Relationship between nucleic acid sequence, structure and function in terms of stabilizing interactions

Petrina R.N. Kamya

A Thesis

in

The Department of

Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy in Chemistry
at Concordia University
Montreal, Quebec, Canada,

July 2009

Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-71159-0
Our file Notre référence
ISBN: 978-0-494-71159-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Relationship between nucleic acid sequence, structure and function in terms of stabilizing interactions

Petrina R.N. Kamya

The relationship between nucleic acid (NA) sequence, structure and function is intricately connected to the stabilizing interactions (primarily hydrogen bonding and π-stacking) that occur between the monomeric subunits that constitute NAs. Therefore, detailed insights into the nature of the stabilizing interactions would permit the full exploitation of the structure-function relationship in NAs. A complete understanding of the role of the stabilizing interactions in NAs involves the fulfillment of two requirements: 1) The ability to determine the electronic structure of the monomeric units (in terms of the electron density distribution as an observable) that make up the fundamental structure of NAs, which is possible through the use of quantum chemical calculations, and 2) The ability to characterize the electronic structure of these monomeric units in the context of realistic NA structures. Ideally, such molecular structures are determined experimentally. These two ideas are combined into a methodology that has been designed, tested and validated in the work presented here. The proof of concept culminates in the ability of the methodology to exploit the structure-function relationship in NAs through procurement of full stabilization profiles of host NA and guest (small molecule inhibitors to the function of the NA) complexes, where the potential inhibitors were designed on the basis of the stabilization profiles of the host and natural ligand complexes.

Acknowledgements

I would like to thank my supervisor Dr. Heidi M. Muchall, for her patience guidance and support.

I would also like to thank my committee members, Dr. Giles Peslherbe and Dr. Paul Joyce for their valuable inputs during the course of this PhD.

Special thanks to past and current group members, especially Elena V. Ivanova and Lei Zhang for their invaluable support over the years, to CERMM system administrators (past and present) thank you for keeping the cluster a-float.

I would like to acknowledge financial support from NSERC. Concordia University for the doctoral thesis completion award, and the international tuition fee remission award. The Quebec Black Medical Association for the QBMA scholarships. The Scripps Research Institute for the workshop attendance fellowship, and the RNA society for the travel award.

Calculations were performed at the Centre for Research in Molecular Modeling (CERMM).

And of course, Tom, words can not even begin to describe my gratitude.

For my Parents.

Table of Contents

List of Figures

List of Tables

List of abbreviations and symbols

2APIN	2-amino-3*H*-pyrimido[4,5-b]indol-4(9*H*)-one
2APO	2-amino-3*H*-pyrimido-4a*H*-pyrrolo[3,2-d]pyrimidine-4(9*H*)-one
A	Adenine
A3_U15_G4_C14	3'-...U15C14...-5' 5'-....A3G4....-3'
AIM	Atoms In Molecules
AT/U	AT or AU base pair
B3LYP	Becke's three parameter hybrid exchange functional with Lee, Yang and Parr's correlation functional
BCP	Bond critical point
BH&H	Becke's half and half functional
C	Cytosine
CCP	Cage critical point
CP	Critical point
DFT	Density functional theory
DNA	Deoxyribonucleic acid
FF	Force Field
G	Guanine
H-bond	Hydrogen bond
HOMO	Highest occupied molecular orbital
HX	Hypoxanthine
IP_v	Vertical ionization potential
MD	Molecular Dynamics
MM	Molecular Mechanics
MP2	Møller-Plesset perturbation theory to the second order
NA	Nucleic acid
NDB	Nucleic acid database
NMR	Nuclear magnetic resonance
PBE0	Parameter-free hybrid density functional theory method developed by Perdew, Burke and Ernzerhof
PDB	Protein data bank

PDB-ID	Protein data bank identification number
QC	Quantum chemical calculations
QM	Quantum mechanics
QTAIM	Quantum Theory of Atoms In Molecules
RCP	Ring critical point
RCSB	Research collaboratory for structural bioinformatics
RNA	Ribonucleic acid
T	Thymine
TD-B3LYP	Time-Dependent Becke's three parameter hybrid exchange functional with Lee, Yang and Par's correlation functional
TD-BH&H	Time-Dependent Becke's Half and Half functional
TD-DFT	Time-Dependent Density Functional Theory
TD-PBE0	Time-Dependent- parameter-free Perdew-Burke-Ernzerhof functional
U	Uracil
UV-Vis	Ultra-violet Visible Spectroscopy
X	Xanthine
XE1	Xanthine Enol tautomer with OH group facing N1
XE3	Xanthine Enol tautomer with OH group facing N3
XK	Xanthine keto tautomer
Ω	Atomic basin
$\nabla^2\varrho$	Laplacian, the second derivative of the electron density
$\nabla\varrho$	First derivative of the electron density
β	Stereochemistry of a natural nucleoside in which the base is in the same side as the C5' of the sugar.
λ_1, λ_2 and λ_3	Eigen values of the Hessian matrix of ϱ
λ_{max}	Wavelength of maximum absorbance
ν	Torsion angles around a five membered ring
ϱ	electronic density
ϱ_{HB}	electronic density at a hydrogen bond critical point
ϱ_{π}	electronic density at a stacking bond critical point
$\Sigma\varrho_{HB}$	sum of the electronic density at all hydrogen bond critical points in a complex

$\Sigma\varrho_{\pi}$	sum of the electronic density at all stacking bond critical points in a complex

Chapter 1

General Introduction

1.1 Fundamentals of nucleic acid structure

All life starts with the assembly of a surprisingly basic list of ingredients to create a genome composed almost entirely of nucleic acids (NAs). It has been established that the unique biological function of any NA is inextricably linked to its sequence but more importantly its three dimensional structure, but the nature of their relationship is only beginning to be understood.

1.1.1 Deoxyribonucleic acid - DNA

The genomes of most living organisms are composed of deoxyribonucleic acid (DNA) and the basic ingredients of DNA are nucleosides. These consist of a deoxyribose sugar attached to a purine or pyrimidine heterocyclic organic base via a β–glycosidic bond, and as a unit they establish an important structural feature of DNA molecules. The precise conformation of the deoxyribose sugar is defined by five endocyclic torsion angles labeled ν_0 to ν_4 as shown in Figure 1.1.[1] In theory there is a continuum of interconvertible puckers separated by energy barriers. However, in practice the presence of the base results in the C2'-endo type of pucker in more than 60% of the deoxyribose sugars in DNA.[2]

Phosphodiester linkages string each base-sugar unit together in sequences that are defined by the adenine (A), guanine (G), cytosine (C) and thymine (T) bases. These sequences form the basic unit of heredity - genes.

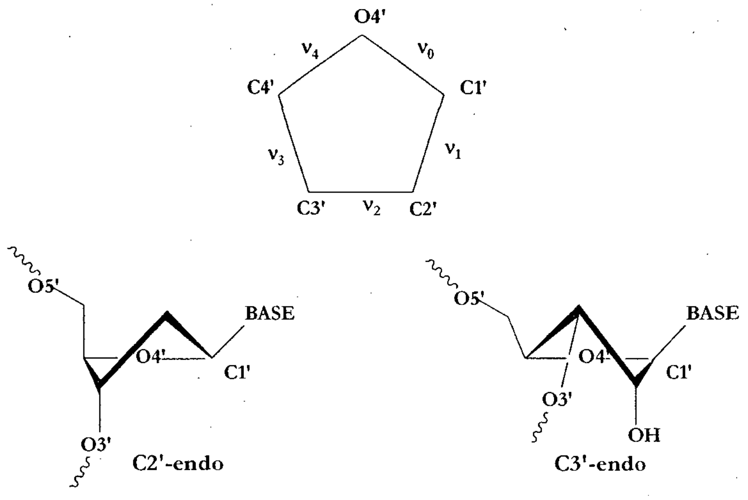


Figure 1.1 The two most common conformations for the sugars of NAs.

DNA can be described as having three structural levels: primary, secondary and tertiary.

The primary structure consists of the nucleotide sequence often written from the 5' to the 3' end as shown in Figure 1.2. It is important for the storage and transmission of genetic information as well as in the formation of specific structural motifs.

Tertiary structure in DNAs involves the interaction of long stretches of DNA with proteins such as histones and nucleosomes as is the case in the packing of eukaryotic DNA into chromatin.

The secondary structure is the most important structural level for DNA function and is any stable recurrent motif produced by the formation of hydrogen bonding (H-bonding) and /or π–stacking between DNA bases or base pairs. Shown on the left-hand side of Figure 1.3 is an example of the most commonly recognized secondary structure: the B-form double helix (B-DNA) discovered by Watson and Crick.[3]

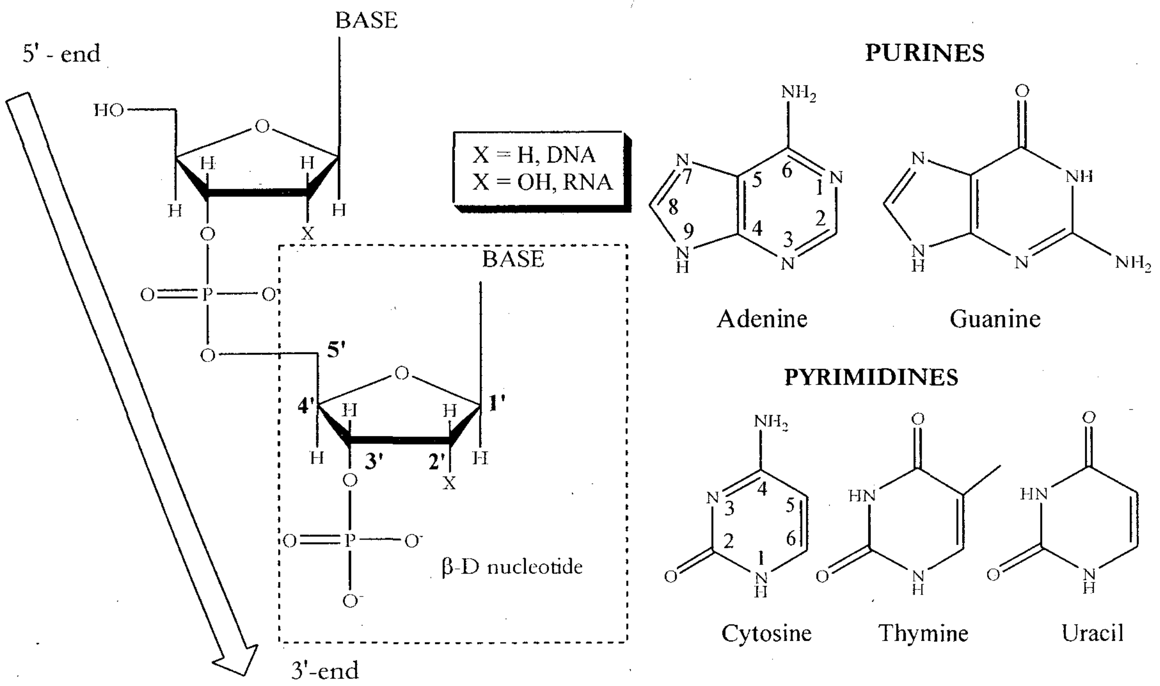


Figure 1.2 A generic nucleotide repeat shown in the 5' to 3' direction as part of a polynucleotide chain. The dashed square outlines one β–D nucleotide unit where X at the 2'-carbon of the sugar is an –OH group in RNA and the sugar is a ribonucleoside and in DNA the X is an H and the sugar is a deoxyribonucleoside. The numbers of the sugar atoms are primed in order to distinguish them from the numbering for the bases shown on the right hand side.

B-DNA is formed from two antiparallel polynucleotide chains that twist along a helical axis creating a right-handed double helix. In aqueous environments B-DNA helices form cooperatively, establishing π–stacking interactions between successive bases on a polynucleotide chain, and hydrogen bonds (H-bonds) between complimentary bases to form Watson-Crick or Canonical base pairs (as shown in Figure 1.6) from two separate chains. A significant feature of the double helix are the grooves that wind around the helical axis. The placement of successive sugars and phosphate groups on the same side of the individual base pairs along a sequence creates a relatively narrow groove known as the minor groove. The minor groove is highly charged and therefore often associated with divalent metal cations for additional stability.[4] While the sugars are on one side of the successive base pairs, on the other side the functional groups belonging to the bases are more exposed, as they are less bulky than the sugar phosphate backbone this side is relatively wider and is called the major groove. Figure 1.3.

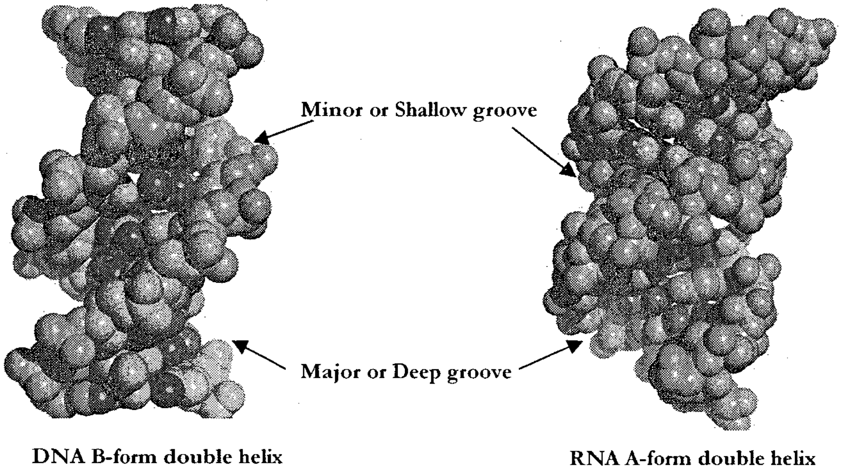


Figure 1.3 B- and A-form nucleic acid double helices showing the locations and morphological differences between the Minor (sugar-phosphate backbone is more exposed) and Major (functional groups are more exposed) grooves in B-DNA and the Shallow and Deep grooves in A-RNA.

Although B-DNA is both the most commonly encountered helical form of DNA and the archetype of how biological function follows from biomolecular structure, it is not the only double-stranded structure or secondary structure available to DNAs. DNA is extremely polymorphic and in the right conditions four additional double helical structures have been identified, these include the A, C, D and Z-form DNAs.[4]

Triplexes and Quadruplexes are higher order non-standard helical forms involving sequences from single or multiple DNA strands.[5] Triplexes form in regions of long polypurine and polypyrimidine hybridized stretches that permit the hybridization of a third strand, giving rise to a 1:1:1 three-stranded polynucleotide complex made up of triply H-bonded bases.[6] Triplex formation in the DNA of certain species has been proposed to regulate the transcription of specific genes. This discovery sparked wide spread interest in the development of "antigene" or artificial gene regulation methods that use purely synthetic molecules to target specific genes.[7] Long runs of guanine bases in nucleic acid sequences have been known to form quadruplexes, which are composed of a series of guanine tetrads (G-tetrad) stacked over each other. Quadruplexes like triplexes are also believed to be important in several biological processes.[8]

DNA is often found in single stranded forms during replication, transcription, recombination and repair where it can fold back on itself, creating unimolecular structures. Figure 1.4 illustrates typical single stranded DNA structural motifs including bulges, hairpin loops, internal loops, and junctions. Examples of mismatched or non-canonical base pairs are shown in Figure 1.5. Non-standard DNA single strand structures are associated with aberrant as well as normal functions. For example, bulges, hairpin and internal loops are often signs that mistakes have been made during replication or damage and are therefore known to serve as signals for DNA repair proteins or functional proteins.[9] Junctions on the other hand are believed to be the central intermediate in the process of homologous genetic recombination which involves the crossing-over of strands from two DNA sequences.[10]

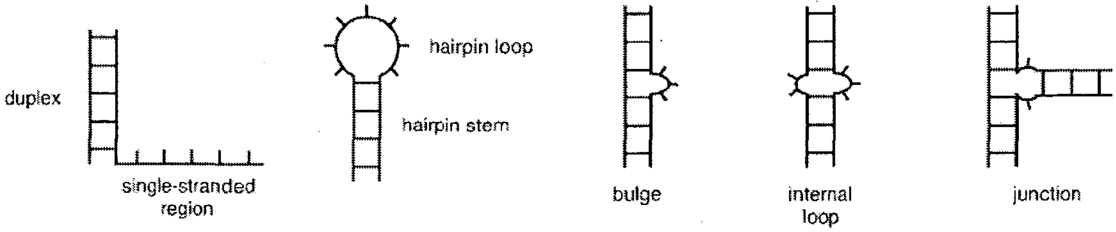


Figure 1.4 Common NAs secondary structural elements.

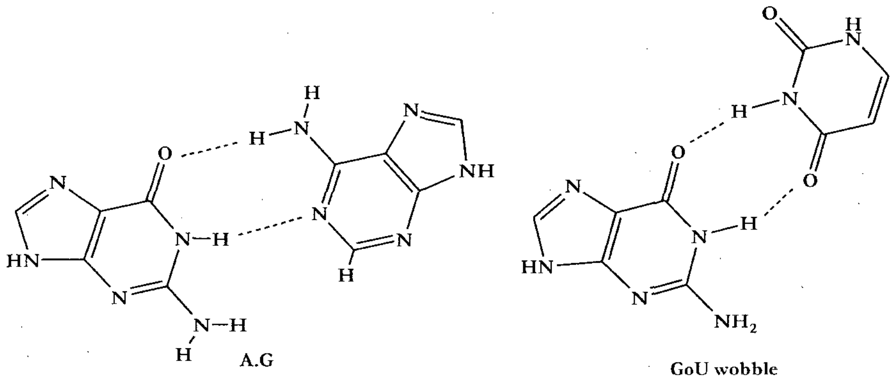


Figure 1.5 Examples of non-canonical or mismatched base pairs in NAs.

1.1.2 Ribonucleic acid – RNA

RNA can form a genome and is only chemically distinguishable from DNA by one sugar and base; (illustrated in Figure 1.2.) Uracil (U) replaces Thymine (T) and a ribose sugar with a 2'-OH group (found 65-85% of the time it is found in the C3'-endo conformation see Figure 1.1) replaces a deoxyribose sugar.[11]

As with DNA, RNA structure can be described by dividing it into three fundamental levels of organization: the primary, secondary and tertiary structure.

The primary structure of an RNA molecule is a sequence of nucleotides that is vital to fulfilling the instructions of a gene. It is also involved in the formation of higher order structural components.[11]

The secondary structure comprises both helical and non-helical structures (also known as secondary structural motifs or elements) that are stabilized by H-bonding and π–stacking interactions between the bases and base pairs.[11] It is widely considered to be central to RNA function.[12,13]

Approximately 60% of RNA secondary structure is comprised of antiparallel double helical RNA that can be formed by both single and double stranded RNA sequences.[11] However, in stark contrast to the polymorphism observed in DNA, only one major polymorph of an RNA double helix has been observed, the A-form. A-RNA is an eleven fold helix with a narrow and deep major groove and a shallow, wide minor groove. These can be referred to as the deep and shallow grooves, respectively, as shown in Figure 1.2. The other 40% of RNA secondary structure consists of stable non-canonical base pairs that are found within the context of an A-form double helix, such as the G°U "wobble" base pair shown in Figure 1.5, and non-helical structures such as stem loops, internal and hairpin loops, bulges, and junctions (Figure 1.4) are defined by Leontis and Westhof as "recurrent and ordered arrays of non-Watson-Crick base pairs."[14] All of these secondary structural motifs have been shown to be extremely important to the specific functions of RNAs, they also form tertiary structural elements that maintain the three dimensional folds found in the crystal structures of all complex RNAs.[15]

Tertiary structural elements are repetitive three-dimensional patterns[16] resulting from stabilizing interactions (base-stacking, base-pairing and base-phosphate interactions) between distinct secondary structural elements.[17] They too are essential to RNA function and to the global architecture of folded RNA molecules.[12,13,17]

A comprehensive analysis of fifty-four high resolution ($R \leq 3$Å) RNA crystal structures revealed a total of seven major groups of tertiary structural elements. [16] The seven groups consist of: the A-minor motif,[18] coaxial stacked helices,[19] ribose zippers,[20] pseudoknots,[21] loop-loop receptors,[22] t-RNA D-loop T-loop,[23] and kissing hairpins.

1.1.3 Non-covalent interactions

Non-covalent interactions govern the formation of secondary and tertiary NA structures and maintain their structural integrity. They are therefore fundamental to NA function. The most important non-covalent interactions are H-bonding and π–stacking interactions, both of which have electrostatic, induction, charge-transfer and dispersion terms that describe the nature of the stabilizing energy.[24]

1.1.3.1 π-stacking interactions

The stabilization energy of π–stacking interactions stems predominantly from dispersion type forces produced by favorable instantaneous multipole/induced multipole charge fluctuations between interacting molecules.[25-27] Since the dispersion energy has a large R^{-6} dependence on inter-molecule distance,[28] a reasonable definition for the π–π stacking effect is "special non local electron correlations between the electrons in the two fragments at small interplane distances."[25]

The formation of the many secondary structural forms of NAs in aqueous solutions is driven by π–stacking between the aromatic heterocyclic bases rather than by H-bonding.[29,30] This implies that the structure of stacked bases and base pairs in NAs is largely governed by the energetics of base stacking.[6]

The relative positions of the π-stacked bases and base pairs in NAs are defined by translation and rotation operations. Translation can be described by the displacement of the bases in the x,y and z directions of a Cartesian coordinate system relative to the helical axis, while the rotations are described by angles made by the bases and base pairs relative to the orientation of the fictitious base pair. The helical parameters that describe the relative orientation of the base pairs in a duplex are illustrated in Appendix A.[6]

1.1.3.2 Hydrogen bonding interactions

H-bonds are vital to formation and maintenance of secondary and tertiary structural motifs in NAs. They are stabilized by electrostatic, induction (charge-transfer) and dispersion energy terms. As the dipole-charge and dipole-dipole contributions to the electrostatic term give H-bonds their all-important directionality, the electrostatic term is the most important contributor to the stabilization energy of H-bonding interactions.[28]

Watson-Crick base pairs (also known as complementary, and canonical base pairs) shown in Figure 1.6 are the result of H-bonding between H-bond donor and acceptor groups found on the Watson-Crick edges of the purine and pyrimidine bases. Their geometric beauty is owed to their isosteric nature which is favored in the context of a double helix that tends to prefer to be uniform irrespective of the sequence.[6] The controversial third, C-H···O H-bond between A•T and A•U base pairs as shown in Figure 1.6 is recognized in this work whenever the interaction has been located.

Any other edge-to-edge combination of the bases that occur within the context of NA structures is known as a non-canonical or mis-matched base pair. A•G and G∘U "wobble" non-canonical base pairs are two such examples and are shown in Figure 1.5.

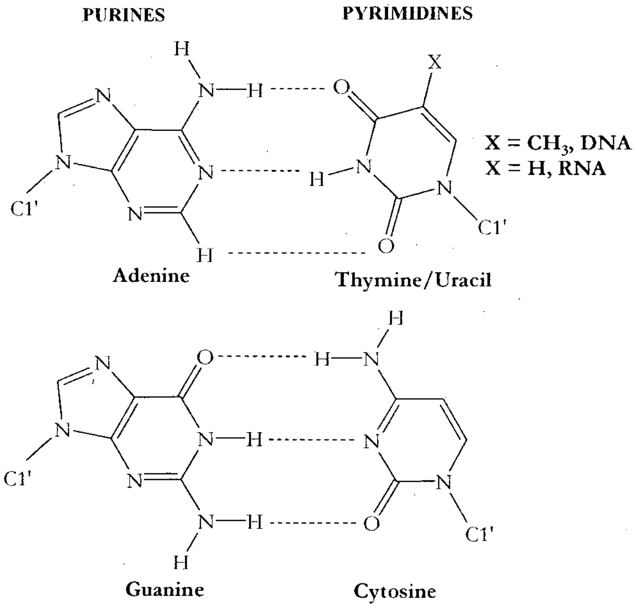


Figure 1.6 Hydrogen bonding interactions between Watson-Crick base pairs.

Base triples are often encountered in junctions and in the formation of triplexes. These often involve the interaction of a Watson-Crick base pair with a third base H-bonding to the Hoogsteen edge of one of the bases in the Watson-Crick pair as shown in Figure 1.7.

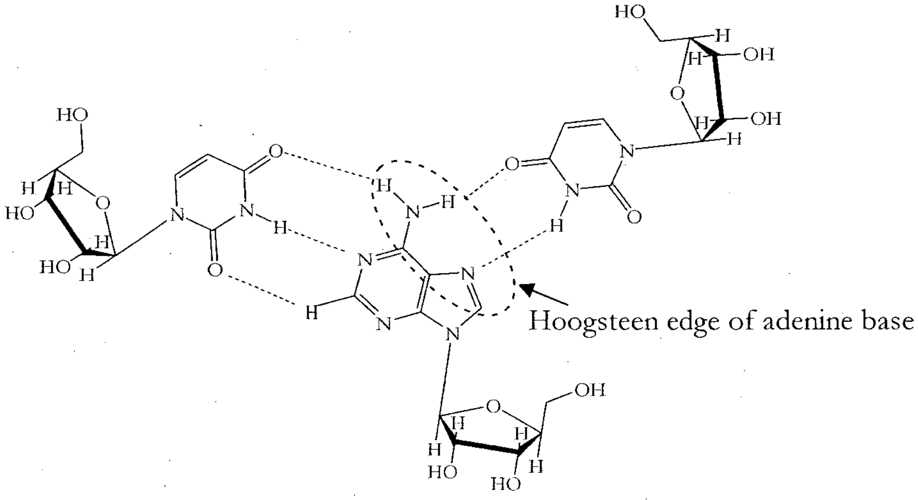


Figure 1.7 Base triple involving (A•U)U.

NAs are assembled from simple ingredients but their structures are complex. The reason for this is that the functional roles of NAs are predominantly governed by their secondary and tertiary structures not the individual bases and base pairs.[12,13,17] For example, of the human genome's 3 billion letters, only 2% of genes code for proteins, the other 98% are now believed to be important in forming biologically functional shapes.[31] This is particularly relevant to the development of novel therapeutic applications that inhibit NA function by targeting NA structure. Such drug design protocols call for the ability to unequivocally characterize NA structure including the nature of the stabilizing interactions in the environment. The identification and quantification of non-covalent interactions is fundamental to this process. However, this poses serious challenges to experimentalists and theoreticians.[24]

Chapter 2

The Problem

2.1 Relationship between NA structure and stabilizing interactions in the literature

2.1.1 Experimental methods

Most details regarding the relationship between NA structure and the stabilizing interactions have been determined using sophisticated experimental approaches. Only X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR) will be briefly described here as they are most relevant to the work covered in this thesis.

2.1.1.1 X-ray crystallography

The use of X-rays to determine molecular structure is based on the similarity between the dimensions of internuclear bonds and the wavelength of X-rays, which are both approximately 1.5Å. [32] When exposed to X-ray radiation, the electron density surrounding each atom in an ordered molecule scatters the incident radiation generating a diffraction pattern. An analysis of these patterns generates a map depicting the electron density distribution of the molecule from which the molecular structure can be deduced.[32] X-ray crystallography has provided most of the highly detailed information regarding NA structure to date.[11] The quality of a crystal structure is often assessed based on its resolution where a resolution greater than 3Å is considered poor.[33]

The information directly obtained from traditional X-ray crystallography is the molecular geometry. Everything else is implied. The presence of H-bonding is often inferred from comparing the relative positions of the nuclei, with an emphasis on distances and angles, to predetermined ideals. This technique can be problematic where weak bonding interactions are concerned, as the ideal geometric situations are more difficult to define. The strength of H-bonds are also gauged on the basis of the deduced geometry of the interaction.[27] The presence and strength of stacking interactions are often determined through an observation of the degree of overlap of the aromatic rings, another subjective rather than objective technique that can be misleading.[25,27,34,35]

2.1.1.2 NMR spectroscopy

NMR is a form of spectroscopy that depends on the detection of a change in spin states of nuclei within a molecule in the presence of a magnetic field. NMR is a powerful tool for the study of the structure and dynamics of NAs in solution, which is an obvious advantage over X-ray crystallography where structure determination is in the solid-state for crystal structures.

The procedure for structure determination starts with the systematic assignment of nuclei resonances. This involves the identification of nuclei that are involved in secondary and tertiary structural formation through H-bonding and π-stacking. Changes to the chemical shifts of the exchangeable imino protons of G and U, and to a lesser extent the amino protons of A, G and C, serve as indicators of H-bonded versus non H-bonded bases. Large coupling constants (J) between the H-bond donor and acceptor groups J= 6-7Hz and smaller J=2-4Hz between the imino hydrogen and an N^{15} labeled acceptor can also serve as signals for H-bonding.[36,37] Changes in the chemical shifts of the imino protons due to ring current shifts can serve as indicators of strength of π–stacking interactions.[38] These are challenging tasks as the signals are often broadened and overlapped which is why a combination of different NMR techniques that are specialized to identify the resonances due to specific nuclei or atom-atom interactions is often required.[39] Overlapping and broadening of chemical shifts are in part a consequence of the large size of NA structures, which limits the applicability of NMR spectroscopy to large nucleotides.[40]

Final structure determination of a NA often involves the use of structural restraints and parameters obtained from a combined analysis of the different NMR spectra in a molecular modeling protocol, where a set of structures that represent an ensemble of conformations of the NA are obtained.[40] A good judge of the quality of an NMR-determined molecular structure is the number of restraints used in the determination of its structure which should be high but not excessively so.[13]

2.1.2 Theoretical methods

Theoretical methods can be used in conjunction with experimental data for the elucidation of experimentally determined geometric parameters as is the case with most NMR determined molecular structures or to improve geometries determined by X-ray crystallographic methods,[40] but this is not always the case. The use of theoretical methods for the prediction of NA structure from sequence as well as for the prediction and understanding of experimentally observable properties has become more main stream with recent improvements in (super)computing technology.[40]

2.1.2.1 Force Field Methods

Force field (FF) based methods use analytical potential energy functions that are defined by the laws of classical physics and a set of parameters that are based entirely or partially on empirical data or quantum mechanical calculations, to describe the physical properties of a system.[41] FF methods are often used to locate local minima on the potential energy surface of a NA or to simulate the dynamics of NAs in solvent.[42]

Opinions regarding the potential applications of FF methods tend to be divided. Some view them as the only possible solution to the application of computational methods towards the study of large biological systems[40] and others as "simplistic" methods that "rely on large-scale compensation of errors," where over-reliance is cautioned against.[27] Never-the-less, the successful application of FF methods towards the study of NA structure is undeniable. Most recent formulations of available force fields are capable of providing a good description of H-bonding and π–stacking interactions from the non-bonded term in the definition of the potential energy of the system, which includes exchange repulsion, dispersion attraction and the electrostatic energy due to atomic charges that allows for reasonably accurate models of the three dimensional structure of oligonucleotides.[25,27,43] However, force field methods can not provide any information regarding the electronic structure of the molecules, which means that the individual contributions from the H-bonding and π–stacking interactions for specific sequences can not be determined either. The timescale of most biological events is longer than a few hundred nano seconds, and even today, simply increasing the simulation time of most molecular dynamic simulations often exposes force field deficiencies, which tend to accumulate over time.[42]

2.1.2.2 Quantum Chemical Calculations

Quantum chemical (QC) calculations are used to describe the fundamental properties of matter based on the theory of quantum mechanics (QM). Size is a major set back in the applicability of QC calculations, therefore most studies on the stabilizing interactions between nucleic acid bases and base pairs involve the use of small model systems where a medium sized cluster is defined as a base pair.[44] Moreover, the geometries of these systems are usually optimized in the gas-phase, often in the absence of any other base pairs or solvent. These types of studies are very important for determining the energetics of π–stacked and H-bonded bases and base pairs in the absence of any external geometric constraints; however this is not a very realistic way of simulating the natural environment of bases in NAs. In addition, an accurate description of the interaction energies often requires the use of computationally demanding highly correlated methods, which places an additional restriction on the size of the systems.[25-27]

It has been shown, though, that QC calculations can produce information regarding the electronic structure that no other experimental or computational technique can provide. [44] Some major achievements in the study of nucleic acids have been obtained using QC calculations; these include the elucidation of the true nature of π-π stacking interactions,[25,26,45] the realization that the N2 and N6 amino groups in G, A and C are not planar,[46] as well as important findings regarding the nature of cation binding to NAs.[47]

2.2 Summary

It is generally agreed that experimental methods provide the most accurate and realistic representations of the three-dimensional structures of nucleic acids. Many force field methods use empirical parameters obtained from X-ray crystal structures, and a QC calculation may begin with a starting geometry obtained from X-ray crystal structures. Databases of DNA and RNA secondary structural motifs are based on X-ray crystal and NMR determined structures. However, experimental data are limited in the amount of detail they can provide regarding the unequivocal identification, characterization and quantification of the weak stabilizing interactions between base pairs. Theoretical methods based on molecular mechanics can now be used to determine the three dimensional structure of nucleic acids, and good progress is being made in the implementation of more reliable methods, but they can not compare to accuracy and precision of QC calculations when it comes to the determination of the electronic structure of the base pairs. QM based methods are the only techniques that can be used to describe the true nature of the stabilizing interactions, as well as quantify the H-bonding and π–stacking interactions separately. However, these studies are yet to be conducted in an environment that is a more realistic representation of the natural environment of the base pairs in a NA structure.

Chapter 3

Objectives and Outline

3.1 Objectives and Outline

Chapter 1 highlighted the inherent dependence of the function of nucleic acids on the sequence of nucleotides but more importantly on the structures and the stabilizing interactions. Chapter 2 introduced the problem associated with current experimental and computational methods that are typically used to gain an understanding of the relationship between the sequence, structure and stabilizing interactions between NA base pairs.

The main objective of this thesis is to design a methodology that can unequivocally identify and quantify the H-bonding and π–stacking interactions between the bases and base pairs in realistic model systems of NAs. To do this, base pair geometries are isolated from experimentally determined molecular structures, by the deletion of water molecules, divalent metal ions (if there are any) as well as the surrounding duplex and the replacement of the sugar-phosphate backbone with a hydrogen. The geometries of the isolated base pairs are unchanged while a 'wavefunction' is obtained and analyzed within the framework of the quantum theory of atoms in molecules (QTAIM)[48] using AIM 2000.[48,49] An introduction to the quantum theory of atoms in molecules in the context which it is used in this work is presented in Appendix B. The key to the success of this methodology is that the model systems remain realistic because the geometries are not changed during the process.

The design of any computational study begins with choosing a suitable model chemistry, here we are particularly concerned with the proper characterization of both H-bonding and π-stacking interactions. This is presented in Chapter 4 of the thesis. The characterization of H-bonding tends not to be a major problem for most computational methods, however the characterization of π-stacking interactions is proven to be rather challenging for a majority of quantum chemical calculations. Chapter 4 therefore involves searching for a model chemistry that can properly describe both H-bonding and dispersion type interactions, which is essential for the proper description of π-stacking interactions, yet is computationally not too demanding so that the system size is not too limiting. The validation procedure involves reproducing the electronic structure as reflected in the ionization potentials and UV-Vis spectra of a selection of substituted and un-substituted [2.2] and [3.3]paracyclophanes using various model chemistries.

Following the selection of a suitable model chemistry, the next logical step is the design and validation of the methodology. This involves selecting an adequate model system and testing the applicability of the methodology to realistic systems. In Chapter 5 the realistic systems involve canonical base pair sequences in DNA and RNA oligonucleotides. The study includes a comparison of the effect of different sequences on the degree of π–stacking and H-bonding in DNAs and RNAs in the context of their experimental geometries. Chapter 6 branches out into non-canonical sequence effects on the stabilizing interactions. Here we focus on the story of tandem G•U pairs in RNAs.

A good indicator of any proficient methodology is the ability to predict the unknown. This is explored in Chapter 7 where the methodology is used to exploit the structure-function relationship of riboswitches and direct the design of potential novel antagonists. The general conclusions of this work are presented in Chapter 8.

Chapter 4

New insights into the use of (TD-)DFT for geometries and electronic structures of restrained π-stacked systems: [n.n]paracyclophanes

Published as:

P.R.N. Kamya and H.M. Muchall. J. Phys. Chem. A, **112**, 13691 (2008).

4.1 Introduction

Despite the fact that weak dispersion-type interactions, such as found in π-stacking, are known to be inadequately represented by most conventional density-functional theory (DFT) methods when compared to higher correlated methods, the success of DFT methods in the reproduction of these very interactions in certain contexts is still reported.[50-57] However, while tremendous efforts have been made towards establishing which computational methods work best for dispersion-type interactions, most of the studies used to illustrate the shortcomings of DFT methods have been conducted on unconstrained, fully optimized, "stacked" benzene rings,[58-61] benzene derivatives,[58-61] and nucleic acid bases.[58,59,62-66] In most (if not all) of these studies it has been reported that DFT methods fail to locate the stacked minimum energy structures on the potential energy surfaces of these systems. This problem has been attributed in part to the asymptotic behavior of the exchange-correlation functionals that results in very weak or non-existent contributions to the correlation energy, which incidentally also affects the proper description of long-range charge-transfer interactions in large aromatic biological molecules.[52,67] Other reasons for this failure have been attributed to the incomplete knowledge of the exact exchange-correlation functional,[68] and the inability of DFT methods to account for static correlation.[69] The term unconstrained is used above to differentiate between these cases where aromatic rings are free to adjust the inter-ring distance, and those where the rings are tethered or constrained. If DFT methods fail to reproduce the stacking in unconstrained π-systems, the problems may be alleviated in tethered systems, for which [n.n]paracyclophanes are the perfect models.

Since the first reported synthesis of [2.2]paracyclophane (**1**, Figure 4.1)[70] [n.n]paracyclophanes have provided key insights into the effects of bringing two conformationally constrained benzene rings into close proximity. As a consequence of the short bridges between the aromatic rings in [2.2]paracyclophane, repulsive as opposed to stabilizing interactions occur between the π-clouds in the inter-ring region that force the π-density to the exterior faces of the aromatic rings, causing a boat-like deformation. The aromatic rings are twisted relative to one another, which partially relieves the torsion strain but in turn introduces additional strains that include a decrease of the inter-ring distance (IRD) and an elongation of the central C-C bridge bonds.[71,72] Increasing the length of the bridges by one carbon as in [3.3]paracyclophane (**2**, Figure 4.2) releases most of the steric strain in the molecule, providing enough space in the inter-ring region for stablizing π-stacking interactions.[73]

The interplay between steric strain and the distribution of π-clouds in [n.n]paracyclophanes generates unique chemical environments that have been exploited in numerous organic and inorganic applications including studies involving cation–π interactions[74] and selective catalysis.[75,76] [n.n]Paracyclophanes have also served as building blocks for various supramolecular compounds and polymers.[77,78] Recently, [2.2]paracyclophane has been used as a model system in computational studies to evaluate various model chemistries on their ability to reproduce the geometry of the global minimum.[50-52,55] Because the geometry is a function of the dispersive inter-ring interactions, an accurate reproduction of the geometry of **1** requires a good handle on electron correlation on the part of the method used. From these studies, it has been suggested that some DFT methods such as B3PW91 and PBE may in fact be able to handle long range dispersion interactions well enough to capture their contribution to the overall geometry of **1**.[50-52]

Our interest in [n.n]paracyclophanes lies in their use as models for stacked oligonucleotide bases. In small [n.n]paracyclophanes, the tether between the interacting aromatic rings restricts their vertical displacement to distances ranging from approximately 2.7 to 3.4Å, thus assuming a similar role as the sugar-phosphate backbone in oligonucleotides. Our ultimate aim is to select a DFT-based model chemistry that is able to characterize the H-bonding and π-stacking interactions in nucleic acid base pair subunits obtained from experimental structures of oligonucleotides. Our rationale for this investigation is, simply, that if the electronic structure, with ionization potentials and excitation energies as observables, of [n.n]paracyclophanes can be reproduced, the model chemistry used to produce these results is adequate for our intended purposes.

The present paper is an investigation of selected DFT methods and one correlated method in their ability to reproduce the geometries, ionization potentials and excitation energies of various [n.n]paracyclophanes. The compounds were selected to reflect a range of geometries and interactions between the two aromatic rings. Figure 4.1 shows [2.2]paracyclophane (**1**) and its derivatives with substitution on one (**1a** to **1f**) and on both rings (**1g** to **1i**). Figure 4.2 shows [3.3]paracyclophane (**2**) and its derivatives (**2a** and **2b**).

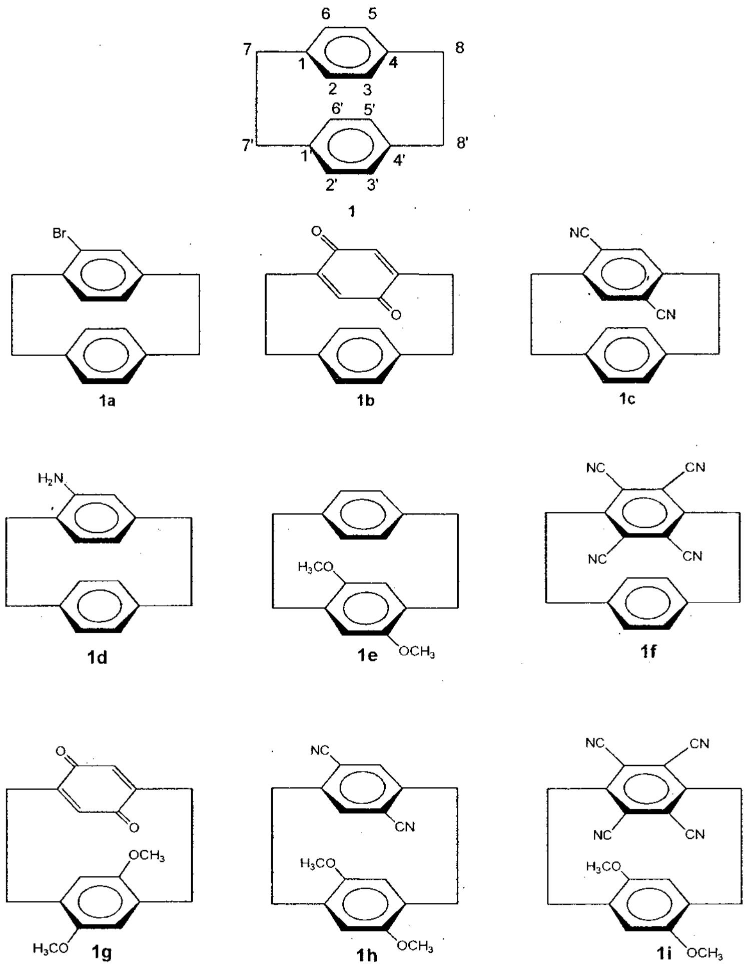


Figure 4.1 [2.2]Paracyclophanes.

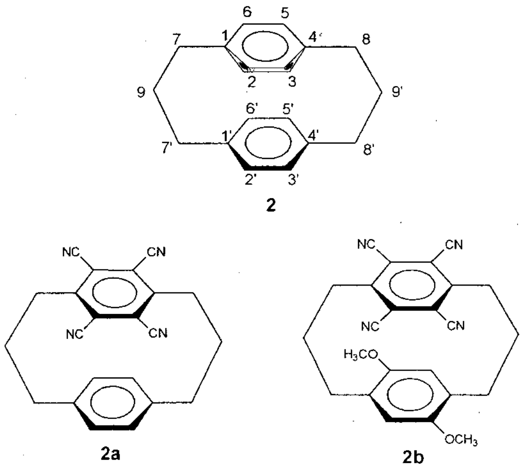


Figure 4.2 [3.3]Paracyclophanes.

4.2 Computational Methods

All calculations were performed using the Gaussian 03 suite of programs.[79] Two common hybrid DFT methods,[41] Becke's three parameter hybrid exchange functional[80] with Lee, Yang and Parr's correlation functional[81] denoted B3LYP,[82] and the parameter-free Perdew-Burke-Ernzerhof[83-85] (PBE0) functional were employed. Both functionals are known to be a good compromise between computational cost and the accuracy of results obtained, compared to higher correlated methods, and have been shown to be capable of quantifying H-bonding interactions between nucleic acid bases, which is important for our future studies.[64] The TD-PBE0 method has also recently been shown to reproduce the lowest energy singly excited state of the stacked cytosine dimer with an accuracy comparable to that of CASPT2.[86] In addition, these DFT methods have been chosen due to the controversy surrounding their abilities to capture dispersion-type interactions. [52,58] Becke's half-and-half functional (BH&H) as it is implemented in Gaussian 03 was employed due to its reported ability to adequately characterize dispersion interactions.[87] It is a hybrid half-and-half functional, where the exchange-correlation energy is calculated from HF and LSDA exchange and LYP correlation energies as shown in Equation 1.

$$E_{XC} = 0.5 \cdot E_X^{HF} + 0.5 \cdot E_X^{LSDA} + E_C^{LYP} \qquad (1)$$

For comparison, a correlated ab initio quantum chemical method, second-order Møller-Plesset (MP2),[88] was included, as it is generally considered better suited for the calculation of dispersion-type interactions despite the fact that it is known to overestimate these types of interactions.[66,89] The particular combination of MP2 with a medium sized basis set has also previously been reported as being ideal for the reproduction of the geometry of [2.2]paracyclophane.[50,55]

Geometry optimizations and frequency calculations were carried out using Pople's double or triple split-valence basis sets with diffuse and polarization functions.[90] All optimized geometries are minima on their potential energy surfaces as indicated by the absence of imaginary frequencies. Bearing in mind that our aim is to identify a usable, medium sized basis set, we evaluated the performance of 6-31G, 6-31+G(d,p) and 6-311+G(d,p) basis sets with B3LYP and PBE0 functionals in reproducing the ionization events of substituted and unsubstituted [2.2]paracyclophanes. As there was a significant improvement when diffuse and polarization functions were included compared to when the basis set increased from double to triple zeta (Figure S4.1 in the Appendix C), all analyses presented in this paper were performed with the 6-31+G(d,p) basis set. First vertical ionization potentials ($IP_{v,1}$) were calculated as the difference in total energies between a molecule and its radical cation, at the molecule's geometry. Higher ionization energies ($IP_{v,1+n}$) were calculated from the orbital energies (ε) by applying Koopmans' theorem ($IP_v \approx -\varepsilon$).[91] According to Equations 2 and 3, the energy difference between $IP_{v,1}$ and the HOMO energy is added onto the energies of the next higher orbitals (ε_{HOMO-n}) as a uniform shift.[92] All total and zero-point vibrational energies are listed in Tables S4.1-S4.3 of the Appendix C.

$$IP_{v,1} - (-\varepsilon_{HOMO}) = \Delta E \qquad (2)$$

$$-\varepsilon_{HOMO-n} + \Delta E = IP_{v,1+n} \qquad (3)$$

When dealing with large molecules, time-dependent density-functional theory (TD-DFT) is often the method of choice for the calculation of excited states, as it has been shown to be reliable for aromatic systems such as substituted phenols,[93] o-chloranil/aniline complexes[86,94] and the cytosine dimer,[86] even though the use of TD-DFT for electronic transitions with significant charge-transfer (CT) character has been questioned.[95-97] We have used TD-B3LYP, TD-PBE0 and TD-BH&H to calculate the first few excitation energies. The UV-Vis spectra were simulated using the SWizard program, revision 4.4, with the Gaussian model.[98] The half-bandwidths were taken to be equal to 3500 cm^{-1}. Molecular Orbitals were plotted from Molekel 4.3.[99,100]

4.3 Results and Discussion

4.3.1 Geometries

4.3.1.1 [2.2]Paracyclophanes

The correct representation of the geometry of [2.2]paracyclophane (**1**) with respect to its point group has been shrouded in controversy for many years.[50-52,55,56,59,71,72] The most recent publication that addresses the issue indicates that the minimum energy geometry of **1** obtained with MP2/6-31+G(d,p) is of D_2 symmetry, reduced from D_{2h} symmetry by torsion strain in accord with the dynamic disorder found in the crystal at room temperature.[50,71]

Table S4.8 of Appendix C lists the geometric parameters of **1** from full geometry optimizations using B3LYP/6-31+G(d,p), BH&H/6-31+G(d,p) and PBE0/6-31+G(d,p), those from MP2/6-31+G(d,p) reported earlier and the X-ray crystal data determined experimentally.[71] The geometric parameters and the atom numbering in Table S4.8 were chosen based on those selected by Caramori et al.[51] In general, the DFT methods perform with an accuracy close to that of MP2, with one notable exception. While B3LYP/6-31+G(d,p) reproduces the bond lengths and angles rather well, it predicts a geometry with close to D_{2h} symmetry and therefore fails to reproduce the most important property, that is the degree of twist in the methylene bridge, which is defined by the dihedral C1-C1'-C7-C7' (see atom numbering in Figure 4.1 and value for torsion angle in Table 4.1). This underestimated torsion angle from B3LYP/6-31+G(d,p) was commented on before, and it was suggested that an inadequate representation of the torsional strain for the eclipsed bridges might be the cause.[51] In contrast, with the PBE0 and BH&H functionals, not only is the twist reproduced, but the calculated torsion angle deviates from the reported experimental value even less than that from MP2. From Table 4.1, BH&H shows the closest agreement with experiment for this parameter; the PBE0 value is closer to experiment if compared to the more recently reported torsion angle of 12.6° from X-ray crystal data at 19K.[52] Finally, both PBE0 and BH&H exhibit a superior performance to B3PW91 (reported earlier).[51]

The addition of donor or acceptor groups to the aromatic rings in **1** relieves the degree of strain by modifying the π-density distribution, which generates changes to the overall geometry.[101,102] Due to the availability of their crystal structures, only **1f**, **1h** and **1i** (Figure 4.1) are included here. Because BH&H and PBE0 performed as well as MP2 in reproducing the geometry of **1**, the more expensive MP2 was not included in this evaluation. Tables S4.9-S4.11 in the Appendix C, with the geometric parameters and the atom numbering taken from Staab et al.,[102] contain the full set of geometries for **1f**, **1h** and **1i.** Table 1 compiles three important geometric parameters for **1**, **1f**, **1h** and **1i**. These are the degree of twist between the parallel aromatic rings, the inter-ring distance (IRD) taken between C1 and C1' or C4 and C4', and the boat-like deformation of the aromatic rings. For ease of comparison, the atom numbers in Table 4.1 for the substituted [2.2]paracyclophanes are as given for **1** in Figure 4.1. The experimental trend in the twist is a general reduction upon substitution, with non-zero dihedrals for all compounds. Again, B3LYP underestimates the degree of twist and therefore determines the wrong point group for **1f**, as it did for **1**, even though full symmetry is not achieved. All methods overestimate the degree of twist in **1i** (Table 4.1). This may be a consequence of the reported intermolecular stacking found in the crystal structure for **1i**,[102] which is absent in the gas phase calculations, a situation similar to that found in crystal and gas-phase geometries of biphenyls.[103-105]

The experimental inter-ring distance (IRD) between the bridgehead carbon atoms is reported to be smaller in the substituted compounds compared to **1**,[102] with the smallest IRD in **1h**.[101] The decrease in IRD that accompanies the introduction of donor and acceptor groups demonstrates an increase in favorable charge-transfer interactions[101] and a decrease in the amount of repulsion between the aromatic rings.[102] All methods reproduce the experimental trend of decreased IRDs relative to **1** (Table 4.1). In general for all molecules considered, B3LYP overestimates the IRDs, possibly due to inadequate treatment of dispersion interactions between the two aromatic rings, as has been reported with the unconstrained stacked systems in the literature.[52,53,58,66] Finally, that there is hardly any difference in the experimental IRD when directly comparing **1f** and **1i**, even though the stronger electron donating groups in **1i** result in a longer wavelength electronic transition compared to **1f**,[102] was ascribed to the rigidity of [2.2]paracyclophanes, which does not permit drastic changes to the geometry due to electronic effects.[102] In this respect, it is remarkable that PBE0 and particularly BH&H indeed determine a smaller IRD for **1i**.

We chose to represent the boat-like deformation of the aromatic rings through the C2-C3-C4-C8 torsion angle and its counterparts, where 180° indicates planarity. All methods reproduce the experimentally reported substantial boat-like deformation in all systems with high accuracy (Table 4.1). The increased non-planarity reported for the tetracyano-substituted ring in **1f**, on the other hand, is not captured by the calculations. This discrepancy may again be due to the intermolecular packing of the molecules in the crystal, favoring a more bent aromatic acceptor ring that would facilitate the stacking.

With respect to the methoxyl substituents in **1h** and **1i**, one other finding is worth reporting. An analysis of the X-ray geometry of **1h**[101] shows that the methoxyl groups deviate from the sp^2 "plane" of the aromatic ring, to a larger degree than those in **1i**.[102] A similar out-of-plane twist has been reported in the X-ray structure of 1,2-dimethoxybenzene.[106] All methods overestimate this "out-of-plane" twist of the methoxyl substituents of **1h** and **1i**, with larger differences for **1i** (Table 4.1). This is once again probably due to changes between the gas phase, where the molecules are not stacked, and the crystal where **1i** shows a higher degree of stacking than **1h**, causing the methoxyl groups to adopt a more in-plane conformation.

Table 4.1 Selected experimental and calculated (6-31+G(d,p) basis set) geometric parameters for **1**, **1f**, **1h** and **1i**.

	Exp.[a]	B3LYP	PBE0	BH&H	MP2[b]
Degree of twist[c]					
1	16.1	1.2	11.5	17.8	22.2
1f	11.5,9.4	0.0	4.4	14.6	
1h	15.8	17.7	19.6	23.1	
1i	8.7[d]	16.8	17.2	19.9	
Inter-ring distance[e]					
1	278.2	283.1	279.8	275.9	277.1
1f	274.4	280.3	276.5	272.5	
1h	273.0[d]	282.0	278.1	273.6	
1i	274.0	280.4	276.2	271.8	
Degree of boat-like deformation[f]					
1	153	152	152	152	152
1f	154,151	153	153	153	
1h	153	152	152	152[d]	
1i	154[d]	152	153[d]	153[d]	
Out-of-plane twist of methoxyl groups[g]					
1h	9.6[d]	11.5	11.1	13.6	
1i	8.0,5.2	13.2	12.0	13.3	

[a] X-ray data from [71](**1**) [102] (**1f** and **1i**) and [101] (**1h**)

[b] From [51]

[c] Given by the torsion angles C1-C7-C7'-C1' and C4-C8-C8'-C4' in degrees (identical if one value is listed).

[d] Averaged value for a difference in distances of 0.3 pm and in angles of 1° or less.

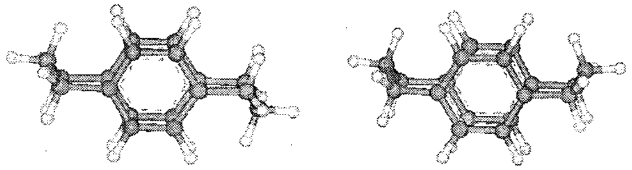
[e] Given by the distances C1-C1' and C4-C4' in pm (identical if one value is listed).

[f] Average "bend" for each ring given by the torsion angles C2-C3-C4-C8, C3-C2-C1-C7, C6-C5-C1-C7, C5-C6-C1-C7 and corresponding torsion angles for the second ring in degrees (identical if one value is listed).

[g] Given by the torsion angles C2'-C3'-O-CH_3 and C5'-C6'-O-CH_3 in degrees (identical if one value is listed).

4.3.1.2 [3.3]Paracyclophanes

The longer bridges between the aromatic rings in [3.3]paracyclophanes provide these molecules with more flexibility than their [2.2]paracyclophane counterparts, thus allowing the effects of substitution on the overall geometry to be more apparent, while increasing the variability of the test set of compounds.[102] Gantzel and Trueblood showed that the crystal structure of [3.3]paracyclophane (**2**) has somewhat deformed aromatic rings that exhibit a bent-boat conformation, with a slight twist of one ring relative to the other. [107] The tri-methylene groups of the tethers are in a conformation that is similar to that in gauche n-butane, and with respect to each other adopt an anti conformation now more widely referred to as "chair" (Figure 4.3).[108]Anet and Brown showed that **2** actually adopts two conformations in solution, the "chair" and the "boat" (syn conformation, Figure 4.3).[108] The relative ratio of chair to boat in $CDCl_3$-$CDCl_2F$ solution at -88°C was determined to be about 1:2.[108] Similarly, **2b** has been determined to exist as 40% chair and 60% boat in the crystal.[102] Gas phase calculations on the relative ratio of chair to boat conformers have not been conducted (to the best of our knowledge), the assumption being that the distribution should be similar to that in solution.[109] We have performed geometry optimizations on [3.3]paracyclophane (**2**) and its derivatives (**2a** and **2b**) using B3LYP, PBE0 and BH&H functionals with the 6-31+G(d,p) basis set. Population distributions for each compound were calculated from zero-point corrected energies at 185 or 298 K, and as with the [2.2]paracyclophane series, we have compared the calculated geometries to experimental data.[102]



2 (chair) **2** (boat)

Figure 4.3 Conformers of **2**.

Table 4.2 shows the experimental and calculated population distributions for **2**, **2a** and **2b**. For **2**, the population of conformers in solution is close to that found in the gas phase with B3LYP and PBE0; BH&H on the other hand predicts a 50:50 distribution. Due to the facts that we are attempting to reproduce solution NMR results in the gas phase, and that the energy difference between the two conformers is small ($\Delta E < 0.8$ kcal mol^{-1}), it is safe to state that all methods perform well. While there are no experimental data on the population distribution of **2a**, all methods agree on a large preference for the boat. For **2b**, only BH&H reproduces the experimental preference for the boat, but again, the energy differences determined with all methods are small, and distributions in the crystal and the gas phase are not necessarily comparable.

Tables S4.12-S4.14 in the Appendix C show the experimental and calculated geometric parameters of **2**, **2a** and **2b**. The important geometric parameters are compiled in Table 4.3. We are once again interested in those parameters that best reflect the electronic structure of the molecules, namely the degree of twist, the IRD and the degree of boat-like deformation of the aromatic rings..

All methods capture the overall release in strain brought about by the longer tether in **2** compared to **1**. The degree of twist between the aromatic rings is smaller, the IRDs are longer and the degree of boat-like bend of the aromatic rings is significantly less (Table 4.3). However, even though the experimental value is small, both B3LYP and PBE0 fail to predict a non-zero value for the degree of twist in **2**, while BH&H does. The experimental trend in the IRD shows that as for the [2.2]paracyclophanes upon substitution, the IRD decreases, and this trend is reproduced by all methods. However, the calculated IRDs for **2** and **2b** show that B3LYP overestimates the distance on average by approximately 10 pm, demonstrating once again that this method underestimates the extent of interaction between the two rings. PBE0 and BH&H predict IRDs that are closer to experiment (Table 4.3).

Compared to **1**, the aromatic rings in **2** are significantly less bent, with values closer to 180°. [107] All functionals reproduce the degree of boat-like deformation with differences from the experimental values of 1° or less (Table 4.3).[102] As in **1i**, the methoxyl substituents in **2b** are not "co-planar" with the aromatic ring (Table 4.3). While B3LYP predicts identical values for the chair and boat, for PBE0 the C-C-O-C twist is less in the boat conformer, whereas for BH&H this twist is less in the chair conformer. There is thus a correlation between the C-C-O-C twist and the stability of the conformers, in line with the fact that for methoxybenzenes (anisoles) the methoxyl group in general lies in the plane of the aromatic ring.[93,104]

Table 4.2 Experimental and calculated[a] (6-31+G(d,p) basis set) population distributions (%) of **2**, **2a** and **2b**.

	Exp.	B3LYP	PBE0	BH&H
2 chair	33[b]	39 (32)[c]	40 (35)[c]	50 (51)[c]
2 boat	66[b]	61 (68)[c]	60 (65)[c]	50 (49)[c]
2a chair		31	13	14
2a boat		69	87	86
2b chair	40[d]	70	67	48
2b boat	60[d]	29	33	52

[a] With zero-point vibrational energy corrections, at 298 K.
[b] Given as chair:boat 1:2 in ref . [108]
[c] With zero-point vibrational energy corrections, at 185 K.
[d] From ref [102]

Table 4.3 Selected experimental and calculated (6-31+G(d,p) basis set) geometric parameters for **2**, **2a** and **2b**.

	Exp.[a]		B3LYP		PBE0		BH&H	
	Chair	Boat	Chair	Boat	Chair	Boat	Chair	Boat
Degree of twist[b]								
2	5.0		0.0	0.0	0.0	0.0	7.3	1.6
2a			5.0	5.7,3.7	6.2	3.3	7.9	4.0[c]
2b		7.1,1.0	1.8	4.2,0.5	1.8	6.4,0.9	1.2	8.9,0.2
Inter-ring distance[d]								
2	313.7		323.9	324.5	319.0	319.4	310.9	312.8
2a			320.4[c]	320.6	314.9	315.1	306.9	307.3
2b		309.9,306.8	319.6	320.3[c]	313.8	314.1[c]	304.9	304.9,302.1
Degree of boat-like deformation[e]								
2	168		167	167	168	168	168	168
2a			168	164,168	169[c]	169[c]	170,168	170,169
2b		170[c]	168[c]	168	170[c]	169	170	171,169
Out-of-plane twist of methoxyl groups[f]								
2b		7.7,5.8	7.6	7.6	5.5	5.7,2.7	2.5	5.6,7.0

[a] X-ray data from ref [107] (**2**) and ref [102] (**2b**).

[b] Given by the improper torsion angles C1-C7-C7'-C1' and C4-C8-C8'-C4' in degrees (identical if one value is listed).

[c] Averaged value for a difference in distances of 1 pm or less and in angles of 1° or less.

[d] Given by the distances C1-C1' and C4-C4' in pm (identical if one value is listed).

[e] Average "bend" for each ring given by the torsion angles C2-C3-C4-C8, C3-C2-C1-C7, C6-C5-C1-C7, C5-C6-C1-C7 and corresponding torsion angles for the second ring in degrees (identical if one value is listed).

[f] Given by the torsion angles C2'-C3'-O-CH_3 and C5'-C6'-O-CH_3 in degrees (identical if one value is listed).

4.4 Ionization energies

The proximity of the two conformationally constrained aromatic rings in [n.n]paracyclophanes is known to influence the electronic structure of these molecules in a way that is depicted in their photophysical behavior.[109] As there is a large collection of photoelectron (PE) spectra of [n.n]paracyclophanes in the literature, a reproduction of the ionization energies of these molecules can be used as a means of further evaluating how well the selected functionals can reproduce their electronic structure. In this section, we compare ionization data from available published PE spectra to calculated IPs where B3LYP, PBE0, BH&H and MP2 methods are considered for **1**, and B3LYP, PBE0 and BH&H for **1a-1e**, **1g**, **1h** and **2**. This particular group of compounds was chosen for their wide range of ionization potentials that reflect the variety in their electronic structures and therefore pose a suitable challenge for the methods considered. All experimental vertical ionization potentials were taken from ref[109] and the original references therein. Tables S4.15-S4.17 of Appendix C contain the numerical IP data for all compounds considered.

Plots showing correlations between calculated and experimental values for the first five ionization events of **1** and **2** (boat conformer) are shown in Figures. 4.4a and 4.4b, respectively.

All methods reproduce the IP_vs of **1** and **2** rather well, with R^2 values in most cases close to 0.99. Yet only B3LYP and PBE0 correlations possess slopes of close to 1.0 (1.03 and 1.05, respectively, for **1**; 1.04 and 1.06, respectively, for **2**), whereas those from BH&H (1.27 for **1**, 1.26 for **2**) and MP2 (1.67) are much steeper, resulting in a progressively more serious overestimation of the higher IPs. The B3LYP performance here is particularly encouraging, as it shows that while it in particular did not perform as well as BH&H and PBE0 when reproducing the geometric parameters of **1** and **2**, it does not fail to grasp the necessary electronic effects required to reproduce their PE spectra.

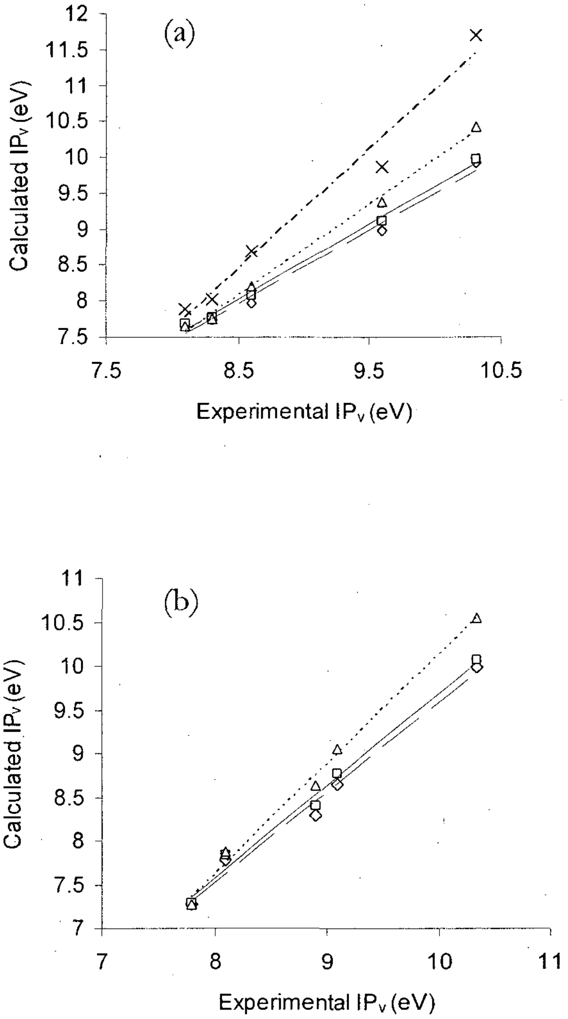


Figure 4.4 Correlation between the first five experimental and calculated vertical ionization potentials (eV) of a) **1** and b) **2** (boat conformer). Correlations are a) – – – B3LYP (◊, R^2 0.9876), —— PBE0 (□, R^2 0.9950), ········ BH&H (Δ, R^2 0.9958), — ▪ — ▪ MP2 (x, R^2 0.9744) and b) – – – B3LYP (◊, R^2 0.9868), —— PBE0 (□, R^2 0.9898), ········ BH&H (Δ, R^2 0.9934).

Substitution in [2.2]paracyclophanes leads to donor-acceptor interactions in the molecule that produce changes in the PE spectra that depend on the placement and nature of the functional groups.[110] Figure 4.5 shows the correlation between the calculated first five ionization events for **1a-1e**, **1g**, **1h** and **2**, and their experimental IPs.

The compounds included in this analysis can be divided into two groups, those with orbital contributions from only one ring and those from both rings. Figure 4.6 shows one example for each. Instinctively, one would expect the DFT methods to be successful in reproducing the low-energy ionizations of the former, as from an orbital perspective these are simply aromatic compounds with para-substitution. However, Figure 4.5 shows that all methods are able to reproduce the experimental IPs of all the compounds, regardless of whether orbital coefficients are found on one or on both rings, and that the π-π interaction of the two rings in the latter case is adequately captured. Finally, we note that the BH&H data points exhibit a much more pronounced scatter than that for the other two functionals, and its correlation shows the above mentioned deviation from a perfect slope (slopes are B3LYP 0.99, PBE0 1.02, BH&H 1.23).

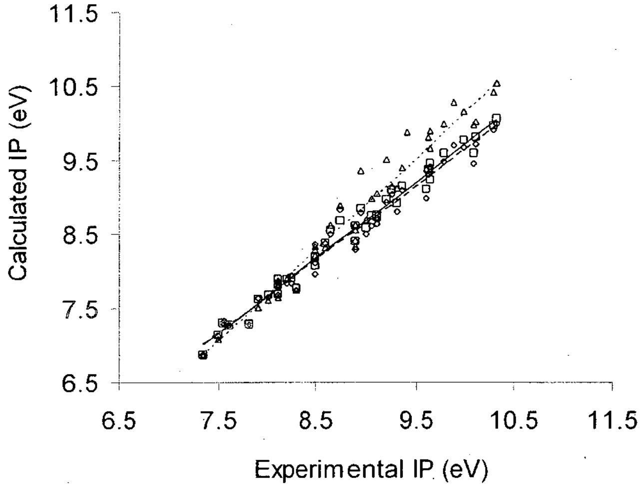
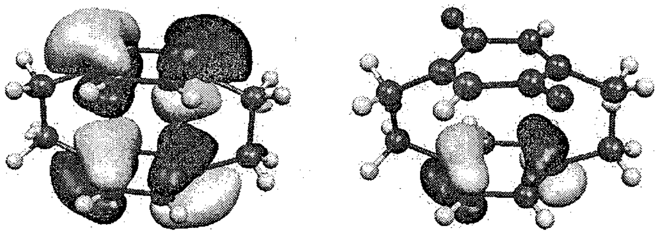


Figure 4.5 Correlation between experimental and calculated vertical ionization potentials for **1**, **1a**-**1e**, **1g**-**1h** and **2**. – – – B3LYP (◊, R^2 0.9691), —— PBE0 (□, R^2 0.9841), ········ BH&H (Δ, R^2 0.9608).



1 1b

Figure 4.6 Highest occupied molecular orbital (HOMO) of **1** and **1b.**

4.5 Excitaiton Energies

The substituted [2.2]paracyclophanes **1f** and **1i**, and [3.3]paracyclophanes **2a** and **2b** (Figures 4.1 and 4.2) have a tetracyanobenzene ring as a common electron acceptor. The variety in this group of compounds stems from differences in the strengths of the donor rings and from the donor-acceptor ring distances. Staab et al. have shown that these two properties have an effect on the geometries and on the charge-transfer properties of the molecules.[102] While we have shown above that the effects of substitution on the geometries of **1f** and **1i** in particular are small, the effects on the charge-transfer properties of these molecules are so dramatic that they can be examined visibly with the naked eye.[102] Compound **1f** with the weakest donor is yellow, **1i** with the stronger donor ring is a deep violet. Keeping the dimethoxyl substitution and increasing the donor-acceptor distance from **1i** to **2b** results in a change in color from deep violet to dark red.[102] Quantitatively, these compounds exhibit "phane-specific" changes, which are displayed in their UV-Vis spectra (in chloroform) and consist of broadening of absorption bands, loss of vibronic structure and the appearance of new absorptions. More specifically, while the charge-transfer transition in **1f** gives rise to a small shoulder at 395 nm, strengthening the donor (**1i**) results in a large bathochromic shift to 520 nm.[52]

A larger distance between the rings results in a somewhat smaller red shift from 416 nm for **2a** to 508 nm for **2b** (a blue shift from **1i**). This shows that the longer donor-acceptor distance causes a similar but less pronounced effect on the charge-transfer transition compared to the substituted [2.2]paracyclophanes.[102]

We calculated the excitation energies of **1**, **1f**, **1i**, **2**, **2a** and **2b** using time-dependent density functional theory (TD-DFT) and simulated the UV-Vis spectra from the TD-DFT output. The parent compounds **1** and **2** are included in this part of the study as their UV-Vis spectra are readily available and their longest wavelength transitions are closest to those of stacked nucleic acid base pairs (240-260 nm).[111] An overlay of the experimental [73] and simulated spectra for **1** is shown in Figure 4.7 Selected simulations for **1**, **2** (boat conformer) and **2b** (chair conformer) are shown in Figure 4.8, with particular focus on the long-wavelength band for each (see insets). Numerical data for the lowest-energy transition in **1**, **1f**, **1i**, **2**, **2a** and **2b** can be found in Table 4.4.

Figure 4.7 shows the overall good agreement between the experimental spectrum and the spectra simulated from TD-PBE0 and TD-BH&H for **1**, the latter functional giving rise to a spectrum that shows more of the experimental features. The TD-B3LYP spectrum is similar to that from TD-PBE0, but as it is even more featureless than that from TD-PBE0, it is omitted for clarity. It is noted that the TD-BH&H spectrum is shifted to higher energies relative to that from TD-PBE0. With respect to the lowest-energy transition, the simulated UV spectra for **1** (Figure 4.8a) obtained with TD-PBE0 and TD-BH&H show good agreement with experimental data (λ_{max} of 302 nm, Table 4.4), in that a shoulder at about 290 nm or a distinct band at about 270 nm, respectively, can be seen (Figure 4.8a inset). In the spectrum from TD-B3LYP, on the other hand, even though an allowed transition at 271 nm is calculated (Table 4.4), this band is not discernible.

As for **1,** the simulated UV spectra of the boat conformer of **2** (Figure 4.8b) show that TD-BH&H produces a spectrum that is shifted to shorter wavelengths, resulting in too low a value for λ_{max} of the lowest-energy transition. In contrast, TD-B3LYP and TD-PBE0 predict the longest absorption wavelength close to the experimental 294 nm (Figure 4.8b inset). All methods, however, are in agreement in so far as only the boat conformer possesses an allowed long-wavelength transition close to the reported literature value (Table 4.4). This difference in spectroscopic behavior of the two conformers is yet to be confirmed experimentally.

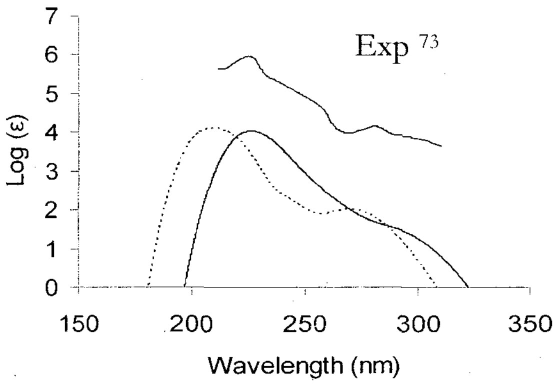


Figure 4.7 Overlay of experimental[73] and simulated (6-31+G(d,p) basis set) TD-PBE0 (——) and TD-BH&H (·········) UV-Vis spectra of **1.**

For the tetracyanobenzene series **1f**, **1i**, **2a** and **2b** as a whole, it is obvious from Table 4.4 that neither TD-B3LYP nor TD-PBE0 provide useful data. Both functionals overestimate the wavelength for the first electronic transition already for **1f** and **2a**, and with the stronger donor rings in **1i** and **2b**, this becomes dramatic. For **2b** with TD-

B3LYP, e.g., λ_{max} is calculated about 280 nm too long. This reflects a serious underestimation in the energy required for the charge-transfer interaction in these compounds, which has been documented before in the calculation of other long-range charge-transfer transitions with TD-DFT.[96]

Figure 4.8c shows an overlay of the calculated spectra for **2b**, where the experimental value for the longest wavelength transition is reported at 508 nm.[102] Interestingly, TD-BH&H produces a λ_{max} of approximately 500 nm, and this good agreement with experiment is true across the series. In fact, the deviation from the experimental value for **1** is about 40 nm, for **1f**, **2** and **2a** about 20 nm, less for **2b**. This suggests that for **1i**, the 495 nm value (see footnote to Table 4.4) should be considered instead of the 377 nm listed, as the larger value again deviates from the experimental value by about 20 nm. If one allows for this, the reproduction of the bathochromic (red) shifts from **1f** and **2a** upon introduction of the stronger electron donating rings in **1i** and **2b** is excellent. For experimental red shifts of 125 and 92 nm in the [2.2] and [3.3] series, respectively, we calculate shifts of 121 and 111 nm with TD-BH&H for the boat conformers.

BH&H has been reported to be capable of reproducing the potential energy surfaces of stacked benzene rings and nucleic acid bases, and the lowest energy conformations of many stacked aromatic compounds.[87] In light of this, the TD-BH&H performance for electronic excitations here is certainly impressive, but maybe not completely unexpected.

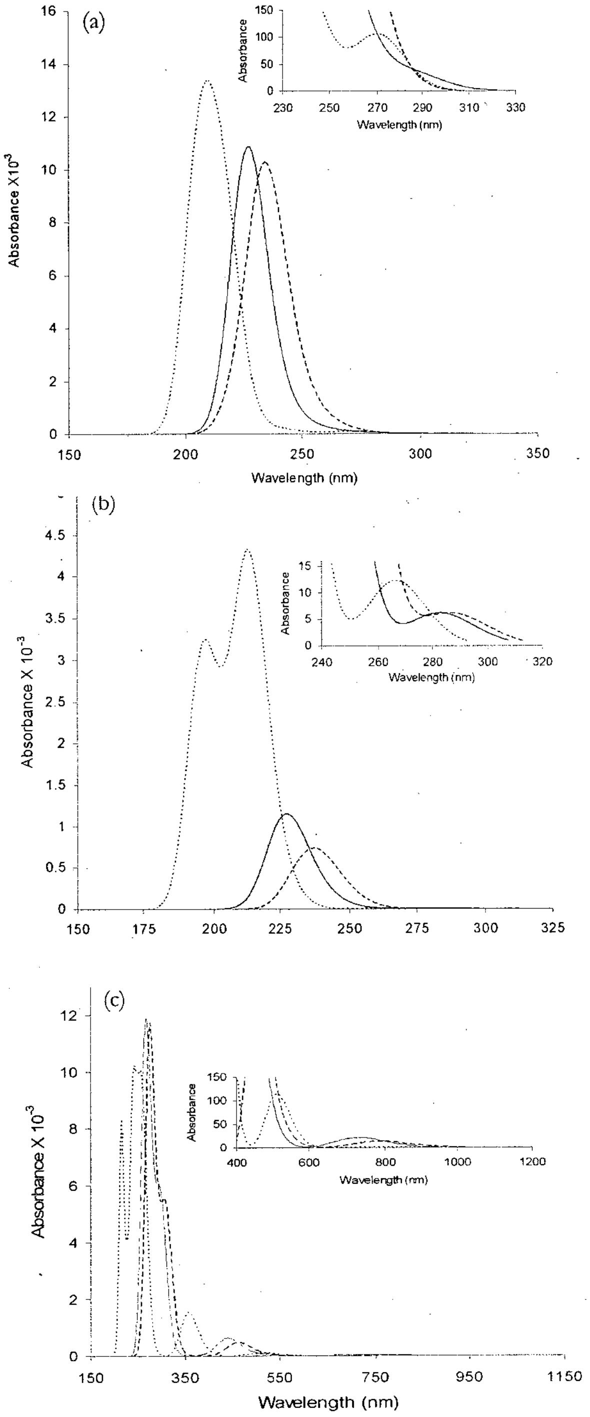


Figure 4.8 Overlay of simulated UV-Vis spectra of a) **1** and b) **2** (boat conformer) and c) **2b** (chair conformer) – – – TD-B3LYP, —— TD-PBE0 and ········TD- BH&H (6-31+G(d,p) basis set).

Table 4.4 Experimental and calculated (6-31+G(d,p) basis set) wavelengths (nm) and oscillator strengths (in italics) for the first electronic transition of **1**, **1f**, **1i**, **2**, **2a** and **2b**.

	Exp.	TD-B3LYP	TD-PBE0	TD-BH&H
1	302[a]	271[b] *0.0009*	286 *0.0005*	271 *0.0017*
1f	395sh[c,d]	419[b] *0.0039*	457 *0.0001*	374 *0.0024*
1i	520[c]	752 *0.0001*	703 *0.0001*	377[b] *0.0041*
2 chair	294[a]	243[b] *0.0073*	237[b] *0.0073*	226[b] *0.0078*
2 boat		287 *0.0001*	283 *0.0001*	267 *0.0002*
2a chair	416[c]	502 *0.0057*	478 *0.0072*	391 *0.0158*
2a boat		496 *0.0093*	472 *0.0127*	384 *0.0271*
2b chair	508[b]	787 *0.0002*	730 *0.0003*	515 *0.0016*
2b boat		786 *0.0012*	728 *0.0030*	495 *0.0179*

[a] From. [73]

[b] Longer-wavelength zero-intensity transitions: B3LYP 291 nm **1**, 479 nm **1f**, 290 nm **2** chair; PBE0 286, 270, 245 nm **2** chair; BH&H 264, 245 nm **2** chair, 495 nm **1i**.

[c] From. [102]

[d] Shoulder.

However, it should be noted that the long-wavelength transitions for compounds **1f**, **1i**, **2a** and **2b** reflect charge-transfer interactions that are much stronger than those encountered in stacked nucleic acid base pairs, which have λ_{max} values closer to those exhibited by **1** and **2**. Table 4.4 shows, as was discussed above, that both TD-B3LYP and TD-PBE0 are capable of reproducing the wavelength of the first electronic transition for these systems.

4.6 Conclusions

We speculated that a potential way of circumventing the problem DFT functionals have with π-stacking interactions was to introduce a tether between the stacked aromatic rings. From this study on [n.n]paracyclophanes, we were able to effectively examine the effects of constraining two interacting aromatic rings on the performance of (TD-)B3LYP, (TD-)PBE0 and (TD-)BH&H for geometries, ionization potentials and excitation energies. The addition of the tether between the interacting rings has improved the performance of PBE0, as is evident from the adequate reproduction of geometries and IPs, whereas B3LYP appears to benefit less. Both TD-B3LYP and TD-PBE0 tend to underestimate charge-transfer excitation energies, giving rise to λ_{max} values in the low-energy region that are grossly exaggerated, while for [n.n]paracyclophanes with weaker donor-acceptor interactions experimental wavelengths are reproduced well. In fact, it is the performance on these latter [n.n]paracyclophanes that is important, as they have lowest-energy transitions close to those of stacked nucleic acid base pairs (260-280 nm). While overall (TD-)BH&H shows a very good performance in this study, it has been reported to overestimate hydrogen bond strengths. Therefore, for the description of oligonucleotide fragments, we recommend the use of (TD-)PBE0, as it not only performs just as well as (TD-)BH&H in most contexts here, but is also known to accurately capture the strength of hydrogen bonds.

Preamble to Chapters 5, 6 and 7

The geometry of a base pair in the crystal structure of a NA is going to be at its minimum energy relative to the forces acting on it from its surroundings. This means that if the base pair is isolated and its geometry is maintained, then the effects that the forces had on the geometry of the base pair while it was part of the NA are also maintained. However, the use of un-optimized experimental geometries in a computational analysis of the electronic structure of nucleic acids has been criticized.[66,112] But this is not a common conviction as there are publications in the literature that show otherwise.[62,113] In addition, performing a single point energy calculation directly on an experimental geometry, using the model chemistry we have chosen, is equivalent to performing a single point energy calculation at a higher level of theory on a geometry that was optimized at a lower level of theory which is common place in QC studies. Finally, all the X-ray crystal and NMR determined molecular structures for the oligonucleotides used in this work have been through a molecular dynamics protocol, which admittedly is capable of detecting many inaccuracies in the experimental geometries.[66]

Chapter 5

Revisiting the sequence and structural effects on the hydrogen bonding and π-stacking interactions in nucleic acids

To be submitted to:
Journal of Physical Chemistry B

5.1 Introduction

Two challenging goals in nucleic acid (NA) research involve deciphering the effects of sequence and structure on the stabilizing interactions, and understanding the nature of the relationship between the stabilizing interactions at the atomic level. The difficulty of these tasks is exacerbated by the complicated intricacy of the interdependence between NA structure and function.[27,114] The most fundamental structural properties of NAs involve the H-bonding and π-stacking interactions that occur between the bases which provide the driving force for folding and stabilize the overall structure. The strength and integrity of these interactions are dictated by the intrinsic nature of the bases and base pairs (the electronic structure), which in turn is a consequence of the forces acting on the bases due to the presence of the sugar-phosphate backbone, and the surrounding bases and base pairs. All of these properties work together to direct the structural parameters such as the twist, roll, and rise of base pairs. [111] Simultaneously understanding their effects on the stabilizing interactions can be a daunting task regardless of the methods used.[24,27]

Experimental methods such as X-ray diffraction and NMR spectroscopy have proven invaluable in the determination of oligonucleotide structures, but they can not unequivocally identify the presence, nature and origin of non-covalent interactions. [27] Computational methods can provide this information but a considerable challenge is the large size of oligonucleotides. This problem can be partially overcome by using empirical force field methods that are based on molecular mechanics (MM) for structure determination. Also, through the application of Newton's equation of motion, molecular dynamics (MD) can combine the benefits of X-ray crystallography and NMR spectroscopy albeit on a much shorter time-scale. Some MM and MD methods have been shown to perform rather well,[115,116] but they have limitations. They often neglect or inaccurately represent various electrostatic effects and do not cover polarization of the bases that is a consequence of the inter-base interactions.[11,117] Furthermore, they fail to reproduce structural effects such as the conformational flexibility in pyrimidine rings and the pyrimidalization of the amino group and can not easily isolate the structural contributions to the overall stability.[27,118] Quantum chemical (QC) calculations, on the other hand, can be used to determine the structure and energetics of base pairs, providing a more comprehensive understanding of the nature of this relationship. However, QC calculations are limited by the size of the systems. As a direct consequence, previous QC studies aimed at characterizing the molecular properties of NAs were conducted on small model systems limited to two, and more recently four, nucleic acid bases where the effects of each property (such as twist, rise, sequence etc.) on the stabilizing interactions between bases were monitored separately.[27,44,62,117,119] In addition it has been shown that some of the more popular density functional theory (DFT) methods including B3LYP[80-82] lack the ability to characterize long-range π–stacking interactions correctly.[44,66,120] To address this, new DFT methods that include dispersion corrections such as Truhlar's MPWB1K[121] have been developed. For a recent overview of new methods see reference.[122]

A promising solution to these problems would be to combine the positive attributes from experimental data and computational techniques. This has previously been employed through a combination of experimentally determined geometries and quantum chemical calculations for the analysis of the stabilizing interactions between nucleic acid bases.[62,123] Hobza et al. adopted this approach after earlier work had shown a significant difference (10%) between the stacking energies of energy minimized and experimental nucleic acid base pairs in the gas-phase.[124] This study illustrated that the strengths of the interactions are very sensitive to the structure, highlighting the benefits of using experimentally determined geometries that capture all effects of the environment of a base pair.

In a similar approach,[113] evaluated closed-shell interactions between (base/base or base/sugar phosphate backbone) for structural elements isolated from experimentally determined NA geometries, using a topological analysis of the electron density.[48] While the quantum theory of Atoms in Molecules provides unambiguous and quantitative data for these weak bonding interactions, the limited number of base pairs studied coupled with the unfortunate choice of the B3LYP functional to evaluate π-stacking interactions, does not allow for general conclusions.[113]

Through a comprehensive analysis of the electron density distribution of base pairs isolated from a total of eleven oligonucleotides (Table 5.1), comprising both DNA and RNA structures determined using X-ray diffraction, we demonstrate that this methodology is unbiased, accurate and sensitive enough to observe even small effects on the stabilizing interactions between the base pairs in DNA and RNA.

Table 5.1 PDB ID, X-ray resolution (Å) and sequence (5'→3') of DNA and RNA duplexes

PDB ID	X-ray res	Sequence 5'→3'	Reference
		DNA	
119D	2.25	CGTAGATCTACG	[125]
1VJ4	1.80	GGTATACC	[126]
126D	2.00	CATGGCCATG	[127]
1SK5	0.89	CTTTTAAAAG	[128]
1IKK	1.60	CCTTTAAAGG	[129]
440D	1.10	AGGGGCCCCT	[130]
		RNA	
1RNA	2.25	UUAUAUAUAUAUAA	[131]
157D	1.80	CGCGAAUUAGCG	[132]
420D	1.90	GCAGAGUUAAAUCUGC	[133]
485D	0.97	GUGAUCGC	[134]
259D	1.46	CCCCGGGG	[135]

It is thus possible to simultaneously observe and quantify the effects of sequence and structure on the stabilizing interactions as well as possibly analyzing the relationship between the stabilizing interactions in NA base pairs. Central to the success of a methodology that uses experimentally determined NA geometries is the knowledge that the form assumed by the electron density is a direct consequence of the forces acting on the system.[48] Therefore, base pairs that are isolated from X-ray or NMR structures possess *all* the environmental effects that are present in the whole oligonucleotide, and an analysis of the electron density can uncover those.

5.2 Methods and Computational Details

Protein Data Bank (PDB) files for all oligomers listed in Table 5.1 were obtained from the Nucleic Acid database.[128] All duplex geometries are X-ray crystal structures (R values 12-19%, resolutions from 0.89-2.25 Å). The sequences chosen vary in length and composition, amongst them two models representing A-tracts (1IKK and 1SK5), a model duplex comprised of AU base pairs (1RNA), two duplexes that contain only GC base pairs (259D), an A-RNA helix and an A-DNA helix 440D, terminated with AT pairs and two duplexes containing mis-matched base pairs 157D, DNA and 485D, RNA with AG and GU mismatches, respectively.

Base pairs of interest were isolated through the deletion of the surrounding duplex and the replacement of the sugar-phosphate backbone with a hydrogen atom. The removal of the sugar-phosphate backbone has been shown to have no significant effect the relative strengths of the H-bonding and the π–stacking interactions between the bases and base pairs.[136] Bickelhaupt et al. compared H-bond lengths and enthalpies in the absence and presence of the sugars and found negligible differences in the two parameters.[137] To test whether this would be true for our systems we included the sugar-phosphate backbone into the calculation of the wavefunction for the H-bonded and π–stacked base pairs, isolated from 1IKK, and observed an insignificant increase in the $\sum\varrho_{HB}$ in the range of 0.003- 0.0003 e/Å^3 for the H-bonds with no change in the trend. We also compared the $\sum\varrho_{HB}$ for base pairs in the presence and absence of their nearest neighbors and found negligible differences (Tables S5.1-5.12, Appendix D)

All wavefunctions were generated from a single point energy calculation in order to maintain the experimental geometries, using the Gaussian 03 suite of programs. [79] We employed the parameter-free hybrid DFT method PBE0 by Perdew, Burke and Ernzerhof,[83-85] and Pople's double split-valence basis set with diffuse and polarization functions (6-31+G(d,p)), which are known to be necessary for the proper description of dispersion-type interactions.[27]

The PBE0 functional was chosen because it performs almost as well as the Becke Half-and-Half (BH&H) functional (as implemented in Gaussian 03), which has previously been shown to quantify π–stacking interactions remarkably well,[87] especially when reproducing the electronic structure of constrained π-stacked aromatic compounds.[120] In addition, PBE0/6-31+G(d,p) is known to characterize H-bonding reasonably well.[138,139] In this study, we found that for both $\sum\rho_{HB}$ and the sum of π-stacking densities, $\sum\rho_{\pi}$, the BH&H values tended to be consistently higher by approximately 0.02e/Å^3, with no change in the trends in relative strength (data not shown).

The quantum theory of Atoms in Molecules provides an approach for the identification and quantification of bonding interactions between any two atoms in terms of critical points in the topology of the electron density which are regions where the first derivative of the density vanishes, ($\nabla\rho_{(cp)} = 0$).[48] The electron density was extracted from the many-particle wavefunction and its topology analyzed using AIM 2000,[49] which produces a representation of the atoms and their bonding interactions, displayed as a molecular graph (Figure 5.1). (See Appendix B for introduction to QTAIM)

Finally to demonstrate the negative effect of a full geometry optimization, optimized AT, AU and GC base pairs from the same model chemistry (PBE0/6-31+G(d,p)) are also included.

5.3 Results and Discussion

5.3.1 Quantification of hydrogen bonding and π-stacking interactions

Bearing in mind that "the form assumed by the distribution of charge in a molecular system is the physical manifestation of the forces acting within the system,"[48] we propose that a molecular graph of the electron density of isolated, un-optimized, base pairs provides us with a means of directly observing sequence and structural effects on the stabilizing interactions between the base pairs, without the need to a simulate these effects, particularly the energetic properties.

It is well known that the density at a bond critical point acts as a direct measure of the strength of the bond,[48] and that there is an inverse correlation between the density at the bond critical point (ρ) and the length of the bond. Various groups have shown an exponential correlation for the range from strong to weak interactions (including those of a van der Waals nature) using data obtained from both calculated and experimental charge density distributions.[140,141]

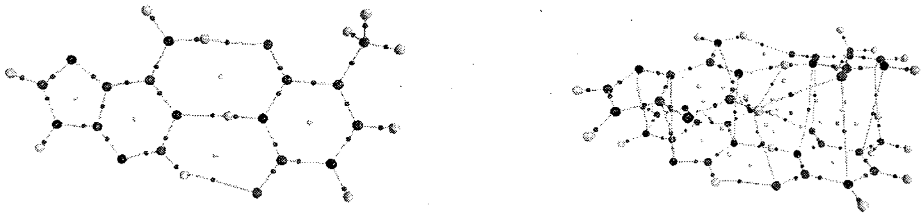


Figure 5.1 Molecular graphs for a) the isolated H-bonded A16T5 base pair and b) π-stacked AT (A16T5_A17T4) base pairs from 1IKK.

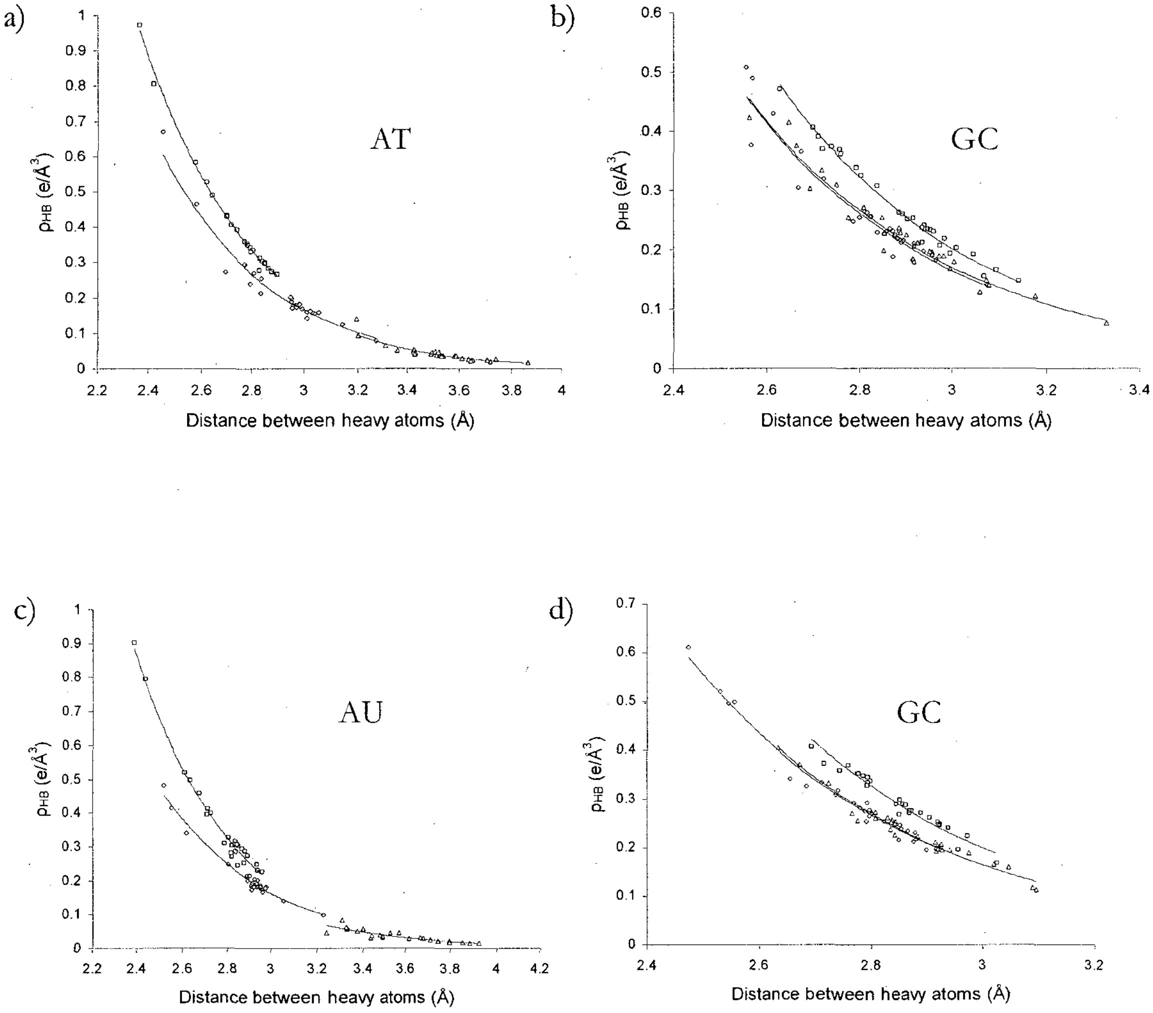


Figure 5.2 Correlation between the heavy atom distance (Å) and ϱ_{HB} $(e/\text{Å}^3)$from a) and b) DNA and c) and d) RNA duplexes. Correlation coefficients are a) AT: ◊ N-H(O) $R^2 = 0.9935$, □ N-H(N) $R^2 = 0.9631$ and Δ C-H(O) $R^2 = 0.9046$ b) GC: ◊ N-H(O) $R^2 = 0.9570$, □ N-H(N) $R^2 = 0.9434$, (O)H-N $R^2 = 0.9764$ c) AU: ◊ N-H(O) $R^2 = 0.9675$, □ N-H(N) $R^2 = 0.9706$, and Δ C-H(O) $R^2 = 0.8393$ and d) GC: ◊ N-H(O) $R^2 = 0.9769$, □ N-H(N) $R^2 = 0.9489$, (O)H-N $R^2 = 0.9558$.

Figure 5.2 and Figure 5.4 show an exponential correlation between the electron density at the bond critical points of both H-bonding (ϱ_{HB}) and π–stacking (ϱ_{π}) interactions and the distance between the interacting nuclei. Which means that the values of both ϱ_{HB} and ϱ_{π}, can be used as a measure of the strength of the interaction. For the

ρ_{HB} correlations, the distance between the heavy atoms, rather than the H-bond distance is plotted to demonstrate that the dependence is not a manifestation of the added hydrogen atoms. A good spread in the data is a reflection of the use of experimental geometries where the geometries and therefore the strength of the H-bonds vary, but also remain within acceptable values (approximately 2.2-4.0Å)[28] for heavy atom distances between H-bonded nuclei.

A clear distinction in the strengths of the H-bonds is apparent in the plots where the N2-H···O2, N1···H-N3, N6-H···O4 in the GC base pairs and N6-H···O4 and N1···H-N3 bonds in AT/U base pairs fall into the same range, while the obviously weaker C2-H···O4 bond in AT/U base pairs have longer bond lengths and smaller values of ρ_{HB}. (see Figure 5.3 for numbering)

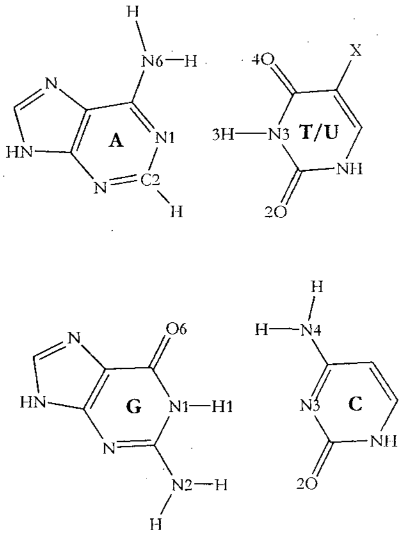


Figure 5.3 Numbering for base pairs a) AT/U where X=CH3 in AT and H in AU base pairs and b) GC.

The correlation coefficients (R^2) for the π-interactions range from 0.9933 (C···H) to 0.8880 (C···O) in DNA (Figure 5.4a) and 0.9921(C···O) to 0.8628 (C···C) in RNA

(Figure 5.4b). Due to the small density range for the C···C interaction in DNA a correlation was not found, yet the RNA data demonstrate that a correlation exists for C···C densities over a wider range.

The ranges observed in the relationship between ρ_π and the distance between the different types of interacting nuclei reflect the effects of sequence and overall structure of the oligonucleotide on these stabilizing interactions, as will be shown below. There is also a larger variation in the types and strengths of stacking densities in DNA compared to RNA, e.g., N···H, C···H and C-H···π interactions are present in DNA between stacked AT/AT and AT/GC base pairs but not for the related bases in RNA. The information obtained from this analysis, namely which π-interactions under the specific geometric constraints imposed by the sequence contribute most towards the stability, can be useful in the design of sequence-specific intercalating molecules.

In the following chapters, specific examples will be provided for how the effects of the structure and sequence of oligonucleotide can be observed through changes in the electron density distribution of its base pairs.

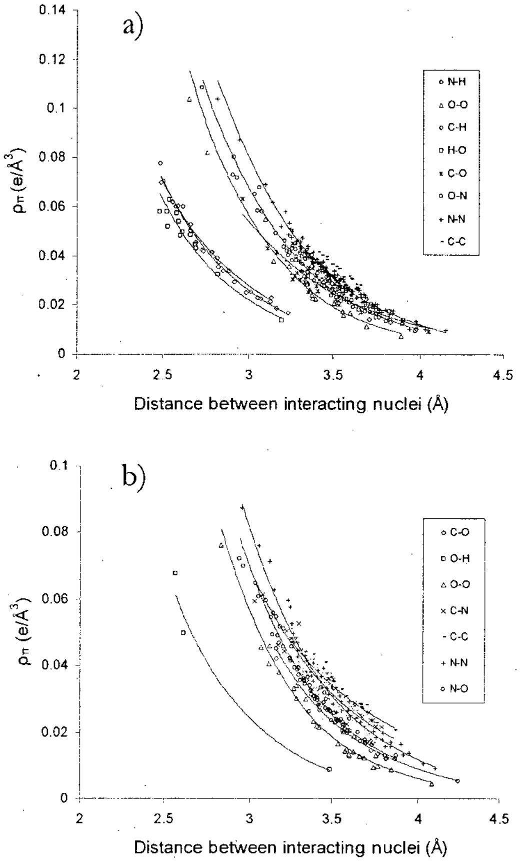


Figure 5.4 Correlation between the distance (Å) and ϱ_π(e/Å^3) at each bond critical point for all interactions between stacked nucleobases from a) DNA and b) RNA duplexes.

5.4 Hydrogen bonding

Most computational studies on the H-bonding in nucleic acid base pairs have focused on the structure and energetics of the global minimum on the potential energy surface of the H-bonded bases. These studies typically involve a full geometry optimization of the base pairs in the gas phase.[44,119,124] While selected studies incorporate some of the effects brought about by the surrounding oligonucleotide and solvent,[117,142] the majority have been conducted in the gas phase in the absence of other base pairs or the sugar-phosphate backbone,[64,124] and as such they do not provide information on how the structure and sequence of an oligonucleotide affect the strength of the H-bonds. Yet this information is crucial, as it is well known that base pairs in an oligonucleotide in general do not possess geometries that are close to the minimum energy structure on their potential energy surface, in particular they are rarely planar when incorporated into an oligonucleotide for example (T13A8 Figure 5.5b).[11,117]

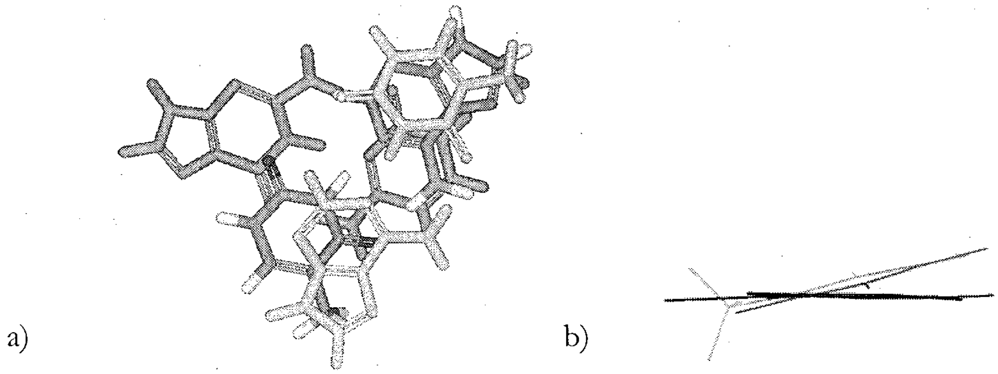


Figure 5.5. Selected base pairs from 126D. a) T13A8 (center) with A12T9 (thick yellow lines) and G14C7 (purple thick lines) on the bottom b) Side view of T13A8.

In fact, until the work of Bickelhaput and co-workers[137] there had been significant disagreement between theory and experiment regarding the H-bond lengths in Watson-Crick pairs, a problem also recently addressed by Dannenberg and co-workers, [143] that was attributed to the absence of experimental conditions in the gas-phase optimizations.[137]

It is shown here that using experimental geometries alleviates the need to simulate the environment and means that even slight differences in the strengths of each particular H-bonding interaction, which are the consequence of sequence and structural effects, are quantifiable using the methodology presented. Bar charts of the densities at the H-bond critical points from all isolated base pairs are compiled by base pair type in Figures S5.1-S5.4 of Appendix D.

5.4.1 AT versus AU base pairs

To date, two different views exist on the relative strengths of H-bonds in AU versus AT base pairs. The view of equally strong H-bonds in AU and AT base pairs is supported by the H-bond lengths as determined from the highest resolved X-ray crystal structures, which are identical within experimental uncertainty.[144] Differences in H-bond strength are also not identified from molecular dynamics studies,[145] and similarly, only small differences in total energy were found with the use of density functionals ($\Delta E_{(AT-AU)}$ = 0.08kcal mol^{-1}).[137,142] However, experimentally it has been shown through the $^{1}J_{NH}$ coupling constants that the N1···N3 H-bonds are stronger in AU than in AT bonds.[116] While the particular systems were not identified, it was stated that the strongest N-H···N interactions occur between polypyrimidine : polypurine tracts.[116] [146]

A comparison of the average the $\sum\varrho_{HB}$ in AT and AU pairs of all the systems in Table 5.1 does not reveal a significant difference, with $\Delta\sum\varrho_{HB}$ (AT-AU) = 0.002 e/Å^3. Accordingly, the optimized base pairs show only a small difference of 0.006 e/Å^3 in the same order as the isolated base pairs. However, Figure 5.6, which shows excerpts from Figures S5.1-S5.4 of Appendix D illustrates that a particular choice of DNA and RNA in an experimental study can easily lead to the conclusion that N-H···N interactions in AU base pairs are stronger than those in AT base pairs (average ϱ_{HB} N-H···N, 0.400 e/Å^3 RNA 420D versus 0.304 e/Å^3 DNA 1SK5, for example). Alternatively, another set of nucleic acids e.g RNA 485D and DNA 1VJ4, would lead to the opposite conclusion (average ϱ_{HB} N-H···N, 0.286 e/Å^3 RNA 485D versus 0.34 e/Å^3 DNA 1VJ4.

Only a large and diverse test set, such as the one chosen here, allows the conclusion that H-bonds in AU and AT base pairs have, on average, the same strength. Between N1-H··N3 interactions in polypyrimidine:polypurine tracts such as in 1SK5 and 1RNA for example ϱ_{HB} is similar to that in tandem AT base pairs in a sequence, however, there is an increase in ϱ_{HB} compared to that in isolated AT base pairs, such as those found in 440D (Figure S5.1 and S5.2 Appendix D).

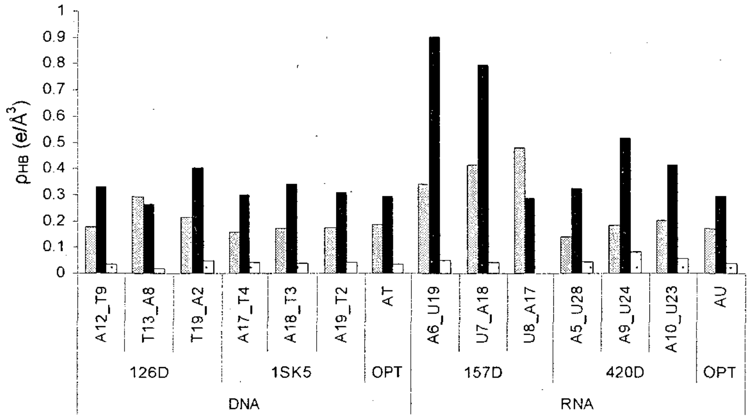


Figure 5.6 Bar chart showing ϱHB from ▨N6-H···O4,■N1-H···N3 and ☐C2-H···O2 in AT and AU base pairs isolated from four duplexes as well as from optimized AT and AU base pairs.

A study on the cooperativity of individual H-bonds in base pairs found that planarity in AT/U base pairs tends to weaken the C2-H···O2 interaction.[143] Since the base pairs in oligonucleotides are rarely planar due to structural restraints caused by neighboring base pairs and the sugar-phosphate backbone[11,117] it is not surprising to find base pair morphologies favoring C2-H···O2 bonds that significantly contribute towards the stability of a duplex.

Figure 5.6 also shows that the relative strengths of the H-bonds can be altered, as is evident in the values for T13A8 from 126D and U8A17 from 157D. The H-bonding in the T13A8 base pair is affected by its geometry in 126D, this evident in the degree of stacking with its nearest neighbors. T13A8 does not stack well on its 3'-side with G14C7 ($\sum\varrho_{\pi}$ = 0.263 e/Å^3) compared to its 5'- side with A12T9 ($\sum\varrho_{\pi}$ = 0.351 e/Å^3). Reasons are a large slide of A8 in the T13A8 base pair on the 3'- side (Figure 5.5a) in addition A8 has a very high propeller twist (Figure 5.5b) that causes an increase in the strength of the N6-H···O4 at the cost of the C2-H···O2 because T13 does not have the same high propeller twist. (Definitions for the relative displacements of base pairs are illustrated in Appendix A) In 157D, the change in the H-bonding pattern is most attributed to the presence of a mis-matched A•G base pair. Its geometry, when incorporated into a regular Watson-Crick duplex, causes significant widening (~2-3Å) that pulls its 5'- side AU base pair apart in the major groove and compresses it in the minor groove, causing the loss of the C2-H···O2 and strengthening of the N6-H···O4 and N1···H-N3 interactions (Figure 5.7).

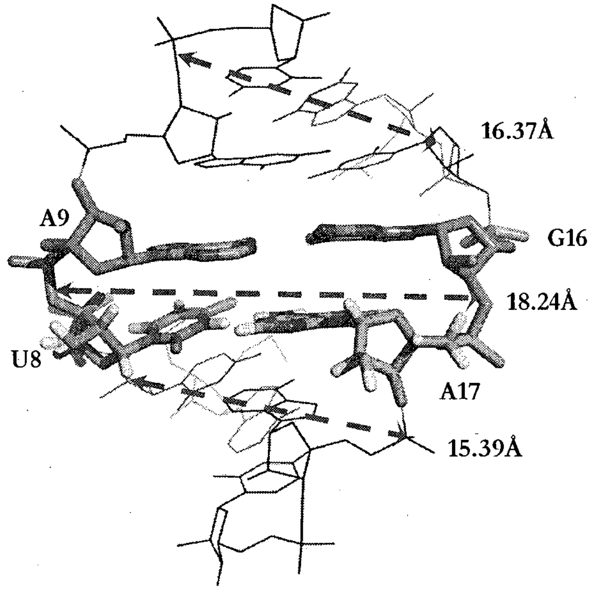


Figure 5.7 Selected base pairs from 157D, showing widening of major groove due to presence of A•G mismatch in duplex.

5.4.2 The "scissoring effect" in GC base pairs

Figure 5.6 (and Figures S5.1 and S5.2 Appendix D) show a mostly consistent pattern for the H-bond strengths for AT and AU base pairs. The pattern consists of a strong central N1···H-N3 interaction, followed by N6-H···O4 and lastly by C2-H···O2 which is often weaker by a factor of ten. The GC base pairs, do not exhibit a preferred pattern (Figure 5.8 and Figures S5.3 and S5.4) with the N-H···O interactions tending to "scissor" or compete for strength. This illustrates that a planar geometry in GC base pairs does not allow three H-bonds to simultaneously achieve their optimal strengths,[143] and demonstrates the variability of these H-bonds compared to their AT/U counterparts. The H-bonding patterns of the base pairs isolated from NAs therefore provide a means of directly observing which GC base pair within a sequence possesses the stronger N-

H···O interaction in the major versus the minor groove and which might be used to predict targets for potential small molecule binders.

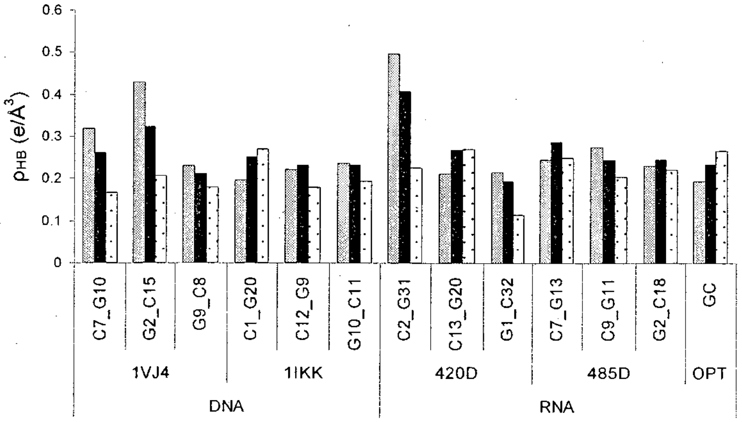


Figure 5.8 Bar chart showing ρ_{HB} from ▨N2-H···O2, ■N1-H···N3 and ☐O6···H-N4 interactions in GC base pairs from four DNA duplexes, as well as from optimized GC base pairs.

5.5 π-Stacking

Molecular graphs for the isolated stacked base pairs reveal bond critical points linked by a network of bond paths between the stacked base pairs that signify the presence of bonding interactions, (Figure 5.1b). Intrastrand and interstrand stacking interactions are identified through the presence of bond critical points and bond paths linking nuclei within the same strand and between nuclei of opposite strands respectively, and where interactions of the O···O, C···O, N···O, N···N, C···H, O···H and N···C type are identified between the stacked base pairs. A subset of the types of interactions found here has been reported earlier.[113]

Interestingly, for the Watson-Crick stacked base pairs from RNA there is a combination of both inter and intrastrand stacking bond critical points compared to

DNA oligomers, where most of the interactions tend to be intrastrand. This is consistent with the structural parameters of base pairs steps in A-RNA and B-DNA helices, which are known to have smaller and larger twist parameters respectively.[111] The distribution in terms of distances as well as values for ϱ_π between the interacting nuclei fall into a similar range for both DNA and RNA, with the bulk of the interactions from the RNA base pairs having a slight shorter range, 3.13-3.82 Å compared to 3.20-4.00 Å in the DNA. (Figure 5.4) It is interesting to note that curves for O···H in RNA and O···H, N···H, C···H and O···O, in DNA all lie below the rest. They therefore fall under shorter interacting distances and are subsequently weaker interactions. (Figure 5.4a) and 5.4b) While the individual values for ϱ_π provide valuable insight into which types of atomic interactions contribute most towards the stability of the stacked pairs as stipulated earlier, it is $\sum\varrho_\pi$ that can provide a clear measure of the actual strengths of the stacking interactions between the base pairs.

5.5.1 AT versus AU π-stacked pairs

Experimental studies on DNA duplexes have shown that the substitution of thymine by uracil weakens the thermodynamic stability of the DNA complex.[147-149] To understand the stabilizing effect of thymine over uracil in a computational study using a modified DFT method referred to as vdW-DF,[117] the potential energy surfaces were mapped for the stacking of planar AT and AU base pairs, keeping rise constant at 3.5Å and varying the twist angle between the stacked base pairs.[117] AA:TT (where AA:TT refers to H-bonded AT stacked over AT, and letters to the left of the colon represent nucleobases on the 5'-3' strand) and AT:AT were found to be more stable than their AU counter parts due to the formation of additional C-H···π interactions from the thymine the methyl group.[117] The TA:TA sequence did not show C-H···π interactions and was found to be of comparable stability with UA:UA.[117] In a related study,[142] the rise and twist were optimized and A-T/T-A pairs (A-T stacked over T-A therefore equivalent to TA:TA) showed additional stability over A-U/U-A which the authors deduced was most likely due to of C-H···π interactions.

Table 5.2 lists the averages of stacking interactions ($\Sigma\varrho_\pi$) for the same sets of base pairs as above obtained from base pairs isolated from the duplexes listed in

Table 5.1. In all three different AT/U sequence combinations interactions between AT stacked pairs are stronger than with AU. The average difference is 0.082 $e/Å^3$ and the largest difference between is AA:TT and AA:UU ($\Delta\sum\varrho_\pi$ 0.144 $e/Å^3$), as reported earlier.[117] The strongest interactions amongst the AT sequences occur in AA:TT, with an average $\sum\varrho_\pi$ of 0.413 $e/Å^3$ and average twist of 38° whereas for AU sequences, UA:UA is strongest with an average $\sum\varrho_\pi$ 0.320 $e/Å^3$ and average twist of 33° (Table 5.2). In agreement with earlier suggestions[117,142] the molecular graphs for the isolated base pairs show additional C-H···π stabilizing interactions in AA:TT and TA:TA.

Table 5.2 Averages for twist (°), rise (Å) and $\Sigma\varrho\pi$ (e/Å^3) for different sequences of π-stacked AT, AU and GC base pairs

	Average twist[a]	Average rise[b]	Average $\Sigma\varrho\pi$
AA:TT	38.86	3.25	0.4131
AA:UU	33.50	3.15	0.2693
AT:AT	31.22	3.19	0.2677
AU:AU	29.69	3.38	0.2216
TA:TA	30.02	3.29	0.3765
UA:UA	32.67	3.23	0.3196
GG:CC	33.63	3.28	0.3214
GC:GC	33.35	3.23	0.3293
CG:CG	35.70	3.50	0.2965

[a,b] Twist and rise values obtained from additional information in the PDB files of the duplexes posted on nucleic acid database.

The methyl group in AT:AT is not involved in stacking interactions (molecular graphs are given in Figures S5.5-S5.7 in Appendix D)and accordingly, the smallest difference in $\sum\varrho_\pi$ occurs between AT:AT and AU:AU stacked pairs ($\Delta\sum\varrho_\pi$ = 0.046 e/Å^3, Table 5.2). The average twist angles are largest for AA:TT and AA:UU, and all are relatively close to their minimum energy values on the potential energy surface determined by Cooper et al.[117,142] Interestingly, the average rises in all sequences are very similar, which shows that the rise is of little importance for the stacking interactions in AT/U pairs, and conclusions drawn from experimental structures with average rises of 3.3Å are in accord with those drawn earlier at 3.5 Å.[142]

5.5.2 GC base pairs in NAs

In contrast to the relative consensus that was reached on stacking in the AT/U sequences, the data on GC stacking in Table 5.2 do not agree with those from a prior study.[117] Table 5.2 shows variability in the twist and $\sum\varrho_\pi$ for AT, but those for the three

GC sequences are very similar. The small range of twist angles and average twist of 36° is at odds with the twist angle ranges of approximately 19° for GC:GC, 35° for CG:CG and 45-55° for GG:CC, when the pairs were again fixed at a rise of 3.5Å. Incidentally, the most stable of these is the CG:CG pair,[117] in accord with the observation on the twist angle from Table 5.2. Thus, GC pairs appear to be more sensitive to the rise than AT/U stacked pairs, which could be attributed to the well known fact that a GC base pair has a higher electrostatic potential compared to an AT/U base pair.[150]

5.6 Conclusions

The isolation of base pairs from experimentally determined molecular structures of oligonucleotides and the subsequent analysis of their electronic structure using the QTAIM allows for an unambiguous analysis of the sequence and structural effects on the weak bonding interactions occurring between NA bases in biologically relevant structural contexts.

The test set of DNA and RNA duplexes used in this study shows a wide but realistic spread in the strengths of the stabilizing interactions. The value added in using the methodology presented here is the direct identification of nuclei that are involved and contribute most towards the stability of the base pairs, this has not (to our knowledge) been done before, particularly for π-stacking interactions.

The density at the H-bonds provides a "blue-prints" of the H-bonding interactions that illustrate where in a particular sequence the H-bonds are weaker or stronger and which areas are more susceptible to weakness. The profiles also reveal consistent patterns in the strengths of H-bonds in AT/U base pairs that are absent in GC base pairs, providing an alternative perspective on the "scissoring effect" of GC base pairs in a duplex and the identification of viable C-H···O interactions in AT/U base pairs.

Previous attempts at understanding the relationship between sequence and stacking interactions required the use of computationally expensive model chemistries and the optimization of step parameters in order to mimic realistic systems. Here the same conclusions can be reached using a simpler methodology that is not limited by the variation of any one particular geometric parameter because experimental geometries are used.

The potential applications of this methodology are far reaching. The next logical step in its validation involves a similar analysis for non-canonical base pairs, and the use of NMR determined molecular structures that account for the dynamism of NA structures in solution.

Concerning the development of the methodology, it would be interesting to investigate the relationship between π-stacking and H-bonding interactions in specific DNA and RNA sequences in terms of the electron density distribution. This is of particular interest as a correlation has been noted in the literature between the length of N-H···N bonds and the strength of stacking interactions, however as it stands, there is no correlation between $\sum\varrho_{\pi}$ and $\sum\varrho_{HB}$.

Chapter 6

An explanation of the thermodynamic behavior of tandem G•U pairs using the electron density

Submitted to:

Journal of Physical Chemistry B

6.1 Introduction

RNA molecules are composed of conserved subunits that define their structure and function. Hydrogen bonds (H-bonds) and π-stacking interactions stabilize these structures allowing the formation of secondary and tertiary structural motifs.[151] Many such motifs feature non-canonical base pairs[152] and of these the G∘U "wobble" is the most common.[153]

Francis Crick discovered the functional importance of G∘U wobble pairs and used their structural flexibility and chemical promiscuity to explain the degeneracy of the genetic code.[154] Since then the significance of G∘U wobbles has been established. Notably through its phylogenetic conservation as the third position of the acceptor helix of nearly all tRNA^{ala},[155] and association with the regulation of the expression of S15 and L30 ribosomal proteins.[156] G∘U wobbles also have been identified as the cleavage site for the Hepatitis delta virus ribozyme.[157] Furthermore, they are known to act as major groove binding sites for fully hydrated divalent metal ions,[156,158] and facilitate the formation of the tertiary structure of the *Tetrahymena* group I intron.[159]

Even though G∘U wobble pairs are nearly isomorphic to Watson-Crick pairs, there are differences between them.[160] The wobbling of the guanine base of a G∘U pair (that is necessary to allow it to H-bond with the uracil)[154] causes the uracil to be pushed into the deep groove (the major groove in DNA; see Figure 6.1, compare a - c) creating a hollow concave surface in the shallow groove (the minor groove in DNA). The nearest neighbors to a G∘U pair also experience unusual stacking that is characterized by significantly more overlap of the U in a G∘U pair towards its 5' -side and the G towards its 3' – side.[161] Finally, the nature of the chemical groups exposed to the deep and shallow grooves also differ from Watson-Crick base pairs (Figure 6.1a - c). These differences ultimately give G○U wobbles a larger variety of structural and functional roles compared to those of other base pairs.

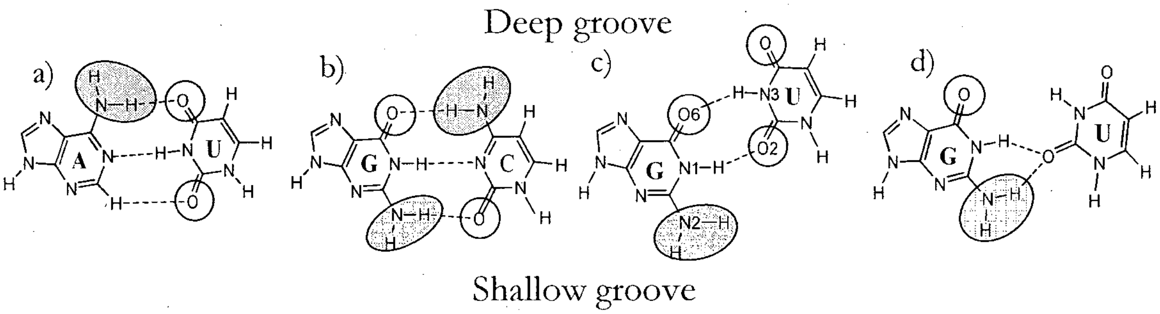


Figure 6.1 Functional groups exposed to the deep and shallow grooves are shown with clear circles identifying H-bond acceptors and grey discs identifying H-bond donors.

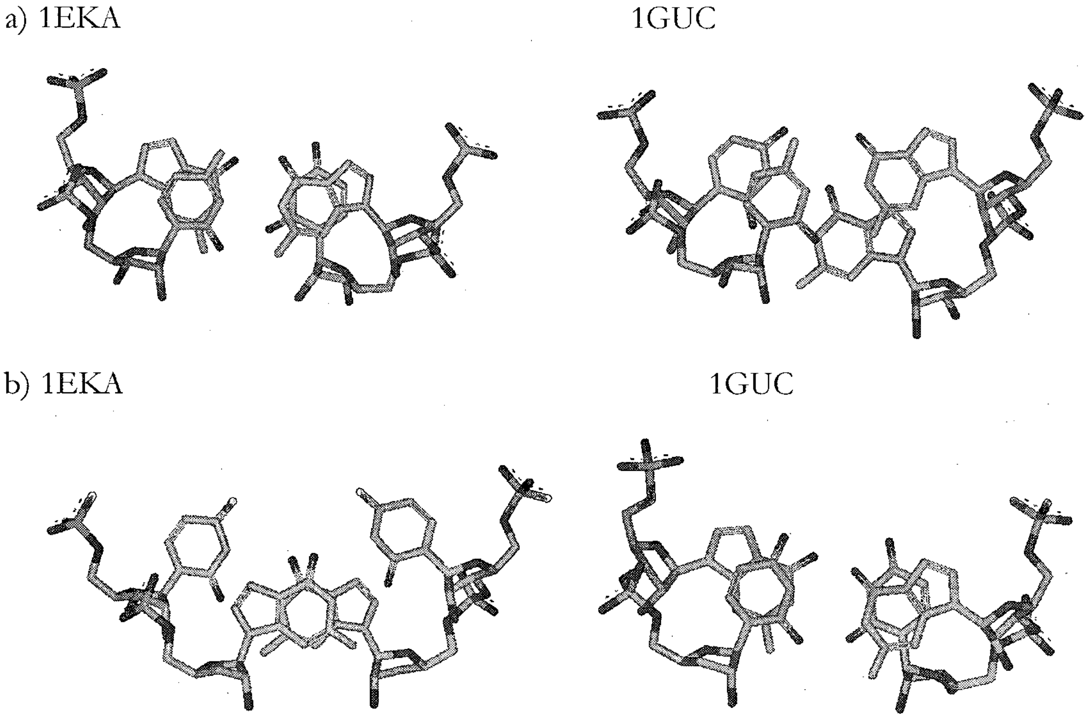
G○U pairs can be either isolated or tandem in an RNA sequence. Leontis and Westhof suggest the use of “○” to refer to the G○U wobble pair and the use of “•” as a generic designation for non-Watson-Crick pairs.[151] As tandem G•U pairs can adopt a variety of geometries including the G○U wobble geometry (Figure 6.1c,d), we will refer to them as G•U pairs as opposed to G○U wobble pairs.

The frequency with which tandem G•U pairs occur in rRNAs is related to two factors; the mismatch sequence, with 5'-UG-3' (motif I) appearing roughly seven times more often than 5'-GU-3' (motif II),[153,162,163] and the nature of the nearest neighbors in the general order 5'G > 5'C > 5'U ≥ 5'A.[153,162,163] Interestingly, for the 5' nearest neighbors C, U and A motif I was found to be on average 1.7 kcal mol^{-1} more thermodynamically stable than motif II, the largest energetic difference observed for the simple reversal of a base pair.[164-169] In fact, the free energy increments ($\Delta G°_{37}$) for motif II sequences were found to be unfavorable for these nearest neighbors; a non-nearest neighbor effect.[165] Collectively, we will refer to this behavior as Situation A.

In Situation B, by contrast, a 5'G nearest neighbor to motif II (5'-GGUC-3') was found to have an almost isoenergetic $\Delta G°_{37}$ when compared to motif I (5'-GUGC-3'). [164-167] However, as sequences containing a 5'G nearest neighbor to motif II rarely occur, this does not contradict the outcome of the phylogenetic diversity studies of tandem G•U pairs. Further study into the sequence and structure relationship in Situation B is none-the-less justified as the 5'-GGUC-3' sequence has been discovered in the P5 helix of the group I intron of *Tetrahymenao thermophilia* [170,171] and in the signal recognition particle RNA from *Humulus japonicus.* [172]

An understanding of the underlying relationship between the thermodynamic stabilities of tandem G•U pairs and the sequence of their nearest neighbors in an RNA duplex is essential because nearest-neighbor interactions are central to structure prediction of RNA duplexes through energy minimization.

To date three inter-related postulates have been used to explain the thermodynamic properties of tandem G•U pairs. The first draws on the unusual π-stacking in the region involving the G•U pairs and their nearest neighbors. [173] In duplexes containing motif I, interstrand overlapping occurs between the guanines of the adjacent G•U pairs while intrastrand overlapping occurs between the tandem G•U pairs and their nearest neighbors. The opposite holds for motif II containing duplexes, (Figure 6.2) [161,165-167,174] which is to the detriment of the strength of the stacking interactions between the nearest neighbors and both G•U pairs, and therefore the thermodynamic stability of the motif II containing duplexes.[173] The second postulate relates to differences in H-bonding geometry and strength between the G•U bases, where those in motif I were thought to have the stronger wobble geometry, and those in motif II were thought to adopt a weaker "chelated" or bifurcated geometry (Figure 6.1c) and d).[101,164,175] Finally, the third postulate suggests that interactions between highly negative electrostatic regions of the guanine bases in G•U pairs in motif II have an unfavorable impact on the thermodynamic stability of the stacked G•U pairs compared to the situation in motif I.[167]


a) 1EKA

1GUC

b) 1EKA

1GUC

Figure 6.2 Base pairs isolated from 1EKA (motif I) and 1GUC (motif II). a) GC above a tandem G•U pair, b) tandem G•U pairs. The sugar-phosphate back bone is shown, hydrogen atoms are omitted.

While all three postulates provide invaluable insight into the thermodynamic behavior of tandem G•U pair sequences, they do not explain the molecular basis behind these observations. This is particularly true of situation B where both motif I and II were found to have isoenergetic ΔG°_{37}. To address this, Pan et al. conducted molecular dynamics simulations on six RNA duplexes containing tandem G•U pairs and focused on the four central base pairs isolated from five snap shots (taken from the first 5ns) of each simulation.[101]

Motif I G•U pairs were found to sample the two-H-bond geometry more often than the one-H-bond (bifurcated) geometry whereas the opposite holds for motif II G•U pairs. This suggested that the thermodynamic stabilities of tandem G•U pairs were predominantly determined by the strength of the H-bonding interactions between the G•U pairs.[101] It was further suggested that stacking interactions between the G•U pairs and between a G•U pair and its nearest neighbors are the driving force for the different H-bond geometries, with weaker H-bonding between G and U bases in motif II only occuring when the nearest neighbors are not involved in G/G interstrand stacking interactions.[101]

In this work, situations A and B are re-visited through an analysis into the changes in the degree of H-bonding and π-stacking in and between motif I and motif II G•U pairs, their nearest neighbors and the base pairs in the rest of the duplex. Specifically, the discussion begins with a general analysis of the H-bonding and π-stacking between the base pairs isolated from all duplexes. This is followed by a comparison of the stabilizing interactions of the G•U base pairs and their nearest neighbors, irrespective of the nearest neighbor sequence, for situation A, and then for each nearest neighbor sequence that has previously been compared in the literature in both situations. This also includes an analysis into the effects of the tandem mismatches on the stability of the base pairs in the rest of the duplexes.

This study uses base pairs isolated from seven of the experimentally determined RNA duplexes that have been previously studied in the literature and have led to the determination of the aforementioned postulates. In addition, two larger RNA complexes 1FFK (2,828 bases and 27 proteins) and 1GID (318 bases) are included in the analysis of situation A (1FFK), and situation B (1GID).

6.2 Methodology and Computational details

Protein Data Bank (PDB) files for all oligomers listed in Table 6.1 were obtained from the Nucleic Acid database,[128] for duplex geometries obtained from X-ray crystallography (resolution of 1.58 Å or better for the smaller duplexes) and NMR spectroscopy. The NMR database entries for 1QET, 1QES and 1GUC are given as a superposition of thirty structures, and we randomly selected ten out of these. Therefore for a given interaction of interest, we report an average from ten structures. We have also included select base pairs isolated from 1FFK and 1GID, where the former corresponds to the crystal structure of the large ribosomal subunit from *Halorcula marismortui* at 2.4Å resolution[176] and the latter corresponds to the crystal structure of the P4-P6 domain of the group 1 *Tetrahymena thermophilia* intron domain determined at 2.5Å.[156] 1FFK and 1GID represent large scale RNA molecules that possess many complex inter-related domains and secondary structural motifs, and the acceptable resolution for 1FFK has been commented on.[176] Data from these RNA structures are included in this analysis to show that the effects of sequence and structure on the stabilizing interactions between base pairs isolated from large oligonucleotides can be captured as well.

Base pairs of interest were isolated by the deletion of all water molecules as well as the surrounding duplex and replacing the sugar-phosphate backbone with a hydrogen atom. It has been shown that the removal of the sugar-phosphate backbone does not affect the relative strengths of the H-bonding and π–stacking interactions between the bases and base pairs.[136]

Table 6.1 Determination, X-ray resolution, number of NMR structures provided, average distances between stacked G•U pairs and references for the RNA duplexes analyzed in this study.

PDB ID	Method	X-ray Res. or # of NMR structures	Sequence	Rise between G•U pairs (Å)	Average $\sum\varrho_\pi$ between G•U pairs (e/Å^3)	$\sum\varrho_\pi$ between G•U pairs and their nearest neighbors (e/Å^3)		Ref.
			Motif I					
1EKA	NMR	1	GAGUGCUC	3.56	0.216	0.345	0.363	167
1QET	NMR	30	GGAUGUCC	2.19	0.507	0.640	0.383	166
315D	X-ray	1.38	GUAUGUAdC	3.04	0.263	0.215	0.216	177
1FFK[a]	X-ray	2.40	GUCUGGAU[b]	2.83	0.295	0.238	0.352	176
			Motif II					
1EKD	NMR	1	GGCGUGCC	3.35	0.310	0.208	0.209	164
1GUC	NMR	30	GAGGUCUC	3.07	0.366	0.339	0.343	167
1QES	NMR	30	GGAGUUCC	3.11	0.432	0.532	0.540	166
332D	X-ray	1.58	GUGUAdC	3.20	0.296	0.203	0.198	177
1GID[e]	X-ray	2.50	GGGUCC[f]	2.83	0.522	0.381	0.359	156

[a,e] 1FFK and 1GID are large RNA oligonucleotides with over one hundred bases.
[b,f] underlined sections from 1FFK and 1GID that were analyzed.

With the sugar-phosphate backbone attached, we observe an increase in the π-stacking and H-bonding densities between the bases, however the relative differences for a set of data with and without the backbone do not change. (manuscript in preparation) As such, the sugar-phosphate backbone is justifiably ignored in these studies. Similarly, the density in the H-bonding region is not affected by the absence or presence of neighboring stacked base pairs.[34]

Using the Gaussian 03 suite of programs,[79] wavefunctions were generated from single point energy calculations to maintain the experimental geometries. We employed the parameter-free hybrid density functional theory (DFT) method PBE0 by Perdew, Burke and Ernzerhof,[83-85] with a Pople double split-valence basis set that included diffuse and polarization functions (6-31+G(d,p)).

The PBE0/6-31+G(d,p) model chemistry is known to characterize H-bonding very well,[138,139] and to perform almost as well as the Becke Half-and-Half (BH&H) functional (as implemented in Gaussian 03) with the same basis set, when reproducing the electronic structure of constrained π-stacked aromatic compounds.[120] This is particularly noteworthy as the BH&H functional is one of a few hybrid-DFT methods that has been shown to quantify π–stacking interactions remarkably well.[87] A comparison of the two model chemistries in the context of this study finds identical performance except for the sum of densities for both H-bonding and π–stacking from BH&H being larger by approximately 0.02e/Å^3 (data not shown).

The electron density was extracted from the many-particle wavefunction, and analyzed within the framework of the quantum theory of Atoms In Molecules (QTAIM)[48] using AIM 2000.[49] AIM provides an approach for the identification of bonding interactions between any two atoms in terms of critical points in the electron density, i.e., points in the topology of the electron density where the first derivative of the density vanishes ($\nabla\rho=0$).[48] Following a topological analysis, atoms and their bonding interactions are displayed as a molecular graph.

6.3 Results and Discussion

6.3.1 Characterization of H-bonding and π–stacking between tandem G•U pairs

It is well established that the value of the electron density (ρ) at a bond critical point located between a pair of atoms serves as a measure of the strength of the interaction where the greater the value of ρ, the greater the strength and the shorter the bond.[48,178,179]

If the density at a H-bond critical point is denoted ρ_{HB}, a collective H-bond strength within a system can be obtained by summing over all ρ_{HB}, $\sum\rho_{HB}$. A representative molecular graph for the bifurcated geometry of the G•U pair is shown in Figure 6.3a where the bond critical points are shown as small red spheres. Figure 6.3c and 6.3d show representative molecular graphs from the bifurcated and three H-bond geometries- which will be discussed in more detail later on.

In general, the calculated values for the electron density at each H-bond critical point lie more or less within the 0.0834 - 0.2480 e/Å^3 calculated range of H-bonding interactions determined at the MP2/6-31G(d,0.25) level, for various geometrically optimized canonical and non-canonical DNA base pairs,[119] and are shown together with $\Sigma\rho_{HB}$ in Table 6.2.

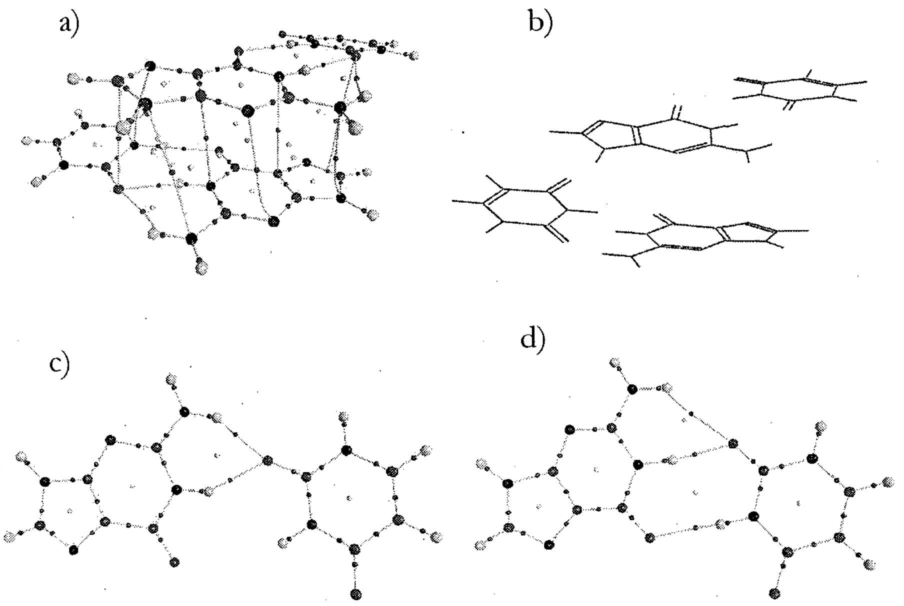


Figure 6.3 Molecular graph of a) G•U4 isolated from 1EKD, showing a 'chelated' geometry, b) U•G4 isolated from 1EKA showing a three H-bond geometry. The nuclei are grey (carbon), blue (nitrogen), red (oxygen) and white (hydrogen). Small red spheres identify bond critical points, the pink lines that link the nuclei to the bond critical points are bond paths, and yellow dots identify ring critical points (3,+1) c) molecular graph for the bifurcated H-bond geometry and d) molecular graph for the 3 H-bond geometry between G•U pairs.

For π–stacking interactions, ϱ_π is determined at a bond critical point between two atoms of two stacked rings, and $\sum\varrho_\pi$ serve as a means of quantifying the strength of inter and intrastrand stacking interactions occurring between the base pairs. Between all stacked G•U pairs studied O···O, C···O, N···O, N···N, and N···C interactions have been identified with a wide spread in the frequency of occurrence in the order 29 N···N, 20 N···C, 16 C···C, 12 N···O, 11 O···O, 9 C···O and 4 O···H interactions. These types of interactions have previously been characterized as stabilizing, closed-shell, van-der-Waals type interactions.[165,179-182]

The values of ρ_π for these interactions tend to be smaller by one order of magnitude compared to the ρ_{HB} values in H-bonded systems.[180] Comparing the scales of the y-axes of Figures 6.4a and 6.4b (ρ_π versus distance) to that of Figure 6.5 (ρ_{HB} versus distance), it is evident that this trend holds for isolated G•U pairs as well.[139]

For various H-bonding and weak van der Waals type interactions it has been shown that ρ displays an exponential correlation as a function of the distance between the interacting nuclei.[140,141] Figures 6.4 and 6.5 show that exponential correlations are also found for both ρ_π and ρ_{HB}, from the isolated G•U pairs, with R^2 values close to 1.00.

To evaluate whether the base pairs from a large macromolecular structure would provide the same results as discussed above, the individual ρ_{HB} and ρ_π as well as the $\Sigma\rho_{HB}$ and $\Sigma\ \rho_\pi$ obtained from base pairs isolated from the large RNA macromolecules 1FFK and 1GID were analyzed. Tables 6.1 and 6.2 show that these data are in the general range of those obtained from the oligonucleotides.

Finally, all interactions are indeed closed-shell as indicated by the positive sign of the second derivative of the electron density ($\nabla^2\rho$) at the bond critical point. (data shown in Tables S6.1-S6.9 in the Appendix E). Also shown in Appendix E are Figures S6.1-S6.6 which are bar charts that show comparisons of the averaged $\sum\rho_{HB}$ and $\sum\rho_\pi$ for all the base pairs in all the duplexes and molecular graphs for all H-bonded G·U pairs are shown in Figures S6.8-6.10.

Table 6.2 Electron density (e/Å^3) for each H-bond in G•U pairs isolated from RNA duplexes containing motif I and motif II, and for an optimized G•U pair (PBE0/6-31+G(d,p)).

		Motif I							
H-bonds	OPT	1EKA		1FFK		1QET[a]		315D	
		UG4	**GU5**	**UG4**	**GU5**	**UG4**	**GU5**	**UG4**	**GU5**
N2-H···O2		0.043	0.048			0.086	0.161		
N1-H···O2	0.258	0.230	0.231	0.232	0.338	0.264	0.264	0.239	0.281
O6···H-N3	0.249	0.148	0.139	0.218	0.281	0.142	0.143	0.214	0.250
$\sum\varrho_{HB}$	0.507	0.420	0.418	0.450	0.619	0.492	0.568	0.453	0.531

	Motif II									
	1EKD		1GID		1GUC[a]		1QES[a]		332D	
	GU4	**UG5**	**GU4**	**UG5**	**GU4**	**UG5**	**GU4**	**UG5**	**GU4**	**UG5**
N2-H···O2	0.134	0.136			0.036[b]		0.050[c]	0.047[c]		
N1-H···O2	0.142	0.139	0.217	0.232	0.156	0.154	0.260	0.256	0.210	0.188
O6···H-N3			0.152	0.195	0.217	0.220	0.129	0.126	0.160	0.197
$\sum\varrho_{HB}$	0.275	0.275	0.368	0.426	0.408	0.374	0.440	0.430	0.370	0.385

[a] Averaged

[b] Only one occurrence in twenty structures.

[c] Avergaed only from those structures where N2-H···O2 was present.

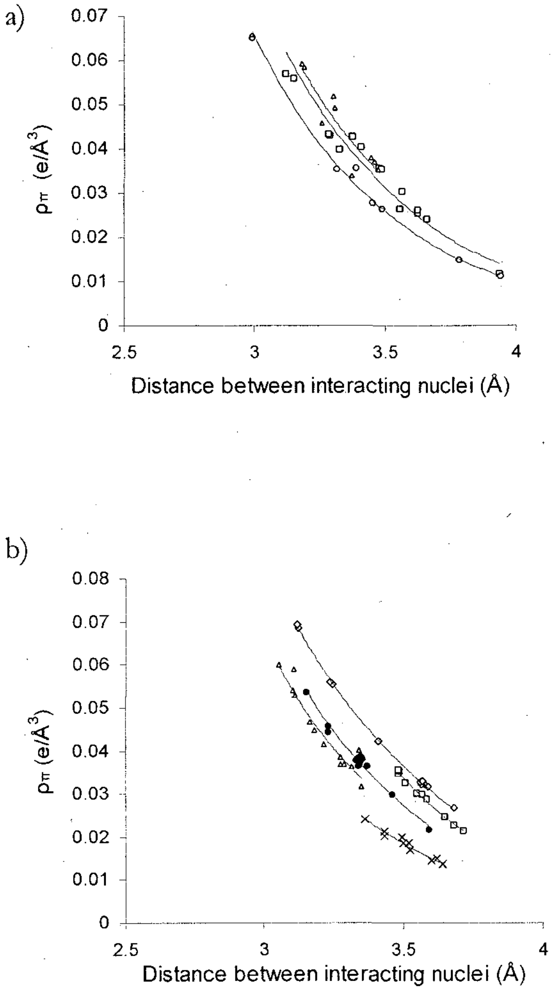


Figure 6.4 Correlations between ϱ_π (e/Å^3) and the distance between the interacting nuclei of stacked G•U pairs isolated from a) motif I and b) motif II containing duplexes □N···N (R^2 motif I = 0.9544 and R^2 motif II = 0.9513), ΔC···N (R^2 motif I = 0.8185 and R^2 motif II = 0.9100), ○C···O (R^2 motif I = 0.9942), ×O···O (R^2 motif II = 0.9921), ◊C···C (R^2 motif II = 0.9973), and •O···N (R^2 motif II = 0.9941).

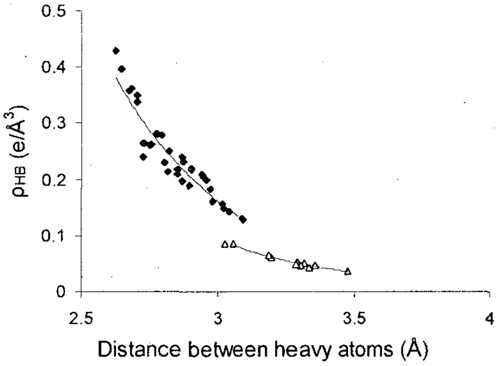


Figure 6.5 Correlation between ϱ_{HB} (e/Å^3) and the heavy atom distance within G•U pairs isolated from RNA duplexes containing motif II and motif I. Δ N2-H···O2, (R^2= 0.9598) ♦ N1-H···O2 O6···H-N3 (R^2 = 0.9162).

6.3.2 Situation A

1QET, 315D and 1FFK, are considered for motif I containing duplexes, and 1EKD, 1QES and 332D, are considered for motif II, Table 6.1. Traditionally, only the relative degree of overlap between the two tandem G•U pairs would serve as a measure of the degree of stacking, with motif II showing more overlap in this region compared to motif I, Figure 6.2a.[174,177,183] Instead, the stacking densities are used and reported as averages for $\Sigma\varrho_\pi$ from the representative duplexes for each motif. The average $\sum\varrho_\pi$ for G•U stacking in motif I and motif II containing duplexes exhibit a difference of 0.008 e/Å^3 (Figure 6.6), with motif I having a slightly higher average (0.355 e/Å^3) than motif II (0.347 e/Å^3). Thus, the stacking of only two bases G/G in motif I is of similar magnitude to the stacking of all four bases in motif II. This is in line with the rise between the G•U pairs (Table 6.1). On average, G•U pairs in a motif I sequence have a much shorter rise (2.69Å) which is very close to the average rise for A-RNA duplexes of

2.80Å,[11] compared to motif II (3.22Å). In addition, calculated interaction energies of stacked nucleic acid bases optimized at the BH&H level have shown that G/G stacked bases are more stable than U/U by 4.45 kcal mol^{-1}, G/G stacking also surpasses C/C by 3.17 kcal mol^{-1} and A/A by 6.50 kcal mol^{-1} stacking[87] which explains the remarkable strength of G/G stacked pairs. Pan et al. calculated the interaction energies of the stacked base pairs using the frozen geometries from the four central base pairs obtained from snap-shots of the molecular dynamics simulations of six RNA duplexes at the MP2 level.[101] Again, stacking interactions between motif I G•U pairs were determined to be stronger than those of motif II.[101] The rationalization was based on favorable intrastrand stacking and strong favorable G/G interstrand stacking between the G•U pairs compared to the G•U pairs in motif II, which showed favorable intrastrand but unfavorable interstrand stacking interactions.[101] In accordance, the molecular graphs reveal both inter and intrastrand π-stacking between the G•U pairs in motif I, whereas between the G•U pairs in motif II the majority of bond paths are intrastrand, and typically involve only one or two interstrand stacking bond critical points that would have negligible contributions towards the stability of the base pairs.

Altogether the results shown here illustrate that $\sum\rho_\pi$ offers an easy, straightforward approach to quantifying stacking interactions that alleviates the ambiguity associated with relying on overlap and correlates very well with previously determined interaction energies.

Based on the observation that the imino protons of the G•U base pairs in 1EKD exchange with water protons at lower temperatures than those in 1EKA, it was proposed that G•U base pairs in motif II possess bifurcated (Figure 6.1d) as opposed to the more stable two H-bonded wobble geometry (Figure 6.1c) of motif I tandem G•U base pairs.[164] This proposition was strengthened through the results from molecular dynamics simulations on six motif I and II containing duplexes that showed that G•U pairs with a motif II sequence sampled the weaker bifurcated H-bonded geometry more often than the wobble geometry, whereas motif I duplexes sampled the wobble more frequently than the bifurcated geometry.[101]

In agreement with these Table 6.2 shows that H-bonding in motif I G•U pairs are on average stronger than that in the motif II sequence, and Figure 6.6 illustrates the significant difference of 0.23 e/Å^3 between the average values for $\sum \rho_{HB}$ in motif I and motif II G•U pairs. However, the molecular graphs of the isolated base pairs do not support the finding that all G•U pairs in a motif II sequence have a bifurcated H-bond geometry (Figure 6.7;and Figures S6.8-S6.10 in Appendix E for all molecular graphs from 1QET, 1GUC, and 1QES). In fact, except for 1EKD, a modified wobble geometry is systematically observed with three H-bonds irrespective of the motif, where the weakest interaction is always N2-H···O2 bond (Figure 6.1c). Only the G•U pairs isolated from 1EKD show the bifurcated geometry (Figure 6.7). Incidentally, the atoms involved in the N2-H···O2 interaction, which are located in the shallow groove, have been linked to a 'ubiquitous' water molecule in certain X-ray crystal structures, for example, those of 315D[177] and 332D.[177] Accordingly, the molecular graphs for the G•U pairs of 332D and 315D do not exhibit the third H-bond.

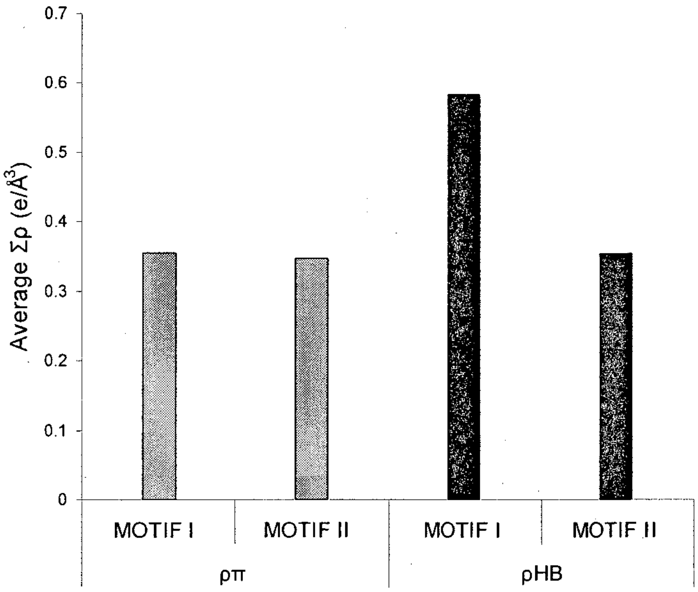


Figure 6.6 Averaged $\Sigma\varrho_{\pi}$ and $\Sigma\varrho_{HB}$ (e/Å^3) for 1EKA, 1QET and 315D (motif I) and 1EKD, 1QES and 332D (motif II).

We therefore propose that the presence of the weak N-H2•••O2 bond in the rest of the G•U pairs shows that in the NMR structures there must be rapid exchange between a water molecule and the N-H2•••O2 H-bond. Only one of the G•U pairs from 1QET has an N-H2•••O2 H-bond that has a value for the ϱ_{HB} (0.161 e/Å^3) that could be considered strong enough to exist in spite of the presence of a water molecule see Table 6.2. The above findings illustrate the advantages of the methodology used here. Clearly, by employing the geometries from experimental structures, H-bond networks can be revealed and analyzed in comparison to geometry optimized base pairs as well as in the light of prior knowledge. The missing N2-H···O2 interaction in the optimized geometry that may be negligible in 1EKA, but certainly not in 1QET (Table 6.2), is evidence against the use of geometry optimizations in these analyses. And it is obvious that the previously held belief that motif II G•U base pairs exhibit a bifurcated H-bond interaction is true for 1EKD but is not supported in general. The molecular graphs for the isolated G•U base pairs are shown in Figure 6.7, with one sample from the ten duplexes for 1QET, 1GUC and 1QES).

It should also be noted that the three H-bond geometry and the presence of water molecules located in the deep and not in the shallow groove have also been reported for the following asymmetric tandem G•U pair sequence 5'-GG-3'.[173]

6.3.3 1EKA and 1EKD (GUGC vs CGUG)

Duplexes containing the sequence 5'-CGUG-3' have been shown to be approximately 3 kcal mol^{-1} less stable than other symmetric tandem G•U pairs with adjacent GC pairs.[163,164,166,167] A difference of 3.4 kcal mol^{-1} in the free energy increment (ΔG°_{37}) associated with the insertion of the G•U pairs to 1EKA and 1EKD[164,167] was linked to the much lower temperature (by 15°C) at which the imino proton resonances for the G•U pairs disappeared in 1EKD compared to those in 1EKA in variable temperature NMR studies. This led to the conclusion that the difference in the thermodynamic stabilities of motif I and II containing duplexes is directly related to the H-bonding between the G•U pairs. Again, motif I was associated with the two H-bond wobble geometry and motif II G•U sequences with the weaker chelated H-bond geometry.[164]

We have compared both $\Sigma\varrho_{\pi}$ and $\Sigma\varrho_{HB}$ from base pairs isolated from 1EKA and 1EKD. Between the G•U pairs, 1EKA shows slightly less stacking density compared to 1EKD, with a difference of 0.094 e/Å^3 which is consistent with the slightly higher rise reported for 1EKA (Table 6.1). Contrary to a prior suggestion that 1EKD is the most stable motif II containing duplex due to the favorable interstrand G/G stacking interactions between the G•U pairs and the GC nearest neighbors,[184] we find that $\Sigma\varrho_{\pi}$ between the nearest neighbors and the G•U pairs is 0.345 e/Å^3 and 0.363 e/Å^3 in the motif I duplex (1EKA) and 0.208 e/Å^3 and 0.209 e/Å^3 in the motif II duplex (1EKD), which is more consistent with previous observations that overlap on the 3'-side of G and 5'-side of U in a G•U pair is more significant than on the 5'-side of G and 3'-side of U.[161] The average difference is 0.15 e/Å^3 which is significant enough to cancel the slight advantage of 1EKD in $\Delta\Sigma\varrho_{\pi}$ between the tandem G•U pairs. However, as proposed from previous NMR studies,[164] the difference between the stability of 1EKA and 1EKD arises primarily from the differences in H-bonding within the G•U pairs (Table 6.2). The large difference between $\Sigma\varrho_{HB}$ (0.838 e/Å^3 in 1EKA compared to 0.5509 e/Å^3 in 1EKD) as well as the significant $\Delta\Sigma\varrho_{\pi}$ between the G•U pairs in motif I and their nearest neighbors reproduce the trend in the thermodynamic stabilities reported experimentally,[164] and is indeed due to the different H-bonding patterns, with 1EKD showing the weaker bifurcated geometry for both G•U pairs.

As the two sequences are not identical, only the GC H-bonds allow further comparison. In agreement with experimental findings,[164] the H-bond strengths are comparable with densities ranging from 0.550-0.573 e/Å^3 for 1EKA and 0.559-0.565 e/Å^3 for 1EKD.

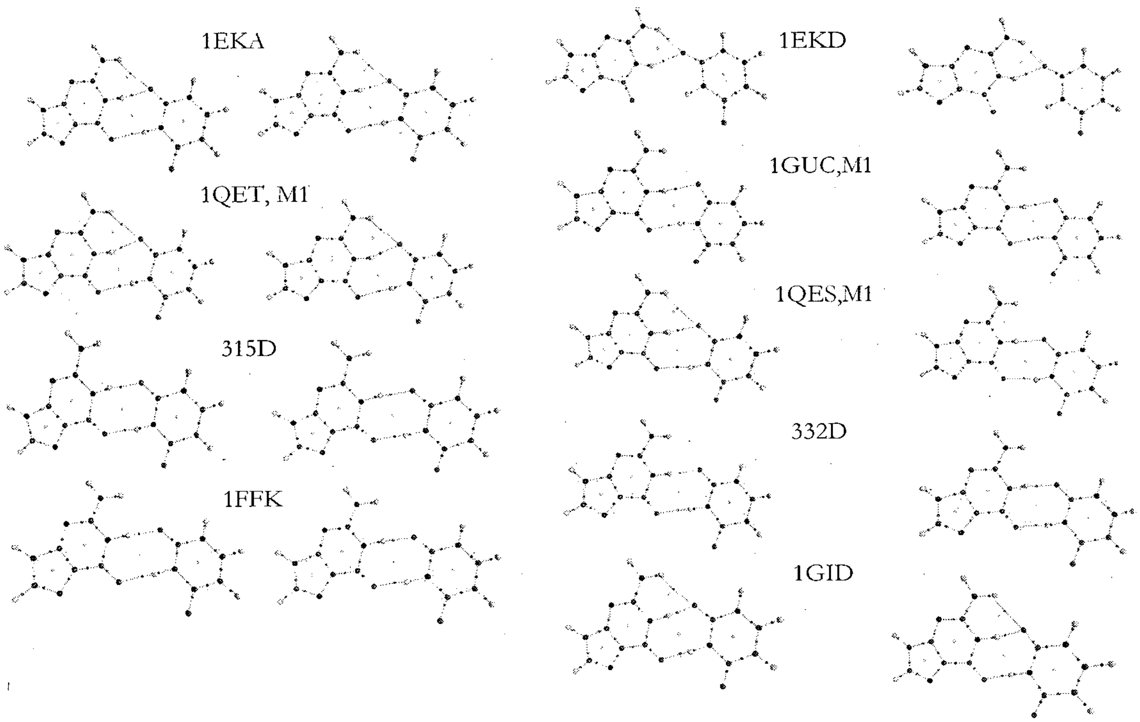


Figure 6.7 Molecular graphs for G•U pairs isolated from the RNA duplexes listed in Table 5.1. M1, (Model 1 of 10 from 1QET, 1GUC and 1QES) is a representative molecular graph.

6.3.4 1QET and 1QES (AUGU vs AGUU)

1QET and 1QES differ in the sequence of their tandem G•U pairs as well as in their average thermodynamic stabilities at 37 °C, where 1QET is 2 kcal mol^{-1} more stable than 1QES.[166] Adjacent to the tandem G•U pairs are AU pairs that were determined to have fragmented and smaller electrostatic potentials compared to GC and G•U pairs and thus were believed to be less involved in the electrostatic contribution to the thermodynamic difference.[166]

Thirty solution structures for both 1QET and 1QES were obtained through NMR spectroscopy and simulated annealing.[166] We randomly selected ten and averaged the data from these ten duplexes for each motif. The difference in $\Sigma\varrho_{\pi}$ between the tandem G•U pairs in 1QET and 1QES is small at 0.075 e/Å^3 (Table 6.1). This is in contrast to the previously suggested difference in the degree of overlap of the negative potentials of the guanines in the G•U pairs and the effect on the thermodynamic stability.[164,167] Similarly, the difference in $\Sigma\varrho_{\pi}$ between the nearest neighbors and the G•U pairs is small, with 0.51 e/Å^{3}in 1QET and 0.53 e/Å^{3}w in 1QES.

As for 1EKA and 1EKD, a significantly larger difference in G•U $\Sigma\varrho_{HB}$ (0.190 e/Å^3) is observed between 1QET and 1QES (Table 6.2) that correlates with the thermodynamic stabilities reported earlier,[166] and was again attributed to a difference in the H-bonding patterns (the two H-bond wobble geometry for 1QET and bifurcated H-bonds for 1QES).[164] Interestingly, all 20 G•U pairs from the 10 1QET structures exhibit three H-bond wobble geometry, as do 14 pairs from the 10 1QES structures. The remaining wobble bifurcated geometry. Bifurcated H-bonds by themselves are not located (Figures S6.8-S6.10 in Appendix E).

The presence of tandem G•U pairs was found to have minimal effects to the overall backbone geometry and stability of the rest of the RNA duplex.[164,166,167,177,185] 1QET and 1QES have identical sequences aside from the motif therefore stabilizing interactions within the rest of the duplex can be compared. Stacking between AU and GC pairs is stronger in 1QES (0.531 e/Å^3) than in 1QET (0.371 e/Å^3). In contrast GC/GC stacking in 1QET has a higher value of the average $\Sigma\varrho_{\pi}$ at 0.584 e/Å^3 compared to that of 1QES (0.484 e/Å^3). Similarly, 1QES has slightly stronger AU H-bonds with $\Sigma\varrho_{HB}$ ranging between 0.475-0.479 e/Å^3 compared to those in 1QET with 0.443-0.445

e/Å^3. However, 1QET has slightly stronger GC H-bonds (0.595-0.602e/Å^3) for 1QET, compared to those in 1QES (0.552-0.612 e/Å^3). These data show that the stabilizing interactions between the Watson–Crick base pairs in 1QET and 1QES are very similar, which is consistent with the differences in Tm for the Watson-Crick pairs of the two duplexes that was determined to be less than 10 °C.[165,166]

To summarize, the data show that the difference in thermodynamic stabilities of 1QET and 1QES correlates with the differences in the G•U H-bonding, where 1QET is shown to exhibit stronger interactions.

6.4 Situation B

In contrast to situation A, the free energies associated with the insertion of G•U pairs in the sequences 5'-GGUC-3' and 5'-GUGC-3' were found to be similar with a $\Delta\Delta G^\circ_{37}$ of approximately 0.4 kcal mol^{-1}.[167] McDowell and Turner determined the solution structures for 1EKA (GAGUGCUC) and 1GUC (GAGGUCUC) using an NMR and simulated annealing protocol, with the aim of gaining insight into the unusual thermodynamics.[167] Cross peak patterns verify that the structures are generally A-form and all nucleotides have anti-glycosidic bonds.[167]

Despite the fact that the thermodynamics in situation B differs from that in situation A, the structures for 1EKA and 1GUC show the same characteristic interstrand G/G overlap for motif I and intrastrand G•U/G•U overlap for motif II.[167] The characteristic overlapping patterns also extend to the nearest neighbors, with more overlap between in 1EKA and less overlap in 1GUC.[167]

Thirty converged structures for 1GUC were reported, and again ten from the thirty were randomly selected and the averaged data is presented. As 1EKA and 1GUC differ only in the sequence of the G•U mismatches, they provide an ideal platform for employing the methodology used above to evaluate situation B comprehensively. $\Sigma\varrho_{\pi}$ between the G•U pairs is larger by 0.150 e/Å^3 in 1GUC, which again correlates with the shorter rise between the G•U pairs in 1GUC (Table 6.1). Interestingly we find that the $\Sigma\varrho_{\pi}$ between the nearest neighbors and the G•U pairs in 1GUC (0.345 e/Å^{3}on both sides of the G•U pairs) is very similar to 1EKA (0.345 e/Å^{3}and 0.363 e/Å^3).

Based on the similarities in the characteristic imino proton chemical shifts and NOE signatures in the NMR data obtained for 1EKA and 1GUC, it was reported that the number of H-bonded imino protons does not affect the thermodynamic stability of the tandem G•U pairs in these systems.[167] Furthermore, both 1EKA and 1GUC sample the two H-bond geometry.[101] These conclusions are confirmed by the data obtained from the AIM study. With the exception of one G•U pair, (M3 in Figure S6.9 of the supporting information in Appendix E) in one of the ten duplexes of 1GUC, all molecular graphs for the G•U pairs isolated from 1GUC have identical (2 H-bonds) patterns, as do those from 1EKA (3 H-bonds). Unlike in situation A, the difference of 0.028 e/Å^3 in $\Sigma\varrho_{HB}$ of the G•U pairs isolated from 1GUC is small. Both 1EKA and 1GUC have strong H-bonds with $\Sigma\varrho_{HB}$ in 1EKA (0.420 e/Å^3) being only slightly larger than that in 1GUC (0.391 e/Å^3). (Table 6.2)

It was previously suggested that the stronger H-bonds between the G•U bases in 1GUC compared to those in other motif II containing duplexes results from weaker interstrand stacking with the G•U nearest neighbors. This weakened stacking allowed for less separation of the G and U bases that otherwise weakens the G•U pairs in other motif II containing duplexes.[101] The molecular graphs for the stacked G•U pairs and their nearest neighbors support this view. The ratio of intrastrand to interstrand bond critical points between the G•U pairs and their nearest neighbors in 1GUC is approximately 5:2 whereas in 1EKD (another motif II containing duplex) it is 2:5. The increased intrastrand stacking interactions, in 1GUC compared to other motif II duplexes promotes the formation of stronger H-bonds which ultimately has led to the stronger H-bonds in 1GUC as observed in 1EKA.

For the rest of the duplexes, data from the Watson-Crick bases in 1EKA and 1GUC show negligible differences in the degree of H-bonding and π–stacking interactions, consistent with experimental findings.[167]

As with situation A, we analyzed base pairs isolated from a large macromolecule with tandem G•U base pairs in situation B: 1GID is the PDB identification number for the crystal structure of a group 1 ribozyme domain determined at 2.4Å. The calculated $\Sigma\varrho_\pi$ between the G•U pairs is very high at 0.522 e/Å^3. Consistent with the data obtained from 1GUC, 1GID exhibits strong π–stacking interactions between the GC/G•U, pairs with values for $\Sigma\varrho_\pi$ reaching 0.381 and 0.359 e/Å^3. As expected, the H-bond patterns for the G•U pairs isolated from 1GID also exhibit a strong three H-bond geometry, in fact the molecular graph shows three H-bonds (Figure 6.7)

6.5 Conclusions

The methodology presented here is capable of reproducing the correlation between thermodynamic stability of tandem G•U pairs and the degree of H-bonding and π-stacking involving the G•U pairs and their nearest neighbors. Specifically, in Situation A the thermodynamic stability of motif I containing duplexes was determined to be higher than that of motif II containing duplexes. The data presented show that motif I, when incorporated in an RNA duplex, favors the formation of strong H-bonds and π-stacking between the G•U pairs and their nearest neighbors. The incorporation of motif II tandem G•U pairs favors the formation of strong π-stacking between the G•U pairs (comparable to that of motif I) at the expense of π-stacking with their nearest neighbors and the strength of the G•U H-bonds. Contrary to previous reports, bifurcated H-bonds are found in only one case. In situation B, the presence of a nearest neighbor G to the G•U pairs, promotes the formation of strong intrastrand π-stacking between the nearest neighbors and the G•U pairs which promotes stronger H-bonds between the G•U pairs, accounting for their comparable thermodynamic stability to motif I duplexes in the same sequence context.

It has also been shown that NMR determined molecular structures as well as large NAs can serve as model systems in the applications of this methodology. This extends the list of viable experimental structures, thus reducing a potential limitation of the methodology.

Finally, the ability to relate an experimental parameter, such as the thermodynamic stability, directly to the degree of H-bonding and π-stacking, as well as to detect the effect of small variations in the structure on the degree of H-bonding and π-stacking individually, suggests that this methodology could potentially be used to investigate host-ligand relationships.

Chapter 7

A stabilization profile for the guanine riboswitch

Submitted to:
Journal of Physical Chemistry B

7.1 Introduction

Riboswitches are mRNA structures that selectively bind target molecules and modulate the expression of genes that are required for the target's metabolism.[186,187] All twenty riboswitch families discovered to date use sophisticated mechanisms that display high levels of variability comparable to protein genetic factors.[186,188] These properties suggest that riboswitches could represent one of the oldest mechanisms for the modulation of gene expression,[188] and lend support to the evolutionary theory that life passed through an "RNA world".[189-192]

The location of the riboswitch on the mRNA, and the nature of the expression platform dictate the method used to exert genetic control.[193,194] In bacteria, most riboswitches are located in the 5'-UTR (un-translated region) of the mRNA, upstream of protein-coding genes related to the metabolism of their target molecules,[195] thus genetic control is often accomplished at the transcription or translation levels.[196]

The purine riboswitch family is one of the most extensively studied model system of riboswitches.[197] The aptamer or ligand binding domain is constructed from three helices, P1 to P3 that are connected through a three-way junction J (where J1/2 would indicate the junction between P1 and P2). The terminal loops L2 and L3 from P2 and P3 respectively, form a series of interconnecting H-bonds which arrange the P2 and P3 helices parallel (Figure 7.1). Cations present between the backbones of the P2 and P3 helices neutralize the repulsive negative charges. Conserved nucleotides in the three-way junction are arranged to define the specific and partially pre-formed binding pocket of the purine riboswitch and stabilize the global helical arrangement of the mRNA.[187,198]

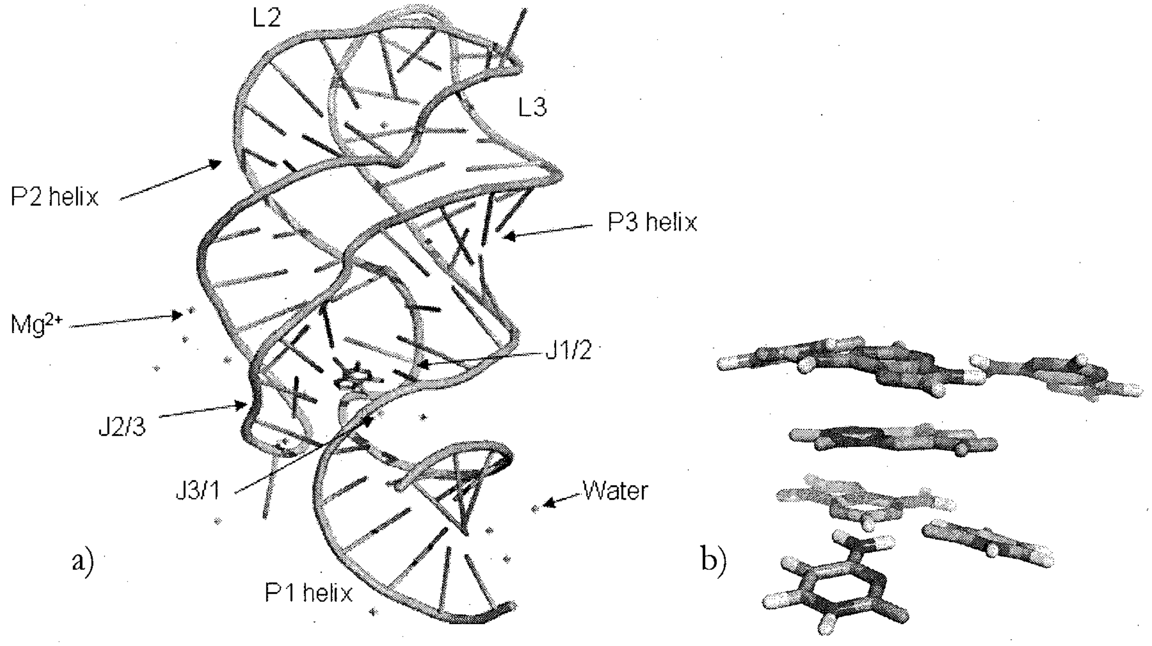
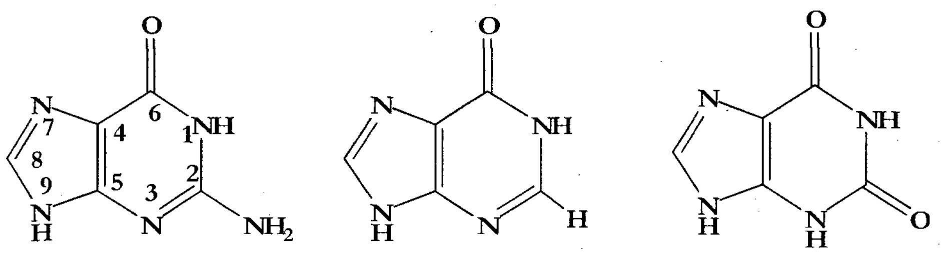


Figure 7.1 a) Crystal structure of guanine (shown as stick model) bound to the *xpt-pbuX* G-box aptamer of the G-riboswitch taken from 1Y27. b) close up of the G-box highlighting base triples directly above and below ligand (green).

The guanine riboswitch (G-riboswitch), a member of the purine riboswitch family, is a *cis*-acting transcription attenuator that works as an OFF switch in response to elevated concentrations of guanine (G), hypoxanthine (HX) and xanthine (X).[187] Upon ligand binding the formation of a terminator over an antiterminator hairpin loop is promoted in the expression platform, which almost inevitably leads to the premature termination of transcription of the cognate genes.[187] Most experiments however were not carried out on the full mRNA that contains both the aptamer and the expression platform but were instead conducted on the short *xpt-pbuX* aptamer that does not contain the expression platform, but binds the ligands nearly as well as the full transcript.[187]

Interestingly, large differences were found in the dissociation constants (K_D) for the binding of guanine (0.004μM) and hypoxanthine (0.76μM)to the *B. subtilis xpt-pbuX* aptamer[199] which were associated with free energies of binding (ΔG) of -12.0 kcal mol^{-1} and -8.5 kcal mol^{-1} respectively.[199,200] The dissociation constant for xanthine was initially reported at 0.05μM,[187] the same value that was found for the binding of HX, both determined for a still shorter 201xpt RNA (a 201 transcript that retains the G-box) at pH 8.5.[187] However, with the xpt-pbuX aptamer, K_D for xanthine is reported as 39μM.[201] Figure 7.1 shows the aptamer bound to G obtained from the PDB file 1Y27;[202] the molecular structures of the metabolites are given in Figure 7.2.

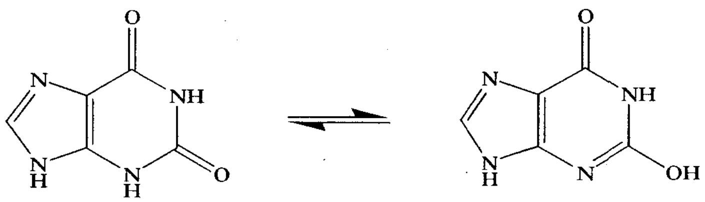
The crystal structures of HX[199] and G[202] bound to the G-riboswitch revealed that the base-triples above and below the ligand that form the floor and the ceiling of the aptamer-pocket (the triples in Figure 7.1b) do not appear to overlap with the ligands.[200] This was rather unusual as base stacking is known to play an important role in ligand recognition and complex stabilization in other RNA aptamers, and led to the suggestion that H-bonding and not π-stacking played the dominant role in RNA-ligand binding.[200] In fact, in-line probing experiments confirmed the importance of the 2'-OH sugar from the U22 sugar, U47, U51 and, most importantly, C74 in forming the essential H-bonds required for specificity and stability of ligand binding to the G-riboswitch.[199,202]



G HX X

Figure 7.2 Molecular structures of guanine (G), hypoxanthine (HX) and xanthine (X).

This binding motif is identical for all three ligands of the guanine riboswitch. Therefore the differences in K_D stem from differences in the strengths of the stabilizing interactions occurring between the bases in the aptamer binding site and the ligands. Sources for the differences are rather apparent upon comparing the three ligands the lack of an N2 amino group in HX and X compromises the strength of the third H-bond with C74 (Figure 7.2),[199,200] and highlights the importance of the substituent at the C2 position.[201] While a C-H···O interaction is still possible with HX, the keto form of X brings two carbonyl oxygens in proximity. Yet, at pH 7.5, a second, tautomeric form (XE) of xanthine exists which allows for O-H···O H-bonding (Figure 7.3). In fact, the crystal structure for X bound to the G-riboswitch was interpreted to show X in the enol form where it can form a third H-bond with the O2 of either C74 or U51,[201] in analogy to the two H-bonds from the NH2 group of guanine (Figure 7.4a)



XK XE

Figure 7.3 Keto-enol tautomerization of xanthine.

Many pathogenic bacteria rely on riboswitches as most regulate the expression of genes required for pathogenicity and survival.[195] This makes riboswitches attractive drug targets,[203] and in fact many riboswitches are largely proposed to be the targets for several known antimicrobial agents.[203,204] In this context, a thorough understanding of the K_D of a target is of utmost importance.

While the participation of C74 and the other bases that H-bond with the ligands in the G-riboswitch is well established, there is yet to be a comparison of the individual stabilization interactions between the ligands and the bases in the binding site of the G-riboswitch. In addition, a rigorous assessment of the stacking interactions in these systems is needed that does not rely on the degree of overlap of the bases.

A methodology that has previously been used to an analysis of the relationship between the sequence, structure and strength of the stabilizing interactions between nucleic acid bases and base pairs[34,35] has been applied to this situation. The pertinent information stems from a topological maps of the electron density distribution that provide detailed pictures of the occurrence and intrinsic properties of weak stabilizing interactions. As the maps are obtained from base pairs that have been isolated from experimentally determined molecular structures, the geometries and subsequently the electron density distributions reflect the consequences of both local and global environment of the base pairs. They also provide a means for unambiguously and quantitatively analyzing the underlying nature of the relationships at hand. A topological map of the electron density creates a stabilization profile, because it provides a blue-print of all stabilizing interactions between the bases, allowing for the separate quantification of each type of interaction.

We present here the stabilization profiles for the three ligands (G, HX and X) bound to the G-riboswitch, with the aims of using the information for the design of novel metabolite analogs that can bind the G-riboswitch.

7.2 Computational Methods

Protein Data Bank (PDB) files for 1Y27[202] and 1U8D[199] were obtained from the Nucleic Acid database,[128] the coordinates for 3GAO[201] were obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (Table 7.1).[205,206]

Using the Gaussian 03 suite of programs,[79] wavefunctions were generated from single point energy calculations to maintain the experimental geometries. The parameter-free hybrid DFT method PBE0 by Perdew, Burke and Ernzerhof,[83-85] with a Pople double split-valence basis set that included diffuse and polarization functions (6-31+G(d,p)). In addition, the three targets G,HX and X (keto and enol forms) were optimized with the same model chemistry for comparison.

Table 7.1 Protein database identification numbers (PDB ID), resolution and references for the three crystal structures of the cognate ligands bound to the *xpt-pbuX* aptamer

Ligand	PDB ID	Resolution (Å)	Reference
G	1Y27	2.40	202
HX	1U8D	1.95	199
X	3GAO	1.90	201

The electron density was extracted from the many-particle wavefunction, and analyzed within the framework of the quantum theory of Atoms In Molecules

(QTAIM)[48] using AIM 2000.[49] QTAIM provides an approach for the identification of bonding interactions between any two atoms in terms of critical points in the electron density, i.e., points in the topology of the electron density where the first derivative of the density vanishes ($\nabla\varrho = 0$).[48] Following a topological analysis, atoms and their bonding interactions are displayed as a molecular graph. Bonding interactions are displayed through bond critical points (small red spheres), ring features through ring critical points (yellow spheres) and cage structures through cage critical points (green spheres).

7.3 Results and Discussion

7.3.1 H-bonding in the binding site of the *xpt-pbuX* aptamer

The G-box in the 5'-UTR of the *B. subtilis xpt-pbuX* mRNA serves as the ligand binding or aptamer domain for guanine and related purines.[187] The *xpt-pbuX* operon encodes genes for xanthine phosphoribosyl transferase and a xanthine transporter, which are essential to purine homeostatis in *B. subtilis*.[207]

The crystal structures of the *xpt-pbuX* guanine-binding domain of the guanine riboswitch (G-riboswitch) of *B. subtilis* bound to HX[187] and G[202] revealed that the highly conserved bases from the J2/3 section of the aptamer form two sets of base triples, above and below the ligand (Figure 6.1b).[199] Of the two base triples that are located on the 3'-side of the pocket one is water mediated and defined by U22-A52·A73† and the

---

† According to Batey et al., the U22-A52·A73 notation denotes Watson-Crick base pairing between U22 and A52 and Hoogsteen pairing between A52 and A73, while the opposite notation (for Watson-Crick base pairs is generally accepted in the literature).

other is defined by A23·G46-G53. In both cases the Watson-Crick edge of the adenine faces the shallow groove. Two base triples (U20-A76·U49 and A21-U75·C50) fasten the J2/3 loop to the P1 helix, completing the binding pocket (Figure 7.1b).[199]

The ligand is bound through a network of hydrogen-bonds (H-bonds) to U22 (its sugar), U47, U51 and C74 which form a base quadruple sitting above the P1 helix.[208] Ligand-induced folding of the three-way junction is necessary in order to encapsulate the ligand completely, as seen in the X-ray crystal structure of the ligand-bound form of the aptamer (Figure 7.1b)[199] Although X binds in exactly the same position as HX and G, (Figures 7.4a, 7.5a and 7.6) the binding of X to the riboswitch is rather complex.[200] X is found as an enolate at a pH greater than 8.5, and below this X exists as either a keto- or an enol tautomer at C2. (Figure 7.3). The enol was proposed to be the bound form, where the hydroxyl group can H-bond with either C74 or U51, denoted XE1 and XE3 respectively (Figure 7.6a).[200]

Interestingly, the PDB file of the crystal structure of X bound to the G-riboswitch, seems to show the keto form with C6-O at 1.227Å and C2-O at 1.242Å. Table 7.2 gives selected bond lengths in the ligands, and it is obvious from comparison with G and HX that the C-O bonds in X belong to carbonyl groups. The calculated enol C-O bonds are, as expected, much longer.[♠] The footnote to Table 7.2 shows the relative energies for the three isomers of X and the keto form is lowest in energy even though it was perceived to be virtually impossible for the keto tautomer of xanthine (XK) to bind.[201]

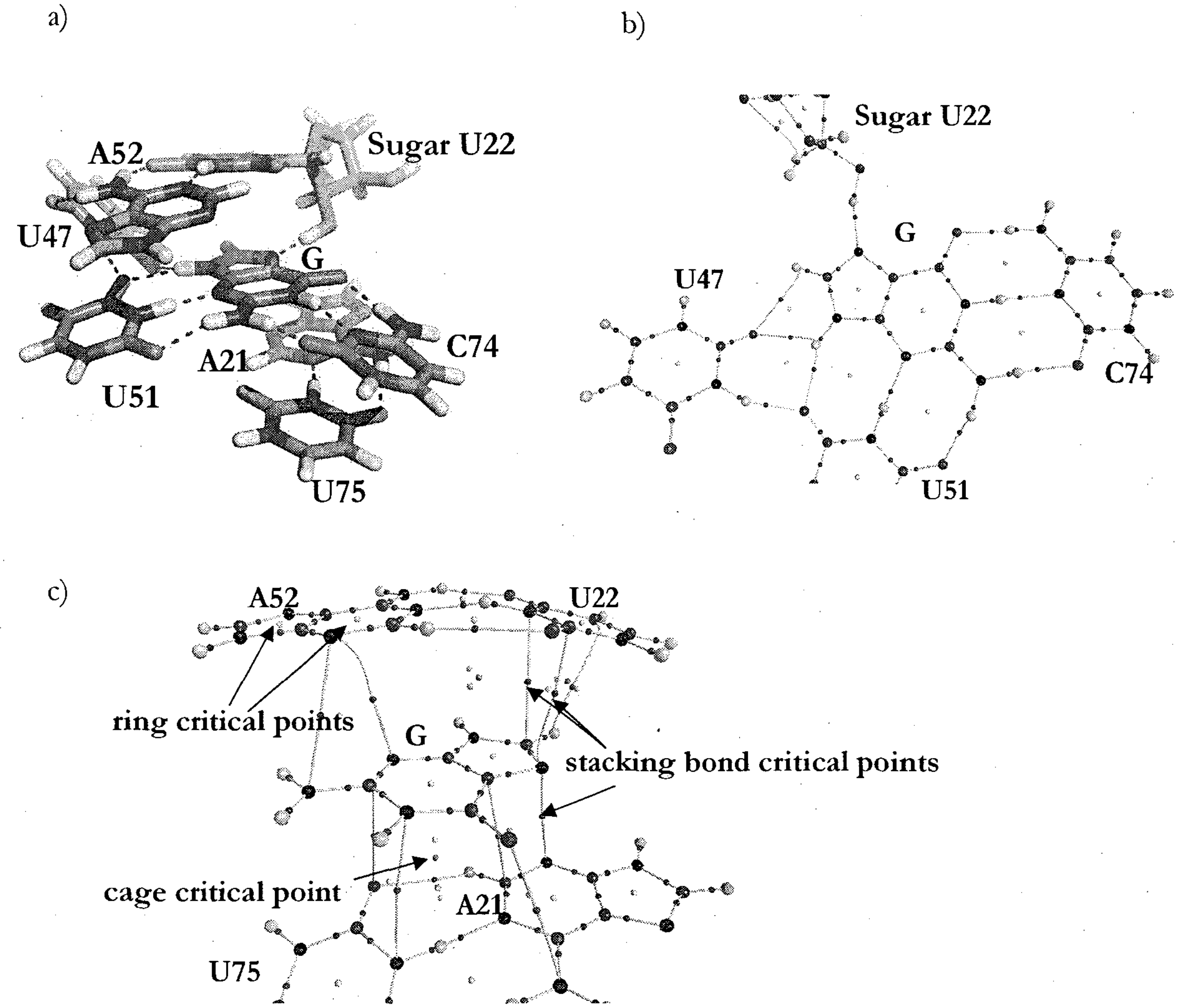


Figure 7.4 a) Stick representation of the bases involved in stabilizing guanine (G) in the binding site from 1Y27. Red dashed lines illustrate the previously predicted H-bonding interactions. [202] b) Molecular graph showing all H-bonding interactions between G and the bases in the binding site, where small red spheres (bond critical points) signify the presence of a bonding interaction. c) Molecular graph showing all stacking interactions between G and the bases in the binding site.

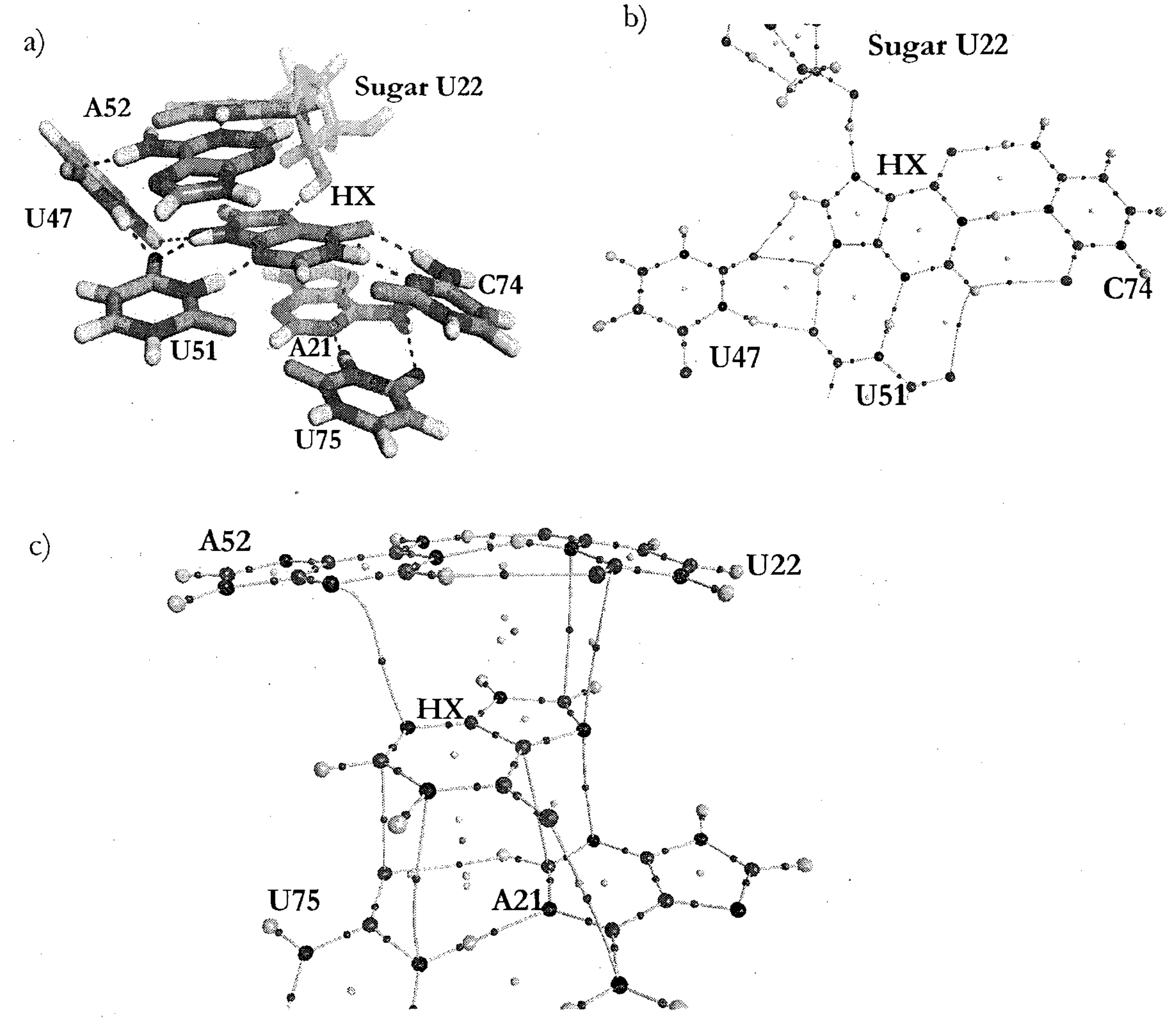
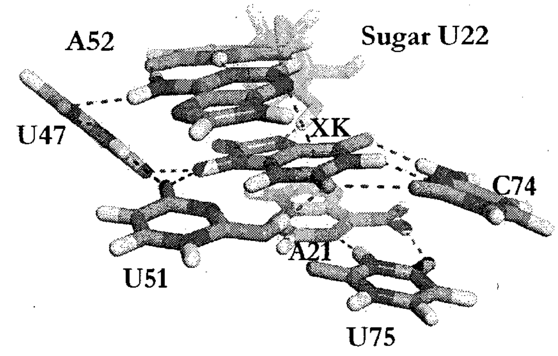


Figure 7.5 a) Stick representation of the bases involved in stabilizing hypoxanthine (HX) in the binding site from 1U8D. Red dashed lines illustrate the previously predicted H-bonding interactions. [199] b) Molecular graph showing all H-bonding interactions between HX and the bases in the binding site, small red spheres (bond critical points) signify the presence of a bonding interaction. c) Molecular graph showing all stacking interactions between HX and the bases in binding site.



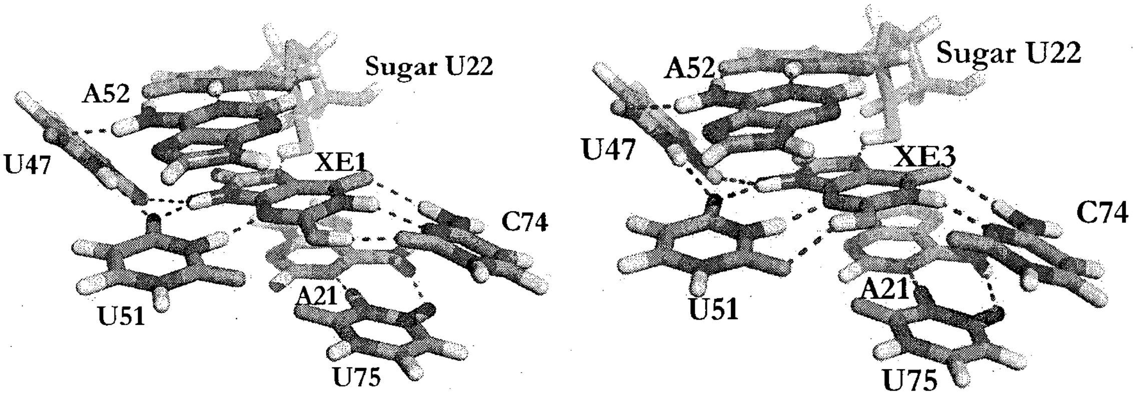


Figure 7.6 Stick representation of the bases involved in stabilizing xanthine (in its keto XK, and enol, XE forms) in the binding site from 3GA0. Red dashed lines illustrate the previously predicted H-bonding interactions for XE1, [193] pymol generated polar contacts. XK bound with enol tautomer of U51.

In the following, all analyses are therefore performed for XK, XE1 and XE3. We modified the X ligand into both XE1 and XE3, and in order to bind XK we modified the U51 to the enol tautomer. In all cases the only changes made were at position C2 of XE1, XE3 and N2 and C2 of U51 for the binding of XK.

Figures 7.4a, 7.5a and 7.6 show the important bases in the binding site of the *xpt-pbuX* aptamer interacting with G, [202] HX [199] and X, respectively. [201] The H-bonding patterns predicted between U22 (its sugar), U47, U51 and C74 and the respective ligands are indicated by red dashed lines. Figures 7.4b, 7.5b and 7.7 show the molecular graphs (stabilization profiles) for the H-bonding interactions. The electron density at the bond critical points for the H-bonds (ϱ_{HB}) provides a tangible and quantifiable means for comparing the strengths of the interactions in all cases.

Table 7.2 Selected bond lengths (Å), for G, HX and X

	C=O			C4-C5			C2-O		
	Opt	Exp.	Δ	Opt	Exp.	Δ	Opt	Exp.	Δ
G[†]	1.217	1.239	0.022	1.392	1.376	0.016			
HX[†]	1.216	1.223	0.007	1.392	1.528[♣]	0.136			
XK[a†‡]	1.213	1.227	0.014	1.379	1.386	0.007	1.217	1.242	0.025
XE1[b†‡]	1.215	1.227	0.012	1.392	1.386	0.006	1.340	1.242	0.098
XE3[c†‡]	1.214	1.227	0.013	1.391	1.386	0.005	1.336	1.242	0.094

[a] Keto tautomer

[b] Enol tautomer with hydroxyl group pointing towards C74

[c] Enol tautomer with hydroxyl group pointing towards U51

[†] $E_{tot\ (au)}$ G = -541.999247 , HX = -486.6832658, XK = -561.8561136, XE1 = -561.8392185, XE3 = -561.8512115

[‡] $E_{rel\ (kcal\ mol-1)}$ XK = 0.0, XE1 = 10.0, XE3 = 3.0

Table 7.3 Sum of density at the H-bond critical points ($\sum\rho_{HB}$) in e/Å^3 for the three ligands bound to the important bases in the binding site of the *xpt-pbuX* aptamer

Ligand	C74	2'-OH Ribose U22	U47	U51	**Total**
G	0.734	0.428	0.097	0.655	**1.914**
HX	0.499	0.394	0.102	0.628	**1.623**
XK[a]	0.399	0.284	0.077	0.388	**1.148**
XE1[b]	0.545	0.279	0.077	0.689	**1.590**
XE3[c]	0.398	0.279	0.077	0.820	**1.574**

[a] Keto tautomer

[b] Enol tautomer with hydroxyl group pointing towards C74

[c] Enol tautomer with hydroxyl group pointing towards U51

[♣] The C4-C5 bond length in the crystal structure of HX appears to be most similar to a single bond at 1.528Å as opposed to a double bond as found in the fully optimized structures. To address this, we compared the geometries of fully optimized HX radical cation, radical anion and full anion, while the bond lengths tended to lengthen none were as long as in the crystal structure (see Appendix F Table S7.1 for a comparison of bond lengths)

Table 7.4 Sum of density at the π-stacking critical points ($\sum\rho_\pi$) in e/Å^3 for the three ligands bound to the important bases in the binding site of the *xpt-pbuX* aptamer

Ligand	A21	A52	U22	U75	**Total**
G	0.123	0.087	0.105	0.054	**0.369**
HX	0.126	0.043	0.085	0.072	**0.327**
XK[a]	0.068	0.046	0.093	0.059	**0.266**
XE1[b]	0.068	0.047	0.093	0.063	**0.271**
XE3[c]	0.068	0.048	0.093	0.063	**0.272**

[a] Keto tautomer
[b] Enol tautomer with hydroxyl group pointing towards C74
[c] Enol tautomer with hydroxyl group pointing towards U51

The values for the sum over all ρ_{HB} from Figures 7.4b, 7.5b and 7.7 are listed in Table 7.3 together with the breakdown of $\Sigma\rho_{HB}$ for the interaction with individual bases. In general, the H-bonding stabilization profiles as depicted in the molecular graphs are very similar. For example, the one bifurcated interaction with U47 has the lowest density value throughout. The overall interaction of U51 with XK is substantially weaker than that for the other ligands, which is a reflection of the necessary enolization of U51.

With the non-modified U51, as found for XE1 and XE3 binding, two H-atoms from two N-H bonds would exhibit a distance of only 0.6Å, which, for a hydrogen atom van der Waals radius of 1.2Å is obviously impossible. As the crystal structure does not show an increased N···N heavy atom distance, the logical conclusion is a necessary enolization of U51, leading to a weaker $\Sigma\rho_{HB}$ for this base interaction. Curiously, the C-O bonds in U51 also have lengths of 1.5Å, marking them as carbonyl bonds. For C74, as expected from a Watson-Crick perspective, the overall strongest interaction is found with G. This also seems to be the main factor for G exhibiting the largest total Σ_{QHB} over all bases.

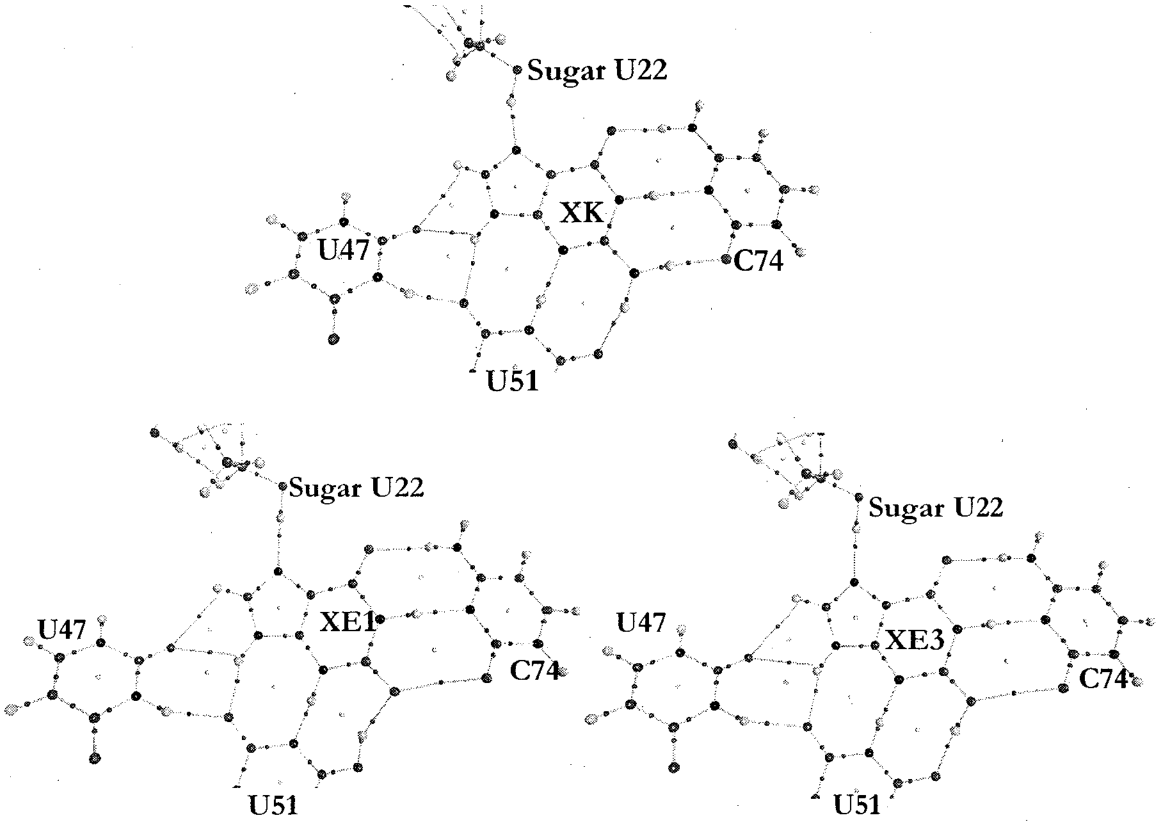


Figure 7.7 Molecular graphs showing the H-bonding interactions between XK, XE1 and XE3 and the bases in the binding site.

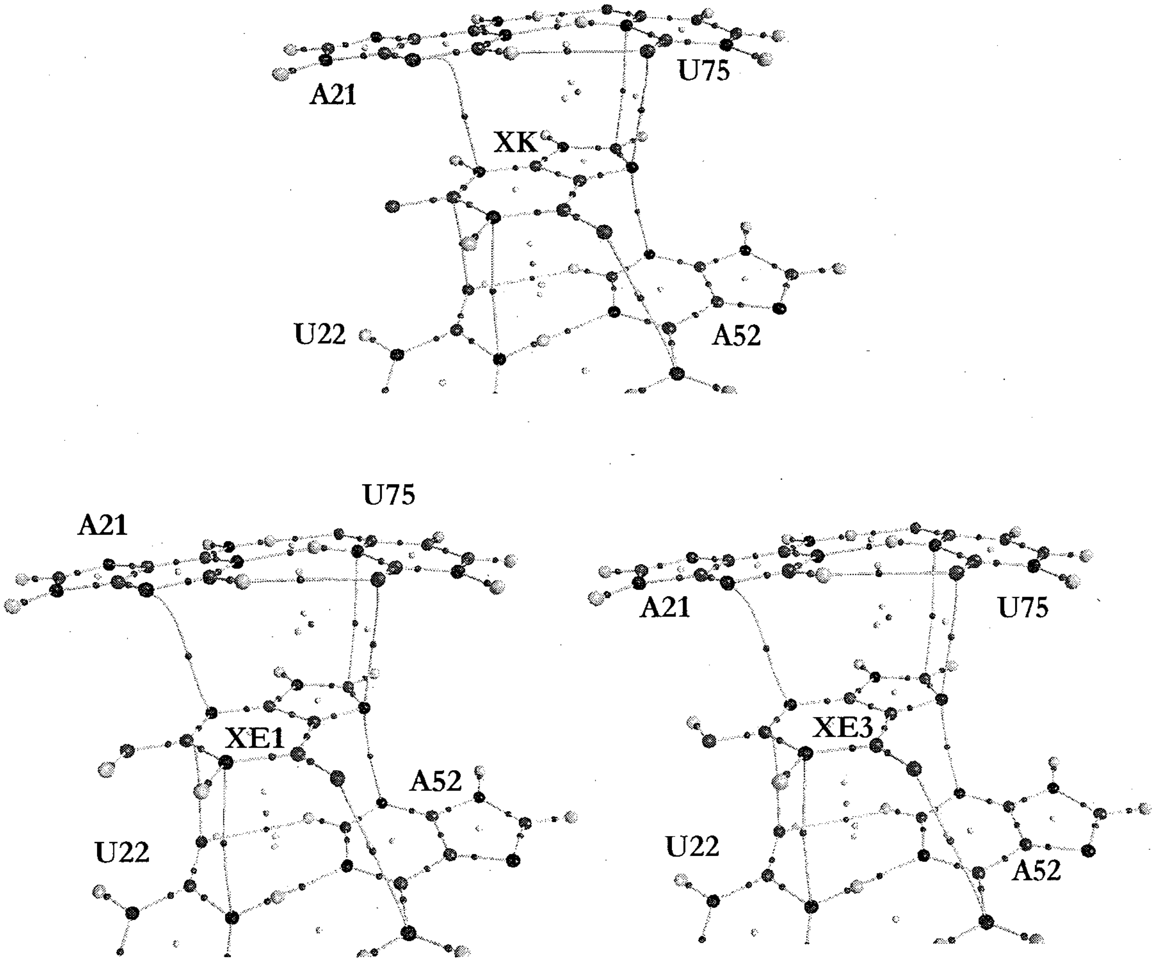


Figure 7.8 Molecular graphs showing the stacking interactions between XK, XE1 and XE3 and the bases in binding site.

As with the dissociation constants,[193] the size of $\Sigma\varrho_{HB}$ decreases in the order G>HX>X irrespective of the nature of X, but binding with XK is weakest from a pure H-bonding analysis. Upon closer inspection, though, $\Sigma\varrho_{HB}$ for HX and the two enol forms is very similar, which in turn is also supported by Kd values from the 201 xpt in-line probing experiments.[187]

Finally, it should be noted that the weak C-H···O bonding interaction between C2 of HX and O4 of Cytosine 74 is clearly identified see Figure 7.5b. The heavy atom distance between C2 and O4 is 3.474 Å and has a ϱ value of 0.308e/Å^3 which is at the previously determined average density for weak H-bonds of 0.317e/Å^3.[209]

7.4 Stacking interactions in the binding site of the *xpt-pbuX* aptamer

Figures 7.4a, 7.5a and 7.6 show that the degree of overlap between the bases above and below the ligands is minimal. However, as demonstrated previously, solely relying on the degree of overlap as a means of gauging the degree of π–stacking can be misleading, and the sum of the stacking densities ($\sum\varrho_\pi$) should be employed instead.[34,35]

Figures 7.4c, 7.5c and 7.8 depict all stacking interactions with the ligands, and it is clear that the bond critical points between the stacked nuclei and subsequently the bond paths linking the ligand and the surrounding bases are few. More importantly, and in contrast to the "no overlap" interpretation,[200] stacking interactions between the ligands and the bases are unambiguously identified. G has the highest number of stacking interactions and also the largest $\sum\varrho_\pi$ (Table 7.4), followed by HX and lastly X. The differences in $\sum\varrho_\pi$ are much smaller than those in $\sum\varrho_{HB}$, and their origin is not as clear cut. As from $\sum\varrho_\pi$ the order in the interaction strength is also G>HX>X, the overall stabilization as given by the sum of $\sum\varrho_\pi$ and $\sum\varrho_\pi$, is still G>HX≥X (for XE) and G>HX≥X for XK.

7.5 Towards the design of novel metabolite analogs

As shown above, both stacking and H-bonding contribute towards the stability of the ligands bound to the G-riboswitch. G shows the strongest H-bonding and π–stacking interactions and subsequently has the lowest K_D value. HX and X have a slightly more complicated story, in that they are either comparable in K_D (0.05μM)[187] or distinctly different in K_D (0.76 μM and 39 μM, respectively)[193] and calculated overall $\sum\rho$ (1.414 e/Å^3 XK, 1.861 e/Å^3 XE1, 1.846 e/Å^{3}XE3). While it seems hardly possible to improve the H-bonding in the G box, one should be bale to improve on the small number of stacking interactions between the ligands and the bases that make up the binding site of the G-riboswitch, and this new information can serve as a good starting point for the design of potential metabolite analogs that can bind and repress the genes encoded by this mRNA. Potential analogs that can take full advantage of the stacking potential in the binding site would need to have a larger π-system, and still maintain the essential Watson-Crick recognition warranted by C74 in the binding site.

Despite the fact that the ligand is completely engulfed in the binding site, it has been shown that binding takes place via a multi-step induced fit mechanism that involves a significant amount of flexibility on the part of the binding site[200] and that the binding site can incorporate larger ligands.[193,201] For example, C74 and, to some extent, U51 have been shown to be flexible enough to accommodate bulky groups at C6 of the ligand.[201] Further evidence of the ability of the riboswitch to accommodate larger ligands comes from the surprising discovery of a new class of purine riboswitch with a virtually identical aptamer binding site as that of the *xpt-pbuX B.subtilis* aptamer, which recognizes 2'-deoxyguanosine.[193] The accommodation of the bulky sugar is achieved through a slight change in conformation of the J2/3 loop.[193]

This is very encouraging for the design of larger ligands. The space located in the binding site between A21, U22 and U47 highlighted with the dashed circle in Figure 7.9a, could accommodate a ligand that has the same Watson-Crick face as guanine, and H-bonding pattern with the other conserved bases in the binding site, but also has a more extensive π–system that would allow for improved stacking between U22 above and A21 below the metabolite analog.

An alternative would be to disrupt the comparatively weak bifurcated H-bond with U47, (see Table 7.2) and build the extension of the π–system towards U51 (which has previously demonstrated conformational maneuverability towards the minor groove) [201] to increase the degree of π–stacking between the ligand and A52, and potentially the H-bonding between U51 and the ligand, (dashed circle in Figure 7.9b). Following these considerations, Figure 7.10 shows the four potential metabolite analogs that could competitively bind the G-riboswitch, and are predicted to have stronger interactions than G.

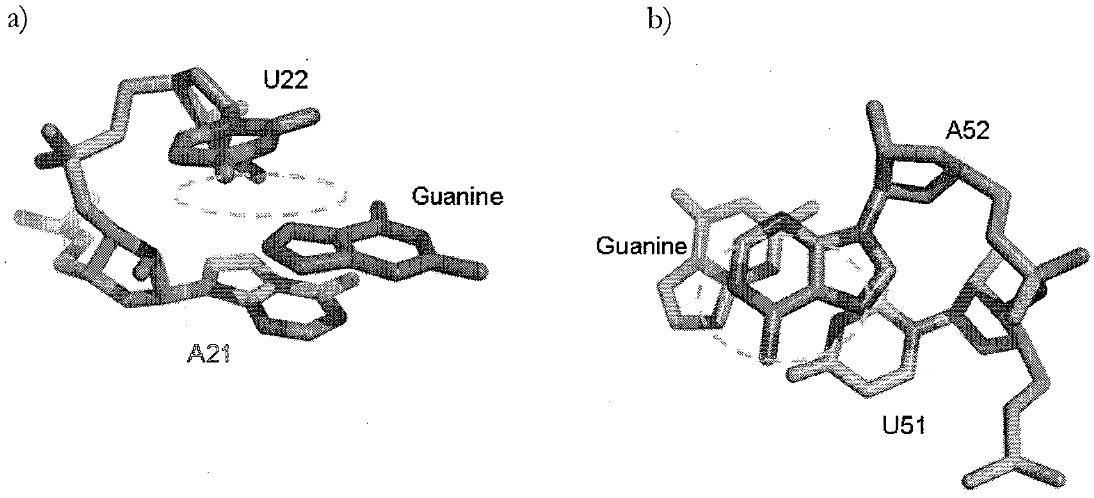
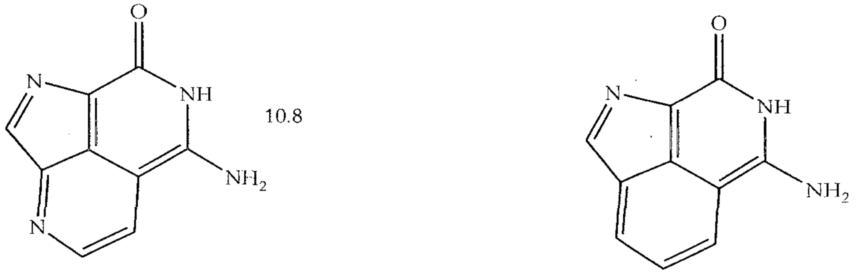


Figure 7.9 a) Stick representation of the first target site for a metabolite analog taken from 1Y27, with the stacking area given with a dashed circle. b) Stick representation of the second target site, with its stacking area circled.

To determine whether the suggested metabolite analogs would exhibit stronger stacking interactions than G in the binding site, the geometries of 2-amino-3*H*-pyrimido-

4a*H*-pyrrolo[3,2-d]pyrimidine-4(9*H*)-one (2APO) and 2-amino-3*H*-pyrimido[4,5-b]indol-4(9*H*)-one (2APIN) were optimized. The structures of the metabolites were then superimposed onto G in the G-box binding site; G was deleted and single point energy calculations were run to generate a wavefunction of the metabolite analog in the binding site. This was possible with 2APO and 2APIN because ligands do not require any additional movements of the bases in the binding site. With 6-aminopyrrolo[2,3,4-de]isoquinolin-8(7*H*)one and 6-aminopyrrolo[2,3,4-de][2,6]napthyridin-8(7*H*)-one, on the other hand, movement of U51 and U22 in the binding site is necessary in order for the analogs to bind without any steric clashing. Unfortunately this can not be checked using the methodology.

2-amino-3H-pyramido-4a*H*-pyrrolo[3,2-d]pyrimidine-4(9*H*)-one **2-amino-3*H*-pyrimido[4,5-*b*]indol-4(9*H*)-one**



6-aminopyrrolo[2,3,4-*de*][2,6]naphthyridin-8(7*H*)-one **6-aminopyrrolo[2,3,4-*de*]isoquinolin-8(7*H*)-one**

Figure 7.10 Structures of potential metabolite analogs for the G-riboswitch.

Figure 7.11 shows the stabilization profiles (molecular graphs) for 2-amino-3*H*-pyrimido-4a*H*-pyrrolo[3,2-d]pyrimidine-4(9*H*)-one (2APO) and 2-amino-3*H*-pyrimido[4,5-b]indol-4(9*H*)-one in the binding site (2APIN). 2APO and 2APIN now show 6 interactions with A52 and U22 (instead of 5 for G) and 7 interactions with A21 and U75 (instead of 5 for G, Figure 6.4c) which result in a significant increase in the degree of π-stacking for 2APO, ($\sum\varrho_{\pi}$ 0.662 e/Å^3) and 2APIN (0.693 e/Å^3) compared to 0.369 e/Å^{3}for G ($\sum\varrho_{\pi}$ 0.369 e/Å^3). Bearing in mind that the difference between G and HX in the *total* $\sum\varrho$ is 0.333 e/Å^3 and assuming that there are no changes to the H-bonding strengths of 2APO and 2APIN in the binding site, these two compounds are very promising metabolite analogs.

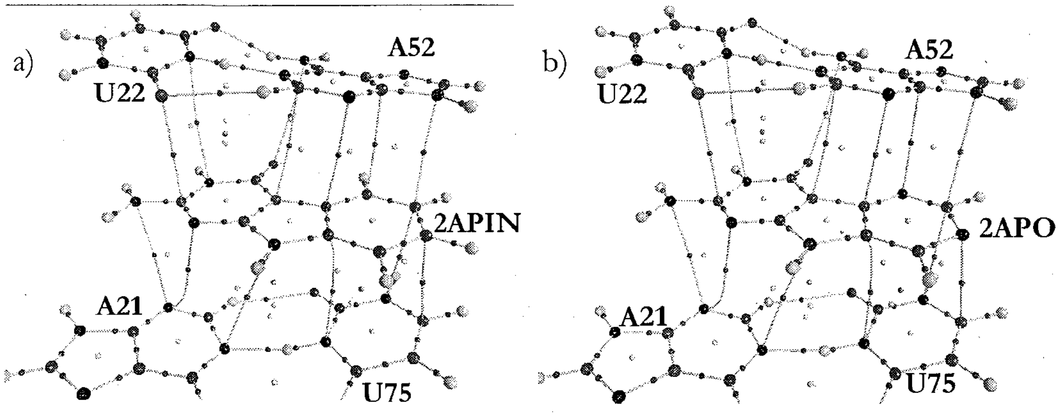


Figure 7.11 Molecular graphs showing the stacking interactions between 2-amino-3H-pyrimido-4a*H*-pyrrolo[3,2-d]pyrimidine-4(9*H*)-one (2APO) and 2-amino-3*H*-pyrimido[4,5-b]indol-4(9*H*)-one in the binding site (2APIN).

7.6 Conclusion

The pattern in experimental K_D for the binding of G, HX and X to the G-box is mirrored in the $\sum\varrho_{HB}$ and $\sum\varrho_{\pi}$. This shows that the methodology presented here is capable of identifying and quantifying all interactions needed for the stabilization of the ligand. The additional information regarding the differences in the degree of π-stacking from one the ligand to another has led to the design of potential metabolite analogs that show stronger stacking interactions in the binding site compared to G.

However, whether the keto or enol tautomer of X is bound to the G-box remains to be clarified. The geometry of X bound to the G-box in the crystal structure more closely resembles the fully optimized XK, and the relative energies of XK, XE1 and XE3 show that both XE1 and XE3 are thermodynamically less stable than XK. The additional anomaly concerning the length of the C4-C5 bond in the crystal structure of HX warrants further investigation.

Chapter 8

Summary, Conclusions and Future work

"DNA sequence is not the be all and end all, shape is primary"

Nick Gilbert, C&E News, May 18, 2009.

The above quote encapsulates one of the most compelling outcomes of the human genome project. The significance of the structure-function relationship of RNAs is well established, but, the possibility that 98% of genes in human DNA could be involved in the formation of functional structures rather than code for thousands of functional proteins was rather surprising. However, fully exploiting the structure-function relationship in NAs necessitates a complete understanding of the fundamental properties of NA structure. This involves gaining insight the relationship between NA sequence, structure and stabilizing interactions.

The methodology presented in this work can be used to do just that. By combining the use of experimentally determined molecular structures, and QC calculations, the methodology encompasses the realism from the experimental geometries, and the insight offered by electronic structure calculations.

This would not be possible if the model chemistry could not correctly characterize H-bonding and π-stacking interactions. In the first results chapter, the performance of a selection of DFT methods (B3LYP, PBE0, BH&H) and MP2 were compared on their ability to reproduce the electronic structure of model systems for stacked and constrained base pairs with a selection of substituted and un-substituted [n.n]paracyclophanes as the models. DFT methods were selected based on their ability to handle relatively large systems, however a careful selection was necessary, as aside from the BH&H method, most are known to fail when characterizing π-stacking interactions in un-constrained aromatic systems. From this study it was found that the PBE0 method performs almost as well as BH&H, and as it is known to characterize H-bonding very well (BH&H is not), it was chosen as the model chemistry for all subsequent studies.

The methodology was then validated against known relationships between NA sequence and the stabilizing interactions, using Watson-Crick base pairs isolated from a selection of DNA and RNA crystal structures. An exponential correlation is known to exist between the interaction energy and the distance of the interacting nuclei. Preliminary results showed that the density in both types of interactions exhibited an exponential correlation with the heavy atom distance, both individually as ρ_{HB} and ρ_{π}, and as the sum ($\sum\rho_{HB}$ and $\sum\rho_{\pi}$). This meant that the quantification of individual interaction strengths as well as the total contribution from either H-bonding or π-stacking in the context of the experimental geometries was possible. Insights into the strength of the H-bonds in AT versus AU pairs as a measure of the electron density helped resolve the issue concerning experimental observations on the relative strengths of AT and AU where it was shown that on average $\sum\rho_{HB}$ is the same, however on a case by case basis, this may differ substantially. Stabilization profiles for the H-bonding between the bases in the duplexes revealed areas where specific H-bonds are comparatively weak, as well as the "scissoring effect" of GC that is not apparent in AT/U base pairs. In addition, these profiles can be used as a means of determining unequivocally whether or not the C-H···O H-bond in AT/U pairs is strong enough to be viable or not. The molecular graphs of isolated and stacked AT/U base pairs revealed that there is indeed a CH···π (that is quantifiable) between stacked AT base pairs in the sequences AA:TT and TA:TA that is not present in stacked AU pairs and that accounts for their greater stability. This analysis also revealed that AT/U stacked base pairs are less susceptible to variations in rise compared to stacked GC pairs, which may be attributable to the higher electrostatic potential in G bases. Together, these data were evidence that the effects of the sequence and structure of the NA were being observed in the isolated base pairs. The apparent effects on the H-bonding in 157D were of particular interest as mis-matches constitute a major portion of RNAs and are known to have numerous important biological functions.

This led to the next chapter that addressed the thermodynamic stabilities of tandem G•U mismatches, where it is known that specific sequences do not follow the nearest neighbor model (a problem when developing folding programs). It was shown that as well as being able to correlate the thermodynamic stabilities of tandem G•U pairs with the degree of H-bonding and stacking, and confirming a rational for the non-nearest neighbor behavior, the ambiguity surrounding the geometry of the H-bonded G•U pairs in motif II is addressed and clarified. In this study, large NAs and NMR determined experimental structures were included as sources for the isolated base pairs. This showed that the methodology is not restricted in the size or source of the systems from which base pairs are isolated. In addition, the results showed that it was sensitive enough to monitor the effects of slight changes in geometry on the degree of H-bonding and π-stacking interactions.

The ultimate test for the methodology would be to show that it can predict an unknown or overlooked property associated with the structure of NAs that could be beneficial. This led to the final chapter where stabilization profiles for the H-bonding and stacking interactions between G, HX, and the keto and enol forms of X bound to the G-riboswitch, clearly identify all the H-bonding and π-stacking interactions occurring between the ligands and the bases in the binding site, permitting a more in-depth comparison of the stabilization of each ligand in the binding site. The profiles also revealed that the presence of weak π-stacking interactions between the natural ligands could be taken advantage of in the design of novel metabolite analogs that could be competitive inhibitors for the G-riboswitch.

Future applications of this methodology can be divided into two trajectories. The first would follow directly from this work. This would involve a more in depth analysis on the nature of the relationship between H-bonding and π-stacking. For example, correlating the ρ_π at N···N interactions between stacked base pairs where the N atoms are also involved in the central N-H···N interaction of the base pairs. This type of detailed information can be useful in the parameterization of folding programs. An analysis into the possible correlation between the electrostatic potential of guanine and the consequences on the stacking densities, with an emphasis on the types of interacting nuclei would also be of interest, and might be able to highlight the source of the vastly different frequencies of occurrence of individual interactions.

Lastly, whether or not the keto or enol forms of HX and X are bound to the G-riboswitch has not completely been resolved. An interesting QC analysis based on the calculation of the activation energies of the keto and enol forms X and HX in the binding site, in order to better understand how easy (or difficult) tautomerization is in the binding would help settle this issue. In addition a re-analysis of the crystal structure of the G-riboswitch bound to the keto form of X with the possibility of enolization of U51 is also recommended.

The second trajectory would involve branching out into host-guest relationships in proteins as there are already numerous crystal structures of proteins. This would be especially useful in cases where the identification of the stabilizing interactions is ambiguous.

References

(1) Bloomfield, V. A. In *Nucleic Acids Structures, Properties, and Functions*; First ed.; Bloomfield, V. A., Crothers, D. M., Tinoco, I. J., Eds.; University Science Books, California: Sausalito, 2000, p 19.

(2) Neidle, S. In *Principles of Nucleic Acid Structure*; First ed.; Elsevier: Amsterdam:Boston, 2008, p 28-31.

(3) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.

(4) Neidle, S. In *Principles of Nucleic Acid Structure*; First ed.; Elsevier: Amsterdamn: Boston, 2008, p 43-67.

(5) Neidle, S. In *Principles of Nucleic Acid Structure*; First ed.; Elsevier: Amsterdam:Boston, 2008, p 88-113.

(6) Turner, D. H. In *Nucleic Acids: Structures, Properties and Functions*; Bloomfield, V. A., Crothers, D. M., Tinoco, I., Jr., Eds.; University Science Books, California: Sausalito, 2000, p 310-311.

(7) Neidle, S. In *Principles of Nucleic Acid structure*; First ed.; Elsevier: Amsterdam:Boston, 2008.

(8) Arthanari, H.; Bolton, P. H. *Chem. Biol.* **2001**, *8*, 221-230.

(9) SantaLucia Jr, J.; Hicks, D. *Ann. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415-C-2.

(10) Neidle, S. In *Principles of Nucleic Acid Structure*; First ed.; Elsevier: Amsterdam:Boston, 2008, p 114-118.

(11) Neidle, S. In *Pricipks of Nucleic Acid Structure*; First ed.; Elsevier: Amsterdam:Boston, 2008, p 204-243.

(12) Pan, T. *Nat. Struct. Biol.* **1998**, *5*, 540-541.

(13) Moine, H.; Ehresmann, B.; Ehresmann, C.; Romby, P. In *RNA Structure and Function.* ; Simons, R. W., Grunberg-Manago, M., Eds.; Cold Spring Harbor Laboratory Press: 1998, p 77-117 pp.

(14) Leontis, N. B.; Westhof, E. *Comparative and Functional Genomics* **2002**, *3*, 518-524.

(15) Neidle, S. In *Principles of Nucleic Acid Structure*; First ed.; Elsevier: Amsterdam:Boston, 2008, p 210.

(16) Xin, Y.; Laing, C.; Leontis, N. B.; Schlick, T. *RNA* **2008**, *14*, 2465-2477.

(17) Batey, R. T.; Rambo, R. P.; Doudna, J. A. *Angew. Chem., Int. Ed.* **1999**, *38*, 2327-2343.

(18) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. *Proc. Nat. Acad. Sci. U.S.A.* **2001**, *98*, 4899-4903.

(19) Kim, S. H.; Suddath, F. L.; Quigley, G. J.; McPherson, A.; Sussman, J. L.; Wang, A. H. J.; Seeman, N. C.; Rich, A. *Science* **1974**, *185*, 435-440.

(20) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678-1685.

(21) Pleij, C. W. A.; Rietveld, K.; Bosch, L. *Nucleic Acids Res.* **1985**, *13*, 1717-31.

(22) Costa, M.; Michel, F. *EMBO* **1995**, *14*, 1276-85.

(23) Sussman, J. L.; Holbrook, S. R.; Warrant, R. W.; Church, G. M.; Kim, S.-H. *J. Mol. Biol.* 1978, *123*, 607-30.

(24) Müller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100*, 143-168.

(25) Grimme, S. *Angew. Chemie Int. Ed.* **2008**, *47*, 3430-3434.

(26) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656-10668.

(27) Šponer, J.; Riley, E. K.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595-2610.

(28) Stone, A. J.; Editor *The Theory of Intermolecular Forces*, International Series of Monographs in Chemistry 32, 1996, Clarendon Press. Oxford University Press. Oxford, New York p272.

(29) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 3734-3739.

(30) Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 503-507.

(31) Amato, I. *Chemical and Engineering News* **2009**, *87*, 47-49.

(32) Neidle, S. In *Principles of Nucleic Acid Structure*; First ed.; Elsevier: Amsterdam:Boston, 2008, p 7-10.

(33) Neidle, S. In *Principles of Nucleic Acid Structure*; First edition ed.; Elsevier: Amsterdam:Boston, 2008, p 2-10.

(34) Kamya, P. R. N.; Muchall, H. M. *To be submitted* **2009**.

(35) Kamya, P. R. N.; Muchall, M. H. *To be submitted* **2009**.

(36) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293-8297.

(37) Pervushin, K.; Ono, A.; FernÃ¡ndez, C. s.; Szyperski, T.; Kainosho, M.; WÃ¼thrich, K. *Proc. Nat. Acad. Sci. . U.S.A.* **1998**, *95*, 14147-14151.

(38) Arter, B. D.; Schmidt, G. P. *Nuc. Aci. Res.* **1976**, *3*, 1437-1447.

(39) Puglisi, E. V.; Puglisi, J. D. In *RNA Struct. Funct.*; Simons, R. W., Grunberg-Manago, M., Eds.; Cold Spring Harbor Laboratory Press: New York, 1998, p 117-146.

(40) Neidle, S. In *Principles of Nucleic Acid Structure*; First Edition ed.; Elsevier: Amsterdam:Boston, 2008, p 11-16.

(41) Jenson, F. *Introduction to computational chemistry*; 2nd edition ed.; John Wiley & Sons Ltd.: West Sussex, 2007.

(42) Šponer, J.; Spacková, N. a. *Methods* **2007**, *43*, 278-290.

(43) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817-3829.

(44) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Mol. Struct.: THEOCHEM* **2001**, *573*, 43-53.

(45) Šponer, J.; Leszczynski, J.; Hobza, P. . *J. Phys. Chem. A* **1997**, *101*, 9489-9495.

(46) Šponer, J.; Hobza, P. *Int. J. Quant. Chem.* **1996**, *57*, 959-970.

(47) Šponer, J.; Leszczynski, J.; Hobza, P. *Biopol.* **2001**, *61*, 3-31.

(48) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Clarendon Press Oxford, 1994.

(49) Biegler-König, F.; Schönbohm, J.; Derdau, R.; Bayles, D.; Bader, R. F. W.; 1.0 ed.; Buro Streibel Biegler-Konig Innovative Software: 1998-2000.

(50) Caramori, G. F.; Galembeck, S. E. *J. Phys. Chem. A* **2007**, *111*, 1705-1712.

(51) Caramori, G. F.; Galembeck, S. E.; Laali, K. K. *J. Org. Chem.* **2005**, *70*, 3242-3250.

(52) Grimme, S. *Chem.Euro.J.* **2004**, *10*, 3423-3429.

(53) Grimme, S.; Antony, J.; Schwabe, T.; Mück-Lichtenfeld, C. *Org. Biomol. Chem.* **2007**, *5*, 741-758.

(54) Habel, M.; Niederalt, C.; Grimme, S.; Nieger, M.; Voegtle, F. *Eur. J. Org. Chem.* **1998**, 1471-1477.

(55) Henseler, D.; Hohlneicher, G. *J. Phys. Chem. A* **1998**, *102*, 10828-10833.

(56) Salcedo, R.; Mireles, N.; Sansores, L. E. *J. Theor. Comput. Chem.* **2003**, *2*, 171-177.

(57) Walden, S. E.; Glatzhofer, D. T. *J. Phys. Chem. A* **1997**, *101*, 8233-8241.

(58) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287-5293.

(59) Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624-1626.

(60) Gonzalez Moa, M. J.; Mandado, M.; Mosquera, R. A. *J. Phys. Chem. A* **2007**, *111*, 1998-2001.

(61) Hobza, P.; Šponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315-25.

(62) Dąbkowska, I.; Gonzalez, H. V.; Jurecka, P.; Hobza, P. *J. Phys. Chem. A* **2005**, *109*, 1131-1136.

(63) Hill, J. G.; Platts, J. A. *J. Chem. Theory Comput.* **2007**, *3*, 80-85.

(64) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247-3276.

(65) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590-6.

(66) Šponer, J.; Riley, K. E.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595-2610.

(67) Grimme, S. In *Rev. Comput. chem.*; Lipkowitz, B. K., Larter, R., Cundari, R. T., Eds.; John Wiley and Sons: Hoboken, 2004; Vol. 20, p 153-211.

(68) Bachorz, R. A.; Bischoff, F. A.; Hoefener, S.; Klopper, W.; Ottiger, P.; Leist, R.; Frey, J. A.; Leutwyler, S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2758-2766.

(69) Goll, E.; Leininger, T.; Manby, F. R.; Mitrushchenkov, A.; Werner, H.-J.; Stoll, H. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3353-3357.

(70) Brown, C. J.; Farthing, A. C. *Nature (London, United Kingdom)* **1949**, *164*, 915-16.

(71) Hope, H.; Bernstein, J.; Trueblood, K. N. *Acta Crystallogr., Sect. B* **1972**, *28*, 1733-43.

(72) Lyssenko, K. A.; Antipin, M. Y.; Antonov, D. Y. *ChemPhysChem* **2003**, *4*, 817-823.

(73) Cram, D. J.; Allinger, N. L.; Steinberg, H. *J. Am. Chem. Soc.* **1954**, *76*, 6132-41.

(74) Frontera, A.; Quinonero, D.; Garau, C.; Costa, A.; Ballester, P.; Deya, P. M. *J. Phys. Chem. A* **2006**, *110*, 5144-5148.

(75) Bräse, S.; Dahmen, S.; Hoefener, S.; Lauterwasser, F.; Kreis, M.; Ziegert, R. E. *Synlett* **2004**, 2647-2669.

(76) Dyson, P. J.; Humphrey, D. G.; McGrady, J. E.; Mingos, D. M. P.; Wilson, D. J. *J. Chem. Soc., Dalton Trans.* **1995**, 4039-44.

(77) Albrecht, M. *Chem. Soc. Rev.* **1998**, *27*, 281-288.

(78) Shibahara, M.; Watanabe, M.; Iwanaga, T.; Matsumoto, T.; Ideta, K.; Shinmyozu, T. *J. Org. Chem.* **2008**, *73*, 4433-4442.

(79) 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, I., Wallingford CT, 2004. 2004.

(80) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-7.

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

(82) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623-11627.

(83) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(84) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Physical Review Letters* **1996**, *77*, 3865-3868.

(85) Perdew, J. P.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982.

(86) Santoro, F.; Barone, V.; Improta, R. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9931-9936.

(87) Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. *J. Comp. Chem.* **2006**, *27*, 491-504.

(88) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(89) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *Chem. Phys. Lett.* **2000**, *319*, 547-554.

(90) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. *J. Chem. Phys.* **1988**, *89*, 2193-218.

(91) Koopmans, T. *Physica* **1934**, *1*, 104-113.

(92) Muchall, H. M.; Werstiuk, N. H. *Can. J. Chem.* **2006**, *84*, 1124-1131.

(93) Zhang, L.; Peslherbe, G. H.; Muchall, H. M. *Photochem. Photobiol.* **2006**, *82*, 324-331.

(94) Bhattacharya, S. *Chem. Phys. Lett.* **2007**, *446*, 199-205.

(95) Dreuw, A.; Head-Gordon, M. *J. Am. Chem. Soc.* **2004**, *126*, 4007-4016.

(96) Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009-4037.

(97) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. *J. Chem. Phys.* **2003**, *119*, 2943.

(98) Gorelsky, S. I.;SWizard program, http://www.sg-chem.net/,CCRI; University of Ottawa, Canada 2008.

(99) Flükiger, P.; Lüthi, H. P.; Portmann, S.; Weber, J. MOLEKEL 4.3; Swiss National Supercomputing Centre CSCS: Manno, Switzerland, 2000-2002
2000-2002.

(100) Portmann, S.; Luthi, H. P. *Chimia* **2000**, *54*, 766-769.

(101) Furo, T.; Mori, T.; Wada, T.; Inoue, Y. *J. Am. Chem. Soc.* **2005**, *127*, 8242-8243.

(102) Staab, H. A.; Krieger, C.; Wahl, P.; Kay, K. Y. *Chem. Ber.* **1987**, *120*, 551-8.

(103) Almenningen, A.; Brunvoll, J.; Popik, M. V.; Sokolkov, S. V.; Vilkov, L. V.; Samdal, S. *J. Mol. Struct.* **1985**, *127*, 85-94.

(104) Bastiansen, O.; Samdal, S. *J. Mol. Struct.* **1985**, *128*, 115-25.

(105) Charbonneau, G. P.; Delugeard, Y. *Acta Crystallogr., Sect. B* **1976**, *B32*, 1420-3.

(106) Gerzain, M.; Buchanan, G. W.; Driega, A. B.; Facey, G. A.; Enright, G.; Kirby, R. A. *J. Chem. Soc., Perkin Trans. 2* **1996**, 2687-2694.

(107) Gantzel, P. K.; Trueblood, K. N. *Acta Cryst.* **1965**, *18*, 958-68.

(108) Anet, F. A. L.; Brown, M. A. *J. Am. Chem. Soc.* **1969**, *91*, 2389-91.

(109) Vögtle, F.; Editor *Topics in Current Chemistry, Vol. 115: Cyclophanes II*, 1983.

(110) Muchall, H. M. In *Modern Cyclophane Chemistry* 2004, p 259-274.

(111) Blackburn, G. M. *Nucleic acids in chemistry and biology*; 2nd ed.; Oxford University Press: Oxford, 1996.

(112) Šponer, J.; Jurecka, P.; Marchan, I.; Luque, F. J.; Orozco, M.; Hobza, P. *Chem. Eur. J.* **2006**, *12*, 2854-2865.

(113) Matta, C. F.; Castillo, N.; Boyd, R. J. *J. Phys. Chem. B* **2006**, *110*, 563-578.

(114) Yildirim, I.; Turner, D. H. *Biochemistry* **2005**, *44*, 13225-13234.

(115) Hobza, P.; Kabelac, M.; Šponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136-1150.

(116) Manalo, M. N.; Perez, L. M.; LiWang, A. *J. Am. Chem. Soc.* **2007**, *129*, 11298-11299.

(117) Cooper, V. R.; Thonhauser, T.; Puzder, A.; Schroder, E.; Lundqvist, B. I.; Langreth, D. C. *J. Am. Chem. Soc.* **2008**, *130*, 1304-1308.

(118) Shishkin, O. V.; Gorb, L.; Hobza, P.; Leszczynski, J. *Int. J. Quantum Chem.* **2000**, *80*, 1116-1124.

(119) Parthasarathi, R.; Amutha, R.; Subramanian, V.; Nair, B. U.; Ramasami, T. *J. Phys. Chem. A* **2004**, *108*, 3817-3828.

(120) Kamya, P. R. N.; Muchall, H. M. *J. Phys. Chem. A* **2008**, *112*, 13691-13698.

(121) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908-6918.

(122) Korth, M.; Grimme, S. *J. Chem. Theory Comput.* **2009**, *5*, 993-1003.

(123) Matta, C. F.; Castillo, N.; Boyd, R. J. *J. Phys. Chem. A* **2005**, *109*, 3669-3681.

(124) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608-15613.

(125) Leonard, G. A.; Hunter, W. N. *J. Mol. Biol.* **1993**, *234*, 198-208.

(126) Shakked, Z.; Rabinovich, D.; Kennard, O.; Cruse, W. B. T.; Salisbury, S. A.; Viswamitra, M. A. *J Mol. Biol.* **1983**, *166*, 183-201.

(127) Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2930-4.

(128) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751-759.

(129) Mack, D. R.; Chiu, T. K.; Dickerson, R. E. *J. Mol. Biol.* **2001**, *312*, 1037-1049.

(130) Gao, Y. G.; Robinson, H.; Wang, A. H. *Eur. J. Biochem.* **1999**, *261*, 413-420.

(131) Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D. *J. Mol. Biol.* **1989**, *209*, 459-474.

(132) Leonard, G. A.; McAuley-Hecht, K. E.; Ebel, S.; Lough, D. M.; Brown, T.; Hunter, W. N. *Structure* **1994**, *2*, 483-494.

(133) Lii, J.-H.; B., M.; Allinger, N. L. *J. Comp. Chem.* **1999**, *20*, 1593-1603.

(134) Masquida, B., Sauter, C., Westhof, E. *RNA* **1999**, *5*, 1384-1395.

(135) Egli, M.; Portmann, S.; Usman, N. *Biochemistry* **1996**, *35*, 8489-8494.

(136) Kurita, N.; Kobayashi, K. *Comp. Chem.* **2000**, *24*, 351-357.

(137) Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117-4128.

(138) Santra, B.; Michaelides, A.; Scheffler, M. *J. Chem. Phys.* **2007**, *127*, 184104/1-184104/9.

(139) Tuma, C.; Daniel, B. A.; Handy, N. C. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3939-3947.

(140) Munshi, P.; Guru Row, T. N. *Cryst. Eng. Comm.* **2005**, *7*, 608-611.

(141) Munshi, P.; Guru Row, T. N. *Acta Crystallogr., Sect. B: Struct. Sci.* **2006**, *B62*, 612-626.

(142) Gil, A.; Branchadell, V.; Bertran, J.; Oliva, A. *J. Phys. Chem. B* **2009**, *113*, 4907-4914.

(143) Asensio, A.; Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6441-6443.

(144) Saenger, W. *Principles of Nucleic Acid Structure, Pt. 1*, Springer-Verlag: New York. 1987.

(145) Perez, A.; Sponer, J.; Jurecka, P.; Hobza, P.; Luque, F. J.; Orozco, M. *Chem. Eur. J.* **2005**, *11*, 5062-6.

(146) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(147) Howard, F. B. *Biopolymers* **2005**, *78*, 221-229.

(148) Ross, P. D.; Howard, F. B. *Biopolymers* **2003**, *68*, 210-222.

(149) Soto, A. M.; Rentzeperis, D.; Shikiya, R.; Alonso, M.; Marky, L. A. *Biochemistry* **2006**, *45*, 3051-3059.

(150) Hunter, C. A. *J. Mol. Biol.* **1993**, *230*, 1025-1054.

(151) Leontis, N. B.; Westhof, E. *RNA* **2001**, *7*, 499-512.

(152) Leontis, N. B.; Lescoute, A.; Westhof, E. *Curr. Opin. Struct. Biol.* **2006**, *16*, 279-287.

(153) Gautheret, D.; Konings, D.; Gutell, R. R. *RNA* **1995**, *1*, 807-14.

(154) Crick, F. H. C. *J. Mol. Biol.* **1966**, *19*, 548-555.

(155) Sprinzl, M.; Horn, C.; Brown, M.; Ioudovitch, A.; Steinberg, S. *Nucleic Acids Res.* **1998**, *26*, 148-153.

(156) Cate, J. H.; Doudna, J. A. *Structure* **1996**, *4*, 1221-1229.

(157) Perrotta, A. T.; Been, M. D. *Nucleic Acids Res.* **1996**, *24*, 1314-21.

(158) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Szewczak, A. A.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1696-1699.

(159) Kieft, J. S.; Tinoco, I. *Structure* **1997**, *5*, 713-721.

(160) Varani, G.; McClain, H. W. *EMBO rep.* **2000**, *1*, 18-23.

(161) Mizuno, H.; Sundaralingam, M. *Nucleic Acids Res.* **1978**, *5*, 4451-61.

(162) He, L.; Kierzek, R.; SantaLucia, J.; Walter, A. E.; Turner, D. H. *Biochemistry* **1991**, *30*, 11124-11132.

(163) Wu, M.; McDowell, J. A.; Turner, D. H. *Biochemistry* **1995**, *34*, 3204-3211.

(164) Chen, X.; McDowell, A. J.; Kierzek, R.; Krugh, R. T.; Turner, H. D. *Biochemistry* **2000**, *39*, 8970-8982.

(165) Boyd, R. J.; Choi, S. C. *Chem. Phys. Lett.* **1986**, *129*, 62-65.

(166) McDowell, J. A.; He, L.; Chen, X.; Turner, D. H. *Biochemistry* **1997**, *36*, 8030-8038.

(167) McDowell, J. A.; Turner, D. H. *Biochemistry* **1996**, *35*, 14077-14089.

(168) Freier, S. M.; Sugimoto, N.; Sinclair, A.; Alkema, D.; Neilson, T.; Kierzek, R.; Caruthers, M. H.; Turner, D. H. *Biochemistry* **1986**, *25*, 3214-3219.

(169) Sugimoto, N.; Kierzek, R.; Turner, D. H. *Biochemistry* **1987**, *26*, 4559-4562.

(170) Cech, T. R.; Damberger, S. H.; Gutell, R. R. *Nat. Struct. Biol.* **1994**, *1*, 273-80.

(171) Michel, F.; Westhof, E. *J. Mol. Biol.* **1990**, *216*, 585-610.

(172) Larsen, N.; Zwieb, C. *Nucleic Acids Res.* **1996**, *24*, 80-1.

(173) Deng, J.; Sundaralingam, M. *Nucleic Acids Res.* **2000**, *28*, 4376-4381.

(174) Jang, S. B.; Hung, L.-W.; Jeong, M. S.; Holbrook, E. L.; Chen, X.; Turner, D. H.; Holbrook, S. R. *Biophys. J.* **2006**, *90*, 4530-4537.

(175) Yang, J.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 9090-9091.

(176) Ban, N. *Science* **2000**, *289*, 905.

(177) Biswas, R.; Sundaralingam, M. *J. Mol. Biol.* **1997**, *270*, 511-519.

(178) Espinosa, E.; Molins, E.; Lecomte, C. *Chem. Phys. Lett.* **1998**, *285*, 170-173.

(179) Grabowski, S. J. *J. Mol. Struct.* **2001**, *562*, 137-143.

(180) Bone, R. G. A.; Bader, R. F. W. *J. Phys. Chem.* **1996**, *100*, 10892-10911.

(181) Boyd, R. J.; Choi, S. C. *Chem. Phys. Lett.* **1985**, *120*, 80-85.

(182) Grabowski, S. J. *J. Phys. Chem. A* **2000**, *104*, 5551-5557.

(183) Utsunomiya, R.; Suto, K.; Balasundaresan, D.; Fukamizu, A.; Kumar, P. K. R.; Mizuno, H. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *D62*, 331-338.

(184) Pan, Y.; Priyakumar, U. D.; MacKerell, A. D. *Biochemistry* **2005**, *44*, 1433-1443.

(185) Biswas, R.; Wahl, M. C.; Ban, C.; Sundaralingam, M. *J. Mol. Biol.* **1997**, *267*, 1149-1156.

(186) Winkler, W. C.; Breaker, R. R. *Chem.Bio.Chem* **2003**, *4*, 1024-1032.

(187) Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R. *Cell* **2003**, *113*, 577-586.

(188) Vitreschak, A. G.; Rodionov, D. A.; Mironov, A. A.; Gelfand, M. S. *Trends in Genetics* **2004**, *20*, 44-50.

(189) Benner, S. A.; Ellington, A. D.; Tauer, A. *Proc. Nat. Acad. Sci. U.S.A.* **1989**, *86*, 7054-7058.

(190) Jadhav, V. R.; Yarus, M. *Biochimie* **2002**, *84*, 877-888.

(191) Jeffares, D. C.; Poole, A. M.; Penny, D. *J. Mol. Evol.* **1998**, *46*, 18-36.

(192) White, H. B., 3rd *J. Mol. Evol.* **1976**, *7*, 101-4.

(193) Edwards, A. L.; Batey, R. T. *J. Mol. Biol.* **2009**, *385*, 938-948.

(194) Montange, R. K.; Batey, R. T. *Ann. Rev. Biophys.* **2008**, *37*, 117-133.

(195) Barrick, J. E.; Breaker, R. R. *Scientific American* **2007**, *296*, 50-57.

(196) Nudler, E.; Mironov, A. S. *Trends in Biochemical Sciences* **2004**, *29*, 11-17.

(197) Kim, J. N.; Roth, A.; Breaker, R. R. *Proc. Nat. Acad. Sci.*U.S.A.**2007**, *104*, 16092-16097.

(198) Barrick, J. E.; Corbino, K. A.; Winkler, W. C.; Nahvi, A.; Mandal, M.; Collins, J.; Lee, M.; Roth, A.; Sudarsan, N.; Jona, I.; Wickiser, J. K.; Breaker, R. R. *Proc. Nat. Acad. Sci.*U.S.A.**2004**, *101*, 6421-6426.

(199) Batey, R. T.; Gilbert, S. D.; Montange, R. K. *Nature* **2004**, *432*, 411-415.

(200) Gilbert, S. D.; Stoddard, C. D.; Wise, S. J.; Batey, R. T. *J. Mol. Biol.***2006**, *359*, 754-768.

(201) Gilbert, S. D.; Reyes, F. E.; Edwards, A. L.; Batey, R. T. *Structure* **2009**, *17*, 857-868.

(202) Serganov, A.; Yuan, Y.-R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J. *Chem. Biol.* **2004**, *11*, 1729-1741.

(203) Blount, K. F.; Breaker, R. R. *Nat. Biotechnol.* **2006**, *24*, 1558-1564.

(204) Sudarsun, N. S.; Cohen-Chalamish, S.; Nakamura, S.; Emilsson, G. M.; Breaker, R. R. *Chem. Biol.* **2005**, *12*, 1325-1335.

(205) Berman, H. M.; Hendrick, K.; Nakamura, H. *Nat. Struct. Biol.* **2003**, *10*, 980.

(206) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nuc. Aci. Res.* **2000**, *28*, 235-242.

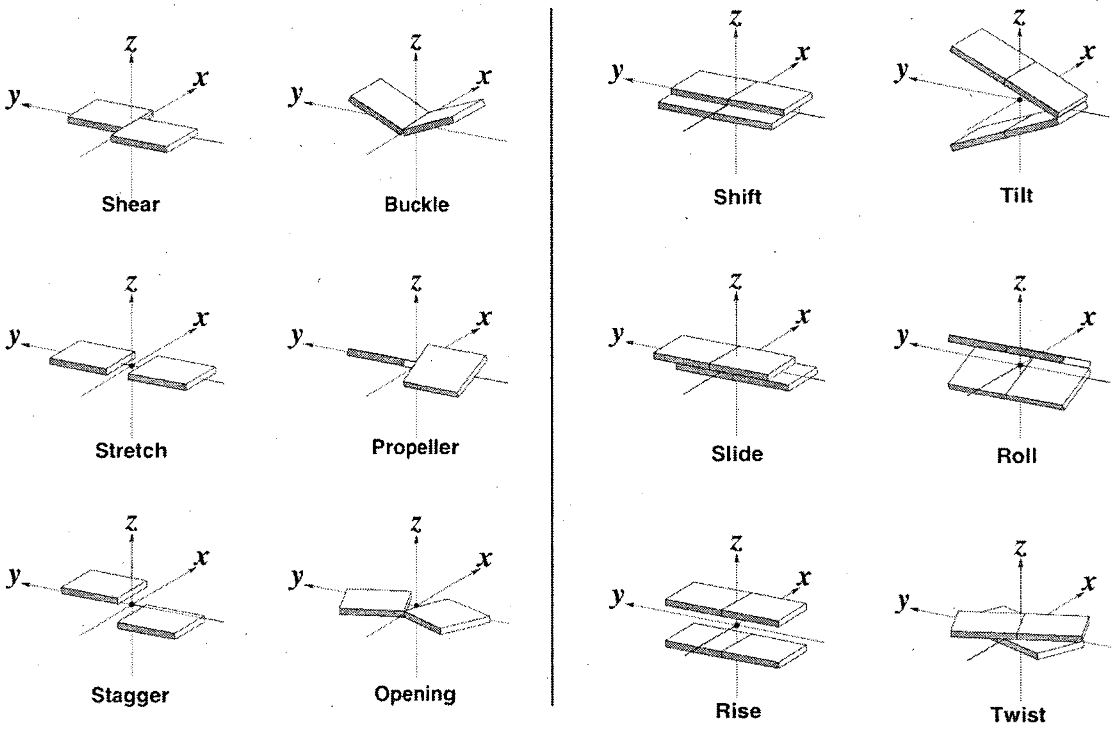
(207) Johansen, L. E.; Nygaard, P.; Lassen, C.; Agerso, Y.; Saxild, H. H. *J. Bacteriol.* **2003**, *185*, 5200-5209.

(208) Stoddard, C. D.; Gilbert, S. D.; Batey, R. T. *RNA* **2008**, *14*, 675-684.

(209) Carroll, M. T.; Bader, R. F. W. *Mol. Phys.* **1988**, *65*, 695 - 722.

Appendix A

Helical parameters that describe the relative orientation of the base pairs in a duplex



Appendix B

An introduction to the quantum theory of atoms in molecules (QTAIM)

One of the most familiar ways of presenting the geometry of a molecule is by using a ball and stick diagram, as shown for the water molecule in Figure B.1, where the spatial distribution of the nuclei is governed by the forces that each nucleus asserts on its neighbor.

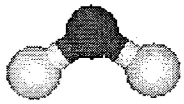


Figure B.1 Ball and stick diagram of water molecule.

An alternative way of looking at a molecule is through the topology of the electron density distribution, which is governed by the attractive forces on the nuclei. A contour plot depicting the electron density distribution in the water molecule in two-dimensions is shown in Figure B.2. The outer contour line corresponds to $\varrho = 0.001$ au (atomic units) which is taken to be the outer limit of the molecule.[48,210] Within this line are concentric circles that are made up of contour lines that connect points of the same electron density. Moving towards the center of the plot, each contour line represents a higher envelope of density, such that if the ball and stick diagram of the water molecule is superimposed onto the contour plot, the region of highest density overlaps with the position of the oxygen atom and the two regions with lower density overlap with the hydrogen atoms, and the overall shape of the molecule is mimicked by the electron density distribution. [48,210]

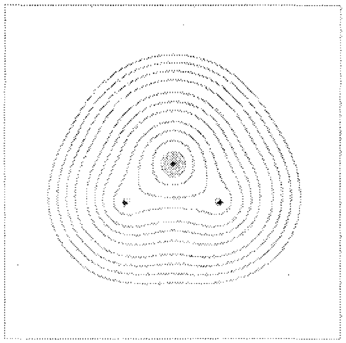


Figure B.2 Contour plot depicting the electron density for the water molecule.

The quantum theory of atoms in molecules extracts the electron density from a wavefunction and uses it as an observable for the determination of the electronic properties of molecules through an analysis of the topology of the electron density.[48,210]

1) Molecular structure

A topological analysis of the electron density of a molecule reveals its molecular structure. Figure B.3 shows a gradient vector field composed of gradient paths, which are lines that originate at the nuclei and move outwards in a direction that is orthogonal to the concentric contour lines and therefore represent the change in density from the region of highest concentration to the lowest concentration-the first derivative of the electron density ($\nabla\rho$).[48,210] A gradient vector field therefore shows exactly where the regions of highest concentration of density are located in a molecule relative to the positions of the nuclei as shown in Figure B.3 with the red stick diagram showing how the water molecule would be positioned. Gradient paths only intersect where they meet, at $\nabla\rho = 0$, this is the definition of a critical point (CP) in a vector field of a scalar quantity.[48,210] Most gradient vector lines shown in Figure B.3 terminate at the nuclei, rendering them a type of CP. CPs are classified by a rank and a signature where the rank describes the number of non-zero curvatures (defined by the eigenvalues of the Hessian, λ_1, λ_2 and λ_3) and the signature, which is the algebraic sum of the curvatures. So a (3,-3) CP describes a point where all curvatures are negative, i.e. a nucleus.[48,210] The blue lines between the nuclei, also gradient vectors, show how the gradient vector field naturally partitions the molecular electron density distribution into regions that define the atoms within the molecules, referred to as atomic basins (Ω); a set of blue lines therefore define the interatomic surface.[48,210]

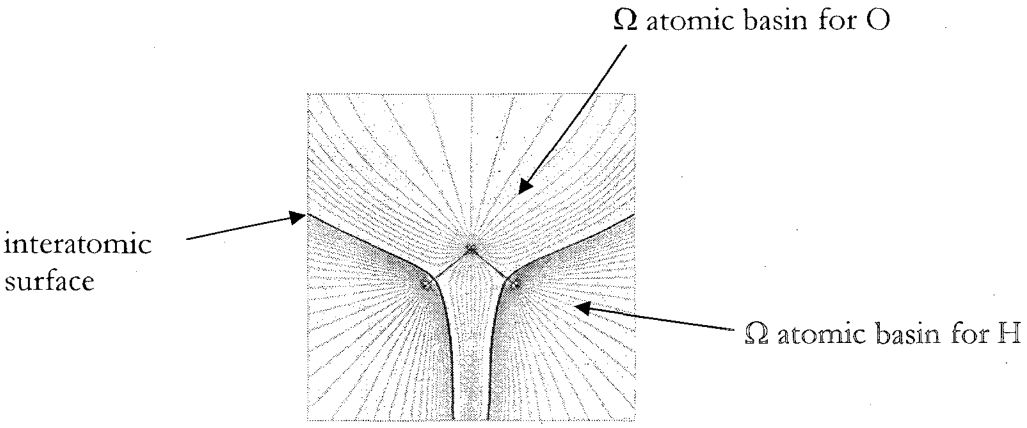


Figure B.3 Gradient vector field for the water molecule.

The blue lines originate at points found between the interacting nuclei.[48,210] These points are also critical points in the topology of the electron density that are defined as having two negative curvatures in the plane that is perpendicular to the molecule and one positive curvature in the plane of the molecule, i.e. a (3,-1) rank and signature.[48,210] This type of critical point is found at the terminus of two trajectories that define the path that connects two interacting nuclei (a bond path) and is called a bond critical point.[48,210] Figure B.3 shows the same water molecule as a molecular graph where the bond paths and bond critical points are identified. The value for the electron density at a bond critical point (usually given in e/Å^3) is directly proportional to the strength of the interaction.[48,210]

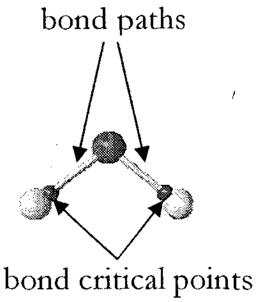


Figure B.3 molecular graph for the water molecule.

Weak bonding interactions are identified in the same way as covalent bonds. For the water dimer in Figure B.4, Figure B.5 shows a superposition of the interatomic surfaces (green lines). The highlighted plane H-O-H ···O on the contour plot corresponds to the highlighted nuclei in Figure B.4. Here the bond critical point (red dot in contour plot Figure C.5) identifies the presence of the hydrogen bond O-H ···O between the two water molecules.

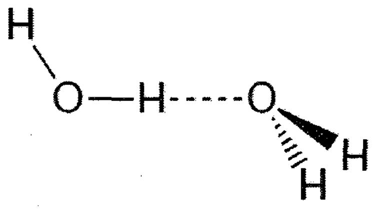


Figure B.4 Water dimer

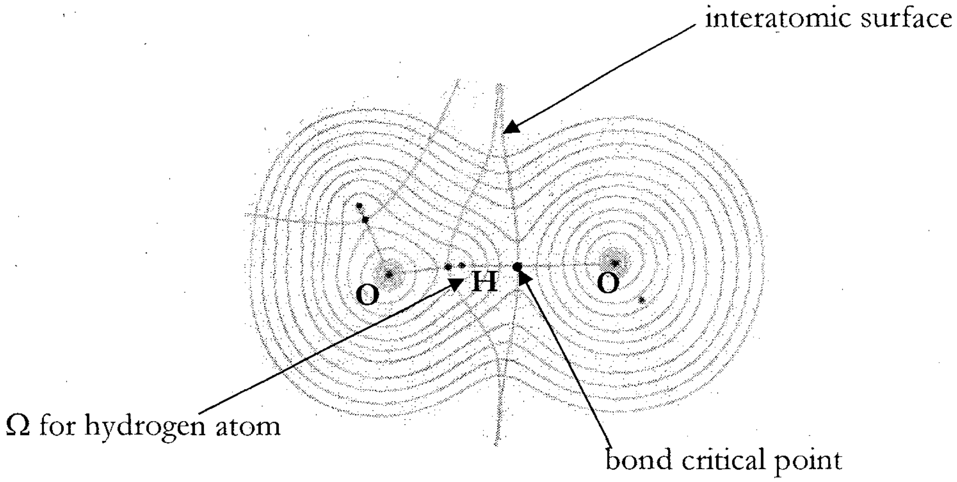


Figure B.5 Superimposed interatomic surfaces onto the two dimensional contour plot of the water dimer in the plane of the O-H ···O bond showing the location of the atomic basins, and the hydrogen bond, bond critical point.

The second derivative of the electron density (the Laplacian, $\nabla^2\varrho$) reveals areas in the molecule of local charge concentration (negative values) and depletion (positive values).[48,210] The $\nabla^2\varrho$ of an atom reveals the Lewis shell model for electron distribution, where solid lines represent areas of charge concentration and dotted lines reveal areas of depletion.[48,210] In Figure B.6, very close to the nucleus is a region of high concentration of charge, $\nabla^2\varrho < 0$ (solid line); moving away from the nucleus the density becomes more diffuse which results in $\nabla^2\varrho > 0$ (dotted lines). This is followed by a second solid line and another area of depletion; finally, the outer region is known as the valence shell charge concentration (VSCC). When an atom is involved in a shared interaction (covalent bonding), the VSCC's belonging to each nucleus are combined as shown in Figure B.6. (the outermost solid line) Thus a bond critical point between covalently bound nuclei is in this region of local charge concentration, therefore the value for $\nabla^2\varrho$ is negative. Weak bonding interactions which are closed shell interactions, occur in areas of relative charge depletion, as indicated by the dotted circles in Figure B.6, and are therefore associated with positive values of $\nabla^2\varrho$.[48,210]

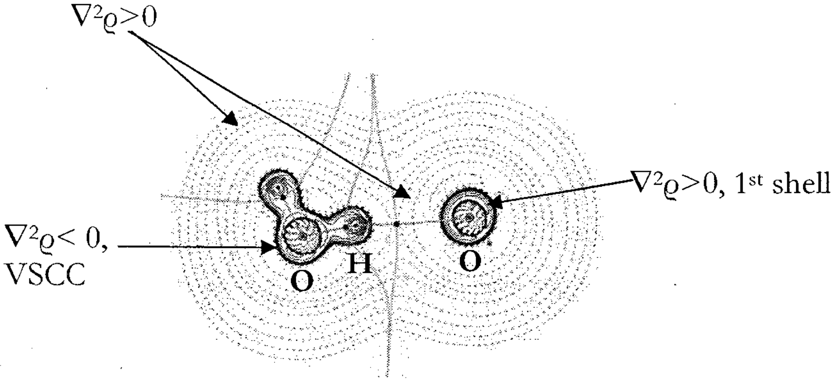


Figure B.6 Plot of the Laplacian of the electron density for the water dimer

2) Additional properties

The ellipticity of a bond (ε) is the ratio of the λ_1, and λ_2, eigenvalues of the Hessian of ϱ, ((λ_1/λ_2) -1) and can be used reveal the type of bond from its curvatures. Cylindrical bonds have $\varepsilon = 0$; $\varepsilon>0$ reveals double bond character.[48,210]

The partitioning of a molecule into atomic basins allows for the partitioning of the molecular electronic properties into atomic contributions. Integrating the electron density within the atomic basin determines the atomic charge, energy, polarization, and volume, all within the context of the molecule.[48,210]

Appendix C

Supporting information for Chapter 4

New insights into the use of (TD-)DFT for geometries and electronic structures of restrained π-stacked systems: [n.n]paracyclophanes

Table S4.1
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2]paracyclophanes from B3LYP/6-31G(d,p).
Table S4.2
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2] and [3.3]paracyclophanes from B3LYP/6-31+G(d,p).
Table S4.3
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2]paracyclophanes from B3LYP/6-311+G(d,p).
Table S4.4
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2]paracyclophanes from PBE0/6-31G(d,p).
Table S4.5
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2] and [3.3]paracyclophanes from PBE0/6-31+G(d,p).
Table S4.6
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2]paracyclophanes from PBE0/6-311+G(d,p).
Table S4.7
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2] and [3.3]paracyclophanes from BH&H/6-31+G(d,p).
Table S4.8
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **1**.
Table S4.9
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **1f**.
Table S4.10
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **1h**.
Table S4.11
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **1i**.
Table S4.12
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **2**.
Table S4.13
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **2a**.
Table S4.14
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **2b**

Table S4.15

Vertical ionization potentials (eV) and orbital energies (eV) of compounds **1**, **1a-e**, **1g**, **1h** and **2** from B3LYP/6-31+G(d,p).

Table S4.16

Vertical ionization potentials (eV) and orbital energies (eV) of compounds **1**, **1a-e**, **1g**, **1h** and **2** from PBE0/6-31+G(d,p).

Table S4.17

Vertical ionization potentials (eV) and orbital energies (eV) of compounds **1**, **1a-e**, **1g**, **1h** and **2** from BH&H/6-31+G(d,p).

Figure S4.1

Comparison of experimental (♦) and calculated $IP_{v,1}$ of a) **1**, **1d** and **1e** and b) **1**, **1a** and **1c** using three basis sets □ 6-31G(d,p), Δ 6-31+G(d,p) and * 6-311+G(d,p) and two functionals – – – B3LYP and —— PBE0. Data points are connected for ease of comparison.

Table S4.1
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2]paracyclophanes from B3LYP/6-31G(d,p).

	E_{tot}	ZPVE	Radical cations E_{tot}
1	-619.323788	0.274054	-619.049395
1a	-3190.431492	0.264224	-3190.153032
1b	-768.544538	0.259544	-768.245334
1c	-803.809784	0.271480	-803.509228
1d	-674.682676	0.290880	-674.429979
1e	-848.373595	0.339622	-848.130159
1g	-997.587561	0.324692	-997.319669
1h	-1032.859350	0.337011	-1032.600613

Table S4.2
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2] and [3.3]paracyclophanes from B3LYP/6-31+G(d,p).

			Radical cations
	E_{tot}	ZPVE	E_{tot}
1	-619.342590	0.273440	-619.060602
1a	-3190.476212	0.263754	-3190.190139
1b	-768.571441	0.258863	-768.263444
1c	-803.834500	0.270938	-803.525340
1d	-674.705864	0.290941	-674.444254
1e	-848.399385	0.338738	-848.147042
1f	-988.304282	0.267375	
1g	-997.630333	0.324126	-997.361862
1h	-1032.889231	0.335904	-1032.621808
1i	-1217.361668	0.332344	
2-boat	-698.013924	0.331380	-697.746483
2-chair	-698.013588	0.331477	-697.746395
2a-boat	-1066.974590	0.325542	
2a-chair	-1066.973687	0.325405	
2b-boat	-1296.029791	0.390654	
2b-chair	-1296.030407	0.390447	

Table S4.3
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2]paracyclophanes from B3LYP/6-311+G(d,p).

			Radical cations
	E_{tot}	ZPVE	E_{tot}
1	-619.454314	0.272624	-619.170294
1a	-3192.99809	0.262796	-3192.710081
1b	-768.725924	0.258133	-768.416248
1c	-803.989870	0.270247	-803.679665
1d	-674.831247	0.290941	-674.567599
1e	-848.566035	0.337665	-848.311890
1h	-1033.100060	0.335005	-1032.831048

Table S4.4
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2]paracyclophanes from PBE0/6-31G(d,p).

	E_{tot}	ZPVE	Radical cations E_{tot}
1	-618.587740	0.276144	-618.311901
1a	-3189.372920	0.266289	-3189.091809
1b	-767.663613	0.261504	-767.362007
1c	-802.868497	0.273565	-802.565274
1d	-673.889699	0.293244	-673.634625
1e	-847.390734	0.342426	-847.145114
1h	-1031.671043	0.339785	-1031.397335

Table S4.5
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2] and [3.3]paracyclophanes from PBE0/6-31+G(d,p).

	E_{tot}	ZPVE	Radical cations E_{tot}
1	-618.602656	0.275594	-618.320254
1a	-3189.413508	0.265929	-3189.126541
1b	-767.685402	0.258863	-767.377220
1c	-802.888011	0.273075	-802.579151
1d	-673.908289	0.292612	-673.646149
1e	-847.411401	0.341600	-847.158939
1f	-987.151597	0.269720	
1g	-996.496390	0.326952	-996.228024
1h	-1031.694939	0.338865	-1031.414543
1i	-1215.961452	0.335407	
2-boat	-697.176094	0.333737	-696.907933
2-chair	-697.175816	0.333826	-696.907815
2a-boat	-1065.724673	0.327219	
2a-chair	-1065.723544	0.327927	
2b-boat	-1294.531886	0.393909	
2b-chair	-1294.532386	0.393746	

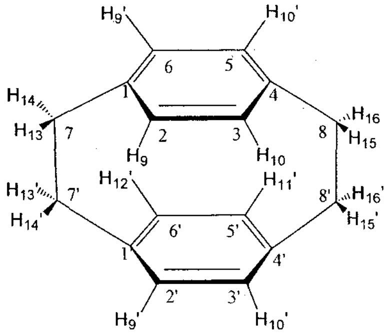
Table S4.6
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2]paracyclophanes from PBE0/6-311+G(d,p).

	E_{tot}	ZPVE	Radical cations E_{tot}
1	-618.704560	0.274601	-618.420619
1a	-3191.929313	0.264590	-3191.640947
1b	-767.828240	0.260160	-767.518837
1c	-803.030275	0.272244	-802.720023
1d	-674.023007	0.291606	-673.759178
1e	-847.565657	0.340114	-847.311728
1h	-1031.883565	0.337222	-1031.607614

Table S4.7
Total energies (E_{tot}, au), zero-point vibrational energies (ZPVE, au) and total energies of radical cations (E_{tot}, au) for [2.2] and [3.3]paracyclophanes from BH&H/6-31+G(d,p).

	E_{tot}	ZPVE	Radical cations E_{tot}
1	-614.549752	0.282249	
1a	-3183.148801	0.272540	-3182.862032
1b	-762.963359	0.268067	-762.657482
1c	-797.799108	0.280249	-797.492613
1d	-669.551590	0.299851	-669.291311
1e	-842.129665	0.350288	-841.877713
1f	-981.025603	0.277444	
1g	-990.546132	0.336295	-990.278341
1h	-1025.378146	0.348223	-1025.112615
1i	-1208.608865	0.345457	
2-boat	-692.587010	0.341539	-692.319654
2-chair	-692.587121	0.341636	-692.320392
2a-boat	-1059.064002	0.337347	
2a-chair	-1059.061961	0.337003	
2b-boat	-1286.643686	0.405122	
2b-chair	-1286.643490	0.405005	

Table S4.8
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **1**.



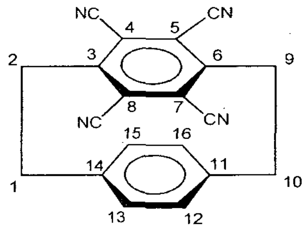
1

	Exp.[a]	B3LYP	PBE0	BH&H	MP2[b]
Bond lengths (pm)					
C1-C2	138.6	140.4	140.0	138.7	140.5
C1-C7	151.1	151.5	150.8	140.0	150.8
C2-C3	138.7	139.6	139.2	137.9	139.2
C3-C4	138.5	140.4	139.9	138.5	139.9
C7-C7'	159.0	161.0	159.0	157.0	159.0
C2-H9	100.0	108.8	108.9	108.5	108.9
C3-H10	101.0	108.8	108.8	108.5	108.8
C7-H13	102.0	109.5	109.6	109.2	109.6
C7-H14	108.0	109.6	109.6	109.2	109.6
Inter-ring distances (pm)					
C1-C1'	278.2	283.1	279.8	275.9	277.1
C2-C2'	309.6	315.7	311.9	306.7	308.6
Angles (deg)					
C1-C2-C3	120.7	120.7	120.9	121.0	120.9
C2-C3-C4	120.7	120.7	120.5	120.3	120.3
C1-C7-H13	112.0	110.1	111.0	111.6	112.0
C1-C7-H14	107.0	110.2	109.2	108.8	108.6
C1-C7-C7'	112.6	113.7	113.2	112.7	112.0
C6-C1-C7	120.9	121.0	120.5	121.4	120.2
C2-C1-C7	120.9	120.9	120.9	121.3	121.2
C2-C1-C6	117.0	116.8	116.9	117.1	117.2
H13-C7-H14	109.0	106.9	106.8	106.7	107.1
Torsion angles (deg)					
C1-C7-C7'-C1'	16.1	1.2	11.5	17.8	22.2
C4-C8-C8'-C4'	16.1	1.2	11.4	17.8	22.2
H13-C7-C7'-H13'	19.4	1.4	13.5	21.0	26.2
C1-C7-C7'-H13'	139.0	123.8	134.8	141.7	146.2
C1-C7-C7'-H14'	103.5	121.1	109.9	102.8	97.8
C4-C8-C8'-H16'	139.0	123.8	134.9	141.7	146.2
C4-C8-C8'-H15'	103.5	121.1	109.9	102.8	97.8
C1-C2-C3-C4	0.0	0.0	0.0	1.0	1.0
C2-C3-C4-C5	14.6	14.9	14.9	14.6	14.9
C4-C3-C2-H9	169.7	170.6	170.6	169.9	170.6

[a] X-ray data from (71).
[b] From (51).

Table S4.9
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **1f**.

	Exp.[a]	B3LYP	PBE0	BH&H
	Bond lengths (pm)			
C1-C2	155.2	161.6	159.9	157.4
C2-C3	151.0	151.1	150.2	140.9
C3-C4	140.5	140.9	140.4	138.8
C4-C5	140.0	141.7	141.1	139.3
C5-C6	138.9	140.9	140.5	138.9
C6-C7	140.3	141.0	140.4	138.8
C7-C8	140.7	141.7	141.1	139.4
C9-C10	156.9	161.6	159.9	157.4
C10-C11	150.4	151.5	150.7	149.5
C11-C12	139.0	140.5	140.0	138.8
C12-C13	138.0	139.7	139.4	138.1
C13-C14	139.7	140.5	140.1	138.6
C14-C15	138.0	140.5	140.1	138.8
C15-C16	138.3	139.7	139.4	138.1
C16-C11	138.4	140.4	140.1	138.6
	Inter-ring distances (pm)			
C3-C14	274.4	280.3	276.5	272.5
C4-C15	307.4	314.4	309.4	302.9
C5-C16	308.7	314.5	309.4	303.3
C6-C11	274.4	280.3	276.5	272.5
C7-C12	306.2	314.4	309.6	302.9
C8-C13	308.2	314.5	309.5	303.3
	Angles (deg)			
C8-C3-C2	121.5	121.5	121.5	121.4
C3-C2-C1	111.8	112.5	112.2	111.6
C2-C1-C14	113.4	113.7	113.3	112.8
C1-C14-C13	121.4	120.9	121.0	121.2
	Torsion angles (deg)			
C3-C2-C1-C14	9.4	0.0	4.3	14.6
C6-C9-C10-C11	11.5	0.0	4.5	14.6



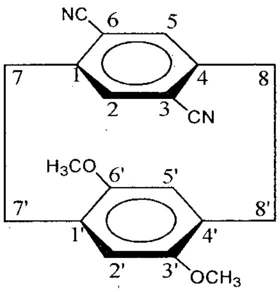
1f

---

[a] X-ray data from (102)

Table S4.10
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **1h**.

	Exp.[a]	B3LYP	PBE0	BH&H
	Bond lengths (pm)			
C1-C2	138.5	138.5	139.2	137.9
C2-C3	139.0	140.6	140.1	138.7
C3-C4	138.4	141.7	141.0	139.5
C4-C5	138.9	139.6	139.2	138.0
C5-C6	138.9	140.6	140.1	138.7
C6-C1	138.2	141.6	141.1	138.5
C1-C7	148.1	151.4	150.7	149.3
C7-C7'	158.2	160.6	159.0	156.7
C7'-C1'	151.2	151.6	150.7	149.3
C1'-C2'	137.7	140.2	139.7	139.4
C2'-C3'	139.4	140.1	139.7	138.3
C3'-C4'	138.3	140.6	140.2	138.6
C4'-C5'	137.8	140.2	139.8	139.4
C5-C6	138.6	140.0	139.6	138.3
C6'-C1'	138.3	140.6	140.2	138.7
C4'-C8'	150.4	151.4	150.8	149.3
C8'-C8	156.6	160.7	159.0	156.6
C8-C4	149.7	151.6	150.7	149.3
	Inter-ring distances (pm)			
C1-C1'	272.8	282.0	278.1	273.6
C2-C2'	303.5	314.0	309.5	303.3
C3-C3'	306.4	318.3	313.5	305.0
C4-C4'	273.1	281.9	278.1	273.6
C5-C5'	301.4	314.1	309.6	303.3
C6-C6'	305.5	318.3	313.4	304.9
	Angles (deg)			
C6-C1-C2	115.9	115.7	115.7	118.0
C1-C2-C3	121.0	121.4	121.2	119.7
C2-C1-C7	122.4	121.1	121.3	119.6
C6-C1-C7	120.7	122.1	121.9	120.6
C1-C7-C7'	112.0	112.6	112.0	112.3
C7-C7'-C1'	112.0	113.3	112.8	111.5
C7'-C1'-C2'	120.5	120.2	120.3	121.6
C7'-C1'-C6'	119.9	120.9	120.6	121.2
C6'-C1'-C2'	118.3	117.4	117.6	115.9
C1'-C2'-C3'	119.3	121.5	121.5	121.2
C3'-C4'-C8'	120.9	120.9	121.0	121.2
C4'-C8'-C8	111.9	113.3	113.0	111.5
C8'-C8-C4	112.5	112.6	112.0	112.3
	Torsion anlges (deg)			
C1-C7-C7'-C1'	15.9	17.7	19.6	23.1
C4-C8-C8'C4'	15.7	17.7	19.6	23.1
	Out-of-plane twist of methoxyl groups (deg)			
C2'-C3'-O-CH_3	9.4	11.4	11.0	13.5
C5'-C6'-O-CH_3	9.7	11.5	11.1	13.6

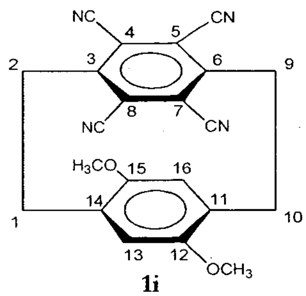


1h

[a] X-ray data from (101).

Table S4.11

Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **1i**.



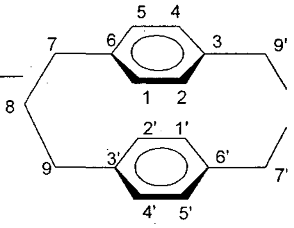
1i

	Exp.[a]	B3LYP	PBE0	BH&H
Bond lengths (pm)				
C1-C2	157.9	160.8	159.3	157.0
C2-C3	150.6	151.3	150.4	149.0
C3-C4	140.1	141.0	140.4	138.9
C4-C5	140.3	141.8	141.2	139.5
C5-C6	140.0	141.1	140.6	139.0
C6-C7	139.7	140.9	140.5	138.9
C7-C8	140.5	141.7	141.2	139.5
C8-C3	139.9	141.0	140.5	157.0
C9-C10	158.2	160.9	159.3	139.0
C10-C11	150.6	151.5	150.8	149.3
C11-C12	139.9	140.2	139.8	138.7
C12-C13	138.3	140.2	139.8	138.3
C13-C14	139.2	140.7	140.3	138.4
C14-C15	139.7	140.2	139.8	138.7
C15-C16	138.5	140.1	139.7	138.4
C16-C11	139.6	140.7	140.3	138.4
Inter-ring distances (pm)				
C3-C14	274.0	280.4	276.2	271.8
C4-C15	307.0	317.0	311.4	302.1
C5-C16	304.0	313.3	307.9	300.5
C6-C11	274.0	280.4	276.2	271.8
C7-C12	308.0	313.3	311.4	302.1
C8-C13	306.0	316.9	307.7	300.5
Angles (deg)				
C8-C3-C2	121.6	121.8	121.7	120.3
C3-C2-C1	111.8	112.1	111.6	112.3
C2-C1-C14	112.8	113.2	112.8	111.2
C1-C14-C13	120.5	120.1	120.2	121.5
Torsion angles (deg)				
C3-C2-C1-C14	8.2	16.7	17.2	19.9
C6-C9-C10-C11	9.1	16.8	17.2	19.8
Out-of-plane twist of methoxyl groups (deg)				
C13-C12-O-CH_3	8.0	13.1	12.0	13.3
C16-C15-O-CH_3	5.2	13.2	12.0	13.3

[a] X-ray data from (102).

Table S4.12

Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **2**.



2

	Exp.[a]	B3LYP		PBE0		BH&H	
	Chair	Chair	Boat	Chair	Boat	Chair	Boat
			Bond lengths (pm)				
C1-C2	138.2	139.6	139.4	139.1	139.3	137.9	137.7
C2-C3	139.2	140.2	140.4	139.7	139.4	138.3	138.7
C3-C4	139.2	140.0	139.8	139.7	140.1	138.3	138.0
C4-C5	138.4	139.5	139.8	139.1	138.9	138.0	138.8
C5-C6	139.1	140.2	139.8	139.7	139.9	138.4	138.0
C6-C1	139.0	140.1	140.4	139.7	139.4	138.4	149.5
C9-C3	151.7	151.8	151.8	151.1	151.0	149.5	152.4
C8-C9	152.5	155.3	155.4	154.1	151.1	152.3	152.4
C1-H(1)	98.0	108.7	108.8	108.8	108.3	108.5	108.6
C2-H(2)	98.0	108.7	108.8	108.8	108.3	108.5	108.5
C4-H(4)	98.0	108.8	108.7	108.8	108.8	108.5	108.5
C5-H(5)	106.0	108.8	108.7	108.9	108.9	108.5	108.5
C7-H(7a)	104.0	109.9	108.8	109.7	109.8	109.4	109.4
C7-H(7b)	92.0	109.7	109.7	109.8	109.8	109.5	109.5
C8-H(8a)	93.0	109.7	109.9	109.8	109.8	109.5	109.5
C8-H(8b)	106.0	109.8	109.7	109.8	109.7	109.4	109.4
C9-H(9a)	101.0	109.9	109.7	109.8	109.8	109.4	109.4
C9-H(9b)	88.0	109.7	109.9	109.8	109.8	109.4	109.5
			Inter-ring distances (pm)				
C6-C3'	313.7	323.9	324.5	319.0	319.4	310.9	312.8
C3-C6'	313.7	323.9	324.5	319.0	319.3	310.9	312.8
C2-C5'	328.9	342.8	332.9	337.5	328.5	325.7	336.6
C5-C2'	328.9	342.8	351.7	337.5	344.5	325.6	319.3
C4-C1'	331.0	341.0	332.9	335.8	328.5	324.8	336.6
C1-C4'	331.0	341.0	351.6	335.8	344.5	324.9	319.5
			Angles (deg)				
C6-C1-H(1)	121.9	119.4	119.6	119.4	119.4	119.2	119.4
C2-C1-H(1)	116.5	119.2	119.1	119.3	119.3	119.5	119.6
C1-C2-H(2)	123.2	119.0	119.1	119.1	119.3	119.0	119.5
C3-C2-H(2)	115.1	119.8	119.6	119.7	119.4	119.8	119.4
C5-C4-H(4)	124.9	119.2	119.2	119.4	119.2	119.5	119.5
C4-C5-H(5)	122.7	119.1	119.2	119.1	119.2	119.5	119.5
C6-C5-H(5)	115.8	119.7	119.5	119.7	119.6	119.5	119.2

Table S4.12 continued.

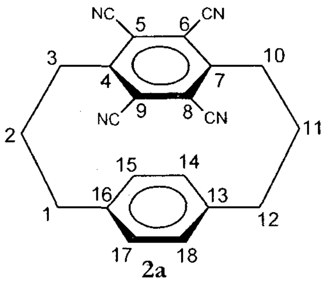
C6-C7-H(7a)	111.0	109.2	109.7	109.7	109.7	110.2	109.8
C6-C7-H(7b)	109.2	109.7	109.2	109.2	109.2	109.3	109.4
C8-C7-H(7a)	108.0	108.9	106.9	107.2	107.3	107.9	107.8
C8-C7-H(7b)	105.9	107.2	109.2	109.3	109.2	109.0	109.2
C7-C8-H(8a)	113.1	109.2	107.1	108.9	107.3	108.8	108.7
C7-C8-H(8b)	108.1	106.9	108.8	107.4	108.9	107.9	107.6
C9-C8-H(8a)	106.3	108.8	108.9	108.9	108.9	109.6	109.1
C9-C8-H(8b)	106.2	107.1	107.1	107.4	107.3	107.7	107.8
C8-C9-H(9a)	108.7	106.9	109.2	109.3	107.2	107.7	107.7
C8-C9-H(9b)	108.4	109.2	106.9	107.2	109.3	109.2	109.4
C3'-C9-H(9a)	110.3	109.8	109.7	109.7	109.7	109.3	109.8
C3'-C9-H(9a)	107.9	109.2	109.3	109.2	109.2	108.7	109.2
H(7a)-C7-H(7b)	106.3	106.2	106.2	106.2	106.2	106.6	106.3
H(8a)-C8-H(8b)	105.5	105.8	105.9	105.9	105.9	105.6	105.8
H(9a)-C9-H(9b)	107.9	106.1	106.2	106.2	106.2	105.7	106.2
			Torsion angles (deg)				
C6-C7-C8-C9[b]	70.1	67.3	67.7	67.5	67.8	70.6	67.6
C7-C8-C9-C3'[c]	65.1	67.3	67.7	67.5	67.8	63.3	69.2

[a] X-ray data from (107).

[b] Identical with C3-C9'-C8'-C7'.

[c] Identical with C9'-C8'-C7'-C6'.

Table S4.13
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **2a**.



2a

	B3LYP		PBE0		BH&H	
	Boat	Chair	Boat	Chair	Boat	Chair
Bond lengths (pm)						
C1-C2	155.1	155.0	154.0	153.9	152.3	152.1
C2-C3	155.8	155.8	154.6	154.6	152.7	152.8
C3-C4	149.6	151.5	150.5	150.6	149.1	149.2
C4-C5	141.0	140.8	140.5	140.3	138.6	138.8
C5-C6	141.6	141.6	141.0	141.0	139.3	139.3
C6-C7	141.0	140.8	140.4	140.3	138.7	138.8
C7-C8	140.7	140.8	140.2	140.3	138.8	138.8
C8-C9	141.6	141.6	141.0	141.0	139.3	139.3
C9-C4	140.7	140.8	140.2	140.3	138.9	138.8
C1-C16	151.9	151.9	151.1	151.1	149.7	149.7
C16-C17	140.0	140.2	139.7	139.8	138.7	138.5
C17-C18	139.9	139.7	139.6	139.3	137.7	138.1
C18-C13	140.0	140.3	139.7	139.9	138.7	138.5
C13-C14	140.2	140.2	140.0	139.8	138.4	138.1
C14-C15	139.4	139.7	139.0	139.3	138.3	138.5
C15-C16	140.5	140.3	140.0	139.9	138.3	138.5
C13-C12	151.9	151.9	151.2	151.1	149.7	149.7
C12-C11	155.1	155.0	153.9	153.9	152.3	152.8
C11-C10	155.8	155.8	154.7	154.6	152.7	152.1
C10-C7	151.4	151.5	150.5	150.5	149.1	149.7
Inter-ring distances (pm)						
C4-C16	320.6	320.2	315.0	314.9	307.3	306.9
C5-C15	350.2	340.8	342.2	334.2	313.0	322.7
C6-C14	350.2	337.7	342.2	331.1	313.1	319.9
C7-C13	320.6	320.2	315.1	314.9	307.2	306.9
C8-C18	328.8	340.8	323.0	334.2	329.9	322.7
C9-C17	328.8	337.7	323.0	331.2	329.9	319.9
Angles (deg)						
C5-C4-C9	116.5	116.4	116.4	116.3	116.6	116.4
C4-C3-C2	125.2	115.7	114.6	115.3	114.0	114.9
C3-C2-C1	113.1	118.5	117.8	118.0	117.0	117.1
C2-C1-C16	114.9	114.7	114.6	114.4	114.3	113.9
C15-C16-C17	117.4	117.4	117.4	117.5	117.6	117.7
C13-C12-C11	114.9	114.7	114.6	114.4	114.3	113.9
C12-C10-C7	118.4	118.5	117.9	118.0	117.1	117.1
C11-C10-C7	115.1	115.7	114.6	115.3	114.0	114.9
Torsion angles (deg)						
C4-C3-C2-C1[a]	64.9[b]	63.0	64.6	62.3	64.1	61.5
C3-C2-C1-C16[c]	67.7	68.0	67.9	68.5	68.5	69.4

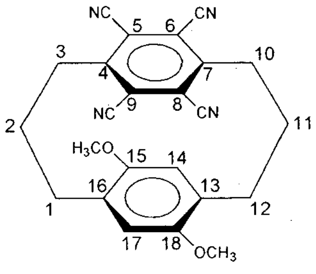
[a] Identical with C7-C10-C11-C12.

[b] 64.0° for C7-C10-C11-C12.

[c] Identical with C10-C11-C12-C13.

Table S4.14.
Experimental and calculated (6-31+G(d,p) basis set) geometrical parameters for **2b**.

	Exp.[a]	B3LYP		PBE0		BH&H	
	Boat	Boat	Chair	Boat	Chair	Boat	Chair
				Bond lengths (pm)			
C1-C2	153.1	155.0	155.1	154.0	154.1	152.1	152.3
C2-C3	153.6	155.8	155.8	154.6	154.6	152.9	152.6
C3-C4	150.7	151.5	151.4	150.5	150.6	149.1	149.2
C4-C5	139.9	140.7	141.0	140.5	140.3	138.7	138.8
C5-C6	139.8	141.7	141.6	141.1	141.1	139.2	139.5
C6-C7	139.7	140.7	140.9	140.5	140.4	138.8	138.8
C7-C8	139.3	141.0	140.9	140.1	140.4	138.8	138.9
C8-C9	139.9	141.6	141.7	141.1	141.0	139.7	139.4
C9-C4	138.8	141.0	140.8	140.2	140.4	138.8	138.8
C1-C16	151.3	152.0	151.6	150.8	150.9	149.7	149.4
C16-C17	138.7	140.2	140.0	139.5	140.3	138.9	138.8
C17-C18	137.9	139.8	140.0	139.9	139.6	138.7	138.3
C18-C13	138.5	140.9	140.7	140.2	139.7	138.4	138.2
C13-C14	139.2	139.9	140.1	139.7	140.3	139.0	138.8
C14-C15	138.1	140.3	139.9	139.4	139.7	137.9	138.3
C15-C16	139.1	140.6	140.8	140.5	139.6	138.1	138.2
C13-C12	151.5	151.6	151.6	151.3	150.9	149.7	149.4
C12-C11	154.5	155.1	155.1	153.9	154.0	152.5	152.2
C11-C10	151.4	155.7	155.6	154.7	154.5	152.5	152.6
C10-C7	151.3	151.4	151.5	150.5	150.5	149.0	149.2
				Inter-ring distances (pm)			
C4-C16	309.9	319.9	319.5	313.6	313.8	304.9	304.8
C9-C17	325.0	329.0	335.8	323.5	335.1	315.9	320.3
C6-C14	325.3	347.5	335.8	337.7	335.1	318.1	320.3
C7-C13	306.8	320.6	319.6	314.6	313.7	302.1	304.9
C8-C18	321.2	350.6	342.4	327.6	328.8	313.1	316.5
C5-C15	325.9	332.8	342.5	340.8	328.8	318.1	316.4
				Angles (deg)			
C5-C4-C9	115.3	116.4	116.2	116.3	116.2	116.5	116.2
C4-C3-C2	114.5	115.0	115.5	114.8	115.2	114.6	114.7
C3-C2-C1	117.9	118.2	118.6	118.2	118.2	116.8	117.4
C2-C1-C16	114.9	114.7	115.5	115.4	115.1	114.0	114.9
C15-C16-C17	117.2	117.8	117.6	117.7	117.9	118.0	118.2
C13-C12-C11	115.0	115.5	115.4	114.5	115.1	116.4	114.8
C12-C11-C10	117.2	118.6	118.7	117.6	118.2	117.6	117.5
C11-C10-C7	114.1	115.3	115.5	114.5	115.1	113.8	114.7
				Torsion angles (deg)			
C4-C3-C2-C1	61.9	64.7	63.7	63.3	63.3	60.4	63.0
C3-C2-C1-C16	68.9	68.9	65.4	69.7	65.1	69.4	64.1
C7-C10-C11-C12	62.9	64.7	63.6	64.7	63.4	61.1	63.0



2b

Table S14 continued.

C10-C11-C12-C13	63.7	65.1	65.4	63.8	65.2	61.3	64.1
	Out-of-plane twist of methoxyl groups (deg)						
C17-C18-O-CH_3	5.8	7.5	7.6	2.7	5.6	7.0	2.4
C14-C15-O-CH_3	7.7	7.6	7.6	5.7	5.4	5.6	2.5

[a] X-ray data taken from ref (102).

Table S4.15
Vertical ionization potentials (eV) and orbital energies (eV) of compounds **1**, **1a-e**, **1g**, **1h** and **2** from B3LYP/6-31+G(d,p).

	Exp.[a] IP_v	$IP_{v,1}$	$-\varepsilon$	Calc. IP
1	8.10	7.67	6.07	7.67
	8.30		6.14	7.74
	8.50		6.37	7.97
	9.60		7.38	8.98
	10.3		8.31	9.91
1a	8.10	7.79	6.23	7.79
	8.10		6.29	7.86
	8.50		6.55	8.11
	9.30		7.25	8.81
	10.10		7.90	9.46
1b	8.60	8.38	6.69	8.38
	9.10		7.04	8.72
	9.20		7.25	8.93
	9.62		7.62	9.30
	9.79		7.81	9.50
1c	8.90	8.40	6.74	8.40
	8.90		6.95	8.61
	9.25		7.38	9.04
	10.12		8.06	9.72
1d	7.50	7.12	5.45	7.12
	7.90		5.96	7.63
	8.20		6.17	7.84
	9.00		6.83	8.50
	10.00		8.01	9.68
1e	7.35	6.87	5.26	6.87
	8.00		6.04	7.65
	8.25		6.23	7.84
	9.05		7.01	8.62
	9.65		7.73	9.34
1g	7.54	7.31	5.69	7.31
	8.66		6.89	8.51
	8.95		7.18	8.79
	9.35		7.48	9.09
	9.41		7.53	9.14
1h	7.60	7.27	5.67	7.27
	8.50		6.76	8.36
	8.75		7.23	8.83
	9.65		7.81	9.41
	9.90		8.10	9.70
2 (boat)	7.80	7.27	5.75	7.27
	8.10		6.26	7.78
	8.90		6.77	8.29
	9.10		7.12	8.64
	10.33		8.47	9.99

[a] From ref (109) and (110).

Table S4.16

Vertical ionization potentials (eV) and orbital energies (eV) of compounds **1**, **1a-e**, **1g**, **1h** and **2** from PBE0/6-31+G(d,p).

	Exp.[a] IP_v	$IP_{v,1}$	$-\varepsilon$	Calc. IP
1	8.10	7.68	6.23	7.68
	8.30		6.32	7.77
	8.50		6.62	8.07
	9.60		7.66	9.11
	10.3		8.52	9.97
1a	8.10	7.81	6.40	7.81
	8.10		6.48	7.62
	8.50		6.78	7.89
	9.30		7.51	8.58
	10.10		8.19	9.77
1b	8.60	8.38	6.88	8.38
	9.10		7.24	8.74
	9.20		7.47	8.97
	9.62		7.87	9.37
	9.79		8.09	9.59
1c	8.90	8.40	6.93	8.40
	8.90		7.14	8.61
	9.25		7.62	9.09
	10.12		8.35	9.82
1d	7.50	7.13	5.63	7.81
	7.90		6.12	7.89
	8.20		6.39	8.19
	9.00		7.08	8.92
	10.00		8.27	9.60
1e	7.35	6.86	5.42	6.86
	8.00		6.22	7.66
	8.25		6.46	7.90
	9.05		7.30	8.74
	9.65		8.63	10.07
1g	7.54	7.30	5.86	7.30
	8.66		7.12	8.56
	8.95		7.40	8.84
	9.35		7.71	9.15
	9.41		7.79	9.20
1h	7.60	7.27	6.03	7.27
	8.50		6.96	8.20
	8.75		7.44	8.68
	9.65		8.00	9.24
	9.90		8.21	9.45
2 (boat)	7.80	7.29	5.91	7.29
	8.10		6.45	7.83
	8.90		7.02	8.40
	9.10		7.39	8.77
	10.33		8.69	10.07

[a] From ref (109) and (110).

Table S4.17
Vertical ionization potentials (eV) and orbital energies (eV) of compounds **1**, **1a-e**, **1g**, **1h** and **2** from BH&H/6-31+G(d,p).

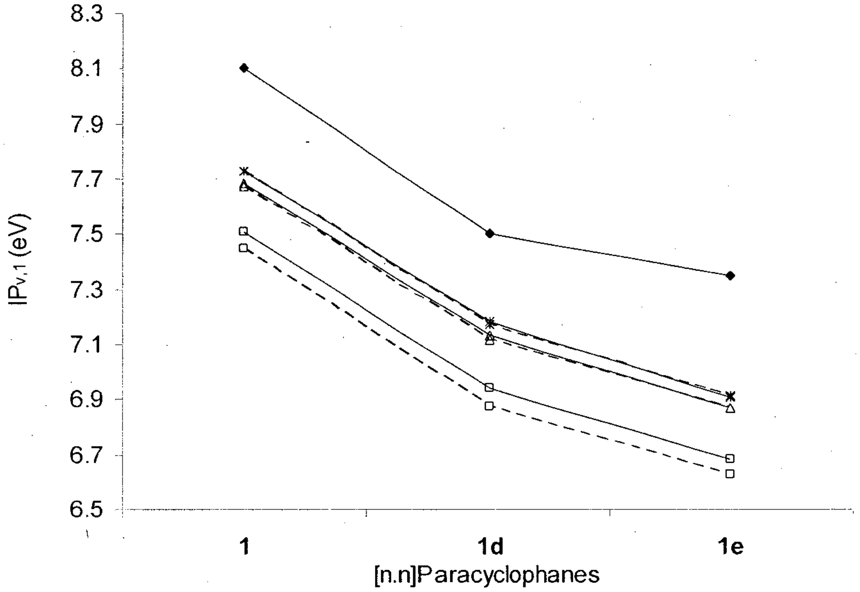
	Exp.[a] IP_v	$IP_{v,1}$	$-\varepsilon$	Calc. IP
1	8.10	7.65	6.92	7.65
	8.30		7.01	7.75
	8.50		7.47	8.21
	9.60		8.64	9.33
	10.3		9.68	10.42
1a	8.10	7.80	7.10	7.80
	8.10		7.19	7.88
	8.50		7.61	8.30
	9.30		8.43	9.12
	10.10		9.29	9.98
1b	8.60	8.32	7.62	8.32
	9.10		8.02	8.72
	9.20		8.80	9.51
	9.62		9.11	9.81
	9.79		9.30	10.00
1c	8.90	8.34	7.65	8.34
	8.90		7.87	8.56
	9.25		8.46	9.15
	10.12		9.33	10.02
1d	7.50	7.08	6.37	7.08
	7.90		6.80	7.51
	8.20		7.19	7.90
	9.00		7.99	8.70
	10.00		9.45	10.16
1e	7.35	6.86	6.20	6.86
	8.00		6.95	7.61
	8.25		7.29	7.95
	9.05		8.32	8.98
	9.65		9.24	9.90
1g	7.54	7.29	6.65	7.29
	8.66		7.99	8.63
	8.95		8.72	9.36
	9.35		8.76	9.40
	9.41		9.25	9.89
1h	7.60	7.26	6.64	7.26
	8.50		7.73	8.34
	8.75		8.28	8.89
	9.65		9.05	9.66
	9.90		9.68	10.28
2 (boat)	7.80	7.27	6.56	7.27
	8.10		7.17	7.87
	8.90		7.92	8.63
	9.10		8.34	9.05
	10.33		9.85	10.55

[a] From ref (109) and (110).

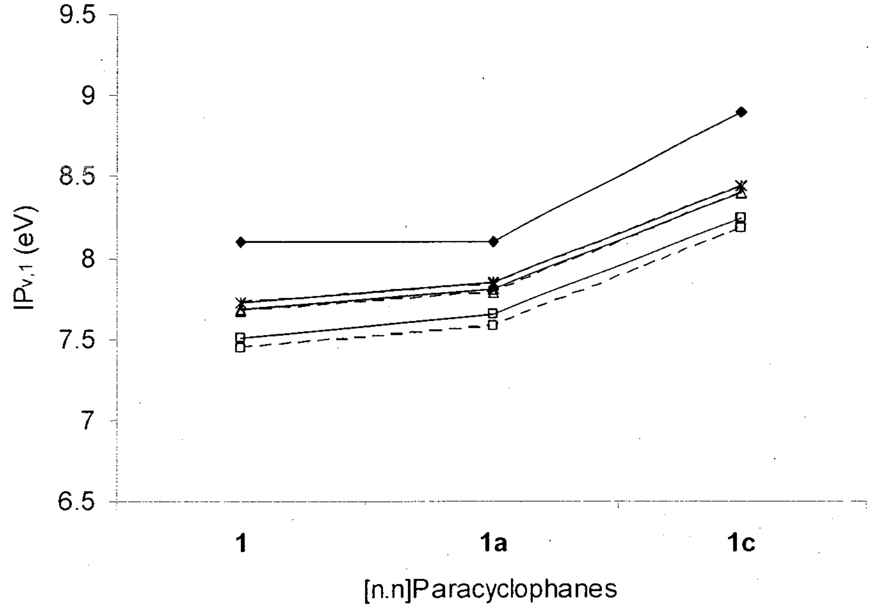
Figure S4.1

Comparison of experimental (♦) and calculated $IP_{v,1}$ of a) **1**, **1d** and **1e** and b) **1**, **1a** and **1c** using three basis sets □ 6-31G(d,p), Δ 6-31+G(d,p) and * 6-311+G(d,p) and two functionals – – – B3LYP and —— PBE0. Data points are connected for ease of comparison.

(a)



(b)



Appendix D

Supporting information to Chapter 5

Revisiting the sequence and structural effects on the hydrogen bonding and π-stacking interactions in nucleic acids

Table S5.1 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **1IKK** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

Table S5.2 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **1RNA** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

Table S5.3 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **126D** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

Table S5.4 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **1SK5** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

Table S5.5 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **1VJ4** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

Table S5.6 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **1RNA** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

Table S5.7 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **259D** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

Table S5.8 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **420D** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

Table S5.9 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **440D** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

Table S5.10 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **485D** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

Table S5.12 Individual values for the density at each H-bond critical point (ϱ_{HB}), from AT and GC base pairs isolated from **157D** when bases are stacked (S) and unstacked (U), the $\Sigma\varrho_{HB}$ and $\Sigma\varrho_{\pi}$ in (e/Å^3).

For all tables $\Sigma\varrho_{HB}$(S-U) = sum of density at HB when base pairs are stacked (S) versus unstacked (U)

Figure S5.1 Density at the HBCPs (ϱ_{HB}) of AT base pairs isolated from DNA duplexes.
Figure S5.2 Density at the HBCPs (ϱ_{HB}) of AU base pairs isolated from RNA duplexes.
Figure S5.3 Density at the HBCPs (ϱ_{HB}) of GC base pairs isolated from DNA duplexes.
Figure S5.4 Density at the HBCPs (ϱ_{HB}) of GC base pairs isolated from RNA duplexes.
Figure S5.5 Molecular graphs of AA:TT stacked base pairs showing the presence of O···O, N···N, N···C, O···N, O···H, and C-H···π interactions.
Figure S5.6 Molecular graphs of TA:TA stacked base pairs showing the presence of O···O, N···N, N···C, O···N, O···H, and C-H···π interactions.
Figure S5.7 Molecular graphs of AT:AT stacked base pairs showing the presence of O···O, N···N, N···C, O···N, O···H, but no C-H···π interactions.

	A16_T5_A17_T4				A18_T3_A17_T4				T13_A8_T14_A7			
1IKK	A16_T5		A17_T4		A18_T3		A17_T4		T13_A8		T14_A7	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.154	0.155	0.078	0.078	0.178	0.178	0.077	0.078	0.171	0.171	0.124	0.125
(N)H-N	0.432	0.431	0.334	0.331	0.275	0.272	0.332	0.331	0.354	0.351	0.265	0.265
C-H(O)	0.064	0.064	0.046	0.046	0.026	0.026	0.046	0.046	0.027	0.028	0.020	0.020
$\Sigma\rho_{HB}$	**0.650**	**0.650**	**0.457**	**0.455**	**0.479**	**0.476**	**0.455**	**0.455**	**0.552**	**0.550**	**0.409**	**0.410**
$\Sigma\rho_{HB}$S-U	**-0.000**		**0.002**		**0.003**		**-0.000**		**0.002**		**0.001**	
$\Sigma\rho_{\pi}$	**0.486**				**0.420**				**0.431**			

	T14_A7_T15_A6				T15_A6_A16_T5				C1_G20_C2_G19			
1IKK	T14_A7		T15_A6		T15_A6		A16_T5		C1_G20		C2_G19	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.125	0.125	0.158	0.158	0.157	0.158	0.155	0.155	0.197	0.197	0.209	0.210
(N)H-N	0.266	0.265	0.348	0.347	0.347	0.347	0.431	0.431	0.251	0.252	0.260	0.260
C/N-H(O)	0.020	0.020	0.039	0.039	0.039	0.039	0.064	0.064	0.270	0.271	0.191	0.189
$\Sigma\rho_{HB}$	**0.411**	**0.410**	**0.546**	**0.544**	**0.543**	**0.544**	**0.649**	**0.650**	**0.718**	**0.720**	**0.659**	**0.658**
$\Sigma\rho_{HB}$S-U	**0.001**		**0.002**		**-0.001**		**-0.001**		**-0.002**		**0.001**	
$\Sigma\rho_{\pi}$	**0.339**				**0.337**				**0.395**			

	C11_G10_C12_G9				C12_G9_T13_A8				A18_T3_G19_C2			
1IKK	C11_G10		C12_G9		C12_G9		T13_A8		A18_T3		G19_C2	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.235	0.236	0.222	0.222	0.223	0.222	0.171	0.171	0.178	0.178	0.210	0.210
(N)H-N	0.232	0.233	0.232	0.232	0.232	0.232	0.354	0.351	0.274	0.272	0.259	0.260
C/N-H(O)	0.194	0.195	0.181	0.180	0.180	0.180	0.027	0.028	0.026	0.026	0.189	0.189
$\Sigma\rho_{HB}$	**0.662**	**0.664**	**0.635**	**0.634**	**0.635**	**0.634**	**0.552**	**0.550**	**0.478**	**0.476**	**0.659**	**0.658**
$\Sigma\rho_{HB}$S-U	**-0.002**		**0.001**		**0.001**		**0.002**		**0.002**		**0.001**	
$\Sigma\rho_{\pi}$	**0.393**				**0.379**				**0.450**			

Table S5.1

1RNA	A3 U26 U2 A27				A5 U24 U4 A25				A7 U22 U6 A23			
	A3_U26		U2_A27		A5_U24		U4_A25		A7_U22		U6_A23	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.182	0.182	0.199	0.200	0.166	0.167	0.181	0.182	0.192	0.192	0.202	0.202
(N)H-N	0.294	0.293	0.303	0.302	0.271	0.270	0.311	0.310	0.316	0.315	0.311	0.309
C-H(O)	0.021	0.021	0.037	0.037	0.045	0.045	0.035	0.035	0.023	0.023	0.041	0.041
$\Sigma\rho_{HB}$	**0.496**	**0.496**	**0.538**	**0.538**	**0.482**	**0.482**	**0.527**	**0.526**	**0.530**	**0.529**	**0.553**	**0.552**
$\Sigma\rho_{HB}$S-U	**0.000**		**0.000**		**0.000**		**0.001**		**0.001**		**0.001**	
$\Sigma\rho_{\pi}$	**0.189**				**0.214**				**0.263**			

1RNA	A9 U20 U8 A21				A11 U18 U10 A19				A13 U16 U12 A17			
	A9_U20		U8_A21		A11_U18		U10_A19		A13_U16		U12_A17	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.183	0.184	0.214	0.214	0.211	0.211	0.191	0.191	0.186	0.187	0.182	0.183
(N)H-N	0.286	0.285	0.231	0.230	0.247	0.246	0.307	0.307	0.281	0.280	0.274	0.273
C-H(O)	0.031	0.031	0.015	0.015			0.028	0.028	0.032	0.032	0.016	0.015
$\Sigma\rho_{HB}$	**0.500**	**0.500**	**0.460**	**0.459**	**0.458**	**0.457**	**0.526**	**0.526**	**0.500**	**0.499**	**0.471**	**0.471**
$\Sigma\rho_{HB}$S-U	**0.001**		**0.001**		**0.001**		**0.001**		**0.001**		**0.000**	
$\Sigma\rho_{\pi}$	**0.221**				**0.198**				**0.252**			

1RNA	A14 U15 A13 U16				U2 A27 A1 U28				U12 A17 A11 U18			
	A14_U15		A13_U16		U2_A27		A1_U28		A11_U18		U12_A17	
					S	U	S	U	S	U	S	U
N-H(O)	0.183	0.183	0.185	0.187	0.197	0.200	0.177	0.177	0.211	0.211	0.183	0.183
(N)H-N	0.253	0.251	0.280	0.280	0.303	0.302	0.304	0.302	0.246	0.246	0.271	0.273
C-H(O)	0.019	0.018	0.032	0.032	0.036	0.037	0.017	0.017			0.015	0.015
$\Sigma\rho_{HB}$	**0.455**	**0.452**	**0.496**	**0.499**	**0.536**	**0.538**	**0.499**	**0.495**	**0.457**	**0.457**	**0.469**	**0.471**
$\Sigma\rho_{HB}$S-U	**0.003**		**-0.003**		**0.002**		**0.004**		**0.000**		**-0.002**	
$\Sigma\rho_{\pi}$	**0.401**				**0.247**				**0.349**			

1RNA	U6 A23 A5 U24				U8 A21 A7 U22				U10 A19 A9 U20			
	U6_A23		A5_U24		U8_A21		A7_U22		U10_A19		A9_U20	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.203	0.202	0.167	0.167	0.214	0.214	0.192	0.192	0.192	0.191	0.184	0.184
(N)H-N	0.308	0.309	0.269	0.270	0.229	0.230	0.314	0.315	0.305	0.307	0.285	0.285
C-H(O)	0.041	0.041	0.045	0.045	0.015	0.015	0.022	0.023	0.027	0.028	0.030	0.031
$\Sigma\rho_{HB}$	**0.551**	**0.552**	**0.481**	**0.482**	**0.458**	**0.459**	**0.528**	**0.529**	**0.524**	**0.526**	**0.499**	**0.500**
$\Sigma\rho_{HB}$S-U	**-0.001**		**-0.001**		**-0.001**		**-0.001**		**-0.001**		**-0.001**	
$\Sigma\rho_{\pi}$	**0.314**				**0.314**				**0.280**			

Table S5.2

	A12_T9_T13_A8				A18_T3_T19_A2				G14_C7_G15_C6			
126D	A12_T9		T13_A9		A18_T3		T19_A2		G14_C7		G15_C6	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.176	0.176	0.292	0.292	0.250	0.250	0.214	0.213	0.253	0.253	0.182	0.182
(N)H-N	0.329	0.330	0.266	0.266	0.330	0.330	0.405	0.405	0.368	0.369	0.191	0.192
C/N-H(O)	0.034	0.034	0.018	0.018	0.025	0.026	0.046	0.046	0.334	0.332	0.120	0.122
$\Sigma\rho_{HB}$	**0.539**	**0.540**	**0.577**	**0.577**	**0.606**	**0.606**	**0.665**	**0.665**	**0.955**	**0.955**	**0.494**	**0.496**
$\Sigma\rho_{HB}$S-U	**-0.001**		**0.000**		**0.000**		**0.000**		**0.000**		**-0.002**	
$\Sigma\rho_{\pi}$	**0.351**				**0.366**				**0.276**			

	C11_G10_T9_A12				C16_G5_C17_G4				C17_G4_A18_T3			
126D	C11_G10		T9_A12		C16_G5		C17_G4		C17_G4		A18_T3	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.188	0.188	0.177	0.176	0.376	0.377	0.182	0.183	0.183	0.183	0.251	0.250
(N)H-N	0.368	0.368	0.329	0.330	0.359	0.361	0.202	0.203	0.204	0.203	0.330	0.330
C/N-H(O)	0.413	0.415	0.034	0.034	0.181	0.184	0.229	0.228	0.228	0.228	0.026	0.026
$\Sigma\rho_{HB}$	**0.969**	**0.971**	**0.541**	**0.540**	**0.917**	**0.922**	**0.613**	**0.613**	**0.614**	**0.613**	**0.606**	**0.606**
$\Sigma\rho_{HB}$S-U	**-0.002**		**0.001**		**-0.005**		**0.000**		**0.001**		**0.000**	
$\Sigma\rho_{\pi}$	**0.282**				**0.286**				**0.157**			

	G15_C6_C16_G5				T13_A8_G14_C7				T19_A2_G20_C1			
126D	G15_C6		C16_G5		T13_A8		G14_C7		T19_A2		G20_C1	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.183	0.182	0.378	0.377	0.293	0.292	0.255	0.253	0.215	0.213	0.140	0.140
(N)H-N	0.193	0.192	0.362	0.361	0.266	0.266	0.371	0.369	0.404	0.405	0.235	0.234
C/N-H(O)	0.121	0.122	0.183	0.184	0.018	0.019	0.332	0.332	0.046	0.046	0.307	0.309
$\Sigma\rho_{HB}$	**0.497**	**0.496**	**0.923**	**0.922**	**0.578**	**0.577**	**0.958**	**0.955**	**0.665**	**0.665**	**0.682**	**0.683**
$\Sigma\rho_{HB}$S-U	**0.002**		**0.001**		**0.001**		**0.003**		**0.000**		**0.001**	
$\Sigma\rho_{\pi}$	**0.328**				**0.263**				**0.280**			

Table S5.3

	A16_T5_A17_T4				A17_T4_A18_T3				A18_T3_A19_T2			
1SK5	A16_T5		A17_T4		A17_T4		A18_T3		A18_T3		A19_T2	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.182	0.181	0.157	0.157	0.156	0.157	0.169	0.169	0.168	0.169	0.175	0.175
(N)H-N	0.300	0.299	0.300	0.298	0.299	0.298	0.344	0.342	0.342	0.342	0.311	0.310
C-H(O)	0.035	0.035	0.039	0.040	0.040	0.040	0.039	0.039	0.038	0.039	0.040	0.040
$\Sigma\rho_{HB}$	**0.517**	**0.516**	**0.496**	**0.495**	**0.495**	**0.495**	**0.551**	**0.549**	**0.549**	**0.549**	**0.527**	**0.525**
$\Sigma\rho_{HB}$S-U	**0.001**		**0.001**		**0.000**		**0.002**		**0.000**		**0.002**	
$\Sigma\rho_{\pi}$	**0.388**				**0.436**				**0.476**			

	A19_T2_G20_C1				C11_G10_T12_A9				T14_A7_T15_A6			
1SK5	A19_T2		G20_C1		C11_G10		T12_A9		T14_A7		T15_A6	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.175	0.175	0.189	0.264	0.196	0.231	0.179	0.179	0.160	0.160	0.160	0.161
(N)H-N	0.311	0.310	0.253	0.253	0.240	0.241	0.336	0.334	0.298	0.303	0.305	0.297
C/N-H(O)	0.040	0.040	0.264	0.189	0.231	0.196	0.043	0.043	0.036	0.036	0.036	0.036
$\Sigma\rho_{HB}$	**0.527**	**0.525**	**0.705**	**0.705**	**0.668**	**0.668**	**0.558**	**0.557**	**0.494**	**0.500**	**0.502**	**0.495**
$\Sigma\rho_{HB}$S-U	**0.002**		**0.000**		**0.000**		**0.002**		**0.006**		**0.007**	
$\Sigma\rho_{\pi}$	**0.291**				**0.248**				**0.485**			

	T12_A9_T13_A8				T13_A8_T14_A7				T15_A6_A16_T5			
1SK5	T12_A9		T13_A8		T13_A8		T14_A7		T15_A6		A16_T5	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.179	0.179	0.176	0.176	0.176	0.176	0.160	0.160	0.161	0.161	0.181	0.181
(N)H-N	0.336	0.334	0.310	0.310	0.312	0.310	0.304	0.303	0.297	0.297	0.299	0.299
C-H(O)	0.043	0.043	0.035	0.035	0.035	0.035	0.036	0.036	0.036	0.036	0.035	0.035
$\Sigma\rho_{HB}$	**0.559**	**0.557**	**0.521**	**0.521**	**0.523**	**0.521**	**0.499**	**0.500**	**0.494**	**0.495**	**0.515**	**0.516**
$\Sigma\rho_{HB}$S-U	**0.002**		**0.000**		**0.002**		**0.000**		**-0.001**		**-0.001**	
$\Sigma\rho_{\pi}$	**0.311**				**0.360**				**0.335**			

Table S5.4

	A4_T13_T5_A12				A6_T11_C7_G10				C7_G10_C8_G9			
1VJ4	A4_T13	A4_T13	T5_A12	T5_A12	A6_T11	A6_T11	C7_G10	C7_G10	C7_G10	C7_G10	C8_G9	C8_G9
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.187	0.187	0.173	0.173	0.197	0.197	0.319	0.319	0.318	0.319	0.228	0.229
(N)H-N	0.358	0.359	0.391	0.392	0.281	0.283	0.263	0.262	0.261	0.262	0.212	0.212
C/N-H(O)	0.044	0.044	0.052	0.053	0.025	0.025	0.167	0.168	0.168	0.168	0.179	0.180
$\Sigma\rho_{HB}$	**0.588**	**0.590**	**0.617**	**0.618**	**0.503**	**0.506**	**0.749**	**0.748**	**0.748**	**0.748**	**0.619**	**0.621**
$\Sigma\rho_{HB}$S-U	**-0.001**		**-0.001**		**-0.003**		**0.001**		**0.000**		**-0.002**	
$\Sigma\rho_{\pi}$	**0.368**				**0.385**				**0.284**			

	G2_C15_T3_A14				T3_A14_A4_T13				T5_A12_A6_T11			
1VJ4	G2_C15	G2_C15	T3_A14	T3_A14	T3_A14	T3_A14	A4_T13	A4_T13	T5_A12	T5_A12	A6_T11	A6_T11
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.429	0.429	0.255	0.255	0.255	0.255	0.187	0.187	0.173	0.173	0.197	0.197
(N)H-N	0.325	0.324	0.310	0.313	0.314	0.313	0.359	0.359	0.393	0.392	0.284	0.283
C/N-H(O)	0.206	0.206	0.028	0.028	0.028	0.028	0.044	0.044	0.053	0.053	0.025	0.025
$\Sigma\rho_{HB}$	**0.960**	**0.959**	**0.594**	**0.596**	**0.597**	**0.596**	**0.590**	**0.590**	**0.619**	**0.618**	**0.507**	**0.506**
$\Sigma\rho_{HB}$S-U	**0.001**		**-0.002**		**0.001**		**0.000**		**0.001**		**0.001**	
$\Sigma\rho_{\pi}$	**0.311**				**0.227**				**0.257**			

	G1_C16_G2_C15			
1VJ4	G2_C15	G2_C15	G1_C16	G1_C16
	S	U	S	U
N-H(O)	0.428	0.429	0.218	0.219
(N)H-N	0.323	0.324	0.303	0.304
C/N-H(O)	0.207	0.206	0.382	0.382
$\Sigma\rho_{HB}$	**0.958**	**0.959**	**0.903**	**0.905**
$\Sigma\rho_{HB}$S-U	**-0.001**		**-0.002**	
$\Sigma\rho_{\pi}$	**0.295**			

Table S5.5

259D	C2_G15_C1_G16 C2_G15 S	 C2_G15 U	 C1_G16 S	 C1_G16 U	C3_G14_C2_G15 C3_G14 S	 C3_G14 U	 C2_G15 S	 C2_G15 U	C4_G13_C3_G14 C4_G13 S	 U	 C3_G14 S	 C3_G14 U
N-H(O)	0.233	0.234	0.291	0.291	0.281	0.281	0.234	0.234	0.253	0.254	0.281	0.281
(N)H-N	0.275	0.275	0.250	0.251	0.262	0.262	0.274	0.275	0.298	0.298	0.261	0.262
N-H(O)	0.262	0.261	0.257	0.256	0.272	0.272	0.262	0.261	0.254	0.253	0.273	0.272
$\Sigma\rho_{HB}$	**0.771**	**0.770**	**0.798**	**0.798**	**0.815**	**0.815**	**0.770**	**0.770**	**0.805**	**0.805**	**0.815**	**0.815**
$\Sigma\rho_{HB}$S-U	**0.001**		**0.000**		**0.000**		**0.000**		**0.001**		**0.000**	
$\Sigma\rho_{\pi}$	**0.317**				**0.297**				**0.291**			

259D	G5_C12_C4_G13 G5_C12 S	 G5_C12 U	 C4_G13 S	 C4_G13 U	G6_C11_G5_C12 G6_C11 S	 G6_C11 U	 G5_C12 S	 G5_C12 U	G7_C10_G6_C11 G7_C10 S	 G7_C10 U	 G6_C11 S	 G6_C11 U
N-H(O)	0.309	0.308	0.254	0.254	0.237	0.237	0.308	0.308	0.292	0.292	0.237	0.237
(N)H-N	0.224	0.223	0.298	0.298	0.247	0.248	0.223	0.223	0.331	0.333	0.248	0.248
N-H(O)	0.158	0.160	0.251	0.253	0.211	0.210	0.160	0.160	0.371	0.370	0.210	0.210
$\Sigma\rho_{HB}$	**0.691**	**0.691**	**0.804**	**0.805**	**0.695**	**0.695**	**0.691**	**0.691**	**0.994**	**0.994**	**0.695**	**0.695**
$\Sigma\rho_{HB}$S-U	**0.000**		**-0.001**		**0.000**		**0.000**		**0.000**		**0.000**	
$\Sigma\rho_{\pi}$	**0.291**				**0.276**				**0.309**			

259D	G8_C9_G7_C10 G8_C9 S	 G8_C9 U	 G7_C10 S	 G7_C10 U
N-H(O)	0.334	0.334	0.291	0.292
(N)H-N	0.288	0.289	0.333	0.333
N-H(O)	0.253	0.252	0.371	0.370
$\Sigma\rho_{HB}$	**0.874**	**0.875**	**0.995**	**0.994**
$\Sigma\rho_{HB}$S-U	**0.001**		**0.001**	
$\Sigma\rho_{\pi}$	**0.288**			

Table S5.6

420D	A3_U30_G4_C29				A5_U28_G6_A27				A9_U24_A10_U23			
	G4_C29	G4_C29	A3_U30	A3_U30	A5_U28	A5_U28	G6_A27	G6_A27	A9_U24	A9_U24	A10_U23	A10_U23
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.269	0.269	0.172	0.173	0.139	0.139	0.293	0.292	0.182	0.182	0.200	0.201
(N)H-N	0.351	0.351	0.396	0.394	0.323	0.325	0.159	0.158	0.518	0.518	0.414	0.412
C/N-H(O)	0.332	0.331	0.045	0.045	0.046	0.046			0.083	0.083	0.056	0.056
$\Sigma\rho_{HB}$	**0.952**	**0.951**	**0.613**	**0.612**	**0.507**	**0.510**	**0.452**	**0.451**	**0.783**	**0.784**	**0.671**	**0.669**
$\Sigma\rho_{HB}$S-U	**0.001**		**0.001**		**-0.003**		**0.001**		**-0.001**		**0.002**	
$\Sigma\rho_{\pi}$	**0.246**				**0.400**				**0.253**			

420D	C2_G31_A3_U30				C13_G20_U14_A19				G1_C32_C2_G31			
	C2_G31	C2_G31	A3_U30	A3_U30	C13_G20	C13_G20	U14_A19	U14_A19	G1_C32	G1_C32	C2_G31	C2_G31
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.497	0.496	0.173	0.173	0.213	0.212	0.247	0.248	0.216	0.215	0.498	0.496
(N)H-N	0.408	0.407	0.392	0.394	0.268	0.268	0.499	0.497	0.196	0.195	0.409	0.407
C/N-H(O)	0.225	0.226	0.045	0.045	0.271	0.271	0.057	0.057	0.112	0.114	0.224	0.226
$\Sigma\rho_{HB}$	**1.130**	**1.130**	**0.611**	**0.612**	**0.752**	**0.750**	**0.803**	**0.802**	**0.524**	**0.524**	**1.130**	**1.130**
$\Sigma\rho_{HB}$S-U	**0.000**		**-0.001**		**0.002**		**0.001**		**0.000**		**0.000**	
$\Sigma\rho_{\pi}$	**0.192**				**0.250**				**0.340**			

420D	A10_U23_A11_G22				A11_G22_U12_A21				G6_A27_U7_A26			
	A10_U23	A10_U23	A11_G22	A11_G22	A11_G22	A11_G22	U12_A21	U12_A21	G6_A27	G6_A27	U7_A26	U7_A26
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.199	0.201	0.263	0.262	0.263	0.262	0.097	0.098	0.293	0.292	0.180	0.182
(N)H-N	0.412	0.412	0.180	0.179	0.180	0.179	0.223	0.225	0.159	0.158	0.397	0.398
C/N-H(O)	0.056	0.056					0.031	0.031			0.059	0.060
$\Sigma\rho_{HB}$	**0.666**	**0.669**	**0.442**	**0.441**	**0.443**	**0.441**	**0.351**	**0.353**	**0.452**	**0.451**	**0.637**	**0.639**
$\Sigma\rho_{HB}$S-U	**-0.003**		**0.001**		**0.002**		**-0.002**		**0.001**		**-0.003**	
$\Sigma\rho_{\pi}$	**0.271**				**0.377**				**0.249**			

420D	G4_C29_A5_U28				G17_C16_C18_G15				U7_A26_U8_A25			
	G4_C29	G4_C29	A5_U28	A5_U28	C18_G15	G17_C16	C18_G15	C18_G15	U7_A26	U7_A26	U8_A25	U8_A25
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.269	0.269	0.138	0.139	0.196	0.195	0.422	0.421	0.182	0.182	0.215	0.215
(N)H-N	0.352	0.351	0.329	0.325	0.168	0.167	0.253	0.251	0.400	0.398	0.456	0.456
C/N-H(O)	0.330	0.331	0.046	0.046	0.115	0.117	0.108	0.109	0.060	0.060	0.062	0.062
$\Sigma\rho_{HB}$	**0.952**	**0.951**	**0.512**	**0.510**	**0.479**	**0.480**	**0.782**	**0.781**	**0.642**	**0.639**	**0.732**	**0.733**
$\Sigma\rho_{HB}$S-U	**0.000**		**0.003**		**0.000**		**0.001**		**0.003**		**-0.001**	
$\Sigma\rho_{\pi}$	**0.255**				**0.313**				**0.270**			

420D	U12_A21_C13_G20				U14_A19_G15_C18				U8_A25_A9_U24			
	U12_A21	U12_A21	C13_G20	C13_G20	U14_A19	U14_A19	G15_C18	G15_C18	U8_A25	U8_A25	A9_U24	A9_U24
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.097	0.098	0.212	0.212	0.248	0.248	0.421	0.421	0.215	0.215	0.182	0.182
(N)H-N	0.228	0.225	0.269	0.268	0.495	0.497	0.252	0.251	0.457	0.456	0.519	0.518
C/N-H(O)	0.031	0.031	0.270	0.271	0.057	0.057	0.109	0.109	0.062	0.062	0.083	0.083
$\Sigma\rho_{HB}$	**0.356**	**0.353**	**0.751**	**0.750**	**0.801**	**0.802**	**0.782**	**0.781**	**0.734**	**0.733**	**0.784**	**0.784**
$\Sigma\rho_{HB}$S-U	**0.003**		**0.001**		**-0.001**		**0.001**		**0.001**		**0.001**	
$\Sigma\rho_{\pi}$	**0.263**				**0.276**				**0.214**			

Table S5.7

	A1_T20_G2_C19				C6_G15_C7_G14				C8_G13_C9_G12			
440D	A1_T20		G2_C19		C6_G15		C7_G14		C8_G13		C9_G12	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.142	0.142	0.481	0.481	0.189	0.189	0.301	0.301	0.273	0.273	0.320	0.321
(N)H-N	0.278	0.276	0.374	0.374	0.219	0.219	0.241	0.241	0.265	0.266	0.280	0.281
C/N-H(O)	0.046	0.046	0.186	0.186	0.197	0.197	0.241	0.240	0.282	0.280	0.205	0.204
$\Sigma\rho_{HB}$	**0.466**	**0.464**	**1.041**	**1.041**	**0.605**	**0.605**	**0.783**	**0.782**	**0.819**	**0.819**	**0.805**	**0.805**
$\Sigma\rho_{HB}$S-U	**0.002**		**0.000**		**0.000**		**0.001**		**0.000**		**0.000**	
$\Sigma\rho_{\pi}$	**0.328**				**0.304**				**0.258**			

	C9_G12_A11_T10				G2_C19_G3_C18				G3_C18_G4_C17			
440D	C9_G12		A11_T10		G2_C19		G3_C18		G3_C18		G4_C17	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.321	0.321	0.201	0.202	0.480	0.481	0.250	0.250	0.250	0.250	0.216	0.216
(N)H-N	0.280	0.281	0.277	0.275	0.374	0.374	0.238	0.239	0.239	0.239	0.244	0.245
C/N-H(O)	0.204	0.204	0.025	0.025	0.187	0.186	0.208	0.207	0.209	0.207	0.290	0.290
$\Sigma\rho_{HB}$	**0.805**	**0.805**	**0.502**	**0.502**	**1.041**	**1.041**	**0.697**	**0.697**	**0.698**	**0.697**	**0.750**	**0.751**
$\Sigma\rho_{HB}$S-U	**0.000**		**0.000**		**0.000**		**0.000**		**0.001**		**-0.001**	
$\Sigma\rho_{\pi}$	**0.279**				**0.332**				**0.316**			

	G4_C17_G5_C16				G5_C16_C6_G15			
440D	G4_C17		G5_C16		G5_C16		C6_G15	
	S	U	S	U	S	U	S	U
N-H(O)	0.216	0.216	0.239	0.239	0.240	0.239	0.189	0.189
(N)H-N	0.245	0.245	0.283	0.284	0.286	0.284	0.220	0.219
(O)H-N	0.290	0.290	0.303	0.301	0.297	0.301	0.195	0.197
$\Sigma\rho_{HB}$	**0.751**	**0.751**	**0.824**	**0.824**	**0.823**	**0.824**	**0.604**	**0.605**
$\Sigma\rho_{HB}$S-U	**0.000**		**0.000**		**-0.001**		**-0.001**	
$\Sigma\rho_{\pi}$	**0.270**				**0.288**			

Table S5.8

	A5_U15_U6_A14				C7_G13_G8_U12				G2_C18_U3_G17			
485D	U6_A14		A5_U15		C7_G13		G8_U12		G2_C18		U3_G17	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.177	0.176	0.204	0.205	0.246	0.245	0.229	0.230	0.231	0.231	0.300	0.299
(N)H-N	0.306	0.307	0.248	0.247	0.289	0.288	0.198	0.199	0.246	0.245	0.277	0.280
C/N-H(O)	0.029	0.029	0.024	0.024	0.246	0.249			0.220	0.222		
$\Sigma\rho_{HB}$	**0.511**	**0.512**	**0.476**	**0.476**	**0.782**	**0.782**	**0.427**	**0.429**	**0.697**	**0.698**	**0.578**	**0.579**
$\Sigma\rho_{HB}$S-U	**-0.001**		**0.000**		**0.000**		**-0.002**		**-0.001**		**-0.001**	
$\Sigma\rho_{\pi}$	**0.341**				**0.230**				**0.313**			

	G8_U12_C9_G11				U3_G17_G4_C16				U6_A14_C7_G13			
485D	G8_U12		C9_G11		U3_G17		G4_C16		U6_A14		C7_G13	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.232	0.230	0.276	0.275	0.300	0.299	0.269	0.268	0.175	0.176	0.245	0.245
(N)H-N	0.197	0.199	0.246	0.244	0.279	0.280	0.238	0.239	0.310	0.307	0.289	0.288
C/N-H(O)			0.204	0.206			0.186	0.188	0.029	0.029	0.248	0.249
$\Sigma\rho_{HB}$	**0.429**	**0.429**	**0.726**	**0.725**	**0.578**	**0.579**	**0.693**	**0.696**	**0.514**	**0.512**	**0.782**	**0.782**
$\Sigma\rho_{HB}$S-U	**-0.001**		**0.001**		**-0.001**		**-0.003**		**0.002**		**0.000**	
$\Sigma\rho_{\pi}$	**0.279**				**0.214**				**0.269**			

Table S5.9

	A4_T21_G5_C20				A6_T19_T7_A18				A10_T15_C11_G14			
119D	A4_T21		G5_C20		A6_T19		T7_A18		A10_T15		C11_G14	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.270	0.271	0.422	0.508	0.466	0.465	0.273	0.272	0.670	0.671	0.489	0.489
(N)H-N	0.432	0.429	0.470	0.470	0.525	0.527	0.581	0.583	0.971	0.972	0.406	0.406
C/N-H(O)	0.051	0.051	0.508	0.423	0.023	0.023	0.035	0.036	0.080	0.080	0.236	0.236
$\Sigma\rho_{HB}$	**0.754**	**0.751**	**1.399**	**1.400**	**1.013**	**1.015**	**0.889**	**0.890**	**1.721**	**1.723**	**1.131**	**1.130**
$\Sigma\rho_{HB}$S-U	**0.003**		**-0.001**		**-0.002**		**-0.002**		**-0.002**		**0.001**	
$\Sigma\rho_{\pi}$	**0.370**				**0.422**				**0.235**			

	G5_C20_A6_T19				T3_A22_A4_T21				T7_A18_C8_G17			
119D	G5_C20		A6_T19		T3_A22		A4_T21		T7_A18		C8_G17	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.509	0.508	0.463	0.465	0.239	0.239	0.271	0.271	0.270	0.272	0.304	0.303
(N)H-N	0.470	0.470	0.529	0.527	0.804	0.805	0.428	0.429	0.586	0.583	0.390	0.390
N/C-H(O)	0.423	0.423	0.023	0.023	0.093	0.093	0.051	0.051	0.035	0.036	0.199	0.199
$\Sigma\rho_{HB}$	**1.401**	**1.400**	**1.015**	**1.015**	**1.136**	**1.137**	**0.751**	**0.751**	**0.892**	**0.890**	**0.892**	**0.892**
$\Sigma\rho_{HB}$S-U	**0.001**		**0.000**		**0.001**		**-0.000**		**0.002**		**0.000**	
$\Sigma\rho_{\pi}$	**0.260**				**0.182**				**0.278**			

	C8_G17_T9_A16				C11_G14_G12_C13				G2_C23_T3_A22			
119D	C8_G17		T9_A16		C11_G14		G12_C13		G2_C23		T3_A22	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.304	0.303	0.295	0.295	0.489	0.489	0.256	0.255	0.710	0.711	0.237	0.239
(N)H-N	0.389	0.390	0.491	0.490	0.407	0.406	0.339	0.337	0.504	0.503	0.801	0.805
N/C-H(O)	0.198	0.199	0.040	0.041	0.235	0.236	0.374	0.375	0.186	0.186	0.093	0.093
$\Sigma\rho_{HB}$	**0.892**	**0.892**	**0.827**	**0.825**	**1.132**	**1.130**	**0.969**	**0.967**	**1.400**	**1.399**	**1.131**	**1.137**
$\Sigma\rho_{HB}$S-U	**0.000**		**0.002**		**0.002**		**0.002**		**0.001**		**-0.006**	
$\Sigma\rho_{\pi}$	**0.377**				**0.167**				**0.335**			

	C1_G24_G2_C23				T9_A16_A10_T15			
119D	C1_G24		G2_C23		T9_A16		A10_T15	
	S	U	S	U	S	U	S	U
N-H(O)	0.279	0.279	0.712	0.711	0.296	0.295	0.672	0.671
(N)H-N	0.374	0.373	0.506	0.503	0.489	0.490	0.968	0.972
N/C-H(O)	0.255	0.254	0.186	0.186	0.041	0.041	0.080	0.080
$\Sigma\rho_{HB}$	**0.907**	**0.906**	**1.404**	**1.399**	**0.826**	**0.825**	**1.721**	**1.723**
$\Sigma\rho_{HB}$S-U	**0.001**		**0.005**		**0.001**		**-0.002**	
$\Sigma\rho_{\pi}$	**0.327**				**0.420**			

Table S5.10

157D	A5 U20 G4 A21				A6 U19 A5 U20				A9 G16 U8 A17			
	A5 U20		G4 A21		A6 U19		A5 U20		A9 G16		U8 A17	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.485	0.486	0.247	0.247	0.340	0.340	0.483	0.486	0.047	0.047	0.480	0.481
(N)H-N	0.409	0.407	0.488	0.487	0.905	0.902	0.408	0.407	0.224	0.224	0.288	0.287
C/N-H(O/N)	0.023	0.023	0.170	0.171	0.050	0.049	0.023	0.023	0.383	0.384		
$\Sigma\rho_{HB}$	**0.918**	**0.916**	**0.905**	**0.904**	**1.294**	**1.291**	**0.914**	**0.916**	**0.654**	**0.655**	**0.768**	**0.767**
$\Sigma\rho_{HB}$S-U	**0.002**		**0.001**		**0.003**		**-0.002**		**0.001**		**0.001**	
$\Sigma\rho_{\pi}$	**0.264**				**0.244**				**0.216**			

157D	G12 C13 C11 G14				U8 A17 U7 A18				C3 G22 G2 C23			
	G12 C13		C11 G14		U8 A17		U7 A18		C3 G22		G2 C23	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.272	0.271	0.342	0.341	0.479	0.481	0.414	0.414	0.273	0.273	0.328	0.326
(N)H-N	0.348	0.347	0.353	0.352	0.286	0.287	0.798	0.795	0.370	0.368	0.339	0.336
C/N-H(O)	0.403	0.405	0.163	0.164			0.041	0.040	0.252	0.255	0.185	0.188
$\Sigma\rho_{HB}$	**1.023**	**1.023**	**0.859**	**0.857**	**0.765**	**0.767**	**1.253**	**1.250**	**0.896**	**0.895**	**0.852**	**0.850**
$\Sigma\rho_{HB}$S-U	**0.000**		**0.001**		**-0.002**		**0.004**		**0.001**		**0.002**	
$\Sigma\rho_{\pi}$	**0.341**				**0.200**				**0.271**			

157D	C11 G14 G10 C15				G2 C23 C1 G24				G4 A21 C3 G22			
	C11 G14		G10 C15		G2 C23		C1 G24		G4 A21		C3 G22	
	S	U	S	U	S	U	S	U	S	U	S	U
N-H(O)	0.343	0.341	0.203	0.202	0.327	0.326	0.611	0.610	0.247	0.247	0.274	0.273
(N)H-N	0.354	0.352	0.272	0.270	0.337	0.336	0.372	0.371	0.489	0.487	0.370	0.368
C/N-H(O/N)	0.161	0.164	0.192	0.194	0.186	0.188	0.216	0.218	0.171	0.171	0.253	0.255
$\Sigma\rho_{HB}$	**0.859**	**0.857**	**0.666**	**0.667**	**0.850**	**0.850**	**1.200**	**1.199**	**0.906**	**0.904**	**0.897**	**0.895**
$\Sigma\rho_{HB}$S-U	**0.002**		**-0.001**		**0.000**		**0.001**		**0.002**		**0.002**	
$\Sigma\rho_{\pi}$	**0.262**				**0.317**				**0.334**			

157D	G10 C15 A9 G19			
	G10 C15		A9 G19	
	S	U	S	U
N-H(O)	0.203	0.202	0.047	0.047
(N)H-N	0.271	0.270	0.224	0.224
C/N-H(O)	0.193	0.194	0.383	0.384
$\Sigma\rho_{HB}$	**0.667**	**0.667**	**0.654**	**0.655**
$\Sigma\rho_{HB}$S-U	**0.000**		**-0.001**	
$\Sigma\rho_{\pi}$	**0.219**			

Table S5.11

Figure S5.1

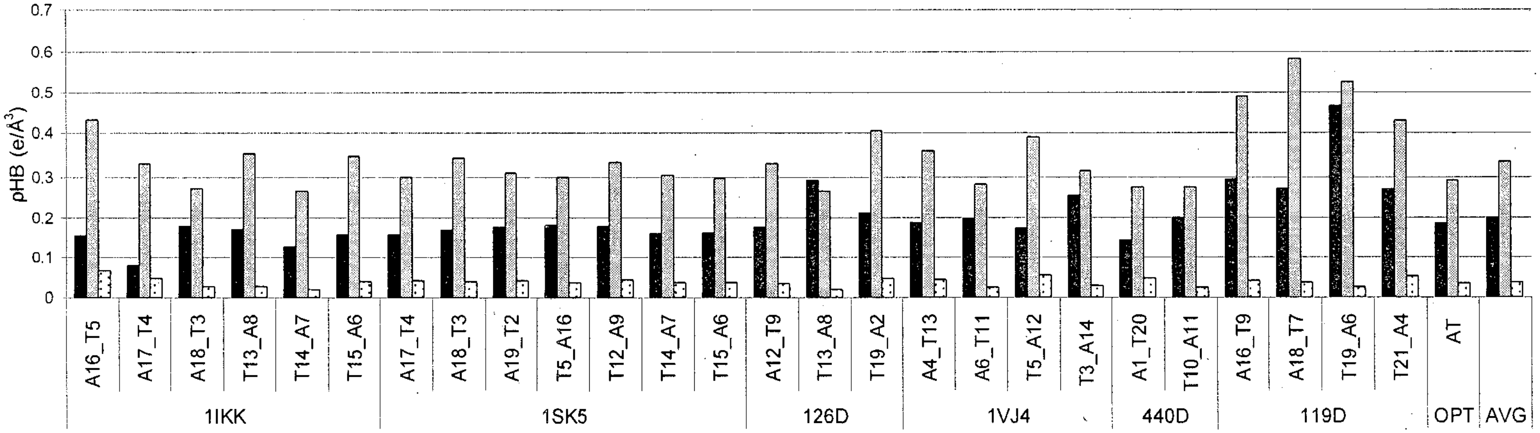


Figure S5.2

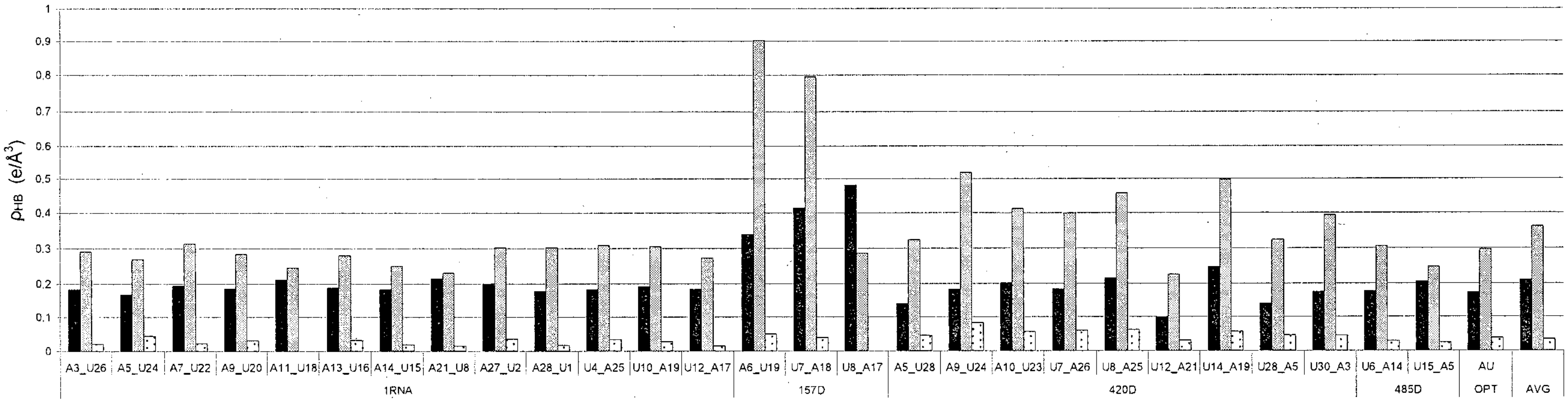


Figure S5.3

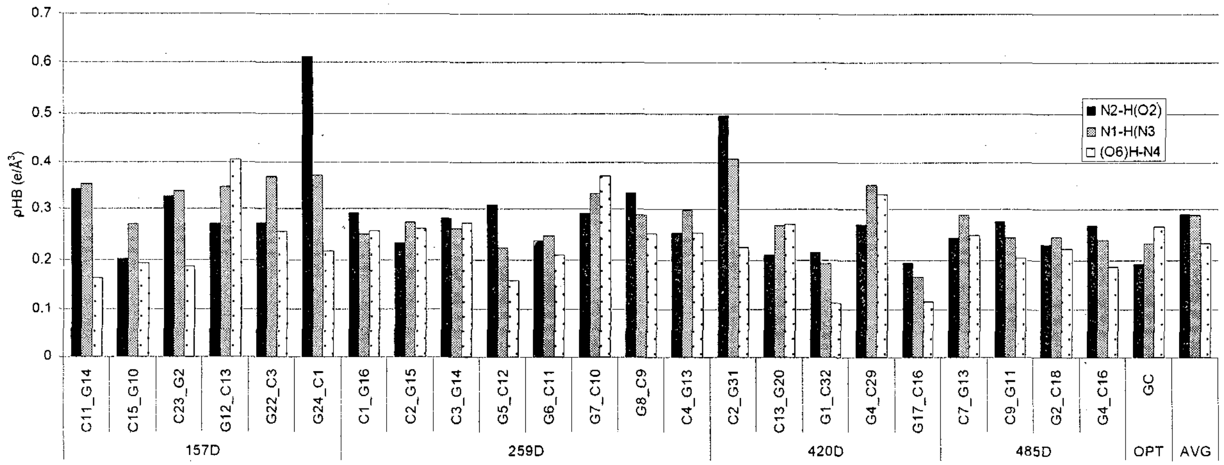


Figure S5.4

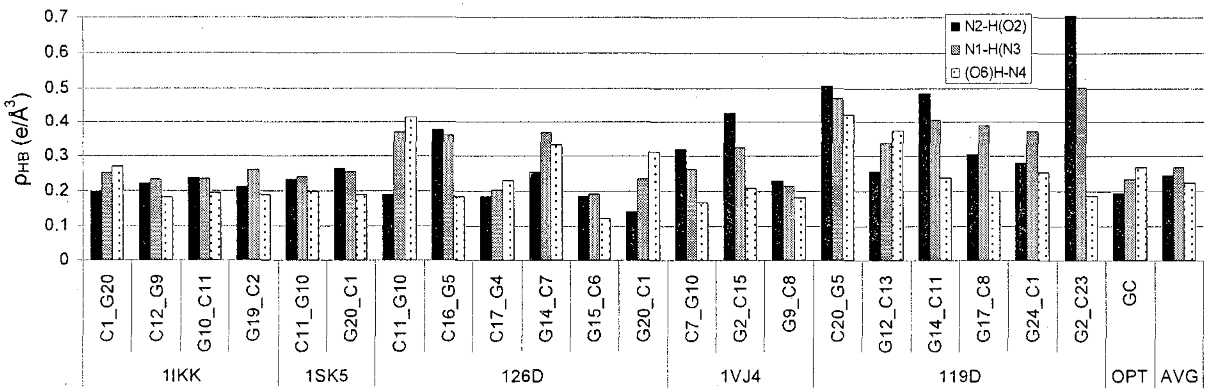


Figure S5.5

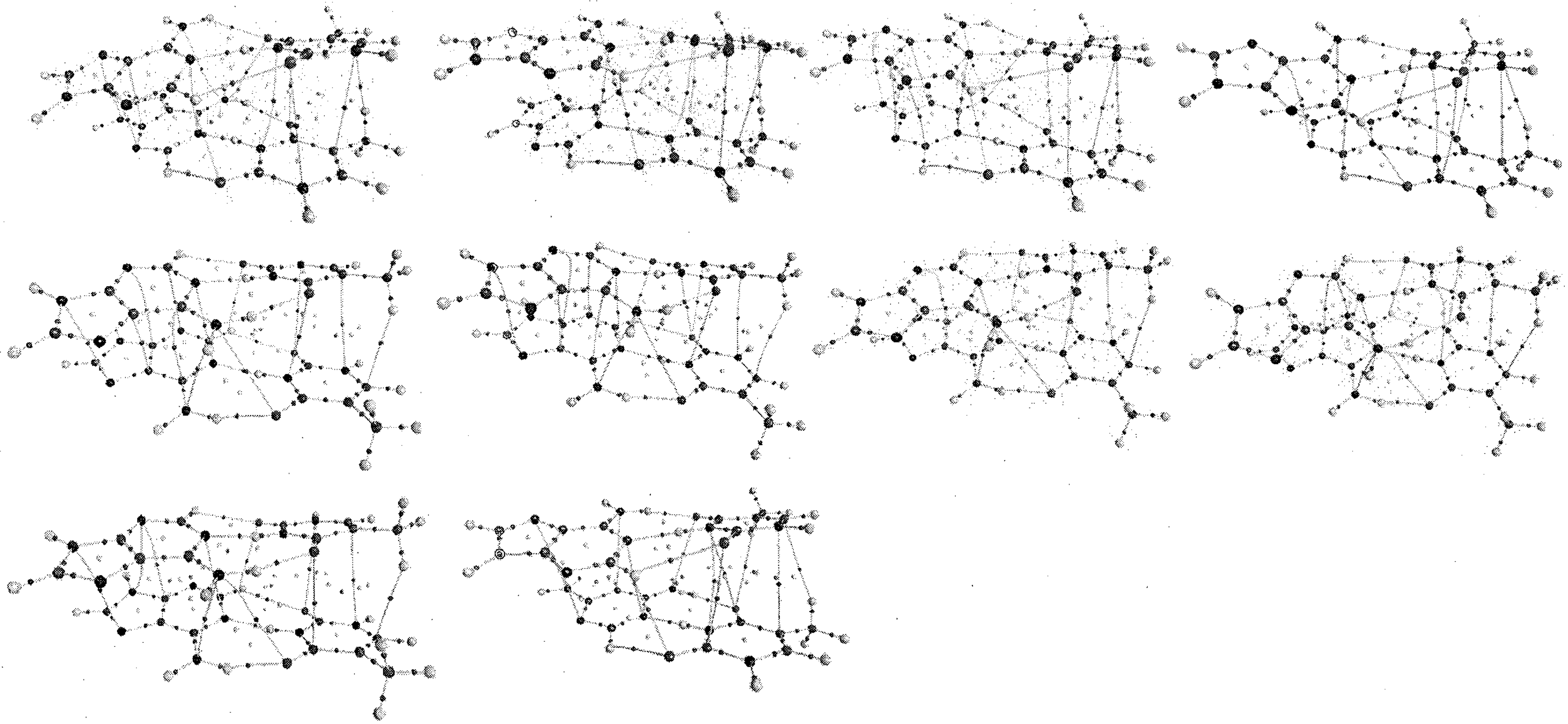


Figure S5.6

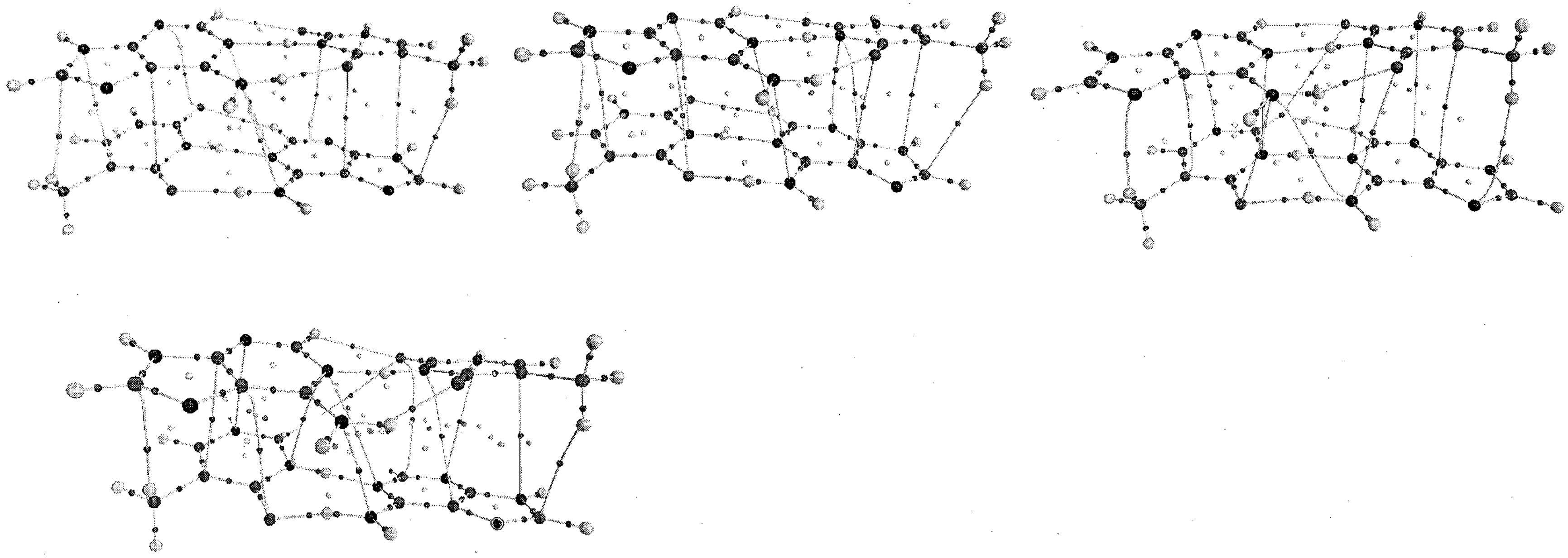
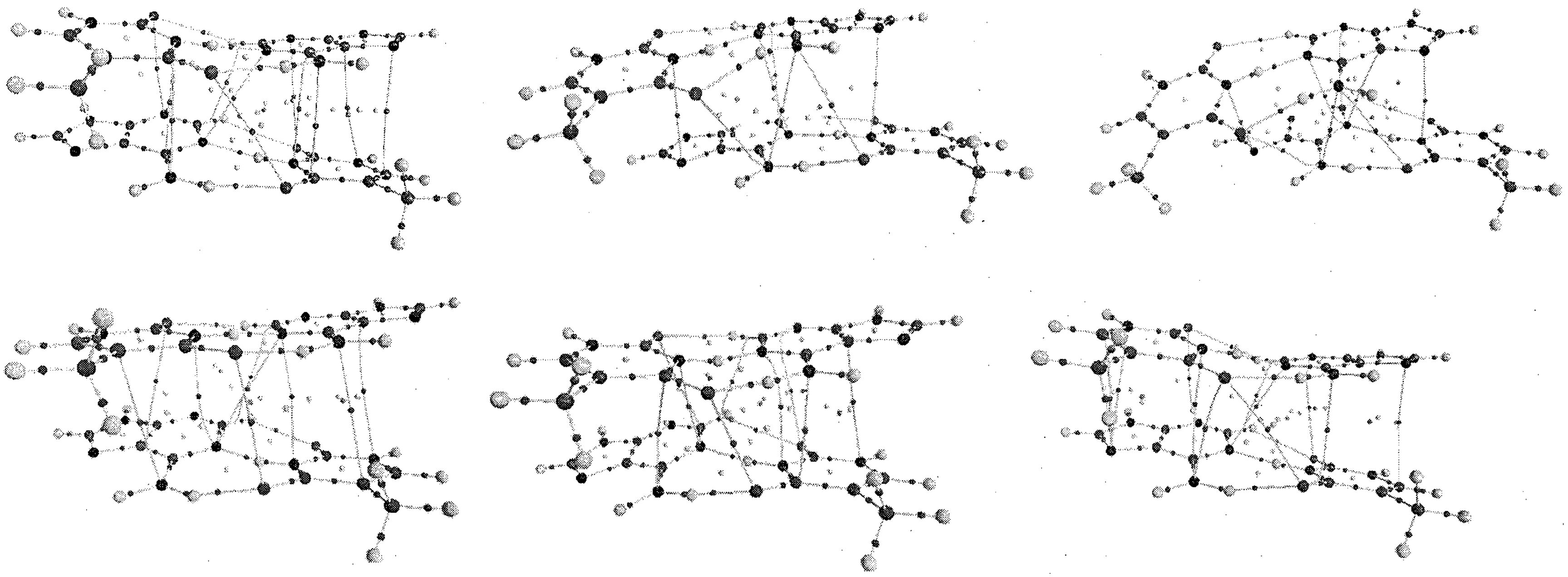


Figure S5.7



Appendix E

Supporting information to Chapter 6

An explanation of the thermodynamic behavior of tandem G•U pairs using the electron density

Table S6.1 ϱ_{π} and $\Sigma\varrho_{\pi}$ (in bold) and $\nabla^2\varrho$ at the bond critical points between stacked base pairs isolated from 1EKA (PBE0/6-31+G(d,p))

$G_1A_2G_3U_4G_5C_6U_7C_8$
$C_8U_7C_6G_5U_4G_3A_2G_1$

G1_C8/A2_U7		A2_U7/G3_C6		G3_C6/U4_G5		U4_G5/G5_U4		G5_U4/C6_G3		C6_G3/U7_A2		U7A2/C8_G1	
ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)
0.019	0.18	0.029	0.31	0.037	0.34	0.011	0.15	0.036	0.33	0.025	0.29	0.024	0.28
0.045	0.48	0.032	0.33	0.029	0.32	0.035	0.35	0.029	0.32	0.031	0.32	0.035	0.35
0.032	0.38	0.049	0.49	0.034	0.45	0.038	0.35	0.036	0.47	0.048	0.47	0.045	0.50
0.015	0.22	0.040	0.44	0.037	0.49	0.043	0.43	0.039	0.52	0.041	0.46	0.045	0.43
0.046	0.50	0.032	0.39	0.020	0.28	0.037	0.35	0.022	0.30	0.032	0.38	0.024	0.27
0.043	0.45	0.041	0.41	0.012	0.22	0.025	0.27	0.033	0.39	0.045	0.44	0.045	0.48
0.032	0.43	0.041	0.47	0.032	0.37	0.015	0.19	0.015	0.24	0.042	0.48	0.043	0.44
0.048	0.46	0.029	0.30	0.040	0.39	0.012	0.16	0.015	0.25	0.032	0.32	0.030	0.40
0.026	0.29			0.033	0.38			0.037	0.37			0.039	0.38
0.034	0.34			0.033	0.36			0.030	0.36			0.021	0.21
0.025	0.30			0.036	0.31			0.039	0.31				
								0.032	0.36				
0.367		**0.294**		**0.345**		**0.216**		**0.363**		**0.298**		**0.351**	

Table S6.2 ϱ_π and $\Sigma\varrho_\pi$ (in bold), and $\nabla^2\varrho_\pi$ at bond critical points between stacked base pairs isolated from 1FFK (PBE0/6-31+G(d,p)).

$U_{647}G_{648}U_{649}C_{650}U_{651}G_{652}C_{653}A_{654}$
$A_{758}C_{757}A_{756}C_{755}G_{754}U_{753}G_{752}U_{751}$

U647_A758/G648_C757		G648_C757/U649_A756		U649_A756/C650_G755		C650_G755/U651_G754		U651_G754/G652_U753		G652_U753/C653_G752		A654_U751/U655_A750		U655_A750/G656_C749	
ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)
0.048	0.61	0.031	0.29	0.059	0.59	0.042	0.46	0.033	0.34	0.030	0.32	0.042	0.46	0.059	0.59
0.055	0.56	0.053	0.51	0.033	0.46	0.037	0.36	0.021	0.27	0.038	0.38	0.037	0.36	0.033	0.46
0.030	0.36	0.027	0.30	0.027	0.33	0.026	0.29	0.041	0.51	0.036	0.49	0.026	0.29	0.027	0.33
0.042	0.43	0.024	0.34	0.036	0.44	0.025	0.31	0.047	0.50	0.045	0.57	0.025	0.31	0.036	0.44
0.041	0.42	0.042	0.55	0.030	0.37	0.051	0.61	0.039	0.41	0.016	0.28	0.051	0.61	0.030	0.37
0.025	0.28	0.044	0.44	0.044	0.41	0.057	0.62	0.051	0.51	0.023	0.37	0.057	0.62	0.044	0.41
0.023		0.018	0.24	0.041	0.42			0.022	0.28	0.043	0.56			0.041	0.42
		0.037	0.37					0.042	0.40	0.040	0.51				
		0.017	0.21							0.038	0.39				
										0.042	0.40				
0.264		**0.295**		**0.270**		**0.238**		**0.295**		**0.352**		**0.238**		**0.270**	

Table S6.3 $\Sigma\rho_\pi$ (in bold), and $\Sigma\nabla^2\rho_\pi$ at bond critical points between stacked base pairs isolated from ten 1QET RNA duplexes (PBE0/6-31+G(d,p)).

$G_1 G_2 A_3 U_4\ G_5\ U_6\ C_7\ C_8$
$C_{16} C_{15} U_{14} G_{13} U_{12} A_{11} G_{10} G_9$

	G1_C16/G2_C15		G2_C15/A3_U14		A3_U14/U4_G13		U4_G13/G5_U12		G5_U12/U6_A11		U6_A11/C7_G10		G7_C10/C8_G9	
Models	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)
M1	0.479	5.12	0.355	3.78	0.381	4.36	0.510	5.50	0.380	4.37	0.372	3.97	0.474	5.04
M2	0.476	5.08	0.369	3.94	0.386	4.41	0.511	5.52	0.383	4.38	0.355	3.79	0.484	5.17
M3	0.482	5.17	0.359	3.83	0.445	5.09	0.506	5.48	0.376	4.30	0.437	4.79	0.492	5.26
M5	0.482	5.15	0.358	3.84	0.381	4.38	0.486	5.25	0.392	4.48	0.353	3.78	0.480	5.13
M10	0.499	5.46	0.371	3.95	0.386	4.41	0.504	5.46	0.381	4.36	0.371	3.94	0.499	5.46
M15	0.461	4.84	0.400	4.34	0.390	4.50	0.512	5.52	0.381	4.37	0.434	4.75	0.484	5.17
M19	0.498	4.84	0.370	3.94	0.382	4.37	0.503	5.44	0.381	4.35	0.365	3.87	0.499	5.49
M23	0.481	5.14	0.370	3.98	0.392	4.49	0.512	5.54	0.388	4.44	0.351	3.75	0.482	5.16
M25	0.481	5.14	0.370	3.98	0.392	4.49	0.512	5.54	0.388	4.44	0.351	3.75	0.482	5.16
M30	0.479	5.11	0.357	3.81	0.388	4.43	0.510	5.50	0.383	4.59	0.357	3.93	0.479	5.12
Average	**0.482**		**0.368**		**0.392**		**0.507**		**0.383**		**0.374**		**0.486**	

Table S6.4 ϱ_π and $\Sigma\varrho_\pi$ (in bold), and $\nabla^2\varrho_\pi$ at bond critical points between stacked base pairs isolated from 315D (PBE0/6-31+G(d,p)).

$G_1 U_2 A_3 U_4 G_5 U_6 A_7 dC_8$
$dC_{16}A_{15}U_{14}G_{13}U_{12}A_{11}U_{10}G_9$

A3U14_G5U12		G5U12_U6A11		U2A15_A3U14		U6A11_A7U10		U4G13_G5_U12		G1C16_U2A15		A7U10_C8G9	
ϱ(e/Å^3)	$\nabla^2\varrho$(e/Å^5)	ϱ(e/Å^3)	$\nabla^2\varrho$(e/Å^5)	ϱ(e/Å^3)	$\nabla^2\varrho$(e/Å^5)	ϱ(e/Å^3)	$\nabla^2\varrho$(e/Å^5)	ϱ(e/Å^3)	$\nabla^2\varrho$(e/Å^5)	ϱ(e/Å^3)	$\nabla^2\varrho$(e/Å^5)	ϱ(e/Å^3)	$\nabla^2\varrho$(e/Å^5)
0.020	0.30	0.034	0.32	0.043	0.43	0.040	0.40	0.043	0.44	0.026	0.24	0.018	0.22
0.028	0.35	0.032	0.40	0.035	0.38	0.032	0.39	0.028	0.35	0.012	0.17	0.032	0.29
0.029	0.27	0.023	0.26	0.037	0.41	0.043	0.47	0.036	0.44	0.028	0.29	0.014	0.24
0.020	0.23	0.019	0.30	0.026	0.28	0.037	0.36	0.046	0.49	0.025	0.35	0.031	0.32
0.021	0.32	0.027	0.32	0.018	0.26	0.025	0.31	0.034	0.40	0.029	0.31	0.057	0.49
0.027	0.32	0.012	0.21	0.030	0.32	0.016	0.22	0.035	0.35	0.046	0.42	0.029	0.34
0.012	0.18	0.028	0.25	0.040	0.41	0.037	0.38	0.011	0.15	0.019	0.27	0.033	0.34
0.036	0.33	0.027	0.35	0.032	0.39	0.034	0.40	0.030	0.31	0.020	0.23	0.024	0.36
0.022	0.25	0.014	0.20							0.031	0.28	0.030	0.33
												0.031	0.28
0.215		**0.216**		**0.263**		**0.263**		**0.263**		**0.236**		**0.299**	

Table S6.5 ϱ_π and $\Sigma\varrho_\pi$ (in bold), and $\nabla^2\varrho_\pi$ at bond critical points between stacked base pairs isolated from 1EKD (PBE0/6-31+G(d,p))

$G_1G_2C_3G_4U_5G_6C_7C_8$
$C_8C_7G_6U_5G_4C_3G_2G_1$

G1_C8/G2_C9		G2_C9/C3_G8		C3_G8/G4_U5		G4_U5/U5_G4		U5_G4/G6_C3		G6_C3/C7_G2		C7_G2/C8_G1	
ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)
0.037	0.38	0.03	0.31	0.016	0.19	0.031	0.29	0.020	0.26	0.038	0.33	0.028410	0.322252
0.040	0.45	0.04	0.32	0.036	0.37	0.036	0.46	0.046	0.46	0.029	0.32	0.028055	0.308590
0.047	0.47	0.03	0.38	0.038	0.41	0.030	0.32	0.032	0.41	0.044	0.47	0.032749	0.332427
0.021	0.28	0.04	0.44	0.032	0.42	0.038	0.50	0.037	0.41	0.044	0.43	0.046740	0.458115
0.050	0.48	0.04	0.48	0.022	0.28	0.022	0.30	0.016	0.19	0.031	0.37	0.042488	0.466086
0.030	0.32	0.02	0.29	0.046	0.45	0.017	0.28	0.035	0.37	0.030	0.36	0.045407	0.449093
0.048	0.46	0.04	0.44	0.018	0.24	0.036	0.46	0.022	0.28	0.046	0.45	0.029134	0.401971
0.029	0.41	0.05	0.45			0.038	0.50			0.041	0.44	0.038908	0.398819
0.029	0.32	0.03	0.36			0.033	0.30			0.017	0.29	0.050612	0.465386
0.031	0.31	0.03	0.31			0.029	0.31			0.030	0.27	0.018064	0.241080
		0.03	0.27							0.029	0.30		
0.364		**0.38**		**0.208**		**0.310**		**0.209**		**0.379**		**0.360567**	

Table S6.6 ϱ_π and $\Sigma\varrho_\pi$ (in bold), and $\nabla^2\varrho_\pi$ at bond critical points between stacked base pairs isolated from 1GID (PBE0/6-31+G(d,p))

G_{200} G_{201} U_{202} C_{203} C_{204}
C_{121} U_{120} G_{119} G_{118} G_{117}

C204_G117/C203_G118		C203_G118/U202_G119		U202_G119/G201_U120		G201_U120/G200_C121	
ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)
0.024	0.26	0.011	0.14	0.040	0.40	0.037	0.34
0.048	0.47	0.050	0.53	0.054	0.58	0.041	0.50
0.038	0.45	0.010	0.15	0.031	0.47	0.060	0.59
0.087	1.11	0.029	0.32	0.060	0.76	0.034	0.41
0.026	0.28	0.048	0.54	0.081	1.05	0.056	0.54
0.018	0.25	0.051	0.53	0.073	0.84	0.056	0.50
0.018	0.22	0.066	0.88	0.065	0.82	0.035	0.45
0.050	0.68	0.044	0.48	0.037	0.53	0.038	0.42
0.067	0.74	0.053	0.62	0.039	0.41		
		0.019	0.31	0.041	0.40		
0.376		**0.381**		**0.521**		**0.359**	

Table S6.7 $\Sigma\varrho_\pi$ (in bold), and $\Sigma\nabla^2\varrho_\pi$ at bond critical points between stacked base pairs isolated from ten 1GUC RNA duplexes, (PBE0/6-31+G(d,p)).

$G_1\ A_2\ G_3\ G_4\ U_5\ C_6\ U_7\ C_8$
$C_{16}U_{15}C_{14}U_{13}G_{12}G_{11}A_{10}G_9$

	G1_C16/A2_U15		A2_U15/G3_C14		G3_C14/G4_U13		G4_U13/U5_G12		U5_G12/C6_G11		C6_G11/U7_A10		U7_A10/C8_G9	
Model number	$\Sigma\varrho$ (e/Å^3)	$\Sigma\nabla^2\varrho$ (e/Å^5)	$\Sigma\varrho$ (e/Å^3)	$\Sigma\nabla^2\varrho$ (e/Å^5)	$\Sigma\varrho$ (e/Å^3)	$\Sigma\nabla^2\varrho$ (e/Å^5)	$\Sigma\varrho$ (e/Å^3)	$\Sigma\nabla^2\varrho$ (e/Å^5)	$\Sigma\varrho$ (e/Å^3)	$\Sigma\nabla^2\varrho$ (e/Å^5)	$\Sigma\varrho$ (e/Å^3)	$\Sigma\nabla^2\varrho$ (e/Å^5)	$\Sigma\varrho$ (e/Å^3)	$\Sigma\nabla^2\varrho$ (e/Å^5)
M1	0.428	4.76	0.463	4.89	0.320	3.38	0.374	4.41	0.324	3.42	0.456	4.76	0.446	4.72
M2	0.430	4.50	0.437	4.59	0.318	3.36	0.376	4.42	0.338	3.63	0.427	4.50	0.430	4.51
M3	0.408	4.84	0.423	4.45	0.391	4.20	0.310	3.60	0.312	3.32	0.459	4.84	0.454	4.79
M5	0.414	4.91	0.413	4.95	0.426	3.43	0.335	4.40	0.350	3.76	0.465	4.51	0.446	4.47
M7	0.448	4.26	0.469	4.38	0.325	3.78	0.3745	4.56	0.351	4.59	0.429	4.26	0.426	4.34
M9	0.443	5.17	0.431	5.21	0.322	3.25	0.375	5.04	0.317	3.72	0.438	5.17	0.431	5.00
M8	0.433	5.14	0.466	4.57	0.352	3.42	0.384	4.48	0.426	3.68	0.414	5.14	0.410	5.01
M4	0.429	4.91	0.434	4.23	0.322	4.60	0.3784	3.93	0.339	3.78	0.471	4.91	0.453	4.72
M6	0.507	4.62	0.486	4.55	0.306	3.39	0.4214	4.42	0.341	3.34	0.472	4.62	0.452	4.53
M30	0.506	4.73	0.497	5.43	0.312	3.32	0.3314	3.88	0.328	3.46	0.449	4.73	0.433	4.58
Average	**0.445**		**0.452**		**0.339**		**0.366**		**0.343**		**0.448**		**0.438**	

Table S6.8 $\Sigma\rho_\pi$ (in bold), and $\Sigma\nabla^2\rho_\pi$ at bond critical points between stacked base pairs isolated from ten 1QES RNA duplexes, (PBE0/6-31+G(d,p)).

$G_1\ G_2\ A_3\ G_4\ U_5\ A_6\ C_7\ C_8$
$C_{16}C_{15}U_{14}G_{13}U_{12}U_{11}G_{10}G_9$

Model umber	G1_C16/G2_C15		C2_C15/A3_U14		A3_U14/U4_G13		G4_U13/U5_U12		U5_G12/A6_U11		A6_U11/C7_G10		C7_G10/C8_G9	
	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)	$\Sigma\rho$ (e/Å^3)	$\Sigma\nabla^2\rho$ (e/Å^5)
M1	0.602	6.58	0.553	5.88	0.520	5.78	0.443	5.35	0.486	5.35	0.523	5.53	0.586	6.41
M2	0.583	6.33	0.533	5.66	0.500	5.57	0.432	5.16	0.501	5.58	0.538	5.77	0.587	6.40
M3	0.586	6.40	0.530	5.62	0.489	5.36	0.406	4.98	0.492	5.40	0.536	5.70	0.593	6.44
M5	0.585	6.36	0.533	5.67	0.499	5.53	0.446	5.39	0.495	5.48	0.532	5.66	0.579	6.28
M6	0.587	6.41	0.523	5.54	0.476	5.26	0.436	5.24	0.481	5.29	0.521	5.52	0.584	6.38
M8	0.593	6.46	0.551	5.85	0.510	5.63	0.409	5.02	0.492	5.42	0.526	5.57	0.531	5.67
M10	0.578	6.28	0.527	5.60	0.485	5.33	0.445	5.35	0.516	5.73	0.540	5.70	0.596	6.47
M15	0.601	6.54	0.545	5.77	0.513	5.67	0.433	5.15	0.483	5.28	0.519	5.48	0.583	6.37
M25	0.594	6.76	0.537	5.67	0.510	5.62	0.433	5.14	0.481	5.25	0.522	5.51	0.587	6.42
M30	0.554	5.95	0.536	5.71	0.498	5.52	0.442	5.36	0.497	5.54	0.500	5.27	0.584	6.30
Average	**0.586**		**0.537**		**0.500**		**0.432**		**0.492**		**0.526**		**0.581**	

Table S6.9 ϱ_π and $\Sigma\varrho_\pi$ (in bold), and $\nabla^2\varrho_\pi$ at bond critical points between stacked base pairs isolated from 332D, (PBE0/6-31+G(d,p)).

$A_{15}U_{14}G_{13}U_{12}$
U_4 G_5 U_6 A_7

A_{15}_U_4/U_{14}_G_5		U_{14}_G_5/G_{13}_U_6		G_{13}_U_6/U_{12}_A_7	
ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)	ϱ (e/Å^3)	$\nabla^2\varrho$ (e/Å^5)
0.035	0.42	0.024	0.27	0.036	0.36
0.033	0.36	0.032	0.30	0.037	0.41
0.029	0.37	0.037	0.49	0.037	0.36
0.031	0.34	0.044	0.53	0.033	0.32
0.025	0.29	0.018	0.31	0.022	0.28
0.019	0.25	0.037	0.47	0.017	0.22
0.032	0.35	0.038	0.48	0.016	0.20
		0.014	0.24		
		0.025	0.28		
		0.027	0.25		
0.203		**0.296**		**0.198**	

Figure S6.1 The averaged $\Sigma\varrho_\pi$ in e/Å^3 from stacked GC/GC base pairs isolated from 1EKA and 1FFK - motif I (grey) and 1EKD, 1QES and 332D - motif II (black).

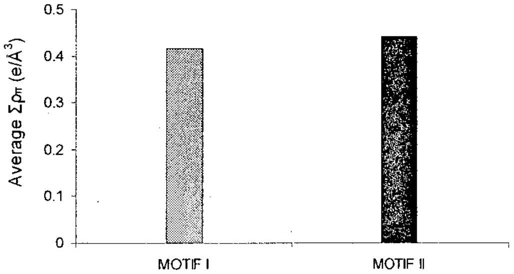


Figure S6.2 The averaged $\Sigma\varrho_\pi$ in e/Å^3 from stacked AU/GC base pairs isolated from 1QET, and 315D - motif I (grey) and 1QES - motif II (black).

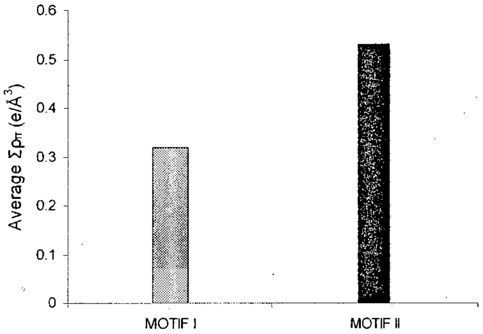


Figure S6.3 The averaged $\Sigma\varrho_\pi$ in e/Å^3 from stacked AU/GU base pairs isolated from 1QET and 315D- motif I (grey) and 1QET and 332D - motif II (black).

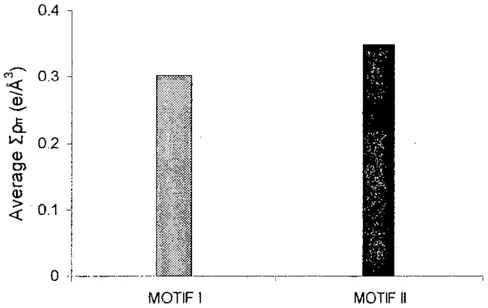


Figure S6.5 The $\Sigma\varrho_{HB}$ in e/Å^3 of GC H-bonds isolated from 315D,1QET and 1FFK -motif I (grey) and 1EKD, 1QES and 332D - motif II (black).

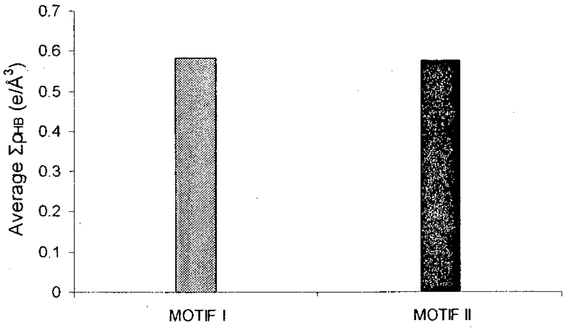


Figure S6.6 The $\Sigma\varrho_{HB}$ in e/Å^3 of AU H-bonds isolated from 1QET and 315D - motif I (grey) and 1QES, and 332D - motif II (black).

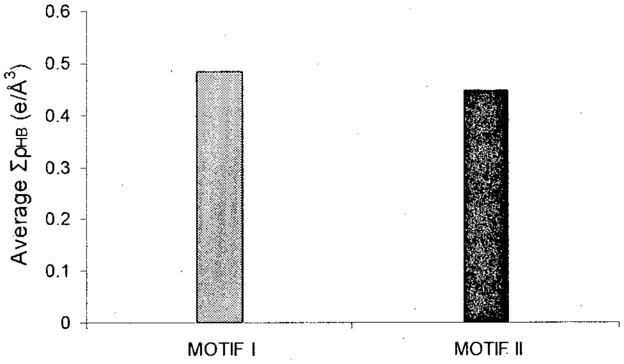


Figure S6.8
Molecular graphs for G•U base pairs isolated from the ten randomly selected RNA duplexes reported for 1QET.

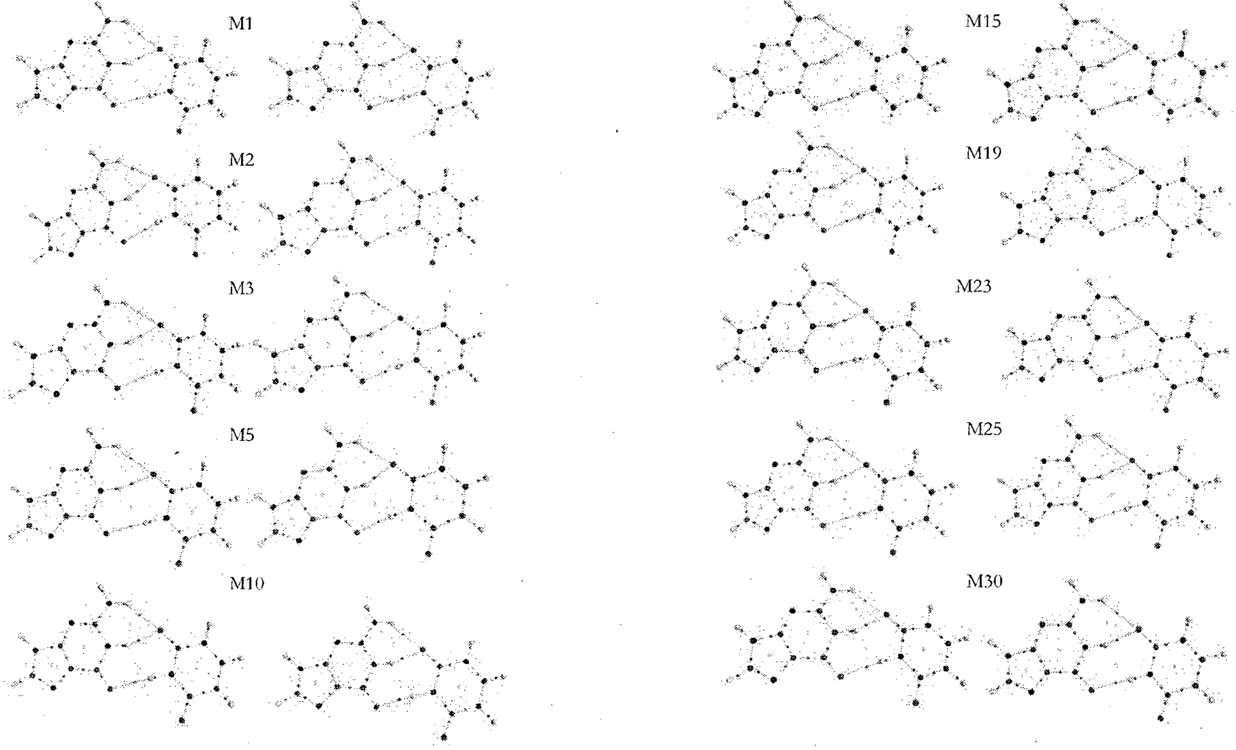


Figure S6.9
Molecular graphs for GU base pairs isolated from the ten randomly selected RNA duplexes reported for 1GUC

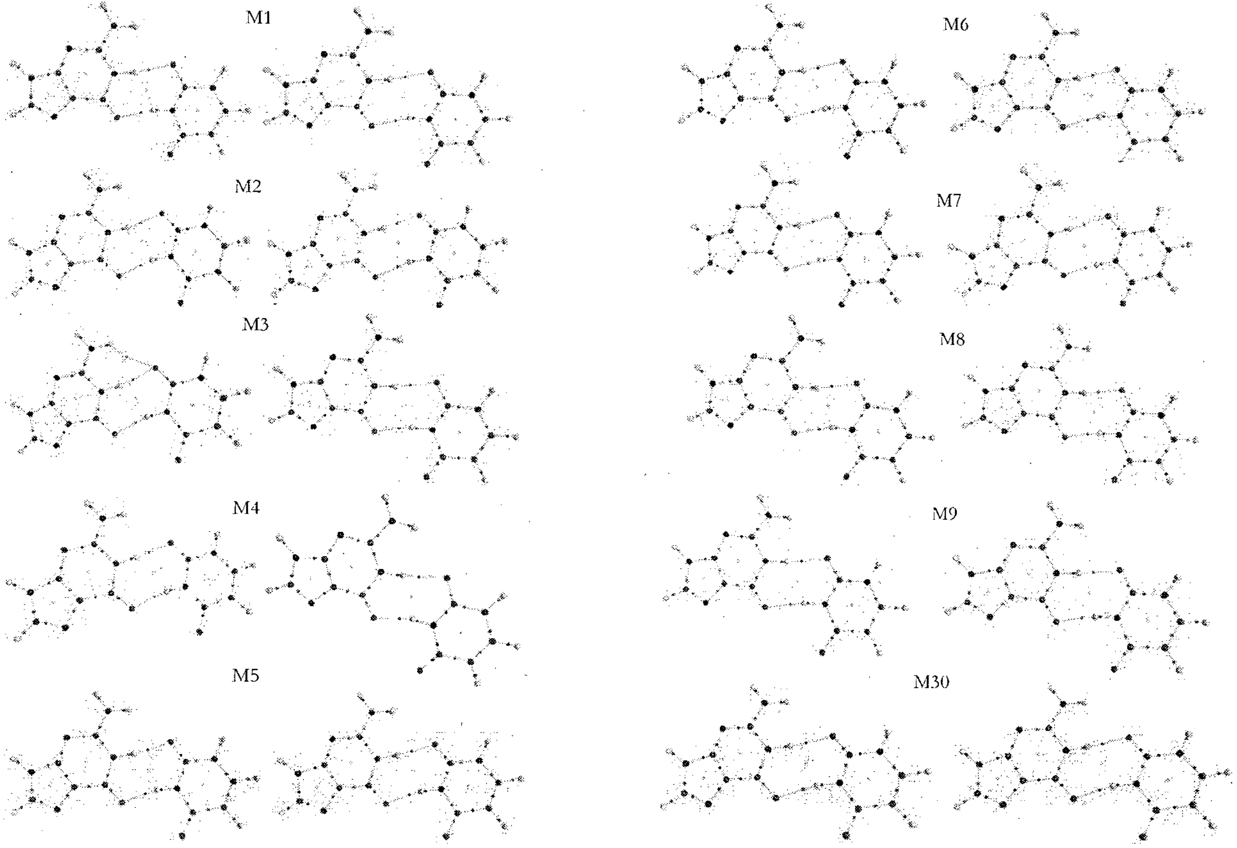
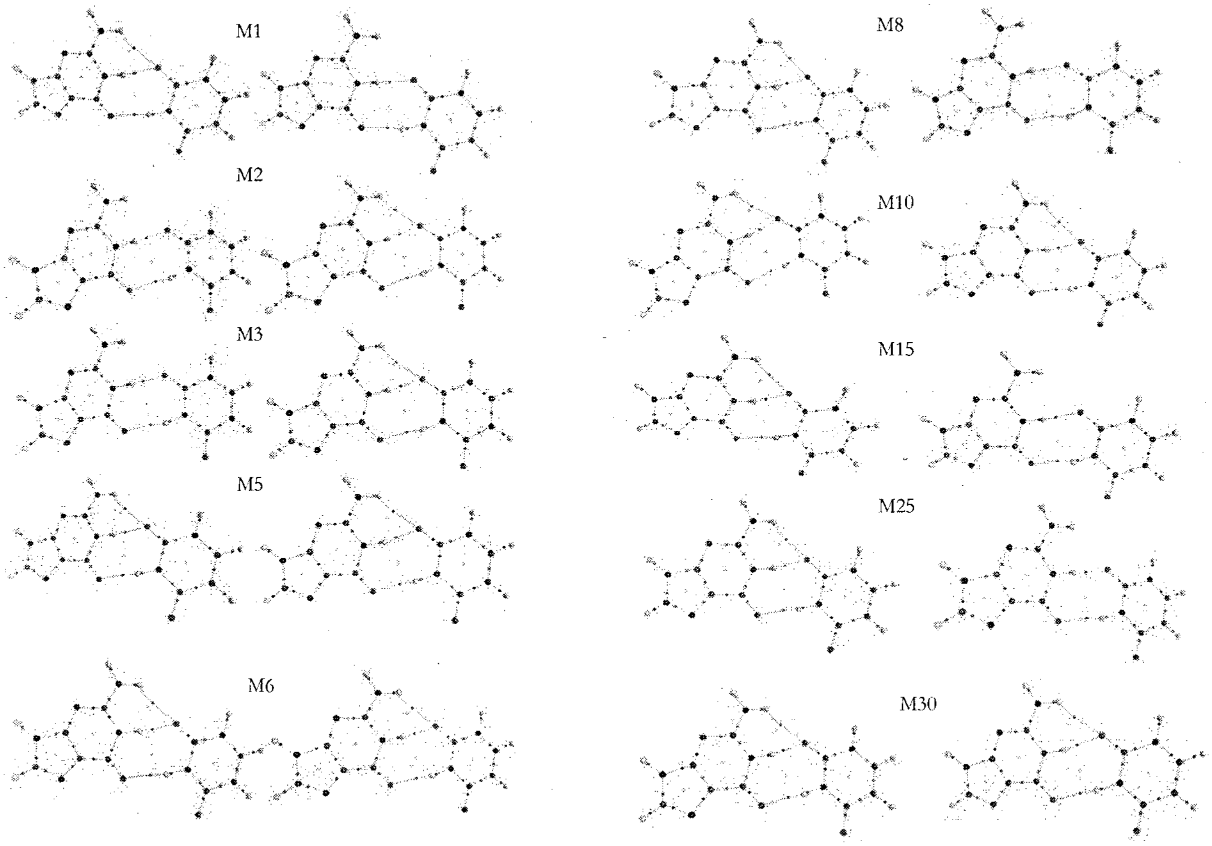


Figure S6.10
Molecular graphs for GU base pairs isolated from the ten randomly selected RNA duplexes reported for 1QES



Appendix F

Supporting information to Chapter 7

A stabilization profile for the guanine riboswitch

Table S7.1

	C=O			C6-N1			N1-C2			C2-N3			N3-C4			C4-C5			C5-C6			C2-O		
	OPT	EXP	Δ	OPT	EXP	Δ	OPT	EXP	Δ	OPT	EXP	Δ	OPT	EXP	Δ	OPT	EXP	Δ	OPT	EXP	Δ	OPT	EXP	Δ
G	1.217	1.239	0.022	1.427	1.391	0.036	1.365	1.373	0.008	1.308	1.330	0.022	1.353	1.361	0.008	1.392	1.376	0.016	1.436	1.421	0.015			
HX	1.216	1.223	0.007	1.423	1.408	0.015	1.359	1.405	0.046	1.299	1.357	0.058	1.357	1.351	0.006	1.392	1.528	0.136	1.442	1.403	0.039			
HX+•	1.206	1.223	0.017	1.431	1.408	0.023	1.333	1.405	0.072	1.335	1.357	0.022	1.314	1.351	0.037	1.443	1.528	0.085	1.456	1.403	0.053			
HX−•	1.240	1.223	0.017	1.390	1.408	0.018	1.430	1.405	0.025	1.375	1.357	0.018	1.317	1.351	0.034	1.420	1.528	0.108	1.438	1.403	0.035			
HX−	1.235	1.223	0.012	1.417	1.408	0.009	1.370	1.405	0.035	1.292	1.357	0.065	1.375	1.351	0.024	1.423	1.528	0.105	1.428	1.403	0.025			
XK	1.213	1.227	0.014	1.417	1.413	0.004	1.380	1.422	0.042	1.390	1.390	0.000	1.366	1.370	0.004	1.379	1.386	0.007	1.449	1.444	0.005	1.217	1.242	0.025
XE1	1.215	1.227	0.012	1.433	1.413	0.020	1.365	1.422	0.057	1.295	1.390	0.095	1.354	1.370	0.016	1.4356	1.386	0.008	1.340	1.444	0.098		1.242	
XE3	1.214	1.227	0.013	1.428	1.413	0.015	1.354	1.422	0.068	1.299	1.390	0.091	1.358	1.370	0.012	1.391	1.386	0.005	1.440	1.444	0.004	1.336	1.242	0.094