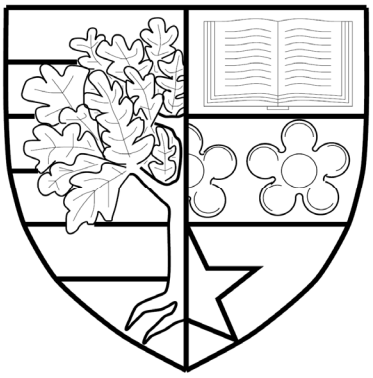
Quantum Chemical Characterization of Biomolecules in the Gas Phase and on Surfaces of Metal Oxides

by

Sanliang Ling



Thesis submitted for the degree of
Doctor of Philosophy

at

Chemistry - School of Engineering and Physical Sciences
Heriot-Watt University, Edinburgh

October 2010

Abstract

During the four years of my PhD study, I performed systematic studies of the conformations of biomolecules ranging from a small amino acid (e.g. glycine) to a medium-sized nucleoside (e.g. 2'-deoxycytidine). To better account for possible effects brought by explicit environments (e.g. radiation, aqueous solution, and so on), we studied biomolecules in different phases, including neutral and charged species, in the gas phase and solid state, and neutral on solid surface. The work being presented in this thesis is original as:

(1) A tool which can automatically generate libraries of conformations for a systematic search of the conformational space of a molecule was developed. When combined with tools developed by our colleagues, our toolbox facilitates a combinatorial computational chemical study of some small biomolecules;

(2) A new method which can suppress barriers between different local minima on a molecular potential energy surface (PES) was developed, and with this new deformed PES, a lot of other techniques (e.g. Monte Carlo and simulated annealing) could be adopted to search for the global minima structure in a much more efficient way;

(3) We performed a highly accurate study of two conformers of glycine up to the coupled-cluster with single and double and perturbative triple excitations (CCSD(T)) with basis sets up to aug-cc-pVQZ level of theory, and we found that the treatment at the CCSD(T) level of theory is necessary to achieve numerical stability of the relative energies with respect to different basis sets at different geometries;

(4) Through a thorough search of the conformational space of 2'-deoxycytidine, we found that its conformations in the gas phase are quite different from those in the solid state, and hopefully this finding could correct some of the previous approaches, in which structural information extracted from solid state experiments was used in computational studies of molecules in the gas phase;

(5) Adsorptions of hydrogen, methanol and glycine on different types of solid surfaces (conductive and semiconductive) were studied, and catalytic performances of these surfaces on breaking chemical bonds were discussed.

The current thesis not only covers the main applications of computational

chemistry tools in the conformational study of biomolecules, it also includes discussions on accuracy and methodology which is involved in these studies. We definitely did not intend to solve all of the problems which people have met in their conformational studies of biomolecules. We just hope that the work being presented here was performed in a much more systematic way, and we hope these studies can give people some insights which might be helpful in their further studies.

Acknowledgements

I would like to express my deepest appreciation to my supervisor Prof. Maciej Gutowski. Thanks for his excellent guidance and generous support during my PhD study. He is one of nicest supervisors I have ever met. His serious-minded attitude of doing scientific research not only reminds me to work harder, but also reminds me to think before acting. He is also an endless source of creative ideas. I have been so lucky to have a supervisor like him.

I am indebted to the many people who have helped and supported me over the last four years.

My colleagues from theoretical and computational chemistry group, and my Chinese friends at Edinburgh, thank you very much for keeping me happy!

Thanks for helpful discussions made with Prof. Zijing Lin, Dr. Donghai Mei, Dr. Maciej Haranczyk, Prof. Jun Li and Prof. Qingfeng Ge.

Last but not least at all, I would like to express my heartfelt gratitude to my parents for their unconditional love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As the main target of this thesis, we studied the conformations of biomolecules of different sizes under different conditions. These conditions include pure gas phase, radiation, ionization, and solid surfaces. We hope the results of our study could make our own contribution to the field of computational chemistry and will help people to understand the most basic behavior of biomolecules and their role in fundamental processes.

1.1 Background

Biomolecules play a central role in the processes which constitute the whole metabolic cycle of humankind. While biology pays more attention to the macroscale phenomenons behind these processes, chemistry pays more attention to the microscale properties of molecules which make it possible for these macroscale phenomenons to happen. Uncovering the molecular mechanisms of these biological processes is always full of challenges, and structures of participating biomolecules are the common starting point which leads these studies, though sometimes they could also be the bottleneck. Biomolecules spread over a very wide range of sizes: from a small amino acid (e.g. glycine) with tens of atoms to a very big deoxyribonucleic acid (DNA) molecule with up to tens of thousands of atoms or even more. Biomolecules are not like aromatic hydrocarbons which contain a lot of benzene rings in the molecular structures. Biomolecules are most likely to be very flexible and thus tend to form different molecular structures under slightly different conditions, which in turn determines their different behavior. As the fast development of analytical chemistry in the last century, a lot of powerful experimental techniques have been developed to study the structure of biomolecules from the gas phase to solid state. These experimental tools contributed quite a lot to people's understanding of the structures of biomolecules, while just being very similar to many of the other experimental techniques with different usages, these analytical techniques also have their own disadvantages. For example, most of them are very expensive and thus many groups could not afford them, some of them have to work under

extreme conditions (e.g. ultra low temperature, ultra high vacuum), and so on. On the other side, when experimental technologies were upgraded from one generation to another, the theoretical approaches to this kind of problem also underwent considerable developments. In particular, quantum mechanical electronic structure theory of atoms, molecules and solids was developed. Through applications of electronic structure theory, theoretical chemists could predict behavior of electrons and nuclei, and thus could provide detailed structural information about molecules and crystals. The electronic structure theory is a great accomplishment for the whole chemistry community just like many other wonderful experimental inventions, which have been recognized for a long period. With the assistance of structural information obtained from the electronic structure theory, people could know structures of biomolecules without rigorous experimental measurements, which makes it possible to allow people to spend more time on other more important issues which could not be explored by theoretical tools currently.

The main aim of the current thesis is to investigate the conformational properties of biomolecules with different sizes under different circumstances, through the usage of modern electronic structure theory, and we hope our studies could help people who are working on experimental studies of these biomolecules to get better understanding of their experimental observations. As most of the current experimental studies of biomolecules were performed in the gas phase and the results of these experimental studies could provide very accurate references for us, we also did quite a lot of our theoretical studies for biomolecules in the gas phase. In this way, it is easy for us to compare our theoretical results with the accurate experimental results so that we can benchmark the current theoretical methods which are widely used by other theoretical chemists as well as further refine the experimental measurements. Another issue we need to consider is the cosmic rays which we are exposed to everyday without sensing, and in fact these cosmic rays have been recognized as the main source of DNA damage which could further turn into cancers or other diseases. In this way, the radiation and ionization induced by cosmic rays to biomolecules were also treated as a very important topic in the current thesis. Finally, we also studied conformations of small molecules on solid surfaces as the solid surface could act as the medium where prebiotic biomolecular synthesis was performed through reactions between small inorganic molecules (H_2, CO, CH_4, N_2, etc.).

We tried to perform these studies in a relatively organized way so that they will look more systematic. However, as the main aim of the current thesis is to provide more accurate and reliable theoretical references for experiments and some of these experimental studies were done in a discrete way by different groups from all over of the world, the connections between different aspects of our theoretical studies which we are talking about here might still be a little bit loose. Nevertheless, we still hope the current thesis could represent some of the best applications of the electronic structure theory across the whole landscape of biomolecule relevant phenomenons.

1.2 Structure of the Current Thesis

The structure of the current thesis was arranged as below.

Firstly, we will talk about the basic ideas of the electronic structure theory in Chapter 2. We will introduce the mostly used electronic structure methods, including the Hartree-Fock (HF) method, the density functional theory (DFT), the Møller-Plesset perturbation theory (MPn), and the coupled cluster (CC) approach. The emphasis will be given to the DFT method as it is the main tool which we used over all the studies included in the current thesis.

In Chapter 3, we continue our introductions on some of the most important methods used throughout computational chemistry. We discuss how to explore the potential energy surface (PES), which include two important issues. One is how to identify a stationary point, such as a local minimum or a first order saddle point, starting from an arbitrary point on the PES. Popular methods of geometry optimizations will be introduced. The other is how to identify the global minimum, also starting from an arbitrary point on the PES. Popular methods of global minima search will be introduced.

In Chapters 4 and 5, as an extension of Chapter 3, we will further introduce some of our own efforts to facilitate conformational search. In Chapter 4, we will discuss how to search for the global minima through the systematic search method, and we will introduce how to create a library of potentially relevant conformers with a tool, which we call SSC, denoting Systematic Screening of Conformers. In Chapter 5, we will introduce a new method on how to modify the original PES

and how to perform a conformational search on the modified PES, of which the barriers separating minima are suppressed.

Finishing discussions about the methodology development, we will turn our attention to the accuracy that the theoretical calculations could achieve on some specific problem. In Chapter 6, we will present our results of a highly accurate study of two conformers of the glycine molecule. Special emphasis will be given to convergence of the results with respect to basis sets and different geometries. Special attention will be paid to relative energies of conformers at different levels of theory. This study will be a very important part of the current thesis as it could provide us with the best benchmark results for both theoretical and experimental studies of two glycine conformers.

Rather than working on only two conformers of glycine, we also work on problems which include many more conformers. In Chapter 7, we will introduce our theoretical results from a conformational study of another representative biomolecule - 2'-deoxycytidine (dC). We modeled this molecule in the gas phase, for both neutral and charged species. This study will feature a systematic search of the conformational space of neutral dC molecule as well as a fruitful search of conformational space of valence anions based on the neutral structures. Possible roles of anions and cations of dC in DNA damage will also be discussed. In the end of this chapter, we will pay attention to a very fundamental question, why the crystalline phase and the gas phase favor different conformations of dC.

In Chapter 8, we will present our theoretical results about the interactions between model solid surfaces, which include surfaces of pure tungsten trioxide and pure rhenium trioxide as well as interfaces of these two transition metal oxides, and some small molecules, such as the hydrogen atom, hydrogen molecule, methanol molecule and glycine molecule. This study will be summarized by a comparison of the catalytic performance of the four model surfaces.

Finally, in Chapter 9, we will summarize what we have found from my PhD study and what could be done in the future.

1.3 Computational Details

Most of the gas phase studies which have just been mentioned above were performed with Gaussian 03 [1]. Some other gas phase studies were also performed with NWCHEM [2]. The MOLPRO [3] code was used to perform high level calculations which were done at the coupled cluster levels of theory. All surface related calculations were performed with VASP [4, 5]. Computing resources were available through (i) the HWcluster at Heriot-Watt University, (ii) the Academic Computer Center in Gdansk (TASK), (iii) a Computational Grand Challenge Application grant from the Molecular Sciences Computing Facility in the Environmental Molecular Sciences Laboratory (EMSL), and (iv) the National Energy Research Scientific Computing Center (NERSC). Pacific Northwest National Laboratory is operated by Battelle for the U.S. DOE under Contract DE-AC06-76RLO1830.

References:

[1] Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; 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.; Bakken, V.; 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.; and Pople, J. A. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.

[2] Bylaska, E. J. d. J., W. A.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Apra, E.; Windus, T. L.; Hirata, S.; Hackler, M. T.; Zhao, Y.; Fan, P.-D.; Harrison, R. J.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Auer, A. A.; Nooijen, M.; Brown, E.; Cisneros, G.; Fann, G. I.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J. A.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Pollack, L.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. "NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.0" (2006), Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA.

[3] MOLPRO, version 2006.1, a package of ab initio programs, H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, and others , see http://www.molpro.net.

[4] Kresse, G.; Hafner, J., Abinitio Molecular-Dynamics for Liquid-Metals. *Physical Review B* **1993,** 47, 558-561.

[5] Kresse, G.; Furthmuller, J., Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Computational Materials Science* **1996,** 6, 15-50.

Chapter 2

Electronic Structure Methods

As the opposite of experimental chemistry, computational chemistry develops and uses physical models to study atoms, molecules and materials. Computational chemistry uses the power of current computers and performs "experiments" on a computer with the electronic structure methods and many other theoretical methods. The electronic structure methods constitute the most important part of computational chemistry as they provide solutions of the Schrödinger equation, which is the basic principle used in computational chemistry to describe behaviors of electrons and nuclei at the microscale. The electronic structure methods are widely used to study chemical and physical properties of molecules and extended systems. The theoretical results for specified properties of a system based on electronic structure calculations could provide very good references and predictions for ongoing or already finished experimental studies. In this chapter, we will talk about different electronic structure methods which are used in current theoretical studies.

2.1 Essentials of Computational Chemistry

Computational Chemistry is based on solving the many-body Schrödinger equation for a molecule or an extended system which could be described as below (suppose we are dealing with the time-independent case) [1, 2],

$$\mathbf{H}\Psi = E\Psi, \tag{2.1}$$

where the wavefunction Ψ depends on the spatial coordinates of all the particles involved, which include both the electrons and nuclei, and $\mathbf{H}$ is the Hamiltonian. Before we give an exact expression for $\mathbf{H}$, it makes sense for us to choose a good coordinate system as it might help us separate the translational motion of the molecule from the internal motions. Since the traditional space-fixed (SF) coordinate system which origins from an arbitrary chosen point with the axes being arbitrarily orientated is not convenient for us to describe the translational motion of the molecule, it is reasonable to choose a body-fixed (BF) coordinate system which is based on the center-of-mass scheme through which we can separate the motion of the center of mass with the following definition of the

center-of-mass position vector

$$\boldsymbol{R}^{BF} = \frac{\sum_i m_i \boldsymbol{r}_i^{SF}}{M}, \tag{2.2}$$

where i represents both nuclei and electrons, m denotes mass of each single particle involved in the molecule (including both nuclei and electrons), $\boldsymbol{r}^{SF}$ denotes position vector in the old space-fixed coordinate system, and M denotes the total mass. With such an approach, we can separate the translational motion of the center of mass from all other internal motions of the molecule.

In the body-fixed coordinate system, the non-relativistic Hamiltonian can be described by the following equation (in atomic units and thereafter),

$$\mathbf{H} = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_{p,i} \frac{z_p}{r_{pi}} + \sum_{i<j} \frac{1}{r_{ij}} + \sum_{p<q} \frac{z_p z_q}{R_{pq}} + H', \tag{2.3}$$

where subscripts p and q represent nuclei, subscripts i and j represent electrons, z indicates charge of nuclei, R indicates distances between nuclei, r indicates distances between electrons or between electrons and nuclei, H' denotes kinetic energy of nuclei and coupling between motions of the nuclei and electrons. From here until the end of this section, we limit our considerations to a diatomic molecule, and then we have

$$H' = -\frac{1}{2\mu}\nabla_{\boldsymbol{R}}^2 - H'', \tag{2.4}$$

where μ is the reduced mass of the nuclei, and H'' contains coupling between motions of the nuclei and electrons [3].

Before we move further, it makes sense for us to separate the total Hamiltonian into two parts as below,

$$\mathbf{H}_e = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_{p,i} \frac{z_p}{r_{pi}} + \sum_{i<j} \frac{1}{r_{ij}} + \sum_{p<q} \frac{z_p z_q}{R_{pq}}, \tag{2.5}$$

$$\mathbf{H}_n = H', \tag{2.6}$$

where $\mathbf{H}_e$ and $\mathbf{H}_n$ are the electronic Hamiltonian and the nuclear Hamiltonian, respectively, and they obey the electronic Schrödinger equation and the nuclear Schrödinger equation,

$$\mathbf{H}_e \Psi_e(\mathbf{r}; R) = E_e(R)\Psi_e(\mathbf{r}; R), \tag{2.7}$$

$$\mathbf{H}_n \Psi_n(\mathbf{R}) = E_n \Psi_n(\mathbf{R}), \tag{2.8}$$

where $\mathbf{R}$ denotes the nuclear coordinates vector, $\mathbf{r}$ denotes the electronic

coordinates vector, and the electronic wavefunction Ψ_e is a function of the electronic coordinates and it is parametrically dependent on the scalar nuclear coordinate *R*.

As the electronic Hamiltonian $\mathbf{H}_e$ in Eq. (2.7) and the nuclear Hamiltonian $\mathbf{H}_n$ in Eq. (2.8) are Hermitian, their eigenfunctions form a complete set of orthonormal functions $\left\{\Psi_e^l(\mathbf{r};R)\right\}$ and $\left\{\Psi_n^k(\mathbf{R})\right\}$, respectively. The cross product of $\left\{\Psi_e^l(\mathbf{r};R)\right\}$ and $\left\{\Psi_n^k(\mathbf{R})\right\}$ forms a complete set of functions dependent on **r** and **R**. Thus the total wavefunction Ψ in Eq. (2.1) can be represented as

$$\begin{aligned}\Psi(\mathbf{R},\mathbf{r}) &= \sum_{l,k} c_{lk}\Psi_e^l(\mathbf{r};R)\cdot\Psi_n^k(\mathbf{R}) \\ &= \sum_l\left[\Psi_e^l(\mathbf{r};R)\cdot\sum_k c_{lk}\Psi_n^k(\mathbf{R})\right] \\ &= \sum_l \Psi_e^l(\mathbf{r};R)\cdot\Xi_n^l(\mathbf{R})\end{aligned} \tag{2.9}$$

where *l* denotes different electronic states. Inserting Eqs. (2.3) and (2.9) into Eq. (2.1), we have

$$\left(\mathbf{H}_e - \frac{1}{2\mu}\nabla_{\mathbf{R}}^2 - H'' - E\right)\sum_l \Psi_e^l\cdot\Xi_n^l = 0, \tag{2.10}$$

and if we project this equation against the Hermitian conjugate of another electronic state $\left\langle\Psi_e^{\,k}\right|$ in the space of electronic coordinates, we have

$$\sum_l\left\langle\Psi_e^k\left|\left(\mathbf{H}_e - \frac{1}{2\mu}\nabla_{\boldsymbol{R}}^2 - H'' - E\right)\Psi_e^l\cdot\Xi_n^l\right.\right\rangle = 0. \tag{2.11}$$

Considering the orthogonality condition $\left\langle\Psi_e^k\middle|\Psi_e^l\right\rangle = \delta_{kl}$, we have

$$\left(E_e^k - \frac{1}{2\mu}\nabla_{\boldsymbol{R}}^2 + H_{kk}{}' - E\right)\Xi_n^k = -\sum_{k\neq l}\left(-\frac{1}{\mu}\left\langle\Psi_e^k\middle|\nabla_{\boldsymbol{R}}\Psi_e^l\right\rangle\nabla_{\boldsymbol{R}} + H_{kl}{}'\right)\Xi_n^l, \tag{2.12}$$

where $H_{kl}{}' = \left\langle\Psi_e^k\middle|H'\Psi_e^l\right\rangle$. Suppose different electronic states are well separated (for most cases this is true) so that the couplings between different electronic states are relatively small. In this way, we can set the right hand side of Eq. (2.12) (which is the so-called non-adiabatic couplings) equal to zero, through which we get

$$\left(E_e^k - \frac{1}{2\mu}\nabla_{\mathbf{R}}^2 + H_{kk}' \right) \Xi_n^k = E\Xi_n^k . \tag{2.13}$$

This is the so-called adiabatic approximation. With the adiabatic approximation, the total wavefunction in Eq. (2.9) can be written as below,

$$\Psi(\mathbf{R},\mathbf{r}) \approx \Psi_e^k(\mathbf{r};R)\cdot\Xi_n^k(\mathbf{R}), \tag{2.14}$$

which means the total wavefunction can be approximated by a product of the electronic wavefunction and the nuclear wavefunction at a specific electronic state that we are interested in. Moreover, as the diagonal correction H_{kk}' in Eq. (2.13) is usually very small, we usually set H_{kk}' to zero, thus we have

$$\left[E_e^k(R) - \frac{1}{2\mu}\nabla_{\mathbf{R}}^2 \right] \Xi_n^k(\mathbf{R}) \approx E\Xi_n^k(\mathbf{R}). \tag{2.15}$$

This is the Born-Oppenheimer approximation, and it is the most common approximation in computational chemistry. According to the Born-Oppenheimer approximation, once we solve the electronic Schrödinger equation, we will get the electronic energy (which is the eigen energy of the electronic Schrödinger equation) for specified nuclear coordinates, and if we solve the electronic Schrödinger equation for all possible nuclear coordinates, we will get a surface which tells how the electronic energy depends on the nuclear coordinates. Because the electronic energy acts as the potential which decides the eigen energy of the nuclear Schrödinger equation, normally we call the electronic energy the potential energy, and we call the surface the potential energy surface (PES).

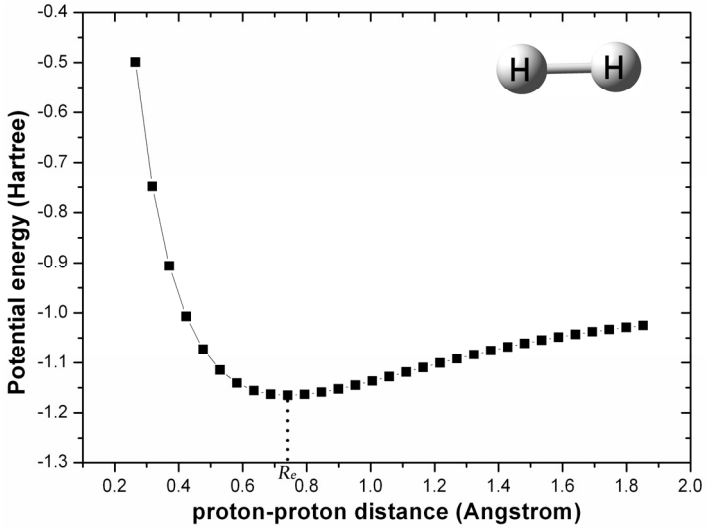


Figure 2.1: Potential energy surface of the hydrogen molecule.

One of the simplest PES pertains to the hydrogen molecule which is shown in Figure 2.1, in which the potential energy is determined at the CCSD/6-31++G** level of theory [4]. From Figure 2.1, we can find that the hydrogen molecule has the lowest potential energy when the two hydrogen atoms are separated with a distance of about 0.7 Ångstrom (see R_e in Figure 2.1), which is the optimal geometry of the hydrogen molecule in the ground electronic state. This is an example showing how the PES decides the structure of the hydrogen molecule.

In most systems, the Born-Oppenheimer approximation works fine. Nevertheless, there are also such occasions where it might fail. For example, we know that the electrons move much faster than the nuclei, and as a result, if the nuclei vibrate or rotate, the electrons can adjust very quickly. However, if for some reason the electrons are loosely bound by the nuclei (e.g. the dipole-bound anion) [5], the electrons can not adjust themselves in a very quick manner when the nuclei vibrate or rotate, and then the Born-Oppenheimer approximation is expected to break down as the non-adiabatic couplings terms in Eq. (2.12) are not negligible any more. Another example is that when the PES generated for different electronic states cross or nearly cross with each other [6], a degeneracy of vibrational levels might occur, which is conflicting with our physical model of the Born-Oppenheimer approximation of which the nuclear motion can only happen on one PES.

2.1.1 Diatomic Molecules

It is very difficult to get the solution of the electronic Schrödinger equation. However, once we solve the electronic Schrödinger equation, it is much easier for us to solve the nuclear Schrödinger equation. Suppose we have already solved the electronic Schrödinger equation and we produced the potential energies, we are now starting to solve the nuclear Schrödinger equation, and our next question would be how to deal with the nuclear wavefunction Ξ_n^k in Eq. (2.15). As the translational motion of the molecule has already been separated from the Hamiltonian in Eq. (2.3), now we only need to consider the rotational and vibrational motions of the nuclei. To simplify the problem, we still need to separate the rotational and vibrational motions of nuclei. An important observation is that E_e^k in Eq. (2.15) depends on scalar R only, while Ξ_n^k in Eq. (2.15) also depends on the orientation of the vector **R**. This allows us to separate

the variables of Ξ_n^k in Eq. (2.15) as below,

$$\Xi_n^k(\mathbf{R}) = Y(\theta,\phi)\frac{\chi_k(R)}{R}, \tag{2.16}$$

of which $Y(\theta,\phi)$ and $\frac{\chi_k(R)}{R}$ are responsible for the rotations and oscillations of the nuclei, respectively. Inserting Eq. (2.16) into Eq. (2.15) and expressing the Laplacian in spherical coordinates, we obtain the following equation,

$$-\left(\frac{R^2}{\chi_k}\frac{d^2\chi_k}{dR^2}\right)+2\mu E_e^k(R)-2\mu ER^2 = \frac{1}{Y}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right). \tag{2.17}$$

From Eq. (2.17), we can see that the left hand side of the equation is only dependent on *R*, while the right hand side of the equation is only dependent on θ and ϕ. Equation (2.17) is only true when both sides equal to a constant (say λ), so we have

$$-\left(\frac{R^2}{\chi_k}\frac{d^2\chi_k}{dR^2}\right)+2\mu E_e^k(R)R^2-2\mu ER^2 = \lambda, \tag{2.18}$$

and

$$\frac{1}{Y}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right) = \lambda. \tag{2.19}$$

Eq. (2.19) is analogous to the Schrödinger equation for a rigid rotor, and it has a solution only if $\lambda = -J(J+1)$, in which $J = 0,1,2,\ldots$ In this way, we have

$$-\frac{1}{2\mu}\left(\frac{d^2\chi_k}{dR^2}\right)+E_e^k(R)\chi_k-E\chi_k = -\frac{1}{2\mu R^2}J(J+1)\chi_k, \tag{2.20}$$

or we can also write it in the form of the eigenvalue problem as below

$$\left(-\frac{1}{2\mu}\frac{d^2}{dR^2}+V_{kJ}\right)\chi_{kvJ}(R) = E_{kvJ}\chi_{kvJ}(R), \tag{2.21}$$

in which

$$V_{kJ}(R) = E_e^k(R)+J(J+1)\frac{1}{2\mu R^2}, \tag{2.22}$$

where *k* is the electronic quantum number, *v* is the vibrational quantum number, and *J* is the rotational quantum number. Equation (2.21) describes the vibrations of the nuclei, while the potential energy V_{kJ} controls the motion of the nuclei on the potential energy surface. From now on, we will only consider small displacements from R_e. Then, we may assume that the second term in the right hand side of Eq. (2.22) does not change at all, thus we have

$$V_{kJ}(R) \approx E_e^k(R) + J(J+1)\frac{1}{2\mu R_e^2}, \tag{2.23}$$

and the second term in the right hand side of the equation is the rotational energy

$$E_{rot}(J) = J(J+1)\frac{1}{2\mu R_e^2}. \tag{2.24}$$

As we know the potential energy of a system increases if a small displacement is applied to the optimal nuclear configuration R_e of this system (see Figure 2.1), for the region where $(R - R_e)$ is very small, we can use a harmonic oscillator to approximate the vibrational motion of the nuclei (see Figure 2.2, in which we use accurate theoretical results from CCSD calculations as reference for comparison).

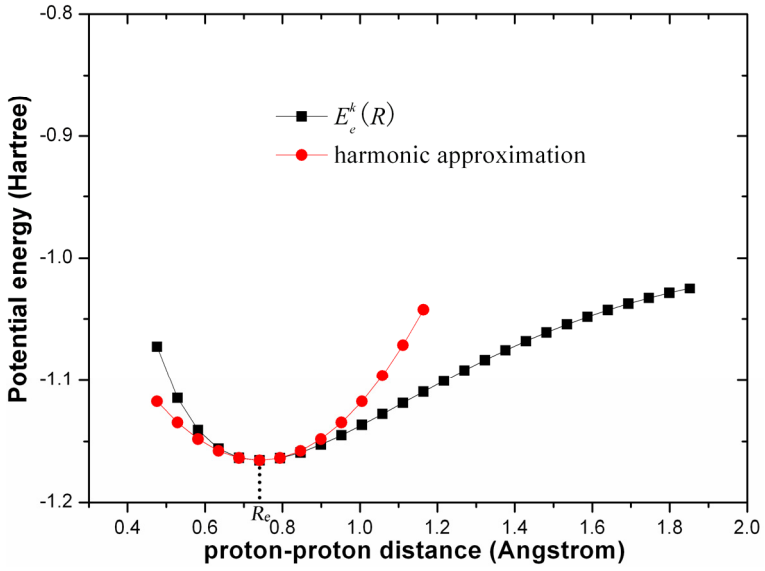


Figure 2.2: Harmonic approximation to the vibrational energy.

Thus we can express the potential energy of this system with the first few terms of a Taylor expansion as below,

$$E_e^k(R) = E_e^k(R_e) + \left.\frac{dE_e^k}{dR}\right|_{R_e}(R - R_e) + \frac{1}{2}\left.\frac{d^2E_e^k}{dR^2}\right|_{R_e}(R - R_e)^2 + \cdots, \tag{2.25}$$

where $\left.\frac{dE_e^k}{dR}\right|_{R_e} = 0$ as the potential energy at R_e is a minimum, and $k_f = \left.\frac{d^2E_e^k}{dR^2}\right|_{R_e}$ is the force constant. Therefore we have

$$E_e^k(R) \approx E_e^k(R_e) + \frac{1}{2} k_f (R - R_e)^2, \tag{2.26}$$

where the first term in the right hand side is the electronic energy at the equilibrium distance R_e,

$$E_{el}(k) = E_e^k(R_e). \tag{2.27}$$

As we used the harmonic oscillator model to approximate the vibrational motion of the nuclei here, we may express the vibrational energy as below,

$$E_{vib}(v) = \left(v + \frac{1}{2}\right)\omega_e, \tag{2.28}$$

in which $v = 0, 1, 2, \ldots$, and $\omega_e = \left(\frac{k_f}{\mu}\right)^{1/2}$. Taking account of the translation energy, and then inserting Eqs. (2.24), (2.27) and (2.28) into Eq. (2.23), we have the final expression for the total energy of a diatomic molecule

$$E_{total} \approx E_{trans} + E_{el}(k) + E_{vib}(v) + E_{rot}(J). \tag{2.29}$$

2.1.2 Polyatomic Molecules

We just talked about separation of vibrations and rotations in diatomic molecules. For polyatomic molecules, they have more degrees of freedom, e.g. the carbon dioxide molecule can vibrate in either the symmetric stretching mode or the antisymmetric stretching mode, or it can also vibrate in the bending mode, which makes the solutions more complicated [7]. For a polyatomic molecule consisting of N atoms, it has a total of $3N$ possibilities of motions, in which each atom contributes three. Considering that there are three degrees of freedom associated with the translational motions and another three degrees of freedom (for nonlinear molecules, or two degrees of freedom for linear molecules) associated with the rotational orientations of the molecule as a whole, and they do not change the potential energy of the molecule, we have to subtract them, which leaves $3N$-5 normal vibrational modes for linear molecules and $3N$-6 normal vibrational modes for nonlinear molecules.

As what we did for diatomic molecules (see Eq. (2.25)), the potential energy of a polyatomic molecule can also be expressed with the first few terms of a Taylor expansion,

$$E_e^k = E_e^k(0) + \sum_i \left(\frac{\partial E_e^k}{\partial x_i}\right)_0 x_i + \frac{1}{2}\sum_{i,j}\left(\frac{\partial^2 E_e^k}{\partial x_i \partial x_j}\right)_0 x_i x_j + \cdots, \qquad (2.30)$$

in which $E_e^k(0)$ denotes the potential energy when atoms in a polyatomic molecule are at their equilibrium positions, and $x_{i,j,\ldots}$ denotes displacements from the equilibrium geometry. Similar to that of a diatomic molecule, when displacements from the equilibrium geometry are very small, the harmonic oscillator model can be used, and the potential energy can be approximated as below,

$$E_e^k \approx E_e^k(0) + \frac{1}{2}\sum_{i,j} k_{ij} x_i x_j, \qquad (2.31)$$

in which $i, j = 1, 2, \ldots, 3N$ denote possible displacements of atoms from their equilibrium positions in the molecule, k_{ij} is a generalized force constant, and $k_{ij} = \left(\frac{\partial^2 E_e^k}{\partial x_i \partial x_j}\right)_0$. The simplification of Eq. (2.31) can benefit from the introduction of mass-weighted coordinates,

$$q_i = m_i^{1/2} x_i, \qquad (2.32)$$

in which m_i is the mass of the atom which has a displacement of x_i. Inserting Eq. (2.32) into Eq. (2.31), we have

$$E_e^k \approx E_e^k(0) + \frac{1}{2}\sum_{i,j} K_{ij} q_i q_j, \qquad (2.33)$$

in which $K_{ij} = \left(\frac{\partial^2 E_e^k}{\partial q_i \partial q_j}\right)_0$. Equation (2.33) is still difficult for practical calculations as there are cross terms with $i \neq j$. To get the eigen energies of different modes of motion, we need to diagonalize the matrix of K_{ij}, which turns Eq. (2.33) into the following equation,

$$E_e^k \approx E_e^k(0) + \frac{1}{2}\sum_i \lambda_i Q_i^2, \qquad (2.34)$$

in which λ_i are effective force constants, $Q_i = \sum_l c_{li} q_l$ turns out to be linear combinations of the mass-weighted coordinates, and there are no cross terms in the above equation. With Eq. (2.34), different modes of motion are decoupled from each other, and the molecular translations and rotations are also separated from molecular vibrations through the fact that the effective force constants corresponding to the molecular translations and rotations equal to zero, which in turn results in distinct eigenvectors Q_i for molecular translations and rotations.

We call Q_i normal coordinates, and they can be determined through diagonalization of the matrix of K_{ij}. Q_i represent different normal modes of motion of the nuclei, which include translations and rotations as well as vibrations. More detailed information about the procedure of determining Q_i is included in Appendix A. Similar to that of a diatomic molecule, the energy levels of the *i*th normal mode of a polyatomic molecule within the harmonic approximation can be expressed as below,

$$E_{i,v_i} = \left(v_i + \frac{1}{2} \right) \omega_i , \tag{2.35}$$

in which the vibrational frequency $\omega_i = \lambda_i^{1/2}$, $v_i = 0,1,2,\ldots$, and then the total vibrational energy of this molecule is

$$E_{vib} = \sum_{i=1}^{3N-6(5)} E_{i,v_i} . \tag{2.36}$$

Once the normal modes and vibrational energy levels are determined, the vibrational partition function can also be determined, and together with the electronic, translational and rotational partition functions, we can obtain the molecular partition function, through which we can calculate the partition function of the entire system, which is an ensemble of molecules assuming the ideal gas model, as well as many of their thermodynamic properties, such as entropy, enthalpy, and Gibbs energy. More detailed information about these thermodynamic properties is included in Appendix A.

2.2 Mathematical Background

There are two fundamental mathematical methods that are used throughout the processes of solving the Schrödinger equation. These are the variational method and the perturbational method. These two methods introduce additional approximations during the processes of solving the Schrödinger equation, which facilitate a fast and yet accurate evaluation of the eigen energy of the Schrödinger equation.

2.2.1 Variational Method

As the first step, our goal is to find the solution for the ground state electronic Schrödinger equation of a system. The variational method is a method which is

based on the variational principle and it facilitates the processes of solving the ground state electronic Schrödinger equation in a mathematical way. The variational method works in such a way that: first we choose a trial wavefunction $\Phi(\mathbf{r},\mathbf{c})$ which includes a set of variational parameters $\mathbf{c} \equiv (c_0, c_1, \cdots, c_P)$, then we calculate the mean value of the Hamiltonian $\mathbf{H}$ with the chosen trial wavefunction $\Phi(\mathbf{r},\mathbf{c})$, and the result can be expressed as below,

$$\varepsilon(\mathbf{c}) = \frac{\langle \Phi(\mathbf{r},\mathbf{c}) | \mathbf{H}\Phi(\mathbf{r},\mathbf{c}) \rangle}{\langle \Phi(\mathbf{r},\mathbf{c}) | \Phi(\mathbf{r},\mathbf{c}) \rangle}. \tag{2.37}$$

Minimizing $\varepsilon(\mathbf{c})$ with respect to different $\mathbf{c}$, finally we will find the optimal set of variational parameters $\mathbf{c}_{opt}$ which give the lowest energy $\varepsilon(\mathbf{c}_{opt})$ that represents an approximation to the ground state energy of the system, while the trial wavefunction $\Phi(\mathbf{r},\mathbf{c}_{opt})$ is an approximation to the ground state wavefunction of this system.

Such a procedure is based on the variational principle which states that $\varepsilon(\mathbf{c})$ is above or equal to the exact energy and the equality is true only if the approximated wavefunction is the exact wavefunction. The variational principle can be proved as shown below. Suppose we already know the exact solutions to the Schrödinger equation,

$$\mathbf{H}\psi_i = E_i\psi_i, \tag{2.38}$$

where *i* represents different quantum states. As we mentioned in the last section, the eigenfunctions of the Hermitian $\mathbf{H}$ form a complete set of orthonormal functions, which means the trial wavefunction can be a linear combination of the functions of this set,

$$\Phi = \sum_i a_i\psi_i, \tag{2.39}$$

and the normalization of Φ requires

$$\langle \Phi | \Phi \rangle = \left\langle \sum_j a_j\psi_j \middle| \sum_i a_i\psi_i \right\rangle = \sum_{i,j} a_j^* a_i \langle \psi_j | \psi_i \rangle = \sum_{i,j} a_j^* a_i \delta_{ij} = \sum_i a_i^* a_i = \sum_i |a_i|^2 = 1 \tag{2.40}$$

Suppose the exact energy (which is also the lowest possible energy) of the system is E_0, the mean value of the energy corresponding to the trial wavefunction Φ is ε, therefore we have

$$\begin{aligned}\varepsilon - E_0 &= \langle \Phi | \mathbf{H} \Phi \rangle - E_0 = \left\langle \sum_j a_j \psi_j \middle| \mathbf{H} \sum_i a_i \psi_i \right\rangle - E_0 \\ &= \sum_{i,j} a_j^* a_i E_i \langle \psi_j | \psi_i \rangle - E_0 = \sum_{i,j} a_j^* a_i E_i \delta_{ij} - E_0 = \sum_i |a_i|^2 E_i - E_0 \\ &= \sum_i |a_i|^2 E_i - E_0 \cdot \sum_i |a_i|^2 = \sum_i |a_i|^2 (E_i - E_0) \geq 0,\end{aligned} \tag{2.41}$$

where the equality is only true when the trial wavefunction Φ equals the exact wavefunction ψ_0. The variational method gives an upper bound to the energy of the ground state.

2.2.2 Perturbational Method

The perturbation method is another mathematical method which can be used to solve the electronic Schrödinger equation. The general idea behind the perturbation method is that we might know the exact solutions for a particular system (the unperturbed system), and this system is very similar to another system (the system with a small perturbation to the unperturbed system) which we want to work with, so we might get the solutions of the second system through observing how the Hamiltonians of the two systems differ from each other. Suppose we want to solve the following Schrödinger equation,

$$\mathbf{H} \psi_k = E_k \psi_k , \tag{2.42}$$

in which the Hamiltonian operator

$$\mathbf{H}(\lambda) = \mathbf{H}^{(0)} + \lambda \mathbf{H}^{(1)} , \tag{2.43}$$

where $\mathbf{H}^{(0)}$ is the unperturbed Hamiltonian operator, $\mathbf{H}^{(1)}$ is the perturbation operator which is small compared with $\mathbf{H}^{(0)}$, and the perturbation parameter $0 \leq \lambda \leq 1$ controls the magnitude of perturbation that we introduce to $\mathbf{H}^{(0)}$. As the solution of the unperturbed system is known, its eigen energy $E_k^{(0)}$ and exact wavefunction $\psi_k^{(0)}$ are also known. In this way, we want to approximate the energy and wavefunction of the real system as a power series in λ as below,

$$E_k(\lambda) = E_k^{(0)} + \lambda E_k^{(1)} + \lambda^2 E_k^{(2)} + \cdots , \tag{2.44}$$

$$\psi_k(\lambda) = \psi_k^{(0)} + \lambda \psi_k^{(1)} + \lambda^2 \psi_k^{(2)} + \cdots . \tag{2.45}$$

Inserting Eqs. (2.43), (2.44) and (2.45) into Eq. (2.42), we have

$$\begin{aligned}&\left(\mathbf{H}^{(0)} + \lambda \mathbf{H}^{(1)}\right)\left(\psi_k^{(0)} + \lambda \psi_k^{(1)} + \lambda^2 \psi_k^{(2)} + \cdots\right) \\ &= \left(E_k^{(0)} + \lambda E_k^{(1)} + \lambda^2 E_k^{(2)} + \cdots\right)\left(\psi_k^{(0)} + \lambda \psi_k^{(1)} + \lambda^2 \psi_k^{(2)} + \cdots\right).\end{aligned} \tag{2.46}$$

Expanding Eq. (2.46) and considering the fact that the coefficients at the same

powers of λ on both sides of Eq. (2.46) have to be equal to each other, we have

$$\mathbf{H}^{(0)}\psi_k^{(0)} = E_k^{(0)}\psi_k^{(0)}, \tag{2.47}$$

$$\mathbf{H}^{(0)}\psi_k^{(1)} + \mathbf{H}^{(1)}\psi_k^{(0)} = E_k^{(0)}\psi_k^{(1)} + E_k^{(1)}\psi_k^{(0)}, \tag{2.48}$$

$$\mathbf{H}^{(0)}\psi_k^{(2)} + \mathbf{H}^{(1)}\psi_k^{(1)} = E_k^{(0)}\psi_k^{(2)} + E_k^{(1)}\psi_k^{(1)} + E_k^{(2)}\psi_k^{(0)}. \tag{2.49}$$

…

Eq. (2.47) is known to be the solution of the unperturbed system, while Eq. (2.48) and Eq. (2.49) contain a lot of unknowns. If we project Eq. (2.48) against $\left\langle \psi_k^{(0)} \right|$, we have

$$\begin{aligned}
&\left\langle \psi_k^{(0)} \left| \left(\mathbf{H}^{(0)} - E_k^{(0)}\right)\psi_k^{(1)} + \left(\mathbf{H}^{(1)} - E_k^{(1)}\right)\psi_k^{(0)} \right\rangle \right. \\
&= \left\langle \left(\mathbf{H}^{(0)} - E_k^{(0)}\right)\psi_k^{(0)} \left| \psi_k^{(1)} \right\rangle + \left\langle \psi_k^{(0)} \left| \left(\mathbf{H}^{(1)} - E_k^{(1)}\right)\psi_k^{(0)} \right\rangle \right. \\
&= 0 + \left\langle \psi_k^{(0)} \left| \left(\mathbf{H}^{(1)} - E_k^{(1)}\right)\psi_k^{(0)} \right\rangle \right. = 0,
\end{aligned} \tag{2.50}$$

thus

$$E_k^{(1)} = \left\langle \psi_k^{(0)} \left| \mathbf{H}^{(1)}\psi_k^{(0)} \right\rangle \right., \tag{2.51}$$

which means the first order correction to the energy equals the mean value of the perturbation with the exact wavefunction of the unperturbed system. Similarly, if we project Eq. (2.49) against $\left\langle \psi_k^{(0)} \right|$, and we take account of the intermediate normalization condition $\left\langle \psi_k^{(0)} \left| \psi_k^{(n)} \right\rangle \right. = \delta_{0n}$, we have

$$\begin{aligned}
&\left\langle \psi_k^{(0)} \left| \left(\mathbf{H}^{(0)} - E_k^{(0)}\right)\psi_k^{(2)} + \left(\mathbf{H}^{(1)} - E_k^{(1)}\right)\psi_k^{(1)} - E_k^{(2)}\psi_k^{(0)} \right\rangle \right. \\
&= \left\langle \left(\mathbf{H}^{(0)} - E_k^{(0)}\right)\psi_k^{(0)} \left| \psi_k^{(2)} \right\rangle + \left\langle \psi_k^{(0)} \left| \mathbf{H}^{(1)}\psi_k^{(1)} \right\rangle \right. - \left\langle \psi_k^{(0)} \left| E_k^{(1)}\psi_k^{(1)} \right\rangle \right. \\
&\quad - \left\langle \psi_k^{(0)} \left| E_k^{(2)}\psi_k^{(0)} \right\rangle \right. = 0 + \left\langle \psi_k^{(0)} \left| \mathbf{H}^{(1)}\psi_k^{(1)} \right\rangle \right. - 0 - E_k^{(2)} = 0,
\end{aligned} \tag{2.52}$$

thus

$$E_k^{(2)} = \left\langle \psi_k^{(0)} \left| \mathbf{H}^{(1)}\psi_k^{(1)} \right\rangle \right.. \tag{2.53}$$

As the solutions of the unperturbed Schrödinger equation form a complete set of functions, the first order correction to the wavefunction can be expanded as a linear combination of the basis functions $\left\{\psi_n^{(0)}\right\}_{n\neq k}$ as below,

$$\psi_k^{(1)} = \sum_{n\neq k} b_n \psi_n^{(0)}, \tag{2.54}$$

in which the coefficients b_n are yet to be determined, and $n = k$ is excluded because of the intermediate normalization condition $\left\langle \psi_k^{(0)} \left| \psi_k^{(1)} \right\rangle \right. = 0$. Inserting Eq. (2.54) into Eq. (2.48), we have

$$\left(\mathbf{H}^{(0)}-E_k^{(0)}\right)\sum_{n\neq k}b_n\psi_n^{(0)}+\mathbf{H}^{(1)}\psi_k^{(0)}=\sum_{n\neq k}b_n\left(E_n^{(0)}-E_k^{(0)}\right)\psi_n^{(0)}+\mathbf{H}^{(1)}\psi_k^{(0)}=E_k^{(1)}\psi_k^{(0)} \tag{2.55}$$

The coefficient b_m can be determined by projecting Eq. (2.55) against $\left\langle\psi_m^{(0)}\right|$, and considering the orthonormality of $\left\{\psi_n^{(0)}\right\}_{n\neq k}$, we have

$$b_m=\frac{\left\langle\psi_m^{(0)}\left|\mathbf{H}^{(1)}\psi_k^{(0)}\right.\right\rangle}{E_k^{(0)}-E_m^{(0)}}. \tag{2.56}$$

Inserting Eq. (2.56) into Eq. (2.54), we obtain the first order correction to the wavefunction,

$$\psi_k^{(1)}=\sum_{n\neq k}\frac{\left\langle\psi_n^{(0)}\left|\mathbf{H}^{(1)}\psi_k^{(0)}\right.\right\rangle}{E_k^{(0)}-E_n^{(0)}}\psi_n^{(0)}. \tag{2.57}$$

Inserting Eq. (2.57) into Eq. (2.53), we obtain the second order correction to the energy,

$$E_k^{(2)}=\sum_{n\neq k}\frac{\left|\left\langle\psi_k^{(0)}\left|\mathbf{H}^{(1)}\psi_n^{(0)}\right.\right\rangle\right|^2}{E_k^{(0)}-E_n^{(0)}}. \tag{2.58}$$

The above process gives the two most important corrections to the energy of the real system, while further corrections are less important and much more complicated to obtain, so we decide to stop here. In Eq. (2.58), if k denotes the ground state and $n\neq k$, we have $E_k^{(0)}<E_n^{(0)}$ thus $E_k^{(0)}-E_n^{(0)}<0$, and then $E_k^{(2)}<0$, which means the second order correction to the energy is negative and it will always stabilize the system.

2.3 Schrödinger Equation and Its Solutions

As discussed above, the essence of solving the many-body Schrödinger equation is to solve the electronic Schrödinger equation. Generally, there are two approaches to the electronic Schrödinger equation. One is through the most popular molecular orbital theory, which is more efficient and more accepted by theorists as it is accurate and easy accessible, and the other is the less popular valence bond theory, which is less efficient but is more recognized by experimentalists as a direct result of the theory itself. These two approaches are different from each other as in the valence bond theory, an electron pair is

localized between two interacting atoms, while in the molecular orbital theory, the molecular orbitals are not only determined by the atomic orbitals of two bonded atoms, they also spread over the whole molecule. In the current study, we employed the widely used molecular orbital methods to study different sorts of problems that we will talk about in the following chapters.

As a very good starting point, we use the Born-Oppenheimer approximation through which we can separate the many-body Schrödinger equation into the electronic Schrödinger equation and the nuclear Schrödinger equation. The solution of the electronic Schrödinger equation, which is the electronic wavefunction, depends only on the spatial and spin coordinates of electrons, and if we solve the electronic Schrödinger equation for different nuclear coordinates in the body-fixed coordinate system, we will get the PES. Once we get the PES, then the nuclear Schrödinger equation can also be solved based on the assumption that the nuclei move in an effective potential obtained by solving the electronic Schrödinger equation for various internuclear positions. The key to solve the many-body Schrödinger equation is to solve the electronic Schrödinger equation. Approaches towards solving the electronic Schrödinger equation are what we call "electronic structure methods", which are the most important tools used across the whole thesis and thus deserve a significant amount of attention.

2.3.1 Slater Determinant and Hartree-Fock Equations

As an electron is a fermion, the electronic wavefunction has to be exchange-antisymmetric. An approximated electronic wavefunction with the exchange-antisymmetric character could be written as below,

$$\Psi_e^A(1,2,\cdots,N)=\frac{1}{\sqrt{N!}}\begin{vmatrix}\psi_1(1) & \psi_1(2) & \cdots & \psi_1(N)\\ \psi_2(1) & \psi_2(2) & \cdots & \psi_2(N)\\ \vdots & \vdots & \vdots & \vdots\\ \psi_N(1) & \psi_N(2) & \cdots & \psi_N(N)\end{vmatrix}, \tag{2.59}$$

where ψ_i are orthonormal one-electron wavefunctions (spin orbitals) with $\frac{1}{\sqrt{N!}}$ being the normalization constant, and numbers in brackets denote total coordinates of different electrons which include both spatial and spin coordinates. Equation (2.59) is what we call the single Slater determinant and it expresses the molecular electronic wavefunction in the form of a determinant which is constructed through one-electron wavefunctions ψ_i with different spatial and

spin coordinates. According to Eq. (2.59), if there are two electrons which have exactly the same spatial and spin coordinates, e.g. $\psi_i(1)=\psi_i(2)$ is true for all $i=1,2,\cdots,N$, and then we have $\Psi_e^A=0$ as a result of a basic property of determinants, which means that electrons with the same spin can not approach each other. This is the result of the *Pauli exclusion principle* which states that no two identical fermions may occupy the same quantum state simultaneously.

Now we already have an approximated electronic wavefunction in the form of a Slater determinant. According to the variational method, which we have just talked about in the last section, our next step would be to find the optimal set of one-electron wavefunctions which gives the lowest energy corresponding to the wavefunction in such a class. To facilitate subsequent mathematical derivation, we decide to introduce the antisymmetrization operator **A** and denote the single Slater determinant in Eq. (2.59) with the following equation,

$$\Psi_e^A=\mathbf{A}[\psi_1(1)\psi_2(2)\cdots\psi_N(N)]=\mathbf{A}\Pi\,, \tag{2.60}$$

in which Π is a product of the diagonal elements in the single Slater determinant, and the antisymmetrization operator is defined as below,

$$\mathbf{A}=\frac{1}{\sqrt{N!}}\sum_P(-1)^p\,\mathbf{P}\,, \tag{2.61}$$

in which **P** is the permutation operator which represents possible permutations of the N electrons and p denotes the number of transpositions of two electrons needed for a given permutation P. It could be proved that **A** is Hermitian and obeys the following rules [8]

$$\begin{aligned}&\mathbf{AH}=\mathbf{HA}\\&\mathbf{AA}=\sqrt{N!}\mathbf{A}\cdot\end{aligned} \tag{2.62}$$

Before we move further, it would be very helpful if we divide the electronic Hamiltonian (see Eq. (2.5)) into three parts as below,

$$\mathbf{H}_e=\sum_i\mathbf{h}(i)+\sum_{i<j}\mathbf{g}(ij)+\mathbf{V}_{nn}\,, \tag{2.63}$$

where

$$\mathbf{h}(i)=-\frac{1}{2}\nabla_i^2-\sum_p\frac{z_p}{r_{pi}} \tag{2.64}$$

is the one electron operator which includes the electronic kinetic energy and attraction between an electron i and all the nuclei,

$$\mathbf{g}(ij)=\frac{1}{r_{ij}} \tag{2.65}$$

is the two electron operator which describes the electron-electron repulsion, and

$$\mathbf{V}_{nn} = \sum_{p<q} \frac{z_p z_q}{R_{pq}} \tag{2.66}$$

is the nuclear repulsion operator. Inserting Eq. (2.60) into Eq. (2.37), we get the energy corresponding to the single Slater determinant which can be expressed as below,

$$\begin{aligned} E_e &= \left\langle \Psi_e^A \middle| \mathbf{H}_e \middle| \Psi_e^A \right\rangle = \left\langle \mathbf{A}\Pi \middle| \mathbf{H}_e \middle| \mathbf{A}\Pi \right\rangle \\ &= \int \left(\mathbf{A}\Pi\right)^* \mathbf{H}_e \left(\mathbf{A}\Pi\right) d\tau_e = \int \Pi^* \mathbf{A}\mathbf{H}_e \left(\mathbf{A}\Pi\right) d\tau_e \\ &= \int \Pi^* \mathbf{H}_e \mathbf{A}\mathbf{A}\Pi d\tau_e = \int \Pi^* \mathbf{H}_e \left(\sqrt{N!}\mathbf{A}\Pi\right) d\tau_e \\ &= \sqrt{N!} \left\langle \Pi \middle| \mathbf{H}_e \middle| \mathbf{A}\Pi \right\rangle = \sum_P (-1)^p \left\langle \Pi \middle| \mathbf{H}_e \middle| \mathbf{P}\Pi \right\rangle. \end{aligned} \tag{2.67}$$

Let us get back to the electronic Hamiltonian $\mathbf{H}_e$ (see Eq. (2.63)). As the nuclear repulsion operator $\mathbf{V}_{nn}$ in Eq. (2.66) is independent on the electronic coordinates, we have

$$\left\langle \Psi_e^A \middle| \mathbf{V}_{nn} \middle| \Psi_e^A \right\rangle = V_{nn} \left\langle \Psi_e^A \middle| \Psi_e^A \right\rangle = V_{nn}. \tag{2.68}$$

For the one electron operator $\mathbf{h}(i)$, there are two possibilities. If **P** is the identity operator, we have (taking electron *1* as an example)

$$\begin{aligned} \left\langle \Pi \middle| \mathbf{h}(1) \middle| \Pi \right\rangle &= \left\langle \psi_1(1)\psi_2(2)\cdots\psi_N(N) \middle| \mathbf{h}(1) \middle| \psi_1(1)\psi_2(2)\cdots\psi_N(N) \right\rangle \\ &= \left\langle \psi_1(1) \middle| \mathbf{h}(1) \middle| \psi_1(1) \right\rangle \left\langle \psi_2(2) \middle| \psi_2(2) \right\rangle \cdots \left\langle \psi_N(N) \middle| \psi_N(N) \right\rangle \\ &= \left\langle \psi_1(1) \middle| \mathbf{h}(1) \middle| \psi_1(1) \right\rangle = h_1, \end{aligned} \tag{2.69}$$

and from the definition of the one electron operator in Eq. (2.64), we can see h_1 represents the mean value of the kinetic energy of an electron in spin orbital ψ_1 plus the potential energy which arises from electrostatic attractions between the electronic density of an electron described with spin orbital ψ_1 and point charges of different nuclei. If **P** is not the identity operator, which means there is permutation between different electrons, we have (taking electron *1* and permutation $\mathbf{P}_{12}$ as an example)

$$\begin{aligned} \left\langle \Pi \middle| \mathbf{h}(1) \middle| \mathbf{P}_{12}\Pi \right\rangle &= \left\langle \psi_1(1)\psi_2(2)\cdots\psi_N(N) \middle| \mathbf{h}(1) \middle| \psi_2(1)\psi_1(2)\cdots\psi_N(N) \right\rangle \\ &= \left\langle \psi_1(1) \middle| \mathbf{h}(1) \middle| \psi_2(1) \right\rangle \left\langle \psi_2(2) \middle| \psi_1(2) \right\rangle \cdots \left\langle \psi_N(N) \middle| \psi_N(N) \right\rangle = 0, \end{aligned} \tag{2.70}$$

thus all integrations of the one electron operator $\mathbf{h}(i)$ in Eq. (2.67) that include

permutations between different electrons, equal zero. For the two electron operator $\mathbf{g}(ij)$, the situation is similar to that of the one electron operator $\mathbf{h}(i)$. First, if **P** is the identity operator, we have (taking electrons 1 and 2 as an example)

$$\begin{aligned}\langle\Pi|\mathbf{g}(12)|\Pi\rangle &= \langle\psi_1(1)\psi_2(2)\cdots\psi_N(N)|\mathbf{g}(12)|\psi_1(1)\psi_2(2)\cdots\psi_N(N)\rangle \\ &= \langle\psi_1(1)\psi_2(2)|\mathbf{g}(12)|\psi_1(1)\psi_2(2)\rangle\cdots\langle\psi_N(N)|\psi_N(N)\rangle \\ &= \langle\psi_1(1)\psi_2(2)|\mathbf{g}(12)|\psi_1(1)\psi_2(2)\rangle = J_{12},\end{aligned} \tag{2.71}$$

in which we call J the Coulomb integral and it is the direct result of electron-electron repulsion between the electron densities associated with the ψ_1 and ψ_2 spin orbitals. The Coulomb integral J is positive, thus it is a destabilizing effect. Second, if **P** is not the identity operator and it represents permutation between two different electrons, we have (taking electrons 1 and 2 and permutation $\mathbf{P}_{12}$ as an example)

$$\begin{aligned}\langle\Pi|\mathbf{g}(12)|\mathbf{P}_{12}\Pi\rangle &= \langle\psi_1(1)\psi_2(2)\cdots\psi_N(N)|\mathbf{g}(12)|\psi_2(1)\psi_1(2)\cdots\psi_N(N)\rangle \\ &= \langle\psi_1(1)\psi_2(2)|\mathbf{g}(12)|\psi_2(1)\psi_1(2)\rangle\cdots\langle\psi_N(N)|\psi_N(N)\rangle \\ &= \langle\psi_1(1)\psi_2(2)|\mathbf{g}(12)|\psi_2(1)\psi_1(2)\rangle = K_{12},\end{aligned} \tag{2.72}$$

in which we call K the exchange integral and it results from using a Slater determinant to describe an *N*-electron system. If a single product of spin orbitals was used (Hartree approximation), then only the Coulomb integral would exist. Considering the coefficient $(-1)^p$ in Eq. (2.67), there is a minus sign before the exchange integral in the final result. At the same time, because of the orthogonality of spin functions, it can be proved that, for electrons with opposite spins, the exchange integral equals to zero, and for electrons with parallel spins, the exchange integral turns out to be positive, thus we could regard K as a correction to J, which reduces the destabilizing effect that J brings to the system. For other situations in which **P** is different from the above two cases, the integral is zero due to the orthogonality of different spin orbitals. In this way, the energy corresponding to the single Slater determinant can be written as below,

$$E_e = \sum_i h_i + \sum_{i<j}\left(J_{ij} - K_{ij}\right) + V_{nn} = \sum_i h_i + \frac{1}{2}\sum_{i,j}\left(J_{ij} - K_{ij}\right) + V_{nn}, \tag{2.73}$$

in which the subscripts *i* and *j* in the first two terms on the right hand side of the above equation denote spinorbitals, and h_i, J_{ij} and K_{ij} were defined according

to Eqs. (2.69), (2.71) and (2.72), respectively.

Considering that the best set of one-electron wavefunctions, which gives the lowest energy E_e, has to remain orthonormal, we use the Lagrange multipliers to handle the variation of energy E_e with respect to the variation of one-electron wavefunctions and we introduce the following Lagrange function and its variation,

$$L = E_e - \sum_{i,j} \lambda_{ij} \left(\left\langle \psi_i \middle| \psi_j \right\rangle - \delta_{ij} \right), \tag{2.74}$$

$$\delta L = \delta E_e - \sum_{i,j} \lambda_{ij} \left(\left\langle \delta\psi_i \middle| \psi_j \right\rangle + \left\langle \psi_i \middle| \delta\psi_j \right\rangle \right) = 0, \tag{2.75}$$

where λ_{ij} are the Lagrange multipliers. To facilitate the variation of energy E_e in Eq. (2.75), we may express Eq. (2.73) in terms of the Coulomb and exchange operators as below,

$$E_e = \sum_i \left\langle \psi_i \middle| \mathbf{h} \middle| \psi_i \right\rangle + \frac{1}{2} \sum_{i,j} \left(\left\langle \psi_j \middle| \mathbf{J}_i \middle| \psi_j \right\rangle - \left\langle \psi_j \middle| \mathbf{K}_i \middle| \psi_j \right\rangle \right) + V_{nn}, \tag{2.76}$$

in which the Coulomb and exchange operators work in the following way,

$$\mathbf{J}_i(1) \left| \psi_j(1) \right\rangle = \left\langle \psi_i(2) \middle| \mathbf{g}(12) \middle| \psi_i(2) \right\rangle \left| \psi_j(1) \right\rangle, \tag{2.77}$$

$$\mathbf{K}_i(1) \left| \psi_j(1) \right\rangle = \left\langle \psi_i(2) \middle| \mathbf{g}(12) \middle| \psi_j(2) \right\rangle \left| \psi_i(1) \right\rangle. \tag{2.78}$$

In this way, we have the variation of the energy E_e

$$\begin{aligned}
\delta E_e &= \sum_i \left(\left\langle \delta\psi_i \middle| \mathbf{h} \middle| \psi_i \right\rangle + \left\langle \psi_i \middle| \mathbf{h} \middle| \delta\psi_i \right\rangle \right) + \frac{1}{2} \sum_{i,j} \left(\left\langle \delta\psi_j \middle| \mathbf{J}_i - \mathbf{K}_i \middle| \psi_j \right\rangle + \right. \\
&\left. \left\langle \psi_j \middle| \mathbf{J}_i - \mathbf{K}_i \middle| \delta\psi_j \right\rangle \right) + \frac{1}{2} \sum_{i,j} \left(\left\langle \delta\psi_i \middle| \mathbf{J}_j - \mathbf{K}_j \middle| \psi_i \right\rangle + \left\langle \psi_i \middle| \mathbf{J}_j - \mathbf{K}_j \middle| \delta\psi_i \right\rangle \right) \\
&= \sum_i \left(\left\langle \delta\psi_i \middle| \mathbf{h} \middle| \psi_i \right\rangle + \left\langle \psi_i \middle| \mathbf{h} \middle| \delta\psi_i \right\rangle \right) + \sum_{i,j} \left(\left\langle \delta\psi_i \middle| \mathbf{J}_j - \mathbf{K}_j \middle| \psi_i \right\rangle + \right. \\
&\left. \left\langle \psi_i \middle| \mathbf{J}_j - \mathbf{K}_j \middle| \delta\psi_i \right\rangle \right) = \sum_i \left(\left\langle \delta\psi_i \middle| \mathbf{F} \middle| \psi_i \right\rangle + \left\langle \psi_i \middle| \mathbf{F} \middle| \delta\psi_i \right\rangle \right)
\end{aligned} \tag{2.79}$$

in which the Fock operator

$$\mathbf{F} = \mathbf{h} + \sum_j \left(\mathbf{J}_j - \mathbf{K}_j \right). \tag{2.80}$$

Inserting Eq. (2.79) into Eq. (2.75) and considering that $\left\langle \psi_i \middle| \delta\psi_j \right\rangle = \left\langle \delta\psi_j \middle| \psi_i \right\rangle^*$ and $\left\langle \psi_i \middle| \mathbf{F} \middle| \delta\psi_i \right\rangle = \left\langle \delta\psi_i \middle| \mathbf{F} \middle| \psi_i \right\rangle^*$, we have

$$\delta L = \sum_i \langle \delta\psi_i | \mathbf{F} | \psi_i \rangle - \sum_{i,j} \lambda_{ij} \langle \delta\psi_i | \psi_j \rangle + \sum_i \langle \delta\psi_i | \mathbf{F} | \psi_i \rangle^* - \sum_{i,j} \lambda_{ij} \langle \delta\psi_j | \psi_i \rangle^* = 0. \tag{2.81}$$

Eq. (2.81) is only true when both the first two terms and the last two terms equal zero. Thus we have

$$\sum_i \langle \delta\psi_i | \mathbf{F} | \psi_i \rangle - \sum_{i,j} \lambda_{ij} \langle \delta\psi_i | \psi_j \rangle = 0, \tag{2.82}$$

$$\sum_i \langle \delta\psi_i | \mathbf{F} | \psi_i \rangle^* - \sum_{i,j} \lambda_{ij} \langle \delta\psi_j | \psi_i \rangle^* = 0. \tag{2.83}$$

If we subtract the complex conjugate of Eq. (2.83) from Eq. (2.82), we get

$$\sum_{i,j} \left(\lambda_{ij} - \lambda_{ji}^* \right) \langle \delta\psi_i | \psi_j \rangle = 0, \tag{2.84}$$

from which we get $\lambda_{ij} - \lambda_{ji}^* = 0$ and $\lambda_{ij} = \lambda_{ji}^*$, which means the Lagrange multipliers λ_{ij} can be regarded as elements of a Hermitian matrix. From Eq. (2.82), we also get the final set of Hartree-Fock (HF) equations as below,

$$\mathbf{F}\psi_i = \sum_j \lambda_{ij} \psi_j . \tag{2.85}$$

With a unitary transformation performed on the above equation, it is possible to diagonalize the matrix which consists of the Lagrange multipliers λ_{ij}, and the resulting diagonal elements ε_i obey the following pseudo-eigenvalue equations,

$$\mathbf{F}\psi_i' = \varepsilon_i \psi_i', \tag{2.86}$$

in which we call the transformed spin orbital ψ_i' a canonical spin orbital and ε_i the orbital energy corresponding to the canonical spin orbital ψ_i'. From now on, we will use ψ_i for the canonical spin orbitals. Therefore the energy corresponding to a single Slater determinant can be expressed with the orbital energies and other components as below,

$$E_{HF} = \sum_i \varepsilon_i - \frac{1}{2} \sum_{i,j} \left(J_{ij} - K_{ij} \right) + V_{nn}, \tag{2.87}$$

in which

$$\varepsilon_i = \langle \psi_i | \mathbf{F} | \psi_i \rangle = h_i + \sum_j \left(J_{ij} - K_{ij} \right), \tag{2.88}$$

and the second term on the right hand side of Eq. (2.87) results from the fact that electron-electron repulsions are counted twice in the sum of orbital energies. From Eq. (2.87), we can also see that the electronic energy is not simply the sum of orbital energies, but it also includes contributions from the Coulomb and exchange integrals as well as nuclear repulsion.

The discussions in this section indicate that electron-electron repulsions are treated in an average way for each electron because we use a single Slater determinant as the trial wavefunction, and this was clearly shown when we were talking about the Coulomb and exchange integrals, which resulted from interactions between one electron and all the other electrons, with their spatial distributions determined by a set of spin orbitals. In this way, the HF method is also understood to be a mean field approximation.

2.3.2 Self-consistent Field Method

From Eq. (2.80), we can see that actually the Fock operator depends on all spin orbitals, which are the solutions that we wanted to find in the very beginning of the last section. To get the orbital energies ε_i, we also need to know the canonical spin orbitals, which can only be determined if all the spin orbitals were known. Such a situation can be easily dealt with using an iterative approach as below:

(1) generate initial guess for spin orbitals;

(2) form Fock operator with these spin orbitals, solve the pseudo-eigenvalue equations, and produce new spin orbitals ψ_i and associated orbital energies ε_i;

(3) evaluate the energy E_e with these new spin orbitals;

(4) use these new spin orbitals to build a new Fock operator and repeat the above process until the difference in E_e's between two successive iterations falls below a certain threshold which has been predetermined.

The above procedure is called as the self-consistent field (SCF) method. It is a typical technique which is widely used to solve the Hartree-Fock equations.

2.3.3 Koopmans' Theorem

In Eqs. (2.73) and (2.87), we gave the energy expression for a molecule in its neutral state. Considering a molecule which has *N* electrons and we remove one electron from a spin orbital ψ_k, and supposing that all other spin orbitals remain unchanged, we have the energy expression for this new system as below

$$E_e^+ = \sum_{i' \neq k} h_{i'} + \frac{1}{2} \sum_{i', j' \neq k} \left(J_{i'j'} - K_{i'j'} \right) + V_{nn} . \tag{2.89}$$

Subtracting E_e from E_e^+ in Eq. (2.73), we get

$$E_e^+ - E_e = -h_k - \frac{1}{2}\sum_i \left(J_{ik} - K_{ik}\right) - \frac{1}{2}\sum_j \left(J_{jk} - K_{jk}\right)$$
$$= -\left[h_k + \sum_i \left(J_{ik} - K_{ik}\right)\right] = -\varepsilon_k ,$$

(2.90)

which means the ionization potential of removing one electron from the spin orbital ψ_k equals to the absolute value of the orbital energy corresponding to the spin orbital ψ_k. A similar result can also be obtained for an anion, which states that the electron affinity of binding an excess electron in an empty orbital (e.g. spin orbital ψ_k) of the neutral molecule equals to the absolute value of the orbital energy corresponding to this spin orbital. This is the so-called *Koopmans' theorem*, which is based on the assumption that spin orbitals are "frozen" during the process of adding an electron to or removing an electron from the neutral molecule, while generally orbital relaxations exist in these processes. On the other hand, just as we mentioned in the last section, the Hartree-Fock method is also an approximated method. These intrinsic disadvantages make Koopmans' theorem to be just an approximated approach for calculating the ionization potential and electron affinity of a molecule.

2.3.4 Electron Correlation Energy

As a variational approach, the SCF method gained some success in the beginning when it was just developed. However, as it was applied to different systems and more complicated problems, people found it to be an unsatisfactory theoretical model, e.g. the SCF method gives the wrong dissociation limit for a hydrogen molecule [9]. The problem of the SCF method is that it is limited to a single Slater determinant as the trial wavefunction, thus it might still miss some of the most important physics that electron holds. In order to obtain better results, one might use a trial wavefunction which has more than one Slater determinant and such a multi-determinant trial wavefunction can be expressed as below,

$$\Psi_e^A = a_0 \Psi_e^{HF} + \sum_i a_i \Psi_e^i , \qquad (2.91)$$

in which a_i are expansion coefficients, Ψ_e^i are excited Slater determinants constructed by exciting electrons from occupied molecular orbitals (MOs) in the HF determinant Ψ_e^{HF} to unoccupied virtual MOs. The advantage of using a multi-determinant trial wavefunction is that such trial wavefunction gives much better descriptions of the motion of electrons being correlated with each other (we

call this behavior of electrons electron correlations), while a trial wavefunction based on a single Slater determinant fails to account for the electron correlations. As a result, the Hartree-Fock energy from the SCF calculation is always higher than the exact energy, and we call the energy difference the electron correlation energy, which can be expressed as below,

$$E_{corr} = E_{exact} - E_{HF}, \tag{2.92}$$

and the electron correlation energy E_{corr} is always negative. Methods based on using a multi-determinant wavefunction to better account for the electron correlations are called electron correlation methods, which include Møller–Plesset perturbation theory, configuration interaction methods, and coupled cluster methods, to name but a few.

2.4 Møller–Plesset Perturbation Theory

The Møller–Plesset (MP) perturbation theory [10] starts from the Hartree-Fock solution to the electronic Schrödinger equation. Based on the perturbational method which we talked about in Section 2.2.2., the Møller–Plesset perturbation theory uses the sum of Fock operators (see Eq. (2.80)) as the unperturbed Hamiltonian operator $\mathbf{H}^{(0)}$ (suppose we are dealing with a system with N electrons),

$$\mathbf{H}^{(0)} = \sum_{m=1}^{N} \mathbf{F}(m), \tag{2.93}$$

in which $\mathbf{F}(m)\psi_i(m) = \varepsilon_i \psi_i(m)$, and spin orbitals ψ_i are occupied and virtual MOs. We will denote the ground state Hartree-Fock wavefunction as $\psi_0^{(0)}$ (the subscript 0 here denotes the ground state). As $\psi_0^{(0)}$ is an antisymmetrized product of the spin orbitals ψ_i, $\psi_0^{(0)}$ can be expressed as a linear combination of terms representing different permutations of electrons among the spin orbitals ψ_i (see Eqs. (2.60) and (2.61)). Considering the solution of the unperturbed system as shown in Eq. (2.47), we have

$$\mathbf{H}^{(0)}\psi_0^{(0)} = \left[\sum_{m=1}^{N} \mathbf{F}(m)\right]\psi_0^{(0)} = \left(\sum_{i}^{occ} \varepsilon_i\right)\psi_0^{(0)}, \tag{2.94}$$

which means our reference energy, which is the energy of the unperturbed system $E_0^{(0)}$, equals the sum of the orbital energies of the occupied MOs, thus we have

$$E_{MP0} = E_0^{(0)} = \sum_i^{occ} \varepsilon_i \,. \tag{2.95}$$

Having got the reference energy, our next step is to find the first order correction to this reference energy. According to Eq. (2.51), the first order correction to the energy equals to the mean value of the perturbation with the exact wavefunction of the unperturbed system, and according to Eq. (2.43), we have the perturbation (let us set the perturbation parameter λ to 1)

$$\mathbf{H}^{(1)} = \mathbf{H} - \mathbf{H}^{(0)} = \sum_{i'<j'} \mathbf{g}\left(i'j'\right) - \sum_{i,j} \left[\mathbf{J}_j\left(i\right) - \mathbf{K}_j\left(i\right)\right]. \tag{2.96}$$

Notice that we exclude the nuclear repulsion operator $\mathbf{V}_{nn}$ from the exact Hamiltonian $\mathbf{H}$ here for the sake of convenience as it only introduces a constant shift to the final energy. Considering that the ground state Hartree-Fock wavefunction $\psi_0^{(0)}$ is also the wavefunction of the unperturbed system, if we insert Eq. (2.96) into Eq. (2.51), we get the energy which includes the first order correction as below,

$$\begin{aligned} E_{MP1} &= E_{MP0} + E_0^{(1)} = E_0^{(0)} + E_0^{(1)} \\ &= \left\langle \psi_0^{(0)} \middle| \mathbf{H}^{(0)} \psi_0^{(0)} \right\rangle + \left\langle \psi_0^{(0)} \middle| \mathbf{H}^{(1)} \psi_0^{(0)} \right\rangle = \left\langle \psi_0^{(0)} \middle| \mathbf{H} \psi_0^{(0)} \right\rangle . \end{aligned} \tag{2.97}$$

Because $\psi_0^{(0)}$ is the ground state Hartree-Fock wavefunction and $\mathbf{H}$ is the electronic Hamiltonian operator, thus we have

$$E_{MP1} = \left\langle \psi_0^{(0)} \middle| \mathbf{H} \psi_0^{(0)} \right\rangle = E_{HF} \,, \tag{2.98}$$

which means that E_{MP1} equals to the Hartree-Fock energy as defined in Eq. (2.87). For the energy with up to the second order correction, the case is much more complicated. According to Eq. (2.58), the second order correction to the energy of the ground state can be expressed as below,

$$E_0^{(2)} = \sum_{s \neq 0} \frac{\left| \left\langle \psi_0^{(0)} \middle| \mathbf{H}^{(1)} \psi_s^{(0)} \right\rangle \right|^2}{E_0^{(0)} - E_s^{(0)}} , \tag{2.99}$$

in which the subscript *s* denotes excited states, and the unperturbed excited state wavefunctions $\psi_s^{(0)}$ are formed from *N* different spin orbitals which include both occupied spin orbitals (let us say spin orbitals *i*, *j*, *k*, *l*, …) and unoccupied virtual spin orbitals (let us say spin orbitals *a*, *b*, *c*, *d*, …). Considering the matrix elements $\left\langle \psi_0^{(0)} \middle| \mathbf{H}^{(1)} \psi_s^{(0)} \right\rangle$ in Eq. (2.99), if $\psi_s^{(0)}$ is the wavefunction of a singly excited state (let us say Φ_i^a and we use Φ_0 to denote the ground state wavefunction of the unperturbed system $\psi_0^{(0)}$ here), we have

$$\left\langle \Phi_0 \left| \mathbf{H}^{(1)} \Phi_i^a \right\rangle = \left\langle \Phi_0 \left| \left[\mathbf{H} - \sum_{m=1}^{N} \mathbf{F}(m) \right] \Phi_i^a \right\rangle = \left\langle \Phi_0 \left| \mathbf{H} \Phi_i^a \right\rangle - \left\langle \Phi_0 \left| \sum_{m=1}^{N} \mathbf{F}(m) \Phi_i^a \right\rangle \right. \right. \right.$$
$$= \left\langle \Phi_0 \left| \mathbf{H} \Phi_i^a \right\rangle - \left[\left(\sum_{k}^{occ} \varepsilon_k \right) - \varepsilon_i + \varepsilon_a \right] \left\langle \Phi_0 \left| \Phi_i^a \right\rangle \right., \tag{2.100}$$

in which ε_i and ε_a are the orbital energies of spin orbitals ψ_i and ψ_a, respectively. According to the *Brillouin theorem* [11], $\left\langle \Phi_0 \left| \mathbf{H} \Phi_i^a \right\rangle = 0$. Because spin orbitals (both occupied and unoccupied) are different eigenfunctions of the Fock operator, they are orthogonal to each other, thus terms composed of $\left\langle \Phi_0 \left| \Phi_i^a \right\rangle \right.$ are all equal to zero. In this way, $\left\langle \Phi_0 \left| \mathbf{H}^{(1)} \Phi_i^a \right\rangle = 0$, which means the matrix elements $\left\langle \psi_0^{(0)} \left| \mathbf{H}^{(1)} \psi_s^{(0)} \right\rangle \right.$ in Eq. (2.99) that contain singly excited determinants do not contribute to the second order correction to the energy $E_0^{(2)}$. Next, matrix elements $\left\langle \psi_0^{(0)} \left| \mathbf{H}^{(1)} \psi_s^{(0)} \right\rangle \right.$ which include triple or higher excitations also equal to zero, according to the *Slater-Condon rules*.[12] Therefore only double excitation terms contribute to the second order correction to the energy $E_0^{(2)}$. If we denote the wavefunction of a doubly excited state with Φ_{ij}^{ab} (and its corresponding energy with E_{ij}^{ab}), we have the second order correction to the energy

$$E_0^{(2)} = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\left| \left\langle \Phi_0 \left| \mathbf{H}^{(1)} \Phi_{ij}^{ab} \right\rangle \right|^2}{E_0^{(0)} - E_{ij}^{ab}}. \tag{2.101}$$

To form a doubly excited state, we need to move two electrons from the occupied spin orbitals to the unoccupied spin orbitals, and the denominator term in the above equation can be expressed as a difference in orbital energies as below,

$$E_0^{(0)} - E_{ij}^{ab} = \varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b. \tag{2.102}$$

Therefore we get the energy which includes up to the second order correction

$$E_{MP2} = E_{HF} + E_0^{(2)} = E_{HF} + \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\left| \left\langle \Phi_0 \left| \mathbf{H}^{(1)} \Phi_{ij}^{ab} \right\rangle \right|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}, \tag{2.103}$$

in which we have

$$\left\langle \Phi_0 \left| \mathbf{H}^{(1)} \Phi_{ij}^{ab} \right\rangle = \left\langle \psi_i(1) \psi_j(2) \left| \mathbf{g}(12) \right| \psi_a(1) \psi_b(2) \right\rangle \right.$$
$$- \left\langle \psi_i(1) \psi_j(2) \left| \mathbf{g}(12) \right| \psi_b(1) \psi_a(2) \right\rangle \tag{2.104}$$

according to the *Slater-Condon rules*. Further corrections can also be obtained in a similar approach.

The second order Møller–Plesset (MP2) perturbation theory is one of the basic methods for calculation of correlation energy, and it can recover most of the electron correlations that the HF method fails to describe. As mentioned above, the MP2 method is based on the assumption that the reference wavefunction, which is the HF wavefunction, is good enough as an approximation to the real wavefunction. Therefore the performance of the MP2 method in describing electron correlations depends on the quality of the HF wavefunction to some extent. The MP2 method is widely used because of its economy and efficiency of including electron correlations.

2.5 Coupled Cluster Methods

The coupled cluster methods [13] represent another approach of recovering electron correlations that the Hartree-Fock method misses. Similar to the Møller–Plesset perturbation theory, the coupled cluster methods also start from the Hartree-Fock solution to the electronic Schrödinger equation, and it uses a multi-determinant trial wavefunction in the following form,

$$\Psi_{CC} = e^{\mathbf{T}}\Phi_0 , \tag{2.105}$$

in which Φ_0 is the ground state Hartree-Fock wavefunction. The exponential function of the cluster operator $\mathbf{T}$ can be expressed as a Taylor expansion as below,

$$e^{\mathbf{T}} = 1 + \mathbf{T} + \frac{\mathbf{T}^2}{2!} + \frac{\mathbf{T}^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{\mathbf{T}^k}{k!}, \tag{2.106}$$

in which the cluster operator $\mathbf{T}$ is given by

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_n , \tag{2.107}$$

and the $\mathbf{T}_i$ operator is defined to generate all the *i*-fold excited Slater determinants from the reference wavefunction which is normally the ground state Hartree-Fock wavefunction. For example, the one-electron excitation operator $\mathbf{T}_1$ and the two-electron excitation operator $\mathbf{T}_2$ are defined as below,

$$\begin{aligned} \mathbf{T}_1\Phi_0 &= \sum_{i}^{occ}\sum_{a}^{vir} t_i^a \Phi_i^a , \\ \mathbf{T}_2\Phi_0 &= \sum_{i<j}^{occ}\sum_{a<b}^{vir} t_{ij}^{ab} \Phi_{ij}^{ab} , \end{aligned} \tag{2.108}$$

in which the amplitudes t are similar to the coefficients a_i in Eq. (2.91), and

similar definitions hold for other $\mathbf{T}_i$ excitation operators. The aim of using such a wavefunction, as shown in Eq. (2.105), is that through mixing Slater determinants, which are constructed with electrons excited from occupied spin orbitals to unoccupied virtual spin orbitals, into the reference wavefunction, electrons can be kept away from each other so that we can recover electron correlation in some sense. Inserting Eq. (2.105) into the electronic Schrödinger equation, we have

$$\mathbf{H}e^{\mathbf{T}}\Phi_0 = Ee^{\mathbf{T}}\Phi_0 . \tag{2.109}$$

Projecting this equation against $\langle\Phi_0|$ and considering the orthonormality of the wavefunctions we are dealing with here, which include both the ground state Hartree-Fock wavefunction as well as the wavefunctions with excited determinants, we have

$$\begin{aligned}
&\left\langle \Phi_0 \middle| \mathbf{H}e^{\mathbf{T}}\Phi_0 \right\rangle = \left\langle \Phi_0 \middle| E_{CC}e^{\mathbf{T}}\Phi_0 \right\rangle = E_{CC}\left\langle \Phi_0 \middle| e^{\mathbf{T}}\Phi_0 \right\rangle = E_{CC}\left\langle \Phi_0 \middle| \left(1+\sum_{k=1}^{\infty}\frac{\mathbf{T}^k}{k!}\right)\Phi_0 \right\rangle \\
&= E_{CC}\left(\left\langle \Phi_0 \middle| \Phi_0 \right\rangle + \sum_{k=1}^{\infty}\frac{1}{k!}\left\langle \Phi_0 \middle| \mathbf{T}^k\Phi_0 \right\rangle \right) = E_{CC}\left(1+0\right) = E_{CC}
\end{aligned} \tag{2.110}$$

Thus the energy with a coupled cluster wavefunction can be expressed as

$$E_{CC} = \left\langle \Phi_0 \middle| \mathbf{H}e^{\mathbf{T}}\Phi_0 \right\rangle = \left\langle \Phi_0 \middle| \mathbf{H}\sum_{k=0}^{\infty}\frac{\mathbf{T}^k}{k!}\Phi_0 \right\rangle . \tag{2.111}$$

According to Eq. (2.63), the electronic Hamiltonian is a two-electron operator, which means terms in Eq. (2.111) that involve excited determinants with more than two electrons being excited from the occupied spin orbitals of Φ_0 vanish as a result of the overlap integral appearing over the orthogonal Φ_0 and such excited Slater determinants. In this way, we have

$$\begin{aligned}
E_{CC} &= \left\langle \Phi_0 \middle| \mathbf{H}\sum_{k=0}^{\infty}\frac{\mathbf{T}^k}{k!}\Phi_0 \right\rangle = \left\langle \Phi_0 \middle| \mathbf{H}\left(1+\mathbf{T}_1+\mathbf{T}_2+\frac{1}{2}\mathbf{T}_1^2\right)\Phi_0 \right\rangle \\
&= \left\langle \Phi_0 \middle| \mathbf{H}\Phi_0 \right\rangle + \left\langle \Phi_0 \middle| \mathbf{H}\mathbf{T}_1\Phi_0 \right\rangle + \left\langle \Phi_0 \middle| \mathbf{H}\mathbf{T}_2\Phi_0 \right\rangle + \frac{1}{2}\left\langle \Phi_0 \middle| \mathbf{H}\mathbf{T}_1^2\Phi_0 \right\rangle \\
&= E_{HF} + \sum_{i}^{occ}\sum_{a}^{vir} t_i^a \left\langle \Phi_0 \middle| \mathbf{H}\Phi_i^a \right\rangle + \sum_{i<j}^{occ}\sum_{a<b}^{vir}\left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a\right)\left\langle \Phi_0 \middle| \mathbf{H}\Phi_{ij}^{ab} \right\rangle .
\end{aligned} \tag{2.112}$$

Considering the *Brillouin theorem*, we get the final expression of the energy with a coupled cluster wavefunction as below,

$$E_{CC} = E_{HF} + \sum_{i<j}^{occ}\sum_{a<b}^{vir}\left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a\right)\left\langle \Phi_0 \middle| \mathbf{H}\Phi_{ij}^{ab} \right\rangle , \tag{2.113}$$

in which

$$\left\langle \Phi_0 \left| \mathbf{H}\Phi_{ij}^{ab} \right\rangle = \left\langle \psi_i(1)\psi_j(2) \left| \mathbf{g}(12) \right| \psi_a(1)\psi_b(2) \right\rangle - \left\langle \psi_i(1)\psi_j(2) \left| \mathbf{g}(12) \right| \psi_b(1)\psi_a(2) \right\rangle \tag{2.114}$$

The remaining problem is to find the amplitudes t in Eq. (2.113), and then we can calculate the energy. This is usually done with some further approximations and then projecting Eq. (2.109) against the complex conjugates of some excited determinants. One approximation we can make before we move further is that we can truncate the cluster operator $\mathbf{T}$ through including only the first few $\mathbf{T}_i$ excitation operators in Eq. (2.107). The coupled cluster singles and doubles (CCSD) method [13] is such an approach which approximates the cluster operator $\mathbf{T}$ with $\mathbf{T} \approx \mathbf{T}_1 + \mathbf{T}_2$. In this situation, Eq. (2.106) can be rewritten as

$$e^{\mathbf{T}_1+\mathbf{T}_2} = 1 + \mathbf{T}_1 + \left(\mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2 \right) + \left(\mathbf{T}_2\mathbf{T}_1 + \frac{1}{6}\mathbf{T}_1^3 \right) + \left(\frac{1}{2}\mathbf{T}_2^2 + \frac{1}{2}\mathbf{T}_2\mathbf{T}_1^2 + \frac{1}{24}\mathbf{T}_1^4 \right) + \cdots \tag{2.115}$$

If we insert Eq. (2.115) into Eq. (2.109) and then project Eq. (2.109) against $\left\langle \Phi_0 \right|$, we get exactly Eq. (2.111). However, if we insert Eq. (2.115) into Eq. (2.109) and then project Eq. (2.109) against $\left\langle \Phi_m^e \right|$, similar to Eqs. (2.110) and (2.111), we get

$$\left\langle \Phi_m^e \left| \mathbf{H}\Phi_0 \right\rangle + \sum_{ia} t_i^a \left\langle \Phi_m^e \left| \mathbf{H}\Phi_i^a \right\rangle + \sum_{ijab} \left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a \right) \left\langle \Phi_m^e \left| \mathbf{H}\Phi_{ij}^{ab} \right\rangle \right. \right. + \sum_{ijkabc} \left(t_{ij}^{ab} t_k^c + \cdots + t_i^a t_j^b t_k^c + \cdots \right) \left\langle \Phi_m^e \left| \mathbf{H}\Phi_{ijk}^{abc} \right\rangle \right. = E_{CCSD} t_m^e \tag{2.116}$$

in which $\left\langle \Phi_m^e \left| \mathbf{H}\Phi_0 \right\rangle = \left\langle \Phi_0 \left| \mathbf{H}\Phi_m^e \right\rangle^* = 0 \right.$, according to the *Brillouin theorem*. If we insert Eq. (2.115) into Eq. (2.109) and then project Eq. (2.109) against $\left\langle \Phi_{mn}^{ef} \right|$, we get

$$\begin{aligned} &\left\langle \Phi_{mn}^{ef} \left| \mathbf{H}\Phi_0 \right\rangle + \sum_{ia} t_i^a \left\langle \Phi_{mn}^{ef} \left| \mathbf{H}\Phi_i^a \right\rangle + \sum_{ijab} \left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a \right) \left\langle \Phi_{mn}^{ef} \left| \mathbf{H}\Phi_{ij}^{ab} \right\rangle \right. \right. \\ &+ \sum_{ijkabc} \left(t_{ij}^{ab} t_k^c + \cdots + t_i^a t_j^b t_k^c + \cdots \right) \left\langle \Phi_{mn}^{ef} \left| \mathbf{H}\Phi_{ijk}^{abc} \right\rangle \right. \\ &+ \sum_{ijklabcd} \left(t_{ij}^{ab} t_{kl}^{cd} + \cdots + t_{ij}^{ab} t_k^c t_l^d + \cdots t_i^a t_j^b t_k^c t_l^d + \cdots \right) \left\langle \Phi_{mn}^{ef} \left| \mathbf{H}\Phi_{ijkl}^{abcd} \right\rangle \right. = E_{CCSD} \left(t_{mn}^{ef} + t_m^e t_n^f - t_m^f t_n^e \right) \end{aligned} \tag{2.117}$$

Notice that terms with products of the amplitudes t, which involve permutations of indices, are neglected from Eqs. (2.116) and (2.117). With Eqs. (2.116) and (2.117), we get a set of simultaneous nonlinear equations, in which the amplitudes

t for singly and doubly excited wavefunctions are unknown. This set of nonlinear equations can be solved through some iterative procedures, and finally we will obtain the amplitudes t. Inserting these amplitudes into Eq.(2.105), we will get to know the coupled cluster wavefunction, and with the coupled cluster wavefunction, we will be able to calculate the CCSD energy with Eqs. (2.111) and (2.113). This is the way how the CCSD method works.

The CCSD method only considers excitations up to doubles, while some other methods extend the consideration to triple or even higher orders of excitations. For example, the CCSD(T) method is such a method which uses the CCSD amplitudes for single and double excitations as well as amplitudes for triple excitations from perturbational theory to construct excited determinants which act as corrections to the single Slater determinant Ψ_e^{HF} (see Eq. (2.91)). Because of its accuracy, the CCSD(T) method is also dubbed "*the gold standard of quantum chemistry*". Analogously to the Møller–Plesset perturbation theory, the performance of coupled cluster methods will also be affected if the HF wavefunction provides poor zeroth order approximation to the real wavefunction.

2.6 Density Functional Theory

The density functional theory (DFT) [14, 15] is an approach which was designed to accelerate *ab initio* calculations and to improve the quality of theoretical predictions based on the HF method. DFT works in such a way that it expresses the electronic energy in terms of the electronic density, and it also provides a way to find the electronic density. Such a treatment reduces the calculations that DFT needs to do by a huge amount, in comparison with the configuration interaction or coupled cluster methods, which makes DFT to be the most popular electronic structure method considering its good balance between accuracy and speed.

2.6.1 The Basis of DFT

The idea of DFT is to calculate the total energy of the system without using the wavefunction but rather with the electronic density distribution, and this was supported by the *Hohenberg-Kohn theorem* which states that the ground state properties of a many-electron system are a unique functional of the electronic

density which depends on only three spatial coordinates [14]. The electronic density distribution of a system with N electrons is defined as below (taking electron *1* as an example),

$$\rho(\boldsymbol{r}_1) = N \sum_{\sigma_1=-\frac{1}{2},\frac{1}{2}} \int d\tau_2 d\tau_3 \cdots d\tau_N \left|\Psi(\boldsymbol{r}_1,\sigma_1,\boldsymbol{r}_2,\sigma_2,\cdots,\boldsymbol{r}_N,\sigma_N)\right|^2, \quad (2.118)$$

in which $\boldsymbol{r}$ includes only the spatial coordinates, τ includes both the spatial and spin coordinates, and the integration of $|\Psi|^2$ is done for all electrons except electron *1*. As the wavefunction $\Psi(\boldsymbol{r}_1,\sigma_1,\boldsymbol{r}_2,\sigma_2,\cdots,\boldsymbol{r}_N,\sigma_N)$ is antisymmetric, which makes $|\Psi|^2$ to be symmetric with respect to exchange of the coordinates of any two electrons, the electronic density distribution defined in Eq. (2.118) does not depend on which electron was considered as the example. The electronic density distribution is given for any point in a three-dimensional (3D) space, and it is a function of the position which is determined by only three spatial coordinates. Providing we know the ground state electronic density $\rho_0(\boldsymbol{r})$ of a many-electron system, we are then able to know the total number of electrons in the system through the integration of $\rho_0(\boldsymbol{r})$ over the whole 3D space, which gives N. Moreover, if we plot $\rho_0(\boldsymbol{r})$ in a 3D space which is big enough to include all nuclei inside, we are also able to know the position and charge of each nucleus through information on the position of every cusp and the slope around this cusp [16]. In this way, we know the composition of this many-electron system, from which we get to know how the electronic Hamiltonian of this system looks like. If we solve the electronic Schrödinger equation of this system, we can also obtain the ground state electronic wavefunction. Considering the fact that all ground state properties of a system can be described by the ground state wavefunction, we conclude that the ground state electronic density can be used as an alternative to the ground state electronic wavefunction for full descriptions of all ground state properties of a system, and this is how the Hohenberg-Kohn theorem is proved.

Compared with a wavefunction which contains $4N$ variables, the current problem becomes much simpler as we only have three variables in the electronic density $\rho_0(\boldsymbol{r})$. Providing we know an expression for the total energy as a functional of the electronic density, our next step would be to find out the ground state electronic density, and the second part of the Hohenberg-Kohn theorem facilitates this step through stating that the ground state electronic energy of a system can be

obtained by a variational approach and the electronic density minimizing the Hohenberg-Kohn functional is the exact ground state electronic density [14]. In the Hohenberg-Kohn scheme, the Hamiltonian of any multi-particle system in an external potential $v_{ext}(\boldsymbol{r})$ can be written as below (the nuclear repulsion operator $\mathbf{V}_{nn}$ is also taken out from the exact electronic Hamiltonian $\mathbf{H}$ here),

$$\mathbf{H} = -\frac{1}{2}\sum_i \nabla_i^2 + \sum_i v_{ext}(\boldsymbol{r}_i) + \sum_{i<j}\frac{1}{r_{ij}}, \tag{2.119}$$

and the corresponding expression of the energy functional is

$$E_{HK}[\rho] = \langle \Psi | \mathbf{H} \Psi \rangle = T[\rho] + E_{ee}[\rho] + \int v_{ext}(\boldsymbol{r})\rho(\boldsymbol{r}) d^3\boldsymbol{r}, \tag{2.120}$$

in which $T[\rho]$ includes the kinetic energies of all electrons, $E_{ee}[\rho]$ represents all electron-electron interactions, and the last term represents the interaction energy between the electrons and the external potential field.

Now we need to find an exact expression for the energy functional, and this was started from the work done by Kohn and Sham [15]. Supposing there is such an ideal *N*-electron system, in which the *N* electrons do not interact with each other, but they interact with another effective external potential instead. Such an effective external potential is to make sure that this ideal system has the same electronic density distribution as the real system in which electron repulsions are treated normally. In this way, if we find the electronic density of the ideal system, we also obtain the electronic density of the real system. Fortunately, the ideal system can be solved by the following one-electron equation,

$$\left(-\frac{1}{2}\nabla^2 + v_{eff}\right)\psi_i^{KS} = \varepsilon_i^{KS}\psi_i^{KS}, \tag{2.121}$$

in which v_{eff} is the effective external potential, and ψ_i^{KS} are Kohn-Sham spinorbitals. The Kohn-Sham spinorbitals can be used to calculate the electronic density according to the following equation,

$$\rho(\boldsymbol{r}) = \sum_i^{occ.} \sum_{\sigma_1 = -\frac{1}{2}, \frac{1}{2}} \left|\psi_i^{KS}(\boldsymbol{r}, \sigma_1)\right|^2. \tag{2.122}$$

Eq. (2.121) is similar to Eq. (2.86) which we deduced in the discussion about the Hartree-Fock method. To simplify the calculation of the energy, Kohn and Sham rewrote Eq. (2.120) as below,

$$E_{KS}[\rho] = \int v_{ext}(\boldsymbol{r})\rho(\boldsymbol{r}) d^3\boldsymbol{r} + T_s + \Delta T[\rho] + \frac{1}{2}\iint \frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{r_{12}} d\boldsymbol{r}_1 d\boldsymbol{r}_2 + \Delta E_{ee}[\rho] \tag{2.123}$$

in which we let

$$\Delta T[\rho] = T[\rho] - T_s[\rho], \tag{2.124}$$

and

$$\Delta E_{ee}[\rho] = E_{ee}[\rho] - \frac{1}{2}\iint \frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{r_{12}} d\boldsymbol{r}_1 d\boldsymbol{r}_2 . \tag{2.125}$$

The advantage of Eq. (2.123) over Eq. (2.120) is that only $\Delta T[\rho]$, which is the difference in the expectation value of the ground state electronic kinetic energy between the real system and our fictitious ideal system without electron repulsions (denoted as T_s), and $\Delta E_{ee}[\rho]$, which is the difference between electron repulsions in the real system and electron repulsions in a continuous distribution of charge with electronic density ρ, are unknown, while the other three terms on the right hand side of Eq. (2.123) can be easily evaluated from ρ. If we put together the two unknowns in Eq. (2.123) and denote them with a quantity $E_{xc}[\rho]$ as below,

$$E_{xc}[\rho] = \Delta T[\rho] + \Delta E_{ee}[\rho], \tag{2.126}$$

in which we call $E_{xc}[\rho]$ the exchange-correlation energy, we have

$$E_{KS}[\rho] = \int v_{ext}(\boldsymbol{r})\rho(\boldsymbol{r}) d^3\boldsymbol{r} + T_s + \frac{1}{2}\iint \frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{r_{12}} d\boldsymbol{r}_1 d\boldsymbol{r}_2 + E_{xc}[\rho]. \tag{2.127}$$

Because the nuclear-electron attraction is the only external potential operated on the electrons, we have

$$v_{ext}(\boldsymbol{r}_i) = -\sum_p \frac{z_p}{|\boldsymbol{r}_p - \boldsymbol{r}_i|} = -\sum_p \frac{z_p}{r_{pi}}, \tag{2.128}$$

and

$$\int v_{ext}(\boldsymbol{r})\rho(\boldsymbol{r}) d^3\boldsymbol{r} = -\sum_p z_p \int \frac{\rho(\boldsymbol{r})}{|\mathbf{r}_p - \mathbf{r}|} d^3\boldsymbol{r} . \tag{2.129}$$

For the kinetic energy of our fictitious ideal system without electron repulsions, according to the *Slater-Condon rules*, we have

$$T_s = -\frac{1}{2}\sum_i^{occ.} \left\langle \psi_i^{KS} \middle| \nabla^2 \psi_i^{KS} \right\rangle . \tag{2.130}$$

Therefore, the only unknown in Eq. (2.127) turns out to be the exchange-correlation energy $E_{xc}[\rho]$. The exchange-correlation energy determines the accuracy of the electronic energy, which is expressed through the electronic density, as shown in Eq. (2.127). With Eq. (2.127), once we know the expression for the exchange-correlation energy $E_{xc}[\rho]$ (either exact or

approximate), it is possible for us to evaluate the electronic energy. Supposing that we already know an expression for the exchange-correlation energy $E_{xc}[\rho]$, our next step is to minimize E_{KS}, which is expressed through Eq. (2.127), and this can be done in a variational approach. Similar to that of the Hartree-Fock method, our solutions, which are the Kohn-Sham spinorbitals ψ_i^{KS}, form an orthonormal set, and this condition has to be satisfied, thus we need to use the concept of Lagrange multipliers again to deal with the variation of $E_{KS}[\rho]$ with respect to the variation of the electronic density ρ. The minimization results in the Kohn-Sham spinorbitals which were defined in Eq. (2.121), and at the same time, we get an expression for the effective external potential, which can be expressed by the following equation,

$$v_{eff}(\boldsymbol{r}_1) = -\sum_p \frac{z_p}{r_{p1}} + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + v_{xc}(\boldsymbol{r}_1). \tag{2.131}$$

Inserting Eq. (2.131) into Eq. (2.121), we get an equation for the Kohn-Sham spinorbitals (taking electron *1* as an example) as below,

$$\left[-\frac{1}{2}\nabla_1^2 - \sum_p \frac{z_p}{r_{p1}} + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + v_{xc}(\boldsymbol{r}_1) \right] \psi_i^{KS}(1) = \varepsilon_i^{KS} \psi_i^{KS}(1), \tag{2.132}$$

in which v_{xc} is the exchange-correlation potential, and is defined as the variation of the exchange-correlation energy $E_{xc}[\rho]$ with respect to the electronic density ρ,

$$v_{xc}(\boldsymbol{r}) = \frac{\delta E_{xc}[\rho(\boldsymbol{r})]}{\delta \rho(\boldsymbol{r})}. \tag{2.133}$$

We call Eq. (2.132) the Kohn-Sham equation. As a one-electron equation, the Kohn-Sham equation is similar to the Hartree-Fock equation (see Eq. (2.86)), and it can also be solved through an iterative procedure. Starting from a proper initial guess for the spinorbitals (usually from a semi-empirical calculation), we can then calculate the initial value of the electronic density ρ, with which we can calculate the effective external potential v_{eff}, and then we solve the Kohn-Sham equation, from which we obtain new spinorbitals and the new electronic density ρ. The new electronic density is used to calculate the new electronic energy (see Eq. (2.127)). We repeat the above process until certain convergence criteria are satisfied. The electron density that gives the lowest electronic energy is the best approximation to the ground state electron density. This is how the electronic density is determined. Therefore the only problem which we need to solve before we get an evaluation of the electronic energy based on the Kohn-Sham scheme is that we need to find an expression for the exchange-correlation energy $E_{xc}[\rho]$. It

is the biggest challenge of the density functional theory. Currently, all approaches use approximated expressions for the exchange-correlation energy functional $E_{xc}[\rho]$. These approximated approaches can be sorted according to the "Jacob's Ladder to Chemical Accuracy", a concept which was proposed by Perdew et al. [17] and which will be introduced in the following section.

2.6.2 The Jacob's Ladder to Chemical Accuracy

From the previous sections, we can see that based on the Kohn-Sham scheme, the multi-electron problem was converted to an effective single electron problem. However, we still need to find an expression for the exchange-correlation functional $E_{xc}[\rho]$, and this will be introduced through a few examples which constitute different rungs of the Jacob's ladder to chemical accuracy.

2.6.3 The Local Density Approximation

The local density approximation (LDA) is the first approximated approach which puts the density functional theory on the table. Kohn and Sham first suggested using the exchange-correlation functional from the homogeneous electron gas (HEG) model to approximate that of a real system [15]. The HEG model is such a system that it is electrically neutral, in which the positive charges do not come from point charges of nuclei, but they are rather smeared out uniformly in the whole box. Based on this model, one solves the electronic Schrödinger equation for this system exactly, and one extracts information about the exchange-correlation energy $E_{xc}[\rho]$ values. These $E_{xc}[\rho]$ values are then used in LDA calculations for molecules and crystals. In this way, we have the so-called local density approximation, in which "local" means at any point in the box, the exchange-correlation potential is a function of the electronic density only at that point, and it is not a function of the derivatives of the electronic density, which means one does not know anything about the electronic density at another point at that moment, and accordingly one does not know whether electronic density at another point contributes to the exchange-correlation potential or not. For situations in which we deal separately with electron densities due to spin-α electrons and spin-β electrons, we also have the local spin density approximation (LSDA).

It is known that in Kohn-Sham DFT, the exchange-correlation functional E_{xc}

can be written as a sum of the exchange energy functional E_x and the correlation energy functional E_c,

$$E_{xc} = E_x + E_c . \tag{2.134}$$

In LSDA, one can show that for the exchange energy functional E_x^{LSDA}, we have [18]

$$E_x^{LSDA} = -\frac{3}{4}\left(\frac{6}{\pi}\right)^{1/3} \int \left[\left(\rho^{\alpha}\right)^{4/3} + \left(\rho^{\beta}\right)^{4/3} \right] d\mathbf{r} . \tag{2.135}$$

There are a few functionals for E_c, among which the one given by Vosko, Wilk and Nusair (VWN) [19] is very popular. Vosko, Wilk and Nusair proposed a new formula for the correlation functional based on their careful analysis of the random phase approximation (RPA) to the correlation energy, an approach which was used to construct highly non-local correlation functionals, and they used a correlation potential determined from accurate quantum Monte Carlo (QMC) calculations [20] for interpolation and parameter fitting, in which both the low-density and high-density limits of the HEG were also incorporated. In this way, they constructed their expression for the correlation functional, which we will not give here due to its complexity. Finally, we get an expression for the VWN exchange-correlation functional E_{xc} as below,

$$E_{xc}^{VWN} = E_x^{LSDA} + E_c^{VWN} . \tag{2.136}$$

Strictly speaking, LDA is the first implementation of DFT in electronic structure calculations. It is the lowest rung on the Jacob's ladder to chemical accuracy. LDA was used to study some metal systems when it was just developed, and because the exchange-correlation interaction is a short-range interaction in solid systems, LDA gained huge success, which led to the wide applications of DFT in early years. The success of LDA in metal systems obscured the Hartree-Fock method, because there is a dilemma for the Hartree-Fock method on how to deal with metal systems, in which the gap between HOMO and LUMO (which is the band gap between the conduction band and the valence band) is equal to zero. On the other hand, as we know, LDA is based on the HEG model, and it might be a good approximation if the electronic density varies very slowly with variation of position. However, for systems in which the electronic density varies fast with variation of position, the LDA is not appropriate any more and can not give good descriptions for the exchange-correlation interactions, thus a lot of generalized gradient approximations were developed.

2.6.4 The Generalized Gradient Approximations

As a very good upgrade of LDA, the generalized gradient approximations (GGA) use both the electronic density and its gradient to express the exchange-correlation energy functional,

$$E_{xc}^{GGA} = E_{xc}^{GGA}\left[\rho(\mathbf{r}), \nabla\rho(\mathbf{r})\right], \tag{2.137}$$

and it makes improvements over LDA by including the gradient of the electronic density to correct the variation of electronic density with variation of positions. Due to this advantage, the applications of GGA were even extended to the study of energies and structures of hydrogen-bonded system, and it was also widely used in other areas of chemistry.

The GGA exchange-correlation energy functional is usually divided into two parts as below,

$$E_{xc}^{GGA} = E_{x}^{GGA} + E_{c}^{GGA}, \tag{2.138}$$

in which E_x^{GGA} and E_c^{GGA} are the exchange and correlation functionals, respectively. Commonly used GGA exchange functionals include Perdew and Wang's 1986 functional [21], denoted as PW86, and Becke's 1988 functional [22], denoted as B88. For the correlation functional, one of the most popular is Lee, Yang and Parr's functional [23], denoted as LYP. While the above mentioned functionals are based on an empirical fitting procedure, there are also some other functionals which are based on fundamental properties of atoms and molecules, and these functionals do not have any empirical parameters, which makes them more universal than those empirical functionals. One of the best examples for this type of functionals is the Perdew, Burke and Ernzerhof's exchange and correlation functional [24], denoted as PBE, which was designed to retain correct features of LSDA, meanwhile, the gradient-corrected non-local features were also considered so that better descriptions of the exchange-correlation interactions could be acquired. The GGA functionals give much better agreement between theory and experiment than LDA. It is the second rung on the Jacob's ladder to chemical accuracy.

2.6.5 Recent Developments of DFT

The developments of LDA and GGA made DFT to be the most popular electronic structure method. However, as DFT was used to study more and more different

problems, the deficiencies of exchange-correlation functionals were also gradually disclosed. For example, most of the LDA and GGA functionals can not give satisfying results for systems in which the dispersion interactions are very important [25-28], and some of these functionals give physically incorrect descriptions for transition metal oxides [29-31]. To solve these problems, many new functionals [32, 33] were developed in the past few years. The meta GGA functionals (such as M06-L [34]) belong to this class. They depend on the kinetic energy density or $\nabla^2\rho$ besides ρ and $\nabla\rho$, and the kinetic energy density is defined as below,

$$\tau = \frac{1}{2}\sum_{i}^{occ.}\left|\nabla\psi_i\right|^2 . \tag{2.139}$$

The meta GGA functionals give much better descriptions of transition metals and noncovalent interactions than traditional GGA functionals, which only depend on ρ and $\nabla\rho$, and the meta GGA functionals constitute the third rung on the Jacob's ladder to chemical accuracy.

In addition to these pure GGA and meta GGA functionals, hybrid functionals which mix some of the Hartree-Fock exact exchange into pure GGA and meta GGA functionals were also found to have better performance in many situations because of the nonlocality of the exact exchange. For example, one of the most popular hybrid GGA functionals, the B3LYP functional [35, 23, 36, 37], is defined as below,

$$E_{xc}^{B3LYP} = \left(1-a_0-a_x\right)E_x^{LSDA} + a_0E_x^{HF} + a_xE_x^{B88} + \left(1-a_c\right)E_c^{VWN} + a_cE_c^{LYP}, \tag{2.140}$$

in which E_x^{HF} is the Hartree-Fock definition of the exchange functional, and the a parameters are fitted to give the best agreement with experimental molecular atomization energies of some particular training sets, which results in $a_0 = 0.20$, $a_x = 0.72$, and $a_c = 0.81$. Another example of hybrid functionals is M06-2X [38], which is a hybrid meta GGA functional, and it is given by the following expression,

$$E_{xc}^{M06\text{-}2X} = \frac{X}{100}E_x^{HF} + \left(1-\frac{X}{100}\right)E_x^{M06\text{-}L} + E_c^{M06\text{-}L}, \tag{2.141}$$

in which $E_x^{M06\text{-}L}$ and $E_c^{M06\text{-}L}$ are the M06-L exchange and correlation functionals, respectively. The parameter X was optimized from training sets with nonmetals, and it was determined to be equal to 54. Detailed definitions of terms involved in Eqs. (2.140) and (2.141) can be found in relevant bibliographies, which have been listed. These hybrid functionals give much better descriptions of

dispersion interactions which were poorly described by traditional DFT methods. However, because they need to calculate the exact Hartree-Fock exchange, they are more time-consuming compared with pure GGA or meta GGA functionals. The hybrid functionals constitute the fourth rung on the Jacob's ladder to chemical accuracy.

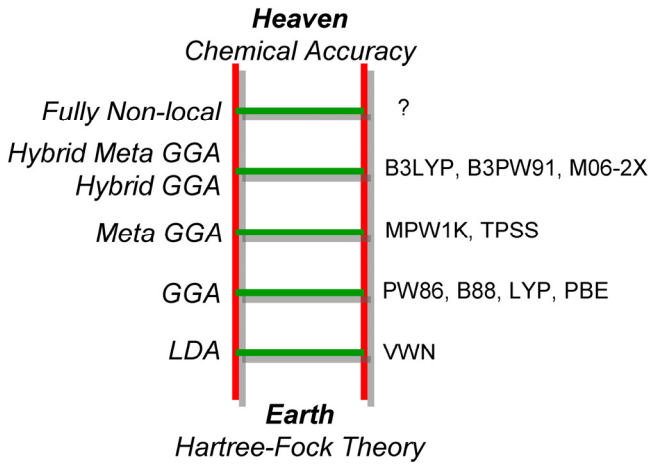


Figure 2.3: Jacob's ladder to chemical accuracy.

The most recent developments of DFT include DFT augmented with an empirical dispersion term [39-41] and double-hybrid DFT [42]. They give very good descriptions of dispersion interactions in weakly bound systems, and both of them have been implemented in the recently released quantum chemistry package Gaussian 09 [43]. One of the biggest problems of current functionals is that they are still local or partly local, which definitely is not the best solution, because from the point of view of physics, nonlocal interactions do spread all over of a many-body system, and intrinsically they can not be well described by these local or partly local functionals. The promising future of DFT relies on such universal functionals, which are fully non-local, so that they can be used for all systems. This is the last rung of the Jacob's ladder to chemical accuracy (see Figure 2.3), and hopefully this can be realized in the near future.

2.7 The Localized Basis Sets

In the previous sections, we talked about different methods which can be used to solve the Schrödinger equation, while these methods are based on such a

precondition that we already know how to express the spinorbitals, and in fact, we have not talked about this problem yet. Therefore the problem we are facing now is that we need to find a mathematical way to express the spinorbitals, and this is what we are going to talk about in this section. In detail, the specific solution to this problem how the spinorbitals are expanded depends on what system we are dealing with, e.g. molecular systems in the free space or extended systems being confined within the periodic boundary conditions (PBC). For the first situation, normally we use molecular spinorbitals which are expanded with so-called localized basis sets, such as the Pople basis sets, the correlation consistent basis sets, and some other split-valence basis sets. For the latter, normally we use crystal spinorbitals which are expanded with plane wave basis sets. In addition, pseudopotentials are also widely used in practical calculations especially for heavy metals to reduce the computational efforts and to include descriptions of relativistic effects. In the current section, we will focus on discussions about the localized basis sets.

To construct a molecular spinorbital, one frequently uses a set of known functions (basis functions, which are the so-called atomic orbitals (AOs)). This approach is dubbed the LCAO-MO method, as a molecular orbital (MO) ψ_i is represented as a linear combination of atomic orbitals (LCAO) χ_s,

$$\psi_i = \sum_{s=1}^{M} c_{si}\chi_s \,, \tag{2.142}$$

in which M is a finite number, c_{si} are the LCAO coefficients, and the atomic orbitals χ_s are normalized basis functions which are already known. The idea of the LCAO method is based on the fact that for a given atom, an electron in the vicinity of this atom should be described by an atomic orbital of this atom, and the LCAO method works in such a way that molecular orbitals are constructed as linear combinations of atomic orbitals which belong to different atoms. With the LCAO method, we are able to expand the unknown MOs with known AOs.

Normally there are two ways to express the atomic orbitals. The first way to define an atomic orbital is to use Slater type orbitals (STOs) which are introduced below,

$$\chi_{\alpha,n,l,m}(r,\theta,\varphi) = AY_{l,m}(\theta,\varphi)r^{n-1}e^{-\zeta r} \,, \tag{2.143}$$

in which ζ is a positive orbital exponent, n, l, and m are quantum numbers, A is a normalization constant, and $Y_{l,m}$ are spherical harmonic functions. The second

way to define atomic orbitals is to use Gaussian type orbitals (GTOs) which have an r^2 exponential dependence, and GTOs can be introduced in terms of polar coordinates,

$$\chi_{\alpha,n,l,m}(r,\theta,\varphi) = AY_{l,m}(\theta,\varphi)r^{(2n-2-l)}e^{-\zeta r^2}, \tag{2.144}$$

or Cartesian coordinates,

$$\chi_{\alpha,l_x,l_y,l_z}(x,y,z) = A\,x^{l_x}y^{l_y}z^{l_z}e^{-\zeta r^2}. \tag{2.145}$$

In the latter, the sum of the quantum numbers $l_{x,y,z}$ determines the type of orbital (*s*-orbital, *p*-orbital, *d*-orbital, …), and we call each single GTO a primitive GTO (PGTO). Both STOs and GTOs can be used to construct an MO, and an STO is superior to a GTO in a few aspects, though it is less popular in practical calculations. One advantage of an STO is that it decays in a similar way to the exact orbitals because of its exponential dependence. Compared with an STO, the r^2 exponential dependence of a GTO makes it problematic that it can not give a proper description for orbital near the nucleus, and also the tail of the wavefunction is usually poorly represented with a GTO. The problem of STOs is that the electron-electron repulsions (the Coulomb and exchange integrals) are difficult to calculate, while on the other hand, these calculations can be performed analytically with GTOs, which makes GTOs more efficient. Because of this advantage, GTOs are widely used as basis functions in almost all electronic structure calculations.

We have just talked about two different types of AOs which can be used to construct an MO, and we call these two different types of AOs basis functions. A basis set is defined as a set of basis functions which are used in the LCAO-MO expansion, and for a given atom, a minimal basis set includes one basis function for each AO which belongs to a fully or partially occupied n*l* shell, e.g. for a carbon atom, with the electronic configuration $1s^2 2s^2 2p^2$, one includes one basis function for the 1*s*, 2*s*, $2p_x$, $2p_y$, and $2p_z$ orbitals, respectively. An improvement of the minimum basis set is to use two basis functions for each occupied AO, which produces a double zeta (DZ) type basis set, or to use three basis functions for each occupied AO, which produces a triple zeta (TZ) type basis set, and so on. These extensions make it a better description of electron distributions possible which might be different in different directions. Additional improvements of the basis sets include additions of polarization functions (usually denoted with "*", operating on heavy atoms only, or "**", operating on both heavy atoms and hydrogen atoms), which introduce higher angular momentum functions into the

basis set and make the basis set more flexible with respect to different bonding environments, and diffuse functions (usually denoted with "+", operating on heavy atoms only, or "++", operating on both heavy atoms and hydrogen atoms), which are usually very flat GTOs and can better represent atomic orbitals in the region that is far from atomic nuclei.

From the point of view of chemistry, because it is the valence electrons which take part in the molecular bonding, it is natural to use more basis functions to represent the valence orbitals. In this way, the split valence basis set was introduced. The split valence double zeta (VDZ) type basis set works in such a way that it doubles basis functions only for valence electrons, and the split valence triple zeta (VTZ) type basis set triples basis functions only for valence electrons. In this way, more basis functions are used for valence electrons in the split valence basis sets, which increases the computational cost, though better descriptions of the chemical bonding can be achieved. The contracted basis sets were developed against this background. In a contracted GTO (CGTO), the PGTOs and their corresponding LCAO coefficients are held fixed during the calculation, and the CGTO is a linear combination of PGTOs, in which we call the fixed LCAO coefficients the contraction coefficients. In this way, the number of unknown LCAO coefficients which are needed to be determined is decreased, which reduces the computational cost quite a lot while the accuracy can still be maintained at a good level if the chosen contraction coefficients are good enough. Thus the next problem is how to choose a set of contraction coefficients which gives MOs as good as possible. Pople et al. [44] developed sophisticated procedures to determine the contraction coefficients for almost all main group elements in the periodic table. For example, one of Pople's basis sets which is used in the current thesis for a glycine molecule ($C_2H_5O_2N$) is 6-31++G**, in which we use one CGTO that is a linear combination of six PGTOs to represent each core orbital, and for each valence orbital, we use two basis functions, including one single PGTO and one CGTO which is a linear combination of three PGTOs, while "++" indicates a set of *sp*-type diffuse functions (which is a total of 4 Cartesian basis functions) being added to heavy atoms together with a *s*-type diffuse function being added to hydrogen atoms, and "**" indicates a *d*-type polarization function (which is a total of 6 Cartesian basis functions) being added to heavy atoms together with a *p*-type polarization function (which is a total of 3 Cartesian basis functions) being added to hydrogen atoms. In this way, a total of 19 basis functions (one for the core orbital, two for each valence orbital, plus four

from diffuse functions and six from the polarization function) were used for each heavy atom in this system (C, O, and N), and a total of 6 basis functions (two for the valence orbital, plus one from the diffuse function and three from the polarization function) were used for each hydrogen atom in this system. The contraction coefficients of this and other Pople's basis sets can be found in the EMSL Basis Set Exchange database [45] and references therein.

In addition to Pople's basis sets, we also have the correlation consistent basis sets which were developed by Dunning et al. [46] aiming at achieving convergence of the calculated *ab initio* electronic energies to the complete basis set (CBS) limit with the assistance of certain extrapolation procedures. For example, one of the correlation consistent basis sets which is used in the current thesis is aug-cc-pVDZ (AVDZ), in which the term "pVDZ" means a polarized valence double zeta type basis set, "cc" means the basis set is designed in such a way to ensure correlation energy is recovered consistently with respect to increase of the cardinal number of the basis set (say, one jumps from DZ to TZ, and the cardinal number is increased by one, which results in a new type of higher-order polarization function being added), and "aug" means one employs additional diffuse functions, in which one extra function with a small exponent was added to each angular momentum. The composition of a series of correlation consistent basis sets which are used in the current thesis is listed in Table 2.1. With the assistance of recently developed extrapolation procedures [47], it is also possible to converge the calculated *ab initio* electronic energies to the CBS limit, and this will be introduced with more details in Chapter 6.

Basis set	Contracted functions	
	First row elements	Hydrogen atom
aug-cc-pVDZ	4s, 3p, 2d	3s, 2p
aug-cc-pVTZ	5s, 4p, 3d, 2f	4s, 3p, 2d
aug-cc-pVQZ	6s, 5p, 4d, 3f, 2g	5s, 4p, 3d, 2f

Table 2.1: Composition of some common correlation consistent basis sets.

As all these localized basis sets which we mentioned above are incomplete, problems appear when we calculate the interaction energy of a molecular complex. The interaction energy is usually calculated as a difference between the energy of complex AB and the sum of energies of monomers A and B as below,

$$\Delta E = E_{AB}\left(AB\right) - E_A\left(A\right) - E_B\left(B\right). \tag{2.146}$$

The performance of this approach depends on the quality of electronic structure method used to calculate all energies on the right-hand side of Eq. (2.146). However, one should recognize that within the complex, basis functions localized on one monomer can also affect the energy of another monomer. This is the so-called basis set superposition error (BSSE). Popular ways to deal with the BSSE include the counterpoise method [48] and the chemical Hamiltonian approach [49]. In addition to the BSSE which presents in intermolecular interactions, it has also been suggested recently that intramolecular BSSE can be as significant as intermolecular BSSE, and a few methods were proposed to correct for intramolecular BSSE [50, 51].

Now, as we have the basic building blocks of a molecular orbital, we need to determine the unknown LCAO coefficients c_{si}, which are included in Eq. (2.142), and this is usually done through an iterative procedure like that which was introduced in Section 2.3.2. From the discussions in the previous sections, we know the Hartree-Fock and Kohn-Sham equations are solved through an SCF approach, and if we insert the MO in the form of a basis set (see Eq. (2.142)) into the Hartree-Fock equation (see Eq. (2.86)) or the Kohn-Sham equation (see Eq. (2.132)), then we get

$$\mathbf{F}\sum_{s=1}^{M} c_{si}\chi_s = \varepsilon_i \sum_{s=1}^{M} c_{si}\chi_s \,. \tag{2.147}$$

Projecting Eq. (2.147) against $\langle\chi_r|$, we have

$$\sum_{s=1}^{M} c_{si}\left(F_{rs} - \varepsilon_i S_{rs}\right) = 0\,, \tag{2.148}$$

in which $r = 1, 2, ..., M$, $F_{rs} = \langle\chi_r|\mathbf{F}\chi_s\rangle$, and $S_{rs} = \langle\chi_r|\chi_s\rangle$. Because F_{rs} depends on ψ_i (either the Hartree-Fock or the Kohn-Sham spinorbital), while at the same time ψ_i depends on the unknown LCAO coefficients c_{si}, Eq. (2.148) can be solved through an iterative procedure like that of the SCF method as below:

(1) generate initial guess for the unknown LCAO coefficients c_{si} (usually through semi-empirical methods), with which we also get an initial guess for the occupied MOs;
(2) form the F_{rs} matrix with these MOs, solve Eq. (2.148), and produce a total of M new MOs, in which we choose the orbitals that have the lowest orbital energies to be occupied by *N* electrons, following the Pauli

Exclusion Principle;

(3) evaluate the energy E_e (see Eq. (2.87) for the Hartree-Fock energy or Eq. (2.127) for the Kohn-Sham energy) with these new occupied MOs;

(4) use these new occupied MOs to build a new F_{rs} matrix and repeat the above process until the difference in E_e's or expansion coefficients c_{si} between two adjacent iterations falls below a certain threshold which has been predetermined.

What should be pointed out here is that the above procedure produces typically not only occupied orbitals but also virtual orbitals, because M is typically much larger than the number of occupied orbitals. Using these virtual orbitals, one can construct additional excited Slater determinants that can be used to recover the electron correlation energy (see Eq. (2.92)) through an advanced electronic structure method, such as MP2 or CCSD.

2.8 Electronic Motion in Extended Systems

In previous sections, we talked about the electronic Schrödinger equation and its solutions for molecular systems. While the molecular systems represent one type of systems which can be well described by electronic structure methods, the extended systems, which we need to deal with in our everyday life as well, represent another type that could be well described by electronic structure methods, such as the density functional theory. The extended systems are different from the molecular systems by the fact that they are characterized by the periodic potential, which makes approaches towards solving the Schrödinger equation under this condition more complicated. In this section, we will talk about the computational strategies used in extended systems through the process how we solve the Schrödinger equation in a periodic potential.

2.8.1 Periodicity and Bloch's Theorem

For bulk crystals, slabs, and polymers, one frequently uses infinite models with periodic boundary conditions. In this way, density functional theory methods for such systems were often implemented in a form which is also subject to the periodic boundary conditions [52, 53]. For the sake of convenience, here we take the crystal as an example for our discussions. The periodicity of a crystal with

periodic boundary conditions is well described by the lattice vectors $\mathbf{R}$ in real space,

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \tag{2.149}$$

in which $n_{1,2,3}$ are integers, and $\mathbf{a}_{1,2,3}$ are the primitive lattice vectors describing a primitive unit cell. The primitive unit cell of a crystal is often defined in such a way to make sure that the crystal can be translationally repeated with the atoms inside this primitive unit cell, and at the same time it has the least volume. To describe behaviors of electrons in a crystal, one of the most popular ways is to use the reciprocal lattice. The reciprocal lattice vectors $\mathbf{G}$ are defined in the reciprocal space, which is the inverse of real space, and $\mathbf{G}$ are defined like the lattice vectors $\mathbf{R}$ in real space as below,

$$\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3, \tag{2.150}$$

in which $m_{1,2,3}$ are integers, and the reciprocal primitive lattice vectors $\mathbf{b}_{1,2,3}$ are defined by

$$\begin{aligned}
\mathbf{b}_1 &= 2\pi\frac{\mathbf{a}_2\times\mathbf{a}_3}{\mathbf{a}_1\bullet\left[\mathbf{a}_2\times\mathbf{a}_3\right]} \\
\mathbf{b}_2 &= 2\pi\frac{\mathbf{a}_3\times\mathbf{a}_1}{\mathbf{a}_1\bullet\left[\mathbf{a}_2\times\mathbf{a}_3\right]}, \\
\mathbf{b}_3 &= 2\pi\frac{\mathbf{a}_1\times\mathbf{a}_2}{\mathbf{a}_1\bullet\left[\mathbf{a}_2\times\mathbf{a}_3\right]}
\end{aligned} \tag{2.151}$$

and it can be readily proven that

$$\mathbf{a}_i\bullet\mathbf{b}_j = 2\pi\delta_{ij}, \tag{2.152}$$

in which $i, j = 1, 2, 3$. A natural result from the definition of reciprocal lattice vectors is

$$\mathbf{R}\bullet\mathbf{G} = 2\pi l, \tag{2.153}$$

in which $l = n_1m_1 + n_2m_2 + n_3m_3$ is always an integer, and it follows that

$$e^{i\mathbf{G}\bullet\mathbf{R}} = 1. \tag{2.154}$$

To include the periodicity of the lattice into the electronic wavefunction, one often introduces Bloch's theorem, which states that the spinorbitals for extended systems have the following property,

$$\psi_{\mathbf{k}}\left(\mathbf{r}-\mathbf{R}\right) = e^{-i\mathbf{k}\bullet\mathbf{R}}\psi_{\mathbf{k}}\left(\mathbf{r}\right), \tag{2.155}$$

in which $\mathbf{r}$ is the electronic position vector defined in real space, and subscript $\mathbf{k}$ characterizes different spinorbitals. According to Bloch's theorem, the values of a spinorbital at two different positions differing by a lattice vector $\mathbf{R}$ are related through a phase factor $e^{i\mathbf{k}\bullet\mathbf{R}}$.

Bloch's theorem can be proved as below. First, let us introduce the translation operator $T_{\mathbf{R}}$, which is defined below,

$$T_{\mathbf{R}} f(\mathbf{r}) = f(\mathbf{r}-\mathbf{R}). \tag{2.156}$$

It can be proved that the translation operator $T_{\mathbf{R}}$ commutes with the electronic Hamiltonian (see Eq. (2.132)), which means that $T_{\mathbf{R}}$ and the electronic Hamiltonian share the same set of eigenfunctions. In this way, we can choose $\psi_{\mathbf{k}}(\mathbf{r})$, which is the eigenfunction of the electronic Hamiltonian, to be an eigenfunction of the translation operator $T_{\mathbf{R}}$ as well,

$$T_{\mathbf{R}}\psi_{\mathbf{k}}(\mathbf{r}) = c_{\mathbf{R}}\psi_{\mathbf{k}}(\mathbf{r}), \tag{2.157}$$

in which $c_{\mathbf{R}}$ is the eigenvalue corresponding to the eigenfunction $\psi_{\mathbf{k}}(\mathbf{r})$. Notice that

$$T_{\mathbf{R}}T_{\mathbf{R}'} = T_{\mathbf{R}'}T_{\mathbf{R}} = T_{\mathbf{R}+\mathbf{R}'}, \tag{2.158}$$

thus we have

$$c_{\mathbf{R}}c_{\mathbf{R}'} = c_{\mathbf{R}+\mathbf{R}'}. \tag{2.159}$$

Because $c_{\mathbf{R}}$ depends on $\mathbf{R}$, to satisfy Eq. (2.159), $c_{\mathbf{R}}$ must be an exponential in $\mathbf{R}$, and considering that $\mathbf{a}_i \bullet \mathbf{b}_j = 2\pi\delta_{ij}$ (see Eq. (2.152)), we can express $c_{\mathbf{R}}$ as below,

$$c_{\mathbf{R}} = e^{-i\mathbf{k}\bullet\mathbf{R}}, \tag{2.160}$$

where we define $\mathbf{k}$ according to the following equation

$$\mathbf{k} = \kappa_1\mathbf{b}_1 + \kappa_2\mathbf{b}_2 + \kappa_3\mathbf{b}_3, \tag{2.161}$$

in which $\kappa_{1,2,3}$ are real parameters. Now we can see the subscript $\mathbf{k}$ in the spinorbitals $\psi_{\mathbf{k}}(\mathbf{r})$ is a vector defined in the reciprocal space. Inserting Eqs. (2.156) and (2.160) into Eq. (2.157), we get

$$T_{\mathbf{R}}\psi_{\mathbf{k}}(\mathbf{r}) = \psi_{\mathbf{k}}(\mathbf{r}-\mathbf{R}) = e^{-i\mathbf{k}\bullet\mathbf{R}}\psi_{\mathbf{k}}(\mathbf{r}), \tag{2.162}$$

which proves Bloch's theorem that we have just introduced in Eq. (2.155). Considering $e^{-i\mathbf{G}\bullet\mathbf{R}} = 1$ (see Eq. (2.154)), we have

$$T_{\mathbf{R}}e^{i(\mathbf{k}+\mathbf{G})\bullet\mathbf{r}} = e^{i(\mathbf{k}+\mathbf{G})\bullet(\mathbf{r}-\mathbf{R})} = e^{-i\mathbf{k}\bullet\mathbf{R}}e^{i(\mathbf{k}+\mathbf{G})\bullet\mathbf{r}} = c_{\mathbf{R}}e^{i(\mathbf{k}+\mathbf{G})\bullet\mathbf{r}}, \tag{2.163}$$

in which $e^{i(\mathbf{k}+\mathbf{G})\bullet\mathbf{r}}$ are eigenfunctions corresponding to the same eigenvalue $e^{-i\mathbf{k}\bullet\mathbf{R}}$. In this way, the spinorbitals $\psi_{\mathbf{k}}(\mathbf{r})$, which are the eigenfunctions of the electronic Hamiltonian, can be expanded through all eigenfunctions of the translation operator $T_{\mathbf{R}}$ that correspond to the same eigenvalue as below,

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} \alpha_{\mathbf{k}}(\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\bullet\mathbf{r}}. \tag{2.164}$$

Spinorbitals for extended systems can be expressed in several different ways, and one of them is to write the spinorbitals in a form as below,

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\,\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \tag{2.165}$$

in which $u_{\mathbf{k}}(\mathbf{r})$ are functions pertaining the translational symmetry of the lattice, and $u_{\mathbf{k}}(\mathbf{r}-\mathbf{R}) = u_{\mathbf{k}}(\mathbf{r})$. Comparing Eqs. (2.164) and (2.165), we can see $u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} \alpha_{\mathbf{k}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}$, where $\alpha_{\mathbf{k}}(\mathbf{G})$ are linear expansion coefficients, and $e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}$ are plane waves. This is the way how spinorbitals for extended system are expanded with plane waves in the formulation of Bloch's theorem.

Spinorbitals for extended systems can also be expanded using localized atomic orbitals in the form of Bloch functions. A Bloch function for a localized atomic orbital is defined by the following equation,

$$\psi_{\mathbf{k},\alpha}(\mathbf{r}) = \sum_{j} e^{i\mathbf{k}\cdot\mathbf{R}_j} \chi_{\alpha}(\mathbf{r}-\mathbf{R}_j), \tag{2.166}$$

in which χ_{α} is a localized atomic orbital labeled by α, $\mathbf{R}_j$ is a lattice vector, and the summation j extends over the whole lattice through $\mathbf{R}_j$. In this way, the spinorbital can be expressed as a linear combination of Bloch functions as below,

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\alpha} c_{\alpha}(\mathbf{k}) \psi_{\mathbf{k},\alpha}(\mathbf{r}), \tag{2.167}$$

in which $c_{\alpha}(\mathbf{k})$ are the linear expansion coefficients. It can be proved that Bloch functions are eigenfunctions of any translation operators $T_{\mathbf{R}}$, and spinorbitals which are expanded with Bloch functions also follow Bloch's theorem (see Eq. (2.155)).

2.8.2 The Brillouin Zone Sampling

According to Bloch's theorem, each single crystal orbital is labeled by a wave vector $\mathbf{k}$ defined in the reciprocal space (see Eq. (2.161)). In extended systems, the values of $\mathbf{k}$ appear to be continuous, and it makes the eigenvalues, which are the orbital energies $\varepsilon_i(\mathbf{k})$, to become continuous bands as well, not like that in molecular systems, in which the energy levels are well separated from each other. As we have pointed out in the last section, the wave vector $\mathbf{k}$ is defined in the reciprocal space and $u_{\mathbf{k}}(\mathbf{r})$ has the periodicity of the lattice, thus the Kohn-Sham equation for extended systems (say we use the DFT method to solve the Schrödinger equation for an extended system) can be solved in one primitive cell of the periodic lattice in reciprocal space, and the First Brillouin Zone (FBZ) has been proved to be a good choice for such a primitive cell in reciprocal space.

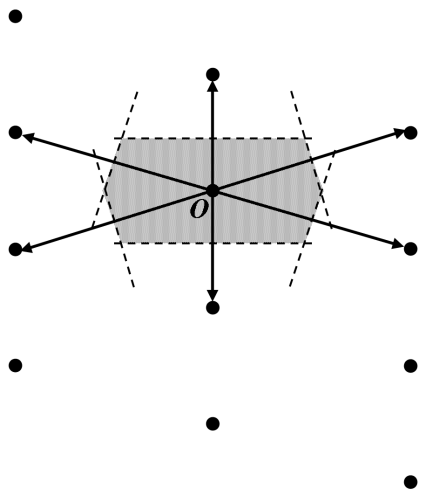


Figure 2.4: Construction of the first Brillouin zone.

A very popular method to construct the First Brillouin Zone is to construct it as a Wigner-Seitz unit cell of the reciprocal lattice, in which we draw perpendicular bisectors between the origin node ***O*** and its nearest and next nearest nodes, and then the First Brillouin Zone is the smallest volume entirely enclosed. An example for construction of the First Brillouin Zone for a 2D lattice is illustrated in Figure 2.4, in which the shadow area indicates the FBZ of this 2D lattice in reciprocal space.

As we have just mentioned above, we only need to solve the Kohn-Sham equation for all wave vectors $\mathbf{k}$ in the FBZ, and then we will be able to know solutions of the Kohn-Sham equation for wave vectors $\mathbf{k}$ outside the FBZ as well. Assume there are two wave vectors $\mathbf{k}'$ and $\mathbf{k}''$, and they are related to each other through the equality $\mathbf{k}''=\mathbf{k}'+\mathbf{G}$, in which $\mathbf{G}$ is a reciprocal lattice vector (see Eq. (150). Considering the way in which the FBZ is constructed, if $\mathbf{k}'$ is a vector inside the FBZ, $\mathbf{k}''$ must be a vector located outside the FBZ. Take a Bloch function as an example, for a Bloch function labeled with $\mathbf{k}''$, we have

$$\begin{aligned}\psi_{\mathbf{k}'',\alpha}\left(\mathbf{r}\right)&=\sum_j e^{i\mathbf{k}''\cdot\mathbf{R}_j}\chi_\alpha\left(\mathbf{r}-\mathbf{R}_j\right)=\sum_j e^{i(\mathbf{k}'+\mathbf{G})\cdot\mathbf{R}_j}\chi_\alpha\left(\mathbf{r}-\mathbf{R}_j\right)\\&=\sum_j e^{i\mathbf{k}'\cdot\mathbf{R}_j}e^{i\mathbf{G}\cdot\mathbf{R}_j}\chi_\alpha\left(\mathbf{r}-\mathbf{R}_j\right)\\&=\sum_j e^{i\mathbf{k}'\cdot\mathbf{R}_j}\chi_\alpha\left(\mathbf{r}-\mathbf{R}_j\right)=\psi_{\mathbf{k}',\alpha}\left(\mathbf{r}\right)\end{aligned}\quad , \qquad (2.168)$$

in which $e^{i\mathbf{G}\cdot\mathbf{R}_j}=1$ (see Eq. (2.154)). Equation (2.168) indicates that Bloch

functions labeled by $\mathbf{k}'$ and $\mathbf{k}''$ are actually identical, which means $\mathbf{k}'$ and $\mathbf{k}''$ are equivalent to each other, and $\mathbf{k}$ values which are enclosed in the FBZ are sufficient to describe the electronic structure of the whole lattice.

The introduction of the FBZ reduces the computational effort quite a lot, while the computational effort can be further reduced if we take advantage of other symmetries beyond the translational periodicity which are very common in many crystals, such as rotations around axes, reflections in planes, and their combinations. This leads to the concept of the Irreducible Brillouin Zone (IBZ), which is usually a very small fraction of the FBZ, and the Kohn-Sham equation is solved for all wave vectors $\mathbf{k}$ in the small volume of IBZ rather than the whole volume of FBZ. Once the Kohn-Sham equation is solved in the IBZ, then solutions of the Kohn-Sham equation in other parts of the FBZ can also be obtained based on the symmetries which the FBZ holds. An example of the Irreducible Brillouin Zone for a 2D square lattice with full symmetry is illustrated in Figure 2.5, in which the shadow area indicates the IBZ of this 2D square lattice in reciprocal space.

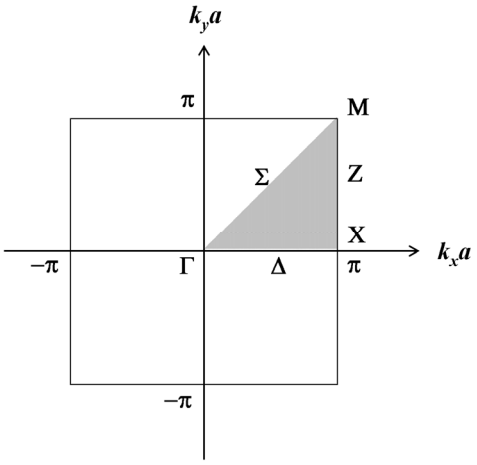


Figure 2.5: The Irreducible Brillouin Zone for 2D square lattice with full symmetry.

According to previous discussions in the DFT section, the energy of a system which is described by the DFT method can be represented as a summation of energies contributed from different terms (see Eq. (2.127)), which include the nuclear-electron attraction V_{ne} (see Eq. (2.129)), the kinetic energy of electrons T_s (see Eq. (2.130)), the Coulomb repulsion V_{ee} and the exchange-correlation

interaction E_{xc} (see Eq. (2.126)). In this way, the total energy of an extended system can be expressed as below,

$$E_{KS}[\rho] = V_{ne} + T_s + V_{ee} + E_{xc}[\rho]. \tag{2.169}$$

Now we need to rewrite these terms in the formulation of an extended system with periodic potential, let us say in a unit cell of this extended system which is represented by an IBZ of this unit cell in the reciprocal space. For the kinetic energy of electrons, it can be written as below,

$$T_s = -\frac{1}{2}\sum_{i,\mathbf{k}}^{occ.}\left\langle \psi_{i,\mathbf{k}} \middle| \nabla^2 \psi_{i,\mathbf{k}} \right\rangle. \tag{2.170}$$

Notice here we introduce another subscript i, which indicates an electronic band, and the summation is done for all occupied electronic bands. For the nuclear-electron attraction, it can be expressed with the following equation,

$$V_{ne} = \sum_{i,\mathbf{k}}^{occ.}\left\langle \psi_{i,\mathbf{k}} \middle| \mathbf{V}_{ext}\psi_{i,\mathbf{k}} \right\rangle, \tag{2.171}$$

in which $\mathbf{V}_{ext}$ represents the external potential from the nuclei. For the Coulomb repulsion, it can be written as below,

$$V_{ee} = \frac{1}{2}\sum_{i,\mathbf{k}\neq\mathbf{k}'}^{occ.}\left\langle \psi_{i,\mathbf{k}}\psi_{i,\mathbf{k}'} \middle| \frac{1}{|\mathbf{r}-\mathbf{r}'|}\psi_{i,\mathbf{k}}\psi_{i,\mathbf{k}'} \right\rangle. \tag{2.172}$$

For the exchange-correlation interaction, it is a functional of the electron density, but the expression of the electron density for an extended system is different from that of a molecular system, and it can be written as below,

$$\rho(\boldsymbol{r}) = \sum_{i,\boldsymbol{k}}^{occ.}\left|\psi_{i,\boldsymbol{k}}(\boldsymbol{r})\right|^2. \tag{2.173}$$

In a practical calculation of the total energy in Eq. (2.169), one often treats valence electrons and core electrons separately, in which contributions involving core electrons are usually approximated with pseudopotentials, and we will give more information about them in the following section.

It should be noticed that in Eqs. (2.170-2.173), all of them are represented as summations of all $\mathbf{k}$ vectors in the IBZ over all occupied states. Though the IBZ can be a very small volume if the lattice is highly symmetrical, the number of possible $\mathbf{k}$ vectors in the IBZ is still huge because the number of unit cells in a crystal is very large. In this way, summations of all $\mathbf{k}$ vectors in the IBZ for the above terms seem to be impracticable. Take the kinetic energy of electrons T_s (see Eq. (2.170)) as an example. For a very small volume $\Delta(\mathbf{k}^3)$ in the IBZ, contribution to T_s from this small volume can be written as below,

$$T_s\Delta\left(\mathbf{k}^3\right)=-\frac{1}{2}\sum_{i}^{occ.}\sum_{\mathbf{k}}^{\Delta}\left[\left\langle\psi_{i,\mathbf{k}}\left|\nabla^2\psi_{i,\mathbf{k}}\right\rangle\right]. \tag{2.174}$$

Eq. (2.174) looks like that the IBZ is cut into many tiny pieces, and the total number of these tiny pieces equals to the volume of the IBZ (which is 1 here since T_s is expressed as per IBZ) divided by $\Delta\left(\mathbf{k}\right)$, which will be a huge number as well, like the number of possible $\mathbf{k}$ vectors in the IBZ. A good approximation to Eq. (2.174) would be that

$$T_s\Delta\left(\mathbf{k}^3\right)=-\frac{1}{2}\sum_{i}^{occ.}\int_{\mathbf{k}}^{\Delta}\left\langle\psi_{i,\mathbf{k}}\left|\nabla^2\psi_{i,\mathbf{k}}\right\rangle d^3\mathbf{k}\right., \tag{2.175}$$

and T_s can be written as a numerical integration of $\mathbf{k}$ vectors over the whole IBZ,

$$T_s=-\frac{1}{2}\sum_{i}^{occ.}\int_{\mathbf{k}}^{IBZ}\left\langle\psi_{i,\mathbf{k}}\left|\nabla^2\psi_{i,\mathbf{k}}\right\rangle d^3\mathbf{k}\right., \tag{2.176}$$

which means summations of $\mathbf{k}$ vectors in Eq. (2.170) can be substituted by numerical integration of $\mathbf{k}$ vectors over the whole IBZ. Similarly, other terms in Eq. (2.169) can also be expressed as numerical integrations of $\mathbf{k}$ vectors. In Figure 2.5, some of the special $\mathbf{k}$ points and lines are also labeled, because integrations of all wave vectors $\mathbf{k}$ in the IBZ are performed along these directions.

The introduction of IBZ further reduces the computational efforts, while it is still not practical to perform an analytical integration of all wave vectors $\mathbf{k}$ in the IBZ, especially for some systems with large volumes of the FBZ and with low symmetry. In this way, we have to adopt some approximations, through which numerical integrations of $\mathbf{k}$ are made possible in the IBZ. Supposing the curve of eigenvalues of the Kohn-Sham equation along some specified directions of $\mathbf{k}$ is very smooth, it would be reasonable for us to equal the integrations of $\mathbf{k}$ in this region to the mean value of $\mathbf{k}$ in this region times the length of the region, in which the smoothness of the curve is preserved. For other situations in which the curve of eigenvalues of the Kohn-Sham equation along some specified directions of $\mathbf{k}$ is very steep, one would need more $\mathbf{k}$ points to better reproduce the result of analytical integration in this region. The efficiency of such approximations depends on how the special $\mathbf{k}$ points are sampled. One of the most popular $\mathbf{k}$ points sampling methods has been proposed by Monkhorst and Pack [54]. Monkhorst and Pack gave a very simple formula for the $\mathbf{k}$ points sampling as below,

$$\mathbf{k} \equiv \sum_{i=1}^{3} \frac{2n_i - N_i - 1}{2N_i} \mathbf{b}_i , \tag{2.177}$$

in which $\mathbf{b}_i$ are the reciprocal primitive lattice vectors, $n_i = 1, 2, \ldots, N_i$, and N_i is an integer which is preselected and determines the number of $\mathbf{k}$ points being used in numerical integrations. The Monkhorst-Pack scheme works in such a way that each reciprocal primitive lattice vector is divided into N_i portions, and then the mean value of each portion is selected to combine with mean values of portions from the other two reciprocal primitive lattice vectors to determine the position of each $\mathbf{k}$ point. As each n_i has a total of N_i different values, the total number of $\mathbf{k}$ points selected to represent all possible $\mathbf{k}$ vectors in the FBZ will be $N_1 \times N_2 \times N_3$. Proper sampling of $\mathbf{k}$ points allows the Kohn-Sham equation to be solved with a finite number of $\mathbf{k}$ points, and as a result, the energy of each spinorbital can also be determined.

2.8.3 Pseudopotentials

In previous sections, we talked about Bloch's theorem and a numerical method of the Brillouin zone sampling. We also mentioned that contributions to the total energy which involve core electrons in extended systems are often approximated by pseudopotentials. The introduction of pseudopotentials was due to the fact that core electrons are essentially inert, or one can say that valence electrons contribute much more than core electrons to most of the physical/chemical properties observed. Because the distances between core electrons and the nuclei are very short compared with the distances between valence electrons and the nuclei, according to Eq. (2.128), the external potential has a $1/r$ singularity and there will be a cusp in the core region, which creates problems as one needs to use a large number of plane waves (assuming one uses plane wave basis sets) to accurately describe the total electron density. However, if one uses pseudopotentials, the number of plane waves needed for a good description of the electronic structure is much less, as pseudo wavefunctions resulting from pseudopotentials are nodeless and smooth in the core region.

Pseudopotentials are constructed to faithfully reproduce behaviors of the ionic potential between valence electrons and the complex of core electrons plus the nuclei outside of the core region, and at the same time to make sure that the pseudo wavefunctions do not have radial nodes in the core region. As the effect of core electrons are replaced by pseudopotentials, the number of electrons one

needs to deal with is reduced, and the number of electronic bands one needs to solve for is also decreased. The advantages of using pseudopotentials reduce the computational efforts quite a lot. They make some previous calculations, which were very difficult at one time, become practicable, and nowadays they are widely used in calculations of extended systems.

Currently, the pseudopotential approximation is mostly implemented together with DFT methods in codes for extended systems. A flow chart which describes the procedure of constructing a pseudopotential for an atom is shown in Figure 2.6. Such a pseudopotential, which is generated from an all-electron calculation performed on an atom, is termed an *ab initio* pseudopotential, and it is distinguished from an empirical pseudopotential which is fitted to experiment.

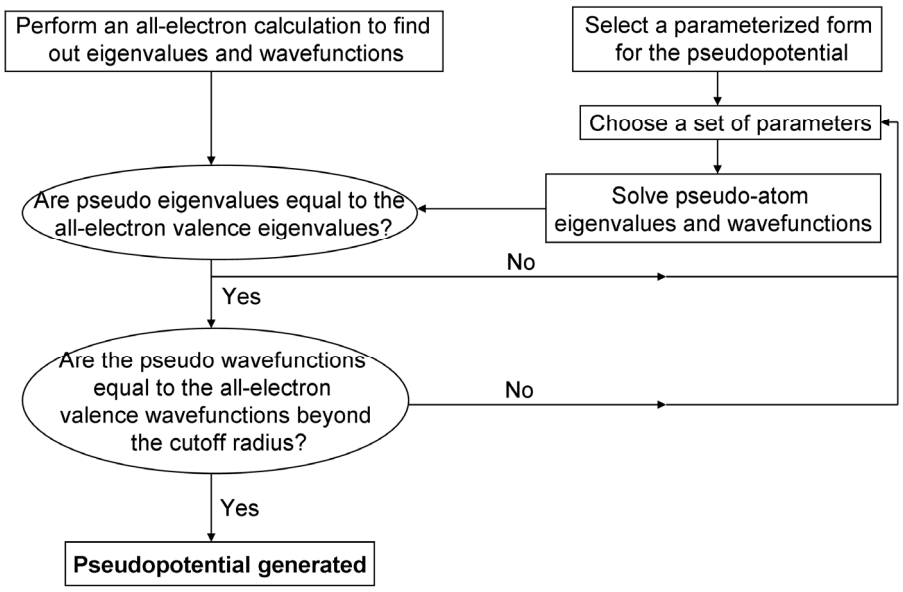


Figure 2.6: Procedure of constructing a pseudopotential for an atom.

There are two types of pseudopotentials, one is local and the other is non-local. Local pseudopotentials use the same potential for different angular momenta, while for non-local pseudopotentials, one uses different potentials for different angular momenta. The quality of a pseudopotential can be measured by many factors, including transferability and softness. The transferability of a pseudopotential indicates its ability to describe valence electrons in different bonding environments, while the softness of a pseudopotential indicates the number of plane waves one needs in a calculation, on the assumption that the same level of accuracy is maintained.

The introduction of pseudopotentials simplifies the calculations of the Coulomb and Exchange operators in the two-electron integrals and reduces the computational cost of such integral calculations, and thus accelerates the convergence of the wavefunction expansion. At the same time, as a method originally developed for periodic calculations and combined with plane-wave basis sets, nowadays pseudopotentials are also used to describe heavy elements in isolated molecular systems, and together with localized basis sets, they facilitate calculations that involve heavy elements.

2.8.4 The Projector Augmented Waves

More recently, the projector augmented wave (PAW) method [55] was proposed as an alternative for the pseudopotential approximation. The PAW method defines a functional for the total energy calculation which involves auxiliary localized functions, and these auxiliary localized functions actually retain the entire set of all-electron core wavefunctions along with smooth parts of the valence wavefunctions. In the current thesis, we use PAW potentials in calculations of extended systems.

The basic idea of the PAW method is to transform the true all-electron wavefunctions into localized auxiliary wavefunctions, which are smooth and can be expanded with rapidly convergent plane waves, and then all physical properties will be evaluated after the true all-electron wavefunctions are reconstructed. Let us denote a one-electron wavefunction as $|\psi_n\rangle$, in which the subscript n indicates a one-electron state defined by an electronic band index, a $\mathbf{k}$ vector, and a spin index. We define a transformation with which the true all-electron wavefunctions $|\psi_n\rangle$ can be transformed into the auxiliary wavefunctions $|\tilde{\psi}_n\rangle$,

$$|\tilde{\psi}_n\rangle = \hat{\mathbf{U}}|\psi_n\rangle, \tag{2.178}$$

in which $\hat{\mathbf{U}}$ is a transformation operator. Accordingly, the true all-electron wavefunctions $|\psi_n\rangle$ can be transformed from the auxiliary wavefunctions $|\tilde{\psi}_n\rangle$ with the following equation,

$$|\psi_n\rangle = \hat{\mathbf{T}}|\tilde{\psi}_n\rangle, \tag{2.179}$$

in which $\hat{\mathbf{T}} = \hat{\mathbf{U}}^{-1}$. The transformation operator $\hat{\mathbf{T}}$ is defined by

$$\hat{\mathbf{T}} = \mathbf{1} + \sum_a \hat{\mathbf{T}}^a, \tag{2.180}$$

in which a is an atom index, and $\hat{\mathbf{T}}^a$ have no effect outside a certain cutoff

radius. Inside the cutoff radius, the true all-electron wavefunctions $|\psi_n\rangle$ are expanded through the partial waves ϕ_i^a, which are determined from an all-electron calculation performed for an isolated atom. Thus we have

$$|\psi_n\rangle = \sum_i P_{ni}^a |\phi_i^a\rangle, \tag{2.181}$$

in which $|\phi_i^a\rangle = (1+\hat{\mathbf{T}}^a)|\tilde{\phi}_i^a\rangle$, with $|\tilde{\phi}_i^a\rangle$ as the auxiliary partial waves corresponding to the partial waves $|\phi_i^a\rangle$, and the expansion coefficients P_{ni}^a are defined by

$$P_{ni}^a = \langle \tilde{p}_i^a | \tilde{\psi}_n \rangle, \tag{2.182}$$

where $|\tilde{p}_i^a\rangle$ satisfy $\sum_i |\tilde{\phi}_i^a\rangle\langle\tilde{p}_i^a| = 1$. In this way, we have

$$\hat{\mathbf{T}}^a = \sum_i \hat{\mathbf{T}}^a |\tilde{\phi}_i^a\rangle\langle\tilde{p}_i^a| = \sum_i \left(|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle\right)\langle\tilde{p}_i^a|, \tag{2.183}$$

and the transformation operator $\hat{\mathbf{T}}$ can be expressed by

$$\hat{\mathbf{T}} = \mathbf{1} + \sum_a \sum_i \left(|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle\right)\langle\tilde{p}_i^a|. \tag{2.184}$$

Finally, the true all-electron wavefunctions $|\psi_n\rangle$ can be expressed as below,

$$|\psi_n\rangle = |\tilde{\psi}_n\rangle + \sum_a \sum_i \left(|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle\right)\langle\tilde{p}_i^a|\tilde{\psi}_n\rangle. \tag{2.185}$$

Inserting Eq. (2.185) into Eq. (2.173), we get the electron density

$$\rho(\boldsymbol{r}) = \sum_n^{Val} |\tilde{\psi}_n|^2 + \sum_a \sum_{i,j} \left(\phi_i^a \phi_j^a - \tilde{\phi}_i^a \tilde{\phi}_j^a\right) D_{i,j}^a + \sum_a \sum_\alpha^{Core} |\phi_\alpha^{a,Core}|^2, \tag{2.186}$$

in which the Hermitian one-center density matrix $D_{i,j}^a$ is defined by

$$D_{i,j}^a = \sum_n \langle\tilde{\psi}_n | \tilde{p}_i^a\rangle\langle\tilde{p}_j^a | \tilde{\psi}_n\rangle. \tag{2.187}$$

With the electron density defined by Eq. (2.186), we are able to determine the total energy of an extended system according to Eq. (2.169). This is the procedure how the total energy can be evaluated through the PAW method. It should be pointed out here that in contrast to pseudopotentials, in which the information on charge density and wavefunctions of the core electrons is missed, the PAW method keeps the full information on all-electron wavefunctions.

2.9 Remarks and Perspectives

The fast developments of different electronic structure methods make computational chemistry not only a counterpart of experimental sciences, but also

a new tool which helps people to discover and create things that people did not know before. As a young science, which has less than 100 years of history, the electronic structure methods in the beginning could only solve problems for systems with up to tens of atoms. However, nowadays with advanced technologies both in computing hardware and software, electronic structure methods can be applied for systems with thousands of atoms. From tens of atoms to thousands of atoms, it might not be such a big step as it appeared to be, but it is really a very big achievement as the process took only less than 50 years, and we do not know what will happen in the next 50 years. Looking forward to the future, we shall not be self-satisfied and stop here, as the systems we need to deal with will be much bigger and the accuracy we need to acquire will be much higher. Fortunately, we are glad to see that the electronic structure methods are still thriving, and a lot of new methods (such as the linear scaling order-N methods [56]) with better performances are being developed. Hopefully these newly developed methods will help people discover and create new molecules and materials that people have never imagined before, and thus make computational chemistry as a science to be side-by-side or even superior to the experimental science.

References:

[1] Piela, L., *Ideas of Quantum Chemistry*. Elsevier: Amsterdam, 2007; p 1086.

[2] Jensen, F., *An Introduction to Computational Chemistry*. Wiley: Chichester, 1999; p 429.

[3] See pages 223-224 of Bibliography 1.

[4] A detailed introduction of this term will be given later in other sections.

[5] Simons, J.; Jordan, K. D., Abinitio Electronic-Structure of Anions. *Chemical Reviews (Washington, DC, United States)* **1987,** 87, 535-555.

[6] Bersuker, I. B., Modern Aspects of the Jahn-Teller Effect Theory and Applications to Molecular Problems. *Chemical Reviews (Washington, DC, United States)* **2001,** 101, 1067-1114.

[7] Atkins, P. W.; Friedman, R. S., *Molecular Quantum Mechanics*. Oxford University Press: Oxford, 2004; p 573.

[8] See Appendix M of Bibliography 1.

[9] See Section 8.5.1 of Bibliography 1.

[10] Moller, C.; Plesset, M. S., Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934,** 46, 618-622.

[11] See Section 10.10.1 of Bibliography 1.

[12] See Appendix M of Bibliography 1.

[13] Bartlett, R. J.; Musial, M., Coupled-Cluster Theory in Quantum Chemistry. *Reviews of Modern Physics* **2007,** 79, 291-352.

[14] Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Physical Review* **1964,** 136, B864-B871.

[15] Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965,** 140, A1133-A1138.

[16] See Section 11.2 of Bibliography 1.

[17] Perdew, J. P.; Schmidt, K. In *Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy*, Density Functional Theory and Its Application to Materials, Antwerp (Belgium), 2001; AIP: Antwerp (Belgium), 2001; pp 1-20.

[18] See Section 6.1 of Bibliography 2.

[19] Vosko, S. H.; Wilk, L.; Nusair, M., Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin-Density Calculations - a Critical Analysis. *Canadian Journal of Physics* **1980**, 1200-1211.

[20] Ceperley, D. M.; Alder, B. J., Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters* **1980,** 45, 566-569.

[21] Perdew, J. P.; Yue, W., Accurate and Simple Density Functional for the Electronic Exchange Energy: Generalized Gradient Approximation. *Physical Review B* **1986,** 33, 8800-8802.

[22] Becke, A. D., Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Physical Review A* **1988,** 38, 3098-3100.

[23] Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B* **1988,** 37, 785-789.

[24] Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996,** 77, 3865-3868.

[25] Perez-Jorda, J. M.; Becke, A. D., A Density-Functional Study of Van Der Waals Forces: Rare Gas Diatomics. *Chemical Physics Letters* **1995,** 233, 134-137.

[26] Perez-Jorda, J. M.; San-Fabian, E.; Perez-Jimenez, A. J., Density-Functional Study of Van Der Waals Forces on Rare-Gas Diatomics: Hartree--Fock Exchange. *Journal of Chemical Physics* **1999,** 110, 1916-1920.

[27] Zhang, Y.; Pan, W.; Yang, W., Describing Van Der Waals Interaction in Diatomic Molecules with Generalized Gradient Approximations: The Role of the Exchange Functional. *Journal of Chemical Physics* **1997,** 107, 7921-7925.

[28] Patton, D. C.; Pederson, M. R., Application of the Generalized-Gradient Approximation to Rare-Gas Dimers. *Physical Review A* **1997,** 56, R2495.

[29] Sauer, J.; Dobler, J., Structure and Reactivity of V2o5: Bulk Solid, Nanosized Clusters, Species Supported on Silica and Alumina, Cluster Cations and Anions. *Dalton Transactions* **2004**, 3116-3121.

[30] Terakura, K.; Oguchi, T.; Williams, A. R.; Kubler, J., Band Theory of Insulating Transition-Metal Monoxides: Band-Structure Calculations. *Physical Review B* **1984,** 30, 4734-4747.

[31] Anisimov, V. I.; Zaanen, J.; Andersen, O. K., Band Theory and Mott Insulators: Hubbard U Instead of Stoner I. *Physical Review B* **1991,** 44, 943-954.

[32] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E., Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Physical Review Letters* **2003,** 91, 146401-4.

[33] Zhao, Y.; Truhlar, D. G., Density Functionals with Broad Applicability in Chemistry. *Accounts of Chemical Research* **2008,** 41, 157-167.

[34] Zhao, Y.; Truhlar, D. G., A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *Journal of Chemical Physics* **2006,** 125, 194101-18.

[35] VOSKO, S.; WILK, L.; NUSAIR, M., Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin-Density Calculations - a Critical Analysis. *CANADIAN JOURNAL OF PHYSICS* **1980**, 1200-1211.

[36] Becke, A. D., Density-Functional Thermochemistry. Iii. The Role of Exact Exchange. *Journal of Chemical Physics* **1993,** 98, 5648-5652.

[37] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *Journal of Physical Chemistry* **1994,** 98, 11623-11627.

[38] Zhao, Y.; Truhlar, D., The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **2008,** 120, 215-241.

[39] Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E., Hydrogen Bonding and Stacking Interactions of Nucleic Acid Base Pairs: A Density-Functional-Theory Based Treatment. *Journal of Chemical Physics* **2001,** 114, 5149-5155.

[40] Grimme, S., Accurate Description of Van Der Waals Complexes by Density Functional Theory Including Empirical Corrections. *Journal of Computational Chemistry* **2004,** 25, 1463-1473.

[41] Grimme, S., Semiempirical Gga-Type Density Functional Constructed with a Long-Range Dispersion Correction. *Journal of Computational Chemistry* **2006,** 27, 1787-1799.

[42] Grimme, S., Semiempirical Hybrid Density Functional with Perturbative Second-Order Correlation. *Journal of Chemical Physics* **2006,** 124, 034108-16.

[43] *Gaussian 09, Revision A.1*, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E.

Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

[44] Hehre, W. J.; Stewart, R. F.; Pople, J. A., Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *Journal of Chemical Physics* **1969,** 51, 2657-2664.

[45] Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L., Basis Set Exchange: A Community Database for Computational Sciences. *Journal of Chemical Information and Modeling* **2007,** 47, 1045-1052.

[46] Dunning, J. T. H., Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *Journal of Chemical Physics* **1989,** 90, 1007-1023.

[47] Helgaker, T.; Klopper, W.; Koch, H.; Noga, J., Basis-Set Convergence of Correlated Calculations on Water. *Journal of Chemical Physics* **1997,** 106, 9639-9646.

[48] van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H., State of the Art in Counterpoise Theory. *Chemical Reviews* **1994,** 94, 1873-1885.

[49] Mayer, I.; ViboK, A., Bsse-Free Second-Order Intermolecular Perturbation Theory. *Molecular Physics* **1997,** 92, 503-510.

[50] van Mourik, T.; Karamertzanis, P. G.; Price, S. L., Molecular Conformations and Relative Stabilities Can Be as Demanding of the Electronic Structure Method as Intermolecular Calculations. *Journal of Physical Chemistry A* **2006,** 110, 8-12.

[51] Jensen, F., The Magnitude of Intramolecular Basis Set Superposition Error. *Chemical Physics Letters* **1996,** 261, 633-636.

[52] Martin, R. M., *Electronic Structure : Basic Theory and Practical Methods*. Cambridge University Press: Cambridge, 2004; p 624.

[53] Kaxiras, E., *Atomic and Electronic Structure of Solids*. Cambridge University Press: Cambridge, 2003; p 676.

[54] Monkhorst, H. J.; Pack, J. D., Special Points for Brillouin-Zone Integrations. *Physical Review B* **1976,** 13, 5188-5192.

[55] Blochl, P. E., Projector Augmented-Wave Method. *Physical Review B* **1994,** 50, 17953-17979.

[56] Artacho, E.; Sanchez-Portal, D.; Ordejon, P.; Garcia, A.; Soler, J. M., Linear-Scaling Ab-Initio Calculations for Large and Complex Systems. *Physica Status Solidi B-Basic Research* **1999,** 215, 809-817.

Chapter 3

Exploring the Potential Energy Surface

In Chapter 2, we talked about the electronic structure methods which were used in theoretical chemistry. Before we really start the electronic structure calculations, there is another important problem that needs to be solved. The electrons interact with the nuclei, but where are the nuclei? One can not perform Born-Oppenheimer calculations without knowing the positions of the nuclei. Fortunately we have the thermodynamics theory which tells us that a structure (which is determined by the nuclear positions) with lower energy is more stable than that with higher energy, thus low-energy structures are more probable to be observed in experiments. With this criterion, once we know the energy of each state (that is each structure with different nuclear positions), we can rank the energy of all possible states from low to high, and the lowest energy structure will most likely be present in experiments. With this lowest energy structure, one can perform electronic structure calculations and compare theoretical results with results from experiments.

3.1 Methods of Geometry Optimizations

Generally, there are two types of systems which we possibly meet during our study, one is a molecular system, which is limited to a finite number of atoms, and the other is an extended system with periodicity and thus theoretically it has an infinite number of atoms. To perform an electronic structure calculation, one would have to know the structure of the system. The structure of a system could either come from experiment or from an artificial guess based on physical and chemical laws. Let us take a molecule as an example. In Chapter 2 we mentioned that within the Born-Oppenheimer approximation, the PES is a function of the nuclear positions. In this way, any artificial guess of the molecular structure is a point on the PES of this molecule, though it does not have to be a stationary point (say a local minimum or a first order saddle point). A local minimum structure is the structure of a molecule at its stable geometry, which means there are no forces acting on the atoms, and if small displacements are applied for different atoms in

the molecule, nuclei will return to the initial positions, because the forces brought by these small displacements tend to push nuclei in a direction towards the initial positions. On the other hand, a first order saddle point corresponds to a possible transition state between two local minimum structures. A first order saddle point also corresponds to a stable geometry with no forces acting on the atoms, but nuclei will not return to the initial positions if small displacements are applied, because the force brought by these small displacements will push the nuclei towards a local minimum. Therefore we start from an initial guess structure, which is a certain point on the PES, and our goal is to determine the nearest stationary point which could either be a local minimum structure or a transition state. We call such a procedure a geometry optimization, and normally a geometry optimization consists of a few steps of atomic displacements.

The concept of a geometry optimization can be well explained in a mathematical way [1]. Let us take a single geometry displacement as an example. Suppose one starts from an initial position $\mathbf{x}_0$ on the PES, and after the geometry displacement, one arrives in a new position $\mathbf{x}_1$, then the potential energy of this new position can be approximated with terms up to the second order of a Taylor expansion around the current point $\mathbf{x}_0$ as below,

$$E(\mathbf{x}_1) \approx E(\mathbf{x}_0) + (\mathbf{x}_1 - \mathbf{x}_0)^T \bullet \mathbf{g}_0 + \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_0)^T \bullet \mathbf{H}_0 \bullet (\mathbf{x}_1 - \mathbf{x}_0), \tag{3.1}$$

in which the superscript T implies a transposed vector, $\mathbf{g} = \nabla E(\mathbf{x})$, $\mathbf{H} = \nabla^2 E(\mathbf{x})$, and $\mathbf{g}_0$ and $\mathbf{H}_0$ are the gradient and the second order derivatives (Hessian matrix) of the potential energy determined at the current position $\mathbf{x}_0$, respectively. A stationary point is defined in such a way that all components of the energy gradient $\mathbf{g}$ are equal to zero, a local minimum further requires that all eigenvalues of the Hessian matrix $\mathbf{H}$ are non-negative, and a first order saddle point requires that one eigenvalue of the Hessian matrix $\mathbf{H}$ is negative while all other eigenvalues are non-negative. To perform a geometry optimization, one needs to determine what atomic displacements should be applied. Depending on which type of stationary points one wants to locate, the direction of atomic displacements differs. Methods for geometry optimizations of both purposes will be given in the following sections.

3.1.1 The Newton-Raphson Method

The Newton-Raphson method is one of the most popular methods which have

been widely implemented in current computational chemistry codes to perform geometry optimizations. Let us insert a unitary matrix $\mathbf{U}$ into Eq. (3.1) so that the Hessian matrix $\mathbf{H}$ can be diagonalized,

$$E(\mathbf{x}_1) \approx E(\mathbf{x}_0) + \mathbf{g}_0^T \mathbf{U}\mathbf{U}^T (\mathbf{x}_1 - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_0)^T \mathbf{U}\mathbf{U}^T \mathbf{H}_0 \mathbf{U}\mathbf{U}^T (\mathbf{x}_1 - \mathbf{x}_0), \quad (3.2)$$

in which $\mathbf{U}^T\mathbf{H}_0\mathbf{U}$ is a diagonal matrix and has eigenvalues λ_k ($k = 1, 2, \cdots, 3N$, supposing we have a total of N atoms in the system). In this way, Eq. (3.2) can be rewritten in the following form,

$$E(\mathbf{x}_1) \approx E(\mathbf{x}_0) + \sum_m G_m^T Q_m + \frac{1}{2}\sum_m Q_m \lambda_m Q_m, \quad (3.3)$$

in which $Q_m = \sum_k U_{m,k}^T (\mathbf{x}_1 - \mathbf{x}_0)_k$, $G_m = \sum_k U_{m,k}^T \frac{\partial E}{\partial (\mathbf{x}_1)_k}$ is the energy gradient along the *m*-th eigenvector of the Hessian matrix $\mathbf{H}$, and the subscript k implies the *k*-th component of the a vector. The position of an adjacent stationary point defined by Eq. (3.3) can be determined by differentiating this equation with respect to Q_m, and we get

$$G_m + \lambda_m Q_m = 0 \Rightarrow Q_m = -\frac{G_m}{\lambda_m}, \quad (3.4)$$

in which Q_m indicates the direction and magnitude of a geometry optimization step that one should take to arrive in a stationary point from the current position. This is the way how the Newton-Raphson method works. It should be pointed out that as far as the potential energy can be well approximated by Eq. (3.1), only one step of geometry optimization is needed to find the nearest stationary point.

The magic of the Newton-Raphson method on geometry optimization is that with only one simple equation (see Eq. (3.4)), the method is capable of finding the nearest stationary point on the PES, which could be either a local minimum point or a first order saddle point, and this can be well illustrated by Figure 3.1, in which the red curve represents a PES. In Figure 3.1, suppose one starts from Position 1, where $G_m < 0$, and suppose it is close to a local minimum (Position 2), then all λ_m will be positive, which leads to $Q_m > 0$ according to Eq. (3.4), thus the optimization will move forward and arrive at Position 2. If one starts from Position 3, then $G_m > 0$ and $\lambda_m > 0$, thus $Q_m < 0$ and the optimization will move backward, which will also arrive at Position 2. However, if one starts from Position 4, which is close to a first order saddle point (Position 5), then we know $G_m > 0$ and one of λ_m will be negative, and for atomic displacement along the eigenvector corresponding to this λ_m, we have $Q_m > 0$, thus the optimization

will move forward in this direction and arrive at Position 5. Analogously, if one starts from Position 6, one will get $Q_m < 0$ along the eigenvector which corresponds to the negative λ_m, and the optimization will move backward and arrive at Position 5. Therefore, wherever one starts from, as far as the potential energy can be well approximated by Eq. (3.1), the success of a geometry optimization is guaranteed within only one step. The type of a stationary point found by the Newton-Raphson method is determined by λ_m, say if all λ_m are non-negative, then a local minimum point will be found, or if only one λ_m is negative while all others are non-negative, then a first order saddle point will be found. Note that to find a first order saddle point using the Newton-Raphson method, one must start from a reasonable initial guess of a transition state with one λ_m being negative, otherwise the optimization might just end up at a local minimum point.

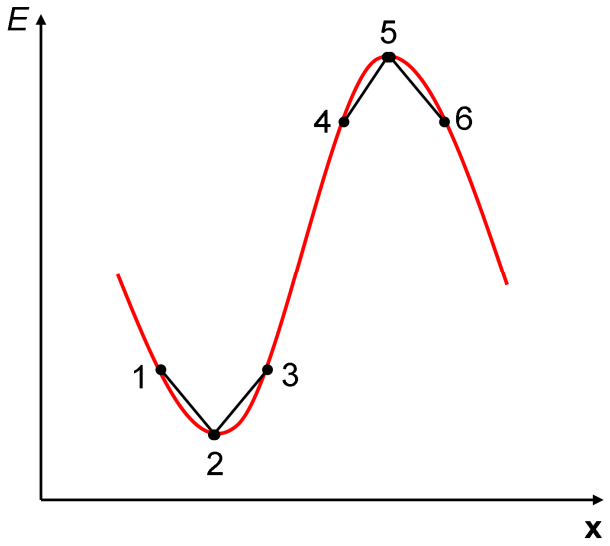


Figure 3.1: The Newton-Raphson method for geometry optimization.

On a real PES, as the potential energy may not be well approximated by Eq. (3.1), the position of a stationary point can be determined through a few iterative steps of geometry modifications defined by Eq. (3.4), and the geometry optimization will stop when the change in energy or geometry of two successive steps falls below a certain threshold. The Newton-Raphson method is very efficient in geometry optimizations. However, as a method which depends on both $\mathbf{g}$ and $\mathbf{H}$, the Newton-Raphson method is very time-consuming. An alternative solution is to use an approximated Hessian which is updated with energy and gradient information obtained during the optimization. Depending on how the Hessian

matrix is updated for each step of geometry optimization, there are a few methods available which could help to reduce the computational cost of the Newton-Raphson method, such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [2-5], and the Berny algorithm [6], which is widely used in Gaussian [7].

3.1.2 Locating Local Minima

As what we have just introduced, the Newton-Raphson method can be used to find a local minimum, providing the starting point is not close to a transition state. We also mentioned that the Newton-Raphson method is very time-consuming, because it needs to calculate both $\mathbf{g}$ and $\mathbf{H}$. Indeed, there are methods for geometry optimization, which only depend on $\mathbf{g}$, such as the steepest descent method and the conjugate gradient method.

The steepest descent method is the most direct method to find a local minimum on a PES. It is based on the fact that the potential energy of a point decreases fastest if one follows the negative gradient of potential energy. The steepest descent method starts from calculation of the energy $E(\mathbf{x}_0)$ and energy gradient $\mathbf{g}_0$ at the initially assumed geometry, in which $\mathbf{g}_0$ could be calculated either analytically or numerically, and then the position of the *i*-th point is determined by

$$\mathbf{x}_i = \mathbf{x}_{i-1} - \gamma_{i-1}\mathbf{g}_{i-1}, \tag{3.5}$$

in which $i \geq 1$, γ is the step length, and it can either be a constant or it can also be a quantity which depends on the magnitude of the energy gradient $\mathbf{g}$. Then the above procedure is iterated, and hopefully it converges to a desired local minimum, supposing a certain convergence criterion is preselected and enough steps of geometry optimizations are allowed to perform.

The conjugate gradient method is more advanced compared with the steepest descent method. As the steepest descent method tries to find a minimum through following the negative gradient of potential energy, there is a possibility that a part of the preceding successful atomic displacement will be drawn back due to an overweighted following displacement. The conjugate gradient method fixes this problem through following the gradient in such a way that displacements of two subsequent searches are "conjugate" with each other. This strategy can be expressed with the following equation,

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \gamma_i \mathbf{d}_i, \tag{3.6}$$

in which $\mathbf{d}_i$ indicates the search direction, and is defined by

$$\mathbf{d}_i = -\mathbf{g}_i + \beta_i \mathbf{d}_{i-1}, \tag{3.7}$$

where β is known as the conjugate gradient parameter, and one of the most popular forms of β is expressed as below [8],

$$\beta_i = \frac{\mathbf{g}_i^T \left(\mathbf{g}_i - \mathbf{g}_{i-1}\right)}{\mathbf{g}_{i-1}^T \mathbf{g}_{i-1}}. \tag{3.8}$$

The conjugate gradient method turns out to converge much faster than the steepest descent method, but it requires more storage space as it uses gradients of both the current step as well as the preceding step. It should be pointed out here that both the steepest descent method and the conjugate gradient method can only locate positions of local minimum points.

3.1.3 Locating Transition States

We just introduced three of the most popular methods which are used to find local minimum structures on a PES starting from an arbitrary point. While the local minima represent one type of important stationary points, there is another type, which is also very important. These are the first order saddle points. If finding a local minimum is challenging, finding a first order saddle point is even more so. We mentioned that the Newton-Raphson method can also be used to find a first order saddle point, if the initial guess for a transition state is good enough. Some other methods for locating transition state structures have also been proposed. For example, there are methods based on interpolation between the reactant and the product, which are actually two adjacent local minimum points on the PES, and one of these methods is the nudged elastic band (NEB) method [9-11].

In the NEB method, a linear chain of images connecting the reactant and the product is generated. Then a spring potential is applied for all the images including the starting and ending points, and the total energy of such a system is expressed as below,

$$S\left(\mathbf{R}_0, \cdots, \mathbf{R}_N\right) = \sum_{i=1}^{N-1} E\left(\mathbf{R}_i\right) + \sum_{i=1}^{N} \frac{k_{sp}}{2} \left(\mathbf{R}_i - \mathbf{R}_{i-1}\right)^2, \tag{3.9}$$

in which $\mathbf{R}$ represent geometries of the $N+1$ images, k_{sp} is the spring constant, $E\left(\mathbf{R}_i\right)$ is the energy of the connecting images, and the second term on the right hand side represents the potential energy of the springs that connect the images. The NEB works in such a way that the $N-1$ connecting images, subjected to a force which is projected on each image [11], are relaxed, and then

the total energy expressed in Eq. (3.9) is minimized until the force acting on each connecting image becomes zero. If the number of connecting images is infinite, then after relaxation, positions of these connecting images form a curve which represents the minimum energy path (MEP) that connects the reactant and the product, and the energetically highest point on this MEP is just the transition state that one is looking for. The NEB method is used to search for transition states of extended systems in the current thesis.

Besides the NEB method, there are also some other methods for transition state searches which are not based on interpolation between the reactant and the product, and one of these methods is the eigenvector following method [12]. Being different from the Newton-Raphson method for transition state search, which requires the starting point to have a negative Hessian eigenvalue, the eigenvector following method could start from any point on the PES and it just follows the direction of a preselected eigenvector (if it starts from a local minimum, then it corresponds to a certain normal vibrational mode), and then potential energy along this direction will be maximized, while potential energy along directions corresponding to other eigenvectors will be minimized. Finally, a first order saddle point connecting the starting point and an associated local minimum through the path corresponding to the preselected eigenvector will be found.

3.1.4 Following Reaction Path

We just mentioned that the MEP connects two adjacent local minima on the PES through a first order saddle point. In transition state theory, one frequently uses the concept of intrinsic reaction coordinate (IRC) [13], which is related to MEP, and which is used to trace the progress of a "reaction" from one local minimum to another. The IRC can be used to verify that the first order saddle point connects two specific local minima. The IRC could also be used to locate positions of the two local minima which a first order saddle point connects. Supposing there is a first order saddle point, a classical particle will depart from this point and it will move on the PES. One calculates the Hessian matrix at this point, one diagonalizes it, and one will get a negative eigenvalue. Supposing an infinitely small displacement is applied for the classical particle along the direction determined by the eigenvector corresponding to the negative eigenvalue, then there will be forces acting on this particle, and it will start to move. Depending on

the displacement applied, which could make the particle to move either forward or backward, the particle will end up at two different local minima. The IRC is uniquely defined as the trajectory of the classical particle, which connects two local minima through the first order saddle point. If the mass-weighted coordinates (see Eq. (2.32)) are used, then the MEP is equivalent to the IRC, which is also the steepest descent path from the first order saddle point to each of the local minima.

A few methods have been proposed to determine the IRC, and one of these methods, which was adopted by Gaussian, was proposed by Gonzalez and Schlegel [14, 15]. In this method, to follow the reaction path, one starts from the first order saddle point, and one determines the initial step in the same way like the classical particle which we discussed above. The reaction path can be constructed in such a way that we assume the reaction path between two successive atomic displacements (say $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$) is an arc of a circle, so that the gradients of these two successive steps (say $\mathbf{g}_i'$ and $\mathbf{g}_{i+1}'$) are tangent to this reaction path, and they intersect at a pivot point (say $\mathbf{x}_{i+1}^*$). Then the displacements of these two successive steps are confined by the following conditions,

$$\mathbf{x}_{i+1}^* = \mathbf{x}_i - \frac{1}{2} s \cdot \frac{\mathbf{g}_i'}{\left|\mathbf{g}_i'\right|}, \tag{3.10}$$

$$\left|\mathbf{x}_{i+1} - \mathbf{x}_{i+1}^*\right| = \frac{1}{2} s, \tag{3.11}$$

in which s is the step length, $\mathbf{g}_i'$ is defined in mass-weighted coordinates, and $\mathbf{x}_{i+1}$ could be found by an energy minimization under the constraint specified by Eq. (3.11). It should be pointed out that no calculation is performed at $\mathbf{x}_{i+1}^*$, and one only needs to calculate the first order derivatives at points along the reaction path during the calculation. With a sufficient number of steps of constrained geometry optimizations along the path defined by Eqs. (3.10) and (3.11), the reaction path from the transition state to both the reactant and the product can be determined, and the activation energy as well as the enthalpy change of a reaction can also be obtained.

3.2 Methods of Conformational Searches

The conformational space of a molecule is spanned by all its internal rotational bonds [16]. Each different combination of these internal rotational bonds gives birth to an individual conformation, while every conformation has its own, in general different, electronic energy, which corresponds to a point on the PES. According to the theory of statistical thermodynamics [17], conformations with lower electronic energies have more favorable Boltzmann factors and they dominate the distribution of conformers. At the same time, only conformers with significant populations contribute to molecular properties measured in experiments. Therefore, to get more information about a molecule from a theoretical study, we need to find the low-energy conformations which determine the physical/chemical properties of the molecule, and this is the reason why we need to perform a search of the conformational space. As we already mentioned, the PES of a molecule is characterized by many local minima and one global minimum, our goal of the conformational search is to find the global minimum structure as well as the most stable local minimum structures.

The shape of the PES of a molecule is dependent to some extent on the method used to calculate the potential energy, and the method could be either the molecular mechanics method or one of the electronic structure methods that we have talked about in Chapter 2. Having chosen a method to calculate the potential energy, the next challenge is to determine how the PES should be explored to obtain the global minimum and the most stable local minima. Practically, it is not a good idea to scan the potential energy of a molecule with respect to its nuclear positions, as molecules have $3N-6(5)$ degrees of freedom, and this procedure is impractical. There are many methods which can explore the PES in a more intelligent and efficient way [18]. These methods can be divided into two types according to their ways of generating the initial geometries. First, there are stochastic search methods, which include the Monte Carlo method, simulated annealing and genetic algorithms. Second, there is the deterministic search method, which is often dubbed the systematic search method.

3.2.1 Stochastic Search Methods

Stochastic search methods generate the initial geometries by random sampling, but the succeeding procedures are different for different methods. Three representative stochastic search methods are the Monte Carlo method, simulated annealing and genetic algorithms.

The Monte Carlo method starts from a given geometry which is sometimes taken from a known low-energy conformation, or it could also start from a randomly selected geometry, and it follows that an energy calculation is performed for this starting geometry to estimate its electronic energy (say E_1). Then a micro-modification or a series of micro-modifications (which are random changes in one or more degrees of freedom of this molecule) are applied to the starting geometry, which results in a new conformation with the electronic energy E_2. To decide if one should accept this new conformation, one uses the Metropolis criterion, which is stated as below:

(1) if $E_2 \le E_1$, then one accepts the new conformation and goes to step 3;

(2) however, if $E_2 > E_1$, then one uses a randomly generated number u, in which $0 \le u \le 1$, and one compares u with another number a, which is determined by $a = e^{-\left(\frac{E_2 - E_1}{k_B T}\right)}$:

   a) if $u \le a$, then one accepts the new conformation and moves to step 3;

   b) if $u > a$, then one rejects the new conformation, one restores the geometry corresponding to E_1, and one goes to step 3;

(3) one starts a new step of micro-modifications, which results in another conformation with electronic energy E_3.

The above steps are repeated, and depending on the micro-modifications applied, the more times the above steps are repeated, the bigger probability to find the global minimum. The Monte Carlo method correctly samples the Boltzmann distribution of different conformations of a molecule with respect to the temperature. It is very fast, because it only requires energy calculations and one does not need to calculate forces. It is a powerful method, and it can also be used for ring systems. A disadvantage of the Monte Carlo method is that there is no real end point for it and one can never make sure that the global minimum conformation has been found.

Simulated annealing is another stochastic method for conformational search [19]. It is coupled with the molecular dynamics (MD) simulation, and is based on a thermodynamic process called annealing. Similar to the Monte Carlo method, it also starts from a known local minimum or a randomly selected geometry. Simulated annealing can be run following different protocols of annealing (such as controls of temperatures and time). A typical simulated annealing is run as below:

(1) perform an MD simulation at a high temperature (say T_1), which results in a new conformation different to a large extent compared with the starting geometry because of the large kinetic energy resulting from the high temperature;
(2) perform another MD simulation at a low temperature (say T_2) based on the conformation resulting from step 1, through which the motion of nuclei slows down and a conformation of the same class with limited geometry relaxation is generated;
(3) another set of MD simulations is performed at a high temperature T_3, which is lower than T_1, and at a low temperature T_4, which is lower than T_2, successively.

The above processes are then repeated until it reaches the limit of simulation time or number of simulation steps. An advantage of simulated annealing approach is that when the temperature of simulation is high, there is a big chance for the initial conformation to go over a large portion of the PES and transform into a new conformation, which is different from the initial one, and such a process enables one to overcome a high transition barrier between two adjacent local minima, while when the temperature of simulation is low, the new conformation which is generated in the high temperature simulation starts to shrink on the PES, and hopefully it stops somewhere close to a local minima and ends up at a more stable conformation. It should be pointed out here that the above procedure of simulated annealing does not guarantee to find the global minimum conformation, but it provides a very good opportunity to find a conformation which is much more stable than the starting point. The simulated annealing method is suitable for all types of conformational spaces, including those of ring systems and clusters, which could not be well explored by the systematic search method that we are going to talk about in the next section. The main drawback of simulated annealing is that it is a very slow method, because one does not only need to calculate the energies, but also the forces, which is time-consuming. In addition, simulated annealing tends to stay in a local minimum at low temperatures (see Chapter 5), and a molecule might break down at high temperatures, which would result in a failed conformational search.

The third stochastic method is based on genetic algorithms. Genetic algorithms are derived from the principles of natural evolution, where an initial (parent) population is a set of conformations that are randomly generated for a molecule, or they are chosen as known low-energy local minima. The conformational search

method which adopts a genetic algorithm is run as below:

(1) generate a random initial population (parents);

(2) conformations in the initial population are characterized by their fitness (the potential energy), in which one chooses the low energy conformations as the best parents;

(3) the best parents breed and produce a new population (children), and at the same time a small amount of mutation is usually allowed in the process;

(4) do energy minimizations for the children (including those from mutation), let these children and their parents be in the same generation, and choose the best members in this generation as the new initial population for the next generation.

At last, like in the Monte Carlo method and simulated annealing, one repeats the above process as many times as possible, and the more times they are repeated, the bigger probability to find the global minimum. Genetic algorithms are typically executed with fast electronic structure methods, such as semiempirical methods. In addition, one does not have to climb over the barrier like in simulated annealing. Similarly to the Monte Carlo method, genetic algorithms have no real end points and one can never be sure that the global minimum has been found.

3.2.2 Deterministic Search Method

The second type of methods for conformational searches is deterministic, which is also called as a systematic search method. The systematic search method generates initial geometries through systematic rotation of dihedral angles of a molecule by discrete increments. It can systematically explore all conformational space, thus all possible stable conformers can be found. However, because there are such a large number of initial geometries to optimize, this method is very time-consuming, and an example is shown in Figure 3.2, from which we can see that, to find the global minimum of a canonical tautomer of arginine by the systematic search method, one would need to optimize a total of 839808 initial geometries [20], which is an impractical task even with semiempirical methods. Therefore the systematic search method cannot be used for large molecular systems, which have many rotational degrees of freedom. In addition, it also has limitations for ring systems. What should be pointed out is that systematic searches are of paramount importance because they provide benchmarks for all other approximate (perhaps faster but less reliable) methods.

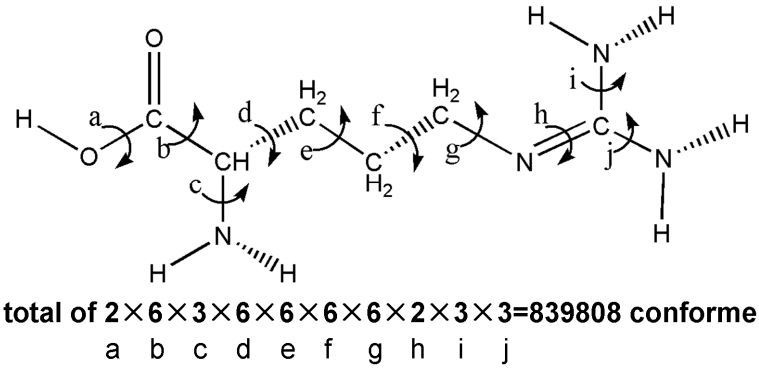


Figure 3.2: A canonical tautomer of arginine.

Our implementation of the deterministic search method will be presented in Chapter 4. This implementation is fully automated, avoids unphysical clashes of substituents, and allows a user to define which degrees of freedom should be probed (see Chapter 7).

3.3 Perspectives

As one of challenging problems in computational chemistry, exploring the PES has attracted a lot of attention in the past few decades, and there is still much work to be done. Looking forward to the future, methods which are faster and can better describe the PES of different systems (say from tens of atoms to thousands of atoms, from small organic molecules to large biomolecules) will be developed, which will make it possible to understand more complex problems. On the other hand, developing new efficient algorithms/methods of exploring PES, such as locating transition states and finding global minimum, is equally important.

In a word, computational chemistry is a complicated science. There are still many fundamental problems to be solved, and both of the issues mentioned above could be a direction where future efforts should be devoted to.

References:

[1] Simons, J., *An Introduction to Theoretical Chemistry*. Cambridge University Press: Cambridge ; New York, 2003; p 461.

[2] Broyden, C. G., Convergence of Single-Rank Quasi-Newton Methods. *Mathematics of Computation* **1970,** 24, 365-382.

[3] Fletcher, R., A New Approach to Variable Metric Algorithms. *Computer Journal* **1970,** 13, 317-322.

[4] Goldfarb, D., A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation* **1970,** 24, 23-26.

[5] Shanno, D. F., Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation* **1970,** 24, 647-656.

[6] Schlegel, H. B., Optimization of Equilibrium Geometries and Transition Structures. *Journal of Computational Chemistry* **1982,** 3, 214-218.

[7] Gaussian 09, Revision A.1, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

[8] Klessig, R.; Polak, E., Efficient Implementations of Polak-Ribiere Conjugate Gradient Algorithm. *Siam Journal on Control* **1972,** 10, 524-549.

[9] Mills, G.; Jonsson, H.; Schenter, G. K., Reversible Work Transition-State Theory - Application to Dissociative Adsorption of Hydrogen. *Surface Science* **1995,** 324, 305-337.

[10] Henkelman, G.; Jonsson, H., Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *Journal of Chemical Physics* **2000,** 113, 9978-9985.

[11] Jonsson, H.; Mills, G.; Jacobsen, K. W., Nudged Elastic Band Method for

Finding Minimum Energy Paths of Transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*, Berne, B. J.; Ciccotti, G.; Coker, D. F., Eds. World Scientific: Singapore, 1998; p 385.

[12] Baker, J., An Algorithm for the Location of Transition-States. *Journal of Computational Chemistry* **1986,** 7, 385-395.

[13] Fukui, K., The Path of Chemical-Reactions - the Irc Approach. *Accounts of Chemical Research* **1981,** 14, 363-368.

[14] Gonzalez, C.; Schlegel, H. B., An Improved Algorithm for Reaction-Path Following. *Journal of Chemical Physics* **1989,** 90, 2154-2161.

[15] Gonzalez, C.; Schlegel, H. B., Reaction-Path Following in Mass-Weighted Internal Coordinates. *Journal of Physical Chemistry* **1990,** 94, 5523-5527.

[16] Leach, A. R., *Molecular Modelling : Principles and Applications*. Prentice Hall: Harlow, 2001; p 744.

[17] Atkins, P. W.; De Paula, J., *Atkins' Physical Chemistry*. 8th ed. ed.; Oxford University Press: Oxford, 2006; p 1064.

[18] Jensen, F., *An Introduction to Computational Chemistry*. Wiley: Chichester, 1999; p 429.

[19] Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by Simulated Annealing. *Science* **1983,** 220, 671-680.

[20] Ling, S. L.; Yu, W. B.; Huang, Z. J.; Lin, Z. J.; Haranczyk, M.; Gutowski, M., Gaseous Arginine Conformers and Their Unique Intramolecular Interactions. *Journal of Physical Chemistry A* **2006**, 12282-12291.

Chapter 4

Constructing Libraries for Systematic Screening of Conformers

Even a relatively small molecule with 10-20 atoms might have a few local minima, which correspond to different conformers. The number of local minima quickly increases with molecular size and the most common algorithms, driven by calculated forces, frequently identify a minimum that is closest to the initial structure, rather than the most stable conformer. Here we discuss how to perform a systematic search of the conformational space for a chain-like molecule. Our approach is fully automated and a user has control which dihedral angles will be probed and with which increments. Moreover, whole fragments of the molecule, which are adjacent to each selected rotational bond, are rotated in a properly selected cylindrical coordinate system and unchemical hybridizations and some "clashes" between neighboring groups, which are common when standard Z-matrices are used, are avoided. A library of potentially relevant conformers is created with a tool, which we call SSC, denoting Systematic Screening of Conformers. Each member of the library is prescreened at a predefined level of theory and the most promising conformers are identified. Finally, they are further evaluated at a higher level of theory to identify the most stable structures and their physicochemical properties. As an example, we demonstrate the results of this approach for 2'-deoxycytidine, as well as methyl p-dimethylaminobenzoate and its two ortho derivatives.

4.1 Introduction

It has been recently demonstrated that combinatorial/computational methods are quite robust to answer fundamental chemical questions. For example, a method has been developed to find the most stable tautomers [1, 2], and the TauTGen code constructs a library of tautomers of a molecular frame built of heavy atoms and hydrogens. Similarly, the ConGENER code [3] constructs a library of

substitution isomers (or the so-called congeners) of a molecule.

The conformational space of a molecule is spanned by its all internal rotational bonds [4]. Each combination of internal rotational bonds gives birth to an individual conformation, while every conformation has its own, in general different, electronic energy, which corresponds to a point on the potential energy surface. According to the theory of statistical thermodynamics [5], conformations with lower electronic energies have more favorable Boltzmann factors and they dominate the distribution of conformers. At the same time, only significantly populated conformers contribute to molecular properties measured in experiments. Therefore, to obtain relevant information about a molecule from a theoretical study, one needs to find the low-energy conformations, as they determine its physicochemical properties.

The potential energy surface of a molecule is characterized by many local minima and one global minimum, and the goal of the conformational search is to find the global minimum structure as well as the most stable local minimum structures. There are stochastic search methods, such as simulated annealing, Monte-Carlo, basin-hopping method [6, 7], which could explore the potential energy surface in an efficient way. However, one never knows for sure that the global minimum has been found by a stochastic search method [7]. The systematic search method, in which one systematically explores the whole conformational space, is the most reliable approach, the accuracy of which is only restricted by the accuracy of the theoretical method employed and the accuracy of sampling of each torsional angle. To perform a systematic search, one would need to generate a library of all reasonable conformers, which are obtained through systematic rotations of dihedral angles of the molecule by discrete increments. This conceptually straightforward task quickly becomes very time-consuming, if one does it manually. It would be much more advantageous to develop a tool which could generate conformers automatically. Indeed, there are tools which are capable of doing this [8, 9]. However, the research community would benefit from a well documented tool, with the source code available in the public domain, so one could modify it for special applications.

The systematic generation of conformations can be done in either the internal coordinates, which are usually written in the Z-matrix format [4], or the Cartesian coordinates. A few algorithms have been suggested for generation of conformers

of a molecule, such as OMEGA [8], and CAESAR [9]. To generate a new conformation, one would need to change dihedral angles. Different schemes have been developed for the purpose of updating dihedral angles, including the simple rotations which is based on a global reference-frame [10], the Denavit-Hartenberg local frames [11], and the atom-group local frames [12]. Zhang and Kavraki [12] compared these three methods, and they concluded that their atom-group local frames are superior to the first two, because their method eliminates bookkeeping and error accumulation and provides complete inheritance of rotation matrices. Later on, Choi [10] improved the efficiency of the simple rotations method by using a series of consecutive operations, including translations and rotations, to update dihedral angles, and she concluded that the improved simple rotations are as efficient as the atom-group local frames.

Due to the inherent nature of Z-matrix coordinates, in which the definition of coordinates of one atom depends on coordinates of another atom that has already been defined, a systematic generation of conformations based on Z-Matrix coordinates encounters problems. An example, which clearly shows the problem when using Z-Matrix coordinates to update a dihedral angle and create new conformers, is given below. In this example, an intuitively obvious Z-matrix is given for a conformer of the canonical tautomer of glycine (See Figure 4.1 for the Z-matrix and for the atomic labels).

If one wants to generate a new conformer of glycine by updating dihedral angles defined by a rotational bond between atoms 1 and 2, one has three choices, because there are three dihedral angles defined by this rotational bond: “dih5”, “dih7”, and “dih8”. Supposing the initial geometry has a C_s symmetry, the dihedral angles “dih7” and “dih8” are equivalent. For this reason, only “dih5” and “dih7” are considered here. If one changes these two dihedral angles by a certain amount (e.g. -60˚ or +60˚), one produces several new conformers, and two of them are shown in Figure 4.1. One can see that both trial structures are chemically meaningless. For example, in the first case we updated “dih5” by -60˚, in which only the positions of atoms 5, 6 and 9 were changed, and it creates a short contact between atoms 5 and 7 and an improper hybridization of atom 1. This is exactly what happens when one manipulates the glycine molecules using this Z-matirx and the commonly used and publically accessible pre- and post-processing program Molden [13]. The problem could be alleviated if atoms 7 and 8 were rotated together with atoms 5, 6 and 9, but it requires special attention from a

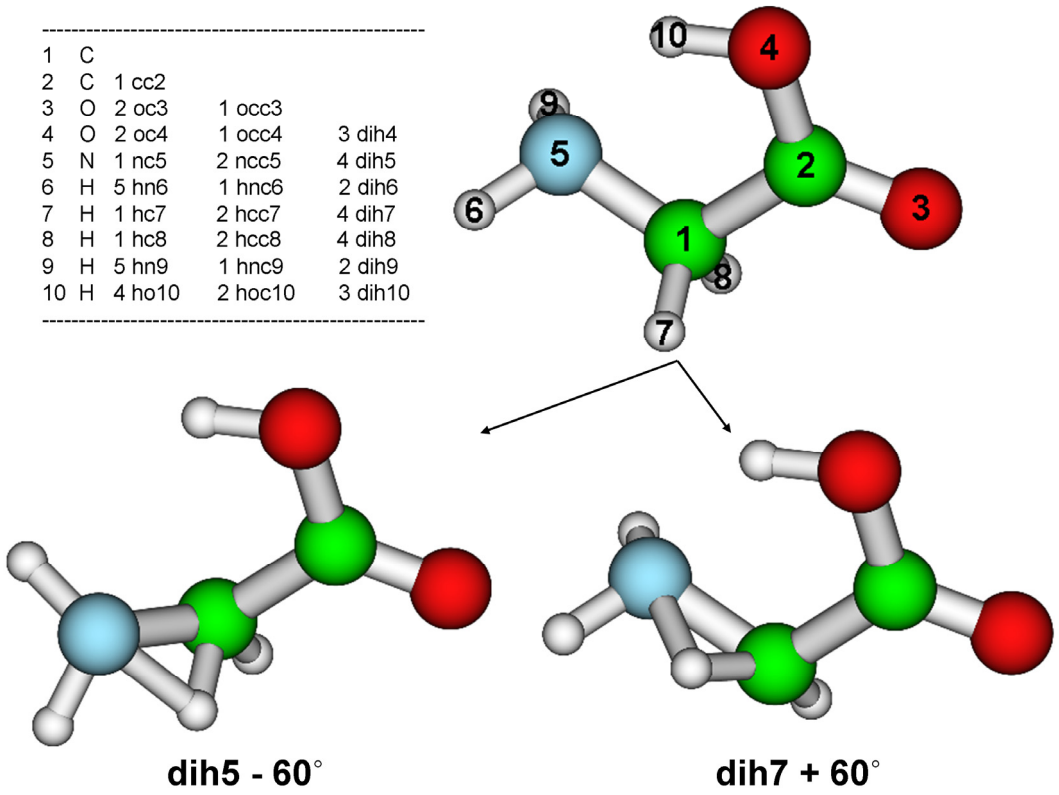


Figure 4.1: Updating molecular geometry based on the Z-Matrix coordinates. Top left – the Z-matrix. Top-right – the initial structure. Bottom left and right – unchemical molecular structures resulting from increments of dih5 and dih7 by -60° and +60°, respectively. Color code: white-hydrogen, blue-nitrogen, green-carbon, red-oxygen.

computational chemist. Alternatively, one can reorder atoms in the coordinate file to avoid this clash, but the approach would require different ordering of atoms for rotations along different dihedral angles. For glycine, this problem is straightforward to resolve, but it becomes serious and difficult to handle in systems that have many large side chains. A decisive solution of this problem would be to perform a concerted movement of all fragments that are connected to the central chemical bond of the dihedral angle under consideration. This approach will be outlined in the next section.

4.2 Methods

We developed a tool, SSC, denoting Systematic Screening of Conformers [14].

SSC is an automatic conformation generator, which we use for systematic searches of conformational spaces of molecules. The user is asked to provide the Cartesian coordinates of a molecule (path of the *.xyz file). To obtain a library of conformations, the user needs to specify which dihedral angles should be activated (the two atoms involved in a rotational bond), and what is the number n_{rot} of rotamers to be considered for this dihedral angle. The value of n_{rot} implies that this dihedral angle will be increased with $360^{\circ}/n_{rot}$ increments. This is all what is needed from a user. Next, SSC will generate possible conformations for each rotational bond recursively, until all specified rotational bonds have been scanned. As an output, geometries (in the form of separate *.xyz files) of all considered conformers will be produced.

SSC can also prepare Gaussian [15] input files for these geometries. To do this, the user will be asked to provide information about the "Route Section" for a Gaussian job, the charge and multiplicity of the molecule. Then Gaussian input files for all geometries available in the library of conformers will be created. Finally, once calculations based on these input files finish, information about converged energies and geometries can be extracted from the output files with the Gaussian Output Tools (GOT) [16], and the global minimum structure as well as the most stable local minima structures will be identified. Further calculations are typically performed at a higher level of theory for the most promising conformers to determine their relative stability and physicochemical properties.

In the current contribution, we update dihedral angles in a step-by-step procedure. The total number of created conformers, n_{total}, is a product of n_{rot} values for all activated bonds:

$$n_{total} = \prod_{i}^{active\ bonds} n_{rot,i} \,. \tag{4.1}$$

The value of n_{total} increases rapidly with the size of the molecule. For example, a molecule with 6 rotational bonds, each probed with 360°/6 increments, would have 46,656 conformers. The value of n_{total} dictates which theoretical model is practical when evaluating the level of fitness of each member of the library. In the case of 2'-deoxycytidine discussed in the next section we could afford to screen the library with quite an accurate electronic structure method, B3LYP/6-31G*. In the case of more complicated molecules the initial screening might be performed using approximate force fields.

Details on how SSC updates a dihedral angle and generates new conformations, is given below. A flowchart, which explains how SSC works, is given in Figure 4.2. SSC reads in the Cartesian coordinates file (*.xyz) provided by the user, a list of rotational bonds that will be active in the conformational search, and the corresponding n_{rot} values. For example, the value of n_{rot} for a rotational bond determined by atoms 1 and 2 (see Figure 4.1) is 6.

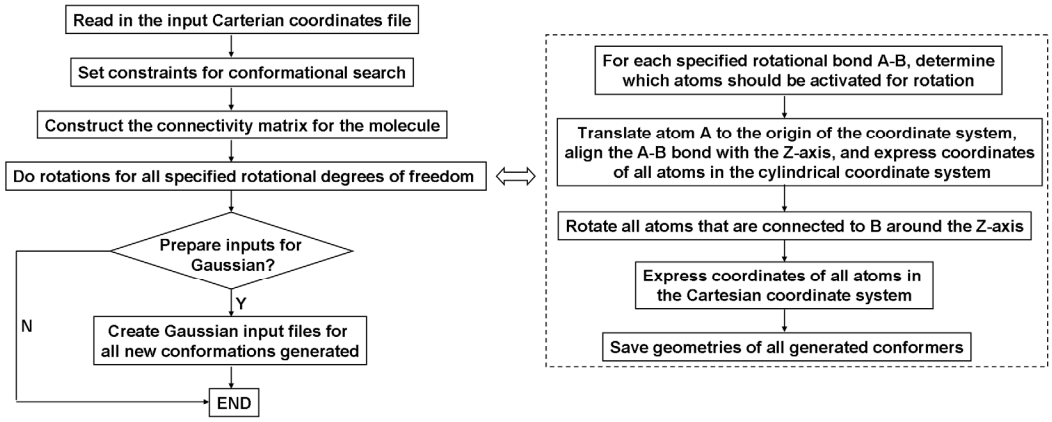


Figure 4.2: Basic working procedure of SSC.

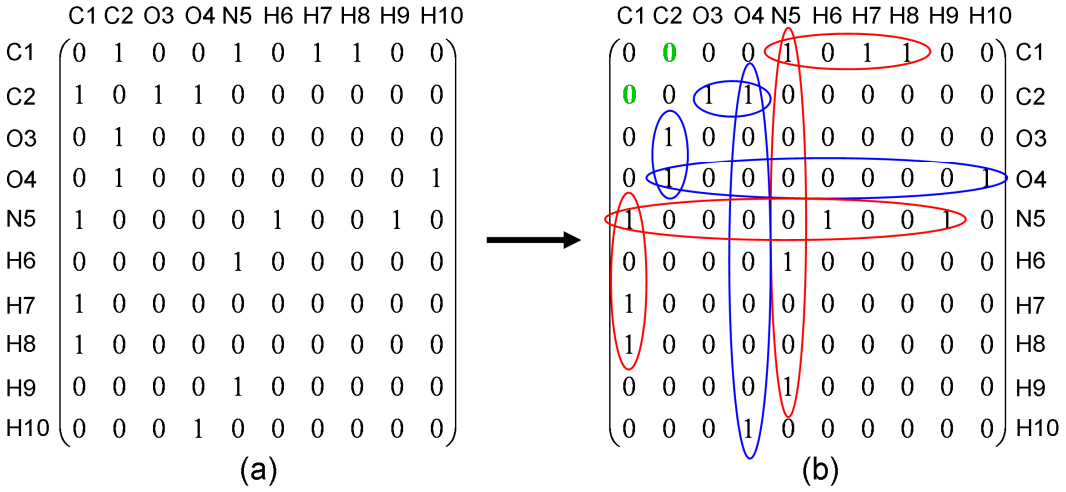


Figure 4.3: The connectivity matrix of the glycine molecule: (a) the original one, (b) the one with atoms C1 and C2 disconnected, in which the modified elements are marked in green, and the red and blue markings indicate molecular fragments connected to C1 and C2, respectively.

Next, a connectivity matrix is constructed for the molecule. Let us take glycine as an example (see Figure 4.1, in which the atom numbering is consistent with the

*.xyz file, which defines the geometry). The connectivity matrix, see Figure 4.3a, is automatically generated in such a way that if two atoms are connected with a chemical bond, we set the related matrix element to 1, or if not, the element is set to 0. The diagonal elements of the matrix are all set to 0. To decide whether two atoms M and N are connected we use a criterion adopted by the Cambridge Structural Database [17]

$$R_{\mathrm{cov}}(M)+R_{\mathrm{cov}}(N)-t \le R_0(MN) \le R_{\mathrm{cov}}(M)+R_{\mathrm{cov}}(N)+t\,, \qquad (4.2)$$

in which $R_0(MN)$ denotes the actual distance between atoms M and N, R_{cov} denotes the covalent radii, as recommended by Cordero et al. [18], and t denotes a tolerance which was set to 0.4 Å. If the value of $R_0(MN)$ fulfills the condition (4.2) then M and N are connected.

The left-most part of (4.2) is also used to identify potential clashes that might develop between various parts of the molecule in the course of systematic rotations. If the distance between any two atoms M and N is smaller than $R_{\mathrm{cov}}(M)+R_{\mathrm{cov}}(N)-t$ then the user is notified about a potential clash and further input is required to decide whether this conformer should be abandoned or not.

Our approach to create conformers that differ by rotations around the A-B bond is the following. First, we create a temporary connectivity matrix based upon the original matrix, but we will disconnect atoms A and B. For example, we show in Figure 4.3b a connectivity matrix for glycine with atoms 1 and 2 disconnected (the modified elements are marked in green). Next, we create two vectors which have at most L elements (where L is the total number of atoms in the molecule). The first vector contains labels of atom A and all these atoms that are directly or indirectly connected to it. The list of these atoms is determined from the temporary connectivity matrix (see red markings in Figure 4.3b). Analogously, the second vector contains labels of atom B and all these atoms that are directly or indirectly connected to it (see blue markings in Figure 4.3b). For example, in the case of rotations around the bond 1-2 of glycine, one produces two vectors [1, 7, 8, 5, 6, 9] and [2, 3, 4, 10], which characterize two branches of glycine, one connected to atom 1 and another connected to atom 2. The software will recognize which branch of the molecule has fewer atoms, and this branch will be rotated to generate new conformations.

Let us assume that the branch connected to atom B is smaller, and we will rotate this branch around the A-B bond. Next steps of our procedure are shown in Figure

4.4. First we translate the whole molecule, such that atom A coincides with the origin of the coordinate system (see Figure 4.4b). Next, we rotate the molecule, such that atom B ends up on the Z-axis, while atom A remains at the origin of the coordinate system (see Figure 4.4c). Now, we use cylindrical coordinates to rotate the B branch of the molecule, while the A branch remains intact. This approach eliminates potential "clashes" between different parts of the B branch, see Figure 4.1. Other potential clashes are handled using the $R_{\mathrm{cov}}(M)+R_{\mathrm{cov}}(N)-t$ threshold discussed earlier. After this rotation, we convert cylindrical coordinates back to Cartesian coordinates and print them out for each conformation into a separate *.xyz file.

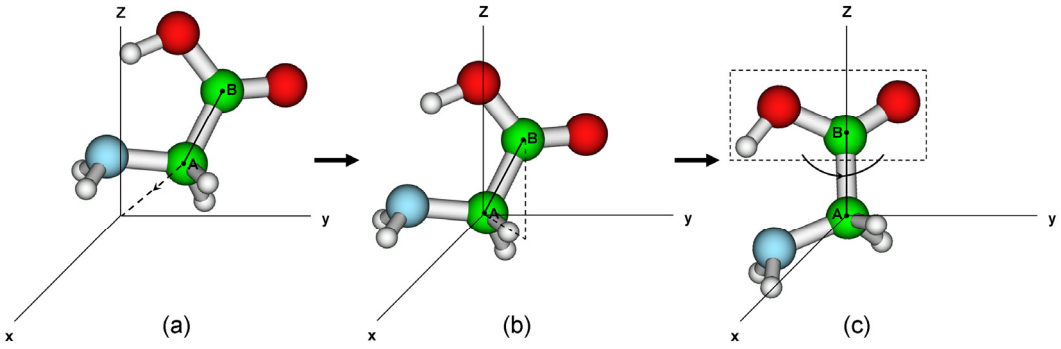


Figure 4.4: Alignment of the A-B rotational bond with the Z-axis.

Finally, we describe our implementation of operations illustrated in Figure 4.4. These steps are illustrated in Figure 4.5, in which all labels and lines in the cylindrical coordinate system are in red color, while others in black color belong to the Cartesian coordinate system. First we translate the whole molecule in such a way that the A atom ends up at the origin of the coordinate system (see Figure 4.5a). From this point we perform two operations in two different cylindrical coordinate systems. First, with Z selected for the longitudinal rotational axis, we rotate the molecule until the rotational bond A-B is in the XZ plane, thus perpendicular to the Y-axis (see Figure 4.5b, $\phi=0$). Next, we convert atomic coordinates to Cartesian coordinates and we define another cylindrical coordinate system with Y as the longitudinal axis. Finally, we rotate the molecule around this axis until ϕ' equals 0 (see Figure 4.5c). We accomplished the requested orientation of the molecule, with the A-B bond along the Z-axis. We return to the original cylindrical coordinates, with Z as the longitudinal axis, and we are ready to do rotations of the B branch (see Figure 4.5d) with predefined increments, $360^{\circ}/n_{rot}$. In the end, we convert the cylindrical coordinates back to the Cartesian

coordinates and print out the Cartesian coordinates of different conformations into separate *.xyz files.

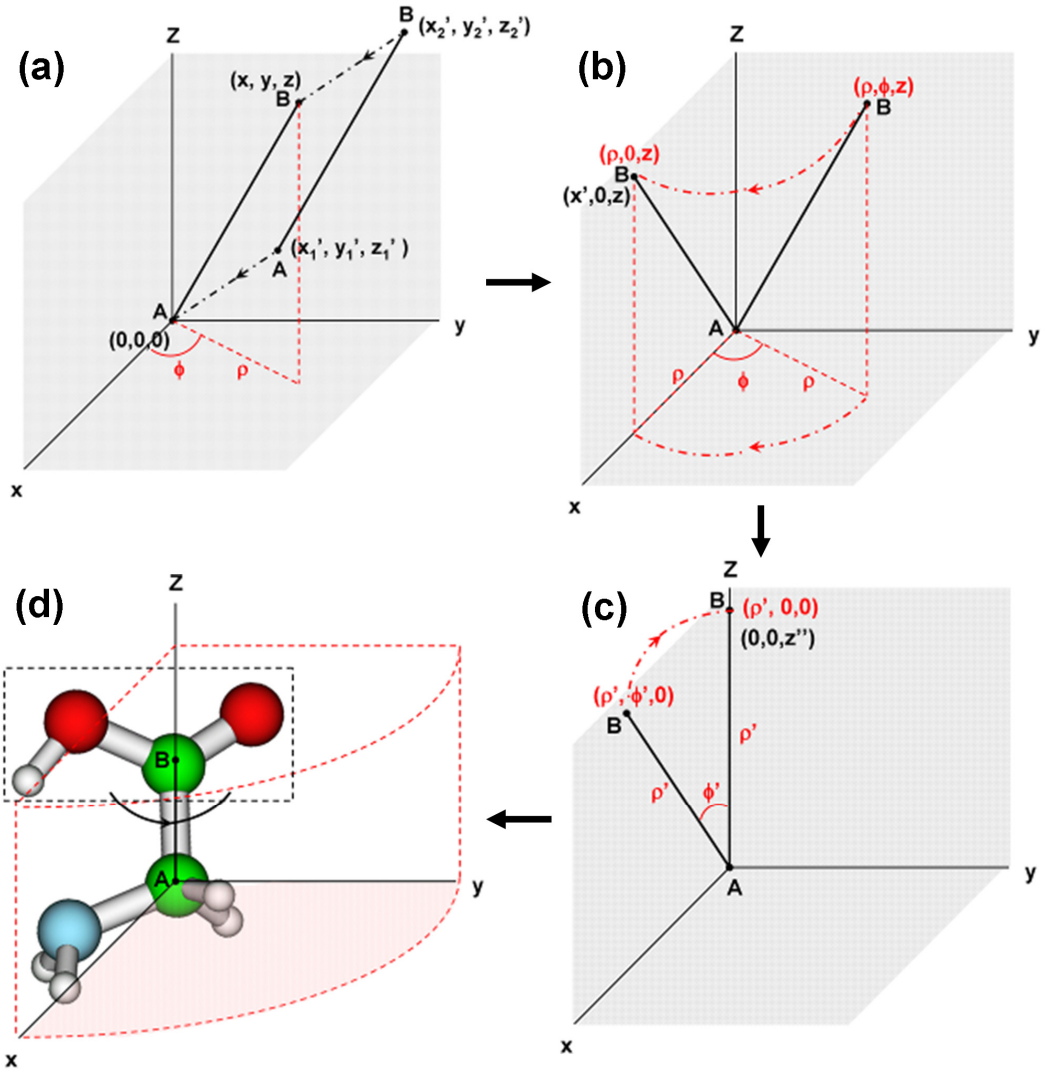


Figure 4.5: A translation and two rotations, which align the rotational bond A-B with the Z-axis.

We do not expect the outcomes of our method to be dependent on the structure of the initial conformer, providing all rotational bonds are sampled and the n_{rot} values are properly selected. One should keep in mind that all initial structures in the library are optimized at the screening stage, compensating for differences in internal geometry parameters of different initial conformers [19, 20].

The MP2 [21] optimizations of the neutral species reported in this study were

performed with NWChem 5.0 [22], and the B3LYP [23, 24] calculations were performed with Gaussian 03 [15].

4.3 Results

4.3.1 Conformational Search of Deoxycytidine

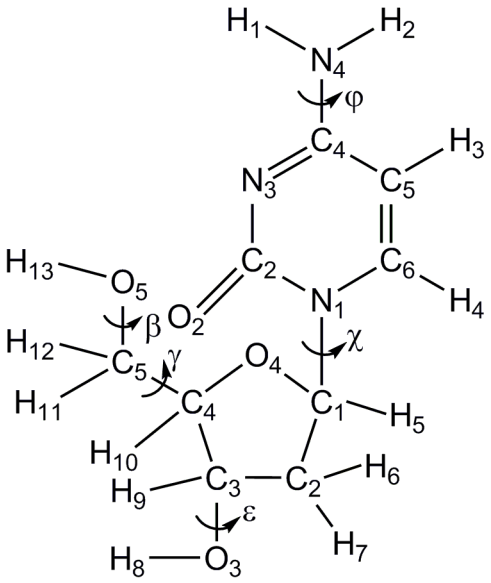


Figure 4.6: Molecular structure of 2'-deoxycytidine.

As an example, SSC has been used to explore the conformational space of 2'-deoxycytidine (dC), a nucleoside consisting of cytosine and deoxyribose. The molecular framework of deoxycytidine is shown in Figure 4.6. It has been reported that in natural nucleic acid helices, the *anti* combination of the base and the sugar is dominant [25]. In a theoretical study performed by Hocquet et al. [26], the authors found that for all eight isolated nucleoside molecules studied, the *anti* conformers were more stable than the *syn* conformers, and for deoxycytidine, the *anti* conformer was more stable than the *syn* conformer by about 5 kcal/mol. In the current contribution, both the *anti* and *syn* combinations of the base and the sugar unit were considered, through rotation of the base along the glycosidic bond by a proper increment. The numbering scheme for all atoms is given in Figure 4.6. In the following discussion, we will add apostrophes on the atomic numbers of all the heavy atoms in the sugar unit to distinguish them from those of the base. All rotational dihedral angles, which are considered in our systematic conformational

search, are also labeled in Figure 4.6. The n_{rot} values for the dihedral angles β, γ, ε, χ and φ considered in our work are 3, 6, 3, 6 and 2, respectively. It should be pointed out here that all previous studies indicated that the amino group of the base is nearly planar [27], and it is far from other parts of the molecule. It implies that any change of the dihedral angle φ has little effect on other degrees of freedom, hence it is reasonable to neglect this degree of freedom at the initial stage of our conformational search. In addition, the puckering of C2' or C3' in the sugar, namely the relative stability of the C2'-endo and the C3'-endo conformations, is still controversial [28]. Obviously, this degree of freedom could not be properly dealt with using the SSC algorithm. Thus, we decided to use a planar sugar conformation in the initial guess structures and let the optimization procedure to distort it according to actual forces acting on atoms, though it might require more time to complete the optimization process. This procedure led in some cases to the C2'-endo and in others to the C3'-endo structures. For the five most stable structures identified in this way we explored whether an opposite puckering of the sugar ring would lead to a more stable structure.

Based on these assumptions, our systematic search of the conformational space of neutral dC is performed in the following way. First, a library of conformers related to the dihedrals angles β, γ, ε and χ is generated by SSC, and these initial structures were optimized at the B3LYP/6-31G* level of theory [15]. The total number of trial structures in the library was 324, but only 66 distinct conformers were identified at this stage (we used an energy difference of 10^{-5} Hartree plus visual evaluation to exclude equivalent minimum energy structures). Secondly, all these 66 conformers were further optimized at the B3LYP/6-31++G** level of theory. Finally, vibrational frequency analysis was performed at the B3LYP/6-31++G** level for the 30 most stable structures. These structures were further optimized at the MP2/aug-cc-pVDZ level to take intra-molecular dispersion interactions into account and to refine the relative stability of the conformers. For the five most stable structures, an opposite puckering of the sugar ring was also considered. Such a multi-step conformational search strategy is attractive because the most promising conformers of dC are characterized at a relatively high level of theory, while computational resources are not wasted on characterization of all possible conformers at the MP2/aug-cc-pVDZ level. Similar strategies were also reported in previous theoretical studies of the structures of peptides [29, 30].

The eight most stable structures we found are displayed in Figure 4.7 and the B3LYP/6-31++G** and MP2/aug-cc-pVDZ relative energies of these eight conformers, with geometries optimized at the corresponding levels of theory, are summarized in Table 4.1. The B3LYP results corrected for zero-point vibration energy and enthalpy and free energy terms, obtained in rigid rotor/harmonic oscillator approximation for T=298 K and p=1 atm, are also listed. The labeling letter N refers to the neutral structure and the numeral following the letter indicates the stability of the neutral species ordered according to the ascending MP2/aug-cc-pVDZ electronic energy.

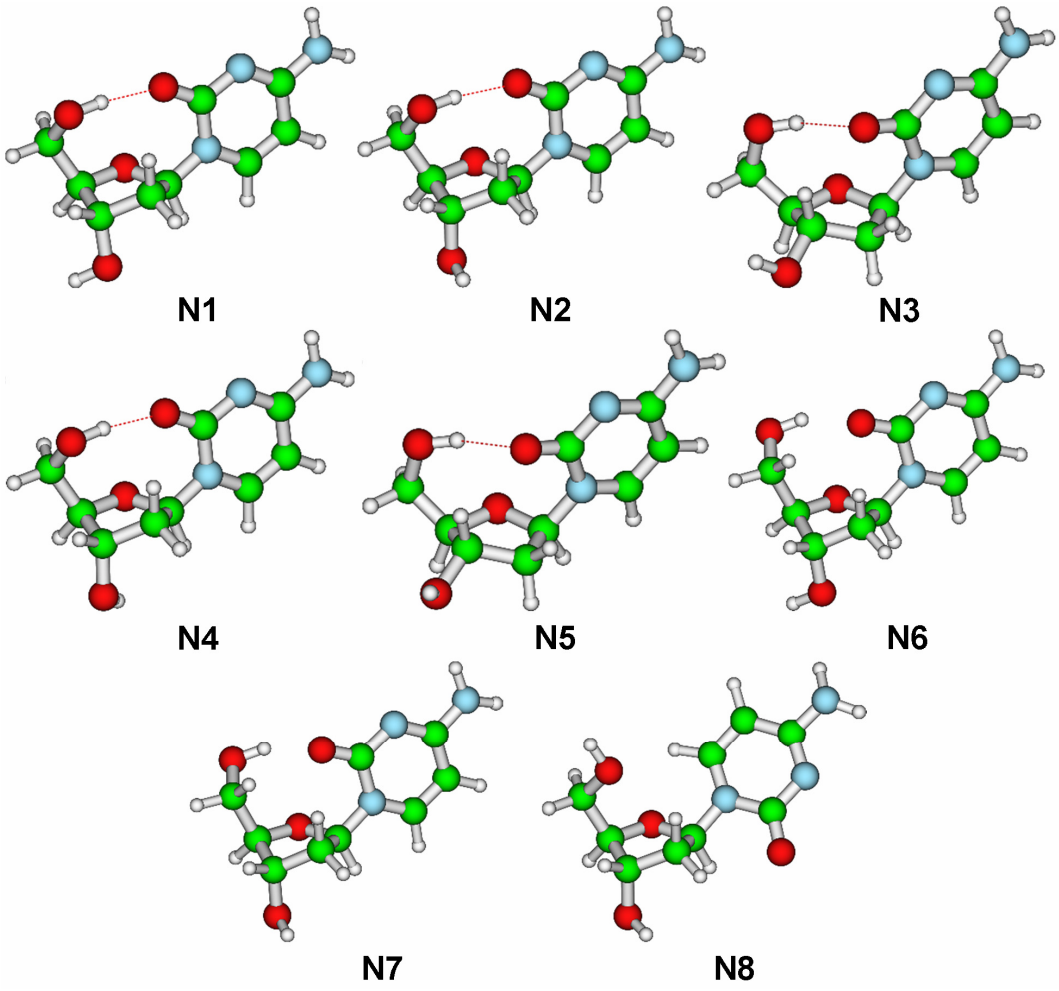


Figure 4.7: MP2 geometries for the eight most stable conformers of 2'-deoxycytidine. The dotted lines indicate the intramolecular hydrogen bonds identified by Molden with default settings.

As was mentioned earlier, our procedure leaves some uncertainty about the

puckering of the sugar ring. For this reason we manually changed the sugar ring puckering for the conformers N1-N5, e.g., if the converged structure was C2'-endo we considered its C3'-endo counterpart, and vice versa. The geometry optimizations of these five structures converged to the structures which already existed in our original set. Some higher energy conformers might favor a puckering of the sugar ring opposite to the one identified by us. In view of small energy differences between oppositely puckered structures, we do not expect these unidentified-yet structures to become competitive with the most stable structures identified so far.

structure	E^{B3LYP}	E^{MP2}	E^{B3LYP} $+\Delta E_{0,vib}$	E^{B3LYP} $+\Delta H_{298,corr}$	E^{B3LYP} $+\Delta G_{298,corr}$
N1	0.000	0.000	0.000	0.000	0.000
N2	-0.003	0.109	0.053	0.021	0.115
N3	1.244	2.009	1.053	1.223	0.804
N4	1.808	2.049	1.689	1.781	1.487
N5	1.349	2.097	1.059	1.290	0.661
N6	3.308	2.733	3.299	3.362	3.305
N7	3.953	3.522	3.900	3.972	3.833
N8	2.983	3.716	2.473	2.808	1.290

Table 4.1: Relative MP2 and B3LYP energies (in kcal/mol) of the most stable conformers of neutral 2'-deoxycytidine. The zero-point vibrational corrections and thermal contributions to enthalpy and free energy are included at the B3LYP level.

The results reported in Tables 4.1 demonstrate the critical importance of systematic conformational searches. Our most important finding is that the neutral dC favors the *syn* conformation in the gas phase, rather than the *anti* conformation, which is common in solid phases [31] and in DNA [32]. Indeed the seven most stable conformers of dC identified by us are *syn*, and the first encountered *anti* structure, N8, is less stable than N1 by 3.6 kcal/mol at the MP2/aug-cc-pVDZ level of theory. This finding is important because many computational studies of dC in the gas phase select an *anti* conformer, e.g. a study on the intramolecular hydrogen bonding in dC [33], a recent study on anion interactions of dC [34], or an interpretation of photoelectron spectra of anionic nucleosides [35].

The formation of intramolecular hydrogen bonds is very important for the relative stability of various conformers of biomolecules [36]. The results reported in Figure 4.7 and Tables 4.1 unravel that the seven most stable neutral conformers are characterized by a hydrogen bond between C2-O2 and O5'-H13, which is plausible for the *syn* conformers. This kind of intramolecular hydrogen bond is impossible for the *anti* conformers, e.g., the conformer N8 supports only a weak hydrogen bond formed between C6-H4 and C5'-O5'. We also found out that the conformers N1 and N2 have very similar molecular structures and stability. They differ only by orientation of a hydroxyl group connected to the C3' atom of the sugar ring.

From Figure 4.7, we can see that the conformers N1 and N3 differ by the puckering of the sugar ring, with N1 being C2'-endo and N3 being C3'-endo, and they differ in stability by 2.0 kcal/mol, see Table 4.1. Thus our results strongly favor the C2'-endo conformation to be dominant in the gas phase, while earlier B3LYP results [26] favored the C3'-endo conformation by 0.53 kcal/mol. Further physicochemical consequences of findings resulting from our systematic search will be discussed in our future report.

4.3.2 Structures of Methyl p-Dimethylaminobenzoate and Its Two Ortho Derivatives

My presentations of the SSC software were welcomed by experimentalists. One of the projects was done jointly with physicists from the University of Gdansk [37]. They were interested in neutral, protonated and diprotonated methyl p-dimethylaminobenzoate and its two ortho derivatives (a total of 9 structures, see Figure 4.8). We used the SSC software to generate a library of conformers for each of these molecules. For the first four molecules (I, IA-B, and II) all rotational degrees of freedom were considered at this stage. Next, we recognized that the contribution from methyl groups to the overall stability of the molecules is relatively small and can be decoupled from contributions associated with other rotational degrees of freedom. Thus for the remaining five molecules (IIA-B, III, IIIA-B) the rotational degrees of freedom of the methyl groups were not considered when creating initial libraries of conformers with SSC. The stability of each conformer was initially determined in the course of geometry optimization with the B3LYP exchange-correlation functional [23, 24] and the 6-31G* basis set [38], including geometrical relaxation of all methyl groups. A typical library

created with SSC contained a few hundreds of conformers and the B3LYP/6-31G* optimization led to a few tens of distinct local minima, see Table 4.2. For these local minima all meaningful conformers derived from rotations of the methyl groups were considered. At the second stage, the structures of the most stable conformers of each molecule were further refined at the B3LYP/6-311G**

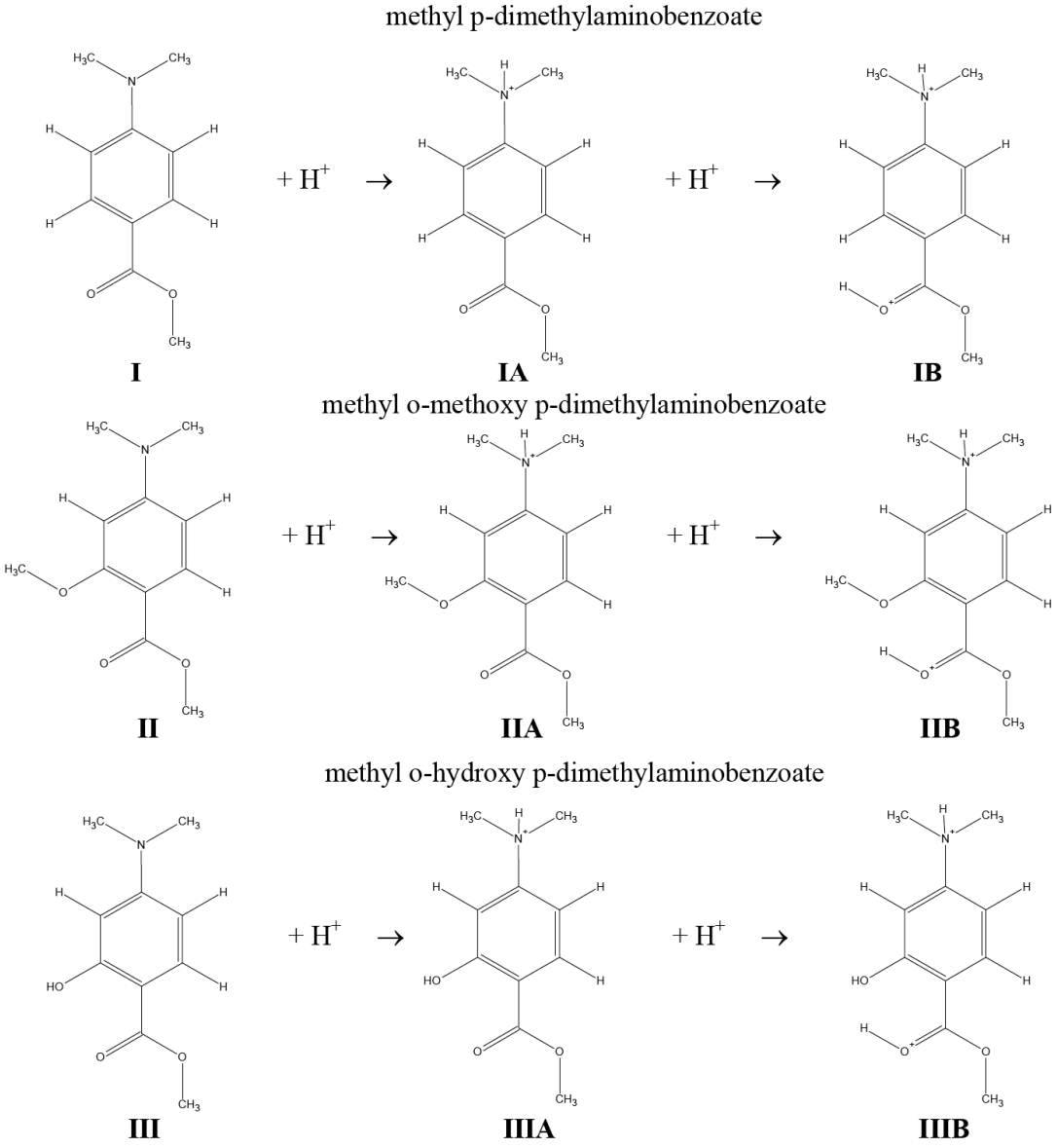


Figure 4.8: Structures of methyl p-dimethylaminobenzoate and its o-methoxy and o-hydroxy derivatives (first column) and their monoprotonated (second column) and diprotonated (third column) forms.

level [39]. The number n of conformers included at the second stage of search was

selected in this way that the (n+1)-th conformer was less stable than the n-th conformer by at least 2 kcal/mol. The details of our two-step procedure are summarized in Table 4.2. For example, for the diprotonated methyl o-hydroxy p-dimethylaminobenzoate (IIIB) the SSC software created 218 initial conformers and the B3LYP/6-31G* prescreening step led to 30 distinct conformers. The two most stable conformers spanned a narrow range of 0.2 kcal/mol and the third conformer was less stable by 7.6 kcal/mol. Hence, only the first two conformers were considered at the second stage of optimization. The effect of solvation with THF was studied by performing additional optimizations within the polarized continuum model (PCM) [40] at the B3LYP/6-311G** level (with ε=7.58). Excited singlet electronic states for the most stable conformers were calculated using the time-dependent B3LYP/6-311G** method [41], with the effect of THF included by using the PCM model. These results are reported in Ref. 42.

structure	First step (B3LYP/6-31G*)		Second step (B3LYP/6-311G**)	
	optimized structures	local minima	optimized structures	local minima
I	432	8	7	1
IA	432	12	4	2
IB	864	15	11	6
II	2592	27	6	3
IIA	179	17	9	8
IIB	329	20	2	2
III	116	12	4	2
IIIA	112	16	5	4
IIIB	218	30	2	2

Table 4.2: Number of optimized structures and local minima for each step of conformation searches.

4.4 Conclusions

We have presented the Systematic Screening of Conformers (SSC) tool, which generates an initial library of conformers for a systematic screening of conformational space of a molecule. After providing an initial structure, the user defines which dihedral angles should be explored and with which increments. By

rotating a whole fragment of the molecule, which is adjacent to the activated bond, unchemical hybridizations and some "clashes" between neighboring groups are avoided. Each member of the library is prescreened at a predefined level of theory and the most promising conformers are identified. The tool offers a significant saving of human time, which otherwise would be spent on generation of hundreds or thousands of initial structures and corresponding input files. We envisage usage of this tool by both computational chemists and those experimentalists, who perform supporting computational studies.

The tool could be used to identify the most stable conformer of a molecule. The method scales steeply with the molecular size. With the nowadays typical computational resources it is applicable to systems with up to 5-6 rotational bonds probed with $2<n_{rot}<6$, if density functional electronic structure methods are used to screen the library. Larger molecules could be studied if initial screening was performed using more approximate force fields. The SSC approach is not limited to molecules in the gas phase. After inclusion of solvent effects at the level of polarizable continuum model, it is applicable to molecules in any solvent.

An application of our tool has been demonstrated for a nucleoside, 2'-deoxycytidine. We demonstrated that *syn* conformers should be dominant in the gas phase rather than *anti* conformers, which are favored in crystalline structures and in DNA. The dominance of *syn* conformers is dictated by an intramolecular hydrogen bond between C2-O2 and O5'-H13. Our results also suggest that the dominant puckering of the sugar ring in the gas phase should be C2'-endo, rather than C3'-endo.

It should be pointed out that SSC is applicable to chain-like molecules, with complicated side chains. We pointed out difficulties encountered when dealing with puckered ring-like fragments, e.g., a sugar unit in nucleosides. For molecules with overall closed ring structures, such as cyclodextrin, SSC would not work, because for a given rotational bond A-B, the molecule could not be divided into two decoupled parts associated with atoms A and B. SSC could not recognize which part of the molecule should be rotated. If such a situation actually happens, SSC will be able to recognize the problem and alarm the user. There might be also problems for molecules with symmetric intramolecular hydrogen bonds, because SSC will assume that the central hydrogen atom is connected to both acceptors, and thus forms a ring. Solutions to these problems will be considered in our future

work. In the near future, the current tool, SSC, will be integrated with TauTGen,[1] a tool for formation of libraries of tautomers, so that the structure of a molecule could be determined automatically taking into account various conformers and tautomers.

References:

[1] Haranczyk, M.; Gutowski, M., Quantum Mechanical Energy-Based Screening of Combinatorially Generated Library of Tautomers. Tautgen: A Tautomer Generator Program. *Journal of Chemical Information and Modeling* **2006,** 47, 686-694.

[2] Haranczyk, M.; Gutowski, M., Combinatorial–Computational–Chemoinformatics (C3) Approach to Finding and Analyzing Low-Energy Tautomers. *Journal of Computer-Aided Molecular Design* **2010,** 24, 627-638.

[3] Haranczyk, M.; Puzyn, T.; Sadowski, P., Congener - a Tool for Modeling of the Congeneric Sets of Environmental Pollutants. *QSAR & Combinatorial Science* **2008,** 27, 826-833.

[4] Leach, A. R., *Molecular Modelling: Principles and Applications*. 2nd ed.; Prentice Hall: Harlow, 2001.

[5] Atkins, P. W.; De Paula, J., *Atkins' Physical Chemistry*. 8th ed.; Oxford University Press: Oxford, 2006; p 1064.

[6] Wales, D. J.; Doye, J. P. K., Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *Journal of Physical Chemistry A* **1997,** 101, 5111-5116.

[7] Wales, D. J., *Energy Landscapes*. Cambridge University Press: Cambridge, 2003; p 681.

[8] Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T., Conformer Generation with Omega: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling* **2010,** 50, 572-584.

[9] Li, J.; Ehlers, T.; Sutter, J.; Varma-O'Brien, S.; Kirchmair, J., Caesar: A New Conformer Generation Algorithm Based on Recursive Buildup and Local Rotational Symmetry Consideration. *Journal of Chemical Information and Modeling* **2007,** 47, 1923-1932.

[10] Choi, V., On Updating Torsion Angles of Molecular Conformations. *Journal of Chemical Information and Modeling* **2006,** 46, 438-444.

[11] Steven, M. L.; Paul, W. F.; Lydia, E. K.; Jean-Claude, L., A Randomized Kinematics-Based Approach to Pharmacophore-Constrained

Conformational Search and Database Screening. *Journal of Computational Chemistry* **2000,** 21, 731-747.

[12] Zhang, M.; Kavraki, L. E., A New Method for Fast and Accurate Derivation of Molecular Conformations. *Journal of Chemical Information and Computer Sciences* **2002,** 42, 64-70.

[13] Schaftenaar, G.; Noordik, J. H., Molden: A Pre- and Post-Processing Program for Molecular and Electronic Structures. *Journal of Computer-Aided Molecular Design* **2000,** 14, 123-134.

[14] SSC: A Tool for Constructing Libraries for Systematic Screening of Conformers. http://sscf.sf.net (accessed July 28, 2010).

[15] Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; 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.; Bakken, V.; 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.; and Pople, J. A. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.

[16] GOT: Gaussian output tools. http://gaussot.sf.net (accessed July 28, 2010).

[17] Cambridge Structural Database. http://www.ccdc.cam.ac.uk/products/csd/radii/ (accessed July 29, 2010).

[18] Cordero, B.; Gomez, V.; Platero-Prats, A. E.; Reves, M.; Echeverria, J.; Cremades, E.; Barragan, F.; Alvarez, S., Covalent Radii Revisited. *Dalton Transactions* **2008**, 2832-2838.

[19] Boström, J., Reproducing the Conformations of Protein-Bound Ligands: A Critical Evaluation of Several Popular Conformational Searching Tools. *Journal of Computer-Aided Molecular Design* **2001,** 15, 1137-1152.

[20] Boström, J.; Greenwood, J. R.; Gottfries, J., Assessing the Performance of Omega with Respect to Retrieving Bioactive Conformations. *Journal of Molecular Graphics and Modelling* **2003,** 21, 449-462.

[21] Moller, C.; Plesset, M. S., Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934,** 46, 618-622.

[22] Bylaska, E. J. d. J., W. A.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Apra, E.; Windus, T. L.; Hirata, S.; Hackler, M. T.; Zhao, Y.; Fan, P.-D.; Harrison, R. J.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Auer, A. A.; Nooijen, M.; Brown, E.; Cisneros, G.; Fann, G. I.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J. A.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Pollack, L.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. "NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.0" (2006), Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA.

[23] Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B* **1988,** 37, 785-789.

[24] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *Journal of Physical Chemistry* **1994,** 98, 11623-11627.

[25] Foloppe, N.; MacKerell Jr, A. D., Intrinsic Conformational Properties of Deoxyribonucleosides: Implicated Role for Cytosine in the Equilibrium among the a, B, and Z Forms of DNA. *Biophysical Journal* **1999,** 76, 3206-3218.

[26] Hocquet, A.; Leulliot, N.; Ghomi, M., Ground-State Properties of Nucleic Acid Constituents Studied by Density Functional Calculations. 3. Role of Sugar Puckering and Base Orientation on the Energetics and Geometry of 2'-Deoxyribonucleosides and Ribonucleosides. *Journal of Physical Chemistry B* **2000,** 104, 4560-4568.

[27] Zierkiewicz, W.; Komorowski, L.; Michalska, D.; Cerny, J.; Hobza, P., The Amino Group in Adenine: Mp2 and Ccsd(T) Complete Basis Set Limit Calculations of the Planarization Barrier and Dft/B3lyp Study of the Anharmonic Frequencies of Adenine. *Journal of Physical Chemistry B*

2008, 112, 16734-16740.

[28] Cheatham III, T. E.; Kollman, P. A., Molecular Dynamics Simulation of Nucleic Acids. *Annual Review of Physical Chemistry* **2003,** 51, 435-471.

[29] Toroz, D.; van Mourik, T., The Structure of the Gas-Phase Tyrosine-Glycine Dipeptide. *Molecular Physics* **2006,** 104, 559-570.

[30] Řeha, D.; Valdés, H.; Vondrášek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S., Structure and Ir Spectrum of Phenylalanyl–Glycyl–Glycine Tripetide in the Gas-Phase: Ir/Uv Experiments, Ab Initio Quantum Chemical Calculations, and Molecular Dynamic Simulations. *Chemistry – A European Journal* **2005,** 11, 6803-6817.

[31] Young, D. W.; Wilson, H. R., Crystal and Molecular Structure of 2'-Deoxycytidine. *Acta Crystallographica, Section B: Structural Science* **1975,** B31, 961-965.

[32] Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V.; Kopka, M. L., The Anatomy of a-DNA, B-DNA, and Z-DNA. *Science* **1982,** 216, 475-485.

[33] Hocquet, A., Intramolecular Hydrogen Bonding in 2'-Deoxyribonucleosides: An Aim Topological Study of the Electronic Density. *Physical Chemistry Chemical Physics* **2001,** 3, 3192-3199.

[34] Tehrani, Z. A.; Fattahi, A., Anion Interactions of Cytosine Nucleobase and Its Nucleosides: Detailed View from Dft Study. *Journal of Molecular Structure: Theochem* **2009,** 913, 277-283.

[35] Stokes, S. T.; Li, X.; Grubisic, A.; Ko, Y. J.; Bowen, K. H., Intrinsic Electrophilic Properties of Nucleosides: Photoelectron Spectroscopy of Their Parent Anions. *Journal of Chemical Physics* **2007,** 127, 084321(1)-084321(6).

[36] Ling, S.; Yu, W.; Huang, Z.; Lin, Z.; Haranczyk, M.; Gutowski, M., Gaseous Arginine Conformers and Their Unique Intramolecular Interactions. *Journal of Physical Chemistry A* **2006,** 110, 12282-12291.

[37] Jozefowicz, M.; Aleksiejew, M.; Abramov, A. V.; Ling, S.; Gutowski, M.; Heldt, J.; Heldt, J. R., Influence of Prototropic Reactions on the Absorption and Fluorescence Spectra of Methyl P-Dimethylaminobenzoate and Its Two Ortho Derivatives. *Molecular Physics* **2010,** Submitted.

[38] Hehre, W. J.; Ditchfield, R.; Pople, J. A., Self-Consistent Molecular Orbital Methods. Xii. Further Extensions of Gaussian---Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *Journal of Chemical Physics* **1972,** 56, 2257-2261.

[39] Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A., Self-Consistent Molecular-Orbital Methods .20. Basis Set for Correlated Wave-Functions. *Journal of Chemical Physics* **1980,** 72, 650-654.

[40] Barone, V.; Cossi, M.; Tomasi, J., A New Definition of Cavities for the Computation of Solvation Free Energies by the Polarizable Continuum Model. *Journal of Chemical Physics* **1997,** 107, 3210-3221.

[41] Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J., An Efficient Implementation of Time-Dependent Density-Functional Theory for the Calculation of Excitation Energies of Large Molecules. *Journal of Chemical Physics* **1998,** 109, 8218-8224.

[42] Jozefowicz, M.; Aleksiejew, M.; Abramov, A. V.; Ling, S.; Gutowski, M.; Heldt, J.; Heldt, J. R. "Influence of prototropic reactions on the absorption and fluorescence spectra of methyl p-dimethylaminobenzoate and its two ortho derivatives". Molecular Physics 2010, to be submitted.

Chapter 5

Deformation of Potential Energy Surface

Searches for the most stable molecular conformer are frustrated by energy barriers separating minima on the potential energy surface (PES). We have suggested that the barriers might be suppressed by subtracting selected force field terms from the original PES. The resulting deformed PES can be used in standard molecular dynamics (MD) or Monte Carlo simulations. The MD trajectories on the original and deformed PESs of ethanolamine differ markedly. The former gets stuck in a local minimum basin while the latter moves quickly to the global minimum basin. A concept of generalized simulated annealing protocols has been formulated.

5.1 Introduction

The conformational space of a molecule is spanned by its internal rotational bonds [1]. Specific combinations of intramolecular torsional angles give birth to molecular conformers, which are represented by minima on the molecular potential energy surface (PES). Each conformer has its own (in general different) electronic energy, vibrational energy levels and rotational constants. According to the Boltzmann distribution, conformers with lower energies dominate the overall population, and therefore the molecular properties measured in experiments. The number of local minima on the molecular PES quickly increases with molecular size, and the most common optimization algorithms (driven by calculated forces) typically identify a minimum which is close to the initial structure, rather than the most stable conformer. It is therefore of ultimate importance to develop methods and algorithms for determination of the most stable molecular conformers.

In this report we are primarily interested in PESs determined using reliable electronic structure methods. This is a broad range of methods from semi-empirical models, through density functional theory (DFT) methods, to Møller-Plesset and coupled cluster treatments of electron correlation. These methods are more time consuming than commonly used force fields [2], but they

offer a broad spectrum of accuracy, with the most advanced methods taking into account all practically important chemical and physical interactions and being able to break chemical bonds. The scaling of these methods is, however, quite steep, e.g. N^3-N^4 for DFT, N^5 for the second-order Møller-Plesset method, etc., where N is proportional to the number of basis functions. The applicability of these reliable methods in structural predictions hinges on efficient global minimum search algorithms.

Finding the global minimum on a PES is frustrated by barriers that separate local minima. The most straightforward approaches that ignore barriers, such as scanning the PES with discrete geometrical increments along all $3N$-6(5) internal degrees of freedom or systematic searches focused on rotational bonds only [1, 3, 4] (see Chapters 4 and 7), suffer from combinatorial explosions of required calculations. These brute force methods provide, however, the most reliable benchmarks for all other, perhaps faster but less reliable, methods.

The most common methods for finding the most stable molecular structures are finite temperature Monte Carlo [5] and molecular dynamics (MD) [6] methods as well as genetic algorithms [4, 7] (for detailed introductions of these methods, see Chapter 3). The molecular dynamics method is of particular relevance here, and it is implemented in various simulated annealing algorithms [8]. The main idea is that a simulation initiated from any initial molecular structure and performed at a sufficiently high temperature (T) for a sufficiently long time has the chance to probe the basins of the global minimum and other low energy minima. The performance of the method depends critically on the height of energy barriers separating local minima and on the overall simulation time [9, 10]. One could think that by increasing the simulation temperature one would provide enough energy to overcome any existing barrier. Unfortunately, molecules decompose into smaller fragments in the course of high T simulations when reliable PESs are used [11, 12]. Performing extremely long simulations at a low T is not an attractive option due to significant computational runtimes. In practice, a compromise is made between the simulation temperature and the overall simulation time and various protocols are used that define T as a function of time. Many valuable results were obtained using simulated annealing algorithms, though the identified low energy structure might be a local minimum rather than the global minimum. Again, the failure of the method is associated with the height of the energy barrier that separates the local minimum from the global minimum.

The past efforts to suppress energy barriers on a PES include the diffusion equation method (DEM), which transforms the original PES with multiple minima into a new PES with only one (the global) minimum [13]. Some other methods which deform the original potential energy surface in a different way include the distance scaling method [14], the shift method [15], the Gaussian density annealing method [16, 17], and the so-called ant-lion strategy of changing the range of the potential [18]. While these methods proved to be robust for complex systems, they require an analytical expression for the original PES. Thus, they are not applicable to PESs determined using electronic structure methods. Another successful approach, the basin-hopping method [19], in which the PES is transformed into a collection of interpenetrating staircases, requires a geometrical energy minimization for each stochastically selected molecular structure. This feature disfavors the method when combined with PESs determined using electronic structure methods, though successful applications have been reported [20, 21].

Here we propose another method to eliminate or suppress barriers that separate various conformers on a PES determined by an electronic structure method. The deformed potential energy surface is obtained by subtracting selected force field terms from the original PES. First we demonstrate that typical conformational barriers result from the torsional and van der Waals (vdW) force field terms. Next, we compare constant temperature molecular dynamics trajectories for ethanolamine on the original PES (determined at the DFT level of theory), and on the deformed PES with the torsional and vdW terms subtracted. We demonstrate that the MD trajectory moves quickly to the global minimum basin on the deformed PES, but it gets stuck in a local minimum basin on the original PES. Finally, we discuss the concept of generalized simulated annealing protocols, in which both the temperature and potential energy can change as time progresses.

5.2 Methods

We start from a potential energy function, E_{elec}, defined by a reliable electronic structure method. In the current case we use density functional theory with the B3LYP hybrid exchange-correlation functional [22, 23] and 6-31G* atomic basis sets [24]. The global and local minima were identified using our Systematic

Screening of Conformers (SSC) tool [25]. DFT/B3LYP is a relatively fast electronic structure method, but still prohibitively slow for structural predictions based on long MD trajectories. In parallel to the electronic structure method we use a potential energy E_{FF} defined by a reliable force field (FF). The potential energy in a typical FF is given by the following expression:

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{cross} + E_{vdw} + E_{coul} \quad , \tag{5.1}$$

where E_{str} is the bond stretching energy, E_{bend} is the bond bending energy, E_{tors} is the torsional energy for rotations of molecular fragments around bonds, E_{cross} describes cross terms, e.g., bond-bond angle, and finally E_{vdw} and E_{coul} are the van der Waals (vdW) energy and the Coulomb energy, respectively [26]. There are two advantages resulting from using a FF. First, the time associated with a computation of E_{FF} is negligible in comparison with the time required to compute E_{elec}. Second, Eq. (5.1) offers a dissection of the total potential energy into distinct components with clear chemical interpretations. In this study we will use a very popular FF, AMBER [27], but other FFs might be used as well.

Our approach hinges on an assumption that *a priori* known terms of Eq. (5.1) contribute to the barriers which separate local minima representing various molecular conformers. We define a modified, or deformed, PES as

$$E_M(\mathbf{x}) = E_{elec}(\mathbf{x}) - \sum_i c_i E_{FF,i}(\mathbf{x}) \tag{5.2}$$

in which E_M indicates the modified potential energy, $E_{FF,i}$ is a specific FF term from the right hand side of Eq. (5.1), c_i is a linear coefficient (most typically $c_i = \pm 1.0$), and $\mathbf{x}$ represents a molecular geometry. The modified potential energy E_M is also used to calculate the modified gradient,

$$\mathbf{g}_M(\mathbf{x}) = \nabla E_M(\mathbf{x}), \tag{5.3}$$

which requires the gradients of E_{elec} and all $E_{FF,i}$s involved in Eq. (5.2). This gradient will be used to represent forces acting on atoms moving on the modified PES, as dictated by the conventional MD equations.

The molecular dynamics simulations were performed for ethanolamine, see Figure 5.1, at 200 K and 400 K, with a time step of 0.5 fs, for a total of 20000 steps on both the original and modified PES. The Beeman algorithm [28] was used to predict accelerations, velocities and coordinates of nuclei. The resulting trajectories were illustrated by the root mean square deviations (RMSD) of the positions of the nuclei with respect to a reference structure, which might be a local or the global minimum structure:

$$RMSD(t) = \sqrt{\frac{\sum_{i=1}^{3N}\left[x_i(t) - x_{i,ref}\right]^2}{3N}}, \qquad (5.4)$$

in which $x_i(t)$ represents the i-th geometrical coordinate at simulation time t, and $x_{i,ref}$ represents the i-th geometrical coordinate at the reference structure. The current and reference structures were aligned to resemble each other using the VMD [29] command "measure fit". While advancing a MD trajectory we collect information about the most stable molecular structures and their complete potential energies. After the MD trajectory is completed we use the identified structures as initial geometries for standard geometry optimizations based on calculated forces.

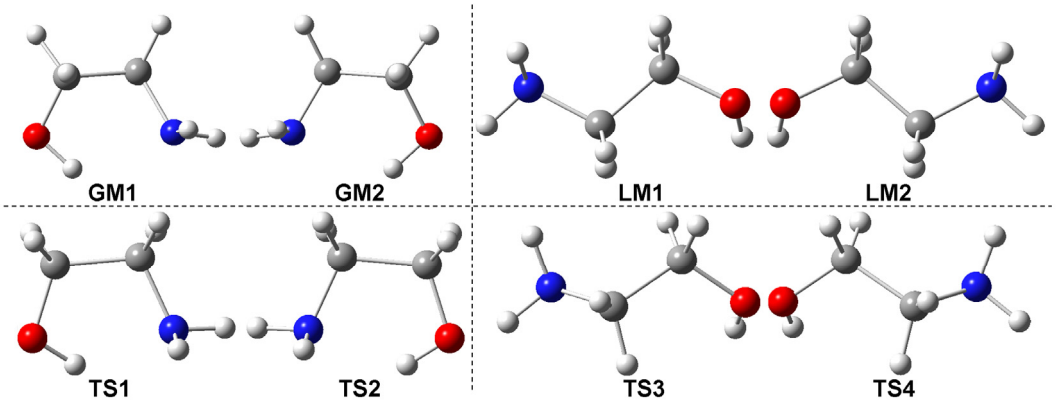
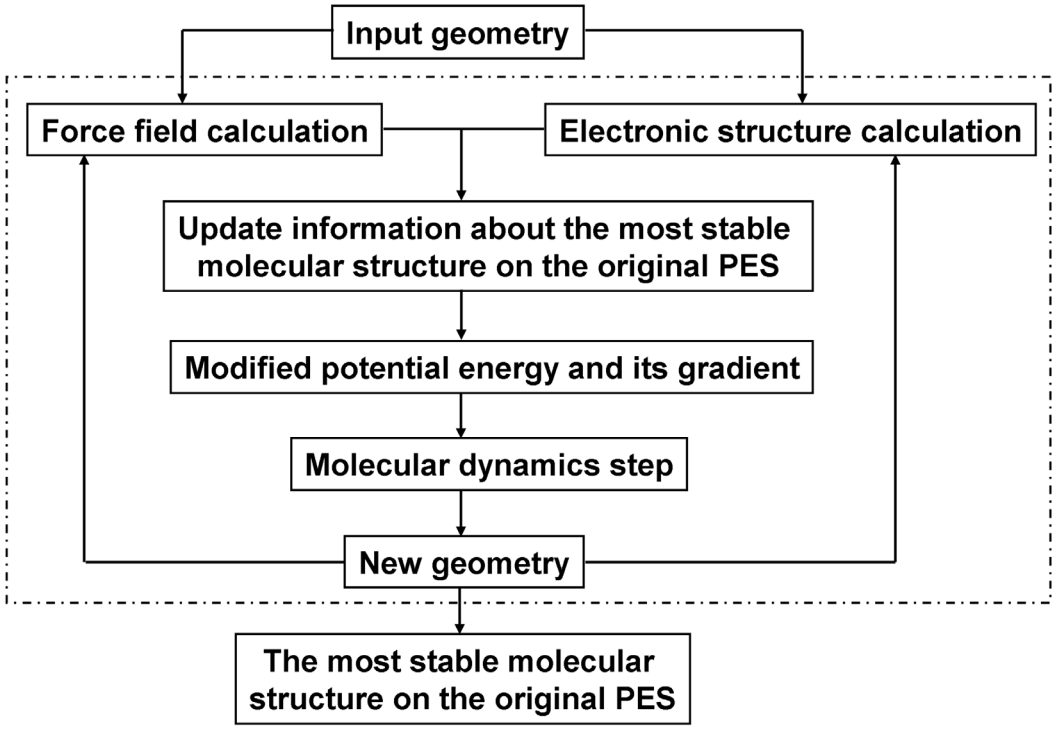


Figure 5.1: Different minima and transition state structures of ethanolamine. Color code: grey – carbon, white – hydrogen, red – oxygen, blue – nitrogen.

All B3LYP calculations were performed with Gaussian 03 (G03) [30]. Molecular mechanics calculations were performed either with G03 or Tinker [31]. An interface has been developed between G03 and Tinker that allowed propagating MD equations on the original and modified PES, see Scheme 5.1. The charge equilibration (QEq) method [32] was used to assign charges on atoms to account for Coulomb energy in the AMBER FF calculations performed with G03. Default atomic charges were used in Tinker.



Scheme 5.1: A flowchart of the interface between an electronic structure method and a force field.

5.3 Results

Ethanolamine has three rotational degrees of freedom and the OH and NH_2 groups can form an intramolecular hydrogen bond, which determines two chiral global minima GM1 and GM2 (see Figure 5.1). In Figure 5.2 we present a potential energy curve for motion around the O-C-C-N dihedral angle. The dihedral angle was increased or decreased, when starting from GM1 and GM2 respectively, with steps of 5˚, and partial geometry optimizations were performed for each fixed value of the angle. Altogether, 72 structures were considered. This procedure unraveled six new stationary points on the PES (Figures 5.1 and 5.2): one pair of chiral local minima (LM1 and LM2) and two pairs of chiral transition states (TS1 and TS2, TS3 and TS4). Subsequent intrinsic reaction coordinate [33] calculations were performed for the fully optimized transition state structures. It was confirmed that both TS1 and TS2 connect GM1 and GM2, TS3 connects LM1 and GM1, and TS4 connects LM2 and GM2. It should be pointed out that there is also a C_s symmetry stationary point (denoted as SP0) between TS1 and TS2 but it was abandoned because of two negative curvatures.

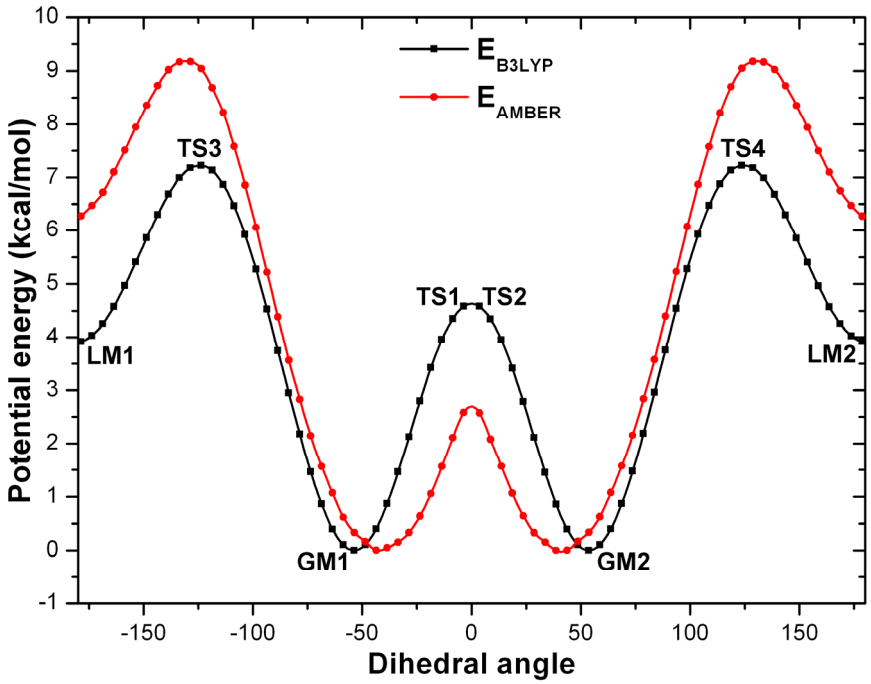


Figure 5.2: Original potential energy curve for ethanolamine as a function of the O-C-C-N dihedral angle. The energies are determined at the B3LYP and AMBER force field levels.

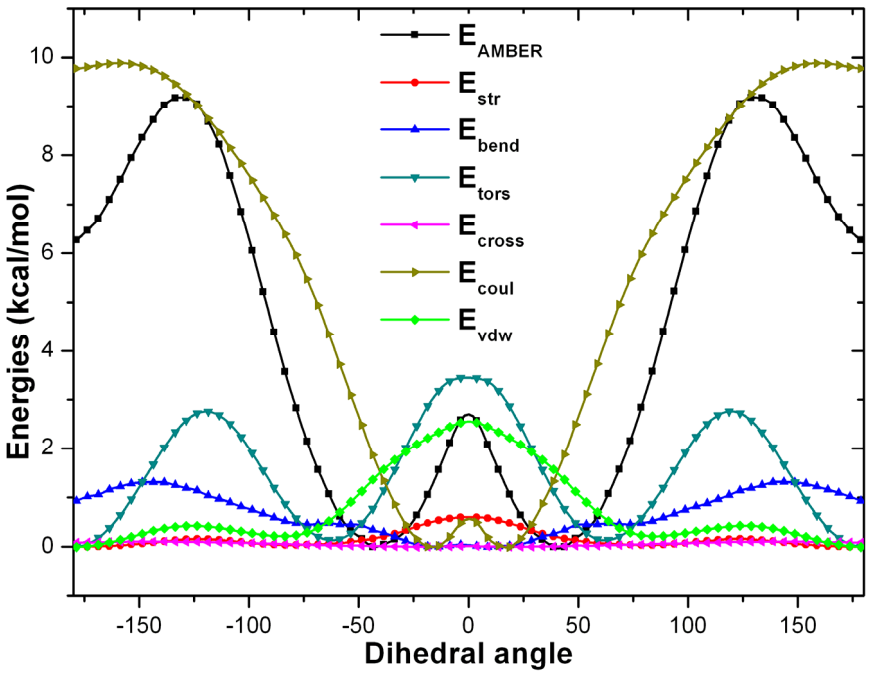


Figure 5.3: The AMBER force field energy and its components as a function of the O-C-C-N dihedral angle.

Based on the 72 partially optimized geometries, we performed single point AMBER FF calculations. A comparison between the B3LYP and AMBER potential energy curves is presented in Figure 5.2 and a dissection of the AMBER energy, given by Eq. (5.1), is illustrated in Figure 5.3. In Figures 5.2 and 5.3 the zero of energy was set to the global minimum of each curve. Figure 5.2 illustrates that the AMBER curve properly reproduces the main features of the B3LYP curve. The angles at which stationary points develop are similar, though some of the barrier heights are inaccurate by ±2-3 kcal/mol.

The results presented in Figure 5.3 illustrate that the barriers on the total potential energy curve result primarily from the torsional and vdW terms, whereas the Coulomb term displays the largest overall variation as a function of the dihedral angle. The last finding is consistent with the dominant role of the intramolecular $OH^{\cdots}NH_2$ hydrogen bond in the GM1 and GM2 structures.

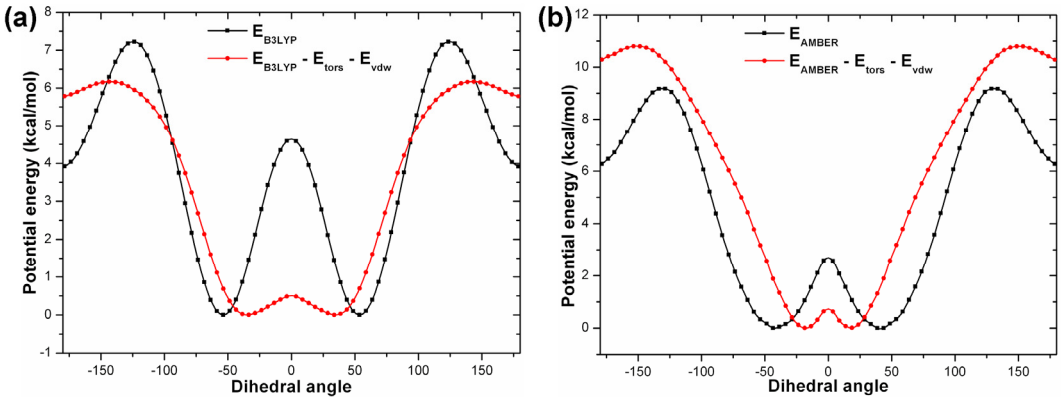


Figure 5.4: Deformation of the original PES of ethanolamine at the (a) B3LYP and (b) AMBER levels.

In Figure 5.4 we compare potential energy curves on the original and modified potential energy surface. The modified surface was obtained by subtracting the AMBER torsional and vdW terms from the B3LYP electronic energy (Figure 5.4a) or from the total AMBER energy (Figure 5.4b). As visualized in Figure 5.4a, the transition barrier between LM1 and GM1 is suppressed from 3.3 to 0.4 kcal/mol. Similarly, the transition barrier from GM1 to GM2 is suppressed from 4.6 to 0.5 kcal/mol. Thus both energy barriers were suppressed by almost 90%. There are two sources of the remaining barriers. First, other energy terms contribute to the barriers, e.g., the Coulomb term. Second, the residual barriers reflect inherent

inconsistencies between the electronic structure and FF models. An important observation is that the positions of the global minima are very similar on the original and modified PESs.

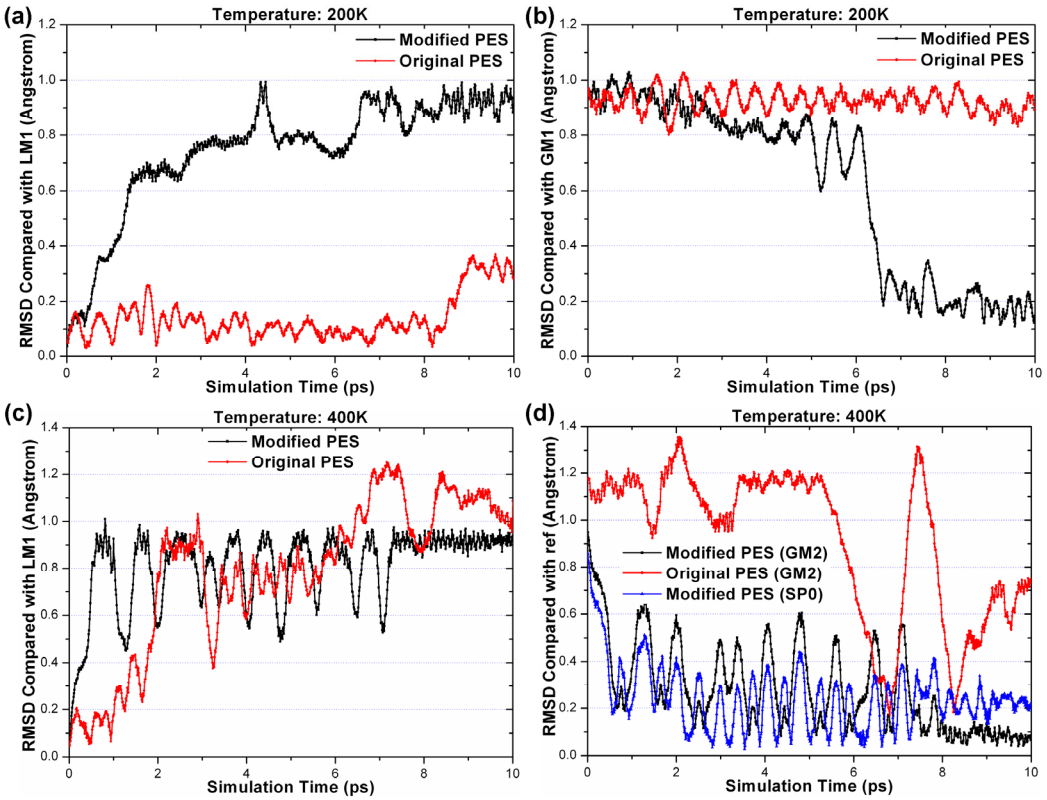


Figure 5.5: The RMSD values as a function of time for trajectories on the modified and original PES at 200 K (a,b) and 400 K (c,d) and calculated with respect to LM1 (a,c), GM1 (b) and GM2 or SP0 (d).

The results of MD simulations for ethanolamine on the original and modified PES potential are summarized in Figure 5.5 for T=200 and 400 K. The simulations were initiated from the local minimum geometry and the presented RMSD values were calculated with respect to the local (Figures 5.5 a,c) and global (Figures 5.5 b,d) minimum structures or SP0 (Figure 5.5d). The RMSD values at T=200 K (Figures 5.5 a,b) demonstrate that the trajectory on the original PES gets trapped in the local minimum. On the other hand, the trajectory on the modified PES settles in the basin of the global minimum after less than 7 ps. For simulations performed at 400 K, the trajectory on the modified PES moves to the basin of the global minimum within 1 ps, then it starts to oscillate between the basins of GM1 and GM2, and finally it locates in the basin of GM2. On the original PES, it takes

about 7 ps.

5.4 Discussion

We have demonstrated that the energy barriers separating different conformers of ethanolamine can be suppressed by subtracting the torsional and vdW terms from the total potential energy, Figure 5.4. This has beneficial effects when running MD trajectories on the deformed PESs. These trajectories move quickly to the global minimum basin, Figure 5.5. Trajectories run on the original PESs might get stuck in the local minimum basin.

Increasing the simulation temperatures helps to overcome the barrier, but there is always some risk associated with this procedure. We found that for MD trajectories run on the original PES at T=2000 K, one observes thermal decomposition of ethanolamine, i.e., a thermal dissociation of the C-H bonds. This is clearly an undesirable scenario. One could suppress these unwelcome events by strengthening the stretching intramolecular degrees of freedom. This would correspond to deformation of the original PES according to Eq. (5.2), but this time we would add some stretching terms to the original PES, i.e., $c_{FF.str} < 0$. One could think about a complex deformation, in which the torsional and vdW terms are subtracted to suppress energy barriers separating various conformers, the intramolecular stretching terms are added to avoid thermal decomposition, and the MD calculations are performed at a temperature as high as a few thousands of K.

The results presented in Figure 5.4 demonstrate that after removal of the torsional and vdW terms there is a deep basin for the global minimum. The existence of this basin results from the intramolecular $OH^{\cdots}NH_2$ hydrogen bond. What would happen if a molecule was complex enough to support several distinct conformers, each with its own hydrogen bond and, therefore, favorable Coulomb interactions? There would be several basins on the PES and the deformations described so far would not suppress the barriers separating them. However, one could suppress these barriers by subtracting the Coulomb term from the original PES, i.e., $c_{FF.coul} > 0$ in Eq. (5.2). This reasoning leads to a concept of generalized simulated annealing protocols. In standard protocols one *a priori* defines temperature for each time interval of the MD simulation. Here we suggest extending this concept by allowing the PES *and* temperature to change as time

progresses. In addition, we suggest that these changes should be made "on the fly" based on local information available from the advancing trajectory. For example, to create an opportunity for an MD trajectory to move from one basin of favorable Coulomb interactions to another, we suggest a temporary removal of the Coulomb term from the original PES. We expect the code to recognize the fact that a basin of attractive Coulomb interactions has been encountered. In response, the code should modify the PES by removing those Coulomb interactions that are characteristic of the encountered basin, and the trajectory should be propagated for a sufficiently long time to leave the basin. It would make sense during this interval to temporarily have the temperature increased, the stretching degrees of freedom strengthened, and the torsional and vdW barriers suppressed. Later, while the trajectory explores a new basin of favorable Coulomb interactions, the Coulomb term of the former basin should be restored; the temperature lowered, etc. The idea of generalized dynamically modified simulated annealing protocols based on deformed PESs will be explored in this ongoing project.

5.5 Summary

We have defined a modified potential energy surface as a linear combination of the original potential energy surface and the terms typically encountered in molecular force fields, such as stretching, bending, torsional, Coulomb, and van der Waals. Depending on the circumstances we deform the surface in different ways. For example, to suppress barriers separating different conformers, we recommend subtracting the torsional and vdW terms from the original PES. To avoid thermal decomposition into smaller molecular fragments in simulations performed at elevated temperatures, we recommend adding stretching terms to the original PES. To create an opportunity for a MD trajectory to get out of a basin of attractive Coulomb interactions, we recommend a temporary removal of selected Coulomb terms.

We formulated a concept of generalized simulated annealing protocols, in which both the temperature and potential energy can change as time progresses. The changes are made "on the fly" based on local information from the advancing trajectory rather than decided *a priori*, before starting the MD trajectory. We developed an interface between Gaussian 03 and Tinker, which allows running MD simulations on the original PES, which typically results from electronic

structure calculations, and on an arbitrarily deformed PES.

We performed a detailed study of the conformational space of ethanolamine. We demonstrated that the torsional and van der Waals energies contribute most to energy barriers separating different minima. A deformed PES, with the torsional and vdW terms subtracted, proved to be useful in MD simulations. The MD trajectory moved quickly to the global minimum basin on the deformed PES, but it gets stuck in a local minimum basin on the original PES.

References:

[1] Leach, A. R., *Molecular Modelling: Principles and Applications*. 2nd ed.; Prentice Hall: Harlow, 2001.

[2] Ponder, J. W.; Case, D. A., Force Fields for Protein Simulations. *Advances in Protein Chemistry* **2003,** 66, 27-85.

[3] Ling, S.; Yu, W.; Huang, Z.; Lin, Z.; Haranczyk, M.; Gutowski, M., Gaseous Arginine Conformers and Their Unique Intramolecular Interactions. *Journal of Physical Chemistry A* **2006,** 110, 12282-12291.

[4] Rak, J.; Skurski, P.; Simons, J.; Gutowski, M., Low-Energy Tautomers and Conformers of Neutral and Protonated Arginine. *Journal of the American Chemical Society* **2001,** 123, 11695-11707.

[5] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E., Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* **1953,** 21, 1087-1092.

[6] Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*. Clarendon: Oxford, 1987; p 450.

[7] Holland, J. H., *Adaptation in Natural and Artificial Systems*. MIT Press: Cambridge, 1992.

[8] Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by Simulated Annealing. *Science* **1983,** 220, 671-680.

[9] Straub, J. E.; Thirumalai, D., Theoretical Probes of Conformational Fluctuations in S-Peptide and RNase A/3′ –Ump Enzyme Product Complex. *Proteins: Structure, Function, and Bioinformatics* **1993,** 15, 360-373.

[10] Straub, J. E.; Thirumalai, D., Exploring the Energy Landscape in Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1993,** 90, 809-813.

[11] Irle, S.; Zheng, G.; Elstner, M.; Morokuma, K., Formation of Fullerene Molecules from Carbon Nanotubes: A Quantum Chemical Molecular Dynamics Study. *Nano Letters* **2003,** 3, 465-470.

[12] Tse, J. S.; Klug, D. D., Structure and Dynamics of Liquid Sulphur. *Physical Review B* **1999,** 59, 34-37.

[13] Piela, L.; Kostrowicki, J.; Scheraga, H. A., On the Multiple-Minima Problem in the Conformational Analysis of Molecules: Deformation of the Potential Energy Hypersurface by the Diffusion Equation Method. *Journal of Physical Chemistry* **1989,** 93, 3339-3346.

[14] Pillardy, J.; Piela, L., Molecular Dynamics on Deformed Potential Energy Hypersurfaces. *Journal of Physical Chemistry* **1995,** 99, 11805-11812.

[15] Pillardy, J.; Olszewski, K. A.; Piela, L., Performance of the Shift Method of Global Minimization in Searches for Optimum Structures of Clusters of Lennard-Jones Atoms. *Journal of Physical Chemistry* **1992,** 96, 4337-4341.

[16] Ma, J.; Straub, J. E., Simulated Annealing Using the Classical Density Distribution. *Journal of Chemical Physics* **1994,** 101, 533-541.

[17] Tsoo, C.; Brooks III, C. L., Cluster Structure Determination Using Gaussian Density Distribution Global Minimization Methods. *Journal of Chemical Physics* **1994,** 101, 6405-6411.

[18] Stillinger, F. H.; Stillinger, D. K., Cluster Optimization Simplified by Interaction Modification. *Journal of Chemical Physics* **1990,** 93, 6106-6107.

[19] Wales, D. J.; Doye, J. P. K., Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *Journal of Physical Chemistry A* **1997,** 101, 5111-5116.

[20] Bulusu, S.; Zeng, X. C., Structures and Relative Stability of Neutral Gold Clusters: Au_n (N = 15-19). *Journal of Chemical Physics* **2006,** 125, 154303-5.

[21] Jiang, D. E.; Walter, M.; Dai, S., Gold Sulfide Nanoclusters: A Unique Core-in-Cage Structure. *Chemistry - A European Journal* **2010,** 16, 4999-5003.

[22] Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B* **1988,** 37, 785-789.

[23] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *Journal of Physical Chemistry* **1994,** 98, 11623-11627.

[24] Hehre, W. J.; Ditchfield, R.; Pople, J. A., Self-Consistent Molecular Orbital Methods. Xii. Further Extensions of Gaussian---Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *Journal of Chemical Physics* **1972,** 56, 2257-2261.

[25] SSC: A Tool for Constructing Libraries for Systematic Screening of Conformers. http://sscf.sf.net (accessed August 12, 2010). Ling, S.; Gutowski, M. J. Chem. Inf. Model. 2010, submitted.

[26] Piela, L., *Ideas of Quantum Chemistry*. Elsevier: Amsterdam, 2007; p

1086.

[27] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995,** 117, 5179-5197.

[28] Beeman, D., Some Multistep Methods for Use in Molecular Dynamics Calculations. *Journal of Computational Physics* **1976,** 20, 130-139.

[29] Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996,** 14, 33-38.

[30] Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; 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.; Bakken, V.; 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.; and Pople, J. A. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.

[31] Ponder, J. W. *Tinker: Software Tools for Molecular Design*, 5.1; Saint Louis, MO, 2003.

[32] Rappe, A. K.; Goddard, W. A., Charge Equilibration for Molecular Dynamics Simulations. *Journal of Physical Chemistry* **1991,** 95, 3358-3363.

[33] Fukui, K., The Path of Chemical-Reactions - the Irc Approach. *Accounts of Chemical Research* **1981,** 14, 363-368.

Chapter 6

Relative Stability of Canonical Conformers of Glycine

We report results of electronic structure calculations on two conformers of canonical glycine. Our study covered explicitly correlated electronic structure methods and a sequence of basis sets from double- to quadruple-zeta quality. The Hartree-Fock and correlation energies were extrapolated to complete basis set limits. In addition, the relative stability of two conformers was studied with the PBE0, B3PW91, B3LYP, BHandHLYP, M06-L, and MPWB1K functionals and a triple-zeta polarized basis set. Our most accurate results suggest that the C conformer with a hydrogen bond between the amine group and the carbonyl oxygen is more stable by 0.60 kcal/mol than the Bn conformer, which has a hydrogen bond between the amine group and the hydroxyl group. The DFT functionals perform quite well, with B3LYP underestimating the latter result by only 0.09 kcal/mol. Also the most recent meta-GGA functionals, M06-L, and MPWB1K, perform well underestimating by only 0.2-0.3 kcal/mol.

6.1 Introduction

Amino acids are amongst the most important molecules of life on this planet [1]. Many experimental and theoretical studies were dedicated to amino acids with the aim to investigate the mechanisms of how they are involved in fundamental biological reactions that constitute the whole metabolic cycle [2-11].

As an important starting point of these studies, conformations of an isolated amino acid need to be determined because they decide how this molecule will interact with other molecules. Extensive theoretical results have been reported on conformations of different amino acids [12-15]. However, due to inherent structural flexibility of amino acids, most of these molecules have quite a large number of geometrical degrees of freedom [13], and to perform a systematic search of the conformational space, one would need to optimize a large number of initial structures. Thus it is impossible to explore the conformational space of

these amino acids thoroughly while still maintaining a high standard of accuracy. For this reason, most theoretical studies stopped at a point of calculating properties of amino acids based on geometries which were optimized at density functional level of theory (DFT). It is still essential to benchmark these DFT results against theoretical results which are determined at a relatively higher standard of accuracy (e.g., coupled cluster (CC) methods).

In past studies on arginine [13], which is a challenging amino acid from the point of view of available conformers and tautomers, it was found that DFT relative energies are unreliable when compared with CC relative energies based on second order Moller-Plesset (MP2) geometries. We concluded that the failure of DFT methods should be attributed to intrinsic intramolecular interactions present in arginine. While most density functionals should be able to deal with regular intramolecular hydrogen bonds, which commonly exist in amino acids, intramolecular dispersion interactions require careful considerations [13], and traditional DFT functional could not give an accurate description to these nonlocal interactions [16, 17]. These findings suggested that MP2 and higher level of theory, such as CC with single and double excitations (CCSD), nowadays commonly extended to include a perturbative contribution from triple excitations (CCSD(T)), should be tested for better-than-DFT geometries of biomolecules [18, 19].

In addition to the electronic structure model, special attention should be paid to the basis set, which is used to expand the occupied and virtual molecular orbitals. A very small basis set might fail to reproduce important intramolecular interactions, which contribute to the stability of various conformers, while an extended basis set can make calculations prohibitively expensive. Thus it is of practical importance to identify a basis set which offers a reasonable compromise between computational cost and accuracy.

We selected glycine as a model system for exploration of the problems discussed above. This molecule received a lot of attentions in the past few decades. In one of the earliest theoretical studies on glycine, Pople et al. analyzed the rotational potential energy surface of glycine at the HF/4-31G level of theory, located three minimum energy structures, and suggested that the most stable conformation has a hydrogen bond between the amine group and the carbonyl oxygen [20], see Figure 6.1a. Later on, Brown et al. found that the hydrogen bond in the most likely

conformation of glycine is actually between the amine group and the hydroxyl group, i.e., the nitrogen acted as a proton acceptor [21], see Figure 6.1b. Many other theoretical [22-28] and experimental [29-33] studies followed. Very recently, Kasalova et al. [26] performed high-level *ab initio* calculations for two lowest-energy conformers of glycine by performing CCSD(T) geometry optimizations with basis sets of triple-zeta quality. They also calculated MP2/6-31G* quartic force fields, which account for anharmonic contributions to rotational constants. With assistance of experimental spectroscopic structural information, they predicted accurate Born-Oppenheimer equilibrium structures for these two conformers.

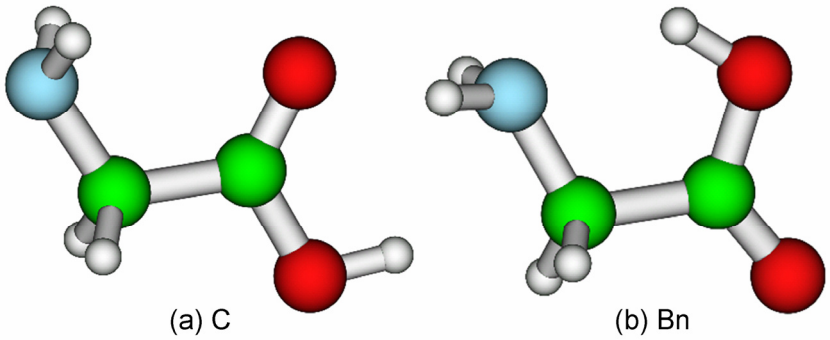


Figure 6.1: The two most stable conformers of glycine.

The main focus of the excellent study of Ref. 26 has been the molecular structure. There is, however, another important problem. What is the relative stability of these two conformers? This information is critical to predict the equilibrium distribution of conformers of glycine in the gas phase. The experimentally measured properties of glycine in the gas phase will be affected by contributions from all populated conformers [34]. In this study we focus our attention on the relative electronic energies, which were not addressed by Kasalova et al. [26].

6.2 Computational Details

The two canonical conformers of glycine are displayed in Figure 6.1. These are the most stable conformer C (Figure 6.1a) and the second most stable conformer Bn (Figure 6.1b). The notation employed here is adopted from a previous study [28]. We have optimized these two conformers at the MP2 [35] and CCSD [36]

levels of theory. For the MP2 optimizations we used the augmented polarized correlation consistent basis sets of double (AVDZ), triple (AVTZ) and quadruple (AVQZ) zeta quality [37, 38], whereas for the CCSD optimizations, only the AVDZ and AVTZ basis sets were employed. For evaluation of the molecular geometries we used the final results from Ref. 26. Subsequent single point energy calculations were performed for these five sets of geometries at the Hartree-Fock (HF), MP2, CCSD and CCSD(T) levels [39].

When studying basis set effects we also included one Pople basis set (6-31++G** [40]) in addition to AVDZ, AVTZ and AVQZ, which result in 125, 170, 400 and 800 Cartesian basis functions, respectively. The total HF and correlation energies, were extrapolated to the complete basis set (CBS) limit according to extrapolation schemes which were proposed by Halkier et al. [41] (see Eq. (6.1)) and Helgaker et al. [42] (see Eq. (6.2)), respectively:

$$E_X^{HF} = E_\infty^{HF} + a \cdot e^{-bX}, \tag{6.1}$$

$$E_X^{corr} = E_\infty + c \cdot X^{-3}, \tag{6.2}$$

where X is the cardinal number of the basis set, and a, b, and c are parameters to be determined.

Finally, we validated various exchange-correlation functionals against our most accurate predictions. The geometries were reoptimized for each exchange-correlation functional. The functionals we considered here come from different rungs of the Jacob's ladder to chemical accuracy, which was proposed by Perdew et al. [43], and include one nonempirical GGA (PBE0 [44]), three hybrid GGAs (B3LYP [45, 46], B3PW91 [47] and BHandHLYP [48]), and two most recent functionals, including one pure meta-GGA (M06-L [49]) and one hybrid meta-GGA (MPWB1K [50]). We hope such choices could cover some of the most representative density functionals which are being widely used today. All DFT calculations were performed with the AVTZ basis set.

The DFT and MP2 calculations were performed with Gaussian 03 [51] and NWCHEM [52], and the coupled cluster calculations were performed with MOLPRO [53].

6.3 Results and Discussions

6.3.1 Quality of Molecular Geometries

Parameters	MP2/AVQZ	CCSD/AVTZ	Ref. 26
r(C-N)	1.443	1.448	1.446
r(C-C)	1.512	1.516	1.511
r(C=O)	1.207	1.202	1.204
r(C-O)	1.351	1.349	1.349
r(C-H)	1.089	1.091	1.088
r(N-H)	1.011	1.012	1.012
θ(C-C-O)	111.1	111.3	111.4
θ(O-C-O)	123.2	123.1	123.1
θ(C-C-N)	115.3	115.3	115.2
θ(CH_2 scissor)	105.9	106.1	105.9

Table 6.1: Comparison of structural parameters of conformer C with the results of Ref. 26. Bond lengths are in Å, and bond angles are in °.

Parameters	MP2/AVQZ	CCSD/AVTZ	Ref. 26
r(C-C)	1.525	1.528	1.524
r(C-N)	1.461	1.465	1.463
r(C=O)	1.204	1.200	1.201
r(C-O)	1.338	1.337	1.337
r(O-H)	0.981	0.974	0.977
r(C-H)	1.088	1.090	1.087
r(N-H)	1.008	1.009	1.009
θ(C-O-H)	104.4	105.6	104.4
θ(C-C-O)	113.7	114.4	113.8
θ(O-C-O)	123.5	123.1	123.5
θ(C-C-N)	111.2	111.6	111.2
θ(CH2 scissor)	107.2	107.2	107.3
θ(NH2 scissor)	107.6	107.4	107.2
τ(NCCO)	10.7	11.1	12.8
τ(CCOH)	-2.0	-1.9	-2.3

Table 6.2: Comparison of structural parameters of conformer Bn with the results of Ref. 26. Bond lengths are in Å, and bond angles and dihedral angles are in °.

C	MP2/AVDZ	MP2/AVTZ	MP2/AVQZ	CCSD/AVDZ	CCSD/AVTZ
RMS(r)	0.014	0.004	0.002	0.013	0.003
RMS(θ)	0.274	0.235	0.166	0.212	0.100

Table 6.3: RMS errors of selected structural parameters of conformer C calculated with respect to the structure from Ref. 26. Bond lengths (r) are in Å, and bond angles (θ) are in °.

Bn	MP2/AVDZ	MP2/AVTZ	MP2/AVQZ	CCSD/AVDZ	CCSD/AVTZ
RMS(r)	0.013	0.004	0.002	0.011	0.002
RMS(θ)	0.204	0.168	0.173	0.583	0.601
RMS(τ)	2.081	1.500	1.500	1.012	1.235

Table 6.4: RMS errors of selected structural parameters of conformer Bn calculated with respect to the structure from Ref. 26. Bond lengths (r) are in Å, bond angles (θ) and dihedral angles (τ) are in °.

The MP2/AVQZ and CCSD/AVTZ geometries are presented in Tables 6.1 and 6.2 for the C and Bn conformers, respectively. The root mean square (RMS) errors for bond lengths, bond angles, and dihedral angles calculated with respect to the "target" molecular geometries of Ref. 26 are reported in Tables 6.3 and 6.4 for C and Bn, respectively. We conclude that the MP2/AVQZ and CCSD/AVTZ geometries are in excellent agreement with those from Ref. 26, with the RMS errors similar to the theoretical uncertainty [26]. The MP2 method fares extremely well in comparison with the computationally more demanding CCSD. The RMS errors for the MP2/AVQZ geometrical predictions are similar to those for CCSD/AVTZ. In the case of bond angles of Bn the MP2 predictions are surprisingly accurate. The good performance of MP2 extends to AVTZ and even AVDZ basis sets.

6.3.2 Convergence of Relative Energies of C and Bn

The computational accuracy will be discussed at four levels, as summarized in Figure 6.2. At the first stage, we will discuss total electronic energies calculated with different methods/basis sets for the fixed CCSD/AVTZ geometries. At the second stage we will discuss convergence of the electron correlation energies (MP2, CCSD, CCSD(T)) as a function of basis sets for two geometries: MP2/AVDZ and CCSD/AVTZ. At the third stage we will discuss convergence of the relative energies of C and Bn as a function of the electronic structure model and basis sets used. All five geometries discussed in Section 6.3.1 will be considered. Finally, we will discuss convergence of the relative energies with respect to the five geometries characterized in Tables 6.3 and 6.4.

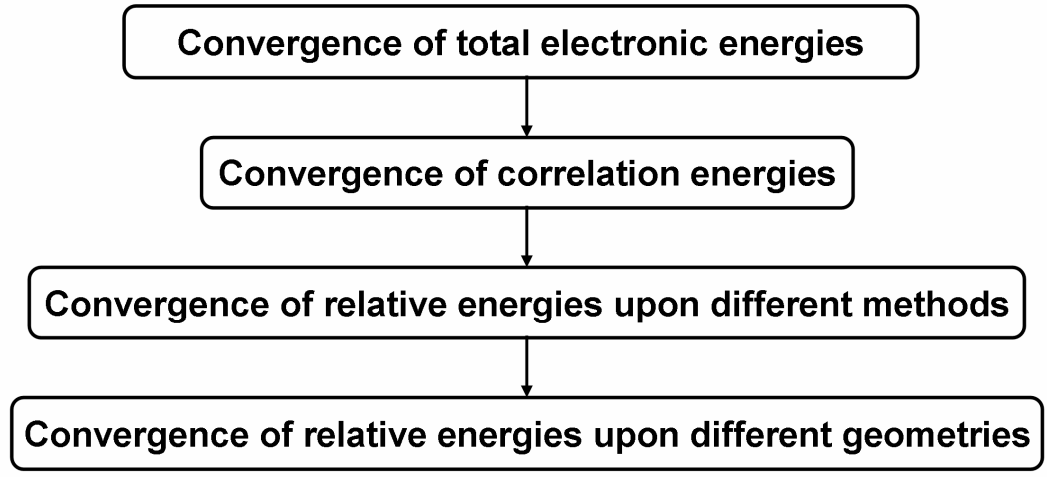


Figure 6.2: Four-level approach to convergence of relative energies of C and Bn.

6.3.3 Convergence of Total Electronic Energies with respect to Basis Sets

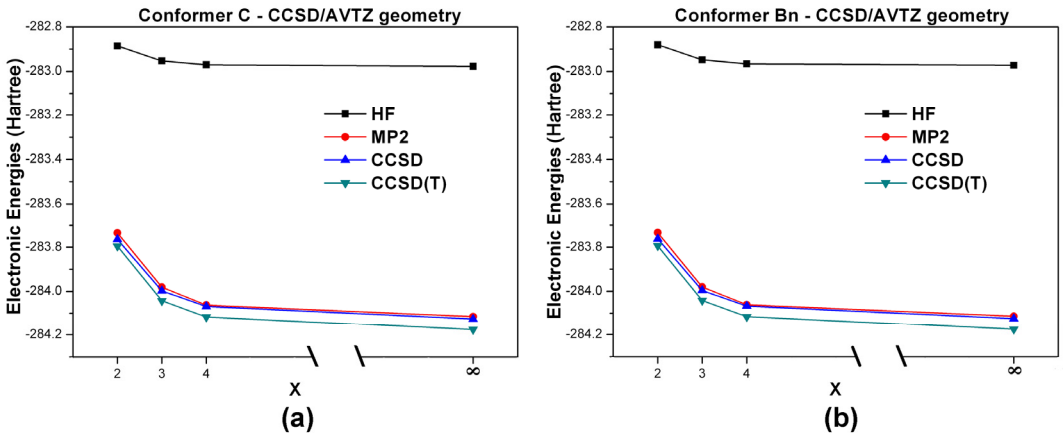


Figure 6.3: Convergence of total electronic energies with respect to basis sets.

In Figure 6.3 we demonstrate convergence of the total electronic energies determined at the HF, MP2, CCSD, and CCSD(T) level as a function of basis sets. The CBS extrapolated electronic energies are also presented. The results are reported for the CCSD/AVTZ geometries. The convergence of the total electronic energies is smooth, with the AVQZ energies differing from the CBS limits by about 0.05 Hartree. A similar difference is typically observed between the AVTZ and AVQZ results. On the other hand, the AVDZ total energies are much less accurate, with the discrepancies between AVDZ and AVTZ being as large as 0.2 Hartree. This illustrates serious uncertainties associated with computational

results obtained with the AVDZ basis. It is also well established that extrapolation of electron correlation energies initiated from the AVDZ results are less accurate. With the energy scale used in Figure 6.3, the AVQZ results seem to be well converged. One should keep in mind, however, that the uncertainty of 0.05 Hartree is equivalent to 31.4 kcal/mol. With the uncertainty in the total electronic energies being one order of magnitude larger than the chemical accuracy, the computational values of relative energies of C and Bn hinge on an extended cancellation of intrinsic errors in total energies of two conformers.

6.3.4 Convergence of Correlation Energies with respect to Different Levels of Theory

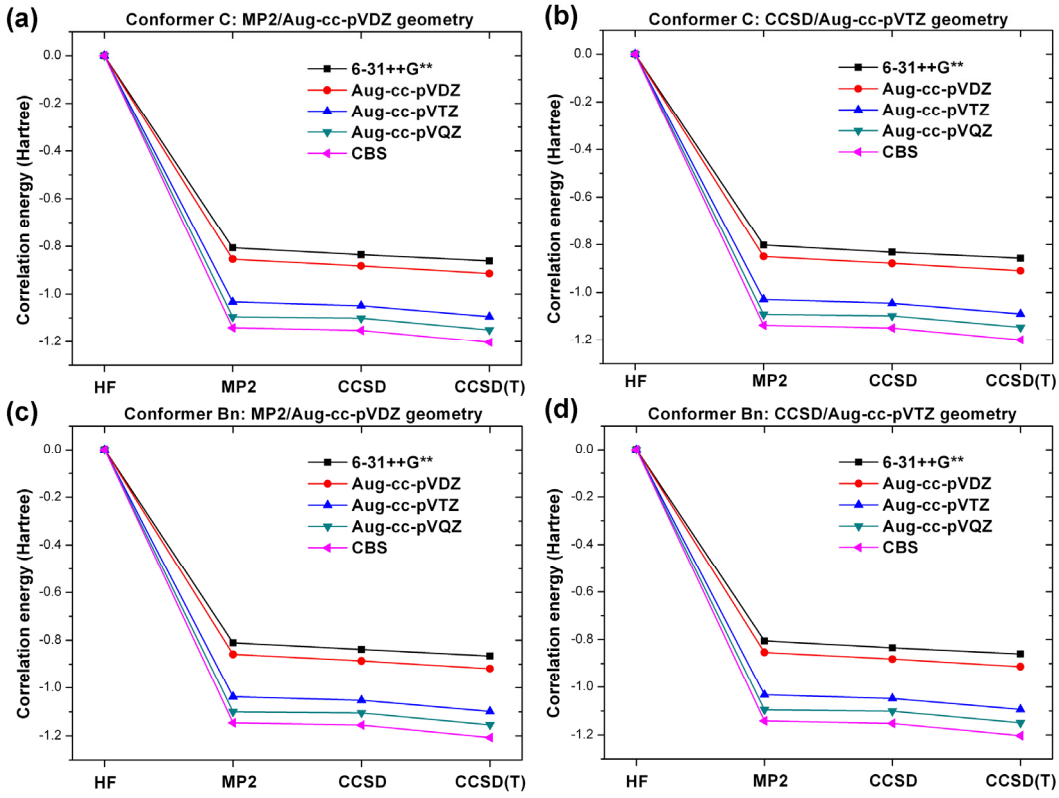


Figure 6.4: Convergence of correlation energies with respect to different levels of theory.

The convergence of the electron correlation energies with respect to the different levels of theory is shown in Figure 6.4. The results are reported for two geometries, MP2/AVDZ and CCSD/AVTZ. The MP2 correlation energies are between 0.8 and 1.2 Hartree, depending on the basis set used. Further contributions to electron correlation resulting from coupled cluster models are

smaller than 0.1 Hartree (i.e., 62.7 kcal/mol!). The correlation energies determined with different basis sets for a fixed level of theory are divided into two groups. The first distinct group includes the 6-31++G** and AVDZ basis sets, and the second group includes basis sets of triple, quadruple, and CBS quality. Clearly, the correlation energies calculated with AVDZ are far from convergence. The difference between the two groups develops clearly at the MP2 level and does not decrease upon further methodological advancements. With the energy scale used in Figure 6.4, the correlation energies seem to be converged. However, the difference between the correlation energies determined at the MP2/AVQZ and MP2/CBS limit is about 0.05 Hartree. i.e., 31 kcal/mol. This further illustrates how difficult it is to converge total correlation energies below the threshold of "chemical accuracy".

There are no significant differences between the convergence plots for the MP2/AVDZ and CCSD/AVTZ geometries. However, the MP2/AVQZ single point energy of C at the MP2/AVDZ geometry is higher than the same energy at the CCSD/AVTZ geometry by about 0.6 kcal/mol.

6.3.5 Convergence of Relative Energies of C and Bn with respect to Methods and Basis Sets

The convergence of the relative electronic energies of C and Bn with respect to different levels of theory and basis sets is illustrated in Figure 6.5, with separate panels for each of five sets of molecular geometries. Notice a change of energy scale from Hartree in Figures 6.3 and 6.4 to kcal/mol in Figures 6.5 and 6.6. The HF relative energies cover a range 2.8-3.4 kcal/mol and systematically overestimate the CCSD(T) relative energies, which cover a range 0.6-1.0 kcal/mol. Therefore the HF method is not suitable for determination of relative energies of various conformers and some treatment of electron correlation is required. The results of Figure 6.5 also illustrate that the relative energies are underestimated at the MP2 level and overestimated at the CCSD level, though these intrinsic errors are much smaller than those of the HF method. It is awarding to notice that the AVTZ, AVQZ, and CBS curves almost overlap in each panel of Figure 6.5. It suggests that convergence of the relative energies with respect to selection of basis sets has been accomplished. Again the AVDZ and 6-31++G** curves are clearly separated from the former three curves, reconfirming the limitations of these two basis sets.

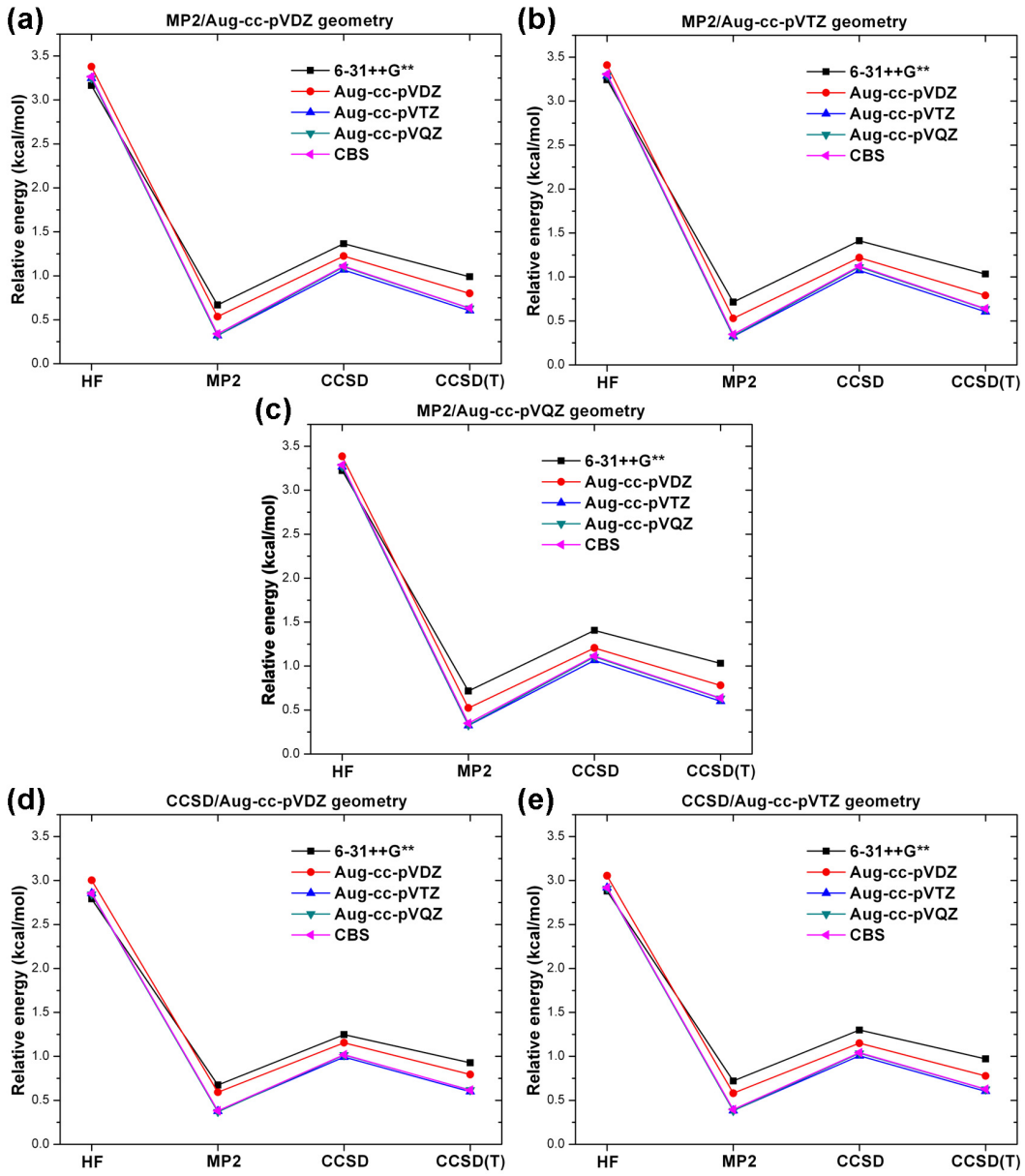


Figure 6.5: Convergence of relative energies with respect to different levels of theory.

6.3.6 Convergence of Relative Energies of C and Bn with respect to Different Geometries

The convergence of the relative energies with respect to the five sets of molecular geometries is shown in Figure 6.6. Each panel corresponds to one electronic structure method. Notice that the energies displayed on the vertical axes cover only a fraction of kcal/mol. The HF method not only seriously overestimates the

relative energy but it is also the most sensitive to minor geometrical changes. On the other hand the CCSD(T) relative energies are quite insensitive to minor geometrical changes, especially with basis sets of the AVTZ and AVQZ quality. The AVQZ curve almost overlaps with the CBS curve, which confirms again the basis set convergence of the reported results. At the same time, the AVTZ basis set tends to underestimate the relative energy by only 0.03 kcal/mol, which is a very promising performance [54, 55].

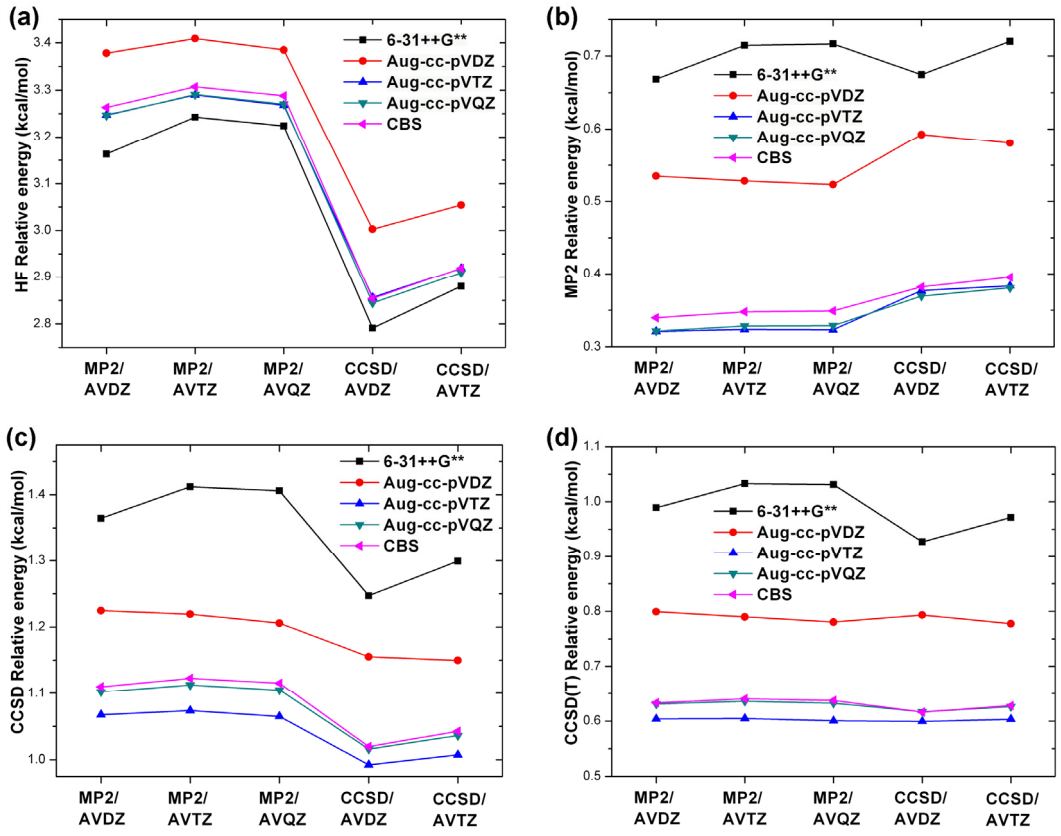


Figure 6.6: Convergence of relative energies with respect to different geometries.

Finally, we recommend the CCSD(T)/AVTZ single point energy calculations for exploration of potential energy surfaces of both canonical conformers of glycine. The geometries of the MP2 stationary points are very similar to the CCSD optimized geometries, at least for this specific amino acid. The performance of the 6-31++G** or AVDZ basis sets is not satisfactory and these basis sets should be used cautiously.

6.3.7 Validations of DFT Functionals

The performance of the PBE0, B3PW91, B3LYP, BHandHLYP, M06-L, and MPWB1K functional in the determination of the relative energies of C and Bn is summarized in Figure 6.7. The DFT results are obtained with the AVTZ basis sets and the molecular structures were optimized for each functional. The DFT results are compared with the HF, MP2, CCSD and CCSD(T) results obtained with the same AVTZ basis set and molecular geometries optimized at the corresponding level of theory, with an exception of CCSD(T), for which the CCSD geometries were used.

Taking the CCSD(T) relative energy of 0.60 kcal/mol as a reference, one can appreciate the performance of the DFT methods, in particular having in mind their low computational cost. The B3LYP functional predicts a relative energy of 0.51 kcal/mol (underestimation of 0.09 kcal/mol), the BHandHLYP functional overestimates by 0.58 kcal/mol and the B3PW91 and PBE0 functionals underestimate by ca. 0.60 kcal/mol. The performance of two most recent functionals, M06-L and MPWB1K, is satisfactory as they underestimate the relative energy by only 0.2~0.3 kcal/mol. The inaccuracy of the DFT functionals should be kept in proper perspective: the CCSD method overestimates the relative energy by 0.41 kcal/mol.

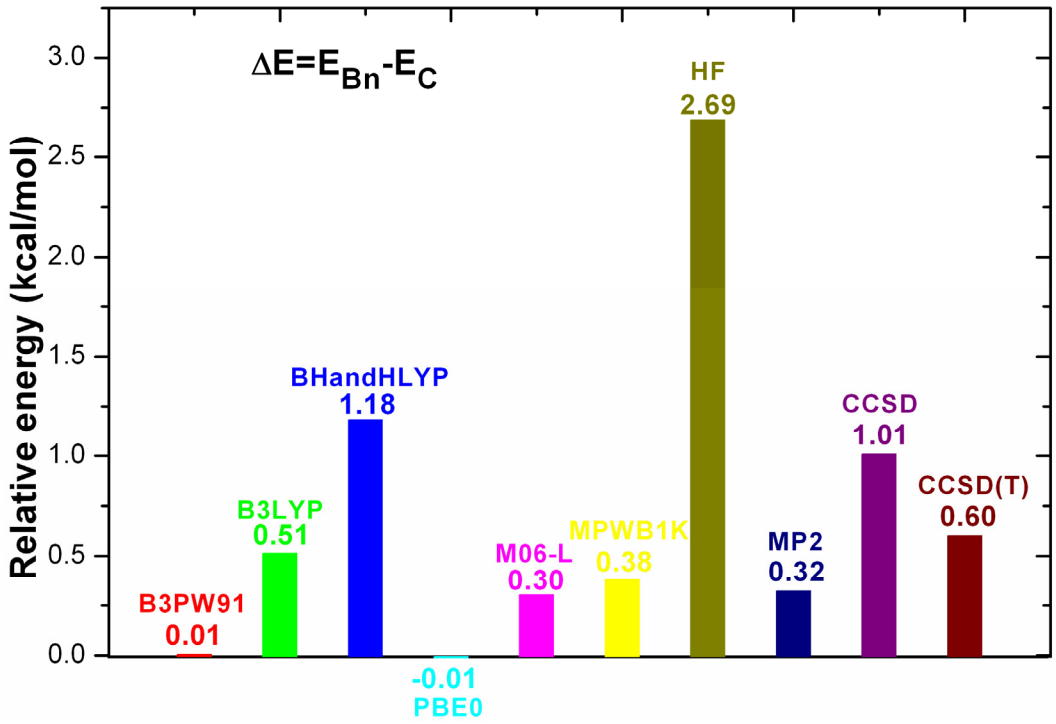


Figure 6.7: Validations of DFT functionals.

6.4 Conclusions

We performed a systematic computational study of the relative stability of two canonical conformers of glycine. Our study covered HF, MP2, CCSD, and CCSD(T) electronic structure methods and a sequence of basis sets from double- to quadruple-zeta quality. The HF and correlation energies were extrapolated to the complete basis set limits. In addition, the relative stability of two conformers was studied with the PBE0, B3PW91, B3LYP, BHandHLYP, M06-L, and MPWB1K functionals and a triple-zeta polarized basis set.

The MP2/AVQZ and CCSD/AVTZ geometries are in very good agreement with the most accurate structural predictions [26]. Our most accurate results suggest that, in terms of CCSD(T) electronic energies, the C conformer is more stable than the Bn conformer by 0.60 kcal/mol. This result is not methodologically saturated, as the CCSD result is 1.01 kcal/mol. The DFT functionals perform quite well, with B3LYP underestimating the CCSD(T) result by only 0.09 kcal/mol. Also the most recent meta-GGA functionals, M06-L, and MPWB1K, perform well underestimating by only 0.2-0.3 kcal/mol.

On the practical side: (1) the aug-cc-pVTZ and aug-cc-pVQZ basis sets are recommended for high quality results, whereas the aug-cc-pVDZ and 6-31++G** basis sets are clearly of lower quality; (2) the CCSD(T) relative energies are sensitive to the basis set selection but the least sensitive to minor geometrical changes; (3) MP2 geometries proved to be surprisingly accurate for the two canonical conformers of glycine, even when relatively small aug-cc-pVDZ basis sets are employed.

Other amino acids and small biomolecules should be studied in a similar fashion to establish the reliability of different electronic structure approaches. The current chapter only covers the two most stable canonical conformers of glycine in the gas phase. In Chapter 8, we will present our results on the adsorption of canonical and zwitterionic glycine on the (001) surface of rhenium oxide.

References:

[1] McKee, T.; McKee, J. R., *Biochemistry: The Molecular Basis of Life*. Oxford University Press: Oxford, 2009; p 717.

[2] Abdoul-Carime, H.; Konig-Lehmann, C.; Kopyra, J.; Farizon, B.; Farizon, M.; Illenberger, E., Dissociative Electron Attachment to Amino-Acids: The Case of Leucine. *Chemical Physics Letters* **2009,** 477, 245-248.

[3] Blanco, S.; Lesarri, A.; Lopez, J. C.; Alonso, J. L., The Gas-Phase Structure of Alanine. *Journal of the American Chemical Society* **2004,** 126, 11675-11683.

[4] Bush, M. F.; Oomens, J.; Saykally, R. J.; Williams, E. R., Effects of Alkaline Earth Metal Ion Complexation on Amino Acid Zwitterion Stability: Results from Infrared Action Spectroscopy. *Journal of the American Chemical Society* **2008,** 130, 6463-6471.

[5] Gao, B.; Wyttenbach, T.; Bowers, M. T., Hydration of Protonated Aromatic Amino Acids: Phenylalanine, Tryptophan, and Tyrosine. *Journal of the American Chemical Society* **2009,** 131, 4695-4701.

[6] Grace, L. I.; Cohen, R.; Dunn, T. M.; Lubman, D. M.; de Vries, M. S., The R2pi Spectroscopy of Tyrosine: A Vibronic Analysis. *Journal of Molecular Spectroscopy* **2002,** 215, 204-219.

[7] Jockusch, R. A.; Price, W. D.; Williams, E. R., Structure of Cationized Arginine (Arg Center Dot M+, M = H, Li, Na, K, Rb, and Cs) in the Gas Phase: Further Evidence for Zwitterionic Arginine. *Journal of Physical Chemistry A* **1999,** 103, 9266-9274.

[8] Snoek, L. C.; Robertson, E. G.; Kroemer, R. T.; Simons, J. P., Conformational Landscapes in Amino Acids: Infrared and Ultraviolet Ion-Dip Spectroscopy of Phenylalanine in the Gas Phase. *Chemical Physics Letters* **2000,** 321, 49-56.

[9] Bowman, V. N.; Heaton, A. L.; Armentrout, P. B., Metal Cation Dependence of Interactions with Amino Acids: Bond Energies of Rb+ to Gly, Ser, Thr, and Pro. *Journal of Physical Chemistry B* **2010,** 114, 4107-4114.

[10] Csaszar, A. G., Conformers of Gaseous Alpha-Alanine. *Journal of Physical Chemistry* **1996,** 100, 3541-3551.

[11] Kassab, E.; Langlet, J.; Evleth, E.; Akacem, Y., Theoretical Study of Solvent Effect on Intramolecular Proton Transfer of Glycine. *Journal of Molecular Structure-Theochem* **2000,** 531, 267-282.

[12] Huang, Z. J.; Lin, Z. J.; Song, C., Protonation Processes and Electronic Spectra of Histidine and Related Ions. *Journal of Physical Chemistry A* **2007,** 111, 4340-4352.

[13] Ling, S.; Yu, W.; Huang, Z.; Lin, Z.; Haranczyk, M.; Gutowski, M., Gaseous Arginine Conformers and Their Unique Intramolecular Interactions. *Journal of Physical Chemistry A* **2006,** 110, 12282-12291.

[14] Yu, W. B.; Liang, L.; Lin, Z. J.; Ling, S. L.; Haranczyk, M.; Gutowski, M., Comparison of Some Representative Density Functional Theory and Wave Function Theory Methods for the Studies of Amino Acids. *Journal of Computational Chemistry* **2009,** 30, 589-600.

[15] Yu, W. B.; Lin, Z. J.; Huang, Z. J., Coexistence of Dihydrogen, Blue- and Red-Shifting Hydrogen Bonds in an Ultrasmall System: Valine. *Chemphyschem* **2006,** 7, 828-830.

[16] Hobza, P.; Sponer, J.; Reschel, T., Density Functional Theory and Molecular Clusters. *Journal of Computational Chemistry* **1995,** 16, 1315-1325.

[17] Cerny, J.; Hobza, P., The X3lyp Extended Density Functional Accurately Describes H-Bonding but Fails Completely for Stacking. *Physical Chemistry Chemical Physics* **2005,** 7, 1624-1626.

[18] Kaminsky, J.; Jensen, F., Force Field Modeling of Amino Acid Conformational Energies. *Journal of Chemical Theory and Computation* **2007,** 3, 1774-1788.

[19] Yu, C.-H.; Norman, M. A.; Schafer, L.; Ramek, M.; Peeters, A.; van Alsenoy, C., Ab Initio Conformational Analysis of N-Formyl -Alanine Amide Including Electron Correlation. *Journal of Molecular Structure* **2001,** 567-568, 361-374.

[20] Vishveshwara, S.; Pople, J. A., Molecular Orbital Theory of the Electronic Structures of Organic Compounds. 32. Conformations of Glycine and Related Systems. *Journal of the American Chemical Society* **1977,** 99, 2422-2426.

[21] Brown, R. D.; Godfrey, P. D.; Storey, J. W. V.; Bassez, M. P., Microwave-Spectrum and Conformation of Glycine. *Journal of the Chemical Society-Chemical Communications* **1978**, 547-548.

[22] Jensen, J. H.; Gordon, M. S., Conformational Potential Energy Surface of Glycine: A Theoretical Study. *Journal of the American Chemical Society* **1991,** 113, 7917-7924.

[23] Csaszar, A. G., Conformers of Gaseous Glycine. *Journal of the American Chemical Society* **1992,** 114, 9568-9575.

[24] Hu, C. H.; Shen, M.; Schaefer, H. F., Glycine Conformational Analysis. *Journal of the American Chemical Society* **1993,** 115, 2923-2929.

[25] Frey, R. F.; Coffin, J.; Newton, S. Q.; Ramek, M.; Cheng, V. K. W.; Momany, F. A.; Schaefer, L., Importance of Correlation-Gradient Geometry Optimization for Molecular Conformational Analyses. *Journal of the American Chemical Society* **1992,** 114, 5369-5377.

[26] Kasalova, V.; Allen, W. D.; Schaefer, H. F.; Czinki, E.; Csaszar, A. G., Molecular Structures of the Two Most Stable Conformers of Free Glycine. *Journal of Computational Chemistry* **2007,** 28, 1373-1383.

[27] Haranczyk, M.; Gutowski, M., Effect of Excess Electron and One Water Molecule on Relative Stability of the Canonical and Zwitterionic Tautomers of Glycine. *Journal of Chemical Physics* **2008,** 128, 125101-12.

[28] Gutowski, M.; Skurski, P.; Simons, J., Dipole-Bound Anions of Glycine Based on the Zwitterion and Neutral Structures. *Journal of the American Chemical Society* **2000,** 122, 10159-10162.

[29] Godfrey, P. D.; Brown, R. D., Shape of Glycine. *Journal of the American Chemical Society* **1995,** 117, 2019-2023.

[30] Lovas, F. J.; Kawashima, Y.; Grabow, J. U.; Suenram, R. D.; Fraser, G. T.; Hirota, E., Microwave-Spectra, Hyperfine-Structure, and Electric-Dipole Moments for Conformers-I and Conformer-Ii of Glycine. *Astrophysical Journal* **1995,** 455, L201-L204.

[31] Suenram, R. D.; Lovas, F. J., Millimeter Wave Spectrum of Glycine - a New Conformer. *Journal of the American Chemical Society* **1980,** 102, 7180-7184.

[32] Ilyushin, V. V.; Alekseev, E. A.; Dyubko, S. F.; Motiyenko, R. A.; Lovas, F. J., Millimeter Wave Spectrum of Glycine. *Journal of Molecular Spectroscopy* **2005,** 231, 15-22.

[33] McGlone, S. J.; Elmes, P. S.; Brown, R. D.; Godfrey, P. D., Molecular Structure of a Conformer of Glycine by Microwave Spectroscopy. *Journal of Molecular Structure* **1999,** 485, 225-238.

[34] Sanz, M. E.; Lopez, J. C.; Alonso, J. L., Six Conformers of Neutral Aspartic Acid Identified in the Gas Phase. *Physical Chemistry Chemical Physics* **2010,** 12, 3573-3578.

[35] Moller, C.; Plesset, M. S., Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934,** 46, 618-622.

[36] Purvis, G. D.; Bartlett, R. J., A Full Coupled-Cluster Singles and Doubles Model - the Inclusion of Disconnected Triples. *Journal of Chemical Physics* **1982,** 76, 1910-1918.

[37] Dunning, J. T. H., Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *Journal of Chemical Physics* **1989,** 90, 1007-1023.

[38] Kendall, R. A.; Dunning, T. H.; Harrison, R. J., Electron-Affinities of the 1st-Row Atoms Revisited - Systematic Basis-Sets and Wave-Functions. *Journal of Chemical Physics* **1992,** 96, 6796-6806.

[39] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M., A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chemical Physics Letters* **1989,** 157, 479-483.

[40] Hehre, W. J.; Ditchfield, R.; Pople, J. A., Self-Consistent Molecular Orbital Methods. Xii. Further Extensions of Gaussian---Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *Journal of Chemical Physics* **1972,** 56, 2257-2261.

[41] Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J., Basis-Set Convergence of the Energy in Molecular Hartree-Fock Calculations. *Chemical Physics Letters* **1999,** 302, 437-446.

[42] Helgaker, T.; Klopper, W.; Koch, H.; Noga, J., Basis-Set Convergence of Correlated Calculations on Water. *Journal of Chemical Physics* **1997,** 106, 9639-9646.

[43] Perdew, J. P.; Schmidt, K. In *Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy*, Density Functional Theory and Its Application to Materials, Antwerp (Belgium), 2001; AIP: Antwerp (Belgium), 2001; pp 1-20.

[44] Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple (Vol 77, Pg 3865, 1996). *Physical Review Letters* **1997,** 78, 1396-1396.

[45] Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B* **1988,** 37, 785-789.

[46] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *Journal of Physical Chemistry* **1994,** 98, 11623-11627.

[47] Perdew, J. P., Density-Functional Approximation for the Correlation-Energy of the Inhomogeneous Electron-Gas. *Physical Review B* **1986,** 33, 8822-8824.

[48] Becke, A. D., A New Mixing of Hartree-Fock and Local Density-Functional Theories. *Journal of Chemical Physics* **1993,** 98, 1372-1377.

[49] Zhao, Y.; Truhlar, D. G., A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *Journal of Chemical Physics* **2006,** 125, 194101-18.

[50] Zhao, Y.; Truhlar, D. G., Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The Mpw1b95 and Mpwb1k Models and Comparative Assessments for Hydrogen Bonding and Van Der Waals Interactions. *Journal of Physical Chemistry A* **2004,** 108, 6908-6918.

[51] Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; 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.; Bakken, V.; 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.; and Pople, J. A. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.

[52] Bylaska, E. J. d. J., W. A.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Apra, E.; Windus, T. L.; Hirata, S.; Hackler, M. T.; Zhao, Y.; Fan, P.-D.; Harrison, R. J.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Auer, A. A.; Nooijen, M.; Brown, E.; Cisneros, G.; Fann, G. I.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J. A.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Pollack, L.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. "NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.0" (2006), Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA.

[53] MOLPRO, version 2006.1, a package of ab initio programs, H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, and others , see http://www.molpro.net.

[54] Peterson, K. A.; Dunning, T. H., Intrinsic Errors in Several Ab Initio Methods: The Dissociation Energy of N2. *Journal of Physical Chemistry* **1995,** 99, 3898-3901.

[55] Feller, D.; Peterson, K. A., An Examination of Intrinsic Errors in Electronic Structure Methods Using the Environmental Molecular Sciences Laboratory Computational Results Database and the Gaussian-2 Set. *Journal of Chemical Physics* **1998,** 108, 154-176.

Chapter 7

Structural Study of 2'-Deoxycytidine

Computational results have been reported for 2'-deoxycytidine (dC), its conformers, selected isomers and tautomers. Neutral molecules and associated anionic and cationic radicals were considered. We also analyzed the effect of the crystalline environment on the conformational structure of dC. The calculations were performed at the density functional level of theory. The conformational space of canonical dC was systematically probed. The neutral canonical dC in the gas phase favors *syn* rather than *anti* conformations. The thermodynamic dominance of *syn* conformations results from the formation of an intramolecular O5'-H13···O2 hydrogen bond. We concluded that intermolecular hydrogen bonds favor *anti* conformers in the experimental structure of the molecular crystal of dC. We identified a cyclic hydrogen bonded motif in the crystalline structure of dC, which involves two N3 proton acceptors and two N4H proton donors.

7.1 Introduction

Recent years brought significant advances in transferring of thermally labile and nonvolatile biomolecules into the gas phase [1-7] and spectroscopic interrogations of these important molecular systems became possible [8-11]. In order to interpret the gas phase spectroscopic data experimentalists use computational results available in the literature, perform calculations by themselves, or engage computational colleagues in fruitful collaborations. At first glance the computational task is straightforward because density functional theory (DFT) or second-order Moller-Plesset (MP2) calculations are nowadays routinely executed for molecules containing several tens of atoms. There is, however, an intriguing caveat: which structure of the molecule should be used in these calculations? Biomolecules typically support several plausible conformers in addition to chemically better distinguished isomers or tautomers. Computational chemists frequently assume an initial structure extracted from X-ray crystal experiments and apply standard optimization methods to an isolated molecule [12-14]. However, the most common optimization algorithms, which are driven by calculated forces, converge to the closest local minimum rather than to the most

stable structure. More sophisticated algorithms, such as basin-hopping [15], simulated annealing [16], genetic algorithms [17], or systematic searches [18] are not routinely used. As a result, computational results might be obtained for chemical structures that are not populated in the gas phase experimental conditions and this problem is illustrated in the current report.

Here we discuss the case of 2'-deoxycytidine (dC), for which the X-ray crystal structure is known [19], and the experimentally determined gas phase acidity [20] and excess electron binding energy [1] were interpreted in terms of computational results. The molecular structures used in these computational studies were biased by X-ray structural prediction [21, 22] or by the structure of DNA [12-14, 23, 24]. Our results unravel that the most stable structure of the neutral gas phase dC differs qualitatively. The reason is that the solid state structure is dictated by a competition between *intra-* and *intermolecular* hydrogen bonds, while the most stable gas phase structure is dictated by an *intramolecular* hydrogen bond. This illustrates that computational results for gas phase molecules based on structures extracted from condense phases might be misleading and agreement with experiment, or lack thereof, might be fortuitous. The current study was facilitated by our new combinatorial/computational tool for systematic screening of conformers (SSC) [25].

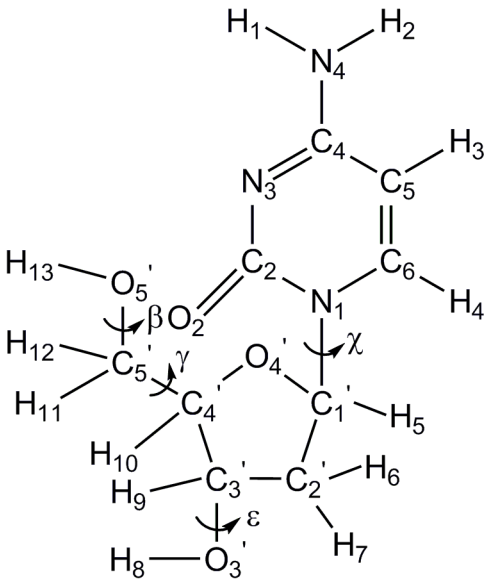


Figure 7.1: Schematic molecular structure of 2'-deoxycytidine.

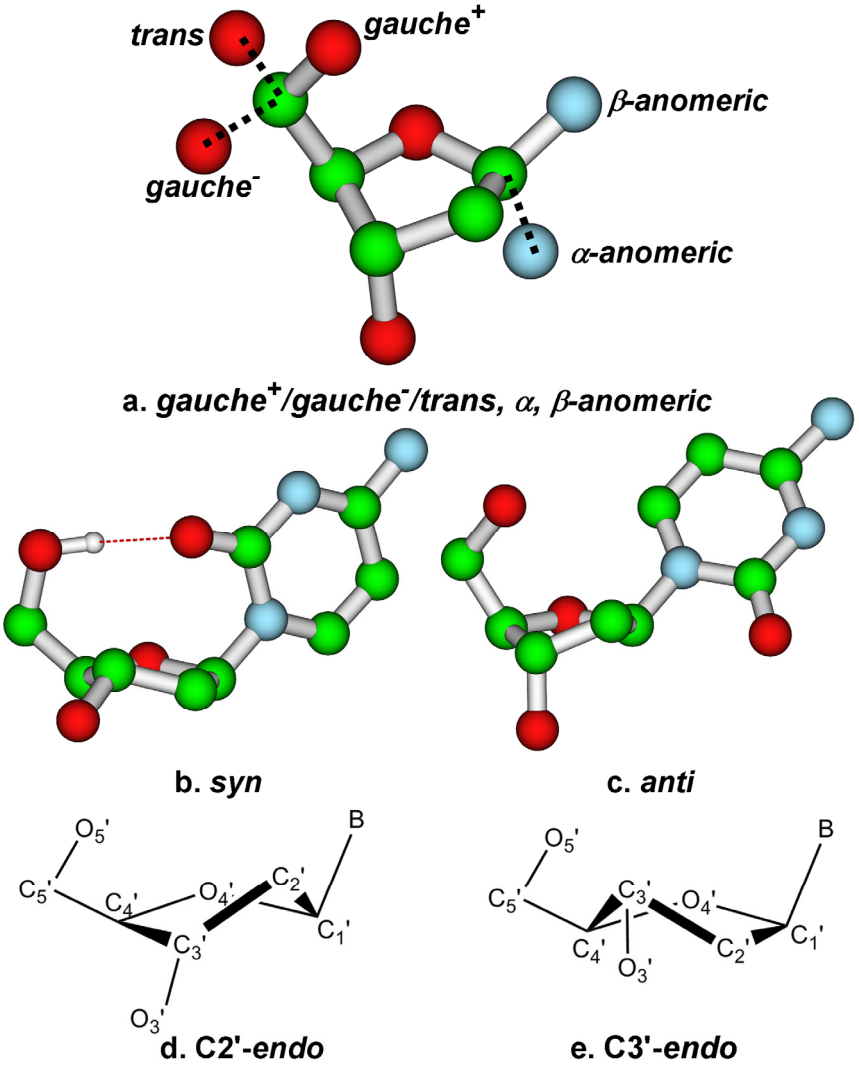


Figure 7.2: A schematic definition of selected degrees of freedom of dC.

For an isolated dC molecule, the structural flexibility could be reflected by its rotational degrees of freedom and various isomers and tautomers, see Figures 7.1 and 7.2. Some rotational degrees of freedom are labeled in Figure 7.1 with lower case Greek letters and describe rotations around the C5'-O5', C1'-N1, C3'-O3', and C4'-C5' bonds. Figure 7.2 illustrates a terminology which is widely accepted in biochemical community. The orientation of the O5'-H13 group, which is characterized by β, determines whether dC is prearranged to form intra- or intermolecular hydrogen bonds. There are three major β orientations observed in DNA, and they are *gauche*$^+$, *gauche*$^-$ and *trans*, see Figure 7.2a. The orientation of the base with respect to the sugar is characterized by χ, with two major conformations being *syn* (Figure 7.2b) and *anti* (Figure 7.2c). The non-planarity of the sugar ring typically takes two forms: C2'-*endo* (Figure 7.2d) and C3'-*endo*

(Figure 7.2e).

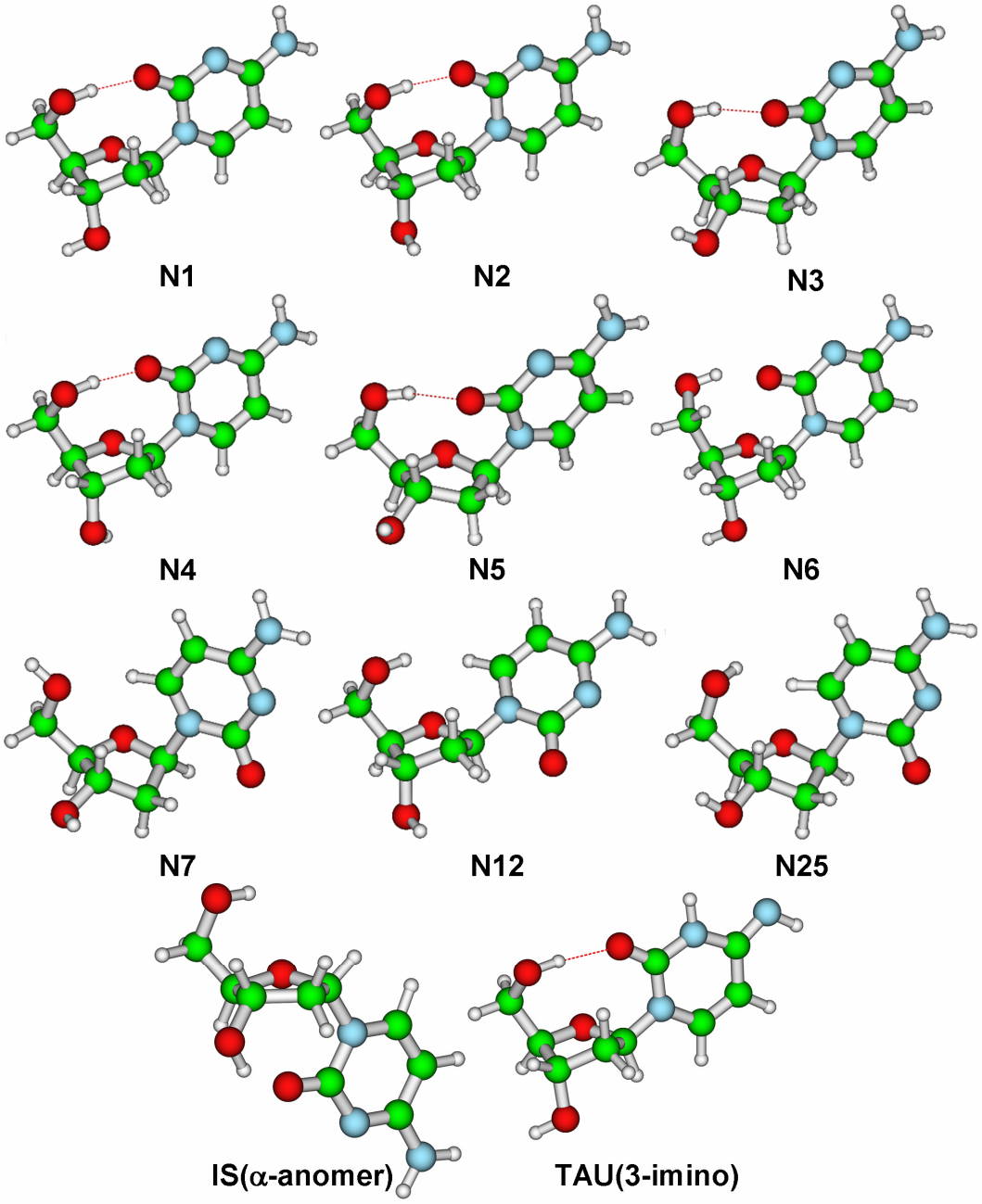


Figure 7.3: MP2 geometries of selected structures of 2'-deoxycytidine. The dotted lines indicate the intramolecular hydrogen bonds.

In addition to the conformational degrees of freedom we consider α- and β-anomers (see Figure 7.2a) and tautomers of cytosine. The β-anomer of dC, in which the base and the C4'-C5' bond are on the same side of the sugar ring, is common in DNA and for this reason is routinely used in computational studies.

Much less is known about the α-anomer of dC, in which the base and the C4'-C5' bond are on opposite sides of the sugar ring. It was reported that the α-anomer of dC could also be obtained during the process of direct synthesis of the β-anomer of dC [26], and very recently a method was developed to separate these two isomers [27]. With regard to different tautomers of cytosine, Xia et al. found in their theoretical study [28],which was performed at the B3LYP/6-311+G**//B3LYP/6-31G* level, that dC with the non-canonical 3-imino tautomer (see the bottom of Figure 7.3) is less stable than the canonical dC by 2.3 kcal/mol, and they found a transition barrier as high as 39.8 kcal/mol.

Foloppe et al. performed a series of theoretical studies on conformational properties of nucleosides [12-14]. Their MP2/6-31G* results indicate that for the *gauche*$^+$ conformations of dC, the *anti* conformation is more stable than the *syn* conformation by more than 7 kcal/mol with either the C2'-*endo* or C3'-*endo* sugar puckers, and the C3'-*endo*/*syn* conformation is more stable than the C2'-*endo*/*syn* conformation by 0.6 kcal/mol [12]. Their MP3/6-31G*//MP2/6-31G* transition barrier from C3'-*endo*/*anti* to C2'-*endo*/*anti* was 3.3 kcal/mol [14]. For the C2'-*endo/gauche*$^+$ conformations, these authors found that the MP2/6-31G* transition barrier from *anti* to *syn* was 8.4 kcal/mol [12]. A characteristic feature of these computational studies is that the β dihedral angle has been fixed at 180° [12]. This constraint was dictated by the structure of DNA, in which the O5' atom is engaged in the sugar-phosphate bond and the H13 atom is missing. Foloppe et al. explored variations of energy with respect to only these geometrical degrees of freedom that are available when nucleosides are embedded in DNA. Their valuable results might be not sufficient to predict the structural properties of nucleosides in the gas phase. Indeed, in two other theoretical studies of free canonical dC [29, 30], in which different combinations of *syn*/*anti*, C2'-*endo*/C3'-*endo*, and *gauche*$^+$/*gauche*$^-$/*trans* conformations were considered without any geometrical constraint, it was found that the *syn* rather than *anti* conformation is the most stable. It confirms that computational results based on molecular geometries dictated by the structure of DNA or the structure of the corresponding molecular crystal might be misleading when interpreting experimental results obtained in gas phase studies.

Kumari et al. measured gas phase acidities of four deoxyribonucleosides by applying the extended kinetic method [20], and the value reported for dC was 1409 kJ/mol. These authors also performed calculations at the B3LYP/6-

311+G**//B3LYP/6-311G** level based on a C3'-*endo*/*anti* conformation of dC. They found that the H2 site of cytosine is the most susceptible to deprotonation and they reported a gas phase acidity of 1409 kJ/mol, in perfect agreement with experiment. However, the agreement might be fortuitous in view of the limited probing of the conformation space of dC and the inaccuracy of the B3LYP functional.

The experimental reports [31-33] from the group of Sanche on single and double strand breaks in DNA induced by low-energy electrons triggered numerous experimental [34] and computational [35] studies on negatively charged biomolecules. The group of Bowen performed pioneering photoelectron spectroscopy (PES) experiments on gas phase anions of nucleic acid bases [36, 37], nucleosides [1],and nucleotides [38]. The measured electron vertical detachment energy (VDE) for dC of 0.87 eV [1] was interpreted in terms of preceding computational results obtained in the group of Schaefer [23]. This group performed extensive DFT calculations on anionic nucleosides [23] and the role of low-energy electrons in the glycosidic bond cleavage [39]. Similarly to Foloppe et al., they also selected conformations of isolated nucleosides as they appeared in X-ray crystal structures of DNA fragments. They characterized the valence bound anion of dC and determined its VDE to be 0.72 eV at the B3LYP/DZP++ level [23]. In view of the fact that B3LYP tends to overestimate the VDEs of nuclei acid bases by about 0.2 eV [40-44] the valence anion of dC deserves further attention.

The glycosidic bond cleavage in anions of dC was studied computationally by several groups [24, 39]. The group of Schaefer determined the activation barrier of 21.6 kcal/mol [39] and a similar result of 22.7 kcal/mol was obtained by Li et al. at the at B3LYP/6-31+G* level [24]. Here we extend these studies to the cationic species, which could also be generated upon the interaction of DNA with sources of high energy radiation.

The structure of the chapter is the following. Our combinatorial approach to the conformational problem and all caveats of our computations are discussed in Section 2. The most stable neutral and anionic structures of dC, their properties, and energetics of the glycosidic bond cleavage are presented in Section 3. The discussion of intra- and intermolecular hydrogen bonds in dC is presented in Section 4, and the chapter concludes with a Summary.

7.2 Methods

A desirable approach would be to explore the conformational/isomeric/tautomeric space of a nucleoside without any constraint, though the problem immediately becomes computationally very intensive. It has recently been suggested that combinatorial/computational approaches, which hinge on efficient automation of repetitive computational steps, might be useful when solving chemical problems [45-47]. In the current study we rely on the Systematic Screening of Conformers (SSC) tool [25] (see Chapter 4), which we use to create an initial library of conformers of canonical dC and we perform a limited exploration of the α-anomer and the 3-imino tautomer.

When building the library of conformers we assumed a planar structure of sugar and we performed initial optimizations for 324 initial structures with the B3LYP hybrid functional [48, 49] (for performance of B3LYP on a similar biomolecular system, see Chapter 6) and 6-31G* basis set [50]. These optimizations led to 65 non-planar minima which were further optimized at the B3LYP/6-31++G** level. Among the 28 most stable structures 16 are C2'-*endo* and 12 are C3'-*endo*. For each of them we created a counterpart with the opposite sugar ring puckering and we optimized these new 28 structures at the same level of theory, i.e., B3LYP/6-31++G**. This approach led to 12 new conformers. Finally, the 39 most stable conformers are further refined at the MP2 [51] level with aug-cc-pVDZ (AVDZ) [52] basis sets to account for the intramolecular dispersion interaction.

The gas phase acidity of dC is defined as the enthalpy change for the reaction

$$\mathrm{dC(g)} \rightarrow \mathrm{[dC\text{-}H]^-(g)} + \mathrm{H^+(g)}, \qquad (5.1)$$

in which the enthalpies of reactants and products were determined in the rigid rotor/harmonic oscillator approximation at T = 298 K and p = 1 atm.

The VIE values are defined as the differences between the electronic energies of dC^+ and dC determined at the optimized geometry of the neutral. The electron propagator theory, which describes electron correlation in a systematic way, has been proved to generate reliable electron binding energies for nucleobases and nucleotides [53]. In the current contribution, the outer-valence Green's function method (OVGF) [54] implemented in Gaussian [55] and the 6-31++G** basis set

was adopted to calculate the vertical ionization energy (VIE) of neutral dC.

Valence-bound anionic states of the canonical dC, the α-anomer (IS), and the 3-imino tautomer (TAU) (see Figure 7.3) were studied at the B3LYP/6-31++G** level of theory and the values of the VDE and adiabatic electron affinity (EA) were determined. The VDE values are calculated as differences between the electronic energies of dC and dC^- determined at the optimized geometry of the anion. The EA values include zero-point vibrational corrections and are based on differences in electronic energies of the neutral molecule and the anion at their respective optimized geometries. For all conformers of anionic canonical dC we assume that the neutral conformer formed in the process of electron photodetachment can relax to the most stable neutral canonical structure. In other words, the barriers separating various conformers of the neutral canonical dC are assumed to be low. For this reason we referenced the reported EA values with respect to the neutral conformer N2 (see Figure 7.3), which is the most stable neutral canonical conformer at the B3LYP/6-31++G** level. For the IS and TAU anions, the EAs are determined with respect to the respective neutrals as the barriers for interconversions are typically larger for isomers than for conformers.

It has been recognized that neutral molecules with dipole moments larger than 2.5 D can support dipole-bound anions [56, 57]. The five most stable conformers of neutral canonical dC have dipole moments larger than 6 D. In addition, the dipole moment of IS is as large as 8.5 D. Thus we also studied dipole-bound anionic states of dC for two selected structures. One is based on the conformer N2 [58] with the B3LYP/6-31++G** dipole moment of 7.1 D, the other is IS. Based on previous experience [59, 60], we supplemented the original AVDZ basis set with additional basis functions with small exponents to accommodate the diffuse charge distribution of the excess electron. The orientation of the dipole roughly coincides with the O2-C2 bond of cytosine. Thus the five-term *s* and five-term *p* sets of diffuse functions, with a geometric progression ratio of 5.0, were centered at the C5 atom. The B3LYP functional overestimates the VDE values for dipole-bound states [61]. Thus, the electronic structure calculations were performed at the MP2 level, which takes into account dispersion-type interactions between the excess electron and electrons of the neutral molecule [62]. The MP2 geometry optimizations were performed for the neutral N2 and IS and for the corresponding dipole-bound anions. The reported values of the EA do not include zero-point vibrational corrections.

The singly occupied molecular orbitals (SOMOs) of the valence and dipole-bound anions are drawn with GaussView. For valence anions, we use an isovalue of 0.02 $Bohr^{-3/2}$. For dipole-bound anions, we have used isovalues that reproduce 80% of the excess electron total charge [63].The OpenCubMan tool has been used to determine the isovalues [64].

The glycosidic bond cleavage was studied for the neutral, anionic and cationic dC. An estimation of the transition state structure was obtained by producing an energy profile for the C1'-N1 bond elongation with a step of 0.1 Å (1.47 Å < $R_{C1'-N1}$ < 3.47 Å) and B3LYP/6-31++G** partial optimizations of the molecular structure. The geometries of the neutral, anionic and cationic N2 conformers were used as starting points of the energy profiles. We selected the highest energy point on the potential energy profile as an initial guess of the transition state structure, and we performed a full transition state geometry optimization using the Berny algorithm [65]. We have verified that the resulting stationary points are characterized by only one imaginary vibrational frequency and intrinsic reaction coordinate (IRC) [66] calculations have been performed to confirm the nature of reactants and products [67]. Finally, all stationary points identified for the bond cleavage reactions of the three species considered were subjected to further MP2/AVDZ geometry optimizations. It should be pointed out that significant geometry relaxations might be observed during the MP2 geometry optimizations. In these cases we adopt the MP2 geometries and energies in our discussions.

We have also performed calculations for the molecular crystal of dC to clarify the competition between the intra- and intermolecular hydrogen bonds. The experimental unit cell [19] contains two dC molecules, both in the *anti* conformation. For comparison, we created a hypothetical crystal structure with two *syn* dC molecules in the unit cell. When constructing the structure we made an effort to engage the *syn* molecules in the most favorable intermolecular hydrogen bonding scheme, though the construction is heuristic. The purpose of working with this hypothetical structure is to contrast the stability of *syn* and *anti* conformers in the gas phase and the crystalline environment. Periodic DFT calculations have been performed with the Vienna Ab initio Simulation Package (VASP) [68, 69] using a plane wave basis set with a cutoff energy of 400 eV. The Perdew-Burke-Ernzerhof (PBE) functional [70] with the projector augmented wave (PAW) method [71, 72] was used to solve the Kohn-Sham equations and

determine the relative stability of the two phases. The Monkhorst-Pack scheme [73] was used to generate 4 × 4 × 4 grids of k-points to sample the Brillouin zones.

The MP2 optimizations of the neutral species with Gaussian sets were performed with NWChem 5.0 [74] or Gaussian 03 [55], and all other gaseous calculations were performed with Gaussian 03 [55]. Molden [75] was used to visualize the molecular structures.

7.3 Results

7.3.1 Structures

structure	E^{B3LYP}	E^{MP2}	E^{B3LYP} $+\Delta E_{0,vib}$	E^{B3LYP} $+\Delta H_{298,corr}$	E^{B3LYP} $+\Delta G_{298,corr}$
N1	0.000	0.000	0.000	0.000	0.000
N2	-0.003	0.109	0.053	0.021	0.115
N3	1.244	2.009	1.053	1.223	0.804
N4	1.808	2.049	1.689	1.781	1.487
N5	1.349	2.097	1.059	1.290	0.661
N6	3.308	2.733	3.299	3.362	3.305
N7	3.953	3.522	3.900	3.972	3.833
N12	2.983	3.716	2.473	2.808	1.290
N25	3.590	4.810	3.078	3.520	1.445
TAU(3-imino)	3.165	2.262	3.822	3.538	4.037
IS(α-anomer)	-0.937	-0.129	-0.661	-0.629	-0.961

structure	μ_N^{B3LYP}	VIE^{OVGF}
N1	6.4	8.585
N2	7.1	8.593
N3	7.0	8.615
N4	8.3	8.699
N5	7.5	8.612
N6	6.1	-
N7	7.0	8.016
N12	4.6	-
N25	6.1	-
TAU(3-imino)	4.9	-
IS(α-anomer)	8.5	-

Table 7.1: Relative MP2/AVDZ and B3LYP/6-31++G** energies (in kcal/mol), dipole moments (μ_N^{B3LYP} in Debye) and OVGF/6-31++G** VIEs (in eV) for selected structures of neutral 2'-deoxycytidine.

The geometries and relative energies of selected structures of neutral dC are shown in Figure 7.3 and Table 7.1, respectively. The 7 most stable conformers of canonical dC are named according to their relative stability at the MP2 level, e.g., the conformer N2 is the second most stable. For the sake of future discussion we also include the N12 conformer, which is related to N2 through a *syn*-to-*anti* transition, N25, which was extracted from the DNA structure and which was used in previous theoretical studies [23, 39], an α-anomer (denoted as IS), and a 3-imino tautomer (denoted as TAU). The main trends in relative stabilities are similar between the MP2 and B3LYP energies, though the exact orderings of conformers differ. The zero-point vibrational and thermal corrections to relative stabilities should definitely be taken into account when considering the population of various conformers in the gas phase. A higher than MP2 level of theory might be required for ultimate predictions.

There are, however, qualitative predictions that are not sensitive to differences between the MP2 and B3LYP methods. First we focus attention on the canonical dC, N1-N7, N12 and N25.The most stable conformers are *syn* rather than *anti*, see Figure 7.3 and Table 7.1. Indeed, the N1-N5 structures are *gauche*$^+$/*syn* and they are stabilized by strong intramolecular hydrogen bonds involving the O5'-H13 proton donor and the O2 proton acceptor. The intramolecular hydrogen bond is weaker in N6 due to the *trans/syn* conformation. The most stable *anti* canonical conformer proves to be N7, which is less stable than N1 by more than 3.5 kcal/mol at both the MP2 and B3LYP levels and with or without zero-point and thermal corrections. It implies that *anti* canonical conformers might be irrelevant in gas phase conditions, even though the results of X-ray crystalline experiments [19] suggest that both dC molecules in the unit cell are *anti*. This point will be further discussed in Section 7.3.5. Here we emphasize that the *syn* conformations permit formation of the O2···H13 hydrogen bond while the most stable *anti* conformation, N7, support only a much weaker hydrogen bond involving the C6-H4 proton donor and the O5' proton acceptor. Two other studies [29, 30] predicted that the *syn* conformations might be more stable than the *anti* conformations, but their relative difference in stability was reduced from 3.4 to 1.0 kcal/mol upon the methodological improvement from the B3LYP/6-31G* to an approximate MP2/6-31G* level [30]. It should be pointed out that the C2'-*endo*/*syn* conformations might be further stabilized by a weak hydrogen bond involving the C2'-H6 proton donor and the O2 proton acceptor.

Another important observation, which applies to the neutral canonical dC and is not sensitive to methodological differences, is that conformers with the C2'-*endo* sugar puckering should dominate in the gas phase. Indeed, among the seven most stable canonical conformers reported in Figure 7.3 and Table 7.1, four of them (N1, N2, N4 and N6) are C2'-*endo* and the most stable C3'-*endo* conformer, N3, is less stable than N1 by 0.8-1.0 kcal/mol. If one considers the *anti* conformations only then an opposite conclusion emerges: N7 is C3'-*endo* and this is in good agreement with past computational results [14, 29, 76].

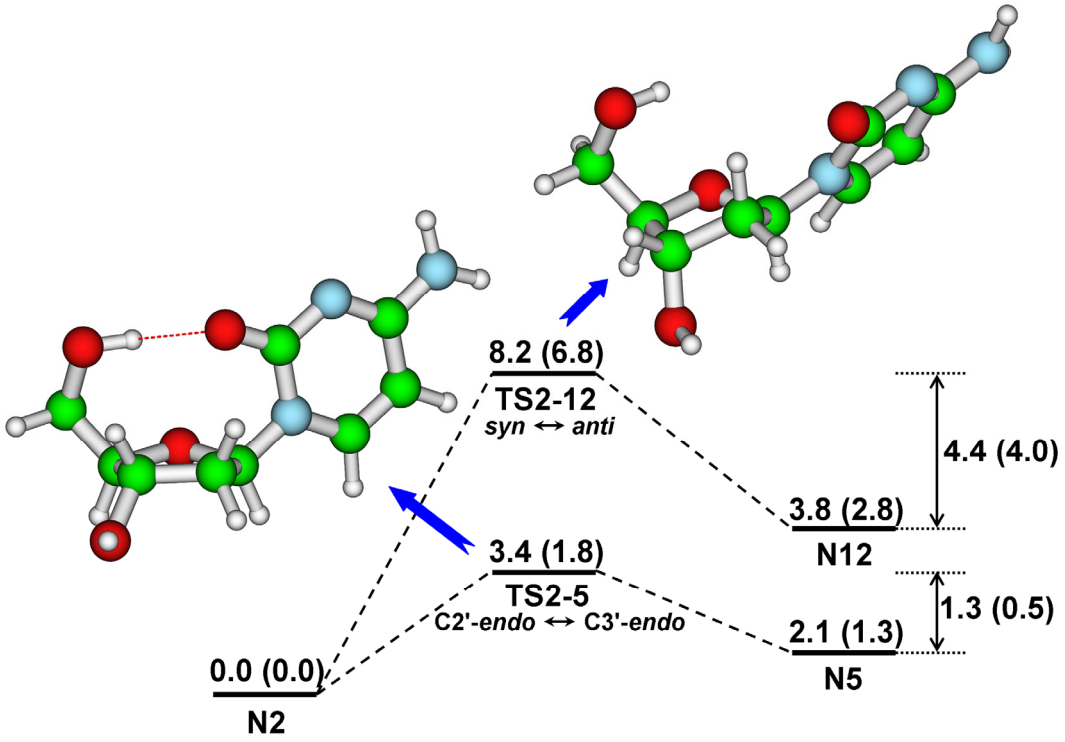


Figure 7.4: Interconversions between the *syn*/*anti*, and C2'-*endo*/C3'-*endo* conformations. Relative energies and transition barriers, which are determined at the MP2/AVDZ and B3LYP/6-31++G** (in parentheses) levels, are in kcal/mol. Only transition state structures are visualized.

We have demonstrated so far that the C2'-*endo*/*syn* conformers are dominant in the gas phase of canonical dC and we discussed their stabilities with respect to the C3'-*endo*/*anti* conformers. Here we discuss the height of barriers for the *syn*-to-*anti* and C2'-*endo*-to-C3'-*endo* transitions. We identified two transition state structures illustrated in Figure 7.4. The first is denoted as TS2-5 and it connects the N2 and N5 conformers through a C2'-*endo*-to-C3'-*endo* transition. The second is denoted as TS2-12 and it connects the N2 and N12 conformers through a *syn*-

to-*anti* transition. The MP2 and B3LYP values of the barrier heights are also reported in Figure 7.4.

It was proposed that there are two possible pathways for the *syn*-to-*anti* transition, one through $\chi = 0^{\circ}$, and another through $\chi = 120^{\circ}$, with the latter characterized by a lower barrier [12]. Our transition state structure TS2-12 is characterized by $\chi = 112^{\circ}$, see Figure 7.4. The MP2 barriers for the forward and reverse reactions are 8.2 and 4.4 kcal/mol, respectively, and the B3LYP barriers are systematically lower. The barrier heights reported by Foloppe et al. [12] are quite different, which results from geometrical constraints imposed by the authors on the orientation of the O5'-H13 group.

Three possible pathways for the C2'-*endo*-to-C3'-*endo* transition were proposed [14]. A pathway through the O4'-*endo* conformation has the lowest barrier and two other pathways proceed through pseudorotations of the sugar unit. Indeed, our TS2-5 is O4'-*endo*, see Figure 7.4, and the MP2 barriers for the forward and reverse reactions are 3.4 and 1.3 kcal/mol, respectively, while the B3LYP barriers are systematically lower. We conclude that the barriers for the C2'-*endo*-to-C3'-*endo* transitions are much smaller than for the *syn*-to-*anti* transitions. The puckering of the sugar ring should be viewed as a dynamical process which favors the C2'-*endo* orientations in the gas phase at standard conditions. The small barrier is reflected by a small imaginary frequency at TS2-5, which is only 49i cm^{-1}.

The α-anomer of dC is more stable than the canonical forms by a fraction of kcal/mol, see Table 7.1 and Figure 7.3. Unlike the most stable conformers of canonical dC, which are either C2'-*endo* or C3'-*endo*, IS is O4'-*endo*. Its stability can be justified only partially by a hydrogen bond between the O3'-H8 proton donor and the O2 proton acceptor. The stability of α-anomers could be further improved by a more systematic probing of their conformational space. It is an intriguing possibility that α-anomers could be formed upon transferring of canonical dC to the gas phase, see Section 7.3.3. The relative stability of α- and β-anomers will be addressed in our future study.

The 3-imino tautomer of dC, denoted TAU in Table 7.1 and Figure 7.3, is *syn* and it is stabilized by an intramolecular hydrogen bond involving the O5'-H13 proton donor and the O2 proton acceptor. However, it is less stable than N1 by 2-3

kcal/mol. This tautomer was also found unstable in the B3LYP study by Xia et al. [28], though these authors considered *anti* conformations only. The stability of the 3-imino tautomer could be further improved by a more systematic probing of its conformational space. A significant barrier of 39.8 kcal/mol for the unimolecular amino-imino tautomerization of dC was found by Xia et al. [28]. This barrier could be significantly lowered in binary events involving an intermolecular proton transfer [77]. Indeed, the anionic imino tautomer of 5-hydroxy-2'-deoxycytidine, an analog of dC, was observed by Suen et al. in a protic solvent in the ultraviolet resonance Raman spectroscopy experiment [78].

7.3.2 Properties

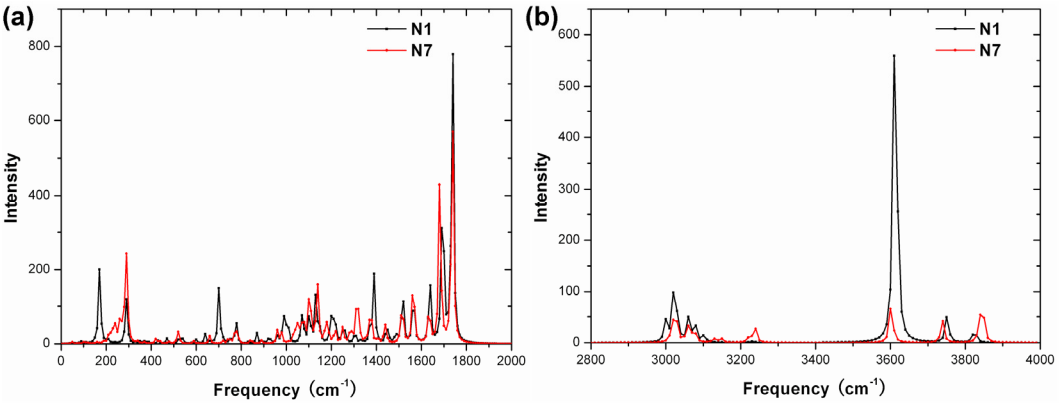


Figure 7.5: IR spectra of conformers N1 (in black) and N7 (in red), determined at the B3LYP/6-31++G** level, in the (a) low frequency, 0~2000 cm^{-1}, and (b) high frequency, 2800~4000 cm^{-1} range.

Our results indicate that a *syn* conformer of canonical dC dominates in the gas phase rather than an *anti* conformer identified in X-ray structural studies. We would welcome an experimental confirmation of our predictions. Such a confirmation could be based on differences in infrared (IR) spectra of the most stable *syn* and *anti* conformers, i.e., N1 and N7, respectively. N1 is characterized by a strong intramolecular hydrogen bond between O5'-H13 and O2, see Figure 7.3. On the other hand, N7 is stabilized by a weak intramolecular hydrogen bond between O5' and C6-H4. Indeed, these two intramolecular hydrogen bonds differentiate their IR spectra, see Figure 7.5. Inasmuch as both conformers provide similar spectra in the low frequency (0~2000 cm^{-1}) region, see Figure 7.5a, they differ markedly in the high frequency (2800~4000 cm^{-1}) region, see Figure 7.5b. In the latter region the IR intensities of different peaks are quite similar for N7,

whereas there is a very strong peak for N1 with the calculated IR intensity 563 km/mol. This peak is five times stronger than other peaks in this region. The strong peak results from a superposition of two transitions with similar vibrational frequencies. The first transition is associated with the symmetric stretching vibration of the amino group, and the other is the stretching vibration of O5'-H13. The latter vibration is red-shifted in N7 by about 250 cm^{-1} in comparison with N1, which is a classical manifestation of the O5'-H13···O2 intramolecular hydrogen bond. We suggest that a strong IR peak in the 2800~4000 cm^{-1} region can be regarded as a fingerprint of the *syn* conformer.

Conformer	Site of Deprotonation	gas phase acidity		
		B3LYP/6-31++G**	Ref.[a]	MP2/AVDZ
N1	H1	1452.3	-	-
	H2	1429.1	1429.4	1411.7
	H8	1510.6	-	-
	H13	1469.7	-	-
N7	H2	1443.7	1445.8	1432.6
?	?		Expt.[b] 1409±2.5	

Table 7.2: Gas phase acidity (in kJ/mol) of conformers N1 and N7. [a] B3LYP/6-311+G**//B3LYP/6-311G** theoretical method which was used in Ref. 20. [b] experimental value taken from Ref. 20.

Kumari et al. [20] reported an experimental deprotonation enthalpy of dC of 1409±2.5 kJ/mol. Their parallel B3LYP/6-311++G** study was focused on the N7 conformer, unraveled that the H2 site of cytosine is the most susceptible to deprotonation, and led to a remarkably accurate computational value of the deprotonation enthalpy of 1409 kJ/mol. In view of the fact that N7 is less stable with respect to N1 by 3.5-3.9 kcal/mol (see Table 7.1) we readdressed the problem and our MP2 and B3LYP results obtained for the most stable *syn* and *anti* conformers, i.e., N1 and N7, respectively, are summarized in Table 7.2. The H2 site of the N1 conformer proves to be the most acidic among the four sites considered and the MP2/AVDZ value of deprotonation enthalpy of 1411.7 kJ/mol is in excellent agreement with the experimental value, while the B3LYP results are systematically larger by 10-20 kJ/mol. Our results also indicate that the earlier computational results were flawed [79] and the corrected B3LYP deprotonation enthalpy of the H2 site of the N7 conformer is 1445.8 and 1443.7 kJ/mol with the 6-311G** and 6-31++G** basis sets, respectively. Not only the N1 conformer is more stable than the N7 conformer but also the $[dC-H]^-$ product is more stable for

the N1 conformer by 21 kJ/mol at the MP2/AVDZ level.

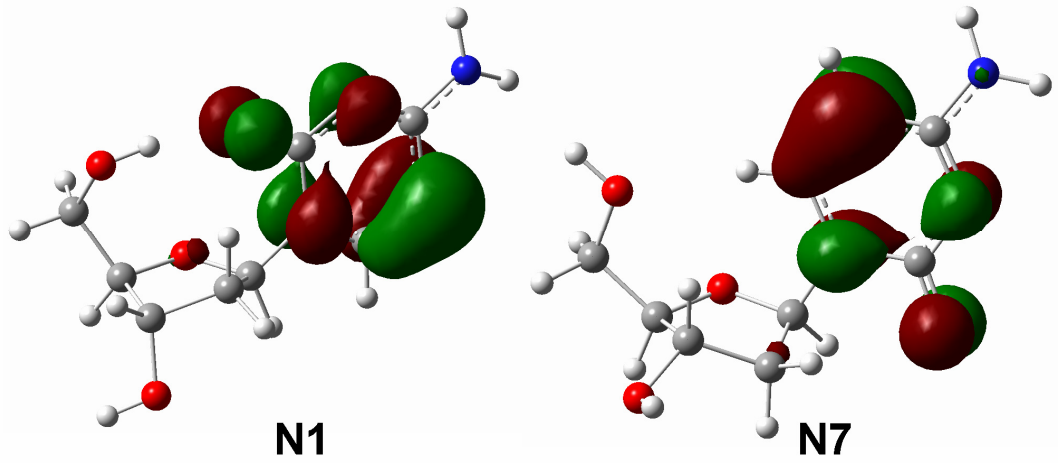


Figure 7.6: The isosurface of HOMO of conformers N1 and N7 determined at the HF/6-31++G** level, with an isovalue of 0.05 $Bohr^{-3/2}$.

The OVGF/6-31++G** values of the VIE are reported in Table 7.1 for selected conformers of dC. The VIE values for the most stable *syn* conformers cluster in a narrow range of 8.59-8.70 eV and are not sensitive to the puckering of the sugar ring. The VIE value for the most stable *anti* conformer (N7) is, however, significantly smaller and amounts to 8.02 eV. The VIE value of dC has not been experimentally measured yet, but our results suggest that its value can be used to discriminate the *anti* and *syn* conformers in the gas phase. The experimental values of the VIE for deoxyribose and cytosine are 10.51 eV [80] and 8.89 eV [81], respectively. Note that due to the existence of various tautomers in the gas phase [81], the measured VIE values might not be related exclusively to the canonical tautomers. The computed VIE value for dC of 8.59 eV is closer to the experimental VIE of cytosine, which implies that the HOMO orbital of dC is localized on cytosine. Indeed, the plots of the Hartree-Fock HOMOs of N1 and N7, see Figure 7.6, demonstrate that these π orbitals are localized primarily on the base. We suggest that the larger value of the VIE for the *syn* conformers results from the hydrogen bond between O5'-H13 and O2. The protic H13 stabilizes the HOMO orbital of the *syn* conformers, which results in their larger VIE values in comparison with the most stable *anti* conformer, N7. This orbital picture is confirmed by a difference in the VIE values determined at the Koopmans' theorem level of 0.6 eV, which is only slightly modified by orbital relaxation and electron correlation effects.

7.3.3 Anions

Structure	EA	VDE
AIS	0.46 (0.43)	0.90
AN4	0.35	0.86
AN2	0.34	0.76
AN1	0.33	0.68
AN5	0.30	0.82
AN9	0.28	0.88
AN25[a]	0.16	0.71
ATAU	0.07 (0.23)	0.28
Expt.[b]	~0.5	0.87

Table 7.3: The B3LYP/6-31++G** values of VDEs and EAs (in eV) for selected valence anions compared with the experimental results. The EAs are calculated with respect to N2. For AIS and ATAU, additional values of EA, which are calculated with respect to IS and TAU, respectively, are given in parentheses.
[a] AN25 corresponds to the structure used in Ref. 23.
[b] Experimental results were taken from Ref. 1.

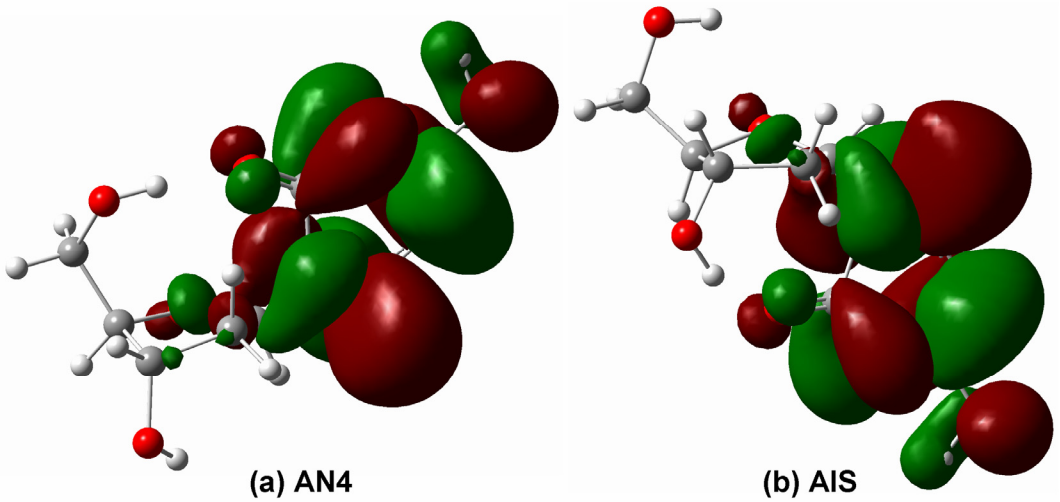


Figure 7.7: The isosurfaces of SOMOs of valence anions of (a) AN4 and (b) AIS determined at the B3LYP/6-31++G** level, with an isovalue of 0.02 Bohr$^{-3/2}$.

The B3LYP/6-31++G** values of the VDE and EA for the valence anions of dC are reported in Table 7.3. We consider there the five most stable valence anions of canonical dC identified in our search. In addition we consider valence anions of: N25, an *anti* conformer considered in a previous computational study [23], the α-

anomer, and the 3-imino tautomer. The names of the valence anions are preceded by the letter "A" and are followed by the name of the neutral molecule the valence anion is related to, e.g., AN4 indicates a valence anion related to the fourth most stable canonical dC, N4. The valence anions are ordered in Table 7.3 according to the descending values of the EA. The plots of the SOMOs for AN4 and AIS are reported in Figure 7.7 and demonstrate that these are indeed valence anions with the excess electron occupying a π^* orbital localized on the base.

AIS and AN4 are the two most stable valence anions in terms of both EA and VDE. Their respective VDE values of 0.90 and 0.86 eV are in good agreement with the experimental VDE value of 0.87 eV [1]. We expect that intramolecular hydrogen bonds between the O2 proton acceptor of cytosine and the O3'H8 and O5'H13 proton donors of AIS and AN4, respectively, contribute to the stability of these anions. In particular the protic hydrogens (H8 or H13) stabilize the excess electron localized on the base, see Figure 7.7. The reported EA values of 0.43 and 0.35 eV, respectively, are significant, and larger than the methodological uncertainty of B3LYP/6-31++G**. Thus our results confirm that valence anions of dC are adiabatically bound [23]. It remains an open question whether the source of valence anions used in Ref. 1 is sufficiently harsh to allow formation of the AIS isomers.

It is known that the B3LYP values of the VDE for valence anions of isolated nucleic acid bases are overestimated by ca. 0.2 eV compared with more accurate CCSD(T) result [40-44]. Thus an excellent agreement between the experimental VDE [1] and the calculated values for AIS and AN4 might be still fortuitous. On the other hand, it is not established yet how the presence of the sugar moiety affects the inaccuracy of B3LYP in predicting the VDE values for anionic nucleosides. These problems need to be resolved in future theoretical studies.

The PES spectrum of dC [1] was interpreted in terms of computational results for the valence anion of N25 [23]. These computational results preceded the PES experiment and their objective was to characterize the valence anion of dC in an environment that is characteristic for DNA rather than to predict the gas phase PES spectrum. The N25 *anti* conformer is not a good model for the gas phase valence anion of dC, see Table 7.3. Its EA is only 0.16 eV, which implies that AN25 is less stable than AN4 and AIS by 4.65 and 7.90 kcal/mol, respectively. Thus the population of AN25 should be negligible at standard conditions in

thermodynamic equilibrium. Also the valence anion ATAU, which is based on the noncanonical tautomer 3-imino, is characterized by small values of EA (0.07 eV) and VDE (0.28 eV). Thus ATAU does not contribute to the PES spectrum of dC.

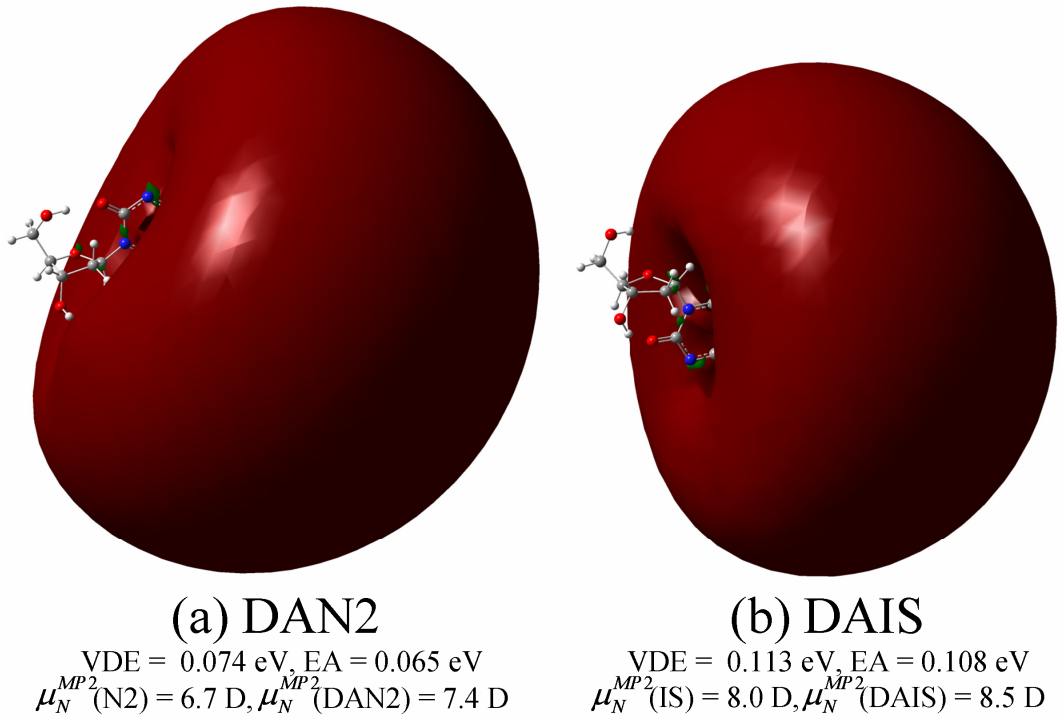


Figure 7.8: The isosurfaces of SOMOs of dipole-bound anions (a) DAN4 and (b) DAIS determined at the HF/AVDZ+5sp level. 80% of the total excess charge is reproduced for each anion. The VDEs and EAs are calculated at the MP2/AVDZ+5sp level. No ZPVE corrections are included in the EA values.

The dipole-bound anions supported by two neutral forms of dC, N2 and IS, are characterized in Figure 7.8 and labeled DAN2 and DAIS, respectively. The N2 and IS neutrals were selected based on their significant dipole moments and overall thermodynamic stability, see Table 7.1. The B3LYP dipole moments are larger than the MP2 counterparts by 0.4-0.5 D, see Table 7.1 and Figure 7.8. The molecular framework distorts upon the excess electron attachment and the dipole moment of the neutral increases by 0.7 and 0.5 D for N2 and IS, respectively, see Figure 7.8. The MP2 values of VDE are 113 and 74 meV for DAIS and DAN2, respectively. Thus these anions are amenable to PES characterization, providing the dipole-bound states could be formed in the anionic beam. The respective MP2 values of EA are 108 and 66 meV, whereas the B3LYP values of EA for valence anions of dC are in the 0.5-0.3 eV range, see Table 7.3. We anticipate that the

B3LYP method overestimates the excess electron binding energy in these valence anions but we are hesitant to speculate about the relative stability of dipole- vs. valence-type anions of dC. The plots of SOMOs for DAN2 and DAIS were obtained with isovalues selected to reproduce 80% of the excess charge. The SOMO of DAIS is more compact, which is consistent with the larger value of the VDE.

7.3.4 Effect of Excess Charge on the Thermodynamics and Kinetics of Glycosidic Bond Cleavage

The glycosidic bond cleavage is an important pathway of DNA damage induced by high energy radiation. Here we discuss the cleavage for the neutral, anionic, and cationic dC. The energetics and geometries of the reactants, transition states and products are shown in Figure 7.9, in which all the geometries are determined at the MP2/AVDZ level. The relative energies are defined with respect to the reactant of each species and are determined at both the MP2/AVDZ and B3LYP/6-31++G** (in parentheses) levels.

For the neutral dC, we find a significant activation barrier of 47.8 kcal/mol and the bond cleavage is endothermic by 16.4 kcal/mol, both are the MP2/AVDZ results. Initially we anticipated formation of two radical moieties and expected that the restricted Hartree-Fock or B3LYP approaches would not be proper to describe the cleavage. To our surprise the restricted approach works fine and the reason is that the cleavage of the glycosidic bond is accompanied by proton transfer from C2'H to the O2 atom of cytosine. In consequence, both products of decomposition are closed-shell moieties with meaningful Lewis structures, with the sugar ring being practically planar. Alternatively, one could interpret the reaction as a heterolytic bond cleavage followed by intermolecular proton transfer, which neutralizes the decomposition products. The IRC calculation from the reported transition state converges to the neutral conformer N5. This conformer is separated from N2 by TS2-5 with a barrier of 3.4 kcal/mol, see Figure 7.4. For this reason we use the N2 conformer as a reactant for neutral dC.

The barrier for cleavage of glycosidic bond in dC^- is 29.1 kcal/mol, hence lower than in the neutral dC by 19 kcal/mol. This is a very significant difference brought in by an excess electron. The barrier is reduced to 22.3 kcal/mol at the B3LYP

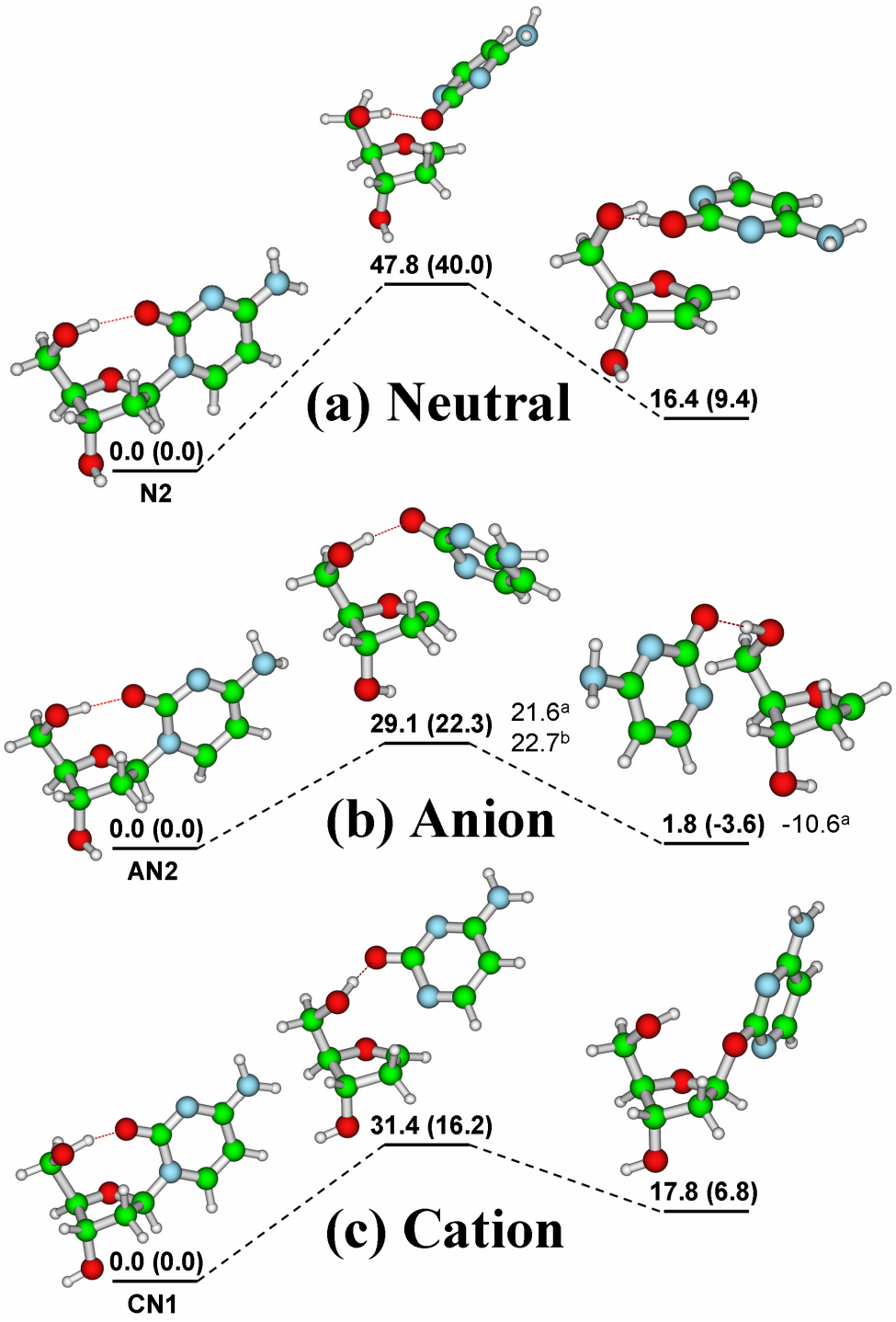


Figure 7.9: Relative energies and transition barriers (in kcal/mol), determined at the MP2/AVDZ and B3LYP/6-31++G** (in parentheses) levels, and MP2 geometries of glycosidic bond cleavage reactions of (a) neutral, (b) anionic, and (c) cationic species of dC. [a] Theoretical value taken from Ref. 39. [b] Theoretical value taken from Ref. 24.

level. This result is similar to those reported earlier, 21.6 kcal/mol in Ref. 39 and 22.7 kcal/mol in Ref. 24, both of which were obtained for *anti* conformers of dC^- at the B3LYP level. The glycosidic bond cleavage is nearly thermoneutral for dC^-, which is another important difference brought in by the excess electron, see also Ref. 39.

Much less is known about the glycosidic bond cleavage in dC^+. The MP2 barrier of 31.4 kcal/mol is comparable with this for dC^-, but the reaction is endothermic by 17.8 kcal/mol, similarly as for dC. The IRC calculation from the reported transition state converges to the N1 conformer of dC^+.

As it could be seen from Figure 7.9, there are some disagreements between the MP2 and B3LYP results about the activation barriers and heats of formation, e.g. the MP2 and B3LYP activation barriers for the glycosidic bond cleavage in dC^+ differ by about 15 kcal/mol. These disagreements might result from intrinsic discrepancies between the two methods. For example, B3LYP fails to describe dispersion effects. In addition, both methods are affected by intramolecular basis set superposition errors (BSSE), which are typically more profound at the MP2 than B3LYP level. Indeed, it was reported that B3LYP and MP2 predict different structures for some similar biomolecular systems due to missing dispersion in B3LYP results and a large BSSE in MP2 results [82, 83]. Similar intrinsic methodological discrepancies and basis set artifacts might affect the structures and energetics of the reactants, transition states and products involved in the glycosidic bond cleavage reactions in the neutral and charged dC. The results reported here can be further improved. For example, basis set artifacts can be suppressed by performing calculations in more complete basis sets. These effects will be explored in our future work.

To conclude, the glycosidic bond cleavage in neutral dC encounters a significant barrier of 47.8 kcal/mol, which is reduced by 16-19 kcal/mol for charged dC, either cationic or anionic. The reaction is nearly thermoneutral for dC^- and endothermic by 16-18 kcal/mol for dC^+ and dC.

7.3.5 Effect of Crystalline Environment on the Structure of dC

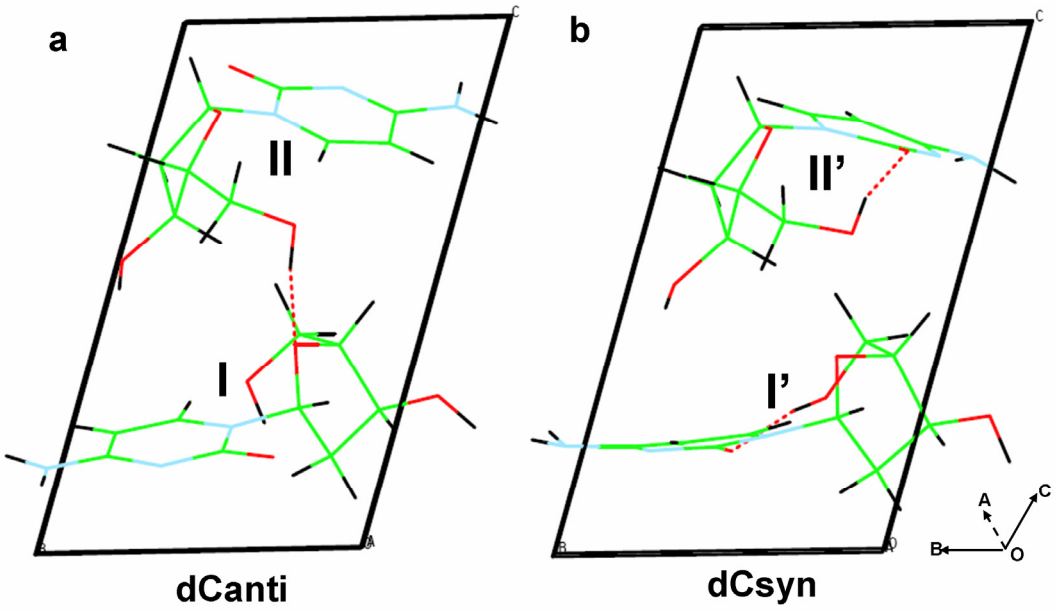


Figure 7.10: Crystal structure of dC, with schematics in the right bottom showing different lattice vectors.

The molecular crystal of 2'-deoxycytidine is characterized by a triclinic unit cell with two dC molecules [19] (see Figure 7.10a): Molecule I with a C2'-*endo*/*anti* conformation, and Molecule II with a C3'-*endo*/*anti* conformation. We label this unit cell as dCanti because of the *anti* conformation of both molecules. In order to understand why *anti* rather than *syn* conformations are favored in the molecular crystal of dC, we constructed a periodic structure based on a similar unit cell, but with *anti* conformers replaced with *syn* conformers, see Figure 7.10b. These are N2, which is C2'-*endo*/*syn* and will be denoted as Molecule I', and N5, which is C3'-*endo*/*syn* and will be denoted as Molecule II' (N2 and N5 served as reactants and products of the C2'-*endo*-to-C3'-*endo* transition, see Figure 7.4). We label this unit cell as dCsyn because of the *syn* conformation of both molecules.

Both periodic structures were optimized with the PBE exchange-correlation functional with lattice vectors and positions of the atoms in the unit cell available for relaxation. As expected, only minor relaxation took place for dCanti, as the initial structure was experimental. The optimized lattice parameters, a = 7.090 Å, b = 6.776 Å, c = 10.819 Å, α = 104.92º, β = 84.35º, and γ = 72.20º, are in good agreement with the experimental values, a = 7.285 Å, b = 6.866 Å, c = 11.074 Å, α = 104.19º, β = 84.53º, and γ = 72.26º, which shows the good performance of the PBE functional for descriptions of such biomolecular crystals.

The main finding is that the electronic energy of the periodic structure based on dCanti is lower than that based on dCsyn by 33.8 kcal/mol per unit cell. As illustrated in Table 7.4, the one-body energies of the isolated Molecules I, II, I', II' are more favorable for dCsyn by 6.3 kcal/mol. Moreover, the relative one-body energies obtained through calculations with the plane wave basis set and a large unit cell (30 Å × 30 Å × 30 Å) are consistent with those obtained through molecular calculations with localized basis functions and using the Gaussian code. The more favorable one-body energies for dCsyn are consistent with the main conclusion of Section 7.3.1 that due to formation of the *intramolecular* O5'-H13···O2 hydrogen bond the *syn* conformers are more stable than the *anti* conformers for an isolated dC molecule. The lower electronic energy of the dCanti periodic structure implies that *intermolecular* interactions favor *anti* conformers in the experimental structure of the molecular crystal of dC.

In Figure 7.11 we illustrate intermolecular hydrogen bonds that develop in the periodic structures dCanti (parts a, c and e) and dCsyn (parts b, d and f). In Table 7.4 we summarize the hydrogen bond acceptors and donors which are involved in hydrogen bonds present in both Figures 7.10 and 7.11. To make intermolecular hydrogen bonds which form along each lattice vector (**A**, **B** and **C**) more visible, we do not show hydrogen bonds which were already present in Figure 7.10. The unit cell was doubled along the **A** vector in parts a and b, whereas in parts c and d it was doubled along the **B** vector but only the interface between adjacent unit cells is displayed, and in parts e and f, we show the interface between two unit cells along the **C** vector.

The main difference in the network of intermolecular hydrogen bonds between dCanti and dCsyn involves a very basic N3 site of cytosine. Due to the *anti* conformation of the molecules in dCanti this site is available to engage in intermolecular hydrogen bonds. Indeed, a favorable cyclic hydrogen bonded motif is illustrated in Figure 7.11e, which involves two N3 proton acceptors and two N4-H1 proton donors of the Molecules I and II. The resulting 8-member ring structure is highlighted in Figure 7.11e. On the other hand, the N3 sites in the dCsyn periodic structure are not exposed enough to engage in intermolecular hydrogen bonds. This results from the "folded" nature of dC in all *syn* conformers, where "folding" is driven by the intramolecular O5'-H13/O2 hydrogen bond. The dC molecules "unfold" in the dCanti periodic structure. Apparently, the expense

of breaking of the intramolecular hydrogen bonds is compensated by formation of intermolecular hydrogen bonds.

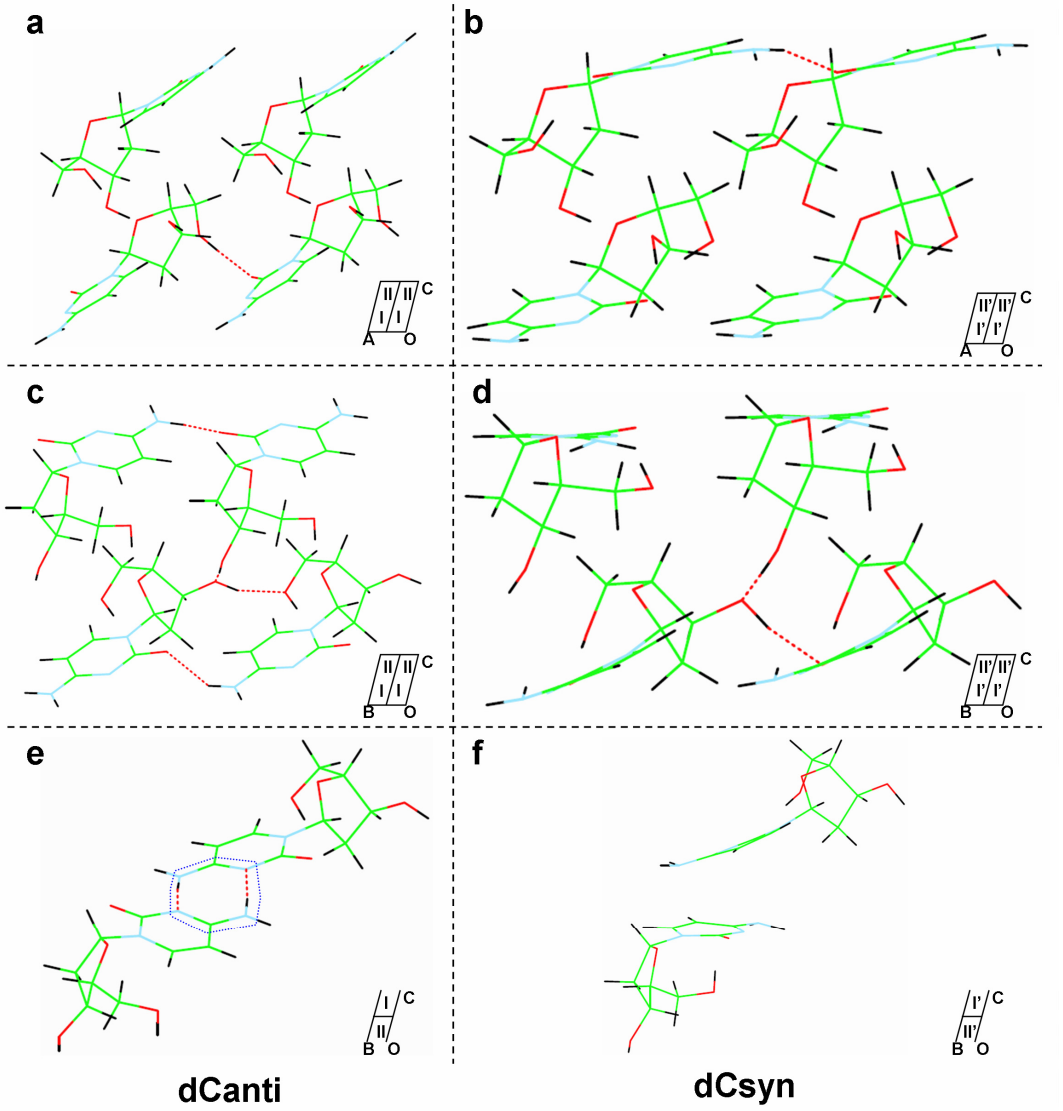


Figure 7.11: Hydrogen bonds in the crystal structure of dC, with schematics in the right bottom showing neighboring unit cells along different lattice vectors.

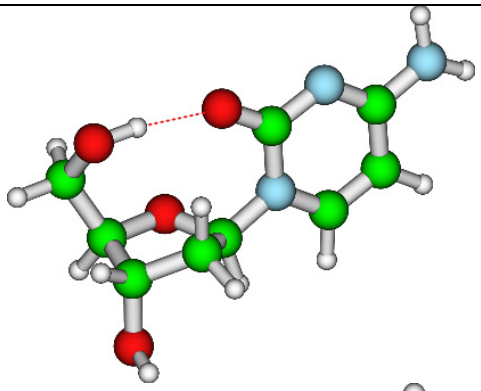
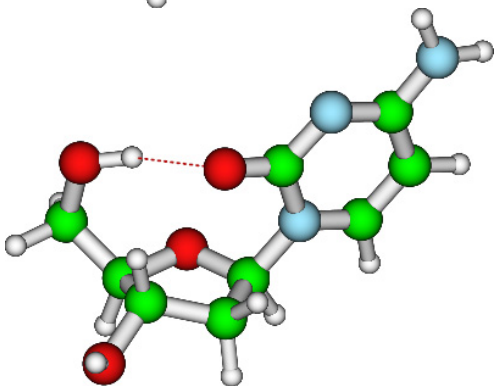
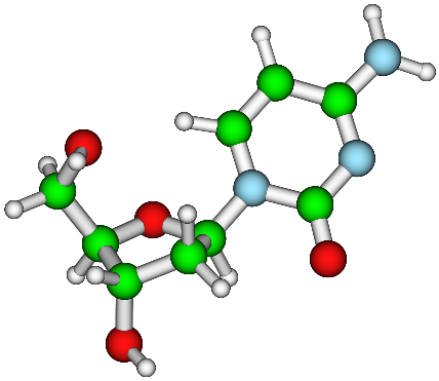
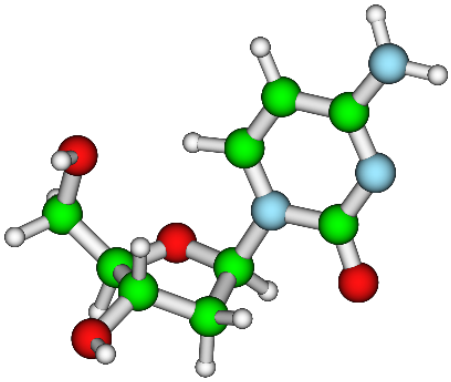
structure	PBE/plane wave	MP2/AVDZ	molecular geometry
N2	0	0	
N5	1.5	2.0	
Molecule I	5.0	4.9	
Molecule II	4.3	3.4	

Table 7.4: Relative energies (in kcal/mol) of conformers N2 and N5, and Molecule I and II with plane wave basis sets and localized Gaussian basis sets.

Figure	Hydrogen bond	
	Donor	Acceptor
10a (dCanti)	O5'-H13 (II)	O4' (I)
10b (dCsyn)	O5'-H13 (I')	O2 (I')
	O5'-H13 (II')	O2 (II')
11a (dCanti_A)	O5'-H13 (I_L)	O2 (I_R)
11b (dCsyn_A)	N4-H2 (II'_L)	O2 (II'_R)
11c (dCanti_B)	N4-H2 (II_L)	O2 (II_R)
	N4-H2 (I_R)	O2 (I_L)
	O3'-H8 (I_L)	O5' (I_R)
	O3'-H8 (II_R)	O3' (I_L)
11d (dCsyn_B)	O3'-H8 (I'_L)	O2 (I'_R)
	O3'-H8 (II'_R)	O3' (I'_L)
11e (dCanti_C)	N4-H1 (I)	N3 (II)
	N4-H1 (II)	N3 (I)

Structure	Number of hydrogen bonds
dCanti	8
dCsyn	5

Table 7.5: Intra- and intermolecular hydrogen bonds in dCanti and dCsyn. "L" and "R" indicate unit cells in the left and right hand side, respectively.

7.4 Summary

We have performed a computational study on 2'-deoxycytidine (dC), its conformers, selected isomers and tautomers. We considered neutral molecules as well as their anionic and cationic radicals. We also analyzed the effect of the crystalline environment on the conformational structure of dC. The calculations were performed at the density functional level of theory. Some molecular systems were also studied at the second order Moller-Plesset level. Electron vertical ionization energies were determined at the OVGF level [54]. The conformational space of canonical dC was explored using our tool for systematic screening of conformers.

Our results indicate that the neutral canonical dC in the gas phase favors *syn* rather than *anti* conformations. The thermodynamics dominance of *syn* conformations results from the formation of an intramolecular O5'-H13···O2 hydrogen bond. We have also demonstrated that the C2'-*endo* sugar puckering is favorable in the gas phase. We identified barriers for the *syn*-to-*anti* and C2'-*endo*-to-C3'-*endo* transitions at 8.2 and 3.4 kcal/mol, respectively, and we concluded that the puckering of the sugar ring should be viewed as a dynamical process which favors the C2'-*endo* orientations in the gas phase at standard conditions. The α-anomer of dC proved to be more stable by a fraction of kcal/mol than the β-anomer. The latter is the dominant building block of DNA. We have demonstrated that the IR spectra of the most stable *syn* and *anti* canonical conformers differ markedly in the region of NH/OH vibrations. The VIEs for the most stable *syn* conformers cluster in a narrow range of 8.59-8.70 eV whereas the VIE of the most stable *anti* conformer is smaller and amounts to 8.02 eV. The difference was interpreted through an orbital stabilizing effect resulting from the intramolecular O5'-H13···O2 hydrogen bond in *syn* conformers. The MP2 value of the deprotonation enthalpy of dC of 1411.7 kJ/mol is in good agreement with the experimental value of 1409±2.5 kJ/mol [20].

Our results have confirmed that the valence anions of dC are adiabatically bound [23]. The most stable valence anions proved to be the α-anomer and a *syn* canonical conformer. Their respective VDE values of 0.90 and 0.86 eV are in good agreement with the experimental VDE value of 0.87 eV. [1] Two dipole bound anions of dC have been characterized, one based on the α-anomer and another on a *syn* canonical structure. The respective MP2 values of VDE are 113 and 74 meV. Thus these anions are amenable to PES characterization, providing the dipole-bound states can be formed in the anionic beam.

The glycosidic bond cleavage has been studied for the neutral, anionic, and cationic dC. The barrier for cleavage is significant in the neutral dC, 47.8 kcal/mol, and it is reduced by 16-19 kcal/mol for charged dC, either cationic or anionic. The cleavage reaction is nearly thermoneutral for dC^- and endothermic by 16-18 kcal/mol for dC^+ and dC. For the neutral dC the cleavage is accompanied by proton transfer from C2'H to the O2 atom of cytosine.

Finally we analyzed the problem why dC molecules favor *anti* conformations in the crystalline phase but *syn* conformations in the gas phase. We have

demonstrated that intermolecular hydrogen bonds favor *anti* conformers in the experimental structure of the molecular crystal of dC. We suggested that the very basic N3 site of cytosine would not be able to engage in intermolecular hydrogen bonds if the crystal was built from *syn* conformers. In the case of *anti* conformers the N3 site is available to engage in intermolecular hydrogen bonds. We identified a cyclic hydrogen bonded structural motif, which involves two N3 proton acceptors and two N4-H1 proton donors.

References:

[1] Stokes, S. T.; Li, X.; Grubisic, A.; Ko, Y. J.; Bowen, K. H., Intrinsic Electrophilic Properties of Nucleosides: Photoelectron Spectroscopy of Their Parent Anions. *Journal of Chemical Physics* **2007,** 127, 084321(1)-084321(6).

[2] Schermann, J.-P., *Spectroscopy and Modeling of Biomolecular Building Blocks*. Elsevier: Amsterdam ; Oxford, 2007; p 479.

[3] Rizzo, T. R.; Park, Y. D.; Levy, D. H., A Molecular-Beam of Tryptophan. *Journal of the American Chemical Society* **1985,** 107, 277-278.

[4] Meijer, G.; de Vries, M. S.; Hunziker, H. E.; Wendt, H. R., Laser Desorption Jet-Cooling Spectroscopy of Para-Amino Benzoic-Acid Monomer, Dimer, and Clusters. *Journal of Chemical Physics* **1990,** 92, 7625-7635.

[5] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization-Principles and Practice. *Mass Spectrometry Reviews* **1990,** 9, 37-70.

[6] Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T., Matrix-Assisted Laser Desorption Ionization Mass-Spectrometry of Biopolymers. *Analytical Chemistry* **1991,** 63, A1193-A1202.

[7] Bald, I.; Dabkowska, I.; Illenberger, E., Probing Biomolecules by Laser-Induced Acoustic Desorption: Electrons at near Zero Electron Volts Trigger Sugar-Phosphate Cleavage. *Angewandte Chemie International Edition* **2008,** 47, 8518-8520.

[8] Wyttenbach, T.; Witt, M.; Bowers, M. T., On the Question of Salt Bridges of Cationized Amino Acids in the Gas Phase: Glycine and Arginine. *International Journal of Mass Spectrometry* **1999,** 183, 243-252.

[9] Price, W. D.; Jockusch, R. A.; Williams, E. R., Is Arginine a Zwitterion in the Gas Phase? *Journal of the American Chemical Society* **1997,** 119, 11988-11989.

[10] Chapo, C. J.; Paul, J. B.; Provencal, R. A.; Roth, K.; Saykally, R. J., Is Arginine Zwitterionic or Neutral in the Gas Phase? Results from Ir Cavity Ringdown Spectroscopy. *Journal of the American Chemical Society* **1998,** 120, 12956-12957.

[11] Alonso, J. L.; Cocinero, E. J.; Lesarri, A.; Sanz, M. E.; Lopez, J. C., The Glycine-Water Complex. *Angewandte Chemie International Edition* **2006,** 45, 3471-3474.

[12] Foloppe, N.; Hartmann, B.; Nilsson, L.; MacKerell, A. D., Intrinsic Conformational Energetics Associated with the Glycosyl Torsion in DNA:

A Quantum Mechanical Study. *Biophysical Journal* **2002,** 82, 1554-1569.

[13] Foloppe, N.; MacKerell Jr, A. D., Intrinsic Conformational Properties of Deoxyribonucleosides: Implicated Role for Cytosine in the Equilibrium among the a, B, and Z Forms of DNA. *Biophysical Journal* **1999,** 76, 3206-3218.

[14] Foloppe, N.; Nilsson, L.; MacKerell, A. D., Ab Initio Conformational Analysis of Nucleic Acid Components: Intrinsic Energetic Contributions to Nucleic Acid Structure and Dynamics. *Biopolymers* **2002,** 61, 61-76.

[15] Wales, D. J.; Doye, J. P. K., Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *Journal of Physical Chemistry A* **1997,** 101, 5111-5116.

[16] Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by Simulated Annealing. *Science* **1983,** 220, 671-680.

[17] Holland, J. H., *Adaptation in Natural and Artificial Systems*. MIT Press: Cambridge, 1992.

[18] Ling, S.; Yu, W.; Huang, Z.; Lin, Z.; Haranczyk, M.; Gutowski, M., Gaseous Arginine Conformers and Their Unique Intramolecular Interactions. *Journal of Physical Chemistry A* **2006,** 110, 12282-12291.

[19] Young, D. W.; Wilson, H. R., Crystal and Molecular Structure of 2'-Deoxycytidine. *Acta Crystallographica, Section B: Structural Science* **1975,** B31, 961-965.

[20] Kumari, S.; Devi, C. L.; Prabhakar, S.; Bhanuprakash, K.; Vairamani, M., Estimation of Gas-Phase Acidities of Deoxyribonucleosides: An Experimental and Theoretical Study. *Journal of the American Society for Mass Spectrometry* **2010,** 21, 136-143.

[21] Colson, A. O.; Besler, B.; Sevilla, M. D., Ab-Initio Molecular-Orbital Calculations on DNA Radical Ions .3. Ionization-Potentials and Ionization Sites in Components of the DNA Sugar-Phosphate Backbone. *Journal of Physical Chemistry* **1993,** 97, 8092-8097.

[22] Lu, L.-H.; Su, C.-C.; Hsieh, T.-J., Density Functional Calculations and Experimental Studies of the Sidearm Remote Controlling Effect in Eight Lariat Crown Ethers. *Journal of Molecular Structure* **2007,** 831, 151-164.

[23] Richardson, N. A.; Gu, J. D.; Wang, S. Y.; Xie, Y. M.; Schaefer, H. F., DNA Nucleosides and Their Radical Anions: Molecular Structures and Electron Affinities. *Journal of the American Chemical Society* **2004,** 126, 4404-4411.

[24] Li, X.; Sanche, L.; Sevilla, M. D., Base Release in Nucleosides Induced by Low-Energy Electrons: A Dft Study. *Radiation Research* **2006,** 165, 721-

729.

[25] SSC: A Tool for Constructing Libraries for Systematic Screening of Conformers. http://sscf.sf.net (accessed Sept 2, 2010). Ling, S.; Gutowski, M. J. Chem. Inf. Model. 2010, submitted.

[26] Fox, J. J.; Yung, N. C.; Wempen, I.; Hoffer, M., Nucleosides. Xii. Direct Synthesis of 2'-Deoxycytidine and Its α-Anomer. *Journal of the American Chemical Society* **1961,** 83, 4066-4070.

[27] Garcia, J.; Diaz-Rodriguez, A.; Fernandez, S.; Sanghvi, Y. S.; Ferrero, M.; Gotor, V., New Concept for the Separation of an Anomeric Mixture of α/B-D-Nucleosides through Regioselective Enzymatic Acylation or Hydrolysis Processes. *Journal of Organic Chemistry* **2006,** 71, 9765-9771.

[28] Xia, F.; Xie, H.; Cao, Z., Density Functional Study of Protonation of Deoxynucleosides: Electrophilic Active Sites and Proton Affinities. *International Journal of Quantum Chemistry* **2008,** 108, 57-65.

[29] Shishkin, O. V.; Pelmenschikov, A.; Hovorun, D. M.; Leszczynski, J., Molecular Structure of Free Canonical 2'-Deoxyribonucleosides: A Density Functional Study. *Journal of Molecular Structure* **2000,** 526, 329-341.

[30] Kolandaivel, P.; Knapp-Mohammady, M.; Suhai, S., Studies on Structure and Conformational Stability of Free Canonical 2'-Deoxyribonucleosides: Approximate Scc-Dftb and Lmp2 Methods. *International Journal of Quantum Chemistry* **2004,** 99, 28-38.

[31] Boudaiffa, B.; Cloutier, P.; Hunting, D.; Huels, M. A.; Sanche, L., Resonant Formation of DNA Strand Breaks by Low-Energy (3 to 20 Ev) Electrons. *Science* **2000,** 287, 1658-1660.

[32] Huels, M. A.; Boudaiffa, B.; Cloutier, P.; Hunting, D.; Sanche, L., Single, Double, and Multiple Double Strand Breaks Induced in DNA by 3-100 Ev Electrons. *Journal of the American Chemical Society* **2003,** 125, 4467-4477.

[33] Martin, F.; Burrow, P. D.; Cai, Z. L.; Cloutier, P.; Hunting, D.; Sanche, L., DNA Strand Breaks Induced by 0-4 Ev Electrons: The Role of Shape Resonances. *Physical Review Letters* **2004,** 93.

[34] Sanche, L., Low Energy Electron-Driven Damage in Biomolecules. *European Physical Journal D* **2005,** 35, 367-390.

[35] Simons, J., How Do Low-Energy (0.1-2 Ev) Electrons Cause DNA-Strand Breaks? *Accounts of Chemical Research* **2006,** 39, 772-779.

[36] Hendricks, J. H.; Lyapustina, S. A.; deClercq, H. L.; Snodgrass, J. T.; Bowen, K. H., Dipole Bound, Nucleic Acid Base Anions Studied Via

Negative Ion Photoelectron Spectroscopy. *Journal of Chemical Physics* **1996,** 104, 7788-7791.

[37] Li, X.; Bowen, K. H.; Haranczyk, M.; Bachorz, R. A.; Mazurkiewicz, K.; Rak, J.; Gutowski, M., Photoelectron Spectroscopy of Adiabatically Bound Valence Anions of Rare Tautomers of the Nucleic Acid Bases. *Journal of Chemical Physics* **2007,** 127.

[38] Stokes, S. T.; Grubisic, A.; Li, X.; Ko, Y. J.; Bowen, K. H., Photoelectron Spectroscopy of the Parent Anions of the Nucleotides, Adenosine-5'-Monophosphate and 2' Deoxyadenosine-5'-Monophosphate. *Journal of Chemical Physics* **2008,** 128, 044314(1)-044314(5).

[39] Gu, J.; Xie, Y.; Schaefer, H. F., Glycosidic Bond Cleavage of Pyrimidine Nucleosides by Low-Energy Electrons: A Theoretical Rationale. *Journal of the American Chemical Society* **2005,** 127, 1053-1057.

[40] Bachorz, R. A.; Rak, J.; Gutowski, M., Stabilization of Very Rare Tautomers of Uracil by an Excess Electron. *Physical Chemistry Chemical Physics* **2005,** 7, 2116-2125.

[41] Haranczyk, M.; Rak, J.; Gutowski, M., Stabilization of Very Rare Tautomers of 1-Methylcytosine by an Excess Electron. *Journal of Physical Chemistry A* **2005,** 109, 11495-11503.

[42] Li, X.; Bowen, K. H.; Haranczyk, M.; Bachorz, R. A.; Mazurkiewicz, K.; Rak, J.; Gutowski, M., Photoelectron Spectroscopy of Adiabatically Bound Valence Anions of Rare Tautomers of the Nucleic Acid Bases. *Journal of Chemical Physics* **2007,** 127, -.

[43] Mazurkiewicz, K.; Bachorz, R. A.; Gutowski, M.; Rak, J., On the Unusual Stability of Valence Anions of Thymine Based on Very Rare Tautomers: A Computational Study. *Journal of Physical Chemistry B* **2006,** 110, 24696-24707.

[44] Haranczyk, M.; Gutowski, M.; Li, X. A.; Bowen, K. H., Adiabatically Bound Valence Anions of Guanine. *Journal of Physical Chemistry B* **2007,** 111, 14073-14076.

[45] Haranczyk, M.; Gutowski, M., Quantum Mechanical Energy-Based Screening of Combinatorially Generated Library of Tautomers. Tautgen: A Tautomer Generator Program. *Journal of Chemical Information and Modeling* **2006,** 47, 686-694.

[46] Haranczyk, M.; Gutowski, M., Combinatorial–Computational–Chemoinformatics (C3) Approach to Finding and Analyzing Low-Energy Tautomers. *Journal of Computer-Aided Molecular Design* **2010,** 24, 627-638.

[47] Haranczyk, M.; Puzyn, T.; Sadowski, P., Congener - a Tool for Modeling

of the Congeneric Sets of Environmental Pollutants. *QSAR & Combinatorial Science* **2008,** 27, 826-833.

[48] Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B* **1988,** 37, 785-789.

[49] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *Journal of Physical Chemistry* **1994,** 98, 11623-11627.

[50] Hehre, W. J.; Ditchfield, R.; Pople, J. A., Self-Consistent Molecular Orbital Methods. Xii. Further Extensions of Gaussian---Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *Journal of Chemical Physics* **1972,** 56, 2257-2261.

[51] Moller, C.; Plesset, M. S., Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934,** 46, 618-622.

[52] Dunning, J. T. H., Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *Journal of Chemical Physics* **1989,** 90, 1007-1023.

[53] Dolgounitcheva, O.; Zakrzewski, V. G.; Ortiz, J. V., Electron Binding Energies of Nucleobases and Nucleotides. *International Journal of Quantum Chemistry* **2002,** 90, 1547-1554.

[54] Ortiz, J. V., Electron Binding Energies of Anionic Alkali Metal Atoms from Partial Fourth Order Electron Propagator Theory Calculations. *Journal of Chemical Physics* **1988,** 89, 6348-6352.

[55] Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; 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.; Bakken, V.; 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.; and Pople, J. A. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.

[56] Crawford, O. H., Negative Ions of Polar Molecules. *Molecular Physics* **1971,** 20, 585 - 591.

[57] Skurski, P.; Rak, J.; Simons, J.; Gutowski, M., Quasidegeneracy of Zwitterionic and Canonical Tautomers of Arginine Solvated by an Excess Electron. *Journal of the American Chemical Society* **2001,** 123, 11073-11074.

[58] N2 is less stable than N1 by only 0.1 kcal/mol at MP2/AVDZ level of theory, but it has a dipole moment larger by 0.7 D.

[59] Gutowski, M.; Jordan, K. D.; Skurski, P., Electronic Structure of Dipole-Bound Anions. *Journal of Physical Chemistry A* **1998,** 102, 2624-2633.

[60] Skurski, P.; Gutowski, M.; Simons, J., How to Choose a One-Electron Basis Set to Reliably Describe a Dipole-Bound Anion. *International Journal of Quantum Chemistry* **2000,** 80, 1024-1038.

[61] Rak, J.; Skurski, P.; Gutowski, M., An Ab Initio Study of the Betaine Anion-Dipole-Bound Anionic State of a Model Zwitterion System. *Journal of Chemical Physics* **2001,** 114, 10673-10681.

[62] Gutowski, M.; Skurski, P., Dispersion Stabilization of Solvated Electrons and Dipole-Bound Anions. *Journal of Physical Chemistry B* **1997,** 101, 9143-9146.

[63] Haranczyk, M.; Gutowski, M., Visualization of Molecular Orbitals and the Related Electron Densities. *Journal of Chemical Theory and Computation* **2008,** 4, 689-693.

[64] Open-source Cubefile Manipulator Program (OpenCubMan) is available free of charge at the SourceForge archive: http://opencubman.sourceforge.net (accessed Sept 2, 2010).

[65] Schlegel, H. B., Optimization of Equilibrium Geometries and Transition Structures. *Journal of Computational Chemistry* **1982,** 3, 214-218.

[66] Fukui, K., The Path of Chemical-Reactions - the IRC Approach. *Accounts of Chemical Research* **1981,** 14, 363-368.

[67] In some cases the IRC calculations converged to a geometry which is close to but not an exact minimum. In these cases we refined this geometry by further optimization.

[68] Kresse, G.; Hafner, J., Abinitio Molecular-Dynamics for Liquid-Metals. *Physical Review B* **1993,** 47, 558-561.

[69] Kresse, G.; Furthmuller, J., Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Computational Materials Science* **1996,** 6, 15-50.

[70] Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996,** 77, 3865-3868.

[71] Blochl, P. E., Projector Augmented-Wave Method. *Physical Review B* **1994,** 50, 17953-17979.

[72] Hobbs, D.; Kresse, G.; Hafner, J., Fully Unconstrained Noncollinear Magnetism within the Projector Augmented-Wave Method. *Physical Review B* **2000,** 62, 11556-11570.

[73] Monkhorst, H. J.; Pack, J. D., Special Points for Brillouin-Zone Integrations. *Physical Review B* **1976,** 13, 5188-5192.

[74] Bylaska, E. J. d. J., W. A.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Apra, E.; Windus, T. L.; Hirata, S.; Hackler, M. T.; Zhao, Y.; Fan, P.-D.; Harrison, R. J.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Auer, A. A.; Nooijen, M.; Brown, E.; Cisneros, G.; Fann, G. I.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J. A.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Pollack, L.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. "NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.0" (2006), Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA.

[75] Schaftenaar, G.; Noordik, J. H., Molden: A Pre- and Post-Processing Program for Molecular and Electronic Structures. *Journal of Computer-Aided Molecular Design* **2000,** 14, 123-134.

[76] Hocquet, A.; Leulliot, N.; Ghomi, M., Ground-State Properties of Nucleic Acid Constituents Studied by Density Functional Calculations. 3. Role of Sugar Puckering and Base Orientation on the Energetics and Geometry of 2'-Deoxyribonucleosides and Ribonucleosides. *Journal of Physical Chemistry B* **2000,** 104, 4560-4568.

[77] Wang, X. C.; Nichols, J.; Feyereisen, M.; Gutowski, M.; Boatz, J.; Haymet, A. D. J.; Simons, J., Ab Initio Quantum Chemistry Study of Formamide-Formamidic Acid Tautomerization. *Journal of Physical Chemistry* **1991,** 95, 10419-10424.

[78] Suen, W.; Spiro, T. G.; Sowers, L. C.; Fresco, J. R., Identification by Uv Resonance Raman Spectroscopy of an Imino Tautomer of 5-Hydroxy-2'-Deoxycytidine, a Powerful Base Analog Transition Mutagen with a Much Higher Unfavored Tautomer Frequency Than That of the Natural Residue 2'-Deoxycytidine. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** 96, 4500-4505.

[79] Notice that the theoretical value of deprotonation enthalpy of N7-H2 from our calculation is different from the one reported in Ref. 20 at the same level of theory which these authors used. This is because these authors double counted the contributions from zero-point corrections, in which they calculated the enthalpy of the deprotonated structure as the sum of the electronic energy (Etot), the zero-point vibrational energy (ZPVE), and the thermal correction to enthalpy (Hoc), and this procedure is absolutely not correct, thus the theoretical results reported there were not correct. The correct value should be 1445.8 kJ/mol.

[80] Ptasinska, S.; Denifl, S.; Scheier, P.; Mark, T. D., Inelastic Electron Interaction (Attachment/Ionization) with Deoxyribose. *Journal of Chemical Physics* **2004,** 120, 8505-8511.

[81] Trofimov, A. B.; Schirmer, J.; Kobychev, V. B.; Potts, A. W.; Holland, D. M. P.; Karlsson, L., Photoelectron Spectra of the Nucleobases Cytosine, Thymine and Adenine. *Journal of Physics B: Atomic, Molecular and Optical Physics* **2006,** 39, 305-329.

[82] Holroyd, L. F.; van Mourik, T., Insufficient Description of Dispersion in B3LYP and Large Basis Set Superposition Errors in MP2 Calculations Can Hide Peptide Conformers. *Chemical Physics Letters* **2007,** 442, 42-46.

[83] Shields, A. E.; van Mourik, T., Comparison of Ab Initio and Dft Electronic Structure Methods for Peptides Containing an Aromatic Ring: Effect of Dispersion and Bsse. *Journal of Physical Chemistry A* **2007,** 111, 13272-13277.

Chapter 8

Effect of Heterostructures on Reactivity

Bulk tungsten trioxide (WO_3) and rhenium trioxide (ReO_3) share very similar structures but display different electronic properties. WO_3 is a wide band gap semiconductor while ReO_3 is an electronic conductor. With the advanced molecular beam epitaxy techniques, it is possible to make heterostructures comprised of layers of WO_3 and ReO_3. These heterostructures might display different reactivity than pure WO_3 and ReO_3. The interactions of two probe molecules (hydrogen and methanol) with (001) surfaces of WO_3, ReO_3, and two heterostructures ReO_3/WO_3 and WO_3/ReO_3 were investigated at the density functional theory level. Atomic hydrogen prefers to adsorb on the terminal O_{1C} sites forming a surface hydroxyl on four surfaces. Dissociative adsorption of a hydrogen molecule at the O_{1C} site leads to formation of a water molecule adsorbed at the surface M_{5C} site. This is thermodynamically the most stable state. A thermodynamically less stable dissociative state involves two surface hydroxyl groups $O_{1C}H$ and $O_{2C}H$. The interaction of molecular hydrogen and methanol with pure ReO_3 is stronger than with pure WO_3 and the strength of the interaction substantially changes on the WO_3/ReO_3 and ReO_3/WO_3 heterostructures. The reaction barriers for decomposition and recombination reactions are sensitive to the nature of heterostructure. The calculated adsorption energy of methanol on WO3(001) of -65.6 kJ/mol is consistent with the previous experimental estimation of -70 kJ/mol. Finally, we found glycine adsorbed on the (001) surface of ReO_3 favors the zwitterionic tautomer over the canonical tautomer.

8.1 Introduction

Metal oxides have been widely used as both catalysts and support materials in heterogeneous catalysis. Generally, insulating metal oxides show limited catalytic activity while conductive metal oxides show high activity with respect to some of heterogeneous reactions such as selective oxidation of alcohols [1-6]. With recent advances in molecular beam epitaxy technology, new types of “heterostructure”

or "heterojunction" materials became available [7]. These heterostructures of metal oxides that are built from layers of structurally similar but electronically different transition metal oxides could be used as new catalysts with tunable activity and selectivity. For example, an epitaxial interface between α-Fe_2O_3 and α-Cr_2O_3 showed non-commutative properties and photocatalytic applications were suggested [8]. Further theoretical study unraveled that the α-Fe_2O_3/α-Cr_2O_3 and α-Cr_2O_3/α-Fe_2O_3 heterostructures have different interfaces. This difference is responsible for non-commutative band offsets and magnetic properties of the interface [9]. Further chemical modifications of the interface were suggested to control the magnitude of band offsets [10]. Tungsten trioxide (WO_3) has a band gap of 2.6 eV [11] while rhenium trioxide (ReO_3) is conductive. Interestingly, monoclinic WO_3 and cubic ReO_3 have commensurate lattice parameters. Thus, heterostructures constructed from WO_3 and ReO_3 may show unique reactivity due to the combination of oxides with different electronic structures.

Supported WO_3 is catalytically active for the dehydration of alcohols [4, 5, 12, 13]. Tanner et al. studied the dehydration of a series of alcohols on the monoclinic γ-WO_3(001) surface using scanning tunneling microscopy (STM) and temperature-programmed desorption (TPD) [12]. They found the 5-fold coordinated metal sites on the WO_3(001) surface are responsible for the oxidative dehydration. No dehydrogenation of alcohols was observed. As the temperature increases, alcohols convert into alkoxides, and then desorb as alkenes. Water molecule is formed by deprotonation of surface hydroxyls [13]. Ma et al. studied the reactivities of ethanol and 2-propanol on the fully oxidized and reduced WO_3(001) surfaces [5]. Their results suggested that both ethanol and 2-proponal molecules are stable until 450 K. Upon further temperature increase, the alkoxy intermediates (dehydrated from alcohols) decompose into alkenes. Methanol only dissociates on the reduced WO_3(001) surface [4]. The reduced WO_3(001) surface shows a slightly higher activity, but does not substantially changes reaction paths. On the contrary, alcohol dehydrogenation occurs on the supported ReO_3 catalysts. A high activity was found for selective methanol oxidation to methylal on V_2O_5- ZrO_2-, Fe_2O_3- and TiO_2-supported ReO_3 [6, 14]. Adsorbed methanol first dehydrogenates into formaldehyde (CH_2O), which could react with lattice oxygen on ReO_3 forming dioxymethylene (CH_2COO). Methylal is produced by oxidative coupling of dioxymethylene with neighboring methoxy (CH_3O) or methanol [2]. The reactivity of methanol on heterostructures of WO_3 and ReO_3 will be explored in the current study.

A number of previous studies indicated that non-stoichiometric hydrogen bronze H_xWO_3/H_xReO_3 ($0.1<x<0.5$) phases are formed from WO_3 and ReO_3 exposed to hydrogen [15, 16]. Adsorption of hydrogen molecules on WO_3 could also lead to metallic tungsten via formation of hydroxyl groups and desorption of water [17]. As such, after hydrogen adsorbs on WO_3 or ReO_3, it will either migrate into the bulk structure forming hydrogen bronze or recombine to form water on the surface. Herein we only focus on the interaction of hydrogen with (001) surfaces of WO_3 or ReO_3, and their heterostructures.

Since the W^{6+} $5d$ band is empty while the Re^{6+} has $5d^1$ electronic configuration, the bulk WO_3 has a band gap of 2.6 eV [11] while ReO_3 is an electronic conductor. The crystalline and electronic structures of ReO_3 and WO_3 have been investigated using density functional theory (DFT) calculations [3, 18, 19]. In agreement with experimental measurements [20, 21], Cora et al. found that ReO_3 is cubic while WO_3 has a distorted cubic structure with off-center displacements of metal ions [3]. The monoclinic structure of WO_3 results from covalent interactions between the metal ion and the nearest oxygen atoms while the presence of the antibonding levels in the conduction band of ReO_3 opposes structural deformations. Yakovkin and Gutowski studied the WO_3(001) surfaces with various types of terminations [18]. They found that the redistribution of density of states near the Fermi level leads to a dramatic decrease of surface electron energy. Consequently, a noticeable distortion and tilting of the surface W atoms were observed in the relaxed surface structure. They also found that the nonpolar c(2×2) O-terminated WO_3(001) slab is more stable than the polar slab with the (1×1) WO_2-termination on one side and the (1×1) O-termination on the other side of the slab [18]. Ge and Gutowski calculated adsorption of methanol on the (001) surfaces of WO_3 and ReO_3. The surface of ReO_3 proved to be much more reactive and favored dissociation of methanol and formation of a methoxy group. The difference in reactivity between the surfaces of WO_3 and ReO_3 was attributed to the partially occupied conduction band of ReO_3, the orbitals of which interact with the orbitals of the hydroxyl group of methanol [19].

Here we will study whether heterostructures made of non-conductive WO_3 and conductive ReO_3 display different reactivity than slabs of pure ReO_3 and WO_3 . Two simple molecules, i.e., hydrogen and methanol are used as probes to test the reactivity. The non-polar c(2×2) O-terminated (001) orientation was chosen for all

four surfaces. The ReO_3/WO_3 heterostructure is constructed by replacing the top layer of $WO_3(001)$ with a ReO_3 overlayer. Similarly, the WO_3/ReO_3 heterostructure is constructed with one WO_3 layer on top of $ReO_3(001)$. The effects of heterostructures on binding energies and reaction barriers were studied at the DFT level of theory.

8.2 Computational Details

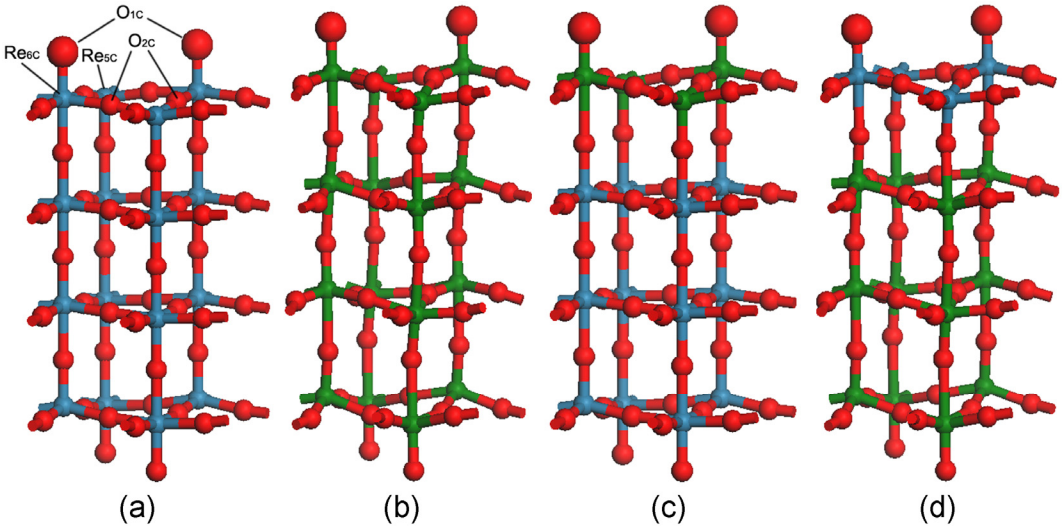


Figure 8.1: Side views of the optimized clean (001) surface structures of (a) ReO_3 ; (b) WO_3; (c) WO_3/ReO_3; and (d) ReO_3 /WO_3. Oxygen atoms are in red; rhenium atoms are in blue; and tungsten atoms are in green.

The periodic DFT calculations were performed with the Vienna Ab initio Simulation Package (VASP) [22, 23] using a plane wave basis set with a cutoff energy of 400 eV. The Perdew-Burke-Ernzerhof (PBE) functional [24] with the projector augmented wave (PAW) method [25, 26] was used to solve the Kohn-Sham equations for crystalline slabs. A Gaussian type of electronic smearing with a width of 0.1 eV was used to improve convergence of electronic self-consistent field calculations. Spin-polarization was needed to describe adsorption of atomic hydrogen. For bulk structure calculations, Monkhorst-Pack k-point grids [27] of (9 × 9 × 9) and (5 × 5 × 3) were used for ReO_3 and WO_3, respectively. The optimized lattice constant of 3.764 Å for the perfect cubic ReO_3 structure is in good agreement with the experimental value of 3.748 Å [20]. The optimized lattice parameters of bulk monoclinic WO_3, a=5.252 Å, b=5.043 Å, c=7.550 Å

and β=93.210º, are also in good agreement with the experimental values of a=5.278 Å, b=5.156 Å, c=7.664 Å and β=91.762º [21].

The (001) surface slab with a c(2 × 2) super cell of ReO_3 and WO_3 were constructed from the optimized bulk structures, see Figure 8.1. The two heterostructure, WO_3/ReO_3 and ReO_3/WO_3, were modeled by replacing the top layer with another oxide, see Figure 8.1. Only minor slab relaxation was found for two heterostructures in the course of geometry optimization. A vacuum layer of 15.0 Å was inserted in the z direction to avoid unphysical interactions between adjacent slabs. For calculations with adsorbates involved, the adsorbate as well as the atoms in the two top layers of the slab are allowed to relax while the atoms in the bottom two layers are fixed. Different k-point grids were tested and a (3 × 3 × 1) k-point sampling was found accurate enough for surface calculations. For density of states (DOS) calculations, k-point grids of (5 × 5 × 1) were used. The Bader's charge analysis [28] was performed using the method developed by Henkelman et al. [29]. All geometry optimizations were performed by using a conjugate-gradient or quasi-Newton scheme as implemented in VASP. The adsorption energy of H, H_2, methanol and glycine is calculated as follows:

$$E_{ads} = E_{adsorbate+surface} - \left(E_{surface} + E_{adsobate}\right) \tag{8.1}$$

where $E_{adsorbate+surface}$ is the total energy of the adsorbate interacting with the surface slab; $E_{surface}$ is the total energy of the optimized surface slab; and $E_{adsorbate}$ is the energy of a single hydrogen atom, a hydrogen molecule or a methanol molecule in vacuum. A negative E_{ads} value indicates the adsorption is energetically favorable.

Transition states were located using the climbing image nudged elastic band (CI-NEB) method [30]. The reaction energy of each reaction path is calculated as the energy difference between the final state and the initial state. The forward and reverse activation barriers of each reaction path are defined as the energy difference between the transition state and the initial and final state, respectively.

8.3 Results and Discussion

8.3.1 Hydrogen Adsorption

In this work, we focus on hydrogen adsorption on the surface, not its diffusion into bulk. We first studied atomic hydrogen adsorption on the four surfaces. Three

possible surface sites, i.e., the single bonded terminal O_{1C}, the bridging double bonded O_{2C}, and the penta-coordinated metal site Re_{5C} (or W_{5C}) are available for hydrogen adsorption. The optimized structures for the ReO_3 (001) surface are shown in Figure 8.2. The structural parameters and adsorption energies are summarized in Table 8.1.

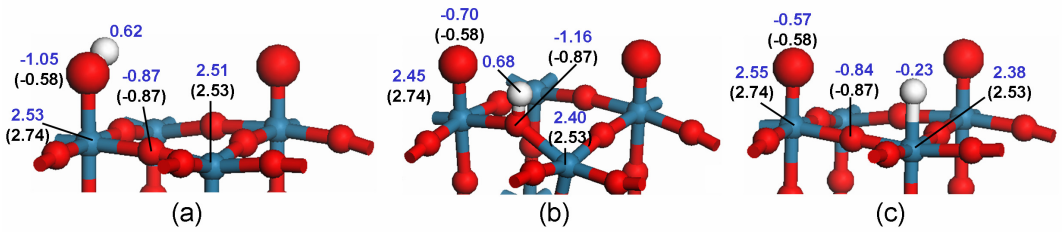


Figure 8.2: Calculated Bader charge changes upon atomic hydrogen adsorption on the ReO_3(001) surface. (a) H on O_{1C} site; (b) H on O_{2C} site; and (c) H on Re_{5C} site. The numbers in black represent the charges before adsorption. The numbers in blue represent the charges after adsorption.

Surface	Adsorption site	E_{ad}	M_{6C}-O_{1C}	M_{6C}-O_{2C}	M_{5C}-O_{2C}	O_s-H/M_s-H
ReO_3	O_{1C}	-296.25	1.93	1.91	1.88	0.98
	O_{2C}	-200.06	1.76	2.08	2.04	0.98
	Re_{5C}	-141.52	1.72	1.91	1.87	1.71
WO_3	O_{1C}	-266.80	1.88	1.88	1.92	0.97
	O_{2C}	-265.09	1.71	2.06	2.04	1.00
	W_{5C}	-7.90	1.71	1.90	1.93	2.34
WO_3/ReO_3	O_{1C}	-278.33	1.88	1.92	1.88	0.96
	O_{2C}	-211.21	1.73	2.10	2.05	0.98
	W_{5C}	-48.87	1.71	1.92	1.89	1.73
ReO_3/WO_3	O_{1C}	-266.82	1.93	1.92	1.87	0.98
	O_{2C}	-207.26	1.73	2.05	2.04	1.00
	Re_{5C}	-152.30	1.71	1.88	1.88	1.72

Table 8.1: Energetics (kJ/mol) and structural parameters (Å) of atomic hydrogen adsorption on four model surfaces. The O_s represents the bonded surface oxygen atom. The M_s represents the closest surface metal atom.

The calculated adsorption energies of a hydrogen atom range from -7.9 kJ/mol at the W_{5C} site of WO_3(001) to -296.3 kJ/mol at the O_{1C} site of ReO_3(001). The oxygen sites (O_{1C} and O_{2C}) are energetically more favorable than the metal sites

(Re_{5C} and W_{5C}) to adsorb a hydrogen atom due formation of surface hydroxyl groups. Compared to the bridging O_{2C} site, the unsaturated terminal O_{1C} is more active. The adsorption energies at the O_{1C} sites are higher than at the O_{2C} sites by about 60~100 kJ/mol, with an exception of WO_3(001), for which the adsorption energies are comparable, 266.8 and 265.1 kJ/mol, respectively. For the most stable hydrogen adsorption configuration (O_{1C} site), the hydrogen binding on the ReO_3(001) surface is the strongest and accounts to 296.3 kJ/mol. It drops to 266.8 kJ/mol for ReO_3/WO_3 illustrating the effect of the heterostructure. The atomic hydrogen adsorption at the M_{5C} site (M=W or Re) exposes differences between tungsten and rhenium. The adsorption on the W_{5C} site of WO_3(001) is very weak (-7.9 kJ/mol) with a W_{5C}-H bond distance of 2.34 Å, indicating a physisorbed bonding state. On the other hand, the adsorption energy at the Re_{5C} site of ReO_3(001) is -141.5 kJ/mol and the Re_{5C}-H bond length is much shorter, 1.71 Å. Overall, atomic hydrogen interacts with the conductive Re(001) surface more strongly than with the insulating WO_3(001) surface. Perusal of the results for WO_3 and WO_3/ReO_3 further illustrates the effect of heterostructure. In comparison with WO_3, WO_3/ReO_3 offers a strong preference for adsorption at the O_{1C}, and the adsorption energy at the W_{5C} site is increased to -48.87 kJ/mol. These differences must be attributed to the ReO_3 substrate.

To get more insights on how the variation of electronic structure affects the hydrogen adsorption, we performed Bader's charge analysis and projected density of states (PDOS) calculations. Figure 8.3 shows the Bader charges for the empty ReO_3(001) slab and for the same slab with a hydrogen atom adsorbed at the Re_{5C}, O_{1C}, and O_{2C} sites. For hydroxyl groups resulting from hydrogen adsorption on the O_{1C}, and O_{2C} sites, the effective charge on hydrogen is positive and effective negative charges on oxygens become more pronounced illustrating partial electron transfer from hydrogen to oxygen. On the other hand, a hydrogen atom adsorbed at the Re_{5C} site acquires an effective negative charge, indication a partial electron transfer from rhenium to hydrogen. In consequence, a hydridic hydrogen is formed, though its adsorption energy is not competitive with the O_{1C} and O_{2C} sites. From PDOS plots shown in Figure 8.3, it is clear that the antibonding 2*p* states of O_{1C} and O_{2C}, and antibonding 5*d* states of Re_{5C} atom are responsible for the hydrogen adsorption.

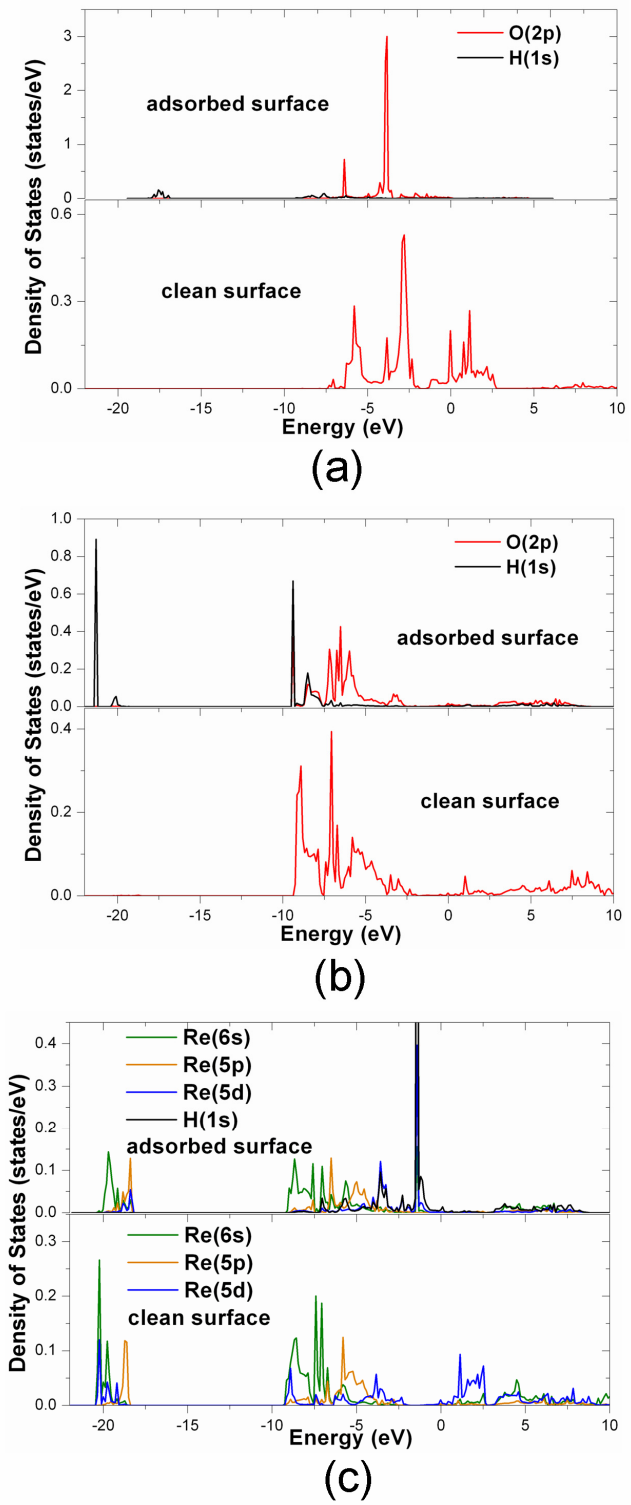


Figure 8.3: DOS plots upon atomic hydrogen adsorption on the ReO_3(001) surface. (a) H on O_{1C} site; (b) H on O_{2C} site; and (c) H on Re_{5C} site. All the energies are shifted to zero at the Fermi level.

Different adsorption configurations were considered for molecular and dissociative adsorptions of a hydrogen molecule: (i) dissociative adsorption at the O_{2C} site, (ii) dissociative adsorption at neighboring O_{2C} sites; (iii) dissociative adsorption with one hydrogen atom at the O_{2C} site and another at the M_{5C} site; (iv) molecular hydrogen adsorption at the M_{5C} site; (v) dissociative adsorption at the O_{1C} site; (vi) dissociative adsorption at the O_{1C} and O_{2C} sites. Our results indicate that the adsorption energies are positive or larger than -3 kJ/mol for the first four cases. For this reason only the last two configurations are further considered and adsorption energies of a hydrogen molecule on the four surfaces are summarized in Table 8.2. A hydrogen molecule dissociatively adsorbed at the O_{1C} can be viewed as a water molecule adsorbed at the M_{5C} site.

Surface	Adsorption sites	E_{ad}	M_{6C}-O_{1C}	M_{6C}-O_{2C}	M_{5C}-O_{2C}	O_{1C}-H	O_{1C}···H(O_{2C})
ReO_3	O_{1C} + O_{1C}	-100.32	2.23	1.90	1.89	0.98	
	O_{1C} + O_{2C}	-68.21	1.98	2.05	2.05	0.97	2.07
	TS		2.09	2.03	1.98	0.97	1.33
WO_3	O_{1C} + O_{1C}	-46.68	2.32	1.90	1.91	0.98	
	O_{1C} + O_{2C}	-29.33	1.89	2.06	2.03	0.97	3.16
	TS		2.08	2.07	1.96	0.97	1.37
WO_3/ReO_3	O_{1C} + O_{1C}	-113.62	2.33	1.93	1.92	0.98	
	O_{1C} + O_{2C}	-67.09	1.94	2.10	2.04	0.97	2.90
	TS		1.97	2.09	2.04	0.97	2.19
ReO_3 /WO_3	O_{1C} + O_{1C}	-58.78	2.21	1.94	1.91	0.98	
	O_{1C} + O_{2C}	-50.30	1.94	2.10	2.05	0.98	3.16
	TS		2.08	2.03	1.99	0.97	1.36

Table 8.2: Energetics (kJ/mol) and structural parameters (Å) of dissociative adsorption of hydrogen molecule on four model surfaces.

We will use a label "State 1" for a hydrogen molecule dissociatively adsorbed at the O_{1C} site and "State 2" for a hydrogen molecule dissociatively adsorbed at the O_{1C} site and at an adjacent O_{2C} site. The adsorption energies and energy barriers for the State 1 ↔ State 2 transformations are illustrated in Figure 8.4. One might expect that the stability of the both dissociatively adsorbed states will be the largest for the metallic slab ReO_3 followed by the ReO_3/WO_3 heterostructure. In the latter, the topmost layer is ReO_3. However, the stability of State 1 evolves as WO_3/ReO_3 > ReO_3 > ReO_3/WO_3 > WO_3 illustrating unexpected properties of the WO_3/ReO_3 heterostructure. The stability of State 2 evolves as ReO_3 > WO_3/ReO_3 > ReO_3/WO_3 > WO_3 illustrating again unexpected properties of the WO_3/ReO_3 heterostructure. State 1 is energetically more favorable than State 2 for all slabs,

with the relative stability exceeding 32 kJ/mol for WO_3/ReO_3 and ReO_3 and being smaller than 18 kJ/mol for WO_3 and ReO_3/WO_3.

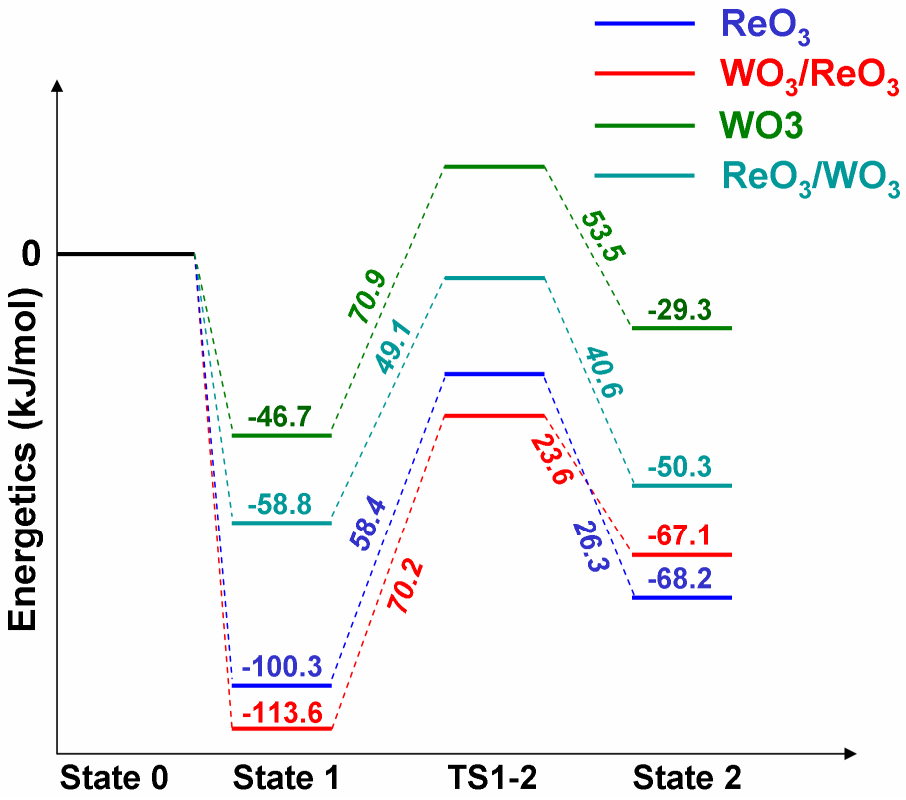


Figure 8.4: Possible mechanisms of adsorptions of a hydrogen molecule on the four model surfaces.

State 1 and State 2 might interconvert between each other. Thus we considered energetic barriers for the State 1 ↔ State 2 transformations. The forward reaction would correspond to breaking an O_{1C}-H bond from State 1 and transferring the hydrogen atom to a nearby O_{2C} atom. The results of NEB calculations for transition states (TS1-2) are summarized in Figure 8.4 and geometries of transition states are summarized in Table 8.2. The energies in Figure 8.4 are referenced with respect to State 0, which represents an empty slab and an isolated hydrogen molecule. As discussed above, the reaction State 1 → State 2 is endothermic for all slabs and the barrier for the forward reaction is the smallest for the ReO_3/WO_3 slab (49.1 kJ/mol) and the largest for the WO_3 slab (70.9 kJ/mol). For the reverse reaction the smallest barrier is for the WO_3/ReO_3 slab (23.6 kJ/mol) and the largest for the WO_3 slab (53.5 kJ/mol).

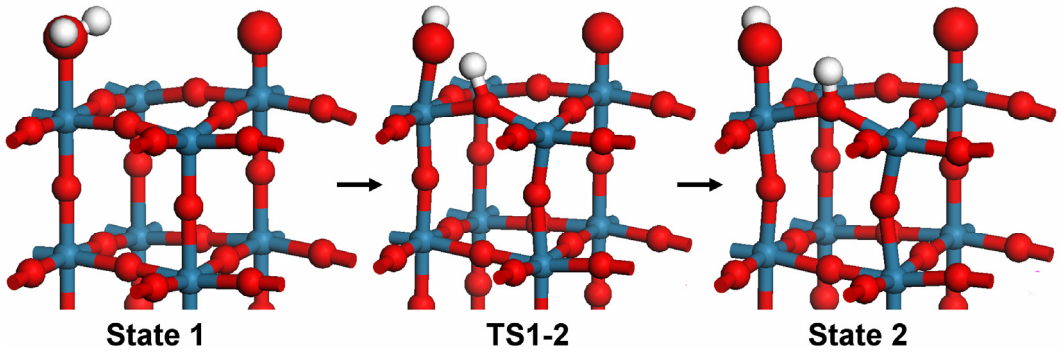


Figure 8.5: Sideview of optimized geometries of different surface states for adsorption of a hydrogen molecule on the pure ReO_3(001) surface.

Taking ReO_3 as an example, we show the optimized structures of State 1, TS1-2 and State 2 in Figure 8.5. This figure and geometries listed in Table 2 illustrate that the slab is only slightly distorted in State 1. As hydrogen transfers through TS1-2 to State 2 a significant geometry relaxation of the slab is observed. In the initial state, which resembles a water molecule adsorbed at the R_{5C} site, the O_{1C}-H distance is only 0.97 Å. Next, the H(O_{1C}) atom approaches the O_{2C} atom and the O_{1C}-H distance increases to 1.33 Å at TS1-2. This structural rearrangement is accompanied by a pronounced surface relaxation. In the final state, one hydrogen atom is bound at the O_{2C} site and its distance from the O_{1C} site is 2.07 Å. The surface remains strongly distorted upon formation of the O_{1C}-H and O_{2C}-H hydroxyl groups. Similar surface relaxation behavior was observed for the State 1 → State 2 reaction on three other slabs.

One might expect that the conductive ReO_3 surface will be the most favorable for the State 1 → State 2 reaction. The calculated barrier is, however, significant, 58.4 kJ/mol, and the reaction is endothermic by 32.1 kJ/mol. A significant improvement is observed for the ReO_3/WO_3 slab, for which the barrier is reduced to 49.1 kJ/mol and the endothermicity to 8.4 kJ/mol! Thus the heterostructure ReO_3/WO_3 displays better properties for dissociation of molecular hydrogen than the ReO_3 surface. Perusal of Figure 8.4 also illustrates activation of the WO_3 surface by putting the RO_3 monolayer on the top and "passivation" of the ReO_3 surface by putting the WO_3 monolayer on the top.

Next we analyze the State 2 → State 1 reaction, which is equivalent to recombination of the O_{1C}-H and O_{2C}-H hydroxyl groups and formation of a water molecules ($H2O_{1C}$) adsorbed on the surface metal site. The reaction is exothermic

for all four slabs, in particular for WO_3/ReO_3 (-46.5 kJ/mol) and ReO_3 (-32.1 kJ/mol). The same two slabs offer the lowest barriers for the recombination step: 23.6 kJ/mol for WO_3/ReO_3 and 26.3 kJ/mol for ReO_3, see Figure 8.4. These barriers are much lower than the corresponding barriers (53.5 and 40.6 kJ/mol) on the pure WO_3 and ReO_3/WO_3 surfaces. Advantageous properties of WO_3/ReO_3 in comparison with ReO_3 illustrate that epitaxial heterostructures might be useful in catalytic applications. Similarly, the reactivity of WO_3 surface can be improved by adding an overlayer of ReO_3.

The results presented above are consistent with experimental findings, which indicated that adsorption of hydrogen on WO_3 could also lead to metallic tungsten via formation of hydroxyl groups and desorption of water [17].

8.3.2 Methanol Molecular Adsorption and Dissociation

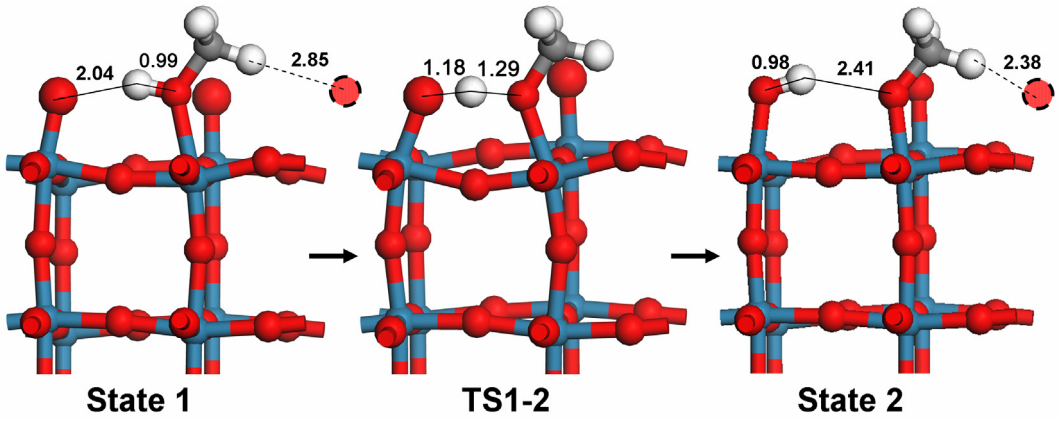


Figure 8.6: Methanol dissociation path on the ReO_3(001) surface. Distances in Å.

Methanol can molecularly adsorb at the M_{5C} site (State 1) and dissociatively adsorb at the O_{1C} and M_{5C} sites (State 2). As shown in Figure 8.6 for ReO_3(001), methanol molecularly adsorbs via the O-M_{5C} bonding. The calculated adsorption energy spans a range from 65.6 kJ/mol for WO_3 to 95.3 kJ/mol for ReO_3, see Figure 8.7 and Table 8.3. The adsorption energies roughly correlate with the M5c-O(OH) distance, which is the shortest for ReO_3 and the longest for WO_3. These distances are significantly longer than the M_{5C}-O_{2C} distances, which illustrate the physisorbed state of methanol. The calculated methanol adsorption energy on WO_3 is -65.6 kJ/mol. This is in good agreement with previous experimental estimation of -70 kJ/mol [4].

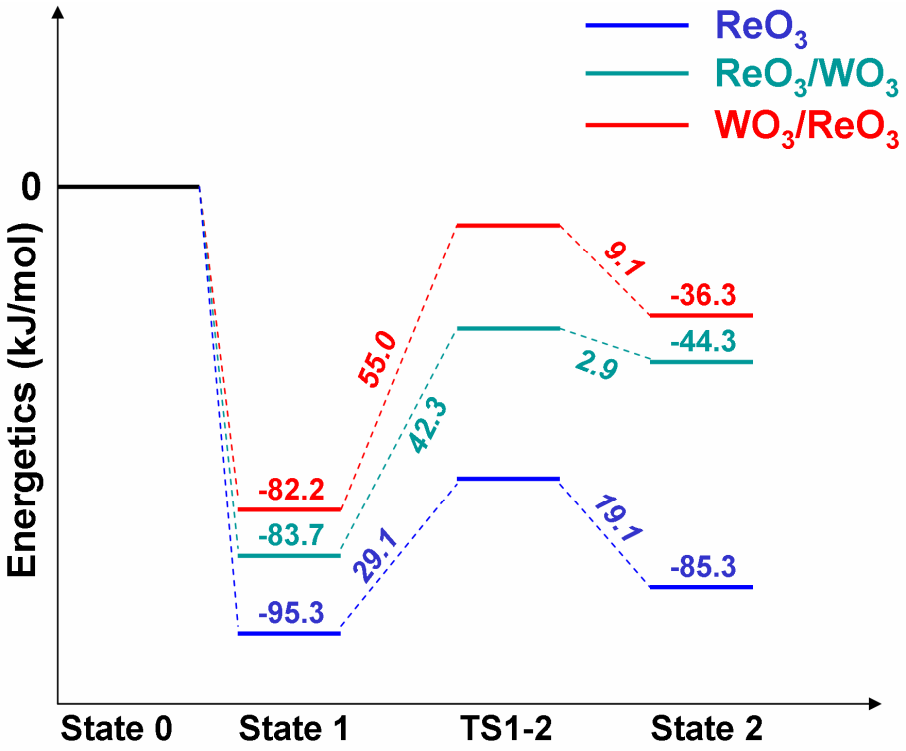


Figure 8.7: Possible mechanisms of adsorptions of a methanol molecule on model surfaces.

Surface	Adsorption mode	E_{ad}	M_{6C}-O_{1C}	M_{6C}-O_{2C}	M_{5C}-O_{2C}	M_{5C}-O(OH)	O_{1C}···H(OH)	O(M)···H(OH)
ReO_3	molecular	-95.30	1.76	1.93	1.87	2.24	2.04	0.99
	dissociative	-85.31	1.94	1.88	1.88	1.94	0.98	2.41
	TS		1.85	1.94	1.87	2.08	1.18	1.29
WO_3	molecular	-65.61	1.73	1.92	1.88	2.38	2.17	0.98
	dissociative	36.89	1.89	1.83	1.96	1.95	0.98	3.09
WO_3/ReO_3	molecular	-82.20	1.74	1.95	1.87	2.29	2.07	0.98
	dissociative	-36.31	1.88	2.06	1.79	1.98	1.02	1.69
	TS		1.87	2.04	1.79	2.00	1.03	1.65
ReO_3 /WO_3	molecular	-83.66	1.76	1.88	1.88	2.32	2.01	0.99
	dissociative	-44.32	1.91	1.86	1.89	2.00	1.00	1.93
	TS		1.86	1.87	1.90	2.08	1.10	1.41

Table 8.3: Energetics (kJ/mol) and structural parameters (Å) of molecular and dissociative adsorptions of methanol on four model surfaces. The O(OH) and H(OH) denote the oxygen and hydrogen atom in the hydroxyl group of methanol, respectively. M denotes Re or W.

We considered dissociation of molecularly adsorbed methanol via O-H bond breaking into a methoxy group followed by formation of a surface hydroxyl HO_{1C}. As discussed in the Introduction, different decomposition paths were observed on the pure ReO_3 and WO_3 surfaces. Supported ReO_3 was found to be catalytically very active [6, 14] while WO_3 is almost inactive for methanol dissociation under UHV condition [4]. We first examined methanol dissociation on the pure WO_3(001) surface. It is found that methanol dissociation is highly endothermic with a reaction energy of +102.5 kJ/mol, see also Ref. 19. This is consistent with experimental findings [4] that methanol does not dissociate on the WO_3(001) surface. We further calculated methanol dissociation on other three slabs.

The stability of State 2 decreases from ReO3 through ReO_3/WO_3 to WO_3/ReO_3 and the adsorption energies span a broad range from -85.3 to -36.3 kJ/mol. The State 1 → State 2 reaction is endothermic for all four slabs, see Table 8.3 and Figure 8.7. We also note the endothermicity of methanol dissociation is surface dependent. For example, methanol dissociation on the ReO_3(001) surface (shown in Figure 8.7) is slightly endothermic (+10 kJ/mol). The endothermicity increases to 39.3 kJ/mol for ReO_3/WO_3 and 45.9 kJ/mol for WO_3/ReO_3. This trend is maintained for the forward barriers, with the smallest for ReO_3 (29.1 kJ/mol) and the largest for WO_3/ReO_3 (55.0 kJ/mol). The small endothermicity and a low forward barrier for ReO_3 is in agreement with previous experimental finding [6, 14] that the supported ReO_3 is a good catalyst for methanol decomposition reaction. We emphasize that the passive WO3 surface is activated when engaged in the WO_3/ReO_3 and ReO_3/WO_3 heterostructures.

8.3.3 Glycine Molecular Adsorption

We also studied molecular adsorption of a glycine molecule, both in the canonical and zwitterionic forms, on the (001) surface of ReO_3. We found that the carbonyl oxygen of glycine binds the M_{5C} site of the surface for both the canonical and zwitterionic glycine, and we show the optimized structures in Figure 8.8. Taking the most stable isolated canonical glycine as our reference, we found that the calculated adsorption energies for the canonical and zwitterionic glycine are -109.0 and -123.9 kJ/mol, respectively, which indicates that adsorption on ReO_3(001) surface is thermodynamically more favorable for the zwitterionic rather than canonical glycine. This is due to the formation of strong surface hydrogen bonds between the amine group of the zwitterionic glycine and the

surface O_{1C} atoms. Such hydrogen bonds are not formed in the case of canonical glycine, because the hydroxyl group forms a surface hydrogen bond with the O_{1C} atom from an adjacent unit cell. This drags the glycine molecule towards the next unit cell and places the amine group far from the surface O_{1C} atoms in the reference unit cell.

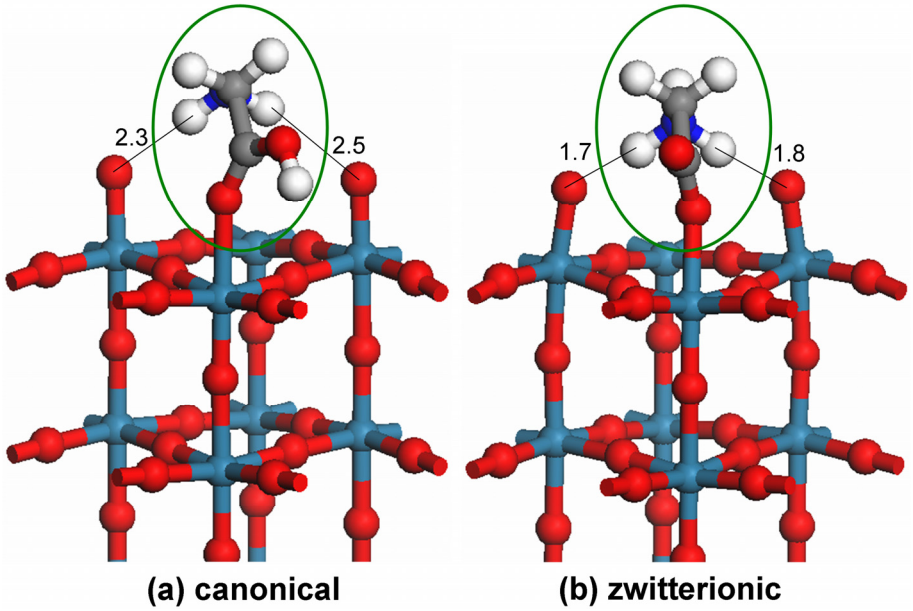


Figure 8.8: Glycine molecular adsorption on the ReO_3(001) surface, (a) canonical tautomer, and (b) zwitterionic tautomer. Distances in Å.

The zwitterionic glycine does not exist in the gas phase. Thus the initial adsorption of molecular glycine on ReO_3(001) will involve the canonical tautomer. Our results indicate that the adsorption of zwitterionic glycine is thermodynamically more favorable by about 15 kJ/mol on ReO_3(001). Therefore a very interesting question would be what is the reaction pathway which connects these two adsorption states. A proposed pathway is shown in Figure 8.9. First the hydroxyl hydrogen of canonical glycine migrates to a nearby O_{1C} atom and forms a surface hydroxyl group. Next, the surface hydroxyl group rotates toward the amine group of glycine in the neighboring unit cell. Finally the surface hydroxyl hydrogen migrates to the amine group and forms an adsorbed zwitterionic glycine. The whole process might be viewed as surface assisted intermolecular proton transfer. The proposed mechanism is currently scrutinized. The adsorption of canonical and zwitterionic glycine on other model surfaces will be studied. We

will also explore the effect of overlayer-like interface structures on barriers associated with the canonical-zwitterion transformations.

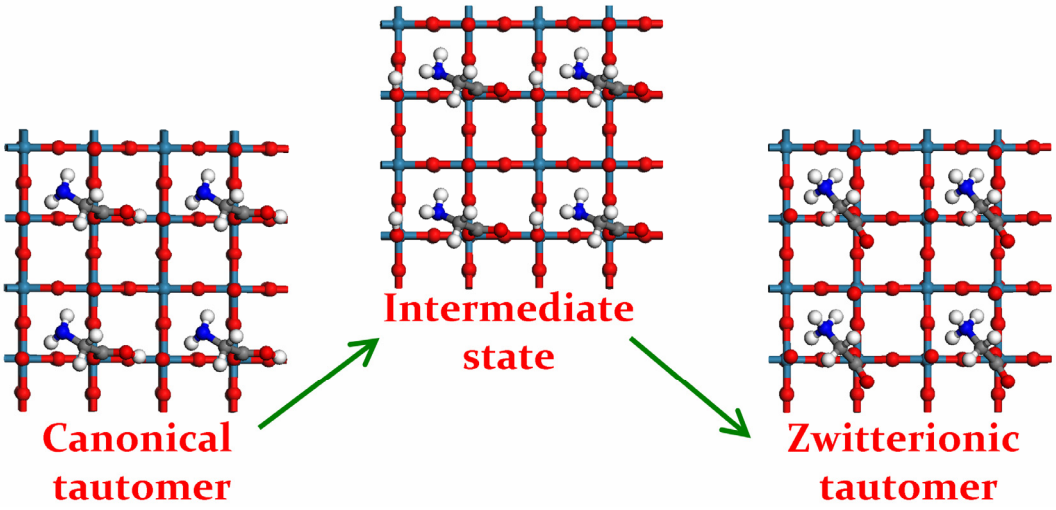


Figure 8.9: Possible pathway of interconversion between adsorption states of canonical and zwitterionic glycine.

8.4 Summary

To understand surface reactivity of epitaxial heterostructures made of metal oxides displaying different electronic structure, a comparative study on adsorption and reactivity of hydrogen and methanol on model (001) surfaces of ReO_3, WO_3, as well as their heterostructures ReO_3/WO_3 and WO_3/ReO_3 has been performed at the density functional theory level. The wide band gap WO_3 and the conductive ReO_3 were chosen due to their structural similarity.

Atomic hydrogen adsorption at the terminal O_{1C} site is the energetically most favorable on all four surfaces. The Bader charge and PDOS analyses clearly suggest that the antibonding $O_{1C}(2p)$ is responsible for hydrogen adsorption.

Dissociative adsorption of a hydrogen molecule at the O_{1C} site leads to formation of a water molecule adsorbed at the surface M_{5C} site. This is thermodynamically the most stable state. A thermodynamically less stable dissociative state involves two surface hydroxyl groups $O_{1C}H$ and $O_{2C}H$. We also determined energy barriers that separate these two states. The interaction of hydrogen with pure ReO_3 is

stronger than with pure WO_3 and the strength of the interaction substantially changes on the WO_3/ReO_3 and ReO_3/WO_3 heterostructures. In particular, the barrier and endothermicity for the forward reaction is reduced by 9.3 and 23.7 kJ/mol upon the replacement of ReO_3 with ReO_3/WO_3. The recombination of the O_{1C}-H and O_{2C}-H hydroxyl groups and formation of a water molecules ($H2O_{1C}$) adsorbed on the surface metal site was found to be exothermic for all four slabs, in particular for WO_3/ReO_3 (-46.5 kJ/mol) and ReO_3 (-32.1 kJ/mol). The same two slabs offer the lowest barriers for the recombination step: 23.6 kJ/mol for WO_3/ReO_3 and 26.3 kJ/mol for ReO_3.

We considered dissociation of molecularly adsorbed methanol via O-H bond breaking into a methoxy group followed by formation of a surface hydroxyl HO_{1C}. In agreement with past experimental observations, our calculations show that methanol does not dissociate on WO_3(001) surface with a barrier exceeding 100 kJ/mol. The calculated adsorption energy of methanol on WO3(001) of -65.6 kJ/mol is consistent with the previous experimental estimation of -70 kJ/mol. However, the reactivity of methanol increases on the ReO_3/WO_3 and WO_3/ReO_3 heterostructures, with the ReO_3 surface remaining the most reactive.

For adsorption of a glycine molecule, we found that the interaction with the ReO_3(001) surface could stabilize the zwitterionic glycine. In fact, the adsorbed zwitterionic glycine is more stable by about 15 kJ/mol than the adsorbed canonical glycine. The results of this study demonstrate that zwitterionic glycine might be experimentally probed not only in solvents but also when adsorbed in surfaces of properly selected metal oxides.

In summary, the results of our comparative theoretical study performed for the four model surfaces indicate the catalytic reactivities of surfaces of conductive and semiconductive transition metal oxides might be tuned by making overlayer-like interfacial structures. The possibilities for improved catalytic activities should be further explored in future theoretical and experimental studies.

References:

[1] Henrich, V. E.; Cox, P. A., *The Surface Science of Metal Oxides*. Cambridge University Press: Cambridge, 1994; p xiv,464p.

[2] Chan, A. S. Y.; Chen, W.; Wang, H.; Rowe, J. E.; Madey, T. E., Methanol Reactions over Oxygen-Modified Re Surfaces: Influence of Surface Structure and Oxidation†. *Journal of Physical Chemistry B* **2004,** 108, 14643-14651.

[3] Cora, F.; Stachiotti, M. G.; Catlow, C. R. A.; Rodriguez, C. O., Transition Metal Oxide Chemistry: Electronic Structure Study of Wo3, Reo3, and Nawo3. *Journal of Physical Chemistry B* **1997,** 101, 3945-3952.

[4] Ma, S.; Amar, F. G.; Frederick, B. G., Surface Heterogeneity and Diffusion in the Desorption of Methanol from Wo3(001) Surfaces. *Journal of Physical Chemistry A* **2003,** 107, 1413-1423.

[5] Ma, S.; Frederick, B. G., Reactions of Aliphatic Alcohols on Wo3(001) Surfaces. *Journal of Physical Chemistry B* **2003,** 107, 11960-11969.

[6] Yuan, Y.; Iwasawa, Y., Performance and Characterization of Supported Rhenium Oxide Catalysts for Selective Oxidation of Methanol to Methylal. *Journal of Physical Chemistry B* **2002,** 106, 4441-4449.

[7] Chambers, S. A., Molecular Beam Epitaxial Growth of Doped Oxide Semiconductors. *Journal of Physics: Condensed Matter* **2008,** 20, 264004.

[8] Chambers, S. A.; Liang, Y.; Gao, Y., Noncommutative Band Offset at Alpha-Cr2o3/Alpha-Fe2o3(0001) Heterojunctions. *Physical Review B* **2000,** 61, 13223-13229.

[9] Jaffe, J. E.; Dupuis, M.; Gutowski, M., First-Principles Study of Noncommutative Band Offsets at α-Cr2o3/α-Fe2o3(0001) Interfaces. *Physical Review B* **2004,** 69, 205106.

[10] Jaffe, J. E.; Bachorz, R. A.; Gutowski, M., Band Offset and Magnetic Property Engineering for Epitaxial Interfaces: A Monolayer of M_2o_3 (M=Al,Ga,Sc,Ti,Ni) at the α-Fe_2o_3/α-Cr_2o_3 (0001) Interface. *Physical Review B* **2007,** 75, 205323.

[11] Butler, M. A.; Nasby, R. D.; Quinn, R. K., Tungsten Trioxide as an Electrode for Photoelectrolysis of Water. *Solid State Communications* **1976,** 19, 1011-1014.

[12] Tanner, R. E.; Meethunkij, P.; Altman, E. I., Identification of Alcohol Dehydration Sites on an Oxide Surface by Scanning Tunneling Microscopy. *Journal of Physical Chemistry B* **2000,** 104, 12315-12323.

[13] Li, M.; Gao, W.; Posadas, A.; Ahn, C. H.; Altman, E. I., Reactivity of 1-Propanol on P(N X 2) Reconstructed Wo3(100) Thin Films. *Journal of Physical Chemistry B* **2004,** 108, 15259-15265.

[14] Yuan, Y.; Shido, T.; Iwasawa, Y., The New Catalytic Property of Supported Rhenium Oxides for Selective Oxidation of Methanol to Methylal. *Chemical Communications* **2000**, 1421-1422.

[15] Berzins, A. R.; Sermon, P. A., Reversible Isothermal Sorption of Hydrogen by Tungsten Trioxide in Presence of Platinum. *Nature* **1983,** 303, 506-508.

[16] Horiuchi, S.; Kimizuka, N.; Yamamoto, A., Absorption of Hydrogen in Reo3. *Nature* **1979,** 279, 226-227.

[17] Bringans, R. D.; Hochst, H.; Shanks, H. R., Hydrogen on Wo3(001). *Surface Science* **1981,** 111, 80-86.

[18] Yakovkin, I. N.; Gutowski, M., Driving Force for the Wo3(001) Surface Relaxation. *Surface Science* **2007,** 601, 1481-1488.

[19] Ge, Q.; Gutowski M. Unpublished results.

[20] Schirber, J. E.; Morosin, B., Compressibility Collapse Transition in Reo3. *Physical Review Letters* **1979,** 42, 1485-1487.

[21] Woodward, P. M.; Sleight, A. W.; Vogt, T., Ferroelectric Tungsten Trioxide. *Journal of Solid State Chemistry* **1997,** 131, 9-17.

[22] Kresse, G.; Hafner, J., Abinitio Molecular-Dynamics for Liquid-Metals. *Physical Review B* **1993,** 47, 558-561.

[23] Kresse, G.; Furthmuller, J., Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Computational Materials Science* **1996,** 6, 15-50.

[24] Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996,** 77, 3865-3868.

[25] Blochl, P. E., Projector Augmented-Wave Method. *Physical Review B* **1994,** 50, 17953-17979.

[26] Hobbs, D.; Kresse, G.; Hafner, J., Fully Unconstrained Noncollinear Magnetism within the Projector Augmented-Wave Method. *Physical Review B* **2000,** 62, 11556-11570.

[27] Monkhorst, H. J.; Pack, J. D., Special Points for Brillouin-Zone Integrations. *Physical Review B* **1976,** 13, 5188-5192.

[28] Bader, R. F. W., Atoms in Molecules. *Accounts of Chemical Research* **1985,** 18, 9-15.

[29] Henkelman, G.; Arnaldsson, A.; Jonsson, H., A Fast and Robust Algorithm for Bader Decomposition of Charge Density. *Computational Materials Science* **2006,** 36, 354-360.

[30] Henkelman, G.; Uberuaga, B. P.; Jonsson, H., A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *Journal of Chemical Physics* **2000,** 113, 9901-9904.

Chapter 9

Summary and Future Work

In the current thesis, we have discussed small biomolecules in the gas phase as well as interactions between surfaces of metal oxides and small molecules, hydrogen and methanol. Our goal includes both methodology development and computational applications.

As the starting point, we developed a new tool, which we call SSC, denoting Systematic Screening of Conformers. A library of potentially relevant conformers can be created with SSC, which facilitates a systematic search of the conformational space of a selected chain-like molecule. Each member of the library is prescreened at a predefined level of theory and the most promising conformers are identified. Finally, they are further evaluated at a higher level of theory to identify the most stable structures and their physicochemical properties. Our tool is fully automated and a user has control which dihedral angles will be probed and with which increments. Moreover, whole fragments of the molecule, which are adjacent to each selected rotational bond, are rotated in a properly selected cylindrical coordinate system and unchemical hybridizations and some "clashes" between neighboring groups, which are common when standard Z-matrices are used, are avoided. Looking forward in the future, a more user-friendly graphical interface could be created, based on the algorithms which we have developed for SSC. In addition, these algorithms might also be implemented into a current available computational chemistry code, to make a one-stop conformational search possible.

We recognize that the systematic search approach might be unfeasible for bigger molecules, if one chooses a relatively accurate method to describe the potential energy surface (PES). To find the global minima of such systems, we would need a more intelligent and powerful method. For this reason, we suggested a concept of generalized simulated annealing protocols. We suggested that the barriers separating minima on a PES might be suppressed by subtracting selected force field terms from the original PES. We used the resulting deformed PES in

standard molecular dynamics (MD) simulations. Taking ethanolamine as our initial model system, we found that the MD trajectories on the original and deformed PESs of ethanolamine differ markedly. The former gets stuck in a local minimum basin while the latter moves quickly to the global minimum basin, which shows the good performance of our idea on searching for the global minimum of ethanolamine on a deformed PES. The current idea should be tested on bigger systems to further check its efficiency.

Besides these methodological developments, we also paid attention to theoretical accuracy. We revisited the two most stable conformers of canonical glycine by highly accurate theoretical calculations up to the CCSD(T)/aug-cc-pVQZ level. Our results showed that considering both the computational cost and accuracy, the MP2/aug-cc-pVDZ method is recommended for geometry optimizations and the CCSD(T)/aug-cc-pVTZ method is recommended for single point energy calculations to determine the relative energies of the different conformers. Our subsequent benchmark calculations also showed that B3LYP gives the best performance among all the density functionals considered in the current contribution for the description of the relative energy of the two glycine conformers, and this finding should be further explored on more complex biomolecules.

Further, we performed a structural characterization for a typical isolated nucleoside, 2'-deoxycytidine (dC), through extensive quantum chemical calculations. Conformers of the canonical tautomer, and selected isomers and non-canonical tautomers have been studied. We considered neutral molecules as well as their anionic and cationic radicals. We also analyzed the effect of crystalline environment on the conformational structure of dC. Our results indicated that the neutral canonical dC in the gas phase favors *syn* rather than *anti* conformations. The thermodynamics dominance of *syn* conformations results from the formation of an intramolecular O5'-H13···O2 hydrogen bond. We have also demonstrated that the C2'-*endo* sugar puckering is favorable in the gas phase. Our results have confirmed that valence anions of dC are adiabatically bound. The most stable valence anions proved to be the α-anomer and a *syn* canonical conformer. The glycosidic bond cleavage has been studied for the neutral, anionic, and cationic dC. The barrier for cleavage is significant in the neutral dC, 47.8 kcal/mol, and it is reduced by 16-19 kcal/mol for charged dC, either cationic or anionic. Finally we analyzed the problem why dC

molecules favor *anti* conformation in the crystalline phase but *syn* conformations in the gas phase. We have demonstrated that intermolecular hydrogen bonds favor *anti* conformers in the experimental structure of the molecular crystal of dC. These results clearly showed the structural flexibility of a relatively small biomolecule. From the point of view of biological relevance, these results should be further validated through comparisons with theoretical and experimental results which are available for bigger biological model systems.

Finally, we considered interactions between surfaces of metal oxides and small molecules, which could be precursors of biomolecules. Comparative studies of adsorptions of atomic and molecular hydrogen as well as methanol on four model surfaces, ReO_3, WO_3, and ReO_3/WO_3 and WO_3/ReO_3 heterostructures, have been performed at the density functional theory level. For adsorption of atomic hydrogen the O_{1C} site is energetically the most favorable for all four surfaces. We also performed Bader's charge analysis and local DOS analysis, which provide a clear physical picture of hydrogen adsorptions, and facilitate a better understanding of the mechanisms behind these elementary processes. For adsorption of a hydrogen molecule, we have found two possible surface states which support adsorption of a hydrogen molecule, one is the molecular adsorption at the O_{1C} site, which is energetically more favorable, and the other is the dissociative adsorption at the O_{1C} site and at one of its neighboring O_{2C} sites. We found that the ReO_3 based surfaces have stronger interactions with the adsorbate hydrogen molecule, compared with the WO_3 based surfaces, and making overlayer-like interfacial structures could lower transition barriers and modify endothermicity. Our adsorption energy for a methanol molecule on WO_3(001) is in good agreement with experimental findings. Our results also confirm that methanol does not dissociate on WO_3(001). For adsorption of a glycine molecule on ReO_3(001), we found that the adsorption of a zwitterionic tautomer is thermodynamically more favorable than the canonical tautomer by 15 kJ/mol, due to the formation of strong surface hydrogen bonds between the amine group of the zwitterionic glycine and the surface O_{1C} atoms. The results of our comparative theoretical study performed for the four model surfaces indicate the catalytic reactivities of surfaces of conductive and semiconductive transition metal oxides might be tuned by making overlayer-like interfacial structures. The possibilities for improved catalytic activities, and whether the surfaces of metal oxides could catalyze and assist prebiotic synthesis of small

biomolecules, should be further explored by theoretical and experimental studies in the future.

Appendix A

Normal Modes and Thermodynamic Information

A.1 Normal Modes

Take a linear triatomic molecule BAB as an example, ignore the bending vibrations, and assume the displacements of three atoms B, A and B along the axis of the molecule are ξ_1, ξ_2 and ξ_3, respectively. For atom pairs in the two bonds of B-A and A-B, considering their relative displacements are $\xi_1-\xi_2$ and $\xi_3-\xi_2$, and they have the same force constants (say k), the potential energy could be written as below according to Eq. (2.31),

$$E_e^k \approx E_e^k(0)+\frac{1}{2}k(\xi_1-\xi_2)^2+\frac{1}{2}k(\xi_3-\xi_2)^2. \tag{A.1}$$

According to Eq. (2.32), $\xi_i=\frac{q_i}{m_i^{1/2}}$, insert it into Eq. (A.1), and consider that $K_{ij}=\left(\frac{\partial^2 E_e^k}{\partial q_i \partial q_j}\right)_0$, we have the force constants matrix as below,

$$\boldsymbol{K}=\begin{pmatrix} k/m_B & -k/(m_A m_B)^{1/2} & 0 \\ -k/(m_A m_B)^{1/2} & 2k/m_A & -k/(m_A m_B)^{1/2} \\ 0 & -k/(m_A m_B)^{1/2} & k/m_B \end{pmatrix}. \tag{A.2}$$

To remove cross terms of the mass-weighted coordinates q_i from Eq. (A.1), we would have to diagonalize the above matrix. The diagonalization could be done through a standard procedure, in which we solve the following secular equation,

$$|\boldsymbol{K}-\lambda\boldsymbol{1}|=\begin{vmatrix} k/m_B-\lambda & -k/(m_A m_B)^{1/2} & 0 \\ -k/(m_A m_B)^{1/2} & 2k/m_A-\lambda & -k/(m_A m_B)^{1/2} \\ 0 & -k/(m_A m_B)^{1/2} & k/m_B-\lambda \end{vmatrix}=0. \tag{A.3}$$

As a result, we get the roots of the above equation for λ as below,

$$\lambda_1=0,\quad \lambda_2=\frac{k}{m_B},\quad \lambda_3=\frac{k}{\mu}, \tag{A.4}$$

in which the effective mass $\mu=\frac{m_A m_B}{m_A+2m_B}$. Considering the eigenvectors of $\boldsymbol{K}$,

denoted as Q_l, are linear combinations of q_i, $Q_l = \sum_i c_{il} q_i$, they could be determined by solving the following set of simultaneous equations,

$$\sum_j \left(K_{ij} - \lambda_l \delta_{jl} \right) c_{jl} = 0, \tag{A.5}$$

in which $i = 1,2,3$ and $l = 1,2,3$. For $l = 1$, which corresponds to $\lambda_1 = 0$, Eq. (A.5) becomes

$$\sum_j K_{ij} c_{j1} = 0. \tag{A.6}$$

Considering the normalization condition $c_{11}^2 + c_{21}^2 + c_{31}^2 = 1$ has to be satisfied, we get

$$c_{11} = c_{31} = \left(\frac{m_B}{m} \right)^{1/2}, \; c_{21} = \left(\frac{m_A}{m} \right)^{1/2}, \tag{A.7}$$

in which $m = m_A + 2m_B$ is the total mass of the molecule. In this way, we get

$$Q_1 = \frac{1}{m^{1/2}} \left(m_B^{1/2} q_1 + m_A^{1/2} q_2 + m_B^{1/2} q_3 \right), \tag{A.8}$$

and Q_1 corresponds to the translational mode of the molecule. For $l = 2$ and $l = 3$, which correspond to λ_2 and λ_3 respectively, the normal coordinates Q_2 and Q_3 could also be determined in the same way as Q_1, and finally we get

$$Q_2 = \left(\frac{1}{2} \right)^{1/2} \left(q_1 - q_3 \right), \tag{A.9}$$

$$Q_3 = \left(\frac{1}{2m} \right)^{1/2} \left(m_A^{1/2} q_1 - 2m_B^{1/2} q_2 + m_A^{1/2} q_3 \right), \tag{A.10}$$

in which Q_2 and Q_3 correspond to the symmetric and antisymmetric stretching vibrations of the molecule, respectively.

A.2. Thermodynamic Information

To calculate the thermodynamic properties of an ideal gas, one needs to start from the determination of a molecular partition function, which carries all thermodynamic information of the entire system. Partition function of the entire system consisting of N molecules could be written as below,

$$Q = \frac{q^N}{N!}, \tag{A.11}$$

in which q is the molecular partition function, and it could be expressed as a product of partition functions corresponding to different types of motions due to the Born-Oppenheimer approximation, which enables one to write the total energy

as a sum of energies resulting from different types of motions. In this way, we have

$$q = q_{trans} \times q_{rot} \times q_{vib} \times q_{elec}, \tag{A.12}$$

in which q_{trans}, q_{rot}, q_{vib} and q_{elec} are translational, rotational, vibrational and electronic partition functions, respectively. A few models could be employed to simplify the expressions of these component parts of the molecular partition function, including particle in a box, harmonic oscillator and rigid rotor. Finally, we get their expressions which are used in Gaussian and we summarize them as below,

$$\begin{aligned}
q_{trans} &= \left(\frac{2\pi m k_B T}{h^2}\right)^{3/2} \frac{RT}{P}, \\
q_{rot} &= \frac{\pi^{1/2}}{\sigma_r}\left(\frac{T^3}{\Theta_{rot,x}\Theta_{rot,y}\Theta_{rot,z}}\right)^{1/2}, \ \Theta_{rot,j} = \frac{h^2}{8\pi^2 I_j k_B}, \\
q_{vib} &= \prod_{K=1}^{3N-6(5)} \frac{1}{1-e^{-\Theta_K/T}}, \ \Theta_K = \frac{h\omega_K}{2\pi k_B}, \\
q_{elec} &= \omega_0,
\end{aligned} \tag{A.13}$$

in which m is the mass of a molecule, k_B is the Boltzmann factor, T is the temperature of the ideal gas, R is the ideal gas constant, h is the Planck constant, P is the pressure of the ideal gas, σ_r is the symmetry number indicating the number of indistinguishable orientations of a molecule, I_j are the moments of inertia along different axis, ω_K are the vibrational frequencies corresponding to different normal vibrational modes K, and ω_0 is the electronic degeneracy of the molecule in the ground state. Once these components of the molecular partition function are known, then the molecular partition function and partition function of the entire system could also be determined, through which the internal energy, the entropy, the enthalpy, and the Gibbs energy could be calculated. Expressions for these thermodynamic quantities as a function of Q are summarized as below,

$$\begin{aligned}
U &= k_B T^2 \left(\frac{\partial \ln Q}{\partial T}\right)_V, \\
S &= k_B \left[\ln Q + T\left(\frac{\partial \ln Q}{\partial T}\right)_V\right], \\
H &= U + RT, \\
G &= H - TS.
\end{aligned} \tag{A.14}$$