Computational Studies on Reactions of DNA Oxidation and Enzyme Catalysis

by

Ning Liu

Submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
August 2006

Your file Votre référence
ISBN: 978-0-494-19589-5
Our file Notre référence
ISBN: 978-0-494-19589-5

DALHOUSIE UNIVERSITY

To comply with the Canadian Privacy Act the National Library of Canada has requested that the following pages be removed from this copy of the thesis:

Preliminary Pages

Examiners Signature Page (pii)

Dalhousie Library Copyright Agreement (piii)

Appendices

Copyright Releases (if applicable)

Table of Contents

List of Figures

Abstract

Recent advances in computational chemistry, especially the impressive development of density functional theory in the last 20 years, have greatly increased our understanding of many complicated reactions, including catalysis by enzymes and oxidative damage to DNA in biological systems.

This thesis describes the application of density functional theory to the elucidation of the reaction mechanisms of several biological systems. The first part describes a study of the reaction mechanisms that lead to the formation of 5-guanidano-4-nitroimidazole and 8-nitroguanine in the peroxynitrite induced oxidation of guanine. Two competitive reaction mechanisms are investigated and it is shown that the guanine radical cation mechanism is preferred over the neutral guanine radical mechanism. The role of water, as catalyst and as a reactant, is also investigated. In the second part, density functional theory is used to study the regeneration of the active site of protein-tyrosine phosphatase. Two mechanisms have been proposed for the formation of the sulphenylamide intermediate and the subsequent reactivation of the catalytic site. Calculations suggest that two competitive mechanisms have similar overall energy barriers and that the preferred route will be determined by the availability of hydrogen peroxide or other oxidizing reagents. The last part describes the reversible isomerization of ribose-5-phosphate to ribulose-5-phosphate catalyzed by ribose-5-phosphate isomerase. The important role of the catalytic base, glutamic ion, in this process is fully addressed. The thesis concludes with some suggestions for future studies on other significant enzymes and on the redox potential of biomolecules.

List of Abbreviations and Symbols Used

ρ	Electron density
ψ_n	Energy eigenstate
Φ	Electronic wave function
Θ	Nuclear wave function
λ	Wavelength
φ	One-electron wave function
$\hat{T}_n$	Kinetic nuclear operator
$\hat{U}_{nn}$	Potential nuclear operator
$\hat{T}_e$	Kinetic electron operator
$\hat{U}_{ee}$	Electron interaction operator
$\hat{V}_{ne}$	Electron-nuclear interaction operator
χ	Spatial basis function
$\hat{F}$	Fock operator
$\hat{H}$	Hamiltonian operator
CC	Coupled cluster
CCD	Coupled cluster doubles
CCSD	Coupled cluster singles and doubles
CI	Configuration interaction
CISD	Configuration interaction singles and doubles
CNDO	Complete neglect of differential overlap
COSMO	Conductor-like screening model
DFT	Density functional theory
DNA	Deoxyribonucleic acid
EA	Electron affinity
GGA	Generalized gradient approximation
Glu	Glutamic
GTO	Gaussian-type orbitals

HF	Hartree-Fock
INDO	Intermediate neglect of differential overlap
IP	Ionization potential
IPCM	Isodensity polarized continuum model
KS	Kohn-Sham
LCAO	Linear combination of atomic orbitals
LDA	Local density approximation
LSDA	Local spin density approximation
LYP	Lee-Yang-Parr
MD	Molecular dynamics
MP	Møller-Plesset
NDDO	Neglect of diatomic differential overlap
PCM	Polarizable continuum model
PES	Potential energy surface
PTP	Protein tyrosine phosphatase
QM/MM	Quantum mechanics/molecular mechanics
R5P	Ribose-5-phosphate
RPI	Ribose-5-phosphate Isomerase
Ru5P	Ribulose-5-phosphate
SCF	Self-consistent field
SCI-PCM	Self-consistent isodensity PCM
SCRF	Self-consistent reaction field
SD	Slater determinant
TDDFT	Time-dependent DFT
UV/VIS	Ultraviolet-visible spectrophotometry
ZDO	Zero differential overlap

Acknowledgements

First of all, I would like to express my sincere thanks to my supervisor Dr. R. J. Boyd. His precise supervision, kind encouragement and constant support have made the last four years so memorable and are greatly appreciated.

Special thanks to Dr. Fuqiang Ban for his tremendous help in my first two years in this group. His advice came from both scientific research and life. I would also like to thank Dr. Balakrishnan Viswanathan for those unforgettable discussions in our office. Those talks were always inspiring and delightful. Many thanks to the help from other members in Dr. Boyd's group during my studies here.

I am grateful to Drs. Philip D. Pacey, Donald F. Weaver, and Neil Burford for their valuable advice. NSERC and the Dalhousie Graduate Fellowship Fund are also highly appreciated for the financial support.

Finally, I would like to thank my family and my girl friend Jing Li for their priceless help and support over the years. I would never have gone so far without you.

Chapter 1

Introduction

Biomolecules are challenging systems for computational methods. They usually contain many different types of atoms with very different electronegativities and properties, such as H, C, O, P, S, as well as several types of metal atoms.[1,2] The molecules can occur in multiply charged states and the different electronegativities lead to inter- and intra-molecular charge transfer.

Biomolecules also exhibit a large variety of different bonding types, including covalent, ionic, and hydrogen bonding and van der Waals interactions. Compared to covalent bond energies, hydrogen bonds are much weaker, ranging from 8 to 80 kJ/mol, which puts considerable demands on the accuracy of computational methods.

The potential energy surfaces (PES) of biomolecules are often highly complex, exhibiting many local minima with small energy differences and separated by small energy barriers.[3] These minima can be very shallow, where large geometrical changes can occur with only a small change in energy.

Moreover, many biomolecules of interest are very large, while chemical reactions are sensitive to the protein environment and solvent effects. Therefore, it is very important to

choose carefully the model to be studied, which has to be able to represent the real system to a certain degree and to be feasible for high-level quantum computational methods as well.

Finally, the description of chemical reactions and transfer processes of protons and electrons is of special interest in theoretical modeling.[4] Therefore, the methods should address issues of bond breaking and formation properly. Furthermore, it is desirable that reaction energies, transition states and reaction pathways are determined with high accuracy, i.e. within about 10 kJ/mol.

For the reasons described above, the balance between accuracy of the theoretical model and the computational cost is a challenging and problematic task when choosing the level of theory to describe medium-sized organic molecules (e.g., guanine contains 78 electrons). The computational cost strongly depends on the choice of appropriate basis sets and the formal scaling behaviour of the computational time with system size for the chosen theoretical methods. High-level *ab initio* methods, like coupled-cluster including singles and doubles (CCSD)[5] or the very similar quadratic configuration interaction (QCISD)[6] method are in principle applicable, and if the triple excitations (T) are included, they should yield the desired accuracy. However, these methods inherently have a steep scaling behaviour with system size, and thus, very soon the limits of current hardware are exceeded. Density functional theory (DFT), however, usually provides fast and sufficiently accurate results for chemical problems of this type. Therefore, in recent years,

DFT methods have flourished dramatically in computational chemistry, especially in the calculation of medium to large sized biomolecular systems.

As a topic of continuing importance in biology and also in chemistry, DNA has been intensively studied with DFT methods in the last decade. Topics in this area include the ionization potential (IP) and electron affinities (EA) of DNA nucleobases, electron transfer processes, hydrogen bonding, etc. The structural properties and EA/IP of nucleobases are relatively well studied by DFT methods.[7,8] The agreement between experimental values and the results of DFT calculations is often very good.[9] Another topic closely related to the ionization potential is the redox potential of biomolecules in aqueous solution. Unfortunately, even though this problem is very important in many fields of biological study, it remains undeveloped. A satisfactory method to treat solvent effects in redox processes still has not been provided.

The treatment of electron transfer processes is also very challenging. It has been shown that DNA double helices transfer electronic charges through long distances when the charge carrier is a hole. The electronic transfer through a DNA segment is very fast (in the range of 10^{-10} s) and irreversible. Trapping of a radical cation occurs most frequently at the guanine of the GC pairs. This reaction causes oxidative damage at guanine, which can lead to dangerous mutations. This process is controversial. A few mechanisms have been proposed and DFT calculations have been carried out to study the structures, charges, electron affinities, ionization potentials and other properties of DNA segments in order to gain insight into the mechanism of oxidative damage.[7,10]

Hydrogen bonding is another interesting but also controversial problem in biological chemistry. It plays, for instance, a key role in the operation of the genetic code. In 1953, Watson and Crick proposed a structure for DNA in which two helical chains of nucleotides are held together by the hydrogen bonds that occur in a selective fashion between a purine and a pyrimidine nucleic base giving rise to the Watson-Crick pairs adenine-thymine (AT) and guanine-cytosine (GC). Theoretical studies with DFT methods indicate that donor-acceptor orbital interactions between the DNA bases and electrostatic interactions make comparable contributions to the formation of the Watson-Crick pairs.[11,12] It is well established that if the environmental effects are considered, DFT methods are able to describe adequately biologically relevant molecules involving hydrogen bonds. [13] It remains a problem whether in general DFT can cope well with systems that involve stacking between DNA base pairs.[14,15]

It is also a challenging task to describe directly reactions associated with the nucleobases or DNA segment, especially the oxidative damage to DNA, with DFT methods. These reactions often involved many steps and sometimes yield multiple products.[16-18] Furthermore, a change in the environment can influence the reactions dramatically. Relatively little research has been focused on this issue.

Of the naturally occurring nucleobases present in human DNA, guanine has the lowest oxidation potential and is therefore most easily oxidized. The reactions are usually quite complicated and give multiple products. In the present work, the one-electron oxidation

of guanine by peroxynitrite is studied with a DFT method and the Onsager solvation model.

As a powerful computational method, DFT has already been used to treat systems with sizes much bigger than that of nucleobases and small peptide chains. Enzymes, the kind of biological systems that usually contain hundreds or even thousands of atoms, including protein backbones, are also widely studied by using DFT methods.

Enzymes are essential to sustain life, because most chemical reactions in biological cells would occur too slowly, or would lead to different products, without enzymes. A malfunction (mutation, overproduction, underproduction or deletion) of a single critical enzyme can lead to a severe disease.

Like all catalysts, enzymes work by lowering the activation Gibbs energy of a reaction, thus allowing the reaction to proceed much faster. An enzyme remains unaltered by the completed reaction and can therefore continue to function. Most chemical catalysts catalyze a wide range of reactions and are not very selective. In contrast, enzymes are usually highly selective, catalyzing specific reactions only.[19,20]

An enzyme can be a monomeric protein, *i.e.* protein containing only one polypeptide chain, or an oligomeric protein, which consists of several polypeptide chains that act together. As with any protein, each monomer is actually produced as a long, linear chain of amino acids, which folds in a particular fashion to produce a three-dimensional

product. Individual monomers may then combine via non-covalent interactions to form a multimeric protein.

Most enzymes are larger than the substrates they act on and only a very small portion of the enzyme, around 10 amino acids, comes into direct contact with the substrate(s).[21,22] This region, where binding of the substrate(s) and the reaction occurs, is known as the active site of the enzyme.

Since enzymes catalyze many difficult and critical reactions in biological systems, the study of enzymes, especially their structures and catalytic mechanisms, has long been an active research topic in biochemistry. There are a number of general themes within the study of enzymes that interest chemists: (1) where and how does proton transfer facilitate bond cleavage, and how is this linked to charge transfer? (2) How are electron transfer and proton transfer coupled or gated? (3) How do the catalytic sites and amino acid residues provide orientation and guidance for the substrate along a reaction pathway? (4) What are the roles of substrate and protein strain? (5) What are the important influences from the protein environment?

To answer those questions, a detailed understanding of the electronic structures of the target enzymes is needed, and *ab initio* computational chemistry methods, especially density functional theory (DFT) methods, have been proven to be very powerful tools in such a realm. DFT methods can be used for calculations on large enzyme systems with good accuracy for structures, properties, and energetics.[23,24] The typical quantitative

accuracy of DFT methods is often quite good for identifying and systematizing the important energetic features of enzymes and for distinguishing between feasible and unlikely reaction pathways. A rough rule is to expect reaction energies and barriers to be accurate to 10-20 kJ/mol (this may be partially due to the fact that there is substantial cancellation of errors) with, for example, the B3LYP method of DFT.[25] Typical barriers or reaction energies are often 0-200 kJ/mol; these are much less than total atomization energies or dissociation energies for the complex as a whole and so are the expected errors.

DFT studies of enzymes usually focus on the active sites, which may contain nitrogen, oxygen, sulphur atoms or metal ions that pay crucial roles in the catalytic processes. In most cases, the backbone proteins are either represented by amino-acid residues or treated separately with molecular mechanics and the latter approach is usually referred to as quantum mechanical and molecular mechanical methods (QM/MM),[26] in which the portion treated with quantum mechanics is called the "quantum cluster".

DFT studies of the active sites of enzymes can be divided into two types: those that characterize the structural properties of a quantum cluster[27] and those that evaluate the reaction mechanisms related to active sites.[28]

In the first type, DFT calculations are carried out to characterize the interaction between the active sites of an enzyme and its substrate, co-factor, or backbone amino-acid residues. In many cases, substrates are not directly bonded to active sites. The water

molecule, which is abundant in biological systems, acts as a bridge between them by means of hydrogen bonding or electrostatic interaction, orienting substrates and transferring protons. In those cases, DFT methods are often combined with molecular dynamics (MD) simulations and give a quantum level description of the parts that are important in the catalytic processes.[21]

The major application of DFT calculations to the study of enzymes is to investigate and evaluate reaction mechanisms either associated with the catalytic process or with the reactivation of the enzyme itself. In order to characterize a particular reaction pathway and to distinguish a favourable mechanism from unlikely ones, intermediates and transition states that are involved in the mechanisms are usually studied with DFT methods.

In the current work, two enzyme systems have been studied by means of DFT calculations. One is the protein tyrosine phosphatase 1B (PTP1B) enzyme. The present computational study focuses on reaction mechanisms for redox regulation of the enzyme's activity. The other is the ribose-5-phosphate isomerase (RPI) enzyme, which catalyzes the isomerization reaction of ribose-5-phosphate (R5P) to ribulose-5-phosphate (Ru5P).

In Chapter 2, a brief review of quantum chemistry will be presented with the emphasis on Hartree-Fock theory and density functional theory methods. Several solvation models, especially the Onsager solvation model, the hybrid quantum mechanical and molecular

mechanical (QM/MM) methods, and molecular dynamics (MD) methods are also briefly described.

Previous studies of peroxynitrite oxidation of guanine and its proposed reaction mechanisms are introduced in Chapter 3 and the results of a density functional theory study of this particular reaction is discussed in detail. The starting point of this study is the one-electron oxidized guanine molecule, a guanine radical cation, which undergoes a one-step deprotonation and forms the neutral guanine radical. The radical cation and neutral radical follow similar reaction pathways and end up with the same final products. In the present study, all geometries and energies of each compound involved in either reaction pathway are calculated. The two reaction pathways associated with each radical are compared from an energy point of view and the validity of the postulated mechanism is examined by means of DFT calculations. Current computational results indicate that the guanine radical cation reaction mechanism is preferred over the neutral guanine radical mechanism.

In Chapter 4, the regeneration of the protein-tyrosine phosphatase 1B (PTP1B) enzyme active site is studied in detail. PTP1B functions to remove the phosphoryl group from tyrosinephosphorylated proteins in insulin signalling. The regeneration of the active site involves a sulphenylamide intermediate derived from the intrastrand cross-linking between the catalytic serine and the neighboring backbone nitrogen. Two mechanisms have been proposed for the formation of the sulphenylamide intermediate and the subsequent reactivation of the catalytic site. In the current work, the proposed

mechanisms have been investigated by the use of the B3LYP method of density functional theory. The present calculations suggest that these two mechanisms have similar overall energy barriers and that the preferred route will be determined by the availability of hydrogen peroxide or other oxidizing reagents.

The reversible conversion of ribose-5-phosphate (R5P) to ribulose-5-phosphate (Ru5P) catalyzed by ribose-5-phosphate isomerase (RPI) is studied by means of DFT calculations in Chapter 5. This isomerization reaction plays a significant role in all forms of life. The reaction starts with ring opening of R5P. A high-energy enediolate intermediate is formed in the isomerization process. In the active site of the RPI enzyme, a glutamic residue acts as a catalytic base, lowering the system energy. Product Ru5P is finally derived with an energy very close to that of the initial R5P system. The present results show that the key intermediate enediolate can be easily derived with the presence of a Glu^- amino acid residue acting as a catalytic base, indicating that the Glu^- amino acid residue is essential in the active sites of all RPI enzymes.

Some possibilities for future work are outlined in Chapter 6. Computational studies of critical enzymes will remain an important direction. Proposed projects may come from some interesting but controversial enzymes, such as the nucleobases. Another unsolved problem is the calculation of the redox potentials of crucial biomolecules. Some basic concepts and difficulties about this issue are discussed in Chapter 6.

Chapter 2

Theoretical Background

2.1 The Schrödinger equation

The Schrödinger equation, proposed by the Austrian physicist Erwin Schrödinger in 1925, describes the time-dependence of quantum mechanical systems. It is of central importance to the theory of quantum mechanics, playing a role analogous to Newton's second law in classical mechanics. According to the postulates of quantum mechanics, the state of any system can be described by a wave function, which can be obtained by solving the Schrödinger equation:

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian operator and E is the energy of the system.

For many real-world problems the energy distribution of the system does not change with time, and it is useful to determine how the stationary states vary with position x (independent of the time t). The Schrödinger equation therefore can be separated into equations for the time and space variation of the wave function under the assumption that the potential energy is independent of time. As such, the time-independent Schrödinger equation can be derived from the time-dependent Schrödinger equation. For every time-

independent Hamiltonian H, there exists a set of quantum states, $|\psi_n\rangle$, known as energy eigenstates, and corresponding real numbers E_n satisfying the eigenvalue equation:

$$\hat{H}\Psi(r,R) = E\Psi(r,R) \tag{2.2}$$

where r and R represent the coordinates of electrons and nuclei, respectively. Such a state possesses a definite total energy, whose value E_n is the eigenvalue of the state vector with the Hamiltonian. This eigenvalue equation is referred to as the time-independent Schrödinger equation.

Essentially, any substance can be described by a number of nuclei and electrons interacting through the electrostatic forces. The Hamiltonian of such a system can be written in the following form:

$$\hat{H} = -\sum_{I=1}^{P}\frac{\hbar^2}{2M_I}\nabla_I^2 - \sum_{i=1}^{N}\frac{\hbar^2}{2m}\nabla_i^2 + \frac{e^2}{2}\sum_{I=1}^{P}\sum_{J\neq I}^{P}\frac{Z_I Z_J}{|R_I - R_J|} + \frac{e^2}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{|r_i - r_j|} - e^2\sum_{I=1}^{P}\sum_{i=1}^{N}\frac{Z_I}{|R_i - r_j|} \tag{2.3}$$

where $R = \{ R_I \}$, I = 1, … ,P, is a set of P nuclear coordinates and $r = \{ r_i \}$, i = 1, … , N, is a set of N electronic coordinates. Z_I and M_I are the P nuclear charges and masses, respectively. Theoretically, all the properties of the system can be derived through solving the Schrödinger equation.

In practice, this task is virtually impossible except for some particular cases, in which numerical solutions are limited to a very small number of particles. Several reasons contribute to this fact. First, in the many-body system, nuclei and electrons follow different statistics and therefore can not be treated uniformly. Second, the full Schrödinger equation can not be decoupled into a set of independent equations and that means it is necessary to deal with $(3P + 3N)$ coupled degrees of freedom. Contemporary *ab initio* calculations rely upon two fundamental approximations: (i) the adiabatic separation of nuclear and electronic degrees of freedom (adiabatic approximation) and (ii) the classical treatment of the nuclei.

2.2 Adiabatic approximation (Born-Oppenheimer approximation[29])

The validity of this approximation is based on the observation that the timescale associated with the motion of the nuclei is much longer than that associated with electrons. This is quite understandable because the mass of an electron as compared to that of a proton (the lightest nucleus) is about 1 in 1836, and therefore for a given energy the velocity of an electron is much larger. Therefore, from the beginning of quantum mechanics it was proposed that the electrons can be adequately described as being able to adjust themselves instantaneously to follow the motion of nuclei and remain in a stationary state.

Under the Born-Oppenheimer approximation, the full wave function factorizes in the following way:

$$\Psi(\mathbf{R},\mathbf{r}) = \Theta_m(\mathbf{R})\Phi_m(\mathbf{R},\mathbf{r}) \tag{2.4}$$

where the electronic wave function $\Phi_m(\mathrm{R},\mathrm{r})$ is the mth stationary state of the electronic Hamiltonian

$$\hat{h}_e = \hat{T}_e + \hat{U}_{ee} + \hat{V}_{ne} = \hat{H} - \hat{T}_n - \hat{U}_{nn} \tag{2.5}$$

$\hat{T}_n$ and $\hat{U}_{nn}$ are the kinetic and potential nuclear operators, $\hat{T}_e$ and $\hat{U}_{ee}$ are the same for the electrons and $\hat{V}_{ne}$ is the electron-nuclear interaction. The corresponding eigenvalue is noted as $\varepsilon_m(\mathbf{R})$. The nuclear wave function $\Theta_m(\mathrm{R})$ obeys the electronic stationary Schrödinger equation

$$\left[\hat{T}_n + \hat{U}_{nn} + \varepsilon_m(\mathrm{R})\right]\Theta_m(\mathrm{R}) = E_m\Theta_m(\mathrm{R}) \tag{2.6}$$

Although the above equation holds true for any electronic eigenstate, the current discussion is focused on the ground state (m=0).

2.3 Classical nuclei approximation

Even with the adiabatic approximation, solving the Schrödinger equation is still a formidable task. Treating both the electrons and nuclei in the quantum framework is currently not applicable for a system with more than six nuclear degrees of freedom.

However, in many cases, the solution of the quantum nuclear equation (2.6) is not necessary. The validity of treating nuclei as classical particles is based on the following observation: The thermal wavelength for a particle of mass M is given by $\lambda\!\!\!^{-}_T = \hbar / \sqrt{Mk_BT}$, so particles in the space separated by more than $\lambda\!\!\!^{-}_T$ do not exhibit quantum phase coherence. The least favourable case is that of hydrogen, and at room temperature $\lambda\!\!\!^{-}_T \approx 0.4\,\text{Å}$, while inter-atomic distances are normally of the order of $1\,\text{Å}$.

Assuming these two approximations, the problem is reduced to solving the many-body electronic Schrödinger equation for a set of fixed nuclear positions. This is a major issue of quantum mechanics and the starting point of an *ab initio* calculation in quantum chemistry.

2.4 The Hartree-Fock theory

Within the framework of the Born-Oppenheimer approximation and classical nuclei approximation, the structure and other physical properties of matter can be derived through solving for interacting electrons in the external field given by a set of atomic nuclei. However, the exact numerical solution is only known for a few simple systems of very limited size. Therefore, the Schrödinger equation is usually solved analytically, and when it comes to an analytic solution, one always has to resort to approximation.

The first approximation for the analytical solution to the quantum many-body problem was proposed by Hartree.[30] It is postulated that the *many-electron* wave function can be written as a simple product of a set of *one-electron* wave functions. Each of these wave functions corresponds to a one-particle Schrödinger equation in an effective potential composed of the Coulombic interaction with nuclei and other electrons, while the later is treated in an average way:

$$\Phi(\mathrm{R},\mathrm{r}) = \prod_i \varphi_i(\mathrm{r}_i) \tag{2.7}$$

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_{eff}^{(i)}(\mathbf{R},\mathbf{r})\right)\varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}) \tag{2.8}$$

with

$$V_{eff}^{(i)}(\mathbf{R},\mathbf{r}) = V(\mathbf{R},\mathbf{r}) + \int \frac{\sum_{j\neq i}^{N} \rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \tag{2.9}$$

where

$$\rho_j(\mathbf{r}) = |\varphi_j(\mathbf{r})|^2 \tag{2.10}$$

is the electronic density associated with particle *j*. In the Hartree method, the energy of the many-body system is given by

$$E_H = \sum_1^N \varepsilon_n - \frac{1}{2}\sum_{i\neq j}^{N} \iint \frac{\rho_i(\mathbf{r})\rho_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \tag{2.11}$$

The subtraction of the second term on the right hand-side of equation (2.11) is needed because the formulation in terms of the effective potential counts each electron-electron interaction twice.

The wave function defined in equation (2.7) is derived through a procedure called the *self-consistent Hartree* approximation. In this procedure, a trial wave function is used to generate a set of variational parameters through minimizing the energy in equation (2.8), and then these parameters are used in the expression for the effective potential (2.9) to solve the Schrödinger equation. This process is repeated until the input and output wave function or potential achieve self-consistency.

The Hartree approximation treats the electrons as distinguishable particles. A major modification to the Hartree method is to introduce the Pauli exclusive principle, and this is done by formulating a many-electron wave function into a Slater determinant.

$$\begin{aligned}&\Phi(\mathbf{R},\mathbf{r})\\&= SD\left\{\varphi_j(\mathbf{r}_i,\sigma_i)\right\}\\&= \frac{1}{\sqrt{N!}}\begin{vmatrix}\varphi_1(\mathbf{r}_1,\sigma_1) & \varphi_j(\mathbf{r}_2,\sigma_2) & \cdots & \varphi_j(\mathbf{r}_N,\sigma_N)\\ \varphi_2(\mathbf{r}_1,\sigma_1) & \varphi_j(\mathbf{r}_2,\sigma_2) & \cdots & \varphi_j(\mathbf{r}_N,\sigma_N)\\ \vdots & \vdots & \ddots & \vdots\\ \varphi_N(\mathbf{r}_1,\sigma_1) & \varphi_j(\mathbf{r}_2,\sigma_2) & \cdots & \varphi_j(\mathbf{r}_N,\sigma_N)\end{vmatrix}\end{aligned} \tag{2.12}$$

Mathematically, switching the position of any two electrons in the system will lead to a change of the sign of the wave function, and the Fermi statistics for electrons is therefore

introduced in an exact manner.[31,32] The approximation is called *Hartree-Fock* (HF) or self-consistent field (SCF) approximation.

It follows from the Schrödinger equation that the energy of a system is given by

$$E = \frac{\int \Psi^* \hat{H} \Psi d\tau}{\int \Psi^* \Psi d\tau} \tag{2.13}$$

The energy calculated from equation (2.13) is the expectation value of the Hamiltonian operator. It is obvious that the energy will be the exact, true energy of the system only if the wave function Ψ and the Hamiltonian $\hat{H}$ are exact. The variation theorem states that *the energy calculated from equation (2.13) must be greater than or equal to the true ground-state energy of the system.* In practice, any molecular wave function inserted into equation (2.13) is always an approximation to the true wave function and so the molecular energy calculated variationally will always be greater than the true energy. The correct energy always lies below any calculated by the *HF* method, and the better the wave function, the lower the calculated energy. Therefore, the molecular wave function can be derived from the minimization of the energy with respect to the spin orbitals. Application of the variational approach leads to the generation of the Hartree-Fock equations

$$\hat{F}\phi_i = \varepsilon_i \phi_i \tag{2.14}$$

$$\hat{F} = -\frac{1}{2}\nabla^2 + \sum_A \frac{Z_A}{r_A} + \sum_j \left(2\hat{J}_j - \hat{K}_j\right) \tag{2.15}$$

where $\hat{F}$ is the Fock operator and ε_i is the orbital energy. The first and second parts of the Fock operator correspond to the kinetic energy of the electron and Coulomb interactions between the electrons and the nuclei. The Coulomb operator $\hat{J}_j$ represents a classical repulsion between two atomic orbitals while the exchange operator $\hat{K}_j$ arises from the antisymmetric nature of the determinantal wave function.

Equation (2.15) can be solved numerically for atoms, but no practical procedures are presently available for obtaining numerical solutions for molecules directly. The general practice is to expand the unknown molecular orbitals as a linear combination of a set of known spatial basis functions[33].

$$\phi_i = \sum_{\alpha} C_{\alpha} \chi_a \tag{2.16}$$

Consequently, the differential equation (2.15) is converted to a set of algebraic equations and solved by standard matrix techniques. This approach is known as the Roothaan-Hall method.[33,34]

Within the Hartree-Fock method, the antisymmetric wave function is approximated by a single Slater determinant. Exact wave functions, however, cannot generally be expressed as single determinants. The single-determinant approximation does not take into account

Coulomb correlation, leading to a total electronic energy different from the exact solution of the non-relativistic Schrödinger equation within the Born-Oppenheimer approximation.

2.5 Correlated calculations

Using a single determinant to represent the electronic wave function, the Hartree-Fock method is inherently unable to describe the explicit electron-electron repulsion, referred to as correlation.[35] Many post Hartree-Fock methods have been developed to treat the correlated electron motion better than does the *HF* method. Some of these methods are Møller-Plesset perturbation theory (MPn, where n is the order of correction), configuration interaction (CI) and coupled-cluster theory (CC). Usually, post-Hartree-Fock methods give more accurate results than Hartree-Fock calculations, although the added accuracy comes with increased computational cost.

Møller-Plesset perturbation theory

The Møller-Plesset (MP) treatment of electron correlation is based on perturbation theory, a very general mathematical approach for the treatment of complex systems in physics. The implementation of this particular approach was described by Møller and Plesset in 1934[36] and developed into a practical computational method by Binkley and Pople.[37] Perturbation theory seeks an approximate solution for a complicated system in terms of a simpler system. The idea is to start with a simple system and gradually turn on an

additional "perturbing" Hamiltonian representing a weak disturbance to the system. If the disturbance is not too large, the various physical quantities associated with the perturbed system (e.g. its energy levels and eigenstates) will be continuously generated from those of the simple system. There is a hierarchy of MP energy levels: MP0, MP1 (these first two designations are not actually used), MP2, etc., which successively give more accurate descriptions of the interelectronic repulsion.

In the Møller-Plesset approach, the unperturbed Hartree-Fock Hamiltonian operator $\hat{H}_0$ is extended by adding a small perturbation V:

$$\hat{H} = \hat{H}_0 + \lambda \hat{V} \tag{2.17}$$

where λ is an arbitrary parameter. If the perturbation is sufficiently small, then the resulting wave function and energy can be expressed as a power series in λ:

$$\Psi = \lim_{n\to\infty} \sum_{i}^{n} \lambda^i \Psi^{(i)} \tag{2.18}$$

$$E = \lim_{n\to\infty} \sum_{i}^{n} \lambda^i E^{(i)} \tag{2.19}$$

Substitution of these series into the time-independent Schrödinger equation gives a new equation:

$$\left(\hat{H}_0 + \lambda V\right)\left(\sum_i^n \lambda^i \Psi^{(i)}\right) = \left(\sum_i^n \lambda^i E^{(i)}\right)\left(\sum_i^n \lambda^i \Psi^{(i)}\right) \tag{2.20}$$

The solution of this equation to zero order gives an energy, which is the sum of the orbital energies for the electrons. Solution to first order (n = 1) corrects this energy and gives the Hartree-Fock energy and wave function. To go beyond the Hartree-Fock treatment it is necessary to go beyond first order. Second (MP2), third (MP3), and fourth (MP4) order Møller-Plesset calculations are standard levels used in calculating small systems and are implemented in many computational chemistry codes. Higher level MP level calculations are possible in some codes, however, they are rarely used.

Systematic studies of MP perturbation theory have shown that it is not a convergent theory at high orders. The convergence properties can be slow, rapid, oscillatory, regular or highly erratic, depending on the precise chemical system or basis set.[38]

Configuration Interaction

The configuration interaction (CI) treatment of electron correlation is based on the simple idea that one can improve the HF wave function, and hence energy, by adding Slater determinants that represent the promotion of electrons from occupied to virtual MOs. The HF and additional Slater determinants each represent a particular electronic configuration, and the actual wave function and electronic structure of the system can be approximated as the result of the interaction of these configurations.

The CI method is based on the Roothaan-Hall method. Suppose one has solved the Roothaan-Hall equations in a finite basis set and obtained a set of $2K$ spin orbitals $\{\phi_i\}$. The determinant formed from the N lowest energy spin orbitals is $|\Psi_0\rangle$. In addition to $|\Psi_0\rangle$, there are a large number of other N-electron determinants that can be formed from the $2K$ spin orbitals. Thus, one can form a set of possible determinants including $|\Psi_0\rangle$, the singly excited determinants $|\Psi_a^r\rangle$ (which is different from $|\Psi_0\rangle$ in having the spin orbital ϕ_a replaced by ϕ_r), the doubly excited determinants $|\Psi_{ab}^{rs}\rangle$, etc., up to N - tuply excited determinants. One can use these many electron wave functions to form a better approximation to $|\Phi_0\rangle$ than $|\Psi_0\rangle$.

$$|\Phi_0\rangle = c_0|\Psi_0\rangle + \sum_{ar} c_a^r|\Psi_a^r\rangle + \sum_{\substack{a<b \\ r<s}} c_{ab}^{rs}|\Psi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c \\ r<s<t}} c_{abc}^{rst}|\Psi_{abc}^{rst}\rangle + \ldots \tag{2.21}$$

This is the form of the full CI wave function.

For n-tuply excitations, one can choose n spin orbitals from those occupied in $|\Psi_0\rangle$ in $\binom{N}{n}$ ways and similarly choose n orbitals from the $2K$-N virtual orbitals in $\binom{2K-N}{n}$ ways. Thus, the total number of n-tuply excited determinants is $\binom{N}{n}\binom{2K-N}{n}$. Even for small molecules and one-electron basis sets of only moderate

size, the number of n-tuply excited determinants is extremely large for all n except 0 and 1.

Given the trial function of Equation (2.21), the corresponding energies can be found by using the linear variation method, which consists of forming the matrix representation of the Hamiltonian in the basis of the *N*-electron functions of expansion (2.21) and finding the eigenvalues of this matrix. This is called the full CI matrix, and the method is referred to as full CI. The lowest eigenvalue will be an upper bound to the ground state energy of the system. The higher eigenvalues will be upper bounds to the excited states of the system.

For all but the smallest molecules, even with a minimal basis set, full CI is a computationally impractical procedure. To obtain a computationally viable scheme one must truncate the full CI matrix. A systematic procedure for accomplishing this is to consider only those configurations that differ from the Hartree-Fock ground state wave function by no more than a predetermined number, *M* for example, of spin orbitals. For full CI, *M* would be equal to the number of electrons, *N*. One simple version of this scheme includes only single and double excitations out of $|\Psi_0\rangle$. The calculation including singly and doubly excited CI is called CISD, similarly, the calculation involving only singly excited CI is CIS, and the calculation involving only doubly excited CI is CID. The effect of single excitations on the correlation energy is generally negligible, whereas the CID and CISD methods can give a major fraction of the correlation energy.

Coupled-cluster Theory

Coupled-cluster theory proceeds by imposing an ansatz on the wave function, where an assumption is made that

$$|\Psi\rangle = e^{\hat{T}}|\Phi_0\rangle \tag{2.22}$$

where $|\Phi_0\rangle$ is an appropriate reference function, such as the Hartree-Fock wave function.

When this is substituted back into the original Schrödinger equation, the result is

$$\hat{H}e^{\hat{T}}|\Psi_0\rangle = Ee^{\hat{T}}|\Psi_0\rangle \tag{2.23}$$

from which we obtain

$$\hat{G}|\Psi_0\rangle = e^{-\hat{T}}\hat{H}e^{\hat{T}}|\Psi_0\rangle = E|\Psi_0\rangle \tag{2.24}$$

It can be shown that the ansatz in equation (2.24) preserves the energies of the original Hamiltonian, while transforming the original Hamiltonian operator to the new operator G. Coupled-cluster theory then suggests that there is an appropriate cluster operator $\hat{T}$, for which the resulting transformed Hamiltonian G is greatly simplified. This means that the matrix representation of G is a diagonal matrix or very nearly diagonal matrix.

The form of the wave function is an ansatz because there is no experimental rationale suggesting that this is the case. Actually it is used because the form of the equation has particular properties that allow complicated quantum systems to be tractable. In particular, because the cluster operator is exponentiated, an operator inside the cluster operator that corresponds to a single excitation of a state simultaneously is able to promote one, two, etc., electrons respectively into virtual spin orbitals. This is because

$$e^{\hat{T}} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \ldots = \sum_{n=0}^{\infty} \frac{\hat{T}^2}{n!} \tag{2.25}$$

and from this Taylor series, it is apparent that higher order states are excited due to the compounding effect of repeated applications of the cluster operator. Depending how many terms are actually included in the summation for $\hat{T}$, one obtains the *coupled-cluster doubles* (CCD), *coupled-cluster singles and doubles* (CCSD) or *coupled-cluster singles, doubles, and triples* (CCSDT) methods:

$$\hat{T}_{CCD} = e^{\hat{T}_2}\Psi_{HF}$$

$$\hat{T}_{CCSD} = e^{(\hat{T}_1+\hat{T}_2)}\Psi_{HF}$$

$$\hat{T}_{CCSDT} = e^{(\hat{T}_1+\hat{T}_2+\hat{T}_3)}\Psi_{HF}$$

The coupled-cluster method gives the exact non-relativistic solution of the Schrödinger equation of the n-body problem if one includes up to the $\hat{T}_n$ cluster operator. However,

the computational effort of solving the equations grows steeply with the order of the cluster operator and in practical applications the method is limited to the first few orders.

2.6 Density functional theory

The total ground state energy of an inhomogeneous system composed by N interacting electrons is given by

$$\begin{aligned} E &= \langle \Phi | \hat{T} + \hat{V} + \hat{U}_{ee} | \Phi \rangle \\ &= \langle \Phi | \hat{T} | \Phi \rangle + \langle \Phi | \hat{V} | \Phi \rangle + \langle \Phi | \hat{U}_{ee} | \Phi \rangle \end{aligned} \tag{2.26}$$

Where Φ is the N-electron ground state wave function. This wave function has to include correlations amongst electrons, and its general form is not known. $\hat{T}$ is the kinetic energy, $\hat{V}$ accounts for the interaction with external fields, and $\hat{U}_{ee}$ is the electron-electron interaction that introduces many-body effects.

$$\begin{aligned} U_{ee} &= \langle \Phi | \hat{U}_{ee} | \Phi \rangle = \langle \Phi | \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq 1}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} | \Phi \rangle \\ &= \int \frac{\rho_2(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \end{aligned} \tag{2.27}$$

where ρ_2 is the two-body density matrix in space. The two-body direct correlation function g($\mathbf{r}$, $\mathbf{r}$') can be defined in the following way:

$$\rho_2(\mathbf{r},\mathbf{r}')=\frac{1}{2}\rho(\mathbf{r},\mathbf{r})\rho(\mathbf{r}',\mathbf{r}')g(\mathbf{r},\mathbf{r}') \tag{2.28}$$

It takes into account the fact that if two electrons are correlated, then the probability of finding electron *a* at a certain position in space depends on the position of electron *b*, and vice versa. In other words, the product of their independent density functions does not adequately describe the real situation. The uncorrelated pair density is too large at small distances and too small at large distances, meaning electrons tend to "avoid each other". With this definition, one can write the electron-electron interaction as

$$\begin{aligned} U_{ee} &= \frac{1}{2}\int\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' \\ &+\frac{1}{2}\int\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}[g(\mathbf{r},\mathbf{r}')-1]d\mathbf{r}d\mathbf{r}' \end{aligned} \tag{2.29}$$

The two-body direct correlation function g(**r**, **r**') includes both exchange and correlation interactions. According to the Pauli principle, the presence of electrons with the same spin in the same spot is forbidden. But it says nothing about electrons with opposite spins. Therefore, $g_X(\mathbf{r},\mathbf{r}')\rightarrow 1/2$ as $\mathbf{r}\rightarrow\mathbf{r}'$.

While the exchange interaction is perfectly solved in Hartree-Fock theory, the calculation of the correlation hole $g_C(\mathbf{r},\mathbf{r}')$ is still a problem in many-body theory. The exact solutions are known only for the homogeneous electron gas.[39]

In density functional theory, the energy of a many-body electronic system can, then, be written in the following way:

$$E = T + V + \frac{1}{2}\int\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + E_{XC} \tag{2.30}$$

where

$$V = \sum_{I=1}^{P}\langle\Phi|\sum_{i=1}^{N}v(\mathbf{r}_i - \mathbf{R}_I)|\Phi\rangle = \sum_{I=1}^{P}\int\rho(\mathbf{r})v(\mathbf{r}-\mathbf{R}_I)d\mathbf{r} \tag{2.31}$$

$$T = \langle\Phi|-\frac{\hbar^2}{2m}\sum_{i=1}^{N}\nabla_i^2|\Phi\rangle = -\frac{\hbar^2}{2m}\int\left[\nabla_{\mathbf{r}}^2\rho(\mathbf{r},\mathbf{r}')\right]_{\mathbf{r}'=\mathbf{r}}d\mathbf{r} \tag{2.32}$$

and E_{XC} is the exchange and correlation energy

$$E_{XC} = \frac{1}{2}\int\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}[g(\mathbf{r},\mathbf{r}')-1]d\mathbf{r}d\mathbf{r}' \tag{2.33}$$

In 1927, Thomas and Fermi brought forward the idea that the total energy may be constructed in terms of only the electronic density.[40] They used the expression for kinetic, exchange and correlation energies of the homogeneous electron gas to construct the same quantities for an inhomogeneous system. This was the first time the *local density approximation*, or LDA, was used. Neglecting exchange and correlation, one can write the TF energy in the following way:

$$E_{TF}[\rho] = C_K \int \rho(\mathbf{r})^{5/3} d\mathbf{r} + \int v(\mathbf{r})\rho(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \quad (2.34)$$

It can be seen that E_{TF} depends only on the electronic density. In other words, it is a *functional* of the density. Within LDA, exchange and correlation could be easily added by using an approximation to the homogeneous electron gas. Although the first DFT method was developed in anticipation that an explicit expression for the universal density functional really exists, this idea was not put on solid mathematical grounds until the Hohenberg-Kohn theorem was proven. The theorem is divided into two parts:

Theorem 1: *The external potential is univocally determined by the electronic density, except for a trivial additive constant.*

Corollary: *Since the electronic density univocally determines the external potential, it also determines the ground state wave function.*

Theorem 2: *The functional for the ground state energy is minimized by the ground state electron density*

These two theorems form the basis of DFT. Using DFT, one can determine the electronic ground state density and energy exactly, provided that the form of the energy functional in terms of the electron density is known. Since the electron density determines

univocally the potential, and this in turn determines univocally the many-body wave function, problems relating to both ground states and excited states should be solvable within DFT. Such a scheme for the ground state was devised by Kohn and Sham. Extensions of Kohn-Sham (KS) theory have also been developed for excited states[41-43] and the time-dependent DFT (TDDFT). [44,45]

There are two major problems associated with the original DFT method proposed by Thomas and Fermi. The biggest difficulty is to deal with correlation, and this is still now an active field of research. The other problem is that the expression of the kinetic energy $\langle \Psi | \hat{T} | \Psi \rangle$ given by Thomas and Fermi is local in the density, while the kinetic operator is inherently non-local.

The later problem was solved by Kohn and Sham[46], who proposed the idea of replacing the kinetic energy of the interacting electrons with that of an equivalent non-interacting system, since the latter is easier to calculate. In the ground state, the spin density matrix can be written as

$$\rho_s(\mathbf{r},\mathbf{r}') = \sum_{i=1}^{\infty} n_{i,s} \varphi_{i,s}(\mathbf{r}) \varphi_{i,s}^*(\mathbf{r}') \tag{2.35}$$

where $\{\varphi_{i,s}(\mathbf{r})\}$ are the natural spin orbitals and $\{n_{i,s}\}$ are the occupation numbers of these orbitals. The kinetic energy can be written exactly as

$$T = \sum_{s=1}^{2} \sum_{i=1}^{\infty} n_{i,s} \left\langle \varphi_{i,s} \left| -\frac{\nabla_i^2}{2} \right| \varphi_{i,s} \right\rangle \tag{2.36}$$

In the following, *the non-interacting reference system* is introduced into the KS model. It is supposed that this system is composed of non-interacting fermions and its ground state coincides with that of the interacting system. This system is described by the Hamiltonian

$$\hat{H}_R = \sum_{i=1}^{N} \left(-\frac{\nabla_i^2}{2} + \upsilon_R(\mathbf{r}_i) \right) \tag{2.37}$$

where the potential $\upsilon_R(\mathbf{r}_i)$ is such that the ground state density of $\hat{H}_R$ equals $\rho(\mathbf{r})$ and the ground state energy equals the energy of the interacting system. This Hamiltonian has no electron-electron interactions and therefore its eigenstates can be expressed in the form of Slater determinants

$$\Psi_s(\mathbf{r}) = \frac{1}{\sqrt{N!}} SD\left[\varphi_{1,s}(\mathbf{r}_1)\varphi_{2,s}(\mathbf{r}_2)\ldots\varphi_{N,s}(\mathbf{r}_{N_s})\right]$$

and the density is written as

$$\rho(\mathbf{r}) = \sum_{s=1}^{2} \sum_{i=1}^{N_s} \left|\varphi_{i,s}(\mathbf{r})\right|^2 \tag{2.38}$$

while the kinetic term is

$$T_R[\rho]=\sum_{s=1}^{2}\sum_{i=1}^{Ns}\left\langle\varphi_{i,s}\left|-\frac{\nabla^2}{2}\right|\varphi_{i,s}\right\rangle \tag{2.39}$$

The single-particle orbitals $\{\varphi_{i,s}(\mathbf{r})\}$ are the N_s lowest eigenfunctions of $\hat{h}_R=\left(\nabla^2/2\right)+\upsilon_R(\mathbf{r})$, that is,

$$\left\{\left(\frac{\nabla^2}{2}\right)+\upsilon_R(\mathbf{r})\right\}\varphi_{i,s}(\mathbf{r})=\varepsilon_{i,s}\varphi_{i,s}(\mathbf{r}) \tag{2.40}$$

The universal density functional can be written in the following form:

$$F[\rho]=T_R[\rho]+\frac{1}{2}\iint\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}'+E_{XC}[\rho] \tag{2.41}$$

This equation defines the exchange and correlation energy as a functional of the density. However, $T_R[\rho]$ is the kinetic energy of the non-interacting reference system and it does not include the correlation part of the true kinetic energy, which is taken into account by adding an additional term in the total KS functional[47]:

$$E_{KS}[\rho]=T_R[\rho]+\int\rho(\mathbf{r})v(\mathbf{r})d\mathbf{r}+\frac{1}{2}\iint\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}'+E_{XC}[\rho] \tag{2.42}$$

Following a procedure similar to that used in the derivation of the Hartree-Fock equations, one can arrive at the *Kohn-Sham* equations:

$$\left[\frac{1}{2}\nabla_i^2-\sum_{nuclei}\frac{Z_A}{\mathbf{r}_{1A}}+\int\frac{\rho(\mathbf{r}_2)}{\mathbf{r}_{12}}d\mathbf{r}_2+\upsilon_{XC}(1)\right]\psi_i^{KS}(1)=\varepsilon_i^{KS}\psi_i^{KS}(1) \tag{2.43}$$

or in the short form

$$\hat{h}^{KS}(1)\psi_i^{KS}(1) = \varepsilon_i^{KS}\psi_i^{KS}(1) \tag{2.44}$$

where the $\{\psi_i^{KS}\}$ are the KS special orbitals. Those orbitals are only mathematical treatments to minimize the kinetic energy under the fixed density constrain. $\{\varepsilon_i^{KS}\}$ are KS energy levels and υ_{XC} is the *exchange correlation potential*, which is defined as the functional derivative of $E_{KS}[\rho(\mathbf{r})]$ with respect to $\rho(\mathbf{r})$:

$$\upsilon_{XC}(\mathbf{r}) = \frac{\delta E_{KS}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} \tag{2.45}$$

If the density function $\rho_0(\mathbf{r})$ and the exchange-correlation energy functional $E_{KS}[\rho_0]$ were known, the KS equation (2.43) would give the exact energy. In DFT theory there is no known systematic way to improve the exchange-correlation functional. Therefore, devising good functionals $E_{KS}[\rho_0]$ is the main problem in density functional theory.

The simplest approximation is *LDA*, which applies to a homogeneous electron gas. This approximation is applicable to solid-state physics but unfavourable in molecular calculations. Better results can be obtained by the *local spin density approximation* (LSDA), in which electrons with different spins in the uniform gas are assigned to different KS orbitals. For species in which all the electrons are strictly paired the LSDA is equivalent to the LDA, while the former can better handle systems like radicals or molecules with severely distorted structures.

The *generalized-gradient approximation* (GGA) uses exchange-correlation energy functionals $E_{KS}[\rho_0]$. Unlike LSDA, GGA utilizes not only the electron density but also its gradient to derive $E_{KS}[\rho_0]$. This is a major improvement from using the uniform electron gas model. Within the GGA, the exchange-correlation energy functional can be written as the sum of an exchange-energy functional and a correlation-energy functional, i.e. $E_{XC} = E_x + E_c$ and the exchange-energy functional and the correlation-energy functional can be treated differently. Examples of gradient-correlated-exchange energy functional include the Becke 88 functional[48] and the G96 functional. Examples of gradient-correlated correlation-energy functionals are the Lee-Yang-Parr (LYP) and the Perdew 1986 (P86) functionals.

The hybrid method combines the Hartree-Fock and density-functional treatments of exchange effects, and gives more accurate energies and structures. In this approach, the Hamiltonian operators are constructed and characterized by the parameter λ,

$$0 \le \lambda \le 1.$$

Systems can be described by the class of Hamiltonian operators,

$$\hat{H}_\lambda = \hat{T}^0 + \lambda \hat{V}_{ee} + \hat{V}_\lambda + \hat{V}_C + \hat{V}_{ext}. \tag{2.46}$$

For any λ we assume we can solve the Schrödinger equation for the ground states

$$\hat{H}_\lambda \Psi_\lambda = E_\lambda \Psi_\lambda. \tag{2.47}$$

After some mathematical operations, we obtain the adiabatic connection formula

$$\langle \Psi_1 | \hat{V}_{ee} | \Psi_1 \rangle = \int_0^1 \langle \Psi_\lambda | \hat{V}_{ee} | \Psi_\lambda \rangle d\lambda \tag{2.48}$$

We try to approximate the integral on the right hand side. Finally we can write E_{XC} as

$$E_{XC} = E_{x;LSD} + a_0\left[E_{x,HF} - E_{x,LSD}\right] + E_{c;.LSD} + a_x\Delta E_{x;NLSD} + a_c\Delta E_{c;NLSD} \quad (2.49)$$

where $E_{x,HF}$ and $E_{x,LSD}$ account for the exchange interaction effect from Hartree-Fock theory and the local-spin-density approximation; $E_{c;.LSD}$ accounts for the correlation effects from local-spin-density approximation; $\Delta E_{x;NLSD}$ and $\Delta E_{c;NLSD}$ are the non-local contributions to the exchange and correlation energies, respectively. Equation 2.49 is due to Becke (1993), and includes three constants, a_0, a_x, and a_c. Therefore, calculations with this approach are given the acronym of B3xxx, where the 'xxx' denotes which non-local-density approximation is used. Most often the GGA of Lee, Yang, and Parr (1988) is applied, resulting in the very popular B3LYP functional.

2.7 Basis sets[49,50]

A basis set is a set of functions (basis functions), linear combinations of which yield molecular orbitals.

$$\psi_i = \sum_{s=1}^{m} c_{si}\phi_s \qquad i = 1, 2, 3, \ldots, m \text{ (m molecular orbitals)} \quad (2.50)$$

Approximating molecular orbitals as linear combinations of basis functions is usually called the LCAO approach. The functions are not necessarily conventional atomic orbitals: they can be any mathematical functions that give useful representations of molecular orbitals. In these cases, the wave functions under consideration are all represented as vectors, the components of which correspond to coefficients in a linear combination of the basis functions. The operators are then represented as matrices, in this finite basis.

Initially, these basis functions were Slater orbitals,

$$R(\mathbf{r}) = A\mathbf{r}^{l} e^{-a|\mathbf{r}-\mathbf{a}|} \tag{2.51}$$

which correspond to a set of functions that decayed exponentially with distance from the nuclei. Slater functions are good approximations to atomic wave functions. However, the evaluation of two-electron repulsion integrals with Slater functions on three or four centers requires excessive computer time. Later, it was realized that these Slater-type orbitals could in turn be approximated by linear combinations of Gaussian orbitals.

$$\phi = x^{k} y^{m} z^{n} \exp\left(-\zeta r^{2}\right) \tag{2.52}$$

Because it is easier to calculate overlap and other integrals with Gaussian basis functions, this led to huge computational savings.

Today, there are hundreds of basis sets composed of Gaussian-type orbitals, (GTOs). The smallest of these are called minimal basis sets, and they usually have the form of STO-*N*G, where *N* indicates the number primitive Gaussians used to compose each basis function. In most molecular bonding, however, it is the valence electrons that play the major roles. In recognition of this fact, it is common to represent valence orbitals by more than one basis function, (each of them can in turn be composed of a fixed linear combination of primitive Gaussian functions). The notation for these split-vale*nce* basis sets is typically *X*-*YZ*G. In this case, *X* represents the number of primitive Gaussians comprising each core atomic orbital basis function. The *Y* and *Z* indicate that the valence orbitals are composed of two basis functions each, the first one composed of a linear combination of *Y* primitive Gaussian functions, the other composed of a linear

combination of Z primitive Gaussian functions. Split-valence basis sets are probably the most commonly used basis sets in conventional *ab initio* and DFT calculation. Sometimes more than one basis function is used to describe each atom orbital in a basis set. Such basis sets are called double, triple, or quadruple-zeta basis sets. For example, cc-pVDZ is a double-zeta basis set that uses two basis functions to describe every valence orbital.

The most common addition to minimal and split-valence basis sets is a set of polarization functions, which are usually added to atoms other than hydrogen and helium. These are auxiliary functions with one additional node. This adds some additional needed flexibility within the basis set, effectively allowing molecular orbitals involving certain atoms, especially hydrogen, to be more asymmetric about the nucleus. For example, "6-31+G(d, p) means a "d" type and a "p" type are added to the heavy atoms and light atoms respectively. This is important when considering accurate representations of bonding between atoms, because the very presence of the bonded atom makes the energetic environment of the electrons spherically asymmetric.

Another common addition to basis sets is diffuse functions. These are Gaussian basis functions with very small exponents that more accurately represent the "tail" portion of the atomic orbitals, which are distant from the atomic nuclei. These additional basis functions can be important when considering anions and other large, "soft" molecular systems. For instance, the addition of diffuse functions on heavy atoms to the popular split-valence basis set "6-31G" would yield "6-31+G".

2.8 Solvent effects

Problems concerning phenomena occurring in solution, and the qualitative and quantitative description of the influence of a solvent on the structural and chemical behaviour of solute molecules, are of interest in many fields of chemistry, biochemistry and physics. Therefore the ability to reliably model solute-solvent interactions is of great importance. Methods for evaluating the effect of a solvent may broadly be divided into two types: those that describe the individual solvent molecules and those that treat the solvent as a continuous medium. The self-consistent reaction field (SCRF) models, which consider the solvent as a uniform polarizable medium with a dielectric constant, belong to the later. Within the SCRF models, the solute charge distribution interacts with the medium, polarising it and creating a reflection charge distribution on the cavity surface (the reaction field), which, in turn, will interact electrostatically with the solute leading to net stabilisation. SCRF methods differ according to the definition of the cavity and the reaction field.

The spherical cavity dipole only SCRF model is known as the Onsager model.[51,52] In this model, the solute is placed in a cavity (spherical or elliptical) immersed in a continuous medium with a dielectric constant ε. A dipole in the molecule will induce a dipole in the medium, and the electric field applied to the solute by the solvent dipole will in turn interact with the molecular dipole to lead to a net stabilization. The Hamiltonian of a solute molecule is given by

$$\hat{H}_{rf} = \hat{H}_0 + \hat{H}_1 \tag{1.53}$$

where $\hat{H}_0$ is the Hamiltonian of the isolated molecule and $\hat{H}_1$ represents the electrostatic solvent effect, describing the coupling between the molecular dipole operator ($\hat{\mu}$) and the reaction field, $\vec{R}$.

$$\hat{H}_1 = -\hat{\mu} \cdot \vec{R} \tag{2.54}$$

The reaction field $\vec{R}$ is proportional the molecular dipole moment, $\vec{\mu}$;

$$\vec{R} = g\vec{\mu} \tag{2.55}$$

The proportionality constant g that describes the reaction field intensity depends on the dielectric constant of the medium, ε, and on the radius of the spherical cavity, a_0:

$$g = \frac{2(\varepsilon - 1)}{(2\varepsilon + 1)a_0^{\,3}} \tag{2.56}$$

When the solvent polarization is included, the energy of the system is given by

$$E = \langle \Psi | H_0 | \Psi \rangle - \frac{1}{2}\vec{\mu} \cdot \vec{R} \tag{2.57}$$

Under the assumption of a spherical cavity, the ion-dipole solute-solvent interaction does not contribute to the electronic energy variationally, which accounts for the efficiency of the Onsager model.

Advanced *ab initio* SCRF models include the *polarizable continuum model* (PCM),[53] the *static isodensity surface polarized continuum model* (IPCM),[54] the *self-consistent isodensity* PCM model (SCI-PCM),[54] the conductor-like screening model (COSMO), [55,56] etc. The latter methods all adopt molecular shaped cavities and describe the reaction

field in terms of the apparent polarization charges or reaction field factors included in the solute Hamiltonian. Thus it is possible to perform calculations of the interaction between the solution wave function and the solvent wave polarization. Those models usually give more accurate results than the Onsager model, but are also more sensitive to initial structures and are computationally more expensive.

2.9 Semiempirical approaches

Semiempirical approaches are normally formulated within the same conceptual framework as *ab initio* methods, but they neglect many smaller integrals to speed up the calculations. In order to compensate for the errors caused by these approximations, empirical parameters are introduced into the remaining integrals and calibrated against reliable experimental or theoretical reference data. This strategy can only be successful if the semiempirical model retains the essential physics to describe the properties of interest.

In current practice, semiempirical methods serve as efficient computational tools, which can yield fast quantitative estimates for a number of properties. This may be particularly useful for correlating large sets of experimental and theoretical data, for establishing trends in classes of related molecules, and for scanning a computational problem before proceeding with higher-level treatments. Compared with *ab initio* or density functional methods, semiempirical calculations are much faster, typically by several orders of

magnitude,[57] but they are also less accurate, with errors that are less systematic and thus harder to correct.

Quantum-chemical semiempirical treatments are defined by the following specifications:

(a) The underlying theoretical approach: Most current general-purpose semiempirical methods are based on molecular obital theory and employ a minimal basis set for the valence electrons. Electron correlation is treated explicitly only if this is necessary for an appropriate zero-order description.

(b) The integral approximation and the types of interactions included: Traditionally there are three levels of integral approximation[58,59] {CNDO (complete neglect of differential overlap), INDO (intermediate neglect of differential overlap), and NDDO (neglect of diatomic differential overlap)}. DDO is the best of these approximations since it retains the higher multipoles of charge distributions in the two-center interactions (unlike CNDO and INDO which truncate after the monopole).

The semi-empirical methods are based on the Hartree-Fock approach. A Fock-matrix is constructed and the Hartree-Fock equations are solved iteratively. The approximations are in the construction of the Fock matrix, in other words in the energy expressions. Another common feature of semi-empirical methods is that they only consider the valence electrons. The core electrons are accounted for in a core-core repulsion function, together with the nuclear repulsion energy.

The next step is to replace many of the remaining integrals by parameters, which can either have fixed values, or depend on the distance between the atoms on which the basis functions are located. At this stage empirical parameters can be introduced, which can be derived from measured properties of atoms or diatomic molecules. In the modern semi-empirical methods the parameters are however mostly devoid of this physical significance: they are just optimized to give the best fit of the computed molecular properties to experimental data.[60] Different semi-empirical methods differ in the details of the approximations (e.g. the core-core repulsion functions) and in particular in the values of the parameters. In contrast to molecular mechanics, only parameters for single atoms and for atom pairs are needed.

The semi-empirical methods can be optimized for different purposes. The MNDO, AM1 and PM3 methods were designed to reproduce heats of formation and structures of a large number of organic molecules. Other semi-empirical methods are specifically optimized for spectroscopy, e.g. INDO/S or CNDO/S, which involve CI calculations and are quite good at prediction of electronic transitions in the UV/VIS spectral region.[61,62]

Some even more approximate methods are still quite useful. In the Hückel and extended Hückel methods the whole sum over two-electron integrals is replaced by a single diatomic parameter (the resonance integral), so that no search for a self-consistent field is necessary (nor possible). These methods have proven extremely valuable in qualitative

and semi-quantitative MO theories of π-electron systems and of organometallic systems.[63]

For π-electron systems, ZDO treatments have been developed that take only π-centers (p-atomic orbitals) into account, in SCF calculation. An example is the Pariser-Parr-Pople method, which involves a CI calculation as well. This method has been very successfully used to predict the optical absorption spectra of conjugated organic molecules.

2.10 Hybrid quantal / classical models[26,64,65]

Combining the strength of both molecular mechanics and quantum mechanics, the QM/MM approach is able to model a very large compound using molecular mechanics and one crucial section of the molecule with quantum mechanics. This is designed to give results that have very good speed where only one region needs to be modeled quantum mechanically and can also be used to model a molecule surrounded by solvent molecules.

Quantitative energy values are one of the most useful results from computational techniques. In order to have a reasonable energy expression when two calculations are combined, it is necessary to know not only the energy of the two regions, but also the energy of interaction between those regions. There have been a number of energy computation schemes proposed. Most of these schemes can be expressed generally as

$$E = E_{QM} + E_{MM} + E_{QM/MM} + E_{pol} + E_{boundary} \qquad (2.58)$$

The first two terms are the energies of the individual computations. The $E_{QM/MM}$ term is the energy of interaction between these regions, assuming that both regions remain fixed. It may include van der Waals terms, electrostatic interactions, or any term in the force field being used. E_{pol} is the effect of either region changing as a result of the presence of the other region, such as electron density polarization or solvent reorganization. $E_{boundary}$ is a way of representing the effect of the rest of the surroundings, such as the bulk solvent.

The van der Waals term is used for describing non-bonded interactions between atoms in the two regions. This is usually represented by a Leonard-Jones 6-12 potential of the form

$$E_{vdW} = A/r^{12} - B/r^{6} \tag{2.59}$$

The parameters A and B are those from the force field being used.

The other term that is very widely used in $E_{QM/MM}$ is a Coulombic charge interaction of the form

$$E_{Coulomb}=Q_iQ_j/r_{ij} \tag{2.60}$$

The subscripts *i* and *j* denote two nuclei, one in the QM region and one in the MM region. The atomic charges for the MM atoms are obtained by any of the techniques commonly used in MM calculations. The atomic charges for the QM atoms can be obtained by a population analysis scheme. Alternatively, there might be a sum of interactions with the QM nuclear charges plus the interaction with the electron density, which is an integral over the electron density.

If the QM and MM regions are parts of the same molecule, having non-bonded interactions is no longer sufficient and it is necessary to describe the bond connecting the two sections. In most cases, this is done using the bonding terms in the MM method being used. This is usually done by keeping every bond, angle or torsion term that incorporates one atom from the QM region.

The above non-bonded energy terms assume that the electron density in each region is fixed. This can be a reasonable approximation for covalent systems, but a poor approximation where the QM region is being stabilized by its environment, as is the case with polar solvent effects. Polarization is usually accounted for by computing the interaction between dipoles included.

The arbitrary separation of QM and MM regions will lead to another problem: it will create atoms with nonfilled valences and in the QM region an atom with a nonfilled valence will behave differently than one with a filled valence. Because of this, one must consider the way in which the quantum mechanical portion of the calculation is truncated. One popular solution is to satisfy the valence with hydrogen or pseudo-halide atoms. These pseudo-halide atoms are parameterized to mimic the behaviour of the MM atom it is substituted for. These extra atoms are called "link atoms" or "junction dummy atoms". The link atoms are not included in the energy expression used to describe the interaction between the regions of the system.

In practice, the source of the greatest instability in the link atom approach is the strong interactions that can develop between the wave function in the regions near the link atom and nearby MM atoms carrying partial atomic charges. This suggests that a worthwhile approach would be to have some sort of buffer layer between the polarizable QM region and the point-charge-represented MM region, i.e. one that is not optimized as part of an SCF procedure. The full system is now partitioned into three regions, which may be called the MM region, the auxiliary region, and the QM region. The new auxiliary region is characterized by nuclei having their normal nuclear charges, and the electron density expressed in terms of some basis functions.

2.11 Molecular dynamics[26,66-69]

In the broadest sense, molecular dynamics is concerned with molecular motion, which is inherent to all chemical processes. This computational method calculates the time dependent behaviour of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. They are also used in the determination of structures from x-ray crystallography and from NMR experiments.

Molecular dynamics simulations generate information at the microscopic level, including atomic positions and velocities. The conversion of this microscopic information to macroscopic observables such as pressure, energy, heat capacities, etc., requires

statistical mechanics, which provides the rigorous mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of the N-body system; molecular dynamics simulations provide the means to solve the equation of motion of the particles and evaluate these mathematical formulas. With molecular dynamics simulations, one can study both thermodynamic properties and/or time dependent (kinetic) phenomena.

In statistical mechanics, a collection of all possible systems that have different microscopic states but have an identical macroscopic or thermodynamic state is called an ensemble. An experiment is usually made on a macroscopic sample that contains an extremely large number of atoms or molecules sampling an enormous number of conformations. In statistical mechanics, averages corresponding to experimental observables are defined in terms of ensemble averages.

$$\langle A \rangle_{enzemble} = \iint dp^N dr^N A\left(p^N, r^N\right) \rho\left(p^N, r^N\right) \tag{2.61}$$

where $A\left(p^N, r^N\right)$ is the observable of interest and it is expressed as a function of the momenta, p, and the positions, r, of the system.

However, ensemble averages are generally extremely difficult to calculate since one must calculate all possible states of the system. According to the ergodic hypothesis, the time average equals the ensemble average:

$$\langle A \rangle_{enzemble} = \langle A \rangle_{time} \tag{2.62}$$

The basic idea is that if one allows the system to evolve in time indefinitely, that system will eventually pass through all possible states.

The molecular dynamics simulation method is based on Newton's second law or the equation of motion. From a knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past. Molecular dynamics simulations can be time consuming and computationally expensive.

The potential energy is a function of the atomic positions (3N) of all the atoms in the system. Due to the complicated nature of this function, there is no analytical solution to the equations of motion; they must be solved numerically. Numerous numerical algorithms have been developed for integrating the equations of motion. All the integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion:

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \ldots$$

$$v(t+\delta t) = v(t) + a(t)\delta t + \frac{1}{2}b(t)\delta t^2 + \ldots$$

$$a(t+\delta t) = a(t) + b(t)\delta t + \ldots$$

where r is the position, v is the velocity (the first derivative with respect to time), a is the acceleration (the second derivative with respect to time), etc.

Systems studied via MD are typically divided into the following categories: empirical potentials, semi-empirical potentials, and *ab initio* methods.

Empirical potentials either ignore quantum mechanical effects, or attempt to capture quantum effects in a limited way through entirely empirical equations. Parameters in the potential are fitted against known physical properties of the atoms being simulated, such as elastic constants and lattice parameters.

Semi-empirical potentials make use of the matrix representation from quantum mechanics. However, the values of the matrix elements are found through empirical formulae that estimate the degree of overlap of specific atomic orbitals. The matrix is then diagonalized to determine the occupancy of the different atomic orbitals, and empirical formulae are used once again to determine the energy contributions of the orbitals.

Ab initio methods use full quantum-mechanical formulas to calculate the potential energy of a system of atoms or molecules. This calculation is usually made "locally", i.e., for

nuclei in the close neighborhood of the reaction coordinate. Although various approximations may be used, these are based on theoretical considerations, not on empirical fitting. *Ab initio* methods produce a large amount of information that is not available from the empirical methods, such as density of states information. Perhaps the most famous *ab initio* package is the Car-Parrinello molecular dynamics (CPMD) package based on density functional theory.

Chapter 3

Modelling Competitive Reaction Mechanisms of Peroxynitrite Oxidation of Guanine

3.1 Introduction

Oxidative damage to DNA is responsible for aging and many diseases including cancers.[70] Among the nucleic acid bases in DNA, guanine has the lowest oxidation potential (1.29 V vs NHE)[71,72] and is therefore the easiest to oxidize to form the guanine radical cation ($G^{\bullet+}$). In the presence of ionizing radiation, photosensitizers or reactive oxygen species (ROS),[71] the electronic hole introduced into duplex DNA shows a strong tendency to migrate from the site of ionization to the guanine base of 5'-guanine of the 5'-GG-3' doublets or the more oxidatively sensitive sites 5'-GGG-3' triplets through DNA π-stacking.[73-75] Guanine radicals can further react with mild reducing agents, such as some redox active amino acids or small peptides,[76-10] or with H_2O and/or O_2, yielding 8-oxo-7,8-dihydroguanine (8-oxoGua) as the most common product.[77] The reactions involved in this irreversible lesion are the hydration of guanine followed by one-electron oxidation.

The guanine radical cation, $G^{\bullet+}$, has an even lower oxidation potential (0.53 V vs NHE[78]) than its guanine precursor and therefore could be easily oxidized to give secondary oxidation products,[79-81] Another two-electron oxidation product of guanine is

2,6-diamino-4-hydroxy-5-formamidopyrimidine (FAPYGua),[82] a minor derivative compared to 8-oxoguanine. Certain DNA-repairing enzymes are able to recognize those lesions and catalyze the cleavage of DNA at the sites of damage.[83]

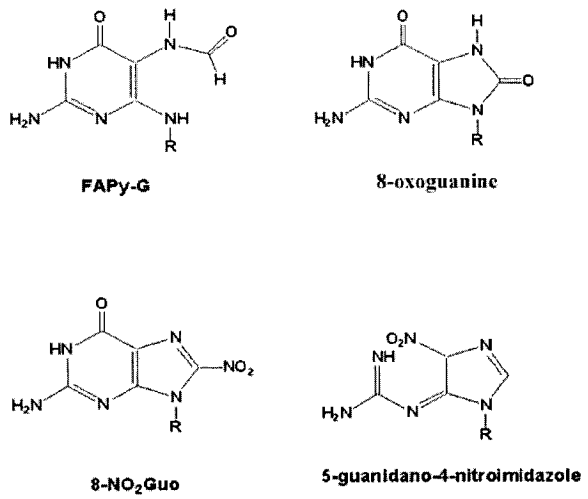


Figure 3.1. Four significant products from guanine oxidation reactions

$G^{\bullet+}$ is a much stronger acid (pK_a=3.9) than guanine itself (pK_a=9.4).[84] Once formed, $G^{\bullet+}$ undergoes rapid deprotonation to generate the neutral radical denoted by **$G^{\bullet}$(-H)**.[85] In a DNA duplex, the proton will be transferred to N3 of cytosine of the GC base pair. Since the pK_a (4.3) of N3-protonated cytosine is only slightly higher than that of $G^{\bullet+}$, $G^{\bullet+}$ formed in DNA duplex should retain more cationic character than would an isolated G nucleoside or nucleotide. The guanine radical cation would persist long enough to react with oxidizing or nitrating reagents.

Various species, such as free hydroxyl radical, sulfate radical, etc. may cause oxidative damage to DNA.[2] Recently, DNA damage caused by reactive nitrogen species has been drawing interest.[86] In inflamed tissues, NO and $O_2^{\bullet-}$ are produced by active phagocyte.

These reactive species interact with each other, resulting in the formation of a highly reactive species, peroxynitrite ($ONOO^-$)[87] at almost the diffusion-controlled rate.[88,89]

$$NO^{\bullet} + O_2^{\bullet -} \rightarrow ONOO^- \qquad k \approx 7 \times 10^9 M^{-1} s^{-1}$$

Although it remains a daunting challenge to trap or assay peroxynitrite because of its high reactivity, there is evidence supporting the notion that peroxynitrite is formed in *vivo*.[90-92] The formation of excess $ONOO^-$ has been implicated in many pathological conditions including reperfusion injury, chronic inflammation, arteriosclerosis, rheumatoid arthritis, neurodegenerative disease, and an increased cancer incidence in tissues under chronic inflammatory conditions.

Peroxynitrite ($ONOO^-$) is a powerful oxidizing and nitrating agent that exhibits high reactivity at physiological pH.[93] It is fairly unreactive and its solution in alkali is stable for weeks if kept frozen. The reactivity of $ONOO^-$ is attributed to the formation of vibrationally high-energy intermediate (HOONO)[93] generated during the breakdown of peroxynitrious acid, ONOOH (pK_a=6.8) and/or the $OH^{\bullet}/NO_2^{\bullet}$ radical pair.[94] At physiological pH, $ONOO^-$ is rapidly protonated to the peroxynitrite acid and rearranges, forming nitrate as the major end-product.[88] In the presence of carbon dioxide, a different reactive species, nitrosoperoxycarbonate anion ($ONO_2CO_2^-$), is formed. $ONO_2CO_2^-$ is postulated to be the reactive intermediate in $ONOO^-$ reactions in the presence of CO_2, with proposed formation of ($CO_3^{\bullet -}$) and nitrogen dioxide radical ($NO_2^{\bullet}$), which are believed to be responsible for the oxidation and nitration of DNA when it is exposed to $ONOO^-$.[95-97] It has been shown that peroxynitrite and neutral peroxynitrious acid might

travel through cell membranes by diffusion or through anion channels.[98,99] Addition of $ONOO^-$ to aromatic compounds leads to both hydroxylation and nitration.

One-electron oxidation products of guanine and $ONOO^-$ include 8-oxoGua, 8-nitroguanine (8-NO_2Gua, denoted **8**)[100] and 5-guanidano-4-nitroimidazole (denoted **6**).[101] Unlike the first two intermediates that readily undergo further oxidation, the last one is a stable and significant reaction product formed only through peroxynitrite-related chemistry. Thus, 5-guanidano-4-nitroimidazole may be meaningful for understanding the DNA damage induced by peroxynitrite.

To describe the mechanisms of the reactions, two competitive reaction pathways (Figure 3.2) have been proposed for the formation of 5-guanidano-4-nitroimidazole and 8-NO_2Gua.[101] In the first one, $NO_2^\bullet$ combines with the guanine radical cation at the C5 (Figure 3.2 a) or C8 position (Figure 3.2 b) of the purine ring. Subsequent reactions lead to the formation of the final products 5-guanidano-4-nitroimidazole and 8-NO_2Gua. At the beginning of the second reaction pathway, the guanine radical cation, $G^{\bullet+}$, deprotonates to yield the neutral radical $\mathbf{G^\bullet(-H)}$; and then a radical combination reaction takes place between $\mathbf{G^\bullet(-H)}$ and $NO_2^\bullet$. The following reaction steps are analogous to those in the guanine radical cation reaction pathway. Intermediates in both reaction mechanisms are labeled **a** through **h,** as is shown in Figure 3.2.

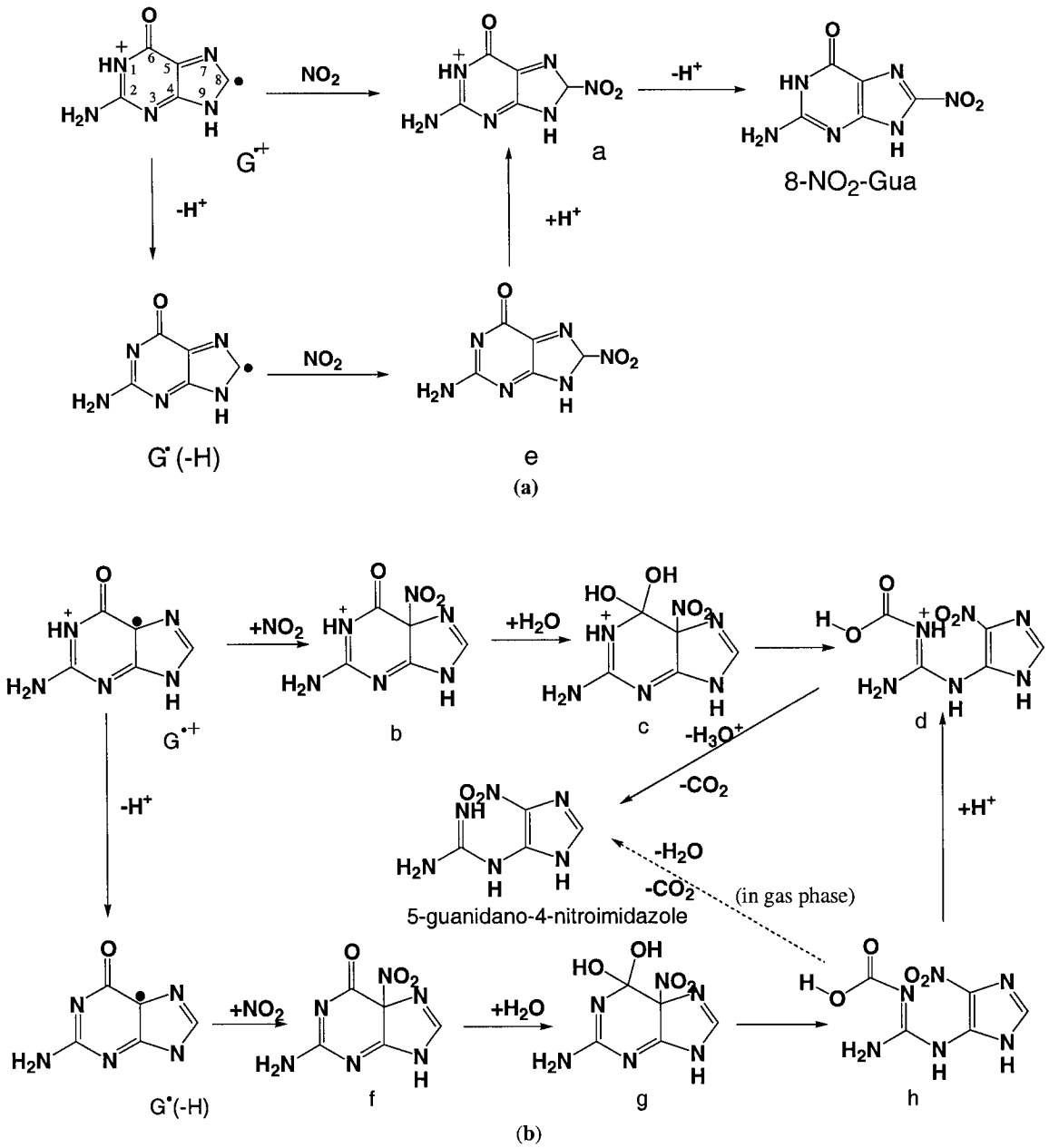


Figure 3.2. Reaction mechanisms leading to the formation of (**a**) 8-NO_2Gua and (**b**) 5-guanidano-4-nitroimidazole (the dashed line in the figure represents direct formation of 5-guanidano-4-nitroimidazole from intermediate **h**, a reaction step that only takes place in gas phase)

This chapter describes a computational investigation of the guanine oxidatively generated lesion introduced by peroxynitrite. Comparison of the two proposed reaction

mechanisms that lead to the formation of two significant products (8-NO_2Gua and 5-guanidano-4-nitroimidazole) indicates that the guanine radical cation mechanism is more favorable than the neutral guanine radical mechanism. The important role of a water molecule in all reaction steps is also presented.

3.2 Computational methods

All geometry optimizations and frequency calculations were performed with the B3LYP hybrid density functional in conjunction with the 6-31G(d,p) basis set using the Gaussian 03 suite of programs. Single point calculations were performed with B3LYP level of theory and Pople's 6-311++G(2df,2p) basis set using the above geometries. The Gibbs free energy correction and zero point vibrational energy correction were derived from the frequency calculation and added to the single point electronic energies to obtain relative energies for all reactants, products, intermediates and transition states. The gas phase results were derived by including the zero-point vibrational energy to the single point calculation, i.e., B3LYP/6-311++G(2df,2P)//B3LYP/6-31G(d,p) +ZPVE. Intrinsic reaction path (IRC) calculations were performed on every transition state to confirm that the transition state connects the minima of interest.

The PCM and CPCM solvation models were used initially to treat solvent effects in this work. However, due to convergence problems in the optimization of several structures, the more robust Onsager solvation model[102] was utilized in all geometry optimization and frequency calculations. Convergence difficulties with the PCM and CPCM solvation models are well-known (see, for example, the work of Vallet et al.[103]). All cavity sizes,

including that of H_3O^+, were taken to be the default values assumed by Gaussian 03. All energies are in kJ/mol.

3.3 Results and Discussion

3.3.1 Calculation results in gas phase

The formation of 8-nitroguanine

In the guanine radical cation pathway, the reaction is initiated by the radical combination reaction of the guanine radical cation and NO_2 at the C8 position of the purine ring, while in the neutral guanine radical reaction mechanism, the neutral guanine radical is formed as a result of the deprotonation reaction that takes place at the N4 position of the guanine radical cation.

a. Guanine radical cation reaction mechanism

a→8-nitroguanine

The guanine radical cation, formed by the one-electron oxidation of guanine has significant unpaired electron density[104] at the O6, C5, and C8 positions. When the NO_2 radical is available, it may immediately combine with the guanine radical cation at the C8 or C5 position to form 8-nitroguanine or 5-guanidino-4-nitroimidazole. As indicated in Figure 3.3, intermediate **a** is formed as the NO_2 radical combines with the guanine radical cation at the C8 position. In aqueous solution, deprotonation takes place at the C8 position and leads to the formation of 8-nitroguanine. The energy of complex **a**...H_2O lies lower than that of **a**+H_2O by –76.2 kJ/mol. A transition state **TSa-8** is found with a relative energy of 3.6 kJ/mol. A complex **8...**H_3O^+ is derived after the transition state

with a relative energy of 67.0 kJ/mol. The relative energy of 8-NO_2Gua + H_2O is –143.8 kJ/mol.

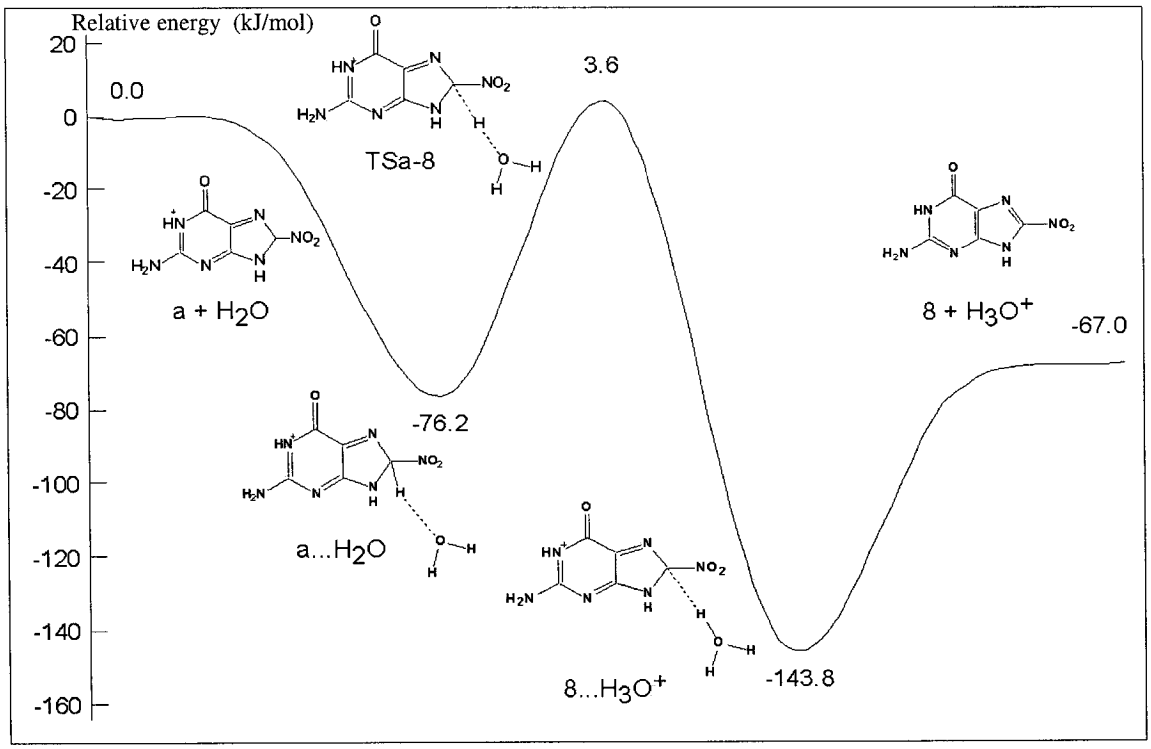


Figure 3.3. Schematic energy profile at 0 K in gas phase for the route **a**→8-NO_2Gua with one water involved in the deprotonation process.

b. Neutral guanine radical reaction mechanism

$G^{\bullet+}$→ $G^{\bullet}$(-H)

As mentioned previously, the guanine radical cation is a relatively strong acid and will readily lose the proton at N1 in the purine ring to an available proton acceptor. This reaction yields the neutral guanine radical that initiates a series of reactions leading to the formation of the final products 5-guanidino-4-nitroimidazole and 8-NO_2Gua. In this deprotonation process (Figure 3.4), the transition state **TSG** and the reactant complex **G**...H_2O have energies –107.3 kJ/mol and –165.8 kJ/mol relative to the energy of the

isolated system of H_2O and guanine radical cation. A complete transfer of the proton to H_2O leads to a relative energy of 44.0 kJ/mol for H_3O^+ + **G$^{\bullet}$(-H)**.

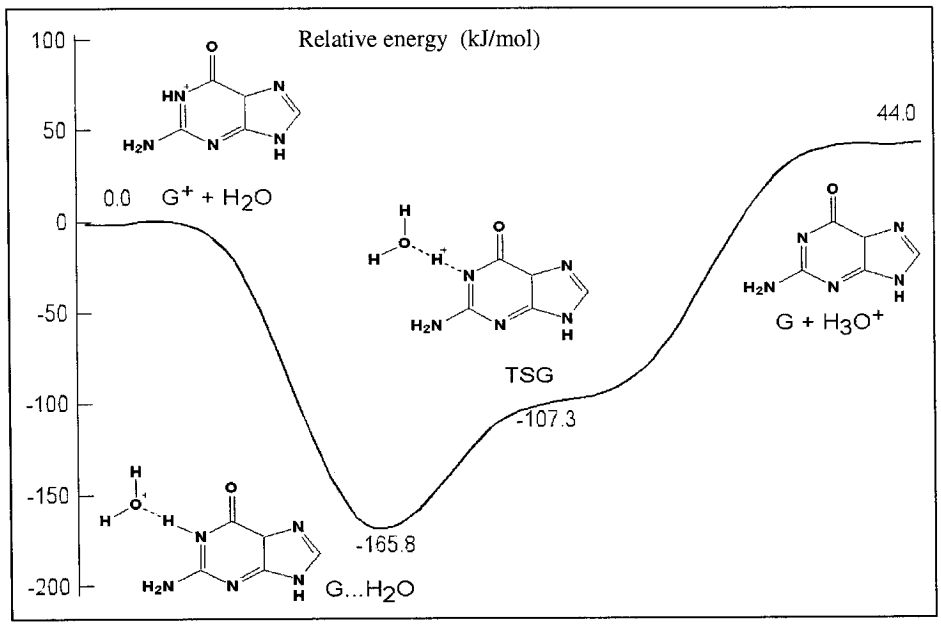


Figure 3.4. Schematic energy profile at 0 K in gas phase for the route **G$^{\bullet+}$→ G$^{\bullet}$(-H)** with one water molecule involved (**G$^{\bullet}$(-H)** has significant unpaired electron density at the O6, C5, and C8 positions[104] and therefore NO_2 may combine with **G$^{\bullet}$(-H)** at either C5 or C8).

e→a

Intermediate **e** is derived once the nitrogen dioxide radical combines with the guanine radical at the C8 position. Unlike intermediate **a** in the guanine radical cation mechanism, **e** is not able to form 8-NO_2Gua in a single-step reaction because one proton has to be removed from the C8 position and another proton attached to N1. It is not likely that these two events could be involved in a single-step reaction due to the planar structure of the compound and the distance between these two positions. The present results show that after a proton is attached to give intermediate **a,** the reaction follows the same pathway as is described in the radical cation reaction mechanism to yield 8-NO_2Gua

(Figure 3.5). Relative to the isolated **e** + H_3O^+ system, the transition state **TSe-a** has an energy of –15.3 kJ/mol. The relative free energies of the complexes **e**...H_3O^+ and **a**...H_2O are –79.0 kJ/mol and –55.1 kJ/mol, respectively. The isolated system **a** + H_2O has a relative energy of 16.0 kJ/mol.

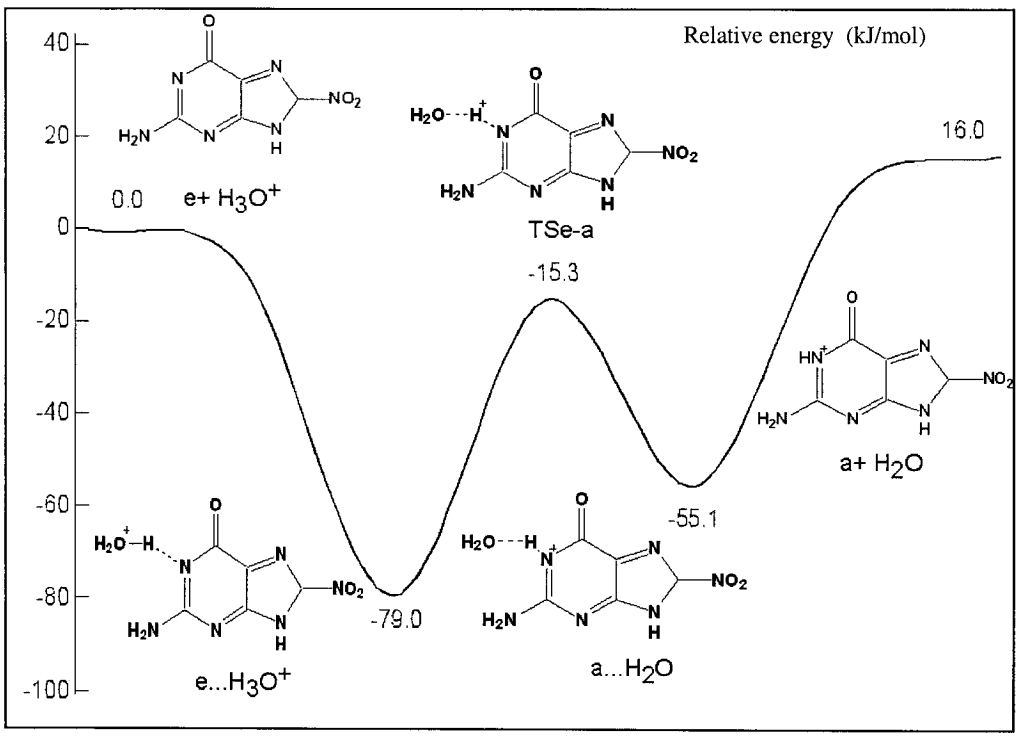


Figure 3.5. Schematic free energy profile at 0 K in gas phase for the route **e**→**a** with one water involved

The formation of 5-guanidino-4-nitroimidazole

a. Guanine radical cation reaction mechanism

b→c

In the case of C5 addition (Figure 3.6), intermediate **b** is formed, which can be attacked by water at the C6 position to yield **c**. In this process, two solvent water molecules are

involved: one water is connected to the electrophilic C6 position via its oxygen atom while the other water molecule serves as a water bridge to facilitate the proton transfer from the first water molecule to the oxygen originally connected to the C6 position. Therefore, a six-membered ring is formed at transition state **TSb-c**. Complex **b**...$2H_2O$ has an energy of –122.4 kJ/mol relative to that of **b**+$2H_2O$; the relative energy of **TSb-c** is –68.9 kJ/mol; the complex formed after the transition state, **c**...H_2O, has a relative energy of –146.0 kJ/mol; the relative energy of separated system **c**+H_2O is –71.0 kJ/mol.

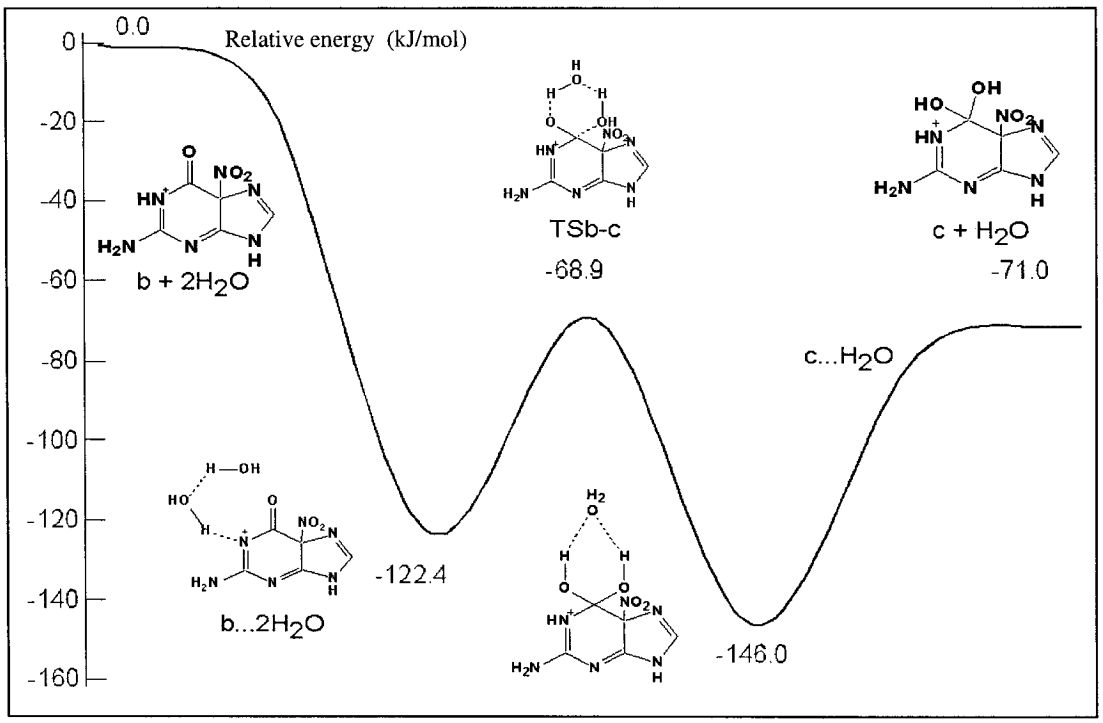


Figure 3.6. Schematic free energy profile at 0 K in gas phase for the route **b**→**c** with two waters involved in this process.

c→d

In the next reaction step (Figure 3.7), an intramolecular proton transfer takes place. The proton is removed from the hydroxyl group that is connected to the C6 position and attached to the N3 position. As shown in Figure 3.7, the proton transfer is an

intramolecular process, which shifts the proton from the hydroxyl group at the C6 position directly to the N3 position with the help of a water bridge. The cleavage of the C5-C6 bond happens simultaneously. Compared to the isolated group **c**+H_2O, the relative energies of the complexes **c**...H_2O and **d**...H_2O are –84.8 and –53.4 kJ/mol, respectively; transition state **TSc-d** has a relative energy of –11.9 kJ/mol, while the relative energy of the product group **d**+H_2O is –66.0 kJ/mol.

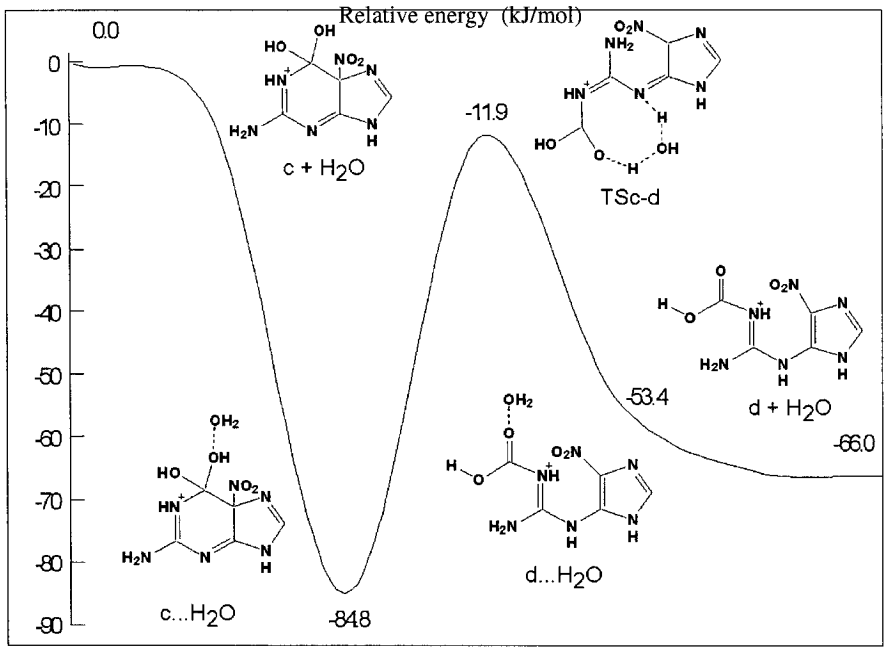


Figure 3.7. Schematic free energy profile at 0 K in gas phase for the route **c**→**d** with a water molecule involved in this process.

d→5-guanidino-4-nitroimidazole

The final product 5-guanidino-4-nitroimidazole is derived through decarboxylation of intermediate **d** (Figure 3.8). A water molecule acts as a proton accepter and helps remove the proton from the carboxyl group of the carbamate **d**. Meanwhile, the C-N bond breaks and a CO_2 molecule is released. In this manner, intermediate **d** is decomposed into: H_3O^+, CO_2, and 5-guanidino-4-nitroimidazole without the formation of any complex after the

transition state **TSd-6**. Relative to the **d** + H_2O system, the complex **d**...H_2O has an energy of –92.0 kJ/mol, the transition state has a relatively higher energy of –11.0 kJ/mol and the product 5-guanidino-4-nitroimidazole + CO_2 + H_3O^+ are much lower at –279.0 kJ/mol.

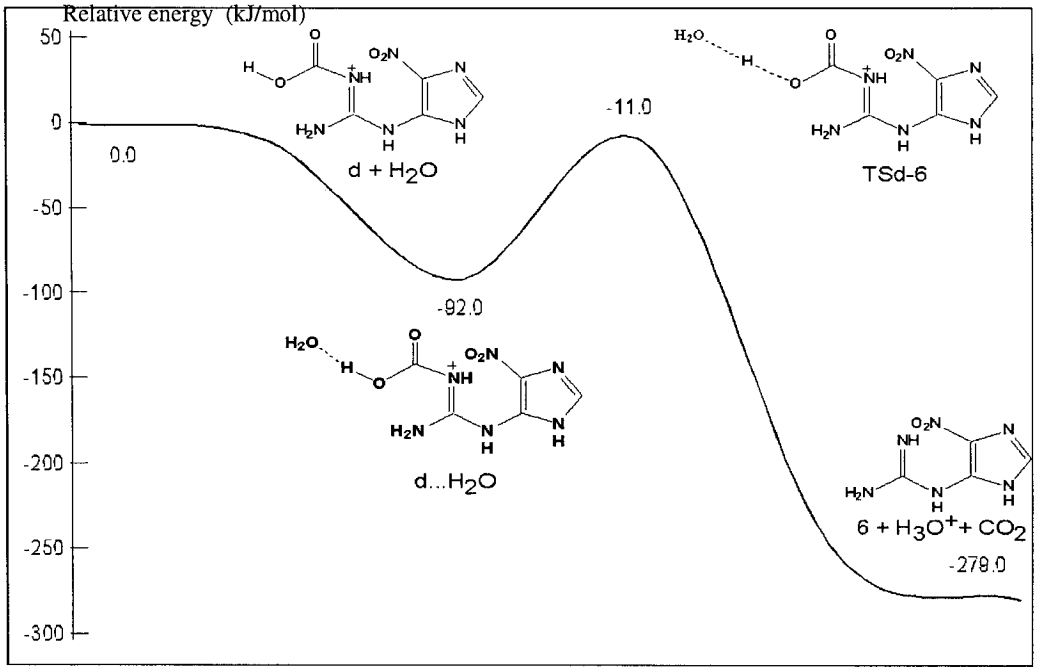


Figure 3.8. Schematic free energy profile at 0 K in gas phase for the route **d**→5-guanidino-4-nitroimidazole (**6**) with one water molecule involved.

b. Neutral guanine radical reaction mechanism

f→g

Analogous to what happens in the radical cation reaction mechanism, the nitrogen dioxide radical may also combine with the **G$^{\bullet}$(-H)** at the C5 position and form an intermediate **f**, which can be attacked by a water molecule at the electrophilic C6 position to yield **g**. In the transition state (Figure 3.9), a six-membered ring is formed with the help of a water-bridge to facilitate the proton transfer. Total energy changes relative to the isolated system of **f**+2H_2O are as follow: the complex **f**...2H_2O has an energy of –

59.0 kJ/mol, the transition state **TSf-g** has a relative energy of 104.6 kJ/mol; the relative energies of complex **g**...H_2O and **g** + H_2O are –74.0 kJ/mol and –24.0 kJ/mol, respectively.

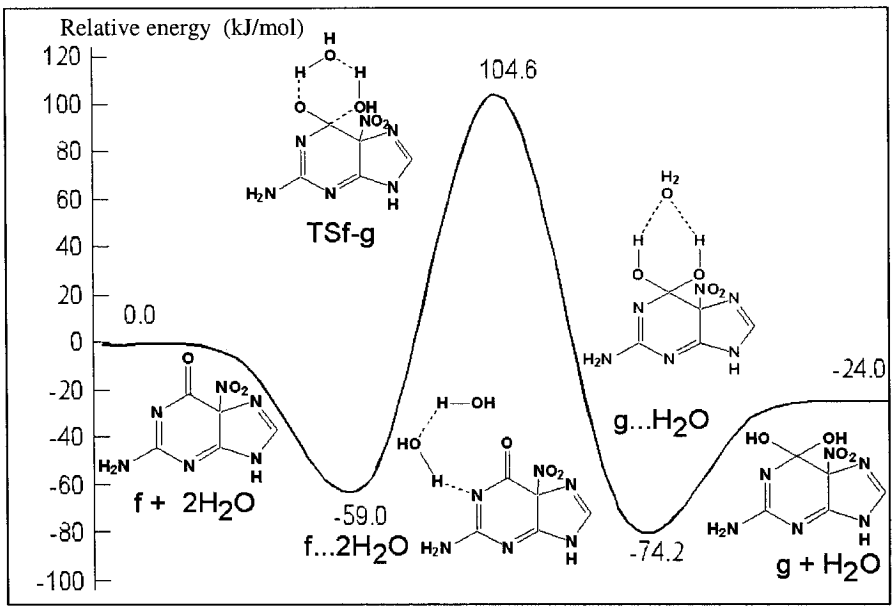


Figure 3.9. Schematic free energy profile at 0 K in gas phase for the route **f**→**g** with two waters involved.

g→h

Subsequently, an intramolecular proton transfer takes place in the newly formed intermediate **g**. A water molecule is involved as well to form the water bridge across the purine ring plane (Figure 3.10). This process is accompanied by the cleavage of the C5-C6 bond and followed by the formation of the carbomate derivative **h**. Relative to **g** +H_2O, complex **g**...H_2O has an energy of –123.9 kJ/mol and transition state **TSg-h** has a relative energy of –96.4 kJ/mol; and the relative energies of complexes **h**...$2H_2O$ and **h**+$2H_2O$ are –187.0 kJ/mol and –172.0 kJ/mol, respectively.

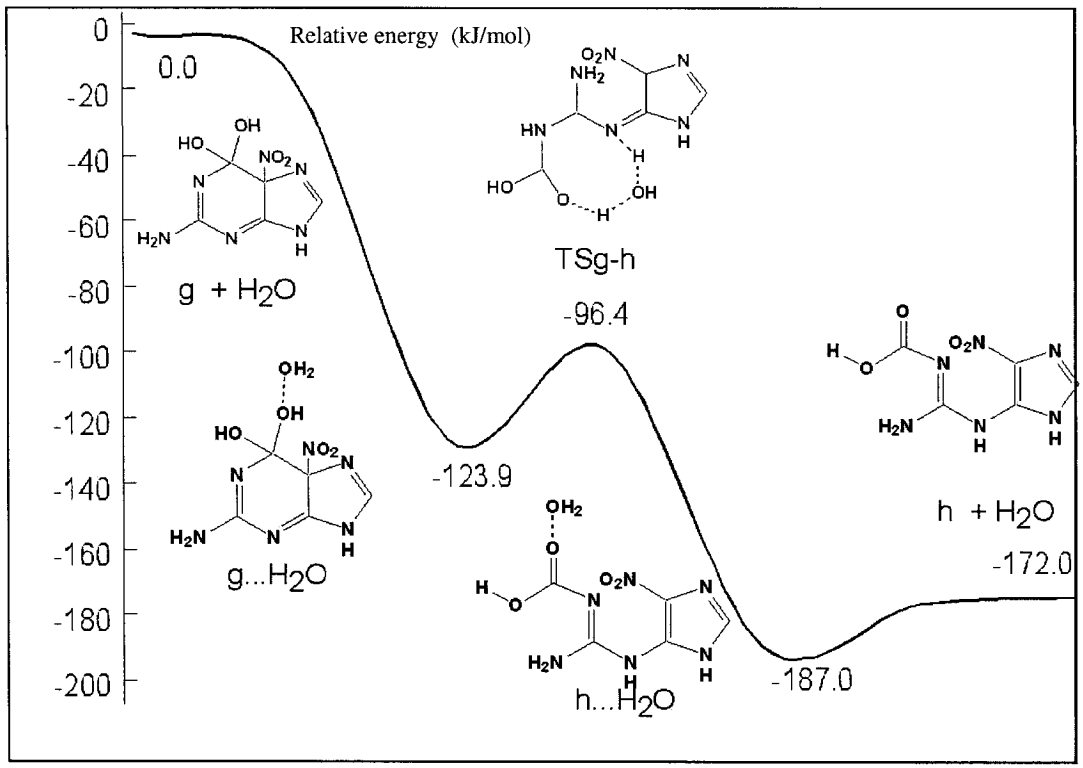


Figure 3.10. Schematic free energy profile at 0 K in gas phase for the route **g**→**h** with one water involved

h→5-guanidino-4-nitroimidazole

In the last step towards the formation of final product 5-guanidino-4-nitroimidazole in the neutral guanine radical reaction mechanism (Figure 3.11), a hydrogen is transferred from the carboxylate –OOH to N1 position and the C–N bond breaks. A complex **h**...H_2O is formed before the transition state, in which a water connects the hydroxyl group and nitrogen through hydrogen bonding. In the transition state **TSh-6**, a six-membered ring is formed to facilitate the intra-molecular proton transfer. Another complex **6**...(CO_2 + H_2O) is formed after the transition state. Finally, the compound decomposes into three parts: 5-guanidino-4-nitroimidazol, a water and a carbon dioxide. Complexes **h**...H_2 and

6...(CO_2+ H_2O) are lower than the isolated **h** + H_2O by –54.3 and –61.8 kJ/mol, respectively. The transition state has a relative energy of –27.7 kJ/mol and that of 5-guanidino-4-nitroimidazol + H_2O + CO_2 is –201.0 kJ/mol.

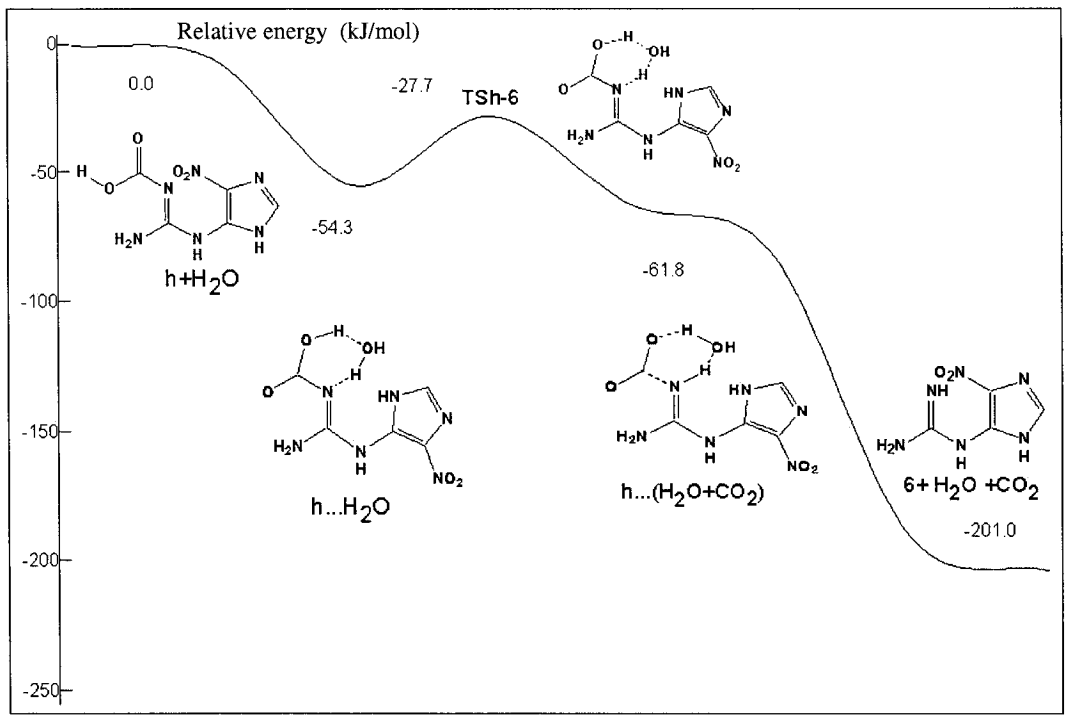


Figure 3.11. Schematic free energy profile at 0 K in gas phase for the route **h**→5-guanidino-4-nitroimidazol with one H_3O^+ involved.

3.3.2 Computation results in solution

The formation of 8-nitroguanine

In the guanine radical cation pathway, the reaction is initiated by the radical combination reaction of the guanine radical cation and NO_2 at the C8 position of the purine ring, while in the neutral guanine radical reaction mechanism, the neutral guanine radical is formed as a result of the deprotonation reaction that takes place at the N4 position of the guanine radical cation.

a. Guanine radical cation reaction mechanism

a→8-nitroguanine

As described in the gas phase calculation results, a deprotonation reaction takes place at the C8 position with one water molecule acting as a proton accepter, leading to the formation of **8-NO_2Gua.** Two complexes **a...H_2O** and **8...H_2O** are present in this reaction step.

For the ZPVE-only corrected case (Figure 3.12), the energy of **a...H_2O** lies lower than that of **a**+H_2O by 43.8 kJ/mol. A transition state **TSa-8** is found with a relative energy of 85.6 kJ/mol. A complex **8...**H_2O is derived after the transition state and followed by the isolated system 8-NO_2Gua + H_2O and their relative energies are –55.5 kJ/mol and –135.4 kJ/mol, respectively.

For the free energy changes at 298.15 K (Figure 3.13), complexes **a**...H_2O and **8**...H_2O are lower than **a** + H_2O by –7.4 and –16.3 kJ/mol, respectively. Transition state **TSa-8** has a relative energy of 126.5 kJ/mol. The isolated **8** + H_3O^+ lies 109.1 kJ/mol below **a** + H_2O. Both complexes are raised by about 40 kJ/mol compared with the ZPVE-only corrected case, while the energy barrier for this reaction step is not changed much.

b. Neutral guanine radical reaction mechanism

$G^{\bullet+}$→ $G^{\bullet}$(-H)

This reaction yields the neutral guanine radical that initiates a series of reactions leading to the formation of the final products 5-guanidano-4-nitroimidazole and 8-NO_2Gua. One complex **G**...H_2O is formed before the transition state **TSG**.

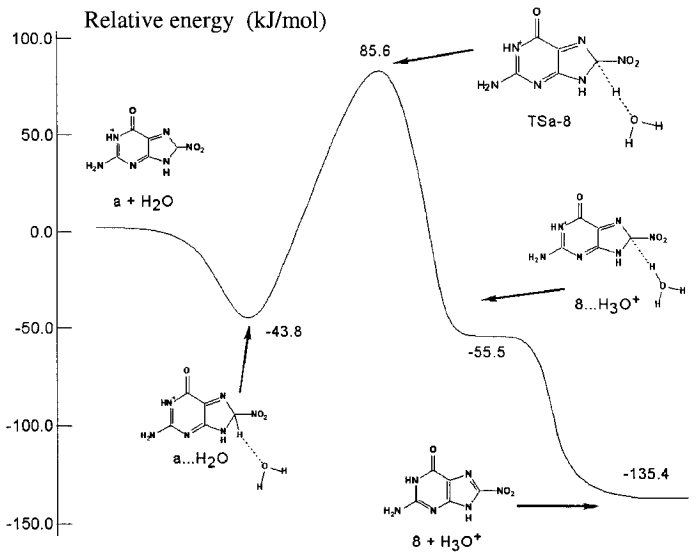


Figure 3.12. Schematic energy profile at 0 K for the route **a→8-NO_2Gua** with one water involved in the deprotonation process.

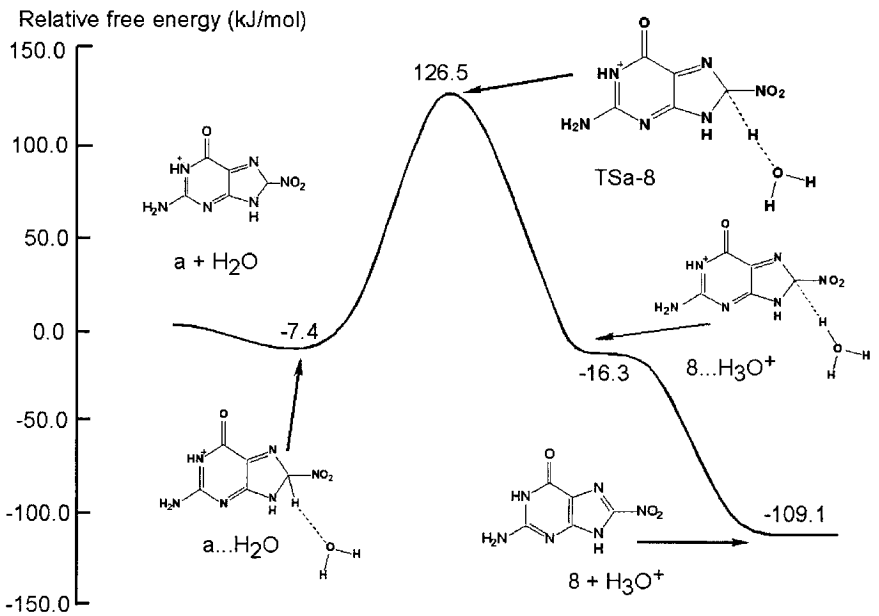


Figure 3.13. Schematic free energy profile at 298.15 K for the route **a→8-NO_2Gua** with one water involved in the deprotonation process.

For the ZPVE-only corrected case (Figure 3.14), the reactant complex **G...H_2O** and the transition state **TSG** have energies –43.4 kJ/mol and 23.9 kJ/mol, respectively, relative to the energy of the isolated system of H_2O + $\mathbf{G^{\bullet+}}$. The relative energy of H_3O^+ + **$G^{\bullet}$(-H)** is -19.2 kJ/mol. The energy barrier of this reaction step is 67.3 kJ/mol.

For the entropy-corrected case at 298.15 K (Figure 3.15), **TSG** has a relative energy of 65.8 kJ/mol compared with the isolated system H_2O + $\mathbf{G^{\bullet+}}$. **G...H_3O^+** lies 6.4 kJ/mol lower than H_2O + $\mathbf{G^{\bullet+}}$. H_3O^+ + **$G^{\bullet}$(-H)** has a relative energy of –19.2 kJ/mol, which is 60 kJ/mol higher than in the ZPVE-only corrected case (Figure 3.14), whereas the energy barrier is not changed much.

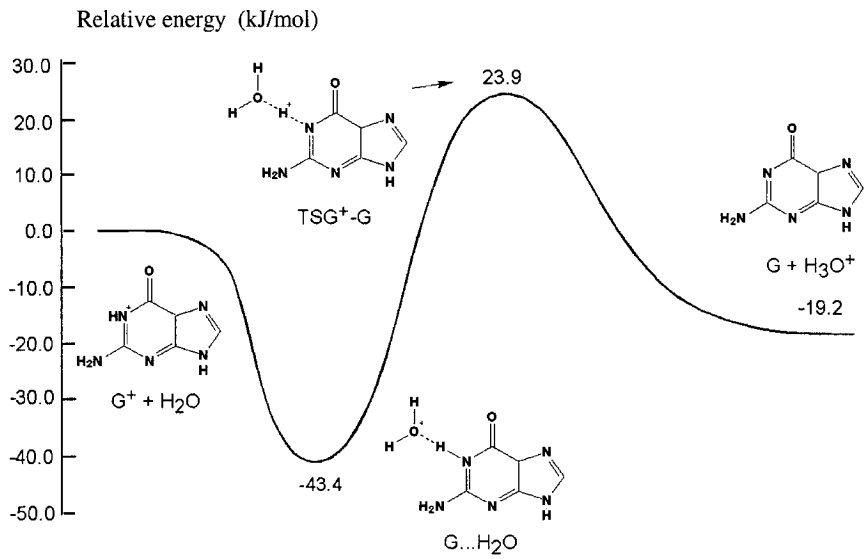


Figure 3.14. Schematic energy profile at 0 K for the route $\mathbf{G^{\bullet+}}$→ **$G^{\bullet}$(-H)** with one water molecule involved (**$G^{\bullet}$(-H)** has significant unpaired electron density at the O6, C5, and C8 positions[104] and therefore NO_2 may combine with **$G^{\bullet}$(-H)** at either C5 or C8).

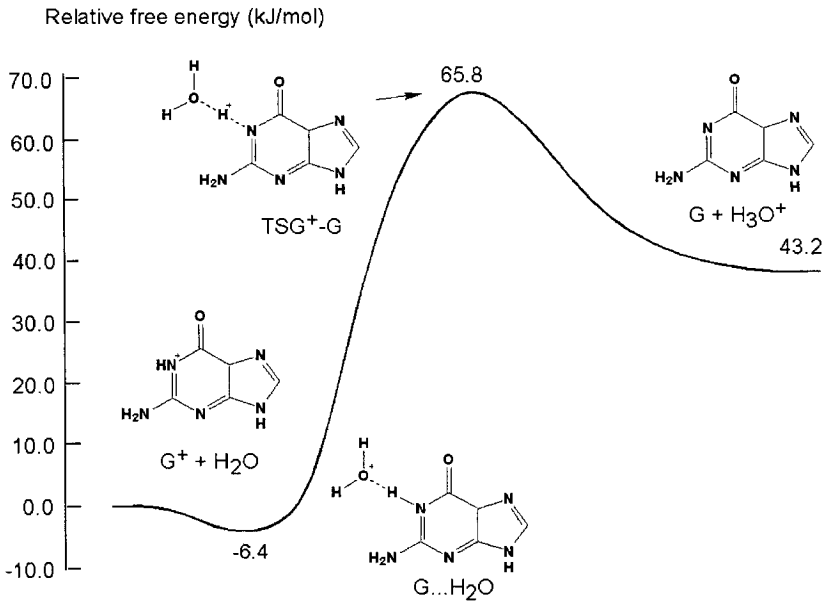


Figure 3.15. Schematic free energy profile at 298.15 K for the route $G^{\bullet+} \rightarrow G^{\bullet}(-H)$ with one water molecule involved

e→a

In this reaction step, a deprotonation reaction takes place at the N1 position of intermediate **e** with a water acting as a proton accepter, leading to the formation **a**. Two complexes **e**...H_3O^+ and **a**...H_2O are formed along the reaction path.

For the ZPVE-only corrected case (Figure 3.16), the transition state **TSe-a** lies 43.3 kJ/mol higher than the isolated **e** + H_3O^+ system. The relative energies of the complexes **e**...H_3O^+ and **a**...H_2O are –93.8 kJ/mol and –100.3 kJ/mol, respectively. The isolated system a + H_2O has a relative energy of –38.4 kJ/mol.

The relative energy of TSe-a becomes –45.3 kJ/mol in the entropy-corrected case at 298.15 K, dropping significantly by almost 90 kJ/mol compared with the ZPVE-only

corrected case (Figure 3.17). The relative energies of **e**...H_3O^+ and **a**...H_2O are –63.1 and –114.7 kJ/mol, respectively. a + H_2O lies 62.1 kJ/mol below the isolated system **e** + H_3O^+. The energy barrier of this reaction step decreases by over 100 kJ/mol compared with the ZPVE-only corrected case, indicating that the entropy contribution to the energy greatly facilitates this reaction step.

The formation of 5-guanidano-4-nitroimidazole

a. Guanine radical cation reaction mechanism

b→c

As described previously, a water molecule attacks intermediate **b** at the C6 position to yield **c**. Two complexes, **b**...$2H_2O$ and **c**...H_2O, are involved in this reaction step.

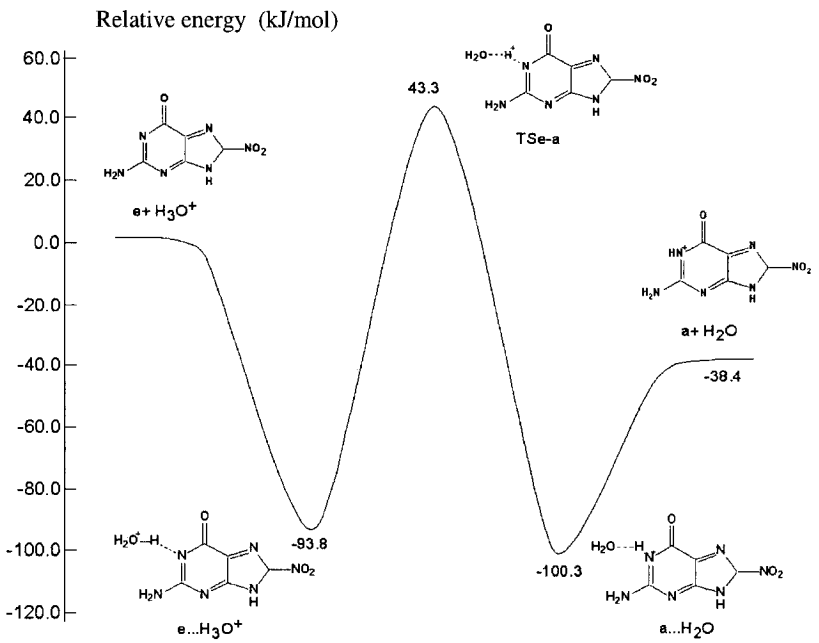


Figure 3.16. Schematic energy profile at 0 K for the route **e→a** with one water involved

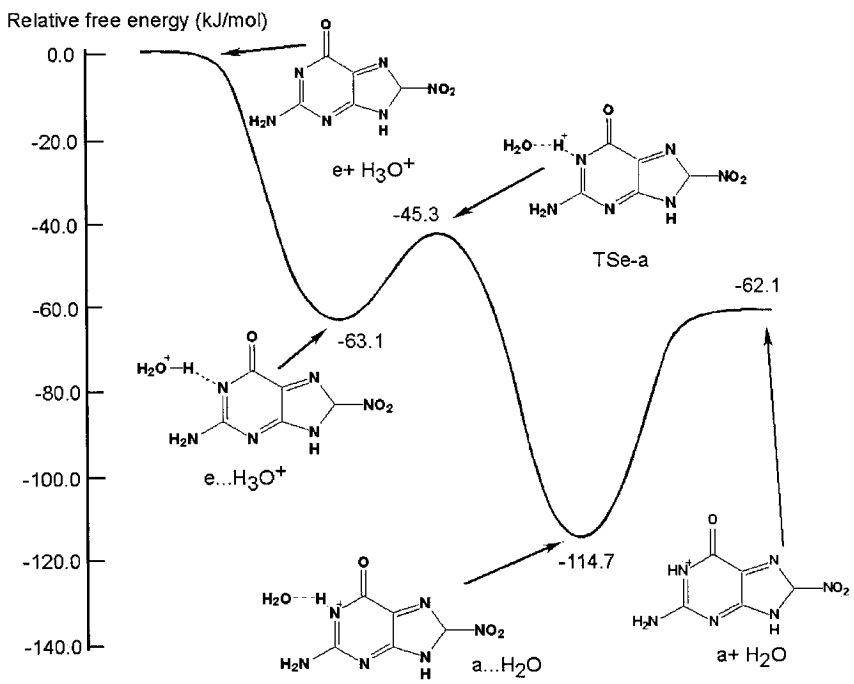


Figure 3.17. Schematic free energy profile at 298.15 K for the route **e**→**a** with one water involved

For the ZPVE-only corrected case (Figure 3.18), complex **b**...$2H_2O$ has an energy of -11.3 kJ/mol relative to that of **b**+$2H_2O$; the relative energy of **TSb-c** is 77.7 kJ/mol; the complex formed after the transition state, **c**...H_2O, has a relative energy of -47.3 kJ/mol; the relative energy of separated system c+H_2O is 14.7 kJ/mol.

Complex **b**...$2H_2O$ is 72.3 kJ/mol higher than **b**+$2H_2O$ in the entropy corrected case at 298.15 K (Figure 3.19), **TSb-c** has a relative energy of 162.1 kJ/mol, and relative energy of **c**...H_2O is 27.3 kJ/mol, all raised about 80 kJ/mol compared with the ZPVE-only corrected case (Figure 3.9), while the energy barrier remains about 90 kJ/mol. The relative energy of c+H_2O is 58.9 kJ/mol.

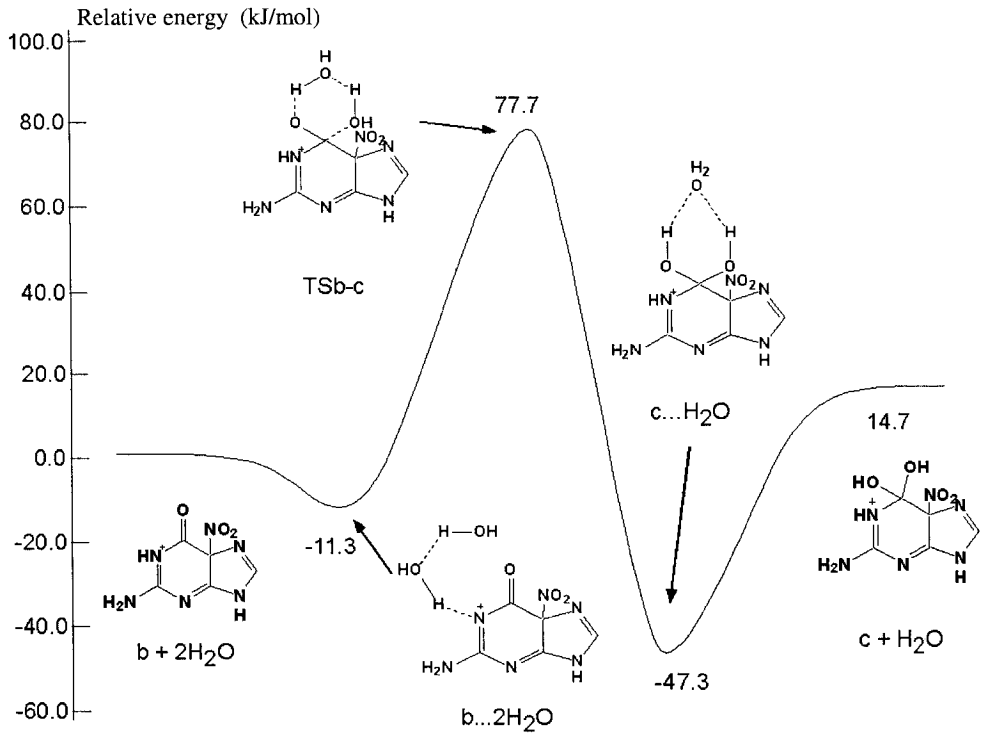


Figure 3.18. Schematic energy profile at 0 K for the route **b→c** with two waters involved in this process.

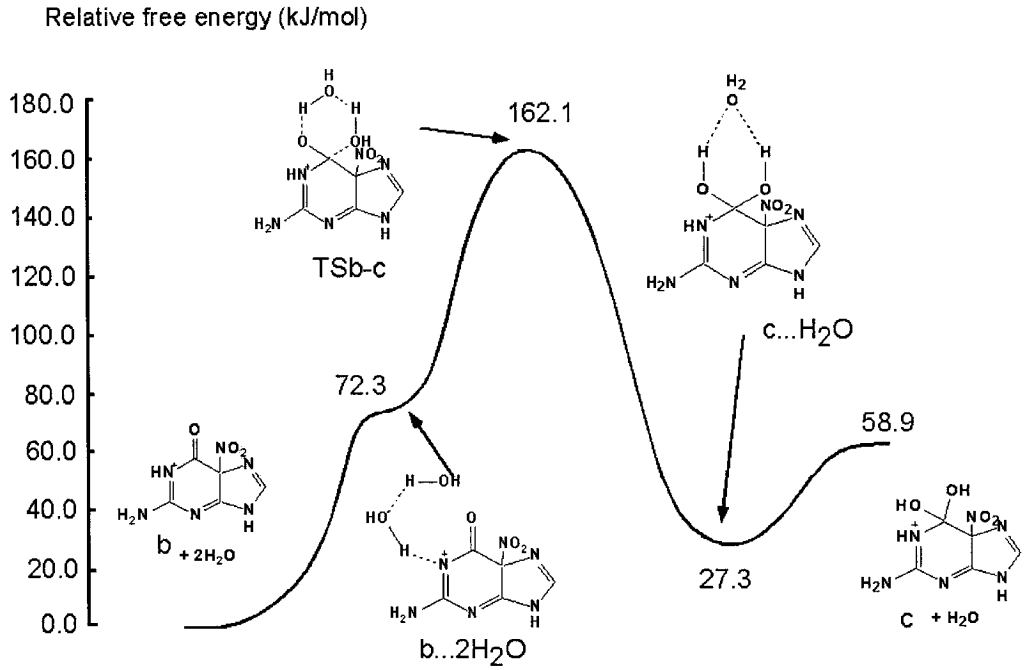


Figure 3.19. Schematic free energy profile at 298.15 K for the route **b→c** with two waters involved in this process.

c→d

In this step, a proton is transferred from the —OH group at the C6 position to the N3 position. This process is companied by the cleavage of the C5—C6 bond.

For the ZPVE-only corrected case (Figure 3.20), complexes **c**...H_2O and **d**...H_2O are –104.5 and –130.8 kJ/mol higher than C + H_2O, respectively; transition state **TSc-d** has a relative energy of –58.6 kJ/mol, while the relative energy of the product group **d**+H_2O is –147.3 kJ/mol.

For the entropy corrected case at 298.15 K (Figure 3.21), the relative energies of **c**...H_2O and **d**...H_2O are 89.8 and 97.9 kJ/mol, respectively. **TSc-d** lies 22.2 kJ/mol below the isolated system **c** + H_2O. The relative energy of **d**+H_2O is 152.8 kJ/mol, not changed much compared with the ZPVE-only corrected case.

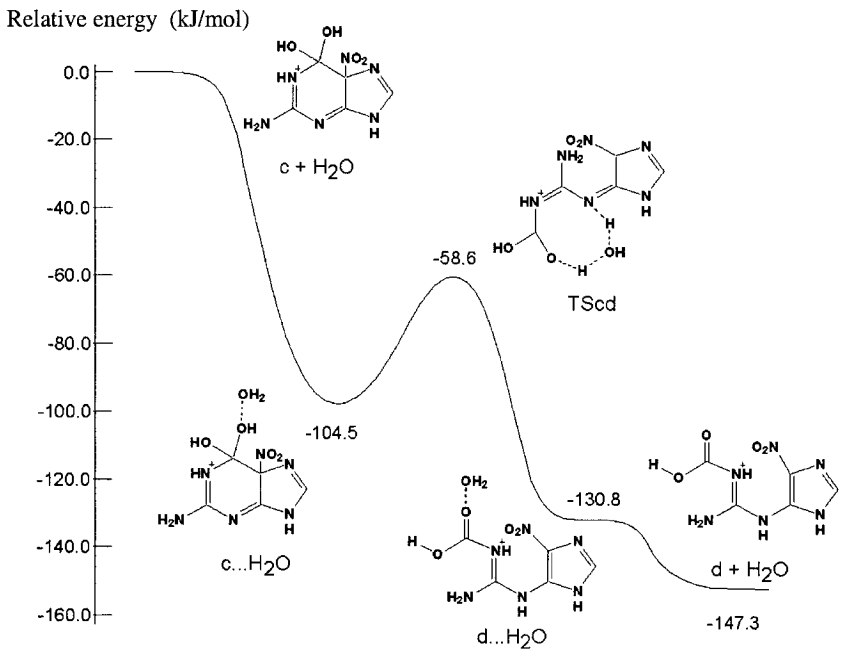


Figure 3.20. Schematic energy profile at 0 K for the route **c→d** with a water molecule involved in this process.

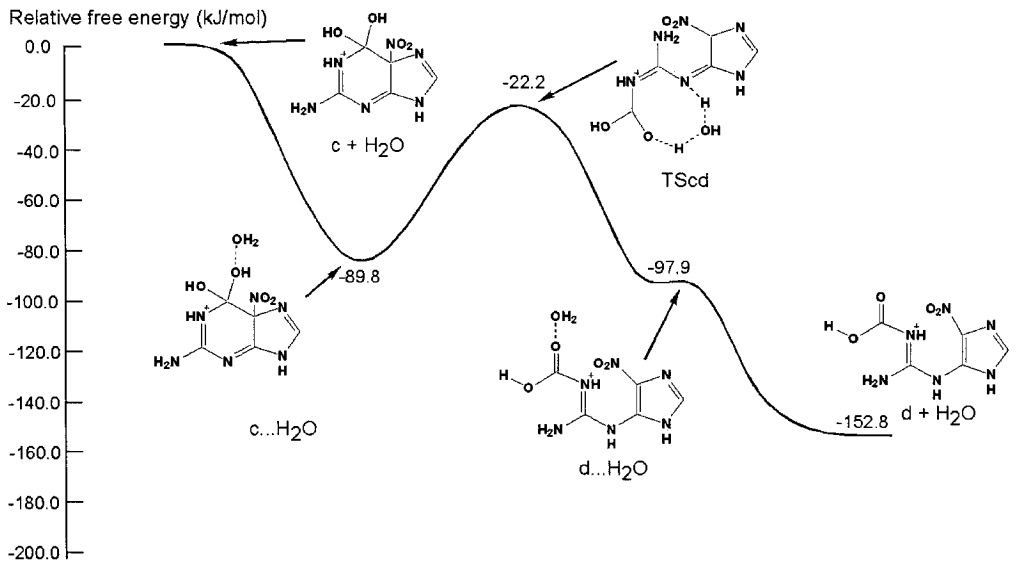


Figure 3.21. Schematic free energy profile at 298.15K for the route **c→d** with a water molecule involved in this process.

d→5-guanidano-4-nitroimidazole

As described in the gas phase results discussion, the final product 5-guanidano-4-nitroimidazole is derived through decarboxylation of intermediate **d**. Intermediate **d** is decomposed into H_3O^+, CO_2, and 5-guanidano-4-nitroimidazole and a complex **d**...H_2O is formed before the transition state **TSd-6**.

For the ZPVE-only corrected case (Figure 3.22), the complex **d...H_2O** has an energy of –59.7 kJ/mol compared to **d** + H_2O, the transition state has a relatively higher energy of –70.3 kJ/mol, making the energy barrier of this reaction step a negative value, and the product 5-guanidano-4-nitroimidazole + CO_2+ H_3O^+ are much lower at –114.5 kJ/mol.

For the entropy corrected case at 298.15 K (Figure 3.23), the relative energies of **d**...H_2O and **TSd-6** are –39.7 and –32.5 kJ/mol, respectively. 5-guanidano-4-nitroimidazole + CO_2+ H_3O^+ lies –66.7 kJ/mol below **d** + H_2O. It can be seen from the above results that the energy barrier is very insignificant, giving this reaction step a descending potential energy surface.

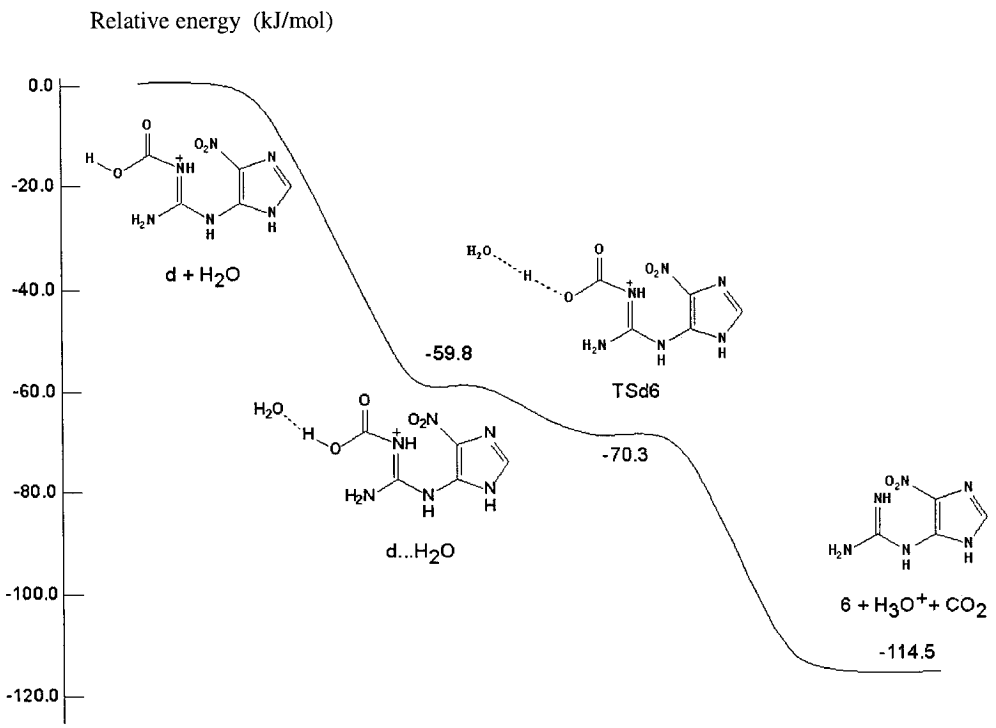


Figure 3.22. Schematic energy profile at 0 K for the route **d**→5-guanidano-4-nitroimidazole with one water molecule involved.

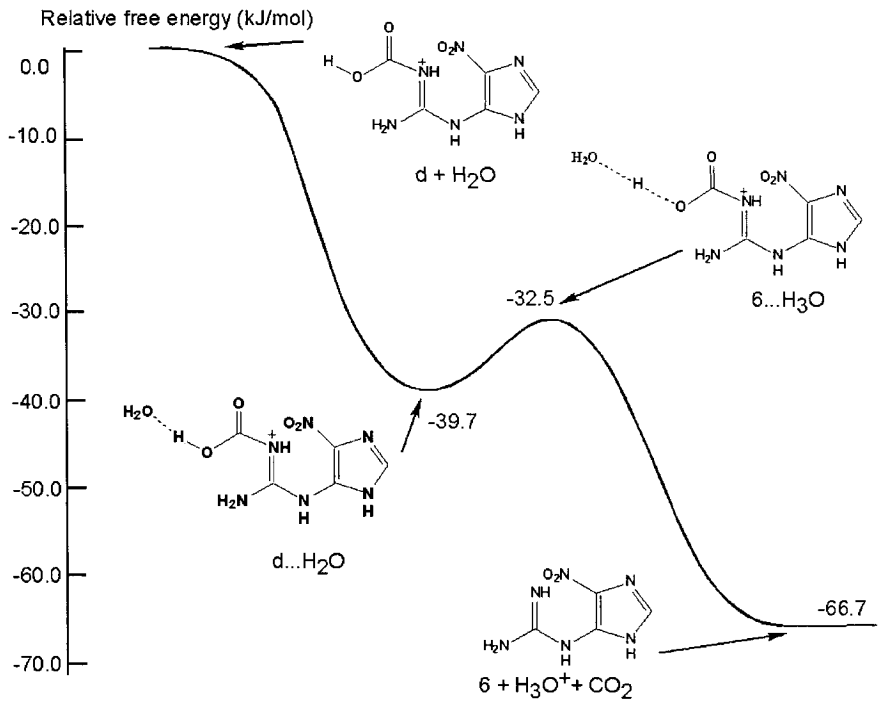


Figure 3.23. Schematic free energy profile at 298.15K for the route **d**→5-guanidano-4-nitroimidazole with one water molecule involved.

b. Neutral guanine radical reaction mechanism

f→g

Intermediate **f** is derived when a nitrogen dioxide radical combines with the **G•(-H)** at the C5 position, and then is attacked by a water molecule at the electrophilic C6 position to yield **g**.

For the ZPVE-only corrected case (Figure 3.24), total energy changes relative to the isolated system of **f**+2H_2O are as follows: the complex **f**...2H_2O has an energy of -72.1 kJ/mol, the transition state **TSf-g** has a relative energy of 148.6 kJ/mol; the relative energies of complex **g**...H_2O and **g** + H_2O are 55.9 kJ/mol and 50.7 kJ/mol, respectively.

For the entropy corrected case at 298.15 K (Figure 3.25), the relative energies of two complexes **f**...$2H_2O$ and **g**...H_2O are 2.5 and 137.3 kJ/mol, respectively. Transition state **TSf-g** lies 234.6 kJ/mol above the isolated **f**+$2H_2O$. The energy difference between **TSf-g** and **f**...$2H_2O$ makes the longest energy barrier of the overall reaction pathway. The isolated **g** + H_2O has a relative energy of 94.5 kJ/mol.

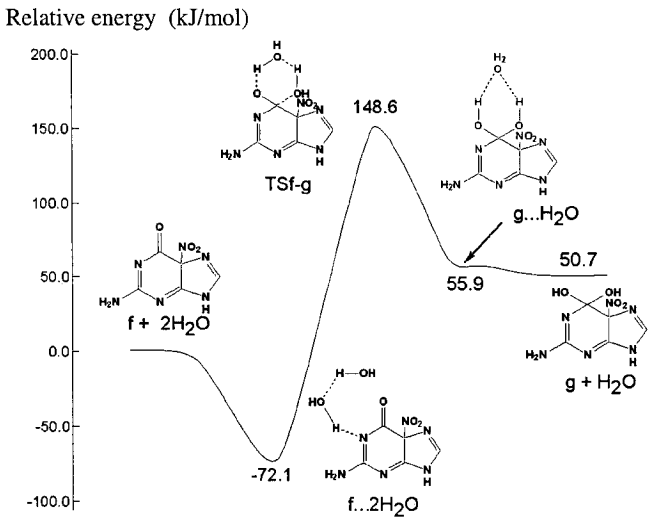


Figure 3.24. Schematic energy profile at 0 K for the route **f**→**g** with two waters involved.

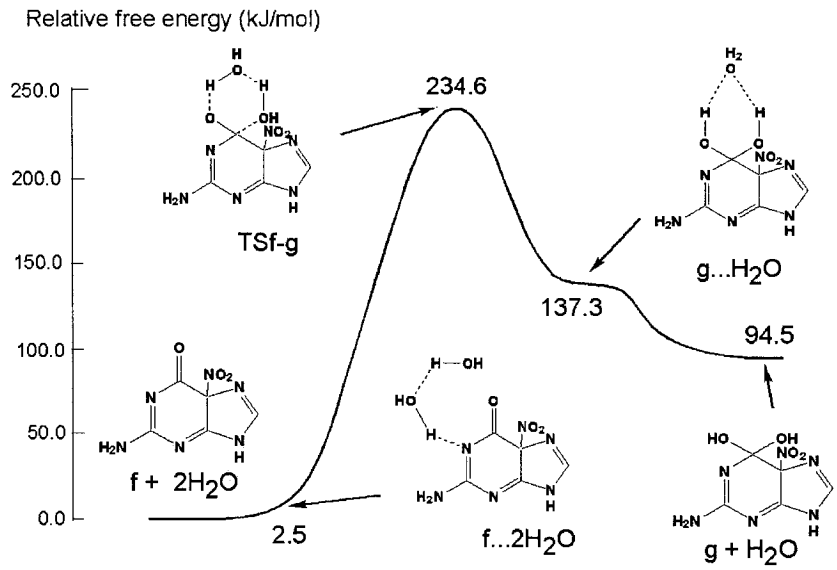


Figure 3.25. Schematic free energy profile at 298.15 K for the route **f**→**g** with two waters involved.

g→h

In this step, a proton is transferred from the —OH group at the C6 position to the N3 nitrogen and the C5—C6 bond breaks at the same time, leading to the formation of intermediate **h**.

For the ZPVE-only corrected case (Figure 3.26), complex **g**...H_2O is –142.2 kJ/mol lower than **g** +H_2O and transition state **TSg-h** has a relative energy of –88.7 kJ/mol. The relative energies of complexes **h**...$2H_2O$ and **h**+$2H_2O$ are –141.7 kJ/mol and –188.7 kJ/mol, respectively.

For the free energy case at 298.15 kJ/mol (Figure 3.27), the relative energies of **g**...H_2O and **h**...$2H_2O$ are –108.9 and –109.5 kJ/mol, respectively. Transition state **TSg-h** has a relative energy of –52.1 kJ/mol. **h**+$2H_2O$ lies –194.3 kJ/mol below the isolated **g** +H_2O.

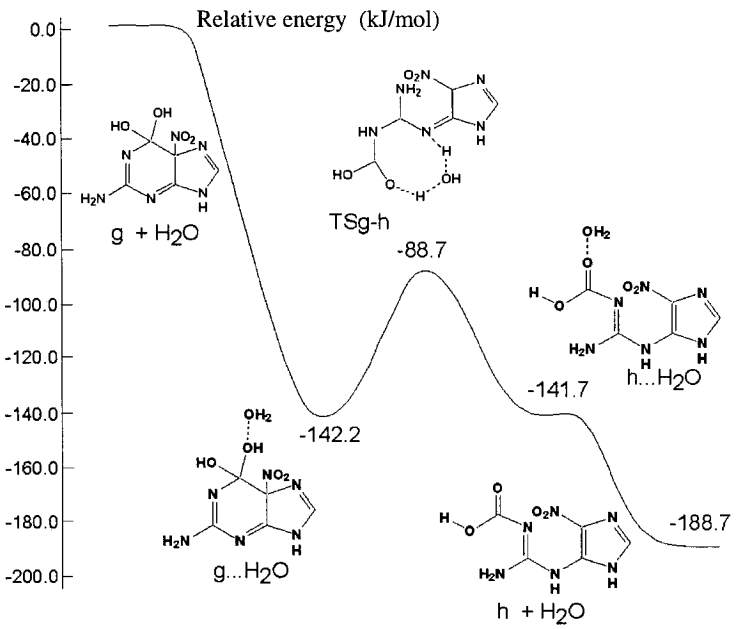


Figure 3.26. Schematic energy profile at 0 K for the route **g→h** with one water involved.

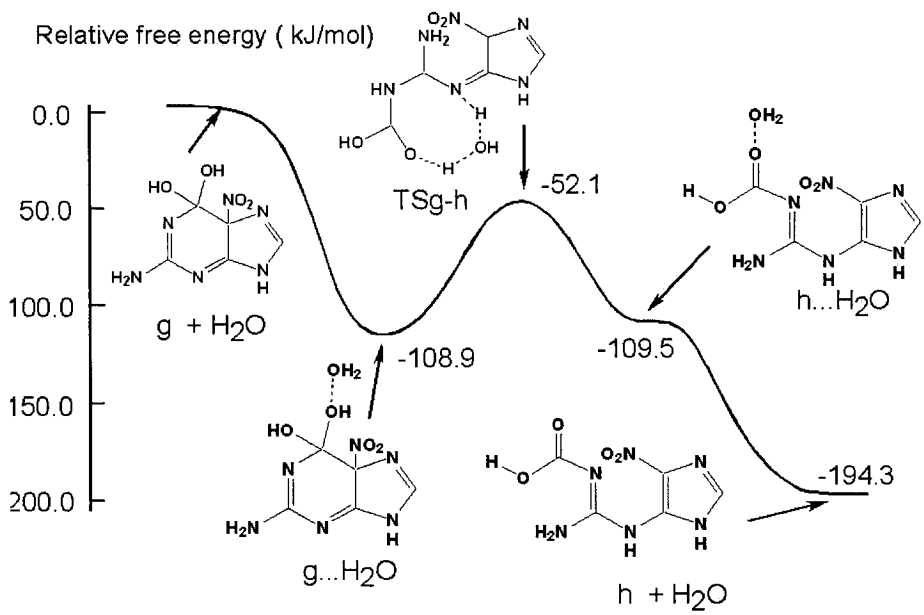


Figure 3.27. Schematic free energy profile at 298.15K for the route **g→h** with one water involved.

h→d

The solvation calculation results show that, 5-guanidano-4-nitroimidazole cannot be derived directly from intermediate **h** in the neutral guanine radical mechanism. A proton must be attached to the N1 position to form intermediate **d** and then the final product 5-guanidano-4-nitroimidazole is derived as a result of the decomposition of **d**. In the protonation process, H_3O^+ serves as a proton donor.

For the ZPVE-only corrected case (Figure 3.28), transition state **TSh-d** is formed with a relative energy of 30.0 kJ/mol and no complex is located before that. The relative energies of the **d**...H_2O complex and products (**d** + H_2O) are -9.4 kJ/mol and –3.4 kJ/mol, respectively.

For the entropy corrected case at 298.15 K (Figure 3.29), the relative energies of **TSh-d** and **d**...H_2O are 46.4 and 2.2 kJ/mol, respectively. **d** + H_2O lies 24.1 kJ/mol below the isolated **h**+H_3O^+.

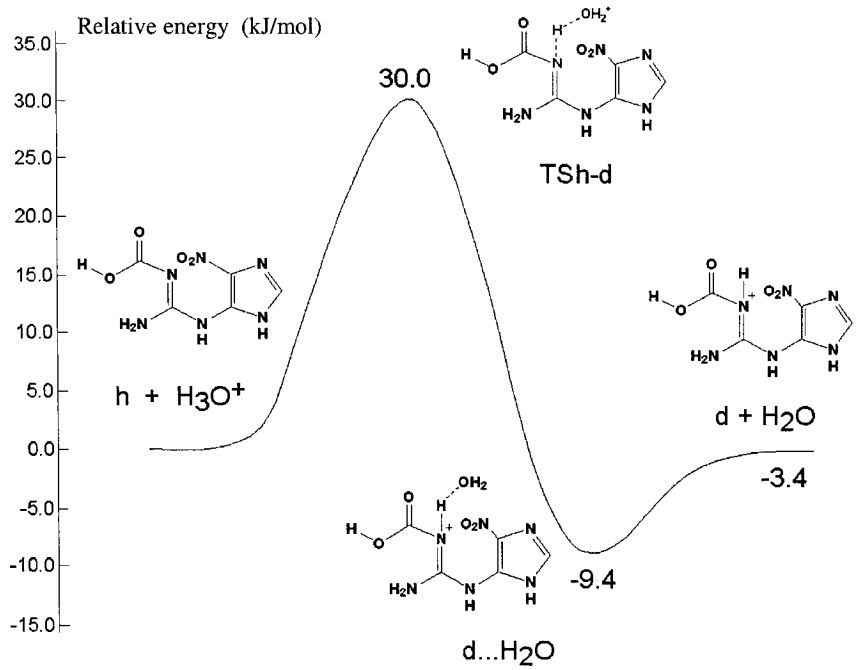


Figure 3.28. Schematic energy profile at 0 K for the route **h**→**d** with one H_3O^+ involved.

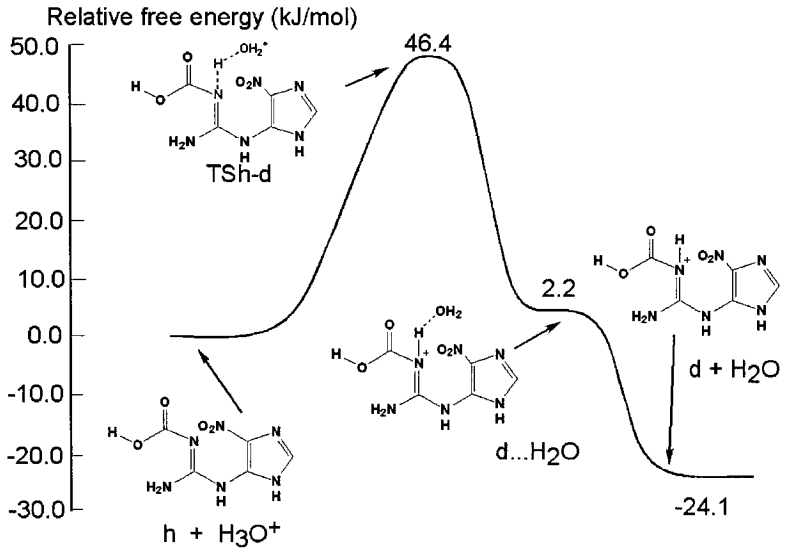


Figure 3.29. Schematic free energy profile at 298.15K for the route **h**→**d** with one H_3O^+ involved.

The following discussions are mainly based on the free energy changes given above.

Immediately following the formation of the guanine radical cation, the radical-radical combination reaction takes place at a diffusion-controlled rate and intermediate **a** is derived. Otherwise, guanine radical cations will undergo deprotonation to yield the neutral radical **G$^{\bullet}$(-H)**.

For the formation of **8-NO_2Gua,** the overall free energy profiles of the guanine radical cation reaction mechanism and the neutral guanine radical mechanism are shown in Figure 3.30. To produce the intermediate **a**, the single-step radical-radical combination reaction in the radical cation reaction mechanism is more favored than the two steps of the reactions in the neutral guanine radical mechanism. The radical-radical combination reaction has no barrier to overcome and is thermodynamically favored by 179.0 kJ/mol. However, the first deprotonation reaction of the neutral guanine radical mechanism has to overcome a barrier of 65.8 kJ/mol via the transition state **TSG.** In addition, the derived neutral guanine radical **G$^{\bullet}$(-H)** has to overcome a second barrier of 17.8 kJ/mol to form intermediate **a**. Since the last step of the reactions to form **8-NO_2Gua** in the two mechanisms are exactly the same, the present results suggest that the guanine radical cation reaction mechanism is preferred over the neutral guanine radical mechanism for the formation of **8-NO_2Gua.**

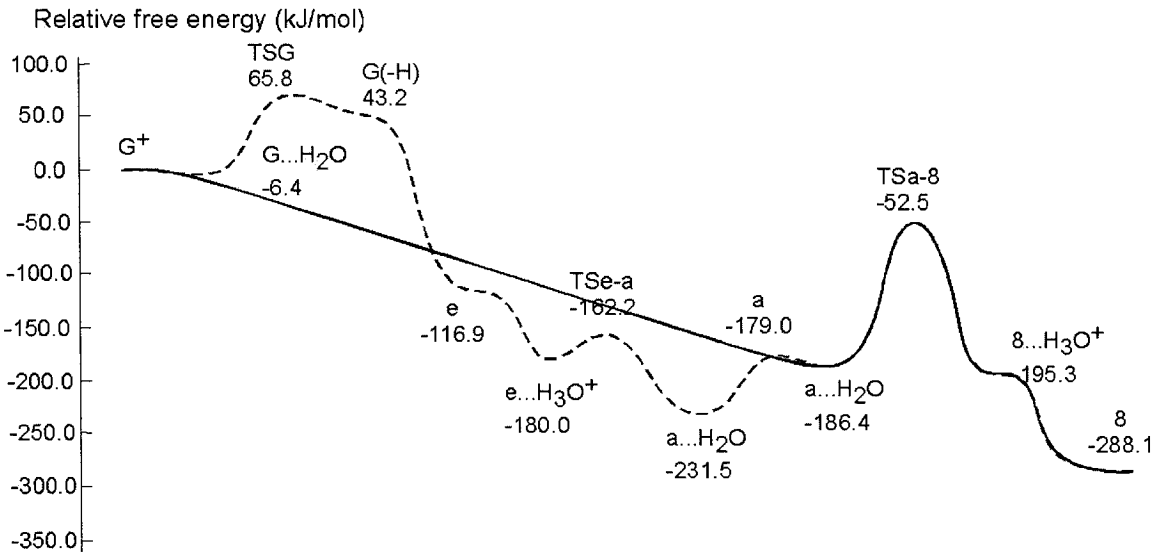


Figure 3.30. Schematic free energy profile at 298.15K for the complete reaction $\mathbf{G^+ \rightarrow 8\text{-}NO_2Gua}$ (The solid line represents the guanine radical cation reaction mechanism while the dashed line represents the neutral guanine radical mechanism).

For the formation of the 5-guanidano-4-nitroimidazole, the overall free energy profiles of the guanine radical cation reaction mechanism and the neutral guanine radical mechanism are shown in Figure 3.31. Overall, the radical cation mechanism is a downhill reaction and has a smoother free energy surface than that of the neutral guanine radical mechanism. In particular, the first-step radical combination reaction in the radical cation mechanism is more favored than the first step deprotonation reaction to form the neutral guanine radical in the neutral guanine radical mechanism. The reaction step that yields 5-guanidano-4-nitroimidazole directly from intermediate **h** was found in the gas phase calculations. However, the transition state of this step could not be located in the solvation calculations, suggesting that this particular reaction step is not favored in solution. In addition, the **f**→**g** step of the reaction in the neutral guanine radical pathway has the largest barrier of 234.6 kJ/mol. Therefore, the guanine radical cation reaction

mechanism is preferred over the neutral guanine radical reaction mechanism for the formation of the significant oxidation product 5-guanidano-4-nitroimidazole.

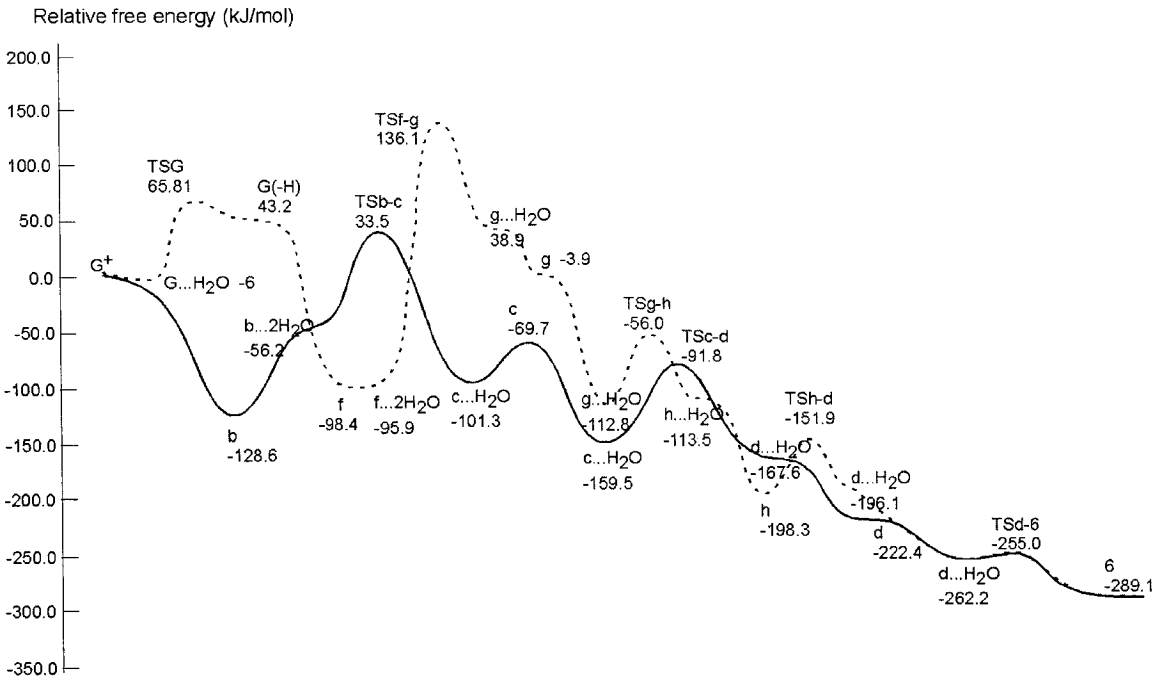


Figure 3.31. Schematic free energy profile at 298.15K for the complete reaction $G^+ \rightarrow$ **5-guanidano-4-nitroimidazole** (The solid line represents the guanine radical cation reaction mechanism while the dashed line represents the neutral guanine radical mechanism).

3.4 Conclusions

The present computational study suggests that the guanine radical cation reaction mechanism is more favorable than the neutral guanine radical reaction mechanism in terms of the formation of both 8-NO_2Gua and 5-guanidano-4-nitroimidazole. Furthermore the present calculations suggest that a water molecule is involved in the guanine radical cation mechanism as a catalyst or as a reactant. Given the role of water in the preferred mechanism, we propose that the guanine radical cation reaction mechanism operates in *vivo*.

3.5 Representation of optimized structures

Graphical representations of the optimized structures are shown in figure 3.32. These figures were drawn with the GaussView software.

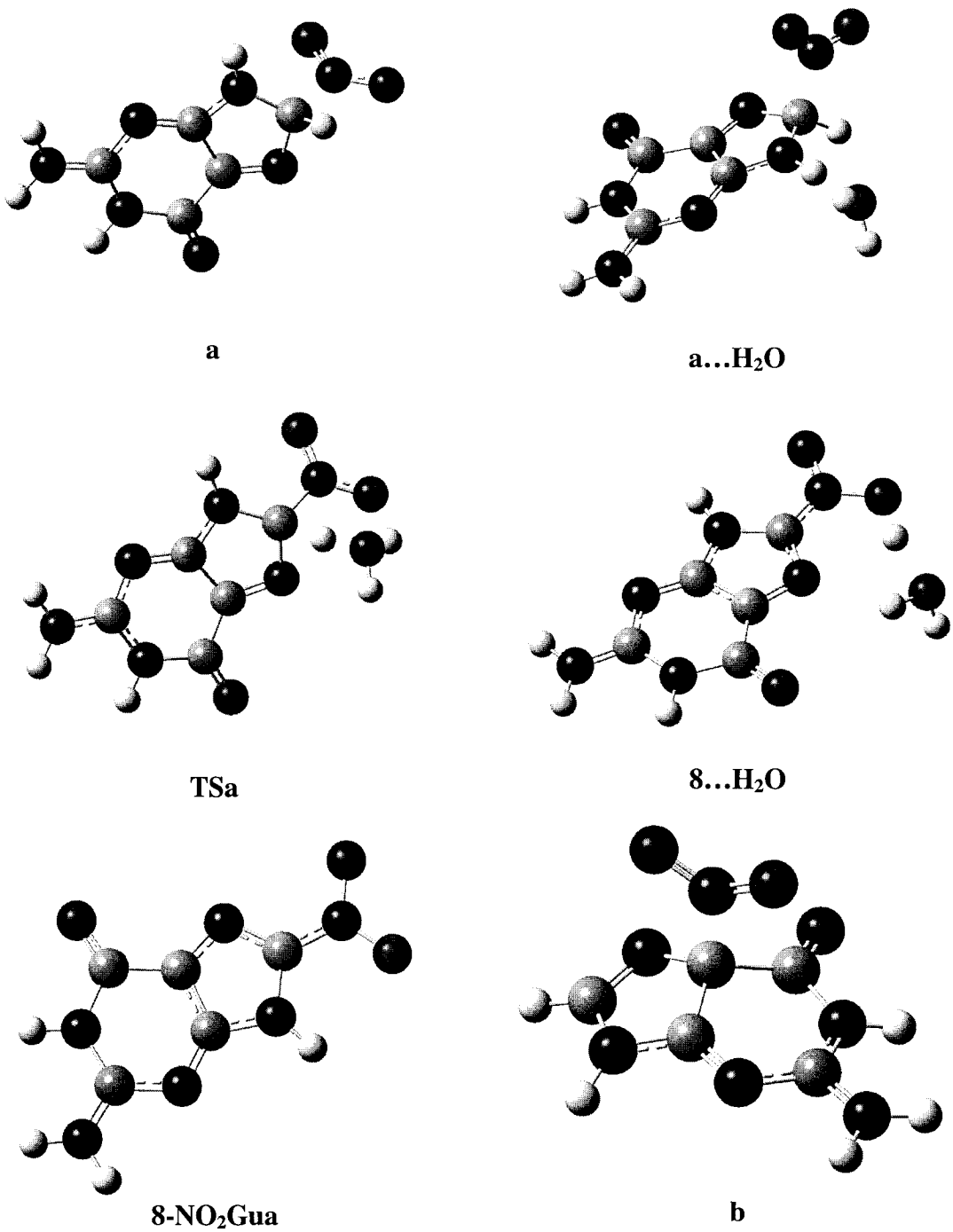
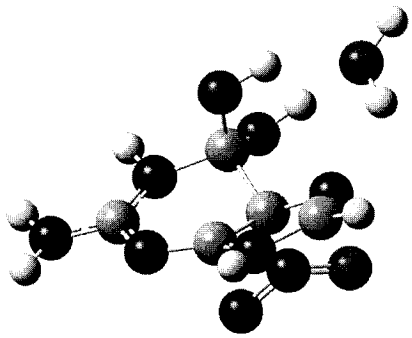
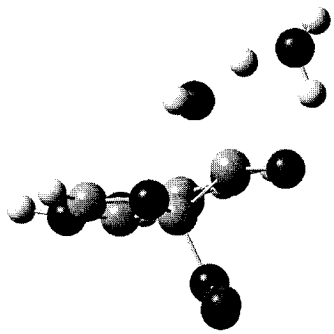


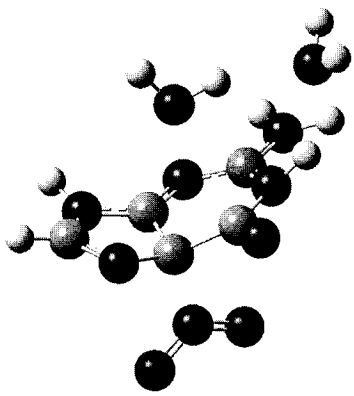
Figure 3.32 Graphical representations of the optimized structures in the peroxynitrite oxidation of guanine reaction



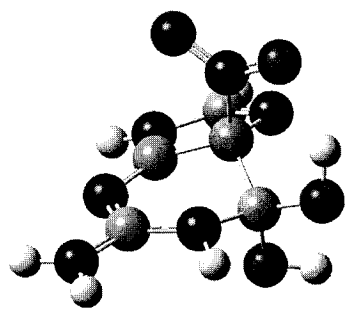
b…2H2O



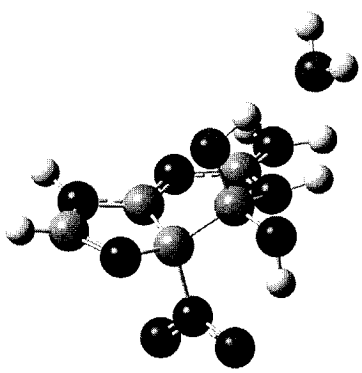
TSb



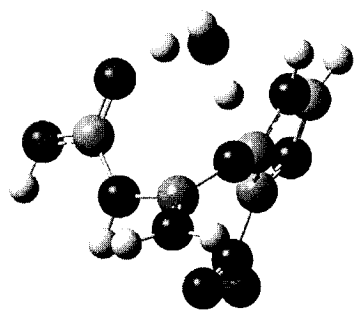
c…H_2O (b-c)



c

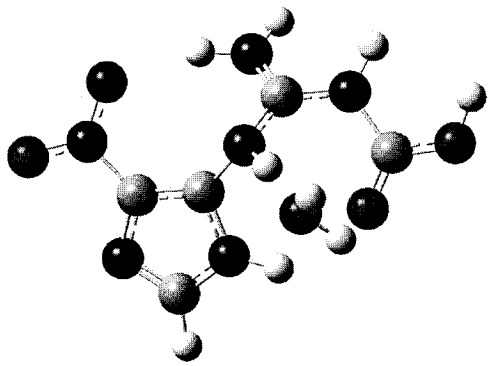


c…H_2O (c-d)

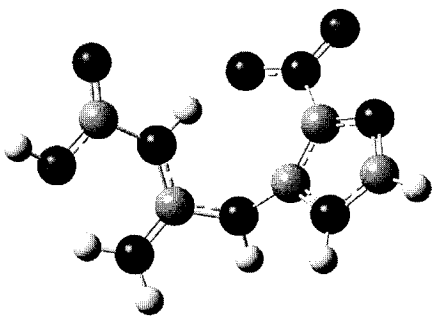


TSc

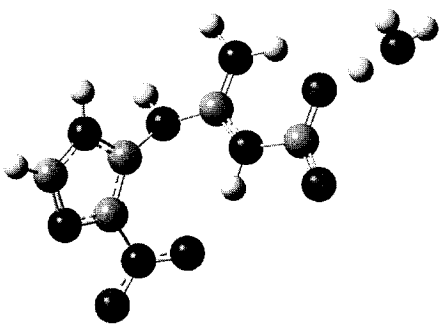
Figure 3.32 Graphical representations of the optimized structures in the peroxynitrite oxidation of guanine reaction (continued)



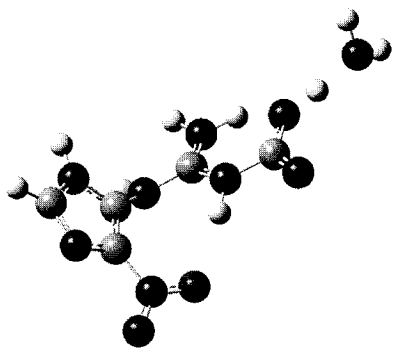
d...H_2O



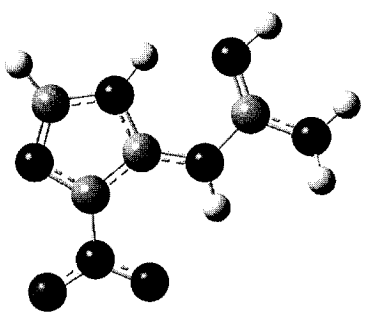
d



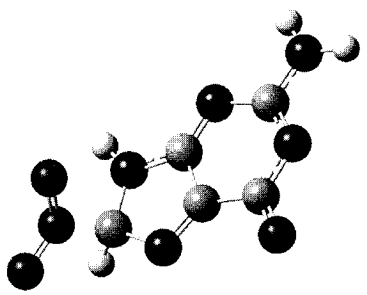
TSd



d...H_2O

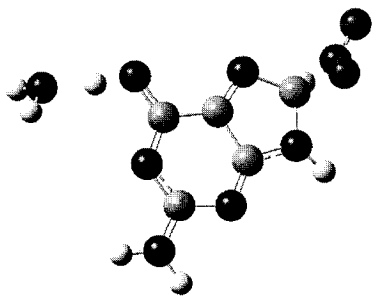


5-guanidino-4-nitroimidazole

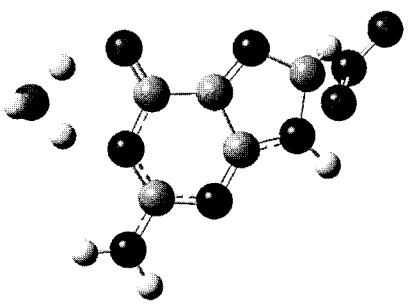


e

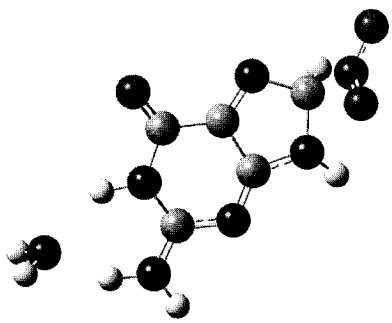
Figure 3.32 Graphical representations of the optimized structures in the peroxynitrite oxidation of guanine reaction (continued)



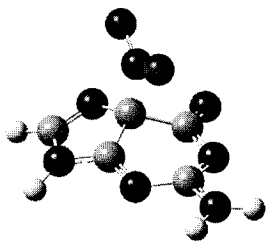
e...H_2O



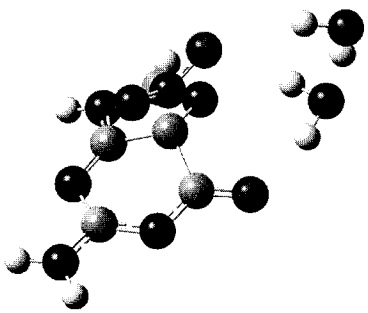
TSe



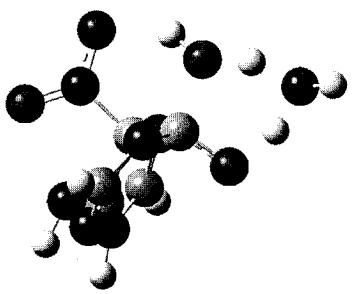
a...H_2O (e-a)



f

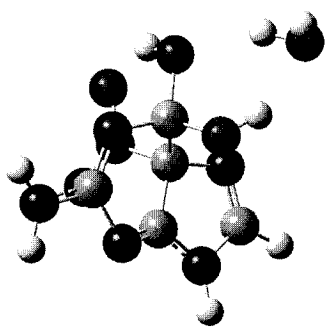


f...$2H_2O$

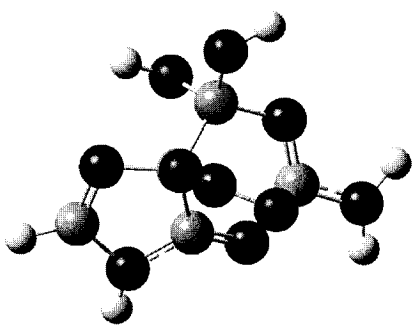


TSf

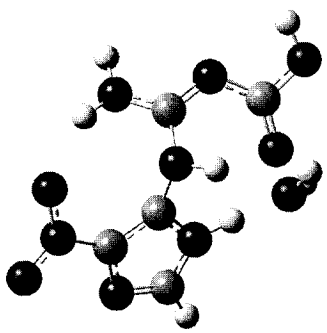
Figure 3.32 Graphical representations of the optimized structures in the peroxynitrite oxidation of guanine reaction (continued)



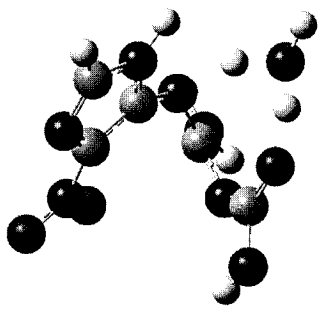
g...H_2O (f-g)



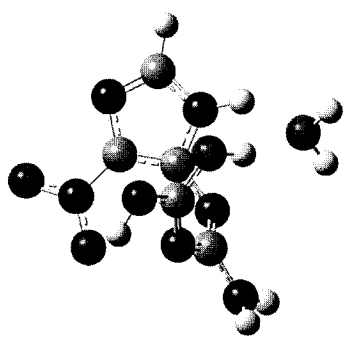
g



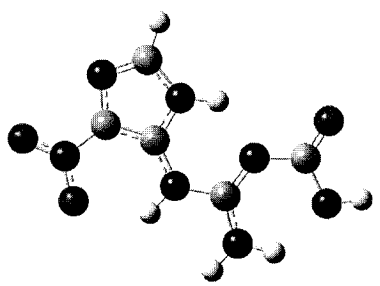
g...H_2O (g-h)



TSg

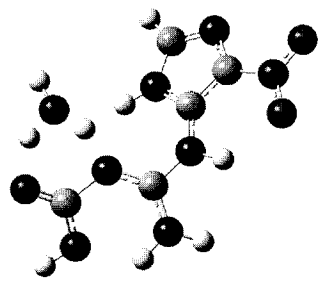


h...H_2O

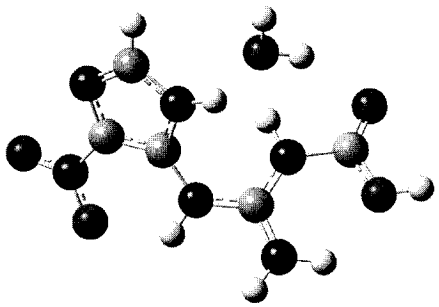


h

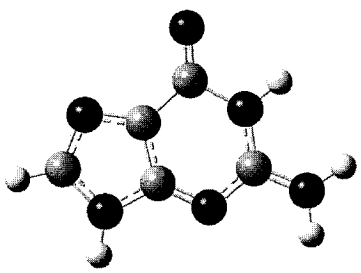
Figure 3.32 Graphical representations of the optimized structures in the peroxynitrite oxidation of guanine reaction (continued)



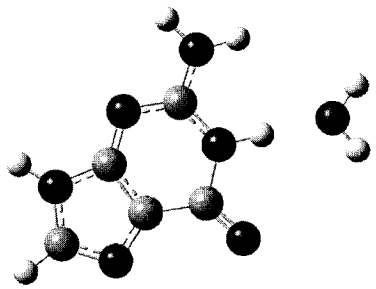
TSh



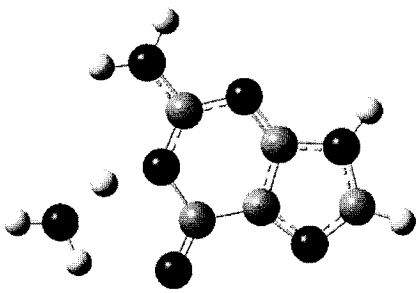
d...H_2O (h-d)



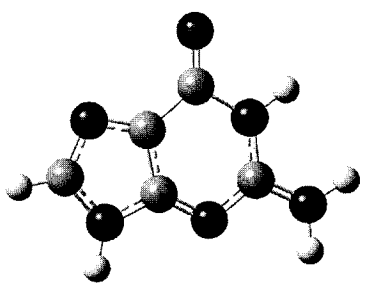
$G^{\bullet+}$



G...H_2O



TSG



$G^{\bullet}$(-H)

Figure 3.32 Graphical representations of the optimized structures in the peroxynitrite oxidation of guanine reaction (continued)

Chapter 4

Modelling the Reaction Mechanisms for Redox Regulation of Protein Tyrosine Phosphatase 1B (PTP1B) Activity

4.1 Introduction

Protein-tyrosine phosphatases (PTPs) are enzymes central in regulating many cellular processes, particularly in response to extracellular signals.[105,106] These enzymes function to remove the phosphoryl group from tyrosinephosphorylated proteins and, together with protein-tyrosine kinases, modulate the cellular level of protein-tyrosine phosphorylation (Figure 4.1).[107,108] In PTPs, a phosphate group is hydrolyzed catalytically from a tyrosine side chain, while in kinases, phosphate is added to the side chain. A proper level and timing of tyrosine phosphorylation is critical for regulating cell growth, differentiation, metabolism, and progression through the cell cycle, as well as cell-to-cell communication

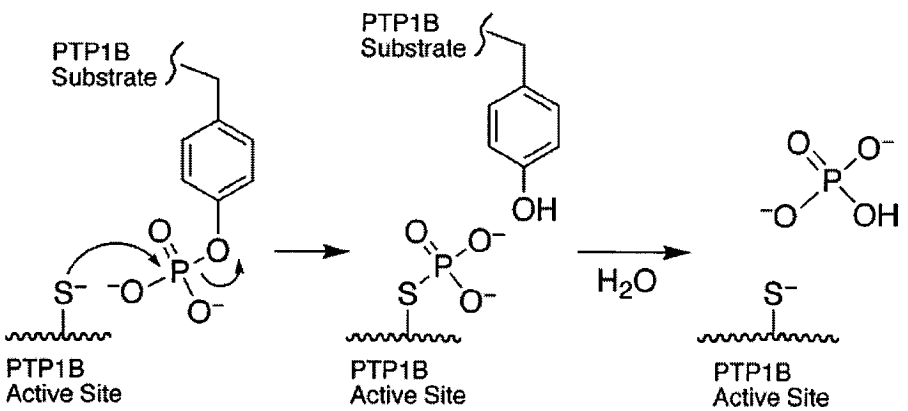


Figure 4.1. Reaction catalyzed by protein-tyrosine phosphatases

and cell death versus survival.[109,110] Genetic and biochemical studies indicate that many of these PTPs are also involved in a number of human diseases,[111-113] which may be partially due to the fact that every PTP has a unique environment that brings a low pK_a to the catalytic cysteine.[114]

PTPs constitute a large family of enzymes: there are about 100 PTPs encoded in the human genome.[115,116] Those include both cytosolic and membrane-bound receptor enzymes.[109,117] The different subfamilies are diverse in sequence, molecular weight, and specificity, but there is considerable experimental evidence that PTPs exhibit a common mechanism with related common structural features.

Protein-tyrosine phosphatase 1B (PTP1B) is a particular PTP whose function in insulin signaling and metabolism has been well established.[118,119] It has been observed that mice lacking PTP1B exhibit resistance to diabetes and do not develop diet-induced obesity,[120,121] suggesting that PTP1B inhibitors may address both insulin and obesity resistance. PTP1B can be inactivated when exposed to reactive oxygen species such as hydrogen peroxide and superoxide[122] and therefore an insulin-stimulated burst of hydrogen peroxide can inactivate PTP1B and lead to an increase in phosphorylation levels of relevant substrate proteins. The active form of the catalytic site can be regained through the inactive protein reacting with the cellular thiol glutathione once hydrogen peroxide is diminished.[123] Recent X-ray crystallographic studies of the redox regulation of PTP1B indicate that an intrastrand protein cross-linking between the catalytic cysteine residue and a neighboring amide nitrogen[124,125] is derived from the oxidative inactivation

of the enzyme. Different reaction mechanisms have been proposed to explain the formation of this 3-isothiazolidinone heterocycle and subsequent reactivation of the catalytic site (Figure 4.2).[126]

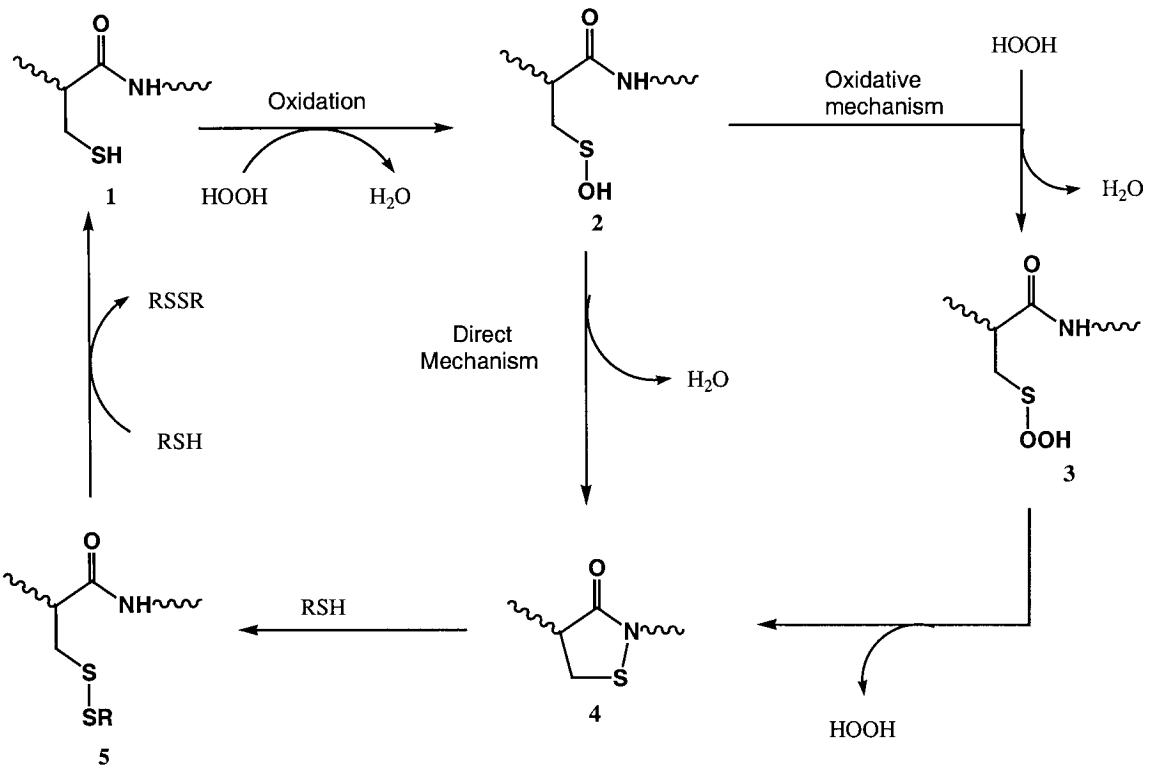


Figure 4.2. Putative mechanism of 3-isothiazolidinon formation and subsequent reactivation of the catalytic site of PTP1B

In the putative mechanism, the catalytic cysteine of PTP1B (E—SH) is oxidized to a sulphenic acid (E—S—OH), denoted as intermediate **2**. In the next step, the sulphenylamide (intermediate **4**) may be formed by a direct mechanism or an oxidative mechanism. In the direct mechanism, the backbone nitrogen of the neighboring serine attacks the S atom of the catalytic cysteine to yield the sulphenylamide directly with subsequent release of water. In the oxidative mechanism, intermediate **2** is further oxidized by hydrogen peroxide to form intermediate **3**, which then reacts to give the sulphenylamide. Because the nitrogen atoms of amide groups are generally considered to

be poor nucleophiles, the sulfenic acid in intermediate **2** must possess sufficient electrophilicity to facilitate the attack by nitrogen of the neighboring amide. Intermediate **5**, the inactive form of the catalytic site, is derived from the sulphenylamide, through a ring-opening process. When intermediate **5** is attacked by a thiol the original form of the catalytic site is regained.

In the present work, quantum chemical calculations are carried out to study the reactivation of the catalytic cysteine in the PTP1B active site and to evaluate the proposed reaction mechanisms. A simplified model, as shown in Figure 4.3, is chosen in this study to represent the active site of PTP1B.

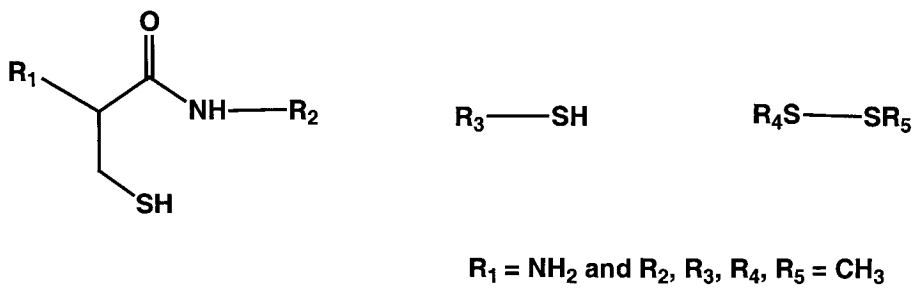


Figure 4.3. A simplified model that represents the active site of PTP1B and compounds involved in the reactivation

4.2 Computational methods

All geometry optimizations and frequency calculations were performed with the B3LYP hybrid density functional in conjunction with the 6-31G(d) basis set by using the Gaussian 03 suite of programs.[127] Harmonic vibrational frequencies and the zero-point vibrational energies (ZPVEs) were obtained at the same level of theory. Relative energies were obtained by performing single-point calculations at the B3LYP level of theory and

Pople's 6-311++G(2df,2p) basis set using above optimized geometries and by including the zero-point vibrational energy, i.e., B3LYP/6-311++G(2df,2p)//B3LYP/6-31G(d) +ZPVE. The entropy contributions to the free energies at 298.15 K were derived from B3LYP/6-31G(d) frequency calculations. Intrinsic reaction path (IRC) calculations were performed on every transition state to confirm that the transition state connects the minima of interest.

The solvation effect on the potential energy surface was investigated by single-point calculations at the B3LYP/6-311++G(2df,2p) level using the conductor-like polarized continuum solvent model (CPCM) (denoted CPCM- B3LYP/6-311++G(2df,2p)) on the geometries obtained by using Onsager model with water as the solvent at the B3LYP/6-31G(d) level (denoted Onsager-B3LYP/6-31G(d)). The zero-point energy corrections obtained from the Onsager-B3LYP/6-31G(d) frequency calculations are included in the calculations of the relative energies in water, i.e., CPCM- B3LYP/6-311++G(2df,2p)//Onsager-B3LYP/6-31G(d) + Onsager-ZPVE. Entropy contributions to the free energies of solvation at 298.15 K were derived from Onsager-B3LYP/6-31G(d) frequency calculations. All energies are in kJ/mol.

4.3 Results and Discussion

4.3.1 Calculation results in gas phase

1→2

In the first reaction step (Figure 4.5), the model of the active site of the enzyme is attacked by a hydrogen peroxide molecule, which connects to the thiol group of the

active site through a water molecule and the complex **1**...(H_2O_2 + H_2O) is formed. For the ZPVE-only corrected case (Figure 4.4), the complex lies 41.6 kJ/mol below the isolated system **1** + H_2O_2 + H_2O. The reaction proceeds through transition state **TS1-2,** which has a relative energy of 93.6 kJ/mol.

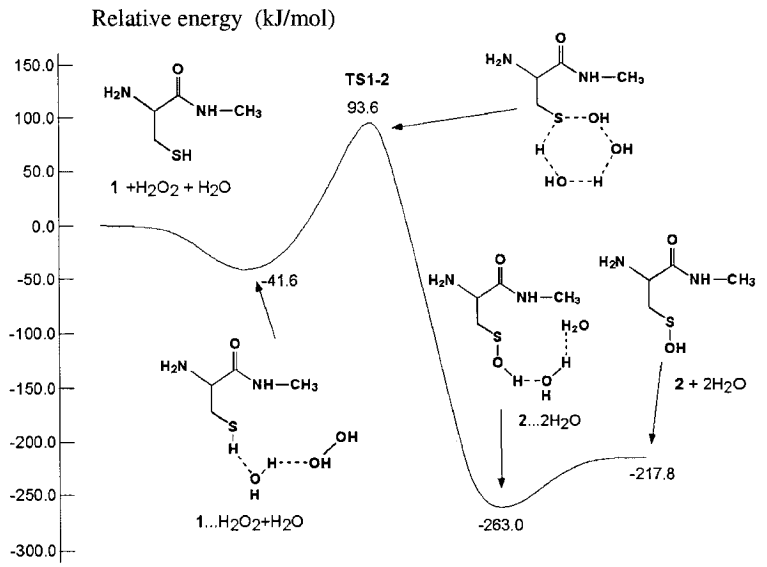


Figure 4.4. Schematic energy profile at 0 K in gas phase for the route **1**→**2**

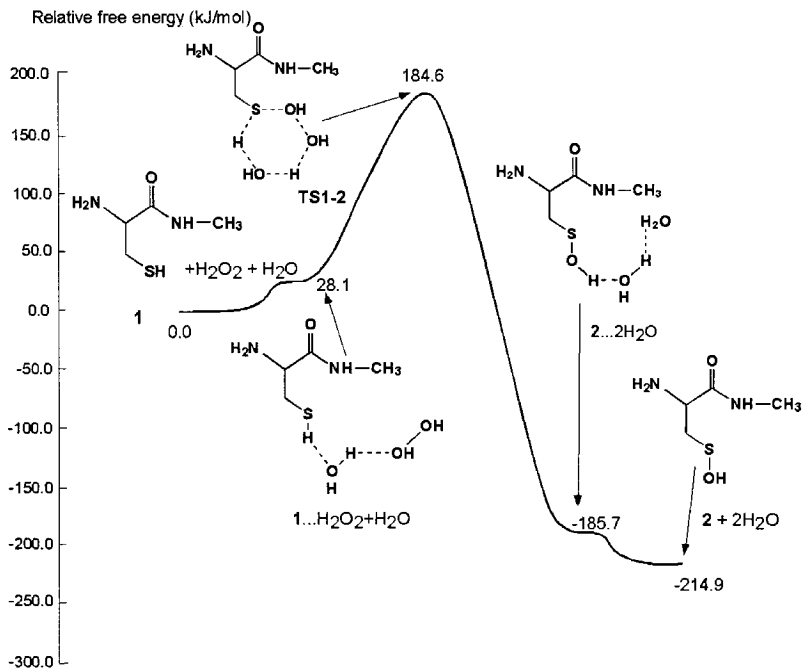


Figure 4.5. Schematic free energy profile at 298.15 K in gas phase for the route **1**→**2**

The transition state is a six-membered ring formed between the thiol group of the active site, the water molecule and the attacking hydrogen peroxide. Following the transition state, another complex **2**...$2H_2O$ is formed in which the thiol group has been oxidized into sulphenic acid (E—S—OH), lying 263.0 kJ/mol lower than the reactants. The isolated system **2** + $2H_2O$ has a relative energy of –217.8 kJ/mol.

For the entropy corrected case at 298.15 K (Figure 4.5), the relative energy of **1**...(H_2O_2 + H_2O) and **TS1-2** are 28.1 kJ/mol and 184.6 kJ/mol, respectively. **2**...$2H_2O$ lies 185.7 kJ/mol below the isolated **1** + H_2O_2+ H_2O. **2** + $2H_2O$ has a relative energy of –214.9 kJ/mol.

2→3

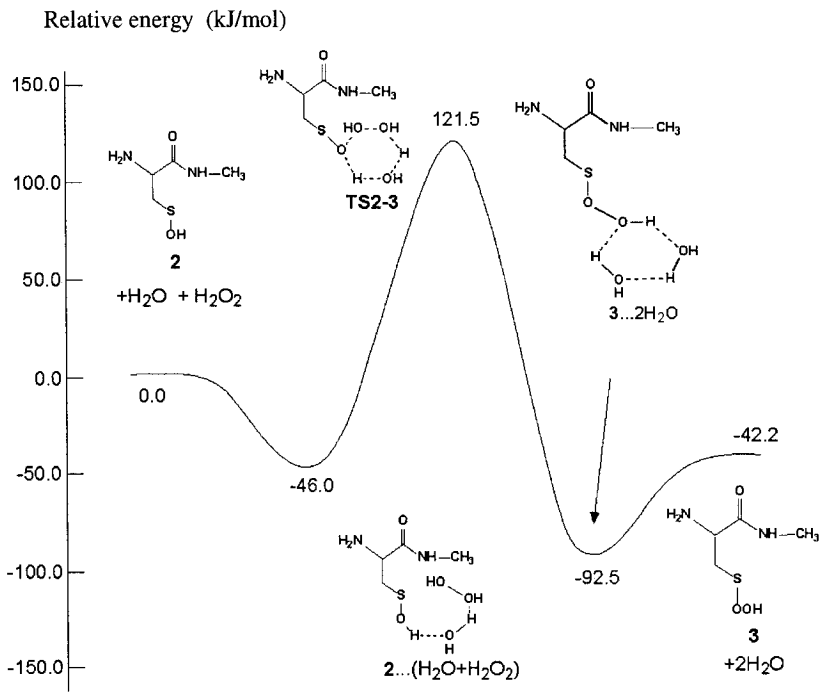


Figure 4.6. Schematic energy profile at 0 K in gas phase for the route **2→3**

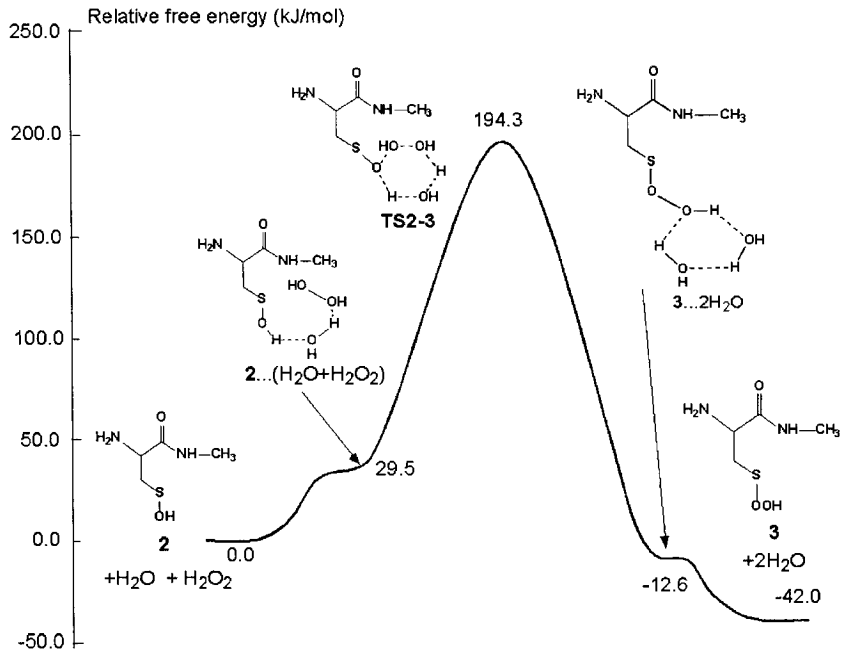


Figure 4.7. Schematic free energy profile at 298.15 K in gas phase for the route **2**→**3**

In this reaction step (Figure 4.6), the newly formed sulphenic acid (E—S—OH) intermediate **2** is oxidized further to the sulphenoperoxoic acid (E—S—OOH) intermediate **3**, by hydrogen peroxide. For the ZPVE-only corrected case (Figure 4.6), one oxygen atom of H_2O_2 is connected to the —OH of intermediate **2** through a water bridge to yield a complex **2**...(H_2O + H_2O_2), with an energy 46.0 kJ/mol lower than that of the isolated system **2** + H_2O + H_2O_2. Transition state **TS2-3** contains a six-membered ring, in which the other oxygen atom of the hydrogen peroxide is connected to the oxygen of the sulphenic acid group. The transition state lies 121.5 kJ/mol higher than the isolated system **2** + H_2O + H_2O_2. Following the transition state, another complex **3**...$2H_2O$ is formed with a relative energy of –92.5 kJ/mol. In this complex, two water molecules and the hydroxyl group of the newly formed sulphenoperoxoic acid group

together form a loose six-membered ring through hydrogen bounding. Intermediate **3** is formed after the release of the two water molecules. The isolated system **3** + $2H_2O$ lies 42.2 kJ/mol lower than **2** + H_2O + H_2O_2.

For the entropy corrected case at 298.15 K (Figure 4.7), **2**...(H_2O + H_2O_2) lies 29.5 kJ/mol above the isolated **2** + H_2O + H_2O_2. The relative energies of **TS2-3** and **3**...$2H_2O$ are 194.3 kJ/mol and –12.6 kJ/mol, respectively. **3** + $2H_2O$ has a relative energy of –42.0 kJ/mol.

3→4

In the reaction from **3** to **4** (Figure 4.8), the neighboring amide nitrogen atom attacks the sulphenoperoxoic acid at the S position, and an isothiazolidine ring is formed, with the hydrogen peroxide acting as a leaving group. Initially, a water molecule is coordinated to the compound, forming two hydrogen bonds, one between the oxygen of the water

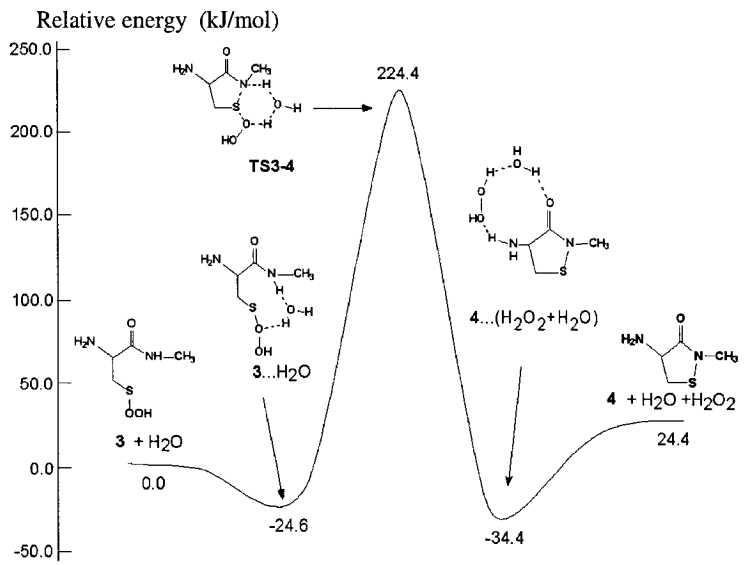


Figure 4.8. Schematic energy profile at 0 K in gas phase for the route **3→4**

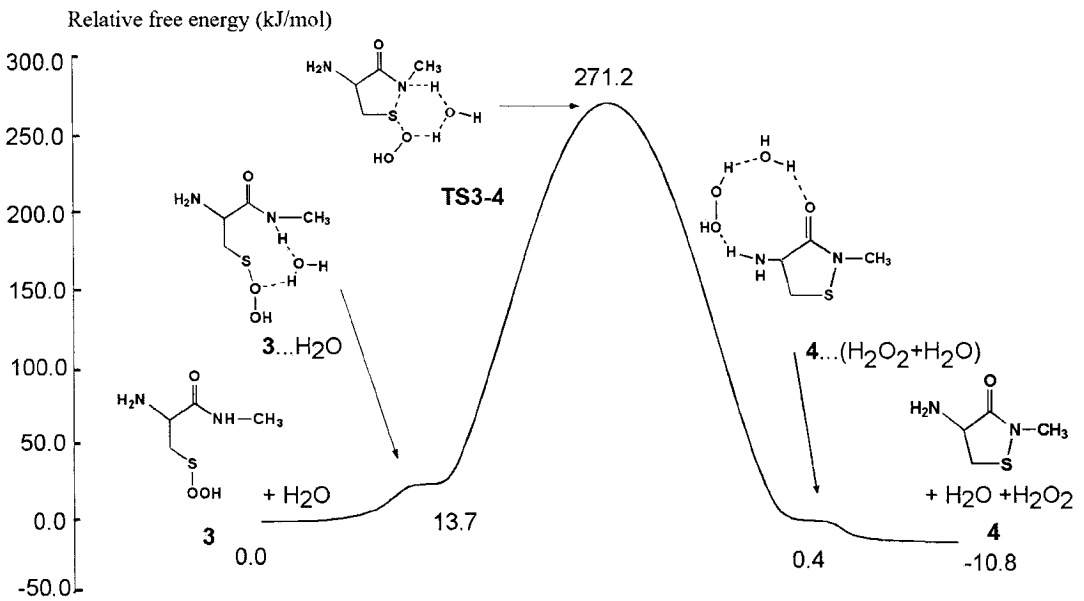


Figure 4.9. Schematic free energy profile at 298.15 K in gas phase for the route **3**→**4**

molecule and the hydrogen of the neighboring N—H, and the other between the hydrogen of the water molecule and the oxygen directly bonded to the S atom. For the ZPVE-only corrected case (Figure 4.8), the complex **3**...H_2O lies 24.6 kJ/mol below the isolated system **3** + H_2O. In the transition state **TS3-4**, the neighboring nitrogen attacks the sulfur and a six-membered ring is formed. In this process, the water molecule donates one hydrogen atom to the —OOH group and accepts one from the nitrogen. The relative energy of **TS3-4** is 224.4 kJ/mol. Once the N—S bond is formed, the N—H and S—O bonds are completely broken. The released H_2O and H_2O_2 are relocated and form a loose loop with the amino group of the cysteine residue and the neighboring carboxylic group. As a result, a complex **4**...(H_2O_2 + H_2O) is formed with a relative energy of -34.4 kJ/mol.

Intermediate **4** is formed when H_2O and H_2O_2 are released. The isolated system of **4** + H_2O + H_2O_2 has an energy 24.4 kJ/mol higher than that of **3** + H_2O.

For the entropy corrected case at 298.15 K (Figure 4.9), the relative energies of **3**...H_2O and **TS3-4** are 13.7 kJ/mol and 271.2 kJ/mol, respectively. **4**...(H_2O_2 + H_2O) has a relative energy of 0.4 kJ/mol. **4** + H_2O + H_2O_2 lies 10.8 kJ/mol below 3 + H_2O.

2→4

In this reaction step (Figure 4.10), the neighboring nitrogen directly attacks the sulfur of the sulphenic acid group to form intermediate **4**, which contains a five-membered isothiazolidine ring. In the complex **2**...H_2O formed before the transition state, a water molecule connects the hydrogen of the neighboring N—H group and the oxygen of the sulphenic acid group through a hydrogen bond. For the ZPVE-only corrected case

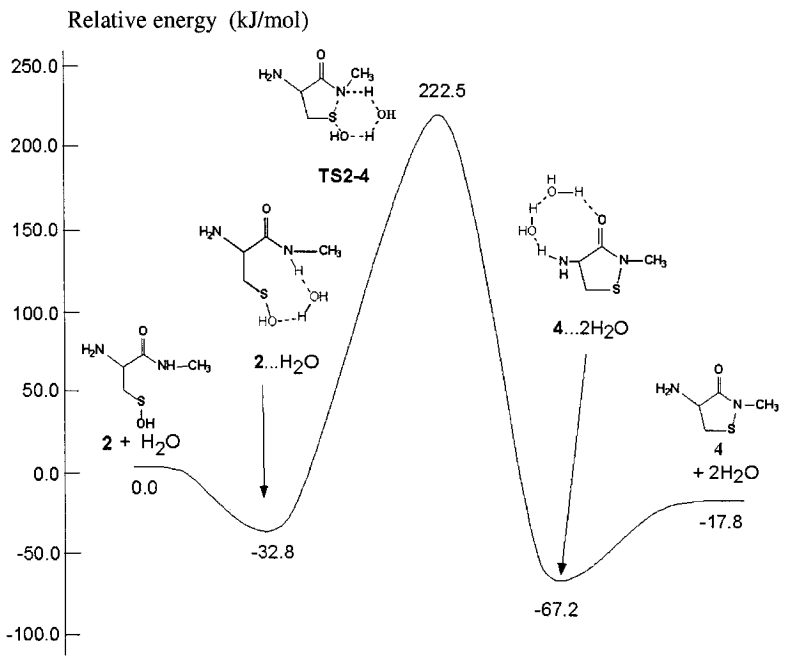


Figure 4.10. Schematic energy profile at 0 K in gas phase for the route **2→4**

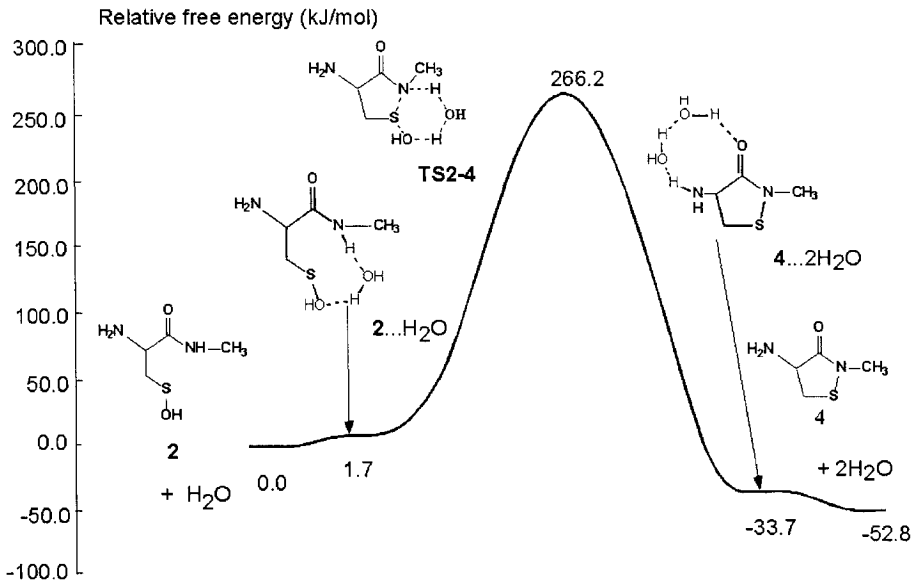


Figure 4.11. Schematic free energy profile at 298.15 K in gas phase for the route **2**→**4**

(Figure 4.10), **2**...H_2O lies 32.8 kJ/mol lower than the isolated system **2** + H_2O. A six-membered ring is formed in the transition state **TS2-4** in which the hydrogen is transferred from nitrogen to the oxygen of the sulphenic acid group and the N—S bond is formed. The relative energy of the transition state **TS2-4** is 222.5 kJ/mol. Another complex **4**...$2H_2O$ is formed after the transition state with a relative energy of -67.2 kJ/mol. The relative energy of the isolated systems **4** + $2H_2O$ is –17.8 kJ/mol.

For the entropy corrected case at 298.15 K (Figure 4.11), **2**...H_2O lies 1.7 kJ/mol above the isolated **2** + H_2O. The relative energies of **TS2-4** and **4**...$2H_2O$ are 266.2 kJ/mol and -33.7 kJ/mol, respectively. **4** + $2H_2O$ has a relative energy of –52.8 kJ/mol.

4→5

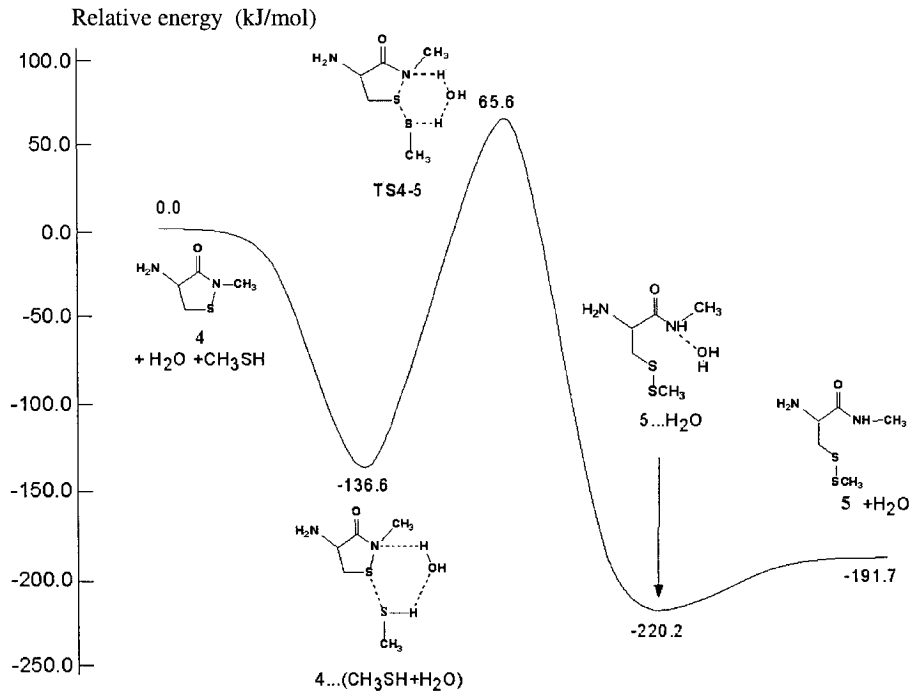


Figure 4.12. Schematic energy profile at 0 K in gas phase for the route **4→5**

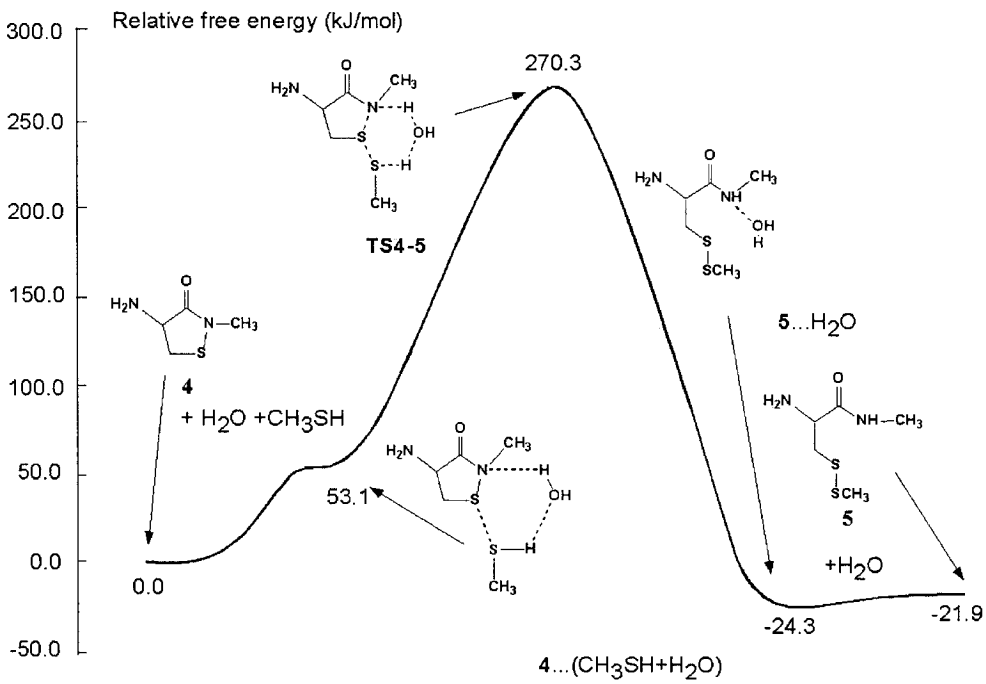


Figure 4.13. Schematic free energy profile at 298.15 K in gas phase for the route **4→5**

In the reaction **4→5** (Figure 4.12), the N—S bond is broken and the inactive form, intermediate **5,** is derived. In the complex **4**...(CH_3SH + H_2O), a water molecule acts as a bridge and connects methanethiol with the nitrogen through hydrogen bonds. For the ZPVE-only corrected case (Figure 4.12), complex **4**...(CH_3SH + H_2O) lies 136.6 kJ/mol lower than the isolated system **4** + CH_3SH + H_2O. The transition state **TS4-5** is formed with an energy 65.6 kJ/mol higher than that of **4** + CH_3SH + H_2O. In the transition state, a hydrogen is transferred from the methanethiol group to the neighboring nitrogen through a water bridge, the N—S bond is broken and the two sulfurs are bonded. Another complex **5**...H_2O is formed after the transition state with relative energy of -220.2 kJ/mol. The relative energy of isolated system **5** + H_2O is –191.7 kJ/mol.

For the entropy corrected case at 298.15 K (Figure 4.13), the relative energies of **4**...(CH_3SH + H_2O) and **TS4-5** are 53.0 kJ/mol and 270.3 kJ/mol, respectively. **5**...H_2O lies 24.3 kJ/mol below the isolated **4** + CH_3SH + H_2O. **5** + H_2O has a relative energy of –21.9 kJ/mol.

5→1

In this last step of the reaction cycle (Figure 4.14), the active site is regained. For the ZPVE-only corrected case (Figure 4.14), CH_3SH attacks the enzyme model and the complex **5**...(CH_3SH + H_2O) is formed with an energy 135.5 kJ/mol lower than that of the reactants. In the transition state **TS5-1**, which has a relative energy of 45.5 kJ/mol, a six-membered ring is formed. A proton is transferred from the attacking CH_3SH to the sulfur of the active site via a water bridge and a S—S bond is formed between the two

CH_3S— groups. Another complex **1**...(CH_3SSCH_3 + H_2O) is found after the transition state with a relative energy of -156.4 kJ/mol. Finally, the active form of the enzyme is regained in the isolated system **1** + CH_3SSCH_3 + H_2O, which has a relative energy of -125.5 kJ/mol.

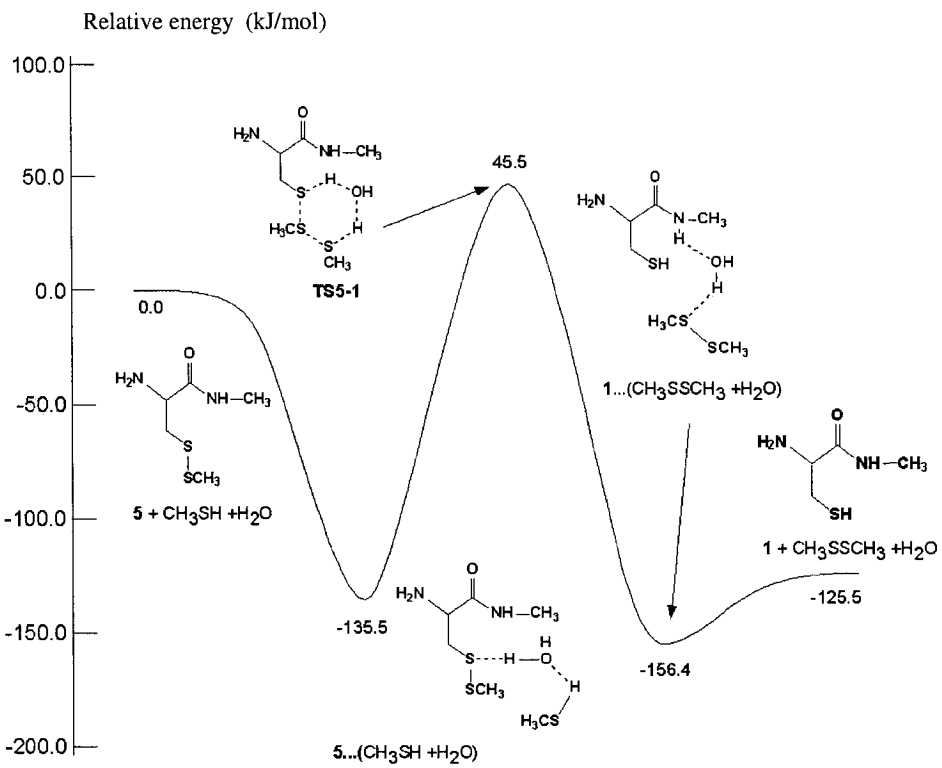


Figure 4.14. Schematic energy profile at 0 K in gas phase for the route **5**→**1**

For the entropy corrected case at 298.15 K (Figure 4.15), **5**...(CH_3SH + H_2O) lies 46.7 kJ/mol above the isolated **5** + CH_3SH + H_2O. The relative energies of **TS5-1** and **1**...(CH_3SSCH_3 + H_2O) are 239.7 kJ/mol and 22.6 kJ/mol, respectively. **1** + CH_3SSCH_3 + H_2O has a relative energy of 15.8 kJ/mol.

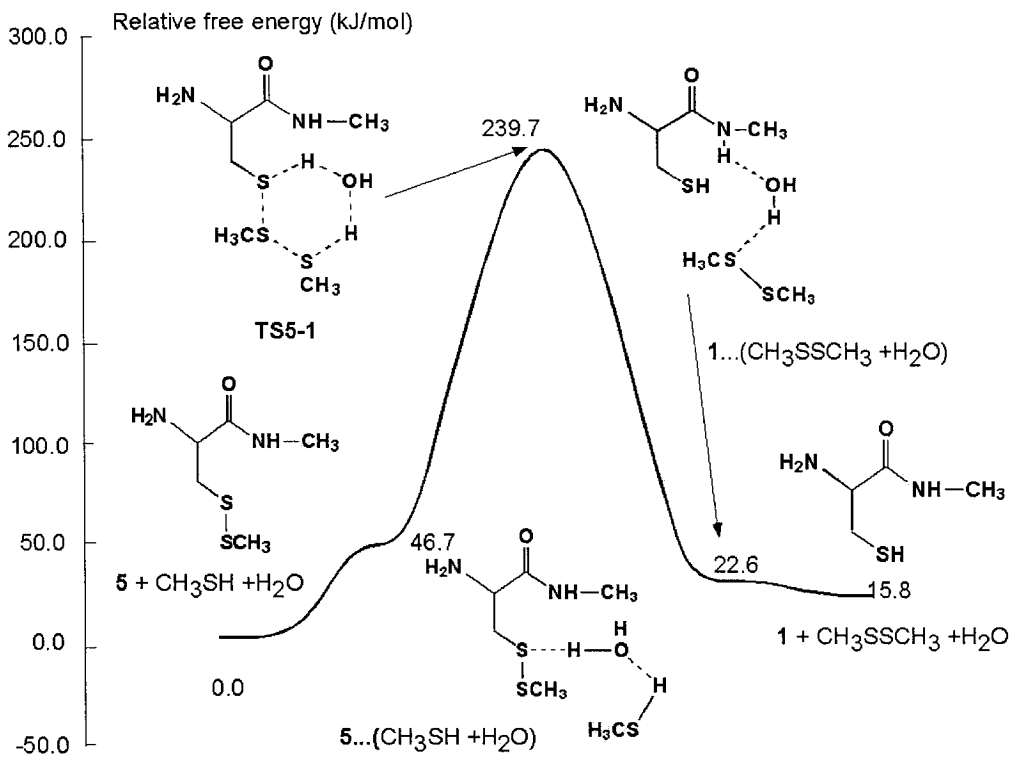


Figure 4.15. Schematic free energy profile at 298.15 K in gas phase for the route **5→1**

4.3.2 Computational results in solution

1→2

As described previously, the active site of the enzyme is attacked by a hydrogen peroxide molecule and the complex **1**…(H_2O_2 + H_2O) is formed in this reaction step.

For the ZPVE-only corrected case (Figure 4.16), the complex lies 1.6 kJ/mol below the isolated system **1** + H_2O_2 +H_2O. The reaction proceeds through transition state **TS1-2,** which has a relative energy of 90.4 kJ/mol. Following the transition state, another

complex **2**...$2H_2O$ is formed with a relative energy of –217.2 kJ/mol. The isolated system **2** + $2H_2O$ has a relative energy of –217.8 kJ/mol.

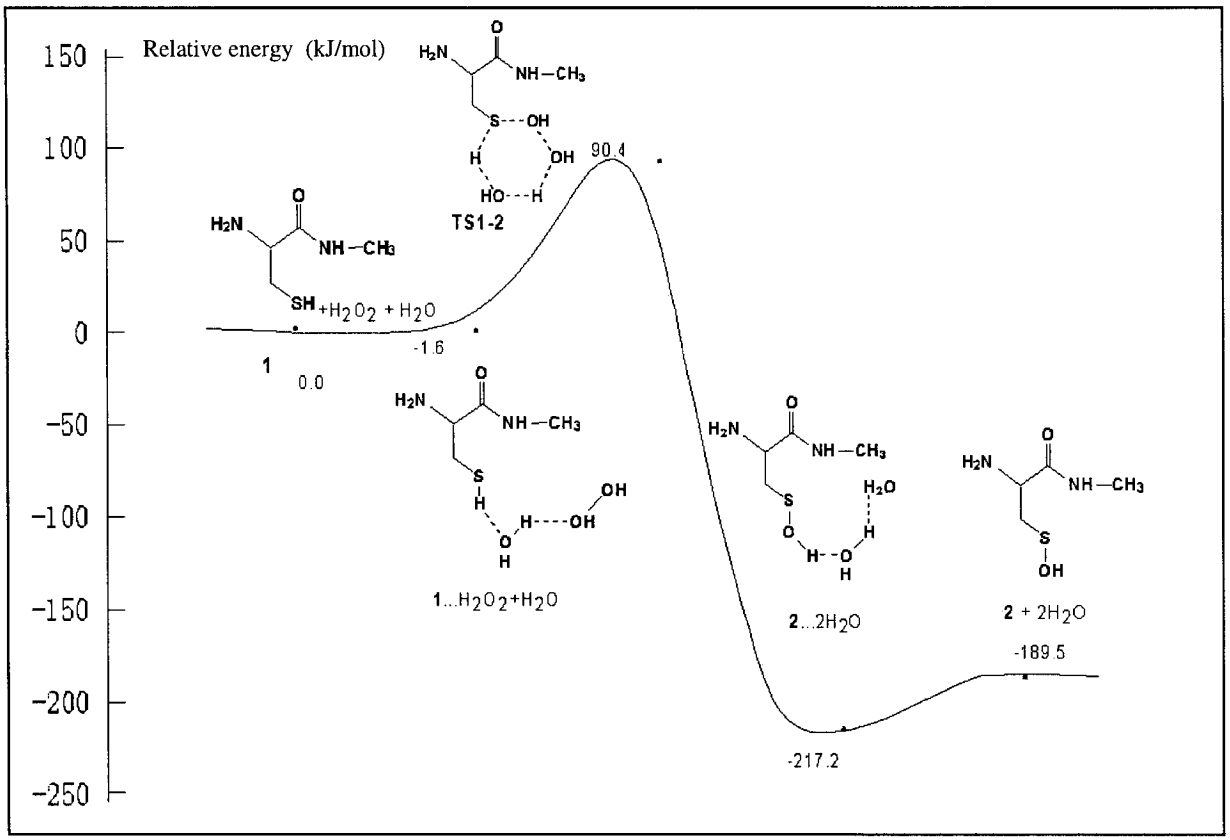


Figure 4.16. Schematic energy profile at 0 K for the route **1**→**2**

For the entropy corrected case at 298.15 K (Figure 4.17), the relative energies of **1**...(H_2O_2 + H_2O) and **TS1-2** are 76.5 kJ/mol and 178.4 kJ/mol, respectively. **2**...$2H_2O$ lies 137.0 kJ/mol below the isolated **1** + H_2O_2 + H_2O. **2** + $2H_2O$ has a relative energy of –185.6 kJ/mol.

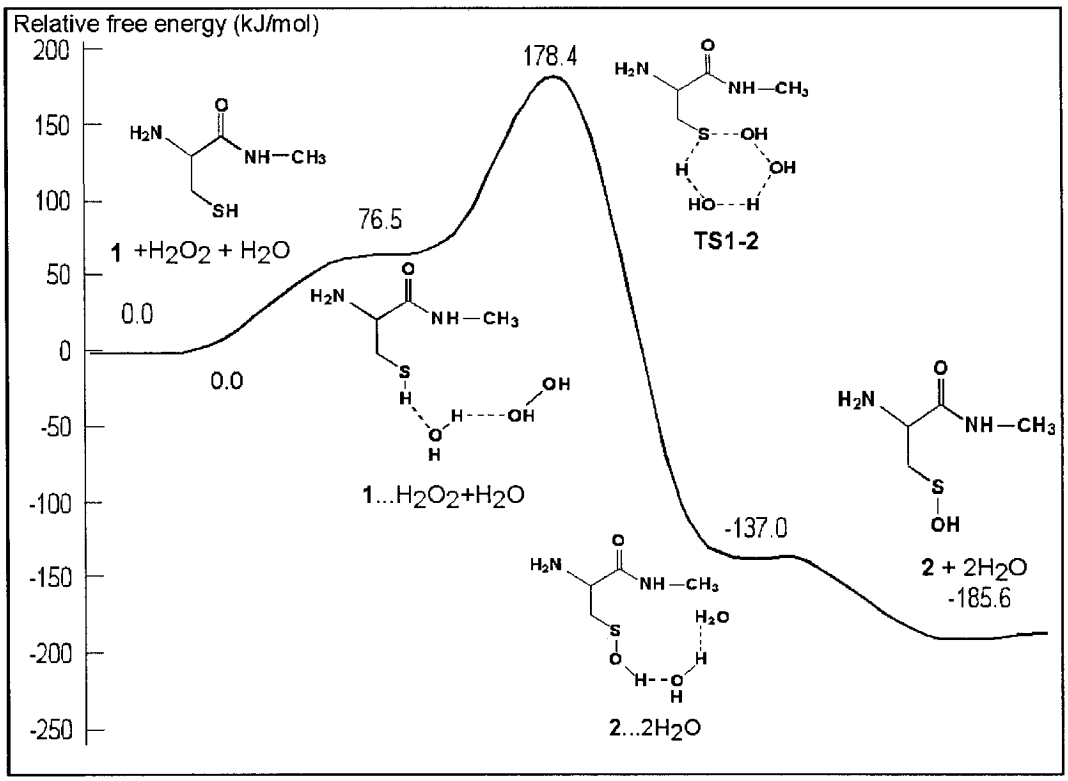


Figure 4.17. Schematic free energy profile at 298.15 K for the route **1**→**2**

2→3

In this reaction step intermediate **2** is oxidized into intermediate **3**. For the ZPVE-only corrected case (Figure 4.18), complex **2**...(H_2O + H_2O_2) lies 4.4 kJ/mol above the isolated **2** + H_2O + H_2O_2. Transition state **TS2-3** has a relative energy of 169.2 kJ/mol. The relative energies of **3**...$2H_2O$ and **3** + $2H_2O$ are –36.9 and 1.4 kJ/mol, respectively.

For the entropy corrected case at 298.15 K (Figure 4.19), the relative energies of **2**...(H_2O + H_2O_2) and **TS2-3** become 83.2 and 246.0 kJ/mol, respectively. **3**...$2H_2O$ has a relative energy of 42.8 kJ/mol. **3** + $2H_2O$ lies 5.9 kJ/mol above **2** + H_2O + H_2O_2.

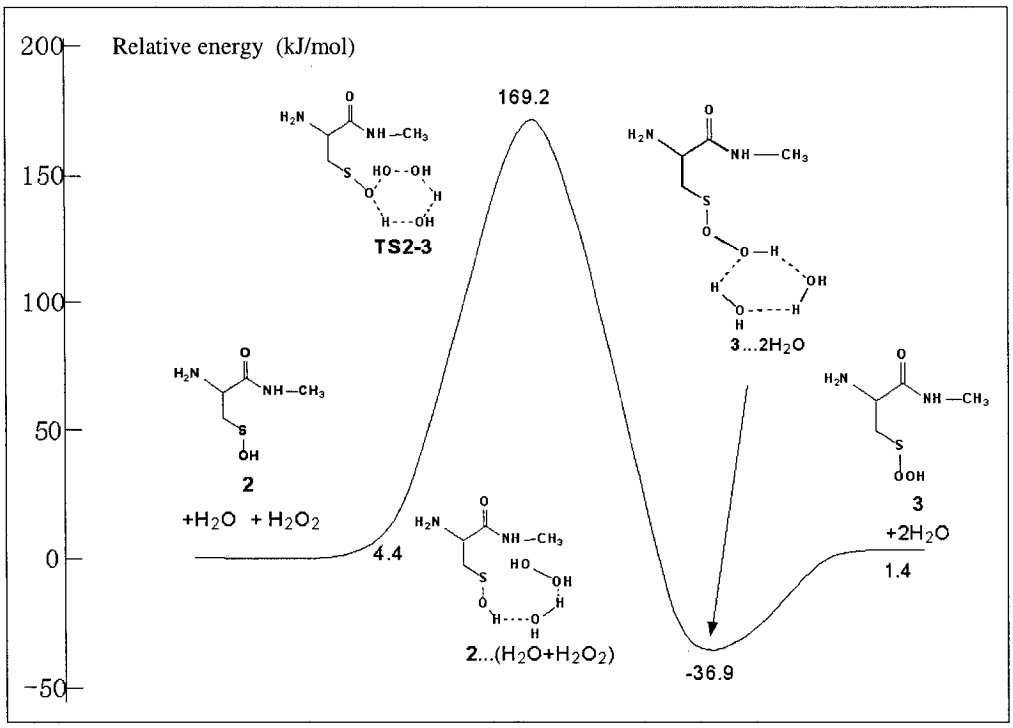


Figure 4.18. Schematic energy profile at 0 K for the route **2→3**

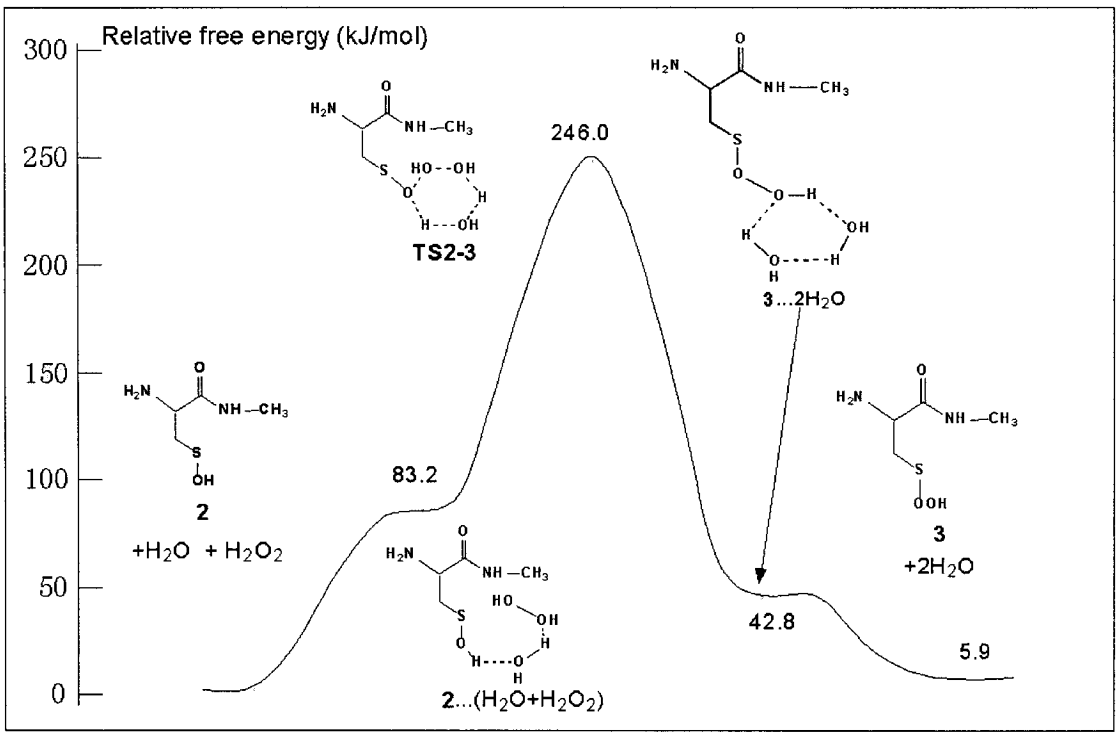


Figure 4.19. Schematic free energy profile at 298.15 K for the route **2→3**

3→4

As described in gas phase results, an isothiazolidine ring is formed in intermediate **4** through this reaction step. In the ZPVE-only corrected case (Figure 4.20), the complex **3**...H_2O lies 32.4 kJ/mol below the isolated system **3** + H_2O. The relative energy of **TS3-4** is 198.7 kJ/mol. Once the N—S bond is formed, the N—H and S—O bonds are completely broken. Complex **4**...(H_2O_2 + H_2O) is formed with a relative energy of –27.0 kJ/mol. The isolated system of **4** + H_2O + H_2O_2 has an energy 24.4 kJ/mol higher than that of **3** + H_2O.

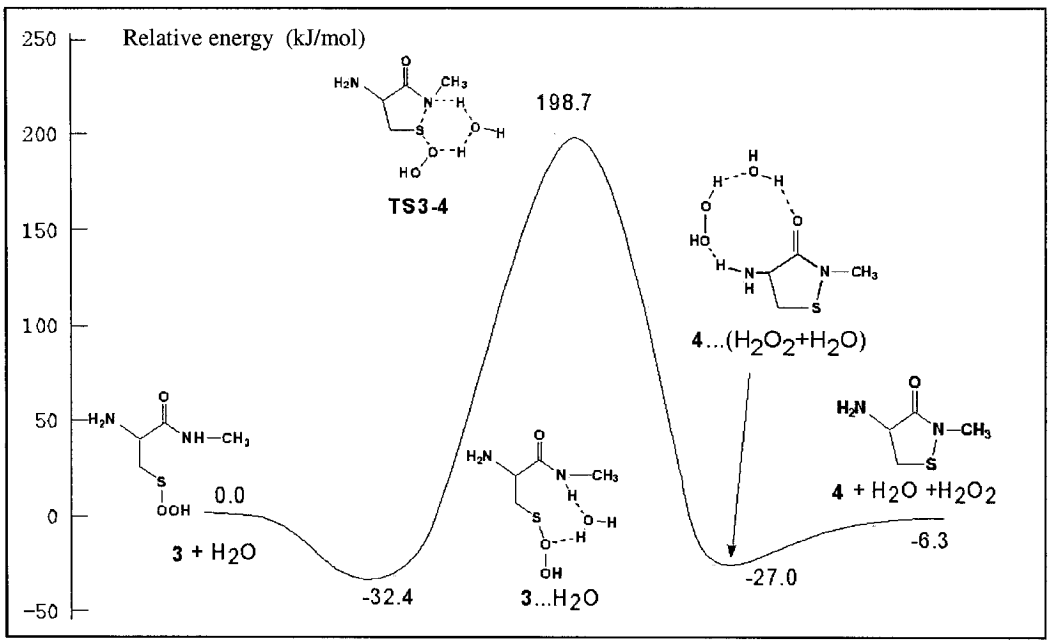


Figure 4.20. Schematic energy profile at 0 K for the route **3→4**

For the entropy corrected case at 298.15 K (Figure 4.21), the relative energies of **3**...H_2O and **TS3-4** become 3.2 kJ/mol and 240.4 kJ/mol, respectively. **4**...(H_2O_2 + H_2O) lies 8.4 kJ/mol below the isolated **3** + H_2O. **4** + H_2O + H_2O_2 has a relative energy of –45.0 kJ/mol.

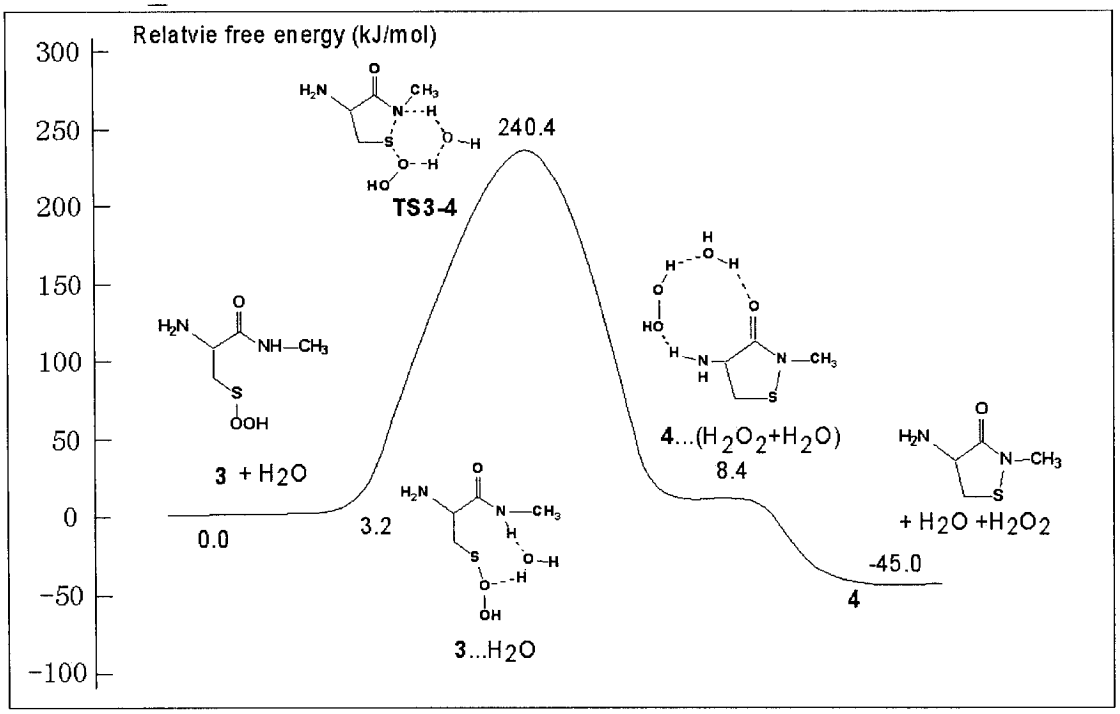


Figure 4.21. Schematic free energy profile at 298.15 K for the route **3**→**4**

2→**4**

In this reaction step, intermediate **4** is directly formed from intermediate **2**. For the ZPVE-only corrected case (Figure 4.22), **2**...H_2O lies 31.3 kJ/mol lower than the isolated system **2** + H_2O. The relative energy of the transition state **TS2-4** is 213.5 kJ/mol. Complex **4**...$2H_2O$ is formed with a relative energy of –46.9 kJ/mol. The relative energy of the isolated systems **4** + $2H_2O$ is –4.9 kJ/mol.

For the entropy corrected case at 298.15 K (Figure 4.23), **2**...H_2O lies 6.2 kJ/mol above **2** + H_2O. The relative energies of **TS2-4** and **4**...$2H_2O$ are 258.6 and –11.7 kJ/mol, respectively. **4** + $2H_2O$ has a relative energy of –39.1 kJ/mol.

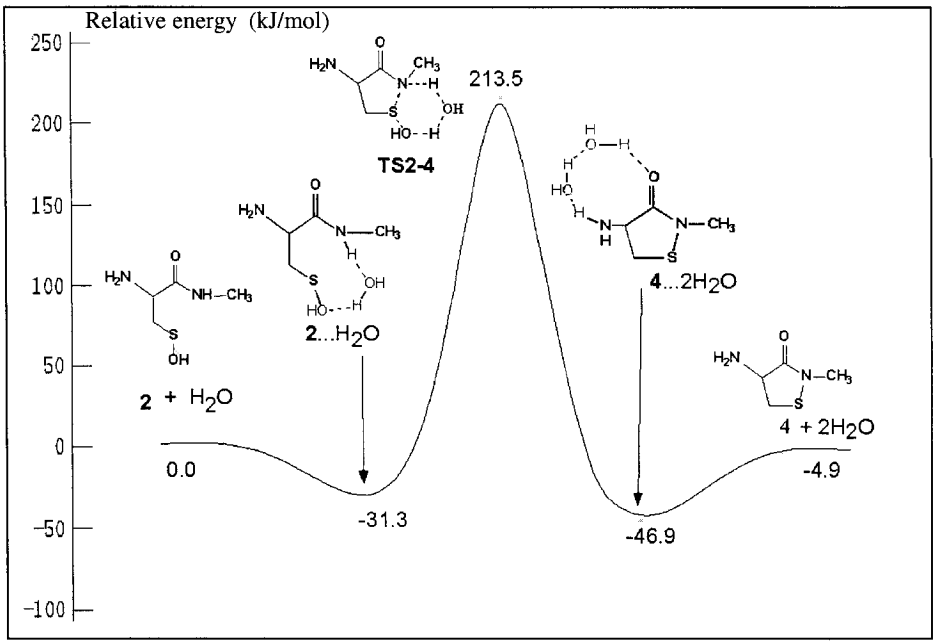


Figure 4.22. Schematic energy profile at 0 K for the route **2→4**

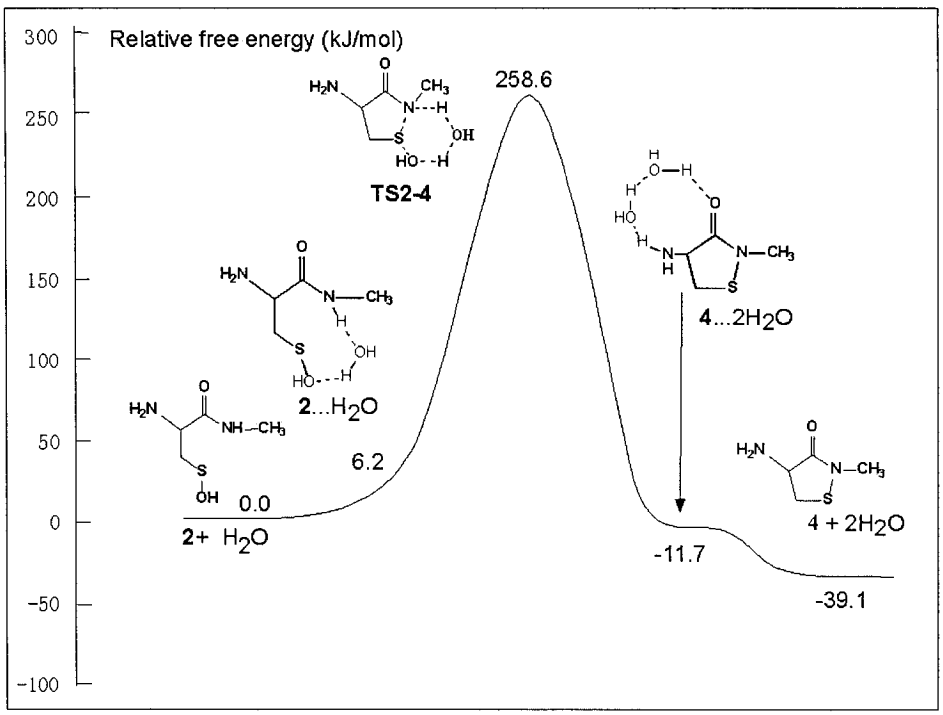


Figure 4.23. Schematic free energy profile at 298.15 K for the route **2→4**

4→5

In the reaction **4→5** the inactive form, intermediate **5,** is derived. For the ZPVE-only corrected case (Figure 4.24), the relative energy of complex **4**...(CH_3SH + H_2O) is 0.0 kJ/mol. The transition state **TS4-5** is formed with an energy 159.8 kJ/mol higher than that of **4** + CH_3SH + H_2O. Complex **5**...H_2O is formed after the transition state with relative energy of –90.0 kJ/mol. **5** + H_2O lies 72.5 kJ/mol below **4** + CH_3SH + H_2O.

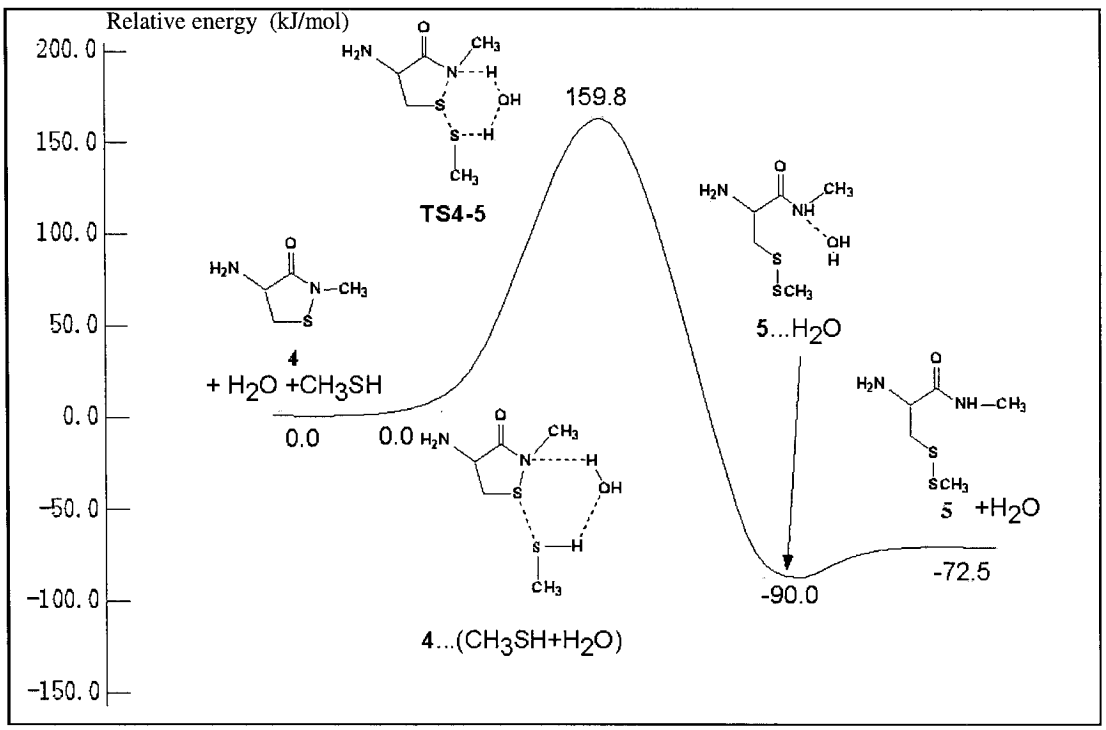


Figure 4.24. Schematic energy profile at 0 K for the route **4→5**

For the entropy corrected case at 298.15 K (Figure 4.25), the relative energies of **4**...(CH_3SH + H_2O) and **TS4-5** are 72.6 kJ/mol and 238.8 kJ/mol, respectively. **5**...H_2O lies 12.6 kJ/mol below **4** + CH_3SH + H_2O. **5** + H_2O has a relative energy of –33.4 kJ/mol.

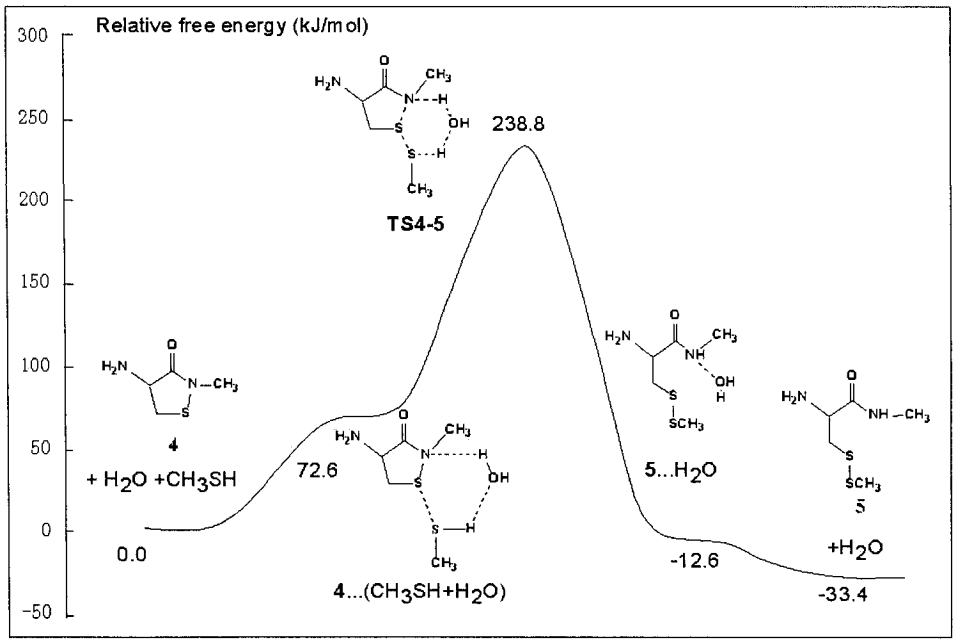


Figure 4.25. Schematic free energy profile at 298.15 K for the route **4→5**

5→1

As described in the gas phase results, the active site is regained in this last reaction step. For the ZPVE-only corrected case (Figure 4.26), complex **5**...(CH_3SH + H_2O) is formed with a relative energy of –5.5 kJ/mol. In the transition state **TS5-1** lies 171.3 kJ/mol above the isolated **5** + CH_3SH + H_2O. Complex **1**...(CH_3SSCH_3 + H_2O) is found after the transition state with a relative energy of –15.3 kJ/mol. Finally, the active form of the enzyme is regained in the isolated system **1** + CH_3SSCH_3 + H_2O with relative energy of -125.5 kJ/mol.

For the entropy corrected case at 298.15 K (Figure 4.27), the relative energies of **5**...(CH_3SH + H_2O) and **TS5-1** are 59.9 kJ/mol and 247.6 kJ/mol, respectively.

1...(CH_3SSCH_3 + H_2O) lies 60.0 kJ/mol above **5** + CH_3SH + H_2O. **1** + CH_3SSCH_3 + H_2O has a relative energy of 3.7 kJ/mol.

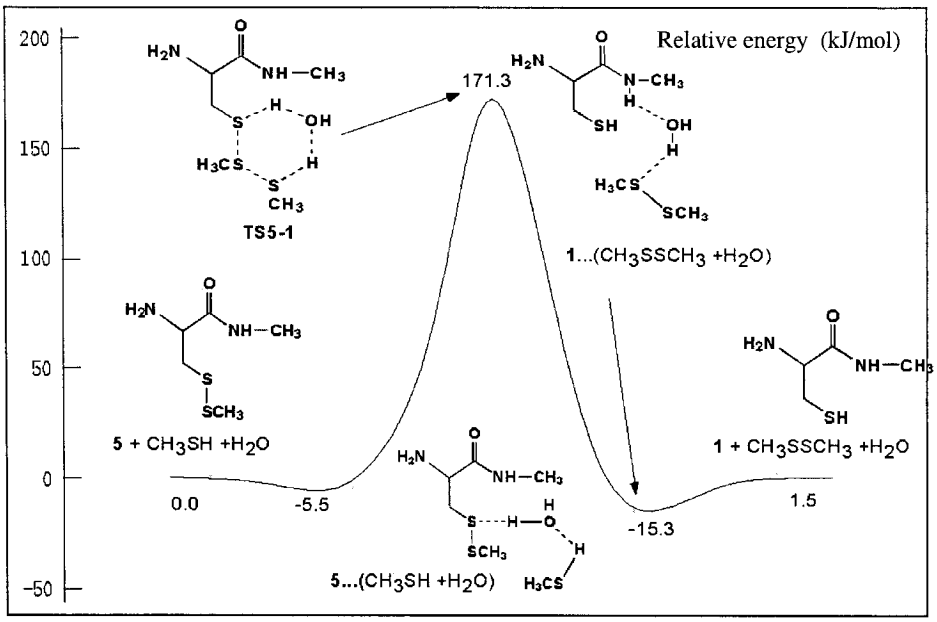


Figure 4.26. Schematic energy profile at 0 K for the route **5**→**1**

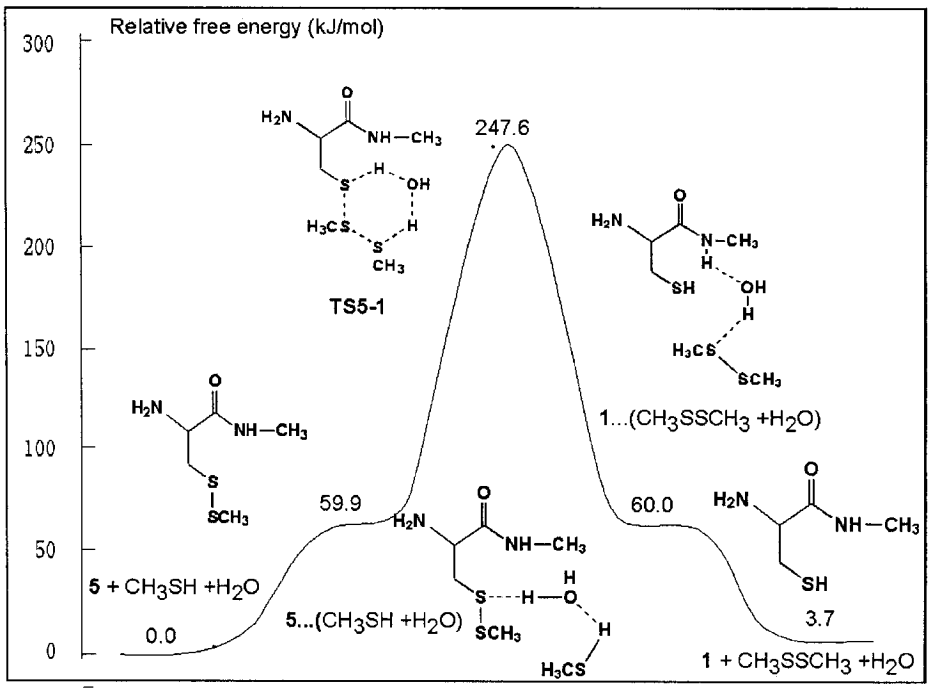


Figure 4.27. Schematic free energy profile at 298.15 K for the route **5**→**1**

The following discussions are mainly based on the entropy corrected data from the solvation calculations.

The current study involves the redox regulation of cysteine-dependent protein tyrosine phosphatases with a simplified model of the system. A sulphenylamide is formed in the transformation between the active form and the inactive form of the enzyme. Two mechanisms have been proposed to describe how sulphenylamide is derived from intermediate **2**. In the oxidative mechanism, a further oxidation step is needed to form intermediate **3** and then the backbone nitrogen in **3** attacks the S atom, leading to the formation of intermediate **4**, the sulphenylamide. In the direct mechanism, the nucleophilic attack by nitrogen leads directly to intermediate **4**.

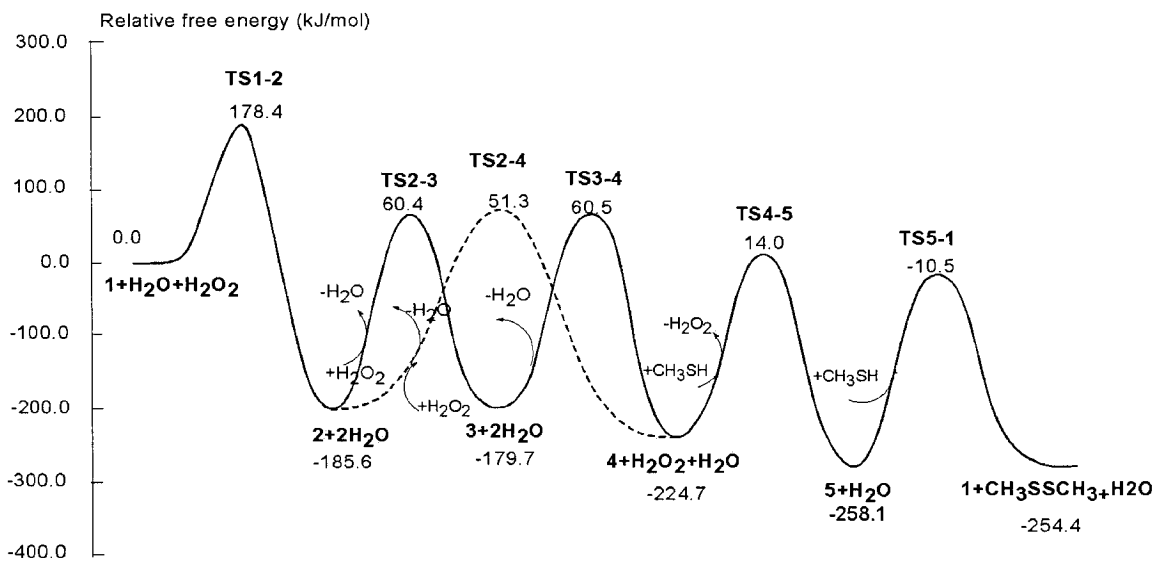


Figure 4.28. Schematic free energy profile at 298.15 K in solution for the overall reaction (The light characters in the graph denote compounds that join in or leave the system in each reaction step)

Figure 4.28 gives the overall energy profile of the reaction loop. If we take the energy barrier as the energy difference between the transition state and the complex formed right before it, then the energy barriers for reaction steps **1→2**, **2→3**, **3→4**, **2→4**, **4→5** and **5→1** would be 101.9 kJ/mol, 162.9 kJ/mol, 237.2 kJ/mol, 116.1 kJ/mol, 187.6 kJ/mol and 252.4 kJ/mol, respectively. In the first reaction step **1→2**, the energy drops by 185.6 kJ/mol, indicating that this reaction step is thermodynamically favorable. The energy barrier of reaction step **2→3** is about 90 kJ/mol lower than that of **2→4**. Comparison of the energy barriers of reaction step **2→4** and **3→4** shows that they are relatively similar in magnitude, although the energy barrier of step **3→4** is slightly bigger than that of **2→4,** the difference is not significant, and thus making **2→4** and **2→3→4** two competitive reaction routes. When the oxidizing agent is abundant, intermediate **2** may be oxidized into **3**, and then proceed to form **4**, however, if H_2O_2 is not readily available, nucleophilic attack of the neighboring nitrogen at the sulfur position may take place to form intermediate **4**. If a thiol compound is available, the reaction may continue to yield intermediate **5,** the inactive form. Additionally, **5** can undergo reduction to give the active form of the site and complete the whole reaction loop. Although **5** + H_2O + CH_3SH and **1** + $2H_2O$ + RSSR are very close to each other in energy, the newly formed RSSR is at a much lower concentration compared to the other compounds present in the system, such as water and hydrogen peroxide, and therefore the direct conversion of **1** to **5** is very unlikely to take place in *vivo*. In the first reaction step **1→2**, one water and one hydrogen peroxide come into the system and two waters are released afterwards; in the second reaction step **2→3**, one water and one hydrogen peroxide are involved and two waters are released at the end the reaction; in the next reaction step **3→4**, one water enters the

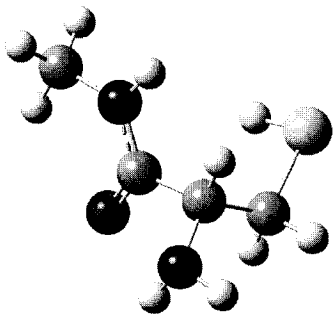
system and one water and one hydrogen peroxide leave the system in the end; in the reaction step **4→5**, one water and one methiol are added into the system and one water leaves after the reaction; in the last reaction step **5→1**, one methiol and one water take part in the reaction and one disulfide compound and one water leave in the end. The net results are that one hydrogen peroxide and two thiol compounds are converted into two water molecules and a disulfide compound and that the total energy goes down by 280 kJ/mol, indicating that the reaction is irreversible.

4.4 Conclusions

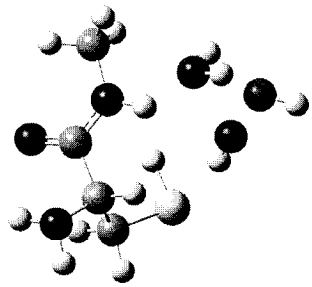
The proposed competitive mechanisms for the reactivation of the catalytic cysteine in the PTP1B active site was investigated using the B3LYP method of density functional theory. The present results indicate that the efficiency of the nucleophilic attack of the neighboring nitrogen at the sulfur position is not affected by the additional oxidation that takes place at the —SOH group and therefore the oxidative mechanism and direct mechanism are competitive with each other. Whether the oxidative mechanism has priority over the direct mechanism or not should therefore depend on the availability of the oxidizing agent H_2O_2. Additionally, due to its downhill nature, the conversion reaction is predicted to be irreversible.

4.5 **Representation of optimized structures**

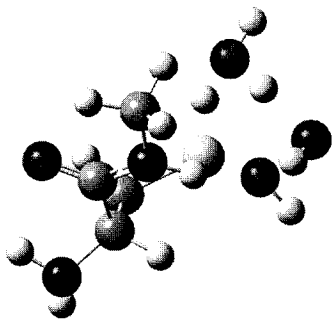
Graphical representations of the optimized structures are shown in figure 4.29 These figures were drawn with the GaussView software.



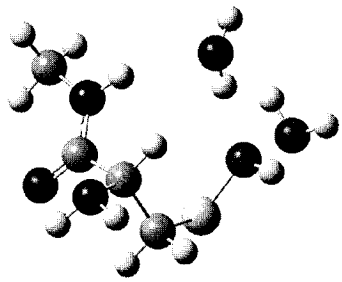
PTP



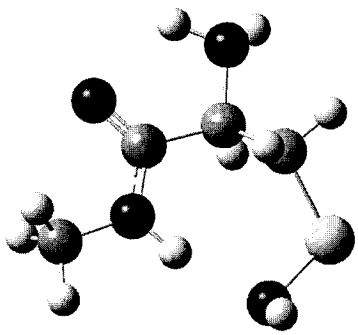
1...(H_2O_2 + H_2O)



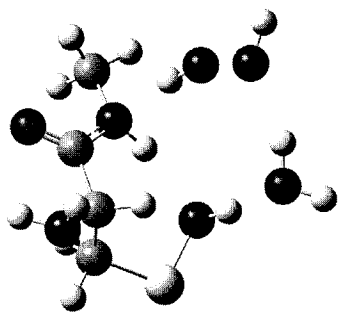
TS1-2



2...2H_2O

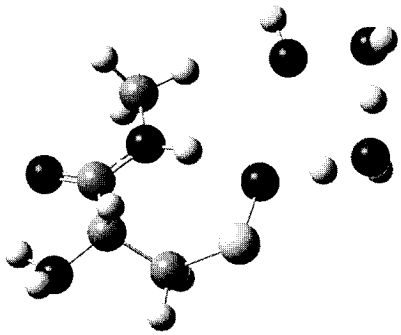


2

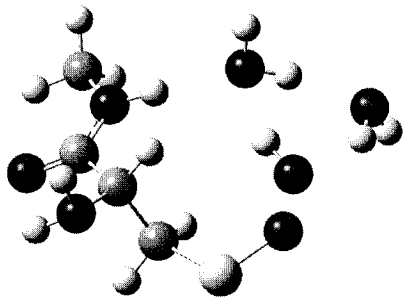


2...(H_2O_2 + H_2O)

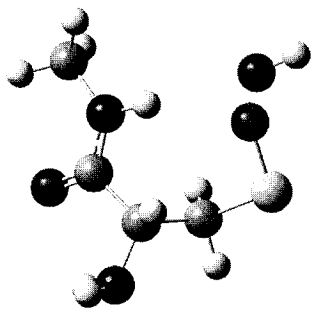
Figure 4.29. Graphical representations of the optimized structures in the redox regulation of PTP1B activity



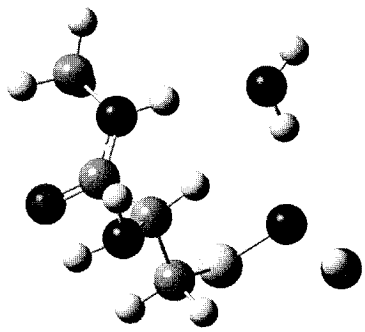
TS2-3



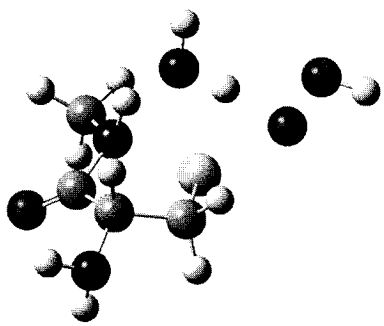
3...$2H_2O$



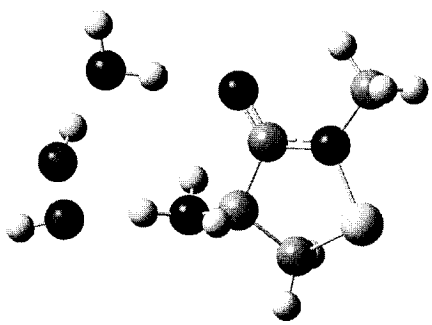
3



3...H_2O

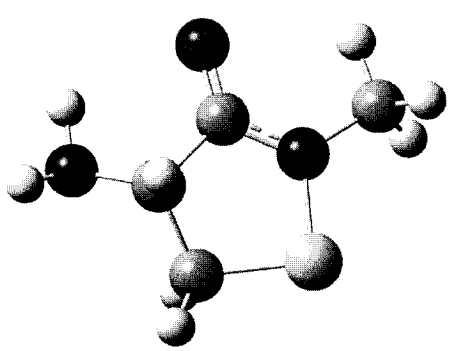


TS3-4

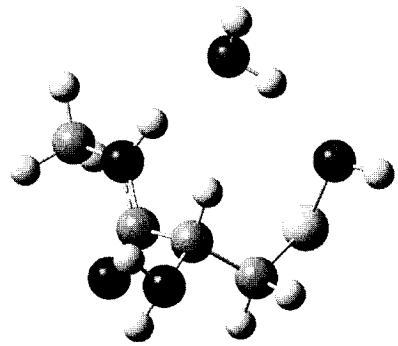


4...(H_2O_2 + H_2O)

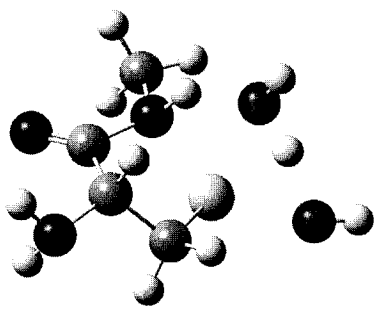
Figure 4.29. Graphical representations of the optimized structures in the redox regulation of PTP1B activity (continued)



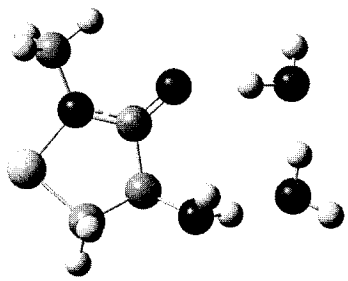
4



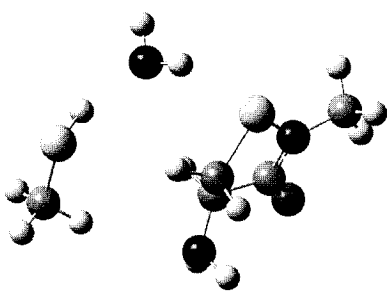
4...H_2O



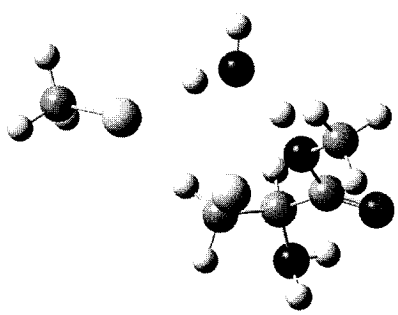
TS2-4



4...$2H_2O$

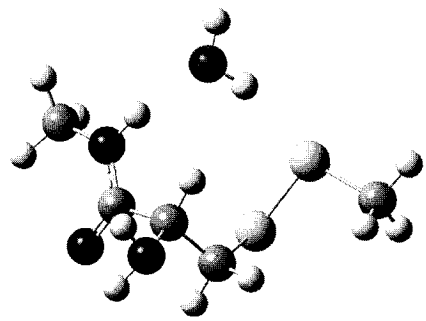


4...($CH_3SH + H_2O$)

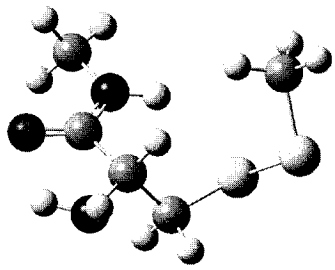


TS4-5

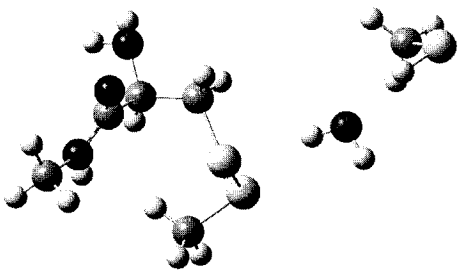
Figure 4.29. Graphical representations of the optimized structures in the redox regulation of PTP1B activity (continued)



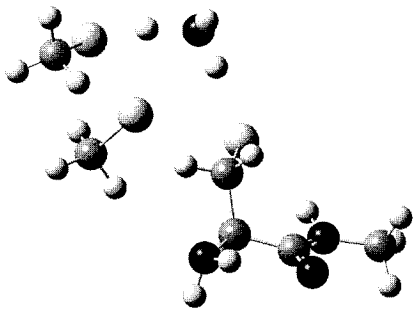
5...H_2O



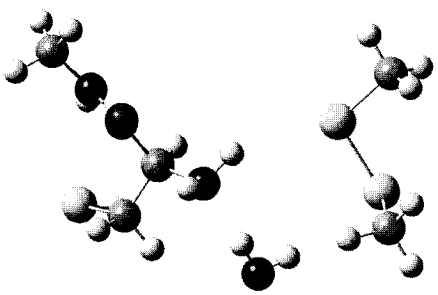
5



5...($CH_3SH + H_2O$)



TS5-1



5...($CH_3S_2CH_3+H_2O$)

Figure 4.29. Graphical representations of the optimized structures in the redox regulation of PTP1B activity (continued)

Chapter 5

Modeling the Reaction Mechanism of the Reversible Isomerization of R5P to Ru5P

5.1 Introduction

Ribose-5-phosphate isomerases (RPIs) enzymes, present in all living cells, catalyze the reversible conversion of ribose-5-phosphate (R5P) to ribulose-5-phosphate (Ru5P). The reaction is involved in the pentose phosphate pathway, in which RPI catalyzes the final step of the conversion of glucose-6-phosphate into ribose-5-phosphates, and in the last step to generate ribulose-1, 5-bisphosphate the accepter of CO_2 in the pentose-phosphate cycle (Calvin cycle).[128] R5P is crucial for the synthesis of various amino acids, pyrimidine and purine nucleotides, and DNA,[129] while Ru5P is the riboflavin precursor.[130] Although both R5P and Ru5P can be generated through several other pathways, it has been shown that the RPI activity is essential for cell viability. Elimination of this enzyme has been shown to be lethal.[131] There is always at least one of the RPI enzymes present in an organism. RPI's sequence is highly conserved in species ranging from bacteria to animals. Studies indicate that its activity is important in bacterial growth[132] and that deficiency in human RPI can lead to extensive brain abnormalities.[133]

Two completely distinctive types of RPI (RPI A and RPI B) have been identified.[132] The RPI A occurs in all forms of life, whereas RPI B has so far only been found in the genomes of some bacteria and protozoa.[134] Even though both enzymes catalyze the same

isomerase reaction, it has been shown that RPI A and RPI B are quite different in structure and active site residues.[135]

Although the RPI activity has been described for a long time, little is known about the catalytic mechanisms of the RPI A and RPI B enzymes. Some crystallography and isotope exchange studies have suggested that the conversion of R5P to Ru5P catalyzed by RPI initiates with a ring opening of the substrate and is followed by isomerization of the open chain form. It is proposed that the conversion proceeds via a proton-transfer mechanism that involves a 1,2-cis-enediolate high-energy intermediate (Figure 5.1).[136] In the proposed mechanism,[137] the furanose form of R5P is reversibly converted into the open chain form and then a deprotonation reaction takes place at the C2 position due to the attack of a base, which in most cases is a glutamic residue from the catalytic site of the enzyme. In the next step, a proton is transferred from the C2 oxygen to the C1 oxygen and then the proton that has been taken away from the C1 position is bonded to the C2 position by the catalytic base and Ru5P is eventually formed (Figure 5.2).

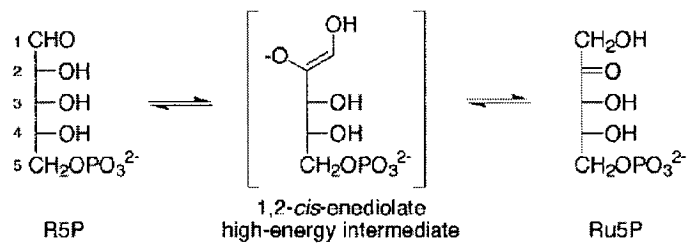


Figure 5.1. Isomerization reaction of R5P to Ru5P

In the current work, quantum chemical DFT methods were carried out to evaluate the proposed reaction mechanism and to provide insight into the isomerization process catalyzed by RPI enzymes. A previous experimental study on RPI enzymes showed that the average distance between the amino acid residues of the catalytic site and substrate is about 3 Å, and that water molecules exist in the catalytic site, acting as bridges connecting the amino acid residues and the substrates.[136] It is then reasonable to suppose that water molecules act as the medium and directly react with the substrate to facilitate the whole catalytic reaction.

The simplified model chosen in the current study mainly deals with the reactions between the sugar substrate and water molecules that directly contact the substrate in the catalytic site. All other amino acid residues are neglected except one glutamic residue that acts as the key catalytic base. Due to the absence of amino acids from the catalytic site that coordinates the phosphate group, the substrate in the present computational model can no longer be fixed as it is in the real situation. Thus the negatively charged phosphate group will affect the proton transfer processes. Therefore, two protons are added to the phosphate group to make it neutral. The calculations presented here show that this modification is both necessary and effective, for the modified phosphate group stays relaxed throughout the whole reaction process.

The present results show that the enediolate intermediate possesses an abnormally high relative energy due to the separation of opposite charges, but the reaction can still take place under mild temperature, because the glutamic residue ion accepts the proton

transferred from the substrate through a water bridge, and therefore the separation of opposite charges that leads to the great energy increase is avoided.

In the present discussion, the R5P molecule, all three intermediates and Ru5P are named **1** through **5** for convenience.

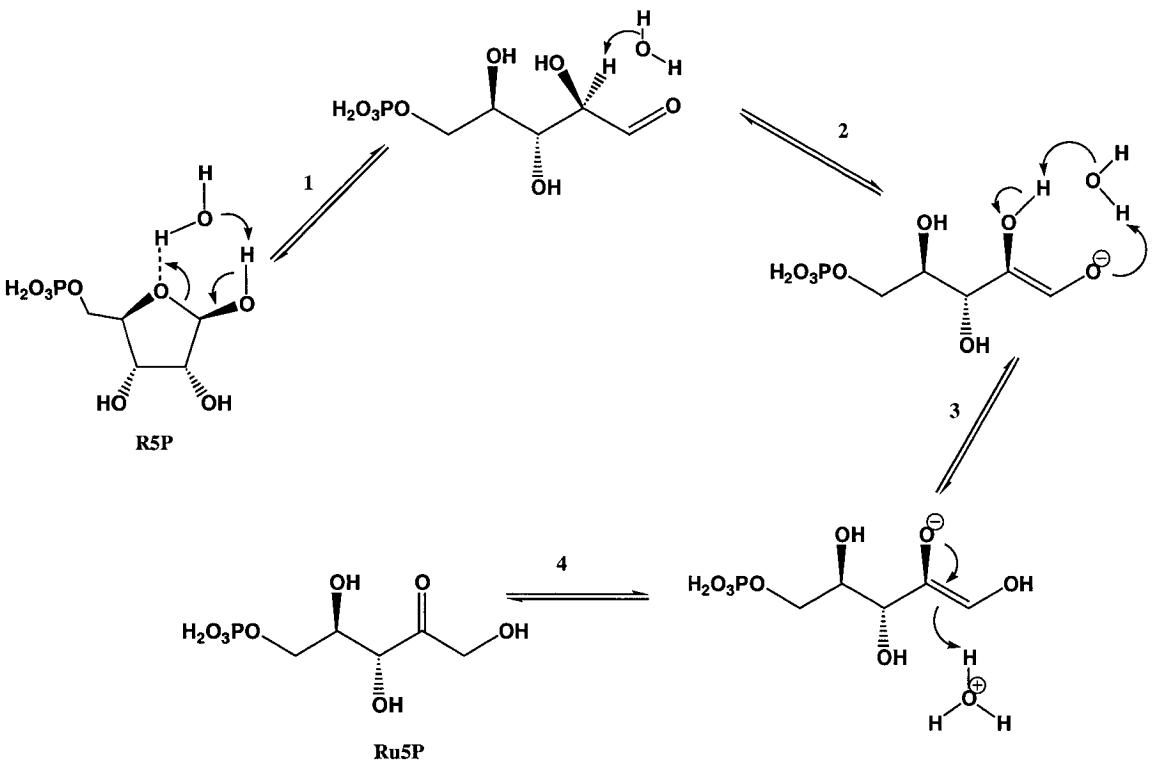


Figure 5.2. Proposed mechanism of the isomerization reaction of R5P to Ru5P catalyzed by RPI enzyme

5.2 Computational methods

All geometry optimizations and frequency calculations were performed with the B3LYP hybrid density functional in conjunction with the 6-31G(d, p) basis set using the Gaussian 03 suite of programs.[127] Single point calculations were performed with B3LYP level of theory and Pople's 6-311++G(2df,2p) basis set using the above geometries and by

including the zero-point vibrational energy, i.e., B3LYP/6-311++G(2df,2P)//B3LYP/6-31G(d, p) +ZPVE. Free energy corrections derived from the frequency calculations were added to the single point electronic energies to obtain relative free energies for all reactants, products, intermediates and transition states. All energies are in kJ/mol. Intrinsic reaction path (IRC) calculations were performed on every transition state to confirm that the transition state connects the minima of interest.

5.3 Results and discussion

5.3.1. Reaction step R5P→ Intermediate **2**

In the first reaction step, ring opening of the cyclic R5P takes place, initializing the isomerization process. For the ZPVE-only corrected case (Figure 5.3), a water molecule is attached to the hydroxyl group at the C1 position through a hydrogen bond, forming a complex **1**...H_2O, which lies 20.4 kJ/mol higher than the isolated system **1** + H_2O. Transition state **TS1-2** is then formed with a relative energy of 122.5 kJ/mol. A six-membered ring is contained in the transition state to facilitate the proton transfer from the hydroxyl group of the C1 position to the C4 oxygen. Meanwhile, the C—O bond between C1 and the C4 oxygen is breaking. Following the transition state, another complex **2**...H_2O, lying 18.4 kJ/mol below **1** + H_2O, is formed in which a water molecule is attached to the oxygen at the C4 position. The derived isolated system **2**+H_2O has a relative energy of -15.6 kJ/mol.

For the entropy corrected case at 298.15 K (Figure 5.4), the relative energies of **1**...H_2O and **TS1-2** are –21.5 kJ/mol and 80.9 kJ/mol, respectively. **2**...H_2O lies 39.3 kJ/mol below the isolated **1** + H_2O, while **2**+H_2O has a relative energy of 9.4 kJ/mol.

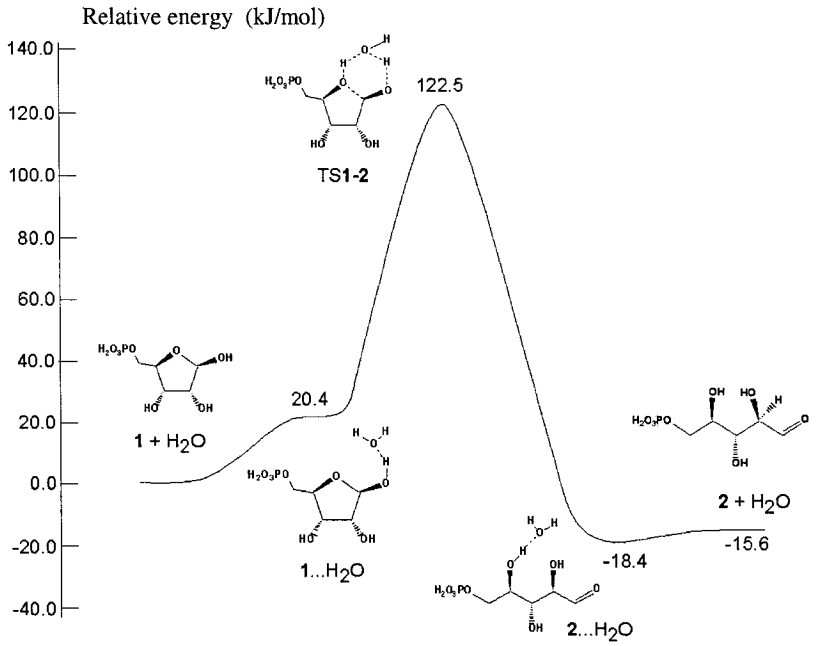


Figure 5.3. Schematic energy profile at 0 K for the route **1**→**2**

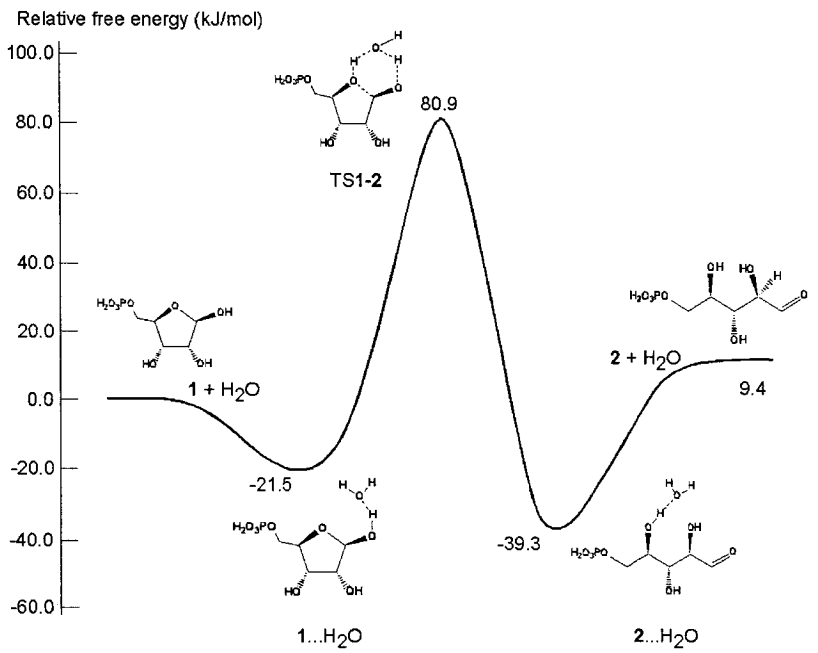


Figure 5.4. Schematic free energy profile at 298.15 K for the route **1**→**2**

5.3.2. Reaction step Intermediate **2**→ Intermediate **3**

In this step, the newly formed intermediate **2** undergoes deprotonation at the C2 position. Initially, a water molecule is attached to the hydrogen of the C2 position through a hydrogen bond, forming a complex **2**...H_2O. For the ZPVE-only corrected case (Figure 5.5), this complex has a relative energy of 2.6 kJ/mol. In the transition state **TS2-3** of this step, the C2 hydrogen is transferred to the water molecule. The transition state lies 96.3 kJ/mol higher than the isolated system **2**+H_2O. Following the transition state **TS2-3**, another complex **3**...H_3O^+ is formed with a relative energy of 22.8 kJ/mol. Without the presence of a catalytic base, this reaction can not proceed any further, because the isolated system **3** + H_2O has an extremely high relative energy of 682.6 kJ/mol.

For the entropy corrected case at 298.15 K (Figure 5.6), the relative energies of **2**...H_2O and **TS2-3** are 8.7 kJ/mol and 108.3 kJ/mol, respectively. **3**...H_3O^+ lies 25.7 kJ/mol above **2**+H_2O and **3** + H_2O has a relative energy of 710.0 kJ/mol.

This result is reasonable because of the separation of opposite charges and the catalytic base Glu^- is not included in this system. Now, as is shown in Figure 5.7, the water acts not as a base, but as a bridge, and the proton is transferred from the C2 position to the Glu^- ion. The protonation of the Glu^- ion lowers the total energy significantly. The isolated system **3** + H_2O + Glu lies 16.0 kJ/mol lower than **2** + H_2O + Glu^- for the entropy corrected case at 298.15 K (Figure 5.7) and 76.3 kJ/mol higher than **2** + H_2O + Glu^- for the ZPVE-only corrected case (Figure 5.6). One can see that with the presence of a Glu^- acting as the catalytic base this reaction step proceeds very easily.

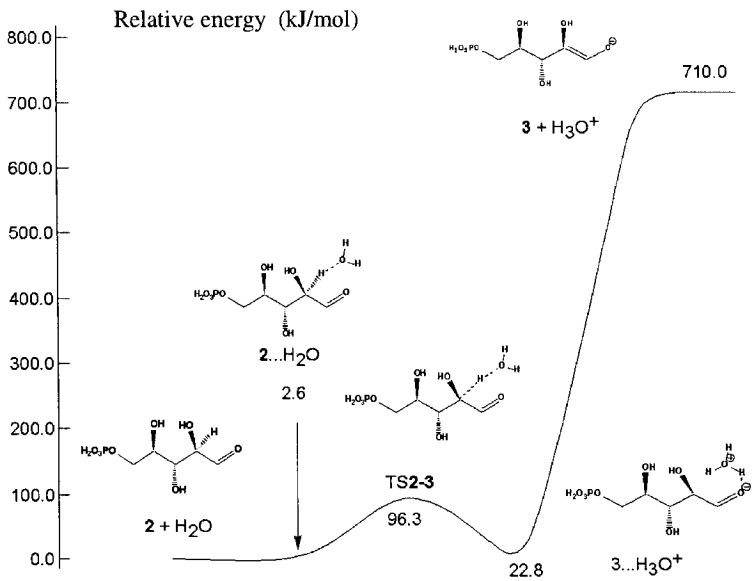


Figure 5.5. Schematic energy profile at 0 K for the route **2→3** (water catalyzed mechanism)

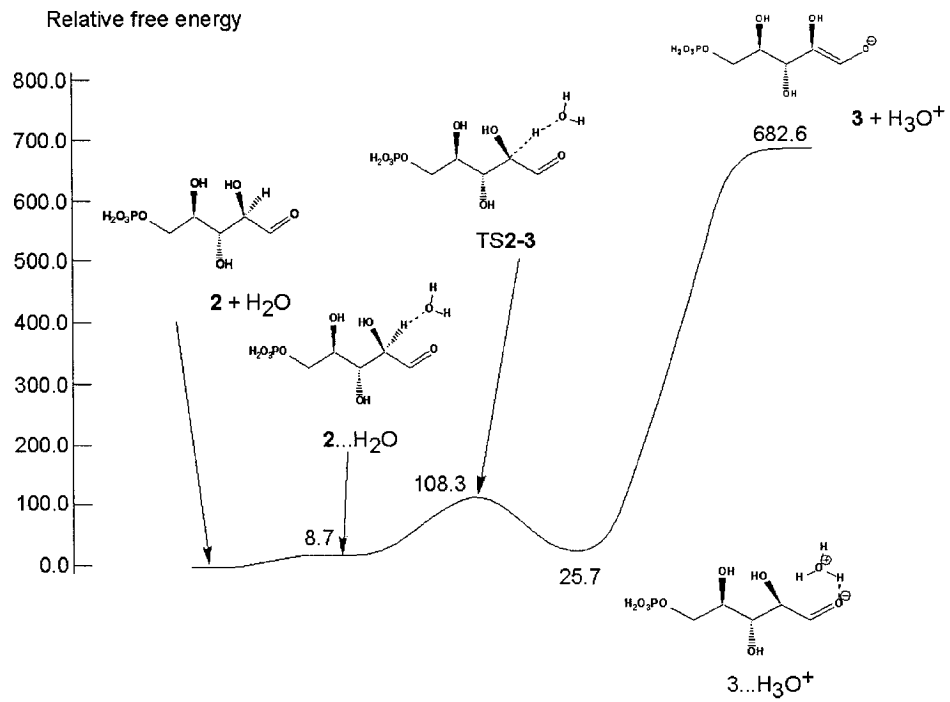


Figure 5.6. Schematic free energy profile at 298.15 K for the route **2→3** (water catalyzed mechanism)

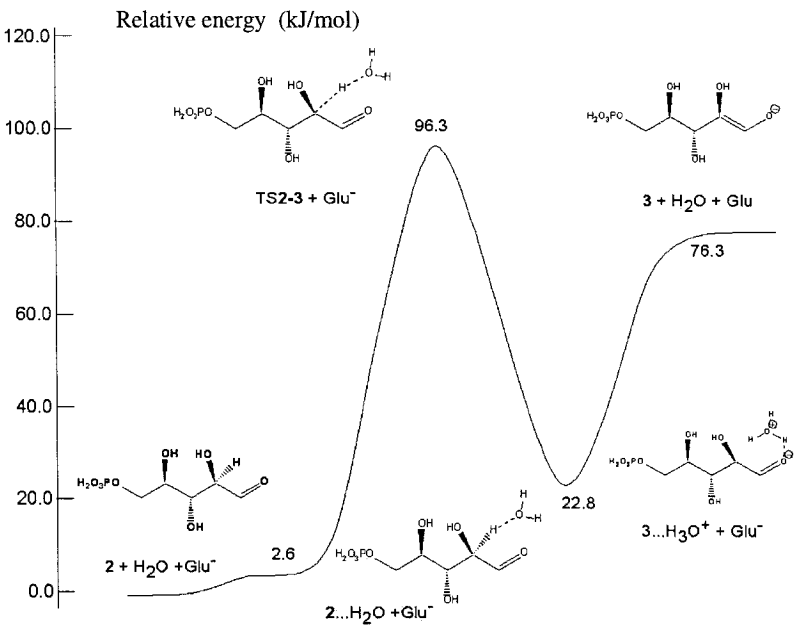


Figure 5.7. Schematic energy profile at 0 K for the route **2→3** (Glu^- catalyzed mechanism)

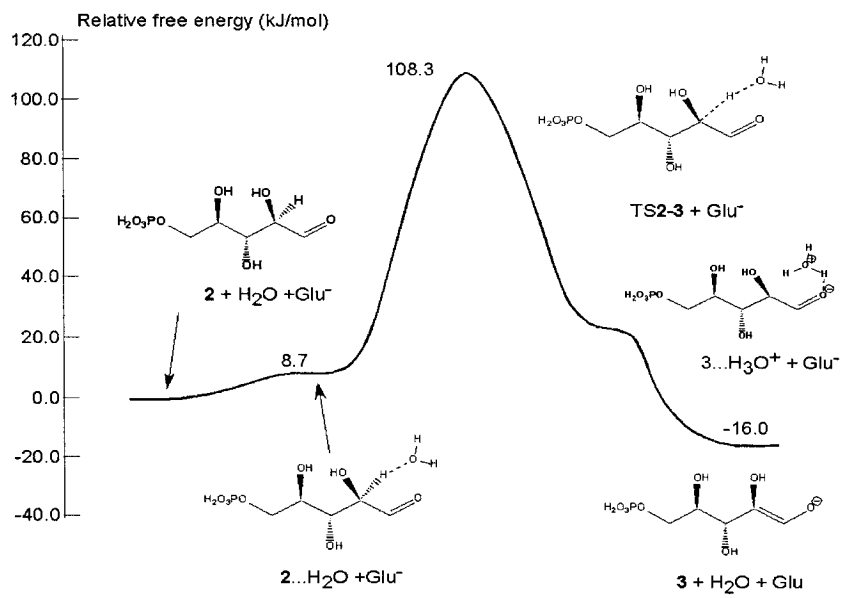


Figure 5.8. Schematic free energy profile at 298.15 K for the route **2→3** (Glu^- catalyzed mechanism)

5.3.3. Reaction step Intermediate **3**→ Intermediate **4**

In this reaction step, a proton is transferred from the hydroxyl group of the C2 position to the oxygen of the C1 position and thus the negative charge is relocated to the C2 oxygen. The Mulliken population analysis of intermediate **3** and **4** shows that C1 becomes more negative as well, which facilitates the protonation at the C1 position, a reaction that takes place in the next step. Initially, a water molecule is attached to the C1 oxygen and a complex **3**...H_2O is formed. For the ZPVE-only corrected case (Figure 5.9), this complex lies 7.0 kJ/mol lower than the isolated **3** + H_2O. Next, transition state **TS3-4** is formed with a relative energy of -3.2 kJ/mol. In the transition state, the water molecule attached to the C1 oxygen transfers a proton from the hydroxyl group at the C2 position to the C1 oxygen. Following the transition state, another complex **4**...H_2O is formed with a relative energy of -30 kJ/mol. Intermediate **4** is derived after the complex **4**...H_2O. The isolated system **4** + H_2O lies 18.4 kJ/mol below **3** + H_2O. As is shown in figure 5.9, the energy barrier in this step is very small, and the reaction is themodynamically favored. These results suggest that, once intermediate **3** is formed, it will quickly undergo proton transfer to give intermediate **4**.

For the entropy corrected case at 298.15 K (Figure 5.10), the relative energies of **3**...H_2O, **TS3-4**, and **4**...H_2O are –7.0 kJ/mol, -3.2 kJ/mol and –30.0 kJ/mol, respectively. The isolated **4** + H_2O lies 18.4 kJ/mol below **3** + H_2O.

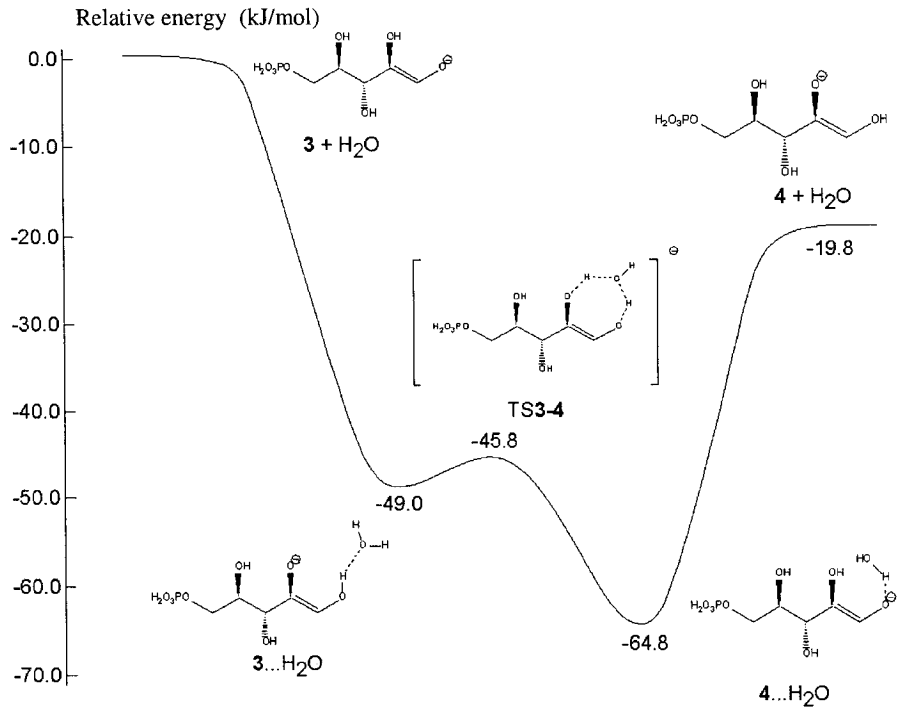


Figure 5.9. Schematic energy profile at 0 K for the route **3**→**4**

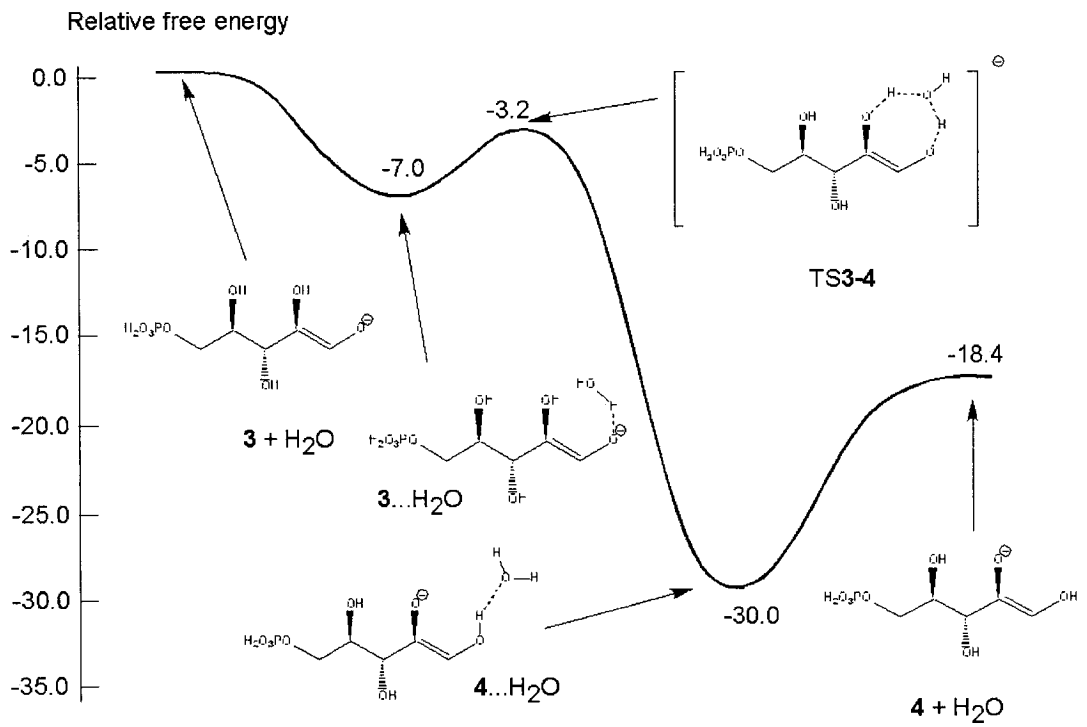


Figure 5.10. Schematic free energy profile at 298.15 K for the route **3**→**4**

5.3.4. Reaction step Intermediate **4**→ Ru5P

Without the presence of the protonated Glu residue, this reaction step begins with the isolated system **4** + H_2O. For the ZPVE-only corrected case (Figure 5.11), a complex **4**...H_3O^+ is formed with a relative energy of -706.6 kJ/mol. In the transition state **TS4-5**, a proton is transferred back to the substrate, but to a different location, the C1 position. The transition state lies 525.7 kJ/mol lower than the isolated system **4** + H_2O. After the transition state, another complex **5**...H_2O is formed with a relative energy of -666.0 kJ/mol. In this complex, the H_3O^+ loses a proton to the substrate and the resulting water molecule attaches to the hydrogen at C1 position. Then, the final product Ru5P is formed and the isolated system **5** + H_2O lies 696.8 kJ/mol lower than **4** + H_3O^+.

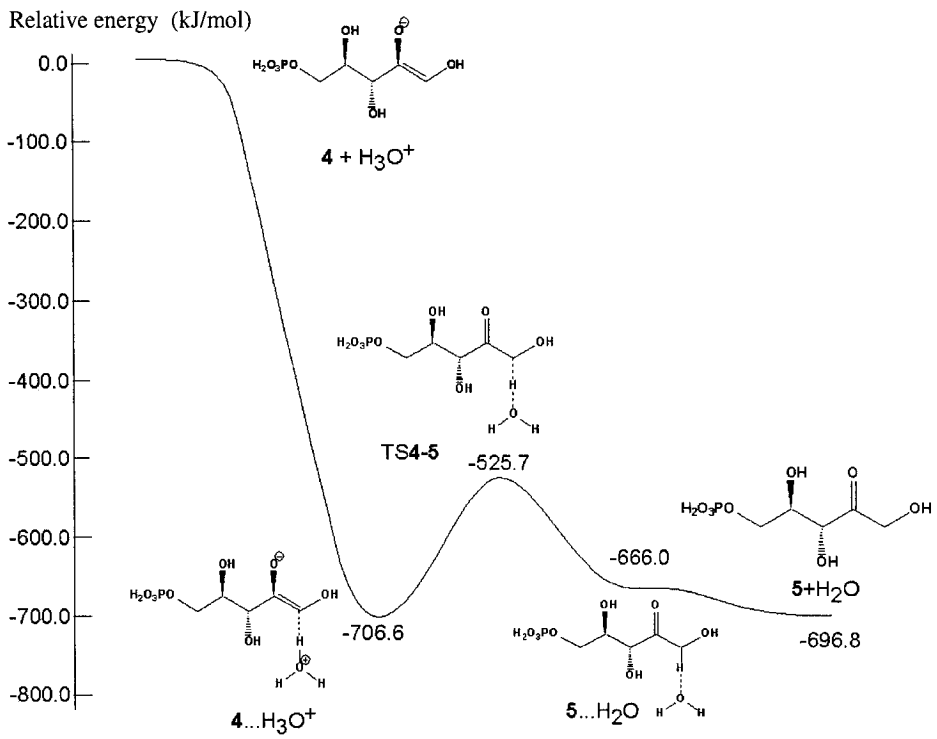


Figure 5.11. Schematic energy profile at 0 K for the route **3**→**4**

For the entropy corrected case at 298.15 K (Figure 5.12), the relative energies of **4**...H_3O^+ and **5**...H_2O are –672.0 kJ/mol and –634.4 kJ/mol, respectively. **TS4-5** lies 484.6 kJ/mol below the isolated **4** + H_2O and **5** + H_2O^+ has a relative energy of –698.0 kJ/mol.

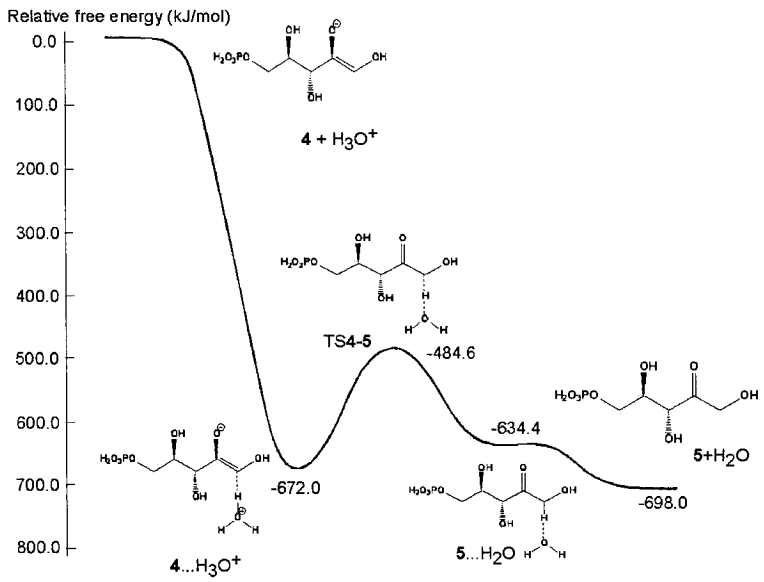


Figure 5.12. Schematic free energy profile at 298.15 K for the route **3**→**4**

Once the protonated Glu takes part in this reaction step, the energy changes become much smaller. A proton is transferred back from the catalytic base Glu to the C1 position of the substrate. For the ZPVE-only corrected case (Figure 5.13), **4**...H_3O^+ + Glu^- lies 73.1 kJ/mol below **4** + H_2O + Glu. The relative energies of **TS4-5** + Glu^- and **5**...H_2O + Glu^- are 107.8 kJ/mol and –32.6 kJ/mol, respectively. **5** + H_2O + Glu^- lies 63.4 kJ/mol below **4** + H_2O + Glu.

For the entropy corrected case at 298.15 K (Figure 5.14), the relative energies of **4**...H_3O^+ + Glu^-, **TS4-5** + Glu^- and **5**...H_2O + Glu^- are 59.4 kJ/mol, 246.4 kJ/mol and 96.9 kJ/mol, respectively. **5** + H_2O + Glu^- lies 33.3 kJ/mol below **4** + H_2O + Glu.

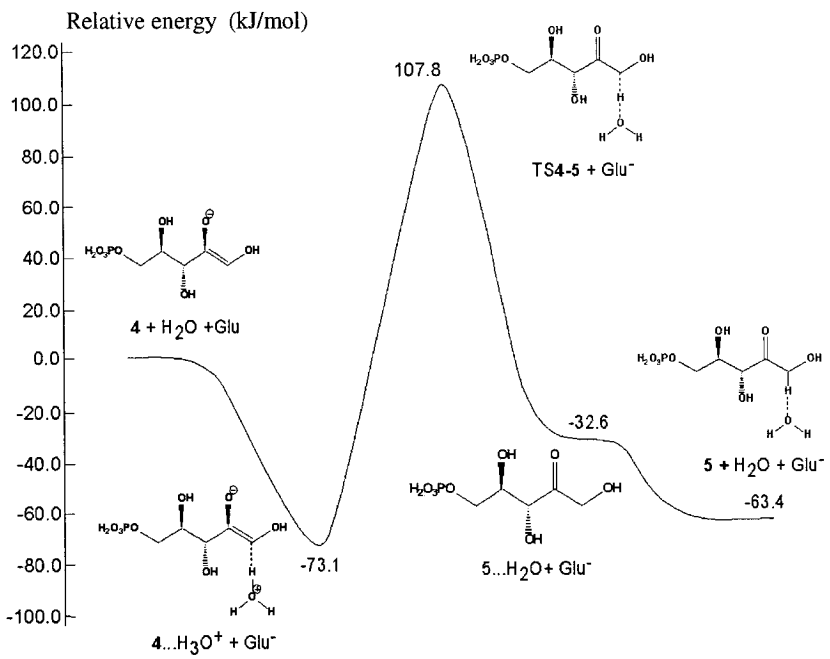


Figure 5.13. Schematic energy profile at 0 K for the route **4**→**5** (Glu catalyzed mechanism)

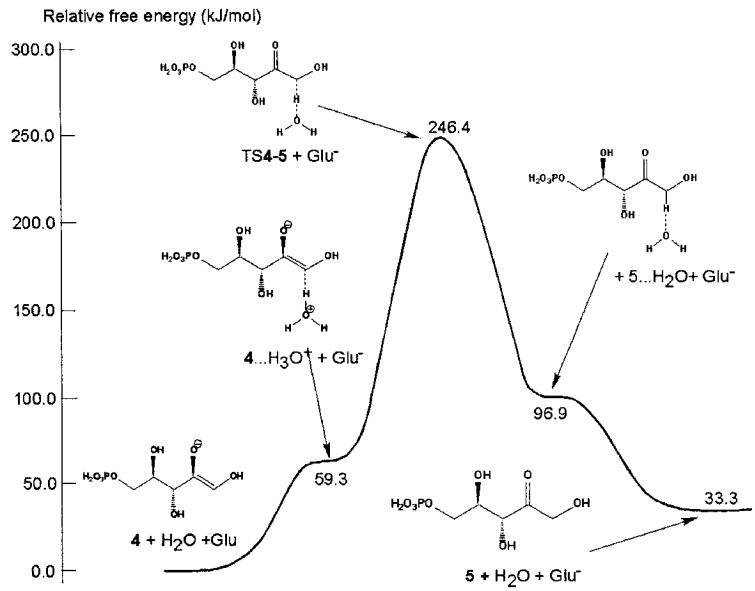


Figure 5.14. Schematic free energy profile at 298.15 K for the route **4**→**5** (Glu catalyzed mechanism)

The following discussion is based on the data of the entropy corrected case at 298.15 K (Figure 5.15, 5.16).

The current study involves the reversible conversion of ribose-5-phosphate (R5P) to ribulose-5-phosphate (Ru5P) catalyzed by ribose-5-phosphate isomerase (RPI) with a simplified model of the system. The reaction starts with ring opening of R5P. A high-energy enediolate intermediate is formed in the isomerization process. In the active site of the RPI enzyme, a Glu^- residue acts as a catalytic base and helps the isomerization of the substrate. The product Ru5P is finally derived with an energy state very close to that of the initial R5P system.

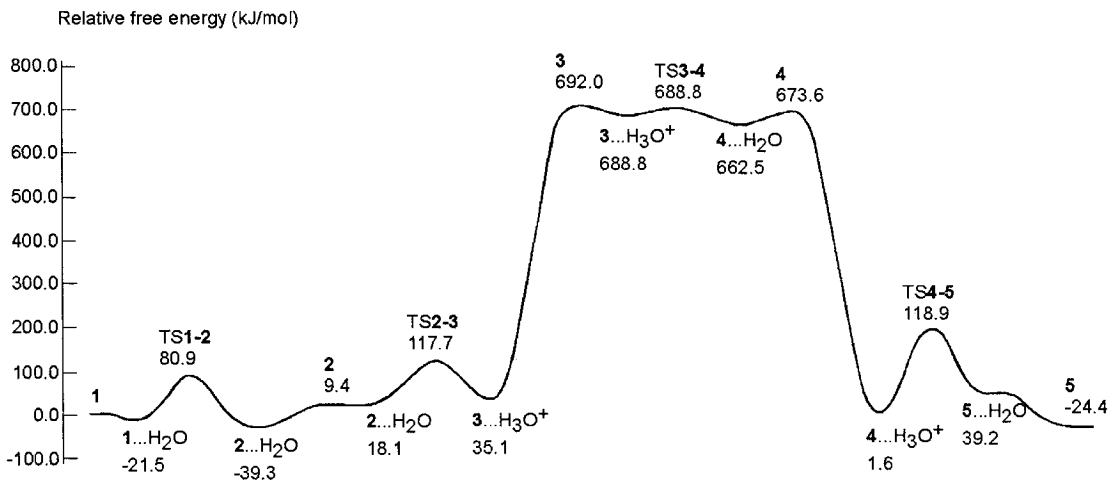


Figure 5.15. Schematic free energy profile at 298.15 K for the reversible conversion of ribose-5-phosphate (R5P) to ribulose-5-phosphate (Ru5P) catalyzed by water

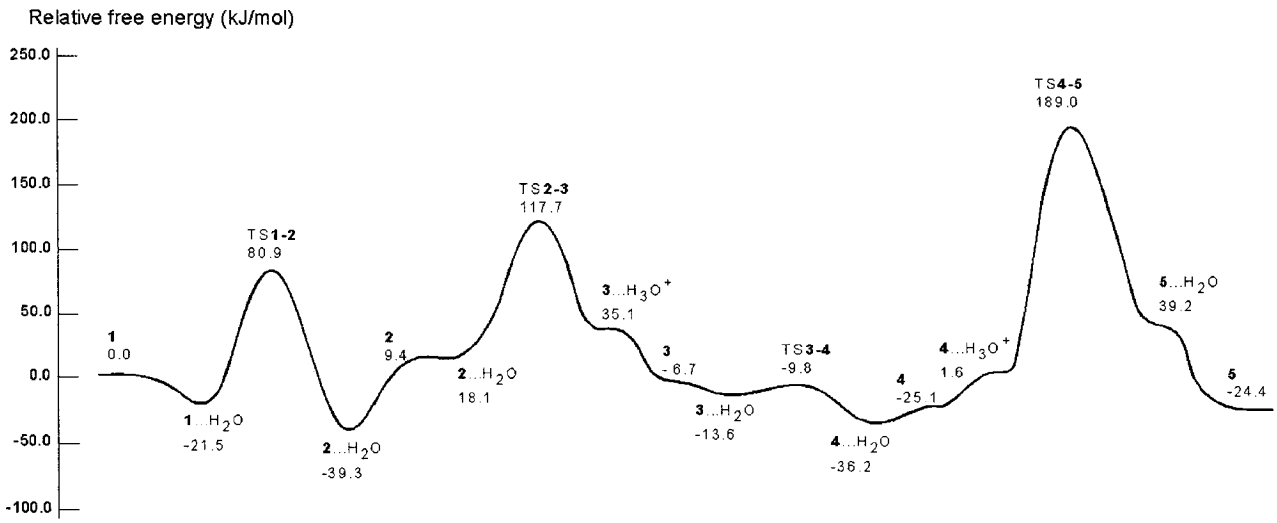


Figure 5.16. Schematic free energy profile at 298.15 K for the reversible conversion of ribose-5-phosphate (R5P) to ribulose-5-phosphate (Ru5P) catalyzed by Glu

Figure 5.15 gives the overall free energy profile of the isomerization reaction without the presence of an amino acid as a catalytic base. In the first step, the ring opening process needs to overcome an energy barrier of about 100 kJ/mol, and the open ring form of R5P, intermediate **2**, is derived. Next, a deprotonation reaction takes place at the C2 position. Although the energy barrier of this step is similar to that of the first step in magnitude, the separation of opposite charges that happens after the transition state **TS2-3** leads to a tremendous energy increase: the relative energy of intermediate **3** is over 600 kJ/mol. Energy changes in the next step are quite small: the energy is only about 4 kJ/mol, suggesting that the reacting speed for this particular step is very fast and that a proton is quickly transferred from the C2 oxygen to the C1 oxygen. In the last reaction step, the proton that has been taken away from the C2 position in the second step is brought back to the C1 position, lowering the system energy by over 600 kJ/mol. The final product Ru5P lies 24.4 kJ/mol lower than the original R5P, confirming that the isomerization reaction is reversible and that the reaction can go either direction depending on the

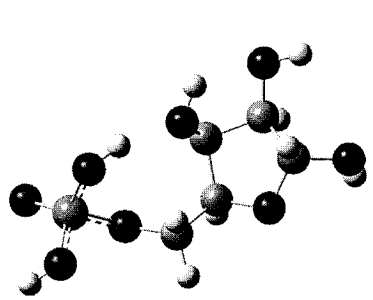
relative concentration of R5P and Ru5P. Figure 5.16 gives the overall free energy profile of the isomerization reaction, in which a Glu^- acts as a catalytic base. Comparing these two graphs one can easily find out that the huge energy increase and decrease that take place in the reaction steps **2→3** and **4→5** become quite small once Glu^- replaces water as the catalytic base that helps the reaction to proceed. This happens because the proton is transferred from the neutral substrate to the Glu^- ion in reaction step **2→3** and then back to the substrate in reaction step **4→5**. The negative charge is transferred between the substrate and the catalytic base thus the separation of opposite charges is avoided, and therefore no dramatic energy changes happen.

5.4 Conclusions

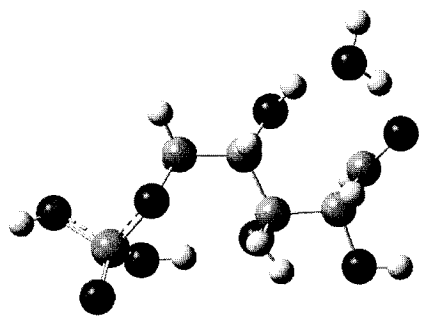
The proposed mechanisms for the isomerization reaction of R5P to Ru5P catalyzed by a RPI enzyme were investigated using the B3LYP method of density functional theory. The present results show that the high energy enediolate intermediate can be easily derived with the presence of a Glu^- amino acid residue acting as the catalytic base, indicating that the Glu^- amino acid residue is essential in the active sites of all RPI enzymes and plays a key role in the catalytic process. Meanwhile, all energy barriers in each reaction step are not significantly higher than 100 kJ/mol. The energy states of R5P and Ru5P are very close to each other, indicating that this isomerization reaction is reversible.

5.5 Representation of optimized structures

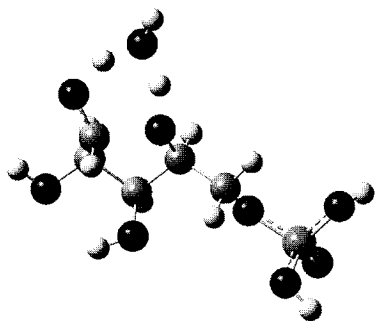
Graphical representations of the optimized structures are shown in figure 5.17. These figures were drawn with the GaussView software.



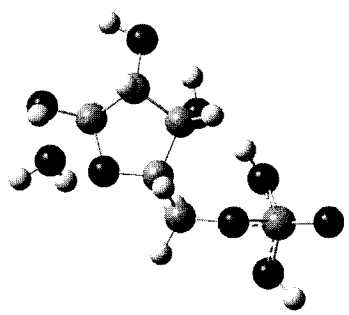
1



1...H_2O

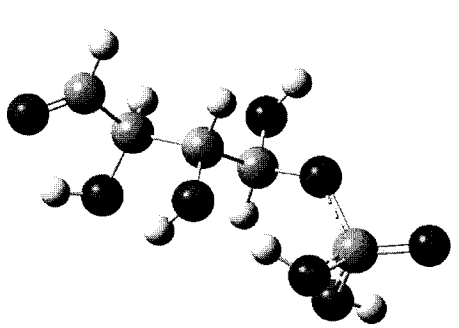


TS1-2

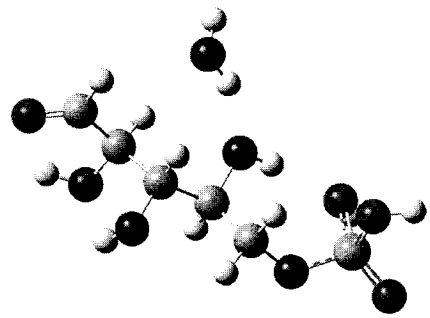


2...H_2O (1-2)

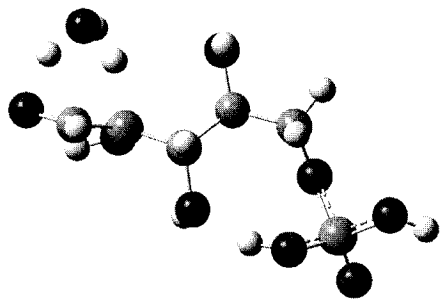
Figure 5.17. Graphical representations of the optimized structures in the reversible isomerization of R5P to Ru5P



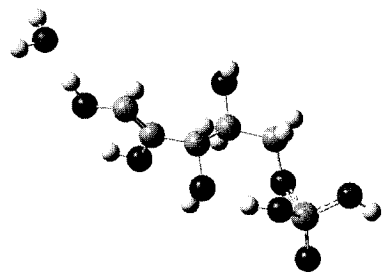
2



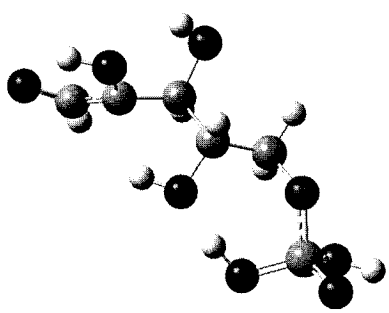
2...H_2O (2-3)



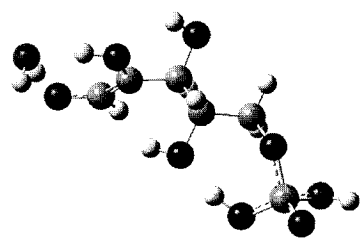
TS2-3



3...H_2O (2-3)

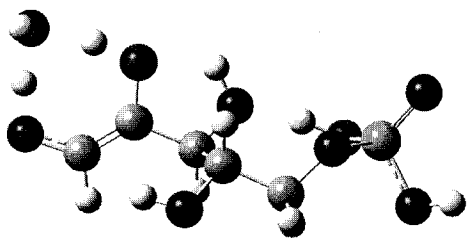


3

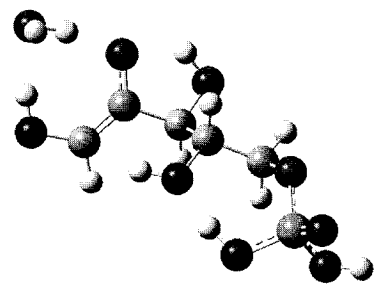


3...H_2O (3-4)

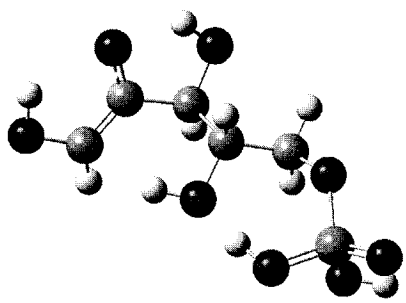
Figure 5.17. Graphical representations of the optimized structures in the reversible isomerization of R5P to Ru5P (continued)



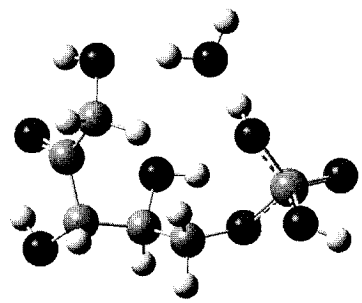
TS3-4



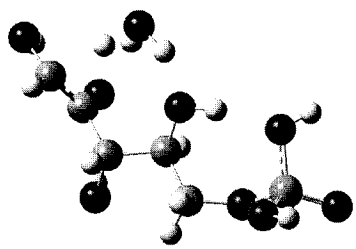
4...H_2O (3-4)



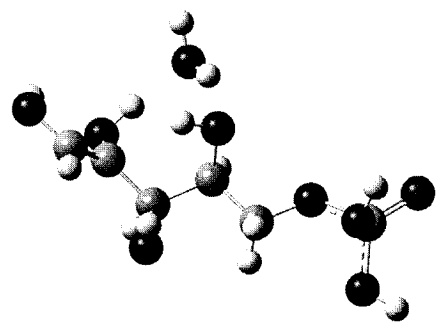
4



4...H_2O (4-5)

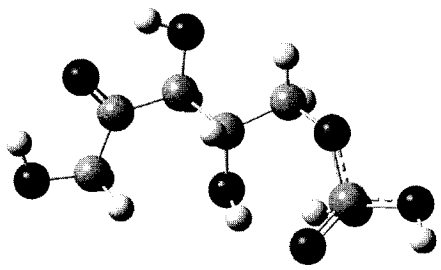


TS4-5



5...H_2O

Figure 5.17. Graphical representations of the optimized structures in the reversible isomerization of R5P to Ru5P (continued)



5

Figure 5.17. Graphical representations of the optimized structures in the reversible isomerization of R5P to Ru5P (continued)

Chapter 6

Concluding Remarks and Future Research

In the proposed computational study, critical enzymes that catalyze important reactions in biological systems will remain an important direction. Possible projects may come from some interesting but controversial enzymes, such as ribozymes, enzymes from ribonucleic acid that catalyze either their own cleavage or the cleavage of other RNAs. Another unsolved problem that may be included in a future computational study is the calculation of the redox potentials of crucial biomolecules.

6.1 Computational study on catalytic mechanisms of critical enzymes

The study of enzymes is an attractive but also controversial issue in computational chemistry studies. There are several facts associated with the nature of enzymes that may contribute to the complexity of a computational study of them: many enzymes are very flexible and often contain charged amino acid side chains and extended H-bond networks from the active site to the protein surroundings; water is often converted to hydroxyl and can enter or leave the active site; amino acid side chains can readily change protonation states or become radicals in some cases.

The combined QM/MM methodology allows reliable electronic structure calculations for enzymatic reactions with a realistic and atomistic description of enzyme environment. Such a combined QM and MM approach can take advantage of the applicability and accuracy of the QM methods for chemical reactions in systems of several dozens of atoms and of the computational efficiency of the MM description for the rest of the enzyme and solvent, which normally consists of thousands of atoms. Therefore, this method will be utilized in future studies of enzymes.

What is mainly needed is a better-integrated quantum plus classical methodology for treating large and complex active sites in an extended protein environment. The related methodology needs to be developed in several directions. There is an ongoing effort to improve DFT methods from a careful analysis of exchange, correlation, and electronic kinetic energy terms. For very large systems, the interface between the quantum and classical subsystems has to be reliably modeled. Active site flexibility requires that the conformations of both quantum and classical subsystems can change in a coordinated way.

Solvent effects is another important problem that one has to consider while developing reliable computational approaches to the study of enzymes. In most cases, explicit water molecules are involved in reactions that take place at the active sites of enzymes. One has to include those water molecules that are usually responsible for electron or proton transfer. While the solvent environment always affects how a reaction may proceed, one has to choose a proper solvation model to simulate the solvent effect. Currently,

continuum solvation models are widely applied. The Onsager and CPCM solvation models have been chosen in the present study, in which the Onsager model was used to obtain optimized structures. The Onsager model is a robust method, and gives fairly good relative energies that are desired especially in dealing with reaction mechanisms. However, due to its over simplified dipole model and spherical cavity, this model gives less accurate results for molecules with small dipole moments or planar structures. The SCRF methods, such as PCM, COSMO, etc., that use explicit point charges to generate the reaction field and adopt molecular shape cavities, usually give better results, but lead to much longer computational times and frequently to convergence problems. Another approach is to optimize a structure with the dipole model and then do a single point calculation with continuum SCRF models such as PCM or COSMO. This approach has been moderately successful for many systems, but it is not the final solution to the solvent effect problem. Therefore, much remains to be done to develop robust, accurate solvation models that are computationally efficient.

Additionally, transition metal ions are often found at the active sites of enzymes and always play a key role in the catalytic processes. Due to their more complicated electron structures, metal ions may exist in various oxidation states. For instance, metal ion oxidation states Fe (II, III, IV), Mn (II, III, IV), and Cu (I, II) are the most prominent in biological systems; the spin states of Fe and Mn are also variable, with the high-spin state being mainly preferred.[138] By contrast, Zn (II) is redox inactive, although its Lewis acid character allows it to participate in many reactions involving charge flow.[139]

Since their discovery some 20 years ago, the study of ribozymes has become enormously active in the fields of molecular biology and biomedical science. Ribozymes are catalytic RNAs that can promote chemical reactions in the absence of proteins. These catalytic RNAs may have played a role in selfreplication of the early states of evolution on earth. Some ribozymes are self-splicing introns. These will catalyze the cleavage and removal of the intervening sequence from an RNA transcript, and then ligate the message from the flanking exons. Ribozymes can also be engineered to cleave other target RNA molecules, and, consequently, they are now widely accepted as agents capable of inhibiting gene expression. They are therefore also associated with very promising links to candidates in gene therapy.

In the proposed future work, DFT will be combined with molecular mechanics methods to investigate various ribozymes including leadzyme, hairpin ribozyme, hammerhead ribozyme, etc.

6.2 Computational study on the oxidation potentials of crucial biomolecules

The redox potential is a key point to understanding in many biological events, such as electron/charge transfer processes, selective oxidation, etc., that take place in the DNA helix. However, a theoretical study of this problem is still an undeveloped topic. As described previously, the EA/IP of nucleobases has already been solved with the deviation from experimental results being close to chemical accuracy. In general, the error associated with the redox potentials of medium to small organic molecules is still at

the 100 kJ/mol level[140]. Therefore, how to choose a proper computational method and how to treat solvent effects will be the crucial steps towards the solution of this problem. Previous attempts[141] include some special techniques such as frozen density functional theory and QM/MM hybrid methods. A standard procedure to calculate the redox potential of biomolecules is not available.

The proposed project will focus on: calculating the redox potential of medium-size biomolecules with or close to chemical accuracy. The target molecules would be the nucleobases that form the DNA helix and medium-size DNA segments. Their oxidation potentials will be studied by DFT methods and post Hartree-Fock methods combined with proper solvation models.

To determine the electrochemical oxidation potentials of biological molecules in aqueous solution, thermodynamic cycles[142] will be used.

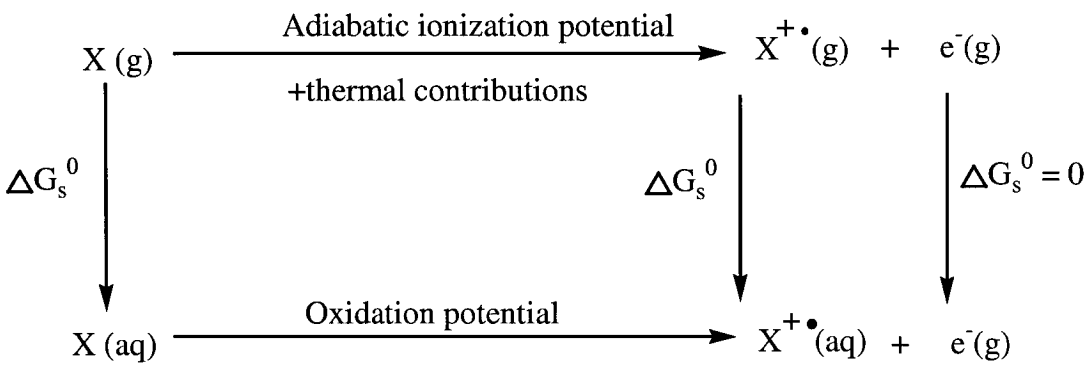


Figure 6.1. Thermodynamic cycle of a redox reaction

Normal hydrogen electrode half-reaction is

$$H+(aq) + e\text{-}(g) \xrightarrow{\Delta G^0 = -4.44eV} \frac{1}{2}H_2(g) \tag{6.1}$$

The energy changes associated with the top of the cycle is the gas-phase oxidation potential, which is closely related to the adiabatic ionization potential (the distinction being that the adiabatic ionization potential is the difference in enthalpies at 0 K, whereas the oxidation potential is a difference in free energies at 298 K). Free energy changes associated with the left and right sides of the cycle are the energies of solvation, and thus the bottom side of the cycle refers to the oxidation potential in solution. Oxidation potentials are usually reported relative to a reference potential, typically taken as the normal hydrogen electrode (NHE); this convention, then, refers to the free energy change for the net reaction resulting from addition of the NHE half-reaction to the half-reaction at the bottom of the free energy cycle. Thus, the aqueous oxidation potential of a closed-shell neutral solute X is determined from Figure 6.1 as

$$\begin{aligned}\Delta G^0_{ox}(X\,vs.\,NHE) = IP(X) + \Delta G_{evr,gas}\left(X \rightarrow X^{+\bullet} + e^{-\bullet}\right) \\ + \Delta G^0_s\left(X^{+\bullet}\right) - \Delta G^0_s(X) - 4.44eV\end{aligned} \tag{6.2}$$

where IP(X) is the adiabatic ionization potential of X; $\Delta G_{evr,gas}\left(X \rightarrow X^{+\bullet} + e^{-\bullet}\right)$ is the difference in thermal contributions to the solute free energy derived from changes in the electronic, vibrational and rotational partition functions upon ionization (note that the difference in the zero-point vibrational energy is defined to be already included in the IP); ΔG^0_s is the aqueous free energy of solvation; and X and $X^{+\bullet}$ are the neutral and oxidized forms of the solute, respectively.

References

(1) Noodleman, L.; Lovell, T.; Han, W.-G.; Li, J.; Himo, F. *Chem. Rev.* **2004**, *104*, 459.

(2) Burrows, C. J.; Muller, J. G. *Chem. Rev.* **1998**, *98*, 1109.

(3) Leszczynski, J. *Computational Chemistry: Reviews of Current Trends*; World Scientific, 2003; Vol. 5.

(4) Cukier, R. I. *Annu. Rev. Phys. Chem.* **1998**, *49*, 337.

(5) Cizek, J. *Adv. Chem. Phys.* **1969**, *14*, 35.

(6) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(7) Wetmore, S. D.; Boyd, R. J.; Eriksson, L. A. *Chem. Phys. Lett.* **2000**, *322*, 129.

(8) Wetmore, S. D.; Boyd, R. J.; Eriksson, L. A. *Chem. Phys. Lett.* **2001**, *343*, 151.

(9) Russo, N.; Toscano, M.; Grand, A. *J. Comp. Chem.* **2000**, *21*, 1243.

(10) Colson, A.-O.; Besler, B.; Close, D. M.; Sevilla, M. D. *J. Phys. Chem.* **1991**, *96*, 661.

(11) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem. Eur. J.* **1999**, *5*, 3581.

(12) Guerra, C. F.; Bickelhaupt, F. M. *Angew. Chem.* **1999**, *38*, 2942.

(13) Reha, D.; Kabelac, M.; Ryjacek, F.; Sponer, J.; Sponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *124*, 3366.

(14) Hjort, M.; Stafstrom, H. *Phys. Rev. Lett.* **2001**, *87*, 228101.

(15) Matta, C. F.; Castillo, N.; Boyd, R. J. *J. Phys. Chem. B* **2006**, *110*, 563.

(16) Ban, F.; Rankin, K. N.; Gauld, J. W.; Boyd, R. J. *Theo. Chem. Accts.* **2002**, *108*, 1.

(17) Oikawa, H.; Nakamura, K.; Toshima, H.; Toyomasu, T.; Sassa, T. *J. Am. Chem. Soc.* **2002**, *124*, 9145.

(18) Glaser, R.; Wu, H.; Lewis, M. *J. Am. Chem. Soc.* **2005**, *127*, 7346.

(19) Fischer, E. *Ber. Dt. Chem. Ges.* **1894**, *27*, 2985.

(20) Koshland, D. E. *Proc. Natl. Acad. Sci.* **1958**, *44*, 98.

(21) Lu, Z.; Yang, W. *J. Chem. Phys.* **2004**, *121*, 89.

(22) Cisneros, G. A.; Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2003**, *125*, 10384.

(23) Lovell, T.; Himo, F.; Han, W.-G.; Noodleman, L. *Coord. Chem.Rev.* **2003**, *211*, 238.

(24) Siegbahn, P. E. M.; Blomberg, M. R. A. *Rev. Phys. Chem* **1999**, *50*, 221.

(25) Noodleman, L.; Lovell, T.; Han, W.-G.; Liu, T.; Torres, R. A.; Himo, F. *Comprehensive Coordination Chemistry II, From Biology to Nanotechnology*; Elsevier: New York, 2003; Vol. 2.

(26) Cramer, C. J. *Essentials of Computational Chemistry Theories and Models*; Wiley, 2002; Vol. 13.

(27) Schenk, G.; Pau, M. Y. M.; Solomon, E. I. *J. Am. Chem. Soc.* **2003**, *126*, 505.

(28) Oikawa, H.; Nakamura, K.; Toshima, H.; Toyomasu, T.; Sassa, T. *J. Am. Chem. Soc.* **2002**, *129*, 9145.

(29) Born, M.; Oppenheimer, J. R. *Ann. der Phys.* **1927**, *84*, 457.

(30) Hartree, D. R. *Proc. Cambridge. Philos. Soc.* **1928**, *24*, 89.

(31) Fock, V. *Z. Phys.* **1930**, *35*, 210.

(32) Slater, J. C. *Phys. Rev.* **1930**, *46*, 618.

(33) Roothaan, C. C. J. *Rev. Mod. Phys* **1951**, *23*, 69.

(34) Hall, G. G. *Proc. Roy. Soc. (London)* **1951**, *A205*, 541.

(35) Raghavachari, K.; Anderson, J. B. *J. Phys. Chem.* **1996**, *100*, 12960.

(36) Möller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(37) Binkley, J. S.; Pople, J. A. *Int. J. Quantum Chem.* **1975**, *9*, 229.

(38) Leininger, M. L.; Allen, W. D.; Schaefer, H. F.; Sherrill, C. D. *J. Chem. Phys.* **2000**, *112*, 9213.

(39) Ceperley, D. M. *Phys. Rev. B* **1978**, *18*, 3126.

(40) March, N. H. *Theory of the Inhomogeneous Electron Gas*; Plenum Publishing: New York, 1983.

(41) Theophilou, A. *J. Phys. C* **1979**, *12*, 5419.

(42) Hadjisavvas, N.; Theophilou, A. *Phys. Rev. A* **1985**, *32*, 720.

(43) Gidopoulos, N. I.; Papaconstantinous, P. G.; Gross, E. K. U. *Phys. Rev. Lett.* **2002**, *88*.

(44) Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *52*, 997.

(45) Petersilka, M.; Gossmann, U. J.; Gross, E. K. U. *Phys. Rev. Lett.* **1996**, *76*, 1212.

(46) Kohn, W.; Sham, L. J. *Phys. Rev. Lett.* **1965**, *85*, 4229.

(47) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.

(48) Becke, A. D. *Phys. Rev.* **1988**, *38*, 3098.

(49) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio Molecular Orbital Theory*; Wiley: New York, 1986.

(50) Davidson, E. R.; Feller, D. *Chem. Rev.* **1986**, *86*, 681.

(51) Onsager, L. *J. Am. Chem. Soc* **1936**, *58*, 1486.

(52) Wong, M. W.; Frisch, M. J.; Wiberg, K. B. *J. Am. Chem. Soc* **1991**, *113*, 4776.

(53) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.

(54) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098.

(55) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.

(56) Cossi, M.; Rega, G. N.; Scalmani; Barone, V. *J. Comp. Chem.* **2003**, *in press.*

(57) Thiel, W. *Adv. Chem. Phys.* **1996**, *93*, 703.

(58) Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; Wiley: New York, 1970.

(59) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.

(60) Stewart, J. J. P. *Reviews in Computational Chemistry*; Wiley-VCH, 1990; Vol. 1.

(61) Carsky, P.; Zaharadnik, R. *Fortschr. Chem. Forsch.* **1973**, *43*, 2.

(62) Beveridge, D. L.; Hinze, J. *J. Am. Chem. Soc.* **1971**, *93*, 3107.

(63) Naray-Szabo, G.; Surjan, P. R.; Angyan, J. G. *Applied Quantum Chemistry*; Reidel, 1987.

(64) Gao, J. *Review in Computational Chemistry*; VCH: New Yok, 1995.

(65) Merz, K. M. J.; Stanton, R. V. *Encyclopedia of Computational Chemistry*; Wiley: New York, 1998.

(66) Hoover, W. G. *Phys. Rev.* **1985**, *31*, 1695.

(67) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.

(68) Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1981**, *74*, 4872.

(69) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.

(70) Kasai, H. *Mut. Res.* **1997**, *387*, 147.

(71) Burrows, C. J.; Muller, J. G. *Chem. Rev.* **1998**, *98*, 1109.

(72) Steenken, S.; Jovanovic, S. V. *J. Am. Chem. Soc.* **1997**, *119*, 617.

(73) Kovalsky, O. I.; Panutin, I. G.; Budowsky, E. I. *Photochem. Photobiol.* **1990**, *52*, 509.

(74) Saito, I.; Takayama, M.; Sugiyama, H.; Nakatani, K. *J. Am. Chem. Soc.* **1995**, *117*, 6406.

(75) Sugiyama, H.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 7063.

(76) Steenken, S.; Jovanovic, S. V. *J. Am. Chem. Soc* **1997**, *119*, 617.

(77) Cadet, J. *IARC Scientific Publications* **1994**, 425.

(78) Yanagawa, H.; Ogawa, Y.; Ueno, M. *J. Biol. Chem.* **1992**, *267*, 13320.

(79) Adam, W.; Saha-Moller, C. R. *Photochem. photobiol.* **1995**, *62*, 231.

(80) Goyal, R. N.; Jain, N.; Garg, D. K. *Bioelec. Bioenerg.* **1997**, *43*, 105.

(81) Raoul, S.; Cadet, J. *J. Am. Chem. Soc.* **1996**, *118*, 1892.

(82) Hems, G. *Nature* **1960**, *186*, 710.

(83) David, S. S.; Williams, S. D. *Chem. Rev.* **1998**, *98*, 1221.

(84) Candeias, L. P.; Steenken, S. *J. Am. Chem. Soc.* **1989**, *111*, 1094.

(85) Hildebrand, K.; Schulte-Frohlinde, D. *Free Rad. Res. Commun.* **1990**, *11*, 195.

(86) Burney, S.; Caulfield, J. L.; Niles, J. C.; Wishnok, J. S.; Tannenbaum, S. R. *Mut. Res.* **1999**, *424*, 37.

(87) Huie, R. E.; Padmaja, S. *Free Rad. Res. Commun.* **1993**, *18*, 195.

(88) Althaus, J. S.; Oien, T. T.; Fici, G. J.; Scherch, H. M.; Sethy, V. H.; Voigtlander, P. F. V. *Res. Commun. Chem. Pathol. Pharmacol.* **1994**, *83*, 243.

(89) Kissner, R.; Nauser, T.; Bugnon, P.; Lye, P. G.; Koppenol, W. H. *Chem. Res. Toxicol.* **1997**, *10*, 1285.

(90) Wink, D. A.; Hanbauer, I.; Krishna, M. C.; DeGraff, W.; Gamson, J.; Mitchell, J. B. *Proc. Natl. Acad. Sci.* **1993**, *90*, 9813.

(91) Goldstein, S.; Czapski, G.; Cohen, H.; Meyerstein, D. *J. Free Rad. Biol. Med.* **1995**, *19*, 785.

(92) Pryor, W. A.; Squadrito, G. L. *Am. J. Physiol.* **1995**, *268*, 699.

(93) Halliwell, B.; Gutteridge, J. M. C. *Free Radicals In Biology And Medicine Third Edition*; Oxford University Press: New York, 1999.

(94) Merenyi, G.; Lind, J. *Chem. Res. Toxicol.* **1998**, *11*, 243.

(95) Uppu, R. M.; Squadrito, G. L.; Pryor, W. A. *Arch. Biochem. Biophys. Res. Commun.* **1996**, *327*, 335.

(96) Lymar, S. V.; Hurst, J. K. *Inorg. Chem.* **1998**, *37*, 294.

(97) Goldstein, S.; Czapski, G. *J. Am. Chem. Soc.* **1998**, *120*, 3458.

(98) Marla, S. S.; Lee, J.; Groves, J. T. *Proc. Nat. Acad. Sci.* **1997**, *94*, 14243.

(99) Ghafourifar, P.; Richter, C. *FEBS Lett.* **1997**, *418*, 291.

(100) Yermilov, V.; Rubio, J.; Becchi, M.; Friesen, M. D.; Pignatelli, B.; Ohshima, H. *Carcinogenesis* **1995**, *16*, 2045.

(101) Niles, J. C.; Wishnok, J. S.; Tannenbaum, S. R. *J. Am. Chem. Soc.* **2001**, *49*, 12147.

(102) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486.

(103) Vallet, V.; Wahlgren, U.; Schimmelpfenning, B.; Szabo, Z.; Grenthe, I. *J. Am. Chem. Soc* **2001**, *123*, 11999.

(104) Hole, E. O.; Nelson, W. H.; Sagstuen, E. *J. Chem. Phys.* **1987**, *86*, 5218.

(105) Kharitonenkov, A.; Chen, Z.; Sures, I.; Wang, H.; Schilling, J.; Ullrich, A. *Nature (London)* **1997**, *386*, 181.

(106) Morinville, A.; Maysinger, D.; Shaver, A. *Trends Pharmacol. Sci.* **1998**, *19*, 452.

(107) Tonks, N. K.; Neel, B. G. *Curr. Opin. Cell Biol.* **2001**, *13*, 182.

(108) Hunter, T. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1998**, *353*, 583.

(109) Jackson, M. D.; Denu, J. M. *Chem. Rev.* **2001**, *101*, 2313.

(110) Zhang, Z.-Y. *Acc. Chem. Res.* **2003**, *36*, 385.

(111) Ren, J.-M.; Li, P.-M.; Zhang, W.-R.; Sweet, L. J.; Cline, G.; Shulman, G. I.; Livingston, J. N.; Goldstein, B. J. *Diabetes* **1998**, *47*, 493.

(112) Minassian, B. A.; Lee, J. R.; Herbrick, J.-A.; Huizenga, J.; Soder, S.; Mungall, A. J.; Dunham, I.; Gardner, R.; Fong, C.-y. G.; Carpenter, S.; Jardim, L.; Satishchandra, P.; Andermann, E.; Snead, O. C., III; Lopes-Cendes, I.; Tsui, L.-C.; Delgado-Escueta, A. V.; Rouleau, G. A.; Scherer, S. W. *Nature Genetics* **1998**, *20*, 171.

(113) Zhang, Z. Y. *Curr. Opin. Chem. Bio.* **2001**, *5*, 416.

(114) Barford, D.; Das, A. K.; Egloff, M. P. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 133.

(115) Alonso, A.; Sasin, J.; Bottini, N.; Friedberg, I.; Friedberg, I.; Osterman, A.; Godzik, A.; Hunter, T.; Dixon, J.; Mustelin, T. *Cell* **2004**, *117*, 699.

(116) Wang, W.-Q.; Sun, J.-P.; Zhang, Z.-Y. *Curr. Topics Med. Chem.* **2003**, *3*, 739.

(117) Kennelly, P. J. *Chem. Rev.* **2001**, *101*, 2291.

(118) Denu, J. M.; Tanner, K. G. *Biochem.* **1998**, *37*, 5633.

(119) Johnson, T. O.; Schulman, F. Y.; Lipscomb, T. P.; Yantis, L. D. *Vet. Pathol.* **2002**, *39*, 452.

(120) Elchebly, M.; Payette, P.; Michaliszyn, E.; Cromlish, W.; Collins, S.; Loy, A. L.; Normandin, D.; Cheng, A.; Himms-Hagen, J.; Chan, C. C.; Ramachandran, C.; Gresser, M. J.; Tremblay, M. L.; Kennedy, B. P. *Science* **1999**, *283*, 1544.

(121) Klaman, L. D.; Boss, O.; Peroni, O. D.; Kim, J. K.; Martino, J. L.; Zabolotny, J. M.; Moghal, N.; Lubkin, M.; Kim, Y. B.; Sharpe, A. H.; Stricker-Krongrad, A.; Shulman, G. I.; Neel, B. G.; Kahn, B. B. *Mol. Cell. Biol.* **2000**, *20*, 5479.

(122) Denu, J. M.; Tanner, K. G. *Biochemistry* **1998**, *37*, 5633.

(123) Downes, C. P.; Walker, S.; McConnachie, G.; Lindsay, Y.; Batty, I. H.; Leslie, N. R. *Biochem. Soc. Trans.* **2004**, *32*, 338.

(124) Salmeen, A.; Andersen, J. N.; Myers, M. P.; Tonks, N. K.; Barford, D. *Mol. Cell* **2000**, *6*, 1401.

(125) van Montfort, R. L. M.; Congreve, M.; Tisi, D.; Carr, R.; Jhoti, H. *Nature (London, United Kingdom)* **2003**, *423*, 773.

(126) Sivaramakrishnan, S.; Keerthi, K.; Gates Kent, S. *J. Am. Chem. Soc.* **2005**, *127*, 10830.

(127) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; J. A. Montgomery, J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revision B.05; Gaussian, Inc.: Pittsburgh PA, 2003.

(128) Zhang, R.; Andersson, C. E.; Savchenko, A.; Skarina, T.; Evdokimova, E.; Beasley, S.; Arrowsmith, C. H.; Edwards, A. M.; Joachimiak, A.; Mowbray, S. L. *Structure (Camb)* **2003**, *11*, 31.

(129) Hove-Jensen, B. *J. Bacteriol.* **1988**, *170*, 1148.

(130) Volk, R.; Bacher, A. *J. Biol. Chem.* **1991**, *266*, 20610.

(131) Miosga, T.; Zimmermann, F. K. *Curr. Genet.* **1996**, *30*, 404.

(132) Sorensen, K. I.; Hove-Jensen, B. *J. Bacteriol.* **1996**, *178*, 1003.

(133) Huck, J. H.; Verhoeven, N. M.; Struys, E. A.; Salomons, G. S.; Jakobs, C.; Knaap, M. S. v. d. *Am. J. Hum. Genet.* **2004**, *74*, 745.

(134) Zhang, R.-G.; Andersson, C. E.; Skarina, T.; Evdokimova, E.; Edwards, A. M.; Joachimiak, A.; Savchenko, A.; Mowbray, S. L. *J. Mol. Biol.* **2003**, *332*, 1083.

(135) Zhang, R.-G.; Andersson, C. E.; Savchenko, A.; Skarina, T.; Evdokimova, E.; Beasley, S.; Arrowsmith, C. H.; Edwards, A. M.; Joachimiak, A.; Mowbray, S. L. *Structure (Lond.)* **2003**, *11*, 31.

(136) Roos, A. K.; Andersson, C. E.; Bergfors, T.; Jacobsson, M.; Karlen, A.; Unge, T.; Jones, T. A.; Mowbray, S. L. *J. Mol. Biol.* **2004**, *335*, 799.

(137) Roos, A. K.; Burgos, E.; Ericsson, D. J.; Salmon, L.; Mowbray, S. L. *J. Biol. Chem.* **2005**, *280*, 6416.

(138) Silva, J. J. R. F. D.; Williams, R. J. P. In *The Biological Chemistry of the Elements*; Press, C., Ed.; Oxford, 1991.

(139) Luchinat, C.; Sola, M. *Encyclopedia of Inorganic Chemistry*; Wiley: New York, 1994.

(140) Zhang, X.; Pugh, J. K.; Ross, P. N. *J. Electrochem Soc.* **2001**, *148(5)*, 183.

(141) Paul, W.; Weber, E. J. *Phys. Chem. Chem. Phys* **2000**, *2*, 1231.

(142) Parker, V. D. *J. Am. Chem. Soc* **1976**, *98*, 98.