear trend between the medium-range NOEs, which are related to the helical structures, and the Leu15-Trp23 interactions. For example, 1ZWA-3 in both the high-pH and medium-pH simulations exhibits strong Leu15-Trp23 interactions (Table 5.1) but has relatively weak medium-pH NOEs. On the other hand, for 1ZWA-8 in the medium-pH simulations, the % of satisfied medium-range NOEs is above average, but the Leu15-Trp23 interaction is very weak.

In order to understand more about the structural variation of hPTH(1–34) in an aqueous solution, we plot the local RMSD (lRMSD) of α-carbons with respect to NMR structures with a five amino acid window (Wishart *et al.* , 1995) (Figure 5-7). The results demonstrate that the N-terminal, C-terminal, and the hinge (around

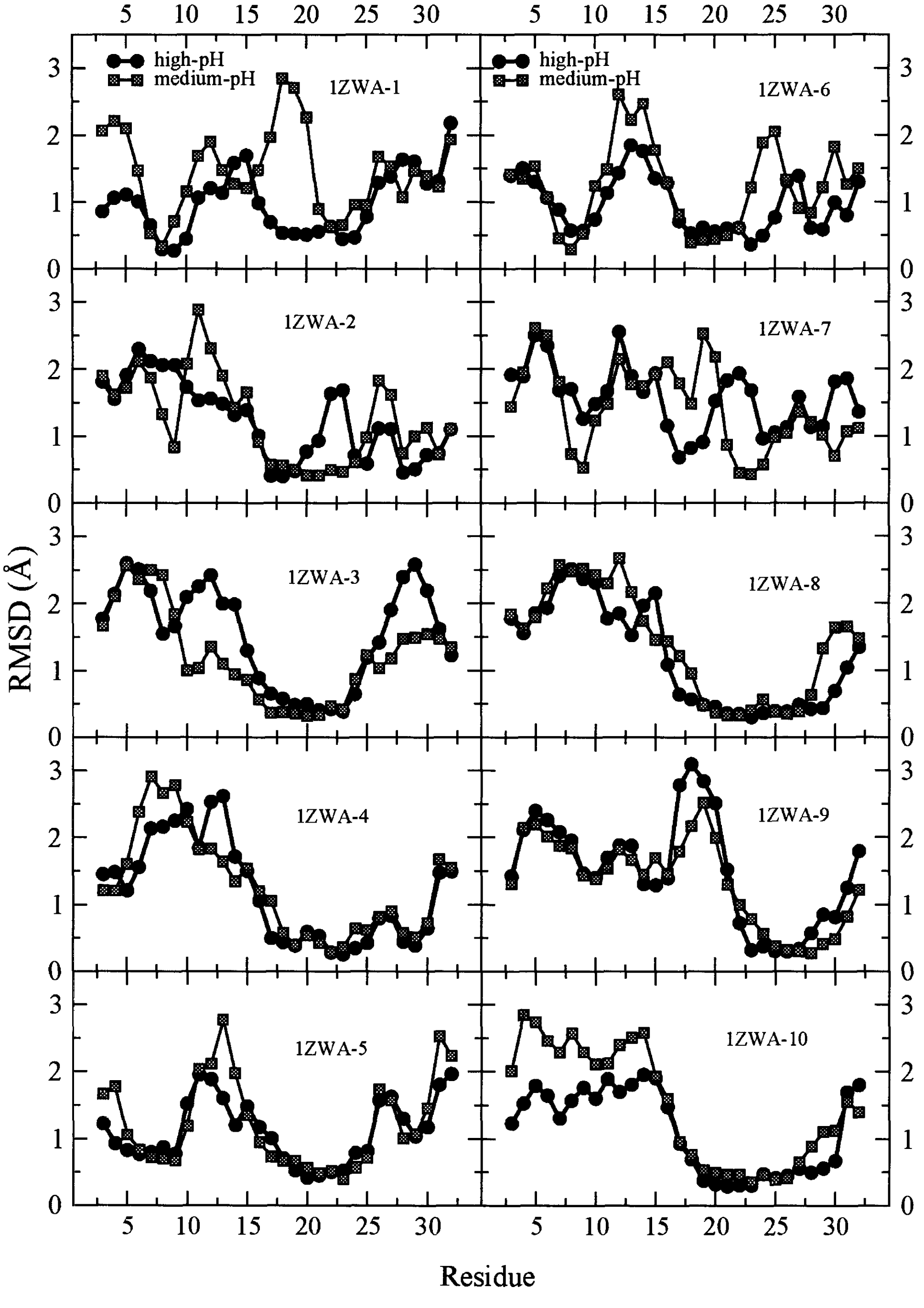


Figure 5-7: Local RMSD (with respect to the NMR structures) of α-carbons with a five amino acid window, averaged over the last 1 ns of each MD simulation.

Gly12) regions of hPTH(1–34) are generally more flexible than the helical regions. The lRMSDs of residues 19–29 are small (< 1 Å) in most simulations, indicating that the C-terminal helix is quite stable. Large lRMSDs (> 1.5 Å) that occur in the C-terminal helix regions in some of the 10 simulations usually involve residues in the beginning of the helix (around Glu19) or around the middle of the C-terminal helix (Glu22 or Arg25) and are consistent with the observation that the helical NOE's around this region are weak (Pellegrini *et al.* , 1998). The N-terminal helix, however, is less stable in the MD simulations than what would be expected from the NMR structures. In the cases of large lRMSDs (> 1.5 Å) in the N-terminal helix region, residues 6–9, the helicity is usually lost. This leads to the smaller numbers of helical residues in MD simulations as shown in Table 5.1. The above results suggest that the N-terminal helix can indeed exist in irregular structures as reported in the NMR studies (Pellegrini *et al.* , 1998; Marx *et al.* , 2000).

In summary, the results of MD simulations are consistent with the structural properties of hPTH(1–34) observed in NMR spectroscopy and CD spectra, i.e., hPTH(1–34) has a short helix in the N-terminal (residues 6–9) but may exhibit unstructured or irregular conformation, has a longer helix (19–29) in the C-terminal that is more stable, and has long range interactions between the hydrophobic residues Leu15 and Trp23. Moreover, we did not observe a strong dependence of the above properties on the degree of protonation of the histidine residues, i.e., high-pH versus medium-pH simulations, indicating that the structural properties of hPTH(1–34) are not sensitive to pH, as was suggested by the CD spectra shown in Figure 5-5.

5.4.2 Characterization of solvent configuration and correlation with the rates of oxidation of Met8 and Met18

In this section, we use the 2SWCN, found to correlate well to the rates of oxidation of methionine in G-CSF, to investigate hPTH(1–34).

The 2SWCN's of the sulfur atoms of Met8 and Met18 averaged over the last 1 ns of each simulation are normalized to the value of freeMet and are shown in Figure 5-8.

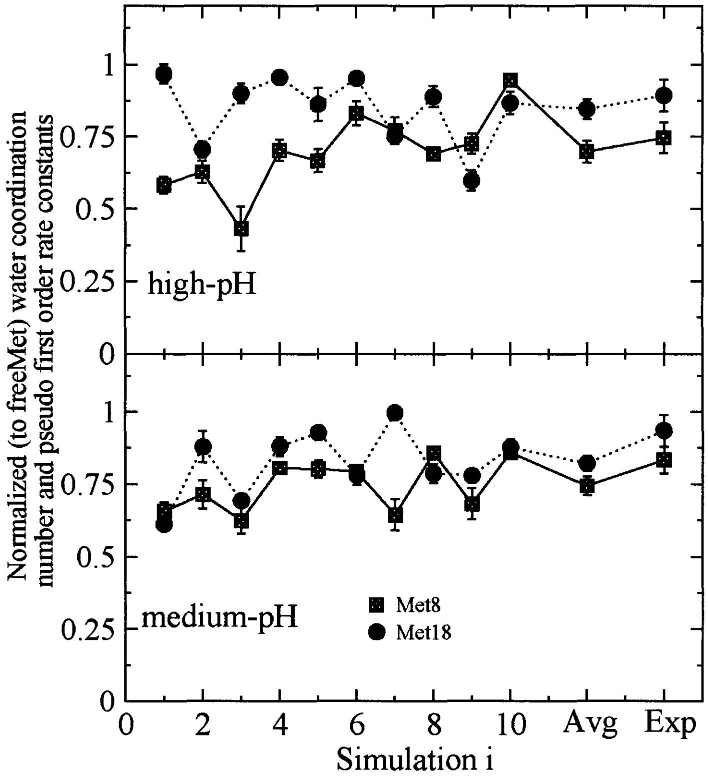


Figure 5-8: Two shell WCN's (normalized to the value of freeMet) of Met8 and Met18 of hPTH(1–34) for each MD simulation. The x-axis is the identity of the starting structure (1–10), and the y-axis is the normalized water coordination number. The values of the average over the ten structures is denoted as "Avg", and the measured rates normalized to the value of freeMet are denoted as "Exp". For high-pH, the rates at pH=8 of Figure 5-1 is plotted, and for medium-pH, the rates at pH=6 is plotted.

On the x-axis of Figure 5-8, the NMR structures (Figure 5-6) from which a simulation is started are labeled by the integer i, i.e., 1ZWA-i, i= 1–10. The average over the 10 simulations, denoted as "Avg" in the x-axis of Figure 5-8, and the normalized rates of oxidation (to the rates of freeMet) measured experimentally (Figure 5-1), are also plotted on Figure 5-8, and are denoted as "Exp". The rates of oxidation of freeMet, Met8, and Met18 measured at pH=8 are compared with the results from the high-pH simulations, and the rates measured at pH=6.3 were compared with the results from the medium-pH simulations.

It can be seen from Figure 5-8 that in most of the high-pH simulations, Met18 has a larger 2SWCN than Met8, except for 1ZWA-7 and 1ZWA-9. Among the medium-pH

Table 5.2: The normalized 2-shell water coordination numbers and the normalized rates of oxidation of Met1 and Met18.

	Met8/Met18[a]		Met8/freeMet[b]		Met18/freeMet[c]	
pH	Exp[d]	Avg[e]	Exp	Avg	Exp	Avg
4	0.92±0.01		0.94±0.01		1.02±0.01	
6[f]	0.89±0.05	0.92±0.07	0.83±0.05	0.74±0.03	0.94±0.05	0.82±0.03
8[g]	0.83±0.04	0.82±0.06	0.75±0.05	0.75±0.03	0.89±0.06	0.93±0.03

---

[a] The ratio between Met8 and Met18.
[b] The ratio between Met8 and freeMet.
[c] The ratio between Met18 and freeMet.
[d] Measured rates of oxidation (Figure 5-1).
[e] The average over all of the simulations.
[f] The simulation results in this row are from medium-pH simulations.
[g] The simulation results in this row are from high-pH simulations.

simulations, only 1ZWA-1 and 1ZWA-8 have 2SWCN's that are in the order opposite to the order of oxidation rates. Figure 5-8 also shows that the results are very different from one trajectory to another in both medium-pH and high-pH simulations, indicating that hPTH(1–34) indeed is flexible and exhibits diverse structures. Figure 5-8 also demonstrates that the 2SWCN's of the sulfur atom of methionine sites are very sensitive to the conformation of the protein molecule and describe the solvent exposure of methionine sites quite well. Although there is significant variation of the 2SWCN's among the 10 simulations, the average values correspond to the measured rates of oxidation for both Met8 and Met18. The average values of the 2SWCN's and the oxidation rates are shown in Table 5.2. Both the rates of oxidation and the 2SWCN's of Met8 and Met18 are not sensitive to pH between pH=4–8. The 2SWCN's have a general trend of slightly underestimating the normalized rates of oxidation, but the ratios between Met8 and Met18 are very close to the experimental ratios. Therefore, even the small difference between the oxidation rates of Met8 and Met18 can still be captured by using 2SWCN's.

From our simulation results, we can explain the reason for the small differences in the rates of oxidation of Met8 and Met18. Met18 of hPTH(1–34) is located in the beginning of the C-terminal helix and right after the loop between Asn10 and Asn16. Since there exists interactions between Leu15 and Trp23 and since the C-terminal helix is generally stable, the sidechain of Met18 is, for the most of time, exposed to solvent (see also Figure 5-6). Met8, on the other hand, is located in the more flexible N-terminal helix and is not within the Leu15 to Trp23 region. Therefore, the sidechain of Met8 can interact with the rest of the peptide more frequently (see also Figure 5-6), leading to lower 2SWCN's and thus, slower oxidation rates compare to Met18. For example, for a structure like 1ZWA-3 as shown in Figure 5-6, Met8 would have a small 2SWCN. Indeed, as shown in Figure 5-8, Met8 of 1ZWA-3 has a small 2SWCN compared to Met18 and 2SWCN's of Met8 in other high-pH simulations.

5.5 Conclusions

The following conclusions can be drawn from the above analyses.

1. The oxidation of free methionine by H_2O_2 proceeds at rates that are nearly independent of pH from pH=2 to 8, as explained by our "water-mediated" mechanism. At pH=1, we hypothesize that the acid-catalyzed mechanism starts to become significant, leading to an increase in the rate of oxidation. Of course, protein formulations generally have moderate ranges of pH, and thus, for them, the water-mediated mechanism will be relevant.

2. Both Met8 and Met18 of hPTH(1–34) can be oxidized at different rates in an aqueous solution. Met18 is oxidized 10–20 % faster than Met8 and both are oxidized more slowly than free methionine between pH=4–8. The rates of oxidation of Met8 and Met18 are also nearly pH independent between pH=4 to pH=8, but at pH=2, the rates of oxidation increase. We hypothesize that this increase is due to the acid-catalyzed mechanism becoming important, as the concentration of hydronium ions increases. These ions can change with the acidic residues very close to Met8 and Met18, effectively enhancing the local concentration of protons.

3. The structural property, 2SWCN, indicates that the local environments of Met8 and Met18 are not sensitive to pH between pH=4 to 8, and CD spectra also show that hPTH(1–34) has a constant helical content of 23 % between pH=1 to pH=8.

These results together are consistent with our water-mediated mechanism of oxidation, for which protein conformation governs oxidation, not acidity.

4. The rates of oxidation of Met8 and Met18 correlate to the 2-shell water coordination number obtained by averaging results from MD simulations on structures obtained from NMR data. Met8 is away from the long range interaction between Leu15 and Trp23 and is located in the flexible N-terminal helix, and Met18 is the starting residue of the C-terminal helix. Met8 thus, has frequent contacts with other residues in the peptide, lowering its exposure to solvent, while Met18 is generally exposed to solvent due to the interactions between Leu15 and Trp23. Therefore, the 2SWCN of Met8 is less than that of Met18, leading to Met8 having a lower oxidation rate than Met18.

5. The relationship between conformational properties and rates of oxidation has been made for a flexible peptide, hPTH(1–34), via extensive sampling of the configurational space of hPTH(1–34). This result is based on the water-mediated mechanism that we developed, and it, thus, can be employed to better understand how different sites in proteins are susceptible to oxidation to varying degrees. This is particularly important in the development of formulations for therapeutic proteins.

Chapter 6

QM/MM and Free Energy Simulations of the Oxidation of Methionine Residues in Aqueous Solutions, Free Methionine and Methionine in G-CSF

The free energy barriers of the oxidation of the amino acid methionine in water and in Granulocyte Colony-Stimulating Factor (G-CSF) are analyzed via QM/MM, constrained molecular dynamics, and transition path sampling simulations. The computed free energy barrier of free methionine amino acid is very close to the measured value (14.7±1.2 versus 15.5±0.02 kcal/mol). The reaction coordinate was found to be the difference between the O-O bond of H_2O_2 and the S-O bond, where the S is the sulfur atom of the methionine residue. It was confirmed by computing the commitor probability distribution and the distribution of constrained forces that this coordinate is not coupled to the activation of other degrees of freedom. The computed free energies of the oxidation of methionine residues in G-CSF indicate that the protein environment has insignificant effects on the reaction barriers of oxidation.

This result further validates our proposal that the access of solvent to methionine sites, as measured by the two-shell water coordination number, governs the kinetics of the oxidation reaction of methionine groups in a protein molecule. We also found that the number of hydrogen bonds between the distal oxygen of H_2O_2 and the water molecules near the methionine increases along the reaction coordinate as oxidation progresses, indicating that the charge separation developed during the S_N2 oxidation by H_2O_2 is stabilized by specific interactions with water molecules, such as hydrogen bonding.

6.1 Motivation

In Chapter 3, we have demonstrated that the mechanism of peroxide induced oxidation of organic sulfides that is consistent with all experimental data is the water-mediated mechanism, in which water molecules stabilize the developed charge separation during the breaking of the O-O bond of H_2O_2 and the formation of the S-O bond via specific interactions, such as hydrogen bonding. An example of the reaction path of the new mechanism is shown in Figure 6-1. It can be seen that breaking of the O-O bond of H_2O_2 and the formation of the S-O bond is the rate-limiting step, not the transfer of a proton as previously thought, since proton transfer occurs after the transition state, as shown in Figure 6-1. We have also shown that 2–3 water molecules are enough to stabilize the transition state of oxidation so that additional water molecules do not lower the barrier of reaction significantly (Chu & Trout, 2004). Although our mechanism satisfies all available experimental data on oxidation, such as the measured activation energies and the pH dependence of the rates of oxidation in the solution, it was developed for small molecules, not proteins, and it does not treat the dynamics effects of the solvent and protein environment.

We have also shown that 2–3 water molecules are enough to stabilize the transition state of oxidation so that the computed activation energy is within the range of the measured data as discussed in Chapter 3. Although our mechanism satisfies all available experimental data such as the measured activation energies and the pH

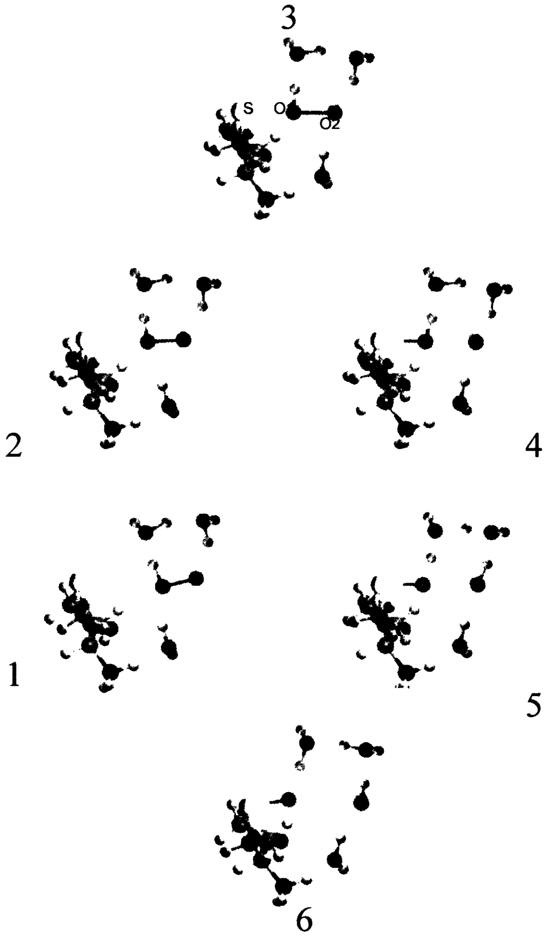


Figure 6-1: A minimum energy path (MEP) for the oxidation of methionine solvated in a water sphere of 18 Å radius. QM atoms are shown in ball and stick representation, and classical water molecules are shown as grey lines. Box 1 is the reactant cluster consisting of a free methionine amino acid, hydrogen peroxide, 3 QM water molecules, and 787 classical water molecules. Box 3 is the transition state, and Box 6 is the product. The MEP was obtained using the nudged elastic band method described in Chapter 2.

dependence of the rates of oxidation in the solution, it is relevant to small molecules, not proteins, and it does not treat the dynamics effects of solvent and protein environment.

The classical molecular dynamics study of a protein with sites that are susceptible to oxidation, Granulocyte Colony-Stimulating Factor (G-CSF) (Chapter 4), demonstrated that even sites that appear in x-ray structures to be buried can be accessed by water. Furthermore, the measured rates of oxidation of these buried sites in G-CSF and the methionine in hPTH(1–34) are found to correlate well with the water coordination number, averaged over two solvation shells as shown in Chapter 4 and in Chapter 5. These together with the results on organic sulfides in Chapter 3, imply that solvent access to methionine sites governs the oxidation process in proteins. However, it leaves many fundamental questions unanswered, such as:

1. What are the governing degrees of freedom of the oxidation process in solution for organic sulfides and in proteins?
2. What role does water dynamics play in the oxidation process?
3. What is the mechanism of oxidation of methionine sites in proteins?

This chapter addresses those questions by applying sophisticated theoretical and computational methods on an important therapeutic protein, G-CSF.

The effects of water and the protein molecule on the oxidation of methionine sites are quantified via employing a combined quantum mechanical and molecular mechanical (QM/MM) Hamiltonian (Das *et al.* , 2002; Gao & Truhlar, 2002). In order to compute the free energy barrier, constrained MD simulations (Sprik & Ciccotti, 1998) are performed along a reaction coordinate that was chosen from minimum energy paths (MEP) (Jónsson *et al.* , 1998; Chu *et al.* , 2003). The commitor probabilities of the configurations at the maximum of the free energy profile are then calculated to examine rigourously whether or not our hypothesized reaction coordinates are indeed the correct ones (Du *et al.* , 1998). In addition to broadening the current knowledge of the mechanism of oxidation of methionine residues in solution and in proteins, the result in this study add to our understanding of the application of transition path sampling methods for studying complex chemical processes in solution (Bolhuis *et al.* , 2000).

6.2 Details of calculations

6.2.1 QM/MM Models for the oxidation of a methionine amino acid and of methionine residues in G-CSF

In order to study the oxidation reaction in an aqueous environment, we perform combined quantum mechanical and molecular mechanical (QM/MM) calculations using all atom models, with at least 5 solvation shells (15 Å) of water molecules to solvate the methionine amino acid and G-CSF. The local region around and including the site being oxidized is treated via quantum mechanics, and the rest of the system is treated using classical force fields.

The QM region has 25 QM atoms, including hydrogen peroxide, 3 surrounding water molecules, the sidechain of the methionine residue being oxidized, and a QM link atom. It is described by the B3LYP density functional and the 6-31G basis set, using the GAMESS-UK program (Guest *et al.* , 1996). This choice of method is convenient because of the high cost of running dynamics on quantum mechanical systems. It is also quite accurate compared with higher level ab initio theories as tested on a model reaction, oxidation of dimethyl sulfide in the gas phase. The difference in activation energies between an MP4//B3LYP/6-31++G(d,p) calculation and a B3LYP/6-31G calculation is only 1.4 kcal/mol.

The classical region is treated using the CHARMM22 all-atom force field (MacKerell *et al.* , 1998). The QM/MM boundary is chosen to be in between the α-carbon and the β-carbon of the methionine residue. The double-link atom (DLA) method is used to divide the QM/MM interface, i.e., a QM link atom (hydrogen) is added to saturate the valence electrons in the QM region, and a classical link atom is also added to endeavor a net charge of zero in the classical region (Das *et al.* , 2002).

All electrostatic interactions between quantum particles, classical particles, and link atoms at the interface are included with the use of a Gaussian blurr function with a width of 2 Å on the classical link atom to avoid short-distance singularities. This method has been shown to remove the undesired artifacts related to the electrostatic properties of a chemical system, such as proton affinity induced by using only one link atom and by ignoring arbitrarily the electrostatic interactions between the link atoms and classical particles at the QM/MM interface (Das *et al.* , 2002). We also compared the QM/MM system with a full quantum treatment of the oxidation of a methionine amino acid with three water molecules in the gas phase, and found that the DLA method is able to reproduce the reaction activation energy within 0.6 kcal/mol.

The van der Waals parameters of the QM atoms are determined via fitting the QM/MM interaction energies and geometries to full QM results which are optimized at the B3LYP/6-31+G* level. These properties are used to treat the interaction between the QM part and the MM part. The model systems include a water dimer and a cluster of H_2O_2 and two water molecules. In the water dimer case, either one of the two water molecules can be chosen as the QM water, giving two QM/MM configurations. The TIP3P model is used as the MM force field, and the QM region is described via the B3LYP/6-31G theory. In the system consisting of H_2O_2 and two water molecules, the H_2O_2 molecule is treated quantum mechanically and the two water molecules are treated classically.

The initial configuration of the free methionine amino acid in the zwitterion form with hydrogen peroxide and three water molecules is obtained from geometry optimizations of gas phase models at 0 K. These configurations are then solvated by placing them in the center of a pre-equilibrated droplet of TIP3P water with a radius of 18 Å. Overlapping TIP3P water molecules are removed. The final system contains 787 water molecules and 2459 atoms in total. All nonbound electrostatic, QM/MM, and van der Waals interactions are calculated explicitly without using a cutoff.

The initial configuration of G-CSF is obtained from the X-ray structure, PDB entry 1CD9 (Aritomi *et al.* , 1999). From the 2.95 ns trajectory in an earlier study (Chu *et al.* , 2004a), a representative frame is chosen and so that the 2 shell water coordination numbers (2SWCNs) of methionine residues of the chosen frame are equal to the averaged value of this property along the trajectory. The protein molecule is then solvated with a shell of 15 Å. The system contains 5280 water molecules and, in total, 18495 atoms including G-CSF. To make the calculations feasible, a 17 Å cutoff is applied to nonbound interactions with a smooth switch function started at 15 Å, and the nonbound list is updated heuristically during energy minimizations and MD simulations.

6.2.2 Minimum energy paths of methionine oxidation

A useful approach to gaining insight into chemical and physical processes occurring in complicated systems is to find a minimum energy path (MEP) connecting two local minima on the potential energy hyper surface of a system. In addition to snapshots of a chemical process, the activation energy at 0 K can be obtained from an MEP. However, for reactions occurring in the liquid state, the dynamics of the solvent can be important, even governing, and free energies are the meaningful quantities to describe the processes. Nevertheless, knowledge of the MEP may still be useful in getting an indication of the underlying reaction coordinate. Moreover, the geometries along the MEP can also be used as the starting configuration for dynamic simulations and for free energy calculations.

The MEPs for the oxidation of free methionine and methionine sites in G-CSF are thus obtained using the nudged elastic band method (Jónsson *et al.* , 1998) and a superlinear minimization scheme based on the adapted basis Newton-Raphson method in CHARMM (Brooks *et al.* , 1983; Chu *et al.* , 2003). The details of calculations are reported in the following paragraph. As shown in the results, it turns out that the MEPs describe the mechanism correctly, Figure 6-1, and ξ=d(O-O)-d(S-O) is the reaction coordinate of oxidation.

A minimum energy path was generated by the nudged elastic band method using a superlinear minimization scheme based on the ABNR (adopted basis Newton Raphson) method (Brooks *et al.* , 1983). For the case of oxidation of free methionine, 25 replicas were generated to represent the reaction path. The reactant cluster (methionine, H_2O_2, and three water molecules) and the product cluster (methionine sulfoxide and four water molecules) were fully optimized at the B3LYP/6-31G level. The transition state in the gas phase is similar to the transition state structure in Figure 1. The calculated energy barrier using B3LYP/6-31G for the gas phase reaction is 11.3 kcal/mol, only 1.4 kcal/mol smaller than that computed at the MP4//B3LYP/6-31++G(d,p) level (Chu & Trout, 2004). Compared to the measured activation energy in aqueous solution, $\Delta E^{\ddagger}=13.2$ kcal/mol, the result of B3LYP/6-31G theory is also reasonable. Therefore, the B3LYP/6-31G method is used to describe the oxidation of methionine residues in water and in G-CSF.

In order to obtain a MEP for the oxidation of a solvated free methionine using NEB, the reactant and the product cluster of the solvated system were first fully relaxed by about 500 ABNR minimizations, until the resulting systems had an RMS (root mean squared) force of less than 0.02 kcal/mol/Å. An initial minimum energy path in the solvated system was then generated by replicating the whole system 25 times, the intermediate structures being chosen by linearly interpolating between the minimized reactant and product cluster. The minimum energy path had a similar reaction pathway to that in the gas phase, i.e., the transfer of proton occurs after the transition state. Before the transition state, the major component of the reaction path is the stretch of the O-O bond and the formation of the S-O bond. Thus, the reaction coordinate for the free energy simulation was chosen to be ξ=d(O-O)-d(S-O). Constrained MD simulations for the oxidation of free methionine are performed at ξ= -2.0, -1,6, -1.4, -1.2, -1.0, -0.8, -0.6, -0.4, -0.25, -0.2, and -0.1 Å to obtain the free energy barrier.

6.2.3 Constrained MD simulations and free energy calculations

The free energy barriers of the oxidation of methionine residues are calculated by integrating the mean force required to constrain the system at specific values along the reaction coordinate, ξ=d(O-O)-d(S-O). We have implemented the constraint in

CHARMM (Brooks *et al.* , 1983) using the SHAKE algorithm, and the formula developed by Sprik et al. (Sprik & Ciccotti, 1998) is used for the correction of the constraint MD ensemble, ($\xi = \xi'$, $\dot{\xi} = 0$), to the blue-moon ensemble ($\xi = \xi'$).

From the MEPs of oxidation, the reactants have ξ values of between -1.8 and -1.6 Å, and the transition states have ξ values around -0.2 Å. Because we are interested in oxidation, only the forward reaction barrier is calculated, i.e., the constrained MD simulations are stopped after the transition state is reached. Constrained MD simulations were performed at 9–10 values of ξ so that the free energy barriers could be computed via integration.

The G-SCF structure was obtained from the classical MD trajectory in an earlier study as mentioned in the paper. The protein molecule was first solvated in water with a solvation shell of 15 Å. The solvated system was then equilibrated by the velocity scaling method every 0.5 ps at 300K for 20 ps with all the atoms of the protein molecule restrained to their original positions by a spring constant of 200 kcal/mol/Å. The transition-state configuration of the oxidant and the three QM water molecules of freeMet were then placed in the vicinity of different methionine sites via a best fit procedure to form the initial structure, and any overlapped classical water molecules were removed. The system was then minimized (until RMS force < 0.01 kcal/mol/Å) and heated to (5 ps) and equilibrated at 300K (50 ps) with the QM energy turned off. The point charges and vdW parameters of H_2O_2 and QM water were set to the classical values during this procedure. During the equilibration, the positions of H_2O_2 and the three QM water were fixed, and the α-carbons of G-CSF were restrained at their original position by a spring constant of 100 kcal/mol/Å. The equilibration procedure was performed in order to remove contacts that the sidechain atoms may have with the H_2O_2 molecule and the three water molecules that will be treated quantum mechanically later. The equilibrated system was then further minimized with the aforementioned restraint conditions until the RMS force of the system was less then 0.01 kcal/mol/Å. The QM Hamiltonian was then turned on and minimized for 500 ABNR steps with the O-O and S-O distance restrained at values around the transition state, 1.95 Å and 2.05 Å respectively, using a restraint spring constant of 5000 kcal/mol/Å. The reactant and the product clusters were then obtained by releasing the restraint around the transition state. Then a MEP of 16 replicas was computed to refine the intermediate states, which were used as initial configurations for the free energy simulations. The same procedure was applied to all methionine residues studied in this work, Met122, Met127, and Met138. Constrained MD simulations were performed for the oxidation of Met127 and Met138 at ξ= -2.0, -1,6, -1.3, -1.0, -0.8, -0.6, -0.4, -0.2, -0.1, and -0.05 Å. For Met122, however, the starting configuration was ξ= -1.6 Å, due to its more restricted surrounding environment.

As described above, the starting structures for the MD simulations are chosen from the replicas on the MEPs. The leap frog (Allen & Tildesley, 1987) integration scheme is used to propagate MD trajectories of the QM/MM Hamiltonian, and a time step of 0.001 ps is employed. For each simulation, the system is first heated to 300 K by randomly assigning velocities from the Boltzmann distribution at an increasing rate of 30 K/0.05 ps. Next, the system is equilibrated for 0.5 ps by scaling velocities every 0.02 ps if the average temperature is beyond a 300±2 K window. The equilibrated system is then run for 10 ps, at the chosen value of ξ at 300 K, and the constraint forces are recorded. For the freeMet simulation containing 787 water molecules and 2459 atoms, the water molecules with a distance to the origin larger than 9 Å are propagated by Langevin dynamics (Pastor *et al.* , 1988; Loncharich *et al.* , 1992) with a collision frequency of 50 ps^{-1} and a bath temperature at 300 K. For the protein system, the velocity scaling method is applied at a frequency of every 0.5 ps if the averaged temperature is beyond a 300±2 K window. Statistical uncertainties of each simulation are estimated by using the standard asymptotic block-averaging method (Allen & Tildesley, 1987; ?).

Committor probably and transmission coefficient calculations

In order to examine the effects of the solvent on the oxidation reaction and whether the chosen ξ is indeed the reaction coordinate, the committor probability distribution (Du *et al.* , 1998; Bolhuis *et al.* , 1998; Bolhuis *et al.* , 2002) of the ensemble defined by the value of ξ at which there is a maximum in the free energy profile is

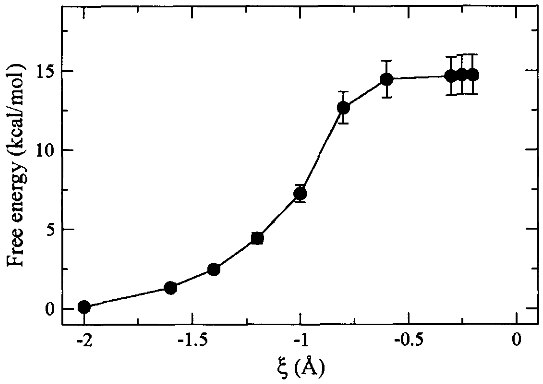


Figure 6-2: The free energy profile of the oxidation reaction of a methionine amino acid solvated in water molecules. The x-axis is the reaction coordinate, ξ (d(O-O)-d(S-O)) in Å, and the y-axis is free energy in kcal/mol, obtained by integrating the mean force along the reaction coordinate from the last 5 ps of the constrained MD simulations.

computed. 10 structures are randomly selected from the 10 ps trajectory and 20 0.1 ps trajectories are shot from each structure, with velocities randomly assigned from the Boltzmann distribution at 300 K. In this way, the probability of the structure going to the product state is obtained. If the ensemble chosen is indeed the transition state ensemble, the histogram of the committor probabilities should peak at 0.5 (Du *et al.* , 1998; Bolhuis *et al.* , 1998; Bolhuis *et al.* , 2002). This also means that the chosen reaction coordinate is correct. The transmission coefficient used for correcting dynamic recrossing events is also obtained, according to the methods of Carter *et al.* (Carter *et al.* , 1989).

6.3 Results and discussions

6.3.1 Oxidation of free methionine in water

The free energy profile for the oxidation of a methionine amino acid along ξ is shown in Figure 6-2. The calculated barrier is 14.7±1.2 kcal/mol, close to the experimentally measured value of 15.5±0.02 kcal/mol (Chu *et al.* , 2004b). The maximum of the free energy profile is located at ξ= -0.2 Å, and proton transfer does not occur during the

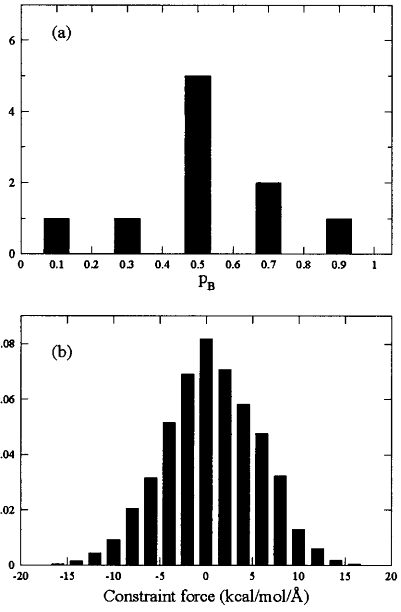


Figure 6-3: The statistics of the transition state ensemble obtained by constrained MD simulations at ξ= -0.2 Å. (a) the probably distribution of the committor probability of going to the product, p_B. In order to compute the committer probability distribution, 20 configurations are randomly chosen from the transition state ensemble. 10 0.1 ps trajectories are then generated for each configuration by randomly assigning the velocities of particles from the Boltzmann distribution at 300K. From the 10 shooting trajectories, p_B is calculated. The p_B distribution is then calculated from the p_B's of the 20 configurations and is coarse-grained with a bin size of 0.25. (b) the probability distribution of the constrained forces at ξ= -0.2 Å.

10 ps simulation period at this point. On the other hand, proton transfer is observed during the constrained MD simulation at ξ= -0.1 Å, after the transition state (ξ= -0.2 Å), during which the constrained forces are predominantly negative. This observation provides evidence that proton transfer occurs after the system passes the transition state in an aqueous solution (Chu & Trout, 2004).

The committor probability distribution at ξ=-0.2 Å is shown in Figure 6-3. It indeed peaks at 0.5, suggesting that the chosen reaction coordinate has not coupled to the activation of other degrees of freedom. In other words, ξ=d(O-O)-d(S-O) is the representative reaction coordinate to describe the oxidation reaction. Moreover, the transmission coefficient is calculated to be 0.7 from the shooting trajectories, indicating that recrossing due to the interaction with the solvent environment has only a small effect on the reaction barrier.

In addition to the commitor probability distribution, the constrained forces during the dynamics also contain information about the coupling of the constrained degree of freedom to the surrounding environment (Gertner *et al.* , 1987; Gertner *et al.* , 1989; Haynes *et al.* , 1994; Haynes & Voth, 1995; Nam *et al.* , 2004). The distribution of constraint forces during the 10 ps simulation is shown in Figure 6-3. It can be seen that the constraint force has a symmetric, Gaussian distribution centered at zero. Although the forces acting on the reaction coordinate at the transition state depend on the dynamics of the surrounding molecules, the distribution in Figure 6-3 suggests that the activation of variables other than ξ is not involved in the oxidation process. If the activation of a different coordinate was involved, a different type of distribution of constraint forces would likely result, for example, a distribution with a minimum centered at zero (Bolhuis *et al.* , 1998). Our analysis of Figure 6-3 leads to the idea that the distribution of constrained forces in an MD simulation can be used as a very quick way of testing whether or not degrees of freedom other than the chosen constrained coordinate are likely to play a role in the reaction. Note, however, that if the transmission coefficient is desired, trajectories from the transition state are still necessary.

As indicated in our previoous study (Chu & Trout, 2004), the charge separation developed during the breaking of the O-O bond of H_2O_2 and the formation of the S-O bond is to a large degree responsible for the reaction barrier for the oxidation of methionine by H_2O_2. For example, in the transition state shown in Figure 6-1, the distal oxygen atom, denoted as O2 in Figure 6-1, becomes more negative and the transferring oxygen, denoted as O1 Figure 6-1, becomes less negative than the oxygen atoms of H_2O_2 in the reactant state (Chu & Trout, 2004). The major role of solvent molecules is to stabilize the charge separation via specific interactions such as hydrogen bonding.

We may wish to ask then, how the number of HB's changes along the reaction coordinate. The average number of HB's between the distal oxygen atom (the O2 atom as shown in Figure 6-1) and the hydrogen atoms of surrounding water molecules are plotted along the reaction coordinate in Figure 6-4. The criterion for an HB is that the hydrogen atom of a water molecule must be within 2.0 Å of the distal oxygen atom, O2. As expected, HB increases from the reactant state to the transition state along the reaction coordinate by about 1.5 HBs. Note that the surrounding QM water molecules form only at most 2 HBs with O2 in all of simulations; an HB value higher than 2 indicates that classical water molecules are also involved in the specific interactions with O2. The MD trajectories themselves also indicate that mobile classical water molecules can interact specifically with the H_2O_2 molecule and can also exchange with the QM water molecules surrounding H_2O_2. Although ξ is the governing degree of freedom, solvent molecules still participate in the reaction and respond to the change of electronic properties of the system.

6.3.2 Oxidation of methionine residues in G-CSF

Experimental data shows that the four methionine residues of G-CSF are oxidized with different rates in the following order: Met1>Met138>Met127>Met122 (Lu *et al.* , 1999; Chu *et al.* , 2004a). The locations of the methionine residues in G-CSF are shown in Figure ??. Among the three methionine residues studied in this work, Met122 is the most buried one, being located in one of the four helices of G-CSF (see Figure ??), and has the lowest rate of oxidation. Met127 is the first residue of the helix in which Met122 resides and is thus more exposed than Met122. It is oxidized 2.4 times faster than Met122. Met138, on the other hand, is located in a loop connecting two helices and is the second most exposed residue among the four Met groups of G-CSF, being oxidized 2 times faster than Met127. We have demonstrated from classical MD simulations and from experiments that the relative rates of oxidation of methionine residues by H_2O_2 in G-CSF can be quantified accurately by two shell water coordination nnumber, 2SWCN (Chu *et al.* , 2004a). The correspondence between a conformational property that measures the degree to which solvent accesses a methionine site and the rate of oxidation of that site suggests that the major factor that determines the rates of oxidation of methionine is the extent of exposure to

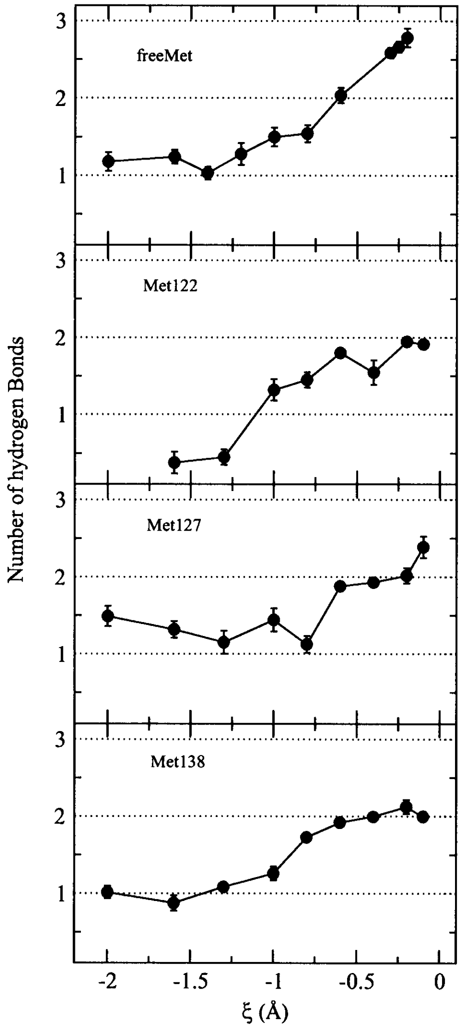


Figure 6-4: The average number of hydrogen bonds (HB's) between the O2 atom of H_2O_2 (Figure 6-1) and the hydrogen atoms of water molecules along the oxidation reaction. The x-axis is the reaction coordinate, ξ=d(O-O)-d(S-O), and the y-axis is the average number of HBs calculated from the 10 ps trajectory.

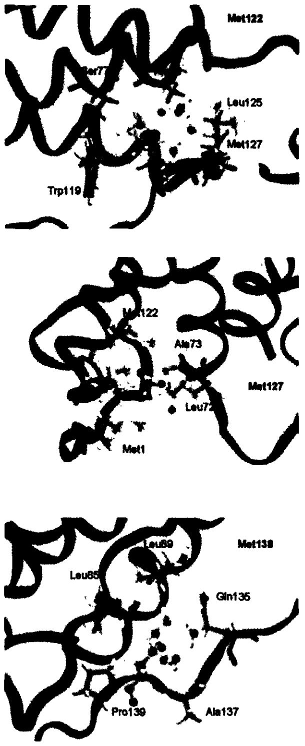


Figure 6-5: The local environment of Met122, Met127, and Met138 of G-CSF. The methionine residue, hydrogen peroxide, and QM water molecules are shown in a ball and stick representation. The residues within 6 Å of a methionine residue is shown in a licorice representation.

solvent molecules, but not the protein environment. This observation is also consistent with the oxidation mechanism that we proposed, that 2–3 water molecules are needed to stabilize the transition state of the oxidation reaction (Chu *et al.* , 2004a). This hypothesis is now examined via computing the free energy barriers of Met122, Met127, and Met138 in G-CSF.

Details of the local environment of Met122, Met127, Met138 are shown in Figure 6-5. Met122 is located in a hydrophobic core composed of Ser77, Ala78, Trp119, Leu125, and Met127. Met127 is at the beginning of a 20 residue helix and is surrounded by Met1, Leu72, Ala73, and Met122. Met138 is surrounded by Leu85, Lue89, and Gln135 and the loop region in which it resides has high flexibility (Chu *et al.* , 2004a). Generally speaking, the local environments of methionine residues in G-CSF

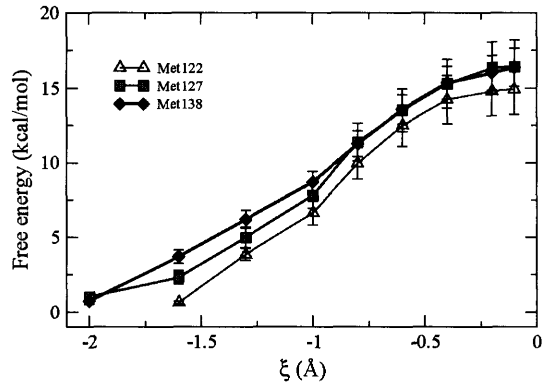


Figure 6-6: The free energy profiles of the oxidation of methionine residues in G-CSF. The x-axis is the reaction coordinate, ξ=(d(O-O)-d(S-O)), in Å, and the y-axis is free energy in kcal/mol, obtained by integrating the mean force along the reaction coordinate from the last 5 ps of the constrained MD simulations.

Table 6.1: Activation free energies (kcal/mol) of free methionine in water and methionine in G-CSF

Methionine	$\Delta F^{\ddagger}$(kcal/mol)
freeMet	14.70±1.22
Met122	14.93±1.70
Met127	16.45±1.75
Met138	16.38±1.25

are hydrophobic and provide various degrees of spatial restriction. Concepts of spatial restriction have been used to address the enzymatic activity of proteins (Garcia-Viloca *et al.* , 2004). To our knowledge, our study is the first attempt to examine the effects of spatial restriction of a protein molecule on the barriers of chemical reactions of the protein itself.

The free energy profiles of the oxidation of different methionine residues in G-CSF obtained by constrained MD simulations are shown in Figure 6-6, and the free energy barriers are listed in Table 6.1. The activation free energies of the three methionine residues are equivalent within statistical uncertainties, indicating that the reaction barriers are indeed not sensitive to the different environments around different methionine sites, once the reactants are near each other.

The HB's between the distal oxygen and the hydrogen atoms of water molecules along the reaction coordinate are shown in Figure 6-4 for the different methionine residues. For Met122, the distal oxygen of H_2O_2 has almost no HB's with water molecules in the reactant state, an indication of the restricted access to its environment. As the reaction progresses, the number of HB's on the distal oxygen increases, reaching a value of 1.8 at the transition state. Both QM and MM water molecules are involved in the hydrogen bonding with H_2O_2.

For Met127 and Met138, the access to solvent molecules is not restricted as much as for Met122, and therefore the number of HBs on the distal oxygen of H_2O_2 in the reactant state of each of these is higher than that for Met122. Along the reaction coordinate, the number of HBs on the distal oxygen also increases, reaching a value of about 2.0 at the transition state, slightly less than that of the freeMet case, but enough such that the reaction barriers are stabilized to the reaction barrier of freeMet.

The above results show that as long as 2–3 water molecules are present around buried methionine sites, the reaction barrier is similar to that of free methionine. The solvent accesses spatially restricted residues via thermo fluctuations, explaining the excellent correlation between 2SWCNs and oxidation rate constants (Chu *et al.* , 2004a). If different environments of methionine sites in the protein molecule affected the reaction barrier in a specific manner, such a correlation would not occur. The key element of the oxidation mechanism is that the ensuing charge separation of the transition state complex during oxidation is stabilized by 2–3 water molecules. Our results do not rule out the possibility that polar atoms of protein molecules could play the role of stabilization via specific interactions if they are close to the sulfur atom in methionine. For G-CSF, however, this arrangement was not observed, and in fact, such a possibility would seem to be rare, at best, because of the tendency for hydrophobic residues to be present in protein cores.

Summary and Conclusions

In this work, QM/MM and constrained MD simulations are employed to study the oxidation of the amino acid methionine by H_2O_2 in an aqueous environment and in G-CSF. The computed reaction free energy of free methionine is very close to the value that we measured. The reaction coordinate, governing degree of freedom, was found to be ξ=d(O-O)-d(S-O). That this is truly the reaction coordinate was verified by computing the commitor probability distribution and the distribution of constrained forces at the maximum of the free energy profile, as shown in Figure 6-3.

This study is the first to explore the effects of both solvent and protein environment on the oxidation reaction of the amino acid methionine using detailed molecular theoretical methods. We found that the reaction barriers of oxidation do not depend on different protein environments surrounding each methionine site in G-CSF. This is consistent with the observation in an earlier study that the measured rates of oxidation correlate with the two shell water coordination number of different methionine sites (Chu *et al.* , 2004a). The hydrogen bonds between the distal oxygen of H_2O_2 and the hydrogen atoms of water molecules are shown to increase along the reaction coordinate, indicating that the charge separation developed during the S_N2 oxidation of methionine by H_2O_2 is stabilized by specific interactions, such as hydrogen bonding, with water molecules, even for the most spatially restricted residue of G-CSF, Met122.

The results of this study lead to an enhanced mechanistic understanding of the oxidation of proteins and should be useful in designing ways to control protein oxidation by peroxides. They also add to our knowledge of how complex environments affect chemical reactions.

Chapter 7

A Comprehensive Picture of "Non-Site Specific" Oxidation of Methionine Residues by Peroxides in Protein Pharmaceuticals

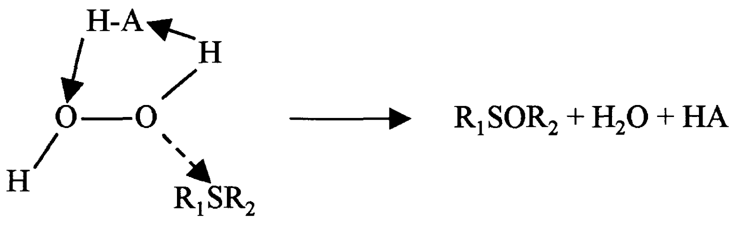
In this chapter, a comprehensive picture of the oxidation of protein pharmaceuticals by peroxides is developed based on the studies that are reported in earlier chapters. We propose a new mechanism, the water-mediated mechanism, for the oxidation of methionine residues, and it has been shown (Chapter 3) to satisfy all available experimental data. Based on the water-mediated mechanism, we found a structural property, average 2-shell water coordination number, that correlates well to the relative rates of oxidation of methionine groups. This was used to study the oxidation of Granulocyte Colony-Stimulating Factor (G-CSF) in Chapter 4 and 1–34 human parathyroid hormone hPTH(1–34) in Chapter 5. We believe that this comprehensive picture should aid researchers in the pharmaceutical sciences to develop solvent formulations for therapeutic proteins in a more rational way.

7.1 Background

Control of oxidation is a critical issue for protein pharmaceuticals. Formulating proteins in aqueous solutions is preferred, and significant effort is expended to develop formulations that will stabilize protein pharmaceuticals against oxidation (Cleland *et al.* , 1993; Cleland & Langer, 1994; Shahrokh, 1997; Wang, 1999). An understanding of the mechanisms of oxidation processes, in addition to the reasons for which different sites in each protein are susceptible to oxidation to different degrees would aid in the development of protein formulations. Each of the earlier chapters has addressed different pieces of the mechanistic puzzle. This chapter puts together the various pieces into a comprehensive picture of non-site specific oxidation.

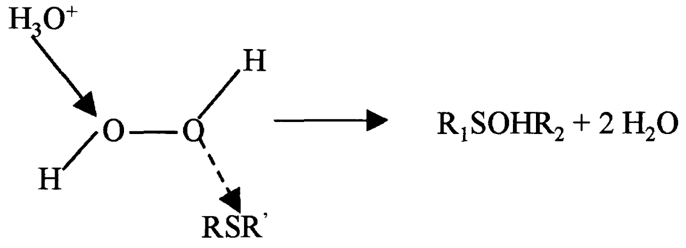
Non-site-specific reactions refer to oxidation enhanced by oxidants such as peroxides or upon exposure to light (Matheson *et al.* , 1979; Vogt, 1995; Levine *et al.* , 1996; Yao *et al.* , 1996; Gao *et al.* , 1998; Patten *et al.* , 1999; Yin *et al.* , 2000), and this thisis focuses on peroxides, for which methionine is the most liable residue. Peroxides can be introduced into a protein formulation through many routes (Cleland *et al.* , 1993; Li *et al.* , 1995; Wang, 1999; Meyer *et al.* , 2002), and forced oxidation via peroxides is the most common procedure used by formulation scientists in order to understand the oxidation of methionine in proteins. (Matheson *et al.* , 1979; Vogt, 1995; Levine *et al.* , 1996; Yao *et al.* , 1996; Gao *et al.* , 1998; Patten *et al.* , 1999; Yin *et al.* , 2000; Griffiths & Cooney, 2002).

The oxidation of methionine by peroxides has historically been hypothesized as proceeding via an acid-catalyzed mechanism (Dankleff *et al.* , 1968; Curci *et al.* , 1970; Bach *et al.* , 1991; Nguyan *et al.* , 1993; Nguyan, 1994; Bach *et al.* , 1994; Li *et al.* , 1995; Okajima, 2001; Ottolina & Carrea, 2001):



(7.1)

where HA is a general acid, for example, a water molecule. HA was thought to serve as an intermediate agent to facilitate the hypothesized rate-limiting step of oxidation, *transfer of proton* from the transferring oxygen atom to the distal oxygen atom, as shown in Eq.(7.1). A further modification of this mechanism was based on *ab initio* calculations that show that the barrier to Eq. 7.1 is too high (Bach *et al.* , 1991; Bach *et al.* , 1994; Okajima, 2001; Ottolina & Carrea, 2001):



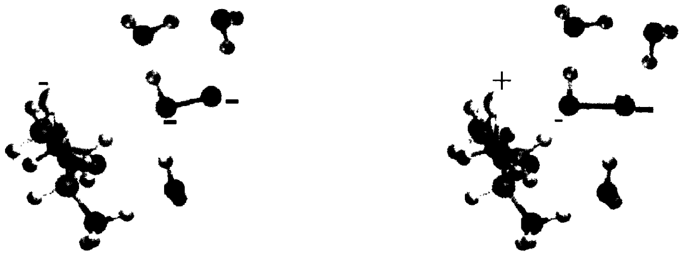
(7.2)

Since these proposals were made, however, experimental studies determined that the oxidation rates of organic sulfides, such as dimethyl sulfide, do not exhibit a significant pH dependence between pH=2–10 as expected. (Note that the effect of pH only becomes significant at pH=1 (Amels *et al.* , 1997), and thus, the mechanism in Eq. 7.2 is valid, at best, only at a very low pH value.) Protein formulations are designed to be at or near the pH at which the protein has maximum stability, usually between pH=4.5–7.5 (Powell, 1996). Therefore, the question remains: what is the mechanism of oxidation at intermediate values of pH?

Linked to this general question is another unknown: why are different methionine residues in a protein oxidized at different rates? As a simple example, methionine residues on the surface of a protein are oxidized at higher rates than the buried ones (Nguyan *et al.* , 1993; Powell, 1996; Nabichi *et al.* , 1998; Lu *et al.* , 1999; Griffiths & Cooney, 2002; Meyer *et al.* , 2002). A obvious explanation is that the rate of oxidation is governed by the accessibility of the sites to oxidants. An typical measure of this is solvent accessible area (Lee & Richards, 1971) calculated from X-ray structures. This measure, however, is inconsistent with basic oxidation data. For example, the solvent accessible areas of partially buried methionine residues, Met122, Met127, and Met138 of Granulocyte Colony-Stimulating Factor (G-CSF) are zero, but these residues are still vulnerable to oxidation (Chu *et al.* , 2004a). Solvent accessible areas also fail to explain the oxidation rates of different methionine residues for α-1 antitrypsin (Griffiths & Cooney, 2002). Therefore, in order to understand oxidation, a structural property that can be linked better to the susceptibility of different sites to oxidation needs to be identified (Meyer *et al.* , 2002).

7.2 Mechanistic understanding of oxidation

Because of the inconsistencies of the acid-catalyzed mechanism (Eq.(7.1) and Eq.(7.2)), the non-site specific oxidation was investigated theoretically in Chapter 3. It was found that what governs the barrier to the oxidation reaction is not the transfer of protons but the charge separation at the transition state developed during the breaking of the O-O bond of H_2O_2 and the formation of S-O bond. Figure 7-1 clearly shows that proton transfer does not play a role in the process up to the transition state, as the acid-catalyzed mechanism would require (Dankleff *et al.* , 1968; Curci *et al.* , 1970; Bach *et al.* , 1991; Nguyan *et al.* , 1993; Nguyan, 1994; Bach *et al.* , 1994; Li *et al.* , 1995; Okajima, 2001; Ottolina & Carrea, 2001). It was also found in Chpater 3 that



Reactant Transition State

Figure 7-1: The reactant and transition state configurations during the oxidation of methionine by H_2O_2. The charge separation at the transition state is shown as "+" and "-". "+" denotes a positively charged atom and "-" for a negative atomic charge. The size of the symbol indicates the relative magnitude of the charge.

2–3 water molecules stabilize the charge separation at the transition state by specific interactions, such as hydrogen bonding (Figure 7-1). Figure 7-2 shows the computed activation energies as a function of the number of water molecules that stabilize the transition state during oxidation. From these data, we conclude that water molecules can adequately stabilize the transition state and that 2–3 water molecules are sufficient. We call the oxidation mechanism based on Figure 7-1 and Figure 7-2 the "water-mediated" mechanism. It is consistent with all available experimental data such as the activation energies of oxidation and the pH dependence of the rates of oxidation (Dankleff *et al.* , 1968; Amels *et al.* , 1997).

The water-mediated mechanism implies that the rate of oxidation of fully exposed sites should not be a strong function of pH. The rates of oxidation of the amino acid methionine (free methionine) in an aqueous solution at different pH values shown in Figure 5-1 indicate that the rates are indeed nearly pH independent between pH=2–8; an increase in the rate of oxidation is seen only at a very low pH of 1. These data provides additional verification of the water-mediated mechanism for pH=2–8.

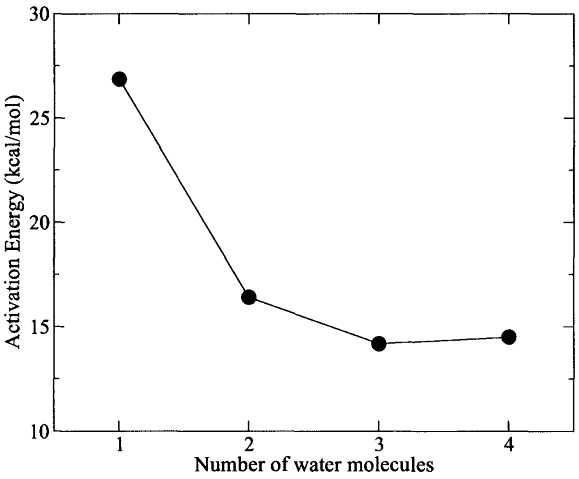


Figure 7-2: The activation energies of oxidation of dimethyl sulfide by H_2O_2 as a function of the number of surrounding water molecules.

7.3 Structural properties and the rates of oxidation

Since solvent accessible areas based on static X-ray structures do not correlate well to oxidation, an alternative property that better describes the susceptibility of different sites to oxidation needs to be found. Based on the water-mediated mechanism, water molecules play an essential role in the oxidation of methionine in a specific manner, and thus they need to be taken into account explicitly. Therefore, we hypothesize that the average number of water molecules within 2 solvation shells govern the rates of oxidation. Despite what the structural data from X-ray studies imply, even buried residues should have a non-zero average number of water molecules because of thermal fluctuations manifest in the dynamic motion of the water molecules and the protein. Since it is difficult to directly follow the dynamics of proteins experimentally, we used molecular dynamics (MD) simulations to sample the protein and solvent structures and compute the average of 2-shell water coordination numbers, 2SWCN's. The explicit definition of 2SWCN is shown in Figure 7-3. It is the number of water molecules within a 5.5 Å (2 solvation shells) cutoff distance from the sulfur atom.

We have investigated the solvation dynamics of two proteins, G-CSF in Chapter 4

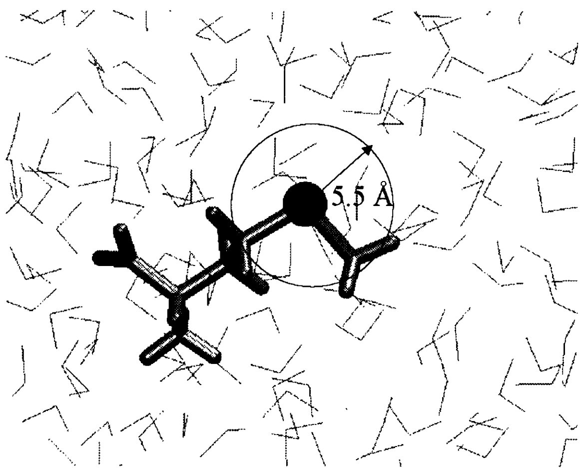


Figure 7-3: A representation of the two-shell water coordination number of the sulfur atom of an methionine amino acid. Grey lines are water molecules.

and hPTH(1–34) in Chapter 5, and we found that the average 2SWCN correlates well to the rates of oxidation. The correlation of the average 2SWCN to the rates of oxidation of methionine residues is shown in Figure 4-11 for G-CSF and Table 5.2 for hPTH(1–34). Note that the four methionine residues of G-CSF have very different rates of oxidation and each has a different pH dependence, but the 2SWCN's can still capture well the difference among partially buried methionine residues and to a lesser degree the pH dependence of rates of oxidation. For hPTH(1–34), the relative rates of oxidation between Met8 and Met18 can also be quantified accurately using the average 2SWCN's with proper sampling of the peptide conformation.

The above results indicate that the degrees of access of water molecules governs the oxidation reaction, consistent with our proposed water-mediated mechanism. It also indicates that the protein environment surrounding a methionine site affects oxidation via spatial restriction to the access to water but does not affect the reaction in a specific manner. We confirmed this mechanism for G-CSF via computing the activation free energy of the partially buried methionine residues in G-CSF and free methionine using combined quantum mechanical and molecular mechanical calculations and constrained molecular dynamics simulations in Chapter 6. The results reported in Table 6.1 indicate that reaction barriers for Met122, Met127, and Met138 of G-CSF are indeed equivalent (~ 15 kcal/mol) to the activation free energy of free methionine when water molecules are present. Our results do not rule out the possibility that polar atoms of protein molecules could also help to stabilize the transition state via specific interactions, if they are close to the sulfur atom in methionine. For G-CSF, however, stabilization of the transition state by polar residues was not observed in the simulations, and in fact, such interactions would be unlikely because of the tendency for hydrophobic residues to be present in protein cores. To our knowledge, the above analyzes are the first attempt to investigate the relation between protein conformation and methionine oxidation from first principles. They provide additional support to the water-mediated mechanism, in addition to a means to predict which residues of a given protein might be most susceptible to oxidation.

7.4 A comprehensive picture of non-site specific oxidation and the effects of excipients

A summary of our comprehensive picture of non-site specific oxidation is:

1. For an intermediate range of pH, O-O bond breaking and S-O bond formation are the governing degrees of freedom for the oxidation of methionine sites by hydrogen peroxide, not proton transfer.

2. As these bonds break or form, there ensues a partial charge separation, which is destabilizing. Specific interactions of water molecules with the partial charges of the transition state help to stabilize it and thus lower the barrier to oxidation.

3. Because water molecules lower the barrier to oxidation, the degree to which they can access methionine governs the rate of oxidation. This can be quantified by the 2 shell water coordination number, which is the average number of water molecules within 5.5 Å.

4. Principles 1–3 above describe the "water mediated" mechanism, which should be applicable for oxidation between roughly pH 2–8. At lower values of the

pH, we hypothesize that the "acid-catalyzed" mechanism starts to contribute to oxidation.

5. For buried methionine groups, in the example of G-CSF, the protein environment does not affect the barrier to oxidation. We anticipate that only the presence of polar groups very close to the sulfur in methionine will affect this barrier and that such a configuration is unlikely to occur in proteins.
6. For methionine groups on the surface of proteins, acid groups that are close to the methionine could increase the value of the pH at which the acid-catalyzed mechanism starts to become important, as in the case of hPTH(1–34).

We believe that this comprehensive picture that we have developed could aid pharmaceutical sciences in terms of developing in a more rational manner solvent formulations that protect therapeutic proteins against oxidation, in addition to being able to predict which sites in a protein might be most susceptible to oxidation. Currently, there are very few publications on the mechanistic basis for effects of excipients on the oxidation of methionine in proteins (DePaz *et al.* , 2000; Yin *et al.* , 2004b; Yin *et al.* , 2004a). This new water-mediated mechanism of oxidation not only resolves the inconsistencies with the traditional acid-catalyzed mechanism but also provides the tools for rational formulation design.

Chapter 8

Recommended Future Work

8.1 Effects of polar residues on oxidation

When polar amino acids, such as Glu, Asp, Gln, Asn, His, etc., are present around a methionine residue, they can participate in stabilizing the charge separation of the oxidation transition state caused by peroxides. When this kind of configuration occurs, this particular methionine site becomes a hot spot for oxidation, and needs special treatment. In order to be able to identify oxidation hot spots according to protein sequence or tertiary structure, the mechanisms of how polar residues affect oxidation need to be quantified. For example, questions that could be addressed are: how close does a polar residue need to be to affect oxidation, how much can a polar residue change the rate of oxidation and why, and what kind of structures could facilitate such effects? If these questions can be answered systematically, the answers could be useful as a basis for developing the second generation of products via improving the stability of a protein pharmaceutical on the market. The hierarchy of procedures in tackling the problems of oxidation developed in this thesis can be employed straightforwardly for this purpose.

8.2 Quantify the effects of excipient molecules on protein conformation and on oxidation

Based on the structure/oxidation relationship using average 2-shell water coordination numbers to correlate the rates of oxidation, one possible strategy of controlling oxidation is to modulate protein conformation via adding excipient molecules. It can be expected that the design aspects would depend on the properties of the individual protein molecule and how it interacts with excipient. Therefore, a thorough knowledge of protein dynamics and excipient effects is necessary to employ such a strategy in a systematic manner. For example, questions such as how can the conformation of a protein molecule be affected by excipients and to what extent needs to be understood. These questions are nontrivial challenges for both theory and experiment, and using molecular simulations could bridge the gap. Particularly, simulation techniques, such as self-guided molecular dynamics and replica exchange MD, that enable efficient sampling of protein conformation can be very useful. These, combined with experimental studies and validations, can be helpful to improve the current status of stabilization against oxidation.

8.3 Investigate other types of oxidative instabilities

There are several other oxidation pathways by which proteins can be degraded that need to be controlled in a formulation. Typical examples include photo-catalyzed oxidation by singlet oxygen, oxidation caused by the superoxide radical, and metal-catalyzed oxidation caused by the hydroxyl radical. The mechanism of each pathway and the relevant amino acid sidechains involved need to be identified. As demonstrated in this thesis, the existing pictures of oxidation can be misleading, and more detailed analyses are necessary, especially when the current mechanisms are mostly conceptual. Furthermore, the relationships between oxidation chemistry and protein structure are still not clear, and it is difficult to measure directly such essential information. *Ab initio* methods, based on molecular orbitals or density functional theories together with the reaction pathway method that is developed in this thesis and free energy simulations, are suitable methodologies to elucidate and quantify these mechanisms and the effects of protein conformations. New stabilization strategies could also be invented given a more detailed understanding of these processes.

8.4 Integrated design of formulation

Although the focus is on oxidation, protein molecules in a formulation can be affected by many degradation pathways simultaneously. Major degradation pathways other than oxidation include unfolding, aggregation, and deamidation. It is indeed crucial to have a mechanistic understanding of each pathway in order to design protein formulations with a rationale, but it is also necessary to incorporate strategies against different degradation pathways all together in a systematic manner. As a result, in addition to the understanding of each degradation process, the interplay among these physical and chemical stresses also needs to be elucidated. For instance, adding preferential exclusion excipients has been used as a strategy to prevent protein unfolding, and adding surfactants is often used to prevent aggregation. These additives can also affect oxidation in a specific manner by modulating the protein conformation, as pointed out in Chapter 7, and it would be ideal to have a design procedure to incorporate these effects in an automatic and self-consistent manner, i.e., when optimizing the stability of one degradation pathway such as protein aggregation, the effects on oxidation should also be controlled. Another example is the aggregation induced by forming intermolecular disulfide bonds. The oxidation of the amino acid cysteine may cause aggregation as well, and this has been found to be a serious problem for G-CSF. There also exists evidence that chemical modifications of a protein molecule are related to protein unfolding and aggregation. Therefore, the connection of different degradation pathways and protein-protein interactions are also important issues that need to be understood in order to maintain protein stability. Molecular simulations can be used to study different aspects of these problems in great detail and thus can be very helpful in order to achieve an integrated approach to the design of formulations.

References

ALLEN, M. P., & TILDESLEY, D. J. 1987. *Computer Simulation of Liquids.* New York: Oxford.

AMELS, RTER, ELIAS, HORST, & WANNOWIUS, KLAUS-JURGEN. 1997. Kinetics and mechanism of the oxidation of dimethyl sulfide by hydroperoxides in aqueous medium Study on the potential contribution of liquid-phase oxidation of dimethyl sulfide in the atmosphere. *Journal of the Chemical Society. Faraday transactions*, **93**(15), 2537–2544.

ANDRICIOAEI, I., DINNER, A. R., & KARPLUS, M. 2003. Self-guided enhanced sampling methods for thermodynamic averages. *Journal of Chemical Physics*, **118**(3), 1074–1084.

ARITOMI, M., KUNISHIMA, N., OKAMOTO, T., KUROKI, R., OTA, Y., & MORIKAWA, K. 1999. Atomic Structure of the Gcsf-Receptor Complex Showing a New Cytokine-Receptor Recognition Scheme. *Nature*, **401**, 713.

ARITOMI, M., KUNISHIMA, N., OKITSU, N., SHIMIZU, M., OTA, Y., & MORIKAWA, K. 2000. Purification, Crystalization and Preliminary X-ray analysis of a complex between granulocyte colony-stimulating factor and its soluble recptor. *Acta crystallographica*, **D56**, 751–753.

BACH, R. D., SU, M. D., & SCHLEGEL, H. B. 1994. Oxidation of Amines and Sulfides with Hydrogen Peroxide and Akyl Hydrogen Peroxide. The Nature of Oxygen-Transfer Step. *Journal of the American Chemical Society*, **116**, 5379–5391.

BACH, ROBERT D., OWENSBY, AMY L., GONZALEZ, CARLOS, SCHLEGEL, H. BERNDARD, & MCDOUALL, JOSEPH J. W. 1991. Nature of the transition structure for oxygen atom transfer from a hydroperoxide. Theoretical comparison between water oxide and ammonia oxide. *Journal of the American Chemical Society*, **113**, 6001–6011.

BARDEN, J. A., & CUTHBERTSON, R. M. 1993. Stabilized Nmr Structure of Human Parathyroid Hormone(1–34). *European Journal of Biochemistry*, **215**(2), 315–321.

BARDEN, J. A., & KEMP, B. E. 1993. Nmr Solution Structure of Human Parathyroid Hormone(1-34). *Biochemistry*, **32**(28), 7126–7132.

BARONE, V., COSSI, M., & TOMASI, J. 1997. A New difinition of Cavities for the Computation of Solvation Free Energies by the Polarizable Continuum Model. *Journal of Chemical Physics*, **107**(8), 3210–3221.

BAZAN, FERNANDO J. 1990. Haemopoietic receptors and helical cytokines. *Immunology Today*, **11**, 350–354.

BERKE, A. D. 1993. Density-functional thermochemistry III, the role of exact exchange. *Journal of Chemical Physics*, **98**, 5678.

BERMS, D. N. 2002. The kinetics of G-CSF Folding. *Protein Science*, **11**, 2504–2511.

BOLHUIS, P. G., DELLAGO, C., & CHANDLER, D. 1998. Sampling ensembles of deterministic transition pathways. *Faraday Discussions*, 421–436.

BOLHUIS, P. G., DELLAGO, C., & CHANDLER, D. 2000. Reaction coordinates of biomolecular isomerization. *Proceedings of the National Academy of Sciences of the United States of America*, **97**(11), 5877–5882.

BOLHUIS, P. G., CHANDLER, D., DELLAGO, C., & GEISSLER, P. L. 2002. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual Review of Physical Chemistry*, **53**, 291–318.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. 1983. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry*, 4(2), 187–217.

Bryant, Z., Pande, V. S., & Rokhsar, D. S. 2000. Mechanical unfolding of a beta-hairpin using molecular dynamics. *Biophysical Journal*, **78**(2), 584–589.

Carter, E. A., Ciccotti, G., Hynes, J. T., & Kapral, R. 1989. Constrained Reaction Coordinate Dynamics for the Simulation of Rare Events. *Chemical Physics Letters*, **156**(5), 472–477.

Caves, L. S. D., Evanseck, J. D., & Karplus, M. 1998. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science*, **7**(3), 649–666.

Chandler, D. 1978. Statistical-Mechanics of Isomerization Dynamics in Liquids and Transition-State Approximation. *Journal of Chemical Physics*, **68**(6), 2959–2970.

Chen, Y, & Barkley, M. D. 1998. Toward understanding tryptophan fluorescence in proteins. *Biochemistry*, **37**(28), 9976–9982.

Choi, Chyung, & Elber, Ron. 1991. Reaction path study of helix formation in tetrapeptides: Effect of side chains. *Journal of Chemical Physics*, **94**, 751–760.

Chu, J. W., & Trout, B. L. 2004. On the Mechanisms of Oxidation of Organic Sulfides by $H_2O_2$2 in Aqueous Solutions. *Journal of the American Chemical Society*, **126**(3), 900–908.

Chu, J. W., Trout, B. L., & Brooks, B. R. 2003. A super-linear minimization scheme for the nudged elastic band method. *Journal of Chemical Physics*, **119**(24), 12708–12717.

CHU, J. W., YIN, J., WANG, D. I. C., & TROUT, B. L. 2004a. Molecular dynamics simulations and oidation rates of methionine residues of Granulocyte Colony-Stimulating Factor at different pH values. *Biochemistry*, **43**(4), 1019–1029.

CHU, JHIH-WEI, YIN, JIN, WANG, DANIEL I. C., & TROUT, BERHNARDT L. 2004b. Oxidation of Methionine Residues and MD simulations of hPTH(1–34). *Biochemistry, submitted.*

CIOBICA, I. M., & VAN SANTEN, R. A. 2002. A DFT study of CHx chemisorption and transition states for C-H activation on the Ru(1120) surface. *Journal of Physical Chemistry B*, **106**(24), 6200–6205.

CIOBICA, I. M., FRECHARD, F., VAN SANTEN, R. A., KLEYN, A. W., & HAFNER, J. 2000. A DFT study of transition states for C-H activation on the Ru(0001) surface. *Journal of Physical Chemistry B*, **104**(14), 3364–3369.

CLELAND, JEFFERY L., & LANGER, ROBERT. 1994. Formulation and Delivery of Proteins and Peptides. *ACS Symposium Series*, 1–19.

CLELAND, JEFFERY L., POWELL, MICHAEL F., & SHIRE, J. 1993. The Development of Stable Protein Formulations: A Close Look at Protein Aggregation, Deamidation, and Oxidation. *Critical Reviews in Therapeutic Drug Carrier Systems*, **10**(4), 307–377.

CURCI, R., DIPRETE, R. A., EDWARDS, J. O., & MONDENA, G. J. 1970. Role of solvent in the oxidation of some organic compounds by peroxyacids. *J. Org. Chem.*, **35**, 740.

CZERMINSKI, RYSZARD, & ELBER, RON. 1990. Reaction path study of conformational transitions in flexible systems: Applications to peptides. *Journal of Chemical Physics*, **92**, 5580–5601.

DANKLEFF, M. A. P., RUGGERO, C., EDWARDS, J. O., & PYUN, H. Y. 1968. The influence of solvent on the oxidation of thioxane by hydrogen peroxide and by tert-butyl hydroperoxide. *Journal of the American Chemical Society*, **90**, 3209.

DARDEN, T., YORK, DARRIN, & PEDERSEN, L. 1993. Particle Mesh Ewald: An N*LOG(N) method for Ewals Sums in Large Systems. *Journal of Chemcial Physics*, **93**(12), 10089–10092.

DAS, D., EURENIUS, K. P., BILLINGS, E. M., SHERWOOD, P., CHATFIELD, D. C., HODOSCEK, M., & BROOKS, B. R. 2002. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *Journal of Chemical Physics*, **117**(23), 10534–10547.

DEPAZ, R. A., BARNETT, C. C., DALE, D. A., CARPENTER, J. F., GAERTNER, A. L., & RANDOLPH, T. W. 2000. The excluding effects of sucrose on a protein chemical degradation pathway: Methionine oxidation in subtilisin. *Archives of Biochemistry and Biophysics*, **384**(1), 123–132.

DU, R., PANDE, V. S., GROSBERG, A. Y., TANAKA, T., & SHAKHNOVICH, E. S. 1998. On the transition coordinate for protein folding. *Journal of Chemical Physics*, **108**(1), 334–350.

ECKARDSTEIN, A., WALTER, M., HOLZ, H., BENNINGHOVEN, A., & ASSMANN, G. 1991. Site-Specific Methionine Sulfoxide Formation is the Structural Basis of Chromatographic Heterogeneity of Apolipoproteins A-I, C-II, and C-III. *Journal of Lipid Research*, **32**, 1465–1476.

EDWARDS, J. O. 1960. Nucleophilic Displacement on Oxygen in Peroxides. *Pages 67–106 of:* EDWARDS, J. O. (ed), *Peroxide Reaction Mechanisms*. NewYork: Interscience Publishers, Inc.

EICHLER, A., & HAFNER, J. 1999. Reaction channels for the catalytic oxidation of CO on Pt(111). *Surface Science*, **435**, 58–62.

ELBER, R., & KARPLUS, M. 1987. A Method for Determining Reaction Paths in Large Molecules: Application to Myoglobin. *Chemical Physics Letters*, **139**(5), 375–380.

ESSMANN, U., PERERS, L., BERKOWITZ, M. L., DARDEN, T., LEE, H., & PEDERSEN, L. G. 1995a. A smooth particle mesh Ewald method. *Journal of Chemcial Physics*, **103**(19), 8577–8593.

ESSMANN, U., PERERS, L., BERKOWITZ, M. L., DARDEN, T., LEE, H., & PEDERSEN, L. G. 1995b. A smooth particle mesh Ewald method. *Journal of Chemical Physics*, **103**(19), 8577–8593.

EVANS, M. G., & POLANYI, M. 1935. Some Applications of the Transition State Method to the Calculation of Reaction Velocities, Especially in Solution. *Transactions of the Faraday Society*, **31**, 875.

EYRING, H. 1935. The Activation Complex in Chemical Reactions. *Journal of Chemical Physics*, **3**, 107.

FARKAS, O., & SCHLEGEL, H. B. 1999. Methods for optimizing large molecules. II. Quadratic search. *Journal of Chemical Physics*, **111**(24), 10806–10814.

FARKAS, O., & SCHLEGEL, H. B. 2002. Methods for optimizing large molecules - Part III. An improved algorithm for geometry optimization using direct inversion in the iterative subspace (GDIIS). *Physical Chemistry Chemical Physics*, **4**(1), 11–15.

FAST, P. L., CORCHADO, J., SANCHEZ, M. L., & TRUHLAR, D. G. 1999. Optimized Parameters for Scaling Correlation Energy. *Journal of Physical Chemistry A*, **103**(17), 3139–3143.

FINK, A. L., CALCIANO, L. J., GOTO, Y., NISHIMURA, M., & SWEDBERG, S. A. 1993. Characterization of the Stable, Acid-induced, Molten globule-like State of Staphylococcal Nuclease. *Protein Science*, **2**(7), 1155–1160.

FINK, A. L., CALCIANO, L. J., GOTO, Y., KUROTSU, T., & PALLEROS, D.R. 1994. Classification of Acid Denaturation of Proteins - Intermediates and Unfolded States. *Biochemistry*, **33**(41), 12504–12511.

FLETCHER, R. 1981. *Practical Methods of Optimization.* Chichester: Wiley.

FORDELL, T., SALO, P., & ALATALO, M. 2002. Self-diffusion on fcc (100) metal surfaces: Comparison of different approximations. *Physical Review B*, **65**(23), art. no. 233408.

FRENKEL, D., & SMIT, B. 1996. *Understanding Molecular Simulation.* From Algorithms to Applications. New York: Academic Press.

FRISCH, M. J., HEAD-GORDON, M., & POPLE, J. A. 1990a. A direct MP2 gradient method. *Chemical Physics Letters*, **166**, 275.

FRISCH, M. J., HEAD-GORDON, M., & POPLE, J. A. 1990b. Semi-direct algorithms for the MP2 energy and gradient. *Chemical Physics Letters*, **166**, 281.

FRISCH, M. J., TRUCKS, G. W., SCHLEGEL, H. B., SCRUSERIA, G. E., ROBB, M. A., CHEESEMAN, J. R., ZAKRZEWSKI, V. G., MONTGOMERY, J. A., STRATMANN, R. E., BURANT, J. C., DAPPRICH, S., MILLAM, J. M., DANIELS, A. D., KUDIN, K. N., STRAION, M. C., FARKAS, O., TOMASI, J., BARONE, V., COSSI, M., CAMMI, R., MENNUCCI, B., POMELLI, C., ADAMO, C., CLIFFORD, S., OCHTERSKI, J., PETERSSON, G. A., AYALA, P. Y., CUI, Q., MOROKUMA, K., MALICK, D. K., RABUCK, A. D., RAGHAVACHARI, K., FORSEMAN, J. B., CIOSLOWSKI, J., ORTIZ, J. V., STEFANOV, B. B., LIU, G., LIASENKO, A., PISKORZ, P., KOMAROMI, I., GOMPERTS, R., MARTIN, R. L., FOX, D. J., KEITH, T., AL-LAHAM, M. A., PENG, C. Y., NANAYAKKARA, A., GONZALEZ, C., CHALLACOMBE, M., GILL, P. M. W., JOHNSON, B. G., CHEN, W., WONG, M. W., ANDRES, J. L., HEAD-GORDON, M., REPLOGLE, E. S., & POPLE, J. A. 1998. *Gaussian 98 (Revision A.1).*

GAO, J., YIN, D. H., YAO, Y., SUN, H., QIN, Z., SCHÖNEICH, C., WILLIAMS, T. D., & SQUIER, T. C. 1998. Loss of conformational Stability in Calmodulin upon Methionine Oxidation. *Biophysical Journal*, **74**, 1115–1134.

GAO, J. L., & TRUHLAR, D. G. 2002. Quantum mechanical methods for enzyme kinetics. *Annual Review of Physical Chemistry*, **53**, 467–505.

GAO, J. L., GARCIA-VILOCA, M., POULSEN, T. D., & MO, Y. R. 2003. Solvent effects, reaction coordinates, and reorganization energies on nucleophilic substitution reactions in aqueous solution. *Pages 161–181 of: Advances in Physical Organic Chemistry, Vol 38*. Advances in Physical Organic Chemistry, vol. 38. London: Academic Press.

GARCIA-VILOCA, M., GAO, J., KARPLUS, M., & TRUHLAR, D. G. 2004. How enzymes work: Analysis by modern rate theory and computer simulations. *Science*, **303**(5655), 186–195.

GARLAND, D., ZIGLER, J. S., & KINOSHITA, J. 1986. Structural Changes in Bovine Lens Crystallins Induced by Ascorbate, Metal, and Oxygen. *Archives of Biochemistry and Biophysics*, **251**(2), 771–776.

GEISSLER, P. L., DELLAGO, C., CHANDLER, D., HUTTER, J., & PARRINELLO, M. 2001. Autoionization in liquid water. *Science*, **291**(5511), 2121–2124.

GERTNER, B. J., BERGSMA, J. P., WILSON, K. R., LEE, S. Y., & HYNES, J. T. 1987. Nonadiabatic Solvation Model for Sn2 Reactions in Polar-Solvents. *Journal of Chemical Physics*, **86**(3), 1377–1386.

GERTNER, B. J., WILSON, K. R., & HYNES, J. T. 1989. Nonequilibrium Solvation Effects on Reaction-Rates for Model Sn2 Reactions in Water. *Journal of Chemical Physics*, **90**(7), 3537–3558.

GONZALEZ, C., & SCHLEGEL, H. BERNDARD. 1990. Reaction Path Following in Mass-Weighted Internal Coordinates. *Journal of Physical Chemistry*, **94**, 5523.

GONZALEZ, C., & SCLEGEL, H. B. 1989. An Improved Algorithm for Reaction Path Following. *Journal of Chemical Pysics*, **90**, 2154.

GRIFFITHS, S. W., & COONEY, C. L. 2002. Relationship between protein structure and methionine oxidation in recombinant human alpha 1-antitrypsin. *Biochemistry*, **41**(20), 6245–6252.

GRUBMULLER, H. 1996. *SOLVATE*.

GUEST, M. F., VAN LENTHE, J. H., KENDRICK, J., SCHOFFEL, K., & SHERWOOD, P. 1996. *GAMESS-UK*.

HAYNES, G. R., & VOTH, G. A. 1995. Reaction-Coordinate-Dependent Friction in Classical Activated Barrier Crossing Dynamics - When It Matters and When It Doesnt. *Journal of Chemical Physics*, **103**(23), 10176–10182.

HAYNES, G. R., VOTH, G. A., & POLLAK, E. 1994. A Theory for the Activated Barrier Crossing Rate-Constant in Systems Influenced by Space and Time-Dependent Friction. *Journal of Chemical Physics*, **101**(9), 7811–7822.

HENKELMAN, G., & JÓNSSON, H. 2000. Improved tangent estimation in the nudged elastic band method for finding minimum energy paths and saddle points. *Journal of Chemical Physics*, **113**, 9978–9985.

HENKELMAN, G., UBERUAGA, B. P., & JÓNSSON, H. 2000. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics*, **113**, 9901.

HILL, C. P., OSSLUND, T. D., & EISENBERG, D. 1993. The Structure of Granulocyte-Colony-Stimulating Factor and its Relationship to Other Growth Factors. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 5167–5171.

HUMPHREY, W., DALKE, A., & SCHOLTEN, K. 1996. VMD-Visulization Molecular Dynamics. *Journal of Molecular Graphics*, **14**, 33–38.

JÓNSSON, H., MILLS, G., & JACOBSEN, K. W. 1998. Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions. *In:* BERNE, B. J., CICCOTI, G., & COKER, D. F. (eds), *Classical and Quantum Dynamics in Condensed Phase Silmulations.* Singapore: World Scientific.

JORGENSEN, WILLIAM L. L. 1998. OPLS Force Fields. *Pages 1986–1989 of:* SCHLEYER, PAUL V.R. (ed), *Encyclopedia of Computational Chemistry*, vol. 3. New York: Wiley.

KABSCH, W., & SANDER, C. 1983. Dictionary of Protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **12**, 2577–2637.

KABSCH, WOLFGANG. 1976. A solution for the best rotation to relate two sets of vectors. *Acta crystallographica. Section A, Foundations of crystallography*, **32**, 922.

KLAUS, W., DIECKMANN, T., WRAY, V., SCHOMBURG, D., WINGENDER, E., & MAYER, H. 1991. Investigation of the Solution Structure of the Human Parathyroid-Hormone Fragment (1-34) by H-1-Nmr Spectroscopy, Distance Geometry, and Molecular-Dynamics Calculations. *Biochemistry*, **30**(28), 6936–6942.

KLOSEK, M. M., MATKOWSKY, B. J., & SCHUSS, Z. 1991. The Kramers Problem in the Turnover Regime - the Role of the Stochastic Separatrix. *Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics*, **95**(3), 331–337.

KOLVENBACH, C. G., NARHI, L. O., PHILO, J. S., LI, T., ZHANG, M., & ARAKAWA, T. 1997. Granulocyte-Colony Stimulating Factor maintains a thermally Stable, Compact, Partially Folded Structure at pH 2. *The journal of peptide research*, **50**, 310–318.

LAVRICH, R. J., PLUSQUELLIC, D. F., SUENRAM, R. D., FRASER, G. T., WALKER, A. R. H., & TUBERGEN, M. J. 2003. Experimental studies of peptide bonds: Identification of the C- 7(eq) conformation of the alanine dipeptide analog N-acetyl- alanine N-methylamide from torsion-rotation interactions. *Journal of Chemical Physics*, **118**(3), 1253–1265. 634AB J CHEM PHYS.

LEE, B., & RICHARDS, F. M. 1971. The Interpretation of Protein Structures: Estimation of Static Accessibility. *Journal of Molecular Biology*, **55**, 379–400.

LEVINE, R. L. 1983. Oxidative Modification of Glutamine Synthetase. *The Journal of Biological Chemistry*, **258**(19), 11828–11833.

LEVINE, R. L., MOSONI, L., BERLETT, B. S., & STADMAN, E. R. 1996. Methionine Residues as Endogenous Antioxidants in Proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 15036–15040.

LI, SHIHONG, SCHÖNEICH, CHRISTIAN, & BORCHARDT, RONALD T. 1995. Chemical Instability of Protein Pharmaceuticals: Mechanisms of Oxidation and Strategies for Stablization. *Biotechnology and Bioengineering*, **48**, 490–500.

LIPARI, G., & SZABO, A. 1982. Model-Free Approach to the Intretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *Journal of the American Chemical Society*, **104**, 4546–4559.

LONCHARICH, R. J., BROOKS, B. R., & PASTOR, R. W. 1992. Langevin Dynamics of Peptides - the Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers*, **32**(5), 523–535.

LOOF, H.D., NILSSON, L., & RIGLER, R. 1992. Molecular Dynamics Simulation of Galanin in Aqueous and Nonaqueous Solution. *Journal of the American Chemical Society*, **114**, 4028–4035.

LU, H. S., FAUSSET, P. R., NARHO, L. O., HORAN, T., SHINAGAWA, K., SHIMAMOTO, G., & BOONE, T. C. 1999. Chemical Modification and Site-Directed Mutagenesis of Methionine Residues in Recombinant Human Granulocyte Colony-Stimulating Factor: Effect on Stability and Biological Activity. *Archives of Biochemistry and Biophysics*, **362**(1), 1–11.

LYNCH, B.J., & TRUHLAR, D.G. 2003. Robust and Affordable Multicoefficient Methods for Thermochemistry and Thermochemical Kinetics: The MCCM/3 and SAC/3. *Journal of Physical Chemistry A*, **107**, 3898–3906.

MACKERELL, A. D., BASHFORD, D., BELLOTT, M., DUNBRACK, R. L., EVANSECK, J. D., FIELD, M. J., FISCHER, S., GAO, J., GUO, H., HA, S., JOSEPH-MCCARTHY, D., NGUYEN, D. T., PRODHOM, B., REIHER, III, W. E., ROUX, B., SCHLENKRICH, M., SMITH, J. C., STOTE, R., STRAUB, J., WATANABE, M., WIORKIEWICZ-KUCZERA, J., YIN, D., & KARPLUS, M. 1998. All-Atom Emipical Potential for Molecular Modeling and Dynamics Studies of Proteins. *Journal of Physical Chemistry B*, **102**, 3586–3616.

MARAGAKIS, P., ANDREEV, S. A., BRUMER, Y., REICHMAN, D. R., & KAXIRAS, E. 2002. Adaptive nudged elastic band approach for transition state calculation. *Journal of Chemical Physics*, **117**(10), 4651–4658.

MARK, G., & CHEVION, M. 1985. Site-Specific modification of Albumin by Bree Radicals. *Biochemical Journal*, **236**, 397–400.

MARX, U. C., ADERMANN, K., BAYER, P., FORSSMANN, W. G., & ROSCH, P. 2000. Solution structures of human parathyroid hormone fragments hPTH(1-34) and hPTH(1-39) and bovine parathyroid hormone fragment bPTH(1-37). *Biochemical and Biophysical Research Communications*, **267**(1), 213–220.

MATHESON, N. R., WANG, P. S., & TRAVIS, J. 1979. Enzymatic Inactivation of Human Alpha-1 Proteinase Inhibitor. *Biochemical and Biophysical Research Communications*, **88**(2), 402–409.

MATHEWS, C. K., & VAN HOLDE, K. E. 1990. *Biochemistry*. Redwood City: The Benjamin/Cummimgs Publishing Company.

MCQUARRIE, DONALD A. 1976. *Statistical Mechanics*. New York: Harper-Collins*Publishers*.

MEYER, JEFFERY D., HO, BERT, & MANNING, MARK C. *2002. Effects of Conformation on the Chemical Stability of Pharmaceutically Relevant Polypeptides.* Pages 85–107 of: CARPENTER, F. JOHN, & MANNING, C. MARK *(eds),* Rational Design of Stable Protein Formulations. *New York: Kluwer Academic/Plenum Publishers.*

MIERTUS, S., & TOMASI, J. *1982. Approximate Evaluations of the Electrostatic Free Energy and Internal Energy Changes in Solution Processes.* Chemical Physics, **65**, *239–245.*

MIERTUS, S., SCROCCO, E., & TOMASI, J. *1981. Electrostatic Interaction of a Solute with a Continuum. A Direction Utilization of Ab Inito Molecular Potentials for the Prevision of Solvent Effects.* Chemical Physics, **55**, *117–129.*

NABICHI, Y., FUJIWARA, E., KUBONIWA, H., ASOH, Y., & USHIO, H. *1998. Kinetic Study of Methionine Oxidation in Human Parathyoid Hormone.* Analytica Chimica Acta, **365**, *301–307.*

NAM, K., PRAT-RESINA, X., GARCIA-VILOCA, M., DEVI-KESAVAN, L. S., & GAO, J. L. *2004. Dynamics of an enzymatic substitution reaction in haloalkane dehalogenase.* Journal of the American Chemical Society, **126***(5), 1369–1376.*

NARHI, L. O., KENNEY, W. C., & ARAKAWA, T. A. *1991. Conformational Changes of Recombinant Human Granulocyte Conlony-Stimulating Factor.* Journal of protein chemistry, **10**, *359–367.*

NGUYAN, TUE H. *1994. Oxidation Degradation of Protein Pharmaceuticals.* ACS symposium series, *59–71.*

NGUYAN, TUE H., BURNIER, J., & MENG, W. *1993. The Kinetics of Relaxin Oxidation by Hydrogen Peroxide.* Pharmaceutical Research, **10**, *1563–1571.*

NOSÉ, SHUICHI. *1991. Constant Temperature Molecuar Dynamic Methods.* Progress of Theoretical Physics Supplement, **103**, *1–46.*

OKAJIMA, TOSHIYA. *2001. Ab initio MO study on the solvent effect for hydrogen shift from hydrogen peroxide (H2O2) to water oxide (O-OH2).* Journal of Molecular Structure (Theochem), **572**, *45–52.*

OLENDER, R., & ELBER, R. *1997. Yet another look at the steepest descent path.* Journal of molecular structure, **398-399**, *63–71.*

OTTOLINA, G., & CARREA, G. *2001. Abinitio study of oxygen atom transfer from hydrogen peroxide to trimethylamine.* Journal of the Chemical Society. D, Chemical communications., *1748–1749.*

PARR, ROBERT G., & YANG, WEITAO. *1989.* Density-Functional Theory of Atoms and Molecules. *New York: Oxford University Press.*

PASSERONE, D., & PARRINELLO, M. *2001. Action-derived molecular dynamics in the study of rare events.* Physical Review Letters, **8710***(10), art. no.–108302.*

PASTOR, R. W., BROOKS, B. R., & SZABO, A. *1988. An Analysis of the Accuracy of Langevin and Molecular-Dynamics Algorithms.* Molecular Physics, **65***(6), 1409–1419.*

PATTEN, S. M. VAN, HANSON, E., BERNASCONI, R., ZHANG, K., MANAVALAN, P., COLE, E. S., MCPHERSON, J. M., & EDMUNDS, T. *1999. Oxidation of Methionine Residues in Antithrombin.* The Journal of Biological Chemistry, **274***(15), 10268–10276.*

PAYNE, M.C., TETER, M.P., ALLAN, D.C., ARIAS, T.A., & JOANNOPOULOS, J.D. *1992. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients.* Review of Modern Physics, **64**, *1045–1097.*

PEARLMAN, D. A., CASE, D. A., CALDWELL, J. W., ROSS, W. S., CHEATHAM, T. E., DEBOLT, S., FERGUSON, D., SEIBEL, G., & KOLLMAN, P. *1995. Amber, a Package of Computer-Programs for Applying Molecular Mechanics,* *Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules.* Computer Physics Communications, **91***(1-3), 1–41. Review.*

PEARLMAN, R., & BEWLEY, T. A. *1993.* Stability and Characterization of Human Growth Hormone,. *Stability and characterization of protein and peptide drugs-case histories. New York: Plenum Press.*

PELLEGRINI, M., ROYO, M., ROSENBLATT, M., CHOREV, M., & MIERKE, D. F. *1998. Addressing the tertiary structure of human parathyroid hormone-(1-34).* Journal of Biological Chemistry*,* **273***(17), 10420–10427.*

POPLE, J. A., RAGHAVACHARI, K., FRISCH, M. J., BINKLEY, J. S., & SCHLEYER, P. v. R. *1983. Comprehensive theoretical study of isomers and rearrangement barriers of even-electron polyatomic molecules HmABHn (A, B = carbon, nitrogen, oxygen, and fluorine).* Journal of American Chemical Society*,* **105***, 6389.*

POTTS, J. T., KRONENBERG, H. M., & ROSENBLATT, M. *1982. Parathyroid-Hormone - Chemistry, Biosynthesis, and Mode of Action.* Advances in Protein Chemistry*,* **35***, 323–396.*

POWELL, MICHAEL F. *1996. A Compendium and Hydropathy/Flexibility Analysis of Common Reactive Sites in Proteins: Reactivity at Asn, Asp, Gln, and Met Motifs in Neutral pH Solution.* Pages 1–140 of: PEARLMAN, RODNEY, & WANG, JOHN Y. *(eds),* Formulation, Characterization, and Stability of Protein Drugs: Case Histories*. New York: Plenum Press.*

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., & FLANNERY, B. P. *1992.* Numerical Recipes in FORTRAN*. New York: Cambridge University Press.*

PURVIS, G. D. III, & BARTLETT, R. J. *1982. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples.* Journal of Chemical Physics*,* **76***, 1910.*

RANKY, WILLIAN O., & NELSON, DOUGLAS C. *1961. Organic Sulfur Compounds.* Page 170 of: KHARASCH, N. *(ed),* Organic sulfur compounds. *New York: Pergamon Press.*

REIDHAAR-OLSON, J. F., DE SOUZA-HART, J. A., & SELICK, H. EE. *1996. Identification of Residues Critical to the Activity of Human Granulocyte Colony-Stimulating Factor.* Biochemistry*,* **35***, 9304–9341.*

SCHÖNEICH, C., ZHAO, F., WILSON, G. S., & BORCHARDT, RONALD T. *1993. Iron-Thiolate Induced Oxidation of Methionine to Methionine Sulfoxide in Small Model Peptides. Intramolecular Catalysis by Histidine.* Biochimica et Biophysica Acta*,* **1158***, 307–322.*

SCHÖNEICH, C., ZHAO, F., YANG, J., & MILLER, B. L. *1997. Mechanisms of Methionine Oxidation in Peptides.* ACS symposium series*,* **4***, 79–89.*

SCUSERIA, G. E., & SCHAEFER, H. F. III. *1989. Is coupled cluster singles and doubles (CCSD) more computationally intensive than quadratic configuration interaction (QCISD)?* Journal of Chemical Physics*,* **90***, 3700.*

SHAHROKH, ZZAHRA. *1997. Developing Pharmaceutical Protein Formulations: Assumptions and Analytical Tools.* ACS Symposium Series*,* **675***, 1.*

SHUL'PIN, G. B., & LINDSAY-SMITH, J. R. *1998. Oxidation with the "H_2O_2-manganese(IV) complex carboxylic acid" reagent - 1. Oxidation of saturated hydrocarbons with peroxy acids and hydrogen peroxide.* Russian Chemical Bulletin*,* **47***(12), 2379–2386.*

SHUL'PIN, G. B., SUSS-FINK, G., & SHUL'PINA, L. S. *2001. Oxidations by the system "hydrogen peroxide-manganese(IV) complex-carboxylic acid" Part 3. Oxygenation of ethane, higher alkanes, alcohols, olefins and sulfides.* Journal of Molecular Catalysis A-Chemical*,* **170***(1-2), 17–34.*

SÖMJEN, D., SCHLÜTER, K. D., WINGENDER, E., MAYER, H., & KAYE, A. M. *1991. Stimulation of Cell-Proliferation in Skeletal Tissues of the Rat by Defined Parathyroid-Hormone Fragments.* Biochemical Journal*,* **277***, 863–868.*

SPRIK, M., & CICCOTTI, G. *1998. Free energy from constrained molecular dynamics.* Journal of Chemical Physics*,* **109***(18), 7737–7744.*

STADMAN, E. R. *1990. Metal Ion Catalyzed Oxidation of Proteins: Biochemical Mechanism and Biological Consequences.* Free Radical & Medicine*,* **9***, 315–325.*

STADTMAN, E. R. *1993. Oxidation of Free Amino Acids and Amino Acid Residues in Proteins by Radiolysis and by Metal-Catalyzed Reactions.* Annual Review of Biochemistry*,* **62***, 797–821.*

STADTMAN, EARL R. *1992. Protein Oxidation and Aging.* Science*,* **257***, 1220–1224.*

STADTMAN, EARL R., & OLIVER, CYNTHIA N. *1991. Metal-Catalyzed Oxidation of Proteins.* J. Biol. Chem*,* **266***, 2005–2008.*

STEINFELD, JEFFREY I. *1999.* Chemical Kinetics and Dynamics*. Upper Saddle River, NJ: Prentice Hall.*

STRICKLAND, L. A., BOZZATO, R. P., & KRONIS, K. A. *1993. Structure of Human Parathyroid Hormone(1-34) in the Presence of Solvents and Micelles.* Biochemistry*,* **32***(23), 6050–6057.*

SUE, S. C., JARRELL, H. C., BRISSON, J. R., & WU, W. G. *2001. Dynamic characterization of the water binding loop in the P-type cardiotoxin: Implication for the role of the bound water molecule.* Biochemistry*,* **40***(43), 12782–12794.*

SZABO, ATTILA, & OSTLUND, S. NEIL. *1989.* Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. New York: McCraw-Hill Inc.*

TRUSHIN, O., GRANATO, E., YING, S. C., SALO, P., & ALA-NISSILA, T. *2002. Mechanisms of dislocation nucleation in strained epitaxial layers.* Physica Status Solidi B-Basic Research*,* **232***(1), 100–105.*

ULISKY, A., & ELBER, R. *1990. A new technique to calculate steepest descent paths in flexible polyatomic systems.* Journal of Chemical Physics*,* **92***(2), 1510–1511.*

VOGT, WALTHER. *1995. Odication of Methionyl Residues in Proteins: Tools, Targets, and Reversal.* Free Radical & Medicine*,* **18***(1), 93–105.*

WANG, WEI. *1999. Instability, Stablization, and Formulation of Liquid Protein Pharmaceuticals.* International Journal of Pharmaceutics*,* **185***, 129–188.*

WEINAN, E., REN, WEIQING, & EIJNDEN, ERIC VANDEN. *2002. String method for the study of rate events.* Physical Review B*,* **66***, 052301.*

WINER, K. K., YANOVSKI, J. A., & CUTLER, G. B. JR. *1996. Synthetic human parathyroid hormone 1–34 vs calcitriol and calcium in the treatment of hypoparathyroidism.* Journal of the American Medical Association*,* **276***(8), 631–636.*

WISHART, D. S., BIGAM, C. G., HOLM, A., HODGES, R. S., & SYKES, B. D. *1995. H-1, C-13 AND N-15 RANDOM COIL NMR CHEMICAL-SHIFTS OF THE COMMON AMINO-ACIDS .1. INVESTIGATIONS OF NEAREST-NEIGHBOR EFFECTS.* Journal of Biomolecular NMR*,* **5***(1), 67–81.*

XU, Y., & MAVRIKAKIS, M. *2002. Adsorption and dissociation of O-2 on Ir(111).* Journal of Chemical Physics*,* **116***(24), 10846–10853.*

YAO, Y., YIN, D., JAS, G. S., KUCZERA, K., WILLIAMS, T. D., SCHÖNEICH, C., & SQUIER, T. C. *1996. Oxidative Modification of a Carboxyl-Terminal Vicinal Methionine in Calmodulin by Hydrogen Peroxide Inhibits Calmodulin-Dependent Acivation of the Plasma Membrane Ca-ATPase.* Biochemistry*,* **35***, 2767–2787.*

YIN, D., KUCZERA, K., & SQUIER, T. C. *2000. The Sensitivity of Carboxyl-Terminal Methionines in Calmodulin Isoforms to Oxidation by H_2O_2 Modulates the Ability to Activate the Plasma Membrane Ca-ATPase.* Chemical Research in Toxicology*,* **13***(103-110).*

YIN, JIN, CHU, JHIH-WEI, RICCI, MARGARET S., BREMS, DAVID N., WANG, DANIEL I. C., & TROUT, BERHNARDT L. *2004a. Effects of Antioxidants on Non-site-specific Oxidation of Methionine Residues in Granulocyte Colony-Stimulating Factor (G-CSF) and Human Parathyroid Hormone (hPTH) Fragment 13–34.* Journal of Pharmaceutical Sciences, submitted.

YIN, JIN, CHU, JHIH-WEI, RICCI, MARGARET S., BREMS, DAVID N., WANG, DANIEL I. C., & TROUT, BERHNARDT L. *2004b. Effects of Excipients on Hydrogen Peroxide Induced Oxidation of Methionine Residues in Granulocyte Colony-Stimulating Factor.* Journal of Pharmaceutical Sciences, submitted.

ZHANG, Y. H., FELLER, S. E., BROOKS, B. R., & PASTOR, R. W. *1995. Computer-Simulation of Liquid/Liquid Interfaces .1. Theory and Application to Octane/Water.* Journal of Chemical Physics*,* **103***(23), 10252–10266.*

ZINK, T., A., ROSS., LUERS, K., CIESLAR, C., RUDOLPH, R., & A., HOLDAK T. *1994. Structure and Dynamics of the Human Granulocyte Colony-Stimulationg Factor Determined by NMR Specstropy. Loop Mobility in a Four-Helix-Bundle Protein.* Biochemistry*,* **33***, 8453–8463.*