Theoretical Investigations of Non-Covalent Interactions: From Small Water Clusters to Large DNA Quadruplexes

by

Alexis Taylor

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

at

Dalhousie University
Halifax, Nova Scotia
March 2010

© Copyright by Alexis Taylor, 2010

DALHOUSIE UNIVERSITY

DEPARTMENT OF CHEMISTRY

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled “Theoretical Investigations of Non-Covalent Interactions: From Small Water Clusters to Large DNA Quadruplexes” by Alexis Taylor in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated:

Supervisor: _____

Readers: _____

_____

_____

DALHOUSIE UNIVERSITY

DATE: March 22, 2010

AUTHOR: Alexis Taylor

TITLE: Theoretical Investigations of Non-Covalent Interactions: From Small Water Clusters to Large DNA Quadruplexes

DEPARTMENT OR SCHOOL: Department of Chemistry

DEGREE: PhD CONVOCATION: May YEAR: 2010

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

_____
Signature of Author

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

The author attests that permission has been obtained for the use of any copyrighted material appearing in the thesis (other than the brief excerpts requiring only proper acknowledgement in scholarly writing), and that all such use is clearly acknowledged.

To all those that have gone before me.

If I have seen further,
it is by standing on the shoulders of giants.

– Sir Isaac Newton

Table of Contents

List of Figures

List of Tables

Abstract

The chemical bonds that hold molecules together are composed of electrons, and in order to study these microscopic systems, electronic structure calculations are often employed. This thesis describes the results from several studies that use computational techniques to investigate a variety of bonding interactions. The systems presented range from small water clusters to large DNA quadruplexes. High-level computational techniques, such as *ab initio* and density functional theory methods, were applied as well as the quantum theory of atoms in molecules (AIM). AIM uses the gradient to analyze the electron density, partitioning the molecule into atomic fragments. Once the system is partitioned, individual atomic contributions to molecular properties can be determined. Furthermore, bonding interactions can be identified by the presence of a specific type of critical point within the topology. These two facets of AIM are exploited throughout this thesis.

The first project presented is a theoretical investigation of the exact electronic structure of hydrated electrons. Whether the excess electron resides within a central cavity or is smeared out over the surface of the cluster remains a contentious issue. In an attempt to investigate this dilemma from a novel viewpoint, AIM was used to analyze the electron density of small anionic water clusters up to ten water molecules. The results suggest that the preferred site of binding is dictated by the relative orientation of the non-hydrogen-bonded hydrogen atoms.

At the other end of the spectrum, the largest systems investigated were several guanine quadruplexes that can form in telomeric regions of DNA. In light of the attention these structures have received as potential therapeutic agents, a clear understanding of their formation is mandatory. The study presented here is a detailed investigation of the electronic energy changes associated with the folding of the quadruplex from the single-stranded telomere. After devising a novel method to display the atomic energy data, several interesting trends in the energy changes were identified. Ultimately, the data presented could help to guide future drug development endeavours, highlighting one of the many practical applications of computational methods.

List of Abbreviations and Symbols Used

A	hydrogen-bond acceptor
AIL	atomic interaction line
AIM	atoms in molecules
AM1	Austin model 1
AO	atomic orbital
B3	Becke's three-parameter exchange functional
BCP	bond critical point
BE	binding energy
BPL	bond path length
BSSE	basis set superposition error
CCP	cage critical point
CI	configuration interaction
CP	counterpoise
D	hydrogen-bond donor
DCBS	dimer-centered basis set
DCR	double chain reversal
DFT	density functional theory
DNA	deoxyribonucleic acid
EDG	electron donating group
EWG	electron withdrawing group
G4-B	basket-type guanine quadruplex
G4-P	propeller-type guanine quadruplex
GC	guanine-cytosine pair
GGA	generalized gradient approximation
GTO	Gaussian-type orbital
HF	Hartree-Fock
IAS	interatomic surface
KS	Kohn-Sham
LCAO	linear combination of atomic orbitals
LSDA	local spin density approximation
LYP	Lee-Yang-Parr correlation functional
MCBS	monomer-centered basis set
MO	molecular orbital
MPA	Mulliken population analysis
MPPT	Møller-Plesset perturbation theory

NCP	nuclear critical point
NHB	non-hydrogen-bonded
NNA	non-nuclear attractor
PDB	Protein Data Bank
PM3	parametric method 3
RCP	ring critical point
RHF	restricted Hartree-Fock
RI-MP2	resolution of the identity MP2
ROHF	restricted open-shell Hartree-Fock
SCF	self-consistent field
ssDNA	single-stranded DNA
STO	Slater-type orbital
UHF	unrestricted Hartree-Fock
ZPVE	zero-point vibrational energy

Ψ	wavefunction
$\hat{H}$	Hamiltonian operator
E	energy
$\hat{T}$	kinetic energy operator
$\hat{V}$	potential energy operator
m_α	mass of nucleus α
∇^2	Laplace operator
Z_α	effective nuclear charge of nucleus α
$r_{\alpha\beta}$	distance between nuclei α and β
r_{ij}	distance between electrons i and j
$\hat{H}_{elec}$	electronic Hamiltonian operator
E_{elec}	electronic energy
ψ_i	spatial orbital
α/β	spin function (up/down)
χ_i	spin orbital
$\hat{P}_{ij}$	permutation operator
ϕ_μ	basis function

$c_{\mu i}$	basis function expansion coefficient
g_k	Gaussian function or primitive
$d_{\mu k}$	Gaussian function expansion coefficient
Φ	trial function
S_{ij}	overlap integral
δ_{ij}	Kronecker delta (if $i = j$, $\delta_{ij} = 1$, if $i \neq j$, $\delta_{ij} = 0$)
$\hat{F}$	Fock operator
ε_i	orbital energy
$\hat{H}^{core}$	core Hamiltonian operator
$\hat{J}$	Coulombic operator
$\hat{K}$	exchange operator
$P_{\mu\nu}$	density matrix elements
$(\mu\nu\|\lambda\sigma)$	two-electron repulsion integrals
$\hat{S}^2$	spin operator
λ	perturbation parameter
$\rho(\vec{r})$	electron density
$E[\rho]$	energy functional
χ_i^{KS}	Kohn-Sham orbital
$\hat{F}^{KS}$	Kohn-Sham operator
ε_i^{KS}	Kohn-Sham orbital energy
E_{XC}	exchange-correlation energy functional
E_X	exchange energy functional
E_C	exchange energy functional
ΔH	enthalpy
ΔS	entropy
ΔG	Gibbs energy
ε_0	electronic energy
ε_{ZPVE}	zero-point vibrational energy
H_{corr}	thermal correction to enthalpy
G_{corr}	thermal correction to Gibbs energy

E_{tot}	correction to internal thermal energy
k_B	Boltzmann constant
T	temperature
S_{tot}	total entropy
$\nabla\rho$	electron density gradient
$\nabla\nabla\rho$	Hessian matrix
ω	rank
σ	signature
Ω	atomic basin
e^-	electron
$G(r)$	electronic kinetic energy density
$V(r)$	electronic potential energy density
$H(r)$	total electronic energy density
ε	ellipticity

Acknowledgements

I would like to take this opportunity to thank the various people who, without their help, I could not have completed this work. First and foremost I would like to thank Dr. Russell J. Boyd for allowing me to pursue my graduate studies under his supervision. Along the way he has provided indispensable guidance as well as many hilarious anecdotes and stories.

Thanks also goes out to the Boyd and Weaver research groups, both past and present. Their support and entertainment over the past four years has made my time at Dalhousie memorable. I would also like to thank my friends outside of Dalhousie, who were always there with welcome distractions.

I would like to acknowledge the members of my supervisory committee: Dr. Axel D. Becke, Dr. T. Bruce Grindley and Dr. Don F. Weaver. Their contributions have helped to shape this thesis into what it is today.

On a more practical note, I would like to thank the Natural Sciences and Engineering Research Council (NSERC), the Killam Trusts, the Walter C. Sumner Foundation and Dalhousie University for financial support. I also need to thank the Atlantic Computational Excellence Network (ACEnet) and the Western Canada Research Grid (WestGrid). Without the additional computational resources made available by these consortia, the majority of this thesis could not have been completed.

Finally, I need to thank my family for their support over the years and without whom I never would have gone into science in the first place. Thank you for opening these doors for me.

Chapter 1: Introduction

Computational chemistry, a relatively young branch of chemistry, can provide a great deal of insight into a large variety of chemical systems. In fact, by using computers, it is possible investigate systems that are often inaccessible using conventional experimental techniques. As computers become faster and more powerful, an increasing number of scientists are supplementing their research with computational results.

There are many different reasons to conduct a computational investigation. Often, before any work is performed in the laboratory, preliminary theoretical calculations are used in order to determine whether a molecule and a given target will interact, for instance a new drug candidate and a receptor. This approach can save a great deal of synthetic effort, reducing not only the amount of physical materials used but also precious manpower. As the world moves into an era of green chemistry and greater environmental consciousness, leaving a smaller footprint is always an admirable goal. This method has been widely applied in the pharmaceutical industry, as well as the design of new materials, and has been particularly successful in both areas. In such cases, computational chemistry actually supplants the need for experimental studies.

On the other hand, computational chemistry is often used to complement experimental work, providing the detailed atomic information needed to support experimental hypotheses. This approach is seen in mechanistic studies, where computational methods are frequently employed to gain insight about short-lived, high-energy intermediates and transition states. Since these transition states cannot be isolated

experimentally, computational techniques provide the only opportunity to thoroughly investigate these putative species. Furthermore, if the species of interest is particularly hazardous or highly reactive, computer simulations may be the only safe method available. For instance, the United States Department of Energy is involved in various computational studies that examine nuclear waste and its possible remediation. Thus, computational techniques allow for the investigation of systems that are often precluded from experimental inquiry.

Sometimes computational methods are used because they can provide a new approach to an old problem. This tactic is employed in this thesis to examine anionic water clusters, perhaps better known as solvated electrons, or more specifically hydrated electrons. Although the solvated electron was first observed in ammonia over 200 years ago, and subsequently in water in the 1960s, the exact electronic structure of these systems remains under debate. However, because this is such a long-standing problem, a great body of work, both experimental and theoretical, exists related to this quandary. The dilemma that has baffled scientists for decades is: what is the exact binding mode of an excess electron in small water clusters? Whether the electron exists as an independent, localized entity, known as a cavity-bound state, or whether it is distributed over several atoms, a surface-bound state, is a contentious issue. Therefore, in order to contribute anything useful to this debate, a novel approach must be employed. To this end, the quantum theory of atoms in molecules (AIM) has been used to probe the nature of an excess electron in small water clusters. There are several key features of this theory, which will be discussed in Chapter 2, that have been utilized in an attempt to

resolve this dilemma. Thus, computational chemistry, specifically the theory of atoms in molecules, provides a novel viewpoint from which to examine an old problem.

In addition to the study of hydrated electrons, several other studies are presented in this thesis, each of which exploits some particular aspect of computational methods. Although at first glance the remaining projects may seem a disjointed collection, they do share at least one common trait. All of the studies presented, including the hydrated electron project, investigate non-covalent interactions ranging from strong hydrogen bonds to weak aromatic-aromatic stacking interactions. For instance, Chapter 5 presents a series of self-assembling helices, a system that was inspired by the elegant structure of deoxyribonucleic acid (DNA). The monomers in the helices consist of alternating hydrogen-bond acceptor (A) and donor (D) groups, and thus these structures are referred to as ADADA helices. In addition to the strong hydrogen bonds that bind the monomers to one another, there is a slew of weak aromatic stacking interactions that stabilize the helices. Initially this project sought to probe the interactions that hold the helices together and to determine if these systems could be modeled with the available resources. However, once it was clear that these helices could be modeled both efficiently and reliably using density functional theory (DFT) methods, the focus of the investigation turned to how various substituents could affect the strength of binding. In this case, although the initial helix structure had been prepared and observed experimentally, computational techniques allow these systems to be investigated at the atomic level and to determine specifically which parts of the molecule contribute to the overall stability of the structure.

The final study presented, an investigation of guanine quadruplexes, is truly the *pièce de résistance* in this thesis. These systems have received a great deal of attention lately due to their potential therapeutic applications. Specifically, the quadruplexes under investigation form in telomeric regions of DNA (i.e., the ends of chromosomes). Again, the theory of atoms in molecules was used; however, in this case, it was the ability of AIM theory to partition the system into atomic fragments that proved particularly advantageous. Individual atomic energies were computed to investigate the energetic consequences of quadruplex formation. This type of detailed atomic data is not available from experiment, nor in fact from conventional computational techniques but rather, it is a feature that is specific to AIM. This study was a massive undertaking: each quadruplex investigated consisted of almost 700 atoms and each individual atomic energy took an average of 100 hours to compute. However, in the end, a clear understanding of how each atom is affected upon quadruplex formation has been gained. This data may ultimately be used to develop drug molecules to stabilize the quadruplex structures for potential therapeutic applications.

Although the examples presented provide only a glimpse of the merits offered by computational methods, it should be clear that these techniques have become invaluable to the fields of chemistry, biochemistry and biology, and that they have novel insights to offer. It is the goal of this thesis to illustrate the utility of computational methods and give an idea of the breadth of chemical problems that can be investigated.

Chapter 2: Theoretical Background

2.1 Introduction

In the early 20th century it became clear that although the laws of classical mechanics were applicable to macroscopic systems, they did not govern the behaviour of microscopic systems. In order to accurately describe and predict the properties of atoms and molecules a new set of laws was needed, and thus quantum mechanics was born. These new laws accounted for the wave-particle duality of electrons and microscopic systems, a feat that classical mechanics had failed to overcome.

The laws of quantum mechanics are extremely elegant. This is reflected in the fact that any microscopic system can be described by a single function, known as the wavefunction (Ψ). This function is only dependent upon the coordinates of the particles in the system and time. Furthermore, it contains all of the information about the system that can possibly be desired. Thus, the goal of quantum mechanics has been reduced to determining the correct wavefunction. Unfortunately, due to the complex nature of this function, an exact solution is impossible for all but the simplest systems.

In order to obtain meaningful information about atoms and molecules using these laws, several approximations and assumptions must be made. Ultimately, each method discussed in this chapter employs a specific set of assumptions and as a consequence each has its own advantages and disadvantages. There are two main types of computational techniques used throughout this thesis: *ab initio* methods and density functional theory (DFT). Both techniques can be used to obtain important information about atomic and molecular systems, however they do so in very different ways. *ab initio* methods, or first

principles methods, do not use any empirical data in their derivation of the approximate wavefunction from which the desired data is extracted. On the other hand, DFT methods assume that the information about the system can be obtained from the electron density and thus never actually solve for the wavefunction. The goal of this chapter is to provide the reader with a fundamental understanding of the methods employed throughout the thesis, while more thorough discussions can be found in various theoretical chemistry textbooks.[1-6]

2.2 The Schrödinger Equation

In 1926 Erwin Schrödinger proposed his namesake equation,[7] which is most often applied to chemical systems in its time-independent form,

$$\hat{H}\Psi(x) = E\Psi(x) \qquad (\mathbf{2.1})$$

where $\hat{H}$ is the Hamiltonian operator and E is the energy of the system. Solutions to equation **2.1** correspond to stationary states of the system under investigation. The solution with the lowest energy, E_0, represents the ground state. Note that the Hamiltonian operator can be decomposed into kinetic ($\hat{T}$) and potential ($\hat{V}$) energy terms,

$$\hat{H} = \hat{T} + \hat{V} \qquad (\mathbf{2.2})$$

Although the Schrödinger equation appears benign enough in its short-hand notation, as shown in equation **2.1**, it is important to realize that the molecular Hamiltonian for an N-electron, M-nuclei system can be written more fully, in atomic units, as,[1]

$$\hat{H}_{total} = -\frac{1}{2}\sum_{\alpha}^{M}\frac{1}{m_\alpha}\nabla_\alpha^2 - \frac{1}{2}\sum_{i}^{N}\nabla_i^2 + \sum_{\alpha}^{M}\sum_{\beta>\alpha}^{M}\frac{Z_\alpha Z_\beta}{r_{\alpha\beta}} - \sum_{\alpha}^{M}\sum_{i}^{N}\frac{Z_\alpha}{r_{i\alpha}} + \sum_{i}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \qquad \textbf{(2.3)}$$

Here α and β refer to nuclei and *i* and *j* to electrons. In spite of the formidable appearance of equation **2.3**, the nature of each term can be easily deciphered. The first and second terms correspond to the kinetic energy of the nuclei ($\hat{T}_n$) and electrons ($\hat{T}_e$), respectively. The last three terms correspond to the potential energy of the interactions between the particles. Specifically, the third term is the internuclear repulsion between nuclei α and β ($\hat{V}_{nn}$), the fourth term is the attraction between nucleus α and electron *i* ($\hat{V}_{ne}$), while the last term is the Coulombic repulsion between electrons *i* and *j* ($\hat{V}_{ee}$). Thus, the total Hamiltonian could be more simply expressed in the following fashion,

$$\hat{H}_{total} = \hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{ne} + \hat{V}_{ee} \qquad \textbf{(2.4)}$$

As mentioned previously, an exact solution to the Schrödinger equation, that is obtaining the correct wavefunction, is impossible for all but the simplest systems. These include the particle in a box problem, the harmonic oscillator and hydrogen and hydrogen-like systems (i.e., one electron systems). The reason equation **2.1** is insoluble for multi-electron systems stems from the fact that the last term, the Coulombic repulsion term, is inseparable. Thus, at best only approximate solutions can be obtained from these equations by making a few careful assumptions.

2.3 The Born-Oppenheimer Approximation

One of the fundamental assumptions made in quantum chemistry is that of Born and Oppenheimer.[8] Simply stated, it assumes that since the nuclei are so much more

massive than the electrons they can be considered stationary while the electrons move quickly in the field of fixed nuclei. Thus, the nuclear and electronic motions are separable. This assumption has the effect of simplifying equation **2.3** since the kinetic energy of the nuclei must necessarily be zero and the internuclear repulsion term becomes a constant, V_{NN}. The total molecular Hamiltonian can then be separated into an electronic and a nuclear component,

$$(\hat{H}_{elec} + V_{NN})\Psi_{elec} = U\Psi_{elec} \qquad (\mathbf{2.5})$$

where $\hat{H}_{elec}$ is the electronic Hamiltonian, Ψ_{elec} is the electronic wavefunction, dependent only upon the coordinates of the electrons for a given nuclear configuration, and U, the total energy, is the sum of the electronic energy and the nuclear repulsion energy,

$$U = E_{elec} + V_{NN} \qquad (\mathbf{2.6})$$

This separation allows the purely electronic Schrödinger equation to be obtained,

$$\hat{H}_{elec}\Psi_{elec} = E_{elec}\Psi_{elec} \qquad (\mathbf{2.7})$$

where E_{elec} is the electronic energy and the purely electronic Hamiltonian, $\hat{H}_{elec}$, is,

$$\hat{H}_{elec} = -\frac{1}{2}\sum_{i}^{N}\nabla_i^2 - \sum_{\alpha}^{M}\sum_{i}^{N}\frac{Z_\alpha}{r_{i\alpha}} + \sum_{i}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \qquad (\mathbf{2.8})$$

It is this form of the Schrödinger equation that is most often employed in computational chemistry.

2.4 The Orbital Approximation

Even with the simplified Hamiltonian (**2.8**), further approximations are necessary in order to obtain meaningful information from the Schrödinger equation. This comes in

the form of the orbital approximation, which assumes that the wavefunction for an N-electron system, $\Psi(1,2,...,N)$, can be written as a product of one-electron functions, known as orbitals (ψ),[1]

$$\Psi(1,2,...,N) = \psi_1(1)\psi_2(2)...\psi_N(N) = \prod_{i=1}^{N}\psi_i(i) \qquad \textbf{(2.9)}$$

This expression for the wavefunction is also known as the Hartree product.

Since the probability density of any N-electron system is given by the product of the wavefunction and its complex conjugate, Ψ^*, the N-electron probability density is reduced to a product of N one-electron probability densities,

$$\Psi^*(1,2,...,N)\Psi(1,2,...,N) = \prod_{i=1}^{N}\psi_i^*(i)\psi_i(i) \qquad \textbf{(2.10)}$$

According to probability theory, if a probability can be written as the product of individual probabilities, the events associated with those probabilities must occur independently of one another. Thus, the orbital approximation is based on an independent electron model, implying that the motion of each electron is independent of all the other electrons.

The Hartree product given above describes the spatial distribution of the N electrons within the system. However, for the complete description of any electron, the spin state of each electron must also be included. This is most easily accomplished by multiplying each spatial function (ψ_i) by a spin function (α or β), producing a unique spin orbital (χ_i). Thus, the wavefunction is in fact written as a product of one-electron spin orbitals.

With the wavefunction conveniently expressed as a product of one-electron functions, it is important to consider the overall symmetry of this expression. Electrons are indistinguishable particles, and as such, no property can be affected simply by the interchange of any two electrons.[1] Consider the electron probability density as shown in equation **2.10**, in order for this result to remain constant following the interchange of two electrons, there are only two possible solutions,

$$\Psi(1,2,...i,j,...N) = \pm\Psi(1,2,...j,i,...N) \qquad (\mathbf{2.11})$$

However, it is known from experiment that electrons, and in fact all fermions, must have antisymmetric wavefunctions with respect to the interchange of any two electrons.[1] Thus, if the interchange of two electrons is affected by a permutation operator, $\hat{P}_{ij}$, the resultant wavefunction must have the opposite sign,

$$\begin{aligned} \hat{P}_{ij}\Psi(1,2,...i,j,...N) &= \Psi(1,2,...j,i,...N) \\ &= -\Psi(1,2,...i,j,...N) \end{aligned} \qquad (\mathbf{2.12})$$

In 1929 Slater proposed that the antisymmetric wavefunction could be obtained from a determinant in which the rows refer to the electrons and each column to a particular spin orbital.[9,10] The result for an N-electron system is,

$$\Psi(1,2,...N) = (N!)^{-1/2}\begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(N) & \chi_2(N) & \cdots & \chi_N(N) \end{vmatrix} \qquad (\mathbf{2.13})$$

where $(N!)^{-1/2}$ is a normalization constant, found by imposing the following condition,

$$\int \Psi^*\Psi \; d\tau = 1 \qquad (\mathbf{2.14})$$

Expansion of the Slater determinant in **2.13** generates all of the permutations of the N electrons amongst the N spin orbitals. Notice that this representation ensures that the wavefunction is antisymmetric since interchange of two electrons (i.e., interchange of two rows) simply changes the sign of the function. It also satisfies the Pauli exclusion principle, which states that no two electrons can have identical spin orbitals. This situation would correspond to two identical columns in the matrix, for which the determinant would vanish.

2.5 Basis Sets

Although the orbital approximation allows for a convenient expression of the many-electron wavefunction as a product of one-electron functions, questions arise about the form of these functions. Typically, the spin orbitals, χ_i, which are often referred to as molecular orbitals (MO), are constructed as a linear combination from a set of known one-electron functions, ϕ_μ, called basis functions,

$$\chi_i = \sum_{\mu=1}^{K} c_{\mu i} \phi_\mu \qquad \textbf{(2.15)}$$

where $c_{\mu i}$ is the coefficient in the expansion for the μ^{th} basis function in the i^{th} MO. A particular set of K basis functions forms a basis set.

Various types and sizes of basis functions are available, as well as the way in which they can be combined. Since the choice of basis set can greatly affect the quality of the MOs, and ultimately the accuracy of the calculation, careful consideration of these factors is essential.

In some cases, the MOs are expanded from a set of atomic orbitals (AO) on the constituent atoms. This approach is commonly known as the linear combination of atomic orbitals (LCAO), but can quickly become impractical for larger systems. Consequently, simpler functions are often used instead of the AOs.

The two most common types of basis functions employed are Slater-type orbitals (STO)[11] and Gaussian-type orbitals (GTO).[12] Both classes of functions exhibit exponential decay, which is necessary for normalization purposes. The key difference between the Slater- and Gaussian-type orbitals stems from the fact that the former decay with $\exp(-\zeta r)$ and possess a cusp at the nucleus while the latter decay with $\exp(-\zeta r^2)$, decaying too rapidly as r increases, and exhibiting a finite value at the nucleus (**Figure 2.1**). The orbital exponents (ζ) can be adjusted to control the size of the orbitals.

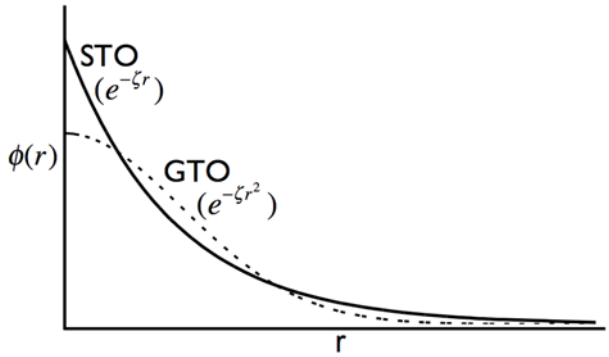


Figure 2.1: Graphical representation of Slater- and Gaussian-type orbitals.

From **Figure 2.1** it should be clear that Slater-type orbitals provide a better description of the molecular orbitals, and thus are considered superior. Unfortunately, the better description comes at an increased computational cost and so GTOs tend to be used more frequently. However, in an attempt to overcome the deficiencies inherent in

GTOs, linear combinations of individual Gaussian functions, called primitives (g_k) are often used,

$$\phi_\mu = \sum_k d_{\mu k} g_k \qquad \textbf{(2.16)}$$

where the coefficients $d_{\mu k}$ are fixed. These basis functions are known as contracted Gaussian basis functions. A single GTO constructed from a set of Gaussian functions will provide better results than any single uncontracted Gaussian function, while reducing the computational effort associated with the use of STOs.

As was mentioned previously, the choice of basis set can greatly affect the results obtained. Generally speaking, larger, more complete basis sets will provide the best results, but at a significant computational cost. Thus, the ultimate goal is to select a basis set that is robust enough to accurately describe the system of interest while keeping the computational demand to a minimum.

2.5.1 Minimal Basis Sets

Minimal basis sets, or single-zeta basis sets, are constructed from a single basis function for each core and valence orbital in the system. This type of basis set represents the smallest basis set that can be used in quantum mechanical calculations. The STO-NG basis set is an example of a minimal basis set in which N primitive Gaussians are fitted to a single STO in a least-squares fashion to form a single basis function. Using primitives with different orbital exponents, ζ, and then contracting them into one basis function gives a better representation than any of the primitives alone. However, even for large N, it is still a minimal basis set and extended basis sets are required for more accurate results.

2.5.2 Split Valence Basis Sets

One easy method to extend a basis set is to increase the number of basis functions used per orbital. Split-valence basis sets employ more than one basis function of variable orbital exponents for each valence orbital and only one basis function for each core orbital. For instance, the valence double-zeta (VDZ) basis set uses two basis functions per valence orbital while the valence triple-zeta (VTZ) uses three, and so on. Another example of these are the *k-lmn*G basis sets developed by Pople and co-workers.[13,14] Here, *k* represents the number of primitive Gaussians used for each core orbital while *l*, *m* and *n* represent the primitives used for the valence orbitals. For instance, the 6-31G basis set uses a set of 6 primitives contracted to one basis function for each core orbital and a split-valence of 3 and 1 primitives for the valence orbitals. While the split-valence basis sets provide a better description of the molecular orbitals because they allow for variable atom size, they still are not able to provide a balanced basis set on their own.

2.5.3 Polarization and Diffuse Functions

Although split-valence basis sets can be extended essentially *ad infinitum*, this is generally not a good procedure. To obtain improved results, polarization functions must be added to the basis set. These are functions of higher angular momentum such as *d*- and *f*- type functions for heavy atoms and *p*- and *d*-type functions for hydrogen and helium atoms. This increases the flexibility of the basis set by allowing the shape of the orbital to change (i.e., become polarized in one direction). The inclusion of polarization functions can be important for systems in which hydrogen is involved such as hydrogen-bonding and proton transfers since these will allow the *s* orbital to become distorted from its regular spherical shape.

Furthermore, the addition of diffuse *s*- and *p*-type functions is useful for systems in which the electron density is spread over a large area. These become particularly important for systems and processes that involve radicals or anions and delocalized π systems.

In terms of the Pople basis set nomenclature, the addition of polarization functions is indicated in brackets after the contraction scheme while diffuse functions are denoted by '+'. For instance, the 6-31++G(d,p) basis set includes a set of *d*-type functions on all heavy atoms and *p*-type functions on hydrogen and helium. The first '+' indicates that diffuse *s*- and *p*-type functions are included on heavy atoms, while the second '+' indicates that diffuse *s*-type functions have been included on hydrogen.

Another group of basis sets that are often used are those of Dunning and co-workers.[15,16] These are the correlation-consistent polarized valence (double/triple/etc) zeta basis sets, cc-pVXZ, where X indicates the degree of splitting. These were developed specifically for use in calculations that include electron correlation.

2.5.4 Basis Set Superposition Error

When calculating the interaction energy between two (or more) monomers, the choice of basis set can have a profound effect. Intuitively, the interaction energy can be calculated as the energy of the complex less the energy of the individual monomers. Ideally, a mathematically complete basis set would be used for these calculations; however, in reality, this is not possible. Due to the use of an incomplete basis set, the interaction energy tends to be overestimated as a consequence of the basis set superposition error (BSSE). This error arises from an artificial lowering of the complex

energy because the monomers in the complex have a better description of their molecular orbitals due to the presence of the other monomer's basis functions.

One popular approach to correct for BSSE, and the one used in this thesis, is the counterpoise (CP) method of Boys and Bernardi.[17] This scheme calculates the energy of a monomer in the presence of 'ghost' basis functions of the other monomers without any nuclei or electrons. While the magnitude of the BSSE is typically quite small, it can have a large effect on the calculated binding energies depending on the strength of the interactions under investigation.

2.6 The Variational Theorem

The traditional goal of quantum chemistry has been to determine the correct form of the wavefunction in order to gain insight about the system under investigation. The Slater determinant (**2.13**) provides a convenient means to determine the antisymmetric wavefunction from the spin orbitals. Furthermore, the spin orbitals themselves can be constructed as a linear combination of basis functions as shown in equation **2.15**. The challenge at this point is how to determine the coefficients, $c_{\mu i}$, in the expansion of the basis functions. This is often done variationally and so at this point it is important to introduce the variational theorem.

More often than not, computational chemistry is concerned with the ground state of a given system. Given the exact ground-state wavefunction, Ψ_0, the corresponding ground-state energy, E_0, can be obtained from the expectation value of the Hamiltonian operator. According to the variational theorem, given a trial function, Φ, that satisfies the boundary conditions of the system, the lowest-energy eigenvalue of the trial function,

E_Φ, will always be greater than or equal to the exact ground-state energy. This can be expressed as,

$$E_\Phi = \frac{\langle \Phi | \hat{H} | \Phi \rangle}{\langle \Phi | \Phi \rangle} \geq E_0 \qquad (\mathbf{2.17})$$

Using this theorem, the exact ground-state wavefunction is approximated from the trial function that gives the lowest energy.

2.7 Hartree-Fock Theory

The Hartree-Fock (HF) method is the simplest *ab initio* method available and provides the basis for many higher order methods. As mentioned previously, in order to obtain meaningful information about the system of interest, the spin orbitals, χ_i, must be determined. The best χ_i are obtained by using the variational method and the Hartree-Fock equations. Note that the spin orbitals are constrained to form an orthonormal set, that is, the overlap integral between two orbitals must equal the Kronecker delta,

$$S_{ij} = \int \chi_i \chi_j \, d\tau = \delta_{ij} \qquad (\mathbf{2.18})$$

2.7.1 The Hartree-Fock Equations

The Hartree-Fock equation[18-21] can be written as,

$$\hat{F}\chi_i = \varepsilon_i \chi_i \qquad (\mathbf{2.19})$$

where $\hat{F}$ is the Fock operator, or the effective HF Hamiltonian and ε_i is the orbital energy of the spin orbital χ_i. The Fock operator can be decomposed into contributions from various operators,

$$\hat{F} = \hat{H}^{core} + \sum_{j=1}^{N} (2\hat{J}_j - \hat{K}_j) \qquad \textbf{(2.20)}$$

where $\hat{H}^{core}$ is the one-electron Hamiltonian corresponding to the motion of a single electron in the field of bare nuclei (this accounts for the motion of the electron and the electron-nuclei interactions). The second term, composed of the Coulomb operator ($\hat{J}_j$) and the exchange operator ($\hat{K}_j$), accounts for the potential experienced by electron *i* due to the presence of other electrons. The Coulomb operator is analogous to classical electron-electron repulsion while the exchange operator has no physical interpretation but arises due to the antisymmetry principle. These two operators are defined as,

$$\hat{J}_j(1) = \int \chi_j^*(2) \frac{1}{r_{12}} \chi_j(2) d\tau_2 \qquad \textbf{(2.21)}$$

$$\hat{K}_j(1)\chi_i(1) = \left[\int \chi_j^*(2) \frac{1}{r_{12}} \chi_j(2) d\tau_2 \right] \chi_j(1) \qquad \textbf{(2.22)}$$

The HF equations are solved iteratively until the χ_i no longer change and are self-consistent with the potential field they generate. For this reason, these methods are said to be self-consistent field (SCF) methods. The set of spin orbitals obtained yields the Hartree-Fock approximation to the true ground-state system. These non-linear equations were very important to the development of quantum chemistry. However, they prove to be difficult to solve other than by numerical methods and thus are not practical for polyatomic systems.

2.7.2 The Roothaan-Hall Equations

In order to transform the problem into a system that is soluble using linear algebra, the Roothaan-Hall method is used. This method expresses the spin orbitals, or molecular orbitals, as linear combinations of atomic orbitals, or basis functions (ϕ_μ) as shown in equation **2.15**. In theory, the set of basis functions employed should form a complete set in order to properly describe the molecular orbitals. However, this would require an infinite number of basis functions and thus this approach is not feasible. Fortuitously, careful selection of the basis functions can provide a decent representation of the molecular orbitals even though a finite basis is employed.[1] In general, the larger the number of basis functions included, the more accurate the description of the molecular orbitals.

Application of the variational method to the Hartree-Fock equations and substitution of equation **2.15** gives the Roothaan-Hall equations,[22,23]

$$\sum_{\nu=1}^{M}(F_{\mu\nu}-\varepsilon_i S_{\mu\nu})c_{\nu i}=0, \qquad \mu=1,2,...,M \tag{2.23}$$

This formulation yields a minimum energy system when the coefficients satisfy the above equation. The Fock matrix, $F_{\mu\nu}$, consists of elements defined by,

$$F_{\mu\nu}=H_{\mu\nu}+\sum_{\lambda=1}^{M}\sum_{\sigma=1}^{M}P_{\lambda\sigma}\left[(\mu\nu|\lambda\sigma)-\frac{1}{2}(\mu\lambda|\nu\sigma)\right] \tag{2.24}$$

where $H_{\mu\nu}$ is a matrix element of the core Hamiltonian operator, $P_{\lambda\sigma}$ is a matrix element of the density matrix such that,

$$P_{\lambda\sigma}=2\sum_{\mathrm{i}=1}^{occ\ MOs}c_{\mu i}c_{\nu i} \tag{2.25}$$

and $(\mu\nu|\lambda\sigma)$ is the two-electron repulsion integral representing the Coulombic repulsion between two local product densities, $\phi_\mu\phi_\nu$ and $\phi_\lambda\phi_\sigma$. The orbital energies, ε_i, and the overlap integral between basis functions ϕ_μ and ϕ_ν, $S_{\mu\nu}$, also appear in the Roothaan-Hall equations. Again, these equations must be solved iteratively by assuming some initial guess for the coefficients. The simplest approach is to use $H_{\mu\nu}$ as an initial guess for $F_{\mu\nu}$ and iterate until both the energy and density matrix have converged.

2.7.3 Restricted and Unrestricted Hartree-Fock Theory

In closed-shell systems, those in which all of the electrons are paired, two electrons are assigned to each spatial orbital with different spins. A Hartree-Fock calculation performed on this type of system is known as a restricted Hartree-Fock (RHF) calculation.

On the other hand, open-shell systems include those cases in which unpaired electrons are present, for example radicals and excited states, and often require special attention. Two general approaches are widely used to deal with open-shell systems. Using Hartree-Fock as an example, both restricted open-shell (ROHF) and unrestricted (UHF) Hartree-Fock are used. The former employs a single set of spatial orbitals, many of which are doubly occupied and one or more are singly occupied. Conversely, the latter employs different spatial orbitals for the α and β spin states using the same basis functions while the coefficients for the two spin states are varied independently. The three different scenarios are illustrated in **Figure 2.2**. Note that electrons are paired in the ROHF system while unpaired in the UHF system. Consequently, the latter predicts slightly lower energies than the former and may provide a better description for some

chemical systems. However, unrestricted calculations can be slow to converge due to the presence of coupled equations. Furthermore, the unrestricted wavefunction is not an eigenfunction of the spin operator $\hat{S}^2$ and spin contamination can become a concern and should always be examined.[1]

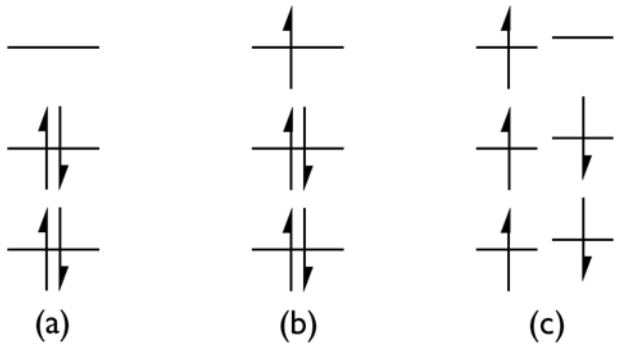


Figure 2.2: Schematic illustration of (a) restricted (R), (b) restricted open-shell (RO) and (c) unrestricted (U) Hartree-Fock methods.

2.8 Electron Correlation

One of the major deficiencies of Hartree-Fock theory is the neglect of explicit electron-electron interactions. This phenomenon is known as electron correlation and it refers to the fact that, in reality, the motion of electrons is correlated. For instance, since electrons repel one another, it is impossible to find two electrons with parallel spin in the same location at the same time. However, HF calculations account for the interactions between electrons only in an average way. Even the best Hartree-Fock calculation with an infinite basis set, known as the HF limit, predicts an energy higher than the true energy of the system.[24] In order to go beyond Hartree-Fock, the instantaneous interactions between electrons must be considered and re-introduced into the calculation. The

difference between the exact energy (E_{exact}) of the system and that obtained at the Hartree-Fock limit (E_{HF}) is the correlation energy (E_{corr}),[1]

$$E_{corr} = E_{exact} - E_{HF} \tag{2.26}$$

Although the correlation energy is small compared to the HF energy, it generally has a magnitude similar to that of chemical bonds, and thus must be accounted for when studying chemical systems and their reactions. Many schemes have been developed to recover this correlation energy, including various perturbation and variational treatments. These methods are known as post-HF methods, some of which are discussed below.

As an aside about post-HF methods, most of these schemes involve some sort of truncation in order to make the calculation feasible. It should be noted that the error associated with using a finite basis set (basis set truncation error) tends to be greater than the error associated with the truncation of the method.[1] Thus, in general it is more important to use an extended basis set with a more severely truncated method than a small basis set with highly correlated method.

2.8.1 Configuration Interaction

One method that is often used to include electron correlation is configuration interaction (CI). CI is a variational method based on the premise that the wavefunction can be expanded as a linear combination of Slater determinants corresponding to the ground state and various excited states. It is the addition of the excited states that accounts for electron correlation.

Consider an N-electron system constructed from a finite basis set of M basis functions. This system has 2M spin orbitals, of which N are occupied, leaving $(2M - N)$

unoccupied or virtual orbitals. The ground-state or Hartree-Fock wavefunction, $|\Psi_0\rangle$, is formed from the N lowest spin orbitals. Various excited states arise from promoting one or more electrons from the N occupied orbitals to any of the virtual orbitals. For this system, there are $\begin{pmatrix} 2M \\ N \end{pmatrix}$ N-electron determinants including the ground state. It is obvious that even for small values of N and M, there is a large number of excited configurations.

As stated previously, the wavefunction can be written as a linear combination of the ground state and all possible excited states,[1]

$$|\Psi\rangle = c_0|\Psi_0\rangle + \sum_{a,r} c_a^r |\Psi_a^r\rangle + \sum_{\substack{a<b \\ r<s}} c_{ab}^{rs} |\Psi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c \\ r<s<t}} c_{abc}^{rst} |\Psi_{abc}^{rst}\rangle + \sum_{\substack{a<b<c<d \\ r<s<t<u}} c_{abcd}^{rstu} |\Psi_{abcd}^{rstu}\rangle + \ldots \quad \textbf{(2.27)}$$

where the coefficients in this expansion, c_i, are treated variationally to find the lowest energy state. The first term in equation **2.27** is the ground-state HF determinant while the following terms correspond to single, double, triple, etc. excitations. The summations indicate that all possible excitations of a single type are included. For instance, all single excitations, ($|\Psi_a^r\rangle$), are included in which one electron from an occupied spin orbital (χ_a) is promoted to an unoccupied virtual orbital (χ_r). Similarly for the double excitations, ($|\Psi_{ab}^{rs}\rangle$), one electron is excited from χ_a to χ_r while another electron is excited from χ_b to χ_s, and so on for the higher excitations.

It should be noted that the use of a mathematically complete basis set and the full configuration interaction (FCI) wavefunction, yields the exact solution to the N-electron system. However, it is impractical to perform full CI calculations with large basis sets and thus truncation schemes become necessary. For instance, inclusion of all single

excitations leads to the configuration interaction singles (CIS) method while inclusion of all double excitations yields the CID method, and so forth. It turns out that the CIS method is useful to predict excitation energies but double excitations must be included for any improvement in the calculated ground-state energy. Several drawbacks are associated with the use of CI methods, most notably slow convergence and the lack of size-consistency for truncated CI methods.[1]

2.8.2 Møller-Plesset Perturbation Theory

Perturbation-based techniques have also been widely applied to systems of interacting particles. The basic premise for this type of calculation is that the real system is only slightly different from an unperturbed system for which the exact solution is known. For instance, Møller-Plesset perturbation theory (MPPT), developed by Møller and Plesset in 1934,[25] has proven particularly useful for the study of atoms and molecules. This method assumes that the true Hamiltonian of the system ($\hat{H}$) is in fact the sum of the unperturbed Hamiltonian ($\hat{H}_0$) and the perturbation ($\hat{H}'$),

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}' \qquad (\mathbf{2.28})$$

where the factor λ allows the perturbation to be gradually turned on, relating the unperturbed and perturbed systems in a continuous fashion.[1] This approach is naturally amenable to the inclusion of electron correlation, assuming it can be added back into the system as a perturbation to the Hartree-Fock wavefunction. Substitution of equation **2.28** into the Schrödinger equation yields,

$$(\hat{H}_0 + \lambda \hat{H}')\Psi = E\Psi \qquad (\mathbf{2.29})$$

In addition, perturbation theory assumes that both the wavefunction and energy can be expanded as a Taylor series in powers of λ,

$$\Psi = \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \lambda^3\Psi^{(3)} + ... + \lambda^k\Psi^{(k)} + ... \quad \textbf{(2.30)}$$

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + ... + \lambda^k E^{(k)} + ... \quad \textbf{(2.31)}$$

where $\Psi^{(k)}$ and $E^{(k)}$ are the *k*th-order correction to the wavefunction and energy, respectively.[1] Substitution of equations **2.30** and **2.31** into equation **2.29** is a formidable result. However, assuming that the perturbation is small, then only the first few terms in the expansion make significant contributions to the wavefunction or energy. Thus, these calculations are often truncated to include a specific level of perturbation. For instance, MP2 includes the second-order corrections while MP3 includes the third order, and so forth. Since the inclusion of perturbation accounts for electron correlation, the different orders of correction include correlation to a different extent. Experience has shown that MP2 recovers most of the correlation energy while MP3 offers little improvement at a greater computational cost and so is rarely used.[1] After MP2, MP4 is the next preferred method but once again is computationally more demanding. Methods beyond MP4 are rarely used due to the cost and the negligible improvements over MP2 or MP4.

As with configuration interaction, there are certain disadvantages associated with perturbative treatments of atoms and molecules. First and foremost, these methods are not variational and thus may predict energies that are lower than the true energy of the system. They can also be slow to converge and in some cases are in fact divergent. That being said, compared to CI they have the advantage of being size-consistent and are less computational demanding and are thus often used when highly accurate results are desired.

2.8.2.1 Resolution of the Identity

While it is desirable to use the most accurate method possible, the use of post-HF methods with large basis sets often proves to be prohibitively expensive. One means of lessening the computational demand is to use auxiliary basis expansions.[26-30] In these schemes, standardized auxiliary basis sets ($|K\rangle$) are used to approximate the products of Gaussian basis functions,

$$|\mu\nu\rangle \approx \sum_{K} |K\rangle C_{\mu\nu}^{K} \qquad (\mathbf{2.32})$$

This has the effect of replacing four-center, two-electron integrals, with three- and two-center integrals, greatly reducing the computational demand. Ultimately, this allows for much faster calculation of energy and gradients (often three- to thirty-fold) with a negligible loss of accuracy (e.g. less than 60 μHartree per atom).[31]

2.9 Density Functional Theory

Although post-Hartree-Fock methods can provide a great deal of information about chemical systems, the large computational cost associated with these methods is often prohibitive. *ab initio* methods seek to determine the correct form of the wavefunction and subsequently extract the desired information. However, these wavefunctions are cumbersome, containing a great deal of redundant information, and are generally lacking in physical meaning. All of these factors have encouraged the search for alternative functions that can provide information about the system of interest. One such result is density functional theory (DFT), which is based on the premise that the energy of a system, and other properties, can be determined from the electron density ($\rho(\vec{r})$) rather than the wavefunction. The appeal of this approach lies in the fact that the

electron density is physically observable and is a function of only three variables (*x, y, z*) regardless of the number of particles in the system. In addition, DFT methods are able to account for electron correlation and have accuracy on par with correlated *ab initio* methods (with some exceptions) at a fraction of the computational cost. Early attempts were made by Thomas[32] and Fermi[33] and later Slater[34] to obtain useful information from the electron density. However, it was the work of Hohenberg and Kohn[35] in 1964 that truly laid the foundation for density functional theory as it stands today.

2.9.1 The Hohenberg-Kohn Theorems

The first Hohenberg-Kohn theorem succinctly proved that the ground-state electron density, $\rho_0(\vec{r})$, uniquely determines the external potential, $\hat{V}_{ext}(\vec{r})$, which in turn determines the Hamiltonian operator, $\hat{H}$. Thus, the electron density can predict the ground-state energy, E_0, as well as a myriad of other properties of both the ground and excited states.[35] The name of this class of methods, density functional theory, stems from the fact that the energy, and indeed many other properties, is a functional of the ground-state electron density,

$$E_0 = E_0[\rho_0(\vec{r})] \tag{2.33}$$

Note that it is only the *ground-state* electron density that can provide the desired information, densities of excited states will not suffice.

The total energy functional can be further decomposed into various energy functional contributions,

$$E\left[\rho_0(\vec{r})\right] = T\left[\rho_0(\vec{r})\right] + E_{ee}\left[\rho_0(\vec{r})\right] + \int V_{ext}\rho_0(\vec{r}) \tag{2.34}$$

where $T\left[\rho_0(\vec{r})\right]$ is the kinetic energy of the electrons, $E_{ee}\left[\rho_0(\vec{r})\right]$ is the energy of the electron-electron interactions and $\int V_{ext}\rho_0(\vec{r})$ is the energy of the electron-nuclei interactions due to the external potential field of the nuclei, V_{ext}.

The second theorem in the 1964 paper relates to the variational principle as it is applied to the electron density. The theorem states that for some trial density, $\tilde{\rho}(\vec{r})$, that satisfies the boundary conditions and is associated with some external potential $\tilde{V}_{ext}$, the energy, $E[\tilde{\rho}(\vec{r})]$, will be an upper-bound to the true ground-state energy,

$$E[\tilde{\rho}(\vec{r})] \geq E_0[\rho_0(\vec{r})] \qquad \textbf{(2.35)}$$

It should be mentioned that this variational theorem holds true only when the exact energy functional is used. Therefore, if the functional is approximated it is possible to obtain energies lower than the true energies.[3] This is the reason why DFT methods are variational in principle but not in practice.

The Hohenberg-Kohn theorems proved that various properties of atomic and molecular systems were obtainable from the electron density rather than via the wavefunction. However, their paper failed to elucidate the nature of the functionals that would return the desired information. There was also no indication of how to determine the density without first finding the wavefunction. Thus, although the paper was theoretically elegant, it lacked practicality.

2.9.2 The Kohn-Sham Theorem

The real breakthrough for DFT methods came a year later in 1965 from Kohn and Sham.[36] They realized that the early density-based methods failed due to poor representations of the kinetic energy. To improve their model, they conjectured the

existence of a reference system of N non-interacting electrons with some effective local potential, $V_S(\vec{r})$, such that the kinetic energy of this system, T_S, is known exactly. The key at this point is to connect the real, interacting system to the reference system. This is accomplished by choosing $V_S(\vec{r})$ such that the density of the reference system, $\rho_S(\vec{r})$, is exactly equal to the density of the interacting system, $\rho_0(\vec{r})$. For any non-interacting system, the ground-state wavefunction can be defined by a Slater determinant. In this case, the determinant is composed of Kohn-Sham (KS) orbitals, χ_i^{KS}, and the resulting density is given by,

$$\rho_S(\vec{r}) = \sum_{i=1}^{N} \left| \chi_i^{KS} \right|^2 = \rho_0(\vec{r}) \qquad \textbf{(2.36)}$$

These KS orbitals can be determined using equations analogous to those used in HF theory,

$$\hat{F}^{KS} \chi_i^{KS} = \varepsilon_i^{KS} \chi_i^{KS} \qquad \textbf{(2.37)}$$

where $\hat{F}^{KS}$ is the Kohn-Sham operator and ε_i^{KS} is the KS orbital energy.

Furthermore, Kohn and Sham refined the energy expression **2.34** by assuming that the true kinetic energy could be defined as the sum of the kinetic energy of the non-interacting system, T_S, and a correction term, T_C. The final result is written as follows,

$$E\left[\rho_0\right] = T_S\left[\rho_0\right] + J\left[\rho_0\right] + E_{XC}\left[\rho_0\right] + \int V_{ext}\rho_0 \qquad \textbf{(2.38)}$$

where the electron-electron interaction energy ($E_{ee}\left[\rho_0(\vec{r})\right]$) is replaced by the classical Coulomb repulsion term, $J\left[\rho_0\right]$, and the remaining non-classical effects, such as the self-

interaction correction, exchange and correlation effects as well as the kinetic energy correction, are folded into the exchange-correlation energy term, $E_{XC}[\rho_0]$.

The exact form of all the terms in equation **2.38** are known except for E_{XC}. Substitution of these expressions and application of the variational principle yields the exact form of the KS operator,

$$\hat{F}^{KS} = -\frac{1}{2}\nabla^2 - \sum_A \frac{Z_A}{r_{Ai}} + \int \frac{\rho(\vec{r}')}{|\vec{r}' - \vec{r}|} + V_{XC}(\vec{r}) \qquad (\mathbf{2.39})$$

where V_{XC}, the potential due to E_{XC}, is defined as the functional derivative of E_{XC} with respect to the density,

$$V_{XC} = \frac{\delta E_{XC}}{\delta \rho} \qquad (\mathbf{2.40})$$

These equations are then solved iteratively in an SCF framework to obtain the KS orbitals, which in turn determine the density and ultimately the energy. However, it is at this point that approximations must be made as to the form of V_{XC}. Essentially all of the unknowns have been confined to E_{XC}, and it is the goal of modern DFT research to determine the correct form of this hitherto unknown functional.

2.9.3 Exchange-Correlation Functionals

Ultimately, the accuracy of any DFT method depends on the accuracy of the approximate E_{XC} functional. The earliest attempts were based on uniform electron gas models, assuming E_{XC} depends only on the density.[36,37] Functionals of this type are often referred to as local spin density approximation (LSDA) functionals,[3]

$$E_{XC}^{LSDA}[\rho_\alpha, \rho_\beta] = \int \rho(\vec{r}) \varepsilon_{XC}(\rho(\vec{r})_\alpha, \rho(\vec{r})_\beta) d\vec{r} \qquad (\mathbf{2.41})$$

where ε_{XC} is the exchange-correlation energy per particle.

The LSD approximation provides better results than HF for certain properties such as equilibrium structures, vibrational frequencies and dipole moments; however, it cannot determine the highly accurate energy data as desired by many computational chemists.[3] In general, the LSDA provides reliable information for those systems that closely resemble a uniform electron gas, namely those in which the density varies slowly with position. However, in reality, atomic and molecular systems do not possess uniform electron densities and thus more sophisticated models are required.

Further development of exchange-correlation functionals was spurred on by the generalized gradient approximation (GGA), which assumes that the functional depends on both the density and the gradient of the density,[1]

$$E_{XC}^{GGA}[\rho_\alpha, \rho_\beta] = \int f(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta) d\vec{r} \qquad \textbf{(2.42)}$$

These functionals generally offer an improvement over the LSDA since they account for the variation of density with position.

To simplify the problem, E_{XC} is often written as the sum of an exchange (E_X) and correlation (E_C) term,[1]

$$E_{XC} = E_X + E_C \qquad \textbf{(2.43)}$$

The exchange-energy functional can then be obtained from the HF exchange term with the Kohn-Sham orbitals in place of the HF orbitals[1] and approximate solutions for E_C are sought. Various exchange and correlation functionals have been developed independently and can be combined in various ways. For instance, one popular GGA

functional is BLYP, where Becke's 1988 exchange functional[38] is paired with the Lee-Yang-Parr correlation functional.[39]

Another set of functionals that are widely used are the hybrid exchange-correlation functionals. These combine the exchange-correlation GGA functionals and an exact exchange functional, ultimately yielding a method that is somewhere between HF and GGA.[1] The most popular hybrid functional, B3LYP,[40,41] uses Becke's 1988 exchange functional (E_X^{B88}) and Lee, Yang and Parr's correlation functional (E_C^{LYP}) as gradient corrections to the LSDA exchange and correlation functionals,

$$E_{XC}^{B3LYP} = (1-a)E_X^{LSDA} + aE_{XC}^{\lambda=0} + bE_X^{B88} + cE_C^{LYP} + (1-c)E_C^{LSDA} \qquad (\mathbf{2.44})$$

where the three parameters, *a*, *b* and *c*, control the contributions of the various functionals.

Both GGA and hybrid functionals predict good equilibrium geometries, vibrational frequencies and dipole moments and are also able to provide accurate energies,[1] a major advantage over the simple LSDA functionals. Although many functionals are currently available that can provide highly accurate results, the development of new functionals continues to be an active area of research.

2.10 Semi-Empirical Methods

Ideally all quantum mechanical calculations would be performed using high-level methods, such as post-HF or DFT. However, some systems are just too large for this to be feasible. The most demanding component of quantum mechanical calculations is the evaluation of two-electron integrals. Semi-empirical methods are able to lessen this demand by neglecting or approximating many of these integrals and using empirical

parameters to correct for the errors introduced. In addition, these methods typically consider only the valance electrons (i.e., those involved in bonding and reactions) and tend to use minimal basis sets. While these calculations have the same general structure as HF calculations, employing a Hamiltonian and a wavefunction, the approximations allow for much larger systems to be investigated than with *ab initio* or DFT methods. Results from semi-empirical calculations are generally reliable as long as the method has been parameterized for the atoms in the system of interest.[24] For instance, Austin Model 1 (AM1)[42] and parametric method 3 (PM3)[43-46] have been parameterized for organic molecules and thus tend to be popular choices to model these systems.

2.11 Thermochemical Data from Calculations

Although computational studies often report results in terms of electronic energies, these same techniques are able to provide thermochemical data that may be compared more directly with experimental results. This can include enthalpies (ΔH), entropies (ΔS) and Gibbs energies (ΔG). A brief overview of how these quantities are obtained in Gaussian 03[47] is presented here, but the reader is referred to Ochterski's paper[48] (and references therein) for a more complete discussion.

In order to obtain the desired thermochemical data, a frequency calculation must be performed on a minimum energy structure. The output of this calculation provides various corrections that must be applied to the electronic energy (ε_0) to determine the thermochemical quantities.

The first of these corrections is the zero-point vibrational energy (ε_{ZPVE}), which accounts for the vibrational energy of a molecule at 0 K. To obtain an accurate estimate of the energy of a molecule (E), this correction must be added to the electronic energy,

$$\text{E = sum of electronic and zero-point energies} = \varepsilon_0 + \varepsilon_{\text{ZPVE}} \quad \textbf{(2.45)}$$

This corrected energy can then be used to calculate relative electronic energies. In an analogous fashion, the enthalpy (ΔH) and Gibbs energy (ΔG) are obtained by adding the appropriate correction to the electronic energy (equation **2.46** and **2.47**).

$$\Delta\text{H = sum of electronic energy and thermal enthalpy} = \varepsilon_0 + H_{corr} \quad \textbf{(2.46)}$$

$$\Delta\text{G = sum of electronic energy and thermal free energy} = \varepsilon_0 + G_{corr} \quad \textbf{(2.47)}$$

H_{corr} already includes $\varepsilon_{\text{ZPVE}}$ and so it is important not to account for this correction twice. H_{corr} is calculated using the Boltzmann constant (k_B), temperature (T) and the correction to the internal thermal energy (E_{tot}) (equation **2.48**). E_{tot} includes contributions from internal translation (E_t), rotation (E_r), vibration (E_v) and electronic motion (E_e).

$$H_{corr} = E_{tot} + k_B\text{T} \quad \textbf{(2.48)}$$

$$E_{tot} = E_t + E_r + E_v + E_e \quad \textbf{(2.49)}$$

G_{corr} also includes $\varepsilon_{\text{ZPVE}}$ and is calculated using H_{corr}, T and the total entropy (S_{tot}) (equation **2.50**). Similar to E_{tot}, S_{tot} includes contributions from the motions mentioned above (equation **2.51**).

$$G_{corr} = H_{corr} - \text{T}S_{tot} \quad \textbf{(2.50)}$$

$$S_{tot} = S_t + S_r + S_v + S_e \quad \textbf{(2.51)}$$

In this thesis, these thermodynamic parameters are calculated for the separated monomers that make up a hydrogen-bonded complex as well as the complex itself. In this case, the former may be considered as reactants and the complex as the product.

Thus, the energy changes that accompany complex formation, or the reaction, are defined as follows:

$$\text{Relative energy of reaction} = E_{prod} - E_{react} \quad (\mathbf{2.52})$$

$$\Delta H_{rxn} = \Delta H_{prod} - \Delta H_{react} \quad (\mathbf{2.53})$$

$$\Delta G_{rxn} = \Delta G_{prod} - \Delta G_{react} \quad (\mathbf{2.54})$$

Finally, entropy changes are determined using the calculated enthalpy and Gibbs energy changes and the Gibbs energy equation,

$$\Delta G = \Delta H - T\,\Delta S \quad (\mathbf{2.55})$$

assuming a temperature of 298.15 K as specified in the frequency calculation.

2.12 The Quantum Theory of Atoms in Molecules

The quantum theory of atoms in molecules (AIM)[4-6,49,50] provides a topological analysis of the electron density, $\rho(\mathbf{r})$. The precepts of this theory allow for the real-space partitioning of molecular systems into fragments that define the atoms in the molecule. This rigorous mathematical partitioning permits the determination of the atomic contributions to various properties such as energy and volume and can also be used to predict bonding.[51]

2.12.1 The Electron Density and Gradient Vector Fields

The major premise of AIM theory is that the electron density can provide detailed information about the chemical system. This density can be obtained either experimentally, from X-ray diffraction studies, or theoretically, from a myriad of computational schemes. Contour plots of ρ are dominated by regions of high electron

density around the nuclei, but fail to yield much additional insight. Nonetheless, a great deal of information is available from the electron density, and the question at this point is how to extract it. The approach used by AIM theory is to analyze the density using the gradient vector,

$$\nabla\rho = \frac{\partial\rho}{\partial x}\hat{i} + \frac{\partial\rho}{\partial y}\hat{j} + \frac{\partial\rho}{\partial z}\hat{k} \qquad (\mathbf{2.56})$$

By evaluating the gradient at some point, then following the gradient vector a short distance and re-evaluating the gradient, a gradient path can be traced showing the path of steepest ascent in ρ.[4] Since the gradient path is constructed from many small vectors, the path is directional with a beginning and end. Typically, in molecules, the gradient path is traced from infinity and terminates at a nucleus, the point where the electron density is a local maximum. Given that the gradient paths often terminate at the nuclei, the nuclei are referred to as attractors. Many of these gradient paths together yield the gradient vector field, a plot of which reveals how the molecule is naturally partitioned into fragments that define the atoms in the molecule.

2.12.2 Critical Points

In addition to the gradient paths that terminate at the nuclei, various critical points within the electron density must be identified to completely partition the system. Critical points are defined as locations within the electron density where the gradient is zero, that is $\nabla\rho = 0$.[4] In fact, the gradient paths always begin and end at points where the gradient is zero. There are four types of critical points found within the electron density: nuclear critical points (NCP), bond critical points (BCP), ring critical points (RCP) and cage critical points (CCP).

Critical points are classified according to the number of non-zero eigenvalues of the Hessian matrix and their signs.[4] This matrix is defined by,

$$\nabla\nabla\rho = \begin{Bmatrix} \frac{\partial^2\rho}{\partial x^2} & \frac{\partial^2\rho}{\partial x\partial y} & \frac{\partial^2\rho}{\partial x\partial z} \\ \frac{\partial^2\rho}{\partial y\partial x} & \frac{\partial^2\rho}{\partial y^2} & \frac{\partial^2\rho}{\partial y\partial z} \\ \frac{\partial^2\rho}{\partial z\partial x} & \frac{\partial^2\rho}{\partial z\partial y} & \frac{\partial^2\rho}{\partial z^2} \end{Bmatrix} \qquad (\mathbf{2.57})$$

The eigenvalues can be determined via diagonalization since this matrix is real and symmetric. Once the eigenvalues are determined, the critical points are classified according to their rank and signature using the shorthand notation (ω, σ). The rank (ω) is simply the number of non-zero eigenvalues and the signature (σ) is the sum of the signs of the eigenvalues.[4] In particular, only those critical points of rank 3 are of interest during AIM analysis. This is because critical points with rank less than 3 are degenerate and are considered topologically unstable and thus the structure itself is unstable.[5] In terms of the critical points previously mentioned, NCP are (3, -3) critical points, BCP are (3, -1), RCP are (3, +1) and CCP are (3, +3). The signature gives an indication of the curvature of the electron density at the critical point. For instance, a (3, -3) critical point is a maximum in all three directions with three negative curvatures while a (3, -1) critical point is a maximum in two directions and a minimum in the third. **Figure 2.3** presents the molecular graph of a hexameric water cluster illustrating the various critical points.

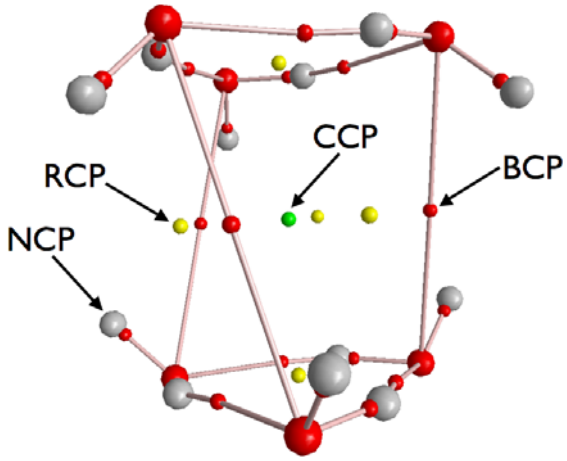


Figure 2.3: Molecular graph of a hexameric water cluster illustrating the various critical points found in the electron density along with the nuclei and bond paths.

One further relationship that should be mentioned with respect to critical points is the Poincaré-Hopf rule. This is an algebraic relationship indicating the number of critical points of each type that can exist within a molecule. The rule is given as,

$$n - b + r - c = 1 \qquad \textbf{(2.58)}$$

where n is the number of nuclear critical points (NCP), b are the BCPs, r are the RCPs and c are the CCPs.[52] This relationship holds for non-periodic systems and can really only be used as an indication that a system is incomplete, since the relationship holds even when two different types of critical points are missing.

2.12.3 Interatomic Surface

As previously discussed, the nuclear critical points are attractors for the gradient paths. Additionally, there is a set of gradient paths that originate at infinity and terminate at the BCPs between the nuclei. These paths define the interatomic surface (IAS), a zero-flux surface satisfying the condition,[4]

$$\nabla\rho(\mathbf{r})\cdot\mathbf{n}(\mathbf{r})=0 \qquad (\mathbf{2.59})$$

where $\mathbf{n}(\mathbf{r})$ is a unit vector normal to the surface. The IAS serves as a boundary between fragments, further defining the atoms that make up the molecule. A given fragment region and its attractor constitute the atomic basin (Ω) and various atomic properties can be determined by performing volume integrals over the basin. For instance, the electronic population of an atom ($N(\Omega)$) is defined as,

$$N(\Omega)=\int_{\Omega} d\tau\rho(r) \qquad \mathbf{(2.60)}$$

Although the integrand appears trivial, due to the complex shapes of the atomic basins, it must be solved numerically.

One of the key advantages of AIM theory is the rigorous partitioning of the system into fragments based on the IASs. Because the individual atoms have well defined boundaries, the atomic properties are additive and must therefore predict the overall molecular property when summed,[5]

$$A_{molecule}=\sum_{\Omega}A(\Omega) \qquad (\mathbf{2.61})$$

where A is any given property of interest.

2.12.4 Bond Paths

As well as the gradient paths that terminate at the nuclei and the BCPs, there is another set of gradient paths that are fundamental in AIM analysis. These paths are traced from the BCP to the nuclei, forming the bond paths that link two nuclei. The bond path defines the atomic interaction line (AIL), a line through the electron density that is a maximum in ρ with respect to any neighbouring line.[53] However, the bond path alone is

not a definitive criterion for bonding, it is also important that the forces on the nuclei vanish. This implies that the bond paths predicted in any minimum energy structure are indeed true bonding interactions.[5] Thus, AIM theory is able to predict bonding between nuclei with no *a priori* knowledge of how the atoms are connected. Furthermore, although the bond path is traced as a single bond path, further analysis of the BCP and its properties can provide an indication of the bond order and ultimately the strength. The network of nuclei and the bond paths represents the molecular graph for a given system.

2.12.5 Non-Nuclear Attractors

On a final note, maxima in the electron density at locations other than that of the nuclei, although rare, do exist and are called non-nuclear attractors (NNA). These were first predicted to exist theoretically in alkali metal clusters[54-56] and were later observed experimentally in metallic beryllium and magnesium.[57,58] NNAs behave topologically just as nuclei; they are attractors in the gradient vector field bounded by a zero-flux surface. As such, NNA are often referred to as pseudo-atoms with associated pseudo-atomic basins and properties. For any system containing a NNA, the Poincaré-Hopf rule must be adjusted to include both the nuclei and NNAs, that is n = NCP + NNA.

Chapter 3: The Topological Nature of Hydrated Electrons

This chapter is reproduced in part with permission from Taylor, A.; Matta, C. F.; Boyd, R. J. *J. Chem. Theory Comput.* **2007**, 3, 1054-1063. Copyright 2007 American Chemical Society.

3.1 Introduction

The exact microscopic structure of the hydrated electron has intrigued scientists for nearly two centuries, and to this day, remains the subject of much experimental and theoretical interest[59-77] (see Taylor et al.[78] and references therein). In broad terms, the solvated electron is a free electron in solution surrounded by a shell of solvent molecules. This free electron can experience a variety of interactions depending on the nature of the solvent. For instance, in a polar solvent such as water, the electron (which is essentially a point charge) could interact with the positive terminus of the solvent molecule, effectively forming some sort of novel hydrogen bond (consider $H\cdots e^-$). Another possibility is that the electron simply interacts with the overall dipole of the solvent molecules. On the other hand, the electron may not even remain independent in the solvent but rather becomes delocalized and is distributed over several solvent molecules. These are the types of suggestions that have been made to try to describe the exact structure of the solvated, or more specifically, hydrated electron.

Humphry Davy first observed solvated electrons in 1808, when he noted that the addition of alkali metals to ammonia formed an intense blue solution.[79] Although Davy

was uncertain of exactly what he had observed, he realized that he had witnessed some new and interesting chemistry. From that point on, these systems were under careful scrutiny as they were extensively characterized. Weyl was able to establish that these solutions behaved as electrolytes,[80] but it was nearly 100 years after Davy's initial observations that Kraus suggested the intense colour was due to the presence of free electrons in the solution.[81] The reaction between the metal and liquid ammonia could be generalized as follows,

$$M_{(s)} \xrightarrow{NH_{3(l)}} M^{+} + e^{-} \qquad (\mathbf{3.1})$$

where M is an alkali metal. Research in this area has been prolific. It has even led to the discovery of a new class of compounds, known as electrides, which are salts in which the free electrons serve as the anions.[82] Solvated electrons are not just an intellectual curiosity, organic chemists exploit their reactivity for practical uses as well. For instance, the Birch reduction uses solvated electrons in ammonia in order to reduce benzene derivatives to cyclic dienes.[83]

Since water has a larger dipole moment than ammonia and greater hydrogen-bonding capabilities, it might be assumed that water would facilitate the ionization of alkali metals. However, it turns out that it is not as easy to generate solvated electrons in water as it is in ammonia. It was not until the 1960s that hydrated electrons (solvated electrons in water) were generated and observed using pulse radiolysis.[84-86] Unfortunately, it was quickly noted that hydrated electrons in bulk solvent are highly reactive and exhibit a short pH-dependent lifetime; the free electrons can be annihilated by recombination with protons from the autoionization of water.

This hurdle was overcome by Haberland and coworkers when they succeeded in generating anionic water clusters of a finite size and observing them using mass spectrometry.[87-90] Ultimately, this facilitated investigations of the hydrated electron since water molecules in small clusters do not experience autoionization and thus the associated hydrated electrons have appreciable lifetimes. Currently there are many experimental techniques used to generate solvated electrons: interaction of alkali metal surfaces with water,[91] injection of energetic electrons into the solvent,[84-86] photoionziation of water,[92-95] supersonic expansion followed by irradiation with a low-energy electron beam,[96,97] and laser vapourization,[98,99] just to name a few. Once the hydrated electrons are produced, a variety of spectroscopic methods can be used to detect them: photoelectron and infrared spectroscopy, as well as mass spectrometry. And yet, in spite of extensive experimental research over many years, questions still remain as to the exact structure of the hydrated electron. Whether the excess electron is localized to a specific region in space or is distributed over several atoms in the cluster remains a point of much contention.

Two general models for the hydrated electron have been proposed: a surface-bound state and a cavity-bound state (**Figure 3.1**). The former is characterized by the distribution of the excess electron over several atoms at the surface of the cluster. This model is also often referred to as the dipole-bound state. The latter predicts that the electron is localized to a specific region within the center of the cluster. To date, experimental techniques have not been able to definitively state which mode of binding is preferred, although there is some experimental support for the cavity-bound model due to

the appreciable volume increase observed upon the injection of electrons into the solvent.[100]

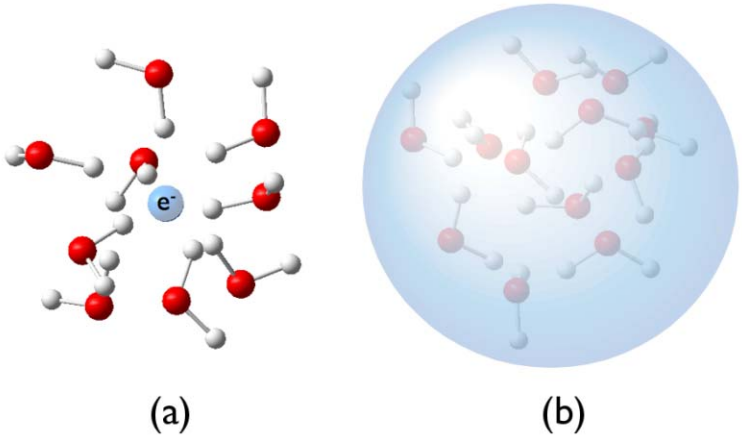


Figure 3.1: Schematic illustration of (a) the cavity-bound and (b) surface-bound states of the hydrated electron.

In order to elucidate the exact electronic structure of anionic water clusters many studies have employed theoretical techniques. Indeed, the theoretical body of work relating to the hydrated electron is vast, dating back to the early 1970s.[101] More recently, the microscopic structure has been probed using various *ab initio* and DFT methods for smaller clusters,[66,69,102-106] while molecular dynamics and Monte Carlo simulations have been performed on larger clusters and the bulk solvent.[59,60,67,70,107,108]

One of the major criticisms of theoretical investigations in general is the inability to directly compare theoretical and experimental results. This stems from the fact that the chemical systems and reaction conditions are often simplified in order to make the calculations feasible. Typically, the calculation is performed assuming gas-phase conditions, which would obviously be a poor representation of a system that exists naturally in solution. However, small anionic water clusters are most often observed

experimentally in the gas phase using mass spectrometry. Thus, in this case, electronic structure calculations can provide direct insight into the experimental results. However, even with an abundance of experimental and theoretical data available, the debate rages on as to which model of the hydrated electron prevails.

Water is ubiquitous throughout the natural environment; it has important roles in a large variety of systems, from the smallest cell to vast interstellar space. Specifically, anionic water clusters have been implicated in various reaction processes that can be detrimental to living systems. For instance, these clusters will react with oxygen to form the superoxide anion,[109] O_2^-, a particularly reactive species which is known to destroy ozone in the upper atmosphere[110] and is involved in aging, cell death and various disease states.[111] Thus, the hydrated electron has remained the focus of much research for many years in the hopes of obtaining a clear understanding of water and its reactivity.

One technique to determine the distribution of an excess electron is the use of population analyses. However, traditional methods, such as Mulliken population analysis (MPA), fail to elucidate information about the nature of the hydrated electron. This scheme assigns electronic populations on the basis of atom-centered basis functions.[24] Consequently, these methods will simply delocalize the excess electron over several atoms in the cluster, predicting a surface-bound model even for those situations in which it is cavity-bound.

On the other hand, the theory of atoms in molecules (AIM) provides a unique approach to the hydrated electron problem. If the excess electron exists as a localized entity within small water clusters, it will appear as a maximum in the electron density. As discussed in Chapter 2, maxima in the electron density at locations other than that of

the nuclei are known as non-nuclear attractors (NNAs). These behave topologically in the electron density just as nuclei and are often referred to as pseudo-atoms. The rigorous criteria used to define atoms in AIM theory are also applicable to NNAs. As a result, they are necessarily bound by a zero-flux surface, defining the pseudo-atomic basin in which electron density may reside. Therefore, if the excess electron adopts a cavity-bound state, it should appear as a NNA with an associated pseudo-atomic basin and electronic population. Otherwise, if it adopts a surface-bound state, the electronic populations of the atoms at the surface of the cluster will simply increase by some fraction of the excess electron. Previous to this study, there was no indication in the literature that AIM theory had ever been used to investigate the hydrated electron. Thus, this work represents a novel contribution to the hydrated electron dilemma.

3.2 Computational Details

All structures were optimized at the UMP2(full)/6-31++G(2d,2p) level of theory and frequencies calculated at that level to ensure local minima had been located. All electronic energies were corrected for zero-point vibrations. DFT results were also obtained at the ROB3LYP/6-31++G(2d,2p) level of theory and were found to give qualitatively similar results with minor differences in geometric parameters, and are not included here. All electronic structure calculations were performed using Gaussian 03.[47] For any system with unpaired electrons, spin contamination can become a concern. However, the spin contamination was never in excess of 0.083% for any of the clusters in this study.

Although there is some precedence in the literature for including floating-center basis functions at the center of cavity-forming clusters,[104,112] this approach was not used

here. It was felt that this type of scheme would bias the results, falsely predicting a central non-nuclear attractor when in fact it is not present in the electron density. Furthermore, by using the theory of atoms in molecules, which is able to predict maxima in the electron density even when a nucleus is not found at that location, the need for these extra basis functions is eliminated.

Molecular graphs were obtained using AIM2000[113] while the AIMPAC suite of programs[114] was used to analyze the electron densities and obtain atomic properties. The spin density isosurfaces, where the spin density is the difference between the α- and β-spin densities, i.e. $\rho^{spin}(\mathbf{r}) \equiv \rho^{\alpha}(\mathbf{r}) - \rho^{\beta}(\mathbf{r})$, were obtained using GaussView.[115] All reported spin density isosurfaces are the 0.001 a.u. isosurface. Higher and lower density isosurfaces were shown to yield the same qualitative results and so the 0.001 a.u. surface was chosen as a representative surface except for **8B**$^-$, which uses the 0.0005 a.u. surface.

3.3 Results and Discussion

The nature of the hydrated electron was probed using electronic structure calculations of select anionic water clusters, $(H_2O)_n^-$, n = 1 to 8, and their neutral counterparts. In order to establish individual atomic contributions to various molecular properties, the theory of atoms in molecules was used. In particular, the distribution of the excess electron was determined from the difference in charge between each atom in the anionic cluster and a neutral cluster in the same geometry. Additionally, the geometries of the anionic clusters were used as initial coordinates for an optimization of a neutral cluster.

At this point, a note should be made about the types of clusters included in this study. The goal of this investigation was to ascertain whether the excess electron prefers a cavity- or surface-bound state, and further to that, what exactly is the minimum requirement for a cavity-bound state, if it exists. As the investigation progressed, it became clear that certain cluster arrangements favoured different distributions of the hydrated electron. For this reason, only two types of structures were studied for each of the different cluster sizes: an anionic cluster that closely resembles the preferred neutral cluster geometry and one in which several non-hydrogen-bonded (NHB) hydrogen atoms are directed towards the center of the cluster, resulting in an empty central cavity. This study is by no means an exhaustive search for all possible cluster geometries, nor was it intended to be.

3.3.1 Anionic Water Monomer, $(H_2O)^-$

The natural starting point for any investigation of small anionic water clusters is a single water molecule. As may be expected, the addition of an electron to an isolated water molecule has little effect on the structural parameters: the O–H bond lengths increase by 0.013 Å while the H–O–H bond angle decreases by 2.8°. However, the energetic consequences of the additional electron are substantial; the neutral water molecule is 89.7 kJ/mol more stable than its anionic counterpart (see **Table 3.1** for relative energies).

The data from the AIM analysis indicate, that, for an isolated water molecule, the excess electron is largely accommodated on the hydrogen atoms, 0.27 electrons (e^-) each, with the remainder on the oxygen atom, 0.44 e^- (**Table 3.2**, page 60). In addition, the

spin population of each atom was computed, with an excess of 0.29 e^- for each hydrogen atom and 0.42 e^- on the oxygen (**Table 3.2**). As a visual aid, the spin population data can be plotted using the 0.001 a.u. spin density isosurface, as shown in **Figure 3.2**. It should be apparent from the spin density plot that the majority of the excess electron is located on the hydrogen atoms, confirming the trend observed in the AIM analysis.

Table 3.1: Relative energies (kJ/mol) of the anionic and neutral water clusters.

System		Anion[a]	Neutral
(H_2O)		0	-89.7
$(H_2O)_2$	**2A**	0	–
	2B	+1.4	-58.2
$(H_2O)_4$	**4A**	0	-46.5
	4B	+13.2	-57.2
$(H_2O)_6$	**6A**	0	+7.4
	6B	-34.9	-57.5
$(H_2O)_8$	**8A**	0	–
	8B	-12.5	-80.7

[a] Note that the relative energies have been defined within each cluster size such that the anionic cluster that forms a central cavity (i.e., **A** types) is assigned a relative energy of zero.

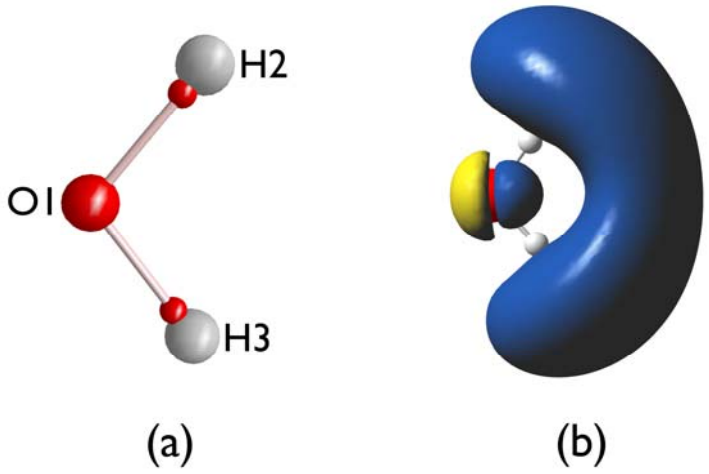


Figure 3.2: Molecular graph of (a) water, illustrating the bond critical points (small red spheres) between the nuclei and, (b) the +0.001 and -0.001 a.u. spin density isosurface on the water anion (blue and yellow lobes, respectively).

3.3.2 Anionic Water Dimers, $(H_2O)_2^-$

Two anionic water dimers were investigated following the optimization of water and its anion. One of these was a cage-like dimer in which the two water molecules align such that the entire dimer forms a flat plane with a central cavity between the monomers (**2A**$^-$). The other anionic dimer (**2B**$^-$) closely resembles the minimum energy structure of the neutral water dimer. Both structures are shown in **Figure 3.3**. It turns out that **2A**$^-$ is marginally more stable than **2B**$^-$ by 1.4 kJ/mol. Although the energy difference itself is negligible, it is interesting to note that the unusual cage-like arrangement is just as stable as the typical hydrogen-bonding scheme upon the addition of a single electron to the water dimer.

AIM analysis of the dimers reveals two very different distributions of the excess electron. The molecular graph of **2A**$^-$ exhibits a long-range O–O bond between the two water molecules. Due to the arrangement of the water molecules, all of the hydrogen

atoms are NHB hydrogen atoms, each accommodating 0.13 e^- while the oxygen atoms possess nearly twice as much with 0.24 e^- (**Table 3.2**). In contrast, **2B**$^-$ accommodates the excess electron on one of the water molecules in particular, the hydrogen-bond acceptor, with the two NHB hydrogen atoms each possessing 0.27 e^- and the oxygen 0.37 e^-. Essentially, **2A**$^-$ localizes the excess electron to the center of the cluster while **2B**$^-$ distributes it over the surface. Although the AIM analysis does not predict a NNA between the water molecules in **2A**$^-$, the two anionic dimers qualitatively typify the two different binding modes of the excess electron. One further point of interest is that in either case, the majority of the excess electron is accommodated on the NHB hydrogen atoms and the oxygen atoms as illustrated by the spin density plots of the dimers (**Figure 3.3**).

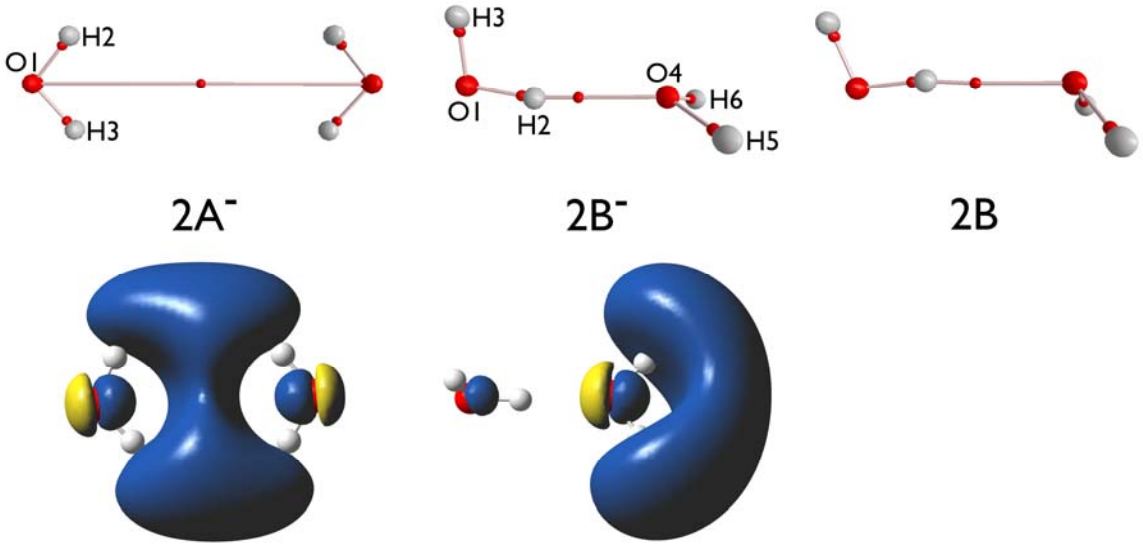


Figure 3.3: Molecular graphs and spin density plots of the anionic water dimers (**2A**$^-$ and **2B**$^-$) and the molecular graph of the neutral water dimer (**2B**).

Both anionic dimers were subsequently optimized as neutral clusters; however, only one common minimum energy structure was obtained. The preferred arrangement for the neutral water dimer, **2B**, is the typical hydrogen-bonding scheme (**Figure 3.3**). This structure is very similar to $\mathbf{2B^-}$ with some small changes in the geometric parameters. The addition of a single electron decreases the length of the hydrogen bond and increases the α angle (illustrated in **Figure 3.4**) in $\mathbf{2B^-}$ relative to **2B**. The neutral water dimer is by far the most stable of the dimer structures investigated; it is approximately 60 kJ/mol more stable than either of the dimer anions (**Table 3.1**).

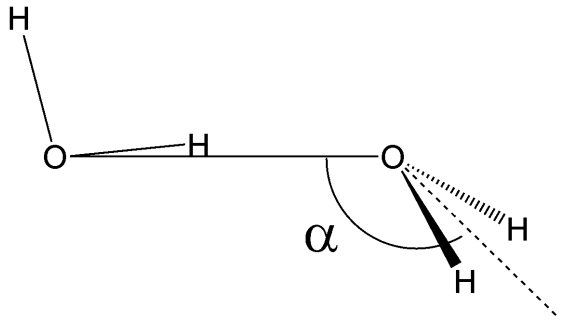


Figure 3.4: Neutral water dimer illustrating the α angle, the angle between the line connecting the oxygen atoms and the line bisecting the hydrogen atoms.

3.3.3 Anionic Water Tetramers, $(H_2O)_4^-$

Both of the anionic tetrameric structures investigated consist of four water molecules hydrogen bonded in a single ring, with each water molecule serving as both a hydrogen-bond donor and acceptor. The key difference between the two systems lies in the orientation of the NHB hydrogen atoms. In $\mathbf{4A^-}$, the NHB hydrogen atoms all point in the same direction on one side of the plane, whereas in $\mathbf{4B^-}$, they alternate sides of the

ring (**Figure 3.5**). Consequently, **4A**$^-$ and **4B**$^-$ possess a four-fold and two-fold axis of symmetry, respectively.

Analysis of the two structures reveals that **4A**$^-$ is more stable than **4B**$^-$ by 13.2 kJ/mol (**Table 3.1**). The relative energy of the two systems is in fact the opposite of what might be expected considering the resemblance of **4B**$^-$ to the preferred arrangement of the neutral tetramer. The fact that the electronic structure calculations predict a lower energy for the unusual **4A**$^-$ arrangement suggests it is somehow better able to accommodate the excess electron compared to the **4B**$^-$ arrangement.

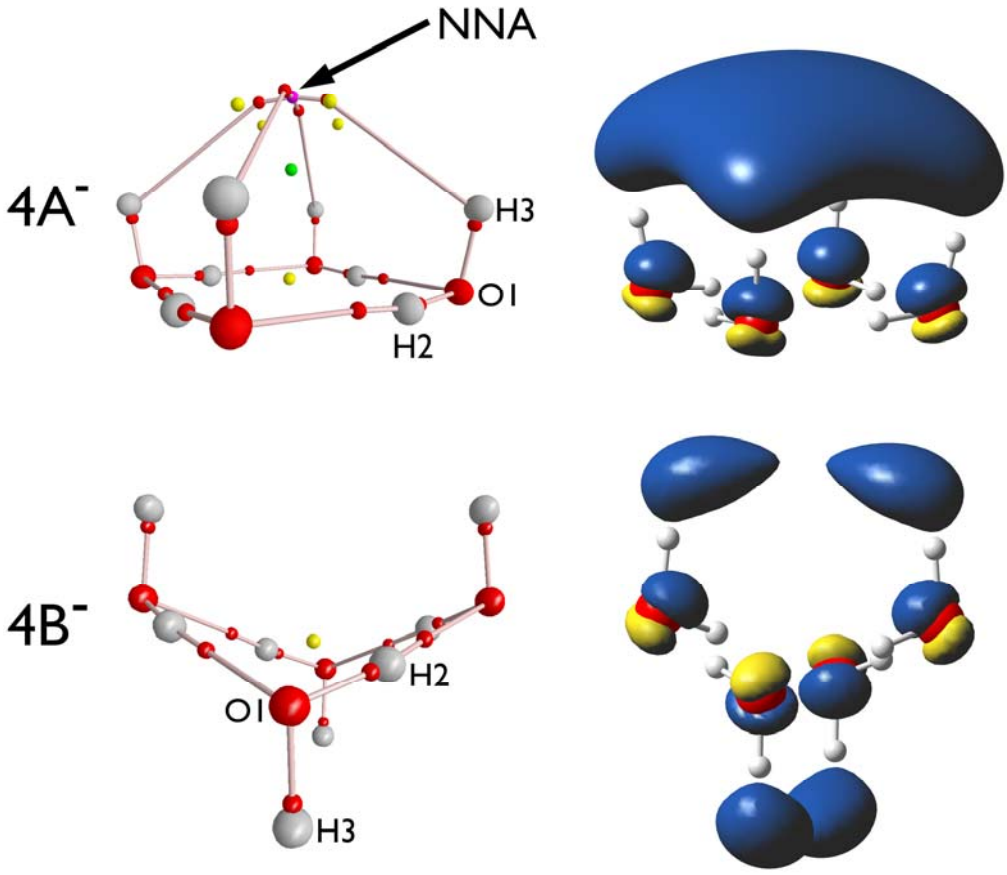


Figure 3.5: Molecular graphs and spin density plots of the anionic water tetramers. The pink sphere in **4A**$^-$ represents the non-nuclear attractor while the yellow and green spheres are the ring and cage critical points, respectively.

The reason for the apparent discrepancy in the energies becomes clear when the topological features of the two anionic clusters are compared. AIM analysis indicates that **4A**$^-$ possesses a non-nuclear attractor, the first such instance in this study. The molecular graph (**Figure 3.5**) reveals the NNA (pink sphere), located above the NHB hydrogen atoms, and also predicts several ring critical points (yellow spheres) and one cage critical point (green sphere), all of which are necessary to satisfy the Poincaré-Hopf relationship. On the other hand, **4B**$^-$ does not possess a NNA and only exhibits one ring critical point in the middle of the tetrameric ring.

Analysis of the non-nuclear attractor and its basin indicates that it has an electronic population of 0.03 e^- (and a spin population of 0.03), contributes 3.2 kJ/mol to the overall energy and occupies a volume of 11.8 a.u. (**Table 3.2**). [As a side note, in terms of the atomic volumes observed in the course of this study, the hydrogen-bonded H atoms typically have atomic volumes of 10 to 15 a.u., while the NHB hydrogen atoms occupy 30 to 90 a.u. and the oxygen atoms between 140 to 200 a.u.] Although the population of the pseudo-atomic basin and its energetic contribution are quite small, the fact that a NNA was detected at all is quite significant.

Further investigation of the atomic properties reveals similar results to those seen in the dimeric cases. The majority of the excess electron is accommodated on the oxygen atoms and the NHB hydrogen atoms. In **4A**$^-$, each NHB hydrogen atom accommodates 0.14 e^- of the excess electron in its atomic basin while the oxygen atoms possess slightly less at 0.12 e^-. Similarly, in **4B**$^-$ the NHB hydrogen atoms again possess 0.14 e^- and the oxygen atoms 0.11 e^- (see **Table 3.2** for charge difference and spin populations). Due to the arrangement of the water molecules, and more importantly the NHB hydrogen atoms,

structure **4A**$^-$ localizes the excess electron to the region above the plane, while **4B**$^-$ distributes the excess electron over the surface of the cluster. This can be seen in the spin density plots for the two systems (**Figure 3.5**).

Optimization of the two anionic structures as neutral tetramers led to two new structures that maintain roughly the same arrangements of the water molecules. The main difference between the neutral and anionic clusters is the position of the NHB hydrogen atoms with respect to the ring. In the anionic clusters the NHB hydrogen atoms are approximately perpendicular to the plane formed by the hydrogen-bonded ring. On the other hand, in the neutral structures the systems tend to flatten out as these hydrogen atoms move away from the center of the cluster (**Figure 3.6**). Additionally, the hydrogen-bonded ring has completely flattened out in **4B** and is no longer V-shaped like **4B**$^-$. The relative energy of the two neutral structures indicates that **4B** is more stable than **4A**, but only by approximately 10 kJ/mol. Both structures are significantly more stable than their anionic counterparts; **4A** by –46.5 kJ/mol and **4B** by –70.4 kJ/mol (**Table 3.1**). Although it is obvious that both neutral structures are more stable than the corresponding anion, it is worth pointing out that the unusual **4A** structure is less destabilized by the addition of a single electron than **4B**.

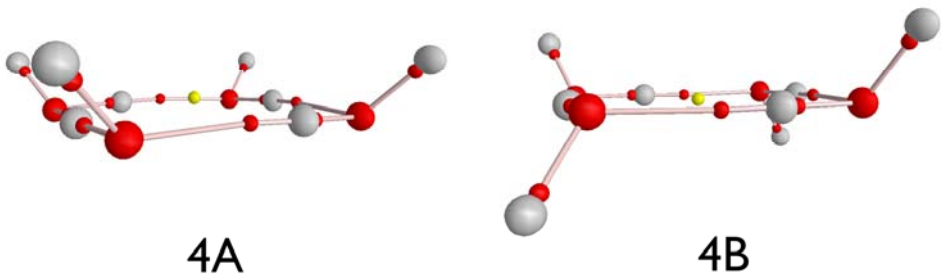


Figure 3.6: Molecular graphs of the neutral water tetramers.

3.3.4 Anionic Water Hexamers, $(H_2O)_6^-$

The two hexameric systems investigated consist of a highly symmetric cluster in which two trimers are stacked upon one another with several NHB hydrogen atoms directed toward the center of the cluster (**6A**$^-$) and a less symmetric cluster which is nearly shaped like a prism but is distorted at one end (**6B**$^-$) (**Figure 3.7**). The two stacked trimers in **6A**$^-$ are not mirror images of one another, but rather one trimer is rotated 60° relative to the other, maintaining a three-fold axis of symmetry.

The molecular graph of **6A**$^-$ exhibits a NNA at the center of the cluster. However, unlike the tetrameric case **4A**$^-$, this NNA possesses a significant fraction of the excess electron, with a population of 0.23 e$^-$, a pseudo-atomic energy of 21.5 kJ/mol and a volume of 93.3 au (**Table 3.2**). The spin density plot of **6A**$^-$ illustrates that the excess electron is essentially confined to the central cavity of the cluster (**Figure 3.7**), satisfying the historical notion of the cavity-bound hydrated electron. The remainder of the excess electron is found on the NHB hydrogen atoms (0.05 e$^-$ each) and the oxygen atoms (0.07 e$^-$ each). On the other hand, the prism-shaped hexamer **6B**$^-$ does not possess any NNAs. Inspection of the spin density plot and the atomic properties determined by AIM analysis indicate that this system largely accommodates the excess electron on a single water molecule with 0.23 and 0.28 e$^-$ on the hydrogen atoms and 0.30 e$^-$ on the oxygen (**Table 3.2**). This specific water molecule has two NHB hydrogen atoms directed away from the cluster, ultimately distributing the excess electron over the surface of the cluster, representing a surface-bound hydrated electron.

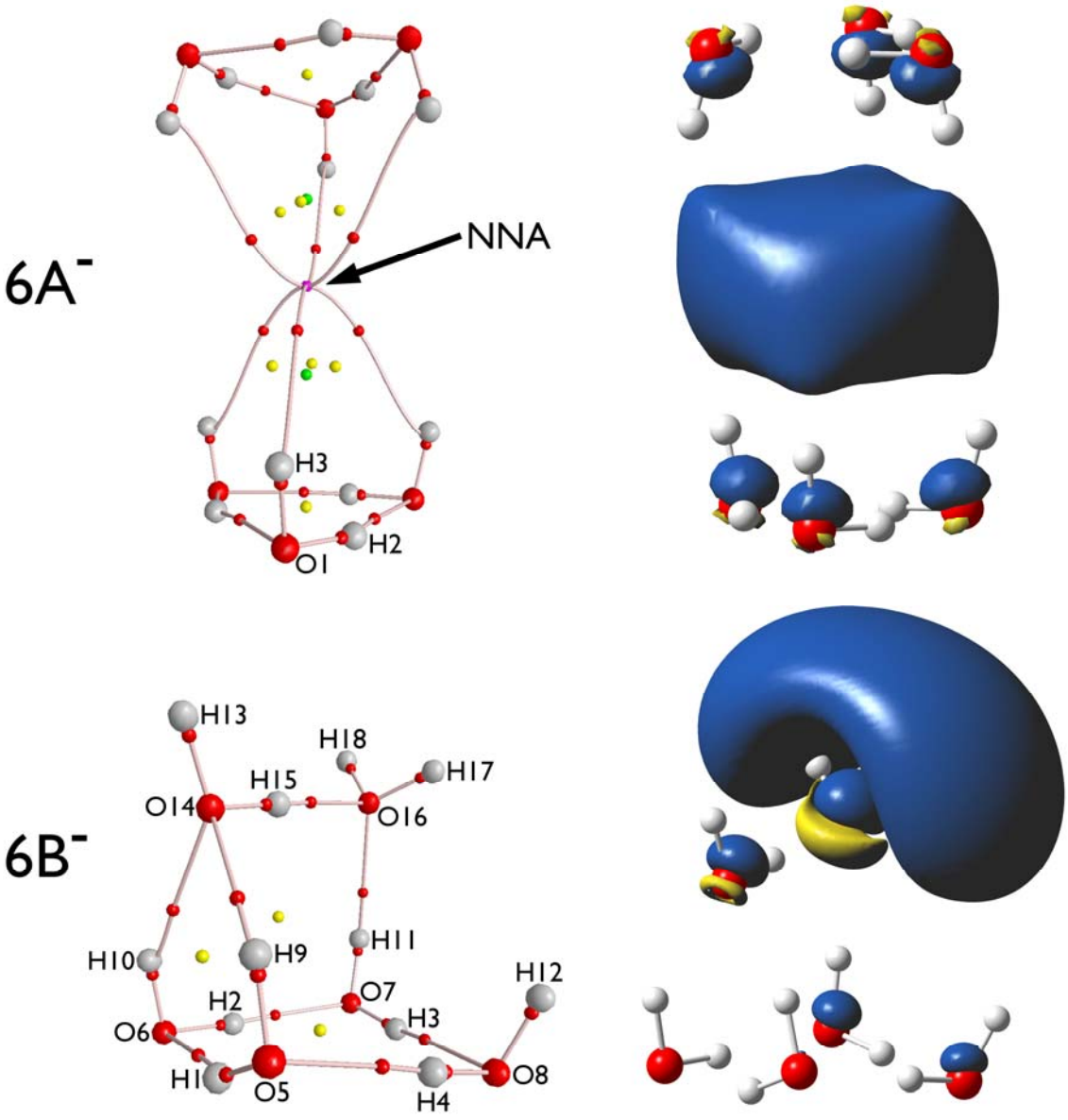


Figure 3.7: Molecular graphs and spin density plots of the anionic water hexamers.

As with the tetrameric case, both structures were subsequently optimized as neutral water clusters. One of the neutral clusters, **6A**, maintains a very similar arrangement of the water molecules compared to its corresponding anion. The cluster still consists of two stacked trimers, however, the two hydrogen-bonded planes are much

closer together in the neutral hexamer and the NHB hydrogen atoms have moved away from the central cavity, flattening out the trimers (**Figure 3.8**). The second neutral cluster, **6B**, experiences more significant changes in its geometry. Rather than forming a prism, the structure takes on a basket-like appearance, with a base of four water molecules hydrogen-bonded in a single ring and two water molecules above the plane diagonal to one another. This structure actually forms three rings and so encloses one cage critical point, whereas $\mathbf{6B}^-$ essentially remains open on one side.

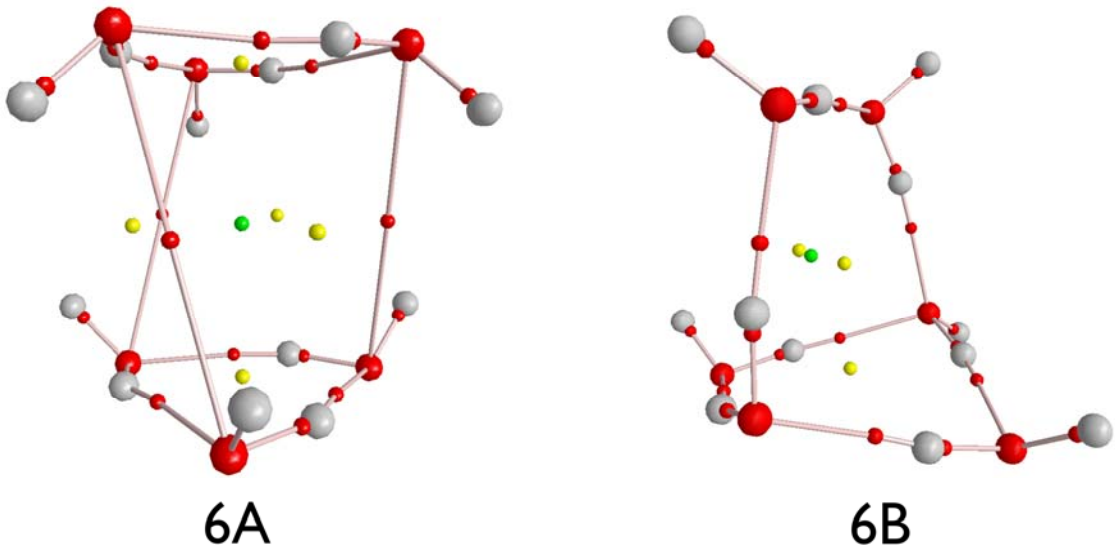


Figure 3.8: Molecular graphs of the neutral water hexamers.

The relative energies of the hexameric clusters, both neutral and anionic, exhibit an interesting trend: $\mathbf{6A}^-$ is actually more stable than its neutral counterpart by 7.4 kJ/mol. However, comparing the two anionic clusters, $\mathbf{6B}^-$ is more stable than $\mathbf{6A}^-$ by 34.9 kJ/mol and the neutral structure **6B** is the most stable hexamer of all, 22.6 kJ/mol more stable than its anion (**Table 3.1**). The fact that the excess electron actually has a

stabilizing effect (or at least a negligible effect) in the **6A⁻/6A** system while it results in a marked destabilization in the **6B⁻/6B** system is noteworthy. This is the only cluster size where this trend is observed.

3.3.5 Anionic Water Octamers, $(H_2O)_8^-$

The largest systems investigated in this study were two octameric clusters formed from stacked tetramers. The more symmetric of the two systems, **8A⁻**, consists of two stacked tetramers like **4A⁻** with the NHB hydrogen atoms directed toward a central cavity and one tetramer is rotated 45° relative to the other. The second cluster, **8B⁻**, consists of two tetramers like **4B** forming a cube such that half of the water molecules have NHB hydrogen atoms directed away from the center of the cluster while the other half do not have any NHB hydrogen atoms and act as double hydrogen-bond donors. Both structures are shown in **Figure 3.9**.

AIM analysis of **8A⁻** reveals a NNA at the center of the cluster, positioned between the NHB hydrogen atoms, analogous to the result obtained for the hexameric system **6A⁻**. The data indicate that this NNA also possesses a significant fraction of the excess electron, with a population of 0.22 e^-, a pseudo-atomic energy of 17.9 kJ/mol and a volume of 89.0 au. In agreement with the previous systems studied, the remainder of the excess electron is largely accommodated on the NHB hydrogen atoms and the oxygen atoms, approximately 0.05 e^- and 0.04 e^-, respectively. On the other hand, **8B⁻** does not exhibit a NNA at the center of the cluster but rather a cage critical point. The excess electron is distributed amongst the NHB hydrogen atoms and the oxygen atoms on the surface of the cluster (see **Table 3.2** for the populations). The spin density plots

graphically illustrate the key difference between the two systems in terms of the distribution of the excess electron, mirroring the trend observed in the AIM analysis (**Figure 3.9**).

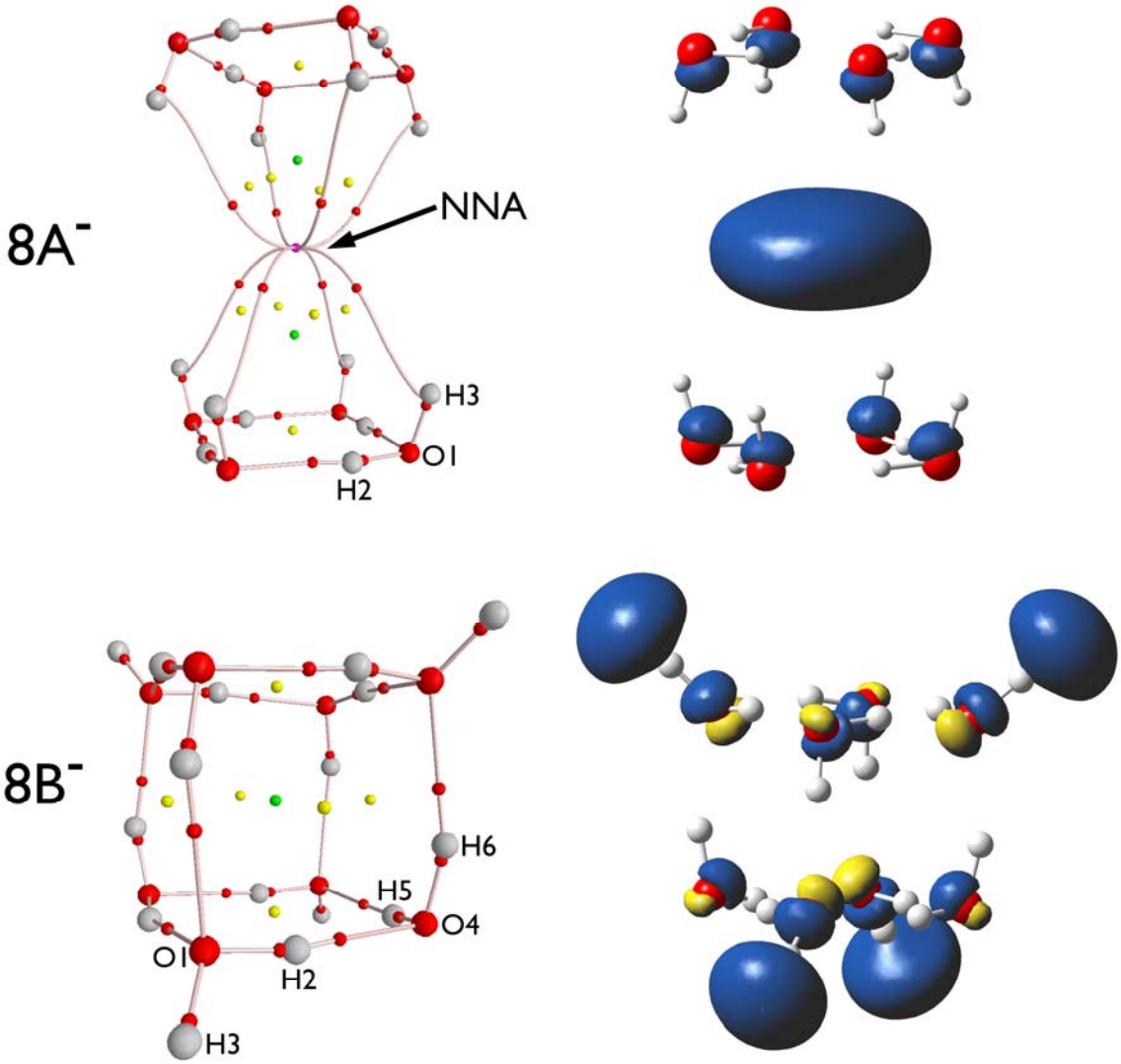


Figure 3.9: Molecular graphs and spin density plots of the anionic water octamers.

Optimization of the two anionic species as neutral clusters yields one common structure, **8B**, which has the same arrangement of water molecules as **8B**$^-$ but with some

minor changes in bond lengths and angles (**Figure 3.10**). The relative energies of all the eight-membered species reveals that the neutral structure is considerably more stable than either anion by at least 60 kJ/mol. More interesting is the fact that **8B**$^-$ is merely 12.5 kJ/mol more stable than **8A**$^-$, a less significant energy difference than may be expected considering the unusual structure adopted by **8A**$^-$. This correlates well with the observations made by Khan,[69] who observed both a cavity- and surface-bound state, depending on the arrangement of the water molecules, that were nearly identical in energy. It is possible that as the clusters get larger, and there are more water molecules available to accommodate the excess electron, the energy difference between the two models decreases.

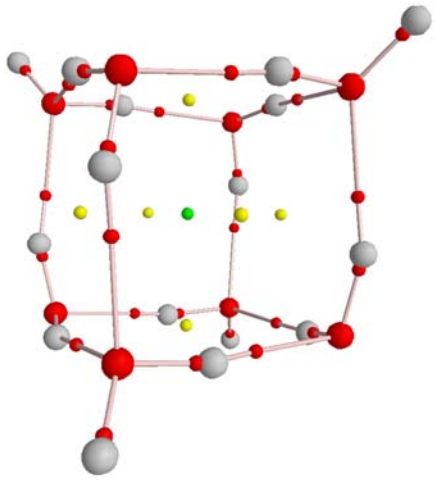


Figure 3.10: Molecular graph of the neutral water octamer.

Table 3.2: Selected atomic properties of the anionic water clusters. Charges and spin populations are in atomic units and energy in kJ/mol.

Cluster	Atom	Atomic Charge	Charge Difference[a]	Spin Population	Energy[b]
$(H_2O)^-$	O1	-1.6047	-0.4449	0.4197	
	H2	0.3072	-0.2727	0.2850	
	H3	0.3070	-0.2729	0.2852	
2A$^-$	O1	-1.3992	-0.2409	0.2036	
	H2	0.4537	-0.1255	0.1442	
	H3	0.4529	-0.1263	0.1448	
2B$^-$	O1	-1.2806	-0.0613	0.0369	
	H2	0.6491	+0.0237	0.0032	
	H3	0.5461	-0.0292	0.0216	
	O4	-1.5476	-0.3749	0.3526	
	H5	0.3223	-0.2736	0.2872	
	H6	0.3239	-0.2720	0.2859	
4A$^-$	O1	-1.3451	-0.1234	0.1020	
	H2	0.6443	-0.0018	0.0022	
	H3	0.4356	-0.1399	0.1609	
	NNA	**-0.0281**	**-0.0281**	**0.0277**	**-3.2**
4B$^-$	O1	-1.3423	-0.1094	0.1028	
	H2	0.6442	-0.0014	0.0037	
	H3	0.4596	-0.1380	0.1420	
	O1	-1.2853	-0.0733	0.0502	
6A$^-$	H2	0.6331	-0.0018	0.0025	
	H3	0.5281	-0.0486	0.0732	
	NNA	**-0.2318**	**-0.2318**	**0.2257**	**-21.5**
6B$^-$	H1	0.6345	-0.0025	0.0001	
	H2	0.6465	+0.0070	0.0002	
	H3	0.6347	-0.0038	0.0005	
	H4	0.6506	-0.0027	0.0002	
	O5	-1.2763	-0.0204	0.0053	
	O6	-1.2759	-0.0272	0.0048	
	O7	-1.2868	-0.0378	0.0145	
	O8	-1.2649	-0.0325	0.0148	
	H9	0.6274	+0.0134	0.0013	
	H10	0.6087	+0.0107	0.0014	
	H11	0.6274	+0.0249	0.0037	
	H12	0.5682	-0.0011	0.0229	
	H13	0.5471	-0.0443	0.0460	
	O14	-1.3159	-0.0821	0.0597	
	H15	0.6506	+0.0161	0.0046	
	O16	-1.4779	-0.3026	0.2830	
	H17	0.3294	-0.2756	0.2935	
	H18	0.3839	-0.2257	0.2362	
8A$^-$	O1	-1.2818	-0.0550	0.0372	
	H2	0.6496	-0.0005	0.0009	
	H3	0.5362	-0.0400	0.0589	
	NNA	**-0.2164**	**-0.2164**	**0.2101**	**-17.9**
8B$^-$	O1	-1.3647	-0.0980	0.1018	
	H2	0.6600	-0.0038	0.0021	
	H3	0.5105	-0.0900	0.0878	
	O4	-1.3253	-0.0553	0.0528	
	H5	0.6353	-0.0008	0.0031	
	H6	0.6348	-0.0021	0.0032	

[a] Charge difference was determined as the difference between the atomic charge in the anion less the atomic charge in the corresponding neutral cluster in the anionic geometry.
[b] Only the total energies of the NNA basins (in kJ/mol) are reported here, since the energies of the other atoms are of too great a magnitude to be reported in kJ/mol.

3.4 Conclusions

The distribution of an excess electron in small water clusters has been investigated using electronic structure calculations and the theory of atoms in molecules. The results indicate that the mode of binding for the excess electron, whether it is surface- or cavity-bound, is highly dependent upon the arrangement of the molecules in the cluster.

Two types of clusters were investigated, those that direct several non-hydrogen-bonded hydrogen atoms towards a central cavity in the cluster (labeled **A**) and those that direct them away from the surface (labeled **B**). The former typically possess higher symmetry than the latter. Those clusters that can form a central cavity (**2A**$^-$, **4A**$^-$, **6A**$^-$, and **8A**$^-$) favour the cavity-bound model, the classic notion of the hydrated electron. The minimum requirement to find a non-nuclear attractor in the electron density appears to be four water molecules with their NHB hydrogen atoms all oriented in the same direction (**4A**$^-$). However, only the larger clusters (**6A**$^-$ and **8A**$^-$) exhibit NNAs with appreciable electronic populations and energies, upwards of 0.2 electrons and 20 kJ/mol, respectively. Amongst all of the cavity-bound clusters investigated, the excess electron was always largely accommodated on the NHB hydrogen atoms and the oxygen atoms, never on the hydrogen-bonding hydrogen atoms.

Conversely, the less symmetric clusters, which resemble the lowest-energy neutral structure for a given cluster size, tend to distribute the excess electron over several atoms at the surface of the cluster, examples of the surface- or dipole-bound model. These systems either distribute the excess electron over the entire surface (**4B**$^-$ and **8B**$^-$) or over one particular water molecule at the surface (**2B**$^-$ and **6B**$^-$). Most

importantly, in any of the surface-bound states observed, the excess electron is accommodated on the NHB hydrogen atoms and the oxygen atoms.

In terms of the calculated electronic energies (**Table 3.1**), it appears that the unusual **A**-type structures can more easily accommodate the excess electron than the **B**-type. This is evidenced by the fact that the energy difference between the anion and its corresponding neutral structure is generally smaller for the **A**-type structures. Furthermore, the difference between the **A**- and **B**-type structures within each cluster size is also quite small.

In conclusion, regardless of whether a surface- or cavity-bound state was found, the distribution of the hydrated electron is highly dependent upon the orientation of the non-hydrogen-bonded hydrogen atoms. For those clusters in which several NHB hydrogen atoms are directed toward a central cavity, the cavity-bound state prevails, otherwise the surface-bound state is observed.

Chapter 4: Characterization of the H···NNA Bond in Anionic Water Clusters

Taylor, A.; Boyd, R. J. *Phys. Chem. Chem. Phys*. **2008**, 10, 6814-6819.

Reproduced by permission of the PCCP Owner Societies.

4.1 Introduction

As outlined in the introduction to Chapter 3, solvated electrons, and specifically hydrated electrons, have been the focus of much experimental and theoretical research for many years now. And yet, there is still no general consensus about the exact structure of the excess electron within small water clusters. The results reported in Chapter 3 suggest that the distribution of the excess electron, whether it is cavity- or surface-bound, is directly dependent upon the orientation of the non-hydrogen-bonded (NHB) hydrogen atoms. For those cases in which several NHB hydrogen atoms are directed towards one location, a non-nuclear attractor (NNA) was observed in the molecular graph, indicating that the cavity-bound state is preferred for that arrangement.[78]

According to the theory of atoms in molecules (AIM), these non-nuclear attractors behave topologically in the same manner as nuclei.[4-6] Thus, they are bounded by a zero-flux surface, which ultimately encloses their pseudo-atomic basin. Electron density may reside within this basin and various pseudo-atomic properties can be determined for the NNA by calculating the appropriate volume integral over said basin. Furthermore, since the NNAs behave just like any other attractor in the electron density, they are capable of forming bonding interactions with their neighbour atoms. The purpose of the

investigation reported in this chapter is to characterize the bonding interaction between the NNA and the NHB hydrogen atoms (H···NNA) that was observed in the molecular graphs presented in Chapter 3.

In order to characterize this novel bond, various aspects of the topological analysis can be exploited further. First of all, it should be re-iterated that the electron density exhibits lines of maximum electron density between two bonded attractors. These lines are traced from the bond critical points (BCP) to the attractors, and rigorously define the presence of a bonding interaction. The strength and order of the bond can be determined by further analysis of the BCP involved.

The two quantities most often used are the electron density at the BCP, ρ_{BCP}, and the Laplacian, $\nabla^2\rho_{BCP}$.[4,51] Considering the electron density first, shared interactions are generally characterized by an accumulation of electron density between the nuclei, while conversely closed-shell interactions typically experience a depletion. The accumulation of electron density observed for shared interactions, which include covalent and polar bonds, results in large values of ρ_{BCP}, while closed-shell interactions like ionic and hydrogen bonds as well as van der Waals interactions exhibit small values of ρ_{BCP}. Thus, the magnitude of the electron density at the bond critical point can be directly correlated with the strength of the bond: the stronger the bond, the larger ρ_{BCP}. In addition, the sign of the Laplacian can give a further indication as to the type of bond present. The Laplacian is determined from the eigenvalues (λ) of the Hessian of the electron density and is defined as the sum of these, $\nabla^2\rho_{BCP} = \lambda_1 + \lambda_2 + \lambda_3$. Note, that the eigenvalues are labeled such that $\lambda_1 \leq \lambda_2 \leq \lambda_3$. Using this definition, shared interactions are characterized

by negative values of the Laplacian at the BCP ($\nabla^2\rho_{BCP} < 0$), while closed-shell interactions possess positive Laplacian values ($\nabla^2\rho_{BCP} > 0$).

The energy of the bonding interaction is composed of both kinetic and potential energy contributions. The exact relationship is described by a local expression of the virial theorem,[4]

$$\frac{1}{4}\nabla^2\rho(r) = 2G(r) + V(r) \qquad \textbf{(4.1)}$$

where $G(r)$ is the electronic kinetic energy density, and is defined as,

$$G(r) = \frac{1}{2}N\int d\tau \nabla\psi^* \cdot \nabla\psi \qquad \textbf{(4.2)}$$

and $V(r)$, the electronic potential energy density, is defined as,

$$V(r) = -r \cdot \nabla \cdot \ddot{\sigma} + \nabla \cdot (r \cdot \ddot{\sigma}) \qquad \textbf{(4.3)}$$

where $\ddot{\sigma}$ is the stress tensor. Note that $G(r)$ is everywhere positive while $V(r)$ is always negative. Thus, equation **4.1** predicts that the sign of the Laplacian is dictated by whether the kinetic or potential energy density is in excess of the 2:1 ratio. However, it has been suggested by Cremer and Kraka,[116] that the total electronic energy density, $H(r)$, should be used to compare the kinetic and potential energy densities. $H(r)$ is defined as the sum of these two terms,

$$H(r) = G(r) + V(r) \qquad \textbf{(4.4)}$$

When $H(r)$ is evaluated at the BCP, any interaction that involves a significant sharing of electrons will exhibit a negative H_{BCP} and the 'covalent character' of the bond will be reflected in the magnitude of H_{BCP}.

One additional parameter that can be used to characterize a bond is its ellipticity, ε. This measures the accumulation of electron density in a plane containing the bond. The value of the ellipticity is obtainable from the eigenvalues of the Hessian (analogous to the Laplacian) and is defined as $\varepsilon = (\lambda_1/\lambda_2 - 1)$. The magnitude of ε indicates how stable the bond is to small geometric changes: the larger ε, the less stable the bond.[5]

Although it is not always a tool used to characterize a bond, the bond path length (BPL) can be an interesting property to examine. Since AIM is able to predict curved bond paths (which are often seen in weak closed-shell interactions or strained cyclic systems), the degree of curvature can be determined by comparing the BPL to the linear distance between the nuclei, R_e. Obviously, for any curved bond path, the BPL will necessarily be larger than R_e.

Since the H···NNA interaction involves the NHB hydrogen atoms, questions arose as to whether this could be regarded as some type of novel hydrogen bond. Koch and Popelier have suggested several criteria within the AIM framework that must be met in order for an interaction to be considered a hydrogen bond.[117] The first criterion, obviously, is that a bond critical point must be found between the nuclei of interest. They also suggest a range of ρ_{BCP} and $\nabla^2\rho_{BCP}$ values that are acceptable. However, these values were determined for neutral systems and are therefore not necessarily applicable to the current study. In addition, the hydrogen atom involved in the bond typically experiences various changes in its atomic properties. These include: loss of electronic charge, energetic destabilization, decreased volume and decreased dipolar polarization and are all measured relative to the hydrogen atom in the isolated monomer. All of the tools described above were used to investigate the H···NNA bond in the small anionic water clusters.

4.2 Computational Details

All geometry optimizations were performed at the UMP2(full)/6-31++G(2d,2p) level of the theory using the Gaussian 03 suite of programs.[47] Frequency calculations were also performed at this level to ensure that the structures were minima and to obtain the zero-point vibrational energy (ZPVE). Various atomic and bond critical point properties were determined using the theory of atoms in molecules as implemented in the AIM2000 program.[113]

4.3 Results and Discussion

The anionic water clusters investigated in this study include the structures reported in Chapter 3 that exhibited NNAs in their molecular graphs (**4A**$^-$, **6A**$^-$, **8A**$^-$) as well as a decameric cluster. The decamer consists of two stacked pentamer rings, with the NHB hydrogen atoms directed toward the center of the cluster, analogous to **6A**$^-$ and **8A**$^-$. The two pentamers are not aligned perfectly, but rather one is rotated 36° relative to the other ring, resulting in a five-fold rotation axis. A single, neutral water molecule was also included in the investigation to be used for comparison when necessary. All optimized structures are shown in **Figure 4.1**. At this point, a note should be made about the labeling scheme used throughout this chapter. The NHB hydrogen atoms are labeled as H_a, while the hydrogen-bonded H atoms (those in the plane of the rings) are labeled H_b.

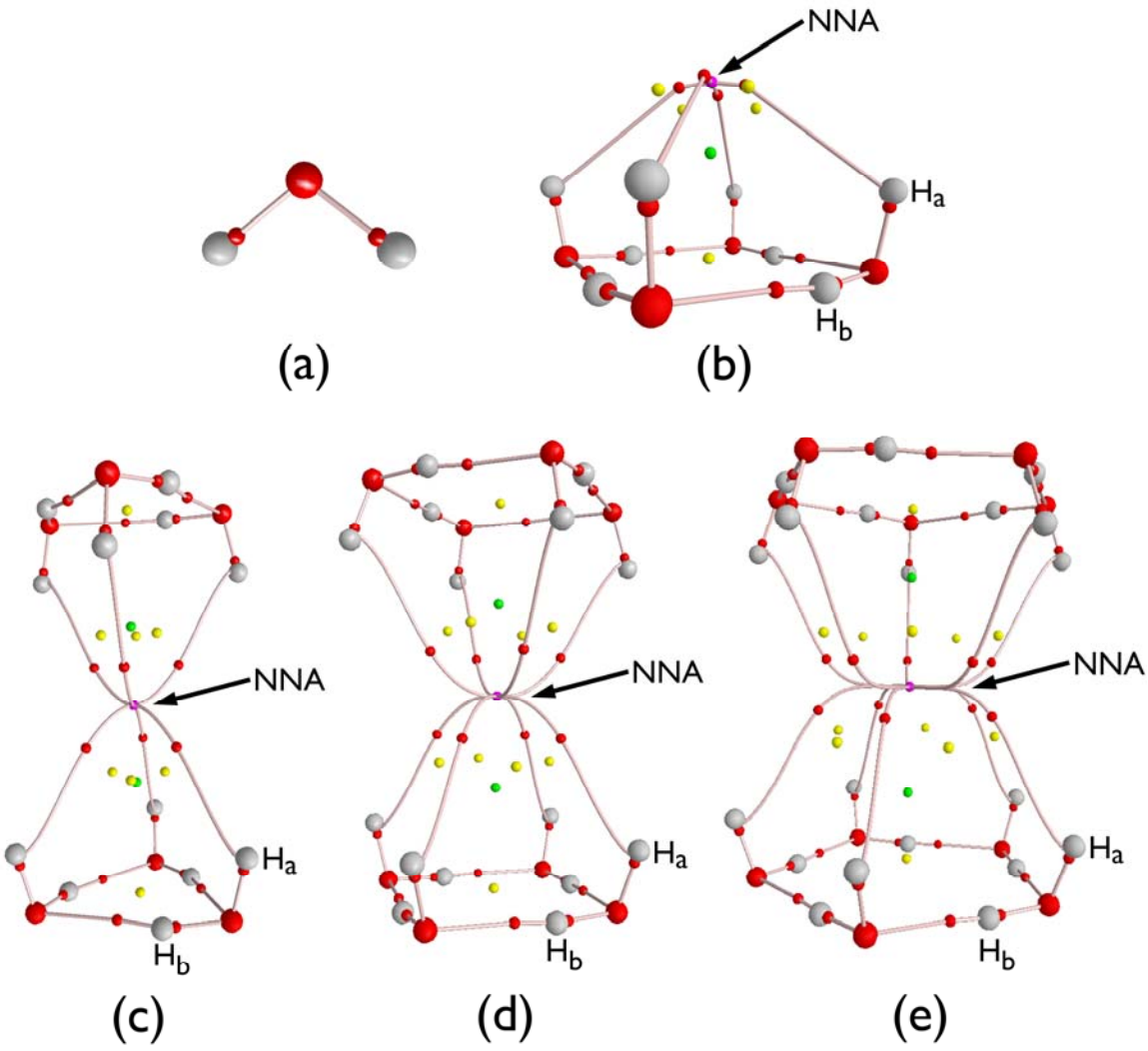


Figure 4.1: Molecular graphs of (a) a neutral water molecule, as well as the anionic water clusters illustrating the non-hydrogen-bonded H atoms (H_a), the hydrogen-bonded H atoms (H_b) and the non-nuclear attractors (NNA): (b) tetramer, (c) hexamer, (d) octamer, and (e) decamer.

Binding energies (BE) for the anionic water clusters were calculated as the difference between the electronic energy of the anionic water cluster and the energy of an appropriate number of water molecules [e.g. $BE_{(tetramer)} = E_{(anionic\ tetramer)} - 4E_{(water)}$]. All energies were corrected for zero-point vibrations and the calculated binding energies are shown in **Table 4.1**. In every cluster, each water molecule participates in a

conventional hydrogen bond as well as an interaction with the NNA via the NHB hydrogen atom. Note from **Table 4.1** that all of the calculated binding energies are negative, indicating that the anionic water clusters are more stable than the separated water molecules. This stabilization is the result of the conventional hydrogen bonds, as well as the interaction with the NNA.

In order to directly compare the calculated binding energies amongst the difference clusters investigated, the binding energies were normalized by accounting for the number of water molecules in a given cluster. The normalized binding energy, BE_{Norm}, is thus defined as the calculated BE divided by the number of water molecules in the cluster. This value gives a better indication of the stabilization that is imparted per water molecule. Inspection of the last column in **Table 4.1** reveals that the tetrameric cluster is significantly less stable than the other clusters under investigation. This result is perhaps not a surprise considering how different the tetrameric cluster is from the others (e.g. it does not possess a central cavity and the NNA only has a very small electronic population and energy).

Table 4.1: Binding energy (BE) and normalized binding energies (BE_{Norm}) for each anionic water cluster (kJ/mol).

Cluster	BE	BE_{Norm}
$(H_2O)_4^-$	-34.60	-8.65
$(H_2O)_6^-$	-95.99	-16.00
$(H_2O)_8^-$	-176.22	-22.03
$(H_2O)_{10}^-$	-226.36	-22.64

The binding energies are plotted versus cluster size in **Figure 4.2**, revealing that there is a linear relationship between these two parameters. This suggests that as the cluster size increases, as is expected.

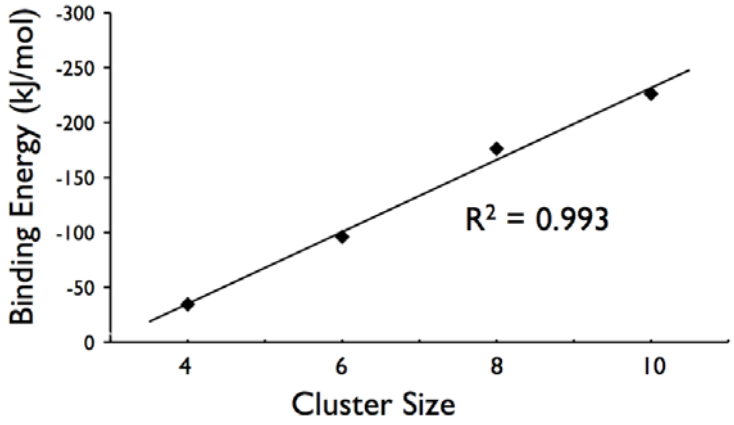


Figure 4.2: Relationship between binding energy and cluster size for the cavity-bound anionic water clusters.

Analysis of the electron density using the theory of atoms in molecules produces the molecular graphs shown in **Figure 4.1**. Further analysis of the individual bond critical points was performed to gain insight about the bonds present. Various BCP properties are provided in **Table 4.2**.

The BCP data suggest several interesting trends for the three different types of bonds present. Examining the values of ρ_{BCP} reveals that the $H_a \cdots NNA$ bonds are approximately two orders of magnitude smaller than the covalent bonds (O–H) and one order smaller than the hydrogen bonds $(O \cdots H_b)$. Since it has already been established that the magnitude of ρ_{BCP} correlates directly with the strength of a bond,[4,51] the $H_a \cdots NNA$ interaction can be described as a very weak bond. Additionally, the small positive $\nabla^2\rho_{BCP}$ values observed for this bond further characterize it as a closed-shell interaction.

Table 4.2: Selected properties of the bond critical points in water and the cavity-bound anionic water clusters. R_e and BPL are in angstroms (Å), ε is dimensionless and the remaining quantities are in atomic units.

System	Bond	R_e	BPL	ρ_{BCP}	$\nabla^2\rho_{BCP}$	G_{BCP}	V_{BCP}	H_{BCP}	ε
H_2O	O–H	0.958	0.921	0.3719	-2.2909	0.0904	-0.7536	-0.6631	0.0244
$(H_2O)_4^-$	O–H_a	0.968	0.926	0.3624	-2.2103	0.0895	-0.7316	-0.6421	0.0200
	O–H_b	0.979	0.942	0.3429	-2.1535	0.0820	-0.7025	-0.6204	0.0203
	O…H_b	1.831	1.873	0.0339	0.0959	0.0262	-0.0285	-0.0022	0.0435
	H_a…NNA	2.572	2.778	0.0015	-0.0003	0.0001	-0.0001	-0.0001	6.1787
$(H_2O)_6^-$	O–H_a	0.963	0.926	0.3643	-2.2402	0.0889	-0.7378	-0.6489	0.0208
	O–H_b	0.974	0.937	0.3517	-2.2127	0.0836	-0.7203	-0.6368	0.0211
	O…H_b	1.942	1.984	0.0272	0.0784	0.0214	-0.0232	-0.0018	0.0735
	H_a…NNA	2.805	2.926	0.0015	0.0001	0.0001	-0.0001	-0.0001	0.0416
$(H_2O)_8^-$	O–H_a	0.963	0.926	0.3660	-2.2403	0.0899	-0.7399	-0.6500	0.0200
	O–H_b	0.979	0.942	0.3424	-2.1619	0.0815	-0.7035	-0.6220	0.0203
	O…H_b	1.804	1.847	0.0358	0.1029	0.0280	-0.0302	-0.0023	0.0489
	H_a…NNA	3.138	3.366	0.0011	0.0001	0.0001	-0.0001	-0.00003	0.1450
$(H_2O)_{10}^-$	O–H_a	0.963	0.926	0.3678	-2.2477	0.0908	-0.7435	-0.6527	0.0200
	O–H_b	0.984	0.942	0.3405	-2.1609	0.0806	-0.7014	-0.6208	0.0200
	O…H_b	1.773	1.815	0.0379	0.1105	0.0299	-0.0322	-0.0023	0.0515
	H_a…NNA	3.498	3.874	0.0008	0.00004	0.00003	-0.00005	-0.00002	0.4292

Comparing the various energetic contributions to the Laplacian at the BCP (i.e., the electronic kinetic ($G(r)$) and potential ($V(r)$) energy densities) can also provide additional information. As mentioned in the introduction, the magnitude of the total electronic energy density $H(r)$ (equation **4.2**) indicates the covalent character of a bond. From **Table 4.2**, the covalent O–H bonds exhibit relatively large H(r) values while the hydrogen bonds and the $H_a\cdots NNA$ interactions are considerably smaller. In fact, since the kinetic and potential energy densities are approximately equal for the $H_a\cdots NNA$ interaction, $H(r)$ is necessarily zero and the bond essentially has no covalent character.

Examining the bond for the unusual $H_a\cdots NNA$ interaction reveals that it is considerably longer than the covalent or hydrogen bonds present in the cluster. The

measured values range between 2.5 to 3.5 Å depending on the cluster, and appears to increase as the size of the cluster increases. Furthermore, the bond itself is noticeably curved. This can been seen from the molecular graphs in **Figure 4.1** or inferred from the fact that there is a measurable difference between the internuclear distance (R_e) and the bond path length as reported in **Table 4.2**. Although there is some difference between these two values for the other bonds present, the difference between R_e and BPL for the $H_a \cdots NNA$ interaction ranges between 0.2 to 0.4 Å, almost a whole order of magnitude larger than the covalent or hydrogen bonds (difference always less than 0.05 Å).

The final characteristic of the bond that was examined was the ellipticity, ε. This value can be used to indicate the relative stability of a bond with respect to small geometric changes. Upon inspection of the final column in **Table 4.2**, several pieces of data particularly stand out. Firstly, all of the covalent O–H bonds exhibit roughly the same ellipticity value of 0.02. Comparing the hydrogen bonds found in the four different clusters, although all of the values have the same order of magnitude, it is interesting to note that there is some variability in the values. The tetramer and octamer (which is composed of two tetramers) exhibit the same value around 0.04 and the decamer is quite similar at 0.05. On the other hand, the hexamer exhibits a larger value of 0.07. This implies that the hydrogen bonds in the hexamer are less stable than those found in the other clusters. This trend may not be surprising considering that the cyclic trimers that form the hexamer are rather strained and the individual water molecules are distorted from the optimal tetrahedral arrangement. This distortion could weaken the hydrogen bond, adding a small amount of instability compared to the other clusters.

However, the variability in the hydrogen bond ellipticities is minor compared to that observed in the $H_a \cdots NNA$ bonds. Ignoring the tetramer for the moment, there is a noticeable increase in the ε value as the size of the cluster increases. The hexameric cluster actually exhibits an ellipticity for the $H_a \cdots NNA$ bond on par with the hydrogen bonds. The octamer and decamer exhibit larger ε values by up to a full order of magnitude. This suggests that the $H_a \cdots NNA$ bond in the hexameric cluster is quite stable while in the larger clusters it is less so. While all of the larger clusters have $H_a \cdots NNA$ ellipticities much smaller than one, the tetramer has a calculated value of 6. Such a large value of ε indicates that the $H_a \cdots NNA$ is very unstable with respect to small geometric changes. Furthermore, inspection of the molecular graphs in **Figure 4.1** reveals that the $H_a \cdots NNA$ bond critical points are in close proximity to nearby ring critical points. As the distance between a BCP and a RCP decreases, they are in danger of merging into one degenerate critical point, as is seen in bifurcation processes.[4] Comparing the measured distances between the BCPs and the RCPs in the clusters reveals that the tetramer is significantly smaller than the rest: tetramer, 0.56 Å; hexamer, 0.83 Å; octamer, 0.87 Å and decamer, 0.98 Å. Thus, on the basis of the calculated ε values, and the relative proximity of the BCPs and RCPs in the tetrameric cluster, this structure would not likely survive small changes in geometry.

The final portion of this investigation looked at the atomic properties of the hydrogen atoms using the theory of atoms in molecules. In order to determine whether the $H_a \cdots NNA$ interaction represents some type of novel hydrogen bond, the criteria established by Koch and Popelier were applied.[117] The atomic properties that were determined were: charge, energy, dipole and volume. These properties for the hydrogen

atoms in the anionic cluster (H_a and H_b) as well as the neutral water molecule can be found in **Table 4.3**.

Table 4.3: Selected atomic properties for the hydrogen atom in water as well as the two types of hydrogen atoms in the anionic, cavity-bound water clusters. All entries are in atomic units.

System	Atom	Charge	Energy	Dipole	Volume
H_2O	H	0.590	-0.363	0.169	21.24
$(H_2O)_4^-$	H_a	0.437	-0.363	0.680	59.48
	H_b	0.644	-0.325	0.121	12.31
$(H_2O)_6^-$	H_a	0.528	-0.357	0.438	42.87
	H_b	0.633	-0.333	0.134	14.23
$(H_2O)_8^-$	H_a	0.536	-0.360	0.322	34.62
	H_b	0.650	-0.323	0.117	11.77
$(H_2O)_{10}^-$	H_a	0.539	-0.362	0.313	27.48
	H_b	0.656	-0.320	0.144	13.86

These results are also graphically summarized in **Figure 4.3**. The criteria set out by Koch and Popelier predict that a hydrogen-bonded hydrogen atom should experience a loss of electronic charge, energetic destabilization, decreased dipolar polarization and decreased volume. The bar graphs in **Figure 4.3** clearly illustrate that the two different hydrogen atoms in the anionic clusters exhibit the opposite behaviour. Compared to the neutral water molecule, the hydrogen-bonded hydrogen atoms (H_b) follow the criteria set out while the NHB hydrogen atoms (H_a), those involved in the $H_a \cdots NNA$ interaction, appear to systematically defy the criteria. Thus, on the basis of these results, the $H_a \cdots NNA$ interaction cannot be classified as a novel hydrogen bond. Rather, the data obtained in this study suggest that it is best described as a weak, closed-shell interaction, within the AIM framework. More specifically, it is probably best described as a dipole-

ion interaction between the positively charged NHB hydrogen atoms and the negatively charged non-nuclear attractor.

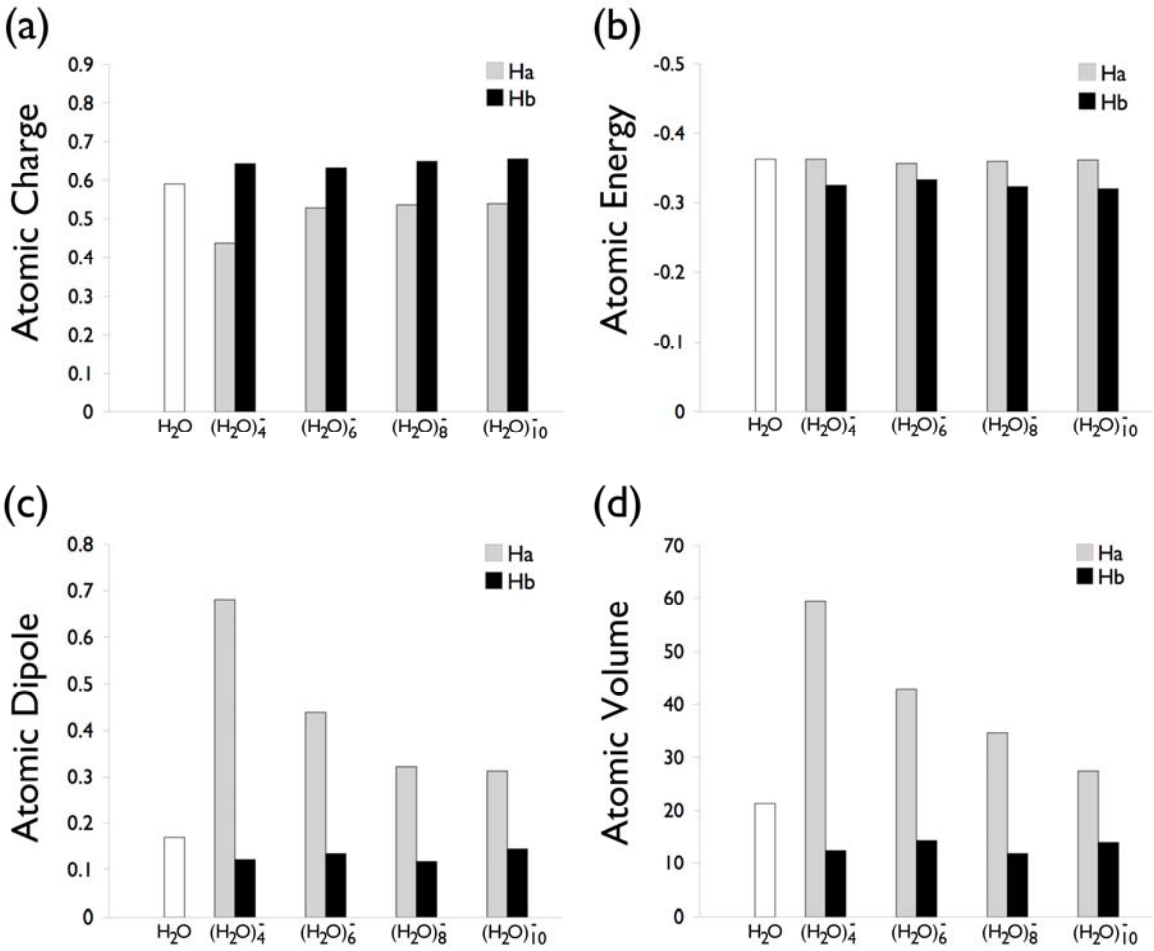


Figure 4.3: Comparison of various atomic properties for the hydrogen in a neutral water molecule (white bar) to the non-hydrogen-bonded H atoms (H_a – gray bars) and the hydrogen-bonded H atoms (H_b – black bars) in the cavity-bound anionic water clusters. The properties investigated include: (a) atomic charge, (b) atomic energy, (c) atomic dipole and (d) atomic volume.

4.4 Conclusions

The interaction between the non-hydrogen-bonded hydrogen atoms and the non-nuclear attractor observed in cavity-bound anionic water clusters has been investigated. This bond was characterized using electronic structure calculations and the theory of

atoms in molecules. The results of the AIM analysis indicate that this novel $H_a \cdots NNA$ interaction is a very weak, closed-shell interaction. The molecular graphs illustrate that it is a noticeably curved bond, occurring over a long distance of 2.6 to 3.5 Å depending on the cluster. The calculated binding energies suggest that the tetramer experiences a considerably smaller amount of stabilization per water molecule compared to the larger clusters. Furthermore, the large ε values for the $H_a \cdots NNA$ interaction and the relatively small separation between the BCPs of the $H_a \cdots NNA$ interaction and the nearby RCPs all suggest that this cluster is quite unstable with respect to small geometric changes. On the other hand, the larger clusters, $(H_2O)_n^-$ for n = 6, 8 and 10, all exhibit relatively large binding energies, with roughly 20 kJ/mol of stabilization per water molecule. The calculated ε values for the hexameric cluster indicate that it has the most stable $H_a \cdots NNA$ interactions; however, its hydrogen bonds are less stable than the other clusters, most likely due to the cyclic strain in the stacked trimers. In addition, the $H_a \cdots NNA$ interaction was investigated using the criteria provided by Popelier, which define a hydrogen bond on the basis of the atomic properties of the hydrogen atom involved. However, the results indicate that this novel interaction cannot be defined as a hydrogen bond. Therefore, the results of this study suggest that the $H_a \cdots NNA$ interaction is probably best described as a weak, dipole-ion interaction between the NHB hydrogen atoms and the NNA. Ultimately, the theory of atoms in molecules has provided additional insight into the hydrated electron dilemma by allowing for the characterization of this novel interaction.

Chapter 5: Self-Assembling ADADA Helices

5.1 Introduction

The helix is a popular motif in nature; from molecular systems such as proteins and DNA to macroscopic particles like the nautilus shell, spiral structures abound. Ever since the structure of the α-helix in proteins was elucidated,[118] and the DNA double helix a few years hence,[119] there has been a large push to develop wholly synthetic, biomimetic materials of this sort. Given that Mother Nature already exploits these molecules for information storage, molecular recognition and catalysis, synthetic chemists see no reason why they shouldn't also. In fact, almost concurrently with Watson and Crick's observations, synthetic helical structures began appearing in the literature. As early as 1955, Natta determined that the isotactic polypropylene formed in his Ziegler-Natta reactions actually adopted a helical structure in its crystalline form.[120] However, upon dissolution the helix was lost. Thus, the next hurdle to overcome was to develop helices that could maintain their structure in solution.

With an eye towards potential applications in materials science, molecular sensing, separation of enantiomers and asymmetric catalysis, various helical molecules have been developed. Indeed, the literature is littered with examples of single-polymer helices.[121-126] However, examples of double-helical structures are much more scarce. As such, in the last decade these double-helical structures have received a great deal of attention. Early work by Lehn relied on metal-ligand interactions to form the helices, or helicates as he called them.[127] In these schemes, an organic scaffold is wound up with another molecule like itself and together they enclose several metal cations along the

central core of the helix. Since then there have been many examples of duplex structures that exploit metal-ligand interactions,[128] hydrogen bonds,[129-132] ion pairing,[133] aromatic-aromatic[134] and hydrophobic interactions[135] or some combination of these. However, in order to truly mimic what nature has already accomplished, the synthetic helices ideally would form using solely non-covalent interactions.

One obvious tool that comes to mind when trying to mimic biological structures is the use of hydrogen bonds. Indeed, there are many examples of duplex structures that successfully exploit hydrogen bonding.[131] However, it is difficult to predict the final three-dimensional structure solely based on the monomeric subunits used.[125] For instance, many of the reported structures are simply ladder- or zipper-like hydrogen-bonded complexes rather than intertwined helices.[130,132] Furthermore, a large portion of the intertwined structures that have been synthesized are based on DNA architecture and thus often use nucleobases or their analogues to achieve the helix motif. Examples of intertwined structures that have been formed in the absence of metal cations and using non-DNA-based abiotic backbones are particularly rare.

However, there are some examples in the literature and the focus of this study is on just such a structure. The current work is a theoretical investigation of an experimentally prepared self-assembling helix. Wisner and colleagues have designed, synthesized and characterized their so-called ADADA helix.[136] The monomer subunits consist of a pentameric ring structure with alternating hydrogen-bond acceptor (A) and donor (D) groups, hence the ADADA moniker. Specifically, the monomer employs a pyridine ring as the H-bond acceptor and 1,4-thiazine-1,1-dioxide as the donor. The chemical structure of the monomer is shown in **Figure 5.1**.

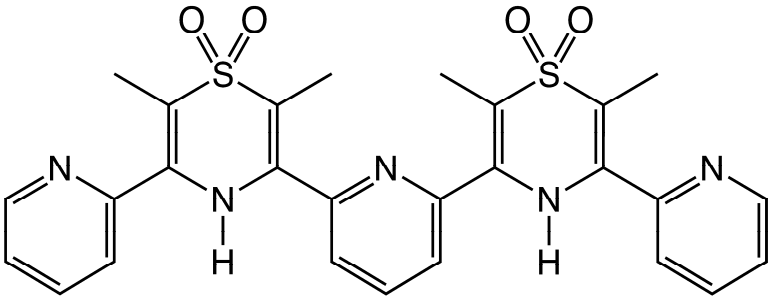


Figure 5.1: Chemical structure of the ADADA monomer.

The overall helix structure is achieved by intertwining two of these monomers. The resultant helix forms four hydrogen bonds along the central core of the helix, shown schematically in **Figure 5.2a**. In addition, various stacking interactions arise in the folded helix since each ring in the monomers stacks approximately face-on with at least one if not two rings in the opposite monomer (**Figure 5.2b**).

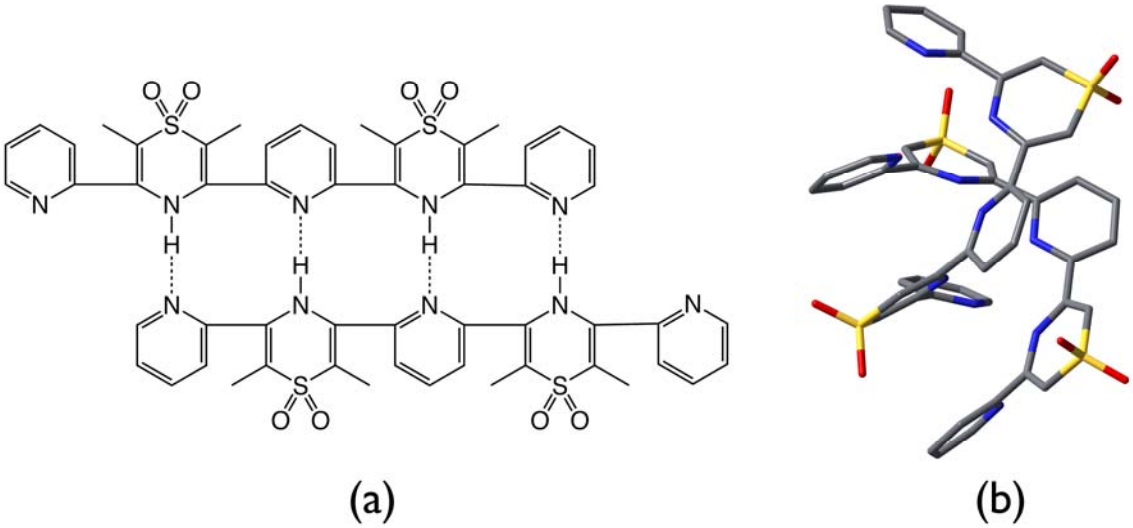


Figure 5.2: Schematic illustration of the ADADA helix illustrating (a) the hydrogen bonds formed and (b) the intertwining and ring stacking that occurs between the monomers (hydrogen atoms have been omitted for clarity).

The goal of the current study is to investigate the various interactions that are formed in the helix and how they contribute to the stability of the structure. The first task was to determine a suitable computational method to study these systems. Following that, the helix structure was analyzed using various electronic structure methods as well as the theory of atoms in molecules (AIM). Specifically, AIM was used to investigate the types and strengths of the interactions present in the helix. Finally, analogous to a previous study on triply hydrogen-bonded complexes,[137] the effect of substituents on the acceptor and donor rings was investigated. Ultimately this information could be used to tune the strength of the hydrogen-bonded helix.

5.2 Computational Details

Geometry optimizations were performed using density functional methods in the Gaussian 03 suite of programs.[47] In order to determine which DFT functional was best able to predict the geometries of the ADADA helices, three functionals were used on a test case in conjunction with the 6-31G(d,p) basis set: B3LYP,[39,40] B3PW91[40,138] and B971.[139] On the basis of this preliminary test, the B3LYP functional was selected for all subsequent optimizations. Frequency calculations were performed at this level to ensure all structures were minima on the potential energy surface and to obtain thermochemical data. Attempts were made to increase the size of the basis set by including diffuse functions on the heavy atoms. However, due to the large size of the system under investigation, these calculations proved too demanding for the available resources.

Accurate energies were obtained using the RI-MP2 method and the cc-pVDZ basis set in Q-Chem.[140] An attempt was made to use this method for geometry optimizations; however, once again this proved to be too computationally demanding.

All energies were corrected for zero point vibrational energies (ZPVE), which were obtained from the frequency calculations. In addition, all binding energies were corrected for basis set superposition error (BSSE) using the counterpoise method of Boys and Bernardi.[17] Unfortunately, calculation of the BSSE in Q-Chem is not automated in the version that was used and thus several additional calculations were required. To calculate the error that arises due to the use of an incomplete basis set, the energy for each monomer (in the complex geometry) must be computed with the monomer- and dimer-centered basis set, MCBS and DCBS, respectively. For instance, for a hypothetical fragment A, the energy would be calculated for the fragment alone as well as with ghost atoms at the nuclear coordinates of fragment B. The ghost atoms have the effect of placing the correct basis functions in the correct locations without including any nuclei or electrons. The first scenario represents the calculation with the monomer-centered basis set, while the latter is the dimer-centered basis set. The BSSE correction, ΔE_{BSSE}, is defined as,

$$\Delta E_{BSSE} = \left(\underset{MCBS}{E(A)} - \underset{DCBS}{E(A)} \right) + \left(\underset{MCBS}{E(B)} - \underset{DCBS}{E(B)} \right) \quad \textbf{(5.1)}$$

where all monomers are in the geometry adopted in the complex. Thus, the overall binding energy (ΔBE), for complex AB, is defined as,

$$\Delta BE = E(AB) - [E(A) + E(B)] + \Delta E_{BSSE} \quad \textbf{(5.2)}$$

where $E(X)$ is the energy of X in its optimized geometry.

Finally, the electron density obtained from the DFT optimization was analyzed topologically using AIM2000.[113]

5.3 Results and Discussion

The first component of this investigation was to determine which method was best suited for predicting the geometries of the helices. Once a functional was selected, a small model system was used to probe the effect of substituents on the individual ring systems. In the end, the results from the small model study were used to guide the substitution on the full, intact helix.

5.3.1 Comparison of DFT Predicted Geometries

For this portion of the study, three DFT functionals were selected (B3LYP, B3PW91 and B971) to investigate a model helix and compare the predicted geometries. The test system used was the full ADADA helix without any substituents on the pyridine or thiazine rings, as shown in **Figure 5.3**.

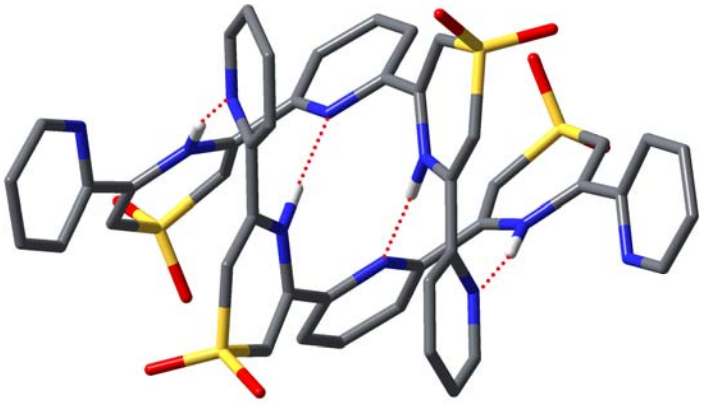


Figure 5.3: Structure of the ADADA helix illustrating the hydrogen bonds between the monomers (red dashed line). All C-H hydrogen atoms have been omitted for clarity.

Various geometric parameters were compared amongst the DFT optimized structures, as well as with the available experimental data. The bond lengths within the

individual monomers and the pyridine and thiazine rings were relatively invariable to the choice of functional used; these bonds typically varied from each other by only thousandths of an angstrom. On the other hand, the inter-monomer hydrogen bonds appeared much more susceptible to the choice of functional and thus were used as an indicator for the quality of the optimization. The parameters used were the inter-monomer N–N distances, the hydrogen bond length (r(N···H)), the hydrogen bond angles ($\angle$(N – H···N)) and the dihedral angles between adjacent rings within the monomers ($\angle$(N–C–C–N)) (**Table 5.1**). The data illustrate that the B3LYP and the B971 functionals produce quite similar results while the B3PW91 functional predicts noticeably shorter hydrogen bonds and N–N distances. Compared to the experimental data, all three methods tend to over- and under-estimate the inner and outer N–N distance, respectively. However, it should be pointed out that the experimental data represents the geometric parameters in the solid state (i.e., the crystal structure). That being said, the calculated gas-phase results are not that different from the crystal structure data considering the large differences in the two environments. On the basis of this comparison, it was decided that DFT methods could be used to model the ADADA helix with reasonable accuracy and computational cost.

As the results in **Table 5.1** show, the B3LYP and B971 functionals performed better for the inner hydrogen bond than the outer, while the opposite was true for B3PW91. Examining the predicted hydrogen bond angles reveals that B3PW91 has a fairly large deviation for the outer hydrogen bond and was thus eliminated as one of the candidates for subsequent calculations. The remaining two functionals appear to perform equally well and ultimately the B3LYP functional was selected largely based on its

notoriety and prevalence throughout the literature. All subsequent geometry optimizations and frequency calculations were performed at the B3LYP/6-31G(d,p) level of theory.

Table 5.1: Various geometric parameters in the ADADA helix as predicted from DFT optimizations and experiment. Bond lengths (*r*) are in angstroms (Å) and angles (∠) are in degrees.

		B3LYP	B3PW91	B971	Experiment[a]
***r*(N–N)**					
	Outer	3.0837	3.0296	3.0828	2.98
	Inner	3.0478	3.0088	3.0513	3.11
***r*(N···H)**					
	Outer	2.0739	2.0103	2.0736	–
	Inner	2.0216	1.9811	2.0239	–
∠(N – H···N)					
	Outer	167.2	169.5	166.7	166
	Inner	176.8	175.0	177.3	175
∠(N–C–C–N)					
		42.5	38.3	42.1	36–53
		45.7	45.8	45.0	
		44.4	47.1	44.4	
		44.6	40.8	44.0	

[a]Ref. 136.

5.3.2 Structure and Energetics of the ADADA Helix

Once a suitable method had been selected (B3LYP/6-31G(d,p)), the structural features of the ADADA helix and its constituent monomers were scrutinized.

5.3.2.1 ADADA Monomer Structure

As mentioned in the introduction, the monomers in the helix consist of a pentameric ring system with alternating hydrogen-bond acceptor and donor groups

(**Figure 5.1**). The monomer was optimized with the method selected in section **5.3.1** and two very different minimum energy structures were obtained, as shown in **Figure 5.4**. The first structure maintains an approximately linear arrangement of the rings, with dihedral angles between adjacent rings of roughly 20-30°. The final structure has the hydrogen-bonding groups pointing in opposite directions (**Figure 5.4a**). On the other hand, the second structure is folded in such a way that all the H-bonding groups are pointed towards one central location. This is achieved by the monomer forming a helix of its own. In this case, several of the rings are coplanar and the two ends twist towards each other. In the final structure the terminal rings are almost stacked upon one another with roughly a 3.65 Å separation (**Figure 5.4b-c**). The arrangement of the rings in the second structure closely resembles one half of the intact helix, just slightly more compressed and with smaller dihedral angles between adjacent rings.

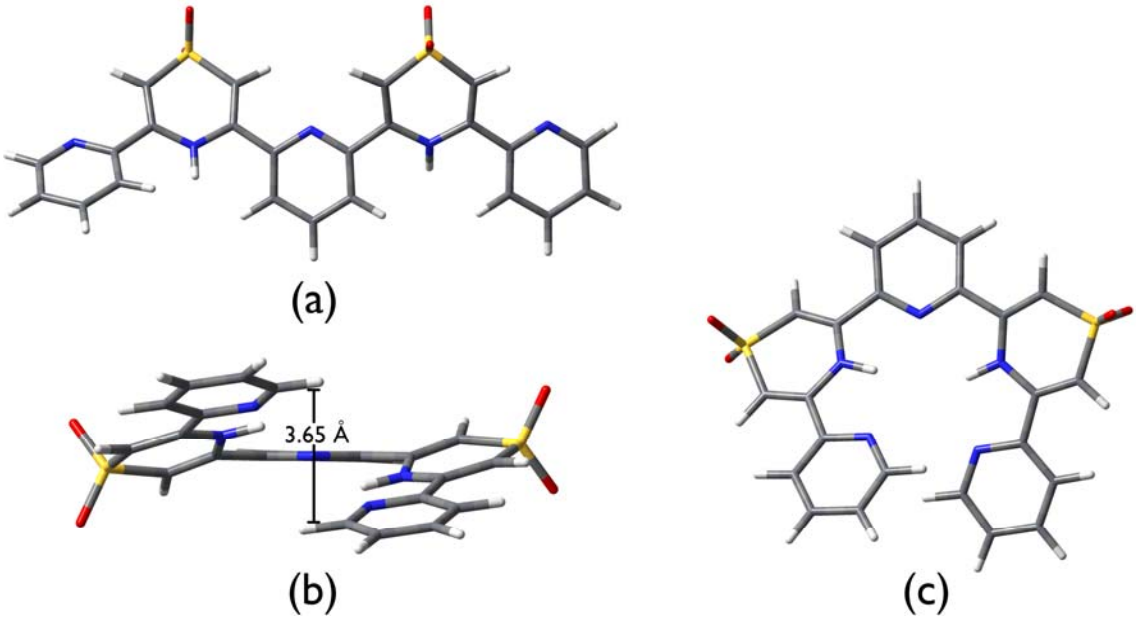


Figure 5.4: Structures of the optimized ADADA monomers: (a) the linear structure and the lower-energy coiled structure from the (b) side and (c) top view.

Comparison of the calculated electronic energies reveals that the coiled structure is more stable than the linear structure by 83.5 kJ/mol. Thus, for the remainder of this chapter, the coiled structure is used as the reference monomer when needed.

5.3.2.2 ADADA Helix Structure and Binding Energy

The major structural features of the ADADA helix were discussed in section **5.3.1**. Recall that the helix forms four hydrogen bonds between the monomers, leaving one external pyridine ring on each monomer without an H-bond. As such, the hydrogen bonds can be divided into two types: inner and outer. In the model helix, the inner H-bonds are slightly shorter, but only by 0.05 Å (**Table 5.1**). In terms of the N – H⋯N bond angles, these exhibit a more substantial difference between the two types. While the inner hydrogen bond forms an angle of 177°, the outer bond is less linear at 167°. Due to the helix structure that is adopted, each ring in the monomers stacks approximately face-on with at least one other ring, while two of the thiazine-type rings are actually sandwiched between two pyridine rings. As described in section **5.2**, accurate binding energies were calculated using the RI-MP2 method. The ZPVE and BSSE corrected binding energy for the model helix is –85.8 kJ/mol. The fact that the binding energy is negative indicates that the helix is more stable than the separated monomers.

5.3.2.3 Methylated ADADA Helix Structure and Binding Energy

Due to the large size of these systems and the computational demand involved in studying them, a simpler model helix was investigated initially. All of the results reported thus far have been for the smaller model. Compared to the experimentally prepared structure, the model system lacks methyl substituents in the *ortho*-positions

relative to the sulfone functional groups. The experimental motivation for these methyl groups was to bias adjacent rings out of coplanarity. However, the initial calculations presented here suggest that these methyl groups are not absolutely necessary since the desired helix structure was obtained even when these groups were omitted. Nonetheless, the full methylated ADADA helix was also investigated.

5.3.2.3.1 Monomer Structure

Although the addition of the four methyl groups to the pentameric monomer results in a qualitatively similar structure to that discussed in section **5.3.2.1**, some noticeable structural changes were observed. For instance, compared to the smaller model system, there is a slight expansion of the monomer helix, with the two ends now separated by 4.02 Å (versus 3.65 Å). Furthermore, while the exterior pyridine rings were coplanar with their neighbour thiazine rings in the smaller model, none of the adjacent rings are coplanar in the methylated system. However, the most significant difference is the distortion of the thiazine ring away from planarity. As shown in **Figure 5.5**, while there is some minor distortion observed in the first monomer (sulfur group is approximately 7° below the plane), the effect is much more pronounced in the monomer with the methyl groups. The sulfur atom is now 20° below the plane while the two methyl groups are 17° above the plane and there is even a small distortion in the nitrogen atom (5° below the plane).

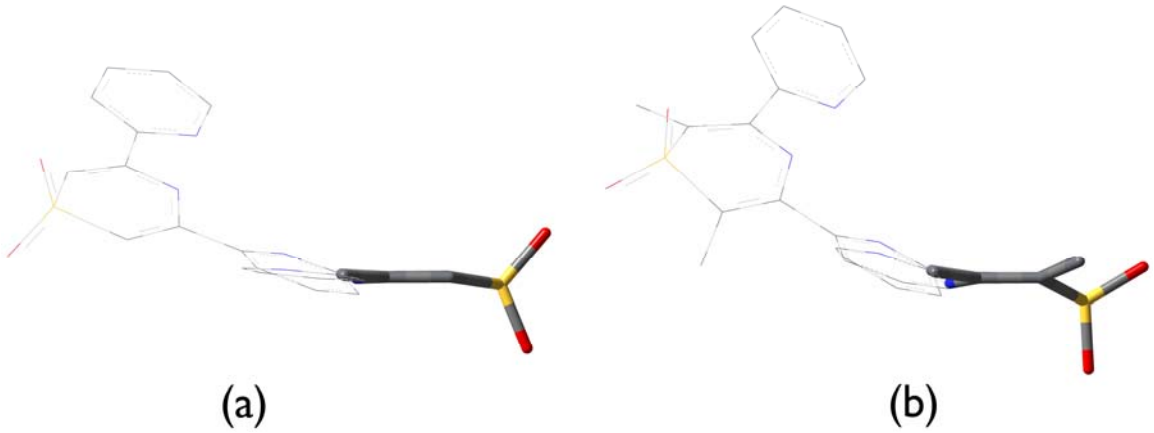


Figure 5.5: Comparison of the thiazine rings in (a) the small model monomer and (b) the methylated monomer. Hydrogen atoms have been omitted and the rest of the monomer has been faded out for clarity.

5.3.2.3.2 Helix Structure

The addition of the methyl groups to the intact helix has a much smaller effect than observed for the isolated monomer. Most notably, the thiazine rings are not greatly distorted from planarity, a feature that is apparently only present in the monomer. This discrepancy between the monomer and the helix structures may be due to the fact that the thiazine rings in the helix experience face-on stacking interactions with adjacent pyridine rings which is not possible in the monomer. Thus, the presence of the nearby pyridine rings may force the thiazine rings to adopt a planar structure.

Another interesting observation is that the N–N distances and the hydrogen bond lengths are the same throughout the methylated helix rather than observing a difference between the inner and outer hydrogen bonds. The methylated helix has a slightly smaller separation between the monomers as evidenced by the decreased hydrogen bond lengths (1.95 Å) and N–N distances (2.98 Å). In addition, the hydrogen bond angles are slightly more linear than in the unmethylated case (inner, 167°; outer, 177°). In terms of the

calculated binding energies, helix formation from the methylated monomers is more favourable with an electronic binding energy of −108.1 kJ/mol.

5.3.2.4 Thermodynamics of ADADA Helix Formation

In the original report of the ADADA helix from Wisner *et al*.,[136] various thermodynamic parameters were determined experimentally. From the experimental data, an enthalpy change (ΔH) of −16 kJ/mol and an entropy change (ΔS) of −42 J/mol·K were predicted. Extrapolating to 298 K and using the above parameters, a Gibbs energy (ΔG^{298}) of −3.5 kJ/mol was obtained. These results suggest that helix formation is an energetically favourable process (as evidenced by the negative ΔG value) and they conclude that helix formation is the result of molecular recognition rather than the result of solvophobic forces (i.e., the enthalpy change dominates the free energy expression).

In an effort to compare the theoretically predicted ADADA helices discussed in this thesis to the experimentally prepared helix, the same thermodynamic parameters were determined. The results for the methylated and unmethylated helices appear in **Table 5.2** along with the experimental predictions.

Table 5.2: Relative energy and thermodynamic parameters associated with ADADA helix formation.

	Relative Energies (kJ/mol)	**ΔH** (kJ/mol)	**ΔS** (J/mol·K)	**ΔG^{298}** (kJ/mol)
Unmethylated Helix *(from section **5.3.2.2**)*	-85.8	-85.0	-244	-12.3
Methylated Helix *(from section **5.3.2.3**)*	-108.1	-103.7	-224	-36.9
Experimental Helix[a] *(methylated)*	–	-16	-42	-3.5

[a]Ref. 136.

Comparing the two theoretically predicted helices, helix formation for the methylated case is much more favourable, with a calculated ΔG^{298} approximately three-fold larger than the unmethylated case. However, comparing the theoretical results to the experimental predictions reveals that the thermodynamic parameters are quite different. The same general trend is observed, predicting negative ΔG^{298} values in all cases, which are dominated by the enthalpy change. However, that is where the similarities end. The calculated thermodynamic parameters for the theoretical helices are five- to ten-fold larger than the experimental values. However, these discrepancies are not necessarily a serious cause for alarm considering that the experimental data was obtained in $CDCl_3$ solution while the calculations were performed in the gas phase. The difference in the calculated values highlights the potentially large effect of the solvent. Ideally, some type of solvent model would have been included in the calculations; however, it is beyond the scope of this work for the time being at least.

5.3.2.5 Molecular Graphs of the ADADA Monomers and Helices

The theory of atoms in molecules (AIM) has been used to characterize the bonding interactions in the individual monomers as well as the intertwined helices. Molecular graphs of the pentameric monomers, both unsubstituted and methylated, reveal that there are several intramolecular hydrogen bonds as well as some unusual interactions between the ends of the monomers (**Figure 5.6**). In all, four intramolecular H-bonds form between adjacent pyridine and thiazine rings. Further analysis of the electron density at the bond critical points (BCPs) suggests that there are two types of H-bonds: outer H-bonds that form between one of the terminal pyridine rings and an adjacent thiazine ring, and inner H-bonds that form between the central pyridine ring and one of

the thiazine rings. The BCP data (**Table 5.3**) indicate that the outer H-bond is stronger since it exhibits a larger value of the electron density at the BCP (ρ_{BCP}). In addition, the relative proximity of the BCP of the inner H-bond to a nearby ring critical point (RCP) as well as the larger ellipticity (ε) value of this bond all suggest it is less stable than the outer hydrogen bonds.

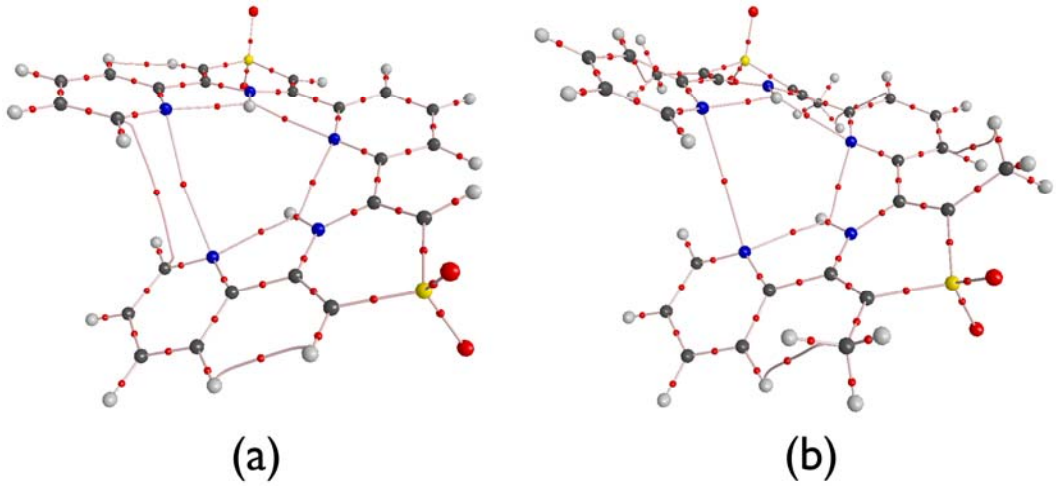


Figure 5.6: Molecular graphs of the (a) unsubstituted and (b) methylated monomer.

In addition, there are also some unusual weak interactions evident in the molecular graphs. In both cases, there is a N–N interaction between the terminal pyridine rings as well as a C–C interaction between the same rings in the unsubstituted monomer. The methylated monomer also exhibits some weak interactions between the methyl substituents and their neighbour rings. Analysis of these critical points reveals that they are very weak, as evidenced by the small ρ_{BCP} values, which are generally a full order of magnitude smaller than observed for the hydrogen bonds (**Table 5.3**). These same bonds

also exhibit very large ellipticity values, once again indicating that they are not particularly stable.

The molecular graphs of the unsubstituted and methylated helices are more complex than the isolated monomers (**Figure 5.7**). The four intramolecular hydrogen bonds in the monomers are replaced by four intermolecular hydrogen bonds in the intertwined helix, leaving one pyridine ring at each end of the helix without a hydrogen-bonding partner. In addition to these strong H-bonds, there is a slew of additional weak interactions such as N–N interactions, the sulfone oxygen interacting with a carbon atom in the nearby pyridine ring as well as C–C interactions between stacked rings. Just like the methylated monomer, the methylated helix also exhibits some weak interactions between the methyl substituents and neighbour atoms.

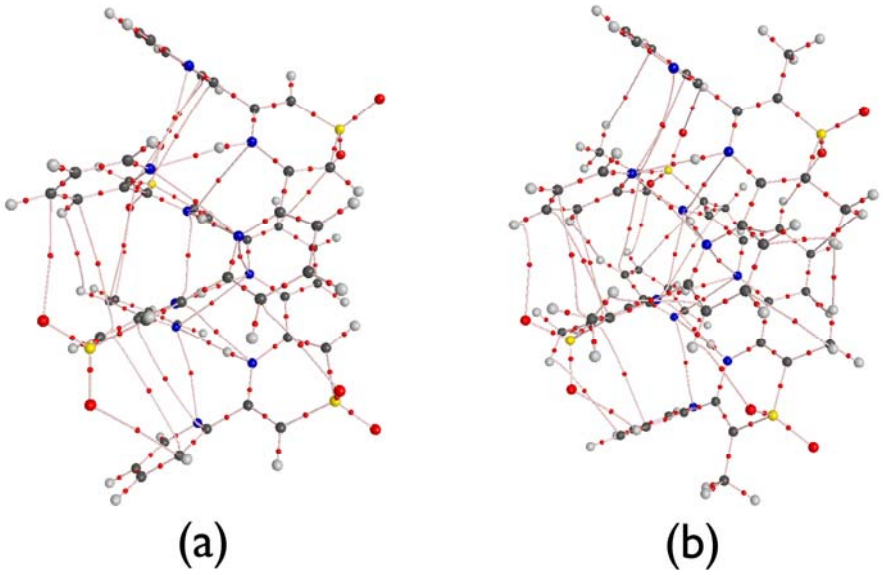


Figure 5.7: Molecular graphs of the (a) unsubstituted and (b) methylated ADADA helix.

Again, the bond critical points were analyzed further by measuring various properties at the BCP, which are reported in **Table 5.3**. As can be seen from the data, all the hydrogen bonds have approximately the same value for ρ_{BCP}, implying they are all the same strength (the inner H-bonds may be marginally stronger). Comparison of the calculated ellipticity values suggest that the hydrogen bonds in the helix are more stable than the intramolecular H-bonds in the monomers. This is perhaps not surprising since the H-bonds in the helix can adopt the preferred linear arrangement, which is not possible in the individual monomers.

All of the other interactions present in the helix are much weaker, with ρ_{BCP} values roughly an order of magnitude smaller. Furthermore, the weaker interactions all tend to have larger ε values, indicating they are not particularly stable relative to small geometric changes. Thus, although these bond critical points for the weak interactions are detected in the electron density, they do not necessarily indicate the presence of a significant interaction but rather suggest that various weak interactions are present in the structure. For the remainder of this investigation, the stronger hydrogen bonds will be the focus.

Table 5.3: Selected properties of the bond critical points (BCPs) in the ADADA monomers and helices. Note that only one entry appears as an example for the weak interactions while the data for all of the hydrogen bonds is provided. All entries are in atomic units except for ellipticity (ε), which is dimensionless.

Bond	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ε
Unsubstituted Monomer			
outer H-bond	0.0238	0.0864	0.2219
	0.0174	0.0740	0.9317
inner H-bond	0.0174	0.0741	0.9307
	0.0238	0.0864	0.2220
N –N	0.0024	0.0095	2.0984
C–C	0.0024	0.0086	6.5395
Methylated Monomer			
outer H-bond	0.0244	0.0877	0.2074
	0.0174	0.0753	1.2345
inner H-bond	0.0174	0.0753	1.2316
	0.0244	0.0877	0.2074
N –N	0.0031	0.0112	0.3968
methyl H–pyridine C	0.0093	0.0383	1.6658
Unsubstituted ADADA Helix			
outer H-bond	0.0239	0.0559	0.0385
	0.0269	0.0619	0.0369
inner H-bond	0.0269	0.0620	0.0369
	0.0239	0.0559	0.0386
N –N between thiazine rings	0.0083	0.0332	0.0769
N –N between pyridine rings	0.0095	0.0261	0.0890
sulfone O–pyridine C	0.0068	0.0244	1.2747
pyridine C–thiazine C (*stacking interaction*)	0.0019	0.0059	1.3626
Methylated ADADA Helix			
outer H-bond	0.0306	0.0726	0.0243
	0.0314	0.0728	0.0312
inner H-bond	0.0314	0.0729	0.0312
	0.0306	0.0726	0.0243
N –N between thiazine rings	0.0071	0.0245	0.5731
N –N between pyridine rings	0.0089	0.0254	0.0615
sulfone O–pyridine C	0.0093	0.0357	0.5340
pyridine C–thiazine C (*stacking interaction*)	0.0022	0.0067	0.9209
methyl H–pyridine C	0.0015	0.0050	0.9045

5.3.3 Substituent Effects on the Pyridine-Thiazine Model System

Akin to a previous study of AAA-DDD triply hydrogen-bonded complexes,[137] the effect of substituents on the hydrogen-bonded network was investigated. Once again, due to the large size of the system, a small test case was used to guide which substitutions would be studied on the full ADADA helix. The model system for this portion of the study was a single pyridine ring hydrogen bonded to 1,4-thiazine-1,1-dioxide, as shown in **Figure 5.8**.

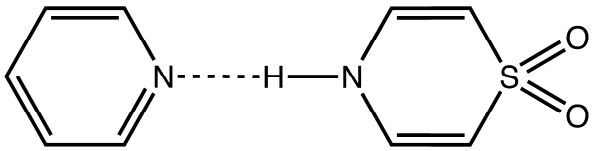


Figure 5.8: Chemical structure of the pyridine – 1,4-thiazine-1,1-dioxide test system.

Due to the structure of the pentameric monomers, not all positions on the rings are open for substitution. For the pyridine ring, the *meta*- and *para*-positions, relative to the nitrogen atom, are available and both were investigated. On the other hand, the only positions available on the thiazine-type rings are the *ortho*-positions next to the sulfone group. In this case, substituents were added to one of the *ortho*-positions first and then to both.

In terms of the substituents themselves, a range of electron withdrawing (EWG) and donating groups (EDG) were used. Since the nitrogen atom in the pyridine ring serves as the hydrogen bond acceptor, increased electron density at this atom should strengthen the hydrogen bond. Conversely, decreased electron density on the nitrogen in the thiazine-type ring should have the same effect since its hydrogen atom is the donor.

Thus, chemical intuition suggests that the combination of EWGs on the thiazine-type ring and EDGs on the pyridine ring should result in the strongest hydrogen bond. Knowing also, as a general rule of thumb, that EDGs tend to be *ortho*- and *para*-directors, it is expected that *para*-substitutions on the pyridine ring should perform better than *meta*-substitutions in terms of strengthening the hydrogen bond. However, for the sake of completeness, both positions were investigated. The substituents used are shown in **Figure 5.9**.

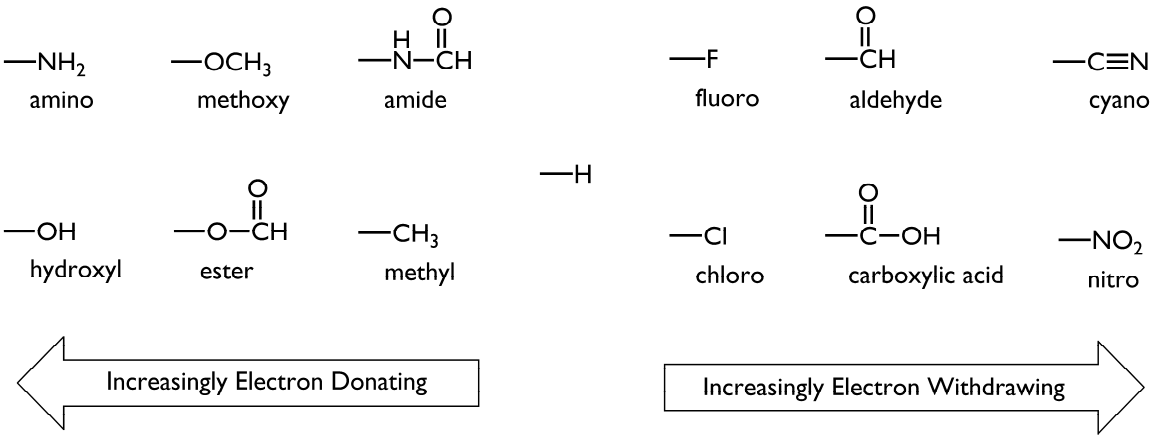


Figure 5.9: Chemical structure of the substituents investigated.

In order to determine the effect of the substituents on the hydrogen bond, several parameters were investigated. First of all, the binding energy for the pyridine-thiazine system was calculated. Since this portion of the study was performed to quickly determine which substituents should be studied on the full helix, the binding energies were calculated at the B3LYP/6-31G(d,p) level of theory and corrected for both ZPVE and BSSE. Although ideally these energies would have been calculated using an MP2 method as done for the full helix, the goal here was simply to elucidate any trends in the

binding energies (i.e., determine which substituents improve binding relative to the unsubstituted case) rather than to obtain an extremely accurate estimate of the binding energy itself.

In addition to the calculated binding energies, the theory of atoms in molecules was used to probe the nature of the hydrogen bonds. Specifically, the electron density at the hydrogen bond critical point (ρ_{BCP}) and the bond length were investigated. These two properties were selected because they have been shown to correlate well with strength of hydrogen bonds.[141-147] Generally, stronger hydrogen bonds are characterized by an increase in the electron density at the bond critical point, as well as a decrease in hydrogen bond length. The binding energies and hydrogen bond properties for the various substituents can be found in **Table 5.4**.

Examining the pyridine data first, compared to the effect of substitution at the *meta*-position, substitution at the *para*-position always results in larger binding energies. In addition, the ρ_{BCP} and hydrogen bond lengths agree with this finding by exhibiting increased electron density and shorter H-bonds in the *para*-substituted cases. In terms of which substituents strengthen the hydrogen bond, the initial suspicion was confirmed. Enhanced binding is observed when the pyridine ring is substituted with electron donating groups, such as: $-NH_2$, $-OH$, $-OCH_3$ and $-CH_3$. Thus, on the basis of these results, *para*-substitution on the pyridine ring with the above four substituents will be investigated on the full helix.

Table 5.4: Calculated binding energy (BE), electron density at the hydrogen bond critical point (ρ_{BCP}) and hydrogen bond length (r) for the pyridine-thiazine test system with various substituents. Energies are reported in kJ/mol, bond lengths in angstroms (Å) and electron density in atomic units.

Substituent	BE	ρ_{BCP}	r(N···H)	BE	ρ_{BCP}	r(N···H)
–H	-36.5	0.0343	1.909			
Substitution on the pyridine ring						
		***para*-position**			***meta*-position**	
$-NH_2$	-44.1	0.0377	1.865	-40.4	0.0360	1.888
–OH	-38.6	0.0354	1.893	-36.9	0.0342	1.910
$-OCH_3$	-40.1	0.0360	1.886	-38.2	0.0347	1.903
–NHC(O)H	-34.7	0.0340	1.911	-31.6	0.0327	1.930
–OC(O)H	-30.7	0.0324	1.933	-28.2	0.0314	1.947
$-CH_3$	-38.7	0.0352	1.897	-38.0	0.0350	1.900
–F	-33.8	0.0333	1.921	-32.4	0.0324	1.933
–Cl	-32.1	0.0326	1.930	-30.8	0.0319	1.939
–C(O)H	-30.4	0.0319	1.941	-29.7	0.0319	1.940
–C(O)OH	-32.1	0.0324	1.933	-32.0	0.0326	1.931
–CN	-26.8	0.0306	1.959	-25.8	0.0301	1.966
$-NO_2$	-25.6	0.0299	1.969	-25.1	0.0297	1.972
Substitution on the thiazine ring						
		***meta*-position**			**both *meta*-positions**	
$-NH_2$	-33.1	0.0331	1.923	-34.1	0.0339	1.914
–OH	-35.9	0.0345	1.905	-34.9	0.0345	1.904
$-OCH_3$	-33.7	0.0333	1.923	-31.5	0.0322	1.937
–NHC(O)H	-40.3	0.0364	1.882	-43.7	0.0383	1.859
–OC(O)H	-41.1	0.0370	1.874	-45.5	0.0397	1.843
$-CH_3$	-34.8	0.0336	1.918	-33.2	0.0329	1.928
–F	-38.6	0.0357	1.890	-40.6	0.0371	1.872
–Cl	-39.7	0.0363	1.882	-42.7	0.0384	1.858
–C(O)H	-42.8	0.0382	1.861	-48.4	0.0419	1.820
–C(O)OH	-41.1	0.0373	1.870	-44.8	0.0399	1.841
–CN	-45.2	0.0391	1.850	-53.7	0.0441	1.798
$-NO_2$	-46.8	0.0404	1.836	-56.3	0.0465	1.775

With respect to the substitutions on the thiazine ring, the data in **Table 5.4** once again indicate that the initial hypothesis is correct; binding is improved when the thiazine

ring is substituted with electron withdrawing groups. The electron density and H-bond lengths follow suit: increased ρ_{BCP} values and shorter hydrogen bonds accompany the larger binding energies. Furthermore, there is an additional enhancement, by as much as 10 kJ/mol, if the ring is substituted at both *meta*-positions rather than just one. In this case, the best results were obtained with the following functional groups: –F, –Cl, –NHC(O)H, –C(O)OH, –OC(O)H, –C(O)H, –CN and $–NO_2$, in order of increasing binding energies. As such, these substituents on the thiazine ring will also be investigated on the full helix.

5.3.4 Substituent Effects on the ADADA Helix

Using the results from the small-model study, various substitutions on the intact ADADA helix were investigated. Since there are four hydrogen bonds in the full helix, substituents may be attached to the monomers such that they are on a ring that is involved in one of two types of hydrogen bonds, either inner or outer. In order to determine if there was any advantage to substituting one ring over another, both cases were investigated. The results have been divided on the basis of whether the substitutions are on the hydrogen-bond acceptor (pyridine ring) or donor (thiazine-type ring). The different substituent locations are shown in **Figure 5.10**.

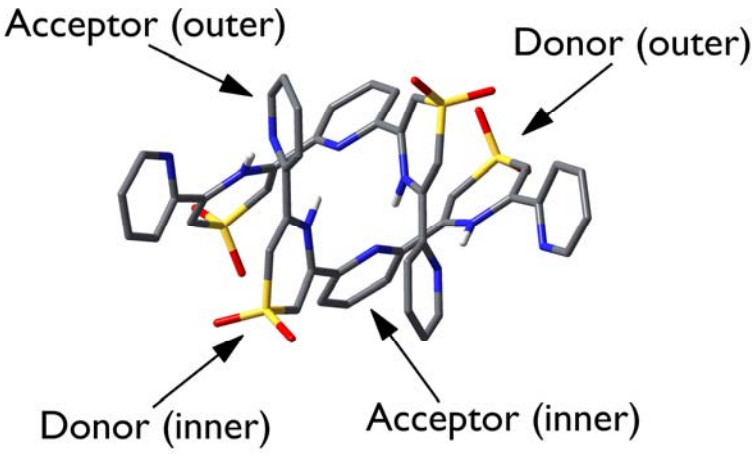


Figure 5.10: Structure of the ADADA helix illustrating the different locations for the substituents on the donor and acceptor rings. All C–H hydrogen atoms have been omitted for clarity.

5.3.4.1 Substituents on the Hydrogen-bond Acceptor Rings

The results from the small-model study suggest several electron donating substituents which may strengthen the hydrogen bonds when they are attached to the pyridine ring. These include an amino ($-NH_2$), an ether ($-OCH_3$), an alkyl ($-CH_3$) and a hydroxyl (–OH) group. These four substituents were investigated on the intact helix in two different positions, affecting either an inner or outer hydrogen bond as shown in **Figure 5.10**. All of the substituted helices were optimized and the binding energies calculated (**Table 5.5)**. Compared to the unsubstituted helix, which has an electronic binding energy of –85.8 kJ/mol, all acceptor-substituted helices exhibit larger binding energies. The best results are obtained with the amino and hydroxyl substituents. Except for the methoxy-substituted case, the helices with substituents on the inner hydrogen-bond acceptor are more stable than the outer.

Table 5.5: Relative energy and thermodynamic parameters associated with ADADA helix formation for helices with substituents on the hydrogen-bond acceptor ring.

Substituent	Position	Relative Energies (kJ/mol)	ΔH (kJ/mol)	ΔS (J/mol·K)	ΔG^{298} (kJ/mol)
$-NH_2$	*Inner*	-99.8	-100.4	-260	-22.8
	Outer	-92.0	-92.3	-264	-13.5
$-OCH_3$	*Inner*	-90.6	-90.1	-247	-16.5
	Outer	-93.2	-92.7	-254	-17.2
$-CH_3$	*Inner*	-89.3	-89.5	-267	-9.9
	Outer	-88.4	-88.6	-266	-9.2
–OH	*Inner*	-110.1	-111.0	-273	-29.5
	Outer	-99.2	-99.2	-253	-23.9

However, the performance of the amino and hydroxyl substituents may be slightly misleading since both of these functional groups form an additional hydrogen bond with one of the nearby sulfone oxygens. In order to determine the effect of the substituents on the hydrogen-bonding network, AIM was used to analyze the electron density at the bond critical points. The data for the hydrogen-bond BCPs is shown in **Table 5.6**. The results indicate that the amino substituent strengthens the central four hydrogen bonds to the greatest extent, as would be expected since it is the best electron donating group among the substituents investigated. Furthermore, examining the hydrogen bonds formed with the sulfone oxygen reveals that the bond formed by the hydroxyl group is quite a bit shorter than the one formed with the amino group (~1.8 Å versus ~2.1 Å). The same trend is seen in the ρ_{BCP} values, where the hydroxyl-sulfone oxygen hydrogen bond is significantly larger than the amino-sulfone oxygen value (0.031 a.u. versus 0.015 a.u.). Thus, it is likely this additional strong hydrogen bond in the hydroxyl case that results in

a larger binding energy than what might be expected based on the electron donating ability of the substituents.

Table 5.6: Selected properties of the hydrogen bond critical points in the acceptor-substituted ADADA helices. Bond lengths (r) are in angstroms (Å) and ρ_{BCP} is in atomic units.

Substituent	H-bond	Inner Substituent		Outer Substituent	
		r(N···H)	ρ_{BCP}	r(N···H)	ρ_{BCP}
$-NH_2$	*Outer*	2.032	0.0260	2.041	0.0256
	Inner	1.927*	0.0329	1.973	0.0298
	Inner	1.983	0.0291	1.931	0.0326
	Outer	2.039	0.0257	1.910*	0.0338
$-OCH_3$	*Outer*	2.099	0.0226	2.079	0.0236
	Inner	2.029*	0.0265	2.048	0.0254
	Inner	2.033	0.0261	2.072	0.0241
	Outer	2.073	0.0239	2.105*	0.0224
$-CH_3$	*Outer*	2.060	0.0246	2.065	0.0243
	Inner	1.989*	0.0288	2.004	0.0279
	Inner	2.015	0.0272	2.011	0.0275
	Outer	2.022	0.0266	2.018*	0.0269
–OH	*Outer*	2.033	0.0259	2.235	0.0170
	Inner	1.932*	0.0325	2.214	0.0180
	Inner	1.963	0.0304	2.128	0.0214
	Outer	2.037	0.0259	2.117*	0.0219

*Indicates that the substituent is attached to the ring involved in the hydrogen bond.

5.3.4.2 Substituents on the Hydrogen-bond Donor Rings

As with the acceptor-substituted helices, the results from the small model study were used to guide the investigation on the full helix. The donor substituents that strengthen the hydrogen bond in the pyridine-thiazine system included: nitro ($-NO_2$), cyano (–CN), aldehyde (–C(O)H), ester (–OC(O)H), carboxylic acid (–C(O)OH), amide (–NHC(O)H) and halogen (–Cl, –F) groups. All but the largest substituents (carboxylic

acid and amide) have been investigated on the full helix. The calculated binding energies for the donor-substituted ADADA helix are shown in **Table 5.7**. The data indicate that only some of the substituents result in improved binding compared to the unsubstituted helix which has a binding energy of –85.8 kJ/mol. The nitro-, cyano- and chloro-substituted helices all exhibit approximately the same binding energy, roughly 5 kJ/mol better than the unsubstituted helix, when the substituent affects one of the inner hydrogen bonds. On the other hand, the improvement is essentially negligible when the substituent affects the outer H-bond. Furthermore, it is interesting to note that some of the functional groups that enhanced binding in the small model system actually have a destabilizing effect in the full helix (e.g. aldehyde, ester and fluoro).

Table 5.7: Relative energy and thermodynamic parameters associated with ADADA helix formation for helices with substituents on the hydrogen-bond donor ring.

Substituent	**Position**	**Relative Energies** (kJ/mol)	**ΔH** (kJ/mol)	**ΔS** (J/mol·K)	**ΔG^{298}** (kJ/mol)
$-NO_2$	*Inner*	-91.1	-90.9	-255	-14.7
	Outer	-87.8	-87.7	-262	-9.5
–CN	*Inner*	-91.3	-90.0	-233	-20.4
	Outer	-86.3	-85.0	-233	-15.4
–C(O)H	*Inner*	-83.0	-82.0	-244	-9.3
	Outer	-85.6	-84.2	-240	-12.6
–OC(O)H	*Inner*	-77.2	-76.3	-238	-5.4
	Outer	-77.0	-75.7	-234	-6.0
–Cl	*Inner*	-90.5	-89.4	-241	-17.7
	Outer	-88.1	-87.2	-242	-15.1
–F	*Inner*	-83.5	-82.1	-233	-12.5
	Outer	-82.8	-81.6	-234	-11.6

Analyzing the hydrogen bonds further (**Table 5.8**), the results indicate that there is a general strengthening of the hydrogen bonds closest to the added substituents, but no major improvements are seen. The relatively minor changes in ρ_{BCP} and hydrogen bond lengths are reflected in the small enhancement seen for the binding energies.

Table 5.8: Selected properties of the hydrogen bond critical points in the donor-substituted ADADA helices. Bond lengths (r) are in angstroms (Å) and ρ_{BCP} is in atomic units.

Substituent	H-bond	Inner Substituents		Outer Substituents	
		r(N···H)	ρ_{BCP}	r(N···H)	ρ_{BCP}
–NO_2	*Outer*	2.118	0.0217	2.080	0.0236
	Inner	2.004	0.0278	1.994	0.0285
	Inner	1.923*	0.0333	1.962	0.0305
	Outer	2.005	0.0273	1.945*	0.0316
–CN	*Outer*	2.133	0.0211	2.106	0.0223
	Inner	1.999	0.0280	2.021	0.0270
	Inner	1.965*	0.0304	1.980	0.0294
	Outer	2.074	0.0238	1.962*	0.0305
–C(O)H	*Outer*	2.127	0.0213	2.093	0.0230
	Inner	1.983	0.0291	1.981	0.0293
	Inner	1.967*	0.0303	1.967	0.0303
	Outer	1.996	0.0280	1.940*	0.0319
–OC(O)H	*Outer*	2.059	0.0247	2.110	0.0221
	Inner	2.034	0.0261	2.022	0.0269
	Inner	1.974*	0.0297	1.959	0.0306
	Outer	2.004	0.0275	1.976*	0.0295
–Cl	*Outer*	2.083	0.0234	2.066	0.0243
	Inner	1.997	0.0282	1.998	0.0283
	Inner	1.984*	0.0292	2.003	0.0280
	Outer	1.999	0.0278	1.969*	0.0299
–F	*Outer*	2.081	0.0235	2.086	0.0233
	Inner	2.014	0.0273	2.023	0.0268
	Inner	2.000*	0.0281	1.996	0.0284
	Outer	2.089	0.0231	2.023*	0.0265

*Indicates that the substituent is attached to the ring involved in the hydrogen bond.

5.3.5 Structure and Energetics of the AAAAA-DDDDD Helix

The original report from Wisner *et al*. investigated helices that formed from monomers with alternating hydrogen-bond acceptor and donor groups. However, they have also developed helices that are constructed from contiguous donors and acceptors rather than the previously discussed ADADA motif. Their results indicate that the helices formed from the contiguous monomers exhibit much greater binding strengths.[148] This type of helix has also been investigated in the current study. In this case, the monomers are still pentameric structures; however, each monomer is constructed from either all pyridine rings (the acceptor, AAAAA) or all thiazine-type rings (the donor, DDDDD). Using the results from the ADADA helices investigated so far, an AAAAA-DDDDD helix was constructed which forms four hydrogen bonds, leaving one ring at each end without a hydrogen-bonding partner. The helix constructed from the contiguous monomers is shown in Figure **5.11** along with the ADADA helix for comparison.

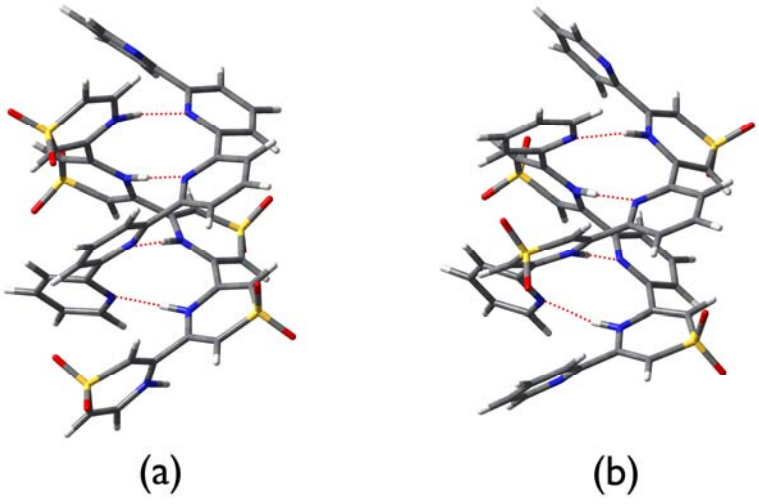


Figure 5.11: Structure of (a) the AAAAA-DDDDD helix and (b) the ADADA helix.

As shown in **Figure 5.11**, the helix constructed from the contiguous monomers is very similar to the ADADA helix. The separation between the stacked rings is approximately the same, roughly 3.8 Å. However, the separation between the hydrogen-bonding faces of the monomers is smaller in the AAAAA-DDDDD helix compared to the ADADA helix. The hydrogen bonds in the former range from 1.85 to 1.99 Å, while the hydrogen bonds in the ADADA helix range from 2.02 to 2.07 Å. Relative to the ADADA helix, the other helix is slightly more twisted or contorted, resulting in a column of four rings that are approximately stacked upon one another while the ADADA helix has at most three rings stacked together. In fact, the shortest hydrogen bond in the helix (which is actually the shortest hydrogen bond observed in any of the helices) is found at the more contorted end of the helix where the final ring in the stack of four is located.

In terms of the energy changes associated with helix formation, the electronic energy and the thermodynamic parameters all predict that the helix constructed from contiguous monomers is significantly energetically more favourable than the unsubstituted ADADA helix. The data for these two helices are provided in **Table 5.9**. The results indicate that the helix formed from the contiguous monomers is almost four times more favourable than the unsubstituted ADADA helix in terms of the electronic energies and enthalpies. The difference between the Gibbs energies is even more significant, with a ΔG^{298} for the AAAAA-DDDDD helix twenty-fold larger than the ADADA helix. Furthermore, it is interesting to note that the entropy change in both cases is roughly the same, which may not be surprising considering the similarity of the two systems. However, the enthalpy changes associated with the two helices are

significantly different, indicating that the bonds formed in the AAAAA-DDDDD helix must be stronger than those in the ADADA helix.

Table 5.9: Relative energy and thermodynamic parameters associated with helix formation for the helices constructed from contiguous monomers.

	Relative Energies (kJ/mol)	**ΔH** (kJ/mol)	**ΔS** (J/mol·K)	**ΔG^{298}** (kJ/mol)
AAAAA-DDDDD Helix	-335.4	-338.1	-293	-250.7
ADADA Helix	-85.8	-85.0	-244	-12.3

Atoms in molecules has been used to characterize the hydrogen bonds present. Analysis of the hydrogen-bond BCPs, reported in **Table 5.10**, indicates that the hydrogen bonds in the AAAAA-DDDDD helix are indeed significantly stronger than those in the ADADA helix. This can be inferred from the fact that the hydrogen bonds are shorter and exhibit larger ρ_{BCP} values. In fact, the longest H-bond in the AAAAA-DDDDD helix is still shorter than the shortest H-bond in the ADADA helix (**Table 5.1**). Collectively, these stronger H-bonds result in a significant enhancement in the binding energy, as well as the enthalpy and Gibbs energy.

Table 5.10: Selected properties of the hydrogen bond critical points in the AAAAA-DDDDD helix. Bond lengths (r) are in angstroms (Å), ellipticity (ε) is dimensionless and all other entries are in atomic units.

Type	r**(N···H)**	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ε
Outer	1.987	0.0287	0.0669	0.0247
Inner	1.914	0.0340	0.0770	0.0302
Inner	1.907	0.0347	0.0776	0.0286
Outer	1.852	0.0393	0.0859	0.0283

5.4 Conclusions

The design and development of self-assembling complexes, especially helices, has been a goal of synthetic chemists since the elucidation of the structure of DNA. The current computational investigation probed the electronic properties of an experimentally prepared ADADA helix. The results indicate that the helix is held together with four strong hydrogen bonds as well as many other weak interactions. Determination of the electronic energy changes, as well as thermodynamic parameters, suggests that helix formation is a favourable process, driven by the formation of the hydrogen bonds. For instance, the unsubstituted helix has an electronic binding energy of –85.8 kJ/mol. Furthermore, the strength of binding can be tuned to some extent by the careful selection of substituents. The hydrogen bonds are strengthened when the pyridine ring (H-bond acceptor) is substituted with an electron donating group such as an amine, while electron withdrawing groups on the thiazine ring (H-bond donor) are preferred.

However, the most significant enhancement in binding is achieved when the helix is constructed from monomers that consist of contiguous hydrogen-bond acceptors or donors. This so-called AAAAA-DDDDD helix was investigated here and exhibited a binding energy almost four-fold greater than the unsubstituted ADADA helix, at –335.4 kJ/mol, a formidable improvement over the ADADA helix. The results from the current study have illustrated the complex and elegant nature of these self-assembling helices.

Chapter 6: Energetic Consequences of Guanine Quadruplex Formation

6.1 Introduction

The human genome contains all of the instructions necessary for cells to function properly; it contains all of the information required for life as we know it. However, this valuable information is precariously stored within DNA, which is subject to damage from a myriad of environmental sources. In order for cells to ensure that their genetic material remains viable, various protection mechanisms must be imposed. One such mechanism is the capping of the ends of chromosomes with non-coding DNA. These long tracts of nucleotides are known as telomeres and consist of tandem repeats of some specific sequence. For instance, human telomeres consist of 5'-TTAGGG-3' while the protozoan *Tetrahymena* possess tracts of 5'-GGGGTT-3'.[149]

One of the main purposes of having telomeric DNA at the end of a chromosome is to have expendable DNA. Due to the end-replication problem, a short segment is lost from the 5' end of the chromosome during each round of replication.[149] This successive shortening would threaten the viability of the cell's genetic material if coding DNA was lost. However, by capping the ends of chromosomes with telomeres, the cell ensures that its genetic information remains intact. Furthermore, the length of the telomere remaining on any given chromosome can be used as an indication of the age of the cell; the shorter the telomere, the older the cell.[150-152] If a cell is getting too old (i.e., the telomere is too short to protect the chromosome), senescence or even cell death mechanisms may be triggered.[150-153] Thus, telomeres contribute to the overall stability of the chromosome.

Telomeres have also been implicated in various disease states. More precisely, it is the enzyme that is capable of maintaining and extending telomeres that is the main culprit. Telomerase is a specialized, reverse-transcriptase that can extend telomeres using an intrinsic RNA template. If the length of a given telomere directly correlates with the life span of that cell, telomerase is capable of subverting the normal cell life cycle. In fact, it has been suggested that telomerase is the causative agent behind the immortality that many tumour cells exhibit. This hypothesis is supported by the finding that over 85% of all tumour cells express high levels of telomerase while negligible activity is detected in somatic cells.[154,155] The different levels of enzyme activity presents an opportunity to selectively target the tumour cells.[156-158] This fact has not gone unnoticed; various strategies to inhibit telomerase can be found throughout the literature (reviewed in Olaussen[159]).

One such strategy exploits the guanine-rich nature of telomeres. Guanine itself is known to self-assemble into quartet structures by forming the usual Watson-Crick hydrogen bonds, as well as an additional Hoogsteen hydrogen bond through N7, as shown in **Figure 6.1**. Furthermore, guanine quartets can stack upon one another to form higher-order structures known as guanine quadruplexes (G-quadruplex), first characterized by Gellert and colleagues in 1962.[160]

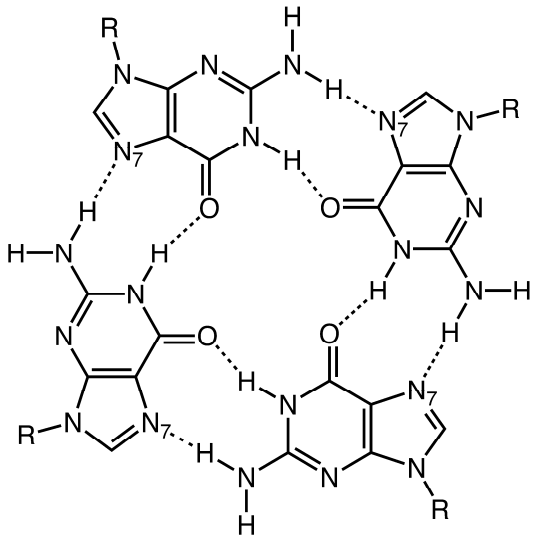


Figure 6.1: Guanine quartet formed from Watson-Crick hydrogen bonds as well as Hoogsteen hydrogen bonds through N7.

At this point, a note should be made about the structure of chromosomes, specifically their ends. Due to the cellular machinery used for replication, the 3' end of the telomere protrudes beyond its complementary cytosine-rich strand. The end result is a guanine-rich, single-stranded overhang, as shown in **Figure 6.2**. In fact, this single stranded portion can be up to hundreds of bases long.[161,162]

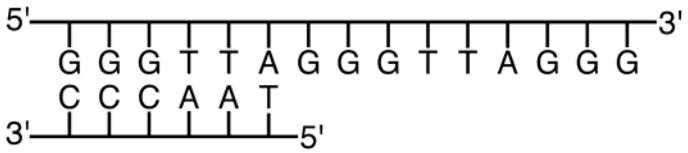


Figure 6.2: Schematic illustration of the single-stranded overhang found at the end of telomeres.

As mentioned previously, guanine tends to self-assemble into quartet structures. Consequently, single-stranded telomere segments are capable of forming a large variety of quadruplex structures using the guanine quartet as its structural motif. Both inter- and

intramolecular structures are possible depending on the number of telomeres that associate to form the quadruplex. In addition, the structures can vary in their strand orientation (parallel versus antiparallel), the glycosidic conformation (*syn/anti*), as well as the arrangement and type of loop connections (diagonal, lateral, double chain reversals). The particular quadruplex structure adopted by a telomere is affected by the presence of cations in the surrounding environment.[163-167] Typically, sodium and potassium cations stabilize these structures by coordinating with the guanine carbonyl groups. With such a large number of variables involved in quadruplex formation, these structures are highly polymorphic as illustrated in **Figure 6.3**.

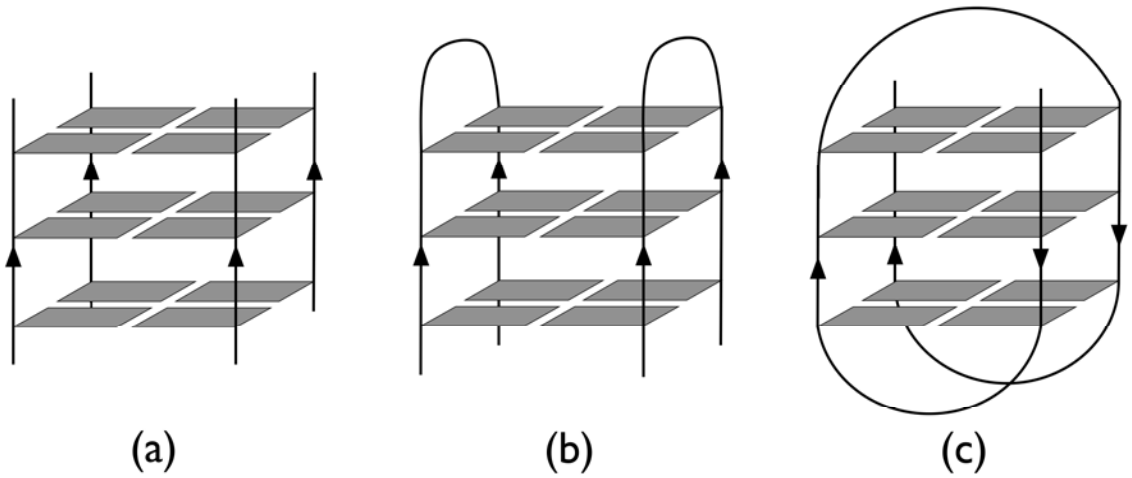


Figure 6.3: Schematic illustration of various quadruplex structures, illustrating their polymorphic nature: (a) an intermolecular, four-stranded parallel quadruplex, (b) an intermolecular dimeric hairpin structure and, (c) an intramolecular basket-type structure.

Although it is well documented that these structures can form *in vitro*,[160,168-174] direct evidence *in vivo* is still lacking. That being said, there is a considerable amount of indirect evidence in the literature. For instance, various proteins have been identified that bind to quadruplexes,[175,176] as well as enzymes that exhibit substrate-specificity for

guanine quadruplexes (e.g. helicases[177] and nucleases[178]). Perhaps the most compelling evidence has been the generation and reaction *in vivo* of a G-quadruplex specific antibody.[179,180]

One of the main reasons that telomeric quadruplexes have received so much attention is the observation that these structures can prevent telomerase from extending telomeres.[181-184] With this inhibition confirmed, there was a large push to develop small molecules that could stabilize quadruplex structures, hopefully enhancing their formation and ultimately resulting in enhanced inhibition. Since the half-life for G-quadruplexes is on the time scale of hours, it was assumed that if the structure could be enticed to form, it should persist.[169] Potential quadruplex-interacting molecules have been developed using a variety of scaffolds, often exploiting large, planar, aromatic systems (e.g. porphyrins, acridines, pyridine, quinines). A number of ligand-quadruplex complexes have been reported and characterized,[185-187] some exhibiting telomerase inhibition[182,188-190] and even anti-cancer activity.[191-195]

Formation of G-quadruplex structures is not restricted to telomeric regions of DNA; guanine-rich tracts abound throughout the human genome. In fact, sequence analysis suggests there may be upwards of 376,000 sequences capable of forming quadruplex structures.[196,197] Furthermore, 40% of all genes have potential G-quadruplex forming sequences within their promoters, suggesting a role for quadruplex structures in gene regulation.[198] Formation of such a structure within the promoter region would prevent transcription and ultimately gene expression. It is also interesting to note that these sequences are more prevalent in proto-oncogenes than in tumour suppressor

genes,[199] providing yet another opportunity to exploit quadruplex structures for therapeutic gains.

Interest in these quadruplex structures is not limited to the pharmaceutical industry alone. For instance, the nanotechnology sector has taken note of the structure and electronic properties of G-quadruplexes. As such, they are being investigated as potential building blocks for nanowires and the like.[200,201] Considering their potential applications ranging from nanotechnology to pharmaceuticals, and their ubiquitous prevalence throughout the human genome, guanine quadruplexes cannot be ignored. In order to fully realize the potential of these structures, a thorough understanding of these quadruplexes must be attained. The current study aims to probe the electronic nature of G-quadruplex structures. Specifically, the energetic changes associated with quadruplex formation is under investigation. Ultimately, the information provided from this study may be used to guide the future development of G-quadruplex stabilizing agents.

6.2 Computational Details

The initial structures of the intramolecular quadruplexes were obtained from Protein Data Bank (PDB) entries 143D and 1KF1. Both structures form in the telomeric regions of human DNA and thus consist of the same sequence, $G_3(T_2AG_3)_3$. However, the quadruplexes are structurally very diverse, differing in the arrangement of the loop segments. The basket-type quadruplex (143D) was acquired in sodium solution using NMR techniques and subsequent refinement with molecular dynamics.[170] Wang and Patel deposited six conformers in the PDB for entry 143D. For the purposes of this study, only one structure was investigated due to the computational demand of the project. Visual inspection of the six available structures reveals that they are all quite

similar with some minor variation in the loop nucleobases. Thus, the quadruplex investigated was selected as the structure with the least number of dangling loop nucleobases (i.e., the most compact structure). On the other hand, the propeller-type quadruplex (1KF1) represents the crystal structure obtained in the presence of potassium ions.[171] Using the coordinates provided in the PDB entries, constrained AM1[42] energy minimizations were performed on the structures such that only the hydrogen atoms and any additional sodium atoms were allowed to optimize. Single point energy calculations were performed on the resultant structures at the B971/6-31+G(d,p) level of theory. The B971 functional[139] was selected on the basis of a previous study by Johnson and DiLabio, which suggests that this functional performs well for systems with hydrogen bonds and van der Waals interactions.[202] All electronic structure calculations were performed with the Gaussian 03 suite of programs.[47] Using the electron densities obtained from the single point calculations, individual atomic energies were calculated using the theory of atoms in molecules, as implemented in AIMPAC.[114]

Due to the large size of the quadruplexes, the systems were decomposed into smaller fragments so that the single point calculations could be completed with the available resources. This approach was previously used by Matta in a study of various biological molecules, and the results from the fragments were found to agree well with the results from the intact molecules.[203] On the basis of Matta's results, a preliminary investigation was performed to determine the effects of fragmenting the nucleic acid system. A segment of duplex telomere DNA, with a 5'-TAGGGTTAGG-3' sequence, was used as the test system. The duplex DNA was treated in the same manner as the quadruplex structures; structural coordinates were obtained from PDB entry 1VFC[204]

with an initial optimization of the hydrogen atoms using AM1 followed by single point calculations at the B971/6-31+G(d,p) level. Using the resultant structure, the third G-C pair was investigated while varying the number of surrounding nucleobases that were included in the single point calculations. The four systems investigated were: the G-C pair alone, the G-C pair with their sugar-phosphate groups attached, the G-C pair with one adjacent nucleobase on each side and the G-C pair with two adjacent nucleobases on each side. The atomic energies were computed using AIMPAC.

In order to determine the energetic consequences of quadruplex formation, a single-stranded reference system was required. In collaboration with colleagues in Dublin, the preferred structure of the single-stranded telomere (5'-TAGGGTTAGG-3') was obtained using molecular dynamics simulations.[205] Using the structure provided by Taylor and Watson, the same computational procedure used for the quadruplex structures was followed for the single-stranded telomere. Again, the individual atomic energies were computed using AIMPAC.

6.3 Results and Discussion

6.3.1 Comparison of Guanine-Cytosine Atomic Energies

The first step in this study was to determine if fragmenting the large DNA structures into smaller systems would affect the computed atomic properties. To this end, several segments of duplex telomere DNA were investigated, in each case including a larger portion of the surrounding nucleobases. Using a telomere segment with sequence 5'-TAGGGTTAGG-3', the third guanine-cytosine pair was examined. The fragments used for comparison were: the G-C pair alone (**GC-1**), the G-C pair with their attached

sugar-phosphate groups (**GC-2**), the G-C pair with one extra nucleobase on each side (5'-G**G**T-3') (**GC-3**) and the G-C pair with two additional nucleobases on each side (5'-G**G**GTT-3') (**GC-4**), as shown in **Figure 6.4**.

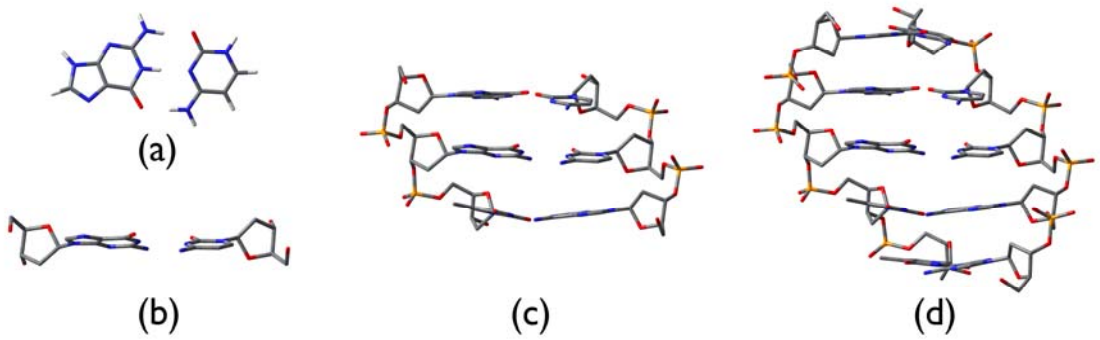


Figure 6.4: Duplex telomere structures investigated: (a) G-C pair, (b) G-C pair with attached sugar-phosphate groups, (c) 5'-GGT-3' and, (d) 5'-GGGTT-3'. Hydrogen atoms have been omitted from (b), (c) and (d) for clarity.

The atomic energies were computed for the G-C pair in each case and are reported in **Table 6.1**. Assuming that the most accurate description of the G-C pair is obtained from the system that includes the largest portion of the surrounding nucleobases, **GC-4** was used as the reference system for comparison. The error in energy was calculated as the difference between the test system and the **GC-4** reference system. However, since the atomic energies vary over a large range (e.g. -75 to -0.4 hartrees) the errors are reported as a percentage of the **GC-4** value so that the errors for different atom types may be compared on the same scale. These values are also reported in **Table 6.1**.

Table 6.1: Atomic energies (hartrees) of guanine and cytosine in the various test systems. The error compared to the **GC-4** value has been computed and reported as a percentage of the **GC-4** value (%E).

	GC-1		GC-2		GC-3		GC-4
Atom	**Energy**	**%E**	**Energy**	**%E**	**Energy**	**%E**	**Energy**
Guanine							
N1	-55.58938	0.115	-55.57988	0.098	-55.53481	0.017	-55.52564
C2	-37.05387	0.067	-37.04933	0.054	-37.03536	0.017	-37.02917
N3	-55.37817	0.104	-55.36855	0.086	-55.33048	0.017	-55.32091
C4	-37.57485	0.047	-37.58341	0.070	-37.56155	0.012	-37.55712
C5	-37.93575	0.062	-37.93780	0.067	-37.91690	0.012	-37.91223
C6	-37.15999	0.061	-37.16089	0.063	-37.14234	0.013	-37.13733
N7	-55.32722	0.087	-55.32539	0.084	-55.29030	0.020	-55.27917
C8	-37.32907	0.009	-37.35583	0.063	-37.33647	0.011	-37.33249
N9	-55.53868	0.023	-55.57506	0.089	-55.53508	0.017	-55.52571
O6	-75.87206	0.110	-75.87628	0.115	-75.80346	0.019	-75.78892
H1	-0.39772	0.152	-0.39736	0.242	-0.39841	0.022	-0.39832
N2	-55.47224	0.106	-55.47411	0.109	-55.42188	0.015	-55.41367
H2a	-0.41209	0.348	-0.41089	0.058	-0.41077	0.028	-0.41066
H2b	-0.44241	0.574	-0.44327	0.770	-0.44044	0.127	-0.43988
H8	-0.59597	1.760	-0.58911	0.589	-0.58516	0.086	-0.58566
Cytosine							
N1	-55.51296	0.080	-55.53296	0.116	-55.47999	0.021	-55.46851
C2	-36.84568	0.014	-36.86343	0.062	-36.84447	0.011	-36.84044
N3	-55.44225	0.094	-55.43957	0.090	-55.40050	0.019	-55.38996
C4	-37.36217	0.048	-37.36049	0.044	-37.34922	0.014	-37.34415
C5	-38.05574	0.031	-38.05516	0.029	-38.05076	0.018	-38.04407
C6	-37.75758	0.008	-37.78465	0.079	-37.75989	0.014	-37.75471
O2	-75.89843	0.082	-75.90608	0.092	-75.84814	0.016	-75.83626
N4	-55.46990	0.089	-55.45960	0.071	-55.42877	0.015	-55.42041
H4a	-0.39206	0.138	-0.39320	0.153	-0.39257	0.007	-0.39260
H4b	-0.45039	0.238	-0.45278	0.291	-0.45322	0.389	-0.45147
H5	-0.60103	0.871	-0.60518	0.188	-0.60888	0.423	-0.60632
H6	-0.59217	2.809	-0.58120	0.905	-0.57499	0.172	-0.57598

At this point a note should be made about the computed atomic energies. Within the AIMPAC suite of programs, atomic properties are calculated via numerical integrations and as such the results are subject to errors due to numerical accuracy. In order to determine the quality of a given integration, the value obtained for the volume

integral of the Laplacian function can be used since it should integrate to zero for any atomic basin.[5] For the purposes of this work, and in fact the entire thesis, integrations were deemed acceptable if the integrated Laplacian value did not exceed an order of magnitude of 10^{-3} a.u. for heavy atoms and 10^{-4} a.u. for hydrogen.[6]

While all of the data presented are within these boundaries, the inherent numerical errors must be taken into account when considering the error percentages reported above. Obviously these numerical errors will have a larger effect on the hydrogen atoms since the atomic values are so much smaller than the heavier atoms. **Table 6.1** illustrates that the hydrogen atoms are often subject to larger percent errors than the heavy atoms; however, the magnitude of the calculated errors is actually quite similar between the hydrogen atoms and the heavy atoms (data not shown).

Neglecting the hydrogen atoms for the time being, and comparing the data for the three model systems, it is apparent from **Table 6.1** and the percent errors that **GC-3** provides a much better description of the G-C pair than the other two systems. This is graphically illustrated in **Figure 6.5**, which plots the percent errors for the three systems. Note that the errors are much smaller across the board for **GC-3** than the other two test systems. On the basis of this data, it was concluded that the larger quadruplex systems could be fragmented such that for any given region of interest, the nearest neighbour nucleobases must be included in the calculation.

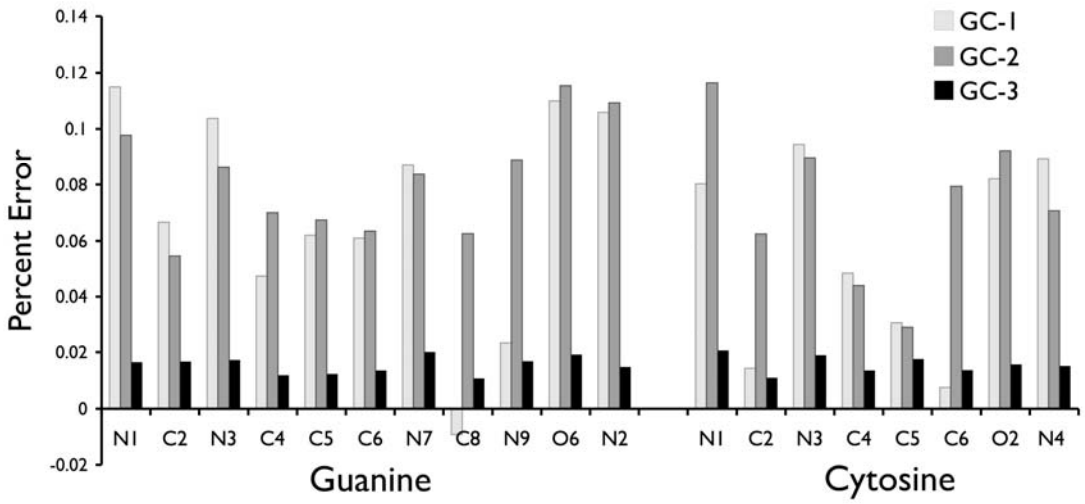


Figure 6.5: Graphical representation of the percent error in the atomic energies of the three test systems relative to **GC-4**.

6.3.2 Single-Stranded Telomere Structure

The final structure of the single-stranded telomere, as obtained from molecular dynamics simulations, remains essentially linear (**Figure 6.6**). Even in the absence of its complementary strand, the telomere appears to maintain stacking interactions between adjacent nucleobases, adopting a slightly helical structure such that these bases may align with one another. The 3' end of the telomere appears slightly puckered, disrupting the co-planar arrangement of the last two guanine nucleobases. This is likely due to the fact that the terminal nucleobase does not have another flanking nucleobase to interact with and so it contorts towards the neighbour that is available. However, this puckering should not affect the results of this study since only the central portion of the telomere (GGGTTA) was used as the reference system. Structural coordinates for the single-stranded telomere are available in **Appendix B**.

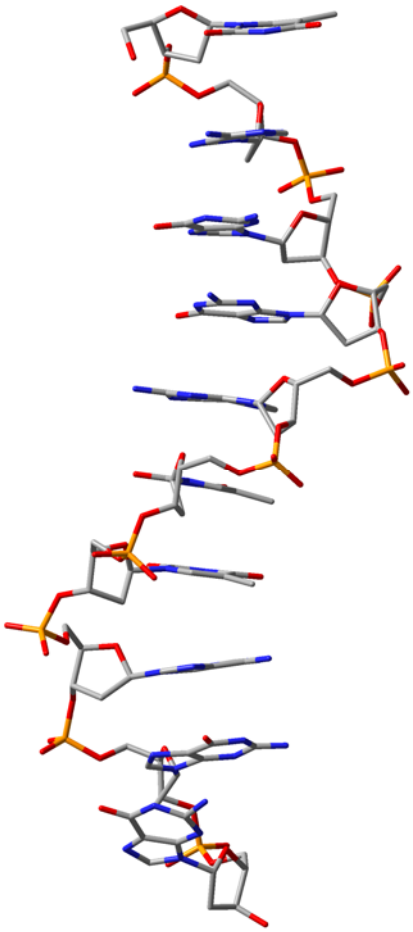


Figure 6.6: Final structure of the single-stranded telomere sequence, 5'–TAGGGTTAGG–3'. The hydrogen atoms have been omitted for clarity.

6.3.3 Basket-type Intramolecular Quadruplex

The first intramolecular quadruplex investigated forms a basket-type structure (**G4-B**), consisting of three stacked guanine quartets linked by one diagonal and two lateral loops as shown in **Figure 6.7**. The overall fold of this quadruplex results in an anti-parallel structure, indicating that the DNA strand orientation around the guanine quartets is variable.

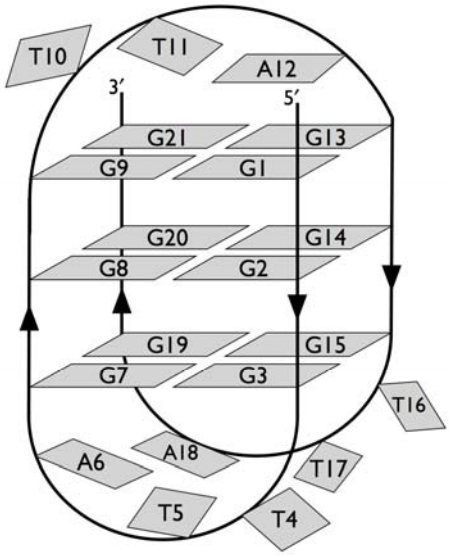


Figure 6.7: Schematic illustration of the basket-type guanine quadruplex and the numbering scheme used for the nucleobases.

The chemical structure of the nucleobases and the numbering schemes used in this thesis are shown in **Figure 6.8**.

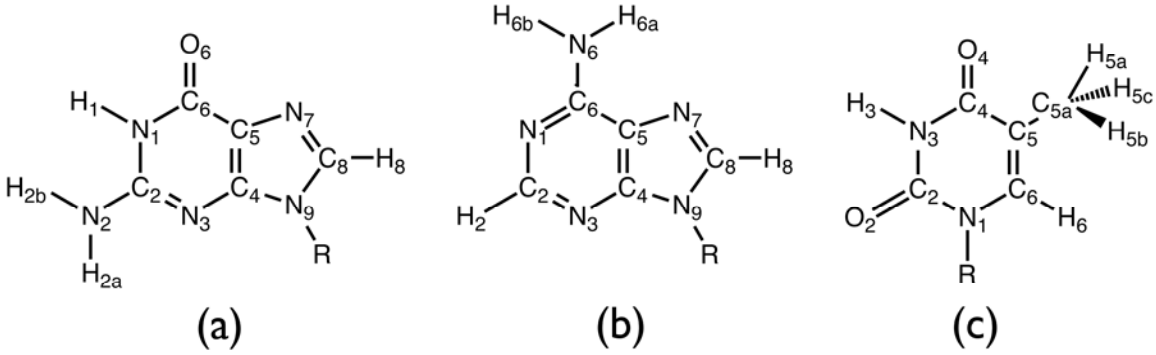


Figure 6.8: Structure and numbering scheme used for (a) guanine, (b) adenine and (c) thymine.

6.3.3.1 Atomic Energy Differences

The atomic energy was calculated for each atom in the quadruplex using the electron density obtained from the single point calculation. The goal of this study was to

determine the energetic consequences of quadruplex formation. This was done by comparing the atomic energy of each atom in the quadruplex to the corresponding atom in the single-stranded telomere. Atomic energy differences (ΔE) were calculated as the difference between the energy of the atom in the quadruplex and the energy of the corresponding atom in the single-stranded telomere, $\Delta E = E_{(\text{atom in quadruplex})} - E_{(\text{atom in ssDNA})}$. If a particular atom is stabilized upon quadruplex formation, the atomic energy difference is negative, while if an atom is destabilized the difference is positive. The detailed atomic data for the basket-type quadruplex are available in **Appendix C**.

Due to the large size of the quadruplex system, questions arose regarding how best to analyze the plethora of data obtained. Simply scanning through the data tables of atomic energy differences gave no indication about local or global regions of stabilization in the folded quadruplex. Therefore, in order to gain information about the quadruplex as a whole, a novel means of displaying the data had to be devised. To this end, the individual atomic energy differences have been plotted on the coordinates of the intact quadruplex in order to visualize the three-dimensional arrangement of the energy changes. Each atom has been colour-coded and size-coded based on the change in energy associated with the unfolded-to-folded transition. Atoms that are stabilized in the folded quadruplex appear as red spheres while those that are destabilized appear as blue spheres. In addition, the size of the sphere corresponds directly to the magnitude of the energy change: the larger the sphere, the larger the atomic energy difference. Those atoms that remain essentially unaffected by the folding process (e.g., $\Delta E \leq 20$ kJ/mol), are represented as small black spheres. The result for the entire basket-type quadruplex is shown in **Figure 6.9**. By representing the atomic energy differences in this fashion, any

trends or patterns have been identified more easily, and are discussed in the following sections.

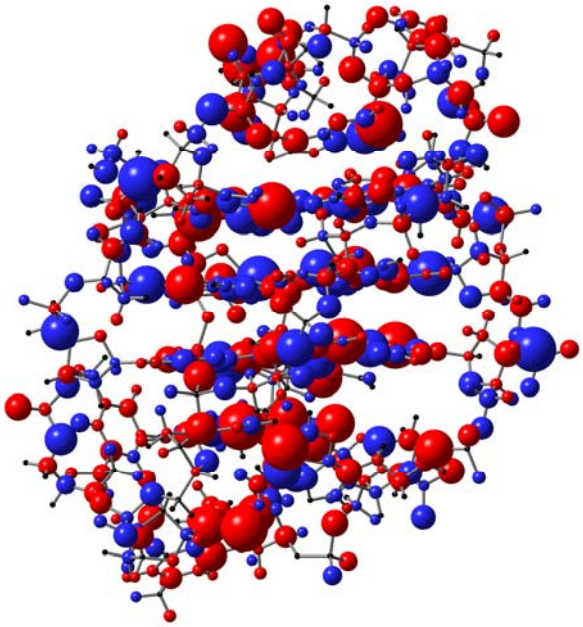


Figure 6.9: Atomic energy differences plotted on the coordinates of the intact quadruplex. Atoms that are stabilized in the folded quadruplex are represented by red spheres, those that are destabilized by blue spheres and those that remain unaffected as black spheres. The size of the sphere corresponds to the magnitude of the energy change.

6.3.3.1.1 Guanine Quartets

Even with the atomic energy differences displayed in a three-dimensional fashion a large amount of data analysis remains. It should be apparent from a quick glance at **Figure 6.9** that quadruplex formation does not result in global stabilization of all the atoms, rather a complex pattern of energy changes emerges. By examining specific regions of the quadruplex on their own, distinct trends were identified. Since the entire structure is held together due to the formation of guanine quartets, this was the first region that was examined more closely. Stripping away the loop regions and sugar-

phosphate backbone reveals that each individual quartet exhibits an alternating pattern of stabilization and destabilization around the hydrogen-bonded ring system. This can be seen more easily when the quartets are unstacked as shown in **Figure 6.10**. For the purpose of this study the 'top' of the quadruplex was defined as the diagonal TTA loop. Thus, the top guanine quartet is composed of nucleobases G1, G9, G13 and G21 while the bottom quartet is composed of G3, G7, G15 and G19 and is closest to the two lateral TTA loops (see **Figure 6.7** for the numbering scheme used).

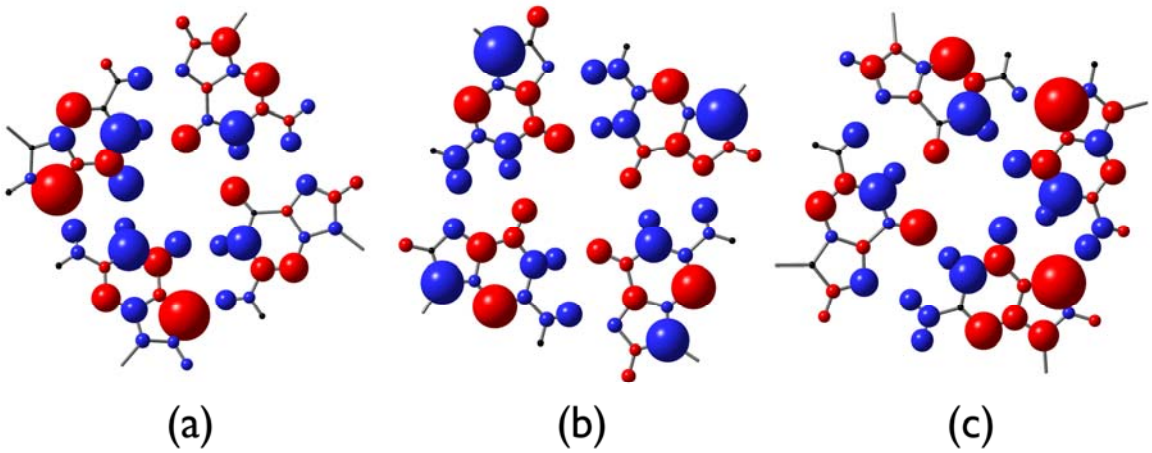


Figure 6.10: Graphical representation of the atomic energy differences for the individual guanine quartets: (a) top, (b) middle and, (c) bottom guanine quartets.

Furthermore, examination of each individual quartet reveals additional trends. For instance, the middle quartet actually exhibits roughly the same pattern of ΔEs on each of the guanines in the quartet. The hydrogen-bonding face (N1, C2 and their attached atoms) is destabilized to a large extent (see **Figure 6.8** for the nucleobase numbering schemes). N9, the nitrogen atom that attaches to the sugar-phosphate backbone, experiences a huge amount of destabilization, ranging between 400-800 kJ/mol for each of the four guanines. The summed atomic energy differences are provided in **Table 6.2**.

Conversely, the central portion of the guanine molecules (N3, C5, C6 and O6) is largely stabilized, especially N3. Interestingly, N7, which participates in the Hoogsteen hydrogen bond, experiences relatively small changes in energy and is destabilized in three of the four guanine molecules. Overall, the atoms in the middle guanine quartet are actually significantly destabilized (by 1800 kJ/mol for the quartet as a whole) compared to the atoms in the single stranded telomere.

On the other hand, the guanine molecules in the other two quartets do not exhibit the same pattern over all four guanines but rather seem to be split into pairs that exhibit similar patterns. For instance, in the top quartet, all four guanine molecules experience a large destabilization and stabilization on N1 and N3, respectively. Other than those two atoms, however, the energy trends diverge for two pairs of guanines. While G1 and G13 are similar to one another, and exhibit large stabilization for N7 and destabilization for O6, G9 and G21 actually exhibit the opposite trends for N7 and O6. The same scenario is seen in the bottom quartet, with all guanine molecules exhibiting a large destabilization for N1 and a large stabilization for N3. Again, the guanine molecules are split into two pairs that exhibit opposing trends: G3 and G15 are stabilized at O6 and destabilized at N7 while the converse is true for G7 and G9. However, unlike the middle quartet, both the top and bottom quartets are stabilized compared to the single-stranded telomere, by approximately –230 and –300 kJ/mol, respectively. As a whole, the central guanine quartet core is destabilized by approximately 1300 kJ/mol.

Table 6.2: Summed atomic energy differences (Σ(ΔE)) for the nucleobases in the basket-type guanine quadruplex as well as the total summation for a given region of interest (e.g. quartet or loop). Energies are in kJ/mol.

Region	Nucleobase	Σ(ΔE)	Total Σ(ΔE)
Top Quartet	G1	-189.1	
	G9	-297.3	
	G13	230.5	
	G21	25.5	
			-230.3
Middle Quartet	G2	496.3	
	G8	630.9	
	G14	248.9	
	G20	425.7	
			1828.8
Bottom Quartet	G3	-154.7	
	G7	-125.8	
	G15	133.8	
	G19	-148.1	
			-294.8
First Lateral Loop	T4	-895.9	
	T5	-1586.0	
	A6	-153.7	
			-2635.5
Diagonal Loop	T10	-680.6	
	T11	-1338.6	
	A12	-201.9	
			-2221.1
Second Lateral Loop	T16	-475.9	
	T17	-1574.8	
	A18	-265.3	
			-2315.9

Perhaps an even more interesting observation was the fact that the regions of stabilization and destabilization actually align with one another when the quartets are stacked in the intact quadruplex. Although it is difficult to see from static images of the quadruplex, **Figure 6.11** shows the central guanine core from a side and top view. From these images, especially the side view, a striated pattern throughout all three quartets can be seen.

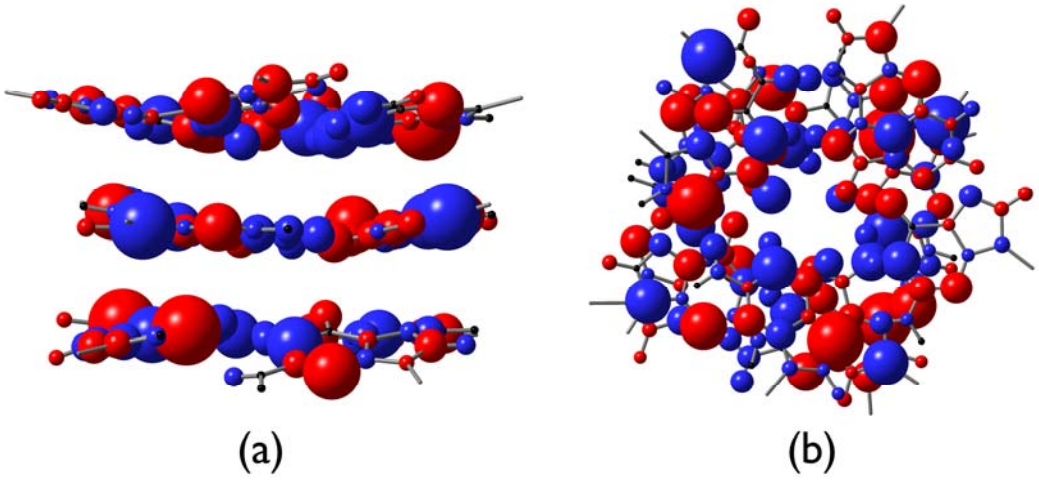


Figure 6.11: Graphical representation of the atomic energy differences for the stacked guanine quartets: (a) side view and (b) top view.

Ultimately this detailed atomic data may prove useful to future drug development endeavours. Designing small molecules to interact with and stabilize quadruplex structures has been a hot topic for several years now. The data provided by this study may help to guide the design of novel compounds by suggesting which regions of the quadruplex need to be further stabilized in order to enhance its formation. The majority of the quadruplex-interacting molecules are based on large, aromatic frameworks that will interact face-on with the guanine quartets. Thus, the information available about the external quartets (i.e., the top and bottom quartets) may be the most instructive for drug design purposes. The following sections discuss the remaining portions of the quadruplex structure, namely the loop regions and the sugar-phosphate backbone. Although most of the small molecules designed to interact with the quadruplex do so with the quartets, it may be important to consider the effect of the loop regions and backbone since they restrict direct access to the quartets. Thus, while these segments have been

largely neglected during drug design to date, it may be possible to exploit these regions to further enhance quadruplex formation.

6.3.3.1.2 Loop Segments

Although all three loop segments are composed of the same sequence of nucleobases (TTA), they all take on distinctly different forms. The basket-type quadruplex has two lateral loops (T4T5A6 and T16T17A18), both located on the bottom face of the quadruplex and one diagonal loop (T10T11A12) that crosses the top face (**Figure 6.7**). In the single-stranded telomere, all of the nucleobases are stacked approximately parallel to one another. However, upon quadruplex formation, many of the loop nucleobases are distorted from this coplanar arrangement. Although the schematic illustration of the basket-type quadruplex (**Figure 6.7**) attempts to indicate the relative orientation of the loop nucleobases, the figure leaves much to be desired in this respect. **Figure 6.12** has been provided to illustrate the basket-type quadruplex in its entirety while highlighting each of the loop segments. However, it is difficult to distinguish all of the important features and nuances from the static image. Thus, the following written description of the positions of the loop nucleobases is provided so that the reader may have a more complete appreciation for the quadruplex structure. Consider, for example, the diagonal loop; all three nucleobases have different orientations with respect to the quadruplex. T10 is directed away from the central quartet core, protruding from the surface of the quadruplex structure. On the other hand, T11 is directed towards the quartet, forming an angle with the plane of the quartet of approximately 45°. The last nucleobase in the loop, A12, maintains the closest resemblance to its position in the single-stranded telomere, stacking in parallel with the

top guanine quartet. Examining the two lateral loops, the positions of the nucleobases differ from those seen in the diagonal loop as well as from each other. The first lateral loop, T4T5A6, shares some similarities with the diagonal loop; however, in this case one of the thymine residues is coplanar with the quartet, while the adenine moiety is at a 45° angle. The second lateral loop, T16T17A18, has one thymine (T16) oriented perpendicularly to the quartet plane while both T17 and A18 are stacked in parallel with the bottom quartet.

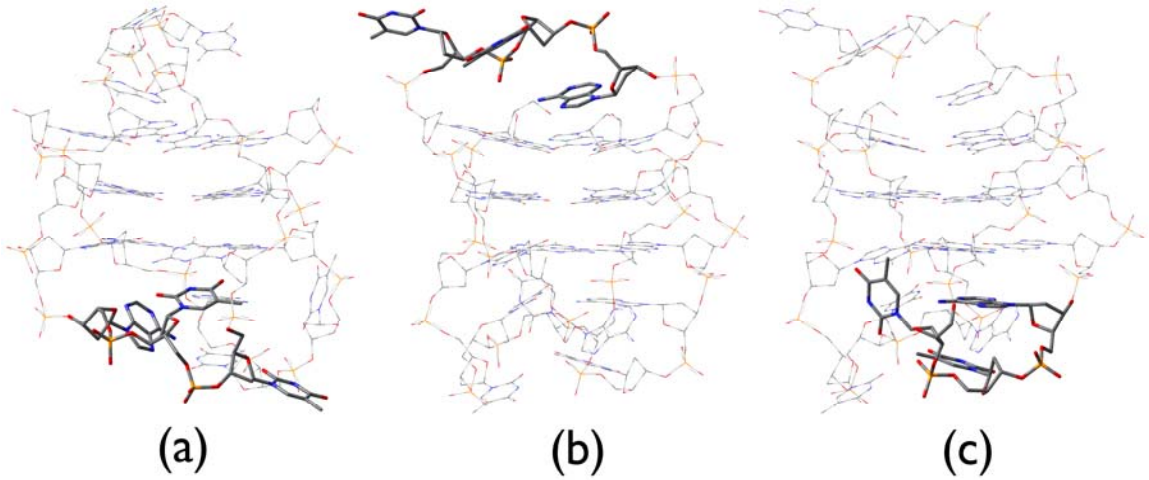


Figure 6.12: Different views of the basket-type quadruplex illustrating the relative position of the loop regions (tubes) to the entire quadruplex (wireframe): (a) first lateral loop (T4T5A6), (b) diagonal loop (T10T11A12) and, (c) second lateral loop (T16T17A18). The hydrogen atoms have been omitted for clarity.

Returning to the focus of this study, which is the energetic changes associated with quadruplex formation, the atomic energy differences observed for the TTA loop regions presented some surprising results. Before any calculations were performed, it was hypothesized that the loop regions would likely exhibit a large amount of destabilization. This assumption was based on the fact that the loop nucleobases were

largely distorted from their coplanar arrangement during the unfolded-to-folded transition, thus eliminating favourable stacking interactions. However, with the atomic energy differences plotted on the intact quadruplex, it was immediately apparent that in fact the opposite was true, the loop regions all exhibit a large degree of stabilization in the quadruplex, as shown in **Figure 6.13**.

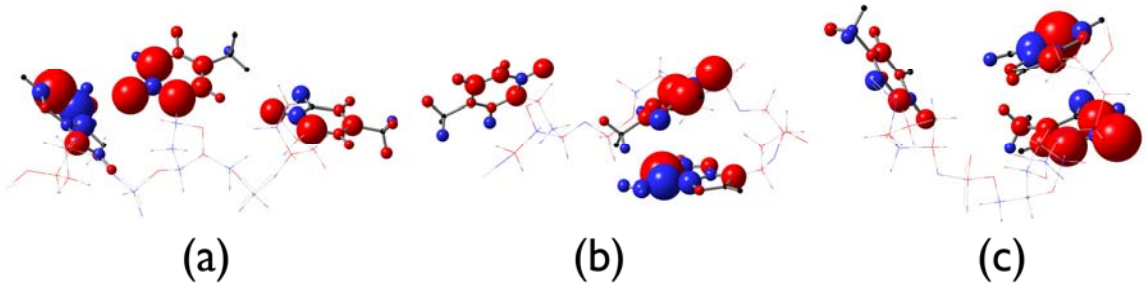


Figure 6.13: Graphical representation of the atomic energy differences for the TTA loop regions: (a) first lateral loop (T4T5A6), (b) diagonal loop (T10T11A12) and, (c) second lateral loop (T16T17A18). The sugar-phosphate backbone is shown in wireframe for clarity.

Examining the nucleobases more closely, it appears that all three loops have roughly the same pattern of stabilization. The first thymine in each loop (T4, T10 and T16) is largely stabilized around the aromatic ring with some minor destabilization observed for C2 and the methyl group. Overall, the atoms in the first thymine nucleobase are stabilized by –475 to –900 kJ/mol per thymine. The summed atomic energy differences for the nucleobases are reported in **Table 6.2**.

The second thymine in the loop segments (T4, T11 and T17) are even more significantly stabilized, between –1340 and –1580 kJ/mol per thymine. Again, there is a large amount of stabilization around the aromatic ring except for C2. Significant destabilization is observed for O4 in T11 and T17, but not T5. Looking at the quadruplex

structure more closely for T5 reveals that the O4 carbonyl group is directed toward the center of the quadruplex and has the potential to interact with several atoms, specifically a near-by amino hydrogen. In the other two loops, the O4 carbonyl is not able to form any additional interactions; in T11 the O4 oxygen is directed away from the quadruplex, completely isolated from other atoms while in T17 the nearest neighbours to O4 are almost 4 Å away. Thus, proximity to other atoms and the ability to form favourable interactions is likely able to explain the discrepancy observed between T5 and T11/T17.

The last component of the loop segments, the adenine moiety, also exhibits a common pattern amongst all three loops. The adenine groups are stabilized to a lesser extent than the thymine, with an overall stabilization between –150 to –265 kJ/mol per adenine (**Table 6.2**). Within each adenine residue, the central portion is largely destabilized (C4, C5 and C6) while there is significant stabilization on N1 and N9.

Interestingly, each nucleobase in the loop regions exhibit a large stabilization (approximately –200 kJ/mol or greater) for the atom that attaches to the sugar-phosphate backbone (i.e., N1 in thymine and N9 in adenine). No such pattern was observed for the guanine residues and in fact the guanine moieties in the middle quartet exhibited extremely large destabilization at N9 (approximately 400 kJ/mol and greater).

Compared to the guanine quartets, of which only a few exhibited overall stabilization, all of the nucleobases in the loop regions were stabilized. Furthermore, considering the individual loops as a whole, the cumulative stabilization amongst the loop nucleobases was upwards of –2000 kJ/mol per loop (see **Table 6.2**). Effectively, this means that the stabilization experienced by the loop nucleobases more than compensates for the destabilization experienced in the guanine quartets. In fact, any one

of the TTA loops alone is stabilized to a greater extent than all of the guanine molecules in the guanine quartet core. Only one component of the guanine quadruplex remains to be examined: the sugar-phosphate backbone. The question at this point is whether the atoms in the backbone will contribute to the overall stabilization achieved largely by the loop nucleobases or not.

6.3.3.1.3 Sugar-Phosphate Backbone

Examining the atomic energy differences of the atoms in the sugar-phosphate backbone reveals that a large portion of these atoms experience relatively small changes in energy upon quadruplex formation. Considering only the atoms in the sugar portion of the backbone, which comprises 273 atoms total, almost 60% of these atoms exhibit energy changes smaller than 50 kJ/mol with an additional 25% under 100 kJ/mol. The plotted atomic energy differences are shown in **Figure 6.14**.

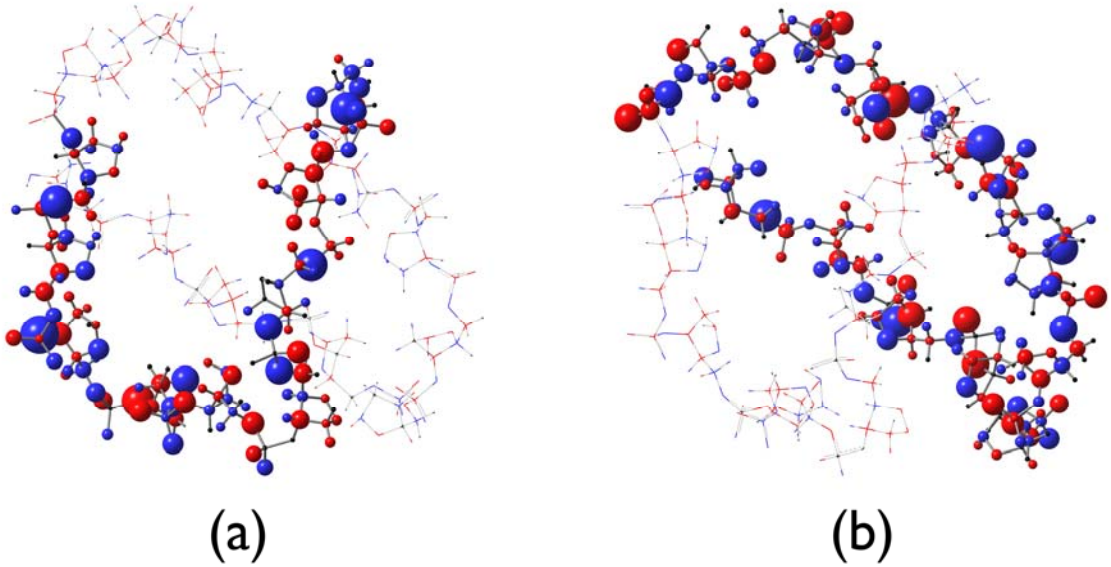


Figure 6.14: Graphical representation of the atomic energy differences for the sugar-phosphate backbone: (a) the first half of the backbone (G1 through G9) and (b) the second half (T10 through G21). In each case the other half of the backbone has been faded out for clarity.

The distribution of the atomic energy differences over the atoms in the deoxyribose rings seems to follow trends dictated by which nucleobase is attached. All of the sugar groups attached to guanine are destabilized except for G2. The overall destabilization for these deoxyribose rings ranges between 200 and 800 kJ/mol whereas G2 is stabilized by almost –600 kJ/mol (**Table 6.3**). Comparing the sugar group of G2 to the other guanine moieties reveals that it has a slightly different structure relative to the attached nucleobase. For G2, the deoxyribose ring is rotated towards the nucleobase such that one face of the ring is directed towards the edge of the guanine nucleobase, resulting in a T-type arrangement. This apparently has a stabilizing effect for the atoms in the ring that is not seen in the other deoxyribose systems. Examining the remaining sugar groups attached to guanine reveals that the observed destabilization is predominantly due to large destabilizations on O5' and O3'. In fact, neglecting the contributions from these two atoms would result in an overall stabilization for the deoxyribose rings. In order to try and explain the large destabilization observed for these oxygen atoms, the atomic structure in the intact quadruplex was inspected. However, no link could be established between the observed destabilization and the local geometry around O5' and O3'.

The atomic energy differences observed in the loop deoxyribose groups are slightly more variable than those of the guanine nucleobases. Again, the energy differences seem to be dependent on the specific nucleobase that is attached. For instance, the deoxyribose group attached to the second thymine base in each loop exhibits small energy changes for the carbon, hydrogen and ring oxygen atoms while large stabilization is seen on the O5' and O3' atoms, dictating the overall large stabilization observed for T5, T11 and T17. The adenine moieties also exhibit relatively small energy

changes for most of the atoms in the deoxyribose ring except for the ether oxygen atom (O4'), which is greatly destabilized. In addition, a large destabilization is observed for O5' while O3' experiences a large stabilization. Overall, the sugar groups attached to the A6 and A12 are stabilized while A18 is destabilized (**Table 6.3**). The first thymine in each loop however does not appear to have a common motif for the three such bases in the quadruplex (i.e., T4, T10 and T16). Although the majority of the ΔE values in the deoxyribose rings are small, there are enough differences to result in a minor stabilization for T4 and a major and minor destabilization for T10 and T16, respectively. A common feature amongst all three of these thymine residues is a major destabilization of O5' and O3', as seen in the guanine deoxyribose rings.

Overall, the sugar groups throughout the entire quadruplex actually serve to destabilize the quadruplex compared to the single-stranded telomere. The 273 atoms that make up the deoxyribose rings contribute a total destabilization of almost 4000 kJ/mol. Interestingly, if the O5' and O3' atoms are neglected from this figure, the remaining atoms in the sugar groups are actually quite stabilized, by –2355 kJ/mol overall. The remaining portion of the backbone, the phosphate groups, have been considered separately from the deoxyribose rings to avoid any double-counting of the atomic energies (since the 3' phosphate group of one nucleotide is actually the 5' phosphate of next). There are some large energy changes observed in the phosphate groups (predominantly stabilizations, see **Table 6.3**), however, most of the large energy changes are attributable to the oxygens rather than the phosphorus atom. There does not appear to be any systematic trend as to which phosphate groups are stabilized and which are destabilized.

Table 6.3: Summed atomic energy differences (Σ(ΔE)) for the ribose groups and the phosphate groups in the basket-type guanine quadruplex. Energies are in kJ/mol.

Nucleobase	Σ(ΔE)	Σ(ΔE)*	Phosphate Group	Σ(ΔE)
G1	830.1	31.5	P^{1-2}	-37.0
G2	-590.6	-644.9	P^{2-3}	-97.4
G3	632.1	-126.5	P^{3-4}	-285.3
T4	-46.4	394.4	P^{4-5}	32.0
T5	-471.0	13.7	P^{5-6}	87.7
A6	-341.3	-126.7	P^{6-7}	-133.4
G7	652.8	-188.5	P^{7-8}	-34.5
G8	-23.1	-84.6	P^{8-9}	-160.6
G9	649.4	-49.5	P^{9-10}	-617.9
T10	303.9	-122.9	P^{10-11}	-246.4
T11	-455.1	51.8	P^{11-12}	-144.6
A12	-260.1	-16.0	P^{12-13}	-308.7
G13	688.9	-292.4	P^{13-14}	-27.3
G14	442.5	20.0	P^{14-15}	0.8
G15	705.2	-10.1	P^{15-16}	-442.9
T16	65.5	-319.6	P^{16-17}	-212.1
T17	-369.2	69.2	P^{17-18}	117.4
A18	189.5	225.1	P^{18-19}	-265.5
G19	460.7	-306.4	P^{19-20}	68.9
G20	180.0	-112.9	P^{20-21}	-125.3
G21	724.2	28.9		
	TOTAL		**1138.5**	

**This figure is the summed atomic energies of the ribose groups excluding O3' and O5'.*

When all is said and done (and counted), the atoms in the folded, basket-type quadruplex are stabilized by –4523.3 kJ/mol relative to the unfolded, single-stranded

telomere. While this may seem like a massive stabilization, considering the size of the system, this really only averages out to a stabilization of roughly −7 kJ/mol per atom. Furthermore, the preceding discussion mentions various atoms that are destabilized upon quadruplex formation. Using the information provided from this work, there is an opportunity to develop potential drug molecules that are aimed at specifically stabilizing those molecules that exhibit large destabilizations. The way in which the wealth of information produced from this study may be used is nearly limitless.

6.3.3.1.4 Glycosidic Conformation

The glycosidic bond in DNA, which links the deoxyribose sugar moiety and the nucleobase, can be used to describe the relative orientation of these two groups. This is done by measuring the dihedral angle (χ) around the glycosidic bond between the plane of the nucleobase and the plane of the ribose ring. In purines, this angle is formed between O4', C1', N9 and C4, while in pyrimidines it is measured through O4', C1', N1 and C2. The nucleobase is said to be in the *syn* conformation if χ is between 0° and ± 90°, otherwise it is in the *anti* conformation. In B-DNA, the *anti* conformation is preferred since it minimizes steric interactions, especially for the pyrimidine bases.

In an effort to explain some of the energy changes observed during quadruplex formation, the glycosidic bond angles were measured in the single-stranded telomere and the basket-type quadruplex (**Table 6.4**). In the single-stranded reference telomere, all of the nucleobases are in the *anti* conformation except G1. Although the measured dihedral angle for G1 technically falls within the range for the *syn* conformation, it is right at the limit of the range. Since this structure was obtained from molecular dynamics simulations, the results may be subject to various errors due to approximations in the

calculations. For the purposes of this study, it was assumed that G1 adopts an *anti* conformation in the single-stranded telomere since it is well documented that nucleobases are usually found in the *anti* conformation in B-DNA.[206]

Table 6.4: Dihedral angles (χ) of the glycosidic bond measured in the single-stranded telomere and the basket-type guanine quadruplex. Angles are in degrees.

Single-Stranded Telomere		**Basket-type Guanine Quadruplex**	
Nucleobase	**χ**	**Nucleobase**	**χ**
G1	-86	G1	-98
G2	-108	G2	+43
G3	-125	G3	-113
T4	-118	T4	-48
T5	-130	T5	-166
A6	-93	A6	-97
		G7	+59
		G8	-101
		G9	+80
		T10	-150
		T11	-136
		A12	+143
		G13	-112
		G14	+78
		G15	-99
		T16	-143
		T17	-123
		A18	-76
		G19	+70
		G20	-116
		G21	+67

The measured dihedral angles indicate that several nucleobases switch from the *anti* to the *syn* conformation upon helix formation, namely G2, G7, G9, G14, G19 and G21. Although it is tempting to try to correlate the large destabilization observed on N9 in G2 with the *anti*-to-*syn* conformational change, further inspection of the measured

dihedral angles indicates that the atomic energy changes on N9 cannot be correlated with the glycosidic dihedral angle. Furthermore, only two nucleobases in the loop regions experience a change in conformation compared to the single-stranded telomere, T4 and A18. In B-DNA, pyrimidines are found only in the *anti* conformation due to unfavourable interactions present in the *syn* conformation. In the guanine quadruplex, however, T4 adopts a *syn* conformation and exhibits a large stabilization not only for the nucleobase as a whole, but also on N1 where the base attaches to the sugar moiety. Thus, it appears that the glycosidic conformation does not influence the observed energy changes in the guanine quadruplexes as it does in B-DNA.

6.3.3.2 Effect of Additional Sodium Cations

As was mentioned in the introduction, mono-valent cations stabilize guanine quadruplexes and can affect the overall structure that is adopted. For the basket-type quadruplex under investigation here, the experimental data was obtained in sodium-containing solution. For this reason, additional quadruplex structures were investigated that included sodium cations during the constrained AM1 optimization.

6.3.3.2.1 Location of Additional Sodium Cations

Although there are some studies that have investigated the preferred binding mode of metal cations to nucleobase quartets,[165,207-211] the exact location and number of cations found within quadruplex structures remains open to debate.[212,213] Using the previous studies as a guide, three additional structures were investigated that included two, three and four sodium cations. Upon completion of the constrained optimizations, the sodium ions adopted positions either between quartets or in the cavity formed by an

external quartet and the loop segments. A schematic illustration of the three systems investigated can be seen in **Figure 6.15**.

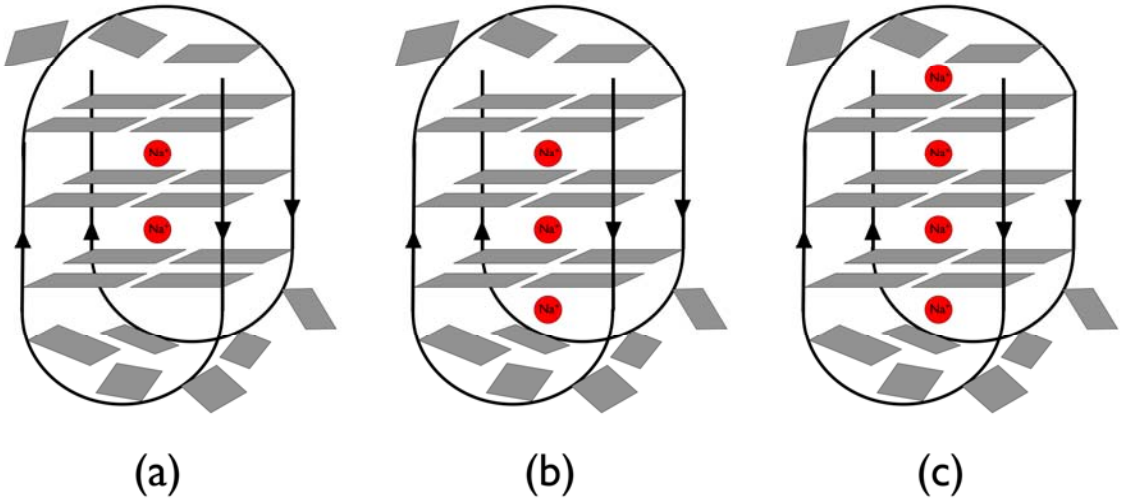


Figure 6.15: Schematic illustration of the location of the sodium cations (red spheres) in the basket-type quadruplex containing (a) two Na^+ ions, (b) three Na^+ ions and (c) four Na^+ ions.

6.3.3.2.2 Difference in Atomic Energy Differences

Individual atomic energies were calculated for each atom in the three systems. Once again, atomic energy differences were calculated relative to the atomic energies of the single stranded telomere and are reported in **Appendix C**. This data was then plotted on the coordinates of the intact quadruplex, resulting in images similar to **Figure 6.9**. However, simply inspecting the resultant plots for any major differences from the quadruplex without sodium cations proved tedious. In order to more easily determine the effect of the added sodium cations, the atomic energy differences between the systems with and without sodium cations were compared. Thus, the difference in atomic energy differences, $\Delta(\Delta E)$, where $\Delta(\Delta E) = \Delta E_{(\text{quadruplex with } Na^+)} - \Delta E_{(\text{quadruplex without ions})}$ were

calculated. This data was then displayed in the same fashion as the atomic energy differences in section **6.3.3.1**. If an atom is stabilized upon the addition of sodium cations ($\Delta(\Delta E) < 0$), the atom appears as a red sphere, if the atom is destabilized ($\Delta(\Delta E) > 0$) it appears as a blue sphere. Those atoms that are unaffected by the inclusion of additional sodium atoms ($\Delta(\Delta E) < 20$ kJ/mol) appear as small black spheres. Again, the size of the sphere corresponds directly with the magnitude of the difference in atomic energy differences.

Examining the $\Delta(\Delta E)$ values plotted in this fashion reveals that the additional sodium cations exert their influence over a relatively short distance, basically affecting only the nearest neighbour atoms. Consider the system with two additional sodium cations located between the quartet planes. The plot reveals that all of the major changes in energy occur in the central core of the quadruplex, with only minor changes observed in the loop regions and throughout the sugar-phosphate backbone (**Figure 6.16a**).

Furthermore, within the guanine quartets that are affected, the largest change is seen on the carbonyl oxygen atoms (O6), every one of which experiences a large stabilization (**Figure 6.16b**). The cumulative effect of the energy changes results in at least a small stabilization for every region of the quadruplex (i.e., nucleobase, sugar group, phosphate), with some large stabilizations observed on the guanine residues (**Table 6.5**). Overall, the basket-type quadruplex that includes two additional sodium cations is −2255 kJ/mol more stable than its counterpart without any sodium ions (**Table 6.6**).

Table 6.5: Summation of the differences in atomic energy differences ($\Sigma(\Delta(\Delta E))$) for the nucleobases in the basket-type guanine quadruplex with varying number of sodium cations. Energies are in kJ/mol.

	$\Sigma(\Delta(\Delta E))$		
Nucleobase	**$2Na^+$**	**$3Na^+$**	**$4Na^+$**
G1	-77.4	-82.4	-99.8
G2	-63.7	-66.1	-111.2
G3	-127.8	-241.9	-217.2
T4	-30.5	-35.2	-50.4
T5	-81.6	-72.6	-127.2
A6	-26.0	-85.6	-102.2
G7	-37.7	-153.1	-154.2
G8	-116.0	-111.1	-118.8
G9	-50.3	-58.5	-137.9
T10	-16.7	-18.5	-20.6
T11	-30.0	-33.1	-51.6
A12	-65.7	-69.8	-169.4
G13	-109.3	-109.3	-191.2
G14	-164.5	-177.3	-206.8
G15	-43.8	-90.2	-87.2
T16	-37.7	-39.8	-66.0
T17	-43.7	-70.6	-99.3
A18	-31.0	15.6	-12.8
G19	-64.9	-80.3	-119.8
G20	-161.9	-193.7	-156.4
G21	-56.4	-60.7	-140.1
TOTAL	-1436.6	-1834.2	-2440.1

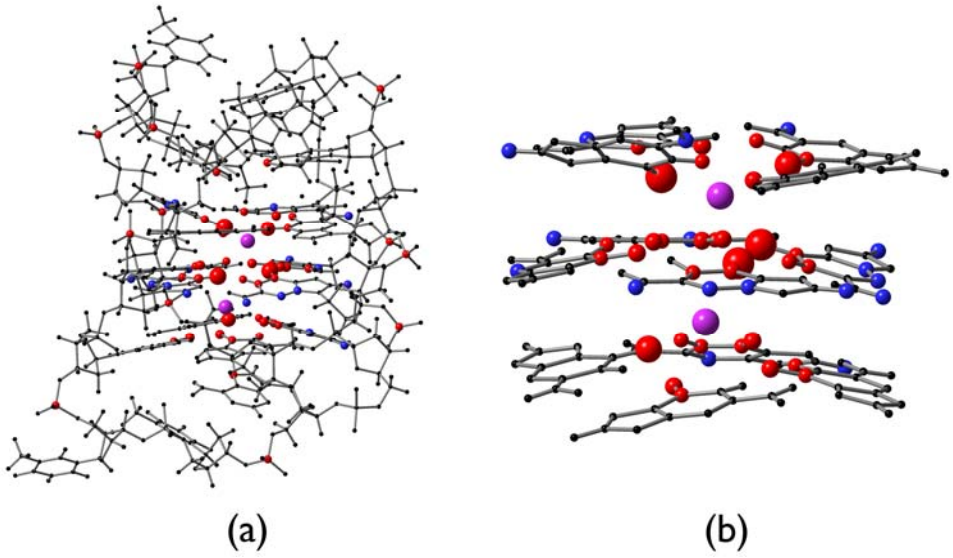


Figure 6.16: Graphical representation of the difference in atomic energy differences, Δ(ΔE), for the basket-type quadruplex with two additional sodium cations (purple spheres): (a) entire quadruplex and (b) close-up of the central guanine core.

Including another sodium cation within the central guanine quartet core, which ends up positioned between the bottom guanine quartet and the two lateral loops, results in a number of additional significant energy changes. The majority of these are located on the first lateral loop (**Figure 6.17**). Otherwise, the magnitude of the Δ(ΔE) values is roughly the same as seen in the two-sodium cation system. The largest total energy difference is observed in the bottom quartet, which is another 300 kJ/mol more stable than the previous system (–565 kJ/mol more than the quadruplex without cations) (**Table 6.6**). Overall, the inclusion of one additional sodium cation results in an increased stabilization of approximately 600 kJ/mol relative to the two-cation system.

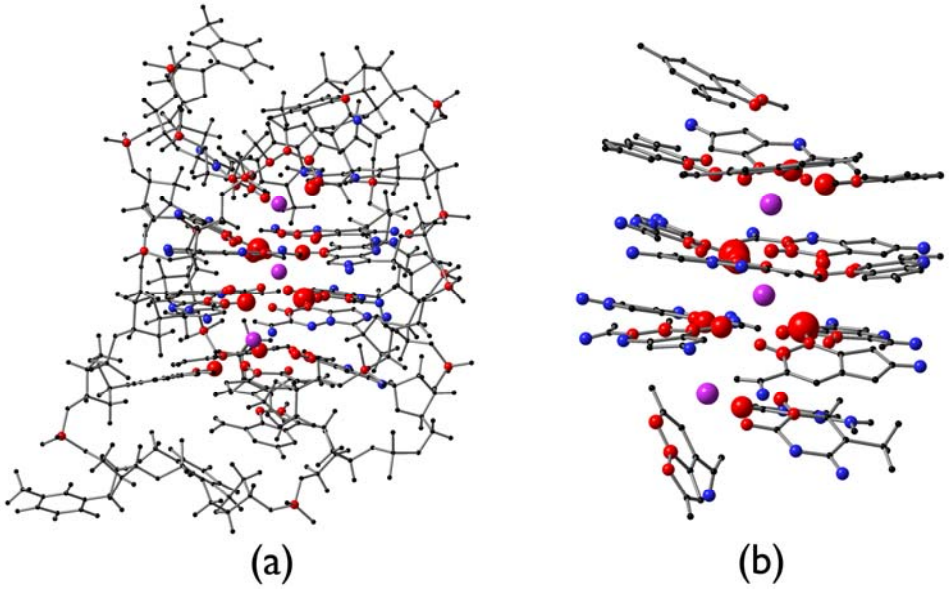


Figure 6.17: Graphical representation of the difference in atomic energy differences, Δ(ΔE), for the basket-type quadruplex with three additional sodium cations (purple spheres): (a) entire quadruplex and (b) close-up of the central guanine core and several loop nucleobases.

Finally, the last system investigated includes one more sodium cation, located between the top quartet and the diagonal loop. Similar changes are observed; the regions closest to the additional cation are affected to the largest extent (**Figure 6.18**). The top quartet is further stabilized by 250 kJ/mol and the diagonal loop by 120 kJ/mol. The total increase in stabilization amounts to 800 kJ/mol more compared to the three-cation system (**Table 6.6**).

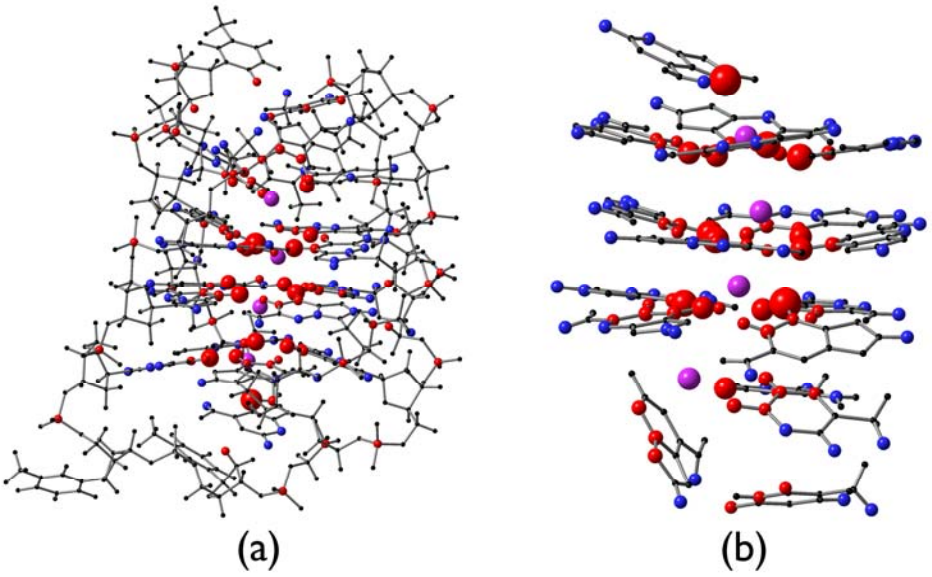


Figure 6.18: Graphical representation of the difference in atomic energy differences, Δ(ΔE), for the basket-type quadruplex with four additional sodium cations (purple spheres): (a) entire quadruplex and (b) close-up of the central guanine core and several loop nucleobases.

Table 6.6: Summation of the differences in atomic energy differences (Σ(Δ(ΔE))) for the quartets, loops and sugar-phosphate backbone in the basket-type guanine quadruplex with varying number of sodium cations. Energies are in kJ/mol.

	Σ(Δ(ΔE))		
Region	**$2Na^+$**	**$3Na^+$**	**$4Na^+$**
Top Quartet	-293.4	-310.9	-569.0
Middle Quartet	-506.1	-548.2	-593.2
Bottom Quartet	-274.2	-565.5	-578.4
Lateral Loop 1	-138.1	-193.4	-279.8
Diagonal Loop	-112.4	-121.4	-241.6
Lateral Loop 2	-112.4	-94.8	-178.1
Backbone	-819.1	-1043.5	-1251.6
TOTAL	-2255.7	-2877.7	-3691.7

Inspection of **Tables 6.5** and **6.6** and **Figures 6.16-6.18** suggest that the inclusion of additional sodium cations can have a large effect on the calculated atomic energies. Some atoms that are destabilized in the first quadruplex structure investigated are actually stabilized in the presence of sodium cations. Taken as a whole, these individual energy changes greatly alter the observed stabilization relative to the unfolded telomere. The global stabilization experienced by the folded, basket-type quadruplex is –6779 kJ/mol, –4701 kJ/mol and –8215 kJ/mol for the two-, three- and four-cation system, respectively.

As was discussed previously, the information provided from this study could be exploited to develop quadruplex-stabilizing drug molecules by tailoring drug candidates to the observed energy patterns. Since the experimental structure was obtained in sodium solution, and any potential therapeutic applications are likely to take place in a similar environment, it is important to include the sodium cations in the model. Thus, any such endeavours should use the detailed atomic data from the cation-stabilized quadruplex structures as their guide.

6.3.4 Propeller-type Intramolecular Quadruplex

The second intramolecular quadruplex investigated forms a propeller-type structure (**G4-P**), again with three stacked guanine quartets. However in this quadruplex the loop regions consist of double-chain reversals (DCR) where a TTA loop segment emerges from one face of the stacked quartets and folds back towards the opposite face. These DCR loop segments result in a quadruplex structure that exhibits parallel strand orientation around the guanine quartets, as shown in **Figure 6.19**. In order to keep the nomenclature similar to that of the basket-type quadruplex, the ‘top’ guanine quartet will

be the quartet closest to the 5' end of the strand while the 'bottom' quartet is closest to the 3' end.

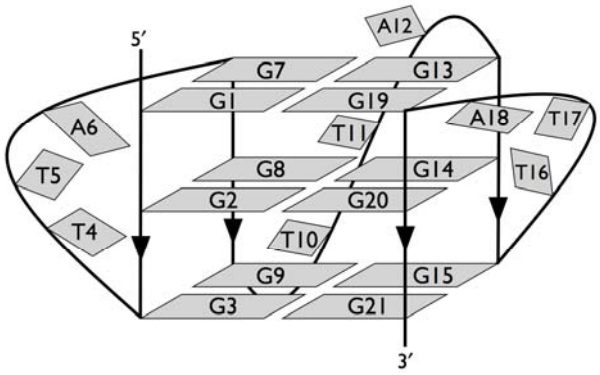


Figure 6.19: Schematic illustration of the propeller-type guanine quadruplex illustrating the numbering scheme used for the nucleobases.

At the time of writing, several fragments of the propeller-type quadruplex system remained under investigation. The following provides the results for the completed fragments.

6.3.4.1 Atomic Energy Differences

The atomic energy differences were calculated and presented in the same manner as described previously for the basket-type quadruplex. The experimental crystal structure of the propeller-type quadruplex included the coordinates of three potassium cations.[171] Thus, unlike the basket-type structure already discussed, there was no need to investigate additional structures with varying numbers of potassium cations; only a single propeller-type structure has been investigated. Two of the potassium cations are located between the guanine quartets with the third above the plane of the top quartet. The completed fragments include the top and bottom guanine quartets and their associated sugar-phosphate backbone. The results for these segments are presented in the following

sections and the detailed atomic energy differences can be found in **Appendix D**. Similar to the basket-type quadruplex (**G4-B**), when the atomic energy differences are plotted on the intact quadruplex, a complex pattern of energy changes emerges (**Figure 6.20**).

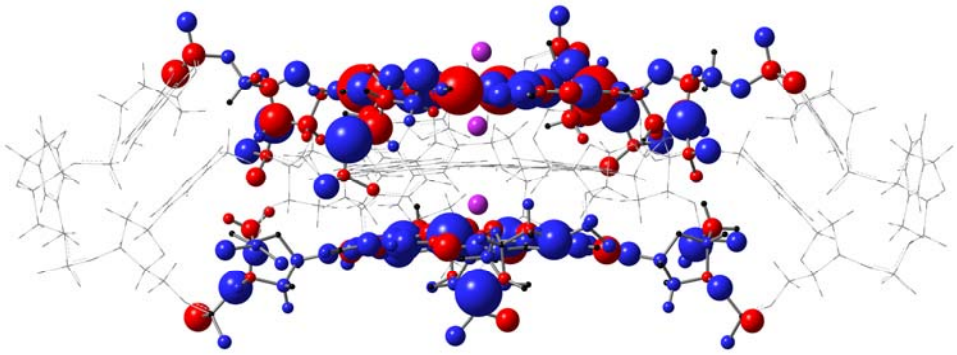


Figure 6.20: Atomic energy differences plotted on the coordinates of the propeller-type quadruplex. Atoms that are stabilized in the folded quadruplex are represented by red spheres, those that are destabilized by blue spheres and those that remain unaffected as black spheres. The size of the sphere corresponds to the magnitude of the energy change. The potassium cations appear as purple spheres and the regions that have yet to be investigated appear as wireframe.

6.3.4.1.1 Guanine Quartets

Examining the top guanine quartet alone reveals that the regions of stabilization and destabilization alternate around the quartet ring (**Figure 6.21a**), similar to what was seen in **G4-B**. However, akin to the middle quartet in the basket-type quadruplex, each guanine actually exhibits the same pattern of energy changes. The hydrogen-bonding face (N1, N2 and their attached hydrogen atoms) is destabilized while N7, the nitrogen atom that participates in the Hoogsteen hydrogen bond, is extremely stabilized (more than 600 kJ/mol). It is also interesting to note that the carbonyl atoms (O6) are significantly destabilized (> 250 kJ/mol), even though they presumably form favourable interactions with the potassium cations. Overall, each guanine nucleobase in the top

quartet of **G4-P** is significantly destabilized, by 600 to 700 kJ/mol per guanine (**Table 6.7**).

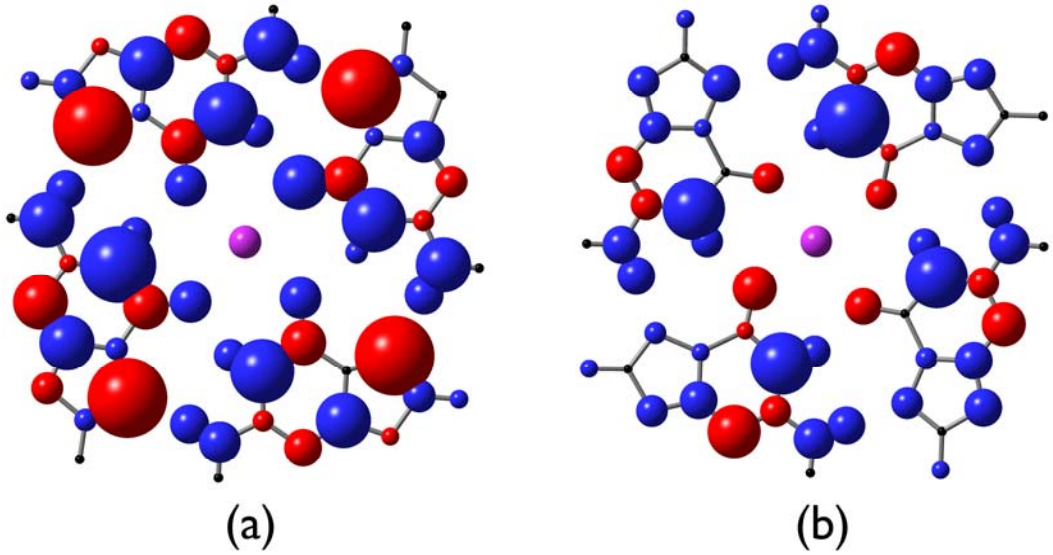


Figure 6.21: Graphical representation of the atomic energy differences for the individual guanine quartets in the propeller-type quadruplex: (a) top and (b) bottom guanine quartets. The purple spheres represent the potassium cations.

Table 6.7: Summed atomic energy differences (Σ(ΔE)) for the nucleobases and the sugar-phosphate groups in the propeller-type guanine quadruplex as well as the total summation for a given region of interest (e.g. quartet). Energies are in kJ/mol.

Region	Nucleobase	$\Sigma(\Delta E)_{base}$	$\Sigma(\Delta E)_{sugar\text{-}P}$	TOTAL
Top Quartet	G1	714.5	747.4	
	G7	665.1	23.2	
	G13	694.2	302.1	
	G19	596.9	394.6	
				4138.0
Bottom Quartet	G3	990.3	775.6	
	G9	863.8	1024.5	
	G15	1072.0	1043.9	
	G21	1167.7	1765.3	
				8703.1

Inspection of the plotted energy differences on the bottom quartet reveals that just like the top quartet, each guanine molecule exhibits the same pattern of energy changes as the others. However, the pattern itself is much different from that observed in the top quartet. The most obvious difference is that **Figure 6.21b** is dominated by blue spheres, indicating that the bottom quartet is predominantly destabilized. There are a few exceptions to this uniform destabilization which include: the carbonyl oxygens (O6), which are stabilized by approximately 150–200 kJ/mol and C2 and N3 which are stabilized by approximately 100 kJ/mol and 200–300 kJ/mol, respectively. As for the destabilizations, the most significant atomic energy difference is seen on N1, which is destabilized by 500–600 kJ/mol. The quartet as a whole still exhibits an alternating pattern of stabilization and destabilization. However, in this case, the stabilized regions are much fewer and much smaller than previously observed. This is the first instance in this investigation where an almost uniform change in energy has been observed. As might be expected, the overall destabilizations for the guanine molecules in the bottom quartet are much larger than those in the top quartet. The summed energy differences range from 860 to 1170 kJ/mol per guanine (**Table 6.7**).

Comparing the quartets shown in **Figure 6.21** suggests that the magnitude of the energy changes tends to be larger in the top quartet than in the bottom. However, since the top quartet has several atoms that experience significant stabilizations, which negate some of the large destabilizations, the overall energy change experienced for the top quartet is smaller than for the bottom.

Although the middle quartet has not yet been investigated, it appears that the alternating regions of stabilization and destabilization in the top and bottom guanine

quartets are roughly aligned in the intact **G4-P**. It will be interesting to see if middle quartet will also follow this trend as was seen in the basket-type quadruplex.

6.3.4.1.2 Sugar-Phosphate Backbone

Similar to the sugar-phosphate groups in the basket-type quadruplex, the atomic energy differences in the backbone attached to the guanine residues were generally smaller than those seen in the nucleobases themselves. The pattern of energy changes in the backbone appears to be dependent upon whether the attached nucleobase is in the top or bottom quartet. Specifically, the backbone attached to a guanine in the top quartet experiences a few more significant energy changes with a large destabilization on O4' and a large stabilization on C3' as well as a large destabilization on O3'. On the other hand, the backbone attached to the bottom quartet guanines only exhibits large destabilizations on O3' and O5' while all the changes in the ribose ring are generally small. Examples of the sugar-phosphate groups attached to the top and bottom quartets are shown in **Figure 6.22**.

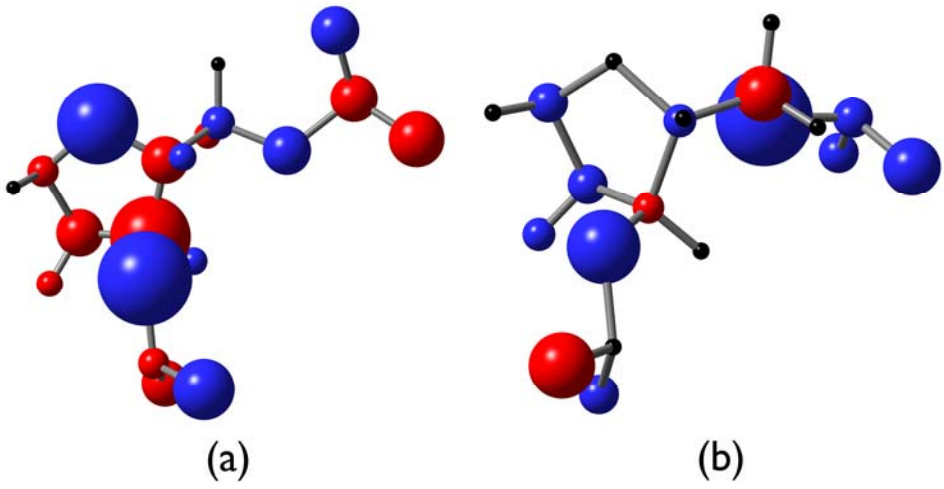


Figure 6.22: Graphical representation of the atomic energy differences for the sugar-phosphate backbone in the propeller-type quadruplex: (a) attached to G13 in the top quartet and (b) attached to G15 in the bottom quartet.

6.3.4.1.3 Glycosidic Bond Conformation

As with the basket-type quadruplex, the glycosidic bond angles were measured to see if any atomic energy changes could be explained by changes in this dihedral angle. The measured angles for the propeller-type quadruplex are presented in **Table 6.8**. In terms of the corresponding angles in the single stranded telomere, all of the guanine molecules in the top quartet (G1, G7, G13 and G19) should be compared to G1 in the telomere which has a measured glycosidic bond angle of −86°. As can be seen from **Table 6.8**, as the glycosidic angle increases for the guanine residues in the top quartet, the atomic energy difference calculated for N9 becomes increasingly negative. It appears that the glycosidic bond angles in the quadruplex are favoured over those adopted in the single-stranded telomere. On the other hand, the guanines in the bottom quartet (G3, G9, G15 and G21) should be compared to G3 in the telomere with a dihedral angle of −125°. In this case, all of the glycosidic bond angles for the guanine residues in the bottom quartet decrease as the atomic energy difference on N9 becomes more positive, the opposite of what is observed in the top quartet. Inspection of the intact structure of **G4-P** suggests that the decreased angle may be the result of the double-chain reversal loop that forms from the subsequent loop bases. Thus, the strained angle may be unavoidable if the quadruplex is to adopt the propeller-type structure. In general, it appears that glycosidic angles between 130−160° are favourable on the basis of the atomic energy of N9.

Table 6.8: Dihedral angles (χ) of the glycosidic bond measured in the propeller-type guanine quadruplex compared to the atomic energy difference for N9 (ΔE_{N9}). Angles are in degrees and energies in kJ/mol.

Top Guanine Quartet			Bottom Guanine Quartet		
Nucleobase	**χ**	**ΔE_{N9}**	**Nucleobase**	**χ**	**ΔE_{N9}**
G1	-160	-109.9	G3	-107	112.5
G7	-136	-50.2	G9	-110	139.3
G13	-120	-24.1	G15	-106	174.7
G19	-129	-34.3	G21	-116	138.5

6.3.5 Comparison Between Basket- and Propeller-type Quadruplexes

Although a large portion of the propeller-type quadruplex has yet to be investigated, comparisons may be drawn between the completed regions and the corresponding regions in the basket-type quadruplex. One of the first points of comparison is the overall structure of the two quadruplexes. While the basket-type structure has its loops above and below the guanine quartet core, and is quite 'tall', for lack of a better word, the loops in the propeller-type quadruplex are positioned around the sides and this structure is quite 'wide'. For instance, **G4-P** is more than 40 Å from the edge of one double-chain reversal loop to the opposite one, but only 11 Å between the top and bottom quartet. On the other hand, **G4-B** is only 21 Å across, but 24 Å from top to bottom. Besides the difference in dimensions, the positioning of the nucleobases is also quite varied amongst the quartets. In the **G4-P** crystal structure, the guanine quartets are much flatter than in the basket structure. Furthermore, each loop in **G4-B** has a different arrangement of the nucleobases while all three DCR loops in **G4-P** are roughly the same with all the nucleobases clustered together. The end result is two very different structures from the same telomeric sequence.

However, the focus of this study was not the structures themselves, since they have been determined experimentally, but rather what are the electronic energy changes associated with quadruplex formation. The calculated atomic energy differences for the two quadruplexes are just as different as their structures. Consider the top guanine quartet in each system. The main difference observed is that all of the guanine residues in **G4-P** exhibit the same distribution of energy changes, while **G4-B** has pairs of guanines with the similar patterns. A quick visual inspection of **Figure 6.10a** and **6.21a** suggests that the energy changes tend to be larger in the propeller structure than the basket. Overall, the guanine nucleobases in **G4-B** are stabilized by –230 kJ/mol whereas they are significantly destabilized in **G4-P** by upwards of 2600 kJ/mol. However, in terms of similarities, both top guanine quartets exhibit alternating regions of stabilization and destabilization evenly distributed over the quartet.

On the other hand, fewer similarities are evident when the bottom quartets are compared. The major difference between the two systems is the almost uniform destabilization of the bottom quartet in **G4-P**, resulting in an overall destabilization of more than 4000 kJ/mol for the nucleobases in the quartet. However, in **G4-B**, larger regions of stabilization are observed, enough so that the bases in the quartet are actually stabilized by –300 kJ/mol.

From the data obtained thus far for the propeller-type structure, it appears that it is much less stable than the basket-type structure. However, it will be interesting to see if the results for the remaining middle quartet, as well as the loop segments, can compensate for the destabilization observed thus far in **G4-P**.

6.4 Conclusions

Individual atomic energies have been computed for two guanine quadruplex structures that form from telomeric tracts of DNA as well as for the single-stranded telomere. The electronic energy changes associated with quadruplex formation have been computed as the difference in atomic energies between the quadruplex and the corresponding atom in the single-stranded reference system. The results indicate that in each case a complex pattern of energy changes emerges rather than any global stabilization or destabilization.

The first structure investigated, a basket-type quadruplex observed experimentally in sodium-containing solution, revealed several interesting trends in the atomic energy differences. First of all, the central guanine core exhibits alternating regions of stabilization and destabilization around each of the guanine quartets that actually align with one another in the intact quadruplex. Overall, the top and bottom quartets were stabilized by roughly 200 to 300 kJ/mol, while the middle quartet is actually destabilized by more than 1800 kJ/mol. However, the nucleobases in the loop segments of the basket-type quadruplex more than compensate for the large destabilization of the middle quartet, each loop being stabilized by more than 2000 kJ/mol. In terms of the energy changes observed in the sugar-phosphate backbone, these tended to be smaller in magnitude than those of the nucleobases. Collectively, the atoms in the backbone actually serve to destabilize the quadruplex by more than 1100 kJ/mol. However, overall the basket-type quadruplex is significantly stabilized compared to the single-stranded telomere, with a total stabilization of –4523 kJ/mol.

Since the experimental structure was obtained in the presence of sodium cations, the effect of additional sodium cations within the basket-type quadruplex was also investigated. The inclusion of two to four sodium cations reveals that these ions have a significant stabilizing effect on the atomic energies. The additional stabilization due to the sodium cations amounted to –2255 kJ/mol for the two-cation system all the way to –3690 kJ/mol for the four-cation system. However, the effect of the sodium cations is relatively short-ranged, affecting only the atoms closest to the additional ions.

In addition to the basket-type structure, a propeller-type quadruplex was investigated, a structure that was originally observed in the solid crystal state. Although only the top and bottom guanine quartets have been investigated so far, the preliminary results indicate that this structure is much less stable than the basket-type structure. However, it is possible that the regions that have yet to be investigated will be able to compensate for the destabilizations observed thus far. Work is underway to complete this portion of the study in order to answer these questions.

Ultimately, this detailed atomic energy data may prove useful for future drug development. There has been a large push to develop small molecules that can stabilize quadruplex structures and enhance their formation. Using the atomic energy differences presented here, potential drug molecules could be tailored to interact with a specific region of the quadruplex and stabilize a region that may be destabilized upon quadruplex formation. Since the potential applications of these drug molecules would be *in vivo*, the data that includes the sodium cations is likely more appropriate than the system without any cations. Further to this, the basket-type quadruplex, which was observed in solution, is likely to be a better representation than the propeller-type crystal structure.

For instance, various aromatic scaffolds (acridines, porphyrins, etc.) have been investigated as potential quadurplex stabilizing agents. It has been assumed that these molecules will interact with the guanine quartet motif. However, the results presented in this thesis suggest that the guanine quartets are not necessarily accessible depending on the quadruplex structure. This allows for the selective targeting of different quadruplex structures. Assuming the basket-type quadruplex is the desired target, the existing drug candidates may be modified to maximize their interaction with the exposed portion of the external quartet as well as adding functional groups that could interact with the TTA loop segments. Since the loop regions in the basket-type quadruplex are already largely stabilized, it will be important to ensure that any drug candidate does not remove this stabilization but rather reinforces it.

The present study has provided a great deal of information about these interesting quadruplex structures. The wealth of data produced is immense and more patterns and trends may yet be identified. Most importantly however, it has provided insight into the folding process, suggesting that folding is not the result of global stabilization of all the atoms, but rather is a delicate balance between the stabilizations and destabilizations of individual atoms.

Chapter 7: Conclusions

This thesis presents the results of several computational investigations ranging from small water clusters to large helices and quadruplexes. This chapter summarizes some of the important results from each study, as well as provides possible suggestions for future work where applicable.

7.1 The Hydrated Electron

The exact structure of an excess electron in water has eluded scientists for some time now. The debate has been heated over whether the electron exists as a localized entity or is simply distributed over the atoms in the water molecules. In an attempt to contribute meaningful insight into the hydrated electron dilemma, small anionic water clusters were investigated using electronic structure methods and the theory of atoms in molecules (AIM). This study represents the first time that the hydrated electron was investigated using AIM. The results suggest that the distribution of the excess electron is largely dependent upon the orientation of the non-hydrogen-bonded (NHB) hydrogen atoms. For those clusters in which several NHB hydrogen atoms are directed towards a central cavity, a non-nuclear attractor (NNA) is detected in the electron density. The presence of the NNA indicates that the cavity-bound state is the preferred mode of binding. On the other hand, when the NHB hydrogen atoms are directed away from the surface of the cluster, the excess electron is simply distributed over the atoms at the surface. In this case, the surface-bound state is preferred. Thus, one mode is not universally favoured over another, but rather the state depends upon the arrangement of the water molecules at any given time.

The presence of the non-nuclear attractor in the electron density led to a secondary investigation about the observed H···NNA interaction. Once again, the theory of atoms in molecules was used, in this case to characterize the novel bond. The results from this portion of the study indicate that the H···NNA bond is a very weak, closed-shell interaction. Although at the outset there were questions about whether this represented some type of novel hydrogen bond, the results of the AIM analysis conclude that it cannot be defined as such. Thus, this bond is probably best described as a weak dipole-ion interaction.

7.2 Self-Assembling Helices

With an eye towards potential applications in molecular recognition, information storage and catalysis, synthetic chemists have sought to develop self-assembling helical structures ever since DNA was discovered. Chapter 5 reports a computational investigation of an experimentally prepared self-assembling helix. The results of this study indicate that the helices are predominantly held together by four strong hydrogen bonds, with several other weak interactions present as well. The monomers in the helix consist of pentameric ring systems with alternating hydrogen bond donor and acceptor groups, resulting in an ADADA helix.

Determination of the energy changes associated with helix formation reveals that the process is favourable. Furthermore, the strength of binding can be tuned by selecting appropriate substituents for the donor and acceptor rings. However, the strongest binding is actually observed when the helix is constructed from monomers composed of either all hydrogen bond donors or acceptors, forming an AAAAA-DDDDD helix.

The foregoing initial survey of the self-assembling helices leaves many possible avenues for future investigation. For instance, the substituent effect portion of the study could be extended to include both donor and acceptor substituents on one helix. Substituent effects could also be investigated on an AAAAA-DDDDD helix.

In fact, the helix constructed from contiguous monomers warrants further investigations since it has been suggested that the stronger binding seen in the AAAAA-DDDDD helix is due to the minimization of repulsive secondary interactions. It would be interesting to determine why the binding is so much stronger in one helix over another.

The helices included in this study were only one size and formed four intermolecular hydrogen bonds. This leaves the possibility of investigating helices of different lengths, with varying number of hydrogen bonds. Furthermore, examining helices that lack the non-hydrogen-bonded rings at the either end of the helix would determine whether these extra rings are important for overall helix stability. It should be apparent from these suggestions that many aspects of the self-assembling helices remain to be investigated.

7.3 Guanine Quadruplexes

Although guanine quadruplexes have received a great deal of attention due to their potential therapeutic applications, relatively little is known about the quadruplexes themselves. The current investigation probed the electronic energy changes associated with quadruplex formation in order to better understand these structures as a whole. Two types of quadruplexes were studied: a basket- and a propeller-type. These two very different structures form from the same telomeric sequence, but under different environmental conditions. The results of the study indicate that the individual atoms in

the quadruplex exhibit a complex pattern of energy changes upon transition from the unfolded to folded state. Furthermore, although the two quadruplexes share a common motif in the guanine quartet, the energy changes observed in the two systems are quite different. Ultimately, the detailed atomic data provided here could be exploited for drug development purposes, tailoring new drug molecules to interact with specific atoms in the quadruplex.

Because of the attention that guanine quadruplexes have received lately, the possibility for future work is almost endless. Obviously, the remaining fragments of the propeller-type quadruplex need to be analyzed. In addition to the two telomeric quadruplexes investigated here, there are other structures available in the literature. For instance, the potassium solution structure is neither a basket- nor propeller-type structure but rather a hybrid somewhere between the two.[174] This third structure could be investigated and the results compared to those presented in this thesis.

As mentioned in the introduction to Chapter 6, the human genome abounds with potential quadruplex forming sequences. It would be interesting to investigate quadruplexes that form in other regions of the genome and compare the results to the telomeric quadruplexes. Furthermore, quadruplex structures have been noted to form in the genomes of other species, presenting another opportunity for comparison.

One of the main reasons that quadruplex structures are the focus of so much research is their potential therapeutic applications. Various studies have investigated the binding of quadruplex-stabilizing agents, which is another possible avenue of investigation. A similar type of analysis could be conducted for the ligand-quadruplex complexes that are available to see where these potential drug molecules stabilize the

quadruplex. With that data in hand, the drug molecules could be modified to further enhance stabilization. Quadruplexes are such a hot topic in the literature currently that the possibilities for further research are immense.

7.4 General Conclusions

Various computational methods have been applied to a range of chemical systems in this thesis. One of the advantages of computational investigations is the ability to gain insight about a system at a level that is often not attainable from experiment. Specifically in this thesis, the theory of atoms in molecules has proven particularly valuable. This theory has provided a novel viewpoint from which to examine the hydrated electron problem. In addition, it has been used to characterize novel bonding interactions in anionic water clusters and the hydrogen bond network in self-assembling helices. Also, its ability to partition a system into atomic fragments has been exploited to provide a detailed atomic view of guanine quadruplex formation. The three projects presented in this thesis illustrate the utility of computational methods and highlight the breadth of systems that may be investigated.

References

1. Levine, I. N., *Quantum Chemistry*. Fifth ed.; Prentice Hall, Inc.: Upper Saddle River, New Jersey, 2000.

2. Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. McGraw-Hill, Inc.: New York, 1989.

3. Koch, W.; Holthausen, M. C., *A Chemist's Guide to Density Functional Theory*. Second ed.; Wiley-VCH: Weinheim, 2001.

4. Bader, R. F. W., *Atoms in Molecules: A Quantum Theory*. Oxford University Press: Oxford, 1990.

5. Popelier, P., *Atoms in Molecules: An Introduction*. Pearson Education Limited: Harlow, England, 2000.

6. *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*; Matta, C. F.; Boyd, R. J., Eds.; Wiley-VCH: Weinheim, 2007.

7. Schrödinger, E., *Ann. Physik* **1926,** 79, 361.

8. Born, M.; Oppenheimer, J. R., *Ann. Physik* **1927,** 84, 457.

9. Slater, J. C., *Phys. Rev.* **1929,** 34, 1292.

10. Slater, J. C., *Phys. Rev.* **1930,** 35, 509.

11. Slater, J. C., *Phys. Rev.* **1930,** 36, 57.

12. Boys, S. F., *Proc. R. Soc. (London) A* **1950,** 200, 542.

13. Binkley, J. S.; Pople, J. A.; Hehre, W. J., *J. Am. Chem. Soc.* **1980,** 102, 939.

14. Ditchfield, R.; Hehre, W. J.; Pople, J. A., *J. Chem. Phys.* **1971,** 54, 724.

15. Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J., *J. Chem. Phys.* **1992,** 96, 6796.

16. Woon, D. E.; Dunning, T. H., Jr., *J. Chem. Phys.* **1993,** 98, 1358.

17. Boys, S. F.; Bernardi, F., *Mol. Phys.* **1970,** 19, 553.

18. Hartree, D. R., *Proc. Cam. Phil. Soc.* **1928,** 24, 89.

19. Hartree, D. R., *Proc. Cam. Phil. Soc.* **1928,** 24, 111.

20. Hartree, D. R., *Proc. Cam. Phil. Soc.* **1928,** 24, 246.

21. Fock, V., *Z. Physik* **1930,** 61, 126.

22. Roothaan, C. C. J., *Rev. Mod. Phys.* **1951,** 23, 69.

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

24. Young, D. C., *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*. John Wiley & Sons, Inc.: New York, 2001.

25. Møller, C.; Plesset, M. S., *Phys. Rev.* **1934,** 46, 618.

26. Feyereisen, M.; Fitzgerald, G.; Komornicki, A., *Chem. Phys. Lett.* **1993,** 208, 359.

27. Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R., *Chem. Phys. Lett.* **1998,** 294, 143.

28. Dunlap, B. I., *PCCP Phys. Chem. Chem. Phys.* **2000,** 2, 2113.

29. Weigend, F.; Kohn, A.; Hattig, C., *J. Chem. Phys.* **2002,** 116, 3175.

30. Jung, Y.; Sodt, A.; Gill, P. M. W.; Head-Gordon, M., *Proc. Natl. Acad. Sci. U. S. A.* **2005,** 102, 6692.

31. *Q-Chem User's Guide, Version 3.1*; Gilbert, A., Ed. Q-Chem, Inc.: Pittsburgh, 2007.

32. Thomas, L. H., *Proc. Cam. Phil. Soc.* **1927,** 23, 542.

33. Fermi, E., *Rend. Accad. Lincei* **1927,** 6, 602.

34. Slater, J. C., *Phys. Rev.* **1951,** 81, 385.

35. Hohenberg, P.; Kohn, W., *Phys. Rev.* **1964,** 136, B864.

36. Kohn, W.; Sham, L. J., *Phys. Rev.* **1965,** 140, A1133.

37. Parr, R. G.; Yang, W., *Desnity-Functional Theory of Atoms and Molecules*. Oxford University Press: New York, 1989.

38. Becke, A. D., *Phys. Rev. A* **1988,** 38, 3098.

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

40. Becke, A. D., *J. Chem. Phys.* **1993,** 98, 5648.

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

42. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P., *J. Am. Chem. Soc.* **1985,** 107, 3902.

43. Stewart, J. J. P., *J. Comput. Chem.* **1989,** 10, 209.

44. Stewart, J. J. P., *J. Comput. Chem.* **1989,** 10, 221.

45. Stewart, J. J. P., *J. Comput. Chem.* **1990,** 11, 543.

46. Stewart, J. J. P., *J. Comput. Chem.* **1991,** 12, 320.

47. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision B.05*, Gaussian, Inc.: Wallingford CT, 2004.

48. Ochterski, J. W., Thermochemistry in *Gaussian*. Gaussian, Inc.: 2000.

49. Bader, R. F. W., *Chem. Rev.* **1991,** 91, 893.

50. Bader, R. F. W.; Nguyen Dang, T. T., *Adv. Quantum Chem.* **1981,** 14, 63.

51. Bader, R. F. W., *J. Phys. Chem. A* **1998,** 102, 7314.

52. Collard, K.; Hall, G. G., *Int. J. Quantum Chem.* **1977,** 12, 623.

53. Keith, T. A.; Bader, R. F. W.; Aray, Y., *Int. J. Quantum Chem.* **1996,** 57, 183.

54. Gatti, C.; Fantucci, P.; Pacchioni, G., *Theor. Chim. Acta* **1987,** 72, 433.

55. Cao, W. L.; Gatti, C.; Macdougall, P. J.; Bader, R. F. W., *Chem. Phys. Lett.* **1987,** 141, 380.

56. Mei, C. J.; Edgecombe, K. E.; Smith, V. H.; Heilingbrunner, A., *Int. J. Quantum Chem.* **1993,** 48, 287.

57. Iversen, B. B.; Larsen, F. K.; Souhassou, M.; Takata, M., *Acta Crystallogr. Sect. B-Struct. Commun.* **1995,** 51, 580.

58. Iversen, B. B.; Larsen, F. K. In *Collected Abstracts*, XVII Congress, International Union of Crystallography, Seattle, Washington, August 8-17; 1996.

59. Barnett, R. N.; Landman, U.; Cleveland, C. L.; Jortner, J., *J. Chem. Phys.* **1988,** 88, 4421.

60. Barnett, R. N.; Landman, U.; Cleveland, C. L.; Jortner, J., *J. Chem. Phys.* **1988,** 88, 4429.

61. Coe, J. V.; Lee, G. H.; Eaton, J. G.; Arnold, S. T.; Sarkas, H. W.; Bowen, K. H.; Ludewigt, C.; Haberland, H.; Worsnop, D. R., *J. Chem. Phys.* **1990,** 92, 3980.

62. Bragg, A. E.; Verlet, J. R. R.; Kammrath, A.; Cheshnovsky, O.; Neumark, D. M., *Science* **2004,** 306, 669.

63. Paik, D. H.; Lee, I. R.; Yang, D. S.; Baskin, J. S.; Zewail, A. H., *Science* **2004,** 306, 672.

64. Hammer, N. I.; Shin, J. W.; Headrick, J. M.; Diken, E. G.; Roscioli, J. R.; Weddle, G. H.; Johnson, M. A., *Science* **2004,** 306, 675.

65. Hammer, N. I.; Roscioli, J. R.; Bopp, J. C.; Headrick, J. M.; Johnson, M. A., *J. Chem. Phys.* **2005,** 123, 7.

66. Herbert, J. M.; Head-Gordon, M., *J. Phys. Chem. A* **2005,** 109, 5217.

67. Turi, L.; Sheu, W. S.; Rossky, P. J., *Science* **2005,** 309, 914.

68. Verlet, J. R. R.; Bragg, A. E.; Kammrath, A.; Cheshnovsky, O.; Neumark, D. M., *Science* **2005,** 307, 93.

69. Khan, A., *J. Chem. Phys.* **2006,** 125, 4.

70. Sommerfeld, T.; Jordan, K. D., *J. Am. Chem. Soc.* **2006,** 128, 5828.

71. Kulkarni, A. D.; Gadre, S. R.; Nagase, S., *Theochem-J. Mol. Struct.* **2008,** 851, 213.

72. Madarasz, A.; Rossky, P. J.; Turi, L., *J. Chem. Phys.* **2009,** 130, 7.

73. Ehrler, O. T.; Neumark, D. M., *Accounts Chem. Res.* **2009,** 42, 769.

74. Turi, L.; Hantal, G.; Rossky, P. J.; Borgis, D., *J. Chem. Phys.* **2009,** 131, 9.

75. Tay, K. A.; Boutin, A., *J. Phys. Chem. B* **2009,** 113, 11943.

76. Ma, L.; Majer, K.; Chirot, F.; von Issendorff, B., *J. Chem. Phys.* **2009,** 131, 6.

77. Griffin, G. B.; Young, R. M.; Ehrler, O. T.; Neumark, D. M., *J. Chem. Phys.* **2009,** 131, 9.

78. Taylor, A.; Matta, C. F.; Boyd, R. J., *J. Chem. Theory Comput.* **2007,** 3, 1054.

79. Thomas, J. M.; Edwards, P. P.; Kuznetsov, V. L., *ChemPhysChem* **2008,** 9, 59.

80. Weyl, W., *Poggendroff's Annalen der Physik und Chemie* **1864,** 121, 601.

81. Kraus, C. A., *J. Am. Chem. Soc.* **1908,** 30, 1323.

82. Dye, J. L., *Science* **1990,** 247, 663.

83. Wade, L. G., Jr., *Organic Chemistry*. Fourth ed.; Prentice Hall, Inc.: Upper Saddle River, New Jersey, 1999.

84. Hart, E. J.; Boag, J. W., *J. Am. Chem. Soc.* **1962,** 84, 4090.

85. Boag, J. W.; Hart, E. J., *Nature* **1963,** 197, 45.

86. Keene, J. P., *Nature* **1963,** 197, 47.

87. Armbruster, M.; Haberland, H.; Schindler, H. G., *Phys. Rev. Lett.* **1981,** 47, 323.

88. Haberland, H.; Langosch, H.; Schindler, H. G.; Worsnop, D. R., *J. Phys. Chem.* **1984,** 88, 3903.

89. Haberland, H.; Ludewigt, C.; Schindler, H. G.; Worsnop, D. R., *J. Chem. Phys.* **1984,** 81, 3742.

90. Haberland, H.; Schindler, H. G.; Worsnop, D. R., *Ber. Bunsen-Ges. Phys. Chem. Chem. Phys.* **1984,** 88, 270.

91. Walker, D. C., *Can. J. Chem.* **1966,** 44, 2226.

92. Jortner, J.; Ottolenghi, M.; Stein, G., *J. Phys. Chem.* **1964,** 68, 247.

93. Schindewolf, U., *Angew. Chem.-Int. Edit.* **1968,** 7, 190.

94. Migus, A.; Gauduel, Y.; Martin, J. L.; Antonetti, A., *Phys. Rev. Lett.* **1987,** 58, 1559.

95. Kimura, Y.; Alfano, J. C.; Walhout, P. K.; Barbara, P. F., *J. Phys. Chem.* **1994,** 98, 3450.

96. Knapp, M.; Echt, O.; Kreisle, D.; Recknagel, E., *J. Chem. Phys.* **1986,** 85, 636.

97. Knapp, M.; Echt, O.; Kreisle, D.; Recknagel, E., *J. Phys. Chem.* **1987,** 91, 2601.

98. Beyer, M. K.; Fox, B. S.; Reinhard, B. M.; Bondybey, V. E., *J. Chem. Phys.* **2001,** 115, 9288.

99. Balaj, O. P.; Balteanu, I.; Fox-Beyer, B. S.; Beyer, M. K.; Bondybey, V. E., *Angew. Chem.-Int. Edit.* **2003,** 42, 5516.

100. Jortner, J., *J. Chem. Phys.* **1959,** 30, 839.

101. Feng, D. F.; Kevan, L., *Chem. Rev.* **1980,** 80, 1.

102. Kim, K. S.; Park, I. J.; Lee, S.; Cho, K.; Lee, J. Y.; Kim, J.; Joannoupulos, J. D., *Phys. Rev. Lett.* **1996,** 76, 956.

103. Kim, J.; Suh, S. B.; Kim, K. S., *J. Chem. Phys.* **1999,** 111, 10077.

104. Weigend, F.; Ahlrichs, R., *PCCP Phys. Chem. Chem. Phys.* **1999,** 1, 4537.

105. Zhan, C. G.; Dixon, D. A., *J. Phys. Chem. B* **2003,** 107, 4403.

106. Lee, H. M.; Suh, S. B.; Tarakeshwar, P.; Kim, K. S., *J. Chem. Phys.* **2005,** 122, 044309.

107. Wallqvist, A.; Thirumalai, D.; Berne, B. J., *J. Chem. Phys.* **1986,** 85, 1583.

108. Landman, U.; Barnett, R. N.; Cleveland, C. L.; Scharf, D.; Jortner, J., *J. Phys. Chem.* **1987,** 91, 4890.

109. Stemmler, K.; von Gunten, U., *Atmos. Environ.* **2000,** 34, 4241.

110. Chameides, W. L., *Science* **1998,** 281, 1152.

111. Beal, M. F., *Trends Neurosci.* **2000,** 23, 298.

112. Shkrob, I. A., *J. Phys. Chem. A* **2006,** 110, 3967.

113. Biegler-Konig, F.; Schonbohm, J.; Bayles, D., *J. Comput. Chem.* **2001,** 22, 545.

114. Biegler-Konig, F. W.; Bader, R. F. W.; Tang, T. H., *J. Comput. Chem.* **1982,** 3, 317.

115. Dennington II, R.; Keith, T.; Millam, J.; Eppinnett, K.; Hovell, W. L.; Gilliland, R. *GaussView, Version 3.09*, Semichem, Inc.: Shawnee Mission, KS, 2003.

116. Cremer, D.; Kraka, E., *Angew. Chem.-Int. Edit. Engl.* **1984,** 23, 627.

117. Koch, U.; Popelier, P. L. A., *J. Phys. Chem.* **1995,** 99, 9747.

118. Pauling, L.; Corey, R. B.; Branson, H. R., *Proc. Natl. Acad. Sci. U. S. A.* **1951,** 37, 205.

119. Watson, J. D.; Crick, F. H. C., *Nature* **1953,** 171, 737.

120. Natta, G.; Pino, P.; Corradini, P.; Danusso, F.; Mantica, E.; Mazzanti, G.; Moraglio, G., *J. Am. Chem. Soc.* **1955,** 77, 1708.

121. Gellman, S. H., *Accounts Chem. Res.* **1998,** 31, 173.

122. Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S., *Chem. Rev.* **2001,** 101, 3893.

123. Nakano, T.; Okamoto, Y., *Chem. Rev.* **2001,** 101, 4013.

124. Cornelissen, J.; Rowan, A. E.; Nolte, R. J. M.; Sommerdijk, N., *Chem. Rev.* **2001,** 101, 4039.

125. Huc, I., *Eur. J. Org. Chem.* **2004**, 17.

126. Yashima, E.; Maeda, K.; Iida, H.; Furusho, Y.; Nagai, K., *Chem. Rev.* **2009,** 109, 6102.

127. Lehn, J. M.; Rigault, A.; Siegel, J.; Harrowfield, J.; Chevrier, B.; Moras, D., *Proc. Natl. Acad. Sci. U. S. A.* **1987,** 84, 2565.

128. Albrecht, M., *Chem. Rev.* **2001,** 101, 3457.

129. Nielsen, P. E., *Accounts Chem. Res.* **1999,** 32, 624.

130. Bisson, A. P.; Carver, F. J.; Eggleston, D. S.; Haltiwanger, R. C.; Hunter, C. A.; Livingstone, D. L.; McCabe, J. F.; Rotger, C.; Rowan, A. E., *J. Am. Chem. Soc.* **2000,** 122, 8856.

131. Prins, L. J.; Reinhoudt, D. N.; Timmerman, P., *Angew. Chem.-Int. Edit.* **2001,** 40, 2382.

132. Sanford, A. R.; Yamato, K.; Yang, X. W.; Yuan, L. H.; Han, Y. H.; Gong, B., *Eur. J. Biochem.* **2004,** 271, 1416.

133. Tanaka, Y.; Katagiri, H.; Furusho, Y.; Yashima, E., *Angew. Chem.-Int. Edit.* **2005,** 44, 3867.

134. Berl, V.; Huc, I.; Khoury, R. G.; Krische, M. J.; Lehn, J. M., *Nature* **2000,** 407, 720.

135. Wang, J.; Meersman, F.; Esnouf, R.; Froeyen, M.; Busson, R.; Heremans, K.; Herdewijn, P., *Helv. Chim. Acta* **2001,** 84, 2398.

136. Li, J. X.; Wisner, J. A.; Jennings, M. C., *Org. Lett.* **2007,** 9, 3267.

137. Newman, S. G.; Taylor, A.; Boyd, R. J., *Chem. Phys. Lett.* **2008,** 450, 210.

138. Perdew, J. P.; Wang, Y., *Phys. Rev. B* **1992,** 45, 13244.

139. Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C., *J. Chem. Phys.* **1998,** 109, 6264.

140. Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C. P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M., *PCCP Phys. Chem. Chem. Phys.* **2006,** 8, 3172.

141. Boyd, R. J.; Choi, S. C., *Chem. Phys. Lett.* **1985,** 120, 80.

142. Boyd, R. J.; Choi, S. C., *Chem. Phys. Lett.* **1986,** 129, 62.

143. Espinosa, E.; Molins, E.; Lecomte, C., *Chem. Phys. Lett.* **1998,** 285, 170.

144. Grabowski, S. J., *J. Phys. Chem. A* **2001,** 105, 10739.

145. Grabowski, S. J.; Sokalski, W. A.; Leszczynski, J., *J. Phys. Chem. A* **2005,** 109, 4331.

146. Domagala, M.; Grabowski, S. J., *J. Phys. Chem. A* **2005,** 109, 5683.

147. LaPointe, S. M.; Farrag, S.; Bohorquez, H. J.; Boyd, R. J., *J. Phys. Chem. B* **2009,** 113, 10957.

148. Wisner, J. A.; Li, J.; Mudraboyina, B. In *Collected Abstracts*, 91st Canadian Chemistry Conference and Exhibition, Edmonton, AB, May 24-28; Canadian Society for Chemistry: 2008; Abstract 867.

149. Blackburn, E. H., *Nature* **1991,** 350, 569.

150. Cooke, H. J.; Smith, B. A., *Cold Spring Harbor Symp. Quant. Biol.* **1986,** 51, 213.

151. Harley, C. B.; Futcher, A. B.; Greider, C. W., *Nature* **1990,** 345, 458.

152. Hastie, N. D.; Dempster, M.; Dunlop, M. G.; Thompson, A. M.; Green, D. K.; Allshire, R. C., *Nature* **1990,** 346, 866.

153. Olovniko.Am, *J. Theor. Biol.* **1973,** 41, 181.

154. Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L. C.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W., *Science* **1994,** 266, 2011.

155. Shay, J. W.; Bacchetti, S., *Eur. J. Cancer* **1997,** 33, 787.

156. Mergny, J. L.; Mailliet, P.; Lavelle, F.; Riou, J. F.; Laoui, A.; Helene, C., *Anti-Cancer Drug Des.* **1999,** 14, 327.

157. Neidle, S.; Parkinson, G., *Nat. Rev. Drug Discov.* **2002,** 1, 383.

158. Rezler, E. M.; Bearss, D. J.; Hurley, L. H., *Curr. Opin. Pharmacol.* **2002,** 2, 415.

159. Olaussen, K. A.; Dubrana, K.; Dornont, J.; Spano, J. P.; Sabatier, L.; Soria, J. C., *Crit. Rev. Oncol./Hematol.* **2006,** 57, 191.

160. Gellert, M.; Lipsett, M. N.; Davies, D. R., *Proc. Natl. Acad. Sci. U. S. A.* **1962,** 48, 2013.

161. Makarov, V. L.; Hirose, Y.; Langmore, J. P., *Cell* **1997,** 88, 657.

162. Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W., *Genes Dev.* **1997,** 11, 2801.

163. Pinnavaia, T. J.; Marshall, C. L.; Mettler, C. M.; Fisk, C. I.; Miles, H. T.; Becker, E. D., *J. Am. Chem. Soc.* **1978,** 100, 3625.

164. Clay, E. H.; Gould, I. R., *J. Mol. Graph.* **2005,** 24, 138.

165. van Mourik, T.; Dingley, A. J., *Chem.-Eur. J.* **2005,** 11, 6064.

166. Dingley, A. J.; Peterson, R. D.; Grzesiek, S.; Feigon, J., *J. Am. Chem. Soc.* **2005,** 127, 14466.

167. Phan, A. T.; Kuryavyi, V.; Patel, D. J., *Curr. Opin. Struct. Biol.* **2006,** 16, 288.

168. Williamson, J. R.; Raghuraman, M. K.; Cech, T. R., *Cell* **1989,** 59, 871.

169. Raghuraman, M. K.; Cech, T. R., *Nucleic Acids Res.* **1990,** 18, 4543.

170. Wang, Y.; Patel, D. J., *Structure* **1993,** 1, 263.

171. Parkinson, G. N.; Lee, M. P. H.; Neidle, S., *Nature* **2002,** 417, 876.

172. Phan, A. T.; Patel, D. J., *J. Am. Chem. Soc.* **2003,** 125, 15021.

173. Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B., *Nucleic Acids Res.* **2005,** 33, 4649.

174. Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J., *Nucleic Acids Res.* **2007,** 35, 6517.

175. Wen, J. D.; Gray, C. W.; Gray, D. M., *Biochemistry* **2001,** 40, 9300.

176. Muniyappa, K.; Anuradha, S.; Byers, B., *Mol. Cell. Biol.* **2000,** 20, 1361.

177. Mohaghegh, P.; Karow, J. K.; Brosh, R. M.; Bohr, V. A.; Hickson, I. D., *Nucleic Acids Res.* **2001,** 29, 2843.

178. Sun, H.; Yabuki, A.; Maizels, N., *Proc. Natl. Acad. Sci. U. S. A.* **2001,** 98, 12444.

179. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A., *Proc. Natl. Acad. Sci. U. S. A.* **2001,** 98, 8572.

180. Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J., *Nat. Struct. Mol. Biol.* **2005,** 12, 847.

181. Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M., *Nature* **1991,** 350, 718.

182. Sun, D. Y.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H., *J. Med. Chem.* **1997,** 40, 2113.

183. Zaug, A. J.; Podell, E. R.; Cech, T. R., *Proc. Natl. Acad. Sci. U. S. A.* **2005,** 102, 10864.

184. Oganesian, L.; Moon, I. K.; Bryan, T. M.; Jarstfer, M. B., *Embo J.* **2006,** 25, 1148.

185. Clark, G. R.; Pytel, P. D.; Squire, C. J.; Neidle, S., *J. Am. Chem. Soc.* **2003,** 125, 4066.

186. Gavathiotis, E.; Heald, R. A.; Stevens, M. F. G.; Searle, M. S., *J. Mol. Biol.* **2003,** 334, 25.

187. Haider, S. M.; Parkinson, G. N.; Neidle, S., *J. Mol. Biol.* **2003,** 326, 117.

188. Izbicka, E.; Wheelhouse, R. T.; Raymond, E.; Davidson, K. K.; Lawrence, R. A.; Sun, D. Y.; Windle, B. E.; Hurley, L. H.; Von Hoff, D. D., *Cancer Res.* **1999,** 59, 639.

189. Shin-ya, K.; Wierzba, K.; Matsuo, K.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H., *J. Am. Chem. Soc.* **2001,** 123, 1262.

190. Cuesta, J.; Read, M. A.; Neidle, S., *Mini Rev Med Chem* **2003,** 3, 11.

191. Binz, N.; Shalaby, T.; Rivera, P.; Shin-Ya, K.; Grotzer, M. A., *Eur. J. Cancer* **2005,** 41, 2873.

192. Burger, A. M.; Dai, F. P.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S., *Cancer Res.* **2005,** 65, 1489.

193. Pennarun, G.; Granotier, C.; Gauthier, L. R.; Gomez, D.; Boussin, F. D., *Oncogene* **2005,** 24, 2917.

194. Moore, M. J. B.; Schultes, C. M.; Cuesta, J.; Cuenca, F.; Gunaratnam, M.; Tanious, F. A.; Wilson, W. D.; Neidle, S., *J. Med. Chem.* **2006,** 49, 582.

195. Zhou, J. M.; Zhu, X. F.; Lu, Y. J.; Deng, R.; Huang, Z. S.; Mei, Y. P.; Wang, Y.; Huang, W. L.; Liu, Z. C.; Gu, L. Q.; Zeng, Y. X., *Oncogene* **2006,** 25, 503.

196. Todd, A. K.; Johnston, M.; Neidle, S., *Nucleic Acids Res.* **2005,** 33, 2901.

197. Huppert, J. L.; Balasubramanian, S., *Nucleic Acids Res.* **2005,** 33, 2908.

198. Huppert, J. L.; Balasubramanian, S., *Nucleic Acids Res.* **2007,** 35, 406.

199. Eddy, J.; Maizels, N., *Nucleic Acids Res.* **2006,** 34, 3887.

200. Kotlyar, A. B.; Borovok, N.; Molotsky, T.; Cohen, H.; Shapir, E.; Porath, D., *Adv. Mater.* **2005,** 17, 1901.

201. Borovok, N.; Iram, N.; Zikich, D.; Ghabboun, J.; Livshits, G. I.; Porath, D.; Kotlyar, A. B., *Nucleic Acids Res.* **2008,** 36, 5050.

202. Johnson, E. R.; DiLabio, G. A., *Chem. Phys. Lett.* **2006,** 419, 333.

203. Matta, C. F., *J. Phys. Chem. A* **2001,** 105, 11088.

204. Hanaoka, S.; Nagadoi, A.; Nishimura, Y., *Protein Sci.* **2005,** 14, 119.

205. Details of the molecular dynamics simulations are as follows. The initial coordinates for the single-stranded telomere were obtained from PDB entry 1VFC. The structure was protonated, solvated and neutralized with sodium counterions using the AMBER9 *Leap* module. DNA was described by the PARMBSC0 force field, water by TIP3P and the sodium atoms by the standard AMBER-99 force field. The system was equilibrated using a multistage protocol. Three independent 30 ns simulations were run starting from the 10, 12 and 14 ps time points of the final equilibration phase. All simulations were performed using the AMBER9 PMEMD module with periodic boundaries and the particle mesh Ewald method was used to evaluate the electrostatics with grid points set no more than 1 Å apart. The cutoff radius for non-bonded interactions was 12 Å with non-bonded list updates every 25 steps. SHAKE was used with a tolerance of 0.00001 to constrain all bonds involving hydrogen allowing for an integration step of 2 fs. Langevin dynamics with a collision frequency of 1 ps-1 was used to maintain a simulation temperature of 300 K. Isotropic position scaling with a relaxation time of 2.0 ps was used to maintain a pressure of 1 atm. These simulations were performed by Justine Taylor and Graeme W. Watson at Trinity College, University of Dublin.

206. Nelson, D. L.; Cox, M. M., *Lehninger Principles of Biochemistry*. Third ed.; Worth Publishers: New York, 2000.

207. Gu, J. D.; Leszczynski, J., *J. Phys. Chem. A* **2002,** 106, 529.

208. Meyer, M.; Suhnel, J., *J. Phys. Chem. A* **2003,** 107, 1025.

209. Meyer, M.; Suhnel, J., *J. Biomol. Struct. Dyn.* **2003,** 20, 507.

210. Davis, J. T.; Spada, G. P., *Chem. Soc. Rev.* **2007,** 36, 296.

211. Setnicka, V.; Novy, J.; Bohm, S.; Sreenivasachary, N.; Urbanova, M.; Volka, K., *Langmuir* **2008,** 24, 7520.

212. Chowdhury, S.; Bansal, M., *J. Phys. Chem. B* **2001,** 105, 7572.

213. Snoussi, K.; Halle, B., *Biochemistry* **2008,** 47, 12219.

Appendix A: Copyright Permission Letters

The following pages are copyright permission letters. The first is from the Royal Society of Chemistry and the second is from the American Chemical Society detailing their policy on Theses and Dissertations.

January 7, 2010

Physical Chemistry Chemical Physics (PCCP)
Royal Society of Chemistry
Thomas Graham House
Science Park
Milton Road,
Cambridge
CB4 0WF

I am preparing my Ph.D. thesis for submission to the Faculty of Graduate Studies at Dalhousie University, Halifax, Nova Scotia, Canada. I am seeking your permission to include a manuscript version of the following paper(s) as a chapter in the thesis:

"Characterization of the bond between hydrogen and the non-nuclear attractor in anionic water clusters"
Alexis Taylor and Russell Jaye Boyd
Physical Chemistry Chemical Physics, Volume 10, pg 6814-6819, 2008.

Canadian graduate theses are reproduced by the Library and Archives of Canada (formerly National Library of Canada) through a non-exclusive, world-wide license to reproduce, loan, distribute, or sell theses. I am also seeking your permission for the material described above to be reproduced and distributed by the LAC(NLC). Further details about the LAC(NLC) thesis program are available on the LAC(NLC) website (www.nlc-bnc.ca).

Full publication details and a copy of this permission letter will be included in the thesis.

Yours sincerely,

Alexis Taylor

Permission is granted for:

a) the inclusion of the material described above in your thesis.

b) for the material described above to be included in the copy of your thesis that is sent to the Library and Archives of Canada (formerly National Library of Canada) for reproduction and distribution.

Name: Gill Cockhead Title: Contracts & Copyright Executive

Signature: Gill Cockhead Date: 8/1/10

PERMISSION GRANTED

The Royal Society of Chemistry is a Signatory to the S.T.M. Guidelines on Permissions

The RSC recommends **the following style** for acknowledgements:

[original citation] — **Reproduced by permission of the PCCP Owner Societies**

American Chemical Society's Policy on Theses and Dissertations

If your university requires a signed copy of this letter see contact information below.

Thank you for your request for permission to include **your** paper(s) or portions of text from **your** paper(s) in your thesis. Permission is now automatically granted; please pay special attention to the implications paragraph below. The Copyright Subcommittee of the Joint Board/Council Committees on Publications approved the following:

Copyright permission for published and submitted material from theses and dissertations ACS extends blanket permission to students to include in their theses and dissertations their own articles, or portions thereof, that have been published in ACS journals or submitted to ACS journals for publication, provided that the ACS copyright credit line is noted on the appropriate page(s).

Publishing implications of electronic publication of theses and dissertation material Students and their mentors should be aware that posting of theses and dissertation material on the Web prior to submission of material from that thesis or dissertation to an ACS journal may affect publication in that journal. Whether Web posting is considered prior publication may be evaluated on a case-by-case basis by the journal's editor. If an ACS journal editor considers Web posting to be "prior publication", the paper will not be accepted for publication in that journal. If you intend to submit your unpublished paper to ACS for publication, check with the appropriate editor prior to posting your manuscript electronically.

If your paper has **not** yet been published by ACS, we have no objection to your including the text or portions of the text in your thesis/dissertation in **print and microfilm formats**; please note, however, that electronic distribution or Web posting of the unpublished paper as part of your thesis in electronic formats might jeopardize publication of your paper by ACS. Please print the following credit line on the first page of your article: "Reproduced (or 'Reproduced in part') with permission from [JOURNAL NAME], in press (or 'submitted for publication'). Unpublished work copyright [CURRENT YEAR] American Chemical Society." Include appropriate information.

If your paper has already been published by ACS and you want to include the text or portions of the text in your thesis/dissertation in **print or microfilm formats**, please print the ACS copyright credit line on the first page of your article: "Reproduced (or 'Reproduced in part') with permission from [FULL REFERENCE CITATION.] Copyright [YEAR] American Chemical Society." Include appropriate information.

Submission to a Dissertation Distributor: If you plan to submit your thesis to UMI or to another dissertation distributor, you should not include the unpublished ACS paper in your thesis if the thesis will be disseminated electronically, until ACS has published your paper. After publication of the paper by ACS, you may release the entire thesis (**not the individual ACS article by itself**) for electronic dissemination through the distributor; ACS's copyright credit line should be printed on the first page of the ACS paper.

Use on an Intranet: The inclusion of your ACS unpublished or published manuscript is permitted in your thesis in print and microfilm formats. If ACS has published your paper you may include the manuscript in your thesis on an intranet that is not publicly available. Your ACS article cannot be posted electronically on a publicly available medium (i.e. one that is not password protected), such as but not limited to, electronic archives, Internet, library server, etc. The only material from your paper that can be posted on a public electronic medium is the article abstract, figures, and tables, and you may link to the article's DOI or post the article's author-directed URL link provided by ACS. This paragraph does not pertain to the dissertation distributor paragraph above.

Appendix B: Structural Coordinates of the Single-Stranded Telomere

The following pages provide the structural coordinates for the single-stranded telomere as determined by molecular dynamics simulations and optimization of the hydrogen atoms with AM1.

H	-0.00000000	-0.00000000	0.00000000
O	0.96953877	-0.00000000	0.00000000
C	1.39170936	1.36368607	0.00000000
H	1.19475636	1.82698157	-1.00162106
H	2.49132096	1.33611182	0.21433608
C	0.65804770	2.06687993	1.09842298
H	1.07448443	3.11864644	1.19434172
O	-0.71940347	2.24464802	0.88752291
C	-1.49716179	1.85735454	1.98639784
H	-1.59657218	2.77174727	2.65160548
N	-2.91718015	1.38264429	1.82785867
C	-3.88106217	1.66611215	2.77300516
H	-3.53093847	2.22402222	3.66468011
C	-5.19346712	1.35618564	2.57394164
C	-6.24679323	1.86326754	3.51073532
H	-6.60737255	2.86242825	3.16291864
H	-7.11181645	1.15576673	3.52650468
H	-5.83987041	1.96266514	4.54596881
C	-5.54030168	0.48320430	1.50595796
O	-6.68369537	0.17277110	1.25190186
N	-4.51187414	0.13520784	0.72951780
H	-4.69404836	-0.55812350	0.02133004
C	-3.18884572	0.51112732	0.79589268
O	-2.37228347	0.22202335	-0.10857181
C	0.80294702	1.33994399	2.42354658
H	1.58686607	0.53530202	2.41204418
C	-0.61729449	0.76155622	2.64743693
H	-0.72129470	-0.21562746	2.11565612
H	-0.86196130	0.62726413	3.73067579
O	1.13466285	2.25427700	3.48245194
P	1.98686874	1.84451223	4.81061784
O	3.12589041	1.07517860	4.39961307
O	2.27954275	3.08373636	5.56246264
O	0.93441691	1.00881523	5.67575642
C	-0.08077944	1.71585958	6.36531779
H	-0.55284402	2.42115277	5.62373899
H	0.40777661	2.31655142	7.18006726
C	-1.13251832	0.79744429	6.94295335
H	-1.98477414	1.47357735	7.28360771

O	-1.71399706	-0.00670286	5.94161129
C	-1.41400275	-1.37438392	6.14924774
H	-2.37072744	-1.89660104	6.47651707
N	-1.12459739	-1.91587821	4.78521042
C	0.05637282	-2.13506690	4.09810824
H	1.02697148	-1.82865071	4.52940437
N	-0.11069738	-2.79112877	2.99279570
C	-1.49823689	-2.91156439	2.86629203
C	-2.36904678	-3.48562220	1.89954560
N	-2.00241319	-3.88894763	0.71337379
H	-2.61021813	-4.43839002	0.15022107
H	-1.02867257	-3.98872360	0.52899666
N	-3.70928538	-3.54492022	2.12902187
C	-4.11330760	-3.11942902	3.33335336
H	-5.19556574	-3.31990483	3.53050700
N	-3.49572486	-2.45831058	4.27429505
C	-2.18600939	-2.37629078	3.94834535
C	-0.69085466	-0.04320850	8.02716822
H	0.27883410	0.27865581	8.50774747
C	-0.41002064	-1.43585218	7.29523067
H	0.64923084	-1.45611551	6.95483001
H	-0.62761490	-2.28471356	7.99056344
O	-1.66083647	-0.07559635	9.04280299
P	-1.53527129	-0.99775214	10.33219968
O	-0.14533727	-1.07624465	10.74636689
O	-2.55960816	-0.45771193	11.34252668
O	-1.96474400	-2.44524297	9.87697986
C	-3.24830786	-2.79168433	9.33746866
H	-3.47175529	-2.05949476	8.51299010
H	-4.00511697	-2.68123029	10.16331948
C	-3.23948576	-4.18013107	8.85018179
H	-4.32167721	-4.51197526	8.76144618
O	-2.68714204	-4.33225465	7.57653837
C	-1.80323930	-5.43516162	7.58697350
H	-2.36536029	-6.38911945	7.37889127
N	-0.76437272	-5.27051725	6.53086736
C	0.52800507	-4.87019447	6.63045433
H	0.93994401	-4.43779893	7.55583505
N	1.25775702	-5.16500580	5.55829479
C	0.28700711	-5.69643761	4.64768918
C	0.34250233	-6.15825813	3.25514748
O	1.26945948	-6.40117423	2.53933492
N	-0.84056959	-6.45877539	2.69192181
H	-0.81158994	-6.84650216	1.76624245
C	-1.95754974	-6.52191540	3.43108621
N	-3.01762982	-6.89333285	2.79324690

H	-2.96756124	-7.69035700	2.19121478
H	-3.89677671	-6.85724451	3.27909807
N	-2.13724087	-6.00541596	4.62436309
C	-0.92049740	-5.66239753	5.21814634
C	-2.47447291	-5.21747186	9.80440209
H	-2.34999830	-4.78828667	10.82802992
C	-1.16684786	-5.38107175	9.00988355
H	-0.52033666	-4.48509431	9.15493380
H	-0.63069485	-6.32287107	9.26312796
O	-3.08244488	-6.44682257	9.83758023
P	-4.23462954	-6.89221231	10.82286583
O	-3.56373181	-7.49082408	11.96644074
O	-5.25856360	-5.81256373	10.98792753
O	-4.74836178	-8.06717961	9.87589340
C	-5.73903453	-7.78557150	8.92287432
H	-5.58338889	-6.73500961	8.55038313
H	-6.74379259	-7.86563158	9.42578894
C	-5.82312608	-8.72129883	7.66003110
H	-6.75284651	-8.39593005	7.10868418
O	-4.76837936	-8.49669512	6.76388370
C	-3.92622337	-9.64919919	6.79207796
H	-4.07431431	-10.20736198	5.82380437
N	-2.48218808	-9.25348064	6.86816941
C	-1.75607972	-8.82406678	7.87131075
H	-2.19555816	-8.61989483	8.87268011
N	-0.49503477	-8.61949095	7.61296912
C	-0.41768931	-8.99924159	6.31111501
C	0.70028114	-9.05430376	5.37018284
O	1.87695518	-8.80259505	5.61966582
N	0.31520666	-9.44543499	4.12184663
H	1.06944916	-9.51804339	3.45921496
C	-0.91354271	-9.81342972	3.78940029
N	-1.10816965	-10.11845746	2.55175429
H	-0.36172652	-10.48974226	2.00136568
H	-1.99330980	-10.51096756	2.29517750
N	-2.00740018	-9.71303551	4.56511525
C	-1.64035262	-9.38444965	5.84078539
C	-5.86921452	-10.20947178	7.98683843
H	-6.34190864	-10.36815789	9.00131906
C	-4.41508161	-10.50111482	8.00010734
H	-3.94149077	-10.16526362	8.95885831
H	-4.18533048	-11.58553234	7.82487342
O	-6.52158274	-11.03617245	7.07891477
P	-6.71585361	-12.64411459	7.23728696
O	-6.75599111	-12.99127954	8.67441355
O	-7.76058280	-13.16378002	6.33891795

O	-5.32252692	-13.19535563	6.73162657
C	-4.99268434	-13.10526496	5.38937516
H	-5.19546237	-12.05790718	5.02633713
H	-5.60624966	-13.83520410	4.79074723
C	-3.55131224	-13.45836531	5.28243510
H	-3.26720070	-13.33965522	4.18746577
O	-2.64749697	-12.62116604	6.01318531
C	-1.42574951	-13.37656798	6.19632124
H	-0.84193179	-13.40730160	5.23301243
N	-0.63343137	-12.62634924	7.18623083
C	-0.96895365	-12.39954015	8.49161192
H	-1.95161635	-12.69764710	8.89287027
N	-0.05288578	-11.74296850	9.10973731
C	0.96850107	-11.58318939	8.20188944
C	2.27287752	-11.11693369	8.40435449
O	2.77881889	-10.48552907	9.33970412
N	3.09423392	-11.34841871	7.36321802
H	4.01774866	-10.95846474	7.46813279
C	2.69787855	-11.82692484	6.17099007
N	3.62981343	-11.85420327	5.26416921
H	4.59361035	-11.84498480	5.51826163
H	3.45130768	-12.32871680	4.40426687
N	1.51263112	-12.35808620	5.96025275
C	0.64054230	-12.18392913	7.02897164
C	-3.20031743	-14.90223737	5.69859700
H	-4.04043414	-15.34292936	6.32468276
C	-2.00930572	-14.72487867	6.55779469
H	-2.31473625	-14.75053106	7.63847174
H	-1.24801115	-15.53536410	6.40108762
O	-3.12815109	-15.66552815	4.54491040
P	-2.96031414	-17.25520385	4.49693810
O	-3.45997192	-17.83358088	5.76825028
O	-3.56197026	-17.80036372	3.22416926
O	-1.35429832	-17.35265777	4.56447554
C	-0.57231655	-16.95838835	3.51302985
H	-0.90695345	-15.93115278	3.17785774
H	-0.71925439	-17.67636320	2.65083333
C	0.92065390	-16.90743991	3.79246660
H	1.45479297	-16.66141554	2.83350788
O	1.25183291	-15.77245268	4.69052052
C	2.16514141	-16.30019673	5.61452214
H	3.21083878	-16.15479757	5.19867201
N	2.18421872	-15.63388036	6.96705720
C	1.16481745	-15.81235525	7.83248381
H	0.28190283	-16.36325566	7.43788138
C	1.15797383	-15.31924469	9.09156800

C	-0.02773915	-15.66164224	9.99224586
H	-0.90607549	-15.96422474	9.37588565
H	0.26514491	-16.49956907	10.67264618
H	-0.28991604	-14.76804175	10.60741912
C	2.32756684	-14.56580586	9.62703391
O	2.42306356	-14.03014573	10.72522399
N	3.35261287	-14.52705910	8.71574673
H	4.20830620	-14.08387512	8.99333012
C	3.34820348	-15.00698147	7.42080898
O	4.37334284	-14.89461799	6.72361245
C	1.44371206	-18.20914468	4.38589524
H	0.63888705	-18.99575406	4.39149038
C	1.85339585	-17.79668152	5.80638980
H	1.00430113	-17.98556378	6.50632626
H	2.75571723	-18.35146230	6.16950106
O	2.58041394	-18.72993573	3.73794537
P	3.46720942	-19.97869055	4.20736609
O	2.59429905	-21.03683463	4.78039535
O	4.41448354	-20.31194464	3.13515522
O	4.40624420	-19.43128732	5.39034905
C	5.49849222	-18.48197438	5.17543295
H	5.04156048	-17.58927988	4.67042994
H	6.23565748	-18.98131104	4.49294543
C	6.15578942	-18.09715803	6.45384309
H	6.95929837	-17.33261481	6.29193284
O	5.20660798	-17.58274317	7.39048924
C	5.63045983	-17.99488769	8.71645464
H	6.53546949	-17.40877170	9.03872044
N	4.52366347	-17.73241650	9.66286677
C	3.29667255	-18.30395210	9.51023697
H	3.13424731	-18.83900232	8.54853819
C	2.32737946	-18.25068588	10.48386274
C	1.00313268	-19.02654493	10.29513825
H	0.41678454	-18.97729489	11.24132843
H	0.41549520	-18.56183995	9.46859956
H	1.23468949	-20.08580490	10.03819916
C	2.56418848	-17.48676422	11.71119005
O	1.80822023	-17.34457016	12.63701236
N	3.88926339	-16.96306608	11.83791826
H	4.09572863	-16.43247487	12.65925198
C	4.86055701	-17.00298038	10.87948129
O	6.00546244	-16.52915059	11.10511104
C	6.72150014	-19.37705028	7.16942699
H	6.66163571	-20.30376241	6.55309212
C	5.90670145	-19.48001009	8.48283239
H	4.97902445	-20.07033747	8.29268744

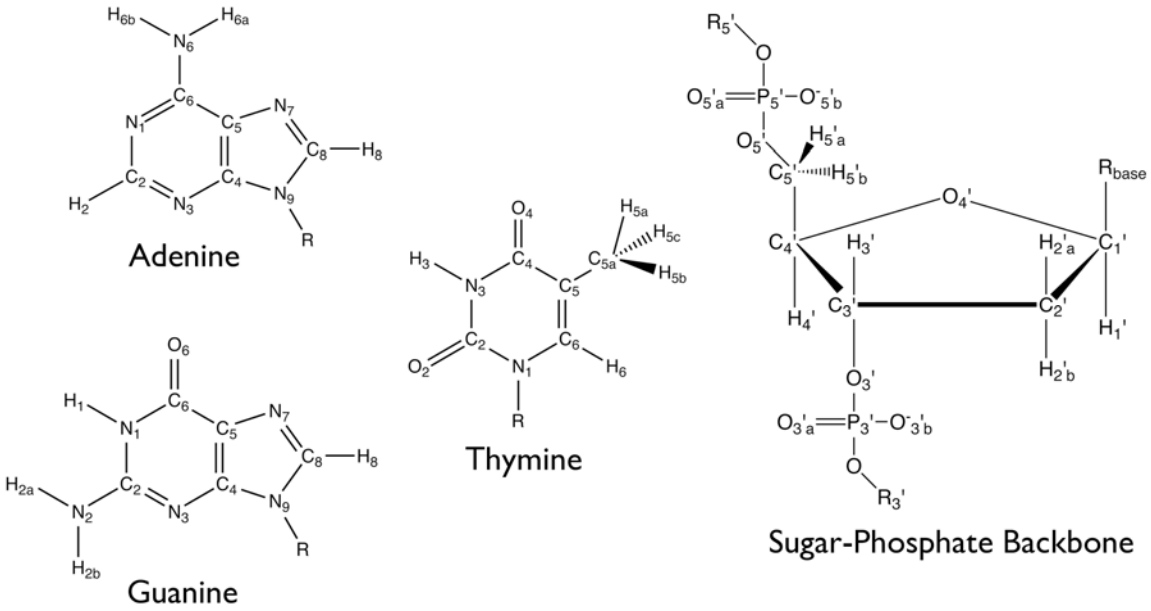
H	6.49573455	-19.93459508	9.31345632
O	8.18024642	-19.13460100	7.42269124
P	9.24413793	-20.22656403	7.88870303
O	8.79565722	-21.59726807	7.44446076
O	10.58948753	-19.80999524	7.56687962
O	8.88853786	-20.08964863	9.45008668
C	9.29218158	-18.92712346	10.12346409
H	8.55015319	-18.10650697	9.91886324
H	10.31383707	-18.63163083	9.76070668
C	9.32783063	-19.17469433	11.64997026
H	9.84145937	-18.31237468	12.15669132
O	8.01072853	-19.18317202	12.11225404
C	7.62714701	-20.43727248	12.63114697
H	7.73563543	-20.39366847	13.76202795
N	6.15075831	-20.67862682	12.42348552
C	5.55407279	-21.34027539	11.35016592
H	6.15552584	-21.71788308	10.50230625
N	4.25565471	-21.40859969	11.46613164
C	4.03498712	-20.82917614	12.71410325
C	2.91176904	-20.54768433	13.42468229
N	1.69305985	-20.88435395	13.08799995
H	0.92578855	-20.38477953	13.49194171
H	1.54176376	-21.11949806	12.12594917
N	2.92662514	-19.86920501	14.66595497
C	4.15181020	-19.46772192	15.04165427
H	4.21444578	-18.88030513	15.98015195
N	5.31649056	-19.72155131	14.41907879
C	5.13829412	-20.40006414	13.29039664
C	9.94792663	-20.52676094	12.07172081
H	10.71505835	-20.91354977	11.35859424
C	8.67155747	-21.40812658	12.06233429
H	8.45219969	-21.69638275	11.00336301
H	8.75377063	-22.32472262	12.68780139
O	10.59831900	-20.33287824	13.36752836
P	11.41302895	-21.48286434	14.23732625
O	11.89823728	-22.45596902	13.21645510
O	12.48847077	-20.91028462	15.07288879
O	10.35629906	-22.16098859	15.25188587
C	9.59862153	-21.36135967	16.13353479
H	9.32369339	-20.41300488	15.58687906
H	10.24293317	-21.10133195	17.02264399
C	8.36301280	-22.04002433	16.62384100
H	7.77045747	-21.36307265	17.30143199
O	7.59118560	-22.31033071	15.48165179
C	7.07677697	-23.56912459	15.57032822
H	6.20383245	-23.59902368	16.30021935

N	6.55927142	-24.08047247	14.29711103
C	7.15163613	-24.37841212	13.13538697
H	8.26169005	-24.40072789	13.06860437
N	6.39155092	-24.55462991	12.09499424
C	5.12465492	-24.51798203	12.67710281
C	3.80696829	-24.68163074	12.14262988
O	3.50543505	-25.04668518	11.04102407
N	2.81037536	-24.34428599	13.02339146
H	1.89695504	-24.31108214	12.59843827
C	2.95319688	-23.96884121	14.28219734
N	1.93024016	-23.83047693	15.02092329
H	1.05519950	-23.54066368	14.63216925
H	2.05516611	-23.50043050	15.96029011
N	4.15547516	-23.83208311	14.86788317
C	5.17904726	-24.09813286	14.00509934
C	8.64671272	-23.38230524	17.27432168
H	9.71117888	-23.51668379	17.60571909
C	8.26291793	-24.31875060	16.13479451
H	9.10695807	-24.37494775	15.39746569
H	8.00454841	-25.34884809	16.47727086
O	7.78828167	-23.56035393	18.35209164
P	8.06202759	-24.71097269	19.49777771
O	9.46812762	-25.17139214	19.46875617
O	7.47741526	-24.20460920	20.71253478
O	7.25347316	-25.94028033	18.93833645
C	5.79218517	-25.92524695	18.91359499
H	5.47806648	-24.95506234	18.45398172
H	5.42709995	-26.03357865	19.96420114
C	5.31121241	-27.11383770	18.07382955
H	4.19318470	-27.19092922	18.25050195
O	5.46415344	-26.86239703	16.67636989
C	5.72239624	-28.12582234	16.04521477
H	4.73512763	-28.54179426	15.69335014
N	6.56423730	-27.87926721	14.85022037
C	7.93201369	-27.77652438	14.78630972
H	8.55247821	-27.66378403	15.68725618
N	8.37587301	-27.65121169	13.53047820
C	7.19579126	-27.78939776	12.77675528
C	7.05404715	-27.58264990	11.36732712
O	7.91352616	-27.38550539	10.51444837
N	5.76399185	-27.60502044	10.94071458
H	5.61076985	-27.24586892	10.00772602
C	4.77961261	-27.96135113	11.73067778
N	3.60155095	-27.87441523	11.14885485
H	3.45811155	-27.20461125	10.41190669
H	2.78495774	-27.99966060	11.70856018

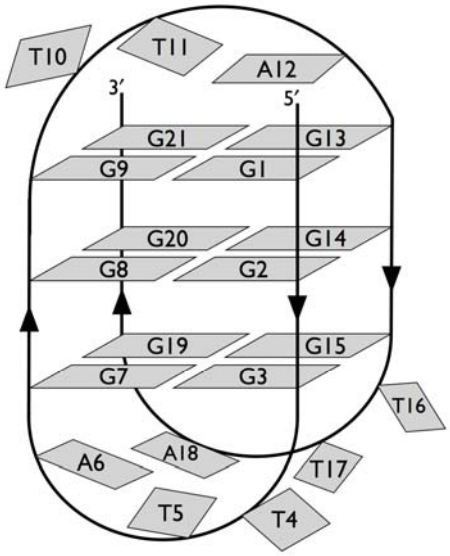
N	4.78850923	-28.07139029	13.05691693
C	6.07538970	-27.93093649	13.56073149
C	5.92251816	-28.52851959	18.39509140
H	6.75485455	-28.42635176	19.13995769
C	6.39388596	-29.07818784	17.07563779
H	7.50902679	-29.04183483	17.02142650
H	6.05114001	-30.12617542	16.90607907
O	4.87545372	-29.34264026	18.94343515
H	4.58448442	-28.88691003	19.74086019

Appendix C: Atomic Energy Differences for the Intramolecular Basket-type Guanine Quadruplex

The following pages present the numbering schemes used for the basket-type quadruplex investigated in Chapter 6 as well as the detailed atomic energy differences, $\Delta E = E_{(\text{atom in quadruplex})} - E_{(\text{atom in ssDNA})}$, in kJ/mol.



Numbering scheme used for the nucleobases and sugar-phosphate backbone.



Numbering scheme used for the nucleobases in the basket-type quadruplex.

	G1				G2				G3			
Atom	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$
NUCLEOBASE												
N1	451.0	430.2	424.9	437.3	240.6	228.6	222.6	220.3	429.4	416.5	370.7	373.5
H1	124.4	114.6	112.6	112.0	146.6	124.4	123.0	107.4	143.7	116.8	99.5	98.2
C2	-19.8	-19.5	-21.0	-17.8	75.0	62.5	62.6	54.3	-138.4	-142.9	-157.1	-160.1
N2	10.6	14.9	16.0	27.1	165.0	173.7	171.6	178.4	-7.8	21.1	-5.4	0.6
H2a	166.2	177.5	178.4	185.4	177.7	174.7	176.3	174.1	93.7	90.3	81.9	78.9
H2b	-29.0	-16.9	-17.4	-6.2	-13.3	5.6	7.0	16.6	6.1	20.7	31.3	30.8
N3	-349.5	-333.8	-330.6	-320.9	-327.2	-305.0	-303.8	-291.5	-419.2	-402.0	-398.0	-398.1
C4	221.1	236.2	237.0	244.4	26.8	39.7	41.2	48.3	76.0	93.3	96.7	97.4
C5	-80.8	-67.3	-69.3	-65.8	-198.9	-199.2	-197.1	-193.3	-93.7	-80.9	-85.2	-86.5
C6	-212.6	-234.5	-233.9	-270.2	-119.4	-154.5	-154.3	-177.4	3.8	-13.6	-32.1	-28.3
O6	325.9	198.4	197.4	128.5	-222.2	-283.5	-287.1	-341.7	-171.2	-372.7	-393.1	-374.5
N7	-801.8	-804.1	-807.9	-813.5	27.6	15.1	14.5	15.1	60.4	70.6	70.7	70.3
C8	33.4	45.5	47.6	59.5	-3.1	10.8	12.5	19.8	-166.5	-153.1	-146.8	-146.8
H8	-8.8	11.7	13.0	28.6	-111.9	-91.2	-90.0	-77.7	63.1	92.6	111.8	114.4
N9	-19.3	-19.3	-18.2	-17.4	632.8	630.8	631.2	632.5	-34.2	-39.2	-41.5	-42.0
BACKBONE												
C1'	-44.0	-50.2	-49.7	-54.7	40.4	24.6	24.4	16.3	18.7	15.5	13.3	14.3
H1'	-26.4	-24.0	-23.5	-21.1	-61.3	-54.5	-54.8	-51.3	47.4	49.0	46.8	46.6
C2'	-31.4	-33.9	-34.8	-36.3	-26.3	-23.6	-23.8	-20.8	-15.0	-17.6	-20.4	-20.8
H2'a	36.4	37.4	37.5	37.9	-113.1	-110.9	-109.9	-105.8	31.6	31.3	32.9	32.8
H2'b	-69.6	-67.2	-66.6	-64.7	-129.1	-116.3	-116.8	-109.9	-25.1	-25.1	-31.0	-32.0
C3'	-241.4	-239.4	-238.7	-234.6	-21.8	-26.2	-25.1	-24.2	-99.4	-95.8	-94.4	-94.6
H3'	19.6	28.4	28.7	37.1	73.6	89.1	90.1	98.5	-14.7	-6.6	0.0	0.0
C4'	49.2	39.6	39.6	31.9	-169.5	-182.7	-183.2	-189.7	51.0	43.1	36.3	36.8
H4'	-8.8	-3.2	-3.2	2.5	72.9	56.2	54.8	44.8	19.8	27.0	33.1	33.3
O4'	228.1	226.3	224.5	223.5	-97.0	-100.2	-100.7	-101.2	16.7	17.7	17.1	17.5
C5'	126.9	125.1	125.0	124.3	-225.9	-221.1	-221.4	-218.0	-84.4	-88.7	-91.8	-91.4
H5'a	26.9	34.0	34.1	41.8	-26.5	-20.0	-19.9	-15.5	29.0	35.1	40.6	40.0
H5'b	-34.0	-34.5	-34.9	-36.0	38.7	39.2	38.5	37.2	-102.3	-99.8	-96.8	-95.8
O5'	212.1	206.3	206.0	202.7	145.1	147.6	147.9	151.8	435.2	425.8	418.8	419.0
P5'	-141.3	-163.0	-162.5	-173.2	-67.4	-96.3	-96.3	-106.1	-52.3	-74.9	-84.0	-83.8
O5'a	44.8	41.3	42.3	40.7	165.5	163.0	163.7	164.3	-21.5	-22.3	-23.8	-23.1
O5'b	-210.0	-212.0	-211.0	-210.2	-135.1	-137.7	-137.3	-136.9	-23.6	-27.3	-28.7	-29.5
O3'	586.5	584.3	583.1	583.3	-90.8	-104.1	-103.7	-106.5	323.4	322.5	319.4	320.0
P3'	-67.4	-96.3	-96.3	-106.1	-52.3	-74.9	-84.0	-83.8	-0.2	-23.4	-44.3	-46.1
O3'a	165.5	163.0	163.7	164.3	-21.5	-22.3	-23.8	-23.1	15.0	12.2	10.0	9.2
O3'b	-135.1	-137.7	-137.3	-136.9	-23.6	-27.3	-28.7	-29.5	-300.1	-301.9	-303.6	-304.1

	T4				T5					A6			
Atom	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$		None	$2Na^+$	$3Na^+$	$4Na^+$
NUCLEOBASE													
N1	-305.7	-305.2	-305.4	-306.3	-425.6	-422.9	-394.1	-389.4	N1	-904.8	-910.1	-898.5	-899.5
C2	131.0	130.4	131.0	129.2	127.8	114.2	105.4	91.5	C2	140.9	134.0	127.5	122.8
O2	-295.3	-313.1	-316.7	-327.0	-432.4	-442.5	-478.6	-485.5	H2	-9.7	-20.2	8.0	0.4
N3	-18.9	-21.0	-22.7	-25.6	-401.8	-432.1	-437.2	-450.1	N3	-25.5	-37.0	-99.0	-116.8
H3	120.0	116.3	115.9	112.8	31.2	9.0	-24.3	-45.4	C4	95.4	95.2	32.8	28.4
C4	-100.7	-101.8	-101.2	-102.8	-70.1	-84.8	-80.6	-94.7	C5	266.7	265.2	263.8	264.4
O4	-48.1	-55.4	-55.1	-54.5	-57.5	-73.9	-48.8	-63.1	C6	524.8	518.2	518.5	517.1
C5	-176.1	-175.1	-175.1	-174.0	-159.9	-160.4	-150.5	-151.2	N6	-142.4	-143.7	-131.3	-126.4
C5a	0.3	-2.7	-4.0	-5.1	-4.4	-6.7	-7.5	-8.7	H6a	38.9	50.9	74.2	85.0
H5a	29.3	31.7	32.3	34.6	-13.1	-0.4	0.5	13.1	H6b	70.1	69.3	89.5	90.4
H5b	-33.2	-31.4	-30.6	-28.8	-16.7	-21.2	-7.0	-11.1	N7	-5.3	-2.9	7.7	15.4
H5c	-83.9	-80.7	-81.0	-77.4	28.6	32.4	33.7	38.0	C8	64.2	63.7	57.5	55.3
C6	-76.6	-82.2	-82.8	-86.8	-170.7	-173.9	-174.1	-176.6	H8	-44.2	-37.9	-23.7	-16.7
H6	-37.8	-36.1	-35.9	-34.3	-21.3	-4.3	4.5	19.9	N9	-223.0	-224.5	-266.4	-275.6
BACKBONE													
C1'	-57.1	-59.8	-59.8	-61.6	102.1	94.6	93.4	86.2		-72.7	-79.4	-110.0	-120.9
H1'	23.3	23.6	21.3	21.4	-42.3	-49.0	-45.6	-51.2		18.8	16.7	24.3	23.9
C2'	-40.6	-42.0	-41.6	-41.4	55.1	53.6	50.0	49.8		-111.2	-111.3	-109.6	-108.0
H2'a	13.0	13.7	15.6	16.3	40.2	45.9	50.0	55.8		0.9	3.2	11.1	14.4
H2'b	-101.8	-96.9	-97.0	-91.9	5.1	3.4	-10.6	-13.2		-14.7	-8.5	1.5	9.5
C3'	-168.5	-170.0	-172.2	-173.4	70.5	70.8	69.5	68.9		-133.7	-132.3	-120.2	-117.4
H3'	29.3	30.5	31.3	32.3	4.0	12.8	14.8	23.2		-17.4	-10.5	3.4	11.6
C4'	90.2	86.9	85.3	84.5	-93.1	-105.2	-105.4	-116.3		-152.2	-161.2	-189.6	-200.9
H4'	62.9	53.2	47.9	39.5	-23.8	-21.0	-17.6	-15.1		3.8	-5.2	5.2	-2.2
O4'	-38.9	-47.1	-48.7	-55.8	-164.1	-165.3	-160.9	-160.5		372.4	361.1	317.7	299.3
C5'	-194.8	-198.3	-198.6	-200.7	45.8	47.1	48.1	50.4		-51.5	-49.7	-44.2	-41.7
H5'a	-18.3	-27.9	-26.0	-35.1	27.9	34.2	35.0	41.5		40.5	46.3	59.4	66.0
H5'b	7.0	11.9	12.8	17.8	-13.7	-11.4	-6.7	-4.4		-9.6	-19.9	-15.4	-23.9
O5'	337.3	331.7	331.0	325.9	-208.7	-212.4	-214.0	-217.2		218.4	216.6	217.6	217.1
P5'	-0.2	-23.4	-44.3	-46.1	-3.6	-26.7	-28.9	-47.0		46.9	22.1	16.5	-5.0
O5'a	15.0	12.2	10.0	9.2	-111.9	-117.5	-117.4	-120.9		-208.7	-210.6	-211.3	-212.0
O5'b	-300.1	-301.9	-303.6	-304.1	147.5	140.6	138.2	135.3		249.5	248.6	246.5	246.7
O3'	10.7	8.1	5.3	4.0	-276.1	-280.8	-281.4	-285.4		-433.1	-437.1	-437.7	-441.9
P3'	-3.6	-26.7	-28.9	-47.0	46.9	22.1	16.5	-5.0		-4.2	-19.5	-48.9	-52.0
O3'a	-111.9	-117.5	-117.4	-120.9	-208.7	-210.6	-211.3	-212.0		220.6	214.7	215.7	216.4
O3'b	147.5	140.6	138.2	135.3	249.5	248.6	246.5	246.7		-392.7	-401.2	-398.9	-398.5

	G7				G8				G9			
Atom	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$
NUCLEOBASE												
N1	455.3	450.2	397.9	394.9	291.0	265.3	256.8	257.3	319.7	299.9	300.5	289.8
H1	136.7	123.2	117.5	112.7	143.0	130.9	132.1	113.1	144.7	132.0	129.3	116.7
C2	-35.5	-39.3	-58.7	-61.5	71.5	62.7	62.9	57.7	-100.8	-107.6	-107.6	-109.0
N2	120.6	136.5	129.7	131.4	115.9	131.4	135.3	135.4	-28.0	-31.6	-32.3	-24.1
H2a	165.1	160.6	172.2	170.3	176.9	187.1	190.3	184.9	122.3	131.1	134.1	131.1
H2b	-25.0	-10.1	4.2	5.4	-15.3	10.1	11.4	17.7	66.3	81.3	82.2	95.7
N3	-153.7	-134.9	-119.7	-119.0	-341.9	-314.6	-313.8	-309.6	-399.9	-382.5	-379.0	-362.8
C4	202.8	217.0	219.4	221.2	26.3	50.2	52.0	53.5	70.5	82.2	83.7	95.0
C5	-101.5	-93.3	-105.3	-103.8	-159.3	-146.7	-145.7	-144.3	-32.1	-25.3	-23.8	-18.4
C6	-208.8	-227.3	-284.8	-284.7	-84.6	-151.3	-155.3	-149.0	34.2	10.5	7.6	-7.0
O6	260.4	161.7	138.0	131.3	-107.6	-265.7	-269.1	-273.0	-222.0	-279.1	-294.6	-401.7
N7	-857.0	-848.2	-858.4	-853.3	-114.7	-123.0	-120.4	-126.2	66.7	58.5	57.0	53.1
C8	72.7	84.9	95.1	98.1	-38.0	-20.4	-18.2	-15.5	-27.5	-16.7	-15.2	-5.3
H8	-21.1	-5.7	12.9	15.1	-100.8	-72.6	-70.8	-63.0	-57.9	-42.2	-40.3	-29.7
N9	-137.0	-138.9	-138.9	-138.1	768.7	771.5	772.3	773.2	-253.4	-258.1	-257.2	-258.3
BACKBONE												
C1'	-36.3	-40.6	-47.8	-47.0	32.7	24.0	25.0	21.5	55.9	52.5	53.9	51.3
H1'	2.4	12.9	25.8	27.3	40.3	45.0	44.7	44.4	-33.4	-26.2	-25.9	-20.6
C2'	-83.3	-85.1	-85.4	-84.8	108.8	103.8	103.7	102.3	-28.2	-30.5	-31.4	-32.8
H2'a	-71.0	-67.3	-59.2	-57.5	-64.0	-64.6	-64.1	-63.2	82.9	70.5	69.2	61.1
H2'b	-25.4	-35.8	-44.3	-45.1	-23.5	-19.1	-19.0	-22.1	-29.2	-21.0	-20.8	-15.4
C3'	-296.4	-291.2	-285.8	-285.3	-31.7	-26.6	-27.5	-26.2	-73.4	-67.7	-67.2	-61.8
H3'	-56.8	-52.9	-47.8	-47.9	-20.7	-6.2	-6.1	-0.1	19.6	25.0	25.5	30.1
C4'	-73.7	-81.4	-90.4	-89.4	-168.1	-183.7	-183.9	-189.0	103.0	94.2	94.1	88.0
H4'	79.9	89.6	99.5	100.5	39.8	50.6	51.1	55.6	8.7	16.8	17.2	24.2
O4'	259.1	258.3	253.4	252.9	166.5	162.8	163.3	166.4	-75.9	-75.5	-78.1	-76.2
C5'	144.3	147.6	140.6	139.1	-172.8	-171.1	-170.5	-168.2	-118.7	-118.2	-118.4	-115.9
H5'a	42.9	34.0	26.0	25.4	-45.0	-48.2	-47.8	-49.0	97.6	85.1	84.4	74.9
H5'b	-74.2	-69.3	-61.1	-61.0	53.0	66.8	66.6	72.3	-58.5	-58.2	-57.3	-56.6
O5'	180.6	168.6	172.0	170.8	142.8	138.7	139.1	141.2	438.2	439.1	438.0	440.4
P5'	-15.2	-30.5	-59.9	-63.0	-27.9	-55.2	-64.6	-67.2	-120.4	-142.1	-141.3	-145.1
O5'a	68.8	62.9	63.8	64.5	112.4	109.1	106.9	106.6	22.6	22.7	23.0	24.3
O5'b	-187.0	-195.4	-193.2	-192.7	-119.0	-121.1	-121.2	-120.9	-62.9	-68.1	-67.6	-69.0
O3'	660.7	662.1	663.9	664.5	-81.3	-82.2	-82.7	-80.9	260.7	259.9	259.0	260.7
P3'	-27.9	-55.2	-64.6	-67.2	-120.4	-142.1	-141.3	-145.1	-158.4	-181.4	-180.8	-192.2
O3'a	112.4	109.1	106.9	106.6	22.6	22.7	23.0	24.3	-131.7	-132.8	-131.5	-131.6
O3'b	-119.0	-121.1	-121.2	-120.9	-62.9	-68.1	-67.6	-69.0	-327.8	-331.0	-330.7	-331.4

	T10				T11					A12			
Atom	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$		None	$2Na^+$	$3Na^+$	$4Na^+$
NUCLEOBASE													
N1	-181.5	-186.9	-183.8	-185.1	-563.2	-570.1	-568.0	-572.2	N1	-772.1	-771.1	-768.0	-752.6
C2	72.0	71.7	72.6	72.6	110.3	111.1	112.7	112.7	C2	136.6	134.4	135.1	140.2
O2	-179.5	-183.5	-186.0	-190.1	-442.5	-441.6	-444.9	-445.5	H2	-32.6	-19.3	-17.6	0.3
N3	-130.7	-129.0	-128.2	-126.3	-406.2	-409.3	-408.5	-411.1	N3	-135.3	-122.6	-125.4	-100.0
H3	-37.9	-34.3	-34.2	-31.1	-13.0	-13.3	-13.3	-12.7	C4	62.9	56.4	54.6	53.9
C4	-168.0	-166.1	-165.0	-163.1	-26.2	-27.4	-25.7	-27.4	C5	210.0	200.0	199.6	211.6
O4	-98.9	-96.6	-100.8	-98.0	197.5	195.8	192.1	192.1	C6	427.2	410.8	410.1	413.3
C5	-36.1	-37.8	-37.1	-37.0	-122.4	-127.2	-127.3	-131.2	N6	125.0	120.0	122.5	130.1
C5a	-0.6	-1.5	-2.3	-2.6	-13.5	-15.7	-17.1	-18.7	H6a	62.8	58.2	57.9	59.2
H5a	7.5	7.5	7.8	7.9	-30.2	-23.4	-23.0	-15.1	H6b	36.3	46.9	47.9	60.5
H5b	83.2	80.5	80.2	78.9	22.0	14.6	13.8	9.2	N7	-23.9	-60.9	-61.0	-234.6
H5c	-86.3	-82.3	-82.5	-78.3	33.6	30.1	30.2	28.3	C8	4.8	-12.2	-16.4	-69.2
C6	-26.2	-33.3	-32.6	-36.2	-127.1	-131.6	-131.5	-135.8	H8	-18.4	-20.4	-17.8	-3.4
H6	102.4	94.2	92.9	87.3	42.4	39.6	38.8	37.3	N9	-285.3	-287.7	-293.1	-280.6
BACKBONE													
C1'	-37.5	-36.8	-35.7	-33.3	61.6	60.9	61.9	64.4		-91.4	-97.1	-98.9	-106.4
H1'	17.1	18.3	18.2	18.2	-33.8	-28.3	-27.7	-21.6		-33.5	-42.0	-41.6	-39.5
C2'	61.2	58.9	58.0	57.8	-21.8	-27.0	-28.2	-33.3		-86.2	-89.0	-90.0	-91.9
H2'a	53.4	53.5	53.8	55.4	40.6	28.5	26.5	8.3		23.9	28.0	28.2	32.5
H2'b	-41.2	-44.8	-45.1	-49.4	15.2	11.4	11.4	5.6		30.2	32.6	32.7	38.5
C3'	-30.1	-38.8	-39.3	-46.1	8.6	4.4	5.1	5.8		-61.8	-59.7	-57.7	-52.0
H3'	28.4	16.1	15.0	6.1	9.9	2.9	1.9	-5.2		-12.9	-3.7	-3.2	7.4
C4'	166.5	169.0	169.7	173.7	-97.6	-99.7	-99.3	-99.5		-102.0	-113.2	-114.2	-125.9
H4'	53.8	53.0	52.7	51.8	-19.7	-12.8	-12.5	-5.1		-6.7	-2.4	-1.5	3.8
O4'	-239.8	-240.1	-242.0	-241.4	36.5	36.9	34.9	39.0		351.2	347.4	345.6	341.9
C5'	-206.1	-208.6	-208.3	-208.1	50.5	43.1	42.0	37.7		-46.8	-48.8	-49.5	-49.2
H5'a	37.7	30.7	30.2	25.3	66.7	65.3	65.1	64.1		35.0	43.1	43.7	53.3
H5'b	13.6	19.1	19.3	24.5	-64.8	-59.6	-59.4	-52.7		-15.0	-16.5	-16.6	-19.4
O5'	317.7	313.6	314.5	313.9	-217.5	-223.9	-223.8	-230.5		212.0	206.3	210.4	207.0
P5'	-158.4	-181.4	-180.8	-192.2	-180.7	-192.3	-191.2	-192.6		-130.2	-154.6	-152.5	-162.9
O5'a	-131.7	-132.8	-131.5	-131.6	-125.7	-135.3	-135.2	-136.0		-242.2	-244.3	-246.0	-248.0
O5'b	-327.8	-331.0	-330.7	-331.4	60.0	49.3	49.6	45.8		227.8	214.4	213.4	211.4
O3'	109.1	107.4	106.2	103.8	-289.4	-289.8	-287.4	-283.5		-456.1	-460.9	-457.2	-459.6
P3'	-180.7	-192.3	-191.2	-192.6	-130.2	-154.6	-152.5	-162.9		-121.8	-143.1	-141.3	-148.6
O3'a	-125.7	-135.3	-135.2	-136.0	-242.2	-244.3	-246.0	-248.0		-102.4	-111.1	-111.5	-114.9
O3'b	60.0	49.3	49.6	45.8	227.8	214.4	213.4	211.4		-41.6	-45.7	-49.1	-49.2

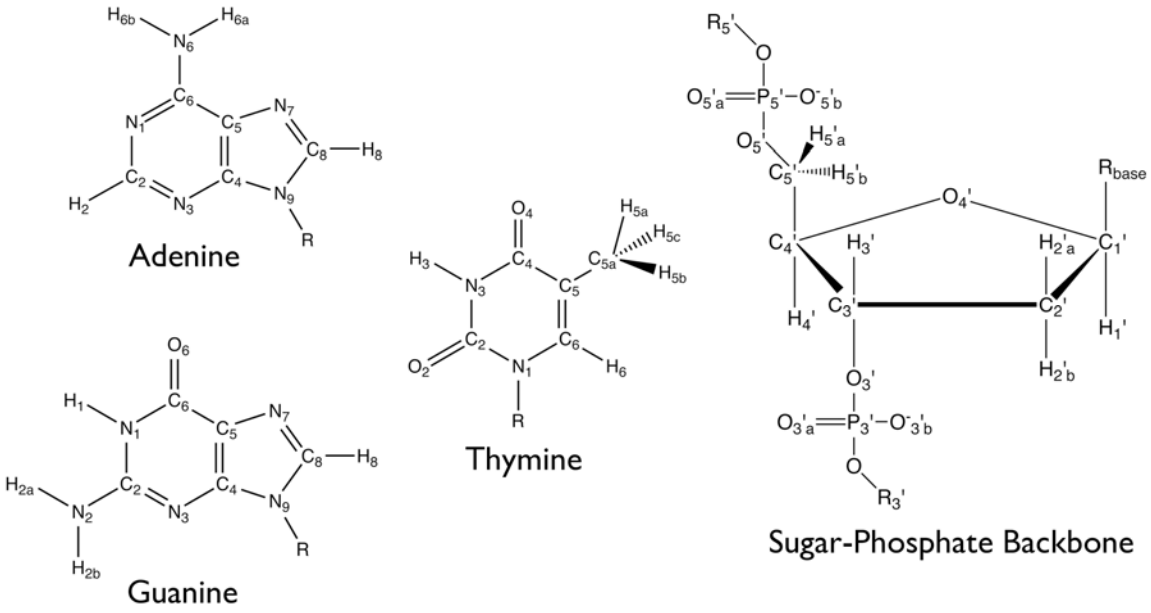
	G13				G14				G15			
Atom	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$
NUCLEOBASE												
N1	508.1	459.3	454.2	415.4	297.9	215.0	206.4	212.9	331.8	315.0	299.5	307.1
H1	145.5	105.4	109.5	73.3	150.4	86.9	84.2	70.2	146.4	120.5	110.1	110.1
C2	-41.9	-47.5	-45.6	-55.8	71.6	49.3	50.4	42.3	-114.6	-120.8	-124.3	-125.2
N2	93.2	91.8	91.2	95.6	91.3	97.0	89.6	98.7	-2.5	7.5	2.2	15.1
H2a	169.5	175.0	178.0	178.0	171.2	159.6	162.8	165.4	161.4	156.0	152.8	152.6
H2b	-21.1	-3.1	-1.8	11.1	0.2	20.7	23.5	33.6	20.5	34.5	43.8	45.1
N3	-297.5	-275.9	-274.2	-258.6	-516.4	-486.4	-483.6	-469.8	-232.6	-209.7	-203.5	-199.6
C4	252.6	268.4	267.7	274.8	27.6	45.3	48.6	56.4	50.2	67.3	69.6	74.3
C5	-60.2	-48.6	-52.1	-47.8	-211.9	-206.5	-202.5	-197.9	-61.0	-53.6	-59.7	-53.8
C6	-205.0	-235.9	-237.2	-270.9	-132.5	-163.2	-164.9	-192.5	42.9	21.7	6.9	1.2
O6	245.8	152.9	152.6	112.6	-174.2	-243.4	-259.6	-314.5	-323.3	-388.8	-392.6	-431.3
N7	-712.4	-712.7	-714.9	-714.9	150.7	148.3	149.1	144.9	222.8	224.7	210.4	216.0
C8	26.2	38.3	39.5	48.8	-14.5	0.8	3.4	12.2	-86.5	-76.3	-69.8	-66.0
H8	36.7	58.3	60.3	78.1	-87.0	-62.5	-60.5	-46.7	-28.9	-12.3	-3.7	0.1
N9	91.0	95.5	94.1	99.6	424.6	423.5	424.7	426.9	7.3	4.2	1.8	1.0
BACKBONE												
C1'	-29.9	-36.1	-35.7	-41.6	36.9	29.5	29.5	24.8	30.4	27.9	25.4	26.3
H1'	-13.4	-12.0	-11.5	-11.6	-31.4	-16.8	-15.6	-6.9	36.7	39.2	40.2	40.6
C2'	-63.6	-65.0	-65.7	-65.8	87.4	83.4	83.1	82.5	75.2	72.9	70.5	70.2
H2'a	-13.1	-11.9	-11.9	-9.8	12.7	-5.9	-6.6	-15.4	30.6	32.3	33.7	34.2
H2'b	-43.3	-40.4	-40.0	-38.6	-67.9	-62.7	-60.8	-56.2	2.3	3.7	1.0	0.2
C3'	-300.6	-299.9	-299.6	-296.9	-18.0	-7.8	-7.6	-1.9	-116.6	-112.8	-110.7	-110.8
H3'	-19.7	-11.6	-11.3	-3.0	39.1	49.0	49.4	55.3	2.8	11.3	17.4	18.1
C4'	-19.3	-28.2	-27.3	-34.1	-139.4	-152.6	-152.4	-158.2	32.3	22.7	15.3	15.3
H4'	1.4	6.1	6.3	10.5	71.9	87.9	88.2	96.7	30.4	34.0	37.1	37.8
O4'	121.5	121.3	118.5	116.5	162.6	162.8	162.3	163.9	-81.0	-83.2	-86.8	-85.0
C5'	138.3	136.0	135.2	136.1	-164.6	-158.4	-158.5	-153.9	-43.3	-43.8	-44.8	-44.4
H5'a	32.8	32.4	32.3	32.4	-8.9	-9.3	-10.2	-11.6	9.0	9.6	10.9	12.1
H5'b	-83.6	-76.6	-76.6	-69.1	39.6	19.2	18.3	8.6	-18.8	-12.6	-7.5	-7.9
O5'	348.2	345.5	349.7	347.7	287.0	287.5	287.6	289.0	482.7	479.2	475.3	475.3
P5'	-110.8	-132.1	-130.3	-137.6	-64.5	-93.3	-93.5	-102.8	-58.3	-85.0	-94.3	-98.1
O5'a	49.4	40.7	40.4	36.9	170.3	167.9	168.3	167.6	24.8	23.0	22.0	21.9
O5'b	-247.3	-251.4	-254.8	-255.0	-133.1	-135.7	-135.0	-135.2	34.3	31.0	29.1	28.9
O3'	633.1	630.3	629.4	629.0	135.5	136.8	136.8	141.4	232.6	230.8	227.4	227.2
P3'	-64.5	-93.3	-93.5	-102.8	-58.3	-85.0	-94.3	-98.1	-58.0	-77.0	-94.1	-95.8
O3'a	170.3	167.9	168.3	167.6	24.8	23.0	22.0	21.9	-90.7	-94.7	-97.3	-97.9
O3'b	-133.1	-135.7	-135.0	-135.2	34.3	31.0	29.1	28.9	-294.2	-296.2	-298.6	-298.5

	T16				T17					A18			
Atom	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$		None	$2Na^+$	$3Na^+$	$4Na^+$
NUCLEOBASE													
N1	-178.9	-180.7	-181.0	-183.9	-431.4	-431.6	-424.3	-425.1	N1	-625.8	-649.2	-638.8	-658.7
C2	70.3	68.9	67.3	65.4	103.3	101.7	96.8	93.8	C2	130.0	115.8	136.9	122.0
O2	-142.7	-149.0	-153.4	-156.7	-605.5	-616.1	-632.4	-638.3	H2	19.7	5.9	-94.4	-96.6
N3	-48.7	-51.8	-53.6	-56.3	-420.1	-428.1	-441.5	-446.6	N3	-85.3	-94.6	-92.1	-99.4
H3	6.7	3.4	2.4	0.2	67.3	64.0	48.4	44.4	C4	38.4	44.2	67.5	72.2
C4	-148.2	-150.5	-150.2	-153.5	-16.5	-21.5	-25.6	-31.9	C5	162.6	156.8	165.8	159.2
O4	-37.4	-50.1	-47.5	-53.9	274.2	260.8	258.0	249.4	C6	368.9	364.6	393.8	388.4
C5	-79.8	-83.1	-81.2	-84.7	-185.5	-189.9	-186.2	-189.1	N6	16.2	30.4	46.5	49.8
C5a	27.4	24.1	23.1	22.4	-131.5	-135.3	-137.2	-139.1	H6a	10.8	16.7	24.8	31.8
H5a	55.5	57.1	57.1	58.9	-23.5	-33.1	-29.2	-35.7	H6b	72.1	67.2	66.5	66.0
H5b	0.5	-7.6	-4.0	-11.7	-1.6	-2.1	-1.8	-2.6	N7	-95.7	-93.5	-81.1	-82.5
H5c	-74.7	-68.6	-68.5	-62.7	49.9	62.1	68.8	79.8	C8	17.0	21.0	26.8	29.6
C6	-75.5	-79.1	-81.1	-83.0	-267.4	-271.2	-271.8	-272.9	H8	-68.2	-55.4	-49.0	-37.6
H6	149.6	153.4	155.0	157.6	13.5	21.8	32.5	39.8	N9	-225.8	-226.2	-222.9	-222.3
BACKBONE													
C1'	-128.6	-129.6	-128.9	-128.9	90.6	89.6	89.0	88.1		-68.0	-70.4	-74.9	-75.8
H1'	35.7	39.1	37.0	40.1	-32.6	-31.3	-33.3	-32.3		66.3	53.5	57.3	46.4
C2'	-7.2	-13.4	-14.4	-18.8	11.8	6.5	4.6	0.6		-35.8	-37.9	-36.6	-38.2
H2'a	58.2	52.8	55.6	50.0	22.5	17.2	19.2	14.7		-9.8	-3.7	-3.2	2.8
H2'b	25.0	17.8	13.1	8.2	15.6	8.7	8.5	2.1		47.2	44.8	51.4	49.0
C3'	-162.6	-165.3	-167.8	-168.9	78.8	77.9	77.5	77.3		14.9	10.6	12.2	9.1
H3'	-37.9	-39.0	-40.5	-41.6	-27.3	-25.6	-22.3	-20.6		-52.1	-44.3	-42.9	-34.8
C4'	104.2	100.2	100.7	98.3	2.8	-2.8	-5.7	-9.0		-26.4	-35.1	-38.8	-46.3
H4'	-16.5	-11.6	-12.7	-7.7	-37.7	-32.1	-31.0	-25.3		-6.5	-3.9	-6.7	-4.5
O4'	-66.0	-65.7	-65.7	-65.4	-92.6	-92.2	-92.3	-90.9		286.0	281.1	274.7	273.2
C5'	-129.9	-134.9	-133.8	-137.4	59.2	54.6	54.3	50.6		3.8	4.2	5.0	6.4
H5'a	2.1	4.2	2.9	5.0	-8.9	-6.3	-4.3	-1.6		7.7	14.1	15.5	22.1
H5'b	4.1	8.0	8.7	12.6	-12.9	-8.5	-7.8	-3.6		-2.2	-1.2	-4.7	-3.9
O5'	326.4	317.8	318.4	310.7	-252.2	-259.7	-262.0	-267.4		293.2	288.2	289.0	285.1
P5'	-58.0	-77.0	-94.1	-95.8	-50.4	-69.2	-71.8	-86.3		66.1	43.2	37.9	19.2
O5'a	-90.7	-94.7	-97.3	-97.9	-255.4	-264.2	-267.6	-272.8		-198.1	-199.8	-200.5	-201.9
O5'b	-294.2	-296.2	-298.6	-298.5	93.7	91.0	90.8	88.6		249.3	247.6	246.1	245.7
O3'	58.7	57.1	54.5	55.5	-186.3	-189.6	-190.1	-191.8		-328.8	-335.6	-336.1	-341.2
P3'	-50.4	-69.2	-71.8	-86.3	66.1	43.2	37.9	19.2		-0.4	-12.5	-27.6	-28.6
O3'a	-255.4	-264.2	-267.6	-272.8	-198.1	-199.8	-200.5	-201.9		237.8	237.2	236.0	236.4
O3'b	93.7	91.0	90.8	88.6	249.3	247.6	246.1	245.7		-545.8	-553.1	-556.8	-558.1

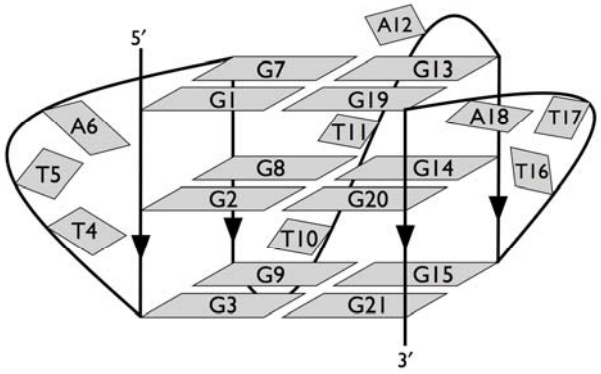
	G19				G20				G21			
Atom	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$	None	$2Na^+$	$3Na^+$	$4Na^+$
NUCLEOBASE												
N1	385.7	390.1	367.2	366.6	369.3	334.8	324.2	320.8	316.9	304.5	304.7	273.7
H1	147.7	125.1	135.5	121.6	148.3	107.1	105.2	108.8	141.7	115.6	115.0	94.3
C2	8.7	-3.4	-4.7	-7.8	58.0	51.9	50.2	42.7	-154.3	-157.8	-157.4	-166.3
N2	150.8	171.9	183.9	194.1	22.5	30.6	31.8	35.6	29.5	30.3	32.7	32.1
H2a	138.5	139.3	145.9	144.6	180.7	179.9	178.9	183.4	142.6	143.4	145.9	139.8
H2b	105.2	129.7	142.6	149.2	-5.2	17.5	19.1	25.5	10.8	27.2	29.2	40.3
N3	-327.1	-307.2	-296.7	-292.1	-414.2	-385.2	-384.7	-379.5	-241.1	-224.3	-223.4	-211.1
C4	199.5	212.9	218.1	221.8	29.6	53.7	58.8	61.2	29.5	41.5	40.4	48.2
C5	-103.2	-93.9	-103.9	-94.9	-180.7	-168.7	-165.7	-167.6	-47.2	-41.4	-41.6	-40.8
C6	-232.9	-261.3	-289.0	-295.9	-73.5	-121.0	-121.2	-140.9	18.8	-3.6	-6.6	-18.9
O6	256.2	138.2	127.0	66.7	-186.4	-360.8	-398.0	-363.0	-265.0	-323.9	-334.4	-383.0
N7	-603.0	-603.4	-624.6	-618.4	55.7	56.9	56.9	51.7	106.3	94.4	95.1	94.2
C8	37.2	49.3	60.5	63.6	-47.1	-29.2	-25.2	-20.5	-71.3	-59.5	-58.2	-51.8
H8	-35.7	-18.7	-6.4	-2.3	-87.2	-56.7	-53.7	-42.7	-71.4	-55.6	-53.1	-40.9
N9	-275.7	-281.6	-283.8	-284.6	582.7	580.1	582.4	581.0	79.8	78.2	76.7	75.7
BACKBONE												
C1'	-81.8	-84.0	-87.5	-87.8	-22.2	-30.3	-28.9	-32.3	60.9	57.5	58.7	56.0
H1'	-32.2	-22.9	-15.0	-12.9	49.6	52.1	51.3	52.6	-65.6	-56.4	-55.4	-47.4
C2'	-92.8	-95.8	-96.7	-96.8	38.3	35.1	35.1	33.9	88.4	87.0	86.4	84.6
H2'a	74.3	63.4	57.4	55.7	-41.7	-38.8	-38.3	-37.3	-3.2	0.3	0.3	1.8
H2'b	-83.7	-81.4	-76.4	-73.8	-76.2	-78.1	-79.3	-82.8	118.2	107.0	105.8	96.6
C3'	-317.6	-312.1	-306.1	-305.0	-32.1	-26.4	-27.6	-23.9	-53.8	-49.2	-47.6	-42.5
H3'	14.0	20.7	27.2	27.9	-24.3	-9.7	-9.7	-3.2	31.1	38.4	39.2	44.6
C4'	-34.3	-41.6	-48.6	-48.9	-153.0	-171.0	-170.5	-177.6	60.7	52.6	52.7	47.2
H4'	-26.1	-17.1	-10.3	-9.8	118.2	130.2	130.7	136.4	-15.8	-9.3	-8.9	-2.7
O4'	252.4	254.7	250.5	250.3	127.1	126.7	128.6	131.3	-147.1	-148.7	-151.3	-151.2
C5'	-30.7	-26.8	-23.3	-23.7	-109.9	-107.7	-106.8	-103.4	-84.0	-82.2	-82.6	-79.6
H5'a	7.1	9.2	4.8	4.2	-14.7	-3.0	-4.0	1.4	13.8	14.0	14.7	14.8
H5'b	45.1	24.7	14.2	10.8	28.0	25.8	26.2	24.8	25.2	15.1	14.4	5.7
O5'	190.7	190.3	190.6	190.5	233.7	231.8	232.6	230.9	538.2	538.3	538.1	540.4
P5'	-11.4	-23.5	-38.6	-39.6	14.4	-14.0	-24.0	-25.7	-106.5	-122.3	-134.1	-142.1
O5'a	85.9	85.3	84.1	84.6	211.6	210.1	209.0	209.0	-53.0	-55.0	-56.0	-56.5
O5'b	-340.1	-347.3	-351.1	-352.4	-157.1	-161.2	-162.8	-165.4	34.2	33.9	33.6	34.5
O3'	576.5	577.3	577.3	577.7	62.2	58.1	57.2	58.9	157.1	160.1	159.6	162.4
P3'	14.4	-14.0	-24.0	-25.7	-106.5	-122.3	-134.1	-142.1	–	–	–	–
O3'a	211.6	210.1	209.0	209.0	-53.0	-55.0	-56.0	-56.5	–	–	–	–
O3'b	-157.1	-161.2	-162.8	-165.4	34.2	33.9	33.6	34.5	–	–	–	–

Appendix D: Atomic Energy Differences for the Intramolecular Propeller-type Guanine Quadruplex

The following pages present the numbering schemes used for the propeller-type quadruplex investigated in Chapter 6 as well as the detailed atomic energy differences, $\Delta E = E_{(\text{atom in quadruplex})} - E_{(\text{atom in ssDNA})}$, in kJ/mol.



Numbering scheme used for the nucleobases and sugar-phosphate backbone.



Numbering scheme used for the nucleobases in the basket-type quadruplex.

Atom	G1	G3	G7	G9	G13	G15	G19	G21
NUCLEOBASE								
N1	669.1	515.0	562.5	594.4	589.5	559.3	543.1	610.8
H1	119.4	106.8	111.9	119.8	97.8	121.6	116.0	132.7
C2	-42.1	-133.9	-58.5	-113.3	-88.4	-144.3	-90.9	-84.6
N2	356.2	169.9	367.0	226.9	313.1	237.2	288.0	266.3
H2a	190.5	168.2	188.4	153.7	177.0	143.5	176.0	161.0
H2b	6.0	16.1	2.9	20.5	2.1	10.5	-4.1	49.4
N3	-374.9	-187.8	-253.6	-337.6	-116.4	-270.4	-191.5	-256.5
C4	335.4	120.7	362.4	187.8	270.7	198.6	332.8	152.6
C5	72.1	37.0	33.6	28.5	40.2	56.9	11.3	32.9
C6	-286.3	4.0	-261.4	-34.8	-294.4	6.7	-268.1	-36.4
O6	295.0	-149.0	244.6	-219.0	328.1	-161.1	255.5	-150.0
N7	-633.9	185.1	-721.5	57.5	-718.2	106.5	-672.3	151.7
C8	94.3	-7.9	109.9	-2.3	95.6	-0.3	106.3	-11.2
H8	23.4	33.6	27.1	42.5	21.7	32.5	29.2	10.6
N9	-109.9	112.5	-50.2	139.3	-24.1	174.7	-34.3	138.4
BACKBONE								
C1'	-38.2	69.0	-71.4	55.0	-87.2	72.6	-62.8	82.6
H1'	-41.4	21.9	-13.7	12.6	-6.8	22.7	-17.9	18.2
C2'	-170.9	86.7	-141.0	78.0	-113.5	94.2	-129.9	61.9
H2'a	2.2	23.6	-30.1	25.3	-24.6	43.5	20.3	9.2
H2'b	-55.1	28.3	-40.3	24.1	-35.2	7.8	-72.9	2.9
C3'	-391.3	-37.8	-331.9	-48.6	-312.1	-26.8	-306.6	-81.6
H3'	46.5	39.5	36.3	45.3	47.4	22.8	50.2	-9.3
C4'	-121.3	9.8	-144.2	18.5	-105.1	29.6	-105.9	13.8
H4'	10.5	21.5	58.3	22.8	38.8	17.8	26.2	33.6
O4'	276.0	-18.9	292.7	-19.3	317.2	20.0	299.4	39.1
C5'	150.6	-127.4	91.6	-131.4	73.3	-119.2	104.9	-43.9
H5'a	26.8	-39.8	-1.4	8.6	-30.9	-3.3	-24.8	41.7
H5'b	-41.5	-37.8	-50.1	-4.4	17.3	2.0	21.1	-7.2
O5'	380.1	288.6	97.7	393.3	125.9	379.7	141.7	383.1
P5'	–	40.0	-172.8	29.5	-129.6	58.0	-116.9	45.2
O5'a	–	106.4	127.7	145.8	124.5	147.8	119.1	99.6
O5'b	–	94.2	-247.6	59.7	-202.1	88.1	-167.9	99.1
O3'	619.7	305.6	495.2	338.3	529.5	270.0	552.8	977.4
P3'	-65.7	22.1	-71.7	28.4	-67.2	19.4	-58.1	–
O3'a	259.3	86.4	251.5	126.7	270.1	61.0	221.5	–
O3'b	-99.0	-206.3	-111.7	-183.7	-127.8	-163.9	-98.8	–