Density Functional Theory Studies on the Chemistry and Properties of Selected Pyrrole Molecules

by

Belquis Mothana

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
December 2006

Your file Votre référence
ISBN: 978-0-494-27200-8
Our file Notre référence
ISBN: 978-0-494-27200-8

DALHOUSIE UNIVERSITY

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

Preliminary Pages

Examiners Signature Page (pii)

Dalhousie Library Copyright Agreement (piii)

Appendices

Copyright Releases (if applicable)

In The Name Of Allah, The Compassionate, The Merciful

To My Family

Table of Contents

List of Figures

List of Tables

List of Schemes

Abstract

The pyrrole heterocycle is an important component of many biological molecules and a key aspect of many synthetic pathways. It is incorporated as a fundamental substructure in a tremendous range of natural products and bioactive molecules; hence the study of the chemistry of pyrroles is of great interest. Computational methods provide a powerful tool for exploring reaction mechanisms, potential energy surfaces, excitation energies and the prediction of many properties of molecules and reactions. Such methods complement experimental results, and provide an insight that is not possible otherwise.

This thesis describes the application of density functional theory to the prediction of nuclear magnetic resonance (NMR) properties of pyrroles and a study of several reaction mechanisms of pyrrole synthesis. In the first part of the thesis, a valid computational scheme for the study of the ^{15}N and ^{13}C NMR properties of nitrogen-containing molecules is provided. The computational scheme is used to elucidate the relationship between electron-withdrawing groups on the nitrogen atom of pyrroles and their ^{15}N and ^{13}C NMR parameters. A correlation between the paramagnetic shift and the ^{15}N chemical shift of *N*-substituted pyrroles indicates that the experimentally observed ^{15}N chemical shift trend arises entirely from variations of the paramagnetic shift contribution to the chemical shift. The second part of the thesis concentrates on selected reaction mechanisms of pyrrole synthesis to provide insight into the biosynthesis of the porphobilinogen (PBG) pyrrole. In particular, the Paal-Knorr and Knorr pyrrole synthesis mechanisms are investigated. The preferred mechanisms are identified on the basis of computed potential energy surfaces for possible reaction pathways. The present results suggest that the hemiaminal cyclization is the preferred pathway for the Paal-Knorr pyrrole synthesis and that the enamine cyclization is the preferred step for the Knorr pyrrole synthesis. A simple mechanistic model for the PBG pyrrole synthesis reaction steps of the C-N and C-C bond formations is also investigated. The preferred mechanism for the PBG pyrrole synthesis is suggested to proceed by the formation of the C-N bond first followed by the C-C bond of the pyrrole ring through a mechanism similar to the Knorr pyrrole synthesis.

List of Abbreviations and Symbols Used

Abbreviations

NMR	Nuclear magnetic resonance
PBG	Porphobilinogen
ALA	5-aminolevulinic acid
PBGS	Porphobilinogen synthase
EWG	Electron-withdrawing group
DFT	Density functional theory
NCS	Natural chemical shielding
HF	Hartree-Fock
SCF	Self consistent field
LCAO	Linear combination of atomic orbitals
STO	Slater-type orbitals
GTO	Gaussian-type orbitals
CI	Configuration interaction
QCI	Quadratic configuration interaction
PT	Perturbation theory
MP	Møller-Plesset
KS	Kohn-Sham
L(S)DA	local (spin) density approximation
GGA	Gradient-corrected approximation
PW86	Perdew and Wang's 1986 exchange functional
B	Becke's 1988 exchange functional
PWx91	Perdew and Wang's 1991 exchange functional
LYP	Lee-Yang-Parr correlation functional
PW91	Perdew-Wang 1991 correlation functional
B3	Becke's three-parameter hybrid exchange functional
TMS	Tetramethylsilane
CHF	Coupled Hartree-Fock

GIAO	Gauge including atomic orbital
IGLO	Individualized gauge for localized orbitals
LORG	Localized orbital local origin
CSGT	Continuous set of gauge transformations
SOLO	Second-order LORG
GIAO-MBPT	GIAO-second-order many-body perturbation theory
MC-IGLO	Multiconfiguration IGLO
MO	Molecular orbital
SOS-DFPT	Sum-over-states density-functional perturbation theory
MAE	Mean absolute error
RMSE	Root mean square error
ZPVE	Zero-point vibrational energy
AIM	Atoms in molecules
NBO	Natural bond orbital
PCM	Polarizable continuum model
SCRF	Self-consistent reaction field
PES	Potential energy surface
TS	Transition state
STQN	Synchronous Transit-Guided Quasi-Newton
IRC	Intrinsic reaction coordinates
QM/MM	Quantum mechanics/molecular mechanics
Boc	Tert-buoxycarbonyl
Ms	Methanesulfonyl
Tf	Trifluoromethanesulfonyl
Me	Methyl
Bn	Benzyl
Ac	Acyl
gCOSY	Gradient-selected correlation spectroscopy
gHSQC	Gradient-selected heteronuclear single quantum coherence
gHMBC	Gradient-selected heteronuclear multiple bond correlation

Symbols

Ψ	wave function
$\hat{H}$	Hamiltonian operator
E	Energy
∇_q^2	Laplacian operator
$\|\Psi\|^2$	probability density
$\hat{T}$	kinetic energy operator
$\hat{V}$	potential energy operator
Z_A	charge of nucleus A
r_{ij}	distance between two electrons
R_{AB}	distance between two nuclei
Ψ_0	ground-state wavefunction
ϕ	trial function
ϕ	atomic orbital
$\hat{F}$	Fock operator
$\upsilon^{HF}(i)$	average potential felt by the i^{th} electron
$\hat{J}_j$	Coulomb operator
$\hat{K}_j$	exchange operator
ε_i	orbital energy
φ_μ	μ^{th} atomic orbital
$C_{\mu i}$	expansion coefficient
ψ_i	spin orbital
$F_{\mu\nu}$	element of the Fock matrix
$S_{\mu\nu}$	element of the overlap matrix
$P_{\lambda\sigma}$	element of the density matrix
$(\mu\nu\|\lambda\sigma)$	two-electron integral
α	alpha spin
β	beta spin
g_s	primitive Gaussian function
ρ_0	probability density
θ_i^{ks}	Kohn-Sham orbitals

E_{xc}	exchange-correlation functional
υ_{xc}	exchange-correlation potential
$\mathbf{I}$	spin quantum number
$\mathbf{B_0}$	magnetic field strength
γ	magnetogyric ratio
$\mathbf{h}$	Planck's constant
$\mathbf{m_I}$	magnetic spin quantum number
σ	chemical shielding
σ_{iso}	isotropic shielding
$\boldsymbol{\sigma}^{PAS}$	shielding tensor in the principal axis system
$\Delta\sigma$	Anisotropy
μ	nuclear magnetic moment
σ_d	diamagnetic shielding
σ_p	paramagnetic shielding
L_0	angular momentum
δ	chemical shift
δ_{iso}	isotropic chemical shift
$\mathbf{A_m}$	vector potential A
Λ_m	multiplicative factor
j^1	current density
δ_d	diamagnetic shift
δ_p	paramagnetic shift
δ_{aniso}	anisotropic shift

Acknowledgments

I would like to extend my sincere thanks to my supervisor Dr. Russell Boyd for his continuous encouragement, support and careful guidance throughout my Ph. D. work. I would also like to thank Dr. Alison Thompson and Dr. Chad E. Hadden for giving me the opportunity to interpret their experimental chemical shifts of substituted pyrroles and for the suggestions and collaboration to publish the results. I would also like to thank Dr. Michael Lumsden for his suggestions and help with the NMR aspect of the thesis. I would also like to thank Dr. Fuqiang Ban for his help in the first two years of my Ph. D.

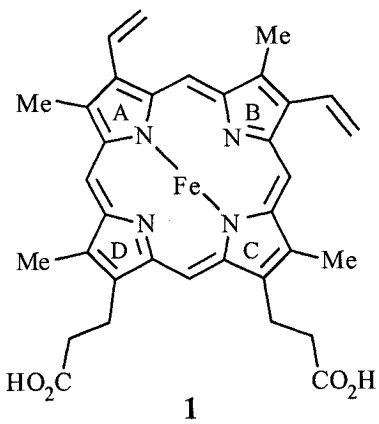
I would also like to thank my committee members: Dr. Alison Thompson, Dr. Philip D. Pacey, and Dr. Donald F. Weaver for their time and guidance for my Ph. D. work.
I also would like to thank all the group members of Dr. Boyd and the computational group of Dr. Weaver for providing a nice environment to work with.

Last but not least, I would like to thank my dear lovely family and my dear husband for all their support and encouragement.

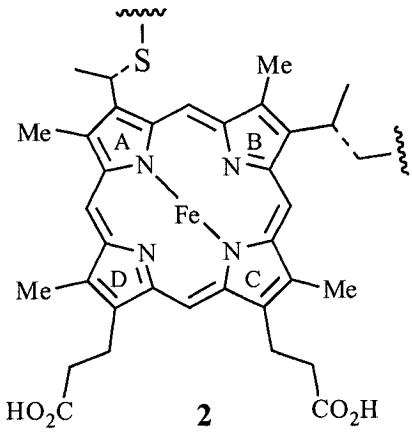
The Natural Sciences and Engineering Research Council of Canada (NSERC), the Killam Trusts, the Walter C. Sumner Foundation and Dalhousie University are gratefully appreciated for their financial support.

Chapter 1: Introduction

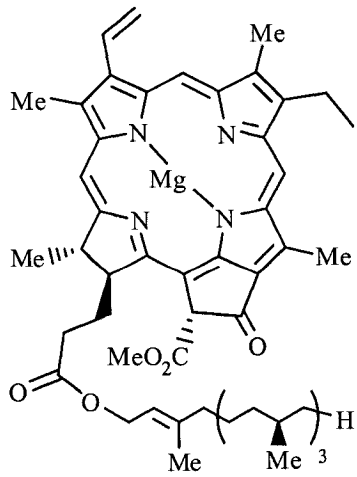
The pyrrole ring is among the most prevalent synthetic and naturally occurring heterocycles.[1] The pyrrole heterocycle is incorporated as a fundamental substructure in a tremendous range of natural products and bioactive molecules.[1,2] The most important class of natural products containing pyrrole rings are the tetrapyrroles, some of which are known as porphyrins. The tetrapyrroles are of great importance since they are responsible for oxygen transport (**1**, heme), electron transport (**2**, cytochrome c) and photosynthesis (**3**, chlorophyll a).[2,3]



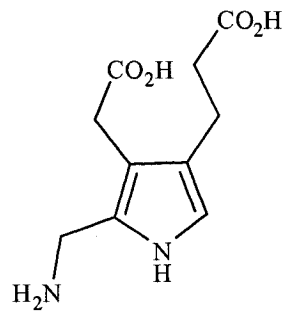
1



2



3

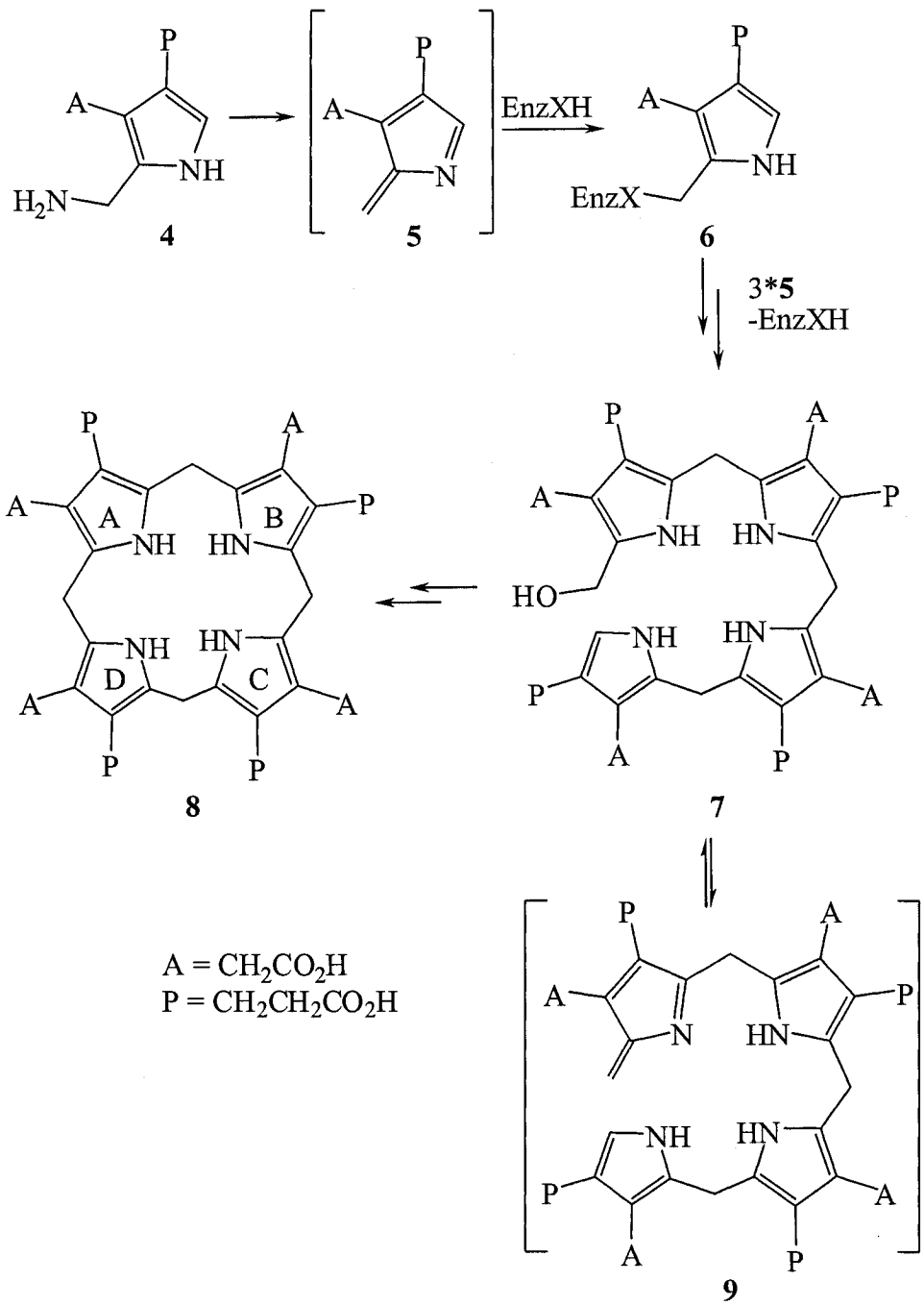


4

The first step in the biosynthesis of all natural tetrapyrroles, including heme, chlorophyll, and vitamin B_{12}, shares the porphobilinogen (PBG, **4**) pyrrole as a common precursor.[4] The biosynthesis of the PBG pyrrole derivative results from the condensation of two 5-aminolevulinic acid (ALA) catalyzed by the enzyme porphobilinogen synthase (PBGS).[4,5]

The mechanism proposed (Scheme 1.1) for the biosynthesis of tetrapyrroles from the PBG **4** pyrrole, the substrate for (hydroxymethyl)bilane synthase (PBG deaminase), is thought to proceed by giving rise to an azafulvene intermediate which becomes covalently bound to the enzyme via a dipyrromethane cofactor.[3,6,7] This sets up the ES_1 complex **6**, and three further pyrrole units are added, involving an azafulvene intermediate **5**, to yield the enzyme bound tetrapyrrole which is released as (hydroxymethyl)bilane **7** by reacting with water.[6] The (hydroxymethyl)bilane **7** then acts as a substrate for the cosynthetase enzyme to bring about the cyclization with inversion of ring-D to yield uroporphyrinogen III **8**.[6] The cyclization is also thought to involve an azafulvene intermediate **9**. The formation of azafulvenes is thought to play a key role in the biosynthesis of (hydroxymethyl)bilane **7** and uroporphyrinogen III **8**, important intermediates in the biosynthesis of porphyrins.

A mechanism involving azafulvenes has been proposed to explain the increased susceptibility of pyrroles to nucleophilic substitution.[1] The presence of electron-withdrawing groups (EWGs) on the nitrogen atom of a pyrrole has been thought to suppress the formation of the azafulvene intermediate.[6,8] The idea of suppressing the azafulvene formation by introduction of EWGs has been used in developing reactive inhibitors of serine protease.[9]



Scheme 1.1 Proposed mechanism for the biosynthesis of tetrapyrroles from PBG **4**.

Since the pyrrole heterocycle is an important component of many biological molecules and a key aspect of many synthetic pathways, the study of the chemistry of pyrroles is of great interest. Hence, the overall objective of this thesis is to study selected

pyrrole chemistry and properties using computational methods. Such methods complement experimental results, and provide an insight that is not possible otherwise.

Computational quantum chemistry is a well established and rapidly growing branch of chemistry. Quantum chemistry has been used in all branches of chemistry, biochemistry, molecular physics and structural biology in order to search for a deeper understanding of the electronic structure and properties of molecules. The application of computational chemistry has been greatly enhanced by the development of efficient density functional theory (DFT) methods[10,11] and the emergence of faster computers. Computational methods provide a powerful tool for exploring areas of chemical interest like substituent effects, reaction mechanisms, potential energy surfaces, excitation energies and are capable of predicting many properties of molecules and reactions. The work presented in this thesis focuses on the prediction of nuclear magnetic resonance (NMR) properties of pyrroles and a study of several reaction mechanisms of pyrrole formation.

1.1 Thesis Summary

A brief account of the theoretical background of quantum chemistry and the methods employed in this thesis is presented in Chapter 2. Basis sets and methods that are used in computational chemistry are briefly discussed. A detailed theoretical background of DFT is presented since it is employed in the calculations of NMR parameters in Chapters 4 and 5 and the studies of reaction mechanisms in Chapters 6 and 7.

In Chapter 3, a brief overview of the theory of NMR parameters is presented. A number of methods used for the calculation of ^{13}C and ^{15}N NMR parameters and their advantages and disadvantages are discussed. Chapter 3 provides the necessary background for the calculations and the interpretations of the NMR data employed in Chapters 4 and 5.

In Chapter 4, the performance of a number of DFT functionals in conjunction with selected basis sets for the prediction of ^{15}N and ^{13}C nuclear shieldings for small nitrogen-containing molecules is assessed. This study was carried out to develop a valid computational scheme for the study of the^{15}N and ^{13}C NMR properties of *N*-substituted pyrrole molecules in Chapter 5. The PBE1PBE/6-311+G(2df,p) level of theory is shown to provide the best choice for calculating the ^{15}N and ^{13}C nuclear shielding constants for molecules containing only singly-bonded nitrogen, while the B3PW91/6-311+G(2df,p) level of theory is almost as good. The levels of theory investigated are less satisfactory for predicting the ^{15}N and ^{13}C nuclear shielding constants of triply-bonded nitrogen-containing molecules.

In Chapter 5, calculations are carried out to investigate the relationship between EWGs on the nitrogen atom of pyrroles and their ^{15}N and ^{13}C NMR parameters. The study stemmed from experimental investigations of the ability of the ^{13}C NMR data of *N*-substituted pyrroles to predict the effect of EWGs on the mechanism of pyrrole nucleophilic substitutions. The study of the effects of EWGs on *N*-substituted pyrroles sheds light on the mechanisms involving azafulvene formation for pyrrole nucleophilic substitutions and biosynthesis of tetrapyrroles. Experimental ^{13}C chemical shifts showed a general increase in deshielding effect with the increase of *N*-EWG strength, whereas

^{15}N chemical shifts of pyrroles do not correlate with the *N*-EWG strength. DFT calculations were carried out to elucidate the relationship between the electronic structures of pyrroles and their ^{15}N and ^{13}C chemical shift data. A correlation between the paramagnetic shift and the ^{15}N chemical shift is observed for the pyrroles indicating that the ^{15}N chemical shift trend for the pyrroles arises entirely from variations of the paramagnetic shift contribution to the chemical shift. It is also shown that a general correlation between the ^{15}N chemical shifts and the EWG strength does not exist. Natural chemical shielding (NCS) analysis shows that the changes in the σ(N5-R)-π^* transitions and changes in the sum of the σ(C1-N5)- π^* and σ(C4-N5)- π^* transitions account for the nitrogen chemical shift trend observed in the pyrroles.

Chapters 6 and 7 describe DFT studies of the reaction mechanisms of the Paal-Knorr and Knorr pyrrole syntheses with the objective of providing an understanding of the mechanism of biosynthesis of the PBG pyrrole. The motivation of the study of the Paal-Knorr and Knorr reaction mechanisms is that the detailed sequence of the steps in the mechanisms and the nature of the intermediates are not clearly understood. In Chapter 6, the nature of the intermediates and transition states in the Paal-Knorr pyrrole mechanism is investigated. Two mechanistic pathways for the reaction are examined: hemiaminal cyclization vs enamine cyclization. The results of the calculated reaction potential energy surface suggest that the hemiaminal cyclization is the preferred pathway for the reaction both in gas phase and in solution. This conclusion is consistent with the experimental results which suggest that the hemiaminal intermediate undergoes cyclization in the rate-limiting step in the Paal-Knorr reaction mechanism.

In Chapter 7, the proposed mechanism of the Knorr pyrrole synthesis is investigated. It is shown that the Knorr mechanism involves high energy barriers for the reaction steps of the hemiaminal, enamine and pyrrole ring formation. The enamine cyclization step is highly exothermic and hence is the driving force for the completion of the reaction. A simple mechanistic model for the PBG pyrrole synthesis reaction steps of the C-N and C-C bond formation is also investigated. The results suggest that the preferred mechanism for the PBG pyrrole synthesis proceed by the formation of the C-N bond first followed by the C-C bond of the pyrrole ring through a mechanism similar to the Knorr pyrrole mechanism.

Finally, concluding remarks and possible directions for future research are provided in Chapter 8.

Chapter 2: Theoretical Background

2.1 The Schrödinger Equation

The goal of quantum chemical approaches is to find the approximate solution of the time-independent, non-relativistic Schrödinger equation

$$\hat{H}\Psi_i(\vec{x}_1,\vec{x}_2,...,\vec{x}_N,\vec{R}_1,\vec{R}_2,...,\vec{R}_M)=E_i\Psi_i(\vec{x}_1,\vec{x}_2,...,\vec{x}_N,\vec{R}_1,\vec{R}_2,...,\vec{R}_M) \tag{2.1}$$

where $\hat{H}$ is the Hamiltonian operator for a molecular system of M nuclei and N electrons in the absence of magnetic and electric fields.[12] $\hat{H}$ is a differential operator representing the total energy of the system:

$$\hat{H}=-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2-\frac{1}{2}\sum_{A=1}^{M}\frac{1}{M_A}\nabla_A^2-\sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}}+\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}}+\sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_AZ_B}{R_{AB}} \tag{2.2}$$

The indices A and B run over the M nuclei and i and j denote the N electrons in the system. The first and second terms represent the kinetic energy of the electrons and nuclei respectively, where the Laplacian operator ∇_q^2 is defined as a sum of differential operators in Cartesian coordinates:

$$\nabla_q^2 = \frac{\partial^2}{\partial x_q^2} + \frac{\partial^2}{\partial y_q^2} + \frac{\partial^2}{\partial z_q^2} \quad . \tag{2.3}$$

M_A is the mass of the nucleus A in multiples of the mass of an electron since the equation is represented using atomic units.[12,13] The other terms define the potential energy part of the Hamiltonian representing the attractive electrostatic interaction between the electrons and the nuclei, the repulsive potential energy due to electron-electron interactions and nucleus-nucleus interactions, respectively. The R_{AB}, r_{iA}, and r_{ij} represent the absolute distances between the nuclei A and B, the nucleus A and the electron i, and the electrons i and j, respectively. Z_A represents the charge on nucleus A.[12] The $\Psi_i(\vec{x}_1,\vec{x}_2,...,\vec{x}_N,\vec{R}_1,\vec{R}_2,...,\vec{R}_M)$ represents the wave function of the i^{th} state of the system, which depends on the $3N$ spatial coordinates and the N spin coordinates of the electrons, which are collectively termed $\{\vec{x}_i\}$, and the $3M$ spatial coordinates of the nuclei, $\{\vec{R}_I\}$. The wave function Ψ_i contains all the information possible for the quantum system. E_i represents the energy of the state described by Ψ_i.[12,13]

The wave function Ψ itself is not observable. A physical interpretation of the wave function that is experimentally observable is the probability density $|\Psi|^2$ which is defined as the square of the wave function.[12,14]

The exact solution of the Schrödinger equation is limited to one-electron systems such as the hydrogen atom and He^+ and cannot be solved exactly for the many-electron systems. Therefore theories that give feasible approximate solutions to the Schrödinger equation of the many-electron systems have been developed.

To simplify the problem of solving the Schrödinger equation, the Born-Oppenheimer approximation[15] is used, which takes advantage of the significant

difference in mass of electrons and nuclei and thus considers the electrons as moving in the field of fixed nuclei since the nuclei's motion is much slower than the electrons. Thus, the complete Hamiltonian given in equation (2.2) reduces to the so-called electronic Hamiltonian:

$$\hat{H}_{elec} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} = \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee}. \qquad (2.4)$$

The solution of the Schrödinger equation with $\hat{H}_{elec}$ is the electronic wave function Ψ_{elec} and the electronic energy E_{elec}. Ψ_{elec} depends on the electron coordinates; the nuclear coordinates enter only parametrically and do not appear explicitly in Ψ_{elec}.[13] The total energy E_{tot} is then the sum of E_{elec} and the constant nuclear repulsion term E_{nuc} (the last term in equation (2.2), i. e.,

$$\begin{aligned} \hat{H}_{elec}\Psi_{elec} &= E_{elec}\Psi_{elec} \\ E_{tot} &= E_{elec} + E_{nuc}. \end{aligned} \qquad (2.5)$$

2.2 The Variational Principle

The variational principle holds a prominent place in most quantum chemical applications. This principle is the foundation for several methods to find approximate solutions of the Schrödinger equation. The variational principle allows one to obtain an approximation of the ground-state energy of a system.[12,13]

In the variational principle, the ground-state wavefunction Ψ_0 can be replaced by a trial function ϕ that satisfies the boundary conditions. Then, according to the

variational principle, the expectation value of the energy obtained with the trial function is greater than or equal to the true ground-state energy

$$E_{trial} = \frac{\langle \phi | \hat{H} | \phi \rangle}{\langle \phi | \phi \rangle} \geq E_0 . \tag{2.6}$$

The closer ϕ is to Ψ_0, the closer E_{trial} is to the true energy E_0.[12,13,16] Thus the strategy for finding the ground-state energy and wavefunction using the variational principle is to minimize the E_{trial} functional by searching through all acceptable N-electron wavefunctions. However, such a search is obviously not feasible.

2.3 The Hartree-Fock Method

Although the Born-Oppenheimer approximation reduces the complexity of the Schrödinger equation, the electronic Schrödinger equation is still extremely complex and intractable for the many-electron systems, due to the electron-electron interactions. The problem is solved by introducing the Hartree approximation, which is known as the independent-particle model. The Hartree approximation assumes that each electron moves independently within its own orbital and only sees an averaged field generated by all the other electrons.[12,14]

Before a discussion of the Hartree-Fock (HF) method, the form of the wavefunction has to be defined. The form of the wavefunction has to satisfy certain criteria. The wavefunction has to satisfy the antisymmetry principle, which requires the wavefunction to be antisymmetric with respect to the interchange of electrons. The

wavefunction must consist of sums of the products of N spin orbitals and the spin orbitals should form an orthonormal set.[14,17]

As discussed above, it is impossible to minimize the energy by searching through all acceptable N-electron wavefunctions. Thus a subset must be defined that can offer a physically reasonable approximation to the exact wavefunction while being computationally feasible. In the HF method, the N-electron wavefunction is approximated by an antisymmetrized product of N one-electron wavefunctions ψ_i. The product is referred to as a Slater determinant:

$$\Psi_0 = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1\alpha(1) & \phi_1\beta(1) & \dots & \phi_N\beta(1) \\ \phi_1\alpha(2) & \phi_1\beta(2) & \dots & \phi_N\beta(2) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \phi_1\alpha(N) & \phi_1\beta(N) & \dots & \phi_N\beta(N) \end{vmatrix} . \tag{2.7}$$

The one-electron functions, $\psi_i = \phi(\vec{r})\sigma(s)$ (where $\sigma = \alpha, \beta$), are called spin orbitals. Spin orbitals have a physical interpretation that $|\phi(\vec{r})\sigma(s)|^2 d\vec{x}$ represents the probability of finding the electron of spin given by σ within the volume element. The $(N!)^{-1/2}$ factor is a normalization constant.[12-14]

Using the variational principal to minimize the E_0 with respect to the choice of spin orbitals, the HF equations can be derived, which determine the optimal spin orbitals. The HF equations for closed-shell systems are as follows:

$$\hat{F}\psi_i = \varepsilon_i \psi_i \qquad i = 1,2,\dots,N \tag{2.8}$$

where $\hat{F}$ is an effective one-electron operator, called the Fock operator, and is of the form

$$\hat{F} = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \upsilon^{HF}(i) = \hat{H}_{core} + \upsilon^{HF}(i) \tag{2.9}$$

where $\upsilon^{HF}(i)$ is the average potential felt by the i^{th} electron due to the other electrons. The essence of the HF approximation is to replace the complicated many-electron problem by a one-electron problem in which the electron-electron interactions are treated in an average way.[12,13] The explicit form of $\upsilon^{HF}(i)$ is as follows

$$\upsilon^{HF}(i) = \sum_{j}^{N}(2\hat{J}_j - \hat{K}_j) \tag{2.10}$$

where the Coulomb operator, $\hat{J}_j$, is defined by

$$\hat{J}_j(1) = \int \psi_j(2)\frac{1}{r_{12}}\psi_j(2)\,d\tau_2 \tag{2.11}$$

and represents the potential that an electron experiences due to the average charge distribution of another electron in spin orbital ψ_j.[12,13] The exchange operator, $\hat{K}_j$, has no classical interpretation and is not a simple function. It has the property

$$\hat{K}_j\psi_i(1) = \left[\int \psi_j(2)\frac{1}{r_{12}}\psi_i(2)\,d\tau_2\right]\psi_j(1). \tag{2.12}$$

The exchange operator $\hat{K}_j$ leads to an exchange of the variables in the two spin orbitals. The occurrence of the exchange term is entirely due to the antisymmetry of the Slater determinant and applies to all fermions.[12,13]

The best approximate ψ_i are eigenfunctions of the HF Hamiltonian operator $\hat{F}$ with eigenvalue ε_i (orbital energy). $\upsilon^{HF}(i)$ depends on the spin orbitals of the other electrons (i.e. $\hat{F}$ depends on its eigenfunctions). Thus, the HF equations are nonlinear

and must be solved iteratively.[13] The procedure to solve the HF equations is called the self consistent field (SCF) method. The SCF method starts with a 'guessed' set of orbitals, with which the HF equations are solved. The resulting new set of orbitals is then used in the next iteration and so on until the orbitals differ by less than a predetermined threshold.[12,13]

The HF equations have been solved numerically for atoms. However, the HF equations are not practical to solve for polyatomic molecules. Thus, the development of the Roothaan-Hall equations made such calculations feasible.

2.4 The Roothaan-Hall Equations

The HF equations require finding the position space and the spin-dependence of each of the orbitals in every single point that makes solving them impossible. Roothaan[18] suggested that instead of varying all the orbitals in all points, only a finite variation can be considered. The idea is that all ψ_i are expanded as a linear combination of atomic orbitals or in terms of basis functions

$$\psi_i = \sum_{\mu=1}^{M} C_{\mu i}\, \varphi_\mu \tag{2.13}$$

where φ_μ is an atomic orbital or a basis function and the $C_{\mu i}$ are the coefficients. Substitution of equation (2.13) in the Hartree-Fock equation leads to the Roothaan-Hall equations:

$$\sum_{\mu=1}^{M} (F_{\mu\nu} - \varepsilon_i S_{\mu\nu})\, C_{\mu i} = 0 \quad \nu = 1,2,\ldots,M \tag{2.14}$$

where ε_i is the one-electron orbital energy of the spin orbital ψ_i, F is the Fock matrix and S is the overlap integral between two orbitals. The form of the Fock matrix is defined by

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda,\sigma}^{M} P_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - \frac{1}{2} (\mu\lambda|\nu\sigma) \right] \tag{2.15}$$

where $H_{\mu\nu} = \langle \varphi_\mu | \hat{H}^{core} | \varphi_\nu \rangle$ is the matrix element of the core Hamiltonian operator with respect to φ_μ and φ_ν. $(\mu\nu|\lambda\sigma)$ is the repulsion integral between the two product densities $\varphi_\mu\varphi_\nu$ and $\varphi_\lambda\varphi_\sigma$, i.e.

$$(\mu\nu|\lambda\sigma) = \iint \varphi_\mu(1)\, \varphi_\nu(1) \frac{1}{r_{12}} \varphi_\lambda(2)\, \varphi_\sigma(2)\, d\tau_1\, d\tau_2 . \tag{2.16}$$

$P_{\lambda\sigma}$ is a matrix element of the density matrix

$$P_{\lambda\sigma} = 2 \sum_{i=1}^{occ.} C_{\lambda i}\, C_{\sigma i} \tag{2.17}$$

where the sum is over the occupied molecular orbitals.[13,14,17] The Roothaan-Hall equations are solved iteratively using the SCF method until self-consistency is obtained. The above equations apply for closed-shell systems. For open-shell systems, different orbitals are assigned to α and β spin and thus two different orbital expansions are used:

$$\psi_i^\alpha = \sum_{\mu=1}^{M} C_{\mu i}^\alpha\, \varphi_\mu , \qquad \psi_i^\beta = \sum_{\mu=1}^{M} C_{\mu i}^\beta\, \varphi_\mu \tag{2.18}$$

The coefficients are varied independently and lead to the open-shell generalization of the Roothaan-Hall equations for open-shell systems.[13]

2.5 Basis Sets

From equation (2.13), it was stated that the molecular orbitals can be formed by a linear combination of atomic orbitals (LCAO) or basis functions. Thus the LCAO method allows the atomic orbitals or basis functions to be represented by a basis set. A basis set is defined as a set of basis functions to describe the molecular orbitals within a system. There are two commonly used basis functions: the Slater-type orbitals (STO) and the Gaussian-type orbitals (GTO). The STOs have an exponential dependence represented as $\varphi_{STO} = N\exp(-\xi r)$, where N is a normalization constant, ξ is the orbital exponent that determines the size of the orbital.[13,19] The STOs give a reasonable description of the atomic orbitals and the nuclear cusp. However, it is very prohibitively complicated to calculate the three and four centered integrals, thus the GTOs are favored. The GTOs have the form $\varphi_{GTO} = N\exp(-\alpha r^2)$, where N is the normalization constant and α is the orbital exponent. The evaluation of the repulsion integrals is rapid using GTOs due to the Gaussian product theorem, which states that the product of two Gaussians on two different centers is a third Gaussian on a center between the first two. The GTOs are less satisfactory at describing the atomic orbitals since they are rounded in the region of the cusp and fall off too rapidly at larger r.[12,19]

In practice, the basis functions are formed by a linear combination of Gaussians. An s-type basis function can be expanded in terms of s-type Gaussians, for example,

$$\phi_\mu = \sum_s d_{\mu s}\, g_s \tag{2.19}$$

where the coefficients $d_{\mu s}$ are fixed.[20] This type of basis set represents a contracted Gaussian basis function, where the g_s's are known as primitive Gaussian functions. In an

uncontracted basis set, the coefficients are variational parameters. Thus contracting a basis set reduces the number of variational parameters and reduces computational time.

The Gaussian package of programs offers a wide range of defined basis sets, which can be classified by the number and types of basis functions they contain. A minimal basis set contains only the minimum number of basis functions needed for each atom.[20] A larger basis set is the split valence basis set that increases the number of basis functions per atom. 3-21G and 6-31G are examples of the split valence basis sets having two or more different sizes of basis function for each valence orbital. To better imitate the atomic orbital, additional Gaussians may be added to increase the accuracy of the basis set. These additional functions are called the polarization and diffuse functions. The polarization functions add orbitals to the basis set with an angular momentum beyond what is required for the ground-state description of each atom, such as adding d functions to carbon atoms.[20] The diffuse functions, which are larger versions of the s- and p-type functions in the valence shell, allow orbitals to occupy a larger region of space. Basis sets with diffuse functions are important for the description of systems where electrons are relatively far from the nucleus, such as anions, molecules with lone pairs, and other systems with significant negative charge.[20]

2.6 Electron Correlation

In the HF method, the electrons are independent of each other and only interact through an average potential. The HF wavefunction accounts for the interactions between electrons in an average way. However, electrons repel each other and the

motion of the electrons is correlated and this electron correlation should be treated in an instantaneous manner. The energy associated with the electron correlation can thus be defined as:

$$E_{\text{corr}} = E_{\text{exact}} - E_{HF} \tag{2.20}$$

where E_{exact} is the exact non-relativistic energy of the system, and E_{HF} is the HF energy obtained using a complete basis set (Hartree-Fock limit).[12,13] Since the HF energy is an upper bound to the exact energy, the correlation energy is negative. The correlation energy is very small, about 1% of the total energy. However, it is very important for many properties, especially for those involving bond breaking or bond formation. Methods which account for the electron correlation are referred to as post-SCF or post-HF methods. Such methods include the configuration interaction (CI), many-body perturbation theory and density functional theory (DFT).

2.7 Configuration Interaction

In the HF method, the wavefunction is approximated by a single Slater determinant. The configuration interaction (CI) method is based on the expansion of the wavefunction as a linear combination of Slater determinants. The first term is taken as the HF determinant and the subsequent determinants are those which differ from the HF determinant by exchanging occupied orbitals with unoccupied orbitals to create excited determinants.[13,14] Thus the CI form of the wavefunction is written as a linear combination of all possible N-electron Slater determinants that are formed from the set of spin orbitals:

$$|\Psi_0\rangle = c_0|\psi_0\rangle + \sum_{ar} c_a^r|\psi_a^r\rangle + \sum_{\substack{a<b \\ r<s}} c_{ab}^{rs}|\psi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c \\ r<s<t}} c_{abc}^{rst}|\psi_{abc}^{rst}\rangle + \ldots\ldots \qquad (2.21)$$

An exact wavefunction can be obtained through an infinite expansion. However some of the terms will not be equal in importance for the improvement of the wavefunction. Most of the correlation energy can be recovered using a limited expansion, if appropriate excitations are included. If all the excitations are included, the method is called full CI. A full CI calculation is computationally an impractical procedure except for the smallest molecules. A computationally viable scheme for the CI method is possible by truncating the CI expansion of the wavefunction at some excitation level and thus reducing the computational expense.[13,14] Some of the widely used truncated CI methods are CIS (includes only singly excited determinants), CID (includes only doubly excited determinants) and CISD (includes only singly and doubly excited determinants). CISD is a useful truncation of the CI expansion and accounts for most of the correlation energy.

One main problem with truncated CI methods is that they are not size-consistent, which might be a factor for large systems. Size-consistency implies that the relative errors in the calculations do not increase in proportion to the size of the molecule. The quadratic configuration interaction (QCI) method developed by Pople[21] is a way to make truncated CI size consistent. The QCI singles doubles (QCISD) is formed by the addition of higher excitation terms, quadratic in the expansion coefficients, which force size-consistency.[13]

2.8 Many-Body Perturbation Theory

The CI method is a variational method that accounts for the correlation energy. The perturbation theory (PT) is another method, which is not a variational method, to obtain the correlation energy. In the PT approach, the total Hamiltonian of the system $\hat{H}$ is divided into the Hamiltonian $\hat{H}^0$, the unperturbed system, which has known eigenfunctions and eigenvalues, and a perturbation, $\hat{H}'$. The Hamiltonian $\hat{H}$ is called the perturbed system. For $\hat{H}^0$ with known eigenfunctions and eigenvalues, the Schrödinger equation is $\hat{H}^0 \Psi_n^{(0)} = E_n^{(0)} \Psi_n^{(0)}$. The $E_n^{(0)}$ and $\Psi_n^{(0)}$ are called the unperturbed energy and unperturbed wavefunction of state *n*. To relate the perturbed system $\hat{H}$ with the unperturbed system with known eigenfunctions and eigenvalues, the perturbation can be imagined to be applied gradually giving a continuous change from the unperturbed to the perturbed system. Mathematically, this can be done by introducing a parameter λ into the Hamiltonian.[13,14] The Schrödinger equation of the perturbed system then becomes

$$\hat{H}\Psi_n = (\hat{H}^0 + \lambda \hat{H}')\Psi_n = E_n \Psi_n . \tag{2.22}$$

Since $\hat{H}$ depends on λ, then the eigenfunctions and eigenvalues of $\hat{H}$ depend on λ. Therefore Ψ_n and E_n can be expanded in a power series with respect to λ:

$$\Psi_n = \Psi_n^{(0)} + \lambda \Psi_n^{(1)} + \lambda^2 \Psi_n^{(2)} + ... + \lambda^k \Psi_n^{(k)} + ... \tag{2.23}$$

$$E_n = E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + ... + \lambda^k E_n^{(k)} + ... \tag{2.24}$$

where $\Psi_n^{(k)}$ and $E_n^{(k)}$ are the kth-order corrections to the wave function and energy.[13,14] If the perturbation is small, then taking just the first few terms of the series will give a good approximation to the true energy and wavefunction.

The most common method that uses the PT approach is the Møller-Plesset (MP) perturbation theory. In the MPPT, the Hartree-Fock Hamiltonian is used as the zeroth-order Hamiltonian. The electron correlation is then treated perturbatively. In practice, inclusion of the second, third, and fourth-order corrections accounts for most of the electron correlation. In the MP2 (second-order) method, the energy correction is calculated through to the second-order. This can be extended using more terms, for example, MP3, MP4, and so on. The truncation of MP calculations is size-consistent unlike that of CI. However, since this method is not variational, the energy obtained might be below the true energy.[13,14]

2.9 Density Functional Theory

2.9.1 The Hohenberg-Kohn Theorem

In 1964, Kohn and Hohenberg[10] showed that, for molecules with a non-degenerate ground-state wavefunction, the ground-state molecular energy and all other molecular electronic properties can be uniquely determined from the ground-state electron probability density, ρ_0, of a system.[22] The electron probability density is a function of only three variables, fewer variables than the wavefunction, and has the ability to calculate the properties of a system. The ground-state electronic energy E_0 is a

functional of the electron density function, $E_0 = E_0[\rho_0]$.[14] The main principle of the density-functional theory (DFT) is calculating E_0 and other ground-state molecular properties from the ground-state electron density.

2.9.2 Kohn-Sham Method

The Hohenberg-Kohn theorem states that in principle, it is possible to calculate all the ground-state molecular properties from ρ but it does not tell how to calculate E_0 from ρ or how to find ρ without finding the wavefunction.[14] Thus, Kohn and Sham[11] developed a series of one-electron equations to practically implement the Hohenberg-Kohn theorem. Kohn and Sham showed that the exact ground-state energy for an N-electron molecule with ground-state probability density ρ is given by

$$E_0 = -\sum_{\alpha} Z_\alpha \int \frac{\rho(\mathbf{r}_1)}{\mathbf{r}_{1\alpha}} d\mathbf{r}_1 - \tfrac{1}{2}\sum_{i=1}^{N} \left\langle \theta_i^{ks}(1) \middle| \nabla_1^2 \middle| \theta_i^{ks}(1) \right\rangle + \tfrac{1}{2} \iint \frac{\rho(\mathbf{r}_1)\,\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1\, d\mathbf{r}_2 + E_{xc}[\rho], \tag{2.25}$$

where the first term is the potential energy of the nuclear-electron attraction and of the repulsion between pairs of nuclei, the second term is the kinetic energy from the electron's motion, the third term is the electron-electron repulsion energy, and the last term is defined as the exchange-correlation energy. Therefore, using the above equation, we can find E_0 from ρ if we obtain the Kohn-Sham (KS) orbitals θ_i^{ks} and know what the functional E_{xc} is.[14]

The KS orbitals are found using the Hohenberg-Kohn variational theorem, which states that the ground-state energy can be obtained by varying ρ so as to minimize the $E_0[\rho_0]$. However, instead of varying ρ, the KS orbitals θ_i^{ks}, which are related to ρ by the relation: $\rho = \sum_{i=1}^{N} |\theta_i^{ks}|^2$, can be varied. The KS orbitals θ_i^{ks} can then be found by solving the one-electron equations:

$$\hat{H}_{ks}\,\theta_i^{ks}(1) = \varepsilon_i^{ks}\,\theta_i^{ks}(1) \tag{2.26}$$

$$\hat{H}_{ks} = -\tfrac{1}{2}\nabla_1^2 - \sum_{\alpha}\frac{Z_\alpha}{r_{1\alpha}} + \int\frac{\rho(\mathbf{r}_2)}{\mathbf{r}_{12}}d\mathbf{r}_2 + \upsilon_{xc}(1) \tag{2.27}$$

where ε_i^{ks}'s are the KS orbital energies. In the KS Hamiltonian operator $\hat{H}_{ks}$, the first three terms are the kinetic, nuclear-electronic, and Coulomb energies, respectively.[14,23] The last term υ_{xc} is the exchange-correlation potential and is the functional derivative of the exchange-correlation energy E_{xc}:

$$\upsilon_{xc}(\mathbf{r}) \equiv \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})}. \tag{2.28}$$

The one-electron KS operator is similar to the Fock operator in the HF equations in that it retains the kinetic, nuclear-electronic, and Coulomb terms of the HF Hamiltonian. However, the exchange operators $-\sum_{j=1}^{n}\hat{K}_j$ in the Fock operator are replaced in the KS Hamiltonian by υ_{xc}, which is a functional of the total charge density that handles the effects of both exchange and electron correlation.[14] Moreover, unlike the HF Hamiltonian, the KS Hamiltonian explicitly includes the dynamic electron correlation

within the potential term υ_{xc}.[14] The presence of the exchange correlation potential υ_{xc} permits the total non-Coulombic electron-electron interaction energy E_{xc} to be presented as a functional of the electronic charge density $E_{xc}[\rho]$.[14] Like the HF and Roothaan-Hall equations, the KS equations are solved in an iterative method via the SCF scheme.

One problem with using the KS equations to find ρ and E_0 is that the exact form of the functional $E_{xc}[\rho]$ is not known. Both E_{xc} in the energy expression in equation (2.25) and υ_{xc} in equation (2.27) are unknown.[14,23] Therefore, various approximate functionals $E_{xc}[\rho]$ were developed. DFT calculations are done using the various approximate functionals and the results are compared to experimental results to test the accuracy of the functionals. One of the drawbacks of the DFT method is the lack of a systematic procedure for improving $E_{xc}[\rho]$ and hence improving the molecular properties.[14,23]

2.9.3 Approximate Functionals of $E_{xc}[\rho]$

To aid in the development of approximate functionals for the KS DFT, the functional E_{xc} is written as the sum of an exchange-energy functional E_x and a correlation energy functional E_c, $E_{xc} = E_x + E_c$. E_x is defined similarly to the exchange energy in HF theory except that the HF orbitals are replaced by the KS orbitals.[14] The HF exchange energy involves the exchange integrals K_{ij}, thus replacing the HF orbitals by the KS orbitals, the exchange-energy functional is written as:

$$E_x = -\tfrac{1}{4}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\langle \theta_i^{ks}(1)\theta_j^{ks}(2)\left| 1/r_{12} \right| \theta_j^{ks}(1)\theta_i^{ks}(2)\right\rangle . \qquad (2.29)$$

A number of approximate functionals of $E_{xc}[\rho]$ have been developed to carry out DFT calculations. One of these functionals, which is attributed to Slater and Vosko, used the homogenous electron gas as a model to develop the following expression for the exchange energy:

$$E_x^S = -\frac{9}{4}\alpha\left[\frac{3\pi}{4}\right]^{1/3}\left(\rho_\alpha^{4/3} + \rho_\beta^{4/3}\right) \tag{2.30}$$

where ρ_α and ρ_β are the respective α- and β-electronic charge densities, and α is a parameter. Using the model of a homogenous electron gas, Vosko, Wilk and Nusair[24] (VWN) also produced expressions for the correlation energy of the general form

$$E_C^{VWN} = \left(\rho_\alpha + \rho_\beta\right)\varepsilon(x\varsigma) \tag{2.31}$$

where $x = \left[\frac{3}{4\pi\rho}\right]^{1/6}$ and $\varsigma = \frac{\rho_\alpha - \rho_\beta}{\rho}$.

The above functionals possess zeroth-order dependence on the respective energies of the electronic charge density and thus are known as local functionals. The above functionals of the exchange and correlation functionals are referred to as the local (spin) density approximation, L(S)DA:

$$E_{xc}^{L(S)DA} = E_x^S + E_c^{VWN}. \tag{2.32}$$

Since real electronic systems are rarely homogenous as assumed by the L(S)DA model, a number of corrections have been developed to deal with the inhomogeneities in the electronic charge density. Most popular approaches to deal with an inhomogeneous charge density involve a gradient correction to the electronic charge density $\nabla\rho$.[14,23]

Gradient-corrected approximations (GGA) correct for the variations of electron density with position by including gradients of ρ^α and ρ^β in the integrand. The gradient-

corrected functionals are usually called nonlocal functionals. The E_{xc}^{GGA} is split into the exchange and correlation parts:

$$E_{xc}^{GGA} = E_x^{GGA} + E_c^{GGA}. \tag{2.33}$$

These functionals are developed by theoretical considerations using known behavior of true functionals E_x and E_c, and empiricism in some cases.[14,23]

Some commonly used gradient-corrected exchange functionals E_x are Perdew and Wang's 1986 functional (PW86), Becke's 1988 functionals (B88, Bx88, Becke88, or B), and Perdew and Wang's 1991 exchange functional (PWx91). Commonly used gradient-corrected correlation functionals E_c include the Lee-Yang-Parr (LYP) functional, the Perdew 1986 correlation functional (P86), the Perdew-Wang 1991 parameter-free correlation functional (PW91) and the Becke correlation functional (B96).[14] Any exchange functional can be combined with any correlation functional. An example of a DFT functional used in some applications is BLYP which combines the Becke1988 exchange functional and the Lee-Yang-Parr correlation functional.

Hybrid exchange-correlation functionals are also widely used. A hybrid functional mixes together the formula (equation 2.29) with gradient-corrected E_x and E_c formulas. A popular hybrid functional is B3LYP, where B3 indicates Becke's three-parameter functional.

Gradient-corrected functionals and hybrid functionals give good equilibrium geometries, vibrational frequencies, dipole moments and generally accurate molecular atomization energies. Overall, the hybrid functionals seem to give the best performance.[14]

The advantage of DFT is that it allows for correlation effects to be included in a calculation that takes roughly the same time as a HF calculation, which does not include

correlation. Some drawbacks of the DFT are that since approximate functionals are used, the KS DFT is not variational, and can yield an energy below the true ground-state energy.

There is no clear way to construct more accurate E_{xc} functionals. Therefore, the usual practice is to try out new functionals one at a time and see whether they improve the results. Therefore much of the current work is devoted to developing improved E_{xc} functionals.

The DFT methods are generally less accurate than, for example, the CI and MP methods. However, DFT can handle large molecules whereas CI and MP calculations are limited to small molecules, making DFT a popular computational tool for studying chemical properties of large molecules.

Chapter 3: Magnetic Properties: NMR Shielding Theory

3.1 Chemical Shielding Theory

Nuclear magnetic resonance (NMR) spectroscopy is an invaluable tool in many areas of chemistry, biology, physics and medicine. In fact, NMR spectroscopy is now one of the most important tools for the characterization of organic, biological, and inorganic chemical systems.

Any non-zero spin quantum number (**I**) nucleus has a magnetic moment associated with it, that can interact with an applied external magnetic field.[25] The Hamiltonian **H** is represented as follows

$$\mathbf{H} = -\gamma \mathbf{B}_0 = -\gamma \mathbf{I} \cdot \mathbf{B}_0 = -\gamma \mathbf{I}_z \mathbf{B}_0, \tag{3.1}$$

where $\mathbf{B_0}$ is the strength of the applied magnetic field along the z-axis and γ is the magnetogyric ratio, a proportionality constant relating the nuclear magnetic moment to its intrinsic angular momentum.[25] In the presence of an applied magnetic field, the energy is evaluated by adding the following term to the energy of the system

$$\mathbf{E} = -\mathbf{m}_\mathbf{I}\, \mathbf{h} \gamma \mathbf{B}_0 / 2\pi, \tag{3.2}$$

where **h** is Planck's constant and $\mathbf{m_I}$ is the magnetic spin quantum number. $\mathbf{m_I}$ can have any of the following values:

$$\mathbf{m_I} = \mathbf{I}, \mathbf{I-1}, \mathbf{I-2}, \ldots, \mathbf{-I+2}, \mathbf{-I+1}, \mathbf{-I}, \tag{3.3}$$

describing the projection of the nuclear moment along the z-axis. For a nucleus with $\mathbf{I} = 1/2$, $\mathbf{m_I}$ can be either 1/2 or -1/2.[25] The energy difference between the two states derived from the above energy equation is

$$\Delta E = h\gamma B_0/2\pi. \tag{3.4}$$

A transition between the two energy levels can be induced by applying a second field $\mathbf{B_1}$ that is perpendicular to $\mathbf{B_0}$ and oscillating at the observation frequency of the nucleus concerned

$$\nu = \gamma B_0/2\pi. \tag{3.5}$$

Since each nucleus has its own characteristic magnetogyric ratio, distinctive frequencies for each type of nucleus at a given field strength, $\mathbf{B_0}$, can be detected.[25] The NMR chemical shift phenomenon was discovered almost 50 years ago, when Proctor and Yu[26] observed that the two different ^{14}N nuclei in NH_4NO_3 resonate at slightly different frequencies from each other. Another example of the NMR chemical shift phenomenon is the ^1H signals for ethanol.[27] Three different resonances were observed for each type of proton in the ethanol molecule instead of one. The observations of the NMR chemical shift phenomenon result from the fact that nuclei in nature are not in a vacuum, isolated and unaffected by other particles.[25] Nuclei are normally found amidst electrons, which are continuously in motion in an atom or a molecule. Application of an external magnetic field, $\mathbf{B^{ext}}$, perturbs the electronic wavefunctions and energies of the atom or molecule. The applied field causes electronic currents that induce a secondary local field, $\mathbf{B^{loc}}$, at any given nuclear site. Hence, a nucleus in a molecule feels a local field that may be less (shielded) or greater (deshielded) in magnitude compared with the $\mathbf{B^{ext}}$, depending

on the behaviour and distribution of electrons around the nucleus.[28] The local field felt by the nucleus can be expressed as follows

$$\mathbf{B}^{\text{loc}} = (1-\sigma)\mathbf{B}^{\text{ext}} \tag{3.6}$$

where σ is defined as the chemical shielding, a molecular electronic property specific to a given nuclear site.[25,28] The shielding σ of a nucleus will vary with its position in the molecule and its chemical environment.

In solid-state NMR, it is easy to realize that the shielding has to be orientation dependent, and the shielding is in fact represented as a second rank tensor quantity.[28] Thus, equation (3.6) can be rewritten as

$$\mathbf{B}_{\alpha}^{\text{loc}} = (1-\sigma_{\alpha\beta})\mathbf{B}_{\beta}^{\text{ext}} \tag{3.7}$$

where $\mathbf{B}_{\alpha}^{\text{loc}}$ is the magnitude of the field felt at the nucleus along the α-direction induced by electronic currents brought about by an applied magnetic field along the β-direction.[28] The shielding is composed of nine independent components, which can be written as a 3×3 matrix:

$$\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix}. \tag{3.8}$$

Symmetry generally reduces the number of independent components. The shielding rank-2 tensor can be broken down into three portions: isotropic, antisymmetric, and symmetric of rank 0, 1, and 2, respectively.[29] Thus the shielding tensor equation (3.8) can be rewritten as

$$\boldsymbol{\sigma}=\sigma_{iso}\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}+\begin{bmatrix}0&\sigma_{xy}^{A}&\sigma_{xz}^{A}\\\sigma_{yx}^{A}&0&\sigma_{yz}^{A}\\\sigma_{zx}^{A}&\sigma_{zy}^{A}&0\end{bmatrix}+\begin{bmatrix}d_{xx}&d_{xy}&d_{xz}\\d_{yx}&d_{yy}&d_{yz}\\d_{zx}&d_{zy}&d_{zz}\end{bmatrix} \tag{3.9}$$

where σ_{iso} is the isotropic shielding, which is simply an average of the principal components of the diagonal symmetric shielding tensor, and is written as follows

$$\sigma_{iso}=1/3(\sigma_{xx}+\sigma_{yy}+\sigma_{zz}). \tag{3.10}$$

The antisymmetry parameters are defined as follows

$$\sigma_{\mu\nu}^{A}=1/2(\sigma_{\mu\nu}-\sigma_{\nu\mu}) \tag{3.11}$$

and the parameters $d_{\mu\nu}$ are defined as follows

$$d_{\mu\nu}=1/2(\sigma_{\mu\nu}+\sigma_{\nu\mu}-2\sigma_{iso}). \tag{3.12}$$

The antisymmetric rank-1 tensor contains only three distinct elements, whereas the symmetric rank-2 tensor has five independent elements.[29] It is difficult to determine all the tensor elements in an experiment. In the standard NMR experiments, only the symmetric part of the shielding (rank-0 and rank-2 tensors in equation 3.9) affects the spectrum to first order. The antisymmetric shielding part does not influence the NMR spectra to first order.

The shielding tensor is usually expressed in the principal axis system, which is the coordinate system for which the rank-2 tensor is diagonal.[29] Thus the shielding tensor, disregarding the antisymmetric rank-1 tensor, is reduced to

$$\boldsymbol{\sigma}^{PAS}=\sigma_{iso}\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}+\begin{bmatrix}d_{xx}^{PAS}&0&0\\0&d_{yy}^{PAS}&0\\0&0&d_{zz}^{PAS}\end{bmatrix} \tag{3.13}$$

The three components in the symmetric tensor are referred to as the "principal components".

The anisotropy and asymmetry of the shielding are two quantities related to the shielding tensor that are usually reported in the literature. The anisotropy and the asymmetry of the shielding tensor are expressed, respectively, as follows

$$\Delta\sigma = \sigma_{33}^{PAS} - 1/2(\sigma_{11}^{PAS} + \sigma_{22}^{PAS}) \tag{3.14}$$

$$\eta = \frac{(\sigma_{22}^{PAS} - \sigma_{11}^{PAS})}{(\sigma_{33}^{PAS} - \sigma_{iso})} \tag{3.15}$$

where it is assumed that $\sigma_{33}^{PAS} \geq \sigma_{22}^{PAS} \geq \sigma_{11}^{PAS}$.[29]

In liquid, gas-phase or solution NMR spectroscopy, the free tumbling motion of the molecules usually averages the shielding tensor without preference to any particular orientation with respect to the applied field.[25,28] Thus, the observed shielding property in liquid, gas-phase or solution NMR is the isotropic shielding, which is the average of the principal components.[25,28]

Both the applied magnetic field and the nuclear moment cause changes in the electronic wavefunctions and energies of the molecule.[28] These changes in energy due to the simultaneous magnetic perturbation can be used to define the shielding from a quantum chemical perspective as being the mixed second derivative of the energy with respect to an applied magnetic field and the nuclear magnetic moment:

$$\sigma_{\alpha\beta} = \left.\frac{\partial^2 E}{\partial \boldsymbol{B}\alpha\, \partial\mu\beta}\right|_{B=0,\,\mu=0}, \tag{3.16}$$

where E is the total molecular eigenenergy.[28]

3.2 Ramsey's Theory

The history of the theoretical analysis of the shielding property parallels the advances in the theoretical evaluation of the energy, *E*, since the shielding is electronic in its origin.[28] The theories of the magnetic shielding begin with Ramsey's theory, which uses the second-order perturbation theory.[30] Ramsey collected all terms in the energy equation that are proportional to both the magnetic moment and the external field obtained after a perturbation treatment and decomposed the shielding into two terms: a diamagnetic term and a paramagnetic term.[30] A magnetic field induces two kinds of electronic current in a molecule: diamagnetic and paramagnetic. The diamagnetic and paramagnetic currents flow in opposite directions and thus give rise to nuclear shielding and deshielding, respectively. The shielding can then be rewritten as a sum of diamagnetic and paramagnetic shielding:

$$\sigma = \sigma_d + \sigma_p \tag{3.17}$$

where σ_d is positive and σ_p is negative.

3.3 Diamagnetic Shielding

Diamagnetic currents arise from the movement of the electrons within atomic or molecular orbitals.[31] The external field causes electrons to circulate in a plane perpendicular to the field direction. The induced current generates a small local field opposed to $\mathbf{B_0}$. The magnitude of the diamagnetic current is determined solely by the ground state electronic wavefunction of the atom or molecule.[25,31] The magnitude of the

diamagnetic current depends on the electron density close to the nucleus and provides the contribution to σ for spherical, closed-shell atoms.[31] The diamagnetic shielding is easy to calculate for atoms and it increases with the number of electrons: 1.7 ppm for hydrogen, 261 ppm for carbon, rising to 10,000 ppm for atoms in the fifth row of the periodic table.[31]

In Ramsey's theory, the diamagnetic shielding, σ_d, is derived from the first-order correction to the energy, which only involves the ground state wavefunction, $|0\rangle$:

$$\sigma_d \approx \langle 0 | (r_0 . r_n) / r_n^3 | 0 \rangle . \tag{3.18}$$

Therefore, the evaluation of the diamagnetic shielding is straightforward since it only requires the knowledge of the positions of the electrons with respect to the nucleus of interest, $\mathbf{r_n}$, and with respect to an origin chosen for the vector potential $\boldsymbol{r_0}$.[28] The diamagnetic shielding only depends on the electron distribution in the ground electronic state.[28,31]

3.4 Paramagnetic Shielding

Paramagnetic currents arise from the movement of electrons within molecules by a circuitous route. For example, in an artificial molecule with two electronic states: a ground state with the form of an atomic p_x orbital having two spin-paired electrons, and an unoccupied excited state resembling a p_y orbital; the electrons circulate from the p_x to the p_y. The molecular orbitals of a molecule in an applied magnetic field are modified owing to the field-induced circulations of the electrons.[31,32] The perturbation can be

expressed by the mixing of unoccupied orbitals into each occupied molecular orbital.[32] Thus an applied field in the z-direction partially overcomes the energy gap between the p_x and p_y and creates a path for the electrons in the p_x orbital to circulate in the xy plane. This induced current generates a magnetic field that augments the external field and deshields the nucleus at the center of the electron density.[31]

The deshielding is clearly linked to the energy gaps involved. The degree of mixing increases with decreasing energy difference between the orbitals and so the deshielding increases.[32] Thus the paramagnetic shielding, σ_p, should be approximately inversely proportional to ΔE, the average excitation energy.[31,32] In principle, the mixing into each orbital has to be taken into account. However, the HOMO-LUMO combination usually yields the largest contribution to the shielding. Other combinations that involve more stable occupied or more energy-rich unoccupied orbitals can also have a significant contribution to σ_p. All the deshielding effects in the same spatial direction can then be added up. The orbital-mixing must obey a selection rule: the combination of an occupied orbital and an unoccupied orbital is magnetically active only when the electron transfer comprises a change in angular momentum.[32] Thus a spherical orbital, like 1s or 2s, produce a magnetically inactive orbital mixing. Similarily, a π-π^* orbital combination in alkenes is magnetically inactive since both orbitals point in the same direction. However, the σ_{CH}-π^* and the π-σ_{CH}^* combinations of alkenes are magnetically active, yielding deshielding contributions perpendicular to the σ-π plane of charge circulation.[32] The π systems possess low-lying unoccupied π^* orbitals that allow efficient mixing with appropriate energy-rich occupied orbitals.[32] These orbital combinations usually offer the most important deshielding contributions. However other orbital combinations of the

occupied π orbital with an unoccupied orbital (other than π*) also contribute to the deshielding. When π and π* orbitals are absent as in alkanes, the ΔE^{-1} term loses importance since the energy gap is large and thus the deshielding part, σ_p, becomes small.[32]

The paramagnetic shielding is also related to the distance between the nucleus and its surrounding electrons, $\mathbf{r}_n$, since the magnetic field at the center of a small current loop is proportional to the inverse cube of its radius.[31] Thus the paramagnetic shielding is proportional to the inverse of the cube of the distance between the nucleus and its surrounding electrons, $\mathbf{r}_n$.

In Ramsey's theory, the paramagnetic shielding is expressed as follows

$$\sigma_p \approx -\sum_q \left| \left(E_q - E_0\right)^{-1} \left(\langle q | L_n / r_n^3 | 0 \rangle \langle 0 | L_0 | q \rangle \right) \right| \tag{3.19}$$

where the perturbed wavefunction is expanded in terms of the unperturbed virtual states $|q\rangle$. The sum over q is a summation over all excited electronic states. E_q is the unperturbed energy of the electronic state q. (E_q-E_0) is the difference between two states. L_n is the angular momentum with respect to the nucleus of interest, whereas L_0 is the angular momentum with respect to the origin of a magnetic vector potential. r_n is the distance between the nucleus and its surrounding electrons.[25,28]

The paramagnetic shielding is difficult to evaluate since it requires the evaluation of the wavefunctions $\langle q|$ and energies E_q of the excited states. It vanishes for electrons in s orbitals, which have zero angular momentum, whereas it becomes quite large when there is an asymmetric distribution of p and d electrons close to the nucleus and these electrons have low-lying excited states.[33]

3.5 Chemical Shift

In the NMR experiment, chemical shielding constants are measured relative to a reference substance, and are given as the chemical shift. *Ab initio* calculations, on the other hand, produce the absolute shielding tensors σ, and their traces σ_{iso}.[34] The shielding tensor is directly related to the chemical shift δ by the following equation:

$$\delta = \frac{\sigma_{ref} - \sigma}{1 - \sigma_{ref}} \tag{3.20}$$

where σ_{ref} is the shielding for a chosen reference compound. For light nuclei, the chemical shift is approximated by:

$$\delta = \sigma_{ref} - \sigma \tag{3.21}$$

since the denominator is close to unity (σ_{ref} is small). The above discussions of the shielding tensors also apply to the chemical shift tensors. The chemical shift usually reported is the isotropic chemical shift, δ_{iso}, which is the average of the three tensor components.[34]

3.6 Chemical Shift Referencing

As discussed above, the chemical shifts in the NMR experiment are referenced to a standard. The accepted chemical shift reference for ^{1}H and ^{13}C is tetramethylsilane (TMS), $Si(CH_3)_4$. However, for ^{15}N several reference standards have been used and there is no single universally accepted reference. The ^{15}N chemical shifts are usually referenced to three different standards. These standards include nitromethane, nitric acid

at several molar concentrations, and various forms of ammonium, including liquid ammonia. Liquid ammonia affords a chemical shift reference that resembles closely the TMS scale for ^{1}H and ^{13}C chemical shift referencing.[34,35]

3.7 Ab Initio Calculations of Shielding

Ramsey's formulation of shielding requires knowledge of ground and excited electronic states and their energies. The method represents the external magnetic field using a vector potential, and thus the results depend on the choice of origin for this vector potential. Therefore, both of the paramagnetic σ_p and diamagnetic σ_d terms depend on the chosen origin for the vector potential. The total shielding σ_{tot} is an observable quantity and should be independent of the vector potential choice of origin. A change of vector potential origin results in additional terms for σ_d and σ_p , which are similar in magnitude and opposite in sign rendering σ_{tot} unchanged. However, in practise, the cancellation does not occur due to the inherent difficulty in calculating the σ_p term.[25]

The ground and excited electronic states can be obtained by an SCF calculation in the absence of B_0. The molecular orbitals, a linear combination of atomic orbitals, describe the electronic states and Gaussian functions are used to represent the atomic orbitals. These zero-field states can be used to calculate the shielding via the perturbation method. The computational time increases as N^4, thus limiting the number of Gaussian functions used and hence, both ground and excited states end up as mere approximations. Also, the evaluation of σ_p is impossible since most states are unknown. Since the σ_d and

σ_p terms have different sensitivities to these approximations, errors are not likely to cancel.[25]

Various challenges arise from the limitations discussed above. One concerns the choice of gauge origin for the external magnetic vector potential. The paramagnetic term needs a very large number of Gaussian functions or basis functions to describe the electronic excited states properly. Furthermore, the coefficients, which define how the atomic orbitals are combined to give the molecular orbitals, are optimized during the SCF step with respect to an origin within the region of space. Thus, molecular orbitals that are far from the chosen origin are not well represented. As a result, the total shielding σ_{tot} becomes dependent on the choice of gauge origin.[25]

3.8 The Gauge Problem

Strategies have been developed to overcome the computational limitations discussed. The methods developed solve and restore the gauge invariance. The gauge problem can be defined by the dependence on the position of the molecule when magnetic fields are present. The origin of the gauge problem comes about from the truncation of the atomic orbital set used to describe the molecular orbitals.[36]

The most common and straightforward solution to the gauge problem is to introduce a dependence in the atomic orbitals to provide the correct first-order solution to the atomic problem in a homogenous field, irrespective of the atom position. This can be done by the multiplication of the atomic orbitals by a complex exponential factor (gauge factor) that depends on the magnetic field, the center coordinates of the atomic orbital

and the cartesian coordinates of the electron.[36] The techniques developed to restore the gauge invariance can be divided in two categories: coupled Hartree-Fock approaches and beyond Hartree-Fock methods.

3.9 Coupled Hartree-Fock (CHF) Approaches

3.9.1 GIAO - Gauge Including Atomic Orbital Method

The GIAO method is one of the quantum mechanical techniques for the calculation of magnetic properties of chemical systems. The GIAO method restores the gauge invariance by introducing field dependence into the basis functions.[36] The basis functions are modified by adding a multiplicative complex factor:

$$\chi_m = \exp(-(i/c)\ \mathbf{A}_m.\mathbf{r})\phi_m \tag{3.22}$$

where $\mathbf{A}_m$ is the vector potential A, at the nuclear position $\mathbf{R}_m$,of the atomic orbital ϕ_m :

$$\mathbf{A}_m = 1/2\,\mathbf{B}_0 \times \mathbf{R}_m\,. \tag{3.23}$$

The result of this modification is that each atomic orbital has its own local gauge origin placed at its center. Hence, this leads to more accurate calculations of the local paramagnetic term. The magnetic field dependence explicit in GIAOs leads to a disadvantage in that the CHF equations with GIAOs require a recalculation of all electron integrals. The integrals obtained from the SCF calculations can not be used since these integrals do not include the extra multiplicative factor.[25,36] Thus, the use of GIAOs is perfect other than the fact that they require time and memory to calculate the expensive

integrals. The improvement of computer software and hardware made the GIAO method accessible to larger molecules.

3.9.2 IGLO – Individualized Gauge for Localized Orbitals Method

The IGLO method was introduced by Kutzelnigg[37] in 1980. This method uses localized molecular orbitals, ψ_m, that are related to the field-independent molecular orbitals, ϕ_m, by

$$\psi_m = \exp(-i\Lambda_m)\phi_m \tag{3.24}$$

where Λ_m is a local multiplicative factor proportional to the field strength

$$\Lambda_m = (e\pi / ch)(\mathbf{R_m} \times \mathbf{B}).\mathbf{r}\,. \tag{3.25}$$

The origin of $\mathbf{R_m}$ is chosen to be at the centroid of localized orbital ψ_m.[25,38]

The IGLO method has an advantage over the GIAO method in that that gauge invariance is achieved without expensive integral calculations since it makes use of integral transformation, commutation and closure relations. Another advantage is that since the IGLO method uses localized molecular orbitals, shielding can be obtained in terms of contributions from chemical bonds and lone pair. However, the IGLO method produces poor results for delocalized systems.[25,38]

3.9.3 LORG – Localized Orbital Local Origin Method

The LORG method was formulated by Hansen and Bouman,[39] five years after the IGLO method was introduced. The LORG method begins by expressing the matrix elements between the ground and excited states in terms of a hypervirial equation

$$\langle 0|F|q\rangle = (1/\omega_{0q})\langle 0|F,H|q\rangle \tag{3.26}$$

where F is any one-electron operator, H is the Hamiltonian and ω is the energy of the transition from the ground state $\langle 0|$ to an excited state $\langle q|$. The ground state is obtained from the HF ground state plus added doubly excited configuration contributions.[25,40] The excited states are expressed as a linear expansion in terms of singly excited HF states. The LORG method is similar to the IGLO method in that it involves localized molecular orbitals, invokes closure relations, and requires integral transformations. Therefore, local bond contributions of shielding can also be extracted in the LORG method.[25,40]

3.9.4 CSGT – Continuous Set of Gauge Transformations Method

The CSGT was introduced in 1992 by Keith and Bader.[41] Instead of calculating the shielding from the use of molecular orbitals, the CSGT method calculates shielding by use of charge densities. The atomic shielding is expressed in terms of the induced first-order electronic current density $\boldsymbol{j}^1$ in the presence of an applied external field and is calculated over each spatially defined atom, Ω, in a molecule as follows

$$\sigma^N(\Omega) = (-1/cB)\int_{\Omega} d\boldsymbol{r}_N[\boldsymbol{r}_N \times \boldsymbol{j}^1(\boldsymbol{r})/\boldsymbol{r}_N^3]. \tag{3.27}$$

The atomic shielding contributions are summed to give $\sigma_{tot.}$.[25,28] Gauge invariance is achieved by accurate calculation of the induced first-order electronic current density and by performing gauge transformation for each point in space. The shielding values produced by the CSGT method are of similar quality to GIAO, IGLO, and LORG.[25,28]

3.10 Beyond Hartree-Fock Methods

3.10.1 SOLO – Second-Order LORG Method

The methods described above were all in the limit of using a single-determinant wavefunction to describe the electronic ground state of the molecule. The electron correlation is only treated as an effective field when describing the repulsion terms in the Hamiltonian. The electron correlation is neglected in GIAO, IGLO, LORG and CSGT methods. The importance of electron correlation appears when there are lone pairs and multiple bonds present which guarantee the presence of low-lying excited states. It is also expected that heavier elements would require some consideration of electron correlation by a virtue of having occupied orbitals at higher levels.

The incorporation of electron correlation in shielding calculations was achieved by Hansen and Bouman[42] in their SOLO approach.[40] The SOLO method extends LORG to second order in electron correlation. SOLO damps out errors associated with the calculations of the long-range contributions to σ_d and σ_p terms. The σ_p term is calculated with correlated wavefunctions and thus SOLO provides improved accuracy compared with LORG and IGLO.[25,40]

3.10.2 GIAO-MBPT(2) – GIAO-Second-Order Many-Body Perturbation Theory

The GIAO-MBPT(2) is a second approach that incorporates electron correlation effects using correlated wavefunctions.[25] Gauss[43] was the first to implement this approach to the GIAO method. The GIAO-MBPT(2) method employs a relaxed density matrix. MP2 is the simplest level of theory to include electron correlation in MPPT. The correlation energy is given by second-order double excitation amplitudes. Since electron correlation is incorporated, computational requirements increase. More storage and memory is required. The inclusion of second-order electron correlation improves the shielding calculations. The use of higher levels of MBPT theory and coupled cluster levels have also been done.[25]

3.10.3 MC-IGLO – Multiconfiguration IGLO

The shielding in molecules that contain low-lying excited states cannot be satisfactorily represented using a single-determinant wavefunction for the ground state. In such systems, there is a significant mixing between the occupied and virtual orbitals and it would be more appropriate to employ several determinants for the ground state. Van Wüllen and Kutzelnigg[44] introduced a new version of IGLO which made use of multiconfiguration SCF wavefunctions called MC-IGLO.[25] The use of multiconfiguration SCF wavefunctions is computationally expensive. However, improved shielding results are obtained for molecules like O_3, where electron correlation is imperative for a proper description of the ground state. The error for the oxygen shielding obtained with SCF-IGLO is 2000 ppm, but MC-IGLO reduces the error to

about 100 ppm.[25] The O_3 molecule is poorly described in the SCF method since electron correlation effects are high due to the presence of multiple bonds and lone pairs.

3.10.4 DFT - Density Functional Theory

Post-Hartree-Fock approaches to the shielding computation: SOLO, GIAO-MBPT(2), and MC-IGLO, have been limited to small molecules owing to the large requirement of computer memory and disk storage. DFT provides a faster and cheaper alternative to other expensive electron correlation methods while including electron correlation effects with less computational resources and efforts. DFT is promising for the computation of the shielding of larger molecules since the calculations take less time and improved shielding is obtained due to the incorporation of electron correlation. The quality of DFT shielding calculations depends on the quality of the exchange-correlation functionals. The quality of exchange-correlation functionals has been tested for small molecules in the literature and more functionals are being developed to increase the ability of these functionals to reproduce magnetic properties, i.e., the shielding constant.[25,29]

3.11 Molecular Orbital Theory with GIAOs

The GIAO method has certain advantages over the other methods (IGLO and LORG) that place the gauge factors on localized molecular orbitals rather than atomic orbitals. There is evidence that the GIAO calculations require somewhat smaller basis

sets than IGLO for the same accuracy.[25,29] Another advantage of the GIAO method is that the localization of MOs is not always unique since some systems have delocalized molecular orbitals.[25,29] In such cases of delocalized systems, the localized theories may give poor results. Moreover, the GIAO method is incorporated in the SCF, DFT, and MP2 methods in the Gaussian software.

3.12 Factors Influencing Shielding Calculations

One of the factors that influence the accuracy of the shielding calculation is the level of theory. The choice of the level of theory depends on the importance of the inclusion of electron correlation. The HF method only uses single determinant wavefunctions and is appropriate for shielding calculations where electron correlation is not significant. However, in molecules where there are low-lying excited states such as in nuclei that have lone pairs and multiple bonds, electron correlation becomes significant and is required to properly describe the ground state. The NMR shielding depends on the hybridization of the lone pair, and minor inaccuracies in the hybridization may result in significant errors in the chemical shielding.[25,36]

NMR shieldings are sensitive to the molecular geometry, and therefore accurate shielding calculations require good geometries. The use of experimental geometries is problematic in that their accuracy is randomly varying, depending on the method and the molecule itself. Optimized geometries, although their accuracy is often no better, provide a homogenous set of data if a single method and basis set is used throughout.[25]

The choice of the basis set is crucial for the accuracy of the shielding calculation. The basis set has to be large enough to describe the effects of the perturbation as well as the correlation effects. For an accurate calculation of the nuclear shielding, a basis set of at least valence triple-zeta quality and with at least one set of polarization functions is needed.[29,36]

NMR shielding calculations requirements depend on the nuclei being investigated. The different nuclei have different sensitivities to the factors discussed above. For example, ^{19}F shieldings require larger basis sets than ^{13}C shieldings. Nuclei that have lone pairs and multiple bonds need more electron correlation since they have low-lying n-π^* transitions.[29,36]

Other factors that influence the shielding calculation are relativistic effects for heavier elements. Relativistic effects have to be considered if heavy elements ($Z > 36$) are present in the molecule. Factors that limit the comparison between the experimental and calculated shieldings are the environmental effects and nuclear motion since most calculations are performed on isolated molecules at the equilibrium geometry.[25,36]

Chapter 4: Validation of a Computational Scheme to Study ^{15}N and ^{13}C Nuclear Shielding Constants

4.1 Introduction

A ^{13}C NMR investigation of the effects of electron-withdrawing groups (EWGs) on a series of *N*-substituted pyrroles showed a correlation between decreasing aromaticity and the strength of EWGs on the nitrogen atom.[8] Deshielding of the experimental ^{13}C NMR chemical shifts of *N*-substituted pyrroles was generally observed with increasing strength of *N*-EWGs.[8] These observations raised the question whether a similar correlation exists for the ^{15}N NMR chemical shifts since the nitrogen is directly attached to the EWG. To obtain a reliable interpretation of the ^{15}N and ^{13}C NMR chemical shifts for the pyrrole systems, quantum chemical calculations can be a useful tool for the prediction and interpretation of experimental chemical shift trends.

The determination of NMR shielding parameters is an active area of research in computational and theoretical chemistry.[45-54] The accuracy of calculated NMR parameters, such as shielding constants, spin-spin coupling constants and nuclear quadrupole coupling constants, using *ab initio* techniques has improved over the past few years.[49,51,55-58] However, the majority of NMR quantum mechanical calculations in the literature have focused on small molecules rather than on larger systems.

For small molecules containing relatively light atoms, satisfactory calculations of NMR shieldings can often be obtained. An exception is for systems in which electron correlation effects are important for a proper description of the ground and excited states. Electron correlation becomes significant in molecules in which there are low-lying excited states such as in molecules that have lone pairs and multiple bonds. For such systems, methods that include electron correlation effects are mandatory for obtaining reliable nuclear shieldings.[59] Conventional wavefunction-based electron correlation methods such as the configuration interaction (CI) and Møller-Plesset (MP) perturbation theory methods are only feasible for small systems. In recent years, the inexpensive density functional theory (DFT) approach has emerged as a promising alternative. It has been shown that DFT successfully predicts various molecular properties and often gives results of comparable, or better, quality than MP2 for a cost similar to Hartree-Fock (HF) calculations.[46] Investigation of the ability of DFT functionals to predict nuclear shielding constants is a very active area in the literature.

The principal difficulty with DFT for the prediction of nuclear shielding constants lies in the description of higher occupied valence orbitals and lower unoccupied orbitals.[48] Moreover, density functional results cannot fully capture the correlation effects since the functionals presently available have no current-density dependence.[54,60,61] New DFT functionals are being developed to improve the prediction of nuclear shielding constants. Several pure and hybrid functionals such as LSDA, BPW91, BLYP, B3PW91, and B3LYP, have been assessed for the prediction of nuclear shielding.[46] The pure and hybrid functionals of DFT with large basis sets yield shieldings that are more accurate than HF results.[46] The generalized gradient approximation (GGA)

DFT functionals produce accurate shieldings for molecules where electron correlation is significant.[46] The PBEPBE and PBE1PBE functionals yield significantly better results than the B3LYP functional for systems with huge correlation effects.[47] DFT results on nuclear shielding constants have been found to be too deshielded. This problem was rationalized in terms of the known deficiency of existing functionals in underestimating the energy differences.[48] It was shown that the functionals HCTH and asymptotically corrected HCTH(AC) functionals produce better predictions of nuclear shielding constants than the GGA functional BLYP, and hybrid functionals B97, B97-1, B3LYP and B3P86, since they improve the description of the low-lying valence excitations of conventional functionals, which tend to overestimate the energies of higher occupied orbitals.[48,62] Hybrid functionals $B3LYP^{0.05}_{GGA}$ and $B97^{0.05}_{GGA}$ that include 5% exact HF exchange using Kohn-Sham eigenvalues and orbitals[50] and the B98 functional by Becke[49] provide dramatic improvement over the conventional DFT functionals B3LYP, B97 and HCTH for the prediction of NMR parameters, showing an accuracy approaching coupled-cluster quality. The DFT functional B98 showed remarkable improvement with respect to B3LYP, especially for different hybridization states and its accuracy is competitive with the MP2 and PBE1PBE results for the prediction of nuclear shielding. Moreover, to overcome the deficiencies of the existing density functionals and the neglect of current-dependent terms, Malkin et al.[63] used the sum-over-states density-functional perturbation theory (SOS-DFPT), which introduces correction terms into the paramagnetic part of the shielding tensor that is usually overestimated in DFT calculations, to improve the agreement with experiment since the paramagnetic term will be reduced.

To validate a method for the calculation of ^{15}N and ^{13}C nuclear shielding in pyrroles, a systematic study was carried out for a selection of small nitrogen-containing molecules using the B3LYP, B3PW91 and PBE1PBE DFT functionals in conjunction with a set of selected basis sets.

4.2 Computational Details

The dependence of the ^{15}N and ^{13}C nuclear shielding constants on the basis sets and functionals was studied for a set of selected small molecules containing at least one N atom: CH_3CN, HCN, N_2, NH_3, CH_3NH_2, NNO, CH_3NO_2, $CH_3SO_2NH_2$, pyrrole, and *N*-methylpyrrole. Thus, the set includes nitrogen atoms in a range of hybridizations and bonding environments. All the DFT calculations were carried out using the Gaussian 98 suite of programs.[64] The geometries of the selected molecules were optimized with three different functionals: B3LYP, B3PW91, and PBE1PBE using the 6-31G(d) basis set. Subsequent nuclear shielding calculations using the gauge including atomic orbital (GIAO) method were performed for the three functionals using various basis sets at the 6-31G(d) geometries. The selected basis sets were 6-311+G(d,p), 6-311+G(2df,p), 6-311+G(3df,p), 6-311++G(d,p), 6-311++G(2df,p) and 6-311++G(3df,p). The importance of diffuse functions on the hydrogen atoms in the last three basis sets was investigated, since previous work has reported that they are necessary to obtain accurate ^{15}N chemical shifts of substituted 2-phenacylpyridines.[65]

4.3 Results and Discussion

4.3.1 Nitrogen Nuclear Shielding

Table 4.1 summarizes the calculated ^{15}N isotropic shielding for the small nitrogen-containing molecules with the basis sets and functionals used in this study. For N_2, upon increasing the basis set for each functional, a range of only 0.5 ppm or less is observed. For CH_3CN, HCN, CH_3NH_2, $CH_3SO_2NH_2$, and NH_3, expansion of the basis set changes the calculated ^{15}N isotropic shielding by 1.0 ppm or less. For the other molecules, CH_3NO_2, NNO, *N*-methylpyrrole, and pyrrole, the effect of expanding the basis set is as much as 4 ppm. Adding diffuse functions to the hydrogen basis sets does not appreciably improve the quality of the results with the maximum effect on the calculated ^{15}N isotropic shielding being only 0.4 ppm. Thus, the inclusion of the hydrogen diffuse functions is not as necessary as it was for the calculation of ^{15}N NMR chemical shifts of substituted 2-phenacylpyridines.[65]

The various basis sets show similar behavior for all functionals used. For $CH_3SO_2NH_2$ and the triply-bonded nitrogen-containing molecules N_2, HCN, and CH_3CN, the calculated ^{15}N shieldings decrease, then increase as the basis set is expanded from 6-311+G(2df,p) to 6-311+G(3df,p). An exception is observed for the B3LYP calculated ^{15}N shieldings for $CH_3SO_2NH_2$, where the ^{15}N shieldings increase as the basis set is expanded. For CH_3NO_2 and *N*-methylpyrrole, the calculated ^{15}N shieldings decrease as the basis set expands. For NH_3 and CH_3NH_2, the calculated ^{15}N shieldings increase as the basis set expands. For the NNO and pyrrole molecules, the calculated ^{15}N isotropic

shieldings of NNO and pyrrole increase from 6-311+G(d,p) to 6-311+G(2df,p), then decrease at the 6-311+G(3df,p) basis set.

For the PBE1PBE functional, the smallest basis set 6-311+G(d,p) gives the most accurate ^{15}N isotropic shielding for N_2, pyrrole, CH_3NO_2 and *N*-methylpyrrole. However, for B3PW91 and B3LYP, 6-311+G(2df,p) gives the most accurate results for pyrrole. The 6-311+G(2df,p) basis set gives the most accurate ^{15}N isotropic shielding for $CH_3SO_2NH_2$, NH_3 and NNO. The 6-311+G(2df,p) and 6-311+G(d,p) basis sets give almost identical shielding (with a difference of only 0.1 ppm) for the *N*-methylpyrrole. The largest basis set, 6-311+G(3df,p), gives the most accurate results for HCN, CH_3CN, and CH_3NH_2. A difference of only 0.2 ppm is obtained for CH_3NH_2 upon increasing the basis set from 6-311+G(2df,p) to 6-311+G(3df,p). A similar basis set dependence is observed for the B3LYP and B3PW91 functionals. Overall, the 6-311+G(2df,p) basis set gives the most reliable ^{15}N isotropic shielding.

Table 4.2 presents the ^{15}N isotropic shieldings obtained with the three functionals and the 6-311+G(2df,p) basis set. The mean absolute errors (MAE) and the root mean square errors (RMSE) are reported in Table 4.2. The mean absolute error (MAE) and root mean square error (RMSE) for the PBE1PBE functional with the 6-311+G(2df,p) basis set are the smallest (-9 and 12, respectively), while the B3LYP errors are the largest (-16 and 18, respectively). The MAE and RMSE for the B3PW91 (-11 and 13, respectively) differ from those of the PBE1PBE by only 2 and -1 ppm, respectively. The largest errors are obtained for the molecules that contain multiply-bonded nitrogen N_2, CH_3CN, CH_3NO_2, HCN and NNO. These molecules have negative shielding values since the paramagnetic shielding term dominates the total shielding. The large errors

Table 4.1 The calculated and experimental ^{15}N isotropic shieldings (in ppm) for nitrogen-containing molecules.

Molecule	Functional	6-311+G(d,p)	6-311++G(d,p)	6-311+G(2df,p)	6-311++G(2df,p)	6-311+G(3df,p)	6-311++G(3df,p)	exptl.
N_2	B3LYP	-88.5	-88.5	-88.7	-88.7	-88.6	-88.6	-61.6[49]
	B3PW91	-86.2	-86.2	-86.7	-86.7	-86.6	-86.6	
	PBE1PBE	-84.8	-84.8	-85.2	-85.2	-85.1	-85.1	
CH_3CN	B3LYP	-31.2	-31.2	-31.5	-31.5	-30.7	-30.8	-8.1[49]
	B3PW91	-29.4	-29.3	-29.8	-29.8	-29.0	-29.1	
	PBE1PBE	-27.7	-27.7	-28.2	-28.2	-27.4	-27.5	
CH_3NO_2	B3LYP	-157.3	-157.3	-159.0	-159.0	-160.5	-160.5	-135.8[66]
	B3PW91	-143.7	-143.7	-145.6	-145.7	-147.3	-147.3	
	PBE1PBE	-141.2	-141.3	-143.1	-143.2	-144.7	-144.7	
HCN	B3LYP	-41.8	-41.9	-42.3	-42.3	-41.2	-41.5	-20.4[49]
	B3PW91	-38.9	-39.1	-39.5	-39.6	-38.4	-38.7	
	PBE1PBE	-37.8	-37.9	-38.5	-38.5	-37.4	-37.7	
N<u>N</u>O	B3LYP	-7.6	-7.6	-6.6	-6.6	-7.8	-7.8	11.3[52]
	B3PW91	-0.9	-0.9	-0.2	-0.2	-1.3	-1.3	
	PBE1PBE	0.9	0.9	1.6	1.6	0.4	0.4	
<u>N</u>NO	B3LYP	78.8	78.8	81.0	81.0	81.0	81.0	99.5[52]
	B3PW91	84.8	84.8	86.9	86.9	86.7	86.7	
	PBE1PBE	87.9	87.9	90.0	90.0	89.8	89.8	
N-Me pyrrole	B3LYP	84.1	84.1	84.3	84.2	83.0	83.0	94.94[67]
	B3PW91	89.9	89.9	89.8	89.7	88.6	88.5	
	PBE1PBE	93.2	93.2	93.1	93.0	91.9	91.8	
CH_3NH_2	B3LYP	235.7	235.6	236.1	236.1	236.4	236.3	250.1[53]
	B3PW91	240.8	240.6	241.3	241.3	241.5	241.4	
	PBE1PBE	243.5	243.4	244.0	244.0	244.2	244.1	
$CH_3SO_2NH_2$	B3LYP	143.0	142.8	143.4	143.3	143.9	143.8	152.9[68]
	B3PW91	150.2	150.0	150.1	150.0	150.6	150.5	
	PBE1PBE	155.0	154.8	154.7	154.7	155.3	155.2	
NH_3	B3LYP	258.1	258.0	258.4	258.4	258.3	258.2	264.5[69]
	B3PW91	259.6	259.6	260.0	260.1	260.2	260.1	
	PBE1PBE	262.1	262.1	262.5	262.5	262.5	262.4	
Pyrrole	B3LYP	91.8	91.7	92.6	92.2	91.1	91.1	98.54[67]
	B3PW91	95.9	95.9	96.6	96.3	95.2	95.2	
	PBE1PBE	98.7	98.7	99.3	99.0	97.9	97.9	

might be due to the underestimation of energy differences which leads to an overestimation of the paramagnetic term.[48] This indicates that the multiply-bonded nitrogen-containing molecules require methods that yield a higher degree of electron correlation than the singly-bonded nitrogen-containing molecules. The errors are smaller for the singly-bonded nitrogen molecules, *N*-methylpyrrole, CH_3NH_2, $CH_3SO_2NH_2$, NH_3, and pyrrole, in which the shielding has a positive value since the diamagnetic shielding is larger than the paramagnetic shielding term. These results indicate that the PBE1PBE/6-311+G(2df,p) level of theory is the best choice for calculating the ^{15}N isotropic shielding for the molecules containing only a singly-bonded nitrogen atom, while the B3PW91/6-311+G(2df,p) method is almost as good.

Table 4.2 ^{15}N isotropic shieldings in ppm of small nitrogen-containing molecules obtained with the 6-311+G(2df,p) basis set.

	Isotropic shielding σ (ppm)			
Molecule	B3LYP	B3PW91	PBE1PBE	Experimental σ (ppm)
N_2	-88.7	-86.7	-85.2	-61.6
CH_3CN	-31.5	-29.8	-28.2	-8.1
CH_3NO_2	-159.0	-145.6	-143.1	-135.8
HCN	-42.3	-39.5	-38.5	-20.4
NNO	-6.6	-0.2	1.9	11.3
NNO	81.0	86.9	90.0	99.5
CH_3NH_2	236.1	241.3	244.0	250.1
N-methylpyrrole	84.3	89.8	93.1	94.94
$CH_3SO_2NH_2$	143.4	150.1	154.7	152.9
NH_3	258.4	260.0	262.5	264.5
pyrrole	92.6	96.6	99.3	98.54
Mean absolute error (MAE)	-16	-11	-9	
Root mean square error (RMSE)	18	13	12	

Comparing these results with a recent extensive study of ^{15}N chemical shifts using SOS-DFPT approach [58], the ^{15}N shielding obtained for pyrrole and *N*-methylpyrrole using the PBE1PBE/6-311+G(2d,f) level of theory gave better results (by 15 and 16 ppm) than those of the SOS-DFPT approach with the local approximation 3 (loc.3). The amine molecules, NH_3 and CH_3NH_2 gave very similar values to those of SOS-DFPT approach. However, for the multiply-bonded molecules CH_3NO_2 and CH_3CN, the SOS-DFPT approach gave much better results since this approach reduces the paramagnetic term that usually dominates the total shielding for these molecules and is overestimated in DFT calculations.

4.3.2 Carbon Nuclear Shielding

Table 4.3 summarizes the calculated ^{13}C isotropic shieldings at the various levels of theory used in this study. Expanding the basis set changes the ^{13}C shielding of $CH_3\underline{C}N$, $H\underline{C}N$, $\underline{C}H_3SO_2NH_2$, pyrrole (C1 and C2), and *N*-methylpyrrole (C1 and C2) by 1.0 ppm or less and by 1 to 2 ppm in the other molecules $\underline{C}H_3CN$, $\underline{C}H_3NO_2$, $\underline{C}H_3NH_2$, and the methyl carbon of *N*-methylpyrrole. The addition of the diffuse functions on the hydrogens affects the ^{13}C shielding results by only 0.4 ppm or less. Therefore, the inclusion of the hydrogen diffuse functions is not necessary for reproducing the ^{13}C isotropic shielding values.

The various basis sets show similar behavior for all functionals used. The calculated ^{13}C shieldings of the methyl group of $\underline{C}H_3CN$, $\underline{C}H_3NO_2$, $\underline{C}H_3NH_2$, and $\underline{C}H_3$ of *N*-methylpyrrole decrease with the expansion of the basis set. For the triply-bonded

carbon (to nitrogen) of H<u>C</u>N and CH_3<u>C</u>N, basis sets larger than 6-311+G(2df,p) give similar results that are appreciably better than those of 6-311+G(d,p). For the carbons on the pyrrole ring (C1 and C2 of pyrrole and *N*-methylpyrrole, the calculated shielding increases from the 6-311+G(d,p) to the 6-311+G(2df,p) basis set, and then decreases with the 6-311+G(3df,p) basis set.

With the PBE1PBE functional, the 6-311+G(d,p) basis set gives the most accurate ^{13}C isotropic shielding for <u>C</u>H_3CN, <u>C</u>H_3NH_2 and the <u>C</u>H_3 of *N*-methylpyrrole, whereas the 6-311+G(2df,p) basis set gives the most accurate results for HCN, $CH_3SO_2NH_2$, pyrrole (C1 and C2) and *N*-methylpyrrole (C1 and C2). The 6-311+G(3df,p) basis set gives the most accurate results for the ^{13}C shielding of CH_3<u>C</u>N, and CH_3NO_2. Similar patterns are observed for the other two functionals B3LYP and B3PW91. Overall, the 6-311+G(2df,p) basis set gives the most reliable ^{13}C isotropic shielding.

Table 4.4 presents the ^{13}C isotropic shieldings obtained with the three functionals and the 6-311+G(2df,p) basis set. The MAE and RMSE are reported in Table 4.4. The MAE and RMSE of the ^{13}C shielding values for the PBE1PBE functional are the smallest (-3 and 3 ppm, respectively), while the B3LYP errors are the largest (8 and -8 ppm, respectively). The MAE and RMSE for the B3PW91 (-4 and 5, respectively) differ from those of the PBE1PBE by only 1 and -2 ppm, respectively. For the PBE1PBE functional, the largest errors for the ^{13}C shieldings occur in the triply-bonded molecules CH_3<u>C</u>N and H<u>C</u>N, in which electron correlation is more demanding to describe. The errors are smaller for the singly- and doubly-bonded carbons. These results indicated that the PBE1PBE functional is the most accurate and the B3LYP is the least accurate for the determination of the carbon shielding values.

Table 4.3 The calculated and experimental ^{13}C isotropic shieldings (in ppm) for nitrogen-containing molecules.

Molecule	Functional	6-311+G(d,p)	6-311++G(d,p)	6-311+G(2df,p)	6-311++G(2df,p)	6-311+G(3df,p)	6-311++G(3df,p)	Exptl.
$\underline{C}H_3CN$	B3LYP	182.1	182.2	181.4	181.5	180.2	180.3	187.7[49]
	B3PW91	184.6	184.7	184.0	184.1	182.7	182.9	
	PBE1PBE	187.0	187.1	186.4	186.5	185.2	185.3	
$CH_3\underline{C}N$	B3LYP	63.8	63.7	64.4	64.3	64.4	64.4	73.8[49]
	B3PW91	66.1	66.0	66.5	66.4	66.6	66.5	
	PBE1PBE	67.3	67.2	67.6	67.5	67.7	67.7	
CH_3NO_2	B3LYP	119.3	119.5	118.6	118.6	117.8	117.8	125.2[70]
	B3PW91	123.8	124.0	123.0	123.1	122.3	122.3	
	PBE1PBE	126.6	126.8	125.9	126.0	125.1	125.1	
HCN	B3LYP	73.0	72.9	74.0	74.0	74.0	73.9	82.1[48]
	B3PW91	74.4	74.3	75.2	75.2	75.2	75.2	
	PBE1PBE	75.3	75.3	76.2	76.2	76.2	76.1	
$CH_3SO_2NH_2$	B3LYP	131.6	131.9	132.1	132.1	131.9	132.0	142.9[68]
	B3PW91	137.0	137.3	137.4	137.4	137.1	137.2	
	PBE1PBE	140.2	140.5	140.6	140.6	140.3	140.4	
CH_3NH_2	B3LYP	150.7	151.0	149.9	149.8	149.1	149.2	158.3[49]
	B3PW91	154.5	154.9	153.8	153.8	153.1	153.1	
	PBE1PBE	157.0	157.3	156.3	156.2	155.5	155.5	
Pyrrole, C1	B3LYP	62.1	62.1	62.4	62.2	61.3	61.3	69.1[70]
	B3PW91	66.6	66.6	66.8	66.7	65.8	65.8	
	PBE1PBE	68.9	68.8	69.0	68.9	67.9	67.9	
Pyrrole, C2	B3LYP	69.0	69.0	69.3	69.2	68.4	68.4	78.8[70]
	B3PW91	72.8	72.9	73.1	73.1	72.3	72.3	
	PBE1PBE	75.2	75.2	75.4	75.3	74.5	74.5	
N-Me pyrrole, C1	B3LYP	57.3	57.3	57.5	57.5	56.5	56.5	NA
	B3PW91	61.8	61.9	62.0	62.0	61.0	61.0	
	PBE1PBE	64.0	64.1	64.1	64.1	63.1	63.1	
N-Me pyrrole, C2	B3LYP	68.5	68.6	68.6	68.7	67.9	67.9	NA
	B3PW91	72.4	72.5	72.4	72.4	71.7	71.7	
	PBE1PBE	74.6	74.7	74.6	74.6	73.8	73.8	
N- Me pyrrole, $\underline{C}H_3$	B3LYP	146.3	146.4	145.6	145.5	144.6	144.6	NA
	B3PW91	149.7	149.8	149.0	148.9	148.1	148.1	
	PBE1PBE	152.0	152.1	151.3	151.2	150.3	150.3	

Table 4.4 ^{13}C isotropic shieldings in ppm of small nitrogen-containing molecules obtained with the 6-311+G(2df,p) basis set.

	Isotropic shielding σ (ppm)			
Molecule	B3LYP	B3PW91	PBE1PBE	Experimental σ (ppm)
$CH_3SO_2NH_2$	132.1	137.4	140.6	142.9
$CH_3\underline{C}N$	64.4	66.5	67.6	73.8
pyrrole, C2	69.3	73.1	75.4	78.8
HCN	74.0	75.2	76.2	82.1
CH_3NH_2	149.9	153.8	156.3	158.3
CH_3NO_2	118.6	123.0	125.9	125.2
pyrrole, C_1	62.4	66.8	69.0	69.1
$\underline{C}H_3CN$	181.4	184.0	186.4	187.7
N-methylpyrrole, C2	57.5	62.0	64.1	NA
N-methylpyrrole, C1	68.6	72.4	74.6	NA
N-methylpyrrole, $\underline{C}H_3$	145.6	149.0	151.3	NA
Mean absolute error (MAE)	-8	-4	-3	
Root mean square error (RMSE)	8	5	3	

4.3.3 Comparison of Nitrogen and Carbon Shieldings

The errors associated with the carbon shieldings are much smaller than those of the nitrogen shieldings. This indicates that recovery of electron correlation effects is more demanding for nitrogen than for carbon, due in part to the presence of lone pairs and the multiply-bonded environment on the nitrogen atom. Moreover, the errors in nitrogen shieldings are larger since the shielding strongly depends on the hybridization of the lone pair; minor inaccuracies in describing it can result in significant errors in the predicted shielding. Furthermore, the paramagnetic term might be overestimated since the DFT functionals cannot accurately describe the energy states of the low-lying

unoccupied molecular orbitals. Since the PBE1PBE functional gives smaller errors, it might indicate a better description of energy states than the other functionals.

This study shows that the 6-311+G(2df,p) basis set is the most efficient and accurate for the calculation of nitrogen and carbon shieldings. The addition of diffuse functions on the hydrogens does not improve the shieldings. The PBE1PBE functional gives the most accurate results for both carbon and nitrogen shieldings, while the B3LYP functional gives the least accurate values. The PBE1PBE functional provides only a small advantage (1 to 3 ppm) over the B3PW91.

4.4 Conclusions

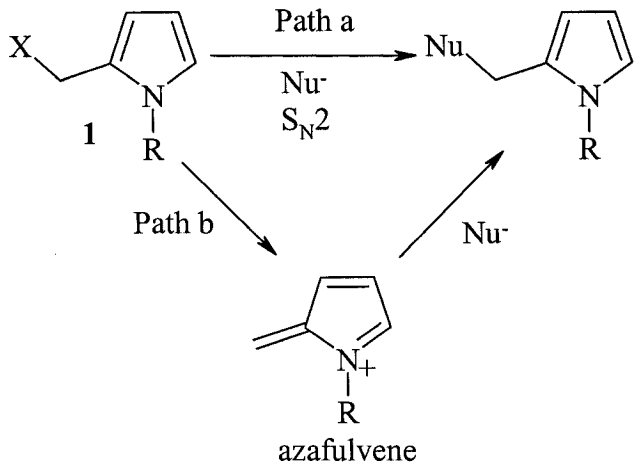
The performance of the B3LYP, B3PW91 and PBE1PBE functionals in conjunction with a selection of extended basis sets for the prediction of the ^{15}N and ^{13}C isotropic shieldings for small nitrogen-containing molecules has been assessed. On the basis of this study, the PBE1PBE/6-311+G(2df,p) level of theory is concluded to be the best choice for calculating the ^{15}N and ^{13}C isotropic shielding constants for the molecules containing only singly-bonded nitrogen and hence most efficient for the pyrrole systems, while the B3PW91/6-311+G(2df,p) level of theory is almost as good. The B3LYP functional is less appropriate for predicting ^{15}N and ^{13}C isotropic shieldings. All three functionals are less satisfactory for predicting the ^{15}N and ^{13}C NMR isotropic shieldings of triply-bonded nitrogen-containing molecules N_2, CH_3CN, and HCN. For these systems, the SOS-DFPT approach provides much more satisfactory results since it improves the estimation of the paramagnetic shielding.[58]

Chapter 5: The Effect of Electron-Withdrawing Groups on ^{15}N and ^{13}C Chemical Shifts: A Density Functional Study on a Series of Pyrroles

5.1 Introduction

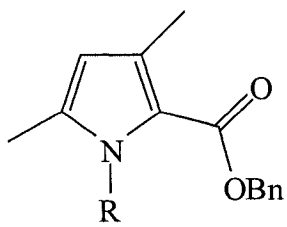
Introduction of an electron-withdrawing group (EWG) onto the nitrogen atom of a pyrrole decreases the aromaticity of the pyrrole-ring and reduces the reactivity of the pyrrole towards electrophilic attack.[6] The resulting deactivated pyrroles can react with dienophiles in Diels-Alder reactions[71-74] and can be readily reduced under Birch conditions.[75] Moreover the synthesis of 2-substituted pyrroles can be stabilized via α-lithiation[76] and pyrroles can undergo regio-controlled acylation.[76,77] These reactions of deactivated pyrroles provide elegant approaches for preparing pyrrole oligomers, terpyrroles, tropanes, alkaloids, and numerous other synthetic targets.[71-81]

Battersby et al.[6] have shown that nucleophilic substitution of pyrroles of type **1** (Scheme 5.1) under Mitsunobu conditions, which may proceed via both S_N2 and azafulvene mechanisms, can be controlled by introducing EWGs on the nitrogen atom of the pyrrolic ring. It was found that nucleophilic substitution is more likely to proceed via the S_N2 pathway when strong EWGs, such as triflyl (Tf), protect the nitrogen, since the triflyl group suppresses the azafulvene formation. The substitution mechanisms are determined by the nature of the transition state, the nature of the ground state, and the inherent activation energies of the competing mechanisms.[6,8]



Scheme 5.1 Mechanism of nucleophilic substitution of pyrroles of type **1**.

Thompson et al.[8] have investigated the nature of the ground state for model pyrroles (see **2**, Figure 5.1) by X-ray crystallography and ^{13}C NMR spectroscopy to elucidate some of the factors that influence the competition between the S_N2 and the azafulvene mechanisms. Analysis of the ^{13}C chemical shifts and X-ray crystal structures of *N*-substituted pyrroles **2** of synthetic interest showed a correlation between the decreased aromaticity and the strength of the EWGs on the nitrogen atom. Deshielding (shifts to lower field) of the ^{13}C chemical shifts is generally observed with the introduction of *N*-EWGs. It was concluded that the effects of EWGs on the likely mechanisms of pyrrole nucleophilic substitution reactions can be predicted using ^{13}C NMR spectroscopic data.

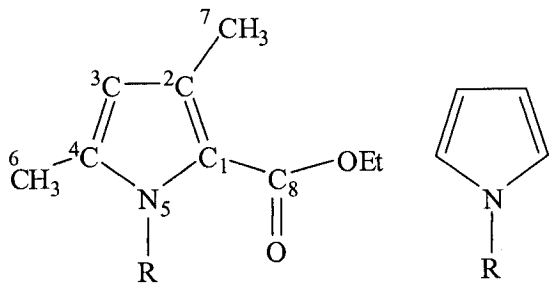


2

a R = H
b R = tert-butoxycarbonyl, COO^tBu (-Boc)
c R = methanesulfonyl, SO_2CH_3 (-Ms)
d R = trifluoromethanesulfonyl, SO_2CF_3 (-Tf)

Figure 5.1 *N*-Substituted pyrroles.

Experimental ^{13}C chemical shift data for analogs of *N*-substituted pyrroles **3**[a] with the four different substituents **3a-d** (Figure 5.2) show a similar correlation between the EWG strength on the nitrogen and the ^{13}C chemical shifts. For example, as the EWG strength increases for the substituents in the order[82] H, Boc, Ms and Tf (see Table 5.1), the ^{13}C chemical shifts for C1 range from 118.0 ppm for R=H to 125.6 ppm for R=Tf. For the extended pyrrole derivatives series **3** (Figure 5.2), the experimental ^{13}C chemical shifts show a general increase in the deshielding with an increase of the *N*-EWG strength (Table 5.1). The EWG strength, which is in agreement with the general trends indicated by Hammett values,[82] increases going down the R column in Table 5.1. The ^{13}C chemical shifts of C1, C2, C3 and C4 all show a general increase in chemical shift in going from **3a** down the column to **3d**.



3

4

a R = H
b R = tert-butoxycarbonyl, COO^tBu (-Boc)
c R = methanesulfonyl, SO_2CH_3 (-Ms)
d R = trifluoromethanesulfonyl, SO_2CF_3 (-Tf)
e R = methyl, CH_3 (-Me)
f R = benzyl, CH_2Ph (-Bn)
g R = acyl, $COCH_3$ (-Ac)

Figure 5.2 Model systems of the pyrrole series **3** and **4** and the numbering of the atoms.

[a] The experimental data for the pyrrole series **3a-g** was obtained from Dr. Alison Thompson at Dalhousie University in collaboration with Dr. Chad E. Hadden at Lilly Research Laboratories.

Table 5.1 Experimental ^{15}N and ^{13}C chemical shifts in ppm for pyrrole series **3a-g**.

	R	N5	C1	C2	C3	C4
3a	-H	155.9	118.0	129.2	111.5	132.9
3e	-Me	153.8	119.5	129.7	110.5	135.6
3f	-Bn	161.8	119.2	130.4	111.5	136.1
3g	-Ac	197.7	120.2	133.7	114.0	135.9
3b	-Boc	178.1	121.6	131.6	113.2	135.3
3c	-Ms	190.1	124.0	131.5	115.4	138.0
3d	-Tf	173.6	125.6	132.8	118.4	138.2

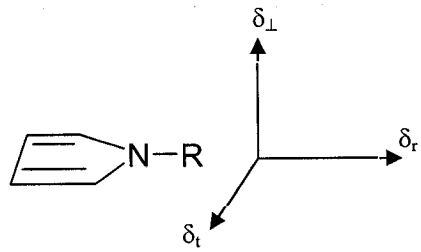
Since the nitrogen atom of the pyrrole is directly attached to the protecting group, it was anticipated that the ^{15}N chemical shifts would be more informative than the ^{13}C chemical shifts in predicting the effects of EWGs on the mechanism of pyrrole nucleophilic substitution reactions and the extent of aromatization. The experimental results in Table 5.1 show that the ^{15}N chemical shifts do not correlate with the increase of the strength of EWGs. To elucidate an interpretation of the experimental results in Table 5.1, DFT calculations were performed on the model systems of pyrrole series **3** and **4** (Figure 5.2) to examine the relationship between the electronic structures of pyrroles and their nitrogen and carbon chemical shifts. Possible conformations of each of the pyrroles were explored to determine the most stable conformation, which is likely the closest to the experimental conformation.[8] The nitrogen chemical shielding and chemical shifts were then calculated for the two pyrrole series **3** and **4** shown in Figure 5.2. The pyrrole series **4** was included in the study to see if the substituents on the carbons of the pyrrole ring have an effect on the ^{15}N chemical shift trend observed.

5.2 Computational Details

All the DFT calculations were carried out using the Gaussian 98 suite of programs.[64] The optimizations and the frequency analyses were carried out using the B3LYP/6-31G(d) level of theory. Relative energies, obtained by subsequent single point energy calculations performed at the B3LYP/6-311+G(2df,p) level using the above geometries, were corrected for zero-point vibrational energy (ZPVE), i.e., B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) + ZPVE. Several possible conformations of the pyrrole series **3a-g** were optimized to determine the most stable conformation, which is the one likely to be closest to the experimental conformation of pyrroles **2**.[8] In chapter 4, it was concluded that the PBE1PBE and B3PW91 functionals with the 6-311+G(2df,p) basis set in combination with the GIAO method are the best choices for calculating the NMR properties of the pyrroles.[83] Therefore, the nitrogen and carbon chemical shielding and chemical shifts of the most stable conformations of the pyrrole series **3a-g** were calculated at these two levels of theory using the B3LYP/6-31G(d) optimized geometries. The relative chemical shifts are always calculated by subtracting the chemical shielding from a reference. In this study, trimethylsilane (TMS) and liquid ammonia are used as the reference compounds to obtain the reference chemical shieldings for the carbon and nitrogen nuclei, respectively. The value of the absolute carbon shielding for TMS is 186.4 ppm and the value of the absolute nitrogen shielding for liquid ammonia is 244.6 ppm.

The NMR calculations using the GIAO method also yield other NMR parameters that aid in the interpretation of chemical shifts, such as the diamagnetic shift δ_d, the

paramagnetic shift δ_p, the anisotropic shift δ_{aniso}, and the shift tensor components (δ_{xx}, δ_{yy}, and δ_{zz}). The orientation of the principal shift tensor components with respect to the molecular frame is critical to the chemical interpretation of the shift tensor values. The spatial orientation of the shift tensor with respect to the molecular frame of the pyrrole systems is presented in Scheme 5.2. δ_{xx} represents the tangential direction of the pyrrole ring, δ_t; δ_{yy} represents the radial direction of the pyrrole ring, δ_r; and δ_{zz} represents the perpendicular direction of the pyrrole ring, $\delta_\perp$. The directions of both δ_t and δ_r are perpendicular to the nitrogen lone pair.



Scheme 5.2 Orientation of the principal components in the pyrrole heterocycle.

The electronegativities of the R groups in the pyrroles **3a-g**, which give a measure of the electron-withdrawing ability of each group, were calculated according to the method developed by Boyd and Boyd[84] using the bond critical point model and the atoms in molecules (AIM) method at the B3LYP/6-31G(d) level of theory. The natural chemical shielding (NCS) analysis, by which the nuclear magnetic shielding tensors computed by the GIAO method are partitioned into magnetic contributions from chemical bonds and lone pairs,[85] was used to deduce the origin of the trend observed for the nitrogen chemical shifts of the pyrrole series. The NCS analysis computes the

contributions to the shielding from well-localized "Lewis" bonds, lone pairs and delocalized "non-Lewis" features, such as bond conjugation and hyperconjugation.[85] The NCS analysis was carried out using the natural bond orbital (NBO) software.[86]

The calculated ^{15}N and ^{13}C chemical shift values for the pyrrole series **3** using the PBE1PBE and B3PW91/6-311+G(2df,p) level of theories agree to within about 1-4 ppm. Although the PBE1PBE functional results are slightly closer to the experimental values, the PBE1PBE functional is not available in the version of the G98 program to which the NBO 5.0 software is linked for the NCS analysis. Thus, the NCS analysis is only carried out at the B3PW91/6-311+G(2df,p) level of theory. Hence, only the results at B3PW91/6-311+G(2df,p) level are discussed in this chapter.

5.3 Experimental Details [b]

The pyrrole compounds **3a**[87], **3b**[88], **3c**[8], **3d**[8], **3e**[88], **3f**[88], and **3g**[89] were prepared according to literature methods. NMR data were acquired on a Varian INOVA 600 NMR spectrometer equipped with a Nalorac 3mm indirect detection probe. The samples were dissolved in 150 mL of deuterochloroform and transferred to a Wilmad 3 mm NMR tube. The samples were thoroughly characterized at 298 K using proton, two-dimensional gradient-selected correlation spectroscopy (gCOSY), heteronuclear direct gradient-selected heteronuclear single quantum coherence (gHSQC) and long-range gradient-selected heteronuclear multiple bond correlation (gHMBC) data. A ^{1}H-^{15}N gHMBC sequence was used to determine the chemical shifts of the nitrogens at natural abundance.

[b] The samples of the experimental data were prepared in the laboratory of Dr. Alison Thompson at Dalhouise University, Halifax, NS. The NMR data were obtained from Dr. Chad E. Hadden at Lilly Research Laboratories, Indiana, USA.

Both ^{1}H and ^{13}C data were referenced to TMS, and ^{15}N was referenced to liquid ammonia.

5.4 Results and Discussion

5.4.1 Electron-Withdrawing Ability of the Substituents

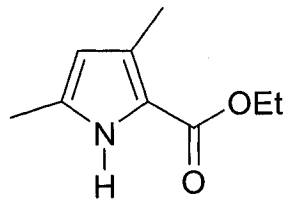
The electronegativities of the R groups give a measure of the electron-withdrawing ability of each group (Table 5.2). The order of the calculated electron-withdrawing ability of these groups is as follows: Me = Bn < Ac < Boc < Ms < Tf. The group electronegativity of the pyrrole ring **3** (3.14) is more electronegative than all of the R groups used in the study. Thus, all the substituents on the pyrrole ring are donating groups with respect to the pyrrole ring. The order of increasing EWG ability on the pyrrole ring then refers to the order of decreasing donating effects of the substituent.

Table 5.2 Calculated AIM group electronegativities of substituents.[84]

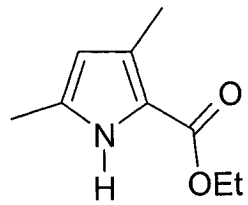
R group	Group electronegativity
-pyrrole ring **3**	3.14
-pyrrole ring **4**	3.13
-H	1.62
-Me	2.54
-Bn	2.54
-Ac	2.56
-Boc	2.63
-Ms	2.72
-Tf	2.73

5.4.2 Conformations of Pyrrole Series 3

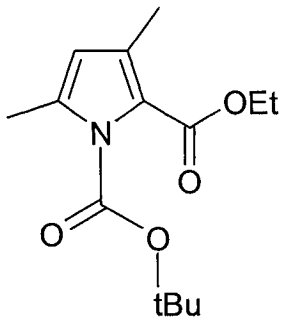
For pyrroles **3a**, **3c**, **3d**, **3e**, and **3f**, two conformations **I** and **II** were examined with different orientation about the ester group (Figure 5.3). For pyrroles **3b** and **3g**, four different conformations (**I**, **II**, **III**, and **IV**) were examined, where each conformation differs by its orientation of the ester group relative to the R group. The optimized geometries, bond lengths and Mulliken charges for pyrrole series **3** are presented in Figure 5.4. Conformation **I** of each pyrrole is the lowest in energy since this conformation gives the least steric interactions between the bulkiest substituents. The optimized conformations **I** of the pyrrole series **3** are very close to the conformations of X-ray crystal structures of analogous pyrroles **2**.[8] The bond length differences are about 0.02 Å and the bond angle differences are about 2°. The bond lengths and bond angles for the pyrroles with the sulfonyl group have larger differences of 0.07 Å and 6°, respectively.[8] The Mulliken charges reported on the nitrogen atom of the optimized conformations increases as the EWG strength increases, indicating less delocalization. The Mulliken charges, however, produce poor estimates of the charges since the method is based on an arbitrary scheme for assigning charges.



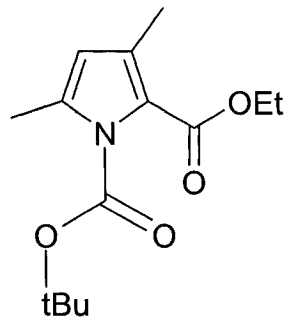
3a-I



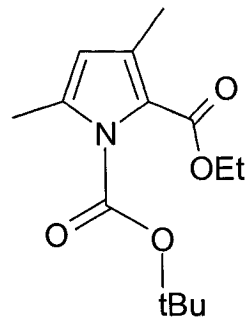
3a-II



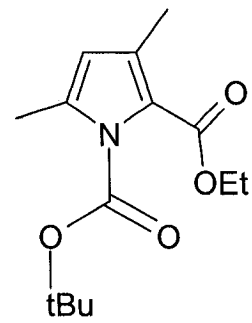
3b-I



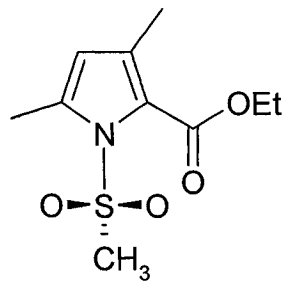
3b-II



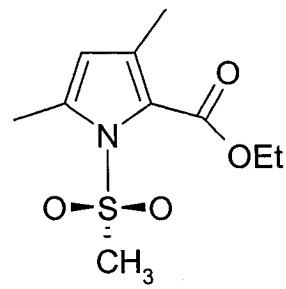
3b-III



3b-IV



3c-I



3c-II

Figure 5.3 Possible conformations for pyrrole series **3** (continued over page).

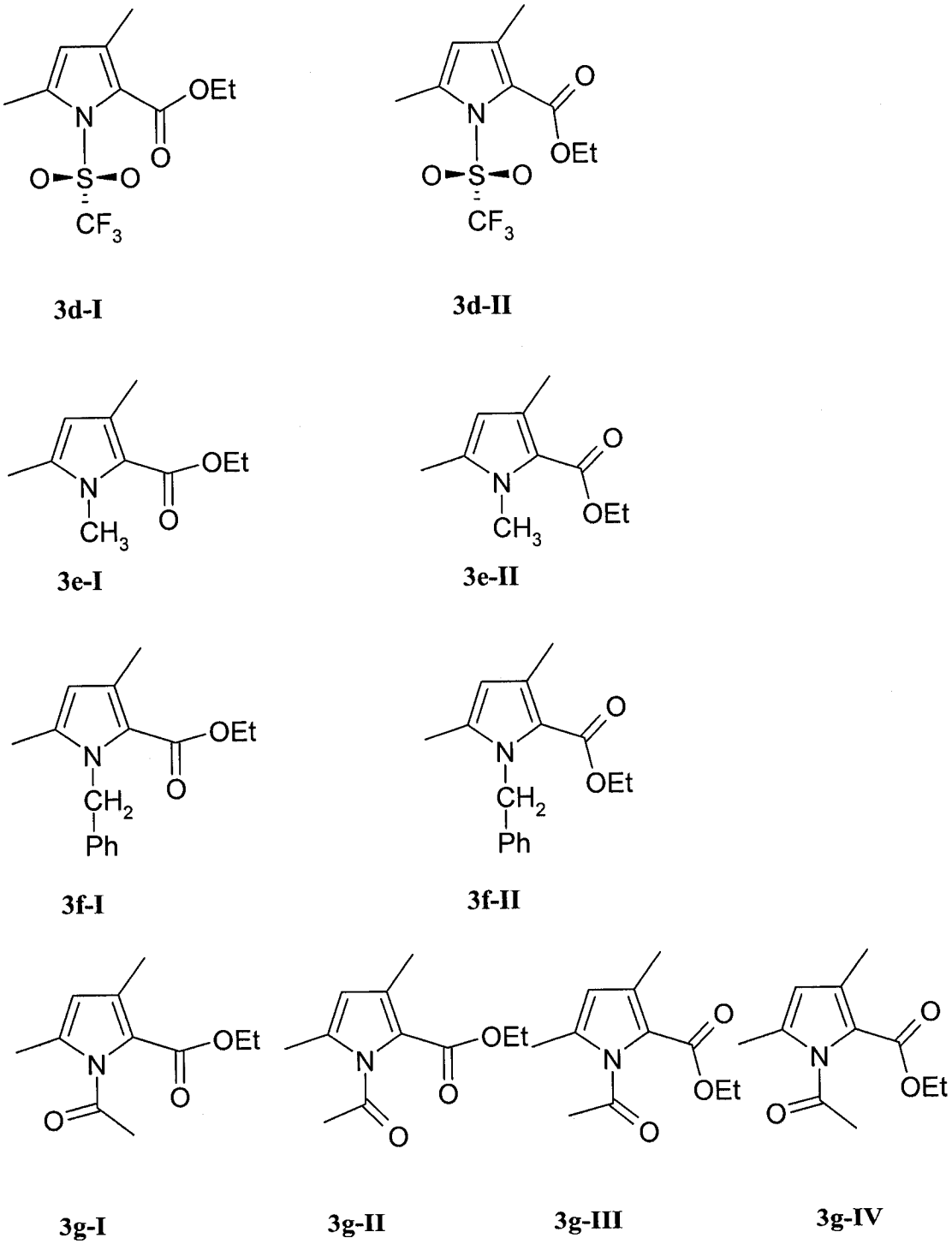
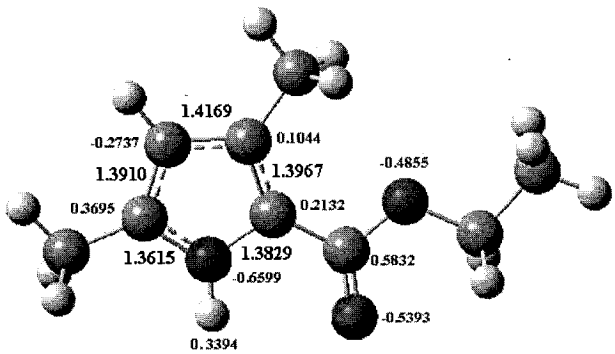
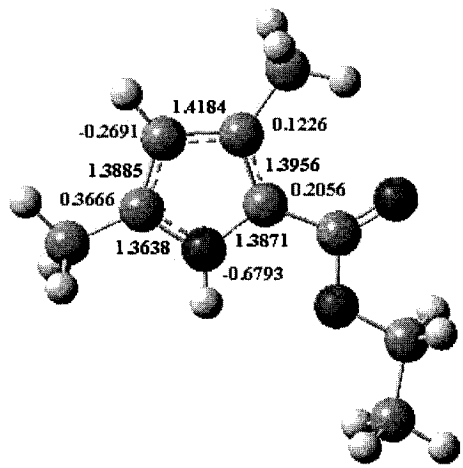


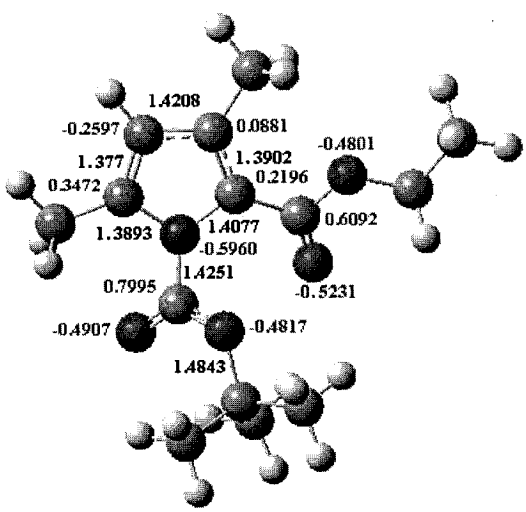
Figure 5.3 Possible conformations studied for pyrrole series **3**.



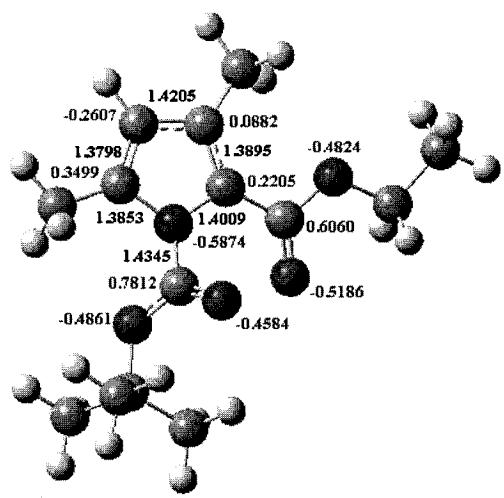
3a-I



3a-II

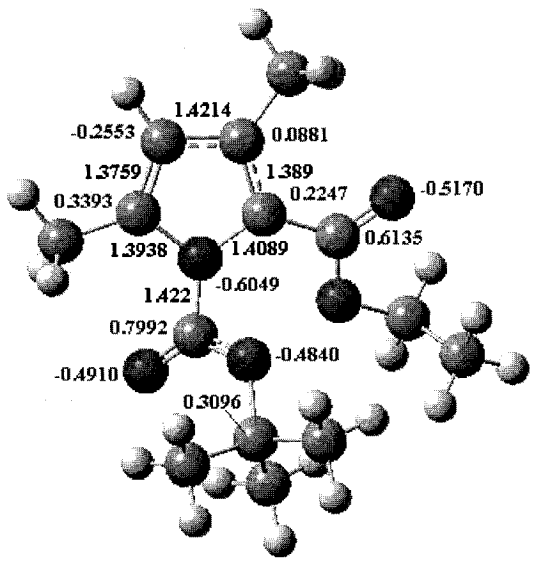


3b-I

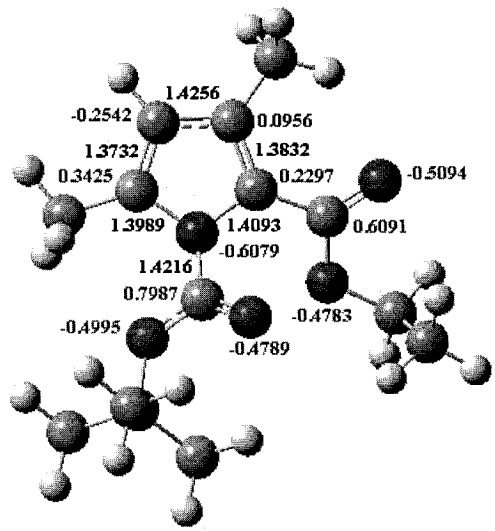


3b-II

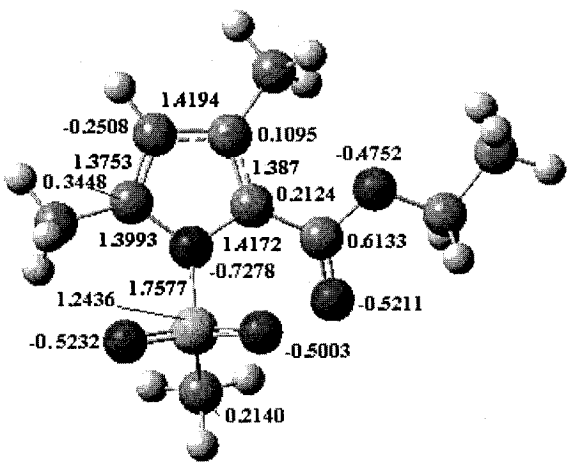
Figure 5.4 The optimized geometries, bond lengths and Mulliken charges of possible conformations of the pyrrole series **3** (continued over page).



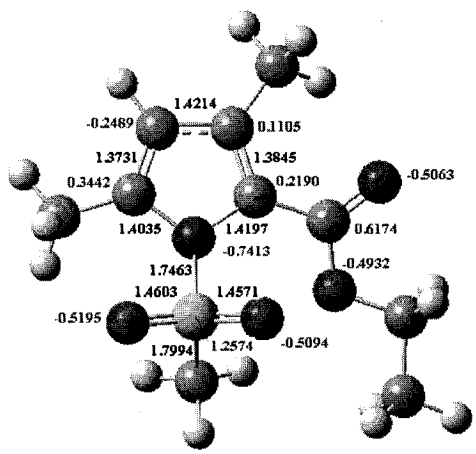
3b-III



3b-IV

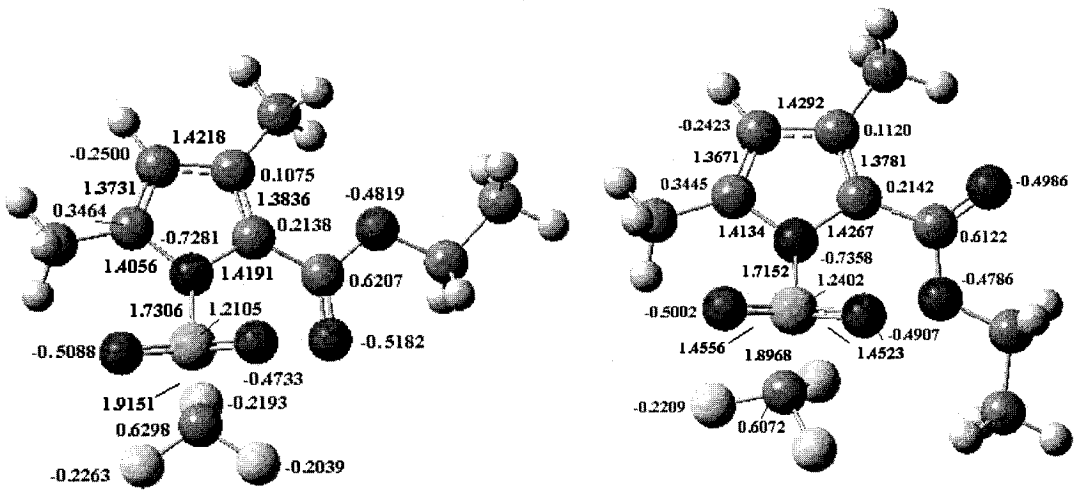


3c-I

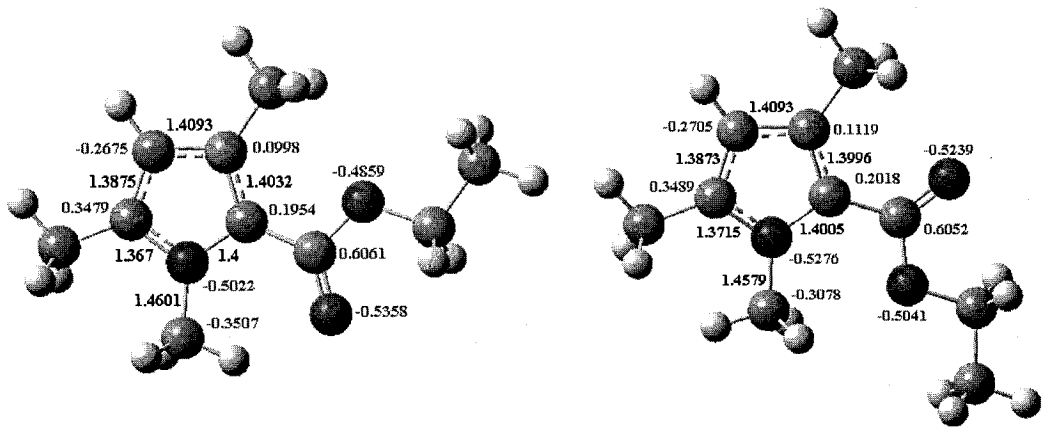


3c-II

Figure 5.4 The optimized geometries, bond lengths and Mulliken charges of possible conformations of the pyrrole series **3** (continued over page).

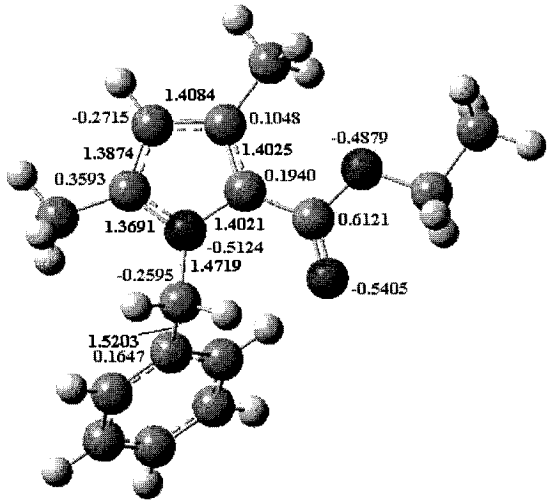


3d-I **3d-II**

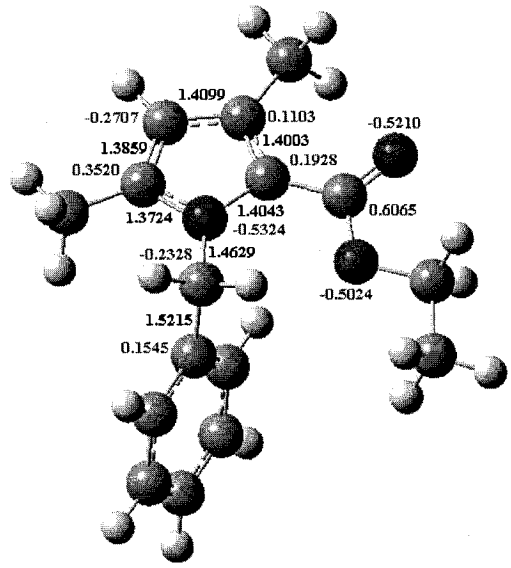


3e-I **3e-II**

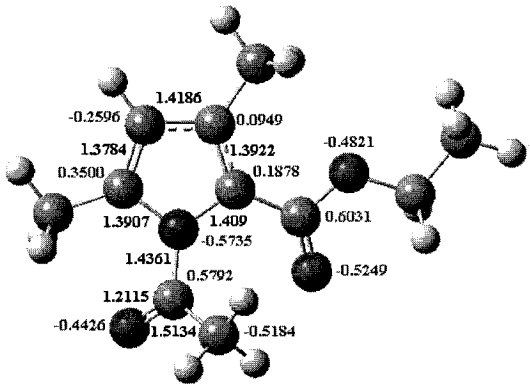
Figure 5.4 The optimized geometries, bond lengths and Mulliken charges of possible conformations of the pyrrole series **3** (continued over page).



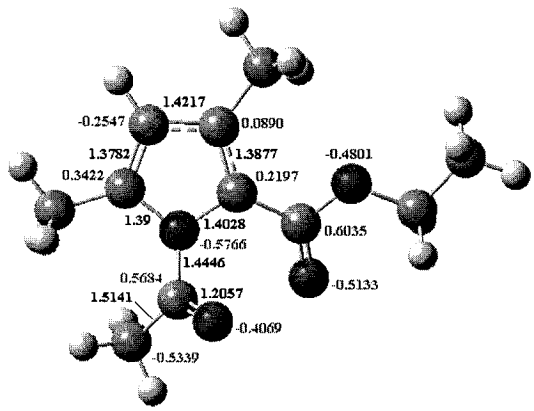
3f-I



3f-II

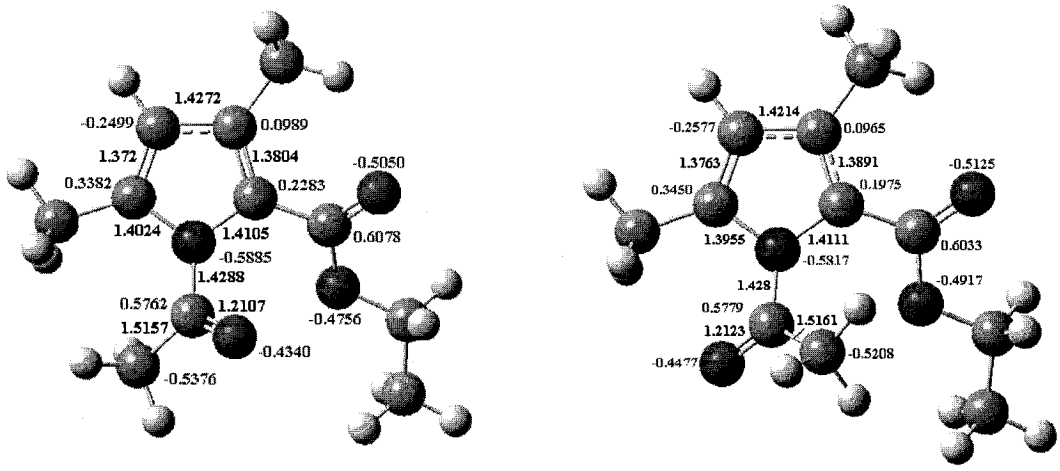


3g-I



3g-II

Figure 5.4 The optimized geometries, bond lengths and Mulliken charges of possible conformations of the pyrrole series **3** (continued over page).



3g-III **3g-IV**

Figure 5.4 The optimized geometries, bond lengths and Mulliken charges of possible conformations of the pyrrole series **3**.

5.4.3 ^{15}N NMR Chemical Shifts for Pyrrole Series 3

Table 5.3 presents the calculated nitrogen isotropic chemical shifts (δ_{iso}), diamagnetic shifts (δ_d), paramagnetic shifts (δ_p), and chemical shift tensors (δ_{xx}, δ_{yy}, and δ_{zz}) for all conformations of the pyrrole series **3** using the B3PW91/6-311+G(2df,p) level of theory. As can be seen in Table 5.3, the chemical shift values and the other parameters are sensitive to the conformational change of the pyrroles. For the substituents H and Me, a significant decrease in δ_{yy} is predicted in conformation **I** compared to conformation **II**. Also, a decrease in δ_{iso} and δ_p is predicted while the anisotropic shift is increased. Similar results are obtained for conformations **I** and **II** of the Bn, Ms, and Tf substituents. For the substituents Ac and Boc, the calculated NMR parameters are also sensitive to conformational changes.

Table 5.3 The B3PW91/6-311+G(2df,p) ^{15}N NMR parameters (in ppm) for pyrrole series **3**. Δε is the relative energy of the conformers.

R	Conformation	δ	δ_{aniso}	δ_d	δ_p	δ_{xx}	δ_{yy}	δ_{zz}	Δε kJ/mol
H	**I**	160.8	125.0	-81.4	486.9	142.3	259.2	81.1	0.0
	II	153.6	138.9	-75.1	473.3	140.3	237.4	83.1	3.5
	Exptl.	155.9	---	---	---	---	---	---	---
Me	**I**	165.4	115.4	-64.5	474.5	149.5	267.4	79.3	0.0
	II	157.6	121.6	-51.9	454.1	147.7	249.5	75.6	6.1
	Exptl.	153.8	---	---	---	---	---	---	---
Bn	**I**	180.1	100.4	-58.3	482.9	192.6	263.8	83.9	0.0
	II	168.7	112.2	-52.4	465.7	173.8	251.8	80.4	4.9
	Exptl.	161.8	---	---	---	---	---	---	---
Ac	**I**	215.3	94.7	-69.7	529.7	308.7	221.9	115.4	0.0
	II	211.7	99.7	-59.8	516.1	307.0	213.1	115.1	4.5
	III	212.9	116.1	-65.0	522.5	292.6	218.8	127.2	14.0
	IV	215.7	114.0	-76.7	537.0	279.0	239.4	128.6	16.4
	Exptl.	197.7	---	---	---	---	---	---	---
Boc	**I**	193.0	123.0	-61.8	499.4	210.0	257.0	111.9	0.0
	II	195.0	126.1	-72.9	512.4	224.6	244.3	116.0	7.1
	III	188.7	126.5	-52.8	486.0	206.2	249.9	109.9	10.6
	IV	189.9	137.3	-50.7	485.1	213.0	238.3	118.3	11.3
	Exptl.	178.1	---	---	---	---	---	---	---
Ms	**I**	219.6	135.8	-78.3	542.6	291.1	220.7	147.1	0.0
	II	214.5	148.2	-58.6	517.7	284.6	208.8	150.3	8.2
	Exptl.	190.1	---	---	---	---	---	---	---
Tf	**I**	208.2	155.7	-77.8	530.6	260.7	214.9	148.9	0.0
	II	200.9	170.9	-49.3	494.8	249.9	201.0	151.8	3.8
	Exptl.	173.6	---	---	---	---	---	---	---

Since conformations **I** are lowest in energy (Table 5.3) and closest to the X-ray structures of pyrroles **2**[8], the calculated data of conformations **I** are used for comparison

with the experimental values of the pyrrole series **3**. The calculated data for some of the other conformations **II**, **III**, **IV** produce chemical shifts that appear to be closer to the experimental values than those of conformations **I**. The closer agreement of the ^{15}N chemical shifts for conformations **II**, **III**, and **IV** to the experimental value is accidental since DFT methods overestimate the paramagnetic shielding term, due to the inaccurate description of excited states.[48] Therefore, the calculated ^{15}N chemical shift values of conformations **II**, **III**, and **IV** may deviate from their true values by the same magnitude as conformation **I** deviates from its true value.[48]

The NMR parameters for conformation **I** of each member of the pyrrole series **3** are plotted in Figure 5.5. As can be seen in Table 5.3 and Figure 5.5, the experimental nitrogen chemical shifts for pyrroles **3a-g** do not correlate with the increasing strength of the *N*-EWGs. Of particular note is the observation that the ^{15}N chemical shifts of pyrroles with Ac and Tf *N*-substituents are unexpectedly high and low, respectively. The experimental ^{15}N chemical shift for the *N*-Tf pyrrole **3d** is more shielded than that of *N*-Ms pyrrole **3c** and the experimental ^{15}N chemical shift of the *N*-Boc pyrrole **3b** is more shielded than that of *N*-Ac pyrrole **3g**. Importantly, the calculated nitrogen chemical shifts of the conformations **I** of the pyrroles reproduce the trend of the observed nitrogen chemical shifts. In particular, the calculated chemical shift of the *N*-Tf pyrrole **3d** is more shielded than that of the *N*-Ms pyrrole **3c**. All of the calculated chemical shift values are overestimated due to the inaccuracies of DFT functionals in describing the energy states that lead to overestimation of the paramagnetic shielding term.[48] The lack of correlation between the ^{15}N chemical shifts and the EWG strength for the small set of

substituents studied is not surprising since the ^{15}N chemical shifts depend highly on the molecular environment and the EWG substituents studied are chemically different.

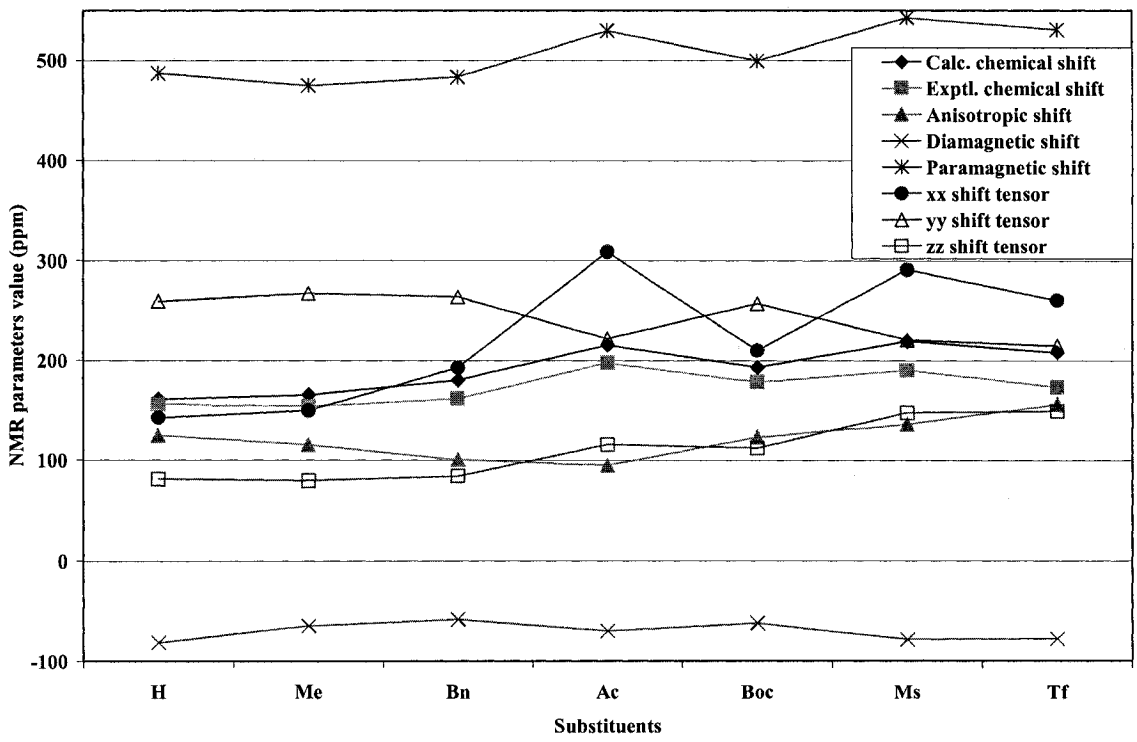


Figure 5.5 A graph of the B3PW91/ 6-311+G(2df,p) calculated NMR parameters and experimental values of conformations **I** of pyrrole series **3**.

The relative nitrogen chemical shift trend was reproduced theoretically. However, small differences in the nitrogen chemical shifts of pyrroles of less than 10 ppm could not be reproduced at the level of theory used. For example, the small differences of 2.1, 4.5 and 7.9 ppm between the experimental chemical shifts of *N*-H pyrrole **3a** and *N*-Me pyrrole **3e,** the *N*-Boc pyrrole **3b** and *N*-Tf pyrrole **3d,** and *N*-Ac pyrrole **3g** and *N*-Ms pyrrole **3c**, respectively, are not reproduced computationally. The

relative trend of the calculated chemical shifts of each pair of the above pyrroles is opposite to that of the experimental trend. This might be due to the overestimation of the paramagnetic shielding in the calculations and the inaccurate geometries, since the deviation of the calculated chemical shifts from the experimental values becomes larger as the R substituents become larger. The chemical shift errors of pyrrole series **3** with the substituents H, Me, Boc, Ac, Bn, Ms and Tf are 5, 12, 15, 18, 18, 30 and 35 ppm, respectively.

The chemical shifts are calculated in the gas phase on an isolated geometry, i.e., the medium effects of solute-solvent interactions, van der Waals effects and nuclear motion were ignored. However, the experimental NMR chemical shift values are a weighted average of the different conformers present in solution at room temperature, and solvent effects may stabilize some of the higher energy conformers. Hence, the calculation of chemical shift values based on a single conformation in the gas phase limits the accuracy of the results. The presence of lone pairs of electrons renders nitrogen shielding susceptible to solute-solvent interactions.[90] Moreover, minor inaccuracies in describing the hybridization of the lone pair can result in significant errors in shielding values.[91] Furthermore, the paramagnetic term might be overestimated since the DFT functionals cannot accurately describe the energy states of the low-lying molecular orbitals.[48] Therefore, only qualitative results for the pyrrole series can be obtained; however, they suffice to provide insight into experimental trends.

In order to understand the origin of the observed nitrogen chemical shift trend, the trend of each of the three nitrogen shift tensor components (δ_{xx}, δ_{yy} and δ_{zz}) of pyrrole series **3** is analyzed to identify the variation in the tensor components that correlates with

the trend in the calculated isotropic nitrogen chemical shift values. The trend for δ_{xx} of the pyrrole series **3** correlates with the trend of the calculated chemical shift, i.e. as the calculated chemical shift increases, δ_{xx} increases and (Figure 5.5). The δ_{xx} component for the Tf-substituent **3d** (260.7 ppm) is more shielded than that for the Ms-substituent **3c** (291.1 ppm) and the δ_{xx} component for the Ac-substituent **3g** (308.7 ppm) is less shielded than that for the Boc-substituent **3b** (210.0 ppm), producing a similar trend to the calculated chemical shifts. The δ_{xx} components are the closest in value (Figure 5.5) to the calculated isotropic chemical shift values except for the Ac, Ms and Tf substituents, which have the largest variations from the experimental chemical shifts. The δ_{yy} values of pyrrole series **3** correlate inversely with the calculated chemical shift trend of the pyrrole series **3**; as the chemical shift is increased, the δ_{yy} component decreases with the exception of the Tf substituent **3d** (Figure 5.5). The δ_{zz} values do not correlate with the calculated chemical shift values. However, the trend observed for the δ_{zz} values (as the EWG strength increases) is interesting in that the δ_{zz} values can be divided into three different groups based on a small difference (less than 5 ppm) of the δ_{zz} values and the similarity of the *N*-substituent group (Figure 5.5). The δ_{zz} values for the substituents: H, Me, and Bn are very close, with a range of 79.3-83.9 ppm. The range for the Ac and Boc substituents is higher, 111.9-115.4 ppm, and the range for the Ms and Tf substituents is the highest, 147.1-148.9 ppm. Thus, *N*-substituents of analogous structure have approximately similar δ_{zz} values and therefore, the origin for the observed nitrogen chemical shift trend originates from the δ_{xx} and δ_{yy} components.

The diamagnetic shifts δ_d for the pyrrole series **3** do not correlate with the observed chemical shifts trend (Figure 5.5). The δ_d for the Me, Bn, Ac, and Boc

substituents have similar values (-64.5, -58.3, -69.7 and -61.8 ppm, respectively), indicating that the electron distribution in the ground electronic states of these pyrroles is similar. The δ_d values for the H, Ms, and Tf substituents (-81.4, -78.4 and -77.8ppm, respectively) are very close to each other, indicating similar electron distributions on the nitrogen atom in their electronic ground states (Table 5.3).

The trend of the nitrogen paramagnetic shifts δ_p for the pyrrole series **3** is similar to that of the calculated chemical shift trend (Figure 5.5). The correlation between the chemical shift and the δ_p trends indicates that the nitrogen chemical shift trend for the pyrroles arises entirely from variations of the paramagnetic shielding contribution.

The calculated nitrogen anisotropic shift of the pyrrole series **3**, with the exception of Ms-substituted pyrrole, inversely maps the observed trend for the chemical shift. In other words, as the chemical shift increases, the anisotropic shift decreases (Figure 5.5). The exception of the Ms-substituted pyrrole to this observation may be related to the larger error of its calculated nitrogen chemical shift compared to that of the Boc-substituent.

In addition, the present study indicates that there is a correlation between the calculated nitrogen chemical shifts and the N-R bond length. Table 5.4 presents the N-R bond lengths for conformations **I** of pyrrole series **3**. The N-R bond length correlates with the chemical shift trend for pyrroles with *N*-substituents of similar structure. The N-C(alkyl) bond lengths increasing from Me to Bn correlates with an increase in their corresponding chemical shift value. A decrease in the N-C(carbonyl) bond length from Ac to Boc correlates with a decrease in their corresponding chemical shift values. The decrease in the N-S bond length from Ms to Tf correlates with the observed decrease in

the corresponding chemical shift values. The change of the calculated chemical shifts with the bond length arises from the paramagnetic shielding term.[90] An increase in the bond length is accompanied by a decrease in the relevant electronic excitation energies which thus furnishes an enhanced paramagnetic contribution to the shielding tensor.[90]

Table 5.4 The calculated bond length and ^{15}N chemical shift for conformations **I** of pyrrole series **3**.

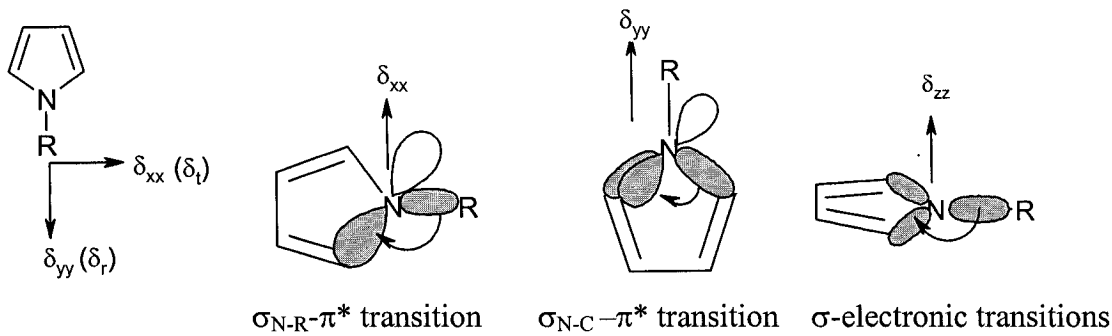
R	N-R*	Bond length (Å)	^{15}N chemical Shift (ppm)
H	N-H	1.0107	160.8
Me	N-C	1.4601	165.4
Bn	N-C	1.4719	180.1
Ac	N-C	1.4361	215.3
Boc	N-C	1.4251	193.0
Ms	N-S	1.7577	219.6
Tf	N-S	1.7306	208.2

5.4.4 NCS Analysis of the ^{15}N Shielding for Pyrrole Series 3

It has been reported in the literature that the chemical shift trends for nitrogen nuclei in different environments arise entirely from variations of the paramagnetic shift contribution.[92] The diamagnetic shift contribution has been observed to be approximately constant.[92] Thus, dramatic changes in the nitrogen shift tensors result from the paramagnetic contribution term upon the changes of the nitrogen hybridization. These changes can be rationalized as follows. The largest shift component, δ_{11}, generally orients perpendicularly to the plane defined by the bonding-antibonding, or HOMO-LUMO, pair of orbitals with the smallest energy gap.[93] The orientation is a consequence of mixing the occupied and unoccupied orbitals by the angular momentum operator, which is specified

in Ramsey's expression.[30] The paramagnetic term in Ramsey's expression is inversely proportional to the energy gap between the pair of orbitals. The term with the smallest energy gap makes the largest paramagnetic contribution. The largest chemical shift component is oriented perpendicularly to the two orbitals involved in the paramagnetic contribution.[66]

A pictorial representation of the orbitals that make the largest contributions[66] to the nitrogen shift tensor components is given in Scheme 5.3. The tangential component, δ_t, is dominated by the $\sigma_{N\text{-}R} - \pi^*$ transitions for the nitrogen nuclei of the pyrrole. The radial component, δ_r, is dominated by the $\sigma_{N\text{-}C} - \pi^*$ transitions for the nitrogen nuclei of the pyrrole. The perpendicular component to the pyrrole ring, $\delta_\perp$, is dominated by the σ-electronic structure. The $\delta_\perp$ value makes up the smallest component since the energy gaps of the orbitals involved are the largest.[66]



Scheme 5.3 Pictorial representations of the electronic transitions making the largest contribution to the chemical shift tensor components in pyrroles.

Table 5.5 and Figure 5.6 report the NCS analysis for conformations **I** of pyrrole series **3**. The major contributions to the shielding from localized bonds, lone pairs and

delocalized features are included. The NCS analysis is time consuming. Due to computer hardware limitations, it was impossible to obtain the NCS analysis for Boc- and Bn-substituted pyrroles. Thus, only a few substituents are included in Table 5.5 and Figure 5.6. The contribution of 1s (N5) is constant for the different EWG pyrroles since the core contribution to the chemical shift is independent of the chemical environment.[94] The σ(N5-R) total contribution correlates with the shielding trend, where the σ(N5-R) corresponds to the shielding contribution in the tangential direction with respect to the nitrogen position (see Scheme 5.2). The tangential direction is the same as the σ_{xx} which correlates with the total shielding, thus the change in the σ(N5-R)-π* transition accounts for the shielding trend observed in the pyrrole series **3**. The shielding does not correlate with σ(C1-N5) and σ(C4-N5). However, the observed shielding correlates with the sum of σ(C1-N5) and σ(C4-N5), i.e. as the shielding increases, the sum of σ(C1-N5) and σ(C4-N5) decreases. The sum of these contributions corresponds to the shielding contribution in the radial direction, σ_{yy} (see Scheme 5.2). The σ_{yy} correlates with the total nitrogen shielding except for the Tf substituent, thus the change in the sum of the σ(C1-N5)- π* and σ(C4-N5)- π* transitions also accounts for the shielding trend observed in the pyrrole series **3**. The total shielding also correlates with the σ(C3-C4) and σ(C4-C6) contributions, i.e. as the shielding increases the σ(C3-C4) and σ(C4-C6) contributions decrease. However, an exception for the correlation between total shielding and σ(C4-C6) exists when comparing the Ac (-0.52 ppm) and Ms (-3.86) substituent contributions. The observed shielding does not correlate with the n_{π}(N5) contribution. Instead, the n_{π}(N5) contribution decreases with increasing strength of the EWG (decreasing aromaticity) (Figure 5.6 and Table 5.5). The n_{π}(O26) and n_{π}(O27)

contributions add to the shielding for **3b**, **3c** and **3d** (Table 5.5). The total Lewis shielding contribution correlates with the total shielding with the exception of the H substituent, which is the most delocalized system in the pyrrole series **3** (Figure 5.6). The non-Lewis shielding contribution correlates with the total shielding, i.e. as the total shielding increases, the non-Lewis shielding contribution increases, with the exception of comparing the non-Lewis shielding contribution of Tf with the Ms compared to the total shielding; in other words, the non-Lewis contribution decreases from the Ms to the Tf pyrrole whereas the total shielding increases.

Table 5.5 Calculated NCS orbital decomposition for ^{15}N isotropic shieldings in ppm for the pyrrole series **3** at the B3PW91/6-311+G(2df,p) level of theory.

	$\sigma_{iso}(N)$				
Orbital / R	H	Me	Ac	Ms	Tf
σ(C1-N5)	-76.3	-55.9	-63.8	-56.7	-80.5
σ(C4-N5)	-111.4	-107.6	-93.6	-40.3	-19.6
σ(N5-R)	25.8	-12.8	-34.2	-91.0	-62.1
σ(C3-C4)	-11.8	-9.0	-5.1	-4.8	-6.0
σ(C4- C6)	-15.1	-7.3	-0.5	-3.9	-5.0
1s (N5)	239.8	239.7	239.7	239.7	239.7
n_π (N5)	41.4	39.3	30.4	20.4	1.0
n_π (O26)	---	---	-18.8	-8.5	-11.7
n_π (O27)	---	---	---	-9.9	-9.2
Sum[σ(C1-N5), σ(C4-N5)]	-187.7	-163.6	-157.5	-97.0	-100.1
Lewis	60.2	72.2	51.7	34.9	49.1
Non-Lewis	35.2	15.3	-17.8	-9.8	-16.3
Total shielding	95.5	87.5	33.8	25.1	32.8

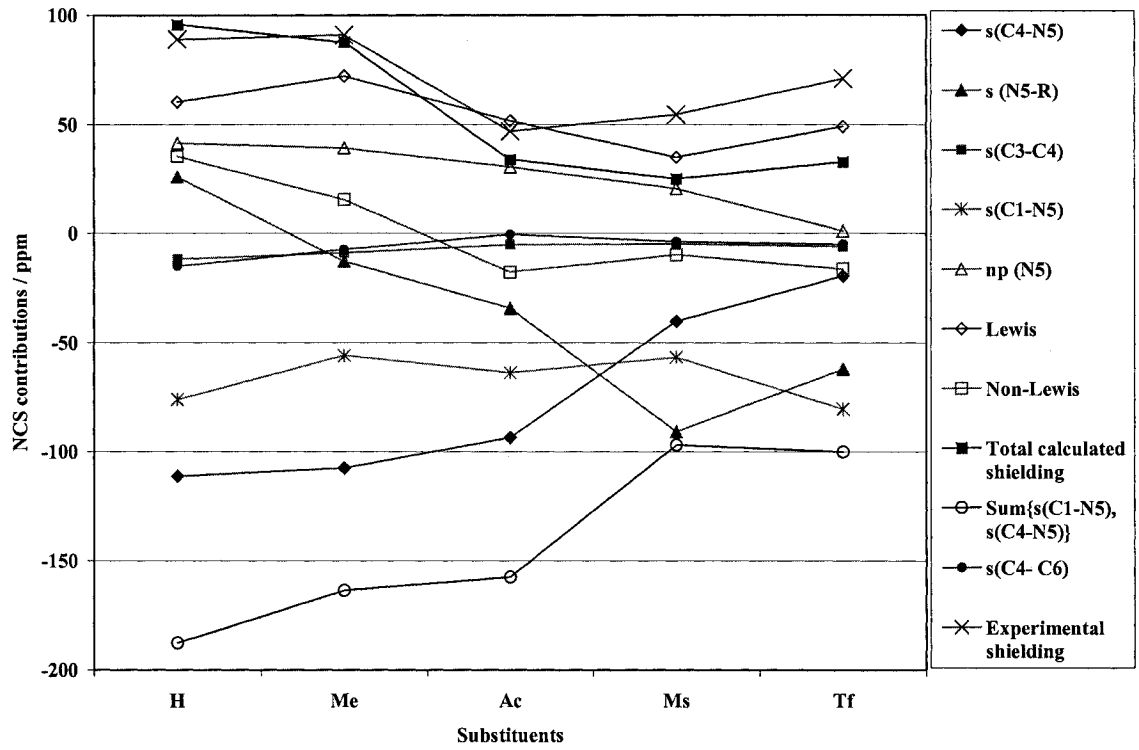


Figure 5.6 A graph of the various NCS contributions for conformations **I** of the pyrrole series **3** versus the substituents. The s in the graph label stands for σ, and np stands for n_{π}.

5.4.5 ^{13}C Chemical Shifts for Conformations I of the Pyrrole Series 3

Table 5.6 presents the ^{13}C NMR chemical shift data for conformations **I** of the pyrrole series **3** at the B3PW91/6-311+G(2df,p) level of theory. Figure 5.7 presents a graph of the calculated and experimental chemical shifts of conformations **I** of the pyrrole carbons. The increase of the experimental ^{13}C chemical shifts for the pyrroles with the substituents: H, Boc, Ms, and Tf correlates with the increase of the electron-withdrawing strength of these substituents determined by the AIM calculations. The calculated ^{13}C chemical shifts corresponding to pyrroles with these substituents reproduce the same correlation. When other substituents (Me, Bn, and Ac) are included, the

observed C1 ^{13}C chemical shifts show a trend that correlates with the trend of the electron-withdrawing strength of the substituents determined by the AIM calculations, while the observed C2, C3 and C4 ^{13}C chemical shifts generally show an increase in the deshielding effect with the increase of the *N*-EWG strength. The calculated C1 ^{13}C chemical shift for the H substituent does not follow the trend of the experimental values since hydrogen bonding effects are ignored, thus resulting in large errors for the chemical shift.[90] The calculated C2 and C3 ^{13}C chemical shifts reproduce the trend of experimental values very well, while the calculated C4 chemical shifts reproduce the experimental trend to a lesser degree due to deterioration caused by the inclusion of the *N*-Ac pyrrole.

Table 5.6 The B3PW91/6-311+G(2df,p) calculated ^{13}C NMR chemical shifts in ppm and experimental values for the pyrrole series **3**.

		^{13}C chemical shifts			
R	Method	C1	C2	C3	C4
H	B3PW91	125.4	134.4	117.5	136.6
	Experimental	118.0	129.2	111.5	132.9
Me	B3PW91	121.7	138.4	116.5	140.9
	Experimental	119.5	129.7	110.5	135.6
Bn	B3PW91	124.9	138.7	117.1	143.1
	Experimental	119.2	130.4	111.5	136.1
Ac	B3PW91	126.8	142.0	118.9	146.2
	Experimental	120.2	133.7	114	135.9
Boc	B3PW91	129.6	137.6	117.5	141.3
	Experimental	121.6	131.6	113.2	135.3
Ms	B3PW91	130.1	140.3	120.6	146.8
	Experimental	124.0	131.5	115.4	138.0
Tf	B3PW91	130.7	141.4	120.7	148.9
	Experimental	125.6	132.8	118.4	138.2

From both experimental and calculated results of the carbon chemical shifts for the pyrrole series, it can be concluded that a general correlation between the ^{13}C chemical shifts and the EWG strength does exist and it was reproduced theoretically. For these pyrrole systems **3a-g**, the use of the isotropic chemical shift instead of its three tensor components loses valuable information about the relationship between the geometry and the ^{13}C chemical shift, since in π systems the anisotropy of shielding is strong. Thus, the use of the tensor components of the ^{13}C chemical shifts gives a better insight into the relationship between geometry and the origin of the ^{13}C chemical shifts.

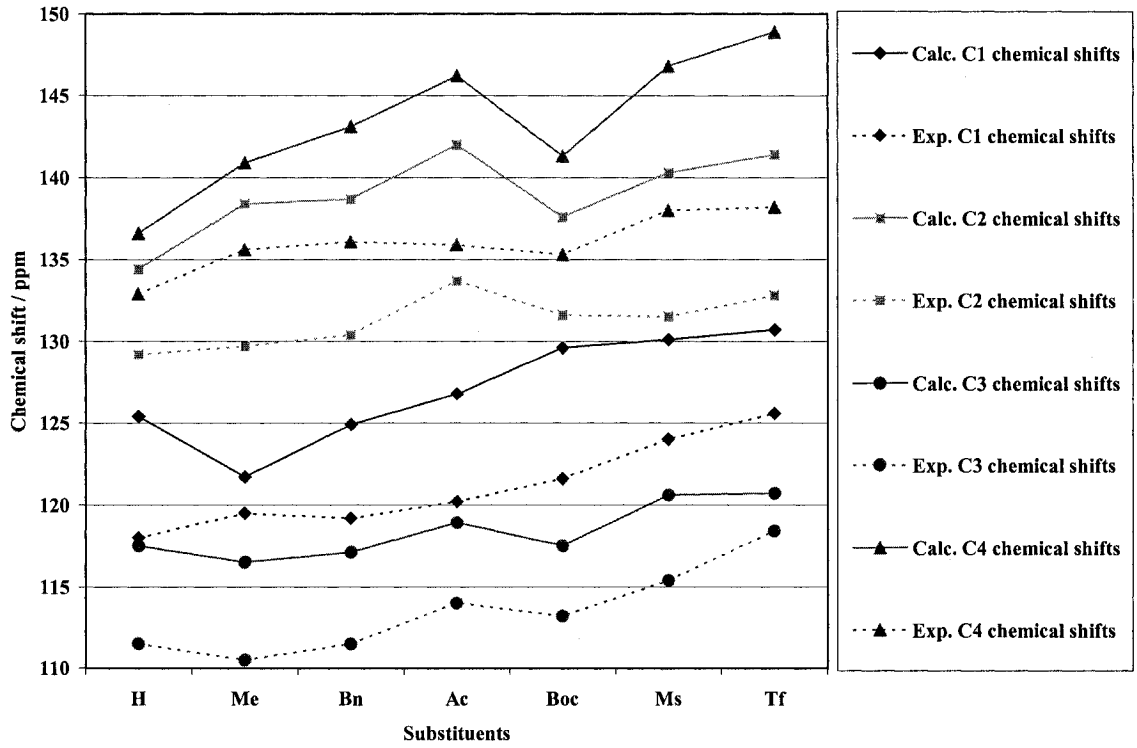


Figure 5.7 A graph of the calculated and experimental chemical shifts of conformations **I** of the pyrrole carbons of pyrrole series **3**. The calculated values were obtained at the B3PW91/6-311+G(2df,p) level of theory. The dotted lines represent the experimental values.

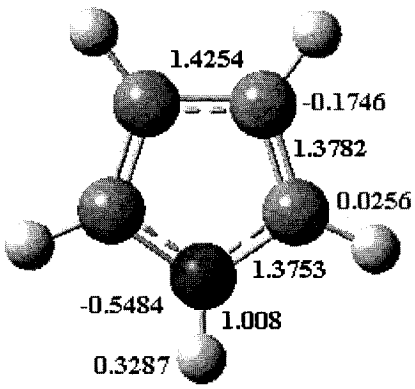
5.4.6 Optimized Geometries for Pyrrole Series 4

The optimized geometries of the pyrrole series **4** are reported in Figure 5.8. The pyrrole molecule itself has a C_{2v} symmetry and the other *N*-substituted pyrroles have C_s symmetry with the exception of the *N*-CH_2Ph pyrrole with C1 symmetry. A planar starting geometry was assumed for all pyrroles.

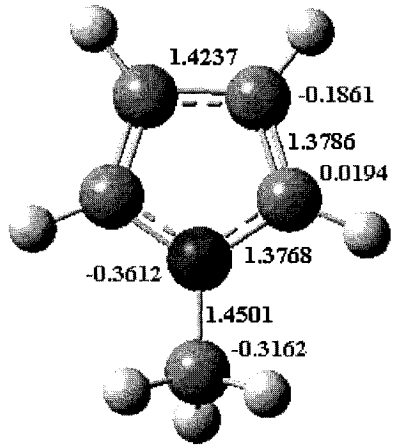
The bond length changes in the pyrrole rings of pyrrole series **4** (Figure 5.8) give an insight to the electron-withdrawing ability of the *N*-substituents. As the strength of the EWG increases, the availability of the nitrogen lone pair decreases and aromaticity decreases. Consequently, longer N5-C1 and N5-C4 bonds result, as can be observed in Figure 5.8. The S-N bond is significantly shorter in the *N*-Tf pyrrole than in the *N*-Ms pyrrole. The stronger N-S bond in the *N*-Tf pyrrole is indicative of the stronger electron-withdrawing ability of the Tf substituent over the Ms substituent.

5.4.7 ^{15}N Chemical Shifts for Pyrrole Series 4

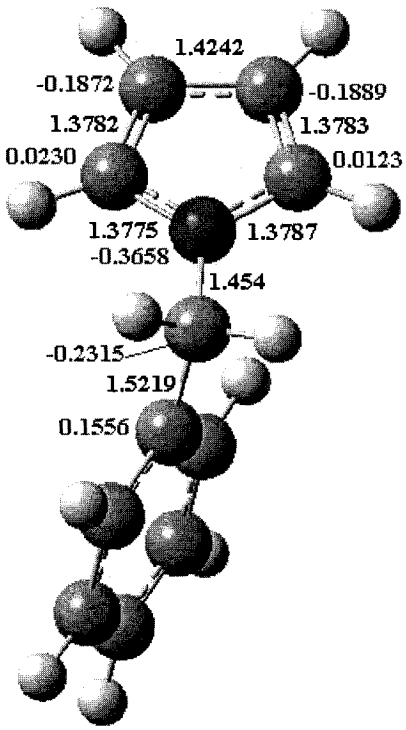
Table 5.7 reports the chemical shift data for the pyrrole series **4** at the B3PW91/6-311+G(2dp,f). Experimental ^{15}N chemical shifts are not available for the pyrrole series **4**. However, they can be compared to the chemical shift trend for the pyrrole series **3**.



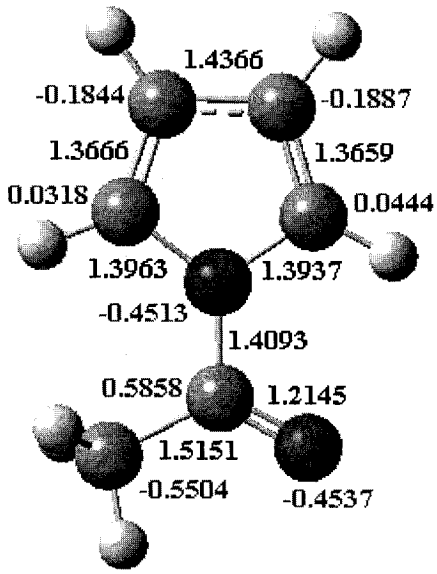
4a



4e

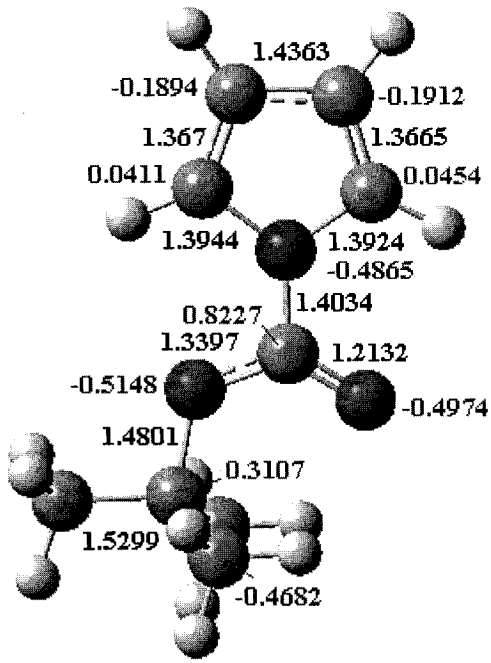


4f

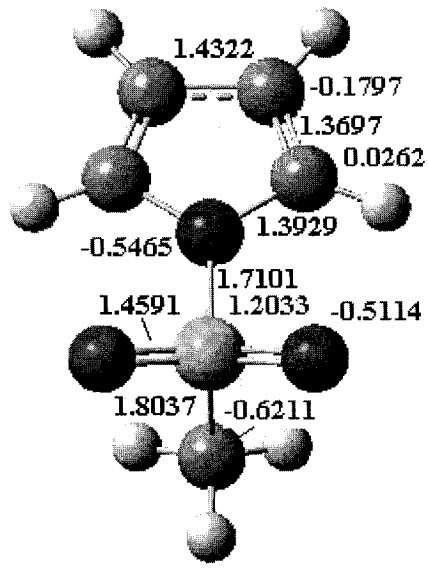


4g

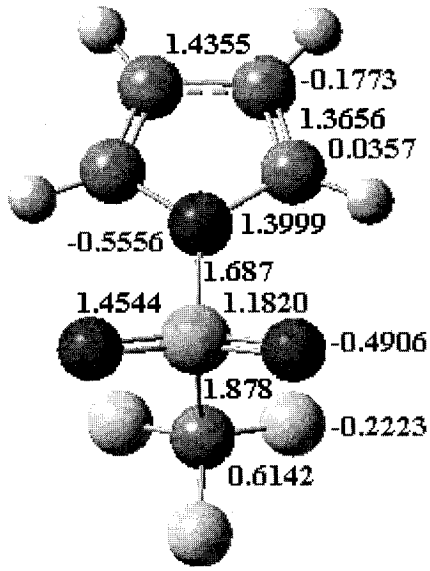
Figure 5.8 The optimized geometries, bond lengths and the Mulliken charges of the pyrrole series **4** (continued over page).



4b



4c



4d

Figure 5.8 The optimized geometries, bond lengths and the Mulliken charges of the pyrrole series **4**.

Table 5.7 Calculated ^{15}N chemical shift data for the pyrrole series **4** at the B3PW91/6-311+G(2dp,f) level of theory.

R	δ	δ_{aniso}	δ_d	δ_p	δ_{xx}	δ_{yy}	δ_{zz}
H	149.1	114.7	-71.5	465.1	141.6	243.1	62.5
Me	156.3	110.8	-63.3	464.2	143.2	258.5	67.1
Bn	168.8	98.6	-50.2	463.6	175.7	259.3	71.5
Ac	215.4	90.5	-65.9	525.9	222.3	311.2	112.7
Boc	193.7	125.8	-60.2	498.6	246.2	220.4	114.5
Ms	217.5	144.1	-59.5	521.6	273.7	228.4	150.5
Tf	201.1	167.5	-54.9	500.6	242.0	211.5	149.7

The nitrogen chemical shifts for pyrrole series **4** were calculated in order to investigate the effect of substituents on the carbons of the pyrrole ring in the pyrrole series **3** on the observed nitrogen chemical shifts. Figure 5.9 presents a graph of the nitrogen chemical shift data for the substituents of the pyrrole series **4**. Compared to the nitrogen chemical shift trend observed for the pyrrole series **3**, the same nitrogen chemical shift trend is still observed for the pyrrole series **4**. Thus, the substituents on the carbons of the pyrrole ring are not responsible for the anomalous results for the Ac and Tf substituents.

Other NMR parameters for the pyrrole series **4** are reported in Table 5.7 and Figure 5.9. The nitrogen diamagnetic shift δ_d stays approximately constant for all the substituents. The nitrogen paramagnetic shift δ_p correlates with the nitrogen chemical shift values in that, as the paramagnetic shift increases, the chemical shift values increase. Therefore, the nitrogen chemical shift trend observed arises from variations of the

paramagnetic shift contribution. The anisotropic shift of the pyrrole series **4** correlates with the observed chemical shift trend in an opposite sense. In other words, as the chemical shift increases, the anisotropic shift decreases (Figure 5.9) with the exception of the Ms and Tf substituents. The discrepancies might be due to larger errors for the Ms and Tf substituents compared to the other substituents.

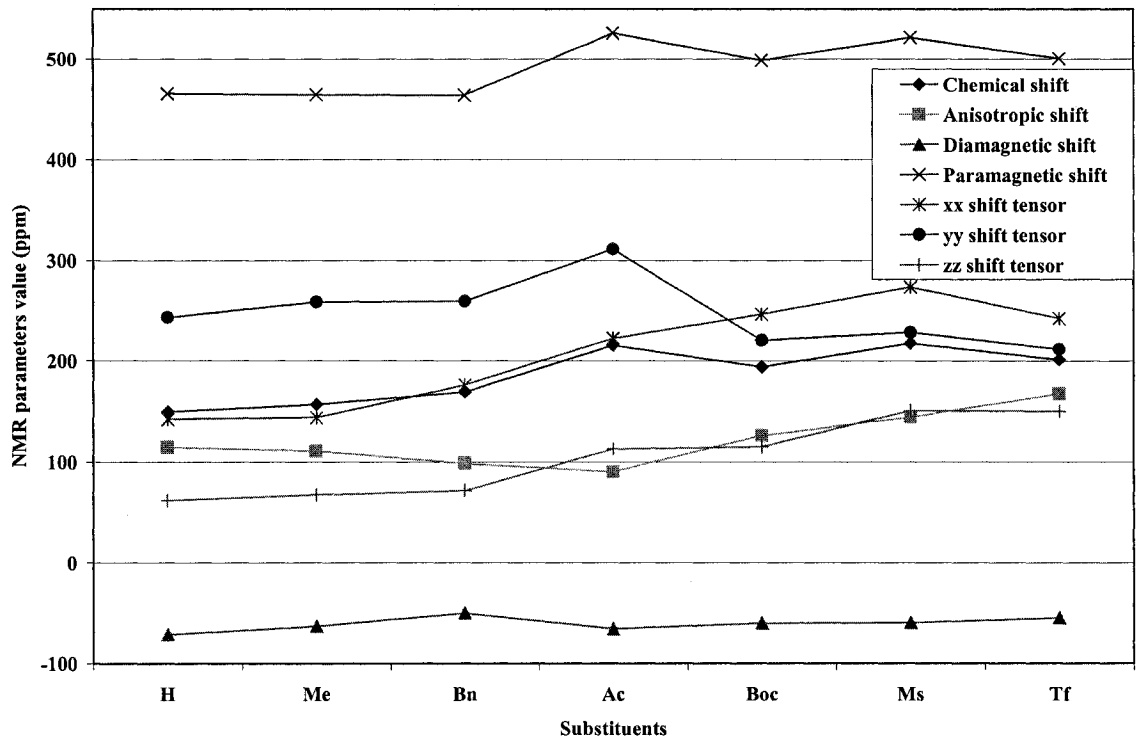


Figure 5.9 A graph of the calculated NMR parameters of the pyrrole series **4** using the B3PW91/ 6-311+G(2df,p) level of theory.

The shift tensor components for the pyrrole series **4** is also reported in Table 5.7 and represented in Figure 5.9. The orientation of the shift tensors is represented in Scheme 5.2. The δ_{xx} tensor component correlates with the chemical shift values except when comparing δ_{xx} of the Ac substituent to that of the Boc substituent; this exception

might be due to larger errors in the nitrogen chemical shifts for the Boc, Ms and Tf substituents. The δ_{yy} tensor component correlates with the chemical shift values. However, the δ_{yy} values of the last four substituents are less deshielding than the δ_{yy} values of the first three substituents (Figure 5.9). Thus, both δ_{xx} and δ_{yy} combined correlate with the chemical shift trend for these pyrroles. The δ_{zz} tensor component yields similar results to the δ_{zz} of the pyrrole series **3** in that the values of δ_{zz} are roughly constant based on the similarity of the EWG substituent (see Figure 5.9).

5.4.8 NCS Analysis of the ^{15}N Shielding for Pyrrole Series 4

The spatial orientation of the shift tensor on the molecular frame of the pyrrole series **4** is also represented in Scheme 5.2. Similar to the pyrrole series **3** analysis, the tangential component, δ_t, is dominated by the $\sigma_{N\text{-}R} - \pi^*$ transitions for the nitrogen nuclei of the pyrrole series **4**. Also, the radial component, δ_r, is dominated by the $\sigma_{N\text{-}C} - \pi^*$ transitions for the nitrogen nuclei of the pyrrole and the perpendicular component to the pyrrole ring, $\delta_\perp$, is dominated by the σ-electronic structure.

Table 5.8 reports the NCS analysis for pyrrole series **4**. The major contributions to the shielding from localized bonds and lone pair, and delocalized features are reported in the table. The NCS contributions are reported for all the substituents. A graph of the various NCS contributions for the pyrrole series **4** versus the substituents is presented in Figure 5.10. The contribution of the 1s (N5) is constant for the different EWG pyrroles since the core contribution to the chemical shift is independent of the chemical environment.[92] The σ(N5-R) total contribution correlates with the shielding trend, where

the σ(N5-R) corresponds to the tangential direction contribution with respect to the nitrogen position (Figure 5.10) except when comparing the Boc and Ms substituents. This might be due to an overestimation of shielding for the Ms and Tf substituents. Therefore the change in the σ(N-R)-π* transition might account for the shielding trend observed in the pyrrole series **4**. The shielding does not correlate with σ(C1-N5) and σ(C4-N5), which are a part of the shielding contributions to the radial direction (σ_{yy}). The observed shielding correlates with the n_{π}(N5) contribution with the exception of the H substituent (Figure 5.10). The total Lewis and non-Lewis shielding contributions correlate with the total shielding.

Table 5.8 Calculated NCS orbital decomposition for ^{15}N isotropic shieldings for the pyrrole series **4** at the B3PW91/6-311+G(2df,p) level of theory.

	σ_{iso}(N)						
Orbital / R	H	Me	Bn	Ac	Boc	Ms	Tf
σ (R-N_5)	-36.21	-39.08	-43.76	-54.81	-60.69	-47.86	-45.68
σ (N_5-C_1)	-54.16	-65.16	-56.73	-72.05	-63.94	-65.45	-61.47
σ (N_5-C_4)	-54.16	-65.16	-59.31	-57.97	-54.47	-65.45	-61.47
n_{π} (N_5)	17.82	32.79	20.72	12.53	30.00	13.38	19.49
1s (N_5)	239.76	239.73	239.75	239.74	239.74	239.69	239.68
Lewis	94.3	88.35	74.46	40.66	64.65	39.76	53.57
Non-Lewis	1.22	-0.06	1.34	-11.45	-13.78	-12.7	-10.05
Total shielding	95.52	88.29	75.8	29.20	50.87	27.07	43.52

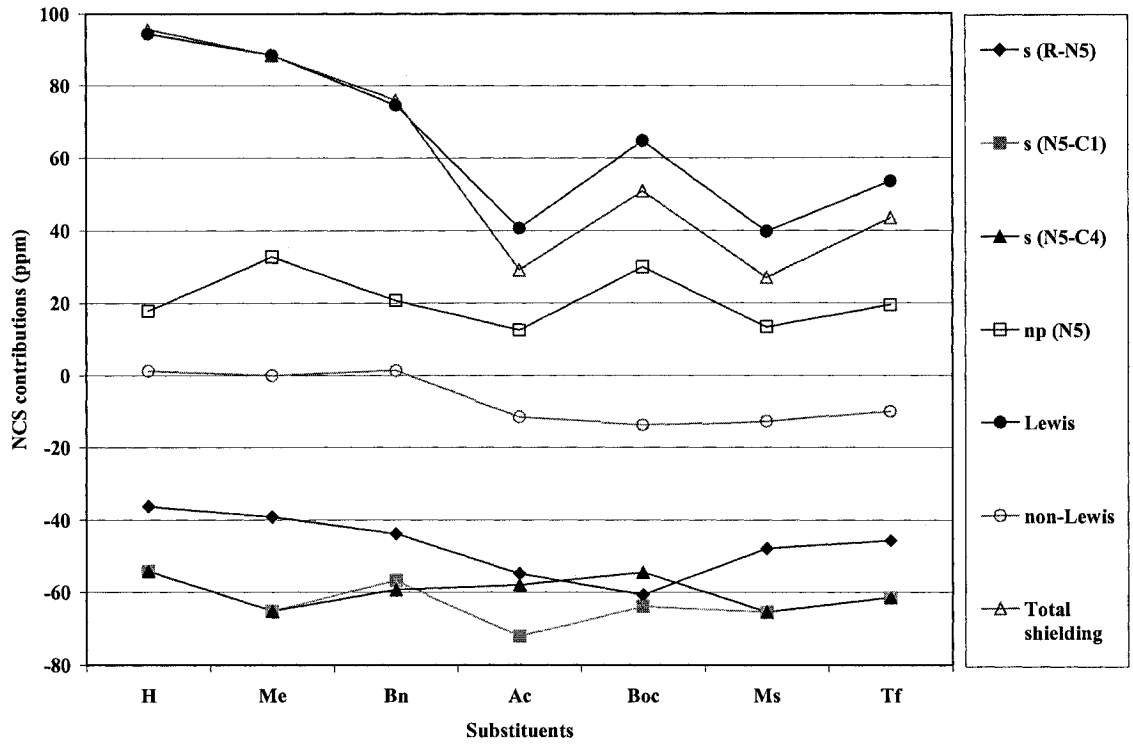


Figure 5.10 A graph of the various NCS contributions for pyrrole series **4** versus the substituents. The s in the graph label stands for σ, and np stands for n_{π}.

5.5 Conclusions

The calculated NMR parameters are very sensitive to conformational changes. The experimental ^{15}N chemical shift trend for the pyrrole series **3** was reproduced computationally. However, small differences (less than 10 ppm) in the nitrogen chemical shifts of pyrroles could not be reproduced. A correlation between the paramagnetic shift and the chemical shift was obtained for the pyrrole series **3** indicating that variations in the paramagnetic shift account for the nitrogen chemical shift trend observed. Dramatic changes in the nitrogen shift tensors of the pyrrole series **3** result from the paramagnetic shift. The σ(N5-R) total nitrogen contribution correlates with the shielding trend, where

the σ(N5-R) contribution corresponds to the tangential direction contribution with respect to the nitrogen position. The tangential direction is the same as the σ_{xx} which also correlates with the total nitrogen shielding, thus the changes in the σ(N5-R)-π* transitions account for the nitrogen chemical shift trend observed in the pyrrole series **3**. The sum of the σ(C1-N5) and σ(C4-N5) contributions corresponds to the shielding contribution in the radial direction, σ_{yy}, and it correlates with the shielding trend. The σ_{yy} correlates with the total nitrogen shielding except for the Tf substituent, thus the change in the sum of the σ(C1-N5)- π* and the σ(C4-N5)- π* transitions also accounts for the chemical shift trend observed in the pyrrole series **3**. Thus for the small pyrrole series studied, a simple correlation between the nitrogen chemical shifts and the EWG strength does not exist since the chemical shift is sensitive to the molecular environment and the EWG substituents set is chemically different.

For the ^{13}C chemical shifts, the C1, C2, C3 and C4 ^{13}C chemical shift trends were reproduced theoretically and show a general increase in chemical shift as the strength of N-EWG increases.

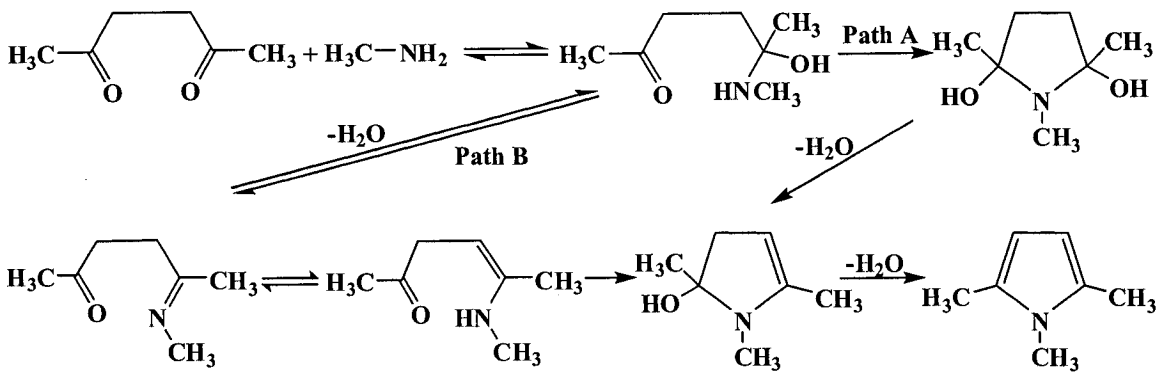
Based on the study of the nitrogen chemical shifts for pyrrole series **4**, it can be concluded that the substituents on the carbons of the pyrrole ring are not responsible for the chemical shift trend observed experimentally for the pyrrole series **3**.

Chapter 6: A Density Functional Theory Study of the Mechanism of the Paal-Knorr Pyrrole Synthesis

6.1 Introduction

The Paal-Knorr[1,95,96] synthesis, one of the most useful preparative methods for generating pyrroles, involves the condensation of 1,4-dicarbonyl compounds with primary amines. In spite of the Paal-Knorr synthetic utility, its mechanism is still not completely understood.[97-99] Although much is known about the individual steps of the Paal-Knorr mechanism, little is known about the detailed sequence of these steps or the nature of the intermediates that occur in the reaction.[97]

The proposed mechanism[97-99] (Scheme 6.1) for the Paal-Knorr synthesis involves two main alternatives: the first pathway involves the formation of the five-membered ring intermediate 2,5-dihydroxypyrrolidine (Path A), and the second one involves the formation of the imine and enamine intermediates (Path B). Path A involves cyclization of the hemiaminal intermediate, whereas Path B involves cyclization of the enamine intermediate. Since the formation of hemiaminals and imines is known to be rapid and reversible,[100] the rate-limiting step must be the ring closure step or the elimination of the diol.



Scheme 6.1 Proposed mechanism for the Paal-Knorr reaction.

The Paal-Knorr mechanism has been debated for a long time in the literature. Earlier studies of the Paal-Knorr mechanism by Borche and Fels[101] suggested that the ring closure occurs in the enamine and its tautomer, the imine, which was eventually detected by Broadbent et al.[102] Sundberg[103] in his 1984 review favoured the 2,5-dihydroxypyrrolidine intermediate arising from the cyclization of hemiaminal, as did Tseng et al.[104-106] on the basis of a detailed kinetic and ^{1}H NMR study of the reaction. In 1987, Katritzky et al.[97] argued on the basis of ^{13}C, ^{1}H and ^{15}N NMR studies that the enamine, which is present in low-steady-state concentrations with the imine, cyclizes in the rate-limiting step. The imine intermediate showed a distinct peak in the ^{13}C NMR which appeared and grew to a maximum and then disappeared as the product pyrrole was produced.[97]

Experimental kinetic studies by Amarnath et al.[98] investigated the stereoisomer reactivity of the reaction to provide insight into the mechanism of the Paal-Knorr mechanism. The reaction of diastereomers of 3,4-dimethyl-2,5-hexanediones in aprotic solvents and in aqueous solutions near neutrality was explored. It was found that the *d,l* diastereomers formed pyrroles faster than the corresponding meso diastereomers, which

contradicts the presence of an enamine intermediate that does not remain saturated at both the 3- and 4-positions through the rate-limiting step. The stereoisomeric difference in reactivity supports the hemiaminal as the intermediate undergoing cyclization. Another recent study by Amarnath et al.[99] examined the Paal-Knorr mechanism between a 4-ketoaldehyde and primary amine. Their deuterium experiments indicated that the reaction rate was not significantly affected by the presence of the deuteriums. The absence of a primary isotope effect and the absence of deuterium loss from the uncyclized ketoaldehyde suggest that the hemiaminal rather than the enamine is the intermediate undergoing cyclization. It was also concluded that the absence of deuterium loss from uncyclized ketoaldehyde suggests that with a primary amine the aldehyde reaches an equilibrium to form the imine, which is observed in the NMR spectra, but does not go further to the enamine.[99]

The Paal-Knorr mechanism has generated considerable interest since numerous biological effects are mediated through pyrrole formation in macromolecules. The formation of pyrrole adducts with proteins and nucleic acid bases has been associated with the active metabolites of several environmental toxicants.[99] A more complete understanding of the Paal-Knorr reaction will facilitate structure-activity relationships that are useful for assessing the potential reactivity and toxicity of new compounds. Further studies may also yield insight into the still debated mechanism of the biosynthesis of the porphobilinogen (PBG) pyrrole, which is a common precursor in the biosynthesis of natural tetrapyrroles such as chlorophyll, heme and vitamin B_{12}.[5]

A theoretical study of the Paal-Knorr mechanism has not been reported previously. Accordingly, a density functional theory investigation of the two pathways

proposed for the Paal-Knorr mechanism was carried out. The purpose of this study is to gain insight into the mechanism and the nature of the intermediates and their inherent activation energies and thereby complement the interpretation of experimental results.

6.2 Computational Details

A molecular complex consisting of 2,5-hexanedione, methylamine, and one water molecule was chosen as the model system to study the pyrrole formation in the Paal-Knorr reaction. It is essential to include a water molecule in the reaction because previous theoretical studies have shown that nucleophilic attack at the carbonyl carbons leads to barrierless reactions, along a pathway that is not available in the presence of a solvent.[107-110] However, a recent theoretical study by Delphine et al.[111] found a high energy barrier for the nucleophilic attack of ammonia on formaldehyde. The addition of one, two, and three water molecules to the reaction was also studied. It was concluded that the addition of one water molecule lowered the activation energy by a factor of two, while the addition of a second water molecule lowered the activation energy from roughly 75 to 54 kJ mol^{-1}. The addition of a third water molecule did not lower the activiation energy significantly (only by 4 kJ mol^{-1}).

Density functional theory calculations were carried out using the Gaussian 03 suite of programs.[112] All geometry optimizations were performed using the B3LYP hybrid density functional and the 6-31+G(d) basis set. The B3LYP functional is composed of Becke's three-parameter hybrid exchange functional (B3),[113,114] as implemented[115] in Gaussian 03, and the correlation functional of Lee, Yang, and Parr[116]

(LYP). Harmonic vibrational frequencies and zero-point vibrational energies (ZPVEs) were calculated at the same level of theory. Relative energies were obtained by performing single-point calculations at the B3LYP level in conjunction with the 6-311+G(2df,p) basis set using the above optimized geometries and by including the zero-point vibrational energy, i.e., B3LYP/6-311+G(2df,p)//B3LYP/6-31+G(d) + ZPVE. The entropy contributions to the free energies at 298.15 K were derived from the B3LYP/6-31+G(d) frequencies. Transition state (TS) optimizations were carried out using the Synchronous Transit-Guided Quasi-Newton (STQN) methods and the Berny algorithm. All TSs were confirmed to be true saddle points by having one, and only one, imaginary frequency. For each TS located, Intrinsic Reaction Coordinate (IRC) calculations were performed along the transition vector defined by the vibration mode of the imaginary frequency in order to confirm that the saddle point structure connected downhill the corresponding forward and backward minima. The IRC method allowed for the identification of the reaction intermediates and TS structures along the reaction path.

The Onsager[117] and polarizable continuum[118-121] (PCM) solvent models of the self-consistent reaction field (SCRF) methods were employed to investigate the solvent effects. The solvent is treated as a polarizable continuum with a dielectric constant and the solute molecule is embedded in a cavity of the dielectric continuum. The simple Onsager model places the solute in a fixed spherical cavity within the solvent reaction field, whereas the PCM model defines a more realistic cavity that is created via a series of overlapping spheres enabling a more accurate description of the solvent effects. For the Onsager calculations, geometry optimizations and frequency analyses were performed at the B3LYP/6-31+G(d) level of theory on the optimized gas-phase structures with a

dielectric constant of 78.39 for H_2O to model the aqueous media in the original reaction conditions. Relative energies were also obtained for the Onsager-solvated mechanism using the same level of theory for the gas-phase mechanism, i.e., Onsager-B3LYP/6-311+G(2df,p)//Onsager-B3LYP/6-31+G(d) + ZPVE. For the PCM model, geometry optimizations and frequency analyses were performed at the B3LYP/6-31+G(d) level of theory on the optimized Onsager structures with a dielectric constant of 78.39 for H_2O. Relative energies were also obtained for the PCM-solvated mechanism using the same level of theory for the gas-phase mechanism, i.e., PCM-B3LYP/6-311+G(2df,p)//PCM-B3LYP/6-31+G(d) + ZPVE. The entropy contributions to the free energies in solution at 298.15 K were derived from both the Onsager-B3LYP/6-31+G(d) and the PCM-B3LYP/6-31+G(d) frequencies. All energies reported are in kJ mol^{-1} and bond lengths in angstroms (Å).

6.3 Results and Discussion

The two proposed alternative pathways for the Paal-Knorr mechanism were studied (Scheme 6.1) for the model system of 2,5-hexanedione, methylamine and a water molecule. The process shown in Scheme 6.1 can be described by four main steps, each involving several intermediates and transition state (TS) structures: hemiaminal formation, hemiaminal cyclization (Path A), enamine cyclization (Path B) and dehydration.

The computed energy profiles for the reaction steps are shown in Figures 6.1-6.12. Figures 6.1-6.6 include only the ZPVE correction, while the entropy contribution at 298.15 K is included in Figures 6.7-6.12. The discussion will be focused on the results of

the ZPVE corrected case since the entropy contribution at 298.15 K shows similar energy profiles.

6.3.1 Gas Phase Calculations

Hemiaminal Formation. Initially, a water molecule forms a hydrogen bonding network between 2,5-hexanedione and methylamine in the gas-phase (Figure 6.1) via an NH...O bond of 2.244Å and an O...HO bond of 1.889Å, forming complex **1b** lying 28.5 kJ mol^{-1} lower in energy than the isolated reactants **1a**. The water bridge assists in the formation of a six-membered ring TS **1c** to facilitate the formation of the C-N bond between 2,5-hexanedione and methylamine as the proton transfers from the water molecule to the carbonyl oxygen of 2,5-hexanedione and from the N of methylamine to the water molecule, with a barrier of 74.9 kJ mol^{-1}. As the proton transfer is completed and the C-N bond of 1.487 Å is formed, a tetrahedral complex **1d** is generated, lying 23.2 kJ mol^{-1} higher in energy than **1b**. Complete removal of the water molecule gives the tetrahedral intermediate **1e**, hemiaminal, plus a water molecule. Two alternative pathways can now take place after the hemiaminal intermediate formation: hemiaminal cyclization (Path A) or enamine cyclization (Path B).

Hemiaminal Cyclization. The isolated hemiaminal **2a** forms a long O...HO hydrogen bond of 2.051 Å between the carbonyl oxygen and the hydroxyl group (Figure 6.2). From **2a**, the reaction proceeds to form a half-chair TS **2b** arrangement with strong hydrogen bonding between the NH and the carbonyl oxygen with a high energy barrier of 146.4 kJ mol^{-1}. The hydrogen transfers to the carbonyl oxygen and the ring cyclizes

forming the other C-N bond to produce the 2,5-dihydroxypyrrolidine intermediate, **2c**, which lies 1.2 kJ mol^{-1} lower in energy than **2a**.

The diol formed from the hemiaminal cyclization can then lose one water molecule to form one of the double bonds of the pyrrole ring as shown in Figure 6.3. A water molecule forms an O...HO hydrogen bond of 1.860 Å to the oxygen of one of the hydroxyls of the diol to produce complex **2e**. The water bridge assists the formation of TS **2f** with a small energy barrier of 60.3 kJ mol^{-1} via a hydrogen bonding network between the abstracted hydroxyl group, the other hydroxyl of the diol and the hydrogen to be abstracted to form the C-C double bond. Complete proton transfer and C=C formation from the TS generates complex **2g** which lies 7.4 kJ mol^{-1} lower in energy than **2e**.

Enamine Cyclization. The hemiaminal formed as in Figure 6.1 can also lose one water molecule to form an imine intermediate, which can tautomerize to form the enamine intermediate. As illustrated in Figure 6.4, a water molecule bridge assists the formation of a six-membered ring TS **3c** with 72.1 kJ mol^{-1} energy barrier to form the imine complex and releases a water molecule. With complete proton transfer, complex **3d** is produced forming the imine intermediate with two water molecules, lying 37.3 kJ mol^{-1} lower in energy than **3b**. Complex **3d**, where two water molecules form hydrogen bonds to the imine moiety, can tautomerize to form the enamine complex through proton transfer assisted by the two water molecules with an energy barrier of 62.8 kJ mol^{-1}. Complete proton transfer generates complex **3f** for the enamine intermediate, lying 33.6 kJ mol^{-1} higher in energy than the imine complex **3d**.

Complete removal of the water molecules yields the isolated enamine complex **3h** which contains an O...HN hydrogen bond of 2.097 Å between the carbonyl oxygen and the NH group (Figure 6.5). A four-membered ring TS **3i** is generated for the cyclization of the enamine to form the C-N bond with a high energy barrier of 161.7 kJ mol^{-1}. Downhill from this TS **3i**, the system evolves to form the five-membered ring complex **3j,** lying 17.7 kJ mol^{-1} higher in energy than the enamine complex **3h**. Complex **3j** is the same complex resulting from the hemiaminal cyclization (Path A) after the loss of one water molecule from the diol. Loss of one water molecule from **3j** yields the pyrrole ring product.

Dehydration. The last step in the Paal-Knorr mechanism is the formation of the pyrrole ring with a loss of another water molecule. This dehydration step occurs through the formation of two TSs as illustrated in Figure 6.6. The dehydration step is assisted by a water molecule that forms an OH...O hydrogen bond of 1.879 Å to the hydroxyl oxygen (complex **4b**). The dehydration proceeds via TS **4c** for the abstraction of the hydroxyl group with an energy barrier of 115.2 kJ mol^{-1}. The water bridge assists in the formation of a six-membered ring TS **4c**. The C-O bond distance of the hydroxyl group elongates to 2.623 Å and the C-N bond of the ring shortens from 1.470 Å to 1.320 Å to form a double bond. The water molecule also forms a long hydrogen bond of 1.977 Å with one of the C4 hydrogens of the ring. The reaction proceeds downhill from TS **4c** to form a high energy complex **4d**, which is only 8.0 kJ mol^{-1} lower than TS **4c**. Complex **4d** proceeds to form a second TS **4e**, where the C4 hydrogen bond is elongated to 1.145 Å to assist in the formation of the C=C of the pyrrole ring. The TS **4e** has a negative barrier that is 3.3 kJ mol^{-1} lower than complex **4d**. This might be due to the deficiency of

the DFT method to locate TSs where the reaction barrier is very small.[122] For hydrogen abstractions, it has also been established that B3LYP underestimates activation barriers.[123-126] The energy barrier for TS **4e** in the entropy-corrected case (Figure 6.12) is not negative and is 2.8 kJ mol^{-1} higher in energy than complex **4d**. The reaction proceeds downhill from **4e** to form the pyrrole ring complex **4f** with two water molecules. The formation of the pyrrole ring is thermodynamically favored and is exothermic. The product complex **4f** is 78.5 kJ mol^{-1} lower in energy than complex **4b**, whereas the isolated products **4g** are 53.3 kJ mol^{-1} lower in energy than **4b**.

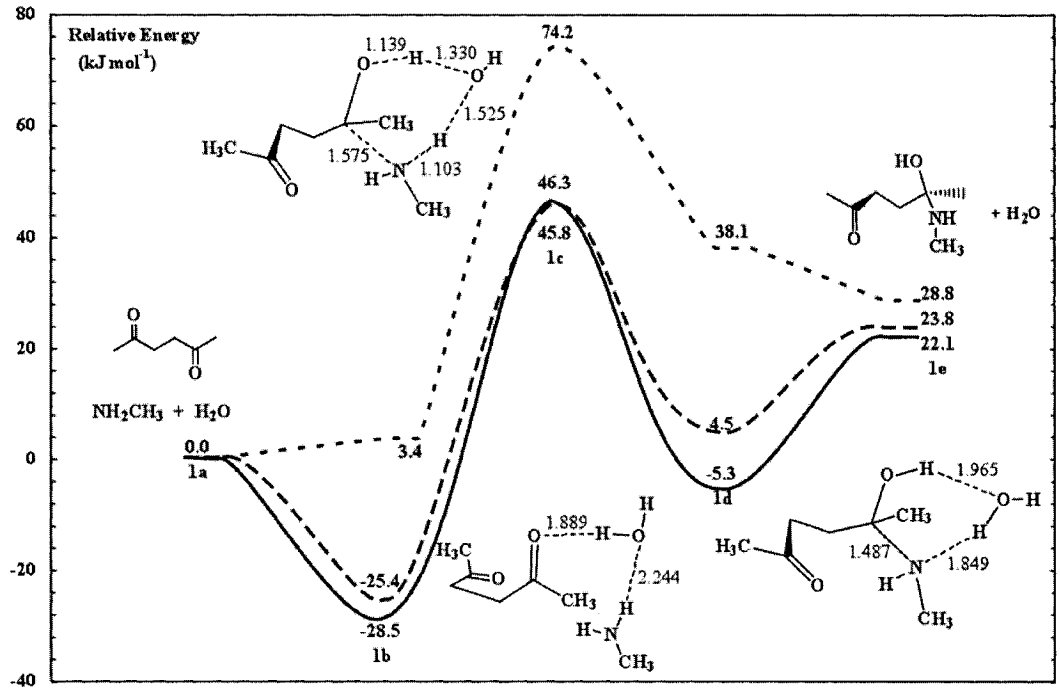


Figure 6.1 Schematic energy profile at 0 K for the hemiaminal formation step of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models. Optimized geometries are taken from the gas phase results.

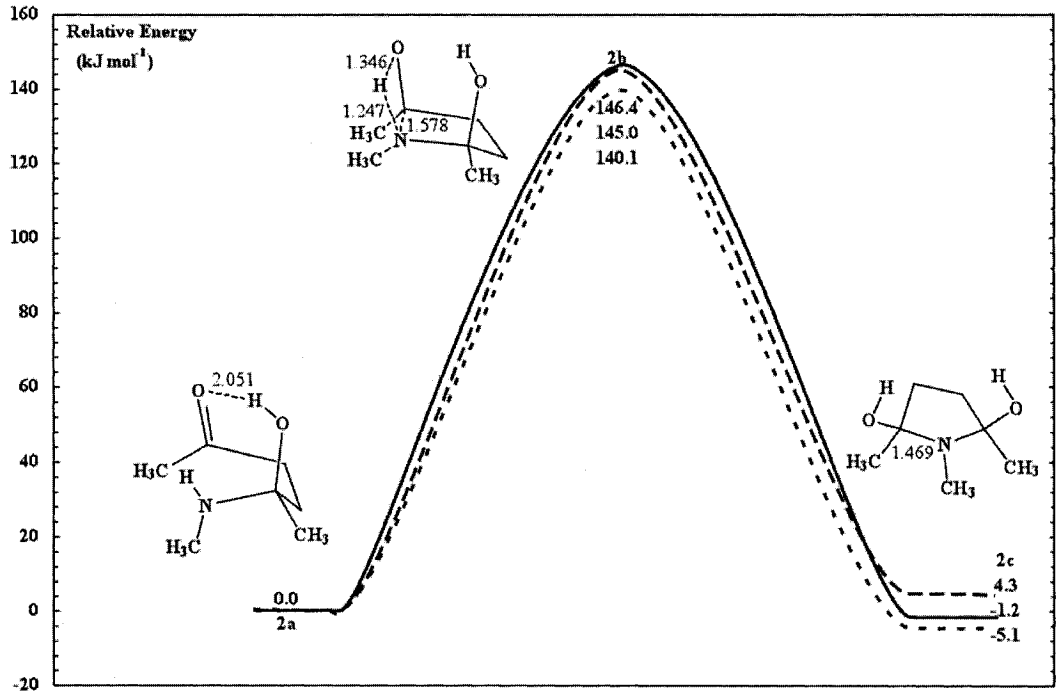


Figure 6.2 Schematic energy profile at 0 K for the hemiaminal cyclization step (Path A) of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models.

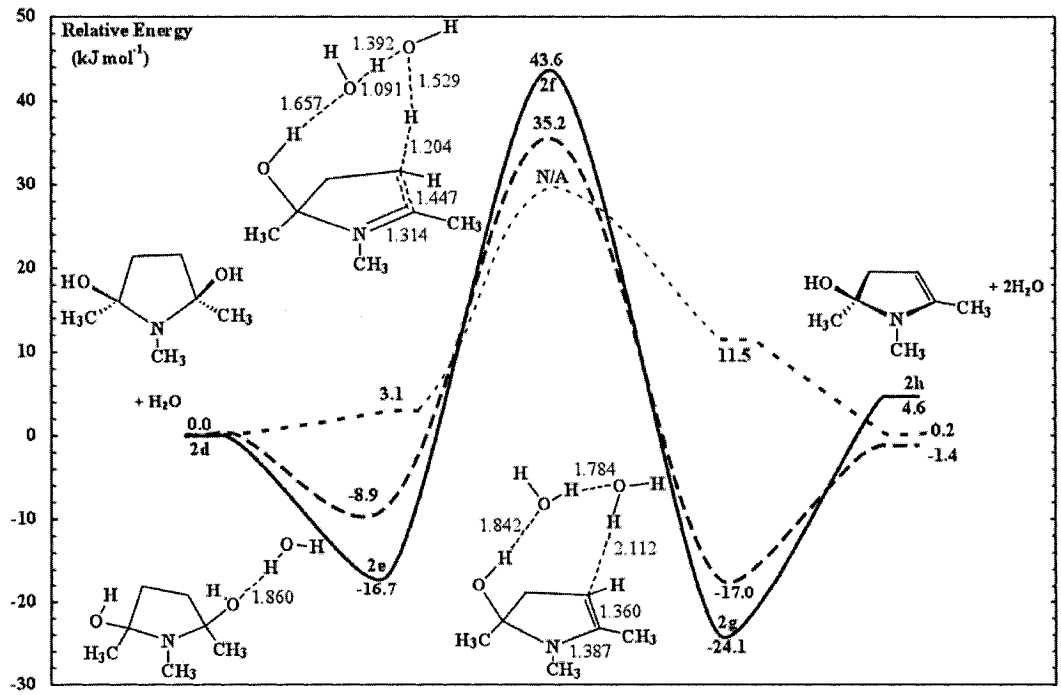


Figure 6.3 Schematic energy profile at 0 K for the elimination of an OH group from 2,5-dihydroxypyrrolidine (Path A) of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models. (---) denotes missing PES line.

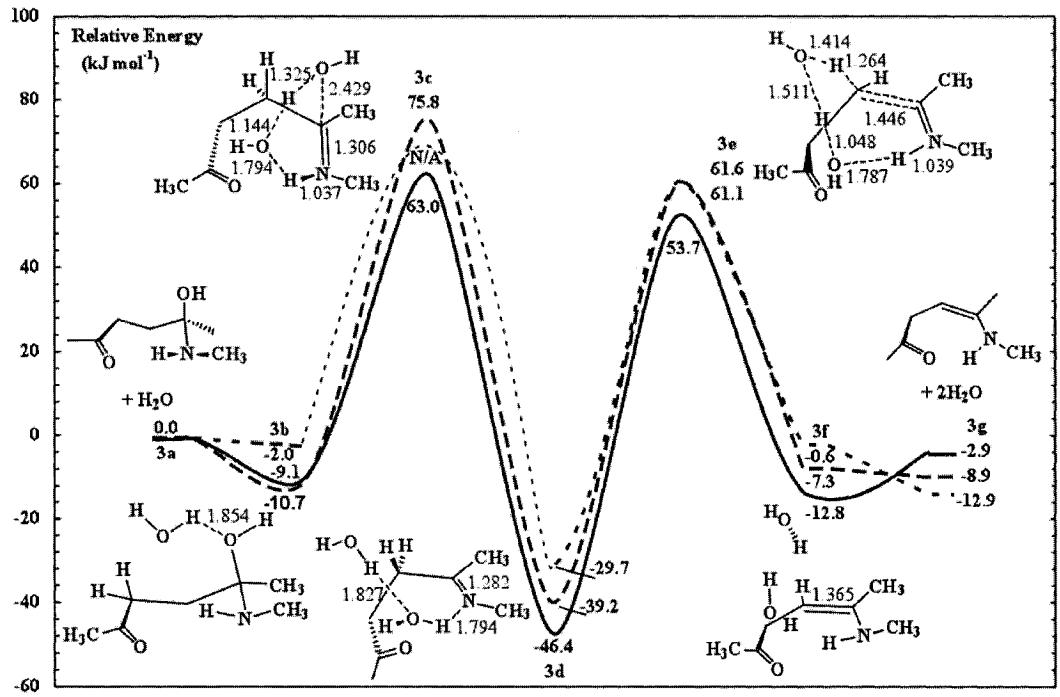


Figure 6.4 Schematic energy profile at 0 K for the imine and enamine formation (Path B) of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models. (---) denotes missing PES line.

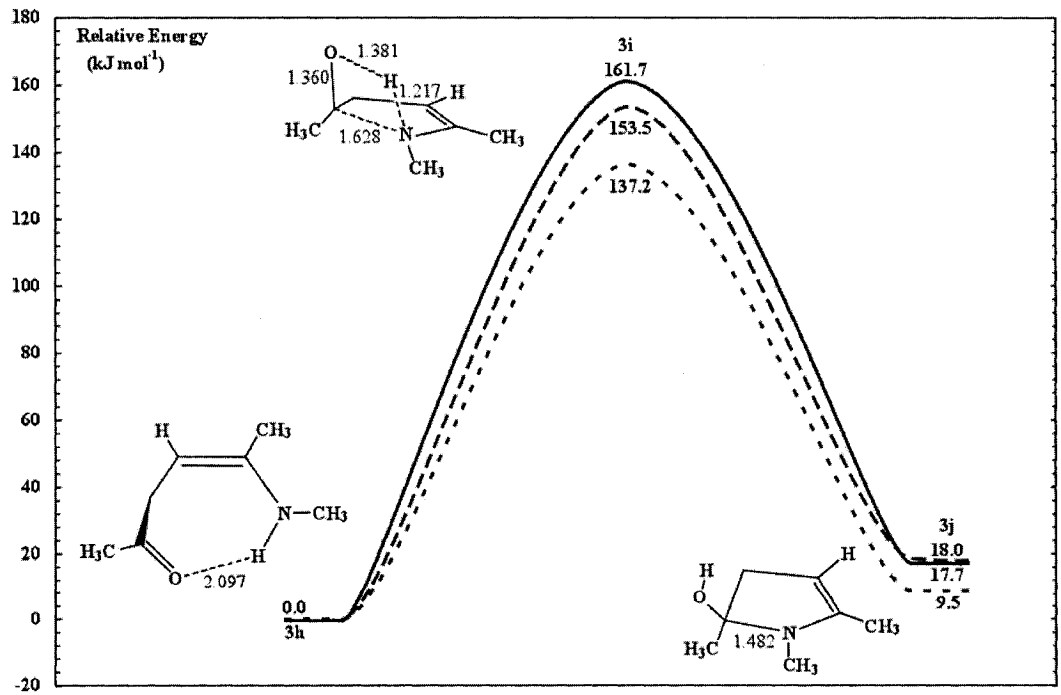


Figure 6.5 Schematic energy profile at 0 K for the enamine cyclization (Path B) of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models.

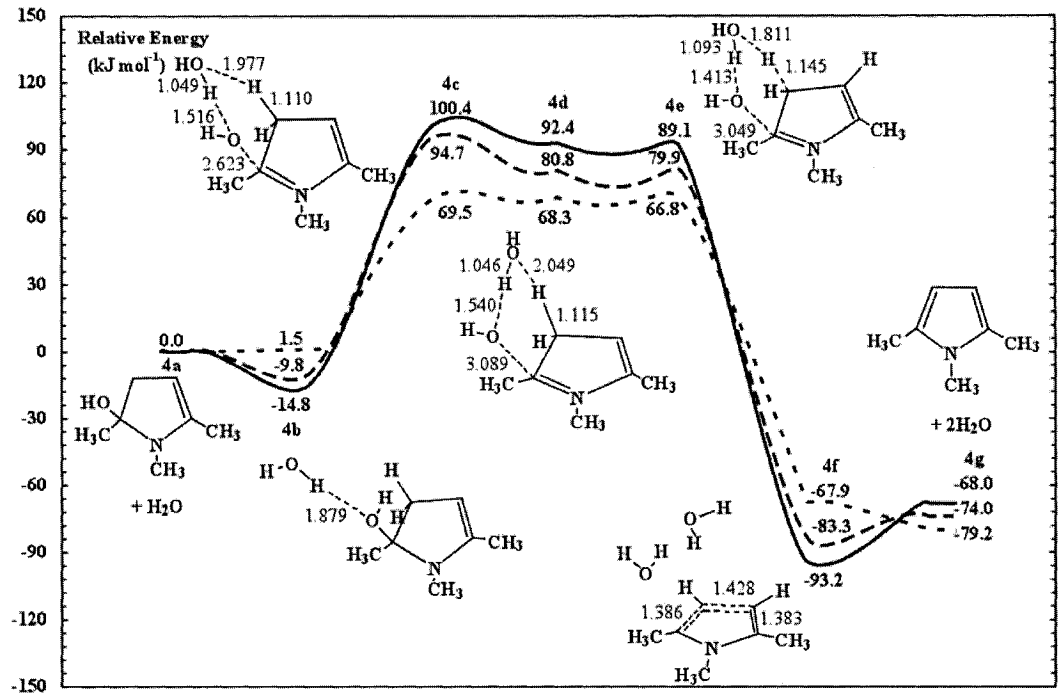


Figure 6.6 Schematic energy profile at 0 K for the dehydration step of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models.

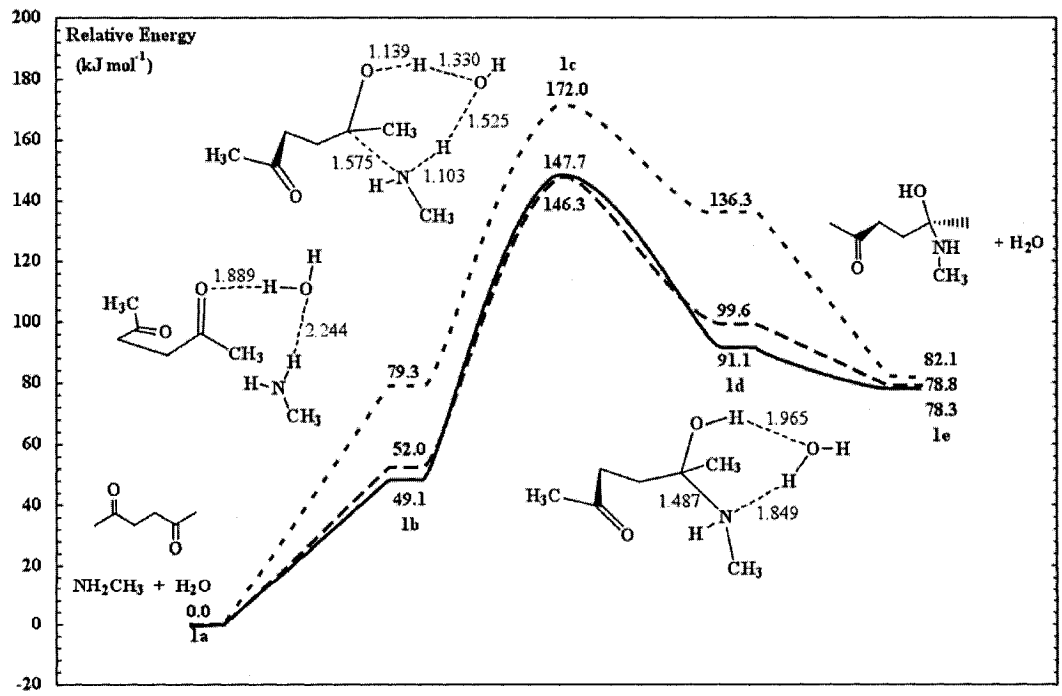


Figure 6.7 Schematic energy profile at 298.15 K for the hemiaminal formation step of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models.

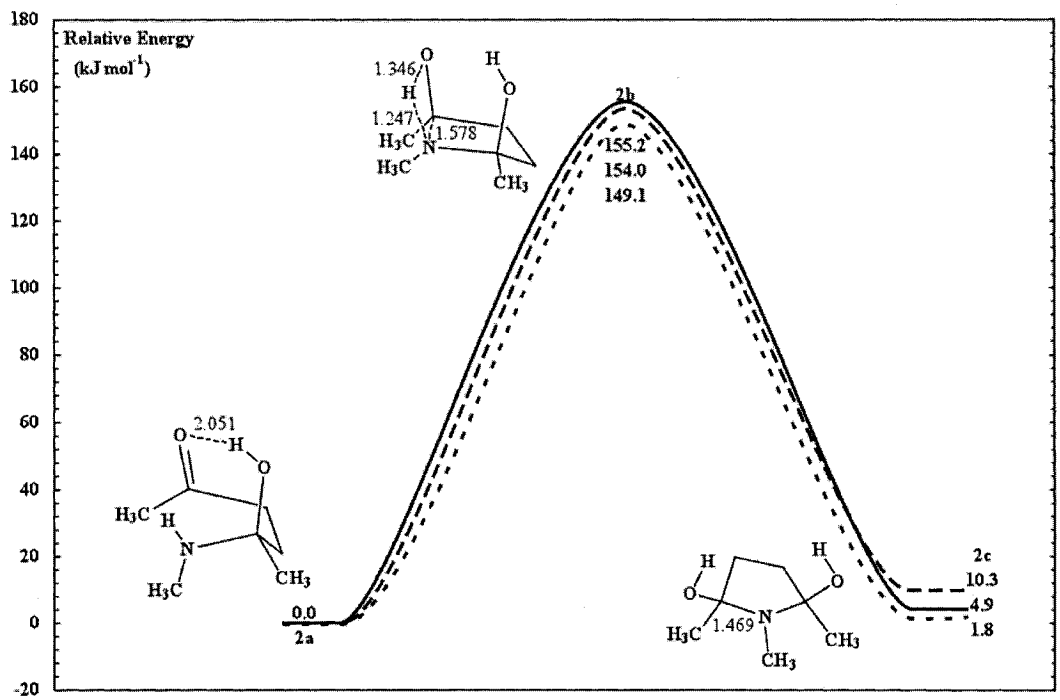


Figure 6.8 Schematic energy profile at 298.15 K for the hemiaminal cyclization step (Path A) of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models.

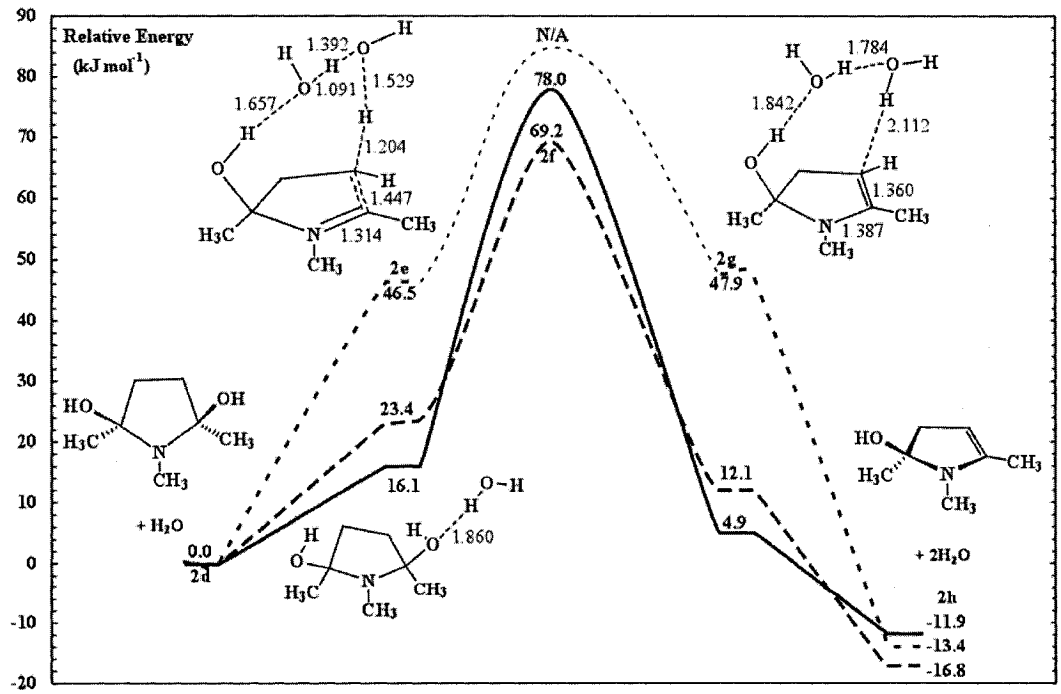


Figure 6.9 Schematic energy profile at 298.15 K for the elimination of an OH group from 2,5-dihydroxypyrrolidine (Path A) of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models. (---) denotes missing PES line.

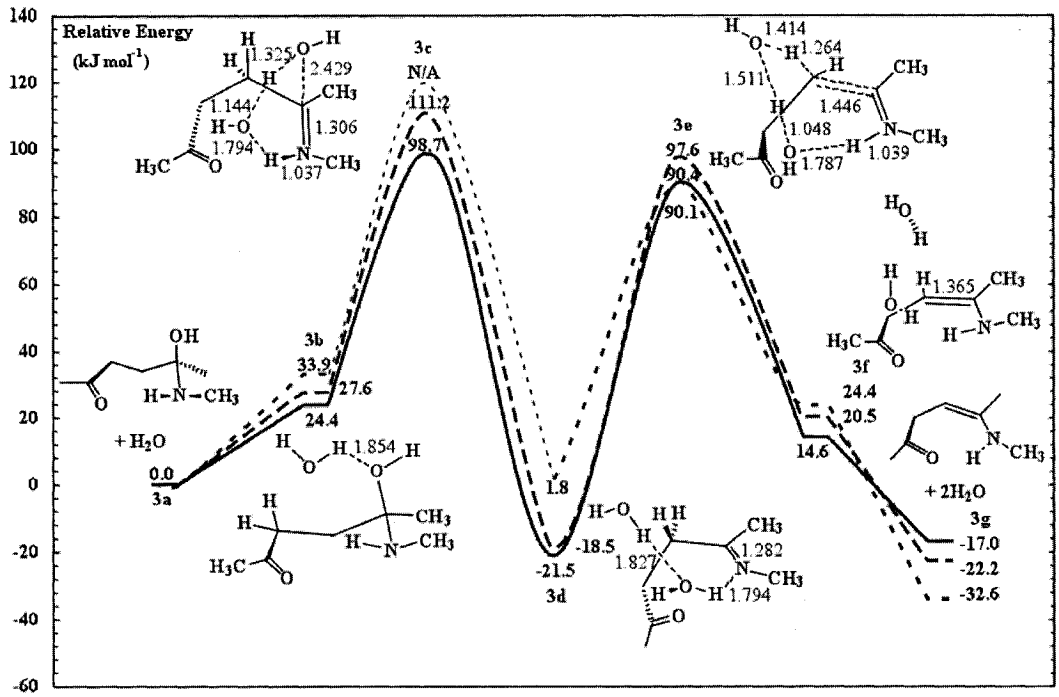


Figure 6.10 Schematic energy profile at 298.15 K for the imine and enamine formation (Path B) of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models. (---) denotes missing PES line.

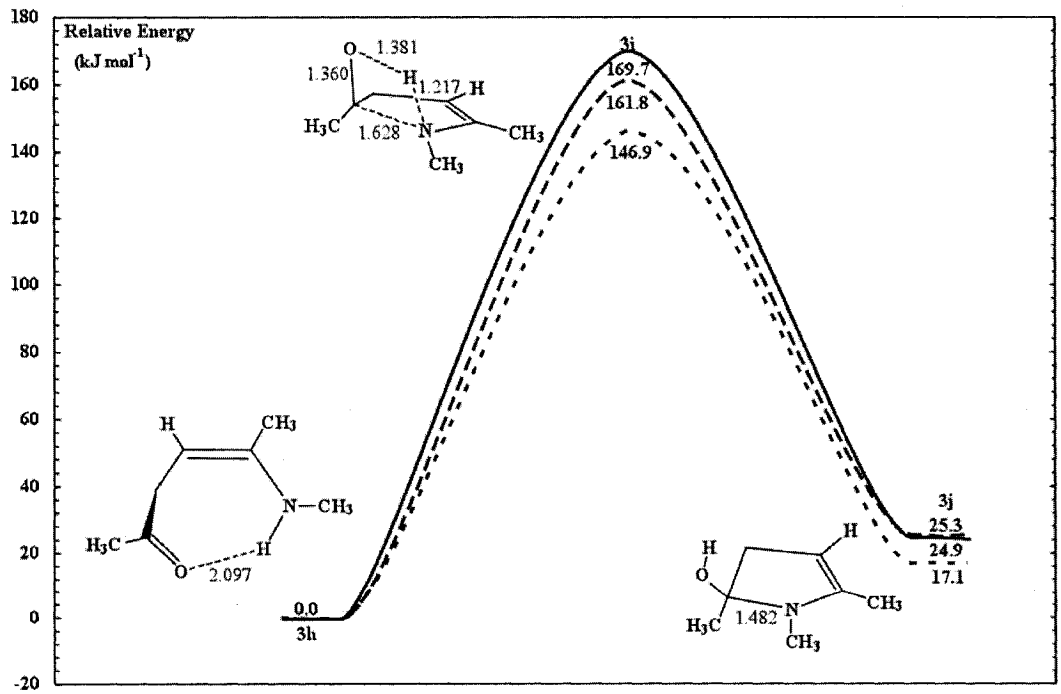


Figure 6.11 Schematic energy profile at 298.15 K for the enamine cyclization (Path B) of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models.

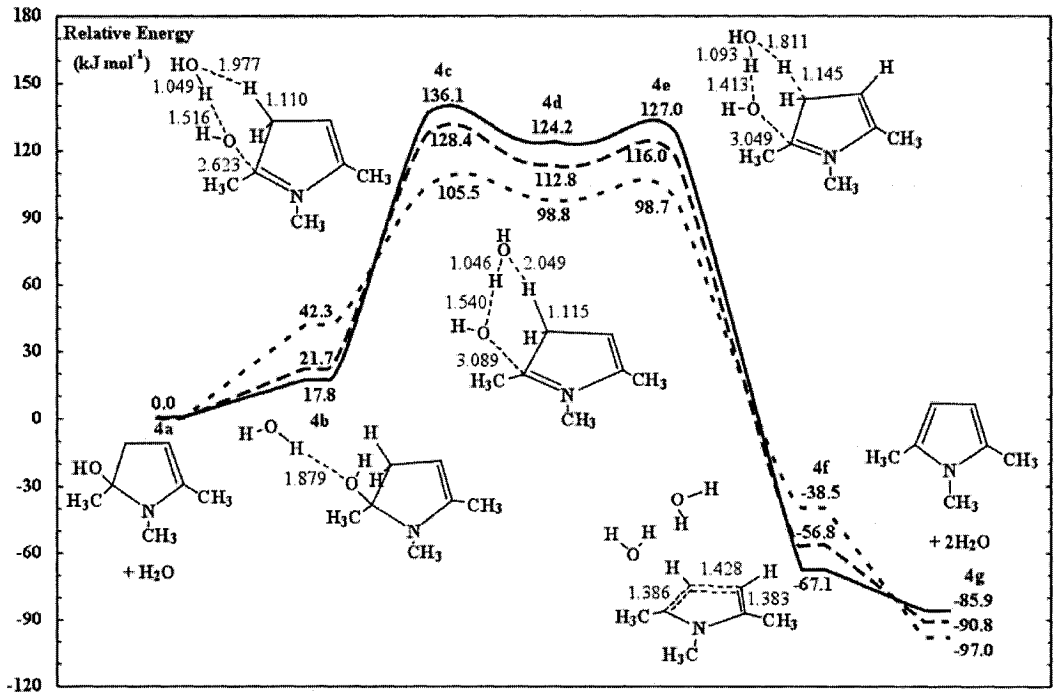


Figure 6.12 Schematic energy profile at 298.15 K for the dehydration step of the Paal-Knorr reaction in the gas phase (—) and in the Onsager (– –) and PCM (---) solvent models.

6.3.2 Onsager Solvent Calculations

Experimental studies on the mechanism were carried out in different environments, i.e. aprotic solvents and aqueous media. The solvent may play an important role in stabilizing intermediates and lowering key barriers in the reaction. The effect of solvent using the Onsager model on the Paal-Knorr mechanism is shown by the dashed lines in Figures 6.1-6.12.

Hemiaminal Formation. The optimized geometries in solution for the hemiaminal formation (Figure 6.1) are very similar to those in the gas phase. The geometry of complex **1b** has slightly shorter hydrogen bonds (NH...O bond of 2.218 Å and an O...HO bond of 1.870 Å). The geometry of the TS **1c** is also very similar to the

gas phase geometry with slightly longer bond lengths (0.002-0.1 Å) in the forming bonds and slightly shorter bond lengths (0.008-0.03 Å) in the breaking bonds. The geometry of the hemiaminal complex **1d** is similar to the gas phase geometry and produces a slightly longer H...O hydrogen bond by 0.05 Å and a slightly shorter N...H hydrogen bond by 0.03 Å. The energy barrier is only 3.7 $kJmol^{-1}$ lower than that of the gas phase.

Hemiaminal Cyclization. The effect of solvent in the hemiaminal cyclization step is not significant (Figure 6.2). The optimized geometries of **2a** and **2c** in solution are very similar to those in the gas phase, i.e. bond lengths for hydrogen bonding and the forming bonds change by less than 0.005 Å. The TS **2b** geometry is also very similar to the gas phase geometry. The forming and breaking bonds in **2b** change by less than 0.01 Å. The energy barrier is not significantly affected and is only 1.4 kJ mol^{-1} lower than that of the gas phase.

The reaction mechanism for the loss of one water molecule from the diol **2c** in solution (Figure 6.3) seems to be affected by solvent. The optimized geometries of **2e**, **2f** and **2g** are very similar to those in the gas phase with small differences. Complex **2e** gives a O...HO hydrogen bond that is shorter by 0.008 Å. The TS **2f** geometry is also very similar to that of the gas phase geometry. The TS **2f** gives a shorter forming C=C bond (by 0.01 Å) and shorter forming hydrogen bonds (less than 0.1 Å difference). The energy barrier is significantly lowered by 16.0 kJ mol^{-1} from that of the gas phase. Complex **2g** gives shorter forming and breaking hydrogen bonds (by about 0.01Å) compared to the gas phase geometry.

Enamine Cyclization. The mechanism for the formation of imine and its enamine tautomer from the hemiaminal intermediate is not significantly affected by

solvent (Figure 6.4). The optimized geometries of **3b**, **3c** and **3d** for imine formation are comparable to those in the gas phase. The geometry of complex **3b** is modified in solution. Complex **3b** forms a longer HO...H hydrogen bond (by 0.01 Å) and a much longer NH...O hydrogen bond (3.860 Å compared to 2.578 Å). Another O...NH hydrogen bond (2.046 Å) between the carbonyl oxygen and the NH group forms in solution. TS **3c** has slightly longer forming hydrogen bonds and shorter breaking hydrogen bonds with differences less than 0.1 Å. The energy barrier is 14.4 kJ mol^{-1} higher than that of the gas phase. Complex **3d** has slightly longer hydrogen bonds with differences less than 0.01 Å. TS **3e** for the formation of the enamine intermediate has a similar geometry to that in the gas phase with slightly shorter forming bonds (differences less than 0.03 Å) and slightly longer breaking bonds (differences less than 0.04 Å). The energy barrier for the enamine formation is not significantly affected by including solvent and is only 0.2 kJ mol^{-1} higher than that of the gas phase. The enamine complex **3f** has a similar geometry to the gas phase geometry. The enamine complex **3f** is 31.9 kJ mol^{-1} higher in energy than the imine complex **3d**.

Similarly to the hemiaminal cyclization, the enamine cyclization is not significantly affected by solvent (Figure 6.5). The enamine complex **3h** forms a longer O...HN hydrogen bond (by 0.01 Å). The four-membered ring TS **3i** gives shorter forming bonds (differences less than 0.02 Å) and longer breaking bonds (differences less than 0.01Å) compared to the gas phase geometry. The energy barrier for the enamine cyclization is 8.2 kJ mol^{-1} lower than that of the gas phase. Complex **3j** gives a similar geometry to the gas phase geometry. The enamine cyclization step is still endothermic by 18.0 kJ mol^{-1}, which is similar to the gas phase (17.7 kJ mol^{-1}).

Dehydration. The dehydration step of the reaction is slightly affected by solvent (Figure 6.6). The optimized geometries are similar to those in the gas phase with slight differences. Complex **4b** forms a shorter hydrogen bond (by 0.009Å) between a water molecule and the hydroxyl group. The TS **4c** for the abstraction of the hydroxyl group gives a similar geometry to that in the gas phase with small changes in the position of the water molecule. However, the same hydrogen bonding network is obtained in solution with differences less than 0.1 Å in the bond lengths of the forming and breaking bonds. The energy barrier for the hydroxyl abstraction is 10.7 kJ mol^{-1} lower than that of the gas phase. The high energy intermediate complex **4d** is also similar to that in the gas phase with small changes in the dihedral angles of the hydrogen bonding network. The differences in the bond lengths of the hydrogen bonding network in **4d** are less than 0.4 Å. The structure of the second TS **4e** for the hydrogen abstraction to form the second C=C bond is comparable to that in the gas phase. The differences in the bond lengths of the hydrogen bonding network to bring about the hydrogen abstraction in **4e** are less than 0.09 Å. The energy barrier of the second TS **4e** is 14.2 kJ mol^{-1} lower than that of the gas phase. The product complex **4f** for the pyrrole ring with two water molecules is very similar to that in the gas phase with differences less than 0.06 Å in the bond lengths of forming bonds.

6.3.3 PCM Solvent Calculations

The solvent effects with the simple Onsager model produced similar results to the gas phase. Therefore, PCM calculations were also carried out to obtain a more accurate description of the solvent effects. The effect of solvent using the PCM model on the

Paal-Knorr mechanism is shown by the dotted lines in Figures 6.1-6.12. Frequency analyses for the complexes involving water molecules yielded low imaginary frequencies (below 58 cm^{-1}). The nature of the displacement of the normal mode corresponding to these imaginary frequencies is not significant and is related to a slight repositioning of water molecules associated with flat regions of the potential energy surface (PES). Therefore, these imaginary frequencies can be ignored. The TSs were confirmed to be true saddle points with one large imaginary frequency that corresponds to the motion of the bond breaking and formation, with the exception of the two TSs in the dehydration step which have small imaginary frequencies of -44 and -11 cm^{-1}, respectively.

Hemiaminal Formation. The optimized geometries for the hemiaminal formation (Figure 6.1) with the PCM model are very similar to those with the Onsager model. The geometry of complex **1b** has slightly longer hydrogen bonds (NH...O bond of 2.234 Å and an O...HO bond of 1.871Å) compared to the Onsager geometries. The geometry of TS **1c** is similar to the Onsager geometry with slightly longer hydrogen bonds (0.07 Å difference) and slightly shorter forming and breaking bonds (0.02 Å difference). The geometry of the hemiaminal complex **1d** produces a slightly longer H...O hydrogen bond (by 0.05 Å) and a slightly shorter N...H hydrogen bond by (0.0008 Å). The energy barrier is only 0.5 kJ mol^{-1} lower than that of the Onsager. The hemiaminal formation step becomes more endothermic compared to the gas phase and Onsager PES.

Hemiaminal Cyclization. As with the Onsager results, the effect of solvent using the PCM model on the hemiaminal cyclization is not significant (Figure 6.2). The Onsager results produced very similar geometries to the gas phase for the hemiaminal

cyclization step. However, the PCM geometries obtained were slightly different. The hemiaminal **2a** forms a longer O...HO hydrogen bond, 0.858 Å longer compared to the Onsager geometry. The PCM half-chair TS **2b** geometry is not as distorted due to the longer O...HO hydrogen bond (3.190 Å), compared to the Onsager geometry (1.908 Å). The geometry of **2c** is very similar to the Onsager geometry. The energy barrier is only 5 kJ mol^{-1} lower than that of the Onsager. The reaction step is exothermic by 5.1 kJ mol^{-1}, while it is endothermic by 4.3 5 kJ mol^{-1} on the Onsager PES.

The PCM reaction mechanism for the loss of one water molecule from the diol **2c** is quite different from that of the Onsager and the gas phase results (Figure 6.3). The optimized geometries of **2e** and **2g** are very similar to the Onsager geometries with differences less than 0.05 Å. All attempts to locate the TS **2f** using the PCM model failed. This reaction step of the hydroxyl abstraction is endothermic since **2g** is 8.5 kJ mol^{-1} higher in energy than **2e**, which is opposite to the results of the Onsager and the gas phase PES.

Enamine Cyclization. The solvent effects using the PCM model on the mechanism of the formation of the imine and its tautomer, enamine, are very comparable to the Onsager results (Figure 6.4). The optimized geometries of **3b** and **3d** for the imine formation are similar to the Onsager geometries, with a slight repositioning of the water bridge (differences less than 0.2 Å). All attempts to locate TS **3c** failed. Similarly to the Onsager results, the formation of the imine is exothermic by 27.7 kJ mol^{-1}. The PCM optimized geometry of TS **3e** for the formation of the enamine intermediate is similar to the Onsager geometry with slightly shorter forming bonds (differences less than 0.04 Å) and slightly longer breaking bonds (differences less than 0.1 Å). The energy barrier for

the enamine formation is 9.0 kJ mol^{-1} lower than that of the Onsager results. The enamine complex **3f** has a similar geometry to the Onsager geometry with differences less than 0.04 Å.

The solvent effect on the enamine cyclization step using the PCM model is comparable to the Onsager result (Figure 6.5). The enamine complex **3h** forms a longer O...HN hydrogen bond (by 0.2 Å). The four-membered ring TS **3i** gives shorter forming bonds and longer breaking bonds (differences less than 0.03Å) compared to the Onsager geometry. The energy barrier for the enamine cyclization is 16.3 kJ mol^{-1} lower than that of the Onsager. The enamine cyclization step is less endothermic by 8.5 kJ mol^{-1} compared to Onsager PES.

Dehydration. The effect of solvent on the dehydration step of the reaction is significantly improved using the PCM model (Figure 6.6). The PCM optimized geometries are comparable to the Onsager geometries. Complex **4b** forms a shorter hydrogen bond (by 0.02Å) compared to the Onsager geometry. The TS **4c** for the abstraction of the hydroxyl group gives a similar geometry to the Onsager geometry with small changes in the position of the water molecules. However, the same hydrogen bonding network is obtained with differences of less than 1.0 Å in the bond lengths of the forming and breaking bonds. The energy barrier for the hydroxyl abstraction is 36.6 kJ mol^{-1} lower than that of the Onsager. The high energy intermediate complex **4d** is also similar to the Onsager geometry with small changes in the dihedral angles of the hydrogen bonding network. The differences in the bond lengths of the hydrogen bonding network in **4d** are less than 0.8 Å.

The geometry of the second TS **4e** for the hydrogen abstraction to form the second C=C bond is comparable to the Onsager geometry. The differences in the bond lengths of the hydrogen bonding network to bring about the hydrogen abstraction in **4e** are less than 0.08 Å. The energy barrier of the second TS **4e** is significantly lowered by 24.4 kJ mol^{-1} compared to the Onsager energy barrier. The product complex **4f** for the pyrrole ring with two water molecules is very similar to the Onsager geometry with slight positioning of the water molecules. Similarly to the gas phase and Onsager PES, the dehydration step is exothermic by 69.4 kJ mol^{-1}. Moreover, significantly lower energy barriers were obtained using the PCM solvent model compared to the Onsager and gas phase results for the dehydration step.

6.3.4 Comparison of Reaction Pathways

The proposed mechanism of the Paal-Knorr reaction involves two pathways. Path A includes the following reaction steps: hemiaminal formation, hemiaminal cyclization and dehydration, respectively; whereas Path B includes the following reaction steps: hemiaminal formation, enamine cyclization and dehydration, respectively.

Inclusion of solvent effects using the Onsager and PCM models leads to lower energy barriers. The solvent effect was significant in lowering the energy barriers for the hydroxyl abstraction from the diol and the dehydration step. However, the inclusion of the solvent produced similar energy profiles as in the gas phase and therefore, the preferred mechanism should be the same both in the gas phase and in solution. This conclusion is supported by the experimental studies of Amarnath et al.,[98,99] which suggest that the rate-limiting step should be the same since the higher reactivity of *d,l* compared

to meso diastereomers was always observed under different conditions (pH, solvent and substitution of amine) that affected the rate of pyrrole formation.

For both pathways in the gas phase (Figure 6.13), the cyclization of hemiaminal and enamine has the highest energy barrier. Therefore, the ring closure step is the rate-limiting step for the mechanism. The cyclization of hemiaminal in Path A adds two steps to the mechanism whereas the cyclization of enamine in Path B adds three steps to the mechanism. The two steps in Path A involve the hemiaminal ring closure and elimination of a hydroxyl group. The hemiaminal ring closure has a high energy barrier (146.4 kJ mol^{-1}) while the hydroxyl group elimination involves a lower energy barrier (60.3 kJ mol^{-1}). Both of these steps are exothermic and favour the formation of the products. In comparison, the three steps in Path B involve imine formation, enamine formation and enamine cyclization with high energy barriers: 72.1, 100.1 and 161.7 kJ mol^{-1}, respectively. Moreover, both the enamine formation and cyclization steps are endothermic and favour the formation of reactants. Based on the energetic results, the calculations suggest that Path A for the cyclization of the hemiaminal is the favoured pathway for the Paal-Knorr reaction. The present results support the recent experimental kinetic studies by Amarnath et al.[99] that the hemiaminal is the intermediate that undergoes cyclization in the rate-limiting step.

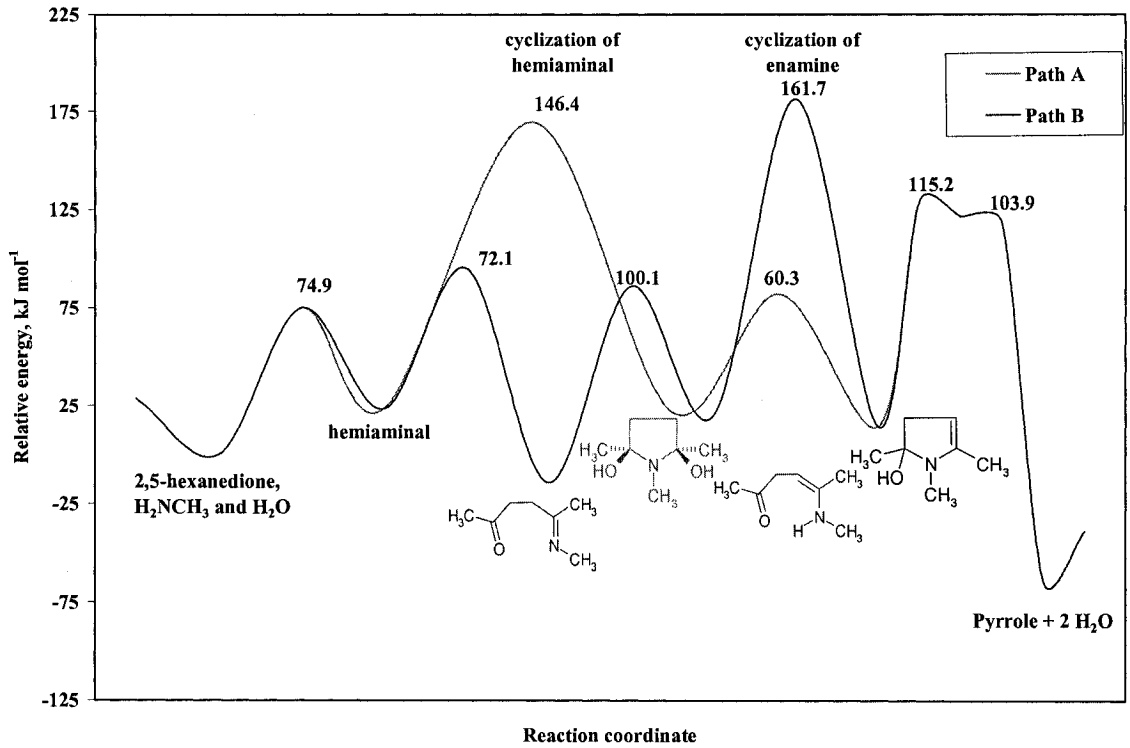


Figure 6.13 A summary graph of the individual steps of both pathways in the gas phase for the Paal-Knorr reaction. The numbers inside the graph refer to the individual barrier height for each step.

6.4 Conclusions

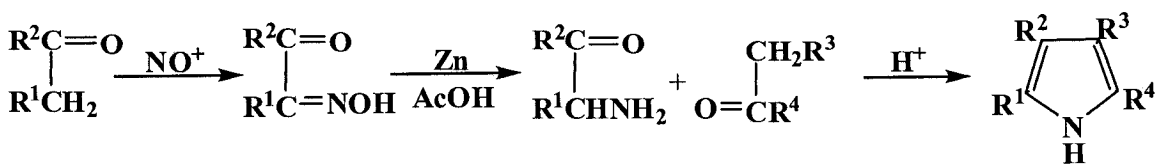
The Paal-Knorr reaction mechanisms of the pyrrole formation have been investigated by the use of the B3LYP density functional theory method. In the model reaction of 2,5-hexanedione and methylamine, a water molecule functions as a catalyst and as a reactant. Water bridges assist in the formation of TSs to facilitate proton transfer in the reaction. The inclusion of solvent effects using Onsager and PCM models leads to lower energy barriers. In particular, solvent effects significantly lower the energy barriers for the hydroxyl abstraction from the diol and the dehydration steps. The calculations suggest that Path A for the hemiaminal cyclization is the preferred pathway

both in gas phase and in solution and that the hemiaminal cyclization step is the rate-limiting step of the reaction. These conclusions are consistent with the recent experimental results[98,99] which indicate that the hemiaminal is the intermediate that undergoes cyclization in the Paal-Knorr reaction mechanism.

Chapter 7: A Density Functional Theory Study of the Mechanism of Pyrrole Formation in the Knorr Pyrrole Synthesis and the Porphobilinogen Synthase

7.1 Introduction

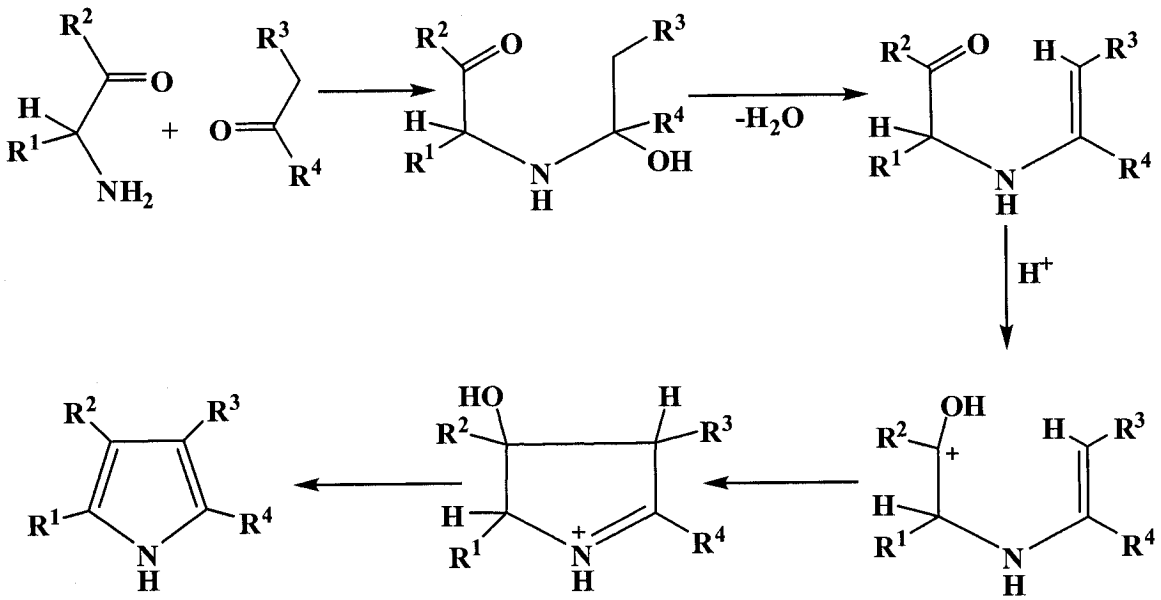
The Knorr pyrrole synthesis method, which involves the condensation of α-aminocarbonyls with ketones or aldehydes, is one of the most important and widely used methods for the synthesis of pyrrole rings.[1,127] The original method was developed by Knorr[96,128] for the synthesis of the diester **5** ($R^2 = R^4 = Me$; $R^1 = R^3 = CO_2Et$) (Scheme 7.1), which is called Knorr pyrrole.[129-131] The Knorr pyrrole was obtained from the condensation of the α-aminoketone **3** produced *in situ* (by nitrosation of the carbonyl compound **1** followed by reduction using zinc in glacial acid) with the carbonyl compound **4** (Scheme 7.1).[129,131] The mechanism and limitations of the Knorr synthesis have been reviewed by Corwin.[132] The limitations of the generalized reaction of **3** + **4** → **5** are summarized as follows: excellent yields of the pyrrole are obtained when R^1 and R^3 are acyl or alkoxycarbonyl groups and R^4 is either an alkyl, aryl, acyl or alkoxycarbonyl group;[132] whereas, a poor yield of the pyrrole is obtained when R^3 is an alkyl group. The nature of R^2 seems to have no effect upon the reaction.



1 2 3 4 5

Scheme 7.1 Formation of Knorr pyrrole.

The mechanism of the Knorr reaction is somewhat complex and has not been studied in detail. The Knorr pyrrole reaction is normally conducted under acidic conditions. The generally accepted proposed mechanism[1,127,133,134] of the Knorr reaction is presented in Scheme 7.2. The proposed mechanism of the Knorr pyrrole reaction starts with a nucleophilic attack of the α-aminoketone on the carbonyl carbon of the ketone or aldehyde to give a tetrahedral intermediate, hemiaminal. Loss of a water molecule produces the enamino ketone intermediate. Under acidic conditions, the enamino ketone intermediate cyclizes to form the pyrrole ring.

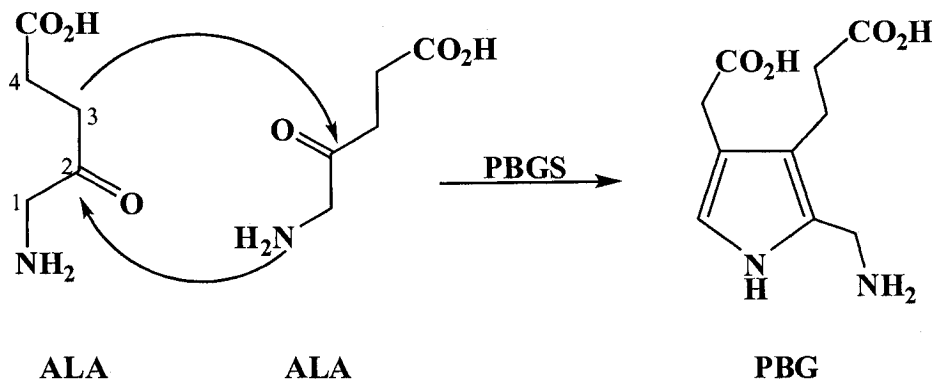


Scheme 7.2 Proposed mechanism of the Knorr pyrrole reaction.

The enamino ketone intermediate was isolated under mild reaction conditions in an early study by Nakano et al.[134] Furthermore, Khetan and George[135] were able to prepare relatively stable enamino ketones by attaching an aryl group to both the nitrogen

and keto functions, and have shown that such intermediates cyclize to give pyrroles up on refluxing in acidic methanol. Sanchez et al.[136] have also shown that enamino esters prepared from 2-amino-2-deoxy-D-glucopyranose and β-keto esters produce pyrroles under mild conditions (i.e. in water, 30 h at room temperature). Moreover, Fabiano and Golding[133] have also shown that compounds such as the enamino ketones undergo cyclization to pyrrole with a half-life of several hours at 10 °C in methanol.

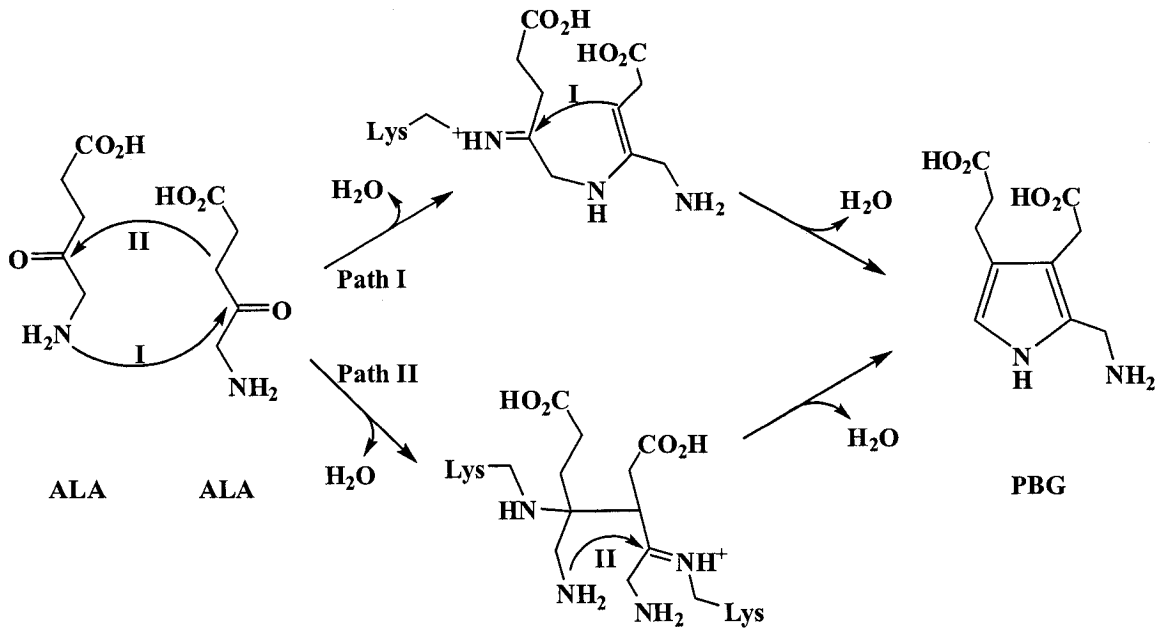
The biosynthesis of the porphobilinogen (PBG) pyrrole derivative results from the condensation of two 5-aminolevulinic acid (ALA) catalyzed by the enzyme porphobilinogen synthase (PBGS) enzyme (Scheme 7.3).[4,5] PBGS enzymes have been isolated from a number of sources and their properties have been characterized.[137] The two features that unify the PBGS enzymes are that they utilize covalent catalysis to form Schiff bases with the substrate during catalysis, and most appear to be metalloenzymes, requiring a bivalent metal ion for activity.[137]



Scheme 7.3 Biosynthesis of PBG pyrrole by PBGS enzyme.

Crystal structures of PBGSs including yeast, *E. coli*, and *Pseudomonas aeruginosa* show that there are two conserved lysine residues side-by-side in the active site.[5] The two lysine residues play a co-ordinate role in the mechanism of PBG synthesis. The dual role of the two lysine residues has been investigated by site-directed mutagenesis and inhibitor studies supported by X-ray crystallography.[137-140] The studies showed that both of the two lysine residues participate in covalent catalysis by forming Schiff bases.[4,5,137]

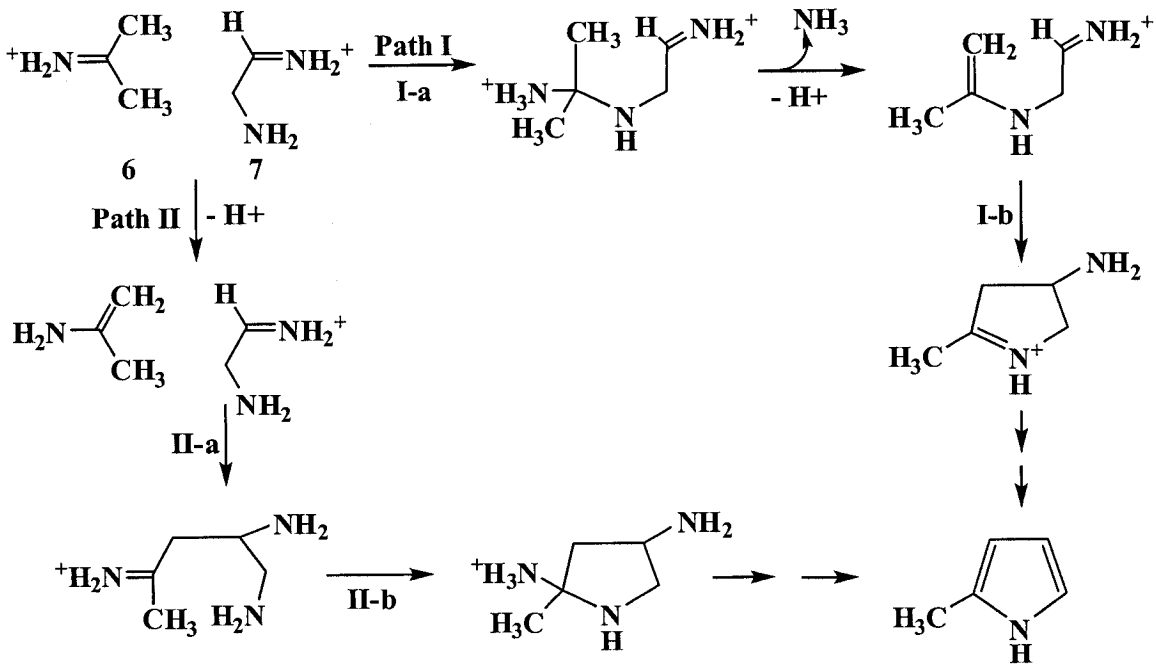
Based on the X-ray structures and other experimental studies of the PBGS enzyme, several suggestions about the catalytic mechanism have been proposed.[4,5,141-144] The mechanism is thought to form Schiff bases between the amino group of a lysine residue and the keto group of each ALA molecule. A sequential formation of both a C-C bond and C-N bond between the two ALA molecules can then take place to form the PBG pyrrole (Scheme 7.4). The order of C-C and C-N bond formations is still debated in the literature. Path **I** of the catalytic mechanism proceeds with the formation of the C-N bond first through a nucleophilic attack of the amine on the Schiff base carbon to form the enamine intermediate. The C-C bond then forms by the cyclization of the ring. The mechanism for Path **I** is similar to that proposed for the Knorr pyrrole synthesis. Path **II** proceeds first by a deprotonation of C3 of one of the ALA to produce an enamine. The enamine formed can then attack the iminium ion of the other ALA to form the C-C bond. The C-N bond can then be formed through a nucleophilic attack of the amine on the Schiff base to cyclize the ring (Scheme 7.4).



Scheme 7.4 Catalytic reaction of PBGS. Two ALA molecules are asymmetrically condensed to form PBG. Two new bonds (C-C and C-N) are formed in the process. Two intermediates are possible, depending on the sequence of the reaction steps.

The Knorr reaction mechanism has not been thoroughly studied in the literature and no theoretical study of the Knorr mechanism has been previously reported. Accordingly, a density functional theory investigation of the proposed mechanism of the Knorr reaction (Scheme 7.2) has been carried out. The model system for the Knorr reaction in Scheme 7.2 is the reaction of amino-acetaldehyde and 2-propanone (i.e. R^1= R^2 =R^3 = H and R^4 = CH_3). A water molecule is included in the reaction model since previous theoretical studies have shown that it is essential to include the solvent to obtain reliable energy barriers.[107-111] The purpose of the study is to gain insight into the mechanism and the nature of the intermediates and their inherent activation energies. Moreover, since the Knorr mechanism is similar to the mechanism of the biosynthesis of PBG, the study of the Knorr mechanism will provide insight into the mechanism of the

PBG formation and serve as a guide to modeling the biosynthesis of PBG. A simple model of the two pathways proposed for the biosynthesis of PBG has been explored (Scheme 7.5) to shed light into the nature of the intermediates of the C-C and C-N bond formations and their inherent activation energies. The simple model of the biosynthesis of PBG mechanism in Scheme 7.5 utilizes the idea of the formation of Schiff bases to the keto groups of the ALA molecules. The two ALA are modeled using the same reactants for modeling the Knorr mechanism with the formation of Schiff bases to the keto groups of 2-propanone and aminoacetaldehyde (see **6** and **7** in Scheme 7.5). The reaction steps for the C-N (**I-a** and **II-b**) and C-C (**I-b** and **II-a**) bond formation for each pathway in Scheme 7.5 are explored.



Scheme 7.5 Proposed reaction model for the biosynthesis of PBG.

7.2 Computational Details

Density functional theory calculations were carried out using the Gaussian 03 suite of programs.[112] All geometry optimizations were performed using the B3LYP hybrid density functional and the 6-31+G(d) basis set. Harmonic vibrational frequencies and ZPVEs were calculated at the same level of theory. Relative energies were obtained by performing single-point calculations at the B3LYP level in conjunction with the 6-311+G(2df,p) basis set using the above optimized geometries and by including the ZPVE, i.e., B3LYP/6-311+G(2df,p)//B3LYP/6-31+G(d) + ZPVE. The entropy contributions to the free energies at 298.15 K were derived from the B3LYP/6-31+G(d) frequencies. Transition state (TS) optimizations were carried out using the Synchronous Transit-Guided Quasi-Newton (STQN) methods and the Berny algorithm. All TSs were confirmed to be true saddle points by having one, and only one, imaginary frequency. For each TS located, Intrinsic Reaction Coordinate (IRC) calculations were performed along the transition vector defined by the vibration mode of the imaginary frequency in order to confirm that the saddle point structure connected downhill the corresponding forward and backward minima. The IRC method allowed for the identification of the reaction intermediates and TS structures along the reaction path.

The Onsager[117] solvent model was employed to investigate the solvent effects. For the Onsager calculations, geometry optimizations and frequency analyses were performed at the B3LYP/6-31+G(d) level of theory on the optimized gas-phase structures with a dielectric constant of 78.39 for H_2O to model the Knorr reaction in aqueous media conditions. Relative energies were also obtained for the Onsager-solvated mechanism

using the same level of theory for the gas-phase mechanism, i.e., Onsager-B3LYP/6-311+G(2df,p)//Onsager-B3LYP/6-31+G(d) + ZPVE. The entropy contributions to the free energies in solution at 298.15 K were derived from the Onsager-B3LYP/6-31+G(d) frequencies. All energies reported are in kJ mol^{-1} and bond lengths in angstroms (Å).

7.3 Results and Discussion

7.3.1 Knorr Pyrrole Mechanism

The proposed mechanism of the Knorr reaction in Scheme 7.2 can be summarized in four main steps of the following order: hemiaminal formation, enamine formation, enamine cyclization and pyrrole ring formation. The computed energy profiles for the reaction steps are shown in Figures 7.1-7.8. Figures 7.1-7.4 include only the ZPVE correction, while the entropy contribution at 298.15 K is included in Figures 7.5-7.8. The discussion will be focused on the results of the ZPVE corrected case since the entropy contribution at 298.15 K produces similar energy profiles. The effect of solvent is shown by the dashed lines in Figures 7.1-7.8.

7.3.1.1 Gas Phase Calculations

Hemiaminal Formation. Initially, a water molecule forms a hydrogen bonding network between the 2-propanone and the amino-acetaldehyde in the gas-phase (Figure 7.1) via an NH...O bond of 2.022 Å and an O...HO bond of 1.844 Å, to form complex **1b** lying 32.8 kJ mol^{-1} lower in energy than the isolated reactants **1a**. The water bridge assists in the formation of a six-membered ring TS **1c** to facilitate the formation of the C-N bond between 2-propanone and amino-acetaldehyde with a high energy barrier of 109.7 kJ mol^{-1}. As the proton transfer is completed and the C-N bond of 1.496 Å is

formed, a tetrahedral complex **1d** is generated, lying 30.6 kJ mol^{-1} higher in energy than **1b**. Complete removal of the water molecule gives the tetrahedral intermediate **1e**, hemiaminal, plus a water molecule. The energy profile for the hemiaminal formation at 298.15 K in Figure 7.5 shows a similar energy profile. However, the energy barrier is higher (134.3 kJ mol^{-1}) and the reaction step becomes more endothermic (51.1 kJ mol^{-1}). Hence, this reaction step is slower at 298.15 K.

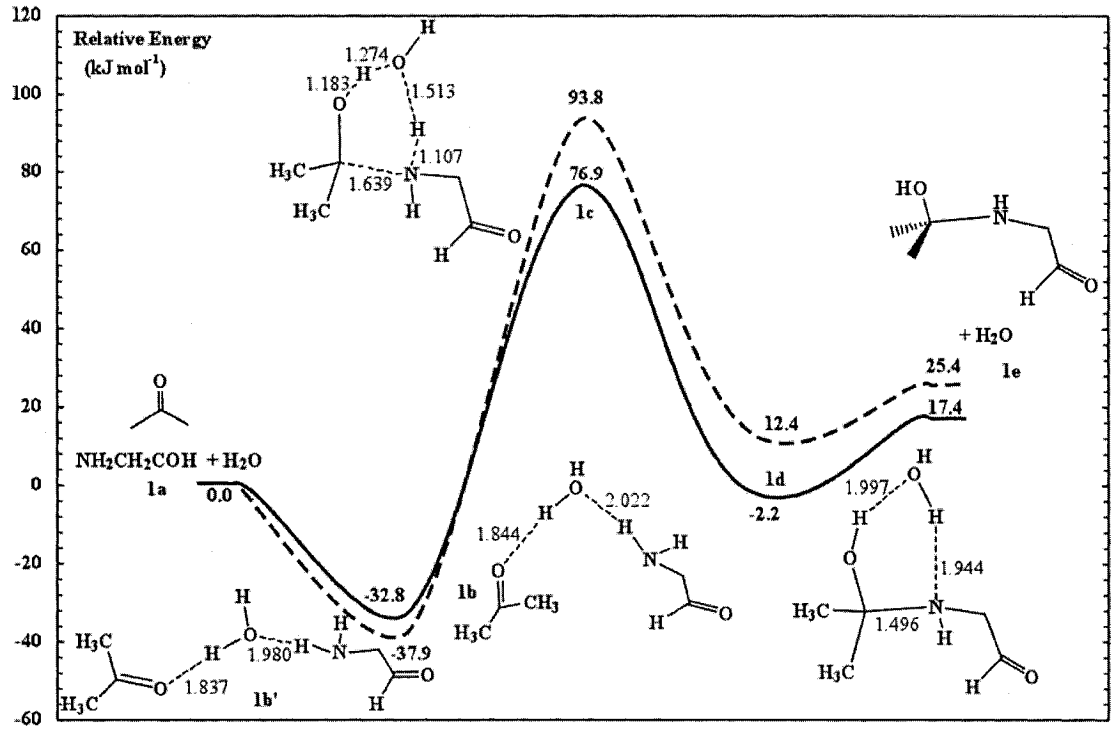


Figure 7.1 Schematic energy profile at 0 K for the hemiaminal formation step of the Knorr reaction in the gas phase (—) and in solution (– –).

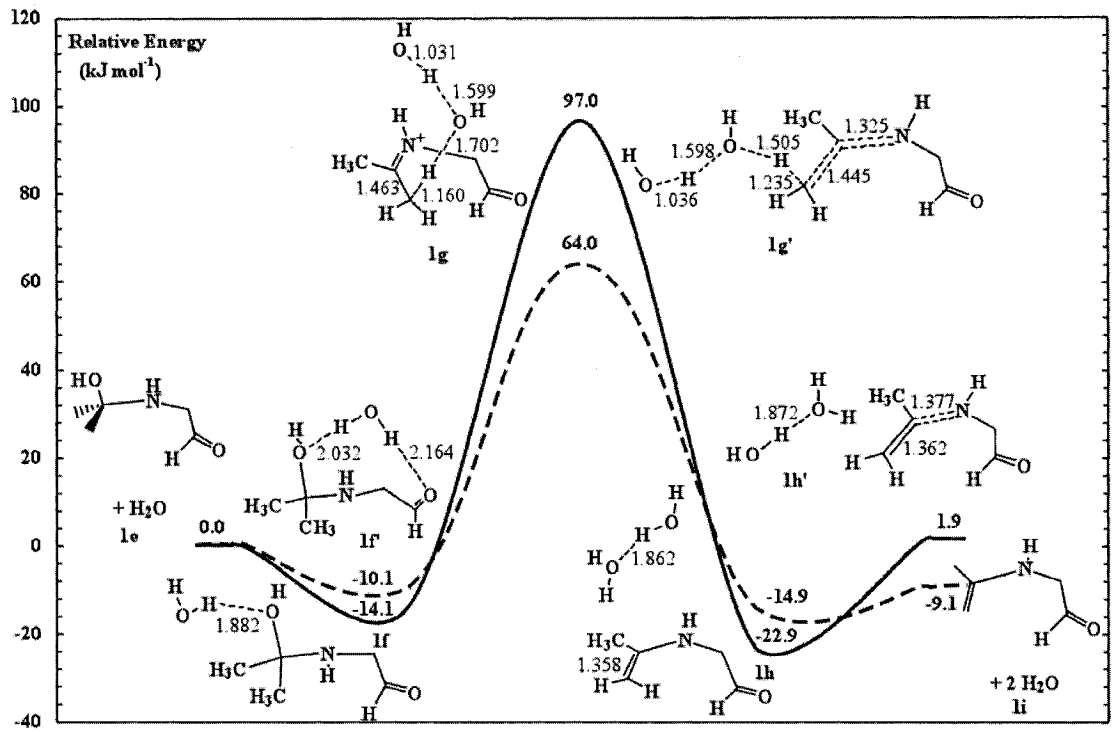


Figure 7.2 Schematic energy profile at 0 K for the enamine formation step of the Knorr reaction in the gas phase (—) and in solution (– –).

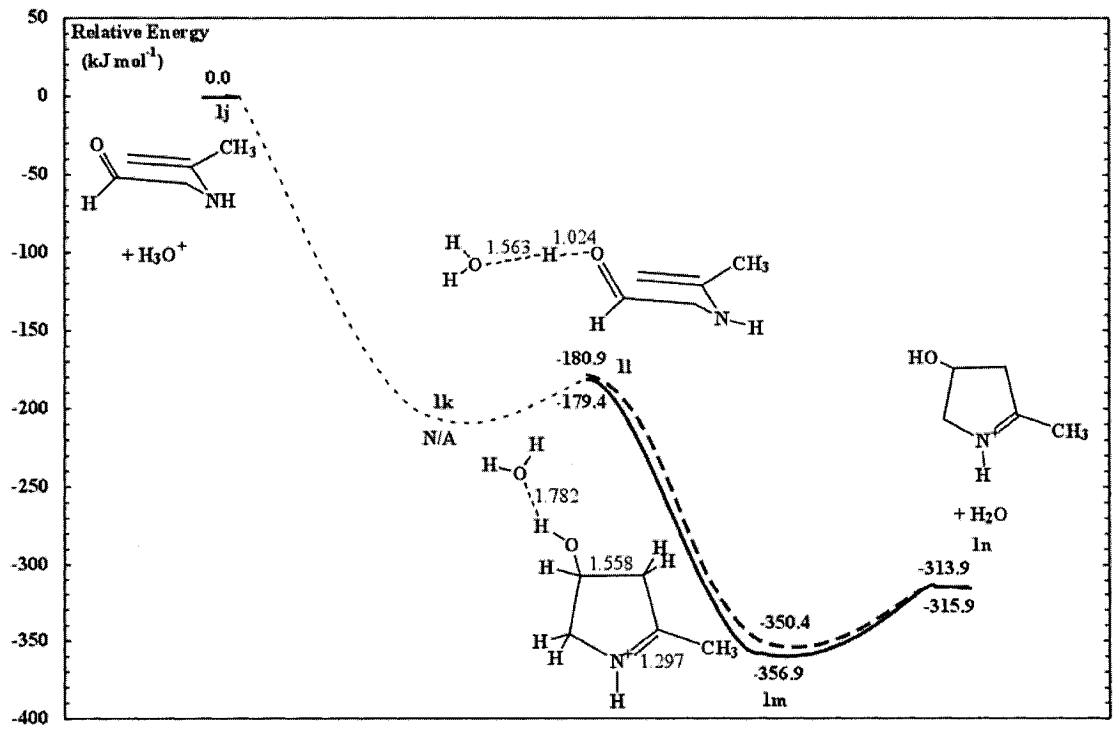


Figure 7.3 Schematic energy profile at 0 K for the enamine cyclization step of the Knorr reaction in the gas phase (—) and in solution (– –). (---) denotes missing PES line.

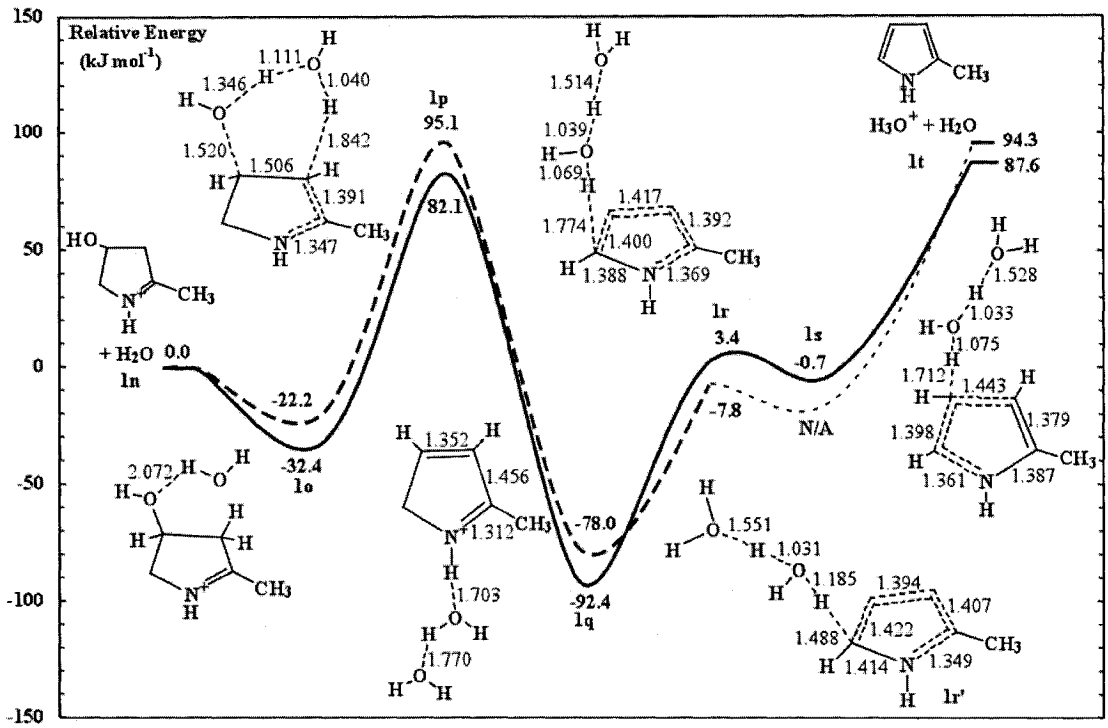


Figure 7.4 Schematic energy profile at 0 K for the pyrrole ring formation step of the Knorr reaction in the gas phase (—) and in solution (– –). (---) denotes missing PES line.

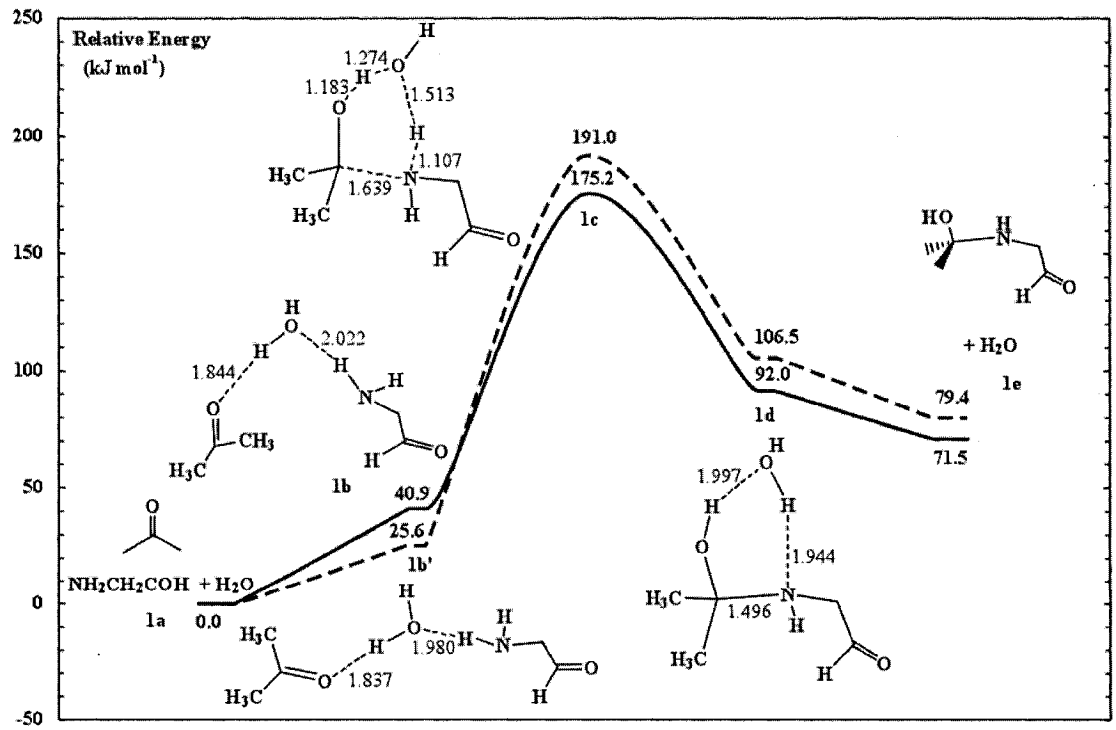


Figure 7.5 Schematic energy profile at 298.15 K for the hemiaminal formation step of the Knorr reaction in the gas phase (—) and in solution (– –).

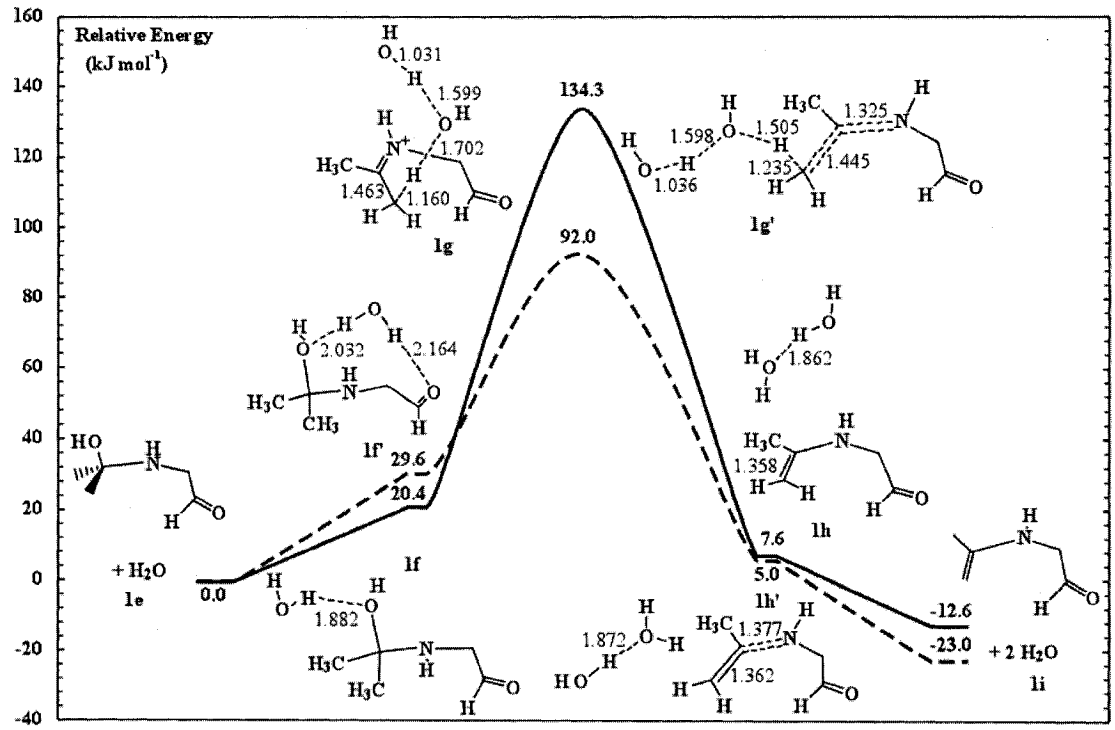


Figure 7.6 Schematic energy profile at 298.15 K for the enamine formation step of the Knorr reaction in the gas phase (—) and in solution (– –).

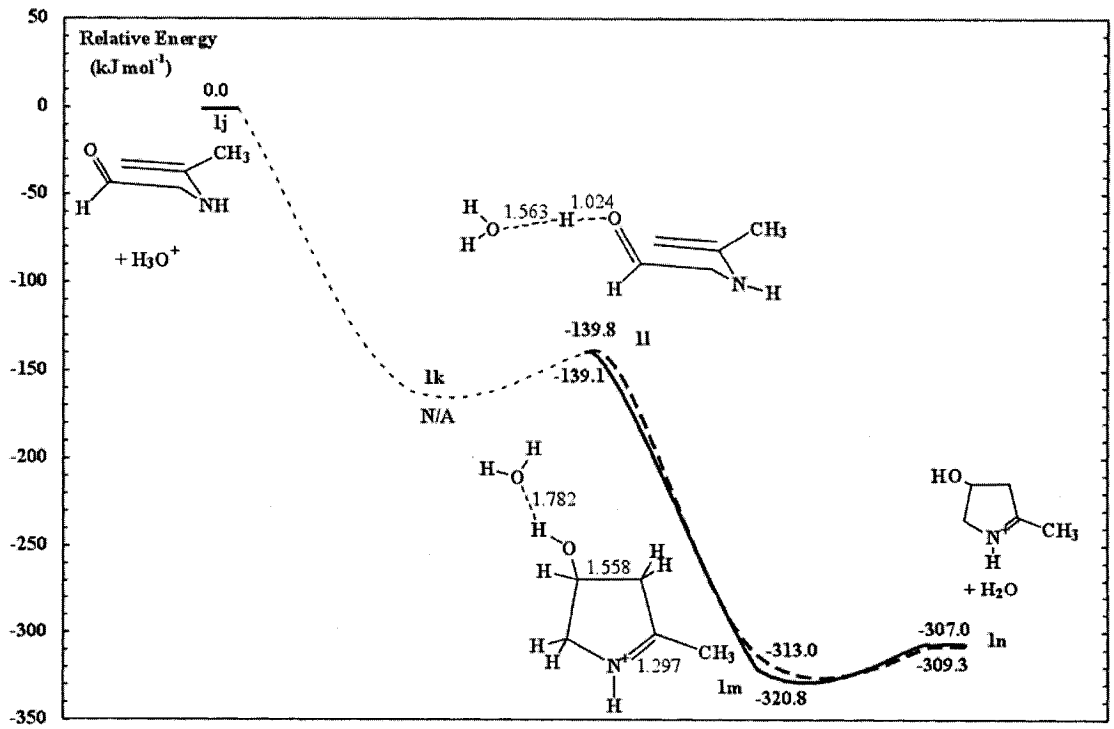


Figure 7.7 Schematic energy profile at 298.15 K for the enamine cyclization step of the Knorr reaction in the gas phase (—) and in solution (– –). (---) denotes missing PES line.

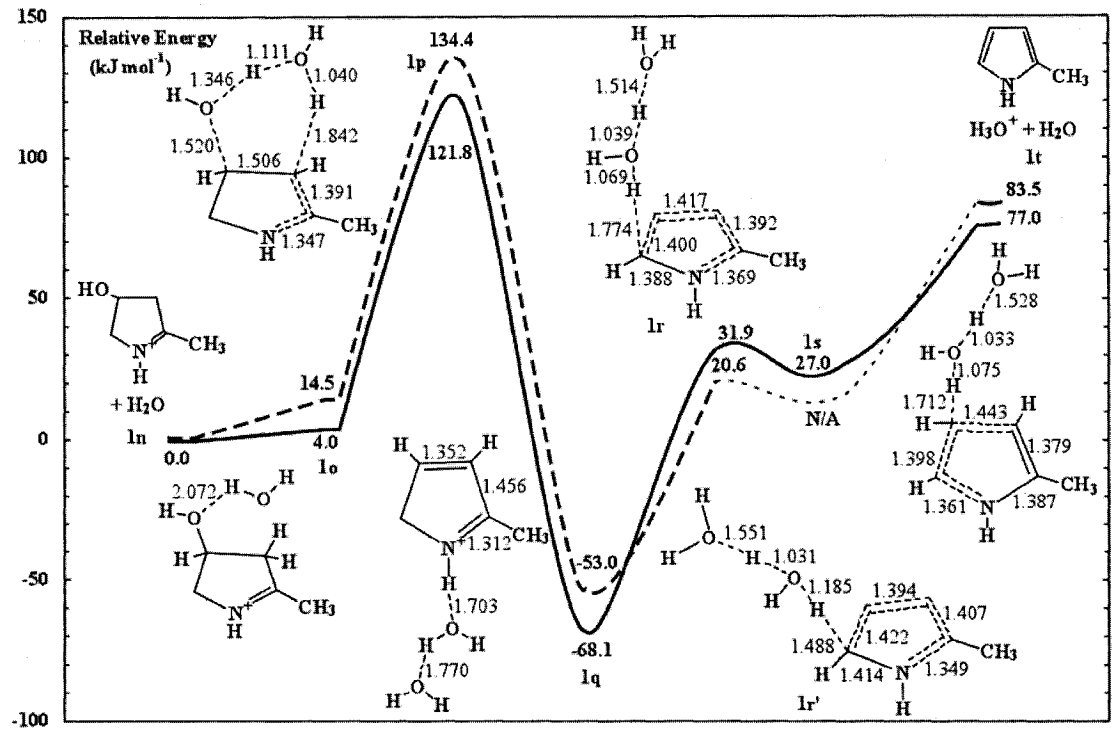


Figure 7.8 Schematic energy profile at 298.15 K for the pyrrole formation step of the Knorr reaction in the gas phase (—) and in solution (– –). (---) denotes missing PES line.

Enamine Formation. The hemiaminal **1e** formed can lose a water molecule to give the enamine intermediate. As presented in Figure 7.2, a water molecule forms a hydrogen bond of 1.882 Å with the hydroxyl oxygen-forming complex **1f** lying 14.0 kJ mol^{-1} lower in energy than the isolated reactants **1e**. The water molecule assists in bridging the distant hydroxyl group and the hydrogen to be abstracted to form the C=C bond through TS **1g**. The loss of a water molecule from complex **1f** to form the enamine intermediate **1h** proceeds with a high energy barrier of 111.1 kJ mol^{-1}. The formation of the enamine intermediate **1h** is exothermic by 8.8 kJ mol^{-1}.

Enamine Cyclization. Since the reaction is normally conducted under acidic conditions, the cyclization of the enamine is modeled using a hydronium ion, H_3O^+ as in Figure 7.3. All attempts to locate a reactant complex **1k** from TS **1l** failed. The IRC

calculation on TS **1l** directly leads to the formation of the cyclized ring on both sides. Moreover, all attempts to locate a reactant complex leads to the formation of the product. The reactant complex might be very close in energy to the TS **1l** making this region of the PES flat and difficult to converge at the level of theory used. A detailed analysis of the PES in a similar study by Bahmanyar and Houk[145] on the hydronium ion-catalyzed reaction of acetaldehyde with *N*-methylvinylamine in the gas phase revealed that the reaction has no barrier and leads directly to the formation of the product. In this study, a TS was located but not a reactant complex. This might be due to the difficulty of DFT to analyze the PES for systems with low energy barriers or flat regions.[122]

The separated reactants **1j** lie 180.9 kJ mol^{-1} higher in energy than TS **1l** due to charge separation. The TS **1l** forms a half-chair structure to bring about the cyclization forming the C-C bond. The carbonyl oxygen takes a proton from the hydronium ion to form the hydroxyl group in the cyclized product **1m**. The cyclized ring complex **1m** lies much lower (176.0 kJ mol^{-1}) in energy than TS **1l**. This reaction step is very favourable both at 0 K and 298.15 K and is highly exothermic. Therefore, the formation of the cyclized ring is the driving force for the completion of the reaction.

Pyrrole Ring Formation. The last step in the Knorr mechanism is the formation of the pyrrole ring with a loss of a water molecule and regeneration of the hydronium ion, H_3O^+, catalyst. This dehydration step occurs through the formation of two TSs as illustrated in Figure 7.4. The dehydration step is assisted by a water molecule that forms an OH...O hydrogen bond of 2.072 Å to the hydroxyl oxygen to yield complex **1o**. The dehydration proceeds via TS **1p** for the abstraction of the hydroxyl group with an energy barrier of 114.4 kJ mol^{-1}. The water bridge assists in the formation of a six-membered

ring TS **1p** to form the C=C bond by loss of a water molecule. The reaction proceeds downhill from TS **1p** to form complex **1q**, lying 60.0 kJ mol^{-1} lower in energy than complex **1o,** making this reaction step exothermic. Complex **1q** proceeds to form a second TS **1r** for the abstraction of the hydrogen on C5 to form the pyrrole ring and regenerate the hydronium ion catalyst with an energy barrier of 95.8 kJ mol^{-1}. The product complex **1s**, which shows the pyrrole ring formed but still bound to the water molecule and the hydronium ion, lies 91.7 kJ mol^{-1} higher in energy than **1q**, making this reaction step highly endothermic. Overall, the pyrrole ring formation step is endothermic by 30.7 kJ mol^{-1}. The separated products **1t** lie even higher in energy than **1q** due to charge separation of the separated products.

7.3.1.2 Onsager Solvent Calculations

Hemiaminal Formation. The optimized geometries in solution for the hemiaminal formation are very similar to those in the gas phase with the exception of the reactant complex **1b**. The reactant complex in solution denoted **1b′** forms a more planar complex **1b′** compared to the gas phase structure **1b** (Figure 7.1). Complex **1b′** forms a NH...O bond of 1.980 Å and an O...HO bond of 1.837 Å. The geometry of the TS **1c** is also very similar to the gas phase geometry with differences of less than 0.03 Å in bond lengths. The geometry of the hemiaminal complex **1d** is also similar to the gas phase geometry with differences of less than 0.02 Å. The energy barrier in solution is 131.7 kJ mol^{-1}, which is 22 kJ mol^{-1} higher than that in the gas phase due to the formation of a lower energy reactant complex **1b′**.

Enamine Formation. The optimized geometries for the enamine formation and loss of water molecule in solution differ from the gas phase geometries in the positioning of the water bridge network. As presented in Figure 7.2, a water molecule forms a long hydrogen bond of 2.032 Å with the hydroxyl oxygen and another hydrogen bond of 2.164 Å with the carbonyl oxygen to form complex **1f'**, lying 10.1 kJ mol^{-1} lower in energy than the isolated reactants **1e**. The TS **1g'** formed in solution is a late TS where the hydroxyl group has been abstracted and is assisting in the abstraction of the hydrogen to form the C=C bond. The energy barrier for the loss of a water molecule to form the enamine is 74.1 kJ mol^{-1}, which is 37 kJ mol^{-1} lower than that in the gas phase. The formation of the enamine intermediate **1h'** is exothermic by 4.7 kJ mol^{-1}.

Enamine Cyclization. The effect of solvent for the enamine cyclization is not significant (Figure 7.3). The optimized geometries in solution for the enamine cyclization are very similar to those in the gas phase with differences less than 0.07 Å in bond lengths. Similarly as in the gas phase, all attempts to locate a reactant complex **1k** from TS **1l** failed to converge. Looking at the optimization steps before convergence failure, the hydronium ion moves away from the enamine intermediate and then the complex falls apart and fails. The IRC calculation on TS **1l** also shows that the PES for the reactant complex **1k** is shallow.

The separated reactants **1j** lie 179.4 kJ mol^{-1} higher in energy than TS **1l** due to charge separation. The cyclized ring complex **1m** lies 171.0 kJ mol^{-1} lower in energy than TS **1l** in solution. The relative energies in solution for this step are not significantly different from those obtained in the gas phase. Similarly to the gas phase results, this reaction step is very favourable both at 0 K and 298.15 K and is highly exothermic.

Pyrrole Ring Formation. The solvent effects are not significant for the last step in the Knorr mechanism (Figure 7.4). The optimized geometries in solution for the loss of the water molecule in the pyrrole ring formation step are very similar to those in the gas phase with differences less than 0.03 Å in bond lengths. Complex **1o** forms an OH...O hydrogen bond of 2.007 Å to the hydroxyl oxygen. The energy barrier for the loss of the water molecule in solution is 117.3 kJ mol^{-1}, which is not significantly different from that of gas phase (114.4 kJ mol^{-1}). Complex **1q** lies 55.9 kJ mol^{-1} lower in energy than complex **1o**, making this reaction step exothermic. The optimized geometry of the TS in solution denoted **1r'** differs from that of the gas phase in the positioning of the water molecules. The energy barrier in solution is 70.2 kJ mol^{-1} which is 25.6 kJ mol^{-1} lower than that of the gas phase. Attempts to find the product complex **1s** in solution failed. Similarly as in reactant complex **1k**, complex **1s** could not be converged. Looking at the optimization steps before convergence failure, the H_3O^+-H_2O complex separates away from the pyrrole ring and then the complex falls apart with convergence problem. The convergence failure might also be due to the fact that this region of the PES is shallow.

7.3.2 PBG Model Mechanism

The energy profile for the simple model of the biosynthesis of PBG mechanism in Scheme 7.5 has been explored for the reaction steps of the C-N (**I-a** and **II-b**) and C-C (**I-b** and **II-a**) bond formation for each pathway. The Knorr-like mechanism, Path **I**, proceeds first with the formation of the C-N bond (**I-a**) through hemiaminal formation to

form an enamine intermediate. Cyclization of the enamine intermediate then forms the C-C bond (**I-b**) of the pyrrole ring. Path **II** proceeds with the formation of the C-C (**II-a**) bond by an enamine attack on the iminium ion of **7** (Scheme 7.5). Cyclization of the ring then takes place through a nucleophilic attack of the amine on the carbon of the Schiff base to form the C-N bond (**II-b**) of the pyrrole ring. The computed energy profiles for the reaction steps are shown in Figures 7.9-7.14. Figures 7.9-7.11 include only the ZPVE correction, while the entropy contribution at 298.15 K is included in Figures 7.12-7.14. The discussion will be focused on the results of the ZPVE corrected case since the entropy contribution at 298.15 K shows similar energy profiles. The effect of Onsager solvent is shown by the dashed lines in Figures 7.9-7.14.

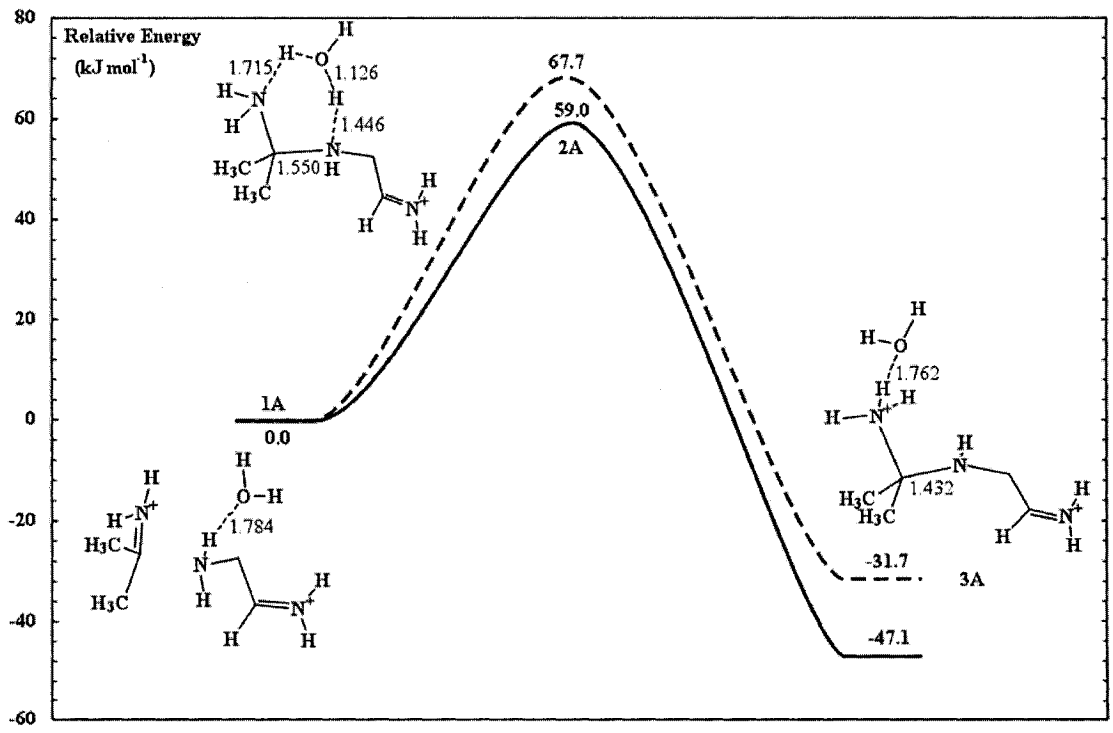


Figure 7.9 Schematic energy profile at 0 K for reaction step **I-a** of the PBG model mechanism in the gas phase (—) and in solution (– –).

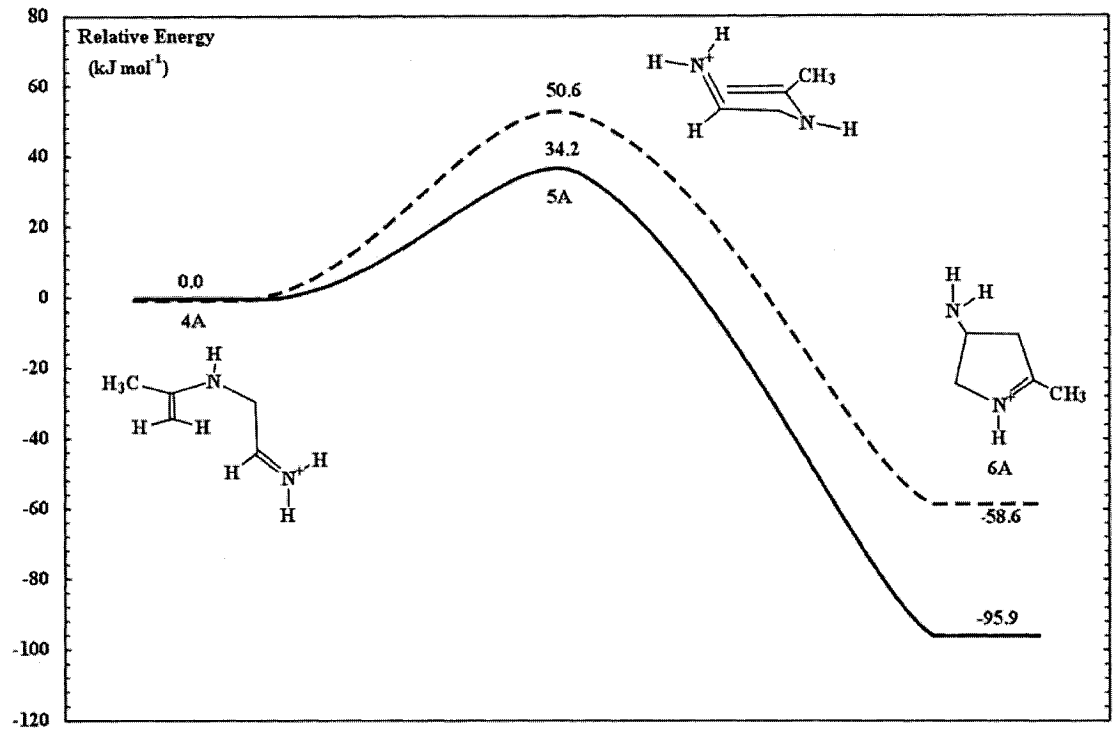


Figure 7.10 Schematic energy profile at 0 K for reaction step **I-b** of the PBG model mechanism in the gas phase (—) and in solution (– –).

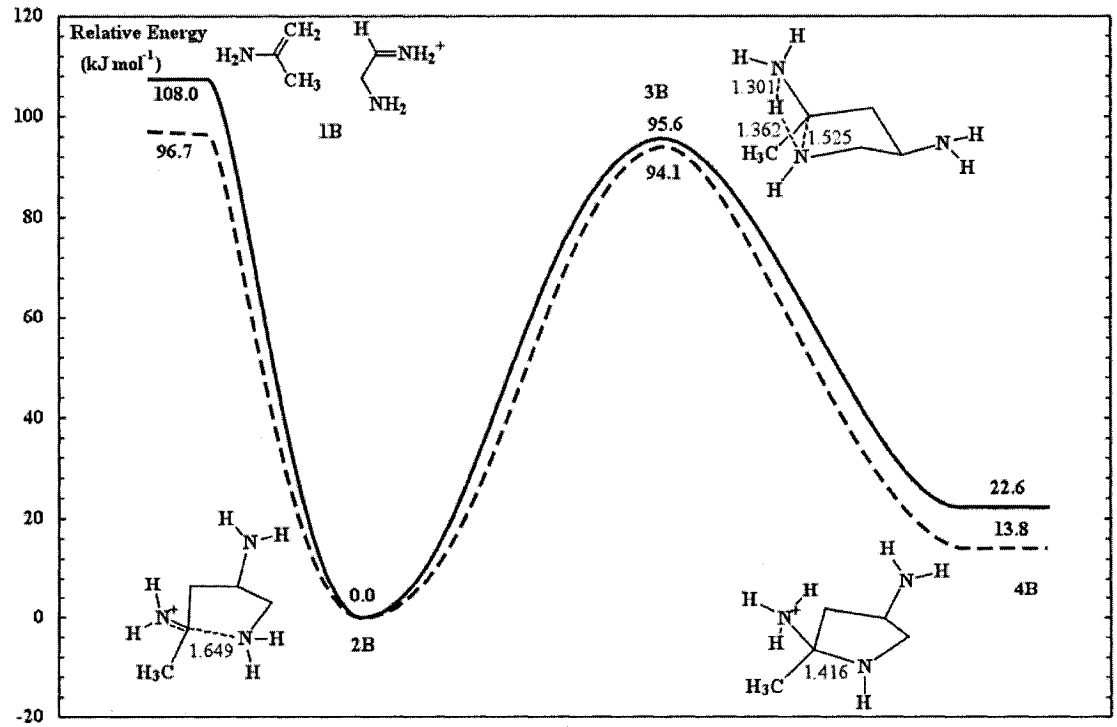


Figure 7.11 Schematic energy profile at 0 K for reaction steps **I-a** and **II-b** of the PBG model mechanism in the gas phase (—) and in solution (– –).

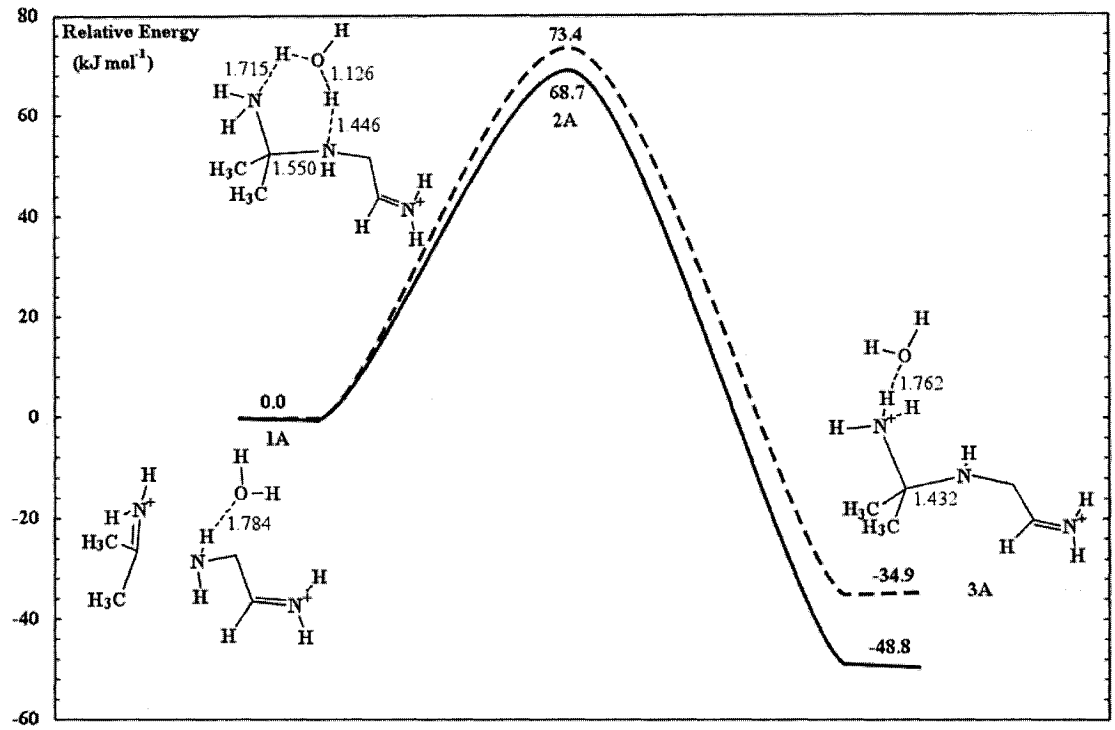


Figure 7.12 Schematic energy profile at 298.15 K for reaction step **I-a** of the PBG model mechanism in the gas phase (—) and in solution (– –).

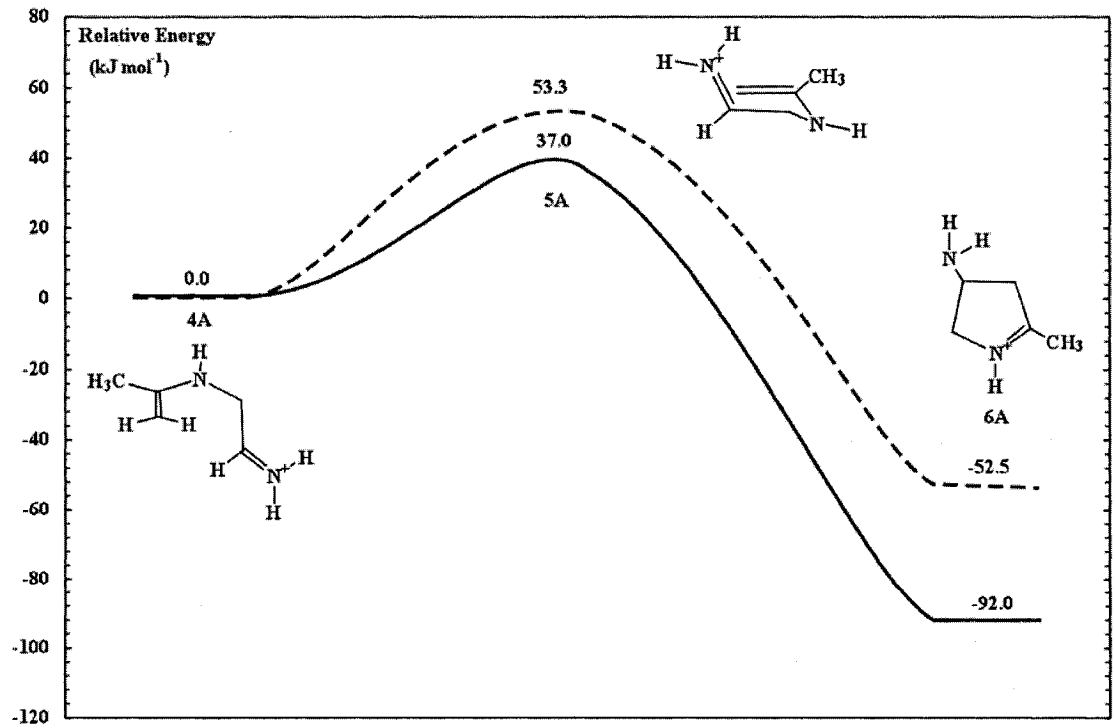


Figure 7.13 Schematic energy profile at 298.15 K for reaction step **I-b** of the PBG model mechanism in the gas phase (—) and in solution (– –).

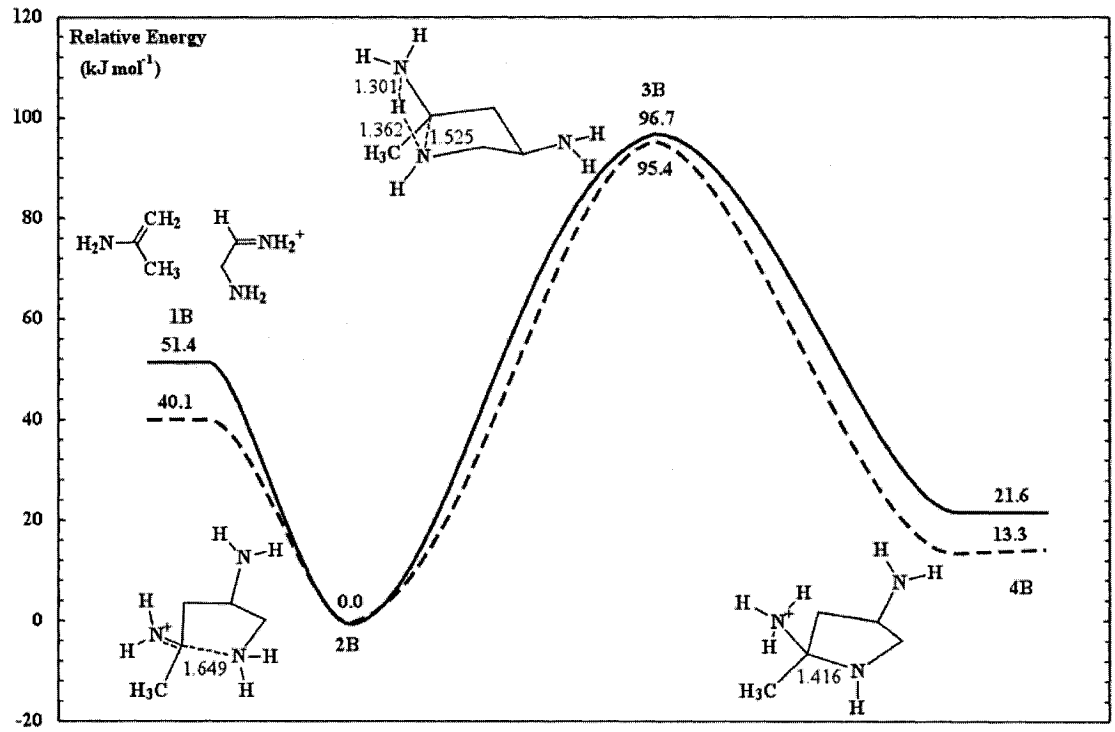


Figure 7.14 Schematic energy profile at 0 K for reaction steps **I-a** and **II-b** of the PBG model mechanism in the gas phase (—) and in solution (– –).

7.3.2.1 Gas Phase Calculations

Reaction step I-a. The first step of the Knorr-like mechanism is the formation of the hemiaminal intermediate to form the C-N bond of the pyrrole ring. The computed energy profile at 0 K for the hemiaminal formation is presented in Figure 7.9. Initially, a water molecule forms a NH...O hydrogen bond of 1.784 Å with the amine moiety to form complex **1A**. The water bridge assists in the formation of a six-membered ring TS **2A** to facilitate the formation of the C-N bond as the proton transfers from the water molecule to the iminium ion nitrogen and from the N of the amine to the water molecule, with a barrier of 59.0 kJ mol^{-1}. When the proton transfer is completed, the C-N bond of 1.432 Å is formed and a tetrahedral complex **3A** is generated, lying 47.1 kJ mol^{-1} lower in

energy than **1A**. This reaction step is favourable and is exothermic by 47.1 kJ mol^{-1}, a result that is opposite to the hemiaminal formation in the Knorr mechanism (Figure 7.1).

The reaction step of hemiaminal formation is similar to the hemiaminal formation in the Knorr mechanism (see Figure 7.1) and the Paal-Knorr mechanism in Chapter 6 (see Figure 6.1). The energy barriers for this step are comparable in the three mechanisms. The energy barrier is highest for the Knorr mechanism (109.7 kJ mol^{-1}) and lowest for the PBG model mechanism (59.0 kJ mol^{-1}). The energy barrier for the Paal-Knorr mechanism is 74.9 kJ mol^{-1}.

Reaction step I-b. An enamine intermediate **4A** can be formed from the hemiaminal intermediate **3A** through deprotonation of one of the methyl groups and release of an NH_3 molecule. The cyclization of the enamine intermediate **4A** forms the C-C bond of the pyrrole ring. The computed energy profile for the cyclization of **4A** at 0 K is presented in Figure 7.10. From the enamine intermediate **4A**, a TS **5A** forms a half-chair structure to bring about the cyclization forming the C-C bond with a small energy barrier of 34.2 kJ mol^{-1}. The cyclized ring complex **6A** lies much lower in energy (95.9 kJ mol^{-1}) than the enamine complex **4A**. Similar to the enamine cyclization step in the Knorr mechanism (Figure 7.3), this reaction step is also very favourable both at 0 K and 298.15 K and is highly exothermic.

Reaction step II-a. In this alternative mechanism, a deprotonation of **6** (Scheme 7.5) forms an enamine intermediate which attacks the iminium ion of **7** to form the C-C bond of the pyrrole ring. All attempts to model the reaction step **II-a** failed to locate a reactant complex and a TS structure. All attempts to locate a reactant complex or TS led directly to the formation of the product **2B** (Figure 7.11) for the C-C bond formation.

This might be due to the difficulty of using DFT to analyze the PES for systems with low energy barriers or flat regions.[122] The product complex **2B** lies 108.0 kJ mol^{-1} lower in energy than the isolated reactants **1B**. Hence, the formation of **2B** is favourable. This reaction step is also similar to the enamine cyclization step in the Knorr mechanism (Figure 7.3) which is also highly exothermic.

Reaction step II-b. From step **II-a**, the formation of the C-C bond of the pyrrole ring produces complex **2B** (Figure 7.11). In this step, the cyclization of the ring takes place by the formation of the C-N bond of pyrrole. A half-chair TS **3B** is produced from complex **2B** for the formation of the C-N bond and the simultaneous proton transfer from the amine to the nitrogen of the iminium ion. The formation of TS **3B** proceeds with a high energy barrier of 95.6 kJ mol^{-1} compared to the small energy barrier (34.2 kJ mol^{-1}) for the cyclization of the enamine complex **4A** (Figure 7.10). The cyclized product complex **4B** formed from TS **3B** lies 22.6 kJ mol^{-1} higher in energy than **2B**. Hence, the reaction step **II-b** for the cyclization is endothermic, whereas reaction step **I-b** for the cyclization of the enamine **4A** is exothermic.

This cyclization step is similar to the hemiaminal cyclization (Figure 6.2) and enamine cyclization (Figure 6.5) of the Paal-Knorr mechanism studied in Chapter 6. The energy barriers for this type of cyclization are high. Both the hemiaminal (146.4 kJ mol^{-1}) and enamine cyclizations (161.7 kJ mol^{-1}) in the Paal-Knorr mechanism proceed with high energy barriers. This is similar to the cyclization step **II-b**, where the energy barrier is the highest (95.6 kJ mol^{-1}) compared to the other steps in the PBG model mechanism.

7.3.2.2 Onsager Solvent Calculations

Reaction step I-a. The optimized geometries in solution for the hemiaminal formation are very similar to those in the gas phase with differences of less than 0.08 Å in bond lengths. The energy barrier in solution is 67.7 kJ mol^{-1}, which is higher than that in the gas phase (59.0 kJ mol^{-1}). The reaction step is still favourable in solution and is exothermic by 31.7 kJ mol^{-1} (Figure 7.9).

Reaction step I-b. The optimized geometries in solution for the cyclization of enamine **4A** are similar to those in the gas phase with differences of less than 0.4 Å in bond lengths. Moreover, some of the dihedrals of the structures of **4A** and **5A** are slightly larger in solution. The energy barrier in solution is 50.6 kJ mol^{-1}, which is higher than that in the gas phase (34.2 kJ mol^{-1}). The cyclized ring complex **6A** lies 58.6 kJ mol^{-1} lower in energy than **4A** (Figure 7.10). The reaction step is still favourable in solution and is exothermic by 58.6 kJ mol^{-1}.

Reaction step II-a. Similar to the gas phase results, several attempts to model the enamine attack on the iminium ion **7** (Scheme 7.5) to form the C-C bond of the pyrrole ring in solution led directly to the formation of the product complex **2B** in Figure 7.11. The product complex **2B** lies 96.7 kJ mol^{-1} lower in energy than the isolated reactants **1B**. Hence, the formation of **2B** is favourable.

Reaction step II-b. The optimized geometries in solution for the ring cyclization to form the C-N bond of pyrrole are very similar to those in the gas phase. Differences are less than 0.01 Å in the forming and breaking bonds. The energy barrier in solution is 94.1 kJ mol^{-1}, which is about the same as in the gas phase (95.6 kJ mol^{-1}). The cyclized

product complex **4B** lies 13.8 kJ mol^{-1} higher in energy than **2B** (Figure 7.11). The reaction step is less endothermic in solution by 8.8 kJ mol^{-1}.

7.4 Conclusions

The Knorr pyrrole reaction mechanism has been investigated using the B3LYP density functional theory method. In the model reaction, a water molecule assists in the formation of TSs to facilitate proton transfer in the reaction. The Knorr mechanism involves high energy barrier steps for the hemiaminal formation, enamine formation and the pyrrole ring formation. The enamine cyclization step is highly exothermic and is the driving force for the completion of the reaction. The inclusion of the solvent effects using the Onsager model did not significantly lower the energy barriers except for the enamine formation. In fact, for the hemiaminal formation, the energy barrier was higher in solution. The inclusion of solvent significantly lowered the energy barrier for the enamine formation from 111.1 to 74.1 kJ mol^{-1}.

A simple mechanistic model for the PBG pyrrole synthesis reaction steps of the C-N (**I-a** and **II-b**) and C-C (**I-b** and **II-a**) bond formation of each pathway has been explored. Path **I**, a Knorr-like mechanism, proceeds with the formation of the C-N followed by the C-C bond of the pyrrole ring. The two steps **I-a** and **I-b** in Path **I** proceed with low energy barriers (59.0 and 34.2 kJ mol^{-1}, respectively) and both steps are exothermic; whereas, in Path **II**, the reaction step **II-a** for the formation of the C-C bond seems to be favoured. The reaction step **II-b** for the ring cyclization to form the C-N bond must pass through a high energy barrier (94.1 kJ mol^{-1}) and is endothermic. The

inclusion of the solvent effects using the Onsager model did not lower the energy barriers. In fact, a slightly higher energy barrier was obtained for step **I-a** and **II-b**. The results of the simple mechanistic model of the PBG were comparable to similar reaction steps in the Paal-Knorr mechanism (Chapter 6) and the Knorr mechanism. These preliminary results suggest that the preferred mechanism for the PBG pyrrole synthesis is through the Knorr-like mechanism, Path **I**. Hence, the order of bond formation in the PBG pyrrole synthesis proceeds through the formation of the C-N bond first followed by the C-C bond of the pyrrole ring.

Chapter 8: Conclusion

The computational studies in this thesis have focused on the NMR chemical shifts of *N*-EWG pyrroles and reaction mechanisms of pyrrole synthesis. This chapter summarizes the conclusions of the thesis and highlights some possible directions for future research.

8.1 NMR Data of *N*-EWG Pyrroles

The current study on the nuclear shielding constants of small nitrogen-containing molecules (Chapter 4) presented a validated computational scheme to study the ^{15}N and ^{13}C shielding constants of *N*-substituted pyrrole (Chapter 5). The reliability of density functional theory in predicting NMR shielding constants has been demonstrated by the good agreement between the calculated values and the experimental data available for the *N*-EWG pyrrole series. The computations of the ^{15}N and ^{13}C chemical shifts of pyrroles have also proven to be a useful tool for the interpretation of the experimental NMR data and deduction of the origin of the observed chemical shift trend. The comprehensive analysis of the NMR parameters using the natural chemical shielding (NCS) analysis illustrates the usefulness and power of theoretical predictions in complementing and interpreting experimental NMR data. The ^{15}N chemical shift trend observed is shown to arise from variations of the paramagnetic shift contribution to the chemical shift. NCS

analysis shows that the σ(N5-R)-π* transitions and the sum of the σ(C1-N5)- π* and σ(C4-N5)- π* transitions account for the nitrogen chemical shift trend observed in the pyrroles.

The *N*-EWG substituents selected for the study in Chapter 5 was chosen on the basis of the available experimental data. Given the limited set of EWG substituents studied it is perhaps not surprising that no correlation was found between the ^{15}N chemical shift and the EWG strength. A useful extension of this work would be to improve the range of substituents by introducing systematic variations in the EWG subtituents. For example, a series of alkyl groups differing in their EWG strength (CH_3, CH_2F, CH_2Cl, CH_2I, ...) could be used. Similarly the carbonyl groups could be varied ($COCH_3$, $COCH_2F$, $COCH_2Cl$, $COCH_2I$, $COCF_3$, ...). For the ester groups, the set could be extended to include CO_2CH_3, CO_2CH_2F, CO_2CF_3, CO_2CH_2Cl, and so on. Finally, the sulfonyl groups could be varied to include SO_2CH_3, SO_2CH_2F, SO_2CHF_2, SO_2CF_3 and so on.

The deviation of the calculated NMR chemical shifts from the experimental values may be mainly due to the limitations of the computational method; basis set incompleteness, lack of exact treatment of electron correlation and ignoring van der Waals, intermolecular and molecule-solvent interactions. Therefore, the investigation of solvent effects using continuum polarizable models might be fruitful in increasing the accuracy of the computed chemical shifts of pyrroles. Furthermore, the accuracy of the calculations of the chemical shift values using the gas phase structure of a single conformation limits the accuracy of the results since the experimental chemical shift values are a weighted average of the different conformers present in solution at room

temperature. Further investigation of nitrogen chemical shifts in pyrrole compounds could involve *ab initio* molecular dynamics simulations of the solvent effects, intermolecular interaction effects and molecular motion. However, such a study will be computationally expensive and is beyond the scope of this thesis.

8.2 Reaction Mechanisms of Pyrrole Synthesis

The mechanistic studies of the Paal-Knorr (Chapter 6), Knorr and the simple model for the biosynthesis of PBG (Chapter 7) have provided insight into the reactions of pyrrole synthesis. The advantage of computational chemistry has been demonstrated by the quantitative predictions of molecular structures, energy profiles and activations energies. Experimental investigations often provide only indirect information about the reaction mechanism, whereas a complete picture of the reaction mechanisms of interest can be obtained with the tools of computational chemistry. Hence, computational methods provide a powerful tool to predict and complement experimental evidence to probe the mechanism of chemical reactions.

The studies of both the Paal-Knorr and Knorr pyrrole syntheses represent the first reported computational studies of these classic reactions. As described in Chapter 6, the energetically preferred pathway for the Paal-Knorr reaction occurs via the cyclization of the hemiaminal intermediate, a result that is consistent with experimental evidence. The reaction was investigated in aqueous media via the Onsager and PCM models. Possible extensions of this work are to explore the mechanism under acidic conditions and to examine the effects of substituents on the relative energetics of the reaction pathway.

The study of the Knorr mechanism (Chapter 7) in acidic conditions provided insight into the energetics involved. Moreover, the Knorr mechanism provided insight into the mechanism of the biosynthesis of PBG pyrrole. The simple model used (reaction of amino-acetaldehyde and 2-propanone) for the Knorr reaction mechanism produced high energy barriers except for the highly exothermic cyclization step of the reaction. The Knorr synthesis and the simple model of the PBG pyrrole synthesis were studied using the simplest Onsager solvent model. Hence, a more reliable solvent model, such as the PCM model, should be explored to accurately model the solvent effects.

The simplest reaction mechanism in Scheme 7.2 (Chapter 7) was investigated, where $R^1 = R^2 = R^3 = H$ and $R^4 = CH_3$. A possible extension of this work is to explore the effect of different substituents on the energetics of the Knorr reaction mechanism. Other useful Knorr syntheses that produce better yields of pyrroles are obtained when R^3 and R^1 are acyl or alkoxycarbonyl groups,[132] hence it would be of interest to study the effect of these substituents on the energetics of the reaction mechanism.

The mechanisms studied in Chapter 6 and 7 involve only the use of one explicit water molecule to model the reaction. Previous work by Delphine et al.[111] showed that the inclusion of two and three water molecule yields even lower energy barriers compared to the inclusion of one water molecules for the nucleophilic attack of ammonia on formaldehyde. Hence, further studies of the reaction mechanisms of pyrrole syntheses with the inclusion of two and three water molecules will be useful. Furthermore, a useful extension of this work would be to consider the tunneling effects since they are important for proton transfer reactions.

The study of the simple model of the biosynthesis of PBG (Chapter 7) provided insight into the long debated reaction mechanism. The simple model used suggests that the preferred mechanism for the biosynthesis of the PBG pyrrole proceeds by the formation of the C-N bond first followed by the C-C bond of the pyrrole ring through a mechanism similar to that proposed for the Knorr pyrrole synthesis. The model investigated for the biosynthesis of PBG is a crude simplification of the original molecules involved and its enzymatic environment. Therefore, a more realistic model of the biosynthesis of PBG will have to be investigated to reveal its mechanistic action. The realistic model should involve the active site and the surrounding conserved residues of the enzyme that affect the mechanistic action and should allow for the study of the pH effects. This larger model could be studied using the quantum mechanics/molecular mechanics (QM/MM) methods. In this method, a small subset (substrate and catalytic residues) of the active site, where bond breaking and forming occurs, should be studied using QM methods. The rest of the model (enzyme structure and explicit solvent molecules) is treated with the classical MM methods using force fields to make the large calculation feasible.

References

(1) Jones, R. A.; Bean, G. P. *The Chemistry of Pyrroles*; Academic Press: London, 1977.

(2) Battersby, A. R. *Nat. Prod. Rep.* **2000**, *17*, 507.

(3) Battersby, A. R.; Leeper, F. J. *Chem. Rev.* **1990**, *90*, 1261.

(4) Frere, F.; Schubert, W.-D.; Stauffer, F.; Frankenberg, N.; Neier, R.; Jahn, D.; Heinz, D. W. *J. Mol. Biol.* **2002**, *320*, 237.

(5) Goodwin, C. E.; Leeper, F. J. *Org. Biomol. Chem.* **2003**, *1*, 1443.

(6) Abell, A. D.; Nabbs, B. K.; Battersby, A. R. *J. Am. Chem. Soc.* **1998**, *120*, 1741.

(7) Scott, A. I. Advances in Detailed Reaction Mechanisms: Mechanisms of Biological Importance; Coxon, J. M., Ed.; JAI Press: Greenwich, 1992; pp 189.

(8) Thompson, A.; Gao, S.; Modzelewska, G.; Hughes, D. S.; Patrick, B.; Dolphin, D. *Org. Lett.* **2000**, *2*, 3587.

(9) Abell, A. D.; Litten, J. C. *Tetrahedron Lett.* **1992**, *33*, 3005.

(10) Hohenberg, P.; Kohn, W. *Phys. Rev. A* **1964**, *136*, 864.

(11) Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133.

(12) *A Chemist's Guide to Density Functional Theory*; 2nd ed.; Koch, W.; Holthausen, M. C., Eds.; Wiley-VCH: Weinheim, 2001.

(13) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Macmillan Publising Co., Inc: New York, 1982.

(14) *Quantum Chemistry*; 5th ed.; Levine, I. N., Ed.; Prentice Hall: New Jersey, 2000.

(15) Born, O.; Oppenheimer, J. R. *Ann. Physik* **1926**, *79*, 361.

(16) McQuarrie, D. A. *Quantum Chemistry*; University Science books, 1983.

(17) Fischer, C. F. *The Hartree-Fock Method for Atoms: A Numerical Approach*; Wiley Interscience Publication, 1977.

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

(19) Springborg, M. *Methods of Electronic-Structure Calculations from Molecules to Solids*; John Wiley & Son Ltd.: Chichester, 2000.

(20) *Exploring Chemistry with Electronic Structure Methods*; 2nd edition ed.; Foresman, J. B.; Frisch, A., Eds.; Gaussian, Inc.: Pittsburgh,PA, 1996.

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

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

(23) Merrill, G. N.; Gronert, S.; Kass, S. R. *J. Phys. Chem. A.* **1997**, *101*, 201.

(24) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(25) de Dios, A. C. *Prog. Nucl. Mag. Res. Sp.* **1996**, *29*, 229.

(26) Proctor, W. G.; Yu, F. C. *Phys. Rev.* **1951**, *81*, 20.

(27) Arnold, J. T.; Dharmatti, S. S.; Packard, M. E. *J. Chem. Phys* **1951**, *19*, 1608.

(28) de Dios, A. C.; Oldfield, E. *Solid State Nucl. Mag. Res.* **1996**, *6*, 101.

(29) Helgaker, T.; Jaszunski, M.; Ruud, K. *Chem. Rev.* **1999**, *99*, 293.

(30) Ramsey, N. F. *Phys. Rev.* **1950**, *78*, 699.

(31) Hore, P. J. *Nuclear Magnetic Resonance*; Oxford University Press: Oxford, 1995.

(32) Dahn, H. *J. Chem. Educ.* **2000**, *77*, 905.

(33) Carrington, A.; Mclachlan, A. D. *Introduction to Magnetic Resonance with Application to Chemistry and Chemical Physics*; Harper & Row Inc.: New York, 1967.

(34) Harris, R. K. *Nuclear Magnetic Resonance Spectroscopy*; Longman Limited: England, 1997.

(35) *Multinuclear NMR*; Jameson, C. J.; Mason, J., Eds.; Plenum Press: New York, 1987.

(36) *Encyclopedia of Nuclear Magnetic Resonance*; Pulay, P.; Hinton, J. F., Eds.; John Wiley: New York, 1996, pp 4334.

(37) Kutzelnigg, W. *Isr. J. Chem.* **1980**, *19*, 193.

(38) *Encyclopedia of Nuclear Magnetic Resonance*; Kutzelnigg, W.; Fleischer, U.; Wüllen, C. V., Eds.; John Wiley: New York, 1996, pp 4284.

(39) Hansen, A. E.; Bouman, T. D. *J. Chem. Phys* **1985**, *82*, 5035.

(40) *Encyclopedia of Nuclear Magnetic Resonance*; Hansen, A. E.; Bilde, M., Eds.; John Wiley: New York, 1996, pp 4292.

(41) Keith, T. A.; Bader, R. F. W. *Chem. Phys. Lett.* **1992**, *194*, 1.

(42) Bouman, T. D.; Hansen, A. E. *Chem. Phys. Lett.* **1990**, *175*, 292.

(43) Gauss, J. *Chem. Phys. Lett.* **1992**, *191*, 614.

(44) Wüllen, C. V.; Kutzelnigg, W. *Chem. Phys. Lett.* **1993**, *205*, 563.

(45) Gauss, J. *J. Chem. Phys.* **1993**, *99*, 3629.

(46) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, *104*, 5497.

(47) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1998**, *298*, 113.

(48) Wilson, P. J.; Amos, R. D.; Handy, N. C. *Mol. Phys.* **1999**, *97*, 757.

(49) Bienati, M.; Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1999**, *311*, 69.

(50) Wilson, P. J.; Amos, R. D.; Handy, N. C. *Chem. Phys. Lett.* **1999**, *312*, 475.

(51) Helgaker, T.; Wilson, P. J.; Amos, R. D.; Handy, N. C. *J. Chem. Phys.* **2000**, *113*, 2983.

(52) Wilson, P. J.; Tozer, D. J. *Chem. Phys. Lett.* **2001**, *337*, 341.

(53) Patchkovskii, S.; Autschbach, J.; Ziegler, T. *J. Chem. Phys.* **2001**, *115*, 26.

(54) Walsh, J. D.; Miller, A.-F. *J. Phys. Chem. B* **2003**, *107*, 854.

(55) Proynov, E.; Chermette, H.; Salahub, D. R. *J. Chem. Phys.* **2000**, *113*, 10013.

(56) Patchkovskii, S.; Autschbach, J.; Ziegler, T. *J. Chem. Phys.* **2001**, *115*, 26.

(57) Helgaker, T.; Jaszunski, M.; Ruud, K. *Chem. Rev.* **1999**, *99*, 293.

(58) Fadda, E.; Casida, M. E.; Salahub, D. R. *J. Phys. Chem. A* **2003**, *107*, 9924.

(59) Ziegler, T. *Chem. Rev.* **1991**, *91*, 651.

(60) Vignale, G.; Rasolt, M. *Phys. Rev. Lett.* **1987**, *59*, 2360.

(61) Wüllen, C. V. *J. Chem. Phys.* **1995**, *102*, 2806.

(62) Lee, A. M.; Handy, N. C.; Colwell, S. M. *J. Chem. Phys.* **1995**, *103*, 10095.

(63) Malkin, V. G.; Malkina, O. L.; Casida, M. E.; Salahub, D. R. *J. Am. Chem. Soc.* **1994**, *116*, 5898.

(64) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; J. C. Burant; S. Dapprich; J. M. Millam; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Rega, N.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; 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.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98, Revision A.11.2; Gaussian, Inc.: Pittsburgh, 2001.

(65) Osmialowski, B.; Kolehmainen, E.; Gawinecki, R. *Magn. Reson. Chem.* **2001**, *39*, 334.

(66) Solum, M. S.; Altman, K. L.; Strohmeier, M.; Berges, D. A.; Zhang, Y.; Facelli, J. C.; Pugmire, R. J.; Grant, D. M. *J. Am. Chem. Soc.* **1997**, *119*, 9804.

(67) Witanowski, M.; Sicinska, W.; Grabowski, Z.; Webb, G. A. *J. Magn. Reson., Ser. A* **1993**, *104*, 310.

(68) Ruostesuo, P.; Hakkinen, A. M.; Mattila, T. *Magn. Reson. Chem.* **1987**, *25*, 189.

(69) Jameson, C. J.; Jameson, A. K.; Oppusunggu, D.; Wille, S.; Burrell, P. M.; Mason, J. *J. Chem. Phys.* **1981**, *74*, 81.

(70) Forsyth, D. A.; Sebag, A. B. *J. Am. Chem. Soc.* **1997**, *119*, 9483.

(71) Singh, S.; Basmadjian, G. P. *Tetrahedron Lett.* **1997**, *38*, 6829.

(72) Huang, D. F.; Shen, T. Y. *Tetrahedron Lett.* **1993**, *34*, 4477.

(73) Davies, H. M. L.; Saikali, E.; Young, W. B. *J. Org. Chem.* **1991**, *56*, 5696.

(74) Tanny, S. R.; Grossman, J.; Fowler, F. W. *J. Am. Chem. Soc.* **1972**, *94*, 6495.

(75) Donohoe, T. J.; Guyo, P. M. *J. Org. Chem.* **1996**, *61*, 7664.

(76) Gharpure, M.; Stoller, A.; Bellamy, F.; Firnau, G.; Snieckus, V. *Synthesis* **1991**, 1079.

(77) Xiao, D.; Schreier, J. A.; Cook, J. H.; Seybold, P. G.; Ketcha, D. M. *Tetrahedron Lett.* **1996**, *37*, 1523.

(78) Rawal, V. H.; Jones, R. J.; Cava, M. P. *J. Org. Chem.* **1987**, *52*, 19.

(79) Wang, J.; Scott, A. I. *Tetrahedron Lett.* **1996**, *37*, 3247.

(80) Tietze, L. F.; Kettschau, G.; Heitmann, K. *Synthesis* **1996**, 851.

(81) Miller, A. D.; Leeper, F. J.; Battersby, A. R. *J. Chem. Soc., Perkin Trans. 1* **1989**, 1943.

(82) Hansch, C.; Leo, A.; Taft, R. W. *Chem. Rev.* **1991**, *91*, 165.

(83) Mothana, B.; Ban, F.; Boyd, R. J. *Chem. Phys. Lett.* **2005**, *401*, 7.

(84) Boyd, R. J.; Boyd, S. L. *J. Am. Chem. Soc.* **1992**, *114*, 1652.

(85) Bohmann, J. A.; Weinhold, F.; Farrar, T. C. *J. Chem. Phys.* **1997**, *107*, 1173.

(86) Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Weinhold, F. NBO 5.0; Theoretical Chemistry Institute, University of Wisconsin: Madison, WI, 2001.

(87) Paine, J. B., III; Dolphin, D. *J. Org. Chem.* **1985**, *50*, 5598.

(88) Ganske, J. A.; Pandey, R. K.; Postich, M. J.; Snow, K. M.; Smith, K. M. *J. Org. Chem.* **1989**, *54*, 4801.

(89) Moon, M. W.; Bell, L. T.; Webster, D. M. *J. Org. Chem.* **1974**, *39*, 315.

(90) Webb, G. A. Shielding: Overview of Theoretical Methods. In *Encyclopedia of Nuclear Magnetic Resonance*; Grant, D. M., Harris, R. K., Eds.; John Wiley: New York, 1996; Vol. 7; pp 4307.

(91) Pulay, P.; Hinton, J. F. Shielding Theory: GIAO Method. In *Encyclopedia of Nuclear Magnetic Resonance*; Grant, D. M., Harris, R. K., Eds.; John Wiley: New York, 1996; Vol. 7; pp 4334.

(92) Mennucci, B.; Martinez, J. M.; Tomasi, J. *J. Phys. Chem. A* **2001**, *105*, 7287.

(93) Solum, M. S.; Facelli, J. C.; Michl, J.; Grant, D. M. *J. Am. Chem. Soc.* **1986**, *108*, 6464.

(94) Gregor, T.; Mauri, F.; Car, R. *J. Chem. Phys.* **1999**, *111*, 1815.

(95) Paal, C. *Chem. Ber.* **1884**, *17*, 2756.

(96) Knorr, L. *Chem. Ber.* **1884**, *17*, 2863.

(97) Katritzky, A. R.; Ostercamp, D. L.; Yousaf, T. I. *Tetrahedron* **1987**, *43*, 5171.

(98) Amarnath, V.; Anthony, D. C.; Amarnath, K.; Valentine, W. M.; Wetterau, L. A.; Graham, D. G. *J. Org. Chem.* **1991**, *56*, 6924.

(99) Amarnath, V.; Amarnath, K.; Valentine, W. M.; Eng, M. A.; Graham, D. G. *Chem. Res. Toxicol.* **1995**, *8*, 234.

(100) Ogata, Y.; Kawasaki, A. *The Chemistry of the Carbonyl Group*; Interscience: London, 1970; Vol. 2.

(101) Borsche, W.; Fels, A. *Ber.* **1907**, *39*, 3877.

(102) Broadbent, H. S.; Burnham, W. S.; Losen, R. K.; Sheeley, R. M. *J. Heterocycl. Chem.* **1968**, *5*, 757.

(103) Sundberg, R. J. In *Comprehensive Heterocyclic Chemistry*; Katritzky, A. R., Rees, C. W., Eds.; Pergamon: Oxford, 1984; Vol. 4; pp 329.

(104) Kuang-Chih, T.; Jia-Sheng, Y.; Ding-Zhang, S. *Acta Chim. Sinica* **1981**, *39*, 215.

(105) Kuang-Chih, T. *Acta Chim. Sinica* **1982**, *40*, 720.

(106) Kuang-Chih, T. *Acta Chim. Sinica* **1966**, *32*, 107.

(107) Salva, A.; Donoso, J.; Frau, J.; Munoz, F. *J. Phys. Chem. A* **2003**, *107*, 9409.

(108) Frau, J.; Donoso, J.; Munoz, F.; Garcia Blanco, F. *J. Comput. Chem.* **1992**, *13*, 681.

(109) Frau, J.; Donoso, J.; Vilanova, B.; Munoz, F.; Garcia Blanco, F. *Theor. Chim. Acta* **1993**, *86*, 229.

(110) Weiner, S. J.; Singh, U. C.; Kollman, P. A. *J. Am. Chem. Soc.* **1985**, *97*, 2219.

(111) Courmier, D.; Gardebien, F.; Minot, C.; St-Amant, A. *Chemical Physics Letters* **2005**, *405*, 357.

(112) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, 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.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revision C.02; Gaussian Inc.: Wallingford CT, 2004.

(113) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(114) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(115) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(116) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

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

(118) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327.

(119) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151.

(120) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.

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

(122) Ban, F.; Lundqvist, M. J.; Boyd, R. J.; Eriksson, L. A. *J. Am. Chem. Soc.* **2002**, *124*, 2753.

(123) Galano, A.; Alvarez-Idaboy, J. R.; Montero, L. A.; Vivier-Bunge, A. *J. Comput. Chem.* **2001**, *22*, 1138.

(124) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 2936.

(125) Durant, J. L. *Chem. Phys. Lett.* **1996**, *256*, 595.

(126) Basch, H.; Hoz, S. *J. Phys. Chem. A* **1997**, *101*, 4416.

(127) Jackson, A. H. Nitrogen Systems: the Azoles. In *Comprehensive Organic Chemistry*; Barton, D., Ollis, W. D., Eds.; Pergamon: Oxford, 1979; Vol. 4; pp 296.

(128) Knorr, L. *Annalen* **1886**, *236*, 290.

(129) Fischer, H. *Org. Synth.* **1935**, *15*, 17.

(130) Fischer, H. *Org. Syn., Coll.* **1943**, *2*, 202.

(131) Fischer, H. *Org. Synth.* **1937**, *17*, 96.

(132) Corwin, A. H. The Chemistry of Pyrrole and its Derivatives. In *Heterocyclic Compounds*; Elderfield, R. C., Ed.; John Wiley & Sons, Inc.: New York, 1950; Vol. 1; pp 277.

(133) Fabiano, E.; Golding, B. T. *J. Chem. Soc., Perk. T 1* **1991**, 3371.

(134) Nakano, H.; Umio, S.; Kariyone, K.; Tanaka, K.; Kishimoto, T.; Noguchi, H. *Tetrahedron Lett.* **1966**, *7*, 737.

(135) Khetan, S. K.; George, M. V. *Tetrahedron* **1969**, *25*, 527.

(136) Sanchez, S. K.; Ventula, A. C.; Scheidegger, U. *Carbohyd. Res.* **1971**, *17*, 275.

(137) Shoolingin-Jordan, P. M.; Spencer, P.; Sarwar, M.; Erskine, P. E.; Cheung, K. M.; Cooper, J. B.; Norton, E. B. *Biochem. Soc. T.* **2002**, *30*, 584.

(138) Erskine, P. T.; Newbold, R.; Roper, J.; Coker, A.; Warren, M. J.; Shoolingin-Jordan, P. M.; Wood, S. P.; Cooper, J. B. *Protein Sci.* **1999**, *8*, 1250.

(139) Erskine, P. T.; Coates, L.; Newbold, R.; Brindley, A. A.; Stauffer, F.; Wood, S. P.; Warren, M. J.; Cooper, J. B.; Shoolingin-Jordan, P. M.; Neier, R. *FEBS Lett.* **2001**, *503*, 196.

(140) Erskine, P. T.; Norton, E.; Cooper, J. B.; Lambert, R.; Coker, A.; Lewis, G.; Spencer, P.; Sarwar, M.; Wood, S. P.; Warren, M. J.; Shoolingin-Jordan, P. M. *Biochem.* **1999**, *38*, 4266.

(141) Kervinen, J.; Jaffe, E. K.; Stauffer, F.; Neier, R.; Wlodawer, A.; Zdanov, A. *Biochem.* **2001**, *40*, 8227.

(142) Jordan, P. M.; Seehra, J. S. *J. Chem. Soc. Chem. Comm.* **1980**, 240.

(143) Jordan, P. M.; Seehra, J. S. *FEBS Lett.* **1980**, *114*, 283.

(144) Jaffe, E. K. *Bioorg. Chem.* **2004**, *32*, 316.

(145) Bahmanyar, S.; Houk, K. N. *J. Am. Chem. Soc.* **2001**, *123*, 11273.