DISSERTATION

STRUCTURE ENERGY RELATIONSHIP OF BIOLOGICAL HALOGEN BONDS

Submitted By

Megan Carter

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2012

Doctoral Committee:

Advisor: P. Shing Ho

Olve Peersen
Eric Ross
Alan Kennan

ABSTRACT

STRUCTURE ENERGY RELATIONSHIP OF BIOLOGICAL HALOGEN BONDS

The primary goal of the studies in this thesis is to derive a set of mathematical models to describe the anisotropic atomic nature of covalent bound halogens and by extension their molecular interactions. We use a DNA Holliday junctions as a experimental model system to assay the structure energy relationship of halogen bonds (X-bonds) in a complex biological environment. The first chapter of this dissertation is reserved for a review on DNA structure and the Holliday Junction in context of other DNA conformations. The conformational isomerization of engineered Holliday junctions will be established as a means to assay the energies of bromine X-bonds both in crystal and in solution. The experimental data are then used in the development of anisotropic force fields for use in the mathematical modeling of bromine halogen bonds, serving as a foundation to model all biological halogen interactions. The DNA Holliday junction experimental system is expanded to compare and contrast halogens from fluorine to iodine. This comprehensive study is used to determine the effects of polarization on the structure-energy relationship of biological X-bonds in solid state and solution phase. The culmination of the work in this thesis, in addition to previously published studies, provides a growing set of principles to guide knowledge-based application of halogens in drug design. These principles are applied to the selection of X-bond acceptors in a protein binding pocket, optimal placement of the halogen on the lead compound, and which halogen is best suited for a particular interaction.

ACKNOWLEDGMENTS

I thank my advisor and mentor Dr. Shing Ho for his patience and enthusiasm in lab and his appreciation for life outside of the lab. I am just beginning to appreciate the nuances of training and how much restraint is required to allow a student to progress while guiding them through unknown scientific processes. I greatly appreciate his willingness to answer questions for hours while still pushing me to think of answers for myself. I would like to thank the Ho lab members, especially those that were patient while I stumbled through the art of mentoring and the changing dynamic of a growing lab.

I owe a large thanks to my friends and classmates for helping me find balance in life. Especially to Jarred, Dana, and Broox for helping me find laughter during struggle. Though she will never read this, I thank Beven for giving me a reason to take walks. Most of all, I want to thank my family and especially my parents for a lifetime of support that became only more critical during graduate school. You taught me to dream but more importantly to work hard and take pride in my accomplishments and that has made all the difference.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1. Literature Review

Halogen bonds, or X-bonds, are electrostatically-driven noncovalent interactions similar to the well-known hydrogen bond, or H-bond. Both H- and X-bonds are non-covalent electrostatic attractive forces. H-bonds involve an attraction between the electropositive hydrogen atom and an electronegative acceptor whereas X-bonds are an attractive interaction between a positive region of electrostatic potential on the crown of the halogen, called the σ-hole, and an electronegative acceptor (Fig. 1.1). Both interactions share a common class of acceptors and typically result in interatomic distances closer than the sum of the interacting atoms' van der Waals radii (ΣR_{vdw}). X-bonds exhibit a strong directionality distinct from H-bonding resulting from the limited area of positive electrostatic potential of the polarized halogen available for interaction (Shields, Murray et al. 2010). The IUPAC task force only recently defined these interactions (Metrangolo and Resnati 2012) though their influence was noted as early as the 19th century (Guthries 1863). These interactions were originally described as charge-transfer bonding by Odd Hassel in work awarded the 1970 Nobel Prize (Hassel 1972). Propelled by the use of X-bonding in molecular engineering and drug design, investigation of halogen interactions, including X-bonds, has grown drastically in the past decade.

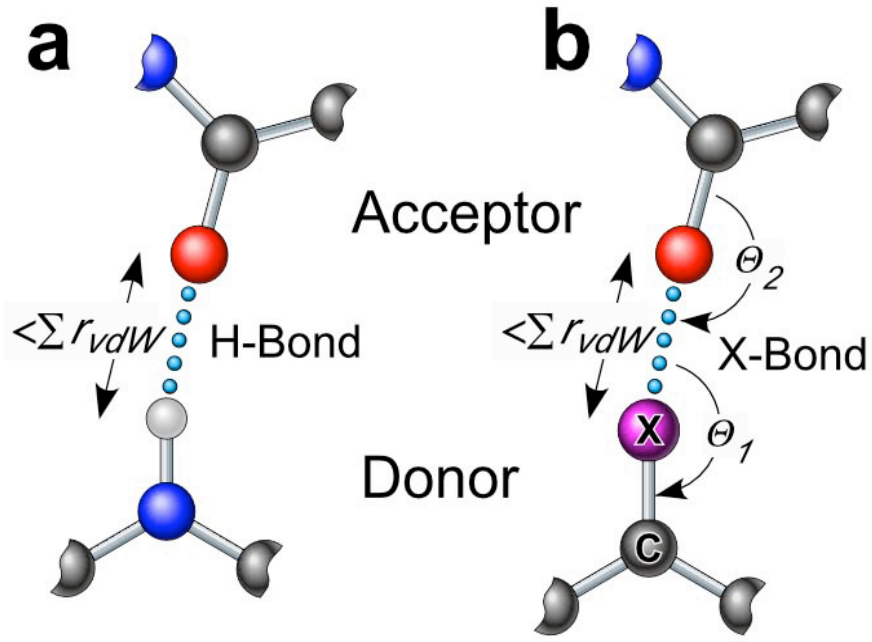


Fig. 1.1. H- and X-bonds. H-bonds (a) and X-bonds (b) are short interactions between an acceptor (A) and donor (D), where the A-D distance is shorter than the sum of their Rvdw (Ouvrard, Le Questel et al. 2003; Auffinger, Hays et al. 2004; Metrangolo 2005). The terms "donor" and "acceptor" refer to the now accepted definitions, in which the donor is the electropositive atom while the acceptor is the electronegative atom in the two interactions (Auffinger, Hays et al. 2004; Metrangolo 2008).

The work presented in this dissertation is focused on improving molecular modeling tools used to predict and optimize halogen interactions. This introduction will describe the current theoretical understanding of halogen interactions, survey the evolving use of X-bonds in molecular engineering and drug design, and present the scientific approach of the following work.

The halogen atomic group includes fluorine (F), chlorine (Cl), bromine (Br), iodine (I), and astatine (At). These elements make up Group 17 of the periodic table and are characterized by high electronegativity and reactivity. Astatine, the last element of the group, is extremely unstable and radioactive and is ignored for the purposes of this study. It initially seems counterintuitive that covalent bonded halogens form stable interactions with other electron rich atoms. Hassel described these events as charge-transfer interactions in the 1970's. However, in recent years dispersion and electrostatic forces have come to the forefront as the primary forces involved in X-bonds (Metrangolo and Resnati 2012). Halogens contradict traditional expectations of interaction because of their tendency to become polarized when covalently bound. The polarization of covalent bonded halogens and resulting anisotropic electrostatic distribution is elegantly described in the σ-hole theory developed by Politzer et. al (Clark, Hennemann et al. 2007; Murray, Lane et al. 2007). The σ-hole model describes the depopulation of the p_z orbital electrons into the sigma bond, resulting in a reduced electron density along the sigma bond. This reduction in electron density exposes the positive nuclear charge, creating an electropositive crown along the σ-bond axis known as the σ-hole (Fig. 1.2). The σ-hole acts as the electropositive X-bond donor, forming an electrostatic attraction to electronegative atoms, or X-bond acceptors.

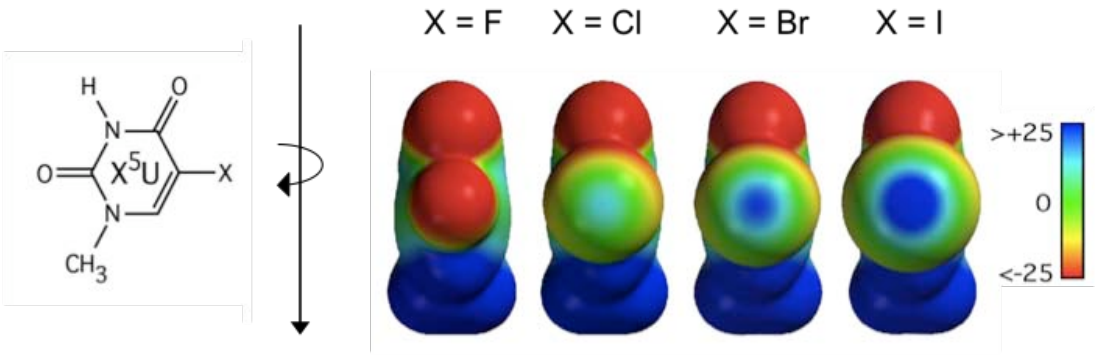


Fig. 1.2. Sigma-Hole Formation. Formation of the σ-hole as demonstrated by ab initio electrostatic potential surface calculations of halogenated uracil (Auffinger, 2004). Positive electrostatic potential is represented in blue and negative electrostatic potential in red presented in the -25 to +25 kcal/mol range.

This interaction is becoming increasingly accepted as the dominating force behind X-bonding, as supported by a strong linear correlation between the QM calculated interaction energies and the electrostatic potential of the σ- hole (Riley, Murray et al. 2009; Shields, Murray et al. 2010). We demonstrate in the following work (Chapter 4) that the depletion of electron density also changes the effective van der Waals radius, or shape of the halogen from the anticipated spherical shape toward a toriodal shape, with an indentation corresponding with the σ-hole. This, in part, allows atoms attracted to the electropositive σ-hole to approach closer than the standard van der Waals radius. The remaining outer shell electrons in the p_x and p_y orbitals maintain a ring of negative charge perpendicular to the σ-bond. The anisotropic and dual electrostatic nature of covalent bonded halogens allows them to act as both electropositive X-bond donors and electronegative H-bond acceptors (Zhou, Qiu et al. 2011; Brammer, Bruton et al. 2001; Voth 2009) and imparts a strong directionality to all halogen electrostatic interactions.

Although X- and H-bonding are similar in many respects, the most significant difference is the directionality inherent to halogen interactions (Voth 2009; Shields, Murray et al. 2010; Murray, Riley et al. 2010; Brammer, Bruton et al. 2001). X-bonds have a strong directional tendency toward a θ_1 angle, the angle of the acceptor approach toward the halogen with respect to the sigma bond, of 180° (Fig. 1.1). H-bonds, however, are much more likely to have non-linear θ_1 angles. These differences can be explained by analysis of σ-hole formation and electrostatic surface potentials of covalently bonded halogen and hydrogen atoms as performed in Shields et al. (Shields, Murray et al. 2010). As mentioned before, polarized halogens have a ring of negative charge maintained by remaining p_X and p_Y valence electrons surrounding the positive σ-hole. A linear approach

of the acceptor, therefore, optimizes contact of the acceptor with the σ-hole of the halogen. A covalently bonded hydrogen atom, on the other hand, has only one electron, which is participating in the covalent bond. The positive electrostatic potential on a covalently bonded hydrogen atom, therefore, approaches a hemispherical shape with increasing electron-withdrawing ability of the substituent. The narrow focus area of the halogen σ-hole imparts directional preference to X-bonds lacking in H-bond formation.

The extent of polarization, and resultant σ-hole formation, increases with the size of the halogen from F<Cl<Br<I and with increasing electron withdrawing ability of the substituent (Fig. 1.2) (Shields, Murray et al. 2010; Riley, Murray et al. 2011). The electronegativity, or attractive force imparted by the nucleus on electrons, decreases from fluorine to iodine with each additional shell of electrons. The electrons in iodine, therefore, are more loosely held and are subject to increased polarization (Clark, Hennemann et al. 2007). Increasing the electron withdrawing ability of the substituent also has a scalable effect on σ-hole formation and X-bond strength. For example, the strength of X-bonds formed between the carbonyl oxygen of acetone and aromatically-bond halogens, Cl, Br, and I were directly correlated to the electron withdrawing ability of the aromatic system (Riley, Murray et al. 2011). In this study the electron withdrawing ability of the halogen substituent, a benzene ring, was increased by systematic substitution of aromatic hydrogens with fluorines. The X-bond was engineered to be up to 100% stronger than that formed by the unmodified halogenated benzene, depending on the number and placement of substitutions. The combination of innate halogen polarizability and sensitivity to substituent electronegativity makes X-bonds an extremely

tunable molecular interaction. Tunability and directionality are two properties that make X-bonds attractive for use in designing novel materials and pharmaceuticals.

The increase in theoretical investigation into the basis of X-bonds during the past decade has gone hand in hand with the increased use of X-bonds in material science, crystal engineering, and drug design (Ritter 2009). Extensive study and review by Resanti and Metrangolo (Metrangolo 2005) and recently echoed by Fourmiqué (Fourmigué 2009) have highlighted the use of aromatic halogen molecules to engineer supermolecular architectures, including elaborate three-dimensional crystal networks. Halogen bonds have been investigated as a means to control the solid-state structures of supermolecular organic conductors and have been shown to form liquid crystals from nonmesomorphic components. The influence of X-bonds is not limited to the material science but is also important to the biochemical field (Auffinger, Hays et al. 2004). Organic halogens are used in pharmaceuticals, herbicides, fungicides, insecticides, flame-retardants, and intermediates in organic synthesis, which can have toxic and carcinogenic effects on humans and are difficult to degrade in nature (van Pee and Unversucht 2003). Over 3500 documented secondary metabolites are halogenated, including the antibiotics vancomycin and chloromaphenicol. Thyroid hormones are naturally iodinated molecules in which multiple iodine X-bonds to carbonyl oxygen play a crucial role in recognition by their associated proteins (Cody and Murray-Rust 1984). Chloride and bromide represent the fourth and fifth most abundant inorganic anions in human plasma and tissues (Wu 2000). The physiological role these abundant anionic halogens play in mammals is still largely unknown. However, bromination of protein tyrosine residues has been established as a form of posttranslational modification common to sites of eosinophil recruitment and

activation in humans in an eosinophil peroxidase dependent manner (Wu 2000). Although this system has been demonstrated to have antibacterial, antifungicidal, antiparasitic, and antiviricidal activities in vitro, it also elicits many pathophysiological features of asthma in cell culture. Similarly, chlorinated proteins, specifically 3-chlorotyrosine, are produced in lungs of premature infants and are found at higher levels during infection (Buss 2003). Halogens clearly have complicated roles, both beneficial and detrimental, in natural biological processes.

Halogens are also highly prevalent in synthetic drug molecules. In fact, "25% of the top 200 brand name drugs by retail dollar in 2009 possess halogen atoms in their molecular structures" (Xu, Liu et al. 2011). Halogen substituents are historically found in small molecule compounds as a means to increase bioavailability, delay the catabolic process, and are common side products of chemical synthesis. In fact, it has been estimated that up to 50% of small molecules used in high-throughput drug screens are halogenated (Voth 2007). Initial examples of drug-protein interactions result from the prevalence of halogenated molecules in high-throughput screens. Protein kinase inhibitors represent a relatively large class of drugs stabilized by halogen bonding interactions. Twelve such complexes were reviewed by Voth et. al. to characterize the role of X-bonding in conferring specificity and affinity for halogenated inhibitors (Voth 2007). Other examples include the IDD594 inhibitor to aldose reductase in which X-bonding increased inhibition and specificity over aldehyde reductase (Muzet, Guillot et al. 2003) and the R221239 inhibitor to HIV reverse transcriptase in which X-bonding to carbonyl oxygen is implicated as a strategy to avoid the effects of common drug resistant mutations (Himmel, Das et al. 2005). The above examples are a product of halogen

presence in historical drug development strategies, however, halogens are becoming increasingly considered as attractive tools for the use in bottom-up drug design.

In recent years there has been a shift toward the rational, or bottom-up, design of drugs as a means to reduce cost and inefficiency inherent to the high-throughput screening approach. This approach focuses on using the knowledge of protein target structural features to design chemical compounds stabilized by specific interactions at or near the target region of the protein (Parisini, Metrangolo et al. 2011). Compound design and lead optimization are generally guided by computational analysis including molecular mechanics and quantum mechanics approaches, optimizing structural complementarity and energetic stabilization. Standard drug docking simulations, however, are useless when modeling halogen interactions because the underlying force fields fail to model σ-hole formation and anisotropic electrostatic properties of covalently bound halogens. Therefore, X-bonds are an attractive, yet underutilized, interaction for use in rational drug design. A few examples of X-bond engineering in drug optimization do exist, however, prompted by the increased recognition of X-bonds and recent advances in σ-hole theory. An inhibitor to blood clotting factor, Xa, was modified to substitute an amidine group forming a H-bond to the carboxyl group of ASP189 with a halogen atom (Matter, Nazare et al. 2009). The iodo-halogen substitution was shown by X-ray crystallography to form a short X-bond and affinities of the inhibitors were shown to increase from Cl < Br < I, though affinities were less than the original lead compound. The lead optimization of a phosphodiesterase type 5 (PDE5) inhibitor was recently used as a case study toward the incorporation of X-bond interaction to increase binding affinity (Xu, Liu et al. 2011). Hybrid quantum mechanics/molecular mechanics

(QM/MM) calculations were used to estimate the potential interaction strength of the putative X-bond between the halogenated (F, Cl, Br, or I) ligand and PDE5. These predicted energies correlated well with the experimentally determined bioactivities of the inhibitors, both with an increasing from F < Cl < Br < I. In this case the X-bond, confirmed with x-ray crystallography, increased binding affinity and decreased the IC_{50} over the original lead compound. In the case of the iodoinated inhibitor, the IC_{50} was reduced to levels comparable to the drug sildenafil already on the market, while maintaining a less complex overall structure. The hybrid QM/MM approach is one way to get around the lack of proper molecular mechanics systems and in this case gave good correlation to experimental results. There still remain multiple advantages to correcting the underlying potential energy and extended force field equations that direct molecular mechanic simulations of halogen interactions. QM calculations remain very time intensive for biomolecules and have cumulative error in atomic coordinates with increasing system size. In addition, once correct force fields are incorporated into molecular simulations conformational and solvent entropy effects on halogen interactions, including hydrophobicity, can be investigated using established free-energy methods.

These examples demonstrate that X-bonds are a practical and effective tool in drug design but require understanding of the underlying properties, including polarization effects and directionality, for effective implementation. The most effective means to disseminate the advances in X-bond theory is to incorporate this knowledge into existing molecular mechanic and drug docking force fields. The major obstacle to this is that the underlying mathematical equations driving computer modeling fail to represent the

anisotropic electrostatic potential and shape, of polarized covalently bound halogens. The classical force fields, namely those used in AMBER and CHARMM packages, continue to model halogens, as with all atoms, with a spherical shape and uniform negative electrostatic potential. This representation prevents modeling of X-bonds all together. Therefore the major contribution I make in the following scientific investigations is to modify the underlying mathematical representation of covalently bound halogens to demonstrate experimentally supported anisotropic atomic properties.

To do this I use DNA Holliday junctions as a well-established model system to assay the structure energy relationship of X-bonds in a complex biological environment. The first chapter of this dissertation is reserved for a review on DNA structure and the Holliday Junction in this context. Next, the conformational isomerization of engineered Holliday junctions will be established as a means to assay the energies of biological halogen bonds. This system is expanded to assay the effects of polarization on the structure-energy relationship of biological halogen bonds in solid state and solution phase. These experimental data are then used in the development of anisotropic force fields for use in the mathematical modeling of biological halogen interactions. In the conclusions we will take a look at the insights gained from the developed mathematical representation and implications for use of X-bonds in drug design.

2. References

Auffinger, P., F. A. Hays, et al. (2004). "Halogen bonds in biological molecules." Proc. Natl. Acad. Sci., USA 101(48): 16789-16794.

Brammer, L., E. A. Bruton, et al. (2001). "Understanding the behavior of halogens as hydrogen bond acceptors." Crystal Growth & Design 1(4): 277-290.

Buss, I., H., Senthilmohan, R., Darlow, B.A., Mogridge, N., Kettle, A. J., and Winterbourn, C.C. (2003). "3-Chlorotyrosine as a Marker of Protein Damage by Myeloperoxidase in Tracheal Aspirates From Preterm Infants: Association With Adverse Respiratory Outcome." Pediatric Res. 53: 455-462.

Clark, T., M. Hennemann, et al. (2007). "Halogen bonding: the sigma-hole. Proceedings of "Modeling interactions in biomolecules II", Prague, September 5th-9th, 2005." J. Mol. Model. 13(2): 291-296.

Cody, V. and P. Murray-Rust (1984). "Iodine X(O, N, S) intermolecular contacts: models
of thyroid hormone–protein binding interactions using information from the
Cambridge crystallographic data files. ." J Mol Struct 112: 189-199.

Fourmigué, M. (2009). "Halogen bonding: Recent advances." Curr Opin Solid State Mat Sci.

Guthries, F. (1863). "On the iodide of iodammonium." J. Am. Chem. Soc. 16: 239-244.

Hassel, O. (1972). Structural aspects of interatomic charge-transfer bonding. Nobel lectures, Chemistry 1963-1970. Amsterdam, Elsevier publishing company.

Himmel, D. M., K. Das, et al. (2005). "Crystal structures for HIV-1 reverse transcriptase in complexes with three pyridinone derivatives: a new class of non-nucleoside inhibitors effective against a broad range of drug-resistant strains." J Med Chem 48(24): 7582-7591.

Matter, H., M. Nazare, et al. (2009). "Evidence for C-Cl/C-Br...pi interactions as an important contribution to protein-ligand binding affinity." Angew Chem Int Ed Engl 48(16): 2911-2916.

Metrangolo, P., Neukirch, H., Pilati, T., and Resnati, G. (2005). "Halogen bonding based recognition processes: A world parallel to hydrogen bonding." Acct. Chem. Res. 38(5): 386-395.

Metrangolo, P. and G. Resnati (2012). "Recommendation from IUPAC Task Group 2009-032-1-100: Definition of the Halogen Bond." Personal Cummunication, In Review.

Murray, J. S., P. Lane, et al. (2007). "Sigma-hole bonding: molecules containing group VI atoms." Journal of Molecular Modeling 13(10): 1033-1038.

Murray, J. S., K. E. Riley, et al. (2010). "Directional Weak Intermolecular Interactions: sigma-Hole Bonding." Australian Journal of Chemistry 63(12): 1598-1607.

Muzet, N., B. Guillot, et al. (2003). "Electrostatic complementarity in an aldose reductase complex from ultra-high-resolution crystallography and first-principles calculations." Proc. Natl. Acad. Sci., USA 100(15): 8742-8747.

Parisini, E., P. Metrangolo, et al. (2011). "Halogen bonding in halocarbon-protein complexes: a structural survey." Chem Soc Rev 40(5): 2267-2278.

Riley, K. E., J. S. Murray, et al. (2011). "Halogen bond tunability I: the effects of aromatic fluorine substitution on the strengths of halogen-bonding interactions involving chlorine, bromine, and iodine." J Mol Model 17(12): 3309-3318.

Riley, K. E., J. S. Murray, et al. (2009). "Br···O Complexes as probes of factors affecting halogen bonding: Interactions of bromobenzenes and bromopyrimidines with acetone." J Chem Theory Comput 5: 155-163.

Ritter, S. (2009). "Halogen bonding begins to fly." Chemical and Engineering News 87(38): 39-42.

Shields, Z. P., J. S. Murray, et al. (2010). "Directional Tendencies of Halogen and Hydrogen Bonds." International Journal of Quantum Chemistry 110(15): 2823-2832.

van Pee, K. H. and S. Unversucht (2003). "Biological dehalogenation and halogenation reactions." Chemosphere 52(2): 299-312.

Voth, A. R., and Ho, P.S. (2007). "The Role of Halogen Bonding in Inhibitor Recognition and Binding by Protein Kinases." Curr. Topics Med. Chem. 7: 1336-1348.

Voth, A. R., Khuu, P., Oishi, K., and Ho, P.S. (2009). "Halogen bonds as orthogonal molecular interactions to hydrogen bonds." Nature Chemistry 1: 74-79.

Wu, W., Samoszuk, M. K., Comhair, S. A. A., Thomassen, M. J., Farver, C. F., Dweik, R. A., Kavuru, M. S., Erzurum, S. C., and Hazen, S. L. (2000). "Eosinophils generate brominating oxidants in allergen-induced asthma." J. Clin. Invest. 105: 1455-1463.

Xu, Z., Z. Liu, et al. (2011). "Utilization of halogen bond in lead optimization: a case study of rational design of potent phosphodiesterase type 5 (PDE5) inhibitors." J Med Chem 54(15): 5607-5611.

Zhou, P. P., W. Y. Qiu, et al. (2011). "Halogen as halogen-bonding donor and hydrogen-bonding acceptor simultaneously in ring-shaped H3N.X(Y).HF (X = Cl, Br and Y = F, Cl, Br) complexes." Phys Chem Chem Phys 13(16): 7408-7418.

CHAPTER 2

DNA STRUCTURE: ALPHABET SOUP FOR THE CELLULAR SOUL[1]

1. Introduction

The story of DNA structure is as varied as it is interesting, the most famous tale being the "discovery" of B-DNA by Watson and Crick. For many biologists, this simple, but elegant structure is all that is needed for a basic, albeit superficial understanding of cellular genetics. A deeper appreciation for how DNA functions comes from the recognition that this is a highly malleable molecule, providing the cell with a plethora of conformations to exploit during replication and transcription. Some of these conformations can give rise to mistakes, while others help to repair those mistakes in the genetic code. In this chapter, we dive into the cellular pot and find a literal alphabet soup of DNA structures. We start our journey by presenting the fundamental principles that serve as the vocabulary to analyze and describe the features of nucleic acid structures. We will explore the conformational variations that lead from double-helices to complexes composed of three or four strands, then consider how conformations interconvert through

[1] P. Shing Ho and Megan Carter
Published as a chapter in DNA Replication-Current Advances
Dr Herve Seligmann (Ed.), ISBN: 978-953-307-593-8, InTech,

various intermediates. Although B-DNA is the standard form in the cell, we suggest that this dance away from the norm is essential for cellular function, giving the cell its genetic soul.

Replication is the process by which the cell creates an exact copy of the genetic information coded in DNA—it is thus intuitive that we would be interested in the actual structure of DNA as a molecule. One would think that, for replication, we need only be concerned with the DNA duplex at the beginning, the single-stranded intermediate state, and the final duplex at the end, since these generally tell us how the information is stored and read, and what the resulting product is. What is becoming clear is that although the general structure of DNA is important in the overall mechanism of replication (Watson & Crick, 1953a), the conformational details are important for understanding how proteins recognize their cognate DNA sequence, and how mutations may be introduced and are repaired. Thus, we must explore and dissect the details in terms of variations that define the particular sequence dependent shape of DNA.

We will not attempt the impossible task of covering every aspect of DNA structure, only those that may be relevant to replication. Also, as crystallographers, we will have a bias towards studies derived from X-ray diffraction and other physical methods, although we will always attempt to relate these back to the biology of replication. In the process, we will explore the details of DNA structure that help elucidate structural principles that contribute to our understanding of the mechanism and fidelity of the replicative process.

2. A brief history of DNA structure

DNA structure has had over 55 years of history and, in that time, has undergone periods of discovery that have pushed the field forward in spurts. The evidence that DNA is the genetic molecule in the cell came from the studies of Avery, MacLeod, and McCarty (Avery et al., 1944), and confirmed by Hershey and Chase (Hershey and Chase, 1952). The seminal experiments of Meselsen and Stahl (Meselson and Stahl, 1958) using heavy atom labeled DNA demonstrated that replication is semiconservative, with each newly replicated daughter strand being paired with one of the two parental strands. These classic studies from the 1940's and 1950's set the stage for a race to determine the molecular structure of DNA, a now familiar story that helps to bring perspective to the discussions in this chapter.

2.1 The race for the structure of DNA: X-ray fiber diffraction studies.

The key element in the race towards the structure of DNA was the availability of X-ray diffraction photographs of DNA fibers, the best of which came from the work of Franklin and Gosling in the lab of John Randall. It was clear at the time that DNA could adopt two different forms, an A-form under low humidity and a B-form at higher humidity. The A-DNA form gave the highest resolution data (Franklin and Gosling, 1953a), but, it was the lower resolution photograph of the "wetter" B-form (Franklin and Gosling, 1953b) (Fig. 2.1) that was more readily interpretable. From this photograph, DNA was clearly seen to be a helical structure (showing the characteristic "helical-X" in the diffraction pattern), with a repeat of 10 units (reflected in the pattern converging after 10 layer lines), and with a distance between repeating units of 3.4 Å (from the d-spacing

of 10th layer line). What was not evident was the number of strands in the helix (indeed, Linus Pauling had initially proposed a three-stranded structure (Pauling and Corey, 1953)), whether it is left- or right-handed, and how the information is read and properly replicated. The interpretation of this data by Watson and Crick (Watson and Crick, 1953b) lead to the iconic right-handed, antiparallel, double-helical model of DNA that we all recognize.

Often missing from this story is that the Watson-Crick model depended not only on the large amount of biochemical and X-ray diffraction data being generated at the time, but also on a proper understanding of the chemical properties of DNA. One of the most important aspects of the Watson-Crick model was the proposal that guanines paired with cytosines and adenines with thymines. For this to occur, however, the nucleotide bases must be drawn in their proper tautomeric forms; however, up to that point, it was not clear, even to the organic chemists, what those forms should be. The initial assignment of guanine and thymine bases in their enol forms had lead to an early parallel model for DNA (Watson, 1968). It was not until the proper tautomers for the common nucleotides were assigned that the now familiar base pairs of G to C and A to T made sense, and, thus, provide a rationale for the well understood Chargaff rules for the complementary composition of nucleotides in the DNA of higher organisms (Chargaff, 1950) and a mechanism by which exact copies of the sequence information along a strand of DNA could result in an exact copy of a duplex through semiconservative replication (Watson and Crick, 1953a).

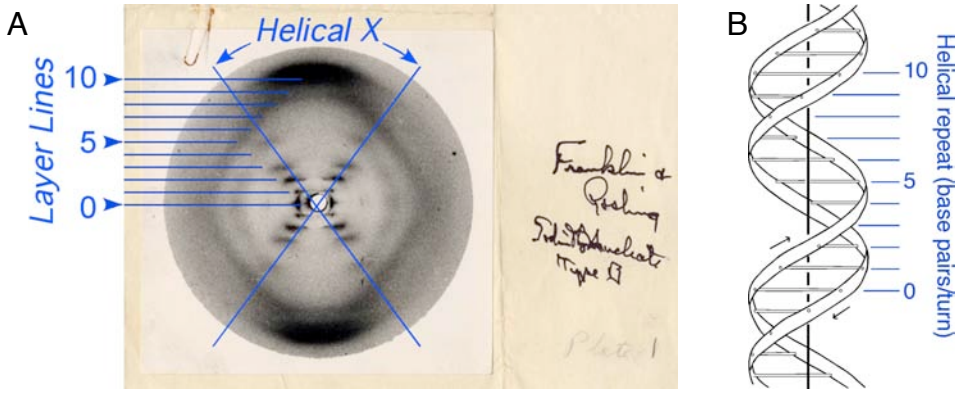


Fig. 2.1. Structure of B-DNA. A. Photograph 51 of B-DNA. X-ray diffraction photograph of a DNA fiber at high humidity (Franklin and Gosling, 1953b). Interpretation of the helical-X and layer lines added in blue. B. Watson-Crick model of B-DNA, adopted from (Watson and Crick, 1953b), with the helical repeat associated with the layer lines labeled.

2.2 The Single-crystal structures of DNA oligonucleotides.

At this point, it should be stressed that Watson and Crick did not "discover" or "solve" the structure of DNA, but had presented a plausible and, basically, correct model that made important predictions that, in the end, led to the birth of modern molecular biology. However, several decades will pass before high resolution single crystals structures of synthetic DNAs emerge to support the essential elements of this model. For example, it was not immediately obvious that the Watson-Crick scheme, particularly for A=T base pairs, was correct—at the time, the crystal structures of adenine bases paired with thymine or uracil bases showed geometries of Hoogsteen-type base pairs (this will be defined in Section 3). It was not until the single crystal structure of the RNA dinucleotide phosphate ApU was determined to a remarkable 0.89 Å resolution (in crystallography, lower numbers refer to higher resolution) by Alexander Rich's group (Rosenberg et al., 1973) that the Watson-Crick form of the A=U (and, thus, the analogous A=T) base pairs were confirmed. The concurrent structure of GpC also confirmed the Watson-Crick form of the G•C base pair (Day et al., 1973) and, together, demonstrated for the first time that nucleotide double-helices (in this case, RNA dinucleotides) were antiparallel and had a right-handed twist.

In the late 1970's, it became possible to chemically synthesize "long" stretches of a defined DNA sequence for crystallographic studies. In 1979, Rich's group (Wang et al., 1979) determined the single crystal structure of the DNA sequence CGCGCG (we write only one strand and drop the "p" for the phosphates for the sake of efficiency, even for double-helical structures). This structure showed DNA to be an antiparallel double-helix with Watson-Crick type base pairs, consistent with the 1953 model. However, it came

with a new twist—this double-helix was left-handed and was called Z-DNA (for the zig-zagged backbone). It was not until 1981, with the single-crystal structure of the sequence CGCGTATACGCG (known as the Drew-Dickerson dodecamer (Drew et al., 1981)), that the Watson-Crick structure for B-DNA was finally "proven" to be correct.

So, what of the dehydrated A-DNA form that Franklin had worked so hard on and struggled with? Soon after the Watson and Crick model of B-DNA, Franklin and Gosling published the structure of the fiber A-DNA form (Franklin and Gosling, 1953a), with a large number of single-crystals of A-DNA being determined and published in the 1980's and 1990's (the "heydays" of DNA crystallography (Mirkin, 2008)). The A-form was subsequently shown to be the native form of RNA duplexes, while DNA/RNA hybrids (primers for replication initiation) can interchange between the A- and B-forms.

Although it is well accepted that the B-DNA form is the most prevalent form in solution and in the cell, there is now a myriad of single-crystal DNA structures, including those assembled as double-, triple-, quadruple-, and even hexa- and octa-stranded complexes. There are hairpins from single-strands, structures with overhangs, etc., and a plethora of forms seen in complexes with proteins. We will discuss some of these in greater detail in Section 4 along with their relevant cellular functions, focusing on replication and the associated processes. First, we must delve into the detailed vocabulary used to describe DNA structure and provide a common language for the remainder of the chapter.

3. A vocabulary lesson for DNA structure

As with any description of a biopolymer, we will start the discussion of DNA structure at the simplest unit (the nucleotide building block), then develop the concepts of structure with increasing size and complexity. In order to reach this stage of complexity, we must first define terms that will be used in discussing DNA structure at all levels.

3.1 General principles

Almost every student today knows that DNA is composed of four basic building blocks, each defined by the unique chemical structure of the aromatic base, and each base attached to a phosphodeoxyribose backbone. The four common deoxynucleotides are categorized as the purine (deoxyadenosine, dA, and deoxyguanosine, dG) or pyrimidine (deoxythymidine, dT, and deoxycytosine, dC) nucleotides. The atoms of sugars are distinguished from those of the bases by a "prime" added to the atom name, so that the sugar carbons are C1', C2', C3', C4', C5', starting with the carbon at the glycosidic bond that attaches the base to the sugar, and so forth around the ring. The deoxynucleotides of DNA lack a O2' oxygen, which distinguishes them from ribonucleotides (RNA). For simplicity, we will simply assume the deoxyform and drop the "deoxy" and "d" prefixes from this point on (Hendrickson et al., 1988).

3.2 What defines a stable DNA structure?

DNA in its functional form is not the isolated nucleotides, but a polymer built from the mononucleotides (G, C, A, T). A DNA polymer is constructed through condensation to form a phosphodiester linkage that bridges the O3' and O5' oxygens of

sequential nucleotides (2.2A). The primary structure, or sequence, of a DNA polymer strand is written in the direction that they are synthesized in the cell, starting at the free O5' oxygen (5'-end) and progresses to the free O3' –end. Two complement strands are brought together in a sequence specific manner to form an antiparallel double-strand, aligning one strand in the 5' to 3' direction and the complement 3' to 5'. Nearly all functional secondary structures of DNA are multi-stranded, most commonly double-stranded. As the sequence of one strand dictates that of its complementary, double-stranded DNA is often considered as a single biological molecule, even though the strands are not covalently linked.

3.2.1 Base pairing

Unlike proteins and RNA, the functional forms of DNA are typically complexes comprised of two or more strands, which are stabilized by base pairing, base stacking, and solvent interactions. Of these, base pairing is best understood for its important role in specifying the sequence of newly synthesized DNA during replication and in general sequence recognition, but is perhaps the most misunderstood for its contribution to DNA stability.

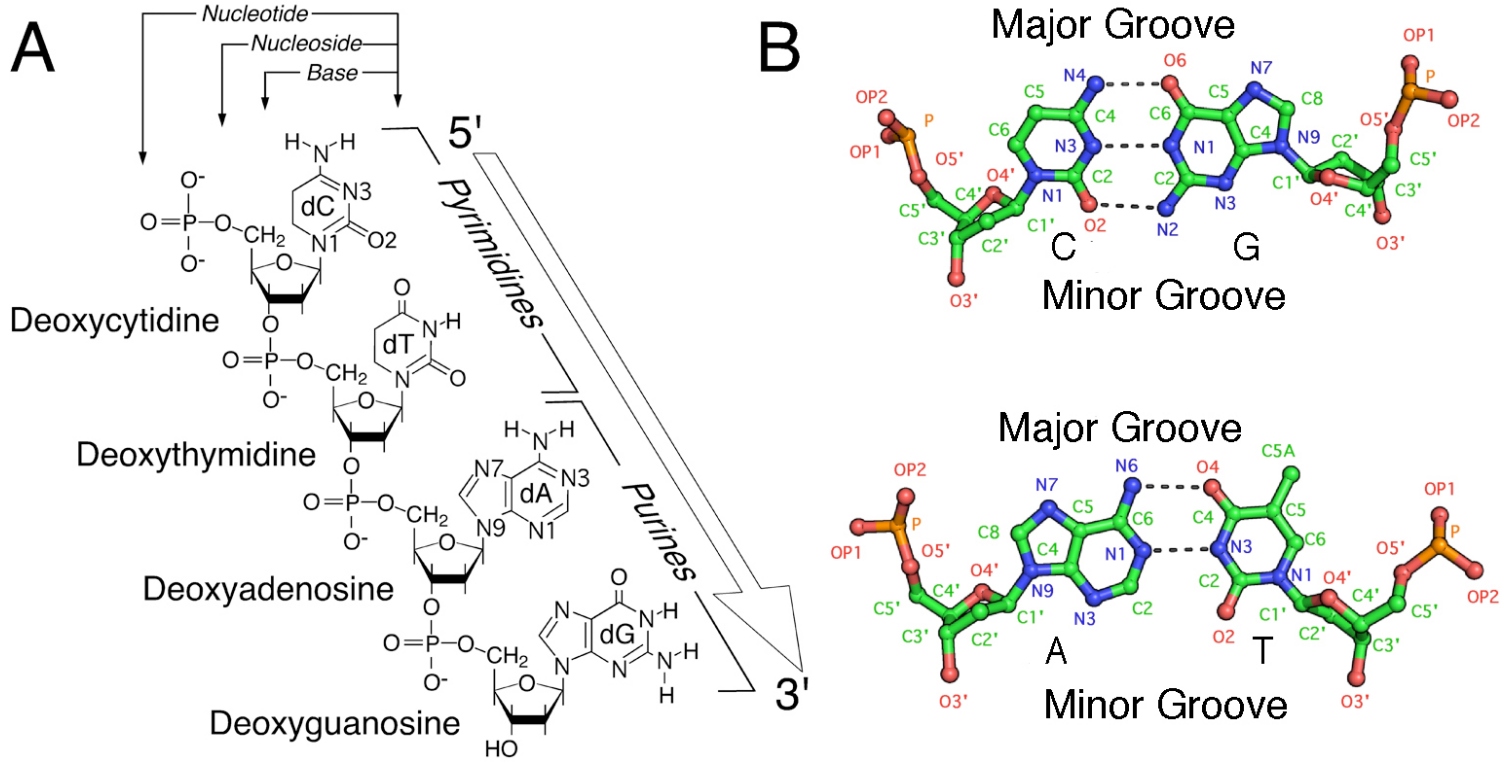


Fig. 2.2. Components of DNA. A. The four common deoxyribonucleotides are connected by phosphodiester bonds to form a single-strand, 5' to 3'. B. Watson and Crick C•G and A•T base pairs with the major and minor grooves labeled.

The most commonly recognized form of DNA, B-DNA, is the double-stranded duplex stabilized by Watson-Crick base pairing (Fig. 2.2B). In standard Watson-Crick G•C and A•T base pairs, hydrogen bonds are formed between the respective donor and acceptor functional groups along what is called the "Watson-Crick" edges of the bases. The geometries of these purine-pyrimidine base pairs are similar in the relative positions of their bases and, consequently, the width of the resulting major and minor groove—the similarity in the geometries of correctly paired bases contributes to the fidelity of the replication polymerases [(Kool, 2001)]. The G•C base pair, however, is stabilized by three hydrogen bonds as opposed to the two that stabilize A•T base pairs; thus, G•C rich sequences tend to have higher stabilization energies and melting temperatures. With only two hydrogen bonds, A•T base pairs offer less resistance to deformations, including twisting of the individual bases from a common plane (called propeller twist, see below). Although the standard Watson-Crick base paired duplex DNA is most universally recognized, it is clear that DNA structures with non-standard pairing of bases are more prevalent and biologically significant than previously thought [(Neidle, 1999)].

Non-standard base pairs play critical roles in the varied structures observed in DNA and RNA. Wobble, mismatched, and reverse base pairs still use the Watson-Crick edges for hydrogen bonding. Reverse Watson-Crick base pairs are found in parallel duplexes, but are not immediately relevant to DNA replication. Wobble base pairing (Fig. 2.3A) is seen in mismatches between G•T and G•U base pairs incorporated into DNA and DNA:RNA complexes and play essential roles in the fidelity of DNA replication and transcription. Such mismatches can lead to genome mutations if not accurately detected and corrected by the proof reading activity of DNA polymerase during replication, or

post-replicative repair systems. Studies suggest that G•T and A^{+}•C are the most frequent mismatches that cause point mutations in cells (Neidle, 1999). The energies of hydrogen bonding in proper and mismatched bases, relative to base stacking and steric effects, however, appear to have little influence on polymerase fidelity (Kool, 2001).

Hoogsteen base pairs take advantage of the Hoogsteen edge of a purine base, which is orthogonal to and, thus, can be accessed without disrupting the Watson-Crick base pairing edge (Fig. 2.3B). Consequently, Hoogsteen interactions allow the assembly of multi-stranded DNA complexes, including triplet helixes and G-quadruplexes.

3.2.2 Base stacking

Although not as intuitive, the stacking of bases into a column is as or more critical to the stability of multistranded DNAs (duplexes, triplexes, tetraplexes, etc) as base pairing. It is estimated that base stacking contributes as much as half of the total stabilizing free energy of a base pair in duplex DNA (Kool, 2001). Van der Waals interactions, electrostatic interactions, and solvent effects define the geometry and associated energies of stacked bases. Van der Waals forces drive bases to stack in a way that best complements their surface topologies. In addition, individual atoms carry permanent partial charges that contribute to either Coulombic attraction or repulsion between bases. This can be modeled as interactions between permanent dipoles, and it is this dipolar interaction, in conjunction with shape complementarities that helps to define the orientation of the stacked bases. The specific orientation of stacked base pairs contributes to the conformational stability of a DNA duplex.

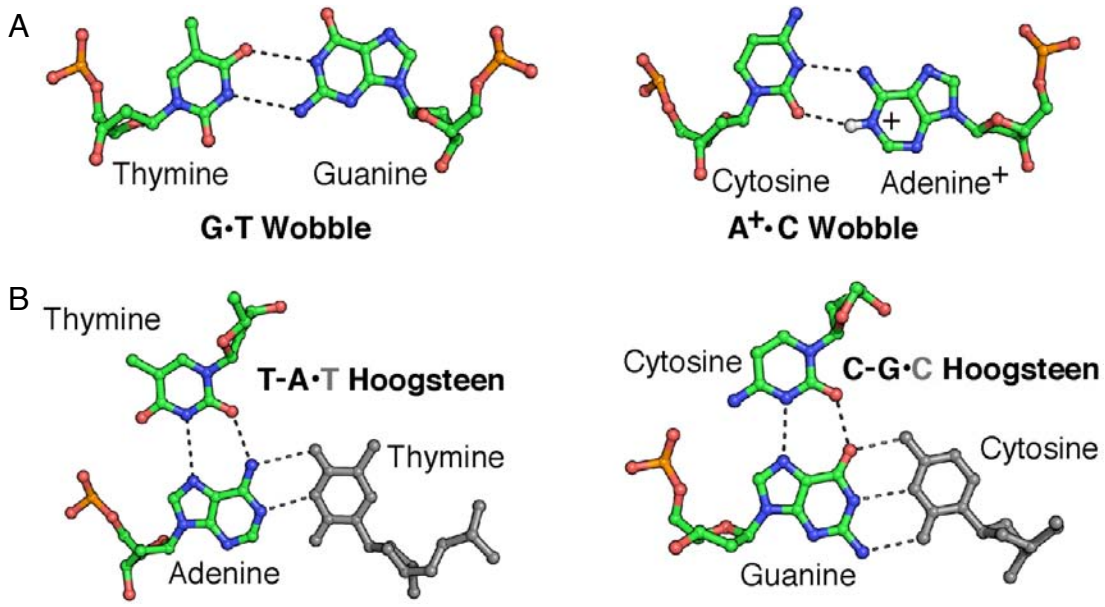


Fig. 2.3. Non-Watson Crick base pairs. A. G•T wobble and A^{+}•C wobble base pairs. B. Thymine Hoogsteen paired to A•T WC base pair, cytosine Hoogsteen paired to G•C WC base pair as observed in triplex strand formation.

Likewise, deformations associated with specific base stacking geometries contribute to the mechanism of indirect sequence specific binding and recognition by proteins. Finally, since the nucleotide bases are aromatic and, therefore, primarily hydrophobic, stacking minimizes the solvent exposure of the base surfaces, thus, leading to the familiar face-to-face stacking of bases and base pairs. It is not surprising, therefore, that DNA conformations that increase exposure of bases are stabilized by organic solvents.

3.2.3 The phosphodeoxyribose backbone

The functional form of DNA links nucleotides together by phosphodiester bonds to form a continuous DNA strand. Phosphodiesters are highly acidic (pKa' ~1.5); thus, at neutral pH, the phosphate group is a monoanion with a formal -1 charge distributed among all four oxygens, with the two non-ester oxygens (OP1, OP2) carrying about twice the charge as the ester bonded oxygens (O5', O3'). As a consequence, the DNA phosphoribose backbone is overall negative and provides an opposing force to the base pairing and stacking interactions that hold a DNA duplex together. Indeed, if the backbone were uncharged, it would be much more difficult to unzip or displace a DNA strand and, consequently, it would take more energy to unwind a duplex to allow replication to start and proceed.

The overall charge of DNA in solution is not simply a sum of -1 for each nucleotide—the backbone charges are counterbalanced by positive cations that accumulate around the DNA. These counterions are simple ions (monovalent Na^{+} and K^{+}, or divalent Mg^{+2} and Ca^{+2} being the most prevalent in a cell), but include cationic polyamines (spermine and spermidine), drugs (ethidium or cis-platin), or proteins (e.g.,

the histone proteins of nuclesomes). In general, DNA in solution is less negatively charged than expected—as a polyelectrolyte, each phosphate of a DNA duplex carries an "effective" charge of approximately -0.6, or ~40% of the charge is counterbalanced by simple cations (Manning, 1977). The remaining net charge, however, acts to destabilize the double-helix. Consequently, structures with closely spaced phosphates are stabilized by increased concentrations of counter cations.

When a protein, such as DNA polymerase, binds to DNA, it must competitively displace the counterions associated with the DNA backbone. For example, nucleosome formation, which helps compact DNA in eukaryotes, is primarily driven by nonspecific interactions of the positive histones with the negative DNA backbone. In order to replicate or transcribe the information of the DNA, the respective polymerase and all of its associated proteins must compete against these non-specific interactions. Thus, the negative charge of the backbone is a platform for sequence independent electrostatic interactions with proteins in the cell (Rohs, et al., 2009).

3.2.4 Solvent Effects

As with any biological molecule, solvent interactions directly influence DNA structure and function. Base pairing and stacking are in part stabilized by the hydrophobic effect. We have already seen how solvent (considered to consist primarily of water and salts) induces base pairs to stack and defines the effective charge of the phosphoribose backbone. Even base pairing is affected by solvent interactions. In forming a base pair, the hydrogen bond donor and acceptor groups of each base must break hydrogen bonds with water molecules first. If the enthalpy of any single hydrogen bond from one base to

another base is essentially the same as they are from the base to water, why then do bases pair and exclude water (at 55.5 M concentration)? The primary answer is that sequestering hydrogen-bonding groups from the competing interactions of water increase the hydrogen bonding potential (Klotz, 1962). One can see from this why base stacking is so important in stabilizing double-, triple, and other multistranded DNA forms that are assembled through hydrogen bonding.

Water, however, is not entirely excluded from, but plays an important role in the structure of DNA. Even in a fully base paired duplex, numerous hydrogen bond donor and acceptor groups of the backbone and bases must be hydrated. There are classes of waters that can, in fact, be considered integral components of a DNA's structure. In a G•T wobble base pair, for example, the number of hydrogen bonds between the bases is reduced by one; however, bridging water molecules help to compensate for this loss (Ho et al., 1985). Similarly, there are well-defined waters lining the minor groove of B-DNA duplexes (the so-called "spine of hydration") (Drew et al., 1981) that exchange slowly with the bulk solvent (Liepinsh et al., 1992) and, therefore, are considered to be integral parts of DNA. Thus, water promotes base stacking, which provides an environment for more stable hydrogen bonds within base pairs. Waters solvate the surfaces of the major groove and form well defined hydrogen bonded networks that bridge the two strands across the minor groove. In order to minimize the opposing repulsion between the phosphates of the DNA strands, cations help to mitigate the negative charges of the phosphoribose backbone (Hamelberg et al., 2001). It is evident, therefore, just how important solvent really is for DNA structure and stability.

Finally, we must briefly discuss how solvent plays a role in DNA function. DNA is a hydrated molecule, until it is bound to a protein, at which point the DNA becomes dehydrated—i.e., a protein must compete against water in order to bind to the DNA. The basic concept of direct read-out of DNA base pairs is a prime example of this. Direct read-out requires a protein to essentially stick its hydrogen bonding side-chain fingers into places where they would not normally belong, the major groove of a DNA duplex, for example. These proteins side chains and the DNA surface that they are trying to read would prefer to remain solvated; however, in order to form a strong complex with DNA, the protein must expel water from both surfaces and, as a result, the complex will become more stable than the sum of the individual parts. This, again, requires a balance between the stability of hydrogen bonds, the resulting decrease in conformational entropy of the protein side chains, and an increase in entropy of the water molecules as they return to the bulk solvent.

3.3 Conformations of the deoxyribose sugar

In addition to charge effects, the phosphoribose backbone helps to define the conformation of DNA via the conformation of the deoxyribose sugar. The detailed conformation of any polymer is defined by the rotations about each freely rotating chemical bond (Fig. 2.4A). We can define three categories of bonds: those of the phosphodiester holding two nucleotides together, those within the five-membered ring of the deoxyribose sugar, and the bond holding the nucleotide base to the sugar. The angles around the bonds that hold two nucleotides together start at the oxygen that links phosphate to the C5'-carbon of the ribose ring.

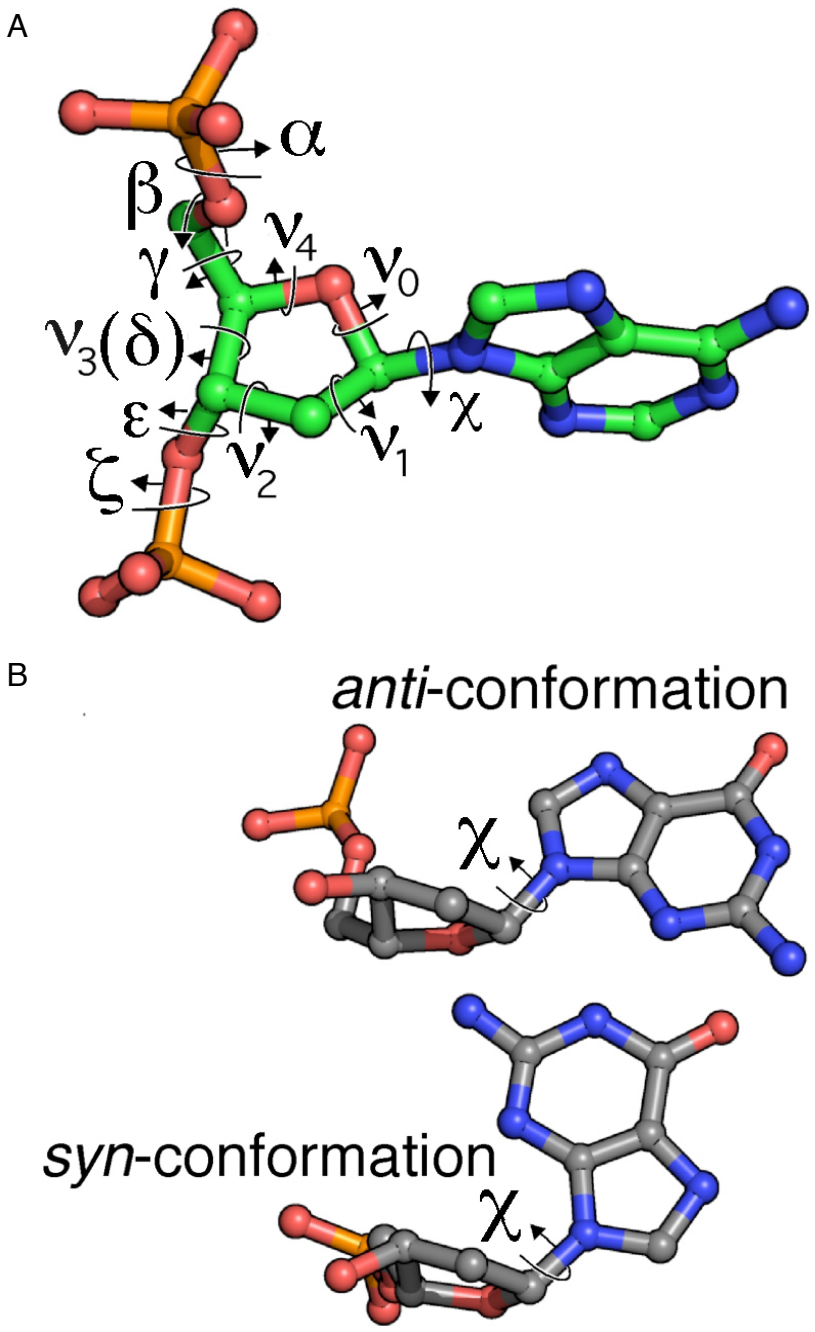


Fig. 2.4. Torsion angles of nucleic acids. A. Torsion angles along the backbone (α to ζ), within the sugar ring (ν_0 to ν_4), and the rotation of the nucleobase relative to the sugar. B. Rotation about the glycosidic bond defines χ-angles for the anti- and syn-conformations of the bases.

Rotation about the P-O5' bond is the α-torsion angle, which is followed by the β-angle for the O5'-C5' bond, and so forth until we get to the ζ-angle that links the O3'-oxygen to the phosphate of the next nucleotide. These bonds adopt angles that help to minimize the repulsion of the negatively charged phosphates within and between DNA strands.

The bonds in the furanose ring are distinguished from those that flow linearly from one nucleotide to the next, and are designated as ν_1 for the C1'-C2' bond, ν_2 for the C2'-C3' bond, and so forth (Fig. 2.4A). The reader would recognize that the ν_3 angle within the ring coincides with the d-angle along the chain. The ring is non-planar, and it is how particular atoms are placed either above or below a reference plane (the "sugar pucker") that facilitates formation of various conformational forms of DNA. The torsion angles are correlated to maintain reasonable bond lengths and angles within the ring, and are described by a single pseudorotation angle Ψ, which defines the sugar pucker. Sugars with atoms puckered above the reference plane (on the same side as the base) are in an endo-form (C2'-endo pucker has the C2'-carbon pointed up and towards the base), while a pucker that places an atom below this plane is in its exo-form (Fig. 2.5). The two general classes of sugar conformations commonly seen in DNA are the C2'-endo and C3'-endo puckers—the interconversion between these forms will be discussed in detail in section 5. The two conformations have profound effects on the overall DNA conformation in that they specify different phosphate-phosphate distances along each strand (~7 Å for C2'-endo and ~6 Å for C3'-endo). Thus, conformations constructed with C3'-endo sugars will require higher concentrations of salts to counter balance the shorter distance between the negatively charged phosphates.

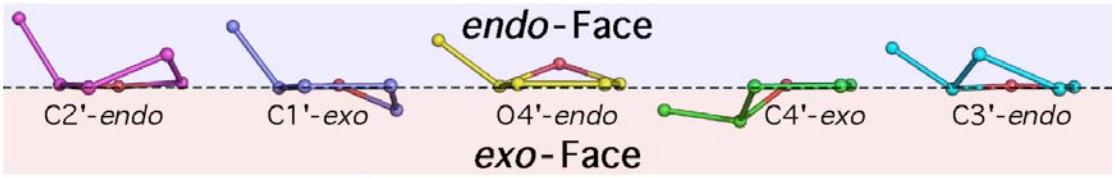


Fig. 2.5. Sugar pucker. Shown are the endo (above) and exo (below) faces of the 5-membered furanose sugar with the nucleotide base extended above the reference plane. Sugars are shown in order of transformation from C2'-endo to C3'-endo.

The base of each nucleotide is attached via the glycosidic bond from the N1 nitrogen of pyrimidines or the N9 nitrogen of purines to the C1'-carbon of the deoxyribose sugar. The rotation about the glycosidic bond, the χ-angle, defines two general conformational classes: the anti conformation ($+90° \leq \chi \leq +180°$), with the base extended away from the sugar, and the syn conformation ($-90° \leq \chi \leq +90°$), with the base essentially lying on top of the sugar ring (Fig. 2.4B). The more compact syn-conformation is more susceptible to steric clashes than the extended anti-form. Although purine rings are generally larger, it has the smaller five-membered ring, as opposed to the six-membered ring of pyrimidines, attached to the sugar. Thus, purines will more readily adopt the compact syn-conformation than pyrimidines, because of reduced steric collisions. Similarly, the syn conformation is less sterically hindered when the sugar is puckered as C3'-endo than C2'-endo. From this, we can now start to appreciate how the interplay between sugar puckers and c-rotations can have profound effects on the structures of DNA and the sequence dependence for their formation.

3.4 Helical parameters

Now that we have assembled well-defined helical structures, how do we describe these structures? We can certainly do this in a very descriptive and qualitative manner, using the classical A- and B-forms as examples. For instance, we can characterize the standard B-form of DNA as a right-handed double-helix held together by Watson-Crick type base pairs that stack directly along a helical axis, resulting in two well defined grooves. However, this raises numerous questions, for example, at which point does a distortion to the Watson-Crick base pair become a wobble base pair, how far off the helix

axis is allowed in this definition, and what if the helix axis is not straight? To address these and other questions, a set of quantitative measures called the "helical parameters" were developed to characterize the regular secondary structures of nucleic acids (both DNA and RNA).

The most commonly recognized parameters for DNA include the helical repeat (number of base pairs in one complete turn) and the helical rise (distance between nucleotides when measured along the helical axis). The repeat defines the angle relating each base pair along the helix axis (the helical twist = 360°/repeat), while the product of repeat and rise is the pitch (distance between one complete turn) of the DNA. These parameters restrict the geometries of the DNA. Indeed, if we consider only the closest physical approach between base pairs (the rise = 3.4 Å, as defined by the thickness of a base), the maximum phosphate-phosphate distance along a strand (measured at ~7.5 Å by single-molecule stretching (Allemand et al., 1998)), and the effective diameter of a duplex (9.5 Å), we see that the largest twist angle between stacked base pairs is ~42°, resulting in a smallest theoretical repeat of 8.5 base pairs per turn. This would be the most tightly or over-wound form of a DNA double-helix. If the phosphate-to-phosphate distance is relaxed to ~7 Å (for a C2'-endo sugar pucker), the helical twist becomes ~36°, which translates to the repeat ~10 bp/turn repeat of B-DNA. Finally, if the sugar adopts a C3'-endo conformation with a ~6Å phosphate-to-phosphate distance, the result is a structure with a helical twist of ~31° and a repeat of 11 – 12 base pairs, similar to that of A-DNA. We can see, therefore, how the sugar pucker defines the intrastrand phosphate-to-phosphate distance, base stacking defines the base-to-base distance, the base pairs define the radius of the DNA, and, finally, how all this comes together to define the way

the DNA double-helix twists into a specific conformation. Of course, these are only very rough approximations of DNA structures—the detailed descriptions require a set of helical parameters in addition to the two described so far.

The helical parameters can be categorized in two general classes to describe the absolute and relative conformations in nucleic acids (Fig. 2.6); base-pair parameters (for single base pairs) and base step parameters (for adjacent base pairs). We note that these classes are not mutually exclusive, but are interrelated. Twist and rise are clearly base step parameters, since they describe the relative angle and distance between two adjacent stacked base pairs. The other base-step parameters that are generally considered relevant include slide, roll, tilt, and shift. It is easy to see that slide can effectively increase the diameter of a DNA duplex and, consequently affect the helical twist and repeat. A-DNA, for example, shows a large slide between base pairs, while B-DNAs have the base pairs essentially stacked on top of each other. Not surprisingly, therefore, A-DNA has a larger overall diameter and, in fact, appears to have a hole down the middle when viewed down its helical axis.

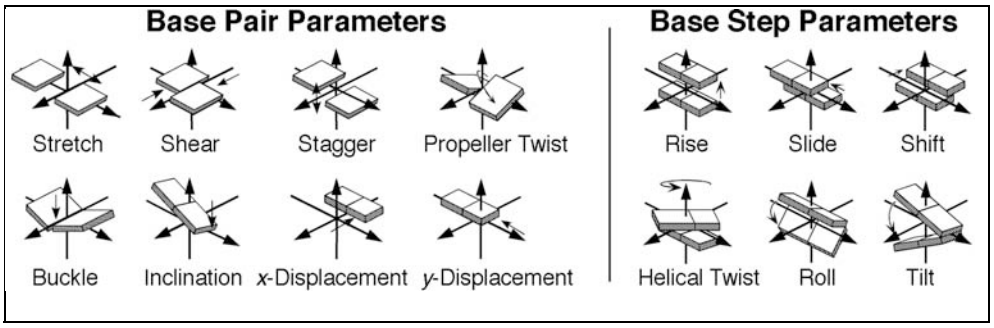


Fig. 2.6. Base Pair and Base Step Parameters. Base Pair Parameters: Translational and rotational relationships of bases within each base pair. Base Step Parameters: Translational and rotational relationsips between two stacked base pairs.

A conundrum in A-DNA is that it has a rise of ~2.5 Å, which would appear to violate the closest approach between stacked base pairs. In this case, the inclination associated with the roll and tilt of the base pairs, in conjunction with the helical twist result in a shortening of the vertical distance between base pairs along the helical axis, even though the stacking distance remains 3.4 Å. Indeed, A-like DNAs that have little or no roll and tilt have helical rises that are ~3.4 Å, as expected (Ng et al., 2000; Vargason et al., 2001).

Base pair parameters include those that relate the position or orientation of the base pair relative to the helical axis (inclination, x-displacement, and y-displacement), or the orientation and positions of the two bases in a pair (propeller twist, shear, stagger, stretch, buckle). It should be obvious that the inclination of a base pair will strongly influence the roll and tilt between base pairs, while slide defines the displacement perpendicular to the base pair (x) and along the base pair (y). Within the base pair itself, the large propeller twist seen in A•T base pairs has been attributed to the flexibility of two hydrogen bonds relative to three observed in G•C base pairs. At the extreme, this results in bifurcated hydrogen bonds, which are considered to be shared between adjacent A•T base pairs (Coll et al., 1987).

Each of these base pair and base step parameters are defined relative to the helical axis that runs down the center of DNA. However, it should be recognized that defining this axis is not entirely straight forward, particularly if the DNA trajectory is bent or curved. There are two approaches to defining helical axes: the global axis and the local axis. The global axis is essentially the continuous curve that best runs down the center of all base pairs in a structure, while the local axis is the best line that defines the center of

any two adjacent base pairs (local axes need not be continuous). Thus, helical parameters are analyzed in the context of global or local axes, and are not interchangeable and may be very different.

Two distinguishing features of double-helical DNAs are the grooves. The widths of the major and minor grooves are measured as the phosphate-to-phosphate distance across the two strands in a direction perpendicular to the trajectory of the strands. These groove widths provide an important means for proteins to interact with the base pairs of the DNA. The wide major groove of B-DNA allows direct read-out of the bases, while the narrow major groove of A-DNA does not—there is, however, an advantage to A-DNA having a wider minor groove, which we will discuss in the next section. It should be immediately obvious from the earlier discussion that the base pair and base step parameters described above conspire to define the groove widths for each form of DNA.

Finally, we can see how a parameter such as twist has such a strong effect on the overall behavior of genomic DNAs. DNA when confined in the cell or the cell's nucleus must be packaged into a compacted supercoiled form and, in the process, this induces stress that will perturb its secondary structure. For simplicity a set of terms have been defined for supercoiled DNA in the context of closed-circular double-stranded DNA such as those found in plasmids, bacterial chromosomes, and viral genomes. These terms can also be applied to linear eukaryotic DNAs that are spatially anchored and stressed through protein binding, DNA unwinding, and DNA compaction. In double-stranded DNA, the number of times the strands wrap around each other along the helical axis is defined as the twist (Tw), with positive Tw associated with right-handed and negative Tw for left-handed duplexes, and unwound duplexes (e.g., melted domains) as Tw = 0. In

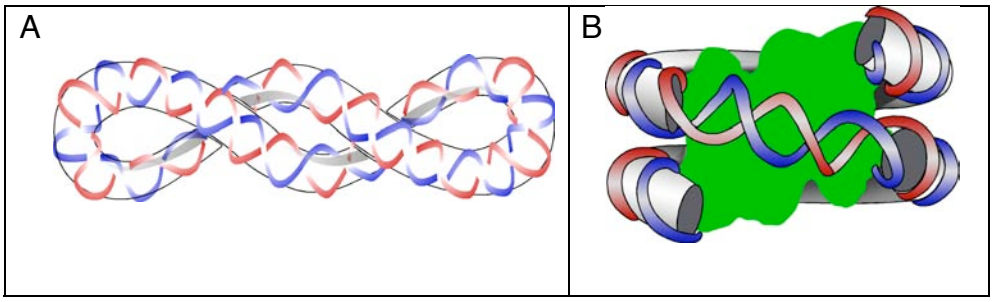


Fig. 2.7. Supercoiled DNA. A: Negative supercoils are manifest as right-handed crossovers in closed circular DNA. B: Negatively supercoiled DNA found in the nucleosome structure wraps approximately twice around the histone core proteins (green) in a left-handed direction (adapted from (Luger et al., 1997).

closed-circular DNA, the ends are joined and not free to turn in accommodating a change in Tw; therefore, a change in twist has additional global effects (Fig. 2.7), resulting in supercoiling, or writhing (Wr), of the double-helix as it wraps around itself.

Together, the twist and writhe define the topological properties of DNA. In truly closed-circular DNA that is unconstrained, twist and writhe are entirely correlated through the linking number (Lk) according to the equation Lk = Tw + Wr. Thus, if we unwind (reduce Tw) in closed circular DNA, the resulting strain must be relieved by increasing Wr (supercoiling). The only way to change Lk is by breaking the bonds of the backbone of one or both of the DNA strands, a process carried out by topoisomerases in the cells. How does all of this play out during replication? Consider the closed circular genome of a bacterium, or a domain of a eukaryotic genome that is locally constrained by nucleosomes and/or matrix attachment regions (MARs). As a DNA helicase plows through the DNA, it will locally unwind and melt the duplex (reduce Tw) for synthesis of the daughter strand. In doing so, the DNA in front of the polymerase will be positively supercoiled, while negative supercoils accumulate in its wake, both energetically unfavorable conditions. To relieve the strain, topoisomerases must relax the supercoils both in front of and behind the replisome.

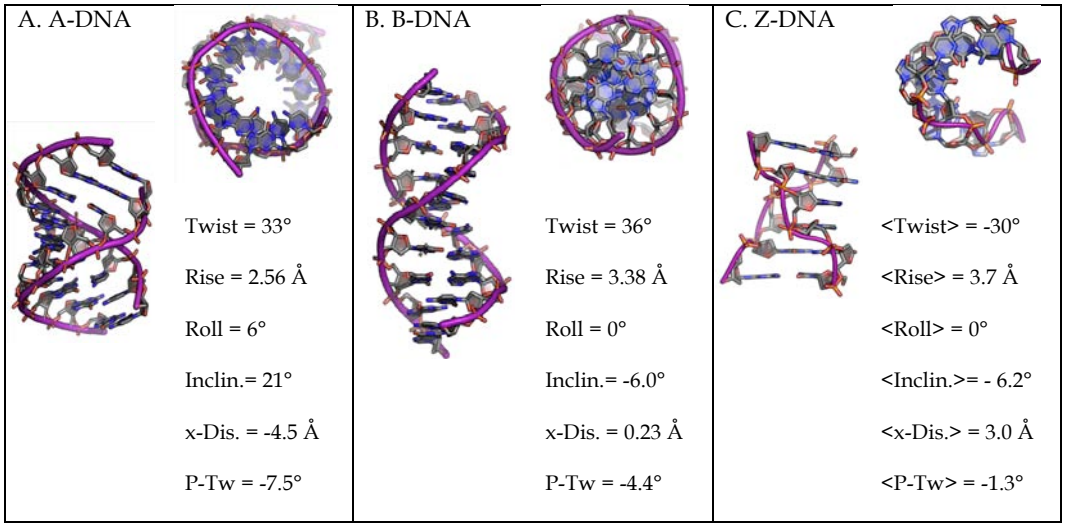


Fig. 2.8. Representatiive double-helical structures of DNA. Structures of A-DNA (Hays et al., 2005), B-DNA (Privé et al., 1991), and Z-DNA (Wang et al., 1979). Abbreviations: Incl. = inclination, x-Disp. = x-displacment, P-Tw = propeller twist. For Z-DNA, the helical paramters are given as averages of the alternating dinucleotide steps.

4. The alphabet soup of DNA structure

DNA is highly polymorphic and, at least at the level of the helical structures, more variable than either proteins or RNA. The various forms of DNA have traditionally been named using the letters of the English alphabet and, from a survey of the literature, it was found that all but four letters have been assigned to at least one unique structural form (Ghosh and Bansal, 2003). We will, in this section, briefly describe a subset of DNA conformations that have been structurally characterized (Fig. 2.8 and 2.9) and the sequence propensities of these structures, starting with B-DNA and working our way through the variations on the double-helix and various multi-stranded conformations. Along the way, we will discuss their potential biological functions, particularly in DNA replication, as appropriate.

4.1 B-DNA: The standard form

B-form DNA is the most recognized and common structural form of DNA in the cell, being considered the conformation adopted by nearly all sequences within a genome. Interestingly, while B-DNA has a distinguishing set of structural properties, it is now understood to be highly variable and malleable. B-DNA is a right-handed, antiparallel double-helix in which the Watson-Crick base pairs are stacked directly along and perpendicular to the helical axis, giving rise to major and minor grooves that are similar in depth. The bases are all in the anti-conformation with a majority of deoxyribose sugars in the C2'-endo form, although the sugar puckers are more variable than in many other conformations (Dickerson, 1999). The highly accessible major groove allows for direct readout of the polynucleotide sequence by proteins through patterns of hydrogen bond

donors and acceptors that are complementary between the amino-acid side chains and each individual base pair. The more narrow minor groove, on the other hand, is characterized by a series of strongly coordinated waters and ions.

Although these properties are general for B-DNA, the structure is highly variable from one sequence to the next and for the same sequence under different conditions. The concept of sequence-based differential deformability recognizes that the B-form of a single sequence can adopt multiple conformations in response to the environment which can affect protein recognition. Therefore, the effect of sequence is important not in terms of any one structure, but instead in its malleability—the ability of that sequence to be deformed and molded as necessary for a particular function. For example, A•T base pairs and long stretches of A/T sequences (A-tract DNAs) seem to deviate significantly from the standard B-structure, showing larger propeller twists, along with narrower and more variable minor groove widths. Narrow minor grooves are shown to have preferential binding by arginine side chains of multiple DNA-binding protein families (Rohs et al., 2009), and represent a specific example of protein recognition based on sequence specific perturbations to the standard B-DNA. A-tract DNA sequences are also associated with large rolls and tilts of their base steps, resulting in rigid bending of the B-DNA duplex (Neidle, 1999). An extreme example of these perturbations is seen with the structure induced in gene promoter sequences by the TATA- binding protein in transcription (called TATA-DNA), which shows a significant tilt and roll of the base pairs, unwinding of the duplex, and widening of the minor groove in a manner similar to that seen with A-DNA (Burley, 1996).

Variations of the B-form have been primarily elucidated by detailed structural studies, particularly X-ray diffraction and NMR, on short oligonucleotides. The question that is often raised is whether these short lengths of DNA may in fact not be relevant (and, in the case of crystals, be otherwise distorted (Dickerson et al., 1994)) relative to sequences embedded in a genomic context. Studies by Tullius' group using hydroxyl-radical foot printing (Greenbaum et al., 2007), have shown significant sequence dependent variation in the solvent accessibility and, thus, the helical structure of protein-free genomic DNA. These structural variations at the genomic level are highly correlated with variations in helical parameters measured in DNA crystal structures (unpublished results) derived from a self-consistent data set (Hays et al., 2005). In conclusion, there is growing recognition that even B-DNA is a highly variable structural form of the DNA double-helix, and that sequence dependent structural variations play a critical role in protein recognition and binding.

4.2 A-DNA: Underwinding for replication fidelity

A-form DNA is also a right-handed antiparallel helical duplex, but is characterized as an underwound structure that is more compact along the helix axis and broader overall across the helix relative to B-DNA. The nucleotide bases, all anti, are shifted by large x-displacements towards the minor groove, creating a shallow, wide minor groove and a channel associated with a deep, narrow major groove. The deoxyribose sugars are consistently C3'-endo, which minimizes the potential steric clashes as the sugar is moved toward the phosphate to accommodate the sliding of the base (Dickerson, 1999).

A-DNA is involved in insuring the fidelity of DNA replication. An analysis of the structure of the Bacillus DNA polymerase in complex with duplex DNA showed a conformational switch from the B- to underwound A-form starting at the site of nucleotide incorporation and extending to four bases upstream (Kiefer et al., 1998). Why is A-DNA induced by the polymerase? There are several perspectives on this answer, from an evolutionary view (the emergence of DNA polymerase from the primoidial RNA world where RNA polymerase reigned) to a functional view. We will discuss the latter in slightly greater detail. The direct read-out mechanism involves sticking amino acid side-chains into the DNA's major groove to read the unique pattern of hydrogen bonding donors and acceptors that specify a particular sequence. One would think that this would be a fairly straight forward way for a polymerase to insure the fidelity of the newly synthesized daughter strand and, thus would want the double-helix to adopt the standard B-form with its wide and accessible major groove. However, DNA polymerases are not sequence specific (i.e., they will synthesize from any template sequence), so the enzyme must distinguish a proper Watson-Crick base pair from various mismatches without knowing what the base pair should be. The characteristic feature of mismatched bases (as in a wobble) is that the structure of the minor groove becomes perturbed (Kool, 2001); thus, by inducing the A-form, the polymerase exploits the structural features of the highly accessible minor-groove to insure that the correct base has been added relative to the template sequence.

4.3 Z-DNA: A left-handed duplex

Z-form DNA is noteworthy as the only characterized left-handed form of the double-helix. The zig-zagged backbone, its namesake, results from the alternation between syn- and anti-conformations, and the respective C3'-endo and C2'-endo sugar puckers. This alternating conformation imposes a sequence preference for alternating purine-pyrimidines, since purines adopt the syn-conformation more readily than do pyrimidines. Thus, the repeating unit is the dinucleotide rather than a single base pair, as in B-DNA. The major groove in Z-DNA is not so much a groove but more a convex outer surface, while the minor groove becomes a deep, narrow and largely inaccessible crevice (Wang et al., 1979).

The biological function of Z-DNA has been widely debated and underappreciated; however, several cellular functions for the Z-form are now supported by experimental evidence (Rich and Zhang, 2003). Z-DNA was initially characterized as a structure induced by high salt conditions (3 M NaCl) (Pohl and Jovin, 1972), leading many to wonder whether it could exist in a cell. Subsequently, it has been shown that cytosine methylation, and other cations such as spermine and spermidine at millimolar concentrations also stabilize Z-DNA (Rich and Zhang, 2003). Most importantly, as a left-handed structure, Z-DNA is the most underwound form of the double-helix and, consequently, serves as a sink for the torsional tension in negatively supercoiled DNA (Rich and Zhang, 2003). This expands the range of cellular situations that could support the formation, at least transiently, of Z-DNA. In one model, RNA polymerase, as it transcribes through a gene, would generate negative supercoils in its wake (Liu and Wang, 1987) and, on the process drive Z-DNA formation upstream of the transcribing

gene. A detailed study of the promoter for human CSF-1 gene showed that up-regulation by the chromatin remodeling protein involves a Z-DNA element (Liu et al., 2001). The authors suggested that Z-DNA upstream of the nuclear factor-1 binding site helped to maintain the gene in its activated, nucleosome-free state (nucleosomes do not bind to the very rigid Z-DNA form (Ausio et al., 1987)). In support of its potential role in the regulation of eukaryotic genes, we have found that Z-forming sequences accumulate near the transcription start site of genes in humans and other eukaryotes (Khuu et al., 2007; Schroth et al., 1992), and that ~80% of the genes in human chromosome 22 have at least one Z-DNA sequence in the vicinity of their transcription start sites (Champ et al., 2004).

The discovery of protein domains having very high specificity for Z-DNA (Rich and Zhang, 2003), in some cases with nanomolar K_D's, have suggested additional functions that include, for example, RNA editing and gene transactivation. Z-DNA sequences have also been implicated in genomic instability, that results in large scale breaks and rearrangements (Kha et al., 2010). Thus, in addition to serving as a sink for superhelical tension, there are several potential functions for Z-DNA that may be either beneficial or deleterious to the cell.

4.4 H-DNA: Three's a crowd

When a single DNA strand invades the major groove of a DNA duplex, a triple helical structure is generated (Fig. 2.9). In order for the duplex to accommodate this third strand, it must unwind to broaden the major groove; thus, such triple-stranded helices are favored in negatively supercoiled DNA (Mirkin, 2008). The invading third strand can be intermolecular or intramolecular.

The interaction between strands involve the Hoogsteen edge of the Watson-Crick base pairs (Fig. 2.3) of the duplex to form base triplets, leading to the name H-DNA for such triplex structures. H-DNA is formed primarily in mirror repeat sequences (sequences that have dyad symmetry within a strand, as in …AGAGGGnnnGGGAGA…, invoked by the sequence preference to form base triplets). Mirror-repeats occur randomly in prokaryotes, but are three to six times more frequent in eukaryotic genomes (Schroth and Ho, 1995). Specific H-DNA forming sequences have been identified in multiple promoter regions with documented effects on gene expression of several disease related genes, c-myc (Kinniburgh, 1989) and c-Ki-ras (Pestov et al., 1991). As with Z-DNA, the repeating sequence motif of H-DNA appears to be a source of genetic instability resulting from double-strand breaks. Wang and Vasquez (2004) reported a ~20 fold increase in mutation frequency upon incorporation of an H-DNA forming sequence found in the c-myc promoter region into mammalian cells. These results suggest that naturally occurring DNA sequences can cause increased mutagenesis via non-standard DNA structure formation.

4.5 HJ, G, and I: The four-stranded DNAs

There are several conformations of DNA that can be assembled from four strands. The three structures discussed here show very different and unique helical forms, starting with a conformation that is most similar to standard B-DNA, and leading through forms that differ dramatically from the original Watson-Crick model (Fig. 2.9).

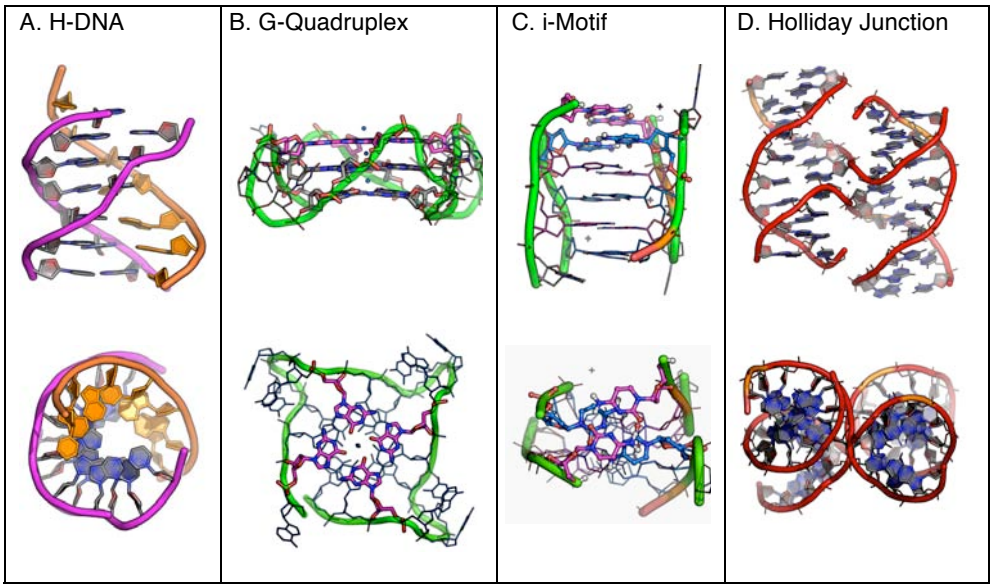


Fig. 2.9. Three- and four-stranded structures of DNA. The structures of triplex H-DNA (Radhakrishnan and Patel, 1993), the Holliday junction (Eichman et al., 2002), human telomeric G-quartet (Parkinson et al., 2002), and the i-motif (Weil et al., 1999), are viewed along (top) and down (bottom) their helical axes.

4.5.1 The four-stranded Holliday junction

Robin Holliday proposed in 1964 that a four-stranded junction would be involved as an intermediate to allow reciprocal exchange of genetic information through recombination across two homologous DNA duplexes (Holliday, 1964). These intermediates, now referred to as Holliday junctions, are essential to several cellular processes including recombination dependent DNA lesion repair, viral integration, restarting of stalled replication forks, and proper segregation of homologous chromosomes during meiosis (Cox et al., 2000; Declais et al., 2003; Dickman et al., 2002; Haber and Heyer, 2001; Nunes-Duby et al., 1987; Subramaniam et al., 2003). The structure of the Holliday junction has been the focus of intense biophysical studies for several decades (Lilley, 1999). Through a set of clever studies in which immobilized junctions are specifically cut by restriction enzymes or probed with fluorescent dyes, DNA junctions were shown to adopt either an extended open-X form under low-salt conditions or a more compact stacked-X conformation as the negatively charged phosphate backbone becomes shielded under high-salt conditions. In the stacked-X form, two continuous DNA strands are connected by two crossover strands, each forming a tight U-turn at the cross-over point, which restricts the migration of the junction. Single molecule studies have shown that junction migration requires a transition to the open-X structure (McKinney et al., 2003), and that this is fairly rapid. As a result, enzymes that catalyze cellular processes that require junction migration (for example, during recombination dependent DNA repair by the RuvABC complex (Dickman et al., 2002)) will recognize and bind the extended and topologically unrestrained open-X structure, while those that do not require junction migration (such as many resolving enzymes in

recombination, including the resolvases from T4 and T7 (Biertumpfel et al., 2007; Hadden et al., 2007)) have active sites that bind to the topologically restrained stacked-X type structure.

Around the end of the 20th century, two groups almost simultaneously solved the single-crystal structures of the DNA Holliday junction (Ortiz-Lombardía et al., 1999; Eichman et al., 2000). Both structures strongly resembled the model derived from the solution studies (McKinney et al., 2003), showing the junction to be essentially two B-DNA double-helices, with standard Watson-Crick type base pairs, linked by two crossing strands that link the duplexes. A unique set of hydrogen bonds helps to stabilize the tight U-turns at the cross-over points (Eichman et al., 2002), and impose a strong sequence dependence in the formation of Holliday junctions, with the inverted repeats GGTACC > GGCGCC > (GATATC = GGGCCC) in their ability to stabilize four-stranded stacked-X junctions (Hays et al., 2005). In addition, the interactions define an ~40° angle relating the two linked duplexes—the structure of an asymmetric junction showed no interactions at the junction center, and an interduplex angle of ~60° (Khuu and Ho, 2009), similar to that determined in solution for analogous constructs (McKinney et al., 2003). The structure of the junction has now been determined with the drug psoralen (Eichman et al., 2001), methylated cytosines (Vargason and Ho, 2002), and various types of cations (Thorpe et al., 2003), all showing effects on the detailed geometry of this four-stranded intermediate (Watson et al., 2004). The effect of sequence on the formation and geometry of junctions lead to a model in which even non-sequence specific resolvases may show sequence preference, not as a result of any specific recognition motif between the protein

and the DNA, but from the thermodynamic propensity of certain sequences to promote formation of the junction (Khuu, 2006).

In replication, Holliday junctions are essential intermediates in double-strand break repair (Cox et al., 2000) in which RecA facilitates invasion of a single-strand into a homologous double-strand sequence, followed by junction migration and resolution by RuvABC (RecG). Homologous recombination also plays a crucial role in rescuing replication forks that stall because of DNA damage. Recombination proteins repair double-strand ends produced when a replication fork encounters a single-strand interruption and help reset replication at stalled forks by converting blocked replication forks into Holliday junctions. Thus, DNA junctions are involved in the repair of damaged DNAs both during and after replication.

4.5.2 G-Quadruplexes

The four-stranded structures assembled from guanine-rich sequences are called G-quadruplexes or G-quartets. Such sequences are found primarily in telomeric DNA repeats (3'-overhangs at chromosome ends (Patel et al., 2007)), but have recently been identified in various other central regions of the genome, including centrometric sequences (Brooks et al., 2010) and in the immunoglobulin switch region. The strands are held together by pairing the Watson-Crick edge of each guanine with the Hoogsteen edge of an adjacent guanine, creating a cyclic arrangement of four guanines into G-tetrads. These tetrads are stacked with a right-handed helical twist, and are stabilized by monovalent cations (Na^+ or K^+) coordinated to the O2 oxygens of the guanines, and sandwiched between the base stacks.

G-quartets can be formed from the association of one, two, or four G-rich DNA strands with various topologies (Mirkin, 2008). Of these, the topologies that can be adopted by single-strands are perhaps most important for G-rich sequences at the 3'-ends (telomeric ends) of chromosomes (characterized as a single–stranded overhang of a guanine-rich sequence that assembles into a nucleo-protein structure). Such sequences have been shown to form G-quadruplex structures, from the DNA in the marconucleus of a ciliate (Mergny et al., 2002) to the exceptionally stable G-quartet formed under physiological conditions by the human telomeric repeats ($(GGGTTA)_3GGG$) (Parkinson et al., 2002). The telomer ends are replicated through the reverse transcriptase function of telomerase, which is itself a protein-RNA complex (Zakian, 2009). The precise length of each telomere controls the cell's ability to replicate, suggesting a regulatory role for their G-quadruplex structures. In normal cells, the length of the telomeric region is reduced during each round of replication until the Hayflick limit is reached, at which point the cell enters apoptosis (Zakian, 2009). The misregulation of telomerase activity can lead to immortality of cells and associated tumorogensis.

Although it is easy to envision formation of a G-quartet structure at the single-stranded end of a chromosome, G-rich repeating sequences with the potential ability to form G-quadruplexes have also been identified at internal sites within genomes (Brooks et al., 2010). Indeed, a recent study by Sarkies, et al. (Sarkies et al., 2010) indicates that the specialized DNA polymerase Rev 1 is involved in replication through G-rich sequences and, when the polymerase is absent, DNA replication and histone recycling becomes uncoupled, leading to the assembly of nucleosomes with newly synthesized histones and, consequently, loss of epigenetic makers at or near these sites. Thus, internal

G-quadruplex sequences are crucial important in the replication of genetic information beyond that of the linear sequence.

4.5.3 I-motifs

In order for a double-stranded G-rich region to extrude into a G-quartet structure, the complementary C-rich strand must also be extruded. The structure that is now associated with C-rich sequences is the four-stranded, intercalated i-motif. The i-motif, or I-form DNA, is fashioned from two parallel C-strands intercalated in a head-to-tail fashion [(Mills et al., 2002). The two duplexes of poly(dC) are stabilized by base pairing the Watson-Crick edges of two cytosines to form hemi-protonated C•C+ pairs.

5. Getting from here to there: Structural transitions in DNA

B-DNA is recognized as the "standard" form in the cell; however, if everything remains standard and static, then life would not be as rich, nor might it exist at all. DNA is thus not only polymorphic, it is also dynamic. In this section, we will explore the mechanisms that drive DNA from the norm as B DNA, focusing on two transitions that present interesting and important insights into how DNA transforms between structural forms.

5.1 Going from B to A

As we have seen, A-type DNA plays an important role in replication as the induced form in the active site of DNA polymerase, allowing the non-sequence specific recognition of base mispairs in the template/daughter duplex. The transition from B- to

A-DNA was one of the earliest characterized, with dehydration of DNA fibers showing a distinct shortening in the helical rise, unwinding of the helical twist, and broadening in the diameter (Franklin and Gosling, 1953a). The transition is also induced in solution by alcohol (a dehydrant), as well as methylation of cytosines (which affects the water structure around the base pairs). The question is, what are the structural and energetic steps involved in this transition? Although this is basically a transition from one right-handed antiparallel double-helix to another, several dramatic structural rearrangments must take place, including a conversion of the sugar pucker, along with large sliding and inclination of base pairs. The details of this conformational shift were observed crystallogaphically at the atomic level on the short DNA sequence GGCGCC (Vargason et al., 2001), which was primarily in the B-form, but, upon cytosine methylation or bromination, adopts a number of conformational states, including true A-DNA forms and a set of logical intermediates between the B- and A-forms (Fig. 2.10). This study generates a structural map for how the sugar conformation works its way around the ring, the order of translational and rotational distortions to the stacked base pairs, and the direction of propagation of a structural transition once initiated.

The transition involves conversion of the sugar from the B-DNA C2'-endo pucker to C1'-exo, then O4'-endo, followed by C4'-exo, and finally to the C3'-endo pucker of A-DNA (Fig. 2.5) (Vargason et al., 2001). Applying ab initio calculations on models of the deoxyribose derived from this study, we found that there is an ~4 kcal/mol energy barrier (primarily bonding energy) at the O4'-endo intermediate step. This is lower than the ~5-6 kcal/mol estimated for planar intermediates required for a direct conversion from C2'- to C3'-endo, and is similar to estimates from experimental (Olson and

Sussman, 1982) and other ab initio calculations (Foloppe et al., 2001) on the barrier (although about 2-fold higher than molecular dynamics estimates (Arora and Schlick, 2003; Harvey and Prabhakaran, 1986)).

Associated with the changes in sugar pucker are perturbations to the base stacking. As the sugars go through a transition from B- towards A-type sugars, the B-A chimeric intermediate (which is half B- and half A-type along each strand) induces a large buckle in the base pairs at the point of transition, which partially unstacks one of the two bases of the pair. The unstacking becomes complete when the sugars assume the full A-type pucker, resulting in an ~10% extension of the spacing between bases, or a rise of ~3.7 Å (Vargason et al., 2000), thereby allowing the large slide and subsequent displacement of the base pairs away from the helical axis that is characteristic of A-DNA. Thus, large shifts between base pairs are predicated on breaking the base stacking interactions, as one would expect. In addition, it shows the transition to A-DNA propagating back towards the 5'-end of each strand. The tilt and roll that causes the inclination and resulting shortened rise of A-DNA are the final steps. The B- to A-DNA transition is unique in that specific intermediates have been trapped to provide an atomic level map for the transition—this is perhaps the most detailed description of a complete structural transition of any biological macromolecule.

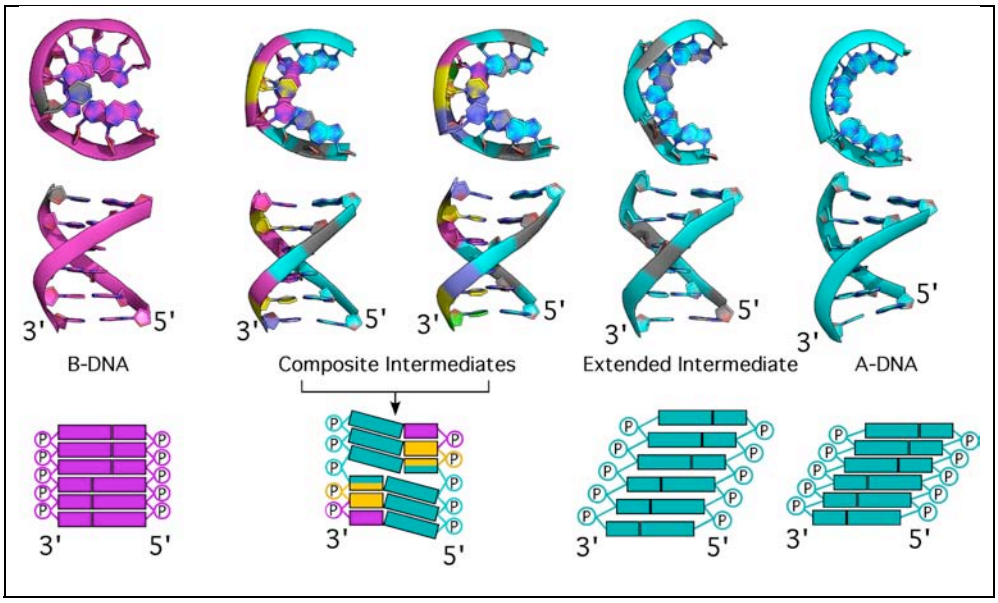


Fig. 2.10. B- to A-DNA transition. The structures of GGCGCC and methylated or brominated variants viewed down (top) and along (bottom) the helix axis. The series of structures show a transition from B-DNA, through a chimeric A-B intermediate and an extended intermediate, and leading finally to A-DNA. Nucleotides are colored according to their sugar puckers, as presented in Fig. 2.5.

5.2 Switching hands: The B- to Z-DNA transition

A more dramatic transition is from the right-handed B- to left-handed Z-DNA (Fig. 2.11), which has been studied extensively in solution and in plasmids. The B-Z transition, however, is not simply taking a right-handed double-helix and twisting it in the opposite direction. The sugar for alternating nucleotides along a strand change from C2'-endo to C3'-endo puckers, concommitant with rotation of the base from the anti- to the syn-conformations. More significantly, the "sense" of the duplex must change—i.e., the direction of the major and minor grooves are swapped (Dickerson, 1992).

In order to accommodate all of these radical changes, there is a junction with an overall zero twist (the B-Z junction) that serves to splice the right- and left-handed twisted duplexes (Peck and Wang, 1983). The structure of this junction was determined in a clever way using a Z-DNA binding protein to stabilize half the DNA in the left-handed form, while allowing the other half to remain in its relaxed B-form (Ha et al., 2005). The structure shows that the bases at the B-Z junction itself have flipped out, which would allow for transition of the sugar pucker and rotation of the bases. It also allows the bases, when they pair again, to change the direction of the grooves, while maintaining stacking between the left- and right-handed columns. The B-Z transition, therefore, can be thought of as initiating with a melting of two base pairs (two B-Z junctions, with a nucleation energy of ~10 kcal/mol (Peck and Wang, 1983)), with each junction subsequently migrating in opposite directions to allow the propagation of the left-handed DNA between them (the propagation energy per base pair being sequence dependent and lowest in alternating GC dinucleotides (Ellison et al., 1985)).

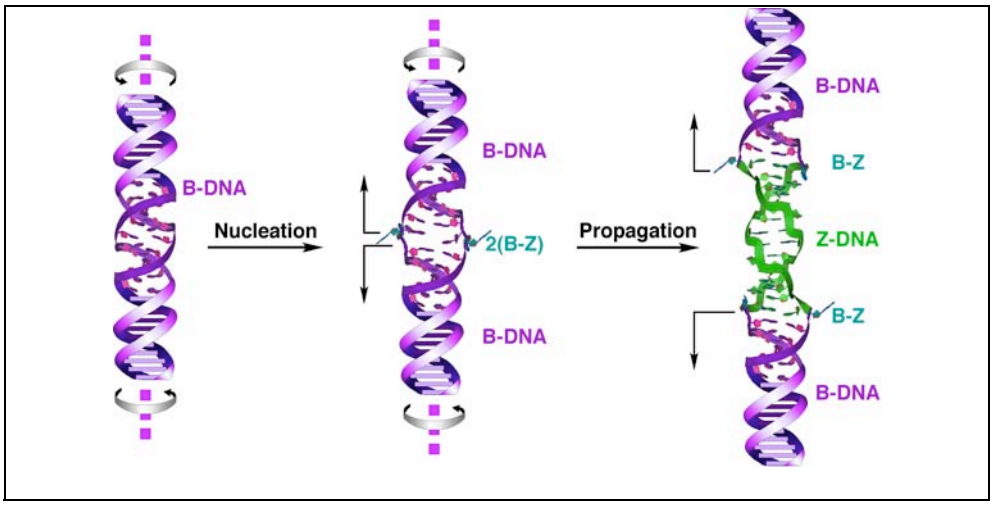


Fig. 2.11. B- to Z-DNA transition. B-DNA, when unwound by negative supercoiling, will first extrude two flipped out base pairs (serving as two B-Z junctions). Further unwinding results in the formation of left-handed Z-DNA as the two junctions migrate in opposite directions.

6. Conclusion

In this review, we have discussed a plethora of structures that come from physical biochemical studies, and show how these structures are defined by sequence and how they transform. Through its history, there has always been a nagging question of "Is this structure relevant?" Clearly, the B-DNA double-helix is relevant, not only to replication, but also to nearly all genetic processes. However, a clearer understanding for the biological roles of the non-B-type DNAs will require a detailed mapping of such structures (Ho, 2009), either experimentally or computationally, across genomes from various organisms.

7. References

Allemand, J.F., Bensimon, D., Lavery, R., & Croquette, V. (1998). Stretched and overwound DNA forms a Pauling-like structure with exposed bases. Proc Natl Acad Sci, USA 95: 14152-14157.

Arora, K., & Schlick, T. (2003). Deoxyadenosine sugar puckering pathway simulated by the stochastic difference equation algorithm. Chem Phys Lett 378: 1-8.

Ausio, J., Zhou, G., & van Holde, K. (1987). A reexamination of the reported B----Z DNA transition in nucleosomes reconstituted with poly(dG-m5dC).poly(dG-m5dC). Biochemistry 26: 5595-5599.

Avery, O.T., Macleod, C.M., & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III. J Exp Med 79: 137-158.

Biertumpfel, C., Yang, W., & Suck, D. (2007). Crystal structure of T4 endonuclease VII resolving a Holliday junction. Nature 449: 616-620.

Brooks, T.A., Kendrick, S., & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. FEBS J 277: 3459-3469.

Burley, S.K. (1996). X-ray crystallographic studies of eukaryotic transcription initiation factors. Philos Trans R Soc Lond B Biol Sci 351: 483-489.

Champ, P.C., Maurice, S., Vargason, J.M., Camp, T., & Ho, P.S. (2004). Distributions of Z-DNA and nuclear factor I in human chromosome 22: a model for coupled transcriptional regulation. Nucleic Acids Res 32: 6501-6510.

Chargaff, E. (1950). Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia 6: 201-209.

Coll, M., Frederick, C.A., Wang, A.H., & Rich, A. (1987). A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. Proc Natl Acad Sci U S A 84: 8385-8389.

Cox, M.M., Goodman, M.F., Kreuzer, K.N., Sherratt, D.J., Sandler, S.J., & Marians, K.J. (2000). The importance of repairing stalled replication forks. Nature 404: 37-41.

Day, R.O., Seeman, N.C., Rosenberg, J.M., & Rich, A. (1973). A crystalline fragment of the double helix: the structure of the dinucleoside phosphate guanylyl-3',5'-cytidine. Proc Natl Acad Sci U S A 70: 849-853.

Declais, A.C., Fogg, J.M., Freeman, A.D., Coste, F., Hadden, J.M., Phillips, S.E., & Lilley, D.M. (2003). The complex between a four-way DNA junction and T7 endonuclease I. EMBO J 22: 1398-1409.

Dickerson, R.E. (1992). DNA structure from A to Z. Methods Enzymol 211: 67-111.

Dickerson, R.E. (1999). Helix structure and molecular recognition by B-DNA. In Oxford Handbook of Nucleic Acid Structure, S. Neidle, ed. (New York, Oxford University Press), pp. 145-197.

Dickerson, R.E., Goodsell, D.S., & Neidle, S. (1994). "...the tyranny of the lattice...". Proc Natl Acad Sci U S A 91: 3579-3583.

Dickman, M.J., Ingleston, S.M., Sedelnikova, S.E., Rafferty, J.B., Lloyd, R.G., Grasby, J.A., & Hornby, D.P. (2002). The RuvABC resolvasome. Eur J Biochem 269: 5492-5501.

Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R.E. (1981). Structure of a B-DNA dodecamer: conformation and dynamics. Proc Natl Acad Sci USA 78: 2179-2183.

Eichman, B.F., Mooers, B.H.M., Alberti, M., Hearst, J.E., & Ho, P.S. (2001). The crystal structures of psoralen cross-linked DNAs: drug dependent formation of Holliday junctions. J Mol Biol 301: 15-26.

Eichman, B.F., Ortiz-Lombardia, M., Aymami, J., Coll, M., & Ho, P.S. (2002). The inherent properties of DNA four-way junctions: comparing the crystal structures of holliday junctions. J Mol Biol 320: 1037-1051.

Eichman, B.F., Vargason, J.M., Mooers, B.H.M., & Ho, P.S. (2000). The Holliday junction in an inverted repeat sequence: sequence effects on the structure of four-way junctions. Proc Natl Acad Sci, USA 97: 3971-3976.

Ellison, M.J., Kelleher, R.J., 3rd, Wang, A.H., Habener, J.F., & Rich, A. (1985). Sequence-dependent energetics of the B-Z transition in supercoiled DNA containing nonalternating purine-pyrimidine sequences. Proc Natl Acad Sci U S A 82: 8320-8324.

Foloppe, N., Nilsson, L., & MacKerell, A.D., Jr. (2001). Ab initio conformational analysis of nucleic acid components: intrinsic energetic contributions to nucleic acid structure and dynamics. Biopolymers 61: 61-76.

Franklin, R.E., & Gosling, R.G. (1953a). Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. Nature 172: 156-157.

Franklin, R.E., & Gosling, R.G. (1953b). Molecular Configuration in Sodium Thymonucleate. Nature (London) 171: 740-741.

Ghosh, A., & Bansal, M. (2003). A glossary of DNA structures from A to Z. Acta Crystallogr D Biol Crystallogr 59: 620-626.

Greenbaum, J.A., Pang, B., & Tullius, T.D. (2007). Construction of a genome-scale structural map at single-nucleotide resolution. Genome Res 17: 947-953.

Ha, S.C., Lowenhaupt, K., Rich, A., Kim, Y.G., & Kim, K.K. (2005). Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. Nature 437: 1183-1186.

Haber, J.E., & Heyer, W.D. (2001). The fuss about Mus81. Cell 107: 551-554.

Hadden, J.M., Declais, A.C., Carr, S.B., Lilley, D.M., & Phillips, S.E. (2007). The structural basis of Holliday junction resolution by T7 endonuclease I. Nature 449: 621-624.

Hamelberg, D., Williams, L.D., & Wilson, W.D. (2001). Influence of the dynamic positions of cations on the structure of the DNA minor groove: sequence-dependent effects. J Am Chem Soc 123: 7745-7755.

Harvey, S.C., & Prabhakaran, M. (1986). Ribose Puckering: Structure, Dynamics, Energetics, and the Pseudorotation Cycle. J Amer Chem Soc 108: 6128-6136

Hays, F.A., Teegarden, A., Jones, Z.J., Harms, M., Raup, D., Watson, J., Cavaliere, E., & Ho, P.S. (2005). How sequence defines structure: a crystallographic map of DNA structure and conformation. Proc Natl Acad Sci, USA 102: 7157-7162.

Hendrickson, W.A., Smith, J.L., Phizackerley, R.P., & Merritt, E.A. (1988). Crystallographic structure analysis of lamprey hemoglobin from anomalous dispersion of synchrotron radiation. Proteins 4: 77-88.

Hershey, A.D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol 36: 39-56.

Ho, P.S. (2009). Thermogenomics: thermodynamic-based approaches to genomic analyses of DNA structure. Methods 47: 159-167.

Ho, P.S., Frederick, C.A., Quigley, G.J., van der Marel, G.A., van Boom, J.H., Wang, A.H., & Rich, A. (1985). G.T wobble base-pairing in Z-DNA at 1.0 A atomic resolution: the crystal structure of d(CGCGTG). EMBO J 4: 3617-3623.

Holliday, R. (1964). A mechanism for gene conversion in fungi. Genet Res 5: 282-304.

Kha, D.T., Wang, G., Natrajan, N., Harrison, L., & Vasquez, K.M. Pathways for double-strand break repair in genetically unstable Z-DNA-forming sequences. J Mol Biol 398: 471-480.

Khuu, P., & Ho, P.S. (2009). A rare nucleotide tautomer in the structure of an asymmetric DNA junction. Biochemistry 48: 7824-7832.

Khuu, P., Sandor, M., DeYoung, J., & Ho, P.S. (2007). Phylogenomic analysis of the emergence of GC-rich transcription elements. Proc Natl Acad Sci U S A 104: 16528-16533.

Khuu, P.A., Voth, A. R., Hays, F. A. and Ho, P. S. (2006). The stacked-X DNA Holliday junction and protein recognition. J Mol Recognit 19: 1-9.

Kiefer, J.R., Mao, C., Braman, J.C., & Beese, L.S. (1998). Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. Nature 391: 304-307.

Kinniburgh, A.J. (1989). A cis-acting transcription element of the c-myc gene can assume an H-DNA conformation. Nucleic Acids Res 17: 7771-7778.

Klotz, I.M. & Franzen, J.S. (1962). Hydrogen bonds between model peptide groups in solution. J Am Chem Soc 84: 3461-3466.

Kool, E.T. (2001). Hydrogen bonding, base stacking, and steric effects in dna replication. Annu Rev Biophys Biomol Struct 30: 1-22.

Liepinsh, E., Otting, G., & Wüthrich, K. (1992). NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution. Nucleic Acids Res 20: 6549-6553.

Lilley, D.M.J. (1999). Structures and interactions of helical junctions in nucleic acids. In Oxford Handbook of Nucleic Acid Structure, S. Neidle, ed. (New York, Oxford University Press), pp. 471-498.

Liu, L.F., & Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proc Natl Acad Sci U S A 84: 7024-7027.

Liu, R., Liu, H., Chen, X., Kirby, M., Brown, P.O., & Zhao, K. (2001). Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. Cell 106: 309-318.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., & Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389: 251-260.

Manning, G.S. (1977). Limiting laws and counterion condensation in polyelectrolyte solutions. IV. The approach to the limit and the extraordinary stability of the charge fraction. Biophys Chem 7: 95-102.

McKinney, S.A., Declais, A.C., Lilley, D.M., & Ha, T. (2003). Structural dynamics of individual Holliday junctions. Nat Struct Biol 10: 93-97.
Mergny, J.L., Riou, J.F., Mailliet, P., Teulade-Fichou, M.P., & Gilson, E. (2002). Natural and pharmacological regulation of telomerase. Nucleic Acids Res 30: 839-865.
Meselson, M., & Stahl, F.W. (1958). The replication of DNA in Escherichia Coli. Proc Natl Acad Sci U S A 44: 671-682.
Mills, M., Lacroix, L., Arimondo, P.B., Leroy, J.L., Francois, J.C., Klump, H., & Mergny, J.L. (2002). Unusual DNA conformations: implications for telomeres. Curr Med Chem Anticancer Agents 2: 627-644.
Mirkin, S.M. (2008). Discovery of alternative DNA structures: a heroic decade (1979-1989). Front Biosci 13: 1064-1071.
Neidle, S. (1999). Oxford Handbook of Nucleic Acid Structure (Oxford, Oxford Press).
Ng, H.L., Kopka, M.L., & Dickerson, R.E. (2000). The structure of a stable intermediate in the A <--> B DNA helix transition. Proc Natl Acad Sci U S A 97: 2035-2039.
Nunes-Duby, S.E., Matsumoto, L., & Landy, A. (1987). Site-specific recombination intermediates trapped with suicide substrates. Cell 50: 779-788.
Olson, W.K., & Sussman, J.L. (1982). How Flexible Is the Furanose Ring? 1. A Comparison of
Experimental and Theoretical Studies. J Am Chem Soc: 270-278.
Ortiz-Lombardía, M., González, A., Eritja, R., Aymamí, J., Azorín, F., & Coll, M. (1999). Crystal structure of a DNA Holliday junction. Nat Struct Biol 6: 913-917.
Parkinson, G.N., Lee, M.P., & Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. Nature 417: 876-880.
Patel, D.J., Phan, A.T., & Kuryavyi, V. (2007). Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. Nucleic Acids Res 35: 7429-7455.
Pauling, L., & Corey, R.B. (1953). A proposed structure for the nucleic acids. Proc Natl Acad Sci U S A 39: 84-97.
Peck, L.J., & Wang, J.C. (1983). Energetics of B-to-Z transition in DNA. Proc Natl Acad Sci U S A 80: 6206-6210.
Pestov, D.G., Dayn, A., Siyanova, E., George, D.L., & Mirkin, S.M. (1991). H-DNA and Z-DNA in the mouse c-Ki-ras promoter. Nucleic Acids Res 19: 6527-6532.
Pohl, F.M., & Jovin, T.M. (1972). Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG-dC). J Mol Biol 67: 375-396.
Privé, G.G., Yanagi, K., & Dickerson, R.E. (1991). Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. J Mol Biol 217: 177-199.
Radhakrishnan, I., & Patel, D.J. (1993). Solution structure of a purine.purine.pyrimidine DNA triplex containing G.GC and T.AT triples. Structure 1: 135-152.
Rich, A., & Zhang, S. (2003). Timeline: Z-DNA: the long road to biological function. Nat Rev Genet 4: 566-572.
Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., & Honig, B. (2009). The role of DNA shape in protein-DNA recognition. Nature 461: 1248-1253.
Rosenberg, J.M., Seeman, N.C., Kim, J.J., Suddath, F.L., Nicholas, H.B., & Rich, A. (1973). Double helix at atomic resolution. Nature 243: 150-154.

Sarkies, P., Reams, C., Simpson, L.J., & Sale, J.E. Epigenetic instability due to defective replication of structured DNA. Mol Cell 40: 703-713.
Schroth, G.P., Chou, P.J., & Ho, P.S. (1992). Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. J Biol Chem 267: 11846-11855.
Schroth, G.P., & Ho, P.S. (1995). Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. Nucleic Acids Res 23: 1977-1983.
Subramaniam, S., Tewari, A.K., Nunes-Duby, S.E., & Foster, M.P. (2003). Dynamics and DNA substrate recognition by the catalytic domain of lambda integrase. J Mol Biol 329: 423-439.
Thorpe, J.H., Gale, B.C., Teixeira, S.C., & Cardin, C.J. (2003). Conformational and hydration effects of site-selective sodium, calcium and strontium ion binding to the DNA Holliday junction structure d(TCGGTACCGA)(4). J Mol Biol 327: 97-109.
Vargason, J.M., Eichman, B.F., & Ho, P.S. (2000). The extended and eccentric E-DNA structure induced by cytosine methylation or bromination. Nat Struct Biol 7: 758-761.
Vargason, J.M., Henderson, K., & Ho, P.S. (2001). A crystallographic map of the transition from B-DNA to A-DNA. Proc Natl Acad Sci U S A 98: 7265-7270.
Vargason, J.M., & Ho, P.S. (2002). The effect of cytosine methylation on the structure and geometry of the Holliday junction: the structure of d(CCGGTACm5CGG) at 1.5 A resolution. J Biol Chem 277: 21041-21049.
Wang, A.H., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G., & Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. Nature 282: 680-686.
Wang, G., & Vasquez, K.M. (2004). Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. Proc Natl Acad Sci U S A 101: 13448-13453.
Watson, J., Hays, F.A., & Ho, P.S. (2004). Definitions and analysis of DNA Holliday junction geometry. Nucleic Acids Res 32: 3017-3027.
Watson, J.D. (1968). The Double Helix: A Personal Account of the Discovery of the Structure of DNA (Atheneum).
Watson, J.D., & Crick, F.H. (1953a). Genetical implications of the structure of deoxyribonucleic acid. Nature 171: 964-967.
Watson, J.D., & Crick, F.H. (1953b). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171: 737-738.
Weil, J., Min, T., Yang, C., Wang, S., Sutherland, C., Sinha, N., & Kang, C. (1999). Stabilization of the i-motif by intramolecular adenine-adenine-thymine base triple in the structure of d(ACCCT). Acta Crystallogr D Biol Crystallogr 55: 422-429.
Zakian, V.A. (2009). The ends have arrived. Cell 139: 1038-1040.

CHAPTER 3

ASSAYING THE ENERGIES OF BIOLOGICAL HALOGEN BONDS[2]

1. Summary

Bromine halogen bonds (X-bonds) had previously been assayed using DNA junctions in which the X-bonds compete against hydrogen bonds (H-bonds) in defining the conformational form of the junction. Here, we show calorimetrically that the stabilizing effect of the X-bond in solution derives primarily from a negative enthalpy (-5 kcal/mol), but is opposed by a negative entropy (-8 cal/mol·K, equivalent to -2 kcal/mol for TΔS at room temperature), resulting in an overall stabilizing free energy of -3 kcal/mol for the X- vs H-bond. Quantum chemical energies for this X-bond are nearly identical to energies derived from the crystallographic and solution assays, confirming that the stabilizing potentials are primarily reflected in the components of the X-bond. A study in which the bromine is replaced by a methyl group (substituents that are similar in size and hydrophobicity) showed that the solvent and steric effects of burying these substituents in the tight pocket of the junction is nearly equivalent in energy to the competing H-bond. Thus, the stabilization of DNA junctions by a bromine X-bond in

[2] Megan Carter and P. Shing Ho
Department of Biochemistry and Molecular Biology,
Colorado State University, Fort Collins, CO 80523

crystals is reflected in the enthalpic stabilization in solution, and that both energies are direct measures of the X-bonding potential of the bromine in a biomolecular system.

2. Introduction

Halogen bonds (or X-bonds, see review by Politzer (Politzer, Murray et al. 2010) for a detailed description) are weak non-covalent interactions that are analogous in many ways to hydrogen bonds (H-bonds) (Metrangolo and Resnati 2001)—they both involve a Lewis acid, the halogen atom, interacting with an electron-rich Lewis base, resulting in interatomic distances that are closer than the sums of their respective van der Waals radii (Fig. 3.1). In the case of an X-bond, the donor is the electropositive crown of a polarized halogen, which is described by the s-hole model for the distribution of valence electrons in Group VII atoms (Murray, Lane et al. 2007) (Fig. 3.2). In this study, we compare and contrast the stabilizing energies of a bromine X-bond against a standard H-bond in solution to help validate the original crystallographic assay (Voth 2007), and apply the crystallographic assay to assess the contribution of solvent and steric effects on stabilizing effects of bromines in this system.

In biological systems, halogens are most commonly observed in secondary metabolites (over 3500 documented halogenated forms (van Pee and Unversucht 2003)) and small molecule inhibitors (including the antibiotics vancomycin and chloromaphenicol, with nearly half the molecules in current screening libraries being halogenated), while halogenation of proteins and nucleic acids have been associated with acute asthma (Wu 2000).

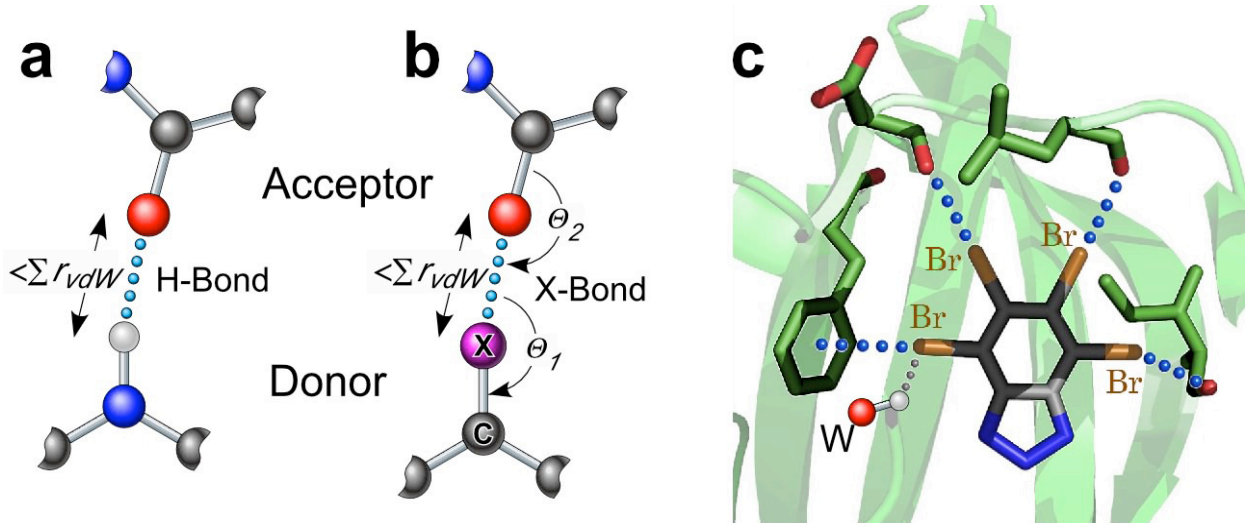


Fig. 3.1. H- and X-bonds. H-bonds (a) and X-bonds (b) are short interactions between an acceptor (A) and donor (D), where the A-D distance is shorter than the sum of their Rvdw (Ouvrard, Le Questel et al. 2003; Auffinger, Hays et al. 2004; Metrangolo 2005). The terms "donor" and "acceptor" refer to the now accepted definitions, in which the donor is the electropositive atom while the acceptor is the electronegative atom in the two interactions (Auffinger, Hays et al. 2004; Metrangolo 2008). c. The inhibitor 4,5,6,7-tetrabromobenzotriazole (De Moliner, Brown et al. 2003) shows 4 X-bonds to phospho-CDK2/cyclin A and one potential H-bond to a water molecule.

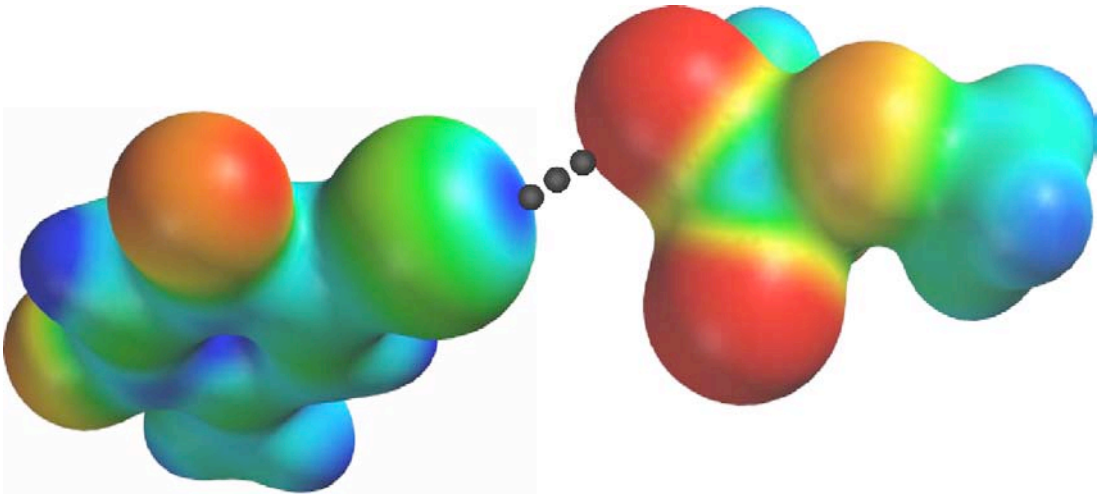


Fig. 3.2. X-bond from a brominated uracil base to phosphate group along an opposing DNA strand in the Br2J junction. Electrostatic potentials were calculated using the MP2 method and rendered with SPARTAN (Wavefunction, Inc., Irvine, CA). Negative electrostatic potentials are shown in red, positive charges in blue, and neutral potentials in yellow. Dots connect the closest interacting atoms of the molecular pair, with the electropositive crown of the bromine representing the σ-hole.

Organic halogens are used in pharmaceuticals, herbicides, fungicides, insecticides, flame-retardants, and intermediates in organic synthesis, which can have toxic and carcinogenic effects and are difficult to degrade in nature (van Pee and Unversucht 2003). Finally, halogens are commonly incorporated as heavy atom substituents to help solve the phasing problem in x-ray crystallography of proteins and nucleic acids.

There is a growing recognition that X-bonds are important in ligand recognition by proteins (Metrangolo 2008; Lu, Shi et al. 2009)—these interactions have been purposely designed to synthesize new inhibitors against the blood-clotting factor Xa (Matter, Nazare et al. 2009) and, recently, a new ATP competitive inhibitor that is highly specific against CDC2-like kinases (Voth 2007; Fedorov, Huber et al. 2011). In addition, we have shown that X-bonds can be engineered to direct the conformation of DNA junctions (Voth 2007), a common scaffold for biomolecular engineering in, for example, DNA origami (Rothemund 2006) and of DNA-based computers (Robinson and Seeman 1987). Thus, X-bonds have great potential as molecular tools for the design and synthesis of new therapeutic agents and biomolecular materials. Unfortunately, not all attempts to engineer X-bonds have been successful, because there are currently no readily accessible computational tools that accurately predict the structure-energy relationships of halogens. For example, halogenated phenylalanine residues were found to be incapable of replacing a functional H-bond in ketosteroid isomerase (Kraut, Churchill et al. 2009), because the halogen was not aligned to form a strong X-bond. In another case, halobenzenes were seen to form X-bonds to the sulfur of a Met residue in an engineered pocket in T4 lysozyme (Liu, Baase et al. 2009), but the interaction energies are weak (~0.6 kcal/mol more stabilizing than simple van der Waals or vdW interactions). Thus, although it is

tempting to include X-bonds in bottom-up, rational design strategies, the efforts have been greatly hampered by the lack of predictive computational tools to accurately model the structure energy relationships of X-bonds in biological macromolecules. To rectify this problem, we must first accurately determine the structure-energy relationship of X-bonds in a biological system. In the current study, we assay the energy of biological halogen bonds using a DNA Holliday junction system, the simplest biomolecular system to date shown to be stabilized by X-bonds.

Four-stranded DNA junctions are ideal for studying X-bonds, because their rigid structures are defined by a small number of specific intramolecular interactions. The structure of four-stranded junctions (Fig. 3.3) has been of interest since it was first proposed by Holliday in 1964 (Holliday 1964). In addition to being the central intermediate in the homologous recombination and recombination dependent cellular events, DNA junctions have also served as the template for the design of artificial crystal lattices and several nanodevices (reviewed by Seeman (Seeman 1999)). The basic structure of DNA junctions under physiological solution conditions (reviewed by Lilley (Lilley 1999; Lilley 2000)) is described as a compact stacked-X form, in which the arms pair and coaxially stack into two nearly continuous double-helices that are related by a 60° angle and interrupted only by the crossing of strands (Fig. 3.3 b & c) (Duckett, Murchie et al. 1988). The strands of the stacked-X junction are aligned antiparallel to each other, thereby forcing the two cross-over strands to form sharp U-turns. How the arms of these asymmetric junctions pair defines different conformational isomers of the stacked-X junction, with the interconversion between isomeric forms being cation dependent (Grainger, Murchie et al. 1998).

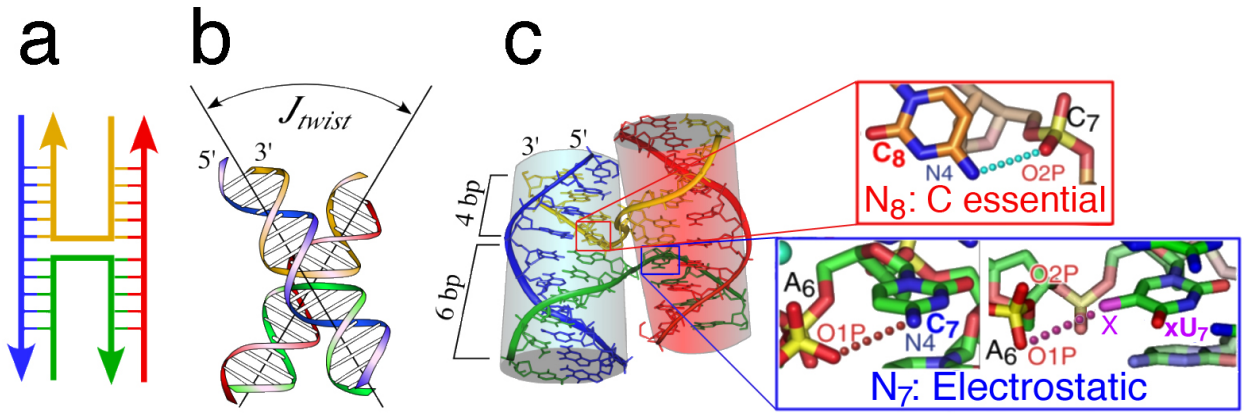


Fig. 3.3. Structure of the DNA Holliday junction. Schematic of the compact stacked-X (a) junction, compared to the solution model (b) (Duckett, Murchie et al. 1988) and single crystal structure of 5'-CCGGTACCGG-3' (c) (Eichman, Vargason et al. 2000). The H-bond from the N4 amino of cytosine C_8 to the phosphate at nucleotide N_7 of the cross-over strand is essential for stabilizing the stacked X-junction (top inset), while a similar interaction at nucleotide N_7 requires only an electrostatic interaction (Hays, Teegarden et al. 2005), including a potential X-bond (Hays, Vargason et al. 2003; Voth 2007).

The crystal structure of the antiparallel stacked-X DNA junction, solved nearly simultaneously by two different laboratories (Fig. 3.3c) (Ortiz-Lombardía, González et al. 1999; Eichman, Vargason et al. 2000), shows the four-stranded complex to be stabilized by a set of H-bonds that link the base pairs of the stacked arms to the phosphate oxygens of the junction cross-over (Fig. 3.3c) (Ho 2001; Hays, Watson et al. 2003). Interestingly, the amino-phosphate H-bonds that stabilize the DNA junctions can be replaced by bromine X-bonds (Hays, Vargason et al. 2003; Auffinger, Hays et al. 2004).

In our earlier work, we exploited the electrostatic interactions in DNA junctions, specifically from the N_7 to the phosphate of the preceding N_6 nucleotide (Fig. 3.3c), to demonstrate that the conformation of the Holliday junction could be directed by an X-bond placed in competition against an H-bond and, in the process, provide an assay to compare stabilizing potential between the two interactions (Voth 2007). In this assay, DNA sequences were designed such that either a cytosine or a halogenated uracil (xU) is at the N_7 position of two complementary strands, with C_8 maintained to insure formation of the junction (Fig. 3.4). The resulting junction can fold to place C_7 in the inside position to H-bond with the phosphate at the crossing strand, leaving the xU_7 at the outside position (H-isomer), or the xU_7 in position to form an X-bond to the phosphate (X-isomer). We designed two constructs, one that competed two bromine X-bonds against two cytosine H-bonds (Br2J) and one that competed one X-bond against 2 H-bonds (Br1J). In both cases, the bromines were seen on the inside position, indicating that the X-bond was more stable than the H-bond, with the shorter X-bond of the Br2J junction being about two-times more stabilizing than the longer interaction of the Br1J construct (Table 3.1).

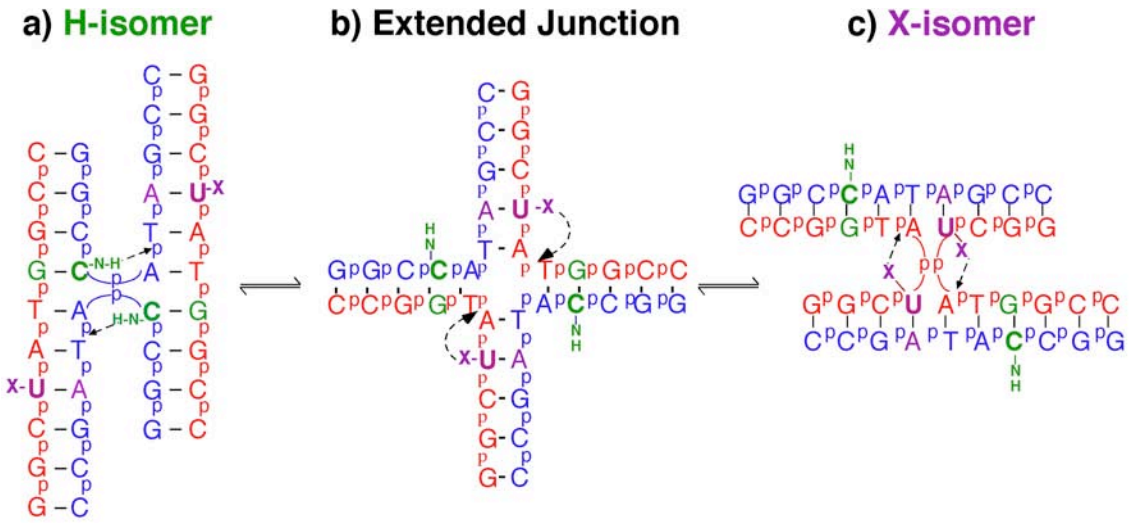


Fig. 3.4. Competition between H- and X-bonds in a DNA junction. Isomeric forms of the stacked-X junction in H-isomer (a) or X-isomer (c), where X is a bromine (Br2J construct). Junction isomerization occurs through an extended intermediate (b) (McKinney, Declais et al. 2003). The isomer form can be distinguished by locating xU, on either the outside strand (H-isomer) or inside strand (X-isomer), and is further confirmed by the identity of the complementary base at either the inside or outside position (if the complimentary base at the inside position is adenine the junction is in the X-isomer, however if the complimentary base is a guanine the junction is in the H-isomer form).

Table 3.1. Geometries and estimated energies of bromine X-bonds vs H-bond in DNA junction from crystallographic assay in which potential halogen bond from a brominated uracil competes against a standard hydrogen bond from a cytosine base to the phosphate backbone in a DNA junction (Voth 2007) (see Fig. 3.4 for details).

Construct	O•••Br Distance	Angle (O•••Br-C)	Energy
Br1J	3.32 Å	163.2°	-2 ± 0.5 kcal/mol
Br2J	2.87 Å	167.2°	-5 kcal/mol

In the current study, we use differential calorimetry (DSC) to determine the enthalpic and entropic contributions to the stabilization free energy of the bromine X-bond vs H-bond in the Br2J junction in solution to validate the results from the original crystallographic assay. In addition, we studied a junction construct in which the bromines are replaced by methyl groups in order to determine the hydrophobic and steric contributions to the crystallographic assay. When coupled with a quantum mechanical model, these results allow us to interpret the various factors that contribute to the stabilizing potential of bromine X-bonds in a model biomolecular system.

3. Theory and Methods

3.1 DNA synthesis and purification

Chemically synthesized DNA oligonucleotides are obtained from Midland Certified Reagent Company on the solid Controlled-Pore Glass (CPG) support with the final dimethoxytrityl (DMT) protecting group attached. Sequences were subsequently purified by reverse phased HPLC followed by size exclusion chromatography on a Sephadex G-25 column after detritylation. The constructs for this study were complementary sequences designed to form four-stranded junctions (Table 3.2).

3.2 Differential scanning calorimetry (DSC) studies on Br2J and H2J junction

Two DNA constructs were designed to form duplexes at low concentrations and four-stranded junctions at high concentrations in order to compare the stabilizing effects of X- vs H-bonds. The complementary DNA sequences for each construct were mixed in equimolar concentrations (concentrations varied from 15 μM to >300 μM) in 50mM

sodium cacodylate and 1mM calcium chloride in order to approximate crystallization conditions, heated to 90°C for one hour and reannealed to room temperature slowly overnight. The energetic parameters for melting the constructs at each concentration were determined by DSC experiments, performed at a constant pressure of ~3.0 atm in a TA Instruments Nano DSC instrument. To obtain a baseline, sample buffer was first analyzed against itself. Each DNA sample was then run against buffer in a heating cycle from 0°C to 90°C at a scanning rate of 1°C/min with an equilibrium time of 900 s. DNA constructs were analyzed at multiple DNA concentrations in order to sample both the duplex and junction conformation, as junction formation has been shown to be concentration dependent (Hays, Schirf et al. 2006). Each experiment was repeated at least three times, although more replicative measurements were performed for the lower concentrations because of the lower signal to noise ratio for each individual run. Data were analyzed using the NanoAnalyze software (TA Instruments, New Castle, DE). The best fit was determined by monitoring the standard deviation of the fit. The data at lower concentrations were best fit using a two state model scaled by a weighting term (Aw), to account for the presence of both double- and single-strand DNAs. Samples in which the Aw term had indicated a much higher than predicted double-stranded concentration suggested the presence of a four-stranded junction component; consequently, the data for these samples were analyzed by applying a two component, two-state model. The similarity in T_m and ΔH_m values for the duplex fractions between the single component analysis of the low DNA concentration data and the two component high DNA concentration data support this interpretation of the analyses. The average ΔH_m for melting the duplex forms of the DNA was, thus, taken as the average of the data for DNA

concentrations from 15 to 20 μ M plus the low temperature component of the two component analysis of data at [DNA] > 100 μM. The ΔH_m of junction was taken from the average of the higher temperature component from the [DNA] > 100 μM analyses. The presence of junctions was evident from the single component analysis of the data from [DNA] from 20 to 100 uM, but did not warrant fitting using the two-component analysis and, therefore, were excluded from the thermodynamic parameters.

3.3 Crystallization and structure solution

Crystallization trials were carried out for the T2J construct by mixing the complementary sequences (Table 3.2) as an equal molar mixture. Initial conditions were searched around the starting condition (0.7mM DNA, 25mM sodium cacodylate pH 7.0 buffer, 10-20mM calcium chloride, and 1.0-1.2mM spermine) that was previously shown to yield crystals of the analogous brominated constructs Br1J and Br2J (Voth 2007). A single crystal of approximately 200 μ in size was selected and frozen in liquid nitrogen for data collection. Diffraction data at liquid nitrogen temperatures was collected at Advanced Light Source (ALS) at the Lawrence Berkeley Laboratories at λ = 0.9 Å and processed using the HKL2000 software (Otwinowski and Minor 1997). The crystal unit cell and space group was isomorphous with those of previous DNA junctions (Table 3.3), indicating that T2J was indeed a junction. Molecular replacement was performed using EPMR with the junction structure of 5'-CCGGTA(BrU)CGG-3'/ 5'-CCGATACCGG-3', [Protein Data Bank ID code 2ORG] as the search model, where the asymmetric unit consisted of one outside and one inside crossing strand of the four-stranded junction. This resulted in an initial model with R = 0.3021, R_{free} = 0.3504, and a correlation of 0.822.

Table 3.2. DNA sequences of constructs to study molecular interactions in junctions. BrU is the nucleotide 5-bromouracil.

Construct name	Complementary Sequences
H2J	5'-CCGGTAUCGG-3'/5'-CCGATACCGG-3'
Br2J	5'-CCGGTA(BrU)CGG-3'/5'-CCGATACCGG-3'
T2J	5'-CCGGTATCGG-3'/5'-CCGATACCGG-3'

Table 3.3 Crystallographic parameters for construct T2J.

Crystallographic parameters	
Space Group	C2
Unit Cell	
a, Å	65.37
b, Å	24.61
c, Å	37.42
β-angle	110.78°
Unique reflections (for refinement)	5940
Resolution, Å	40 - 1.7
Completeness, % (highest resolution shell)	92.3 (97.7)
I/sig, I*	30.23 (5.18)
R_{merge}, %*	0.071 (0.317)
Refinement Statistics	
R_{cryst}, (R_{free}), %	23.92 (29.70)
No. of atoms: DNA (solvent)	808 (129)
<B-factor> DNA (solvent)	9.87 (19.48)
RMSD bond length, Å	1.34
RMSD bond angle	1.91°
PDB ID code	3TOK

*Values for the highest resolution shell are shown in parentheses

Structure factors and atomic coordinates have been deposited in the Protein Data Bank.

Refinement was carried out in CNS (Brünger, Adams et al. 1998) with rigid body refinement, simulated annealing, several rounds of positional and individual B-factor refinement, and addition of solvent. The crucial base pairs, those whose identity would determine isomer form, were left ambiguous during refinement. During multiple rounds of refinement, it became clear that the structure was composed of two different models with a near equal contributions to the electron density. A dual model was constructed by generating symmetry-related coordinates from the two strands of the asymmetric unit, with the atoms in each strand set to half occupancy. This four-stranded complex was further refined with the van der Waals interactions for crystal packing turned off in order to prevent unintended clashes.

Once the refinement of this initial model converged, we set out to determine the occupancy of each model to the overall structure using an occupancy titration. In this method, the occupancies of the 5-methyl group at nucleotide N_7 of the inside strand or the analogous nucleotide N17 at the outside strand, in conjunction with the N2 nitrogen of the complimentary bases (N14 complimentary to N7 or N4 complimentary to N17) were varied from 0 to 100%, and the crystallographic R and R_{free} monitored as a function of the occupancy. Mock occupancy titrations, in which the starting occupancies were not changed were performed in order to determine a background R and R_{free} change associated with the progressive refinement of a structure. This served as the baseline, which was subsequently subtracted from the experimental R and R_{free} titration to yield the final titration curve. Each final curves was fitted by a polynomial equation, and the analytical minimum for each curve determined from the first derivative of these equations, with the minima defining the percent contribution of the two competing

models to the overall structure. As controls, increments of 5-methyl group and N2 nitrogen were similarly added to bases that should have no density for these groups, C2-G19 (outside position) and C18-G19 (inside position). Atomic coordinates and structure factors will be deposited in the PDB (Berman, Westbrook et al. 2000) upon acceptance.

3.4 Quantum chemical calculations on Br1J and Br2J X-bonds

Ab inito calculations were performed with the program GAMESS (Schmidt, M.S. Gordon et al. 1993) by using the WebMO interface (Schmidt 2005) for importing and constructing models. Minimal molecular models of the Br1J and Br2J X-bonding interactions were constructed starting with the atomic coordinates of the published structures (PDB codes 2ORF and 2ORG, respectively) (Voth 2007). These models for the X-bond donor consisted of the bromouracil nucleotide terminated with a hydrogen added to the N1 carbon of the base. The X-bond acceptor was as a hypophosphite anion, constructed from the phosphate group of the interacting DNA strand, and replacing O3' and O5' oxygens with hydrogens. Energies were calculated using the Møller-Plesset perturbation theory (MP2) calculations with a 6-31G(d)) basis set and in cyclohexane solvent (to estimate the dielectric of the junction interior environment). The X-bonding energies of interaction are calculated as the MP2 energies of the bromouracil-hypophosphite complex minus the sum of the energies for the individual components (each corrected for basis set superposition errors).

4. Results and Discussion

4.1 Stabilizing energy of bromine X-bonds in solution.

The crystallographic assay on bromine X-bonds showed a definitive effect of geometry on the stabilizing potential of the interaction relative to a standard H-bond (Table 3.1), with a shorter interaction being more favorable than one that is longer and approaching the sums of the standard van der Waals radii of the donor/acceptor pair (Voth 2007), as one would expect. The primary assumption made in this assay was that the distribution of conformational isomers (X-isomer/H-isomer for the X-bonded vs H-bonded conformers, respectively) in the crystal samples the population in solution. In addition, we made the assumption that the two isomeric forms have near identical conformational entropies in the crystal and, therefore, the observed isomer distributions reflect primarily the enthalpic differences (ΔH) between the competing X- and H-bonding interactions. We test these assumptions here by comparing the effect of the bromine X-bond of the Br2J construct to that of the H-bond in a hydrogen bond only construct (H2J, which is identical to the Br2J sequence with the bromines removed) on the stability of analogous four-stranded DNA junctions in solution.

This solution assay takes advantage of the concentration dependent transition of these DNA constructs from duplex to junction, with a mid-point for transition at ~100 to 200 μM of duplex DNA (Hays, Schirf et al. 2006). The study is designed to determine the difference in stabilizing potential of the bromine X-bond to the H-bond by measuring and comparing the energies required to melt the Br2J and H2J constructs from the junction to their single-stranded forms (at DNA concentrations > transition concentration). In order to account for effects of the bromine on the stability of the duplex arms of the junction,

we also measure and compare the melting energies of the two constructs in their duplex forms (DNA concentrations < transition concentration). The energy of the bromine X-bond relative to the H-bond in solution ($\Delta E_{XB\text{-}HB}$) is thus simply the difference in the melting energy at high concentrations for Br2J ($E_{Br2J\text{-}HC}$) and H2J ($E_{H2J\text{-}HC}$) minus the difference in melting energies at low concentrations ($E_{Br2J\text{-}LC}$ and $E_{H2J\text{-}LC}$, respectively) (Eq. 1), with the assumption that the energies of the single-stranded forms are essentially equivalent for these DNA sequences.

$$\Delta E_{XB-HB} = \left(E_{Br2J-HC} - E_{H2J-HC}\right) - \left(E_{Br2J-LC} - E_{H2J-LC}\right) \qquad \text{Eq. 1}$$

For these studies, solutions of the Br2J and H2J constructs are first annealed at the concentration at which they will be studied. The enthalpies required to melt (ΔH_m) and the melting temperatures (T_m) of the constructs to their single-strands are measured at increasing concentrations from 15 μM to >100 μM by differential scanning calorimetry (DSC). At low concentrations (15 – 20 μM) of both the Br2J and H2J junctions, the melting profiles appear to be of a single species, and can be fit with a simple two-state transition model (Fig. 3.S1)—we assign this to a duplex to single-strand DNA transition. As the DNA concentration is increased, the apparent T_m also increases; however, at the high concentrations (>100 μM), the melting profiles can no longer be fit by single two-state models, but are better fit by composites of two separate two-state models, with the lower temperature component having the same T_m and ΔH_m as measured at 15 and 20 μM. We, therefore, assigned this low temperature component to the duplex to single-strand transition. The higher temperature component, which increases with increasing concentration of DNA, was assigned to the four-stranded junction to single-strand transition.

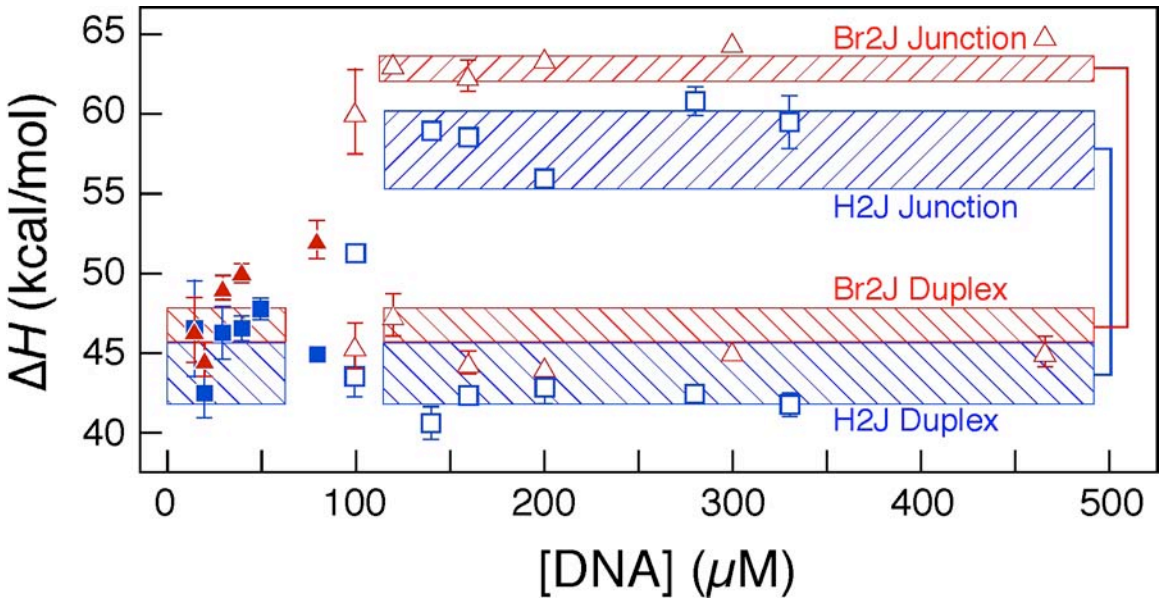


Fig. 3.5. Enthalpies of melting (ΔH_m) measured by differential scanning calorimetry (DSC) results for H2J (blue squares) and Br2J (red triangles) constructs at increasing DNA concentrations. The solid symbols indicate DSC data analyzed with a single component two-state model while open symbols indicate data analyzed by a two-component two-state model. Boxes represent data for the duplex (left hashed boxes) and junction (right hashed boxes) used to calculate the averages and standard deviations of ΔH_m. The height of each box represents the standard deviation for all the data included in the average.

At intermediate concentrations (80 to 100 μM DNA), the DSC scans were fit to a single two-state transition, indicative of primarily duplex melting, but had increasing T_m and ΔH_m values, suggesting some contribution from the junction form; consequently, these data were not included in the subsequent analyses of the thermodynamic parameters. Thus, in this system at high concentrations, the Br2J and H2J DNA melting profiles in solution are convolutions of the the duplex and junction forms, but these profiles can be deconvolved to the contributions of each of the two individual components (Fig. 3.5).

From the DSC studies, the ΔH_m and T_m are measured directly for the Br2J and H2J constructs (Table 3.4). The T_m and ΔH_m observed for the melting of Br2J as a duplex were slightly higher than that of H2J, indicating that the bromine contributes ~ 3kcal/mol to the stabilization of DNA duplex. The difference in melting enthalpy for the junction vs duplex forms of the Br2J construct is ~2 kcal/mol larger than that that for the H2J construct, which indicates that the bromine helps to stabilize the junction. The entropic contributions, calculated as $\Delta S_m = \Delta H_m/T_m$, showed a 4 cal/mol·K higher difference ΔS_m for melting the junction vs duplex of Br2J compared to H2J. In order to directly compare these energies, we extrapolated these values to a common reference temperature (T_{ref}) (Table 3.5), which we defined as the T_m of the H2J duplex (50.6°), using the standard relationships in Eqs. 2 and 3.

$$\Delta H_{m(T_{ref})} = \Delta H_{m(T_m)} + \Delta C_p (T_{ref} - T_m) \quad \text{Eq. 2}$$

$$\Delta S_{m(T_{ref})} = \Delta S_{m(T_m)} + \Delta C_p \ln\left(\frac{T_{ref}}{T_m}\right) \quad \text{Eq. 3}$$

At the T_{ref}, the contribution of the bromine to the overall melting enthalpy ($\Delta\Delta H_m$) was calculated to be 2 kcal/mol and to the overall entropy of melting ($\Delta\Delta S_m$) to be 4.6

cal/mol·K, which, at this temperature results in an overall free energy difference ($\Delta\Delta G_m$) of essentially 0 kcal/mol (Table 3.5). This indicates that at or near the melting temperatures, the population of X- and H-isomers are essentially identical, even though there is an enthalpic contribution to stabilizing the junction by the bromine. This can be interpreted as an example of enthalpy-entropy compensation (Dunitz 1995), where a more stabilizing interaction results in a more rigid structure and, thus, a structure with lower conformational entropy. This also means that the $\Delta\Delta H_m$ and $\Delta\Delta S_m$ values are for a population of junctions that is an equal mix of X- and H-isomers. The actual contribution of the X-bond to the distribution of X- over H-isomer, therefore, should be approximately twice the observed melting energies (-3.6 kcal/mol for $\Delta\Delta H$ and -9.2 cal/mol·K for $\Delta\Delta S$).

If there is enthalpy-entropy compensation, why did the crystallographic assay show predominantly the X-isomer form rather than an equal population? To address this issue, we must consider that crystals are not grown at temperatures near the T_m, but at room temperature or below. In addition, we would again, not expect a significant difference in entropy within the crystal for the two conformers. When extrapolated to 25°C (assuming the heat capacities are relatively temperature independent in this range (Wu, Nakano et al. 2002)), we see that the difference in enthalpy for the X- vs H-bond ($\Delta\Delta H_{XB\text{-}HB}$) is about -5 kcal/mol, while the entropic difference is near -8 cal/mol·K, leading to an overall free energy difference ($\Delta\Delta G_{XB\text{-}HB}$) of -3 kcal/mol, or that the Br2J near the temperature for crystallization generally favors the X-isomer, as observed. Thus, the thermodynamic results of the crystallographic and DSC assays are consistent with each other, with both resulting in a $\Delta\Delta H_{XB\text{-}HB} \approx$ -5 kcal/mol.

Table 3.4. Thermodynamic parameters for the difference in energy between the junction and duplex forms of the Br2J ($Br2J_{(J\text{-}D)}$) and the H2J ($Br2J_{(J\text{-}D)}$) constructs at the reference temperature (50.6 °C), and the subsequent X- vs H-bond interactions at the reference and room temperatures. The differences in enthalpies (ΔΔH) and entropies (ΔΔS) were extrapolated from the DSC determined values (Table 3.4) to each temperature using the heat capacities of each form. The free energy differences at each temperature (ΔΔG) were calculated by the standard relationship $\Delta\Delta G = \Delta\Delta H - T\Delta\Delta S$.

Construct	T (°C)	ΔΔH (kcal/mol)	ΔΔS (cal/molK)	ΔΔG (kcal/mol)
$Br2J_{(J\text{-}D)} - H2J_{(J\text{-}D)}$	50.6	-2 ± 1	-4.6 ± 3.7	0 ± 1
X-Bond – H-Bond	50.6	-3.6 ± 1	-9.2 ± 3.6	0 ± 1
X-Bond – H-Bond	25	-5.4 ± 1.1	-7.9 ± 3.6	-3 ± 1

4.2 Contribution of solvent and steric effects on isomer distributions in junctions.

Although we expect X-bonding to contribute to stabilization of the X-isomer over the H-isomer in the Br2J construct, bromines are also hydrophobic and occupy space. The question is, how much do the solvent and steric effects of the bromine contribute to the observed isomer distribution in the brominated DNA junction constructs? To address this question, we applied the crystallographic assay to a DNA construct in which the bromines in Br2J have been replaced with methyl groups (replacing the bromouracil bases with methyluracil, or thymine bases at the N_7 nucleotide position) to generate the T2J construct.

Methyl groups are very similar in hydrophobicity and size to bromines, but do not have strong electrostatic properties. The effective radius of a methyl group is ~2.0 Å (Case, Cheatham et al. 2005) as compared to the reported 1.85 Å van der Waals radius of a bromine (Bondi 1964). Furthermore, the partition coefficient (Hansch 1979) of methylbenzene (toluene) from water to octanol (2.7), which reflects the solvent free energy of the compound (Eisenberg and McLachlan 1986; Kagawa, Stoddard et al. 1989), is very similar to that of bromobenzene (3.0), and both are significantly more positive and thus more hydrophobic than benzene (~2.1). Studies of T2J, therefore, allow us to determine how a substituent that is similar in size and hydrophobicity to bromine affects the X- to H-isomer ratio in our crystallographic assay for the energies of X-bonds.

Crystals of T2J were isomorphous with H2J, Br2J, and Br1J crystals, indicating that this DNA construct also forms a junction, and was solved as such. During refinement of T2J, it was immediately obvious that the structure could not be represented by a single conformational model.

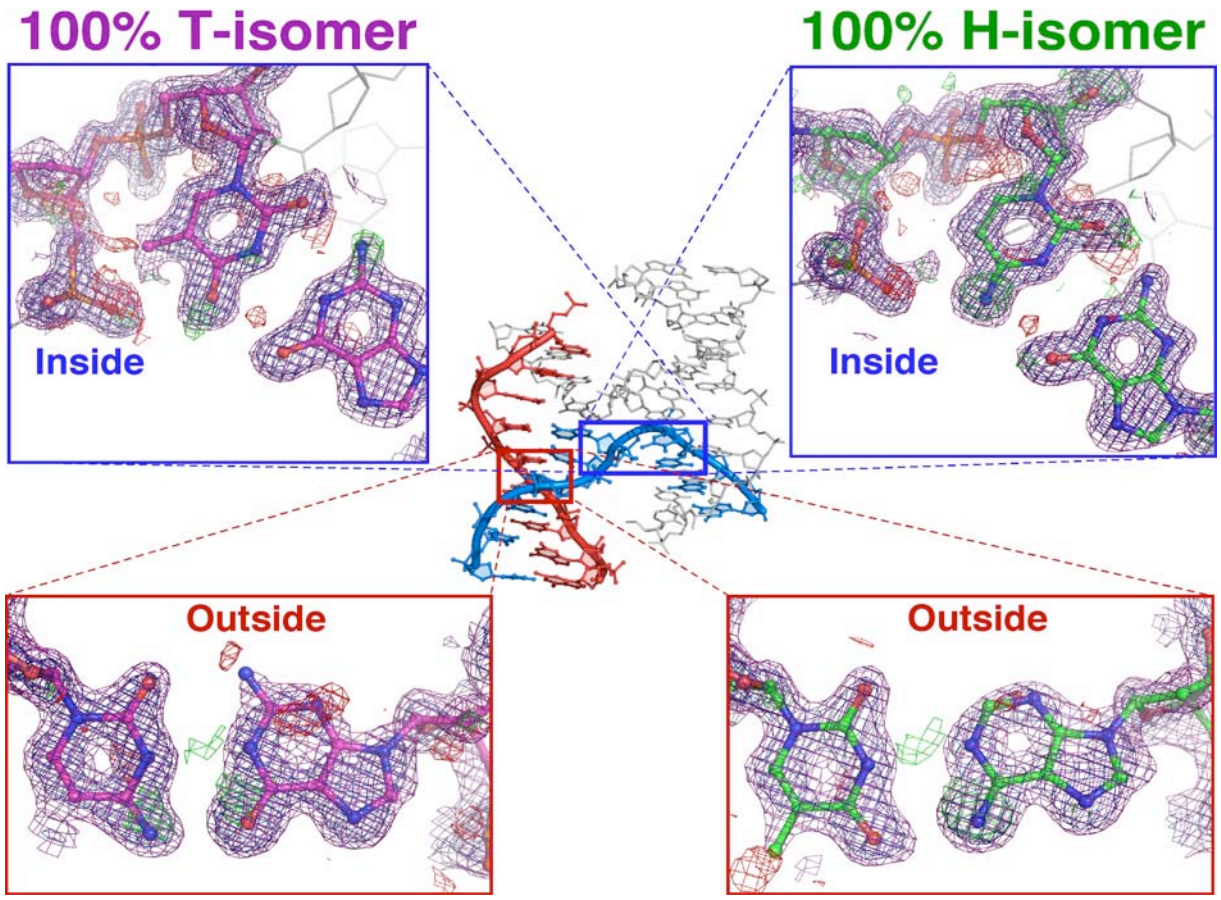


Fig. 3.6. Single-crystal structure of the T2J DNA construct. The structure is a four-stranded DNA junction, in which the complex of one outside (red) and one crossing strand (blue) constitutes the asymmetric unit (center structure), with the entire four-stranded junction generated by symmetry. The inset panels show the 2Fo-Fc electron density (blue wire, contoured at 2σ) and the Fo-Fc difference density (green for positive and red for negative, contoured at 2.5σ). The two left panels show the base pairs at the inside cross-over strand (top) and outside strand (bottom) refined as 100% T-isomer (magenta carbons), modeled as A•T base pairs at the inside crossing strand and C•G at the outside strand. The two right panels show the base pairs at the inside cross-over strand (top) and outside strand (bottom) refined as 100% H-isomer (green carbons), modeled as C•G base pairs at the inside of the junction and A•T base pairs on the outside.

The 2Fo-Fc electron density and Fo-Fc difference maps indicated a mixed population of X- and H-isomer with a majority of H-isomer (Fig. 3.6). When the structure was refined as 100% T-isomer, with the base pairs at the N_7 and complementary N_{14} positions modeled as a T•A pair, there was significant residual negative density observed in the difference electron density map at the methyl group of the thymine of the junction cross-over and residual positive density near the C2 carbon of the complementary adenine base, indicating that this inside position has significant contributions from the C•G base pair of the competing H-isomer. Similarly, the base pair of the outside strand, modeled as a C•G base pair, showed a small amount of residual density near the C5 carbon of the cytosine base, indicating contribution from a T•A base pair. When the structure is refined in the H-isomeric form, the difference maps were cleaner, but the negative residual density around the methyl group of the thymine of the outside strand suggests contributions of a C•G base pair from the competing T-isomer.

The T-isomer placed the methyl group of the T_7 thymine base to within 3.22 Å of and at an angle of 150.42° ($O\bullet\bullet\bullet C_{methyl}\text{-}C_5$) to the phosphate oxygen of the junction's crossing strand. This geometry is very similar to that of the bromine X-bond in the Br1J construct (the deviations reflect the 0.15 Å difference in the effective radii of the two substitutents), suggesting that there is very little difference in the steric effects on the isomer distribution for bromines and methyl groups, as expected. In contrast, the H-isomer positioned the N4 nitrogen of the C_7 cytosine to within 3.15Å and at angle of 102.53° from this same oxygen, indicative of a weak H-bond.

To quantify the isomer ratio, the occupancies of the methyl group of the T7 nucleotide and N2 nitrogen of the complimentary purine base were systematically

increased while monitoring the crystallographic R- and R_{free} values (Fig. 3.S2), similar to that used previously to estimate the isomer ratios in Br1J (Voth 2007). The resulting occupancy titration showed an approximate isomer distribution of 20 ± 5% T-isomer and 80 ± 5% H-isomer, equivalent to an energy difference of 0.8 kcal/mol in favor of the H-isomer. Thus, the anticipated stabilizing effect of burying a hydrophobic group, such as the methyl or a bromine, in the pocket of the junction in the X- (or T-) isomer is not sufficient to overcome the stabilizing electrostatic H-bond interaction of the H-isomer.

5. Conclusions

We have extended our previous study(Voth 2007) that applied a crystallographic assay on DNA junctions as a means to compare the stabilizing potential of bromine X-bonds against a standard H-bond. In the current study, we show that the calorimetric energies for the two interactions can be compared and contrasted in solution. The energies are shown to be consistent for this system between solution and crystal environments, thereby validating the original assay when extrapolated to room temperature. We also presented the structure of the T2J construct, which serves as a control to determine the contribution of hydrophobicity and steric effects on the preference for the X- vs H-isomer. Together, the results of the study provide greater insight into the energetics of X-bonding in a biological context.

The stabilizing potential for the bromine X-bond is seen to be greater than that of an H-bond in the DNA junction, with a large negative enthalpy (~5 kcal/mol) accounting for much of the stabilizing free energy of interaction.

Table 3.5. Comparison of energies for the Br1J and Br2J geometries (Table 3.1) estimated from the crystallographic assay[4], DSC in solution, and from MP2 calculations on $BrU{\bullet}H_2PO_2^{-1}$ as a model for the X-bonding pair.

Construct	Energies (kcal/mol)		
	$\Delta E_{XB\text{-}HB}$ Crystal Assay	$\Delta\Delta H_{XB\text{-}HB}$ (DSC)	MP2 Energies
Br1J	-2 ± 0.5	ND	-2.86
Br2J	-5	-5.4 ± 1.1	-6.40

This enthalpic stabilization can be attributed to the contribution of electrostatic and dispersion potentials to the bromine X-bond. To test this interpretation, we constructed a minimum model system that represents the X-binding interaction in the DNA, with the donor defined as the bromouracil base and the acceptor as a hypophosphite anion ($H_2PO_2^-$ [1]), with their positions and orientations defined by the conformations of the Br1J and Br2J constructs (Table 3.6). MP2 calculations on these models resulted in energies that very closely match those derived from the crystallographic (Voth 2007) and current solution studies, supporting the thesis that the X-bonding potential is primarily driven by electrostatic and dispersion energies.

What of the H-bond energy in this calculation? We propose here that the H-bonding energy is essentially compensated for by the hydrophobicity of the bromine, suggesting that the crystallographic and calorimetric studies are directly assaying the energetics of the X-bonding interaction. The T2J structure tells us that the hydrophobic and steric effects for a substituent that is analogous to bromine, but without the ability to participate in a strong electrostatic interaction, has a lower stabilizing potential compared to the H-bond. We would expect that sequestering a hydrophobic group such as the methyl of T2J or the bromine of Br2J into the junction would help stabilize the X-isomer (and equivalent T-isomer of T2J)—solvent free energy calculations based on differences in the exposure of the bromines in the Br1J junction suggested that the hydrophobic effect would contribute as much as 0.5 kcal/mol in favor of the X-isomer. In contrast, we would expect that placing a bulky substituent (methyl or bromine) in the tight pocket of the junction would sterically disfavor the X-isomer over the H-isomer form in the absence of the X-bonding interaction. The T2J results would indicate that, in this case, the

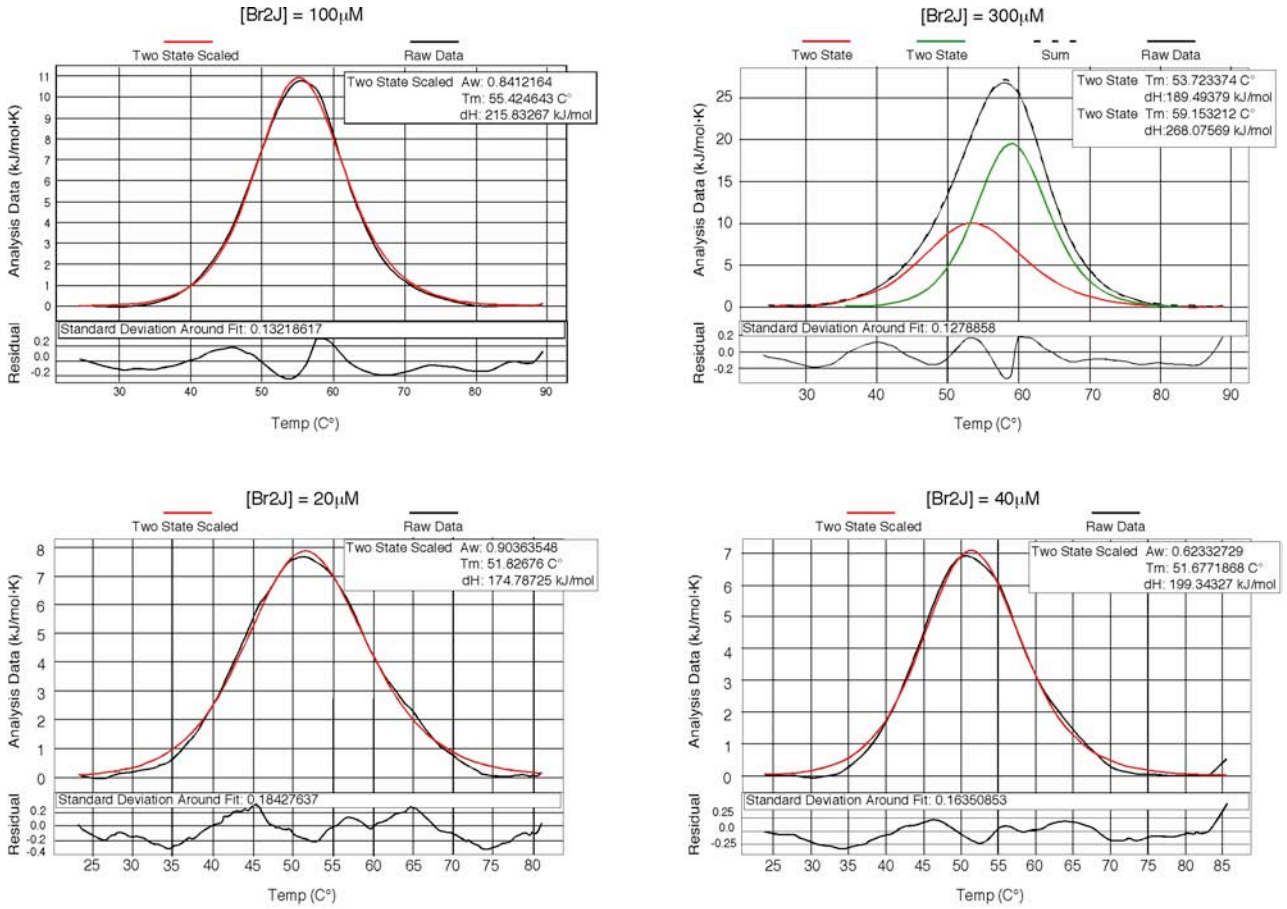
destabilizing steric effects win out, but only slightly, over the solvent effects, and that, together, they lose to the H-bond, but, again, only slightly (by ~0.8 kcal/mol). The slightly smaller bromine atom in comparison to the methyl group (the 0.15 Å difference in the effective radii is seen reflected in the shorter O•••Br distance in Br1J compared to the O•••CH_3 distance of T2J) is expected to reduce the steric effects and, in the end, could further minimize these small differences. Indeed, a reduction of the destabilizing steric effects by ~0.5 kcal/mol would balance the competing effects. We can conclude, therefore, that the enthalpic differences between X- and H-bonds on the stability of the DNA junction system in solution and in the crystal are, to the first approximation, directly measuring the X-bonding potential of the bromine donor to the negatively charged oxygen acceptor of the DNA phosphate.

6. Acknowledgements

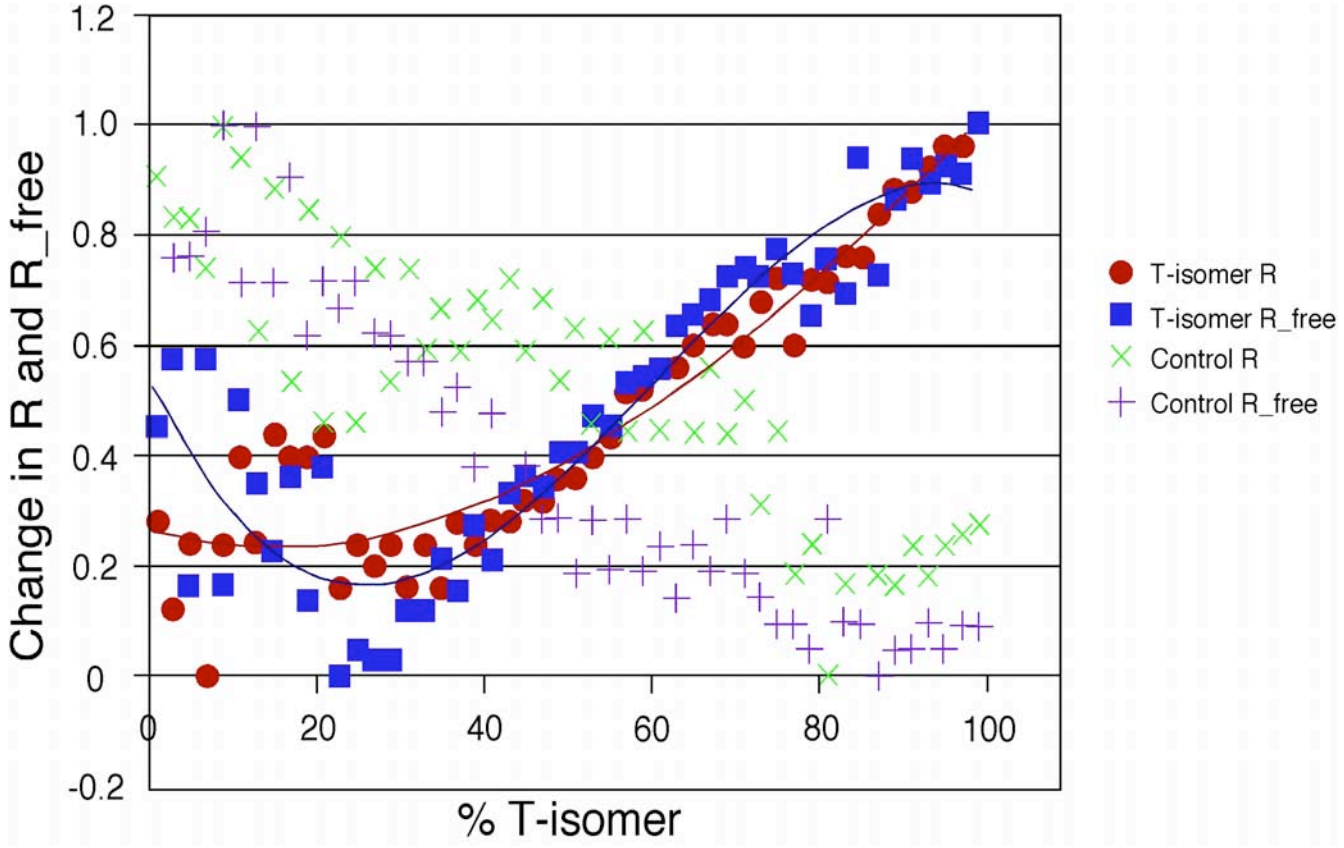
This work was supported by funds from Colorado State University.

7. Supplementing Information Available

Two figures (Fig. 3.S1: Figure showing analysis of thermodynamic data from differential scanning calorimetry; and Fig. 3.S2: Figure showing the crystallographic estimate for the percent of isomer forms in the T2J construct as a DNA junction) have been included as supplemental material in support of this study. This information is available free of charge via the Internet at http://pubs.acs.org



S1. Thermal profiles from differential scanning calorimetry analysis of DNA junctions. Lower [DNA] samples (20-100µM) are best fit with a single two-state model. However, increasing influence of junction formation is evident first as a deviation from a Gaussian profile near 55 C° and between 80-100µM [DNA] as an increase in the apparent Tm. Consequently data from 80-100µM samples was excluded from further analysis. At concentrations >100µM data is best fit with two separate two-state models to account for the mixed population of duplex and junction.



S2. Occupancy Titration of T2J. The occupancy of C57 methyl at the crossover was systematically decreased from 100% to 0% (decreasing T-isomer) as the occupancy of the N2 nitrogen of the complimentary base to the crossover position was increased from 0% to 100% (increasing H-isomer). The change in R and R_{free} were monitored. As controls, increments of C57 and N2 density were added to bases that should have no density in either isomer forms.

8. References

Auffinger, P., F. A. Hays, et al. (2004). "Halogen bonds in biological molecules." Proc. Natl. Acad. Sci., USA 101(48): 16789-16794.

Berman, H. M., J. Westbrook, et al. (2000). "The Protein Data Bank." Nucleic Acids Res 28(1): 235-242.

Bondi, A. (1964). "van der Waals volumes and radii." J Phys Chem A 68(3): 441-451.

Brünger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr 54 (Pt 5): 905-921.

Case, D. A., T. E. Cheatham, 3rd, et al. (2005). "The Amber biomolecular simulation programs." J Comput Chem 26(16): 1668-1688.

Duckett, D. R., A. I. H. Murchie, et al. (1988). "The structure of the Holliday junction, and its resolution." Cell 55: 79-89.

Dunitz, J. D. (1995). "Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions." Chem Biol 2(11): 709-712.

Eichman, B. F., J. M. Vargason, et al. (2000). "The Holliday junction in an inverted repeat sequence: sequence effects on the structure of four-way junctions." Proc. Natl. Acad. Sci., USA 97: 3971-3976.

Eisenberg, D. and A. D. McLachlan (1986). "Solvation energy in protein folding and binding." Nature 319(6050): 199-203.

Fedorov, O., K. Huber, et al. (2011). "Specific CLK inhibitors from a novel chemotype for regulation of alternative splicing." Chem Biol 18(1): 67-76.

Grainger, R. J., A. I. H. Murchie, et al. (1998). "Exchange between stacking conformers in a four-way DNA junction." Biochemistry 37: 23-32.

Hansch, C. L., A. (1979). Substituent Constants for Correlation Analysis in Chemistry and Biology. New York, Wiley.

Hays, F. A., V. Schirf, et al. (2006). "Solution formation of Holliday junctions in inverted-repeat DNA sequences." Biochemistry 45(8): 2467-2471.

Hays, F. A., J. M. Vargason, et al. (2003). "Effect of sequence on the conformation of DNA holliday junctions." Biochemistry 42(32): 9586-9597.

Hays, F. A., J. Watson, et al. (2003). "Caution! DNA crossing: crystal structures of Holliday junctions." J Biol Chem 278(50): 49663-49666.

Ho, P. S., and Eichman, B. F. (2001). "The crystal structures of DNA Holliday junctions." Current Opin. Struct. Biol. 11: 302-308.

Holliday, R. (1964). "A mechanism for gene conversion in fungi." Genet. Res. 5: 282-304.

Kagawa, T. F., D. Stoddard, et al. (1989). "Quantitative analysis of DNA secondary structure from solvent-accessible surfaces: the B- to Z-DNA transition as a model." Biochemistry 28(16): 6642-6651.

Kraut, D. A., M. J. Churchill, et al. (2009). "Evaluating the potential for halogen bonding in the oxyanion hole of ketosteroid isomerase using unnatural amino acid mutagenesis." ACS Chem Biol 4(4): 269-273.

Lii, J. H. and N. L. Allinger (2008). "The Important Role of Lone-Pairs in Force Field (MM4) Calculations on Hydrogen Bonding in Alcohols." J. Phys. Chem. A 112(46): 11903-11913.

Lilley, D. M. J. (1999). Structures and interactions of helical junctions in nucleic acids. Oxford Handbook of Nucleic Acid Structure. S. Neidle. New York, Oxford University Press: 471-498.
Lilley, D. M. J. (2000). "Structures of helical junctions in nucleic acids." Quart. Rev. Biochem. 33: 109-159.
Liu, L., W. A. Baase, et al. (2009). "Halogenated benzenes bound within a non-polar cavity in T4 lysozyme provide examples of I...S and I...Se halogen-bonding." J Mol Biol 385(2): 595-605.
Lu, Y., T. Shi, et al. (2009). "Halogen bonding--a novel interaction for rational drug design?" J. Med. Chem. 52(9): 2854-2862.
Matter, H., M. Nazare, et al. (2009). "Evidence for C-Cl/C-Br...pi interactions as an important contribution to protein-ligand binding affinity." Angew Chem Int Ed Engl 48(16): 2911-2916.
Metrangolo, P. and G. Resnati (2001). "Halogen bonding: a paradigm in supramolecular chemistry." Chemistry 7(12): 2511-2519.
Metrangolo, P., Resnati, G. (2008). "Chemistry: Halogen versus hydrogen." Science 321: 918-919.
Murray, J. S., P. Lane, et al. (2007). "Sigma-hole bonding: molecules containing group VI atoms." Journal of Molecular Modeling 13(10): 1033-1038.
Politzer, P., J. S. Murray, et al. (2010). "Halogen bonding: an electrostatically-driven highly directional noncovalent interaction." Physical Chemistry Chemical Physics 12(28): 7748-7757.
Ortiz-Lombardía, M., A. González, et al. (1999). "Crystal structure of a DNA Holliday junction." Nat. Struct. Biol. 6: 913-917.
Otwinowski, Z. and W. Minor (1997). "Processing of x-ray diffraction data collected in oscillation mode." Methods Enzymol. 276: 307-326.
Robinson, B. H. and N. C. Seeman (1987). "The design of a biochip: a self-assembling molecular-scale memory device." Protein Eng 1(4): 295-300.
Rothemund, P. W. (2006). "Folding DNA to create nanoscale shapes and patterns." Nature 440(7082): 297-302.
Schmidt, J. R., and Polik, W.F. (2005). "WebMO." 6.0. from http://www.webmo.net.
Schmidt, M. W. K., K. Baldridge, J.A. Boatz, S.T. Elbert,, J. J. J. M.S. Gordon, S. Koseki, N. Matsunaga,, et al. (1993). "General atomic and molecular electronic structure system." J. Comput. Chem. 14: 1347-1363.
Seeman, N. C. (1999). "DNA engineering and its application to nanotechnology." Trends Biotechnol 17(11): 437-443.
van Pee, K. H. and S. Unversucht (2003). "Biological dehalogenation and halogenation reactions." Chemosphere 52(2): 299-312.
Voth, A. R., and Ho, P.S. (2007). "The Role of Halogen Bonding in Inhibitor Recognition and Binding by Protein Kinases." Curr. Topics Med. Chem. 7: 1336-1348.
Voth, A. R., Hays, F.A., and Ho, P.S. (2007). "Directing macromolecular conformation by halogen bonds." Proc. Natl. Acad. Sci., USA 104: 6188-6193.
Wu, P., S. Nakano, et al. (2002). "Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation." Eur J Biochem 269(12): 2821-2830.

Wu, W., Samoszuk, M. K., Comhair, S. A. A., Thomassen, M. J., Farver, C. F., Dweik, R. A., Kavuru, M. S., Erzurum, S. C., and Hazen, S. L. (2000). "Eosinophils generate brominating oxidants in allergen-induced asthma." J. Clin. Invest. 105: 1455-1463.

CHAPTER 4

SCALABLE ANISOTROPIC SHAPE AND ELECTROSTATIC MODELS FOR BIOLOGICAL BROMINE HALOGEN BONDS[3]

1. Summary

Halogens are important substituents of many drugs and secondary metabolites, but the structural and thermodynamic properties of their interactions are not properly treated by current molecular modeling and docking methods that assign simple isotropic point charges to atoms. Halogen bonds, for example, are becoming widely recognized as important for conferring specificity in protein-ligand complexes, but, to this point, are most accurately described quantum mechanically. Thus, there is a need to develop methods to both accurately and efficiently model the energies and geometries of halogen interactions in biomolecular complexes. We present here a set of potential energy functions that, based on fundamental physical properties of halogens, properly model the anisotropic structure-energy relationships observed for halogen interactions from

---

[3] Megan Carter[1], Anthony K. Rappé[2], and P. Shing Ho[1,*]
[1]Department of Biochemistry and Molecular Biology,
[2]Department of Chemistry, Colorado State University,
Fort Collins, CO 80523, USA.
Reproduced with permission from Journal of Chemical Theory and Computation, ACS Publications.

crystallographic and calorimetric data, and from ab initio calculations for bromine halogen bonds in a biological context. These energy functions indicate that electrostatics alone cannot account for the very short-range distances of bromine halogen bonds, but requires a flattening of the effective van der Waals radius that can be modeled through an angular dependence of the steric repulsion term of the standard Lennard-Jones type potential. This same function that describes the aspherical shape of the bromine is subsequently applied to model the charge distribution across the surface of the halogen, resulting in a force field that uniquely treats both the shape and electrostatic charge parameters of halogens anisotropically. Finally, the electrostatic potential was shown to have a distance dependence that is consistent with a charge-dipole rather than a simple Coulombic type interaction. The resulting force field for biological halogen bonds (ffBXB) is shown to accurately model the geometry-energy relationships of bromine interactions to both anionic and neutral oxygen acceptors, and is shown to be tunable by simply scaling the electrostatic component to account for effects of varying electron-withdrawing substituents (as reflected in their Hammett constants) on the degree of polarization of the bromine. This approach has broad applications to modeling the structure-energy relationships of halogen interactions, including the rational design of inhibitors against therapeutic targets.

2. Introduction

Accurate methods to model noncovalent molecular interactions are crucial to "bottom-up" strategies in biomolecular engineering. Current molecular mechanics (MM) force fields are powerful tools for modeling biomolecular systems and have, for example, been successful in accurately predicting affinities of ligands in various protein complexes (Steinbrecher and Labahn ; Huang and Jacobson 2007; Boyce, Mobley et al. 2009). Halogens are typically considered to be hydrophobic substituents that are electron-rich and, as a consequence, should repel electronegative atoms; however, halogens as covalent substituents in organic and biomolecular molecules are now recognized as displaying simultaneously electronegative and electropositive potentials, allowing them to serve both as hydrogen bond (H-bond) acceptors (Politzer, Murray et al. 2007; Lu, Wang et al. 2009) and as halogen bond (X-bond) donors (Auffinger, Hays et al. 2004; Metrangolo 2005). Their electrostatic properties, therefore are highly directional and should be treated as such. In this study, we have derived a set of potential energy functions that describe the aspherical shape and anisotropic distribution of electrostatic potentials of bromines, thereby providing framework for the structure-energy relationships that can accurately model the ability of halogens to simultaneously participate in X- and H-bonding interactions.

Halogen bonds, formerly called "charge transfer bonds" (Hassel 1972), have seen a resurgence of interest as a tool to engineer new molecular materials (Metrangolo 2008), including, in medicinal and biophysical chemistry, the design of new protein inhibitors (Lam, Clark et al. 2009; Matter, Nazare et al. 2009; Xu, Liu et al.) and for directing DNA conformations (Voth 2007). X-bonds are closely related to the better-known hydrogen

bonds (Fig. 4.1a and b) (Ouvrard, Le Questel et al. 2003) in that both have donor-acceptor distances that can be significantly shorter than the sum of their respective van der Waals radii ($\sum r_{vdW}$) and share a common set of acceptors (we adopt definitions that are analogous to that of H-bonds, where the X-bond donor donates a positive charge from polarization of the halogen and the acceptor is an electron-rich atom or group involved that pairs with the donor (Auffinger, Hays et al. 2004; Metrangolo 2008)).

The X-bond "donor" is an electropositive cap resulting from polarization of the halogen (X) along the C-X σ-bond (Fig. 4.1c). The "σ-hole" model (Murray, Lane et al. 2007; Politzer, Murray et al. 2007) provides a simple, primarily electrostatic description of this phenomenon (Lii and Allinger 2008): in this model, the valence electron in the outer shell p_z orbital participates in formation of the covalent σ-bond, leaving the orbital depopulated and, thus, exposing the nuclear charge that is the electropositive crown opposite the covalent bond. The degree to which the orbital is polarized follows the series I > Br > Cl > F, which defines the order of stabilizing energies of the X-bond (Lommerse, Stone et al. 1996). The p_x- and p_y-orbitals, however, remain fully occupied, thereby providing an electronegative annulus around the halogen that serves as a potential H-bond acceptor perpendicular to the σ-bond. Thus, the polarized halogens are amphoteric, serving both as X-bond donors in one direction and classical H-bond acceptors in the perpendicular direction (Politzer, Murray et al. 2007; Lu, Wang et al. 2009). There is an understanding, however, that X-bonding is not solely an electrostatic effect, but that dispersion and, to a lesser extent, charge transfer also contribute at least to the energetics of the interaction.

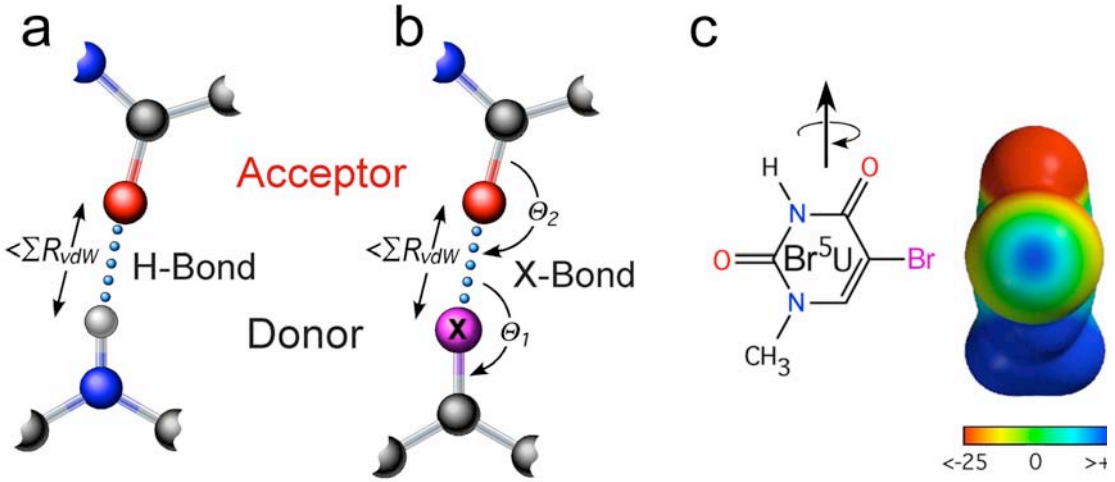


Fig. 4.1. Hydrogen and halogen bonds. a. The hydrogen bond (H-bond) is a non-covalent interaction in which the approach of the hydrogen donor to the acceptor atom (red) is closer than the sum of their van der Waals radius (ΣR_{vdW}). b. Similarly, the halogen bond (X-bond) brings the halogen donor (magenta) closer to the acceptor (red) than their ΣR_{vdW}. As a highly directional interaction, the X-bond is also defined by the angle of approach of the acceptor to the donor (Θ_1) and the donor to the acceptor (Θ_2). c. The bromine substituent of a 5-bromouracil (Br^5U) base is modeled to show the positive crown resulting from polarization along the C-Br bond. Electrostatic potentials (from -25 kcal/mol to + 25 kcal/mol) were estimated from DFT calculations (Auffinger, Hays et al. 2004).

X-bond acceptors in biological systems include both charged and uncharged oxygens, amino and imino nitrogens, sulfurs, and aromatic rings (Auffinger, Hays et al. 2004; Voth 2007; Parisini, Metrangolo et al. 2011; Riley, Murray et al. 2011). Their energies of interaction depend strongly on the geometries relating the donor and the acceptor (Auffinger, Hays et al. 2004; Voth 2007). X-bonds are highly directional (Lommerse, Stone et al. 1996), with geometries defined by the angular approach of the acceptor towards the halogen (Θ_1) and of the halogen to the acceptor (Θ_2) (Fig. 4.1); Θ_1 is generally in the direction of halogen polarization (Lommerse, Stone et al. 1996; Ouvrard, Le Questel et al. 2003; Auffinger, Hays et al. 2004; Lu, Shi et al. 2009), while Θ_2 aligns with the nonbonding or π-electrons of the acceptor (Voth 2009).

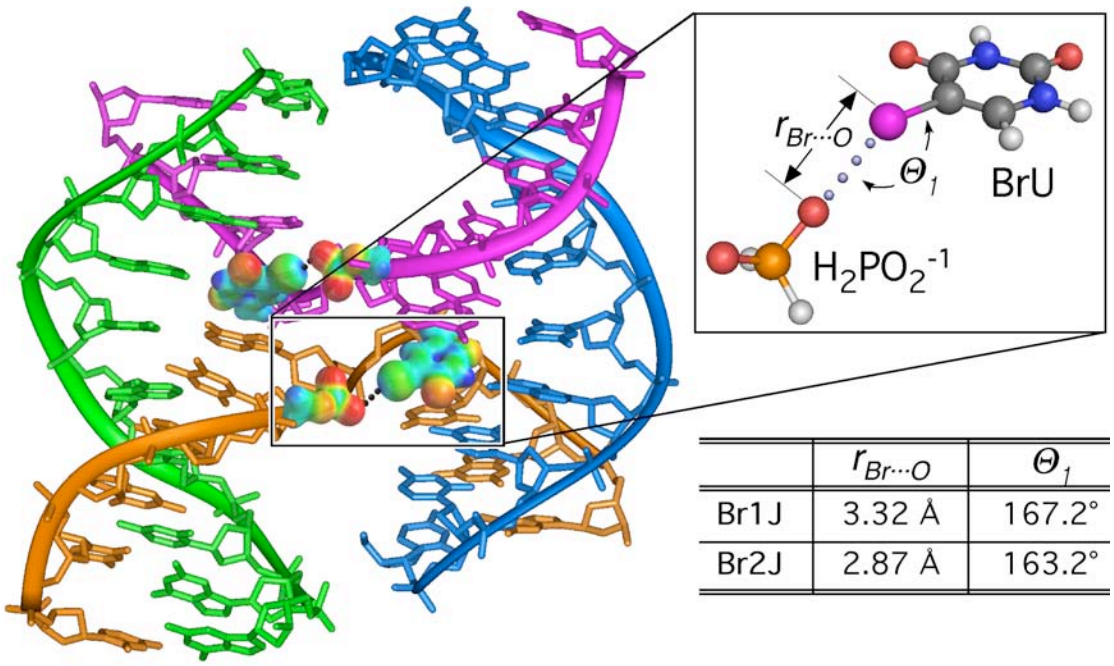
The contribution of polarization to halogen interactions (Politzer, Murray et al. 2007; Voth 2009) is most accurately modeled through quantum mechanical (QM) and semi-empirical QM methods (Ibrahim 2011), but is poorly treated by MM algorithms that treat atoms classically as single point charges (Dobes, Rezac et al. 2011). QM calculations applying density functional theory (DFT) on the ultrahigh resolution aldose reductase/inhibitor structure (Muzet, Guillot et al. 2003) had identified an X-bond that accounts for the high specificity of the inhibitor for this enzyme, while the semi-empirical PM6-DH2X method has been shown to accurately model both the geometries and binding energies of several kinase inhibitors (Dobes, Rezac et al. 2011). Finally, hybrid QM/MM approaches (Vreven, Morokuma et al. 2003; Case, Cheatham et al. 2005) have seen some success in modeling X-bonds in protein-ligand complexes (Lu, Shi et al. 2009). However, QM calculations on biomolecules remain very time intensive and are subject to cumulative errors in atomic coordinates. There is, therefore, a great need for an

accurate molecular mechanics force field for biological X-bonds that is consistent with and, therefore, can be incorporated into current MM algorithms to facilitate our ability to exploit halogens as design elements in engineering biomolecular interactions.

Recently, there have been attempts to model the positive crown of halogens and their associated X-bonds using a positive extra-point (PEP) approach, where an additional partial positive point charge is added, displaced at some distance from the atomic center, while maintaining the overall negative charge of the halogen. The work of Ibrahim (Ibrahim 2012) on small molecular systems as well as protein-inhibitor complexes demonstrate that such an approach can be useful in modeling X-bonds in multiple systems. From molecular dynamics studies applying the AMBER MM force field, the PEP method was capable of reproducing the X-bond lengths (to within 0.1 to 0.29 Å, with an rmsd of 0.2 Å) and energies (within 0.1 to 0.37 kcal/mol, rmsd 0.27 kcal/mol) from MP2 calculations of bromobenzene donors to various acceptors. In addition, the PEP approach was capable of calculating inhibitor binding energies to CK2 kinases that correlated well (R-values of 0.92 to 0.96) with the experimental values, with acceptors approaching the halogens at near linear angles. The absolute energies, however, were significantly more negative (by ~20 kcal/mol) than the experimental values, and the calculated X-bond lengths were significantly longer than those seen in the X-ray structures (this could be attributed to waters that were crystallographically observed in the active site, but absent from the AMBER models). Thus, the field of developing MM models for X-bonds remains open to improvements and new approaches.

In the current studies, we derive a set of simple directional potential energy functions to model the shape and electrostatic properties of halogens, which, together,

constitute a set of potential energy functions for an ffBXB, force field for biological halogen bonds. This differs from attempts at modeling X-bonds using a purely electrostatic PEP approach (Sponer, Riley et al. 2008; Ibrahim 2011) in that the ffBXB attempts to also model both the repulsive steric and attractive dispersion contributions to the physicochemical properties of halogen interactions—although the σ-hole model for X-bonds does not explicitly consider steric and dispersion terms, we show that the size and shape of the halogen is aspherical, which we interpret to be attributable to the depopulation of the atomic p_Z-orbital, a hallmark of the σ-hole model. The ffBXB functions are parameterized against the AMBER force field (Case, Cheatham et al. 2005) and applied to the structure-energy relationships of X-bonds derived from a four-stranded DNA junction system studied in crystals (Voth 2007) and in solution (Gribble 2003). We have developed four-stranded DNA junctions as a unique model system to assay the energies of X-bonds in a biological context, where a halogen interaction competes against a classic H-bond in stabilizing the complex (Fig. 4.2). The energies of two geometries (the longer Br1J and shorter Br2J X-bonds) have been characterized for bromine X-bonds ($Br\cdots O^{-1}$) in this system, and have shown that the steric and hydrophobic properties of the bromine essentially compensate for the stabilizing potential of the competing H-bond; thus, the experimental energies can be applied directly as an empirical test for a set of QM calculations as well as the ffBXB functions derived from them.



	$r_{Br\cdots O}$	Θ_1
Br1J	3.32 Å	167.2°
Br2J	2.87 Å	163.2°

Fig. 4.2. Model for bromine X-bond in DNA junction. A four-stranded DNA junction stabilized by an X-bond from the bromouracil base at the nucleotide N7 position to the phosphate backbone at N6 of the crossing strand of the junction. The inset shows the bromouracil (BrU) to hypophosphite ($H_2PO_2^-$) pair used to model the X-bond interaction within the DNA, along with the distance (r) and angle (Θ_1) definitions for the geometry of the interaction between the bromine X-bond donor and the oxygen acceptor. Two specific geometries are observed for the X-bond in the DNA junction (Voth 2007), with the longer Br1J is less stable than the shorter Br2J form.

We then apply the ffBXB to map the geometry-energy relationships of X-bonds with formally neutral oxygens (the primary interactions seen in complexes of halogenated ligands with the peptide backbone of protein (Auffinger, Hays et al. 2004; Voth 2009; Parisini, Metrangolo et al. 2011)), to predict the geometries of potential H-bonds to bromines (demonstrating the ability of the potential energy functions to model the amphoteric properties of halogens), and to develop parameters for PEP models that approximates and, thus, can be compared and contrasted with the structure-energy relationships of the ffBXB. The quality of each model is measured by their correlation with QM and experimental energies for the X-bonds in the DNA junction system as well as QM energies for a bromobenzene···acetone pair, which serves as a model for X-bonds in protein systems.

3. Theory and Methods

The primary goal of the current study is to derive a set of potential energy functions that accurately describes the short-range and angular dependent properties of halogen bonding interactions in the context of a biological system. The biological system we chose to model is that of a four-stranded DNA junction (Fig. 4.2), which has been shown to be stabilized by and whose conformation can be controlled by X-bonding (Gribble 2003; Voth 2007). For this set of studies, we focus on bromine, which has been the most extensively studied in the DNA system in terms of the experimentally derived structure-energy relationships, and for which the steric and hydrophobicity contributions to the interaction largely balance the energy of the competing H-bond (Gribble 2003). In this system (Voth 2007), the energy is directly correlated with specific distances and

geometries of the interaction in a crystal, with the crystal-state energies seen to correlate well with those in solution (Gribble 2003).

Our initial attempt to model the structure-energy relationship of X-bonds started by determining whether a standard set of potentials applying single-point charges in a Coulombic function and van der Waals functions could describe the geometries seen in the DNA junction system. We found that a stabilizing potential could be calculated for an X-bond with the acceptor approaching the bromine in a linear orientation to the σ-bond; however, in order to allow an oxygen to approach the bromine to within 2.9 Å (near the optimum distance for the interaction), the halogen had to be assigned an unusually large positive charge (+2e), or the potential energy well for the van der Waals interaction of the Br···O pair needed to be set at <10% of the standard values (data not shown). In either case, the energies from this simplistic model did not fit well with the observed experimental results, showing a 3.3 Å X-bond to be ~4 kcal/mol more favorable than the shorter 2.9 Å interaction. We, therefore, need to treat the basic physicochemical properties of halogens as substitutents in molecular systems in a different way.

To develop a more accurate model for halogens, we started with a set of very basic questions: What is the physical shape of a bromine substituent in a covalent molecule, what determines this shape, and can the shape be modeled empirically? We then determine whether the same principles that dictate the shape can be applied to model the anisotropic electrostatic properties of the halogen. The resulting potential energy functions that describe the shape and electrostatic properties are then parameterized against QM calculations on a minimum molecular model for the enthalpic components of

the interaction in the experimental DNA system and, finally, extrapolated to broader classes of interactions of halogens in other biological contexts.

3.1 Effective shape of bromine

The effective shape of an atom in a molecule can be described by the two competing terms of the van der Waals (vdW) interaction (E_{vdW}): the attractive dispersive London force acting at long distances are opposed by the steric repulsive forces at short distances. We mapped the effective atomic radius of a bromine starting with a simple Br_2···He interaction pair, placing the He atom at distance intervals from 2.5 to 5 Å for angles from 90° to 180° (in 22.5° steps). Since He is a small, nonpolarizable atom, high level ab initio calculations on this model complex allows us to focus on the distance and angle dependence of the competing attractive and repulsive (Gribble 2003) components, similar to previous approaches used to map the shape of chlorine (Peebles, Fowler et al. 1995). Second-order Møller-Plesset (MP2) calculations (Møller and Plesset 1934) show the total E_{vdW} for the Br···He interaction to be strongly dependent on the angle of approach of the He atom relative to the Br-Br covalent bond, Θ_1 (Fig. 4.3a). These results indicate that the shape of bromine is aspherical, with the effective van der Waals radius (R_{vdW}) being ~0.5 Å shorter when approaching along as opposed to perpendicular to the Br-Br bond.

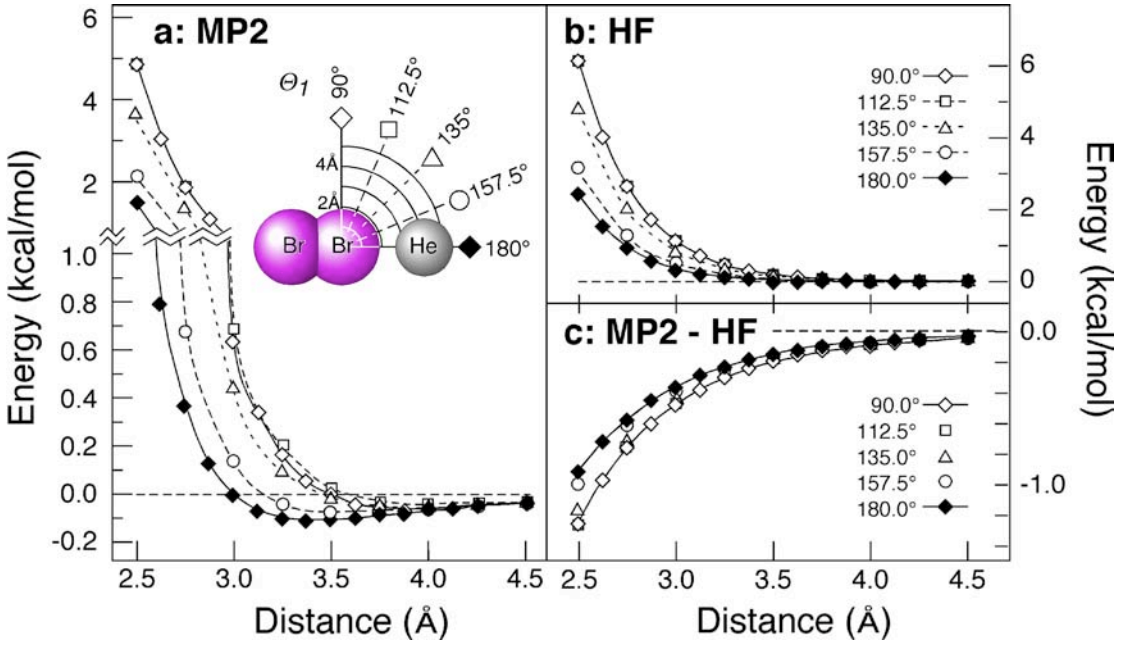


Fig. 4.3. Quantum chemical calculations (applying augmented cc-pVTZ basis) mapping distance-angle relationship of He interacting with Br_2. a. (MP2) calculations applied to He···Br distances from 2.5 to 5.0 Å (data to 4.5 Å shown), and He···Br-Br angles (Θ_1) from 90° to 180°, in 22.5° increments. b. Hartree-Fock (HF) calculations on the same system as a. c. Difference between the MP2 and HF calculated energies. Curves connecting each point have been added to show trends and have no theoretical basis in this figure.

To establish the root of the aspherical shape of the bromine, a Hartree-Fock (HF) calculation was applied to this $Br_2\cdots He$ model to probe the repulsion energies in the absence of any dispersive interactions. This analysis showed that the repulsive term is also highly dependent on Θ_1 (Fig. 4.3b). At the standard van der Waals distance ($\sum R_{vdW}$ = 3.25 Å), there is an ~0.33 kcal/mol difference between the linear (Θ_1 = 180°) and the perpendicular (Θ_1 = 90°) approach. Similarly, the distance at which the repulsion energy is 1 kcal/mol extends from ~2.7 Å for the linear to ~3.0 Å for the perpendicular approach. By subtracting the HF energies from the MP2 calculated energies, we can estimate the contribution of dispersion to the overall E_{vdW} (Fig. 4.3c). The dispersion component is seen to be relatively independent of Θ_1, with a difference of <0.4 kcal/mol between the linear and perpendicular approach at 2.5 Å distance. Thus, the angle dependence of the interacting atoms derives primarily from the repulsion term, allowing us to treat the dispersive component as essentially isotropic in terms the effective R_{vdW} for the bromine.

In order to relate these results to the effective shape in terms of the van der Waals radius (R_{vdW}) of the bromine atom, we fit the overall MP2 calculated E_{vdW} to a modified Lennard-Jones type potential energy function (V_{LJ}) in which the repulsive ($1/r^{12}$) component is treated as a function of the Θ_1-angle, while the dispersion ($-1/r^6$) component is angle independent (Eq. 1). For the angle dependent term, we define an effective average van der Waals radius for the bromine ($<R_{vdW(Br)}>$) that is applied to both components, but with a $\Delta R\cos[va]$ added only to the repulsive component, where ΔR is a perturbation to the $< R_{vdW(Br)} >$, ν is the period of the cosine function for the Θ_1-dependence, and $a = 180° - \Theta_1$.

$$V_{LJ} = 4\varepsilon\left[\left(\frac{R_{vdW(He)} + \left\langle R_{vdW(Br)}\right\rangle - \Delta R\cos(\nu\alpha)}{r}\right)^{12} - \left(\frac{R_{vdW(He)} + \left\langle R_{vdW(Br)}\right\rangle}{r}\right)^{6}\right] \qquad \text{Eq. 1}$$

A nonlinear least squares fit of the V_{LJ} function in Eq. 1 to the distance-angle dependence of E_{vdW} from the MP2 calculations results in a set of values for $<R_{vdW(Br)}>$, ΔR, and ν that describes the overall shape of the bromine in Br_2. The resulting $<R_{vdW(Br)}>$ = 1.816 Å is only slightly shorter than the accepted isotropic 1.85 Å value (Bondi 1964), while the ΔR of 0.157 Å indicates that there is a significant flattening of the Br atom along the σ-bond ($\Theta_1 = 180°$) and bulging approximately perpendicular to the bond. The minimum and maximum effective $R_{vdW(Br)}$ range from 1.66 Å to 1.97 Å, equivalent to an ~16% (0.314 Å) difference between the smallest and largest effective radius (Fig. 4.4). A similar polar flattening of 0.37 Å was reported from an analysis of the crystal structures of Br_2 complexes (Nyburg 1979). The value of $\nu = 2.53$ indicates that the bulge is at Θ_1 = 108.9° rather than at 90° expected for the ideal orientation of the p_x and p_y orbitals (ideally, $\nu = 2.0$), relative to the covalent Br-Br bond, suggesting that these orbitals are canted ~19° from perpendicular and towards the depopulated p_z-orbital. This non-perpendicular bulge is consistent with recent calculations on the multipolar electron densities showing the maximum charge concentration for bromine from 95° to 110° relative to the C-Br bond (personal communication Prof. E. Espinosa, Université de Nancy, France).

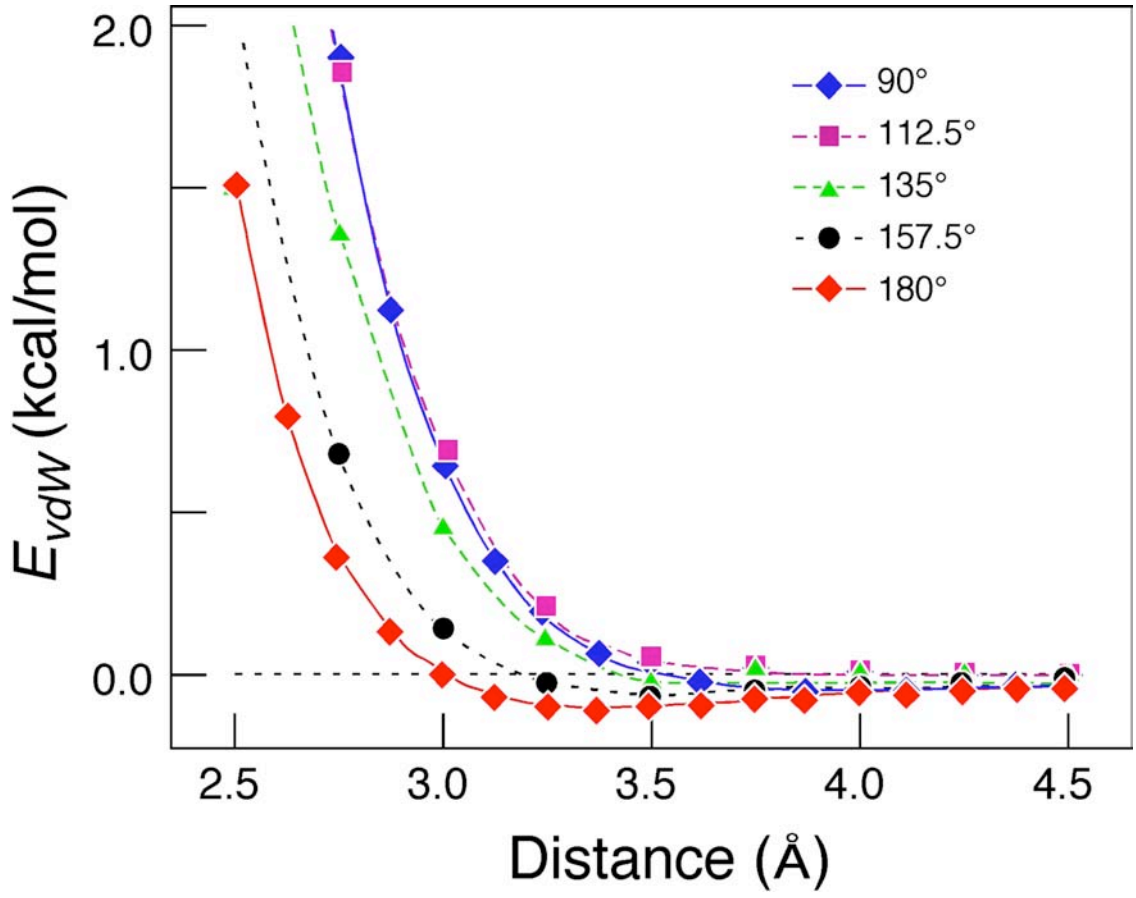


Fig. 4.4. QM calculated van der Waals energy (E_{vdW}) fitted as a directional Lennard-Jones potential (Eq. 1). The fitted parameters result in an overall correlation coefficient of 0.987, with a standard deviation between the MP2 calculated and fitted curves of 0.155 for $E_{vdW} \leq 1.0$ kcal/mol.

3.2 Effective partial charge and electrostatic potential function for bromine

The basic premise of the σ-hole model for halogen bonding is that the p_z-orbital of covalently bonded halogen is depopulated when forming the σ-bond; consequently, an electropositive cap is created from polarization of the electrostatic potential along this bond, and it is this cap that interacts with electron-rich acceptors, such as negative and neutrally charged oxygens, sulfurs, and nitrogens. Overall, bromines are considered to be slightly negatively charged, but can carry a partial positive charge when bound to a strongly electron withdrawing atom or group (Armstrong 2012) (as in BrCl). Thus, the aspherical shape of the bromine described above would suggest an anisotropic charge distribution across the atomic surface of bromine, with the effective charge being most positive for a linear approach towards the σ-bonded Br and most negative for an approximate perpendicular approach. This is indeed what has been observed in QM calculations of charge distributions across various halogens (Lii and Allinger 2008; Lu, Wang et al. 2009).

A simple model to describe such an anisotropic charge distribution would be to apply the same cosine function that was used to model the aspherical shape of the bromine to define the effective partial charge of the bromine (Z_{Br}) as a function of the approach angle ($a = 180° - \Theta_1$) (Eq. 2).

$$Z_{Br} = A\cos(na) + B \qquad \text{Eq. 2}$$

In this form, the cosine function is identical to that in Eq. 1, with the parameter A introduced to scale the amplitude of the cosine function, and B to define the baseline for where the positive and negative charges cross.

The period (ν) of the function and the ratio of A/B were determined by applying Eq. 2 to the previously reported electrostatic potential distribution for a bromine of bromobenzene (Lu, Wang et al. 2009) (Fig. 4.5). The fitted period of the cosine function places the most negative partial charge at ~99.3°, or nearly perpendicular to the covalent bond, indicating that the orientations of the p_x- and p_y-orbitals are dependent on the context of the bromine as a substituent—the phenyl group is less electronegative than a bromine substitutent and, therefore, the p_z-orbital electrons are expected to be less depopulated in bromobenzene than for Br_2.

The overall charge of the bromine in this anisotropic model was estimated by considering the charge at each angle a (from 0° to 90° in 5° increments) as defined by Eq. 2, and scaling that to the area associated with the surface in the increment from a_1 to a_2 according to the following Eq. 3.

$$SA(\alpha_1 \rightarrow \alpha_2) = 2\pi \int_{\alpha_1}^{\alpha_2} \left(\left\langle R_{o(Br)} \right\rangle - \Delta R \cos \nu\alpha \right) \sin\alpha \; d\alpha =$$

$$\pi \left(2 \left\langle R_{o(Br)} \right\rangle \cos\alpha - \Delta R \left(\sin\left[(1+\nu)\alpha \right] + \sin\left[(1-\nu)\alpha \right] \right) \right]_{\alpha_1}^{\alpha_2} \qquad \text{Eq. 3}$$

The resulting overall normalized partial charge of the bromine is estimated to be Z_{Br} = -0.14e, similar to results from MP2 calculations. The bromine is seen to be effectively positively charged for $\Theta_1 \geq 130°$, allowing it to serve as an X-bond donor in this range, and negative for $\Theta_1 \leq 130°$, serving as an H-bond acceptor.

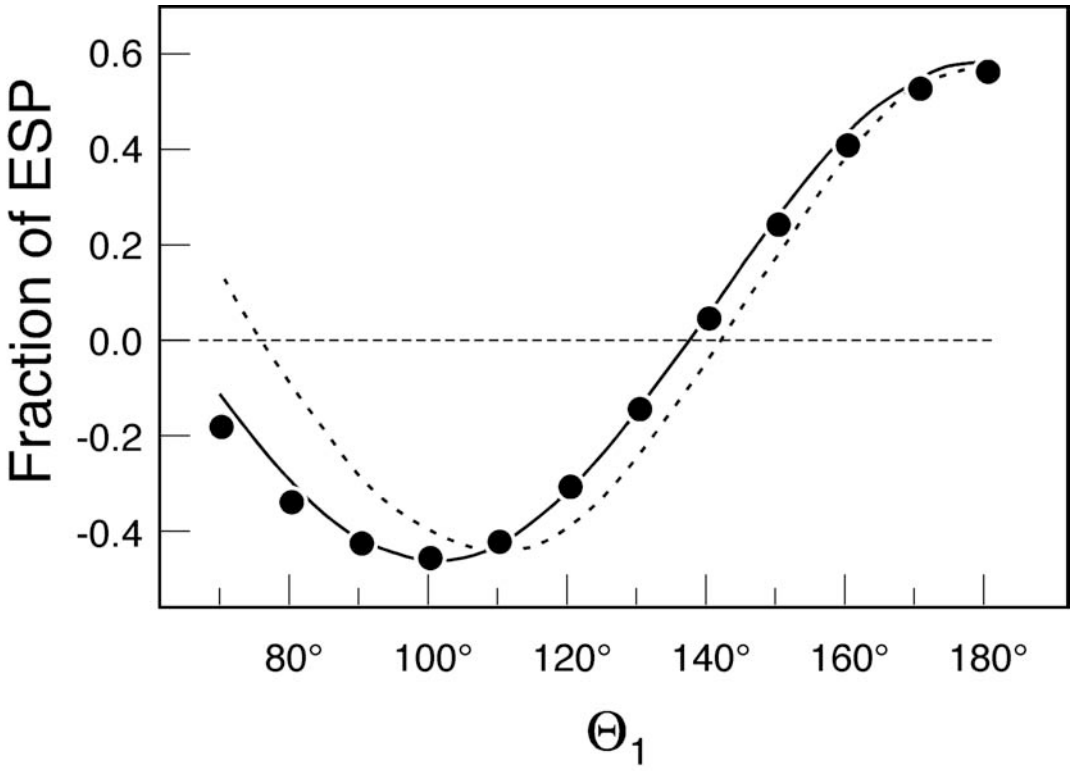


Fig. 4.5. Electrostatic potential of bromine as a function of Θ_1. The fraction of the electrostatic potential (ESP, solid circles) calculated for the bromine in a bromobenzene model calculated using MP2/aug-cc-pVDZ (Lu, Wang et al. 2009), and scaled to the difference in the maximum and minimum ESP values. Fraction of partial charge of the bromine fitted to Eq. 2 with n = 2.231 ± 0.008, and a ratio of B/A = 0.124 ± 0.005 (solid curve). The dotted curve represents the fraction partial charge predicted for n = 2.535 associated with the shape analysis (Fig. 4.4).

The effective Z_{Br} can be placed in the context of a general distance (r) dependent electrostatic potential energy function (V_{Elec}, analogous to a Coulombic potential) when paired with an acceptor with charge Z_A, as in Eq. 4 (where D is the dielectric constant and e is the charge of a proton). In this general form, we make no assumptions concerning the exponential power term n for the dependence of the potential energy on $1/r^n$.

$$V_{Elec} = \frac{Z_{Br} Z_A e^2}{D r^n} \qquad \text{Eq. 4}$$

4. Results

4.1 Potential energy function for bromine

The overall potential energy for the nonbonding interactions between a bromine donor and an acceptor atom (V_{Br}) is given as the sum of the vdW and electrostatic potentials. In order to compare this potential energy function to the PEP approach, we have rewritten the equations to be consistent with the AMBER force field (Eq. 5, where $\epsilon = \sqrt{\varepsilon_1 \varepsilon_2}$, with ϵ_1 and ϵ_2 being the energy contributions and R_o being the effective atomic radii of each of the two interacting atoms potential energy minimum), and will parameterize V_{Br} function against AMBER ff99 (Case, Cheatham et al. 2005) force field, because of its broad use in simulating macromolecular structures and its adaptability to several molecular mechanics modeling programs (Wang, Wolf et al. 2004) (including CHARMM (Brooks, Brooks et al. 2009) and GROMACS (Van Der Spoel, Lindahl et al. 2005)).

$$V_{Br} = V_{LJ} + V_{Elec} = \varepsilon \left[\left(\frac{\sum R_o}{r} \right)^{12} - \left(\frac{2 \sum R_o}{r} \right)^{6} \right] + \frac{Z_{Br} Z_A e^2}{D r^n} \qquad \text{Eq. 5}$$

For the bromine, which follows the σ-hole model, the total potential function is dependent on a = 180°- Θ_1 according to Eq. 6, where $<R_{o(Br)}>$ is now the average radius of the bromine at the energy minimum.

$$V_{Br} = \varepsilon\left[\left(\frac{R_{o(A)} + \left\langle R_{o(Br)}\right\rangle - \Delta R\cos\nu\alpha}{r}\right)^{12} - \left(\frac{2\left(R_{o(A)} + \left\langle R_{o(Br)}\right\rangle\right)}{r}\right)^{6}\right] + \frac{\left[A\cos(\nu\alpha) + B\right]Z_A e^2}{Dr^n}$$

Eq. 6

4.2 Parameters for bromine potential energy functions

In order to parameterize V_{Br} in Eq. 6, we start by defining a biomolecular system for which there are good estimates for the distance-angle dependence of the non-bonded interaction energy. We then use the structure-energy relationship for this X-bonded system calculated by QM methods to determine values for the parameters in Eq. 6. Finally, crystallographic and solution-state studies on the structure and energies of the interactions in this system are used to validate both the QM calculations and the parameterized potential energy equation. For this study, we selected the bromouracil-phosphate ($BrU\cdots PO_4^{-1}$) interaction from the competition assay in a four-way DNA junction (Gribble 2003; Voth 2007) as the biomolecular system—the energies of interaction in this system have been experimentally determined for two unique X-bond geometries (Br1J and Br2J, Fig 2). The overall strategy is to i) define a minimum model system for the relevant X-bond interaction in the DNA, ii) calculate MP2 energies for the interaction at various distances and angles, iii) derive parameters for the potential energy functions in Eq. 6 to be consistent with the AMBER ff99 (Case, Cheatham et al. 2005) force field for all non-halogen interactions to create a force field for the bromine halogen

bond, and iv) compare the calculated energies to the MP2 calculated energies and to those for the two experimental geometries.

The system for QM calculations was reduced from the complete DNA junction (with over 600 non-hydrogen atoms) to a simple bimolecular pair of an isolated BrU base interacting with a hypophosphite ($H_2PO_2^{-1}$) anion (Fig. 4.2). The $H_2PO_2^{-1}$ model for the phosphate group in the DNA system was selected because the MP2 calculated partial charge of the oxygens closely mirrored the partial charges in the AMBER force field (approximately -0.85e). Although the system is now reduced to only 17 atoms, it is still too computationally costly to perform the highest-level quantum calculations for the number of distances and angles required to fully define a geometry-energy relationship. Thus, we applied MP2 calculations using the 6-31G(d) basis set, with cyclohexane as the solvent (D = 2) and a counterpoise correction for the basis set superposition error (Paizs and Suhai 1998), on the model with the closest bromine to oxygen approach of the BrU•$H_2PO_2^{-1}$ pair varied from 2.4 to 3.4 Å and for Θ_1-angles from 90° to 180°. We note that the relative energies calculated for the two defined geometries of Br1J and Br2J were very similar among various basis sets and different values for D, even though the absolute energies differ.

The resulting MP2 calculations show an angle dependence for the energies at each distance which is sinusoidal with a cross-over from positive to negative energies at $\Theta_1 \approx$ 140° (Fig. 4.6a), mirroring the relationships for the electrostatic potential energies (Fig. 4.5). Indeed, at a distance r = 2.8 Å, the shape and cross-over from positive to negative energies is nearly an exact reflection of the angle dependence for the partial charge of the

bromine, consistent with a strong contribution of electrostatics on the interaction, as expressed by the σ-hole model.

We used the MP2 calculated energy-geometry relationship to determine the parameters for V_{Br}, as defined by Eq. 6. For this process, we first applied the AMBER type Lennard-Jones and electrostatic potentials to calculate the atom-to-atom interaction energies between the atoms in the BrU and $H_2PO_2^{-1}$ molecular pair, excluding those involving the bromine, for each geometry of the model. Since we do not include explicit solvent in the model, a distance dependent dielectric (Ferrara, Gohlke et al. 2004) of the form 4r was applied to all of the electrostatic potential energy calculations. These non-Br energies ($E_{non\text{-}Br}$) were subtracted from the MP2 calculated energies (E_{MP2}) to yield a residual that describes explicitly the energy of the bromine interacting with the atoms of the $H_2PO_2^{-1}$ component ($E_{Br} = E_{MP2} - E_{non\text{-}Br}$). This E_{Br} was subsequently used to determine the parameters for the V_{Br} potential function in Eq. 6.

The function for V_{Br} includes seven unique parameters: $\langle R_{vdW(Br)} \rangle$ and ΔR to describe the average van der Waals radius and perturbation to that radius, and e_{Br} to define the bromine contribution to the minimum van der Waals energy; A and B for the partial charge and n for the exponential dependence of the electrostatic potential on r; and ν for the period of the cosine function that describes the aspherical shape and charge distribution for the bromine (Table 4.1), which will be fitted against 30 MP2 calculated energy-geometry relationships. We considered the $\langle R_{vdW(Br)} \rangle$ from the shape analysis to be very robust and, therefore, converted this to an $\langle R_{o(Br)} \rangle = 2.04$ (recalling that the two radii are related by $\sqrt[6]{2}$) and fixed its value, leaving only six parameters to fit (we note that the $\langle R_{o(Br)} \rangle$ remained close to this value when allowed to float).

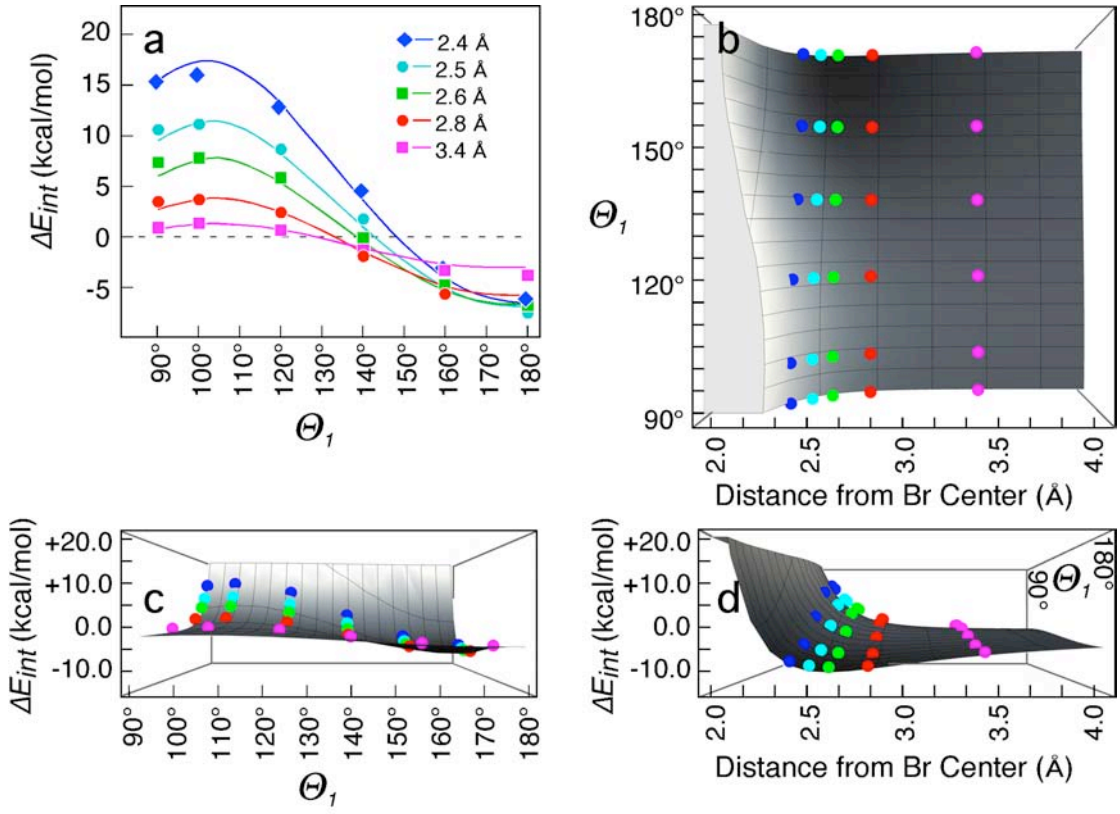


Fig. 4.6. Energies of $BrU\cdots H_2PO_2^{-1}$ interaction as a function of angle and distance. a. Results from MP2/6-31G(d) quantum calculations for the interacting system are shown as closed symbols. Curves are calculated energies from the V_{Br} function (Eq. 6) for the angles and distances associated with the QM calculations. The resulting QM calculated energies are mapped onto an energy landscape from the V_{Br} function (Eq. 6) of the ffBXB model (b, viewed onto the angle-distance surface; c, into the energy-angle plane and d, into the energy distance plane).

The MP2 energies were best fitted with n = 2.29 for the $1/r^n$ term of V_{Elec} in Eq. 4, suggesting that the electrostatic component is not a classic Coulombic potential (where n = 1). To confirm this, we calculated the MP2 energies for BrU interacting with a formally neutral $H_2PO(OH)$ at three angles (100°, 140°, and 180°) and five distances (2.4 Å, 2.5 Å, 2.6 Å, 2.8 Å, and 3.4 Å) for each angle, and compared them to the corresponding E_{Br} energies. We then subtracted the E_{Br} for the protonated neutral form from the anionic form of hypophosphite for each angle to determine the distance relationship for the effect of the charge on the X-bond energy. In this case, the average value for n was determined to be 2.4 ± 0.5. Thus, the value of $n \approx 2.5$ suggests that the electrostatic component falls between a charge-dipole (n = 2) and is analogous to a dipole-dipole (n = 3) interaction, a reasonable description of the polarization effects that define, in this case, the distribution of charge across the surface of the bromine relative to the angle of approach by the acceptor atom. The value $\nu = 2.31$ orients the p_x and p_y orbitals ~12° from perpendicular, and the ratio of A/B = 2.2 defines an overall slightly positive (+0.14e) charge across the surface of the bromine (Fig. 4.7), compared to the +0.047e from the MP2 calculations on BrU.

Table 4.1. Parameters for the angle-dependent ffBXB functions describing the anisotropic shape and electrostatic potential energy functions for the bromine of BrU (Eq. 6). Errors are indicated for fitted parameters.

Shape Parameters				
$<R_{o(Br)}>$	ΔR (Å)	e_{Br} (kcal/mol)		
2.04 Å	0.060 ± 0.022	0.019 ± 0.002		
Electrostatic Parameters				Angle
A	B	n		n
2.84 ± 0.82	1.53 ± 0.45	2.29 ± 0.29		2.31 ± 0.02

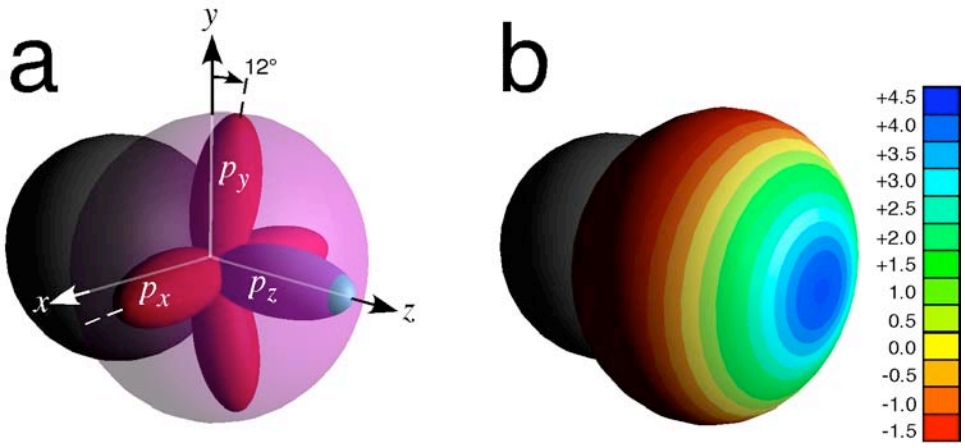


Fig. 4.7. Atomic structure of Br2 as modeled by V_{LJ} (Eq. 1) and Z_{Br} (Eq. 2) functions applying the parameters in Table 4.1. a. One bromine of a Br_2 molecule is shown with the outer p-orbitals (fully occupied p_x- and p_y-orbitals in red, p_z-empty orbital in blue) relative to their respective Cartesian reference axes. The polar flattening of the effective atomic radius along the z-axis is associated with the depopulated p_z orbital. b. Distribution of partial charge across the bromine surface, ranging from +5 to -2.0e.

4.3 Comparisons of QM and ffBXB calculated to experimental X-bonding energies

With the parameters in Table 4.1, the V_{Br} potentials could be combined with the standard AMBER potentials to define a complete force field for the bromine X-bond (the ffBXB), which can be applied to calculate the MM energies of the BrU···$H_2PO_2^{-1}$ interactions at all geometries (Fig. 4.6b - d). The resulting ffBXB energies fit the total MP2 calculated energies very well (R = 0.96). The QM and ffBXB approaches can be validated by comparing the calculated energies of interaction of BrU with the anionic hypophosphite or a dimethylphosphate (DMP^{-1}) acceptors, the latter being a more complete model for the phosphodiester linkage between nucleotides, to the experimental X-bonding energies of the Br1J and Br2J conformations determined in the DNA junction system. Both the QM and ffBXB model calculations, when applied to the BrU···$H_2PO_2^{-1}$ or BrU···DMP^{-1} models in the Br1J or Br2J junction geometries resulted in interaction energies that are well within the errors of the energies determined experimentally in the crystal system (Voth 2007) and in solution (Gribble 2003) (Table 4.2). Thus, the resulting ffBXB functions replicate both the MP2 and experimental X-bonding energies of the DNA system used to derive the model.

Table 4.2. Comparison of experimental and calculated enthalpies (kcal/mol) for bromine X-bonds in the Br1J and Br2J conformations of DNA junctions (Voth 2007). Interaction enthalpies for X- minus H-bond (ΔH_{X-H}) determined from a crystallographic competition assay (Voth 2007) and by differential scanning calorimetry in solution (Gribble 2003) are compared to X-bond energies from QM and ffBXB calculations applied to X-bonds of the BrU···$H_2PO_2^{-1}$ or BrU···Dimethylphosphate (DMP^{-1}) model systems.

Form	Experimental ΔH_{X-H}		Calculated Energies ($H_2PO_2^{-1}$/DMP^{-1})	
	Crystal Assay	Calorimetric	QM	ffBXB
Br1J	-2.0 ± 0.5	-	-1.44/-1.53	-1.97/-2.47
Br2J	-	-3.5 ± 1.3	-3.02/-3.06	-2.86/-4.63

4.4 Potential energy landscapes for bromine interactions

The X-bond in the DNA junction is representative of those seen in other nucleic acid systems, including multistranded DNAs (Sunami, Kondo et al. 2004; Sunami, Kondo et al. 2004) and RNA (Gilbert, Reyes et al. 2009), with the bromine interacting with a single, formal negatively charged oxygen acceptor. For protein systems, this would also serve as an approximate model for halogen interactions with the charged oxygen acceptors of aspartate and glutamate side chains. It is useful, therefore to derive a more general map for bromine interacting with a formally charged anionic acceptor. In this case, the ffBXB function shows the polar flattening associated with the V_{vdW} potential function (Fig. 4.8a) and the anisotropic charge distribution of the V_{elec} potential function (Fig. 4.8b).

The resulting total V_{Br} potential predicts a relatively deep potential energy well (-10.8 kcal/mol) at ~2.5 Å from the Br center and aligned along the σ-bond axis (Θ_1 = 180°), as predicted (Fig. 4.8c). The depth of this well for an isolated anionic oxygen is approximately 50% larger and placed ~1.5 Å shorter than that of the complete BrU···$H_2PO_4^{-1}$ pair, as calculated by the QM and ffBXB approaches. The differences can be attributed to non-covalent interactions between the additional atoms of the molecular system. A negative stabilizing potential is seen to extend to $\leq 130°$, indicating that, although directional, Br X-bonds are stabilizing over a $\geq 90°$ range (±45° from linear). The zero point energy is at ~2.2 Å, while the stabilizing potential to -1 kcal/mol extends to >5 Å from the halogen center.

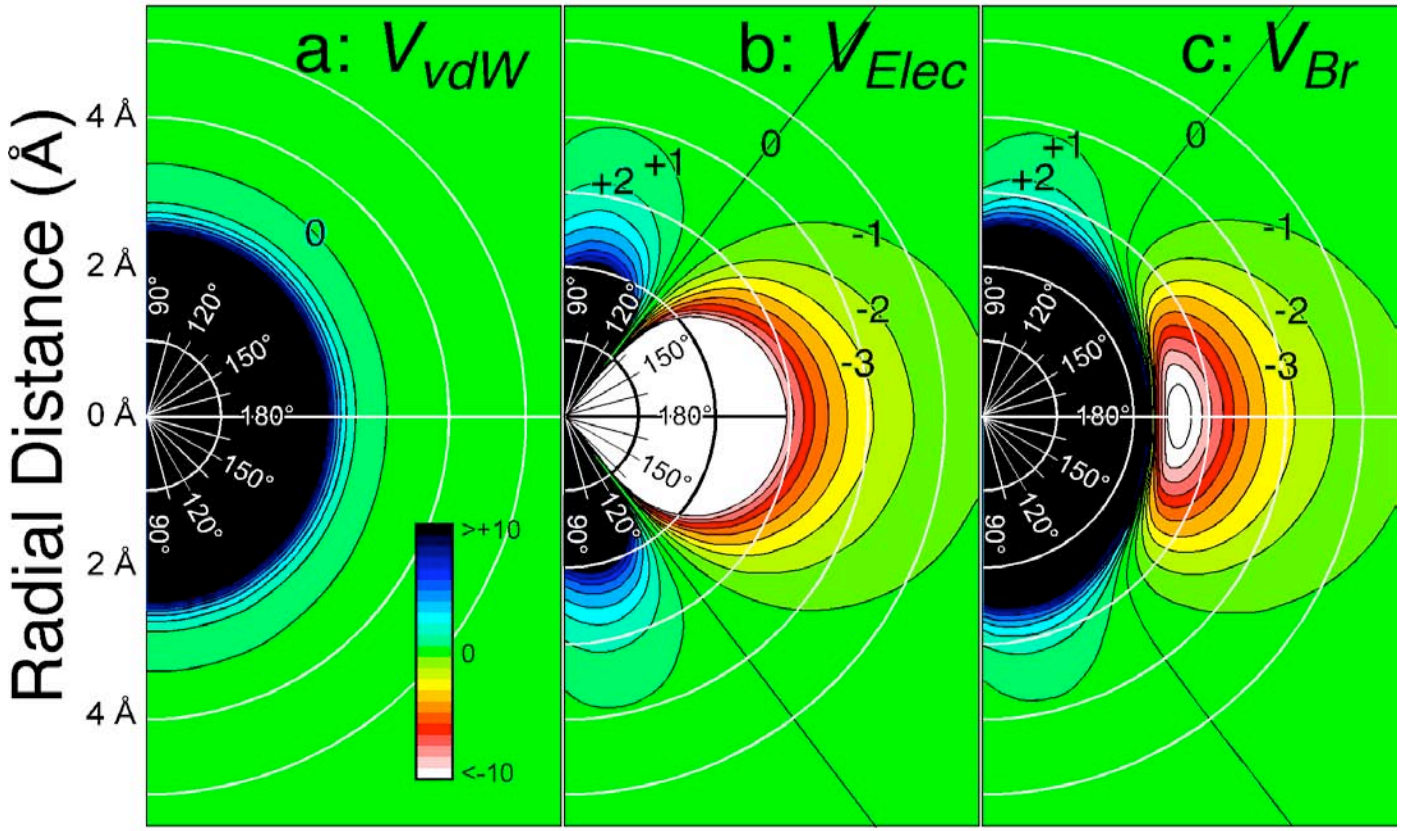


Fig. 4.8. Potential energy maps calculated form the ffBXB of bromine interactions with formally charged oxygen. The V_{vdW} (a) from Eq. 1, V_{elec} (b) from Eq. 2, and total V_{Br} (c) from Eq. 6 are mapped onto polar plots, with concentric circles defining 1 Å radial distances from the bromine center, and angles relative to the C-Br -bond labeled. The formally anionic O^{-1} is assigned an effective partial charge of -0.85e to be consistent with the AMBER ff99 force field.

Formally charged oxygens represent only one type of X-bond acceptor seen in biological systems. The majority of biological X-bonds are to the carbonyl oxygens of the peptide bond in proteins (Auffinger, Hays et al. 2004; Voth 2009; Parisini, Metrangolo et al. 2011), although the oxygens of alcoholic and acidic side chains and sulfurs of methionine and cysteine residues can also serve as X-bond acceptors (Liu, Baase et al. 2009; Vallejos, Auffinger et al. 2010). In addition, halogens are seen to be amphoteric, capable of serving as hydrogen bonds acceptors. To determine how the ffBXB can be applied to other types of interactions, we compare the potential maps for charged oxygens (Fig. 4.8c) to formally uncharged oxygens (partial charge -0.49e) and to a hydrogen (+0.5e) that can serve as an H-bond donor (Fig. 4.9b). In each case, the calculations used the R_{vdW}, partial charge, and e values for the acceptor atom, as defined by AMBER ff99 (Case, Cheatham et al. 2005).

The ffBXB potential map for a bromine to neutral oxygen interaction shows an energy minimum of -5.4 kcal/mol at 2.7 Å. The stabilization energy is approximately half of that for an anionic oxygen, consistent with the ~2-fold difference in the AMBER assigned charge to the acceptor. Furthermore, the zero point distance is similar for both oxygens, while the optimum distance of interaction is only 5% longer for the neutral as opposed to the anionic X-bond acceptor and the -1 kcal/mol distance is contracted to <5 Å.

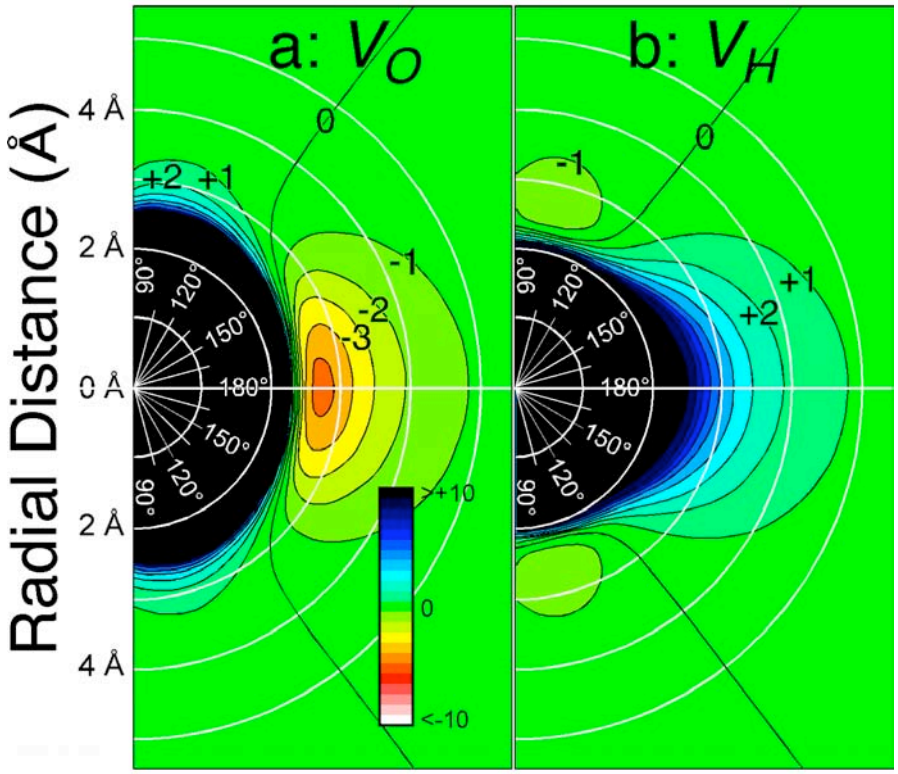


Fig. 4.9. Energy landscapes calculated from the ffBXB functions for Br interactions with neutral oxygen (a) and hydrogen atoms (b). Energies are in kcal/mol. Although both the oxygen and hydrogen atoms are formally neutral, they were assigned partial charges of -0.49e and +0.5e, respectively, consistent with the AMBER ff99 force field.

The ffBXB potential energy map for the Br···H interaction shows that the bromine can also serve as an H-bond acceptor in a direction approximately perpendicular to the X-bonding potential, as seen in crystal structures (Lu, Wang et al. 2009). The depth of the minimum energy well is calculated to be -1.7 kcal/mol at a distance of ~2.5 Å from the bromine center and ~11° from perpendicular as a result of most negative potential tipped slightly from Θ_1 of 90°. The electrostatic potential is sufficiently strong to pull the hydrogen to a distance ~0.4 Å shorter than the sum of the R_{vdW} for the bromine and hydrogen; therefore, this interaction can be classified as a classic H-bond.

5. Discussion

A set of potential energy functions is presented here that describe the aspherical shape and anisotropic distribution of electrostatic charge across the surface of a bromine substituent in molecular compounds. The functions very accurately reproduce the experimental and QM calculated geometry-energy relationships of various interactions with halogens, including X-bonds to charged, uncharged, and aromatic acceptors, and H-bonds to electropositive donors. The hallmark of the ffBXB function is that it is derived from fundamental physicochemical properties of the halogen. The electrostatic function (V_{elec}) is clearly more akin to a dipole-dipole interaction in terms of the angular and distance dependence than to the standard Coulombic potential. This restricts both the range of angles and distances at which these interactions extend when compared to a true point charge electrostatic interaction.

5.1 Application to PEP approach to modeling halogen interactions

The alternative PEP approach to modeling X-bonding potentials is to add an extra positive charge with an R_{vdW} = 0 to mimic the electropositive crown resulting from polarization of the halogen. Previous studies that apply this model in a molecular mechanics approach have been successful in generally modelling interactions between X-bond donors and acceptors that correlate well with measures of affinity in protein-ligand complexes (Sponer, Riley et al. 2008; Ibrahim 2011), with distances between the interacting atoms within ~0.3 Å of the corresponding distances observe in their crystal structures. We should note that the PEP models derived here are very simplistic, and primarily serve to compare and contrast this model to the ffBXB functions for the model systems in this particular study, and should not be construed as being generally applicable to other systems (we leave the development of a more general model to those who are more invested in this approach).

To develop a PEP model, we define a partial negative charge (ZBr) centered at the bromine atom and an added positive charge (Z_ψ) at some distance (rBr-y) to model the σ-hole resulting from polarization of the halogen. Values for these three parameters were determined by fitting the energy-distance profile for the $Br\cdots O^{-1}$ interaction at Θ_1 = 180° (Fig. 4.8c) using standard Coulombic (with a distance dependent dielectric 4r, applying partial charges to non-halogen atoms as they were in the ffBXB model) and van der Waals potentials.

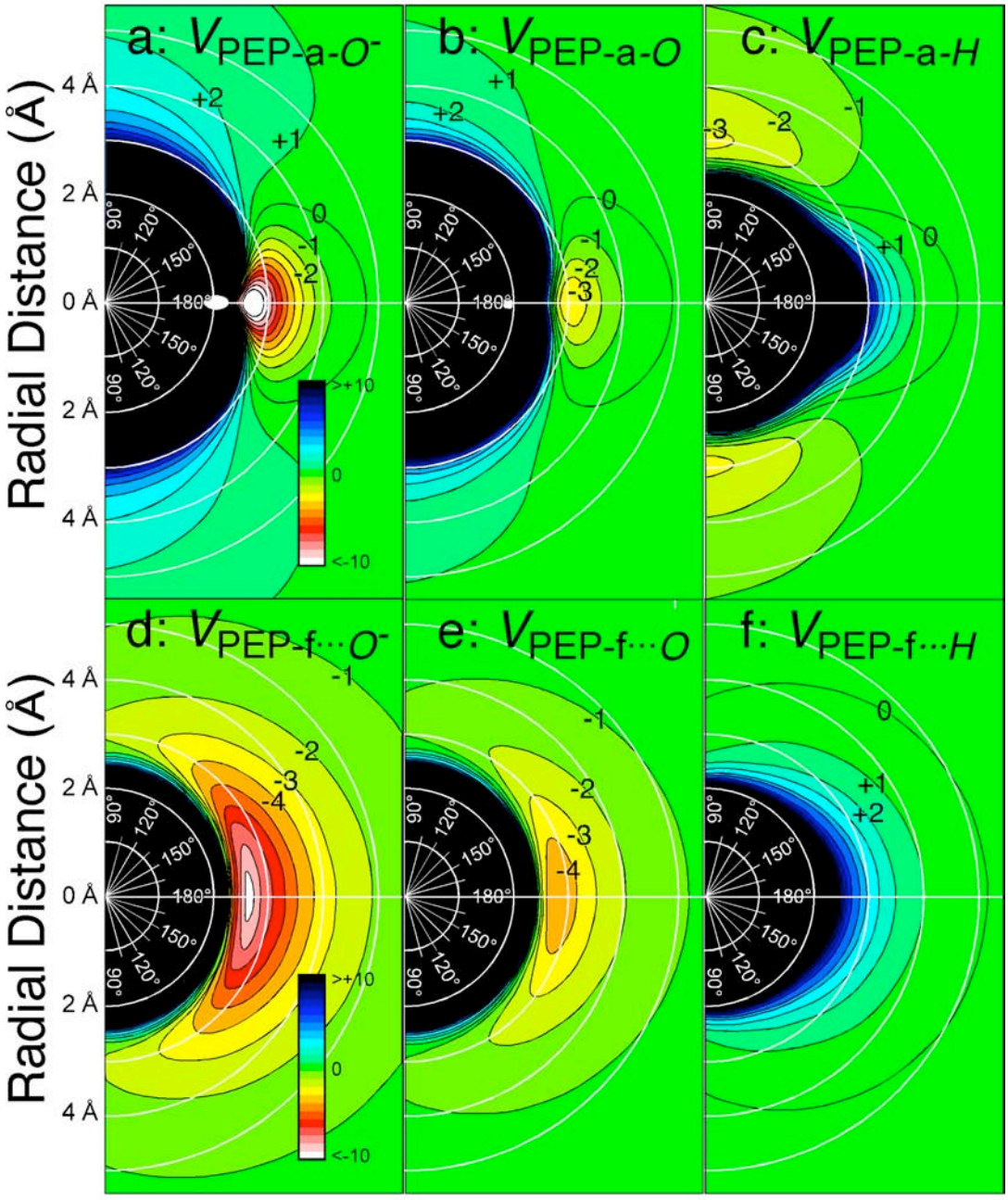


Fig. 4.10. Potential energy landscape applying PEP models for the polarized bromine, applying the parameters from Table 4.3 to the AMBER ff99 force field. a-c. Potential energy maps for bromine to anionic oxygen acceptor ($Br\cdots O^{-1}$), bromine to neutral oxygen ($Br\cdots O$), and bromine to neutral hydrogen ($Br\cdots H$) interactions using the PEP-a model (using the standard R_o and e_{Br} from the AMBER force field), with partial charges of -0.85 for the anionic oxygen, -0.49 for the formally neutral oxygen, and +0.5 for the neutral hydrogen as an H-bond donor. Energies are in kcal/mol. Concentric circles indicate the radial distances fro the center of the bromine atom, with angles indicating the angle of approach of the acceptor atoms towards the C-Br bond. d-f. The same as a-c, except using the PEP-f model with R_o and e_{Br} from the ffBXB model.

With the steric R_o and e_{Br} parameters set to the standard values from AMBER ff99 (Case, Cheatham et al. 2005), the resulting parameters (PEP-a, Table 4.3) define the negative charge of the bromine to be overall 2.5-times that of the positive charge associated with the σ-hole, with the added charge at the standard r_{vdW} of the bromine (~1.85 Å from the Br center). The charges are thus those that best fit the QM energy profiles for this particular system.

The energy landscape calculated for the Br···O^{-1} pair (Fig. 4.10a) is qualitatively very similar to that calculated using the ffBXB functions (Fig. 4.8c) in terms of the depth of the energy well (-12 kcal/mol) at the optimum distance of interaction (~2.7 Å). Thus, the PEP-a approach models the general features of the ffBXB energy landscape for the Br···O^{-1} interaction reasonable well (at Θ_1 = 180°, the PEP-a and ffBXB energies are correlated by an R = 0.932), although the energy well is deeper and narrower than that of the ffBXB model.

When applied to the Br1J and Br2J conformations of the BrU···$H_2PO_2^{-1}$ or BrU···DMP^{-1} interacting pairs, however, the PEP-a model predicts positive energies of interaction for both conformations (Table 4.3), with the shorter Br2J being significantly more positive than Br1J. The energies of interaction for the reference anionic oxygens that are aligned nearly linearly with the C-Br bond are negative for both conformations, but the steric clash of the bromine with all the remaining atoms make the overall energies positive. This suggests that the standard R_o and e_{Br} values in the AMBER force field as applied here do not properly model the van der Waals interactions of the bromine in this more complex system.

Table 4.3. Parameters for PEP approach to model ffBXB potential energy maps for Br···O^{-1} interactions, based on the R_o and e_{Br} from the AMBER ff99 force field (PEP-a model) and from the ffBXB parameters (PEP-f model). Interaction energies for X-bonds in the BrU···$H_2PO_2^{-1}$ or BrU···Dimethylphosphate (DMP^{-1}) model systems, with a dielectric constant D = 4r, are compared between the two models.

Model	Steric Parameters		Electrostatic Parameters		
	R_o (Å)	e_{Br} (kcal/mol)	Z_{Br} (e)	Z_ψ (e)	$r_{Br\text{-}\psi}$(Å)
PEP-a	2.22	0.32	-1.05 ± 0.26	+0.39 ± 0.04	1.856 ± 0.15
PEP-f	2.04	0.019	-0.17 ± 0.31	+0.54 ± 0.21	1.01 ± 0.13

Model	X-Bond Energies ($H_2PO_2^{-1}$/DMP)	
	Br1J	Br2J
PEP-a	1.57/3.88	5.66/6.12
PEP-f	-3.40/-4.96	-4.15/-5.99

To test this possibility, we redetermined a set of PEP parameters using the R_o and e_{Br} values derived from the ffBXB approach (PEP-f), resulting in electrostatic terms that help to counterbalance the reduced steric interactions. These parameters correlate very well with the ffBXB energies at $\Theta_1 = 180°$ (E_{min} = -9.81 kcal/mol at 2.6 Å, R = 0.999 for the fit). Applying the PEP-f model to the Br1J and Br2J models with the $H_2PO_2^{-1}$ and DMP^{-1} acceptors show the energies of interaction are negative for both conformations, with the shorter X-bond being more favorable than the longer interaction, as expected. Results from the PEP-f and ffBXB analyses suggest that the size of the halogen and energy terms for the van der Waals interaction need to be reduced relative to the standard AMBER definitions in order to properly describe the interactions in the experiment X-bonded DNA junction system.

A comparison of the overall landscape for the Br···O^{-1} interaction (Fig. 4.10a and d) show that although the depth and positions of the energy well for the two PEP models are similar, the well for the PEP-a model is very narrow (with the -1 kcal/mol contour extending from 180° to 150°) while the PEP-f well is very broad (having a negative energy completely encompassing the bromine atom). This same trend is seen for the interaction to the formally neutral oxygen (Br···O, Fig. 4.10b and e). The extension of which can be attributed to the very small negative charge assigned to the bromine in the PEP-f model. The result is that for the Br···H interaction, the PEP-a model predicts energy wells of -3.5 kcal/mol at Θ_1= 90°, while no favorable interactions are predicted by the PEP-f model (Fig. 4.10 c and f). Thus, the PEP models derived here seem to trade accuracy in X-bonding behavior at $\Theta_1 \approx 180°$ in the PEP-f model for potential to form H-bonds at Θ_1= 90° in the PEP-a model. There probably exists a model between PEP-a and

PEP-f that could account for both, but it is not clear how such a model can be derived using the current experimental system.

5.2 Comparison of ffBXB and PEP models for bromobenzene to acetone interactions

At this point, we can ask how the two models compare in their ability to model a bromine X-bond in a molecular system in which they were not initially optimized for. Since most biological X-bonds are to the carbonyl oxygens of the peptide backbones in proteins, this comparison can be made to the energies of interaction (E_{Int}) for a bromobenzene to acetone model system from high-level MP2 calculations (Fig. 4.11). These energies have been shown by Riley, et al. (Riley, Murray et al. 2011) to be tunable by varying the electron-withdrawing capability of the benzene ring by adding fluorines at various positions relative to the bromine.

If the quality of the two empirical models were judged solely on their abilities to reproduce the MP2 energies of the bromobenzene-acetone interaction, one would conclude that the PEP-a approach is better than either the ffBXB or PEP-f models (Table 4.4). The PEP-a calculated E_{Int} using the scaled partial charges from Table 4.3 (to yield an overall neutral donor molecule) is within 0.16 kcal/mol and positioned within 0.22 Å of the minimum of the MP2 energy curve—these deviations are similar to those obtained by Ibrahim for protein-ligand complexes. In contrast, although the E_{Int} from the ffBXB model falls well within the MP2 energy curves for bromobenzene and its various fluorinated derivatives, it does not exactly match any single MP2 curve. The depth of the energy well falls between those of the o-difluorobromobenzene and pentafluorobenzene energies (Table 4.4) at a distance ~0.1 Å shorter than that for pentafluorobenzene.

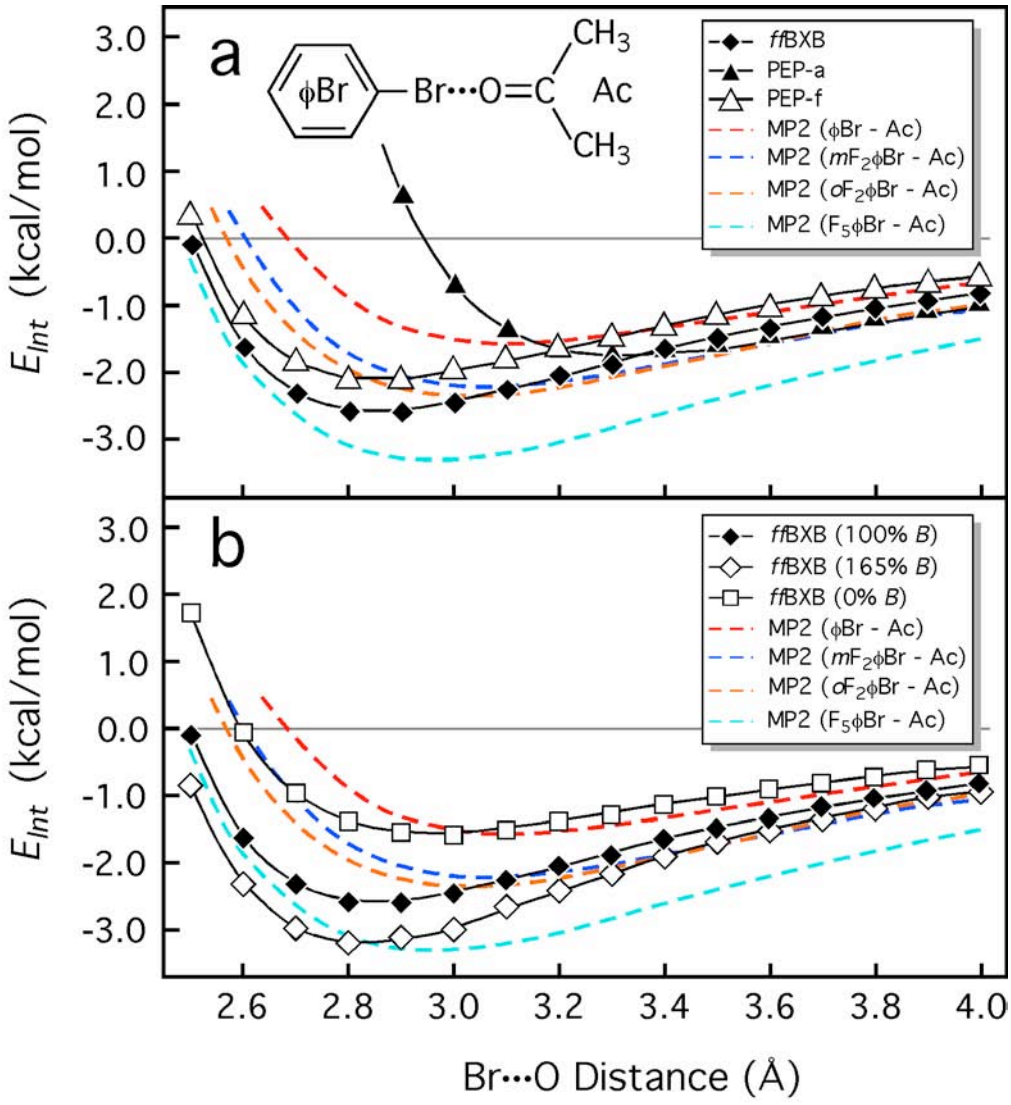


Fig. 4.11. Comparison of energies for the interaction of acetone with bromobenzene and its fluorinated derivatives. a. Comparison of MP2 to ffBXB and PEP-a and PEP-f models calculated energies of interaction (E_{Int}). Energies as a function of the carbonyl oxygen to bromine distance (at a 180° angle of approach) for the molecular interaction of acetone to bromobenzene (fBr, dashed red line), meta-difluorobromobenzene (m-F_2fBr, dashed blue line), ortho-difluorobromobenzene (o-F_2fBr, dashed orange line), and pentafluorobromobenzene (F_5fBr, dashed cyan line) are redrawn from Riley, et al. (Egner, Kratzschmar et al.). These energies are compared to those calculated from the ffBXB model (at 0.1 Å intervals in the O···Br distance) using the parameters from Table 4.1 (solid diamonds), and to the PEP-a (solid triangles) and PEP-f (solid diamonds) models. b. Interaction energies from MP2 calculations compared to those from the ffBXB model with the electrostatic B term scaled by 100% (solid diamonds), 165% (open diamonds) or by 0% (open squares).

Table 4.4. Interaction energies calculated for acetone with bromobenzene (fBr) and its fluorinated variants as X-bond donors. The minimum energies of interaction (E_{min}) and the Br···O distances for the E_{min} (R_{min}) for acetone interacting with fBr, m-difluoro-fBr (m-F_2-fBr), o-difluoro-fBr (o-F_2-fBr), and pentafluor-fBr (F_5-fBr) from Riley, et al., are compared to those calculated by the ffBXB and the PEP methods. In addition, the overall partial charges of the bromines (Z_{Br}) in each of the fluorinated variants were estimated from MP2, ffBXB, or PEP approaches. Finally, the substituent effects of the each donor compound (as reflected in the summed Hammett s constants for the fluorine substituents (∑s)) are compared, applying reported values for the meta- and para-positions (McDaniel and Brown 1958), estimated for the ortho-position from the Z_{Br} values (in parentheses), or from E_{min} values (in brackets), see text for description.

X-Bond Donor	E_{min} (kcal/mol)	R_{min} (Br···O)	Z_{Br}	ås
fBr	-1.58	3.10 Å	-0.0412e	0.0
m-F_2-fBr	-2.22	3.05 Å	0.0012e	0.68
o-F_2-fBr	-2.37	3.00 Å	0.0255e	(0.93)
F_5-fBr	-3.34	2.96 Å	0.0675e	(1.67)
fBr (ffBXB, 100% B)	-2.75	2.85 Å	-	[1.36]
fBr (ffBXB, 0% B)	-1.58	2.97 Å	-	[0.137]
fBr (ffBXB, 165% B)	-3.34	2.80 Å	-	[1.98]
fBr (PEP-a)	-1.74	3.32 Å	-	[0.31]
fBr (PEP-f)	-1.74	3.32 Å	-	[0.31]
m-F_2-p-F-fBr	-	-	0.0043e	0.74
Br-Uracil	-	-	0.0470e	(1.44)
o-F_2-m-F_2-fBr	-	-	0.0450e	(1.40)

We must recognize, however, that the ffBXB and PEP-f parameters in Tables 2 and 3 were derived not for bromine attached to a benzene, but to a uracil base, which is apparently much more electron-withdrawing than benzene. This is not surprising, as bromine attached to a heterocyclic ring (as in bromopyrimidine, a model that is more analogous to uracil) has been shown to form a stronger X-bond to acetone than bromobenzene (Riley, Murray et al. 2009). By comparing the ffBXB energy to those of the various fluorinated X-bond donors, we can estimate that the uracil base is equivalent in electron-withdrawing potential as a tetrafluorinated benzene.

To quantify the effective electron withdrawing ability of the uracil base in our initial model system, we first calculated the effective charge of the bromine (Z_{Br}) in bromobenzene, in 3,5-difluorobromobenzene, and in 3,4,5-trifluorobromobenzene, and related these values to the standard Hammett σ coefficient (Hammett 1937) as the measure of the inductive effects for fluorine substituents at the meta- and para-positions. As expected, increasing the electron-withdrawing property of the benzene ring with added fluorines exaggerated the polarization and, consequently, the effective overall positive charge of the bromine atom (Table 4.4). The resulting linear relationship $\sum\sigma = 16.5Z_{Br} + 0.684$ ($R^2 = 0.999$, where $\sum\sigma$ is the sum of the s constants) allowed us to estimate a $\sigma = 0.46$ for an ortho-fluorine substituent. The inductive effects are linearly related to the energies of interaction of the fluorinated variants of bromobenzene with acetone according to the equation $\sum\sigma = -1.05E_{Int} - 1.52$ ($R^2 = 0.985$). By comparison, the Z_{Br} from MP2 calculations on BrU estimates $\sum\sigma = 1.44$, while the E_{Int} calculated for the bromobenzene-acetone pair from the ffBXB parameters is equivalent to an effective $\sum\sigma = 1.36$. The uracil base, therefore, apparently has an equivalent electron-withdrawing

capacity as a tetrafluorinated benzene, and is accurately modeled by the ffBXB parameters.

The ffBXB and energies can be readily tuned to that of the pentafluorobenzene by scaling the electrostatic B term of the model by 165% and to that of the unfluorinated bromobenzene by scaling this term to 0 (Fig. 4.11b). With B = 0, we see that the ffBXB model fares very well in comparison with the PEP-a approach, with a Br···O distance for the E_{Int} optimum that is 0.13 Å shorter than the MP2 curve. The E_{Int} for each of the scaled B-terms can be linearly related to the magnitude of the electron-withdrawing capacity of each of the corresponding fluorinated derivatives by as $B = 1.38\sum\sigma - 0.173$ ($R = 0.9995$). The effect of substitutents around the benzene ring on the polarizability of the bromine can, therefore, be readily modeled through a standard measure of the electron withdrawing ability of the molecule that the halogen is attached to. An analogous analysis allows the Z_ψ term of the PEP-f model to be scaled relative to the $\sum\sigma$ parameters according to the relationship $Z_\psi = 0.156\sum\sigma + 0.385$ ($R = 0.9997$), resulting in curves similar to those of the ffBXB relationships.

The ffBXB optimum energies consistently fall at distances 0.13 to 0.16 Å shorter than of those from MP2 calculations. One can argue that the reduced polarizability of the bromine in the bromobenzene model should result in a slightly larger van der Waals radius for the halogen, particularly in the direction of the σ-hole. In this case, increasing $\langle R_{vdW(Br)} \rangle$ by 0.1 Å would place all of the energy minima to within 0.05Å of the MP2 calculated curves for the bromobenzene···acetone pair.

6. Conclusions

The ffBXB model now provides a complete description of the geometric constraints that allow us to explore the great potential of X-bonds as well as H-bonds as molecular tools for the design and synthesis of new halogenated therapeutic agents and biomolecular materials. The resulting set of potential energy functions very accurately model the structure-energy relationships for bromine to anionic oxygen X-bonds calculated from QM analyses and observed experimentally in a DNA junction biomolecular system, and can be extended to predict the X-bonding and perpendicular H-bonding potential of halogens. The ffBXB functions can be directly incorporated into current molecular mechanics force fields, in much the same manner that angular dependent H-bonds have been incorporated (Lii and Allinger 2008), or they can be used to derive parameters for a more conventional PEP approach to simulating the polarization effects of halogens.

Both the ffBXB and PEP approaches can, to varying degrees be applied to simulate bromine halogen interactions in biomolecules other than the DNA system used in the current study, including X-bonds to other types of acceptors and H-bonds. The simulations of X-bonding to the carbonyl oxygen of acetone mimics the energies and geometries observed in protein-ligand systems and, therefore, demonstrate the utility of such empirical force fields for inhibitor design. In this case, we show that the electrostatic potentials for the interactions can be tuned by considering the electron withdrawing potential of the molecule that is halogenated. We can, thus, propose that in designing a new halogenated version of a lead inhibitor, the electrostatic parameters can be initially defined according to standard measures of this property, which would result in much

more accurate energies of interactions and, consequently, more accurate dissociation constants specific for that particular system. In the ffBXB model, this is achieved by a simple and straightforward scaling of the electrostatic B term. For the PEP approach, however, it is clear that the standard descriptors of the size of the bromine as currently implemented in, for example, the AMBER force field may not properly model the effects on the associated steric and dispersive forces.

Obviously, these potential energy functions do not explicitly treat either entropy or solvent effects on the energies of molecular halogen interactions. For example, bromines are known to be hydrophobic substituents, which may initially appear to be contradictory to the strong electrostatic contributions to the H- and X-bonding interactions the halogen is expected to make with water; however, both these interactions are predicted by the energy functions to be highly directional, which would limit the configurational space available and, thus the entropy of each interacting water molecule. According to the Lum, Chandler, and Weeks model (Lum, Chandler et al. 1999), this reduced solvent entropy would contribute to the hydrophobicity of halogen substituents. Once incorporated into current force fields commonly used for molecular simulations and molecular docking, we expect that both conformational and solvent entropy effects can be modeled using established free-energy methods (Steinbrecher and Labahn ; Huang and Jacobson 2007; Boyce, Mobley et al. 2009), thereby providing a test for this hypothesis.

Finally, although the ffBXB functions were derived specifically for bromine in the current study, the results provide a strategy to parameterized the functions for all other halogens (chlorine and iodine in particular), and, potentially, for other Group V and VI atoms that show significant σ-hole polarization (Clark, Hennemann et al. 2007; Murray, Lane et al. 2007).

7. Acknowledgments

This work was supported in part by funds from Colorado State University and a grant from the National Science Foundation (CHE-1152494).

8. References

Armstrong, J. (2012). General, Organic, and Biochemistry: An Applied Approach. Belmont, CA, Brooks/Cole.

Auffinger, P., F. A. Hays, et al. (2004). "Halogen bonds in biological molecules." Proc. Natl. Acad. Sci., USA 101(48): 16789-16794.

Bondi, A. (1964). "van der Waals volumes and radii." J Phys Chem A 68(3): 441-451.

Boyce, S. E., D. L. Mobley, et al. (2009). "Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site." J Mol Biol 394(4): 747-763.

Brooks, B. R., C. L. Brooks, 3rd, et al. (2009). "CHARMM: the biomolecular simulation program." J Comput Chem 30(10): 1545-1614.

Case, D. A., T. E. Cheatham, 3rd, et al. (2005). "The Amber biomolecular simulation programs." J Comput Chem 26(16): 1668-1688.

Clark, T., M. Hennemann, et al. (2007). "Halogen bonding: the sigma-hole. Proceedings of "Modeling interactions in biomolecules II", Prague, September 5th-9th, 2005." J. Mol. Model. 13(2): 291-296.

Dobes, P., J. Rezac, et al. (2011). "Semiempirical quantum mechanical method PM6-DH2X describes the geometry and energetics of CK2-inhibitor complexes involving halogen bonds well, while the empirical potential fails." J Phys Chem B 115(26): 8581-8589.

Egner, U., J. Kratzschmar, et al. (2005). "The target discovery process." Chembiochem 6(3): 468-479.

Ferrara, P., H. Gohlke, et al. (2004). "Assessing scoring functions for protein-ligand interactions." J. Med. Chem. 47(12): 3032-3047.

Gilbert, S. D., F. E. Reyes, et al. (2009). "Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs." Structure 17(6): 857-868.

Gribble, G. W. (2003). "The diversity of naturally produced organohalogens." Chemosphere 52(2): 289-297.

Hammett, L. P. (1937). "The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives." J. Am. Chem. Soc. 59(1): 96-103.

Hassel, O. (1972). Structural aspects of interatomic charge-transfer bonding. Nobel lectures, Chemistry 1963-1970. Amsterdam, Elsevier publishing company.

Huang, N. and M. P. Jacobson (2007). "Physics-based methods for studying protein-ligand interactions." Curr Opin Drug Discov Devel 10(3): 325-331.

Ibrahim, M. A. (2012). "AMBER empirical potential describes the geometry and energy of noncovalent halogen interactions better than advanced semiempirical quantum mechanical method PM6-DH2X." J Phys Chem B 116(11): 3659-3669.

Ibrahim, M. A. A. (2011). "Molecular Mechanical Study of Halogen Bonding in Drug Discovery." JOURNAL OF COMPUTATIONAL CHEMISTRY 32(12): 2564-2574.

Ibrahim, M. A. A. (2011). "Performance Assessment of Semiempirical Molecular Orbital Methods in Describing Halogen Bonding: Quantum Mechanical and Quantum

Mechanical/Molecular Mechanical-Molecular Dynamics Study " JOURNAL OF CHEMICAL INFORMATION AND MODELING 51: 2549-2559.
Lam, P. Y. S., C. G. Clark, et al. (2009). Structure-based drug design utilizing halogen bonding: Factor Xa inhibitors. The 238th ACS National Meeting. P. Metrangolo, and Resnati, G. Washington, DC, ACS. ORGN: 58.
Lii, J. H. and N. L. Allinger (2008). "The Important Role of Lone-Pairs in Force Field (MM4) Calculations on Hydrogen Bonding in Alcohols." J. Phys. Chem. A 112(46): 11903-11913.
Liu, L., W. A. Baase, et al. (2009). "Halogenated benzenes bound within a non-polar cavity in T4 lysozyme provide examples of I...S and I...Se halogen-bonding." J Mol Biol 385(2): 595-605.
Lommerse, J. P. M., A. J. Stone, et al. (1996). "The nature and geometry of intramolecular interactions between halogens and oxygen or nitrogen." J. Am. Chem. Soc. 118: 3108-3116.
Lu, Y., T. Shi, et al. (2009). "Halogen bonding--a novel interaction for rational drug design?" J. Med. Chem. 52(9): 2854-2862.
Lu, Y., Y. Wang, et al. (2009). "C-X...H contacts in biomolecular systems: how they contribute to protein-ligand binding affinity." J. Phys. Chem. B 113(37): 12615-12621.
Lum, K., D. Chandler, et al. (1999). "Hydrophobicity at Small and Large Length Scales." J Phys Chem B 103(22): 4570-4577.
Matter, H., M. Nazare, et al. (2009). "Evidence for C-Cl/C-Br...pi interactions as an important contribution to protein-ligand binding affinity." Angew Chem Int Ed Engl 48(16): 2911-2916.
McDaniel, D. H. and H. C. Brown (1958). "An Extended Table of Hammett Substituent Constants Based on the Ionization of Substituted Benzoic Acids." J. Org. Chem. 23(3): 420-427.
Metrangolo, P., Neukirch, H., Pilati, T., and Resnati, G. (2005). "Halogen bonding based recognition processes: A world parallel to hydrogen bonding." Acct. Chem. Res. 38(5): 386-395.
Metrangolo, P., Resnati, G. (2008). "Chemistry: Halogen versus hydrogen." Science 321: 918-919.
Møller, C. and M. S. Plesset (1934). "Note on an Approximation Treatment for Many-Electron Systems." Phys Rev 46(7): 618-622.
Murray, J. S., P. Lane, et al. (2007). "Sigma-hole bonding: molecules containing group VI atoms." Journal of Molecular Modeling 13(10): 1033-1038.
Muzet, N., B. Guillot, et al. (2003). "Electrostatic complementarity in an aldose reductase complex from ultra-high-resolution crystallography and first-principles calculations." Proc. Natl. Acad. Sci., USA 100(15): 8742-8747.
Nyburg, S. (1979). "Polar flattening: Non-spherical effective shapes of atoms in crystals." Acta Cryst A35: 641-645.
Ouvrard, C., J. Y. Le Questel, et al. (2003). "Halogen-bond geometry: a crystallographic data-base investigation of dihalogen complexes." Acta Cryst. B59: 512-526.
Paizs, B. and S. Suhai (1998). "Comparative study of BSSE correction methods at DFT and MP2 levels of theory." J. Comput. Chem. 19(6): 575-584.

Parisini, E., P. Metrangolo, et al. (2011). "Halogen bonding in halocarbon-protein complexes: a structural survey." Chem Soc Rev 40(5): 2267-2278.

Peebles, S. A., P. W. Fowler, et al. (1995). "Anisotropic repulsion in complexes B•••C12 and B•••HCI: the shape of the chlorine atom-in-a-molecule." Chemical Physics Letters 240: 130-134.

Politzer, P., J. S. Murray, et al. (2007). "Halogen bonding and the design of new materials: organic bromides, chlorides and perhaps even fluorides as donors." J. Mol. Model. 13(6-7): 643-650.

Politzer, P., J. S. Murray, et al. (2007). "Sigma-hole bonding and hydrogen bonding: Competitive interactions." Int. J. Quantum Chem. 107: 3046-3052.

Riley, K. E., J. S. Murray, et al. (2011). "Halogen bond tunability I: the effects of aromatic fluorine substitution on the strengths of halogen-bonding interactions involving chlorine, bromine, and iodine." J Mol Model 17(12): 3309-3318.

Riley, K. E., J. S. Murray, et al. (2009). "Br···O Complexes as probes of factors affecting halogen bonding: Interactions of bromobenzenes and bromopyrimidines with acetone." J Chem Theory Comput 5: 155-163.

Sponer, J., K. E. Riley, et al. (2008). "Nature and magnitude of aromatic stacking of nucleic acid bases." Phys Chem Chem Phys 10(19): 2595-2610.

Steinbrecher, T. and A. Labahn "Towards Accurate Free Energy Calculations in Ligand Protein-Binding Studies." Curr. Med. Chem. 17: 767-785.

Sunami, T., J. Kondo, et al. (2004). "Structures of d(GCGAAGC) and d(GCGAAAGC) (tetragonal form): a switching of partners of the sheared G.A pairs to form a functional G.AxA.G crossing." Acta Crystallogr D Biol Crystallogr 60(Pt 3): 422-431.

Sunami, T., J. Kondo, et al. (2004). "Structure of d(GCGAAAGC) (hexagonal form): a base-intercalated duplex as a stable structure." Acta Crystallogr D Biol Crystallogr 60(Pt 1): 90-96.

Vallejos, M., P. Auffinger, et al. (2010). Halogen interactions in biomolecular crystal structures. Int. Tables Cryst., Vol. F. D. M. Himmel: Ch 23.26.

Van Der Spoel, D., E. Lindahl, et al. (2005). "GROMACS: fast, flexible, and free." J Comput Chem 26(16): 1701-1718.

Voth, A. R., and Ho, P.S. (2007). "The Role of Halogen Bonding in Inhibitor Recognition and Binding by Protein Kinases." Curr. Topics Med. Chem. 7: 1336-1348.

Voth, A. R., Hays, F.A., and Ho, P.S. (2007). "Directing macromolecular conformation by halogen bonds." Proc. Natl. Acad. Sci., USA 104: 6188-6193.

Voth, A. R., Khuu, P., Oishi, K., and Ho, P.S. (2009). "Halogen bonds as orthogonal molecular interactions to hydrogen bonds." Nature Chemistry 1: 74-79.

Vreven, T., K. Morokuma, et al. (2003). "Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints." J Comput Chem 24(6): 760-769.

Wang, J., R. M. Wolf, et al. (2004). "Development and testing of a general amber force field." J Comput Chem 25(9): 1157-1174.

Xu, Z., Z. Liu, et al. (2011). "Utilization of halogen bond in lead optimization: a case study of rational design of potent phosphodiesterase type 5 (PDE5) inhibitors." J Med Chem 54(15): 5607-5611.

CHAPTER 5

EFFECT OF POLARIZATION ON THE STRUCTURE-ENERGY RELATIONSHIP OF BIOLOGICAL HALOGEN BONDS[4]

1. Summary

Interest in the non-covalent interactions involving halogens, particularly halogen-bonds (X-bonds), has grown dramatically in the past decade, propelled by the use of X-bonding in molecular engineering and drug design. X-bonds have been proposed as practical and effective tools in rational, or bottom up, drug design and have seen some success in a few preliminary cases. However, it is clear that a more complete analysis of the structure-energy relationship must be established for X-bonds in biological systems in order to fully exploit them in such biomolecular engineering applications. We present here the single-crystal structures of DNA Holliday junctions containing uracil bases modified by F, Br, Cl, or I, crystallographic titrations to estimate the enthalpic energies of the X-bonds relative to the competing hydrogen bonds (H-bonds) in the crystal system, and differential scanning calorimetry studies to compare the enthalpic and entropic

---

[4] Megan Carter[1], Andrea R. Voth[2], Matthew Scholfield[1], Brittany Rummel[1], Lawrence Sower[3], and P. Shing Ho[1*]
[1]Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO 80523; [2] National Institutes of Health, Bethesda, MD 20892; [3]School of Medicine, Loma Linda University, Loma Linda, CA 92350.
Formatted for submission to JACS

energies of bromine and iodine X-bonds in solution. The culmination of this study demonstrates that enthalpic stabilization of X-bonds increases with increasing polarization from F < Cl < Br < I, as consistent with the σ-hole theory of halogen polarization. However, total free energy of stabilization is determined in part by entropic contributions and these must be considered to effectively predict the outcome of halogen interactions. For this system, we find that bromine has the optimal balance between enthalpic and entropic terms to form the lowest free energy X-bonding interaction. The X-bond formed by iodine is a stronger molecular interaction in enthalpic terms, but with an entropic penalty. Thus, the overall free energy of an X-bonding interaction balances the stabilizing electrostatic effects against the competing effects on the local dynamics of a system.

2. Introduction

Halogen bonds, or X-bonds, are electrostatically-driven noncovalent interactions between a negatively charged Lewis base, or acceptor, and the positive region of a polarized halogen, defined as the X-bond donor (Fig. 5.1) (Metrangolo and Resnati 2012). In the past decade there has been a substantial increase into the theoretical understanding and application of X-bonds in material science, crystal engineering, and drug design (Ritter 2009; Metrangolo and Resnati 2001; Fourmigué 2009; Metrangolo 2005; Zhou, Huang et al. 2012; Xu, Liu et al. 2011). X-bonds have been implicated as practical and effective tools for the use in rational, or bottom up, drug design and have been used successfully in a few preliminary cases (Xu, Liu et al. 2011; Matter, Nazare et al. 2009). Unfortunately, current molecular mechanic calculations inaccurately model halogen interactions because

underlying classical force fields fail to model the anisotropic nature of polarized halogens, a property directly linked to X-bond formation. While the development of anisotropic force fields for biological halogen bonds (ffBXB) (Carter, Rappé et al. 2012) is promising, they are actively being incorporated into current programs and not yet available. Thus, prediction of interaction strength is currently limited to molecular systems accommodated by quantum mechanical calculations. Quantum mechanical and combined quantum mechanical/molecular mechanical calculations have shown some success in predicting interaction strength and relative binding affinities (Lu, Shi et al. 2009) but these calculations become computationally difficult and have accumulated error for large molecular systems. In addition, these prediction methods fail to consider entropic contributions to overall binding affinity. For X-bonds to progress as tools in bottom-up drug design the entropic cost of halogen incorporation must be investigated both experimentally and computationally. In this study we compare and contrast the enthalpic and entropic stabilization energy of halogens, from fluorine to iodine, in their ability to form X-bonds and, in this way, provide insights into not only the structure energy relationship of X-bonds but also which of the halogens may be optimal for this interaction in a biological context.

The halogen class of elements, including fluorine (F), chlorine (Cl), bromine (Br), and iodine (I), are traditionally considered to carry an overall negative charge, but become polarized when forming a covalent σ-bond to another atom. Polarization of Group 16 and 17 elements is effectively described by the σ-hole model developed by Politzer et al. (Clark, Hennemann et al. 2007; Politzer, Murray et al. 2007).

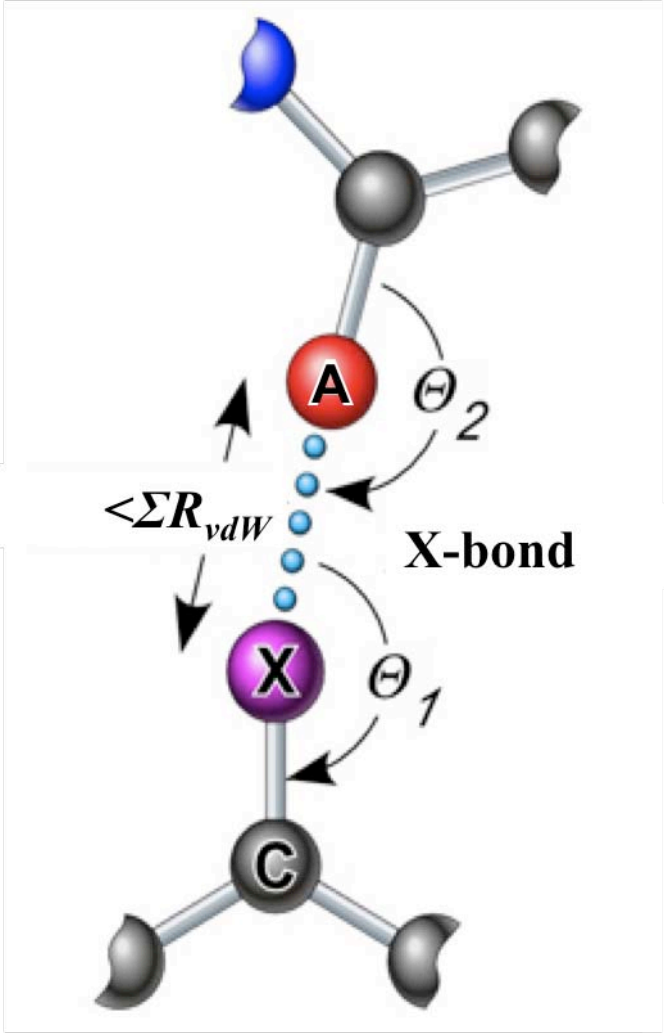


Fig. 5.1. X-bonds. X-bonds are short interactions between an electronegative acceptor (A) and donor (X). The electropositive σ-hole of a polarized halogen acts as the donor in X-bonds. The interatomic X···A distance is shorter than the sum of their respective R_{vdW}.

The σ-hole model details the depopulation of the p_z orbital electrons into the σ-bond, resulting in a reduced electron density along the σ-bond. This exposes the nuclear charge, creating a region of positive electrostatic potential called the σ-hole at the crown of the halogen along the σ-bond (Fig. 5.2). The remaining electrons in the p_x and p_y orbitals maintain a ring of negative charge roughly perpendicular to the σ-bond (Carter, Rappé et al. 2012). The extent of σ-hole formation is dependent on the polarizability of the halogen, which increases from F<Cl<Br<I (Clark, Hennemann et al. 2007). Increasing the electron withdrawing ability of the covalently bound substituent can also increase relative polarization and σ-hole formation (Riley, Murray et al. 2011).

The anisotropic electrostatic distribution resulting from halogen polarization has been directly associated to the strong directional preferences (Voth 2009; Murray, Riley et al. 2010; Shields, Murray et al. 2010) and resulting interaction energy of X-bonds (Riley, Murray et al. 2009; Shields, Murray et al. 2010).

In order to assay the structure energy relationship of halogens in a biological context Voth et al. had previously shown that bromine X-bonds could be engineered to stabilize the formation of DNA Holliday junctions (Voth 2007). Holliday junctions are four-stranded DNA complexes involved in multiple cellular processes, including genetic recombination, DNA lesion repair, viral integration, restarting of stalled replication forks, and proper segregation of homologous chromosomes during meiosis (Ho 2011). Four-stranded DNA junctions are ideal for studying X-bonds because their rigid structure is defined by a small number of specific intramolecular interactions.

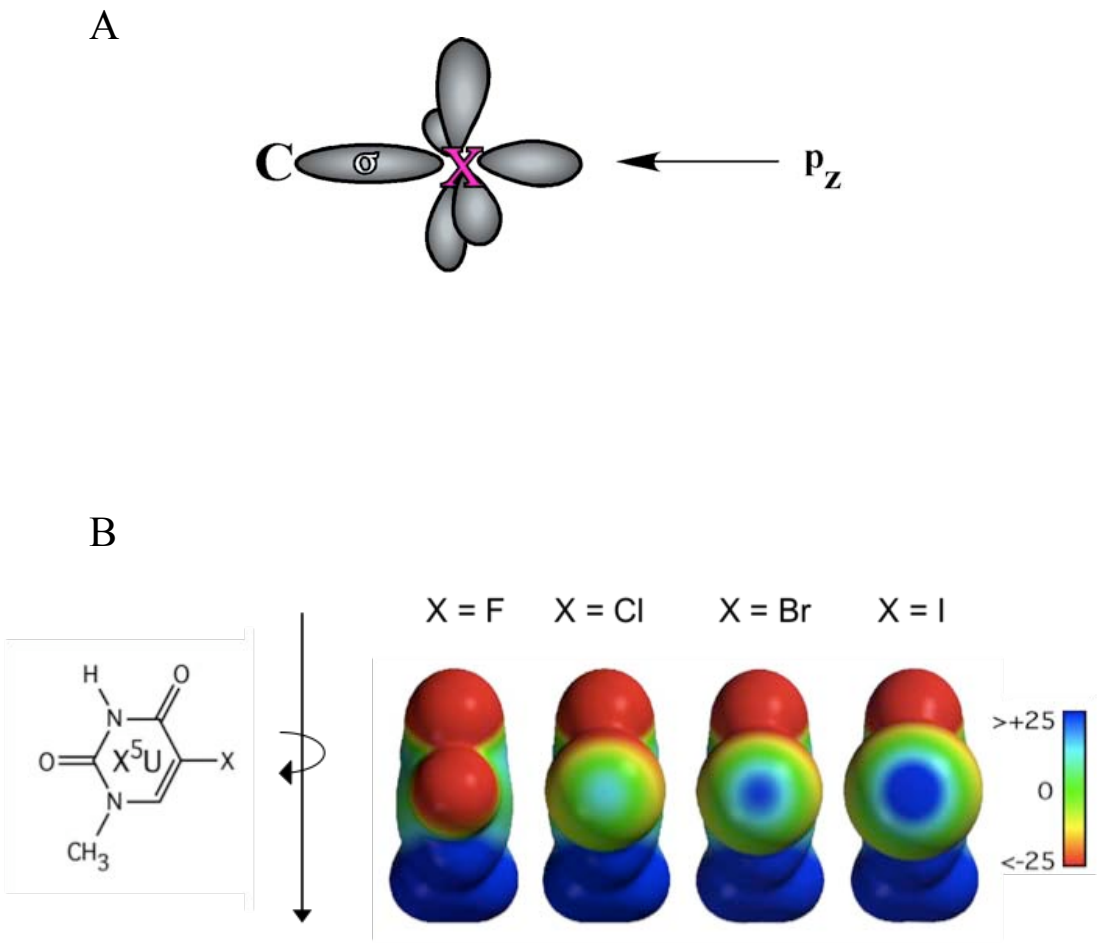


Fig. 5.2. Polarization and Sigma-Hole Formation. (A) σ-hole theory describes reduction of p_z orbital electron occupancy as electrons are pulled into the sigma bond, thus exposing a positive nuclear charge. (B) Formation of the σ -hole as demonstrated by ab initio electrostatic potential surface calculations of halogenated uracil [18]. Positive electrostatic potential is represented in blue and negative electrostatic potential in red presented in the -25 to +25 kcal/mol range.

The stacked-X junctions studied here consist of two continuous DNA strands connected by two crossover strands, each forming a tight U-turn (Fig. 5.3A) (Duckett, Murchie et al. 1988) (Ho 2001). Holliday junctions used in this and related studies are formed from the sequence d(CCnnn$N_6N_7N_8$GG) where $N_6N_7N_8$, defines the trinucleotide core responsible for junction stabilization of the U-turn necessary of crossover strands (Fig. 5.3) (Eichman, Vargason et al. 2000). These junctions, under physiological salt conditions, adopt a stacked-X form, with arms stacked to form near continuous standard B-DNA duplexes (Ho 2001). An H-bond from cytosine at position eight (C_8) to the phosphate of the preceding base is required for junction formation. The junction is further stabilized by an interaction from the base at position seven to the phosphate at position six (Hays, Vargason et al. 2003).

This secondary interaction site can be stabilized by H-bonding or X-bonding and therefore provides framework to compare the relative strengths of X- vs H-bonds observed in crystal structures. Solution state energies were consistent with those estimated from the crystallographic competition assay (Carter and Ho 2011; Voth 2007).

In the current study, we investigate the effect of polarization on the structure-energy relationship of X-bonds in a biological system by comparing the ability of F, Cl, Br (Voth 2007), and I X-bonds to affect the conformational stability of DNA Holliday junctions. We show here that the increase in polarizability of the halogen from F < Cl < Br < I is correlated with an increase in the stabilizing potential of the engineered X-bond, as predicted by the σ-hole model.

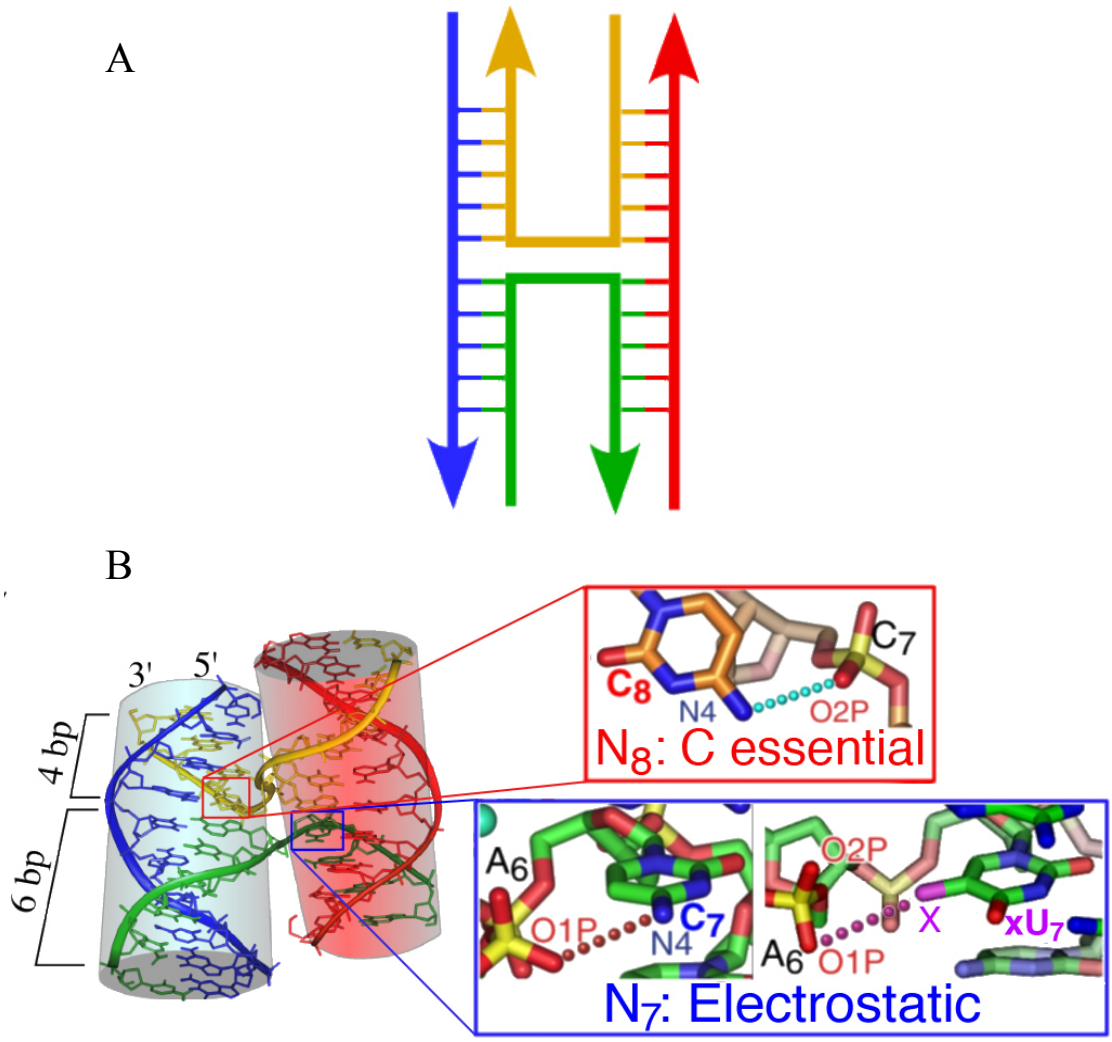


Fig. 5.3. Structure of DNA Holliday junction. (A) Schematic of the compact stack-X junction with two continuous (blue and red) and two crossover (yellow and green) strands. (B) Crystal structure of 5'-CCGGTACCGG-3' highlighting junction stabilizing interactions. The H-bond from the amino group of cytosine at position N_8 to the phosphate at position N_7 is essential for junction stabilization. A similar interaction from the nucleotide at N_7 to the phosphate of nucleotide at P6 supplies accessory stabilization with either a H-bond or X-bond.

The correlation between the structures and their energies allow us to distinguish between the enthalpic and entropic contributions to the interactions, which, ultimately, provides a better understanding for what makes a good "halogen bond" in a biological system.

3. Theory and Methods

3.1 DNA synthesis and purification

Chemically synthesized DNA oligonucleotides are obtained from Midland Certified Reagent Company on the solid Controlled-Pore Glass (CPG) support with the final dimethoxytrityl (DMT) protecting group attached. Sequences were subsequently purified by reverse phased HPLC followed by size exclusion chromatography on a Sephadex G-25 column after detritylation. The constructs for this study were complementary sequences designed to form four-stranded junctions competing either one halogen (X1J) or two halogens (X2J) against two hydrogen bonds (Table 5.1).

Table 5.1. DNA Holliday junction constructs. DNA construct denotation, sequence, and X:H ratio. The trinucleotide core, responsible for sequence dependent junction stabilization, is listed in bold for each sequence. Halogenated uracils are denoted as ^{x}U where X is fluorine, chlorine, bromine, or iodine attached at position 5 of the uracil base.

Construct	Sequences	X:H
H2J	2(CCGATACCGG) + 2(CCGGTAUCGG)	0:2
F2J	2(CCGATACCGG) + 2(CCGGTA^{Fl}UCGG)	2:2
Cl2J	2(CCGATACCGG) + 2(CCGGTA^{Cl}UCGG)	2:2
Cl1J	2(CCGATACCGG) + 1(CCGGTA^{Cl}UCGG) + 1(CCGGTAUCGG)	1:2
Br2J	2(CCGATACCGG) + 2(CCGGTA^{Br}UCGG)	2:2
Br1J	2(CCGATACCGG) + 1(CCGGTA^{Br}UCGG) + 1(CCGGTAUCGG)	1:2
I2J	2(CCGATACCGG) + 2(CCGGTA ^{I} UCGG)	2:2

3.2 Crystallization and structure solution

The crystal structures of the Br1J and Br2J constructs were previously reported by Voth et al. and are referenced here for comparison (Voth 2007). The F2J, Cl1J, Cl2J, I1J, and I2J constructs were crystallized by sitting drop vapor diffusion from solutions that contained 0.7mM DNA, 25mM sodium cacodylate pH 7.0 buffer, 10-25mM calcium chloride, and 0.8-1.2mM spermine, equilibrated against a reservoir of 30-40% aqueous MPD. Diffraction data were collected at Advanced Light Source (ALS) at the Lawrence Berkeley Laboratories at λ = 0.9 Å under liquid nitrogen temperatures. All data was processed using DENZO and SCALEPACK from the HKL2000 software (Otwinowski and Minor 1997). Structures were solved by molecular replacement using EPMR (PDB codes of the search models, and the associated correlation coefficient and R_{cryst} for each initial model are listed in (Table 5.2). The C2 symmetry and unit cell volumes of Cl1J, Cl2J, I1J, and I2J indicated that their asymmetric units are defined as two DNA strands, one continuous and one crossover strand, with the full DNA junction generated by the crystallographic two-fold symmetry at the center of the four-stranded junction. The F2J construct also crystallized in the C2 space group; however, the 2-fold axis is shifted, resulting in a doubling of the c-axis, which results in an asymmetric unit cell consisting of a full four-stranded junction (Eichman, Vargason et al. 2000). The DNA-DNA contacts that define the crystal lattices are the nearly identical for all of the structures in the current study and with the previous studies (Hays, Vargason et al. 2003; Voth 2007; Eichman, Vargason et al. 2000). Crystallographic and refinement statistics for all current structures are reported in supplemental data (Table 5.2).

Table 5.2 Crystallographic Parameters

	F_2J	Cl_1J	Cl_2J	I_1J	I_2J
Space Group	C2	C2	C2	C2	C2
Unit Cell					
a, Å	65.213	65.58	65.69	65.74	64.96
b, Å	23.917	24.35	23.57	25.19	24.77
c, Å	77.45	37.24	37.29	37.17	37.62
β-angle	114.802	110.85	110.92	100.88	111.59
Unique reflections (for refinement)	3398	4947	3323	3562	5772
Resolution, Å	50 – 2.38	50 – 1.7	50 – 1.94	50-1.9	50 – 1.7
Completeness, %*	74.2 (51.2)	79.2 (41.4)	81.0 (54.2)	76.4 (69.9)	95.4 (90.3)
I/sig, I*	48.93 (5.6)	28.25 (1.25)	29.54 (3.25)	11.31 (2.20)	14.14 (3.45)
R_{merge}, %*	(26.5)	(32.3)	(26.8)	4.2 (24.4)	6.9 (25.4)
Refinement Statistics					
R_{cryst}, (R_{free}), %	22.5 (29.1)	27.9 (32.1)	26.4 (30.0)	23.6 (26.3)	23.9 (25.5)
No. of atoms: DNA (solvent)	808 (99)	404 (158)	404 (95)	404 (73)	404 (74)
<B-factor>, DNA (solvent)	16.3 (11.8)	16.4 (24.7)	13.0 (21.4)	30.0 (35.1)	17.3 (21.7)
RMSD bond length, Å	0.002	0.001	0.005	0.005	0.007
RMSD bond angle	0.6	0.4	1.0	0.9	1.1

*Values for the highest resolution shell are shown in parentheses

Constructs of DNA junctions containing potential X-bonding halogenated and potential H-bonding non-halogenated strands have previously been shown to adopt one of two isomeric forms, the X-isomer or H-isomer (Fig. 5.4). In the X-isomer the halogenated uracil, xU_7, is located on the inside crossover strand and forms a stabilizing X-bond to the phosphate oxygen. In the H-isomer the xU_7 is located on the outside continuous strand and cytosine, C_7 is at the inside location forming a stabilizing H-bond to the phosphate oxygen. Refinement for all constructs was carried out in CNS (Brünger, Adams et al. 1998) with rigid body refinement, simulated annealing, several rounds of positional and individual B-factor refinement, and addition of solvent. Refinement was performed with models ambiguous for isomeric form, not specifying the presence of the halogen (or the possible N2 nitrogen of the guanine base complimentary to the variable cytosine) at either the inside or outside position .

Once the refinement of this initial model converged, we set out to determine the occupancy of each model to the overall structure using an occupancy titration. X- and H-isomer overlaying structures were generated and occupancies were varied in a correlated manner from 0 to 100%. Constructs competing one X-bond against two H-bonds (X1J) were titrated with only one X-bond modeled for the X-isomer. Each iteration of occupancy was followed by a single round of B-factor refinement and the crystallographic R and R_{free} were monitored as a function of the occupancy. Mock occupancy titrations were performed in which iterative rounds of B-factor refinement were used to determine a background R and R_{free} change associated with the progressive

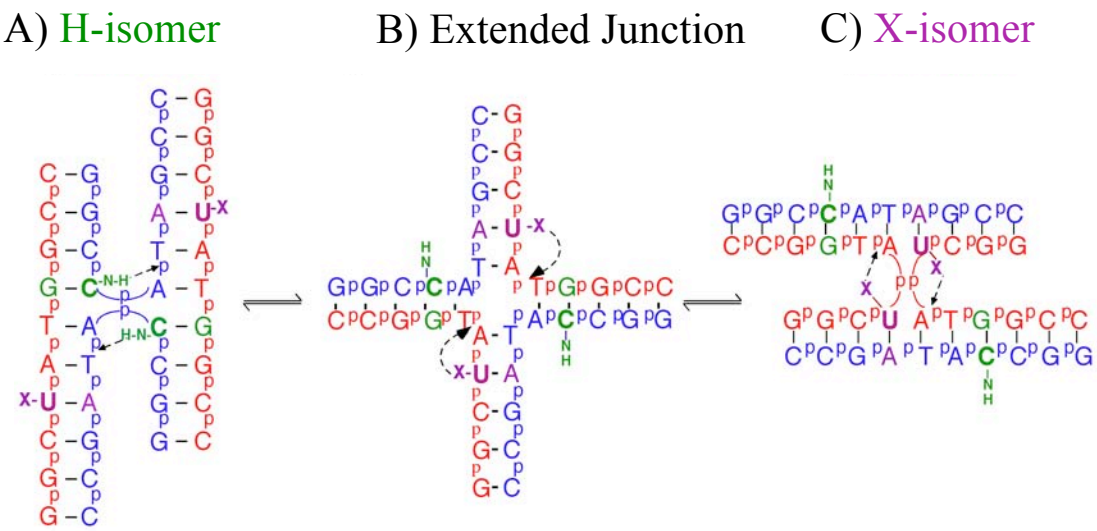


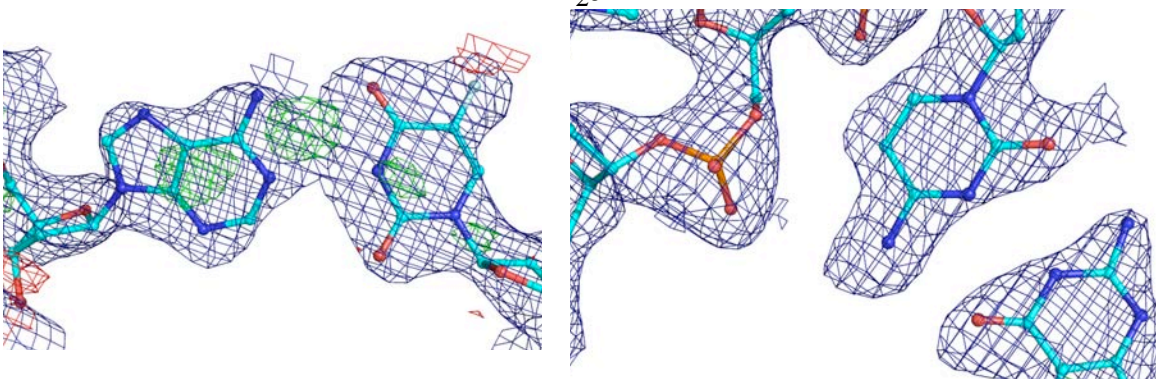
Fig. 5.4. Isomeric Competition. Isomeric forms of the stacked-X DNA junction where the H-isomer (A) is stabilized by a H-bond from cytosine at C_7 to the preceding phosphate oxygen and the X-isomer (C) is stabilized by a X-bond from halogenated (F, Cl, Br, or I) uracil, xU_7, to the preceding phosphate oxygen. Junction isomerization occurs through extended intermediate (B). Isomeric form is distinguished by location of the xU_7 at either the inside (X-isomer) or outside (H-isomer) position.

refinement of a structure with no change in occupancy. This served as the baseline, which was subsequently subtracted from the experimental R and R_{free} titration to yield the final titration curve. Titration data with a clear minima were fit in Kalidograph with linear models for the decrease and, separately the increase in R_{free}. The intersection of the two lines indicates the best fit of the model to the data. The error in this analysis is estimated by the spread in the minima as indicated by the models respective intersections with the average of the minima data points. Titration results for I1J and I2J both trend toward 100% without clear over fitting and therefore were not analyzed in this manner. Occupancy titration results reported here for Br_1J, Br_2J are in agreement with the previously published data, but have slight variance within the associated error. Occupancy titration results were also in agreement with electron density inspection of the inside and outside positions. To demonstrate this the inside crossover location (bases $N_6N_7N_8$) are imaged with 2Fo-Fc and Fo-Fc electron density (Fig. 5.5). Atomic coordinates and structure factors will be deposited in the PDB upon acceptance.

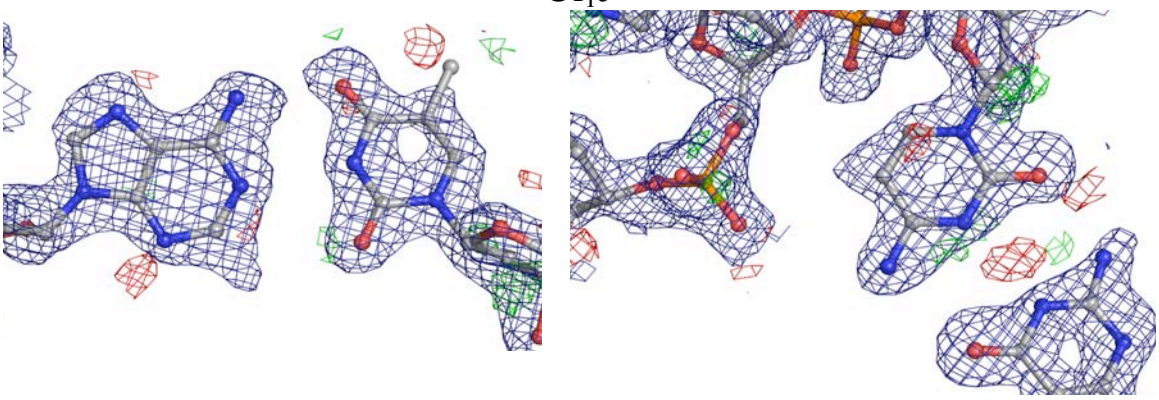
3.3 Differential scanning calorimetry studies

Differential scanning calorimetry (DSC) studies were performed to determine the stabilizing energies of the H2J, Br2J, and I2J constructs in solution as previously described (Carter and Ho 2011). The complementary DNA sequences of each construct were mixed in equimolar concentrations (varied from 15 μM to >300 μM) in 50mM sodium cacodylate pH 7.0 buffer and 1mM calcium chloride, in order to approximate crystallization conditions. The solutions were heated to 90°C for one hour and slowly reannealed to room temperature overnight. The energetic parameters for melting the constructs at each concentration were determined using a TA Instruments Nano DSC

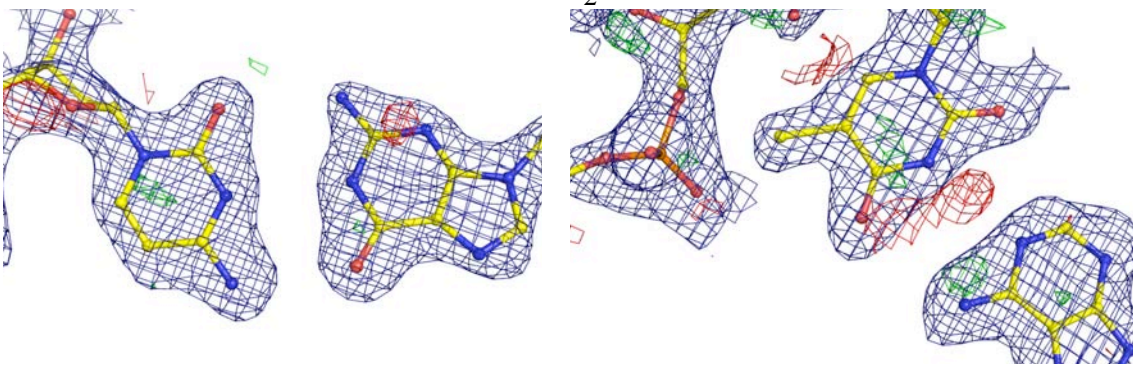
F_2J



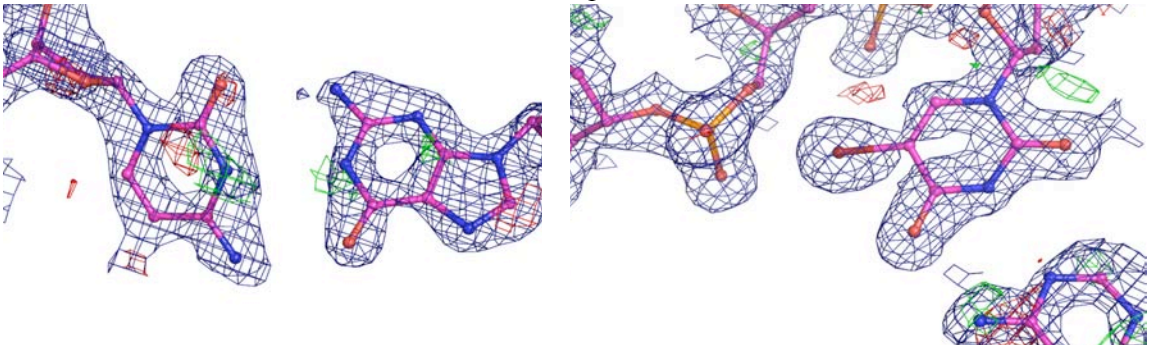
Cl_1J



Cl_2J



Br_1J



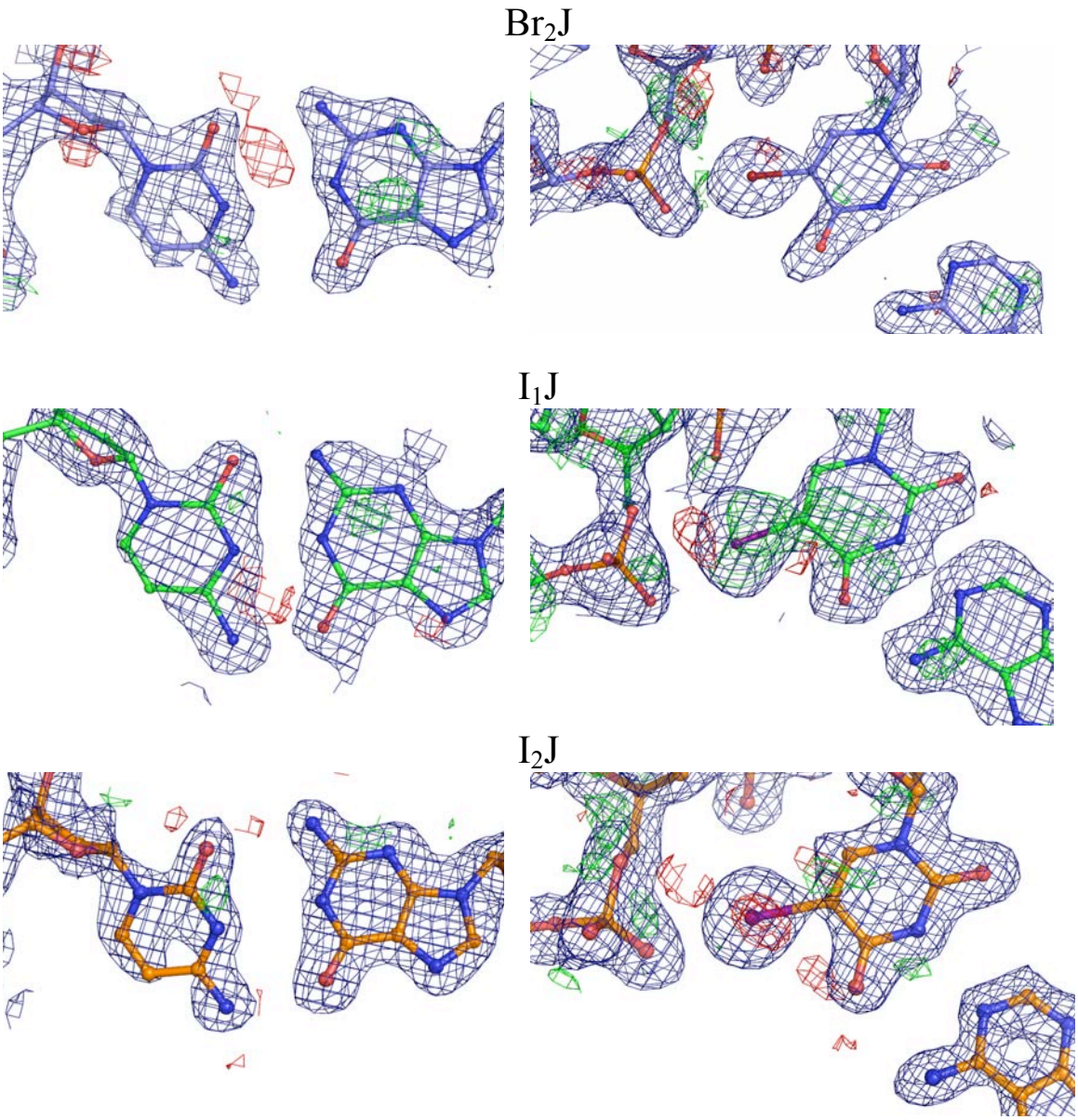


Fig. 5.5. Structure and Electron Density Maps. The defining inside (N_{14}:N_7••OP_6) positions of the DNA Holliday junctions for each construct are shown with the isomer dominate identity and occupancy as determined from occupancy titration experiments: F2J- 60% H-isomer, Cl1J- 70% H-isomer, Cl2J 63% X-isomer, Br1J – 80% X-isomer, Br2J – 100% X-isomer, I1J – 90% X-isomer, I2J – 100% X-isomer. Electron density are shown at 2Fo-Fc 1σ in blue, positive Fo-Fc in green, and negative Fo-Fc in red. Increasing halogen occupancy stabilizing junction formation at the inside location is evident from F<Cl<Br<I.

with the pressure held constant at 3.0 atm. Each DNA sample was run against buffer in a heating cycle from 0°C to 90°C at a scanning rate of 1°C/min with an equilibrium time of 900 s, and repeated at least three times. DNA constructs were analyzed at multiple DNA concentrations in order to sample both the duplex and junction conformation, as junction formation has been shown to be concentration dependent (Hays, Schirf et al. 2006). Data were analyzed using the NanoAnalyze software from TA Instruments (TA Instruments, New Castle, DE), with the best fit determined according to the standard deviation of the fit. The data at lower concentrations were best fit using a two state model scaled by a weighting term (Aw), to account for the presence of both double- and single-strand DNAs. Samples in which the Aw term had indicated a much higher than predicted double-stranded concentration suggested the presence of a four-stranded junction component; consequently, the data for these samples were analyzed by applying a two component, two-state model. The similarity in T_m and ΔH_m values for the duplex fractions between the single component analysis of the low DNA concentration data and the two component high DNA concentration data support this interpretation of the analyses. The presence of junctions was evident from the single component analysis of the data from [DNA] from 20 to 100 uM, but did not warrant fitting using the two-component analysis and, therefore, were excluded from the thermodynamic parameters. The ΔH_m for melting the duplex form of each construct was taken as the average ΔH_m for DNA concentrations from 15 to 20 μM along the low temperature component from the two component analysis of data at [DNA] > 100 μM. The ΔH_m of the junction form of each construct was taken as the average of the higher temperature component from the [DNA] > 100 μM analyses.

4. Results and Discussion

We have previously shown that bromine X-bonds could effectively compete against H-bonds to direct the conformation of DNA Holliday junctions, demonstrating that such X-bonds are approximately 2 to 5 kcal/mol more stabilizing than a classical H-bond (Carter and Ho 2011). Here, we extend the analysis to compare and contrast the structures and energies of potential X-bonds involving fluorine (F), chlorine (Cl), bromine (Br), and iodine (I) against a junction construct that is fully stabilized by two H-bonds (the H2J construct) in stoichiometries of 1X:2H (X1J) or 2X:2H (X2J). In the current study, the single-crystal X-ray diffraction of X-bonds are used to compared the structures of each halogenated DNA junction (F2J, Cl1J, Cl2J, I1J, I2J, and the two previously studied Br1J and Br2J constructs) and to estimate their stabilizing potential against two competing H-bonds. For the Br and I variants, we also determine the calorimetric X-bonding energy in solution in order to compare and contrast the contributions of enthalpic and entropic terms to the interactions, and to relate these back to the observed structural effects.

4.1 Crystallographic Studies

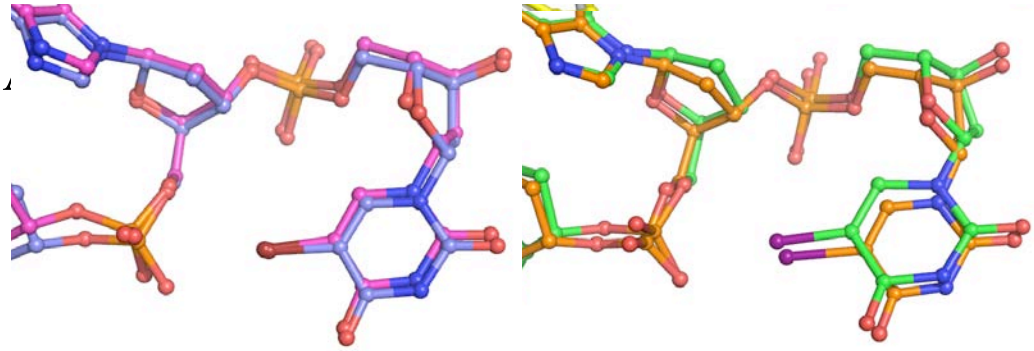
The single-crystal structures of all constructs conform to the general form of the stacked-X DNA junction (Ho 2001). The H-bond from the N4 nitrogen of cytosine C_8 to the preceding phosphate oxygen, essential for forming the tight U-turn of the junction crossing-strands, was observed in both isomer forms of these junctions. The DNA junction in this study is capable of adopting either a conformation in which H-bonds from the C_7 cytosine to a phosphate oxygen of the preceding base (designated the H-isomer) or

potential X-bonds from the xU_7 halogenated uracil, to the preceding phosphate oxygen (X-isomer) at the analogous nucleotide positions help to stabilize the tight U-turn of the DNA backbone of the crossing strand. The isomeric form of each halogenated construct can be identified crystallographically by determining the position of the halogenated uracil and the associated complimentary base in their respective single-crystal structures. The X-isomer places the halogen at the inside position of the crossing strand, while the H-isomer places the halogen at the outside position of the non-crossing strand of the junction. The X- and H-isomers are seen here and from previous studies to be isomorphous and, thus, accommodated by the same crystal lattice interactions. In addition, the junction stabilizing H- and X-bonding interactions are far removed from the direct intermolecular DNA lattice contacts. Therefore, the ratio of X- to H-isomers in the crystal reflects the solution population and, consequently, can be used to assess the difference in energy between the competing X- and H-bonding interactions.

The H2J construct does not contain a competing halogenated uracil, but, instead, shows an additional H-bond from the amino nitrogen of C_7 to phosphate oxygen of the preceding base (with an N···O interatomic distance of 2.84 Å); thus, the H2J junction forms only the H-isomer and this serves as the H-bonded control. In addition, a similar H-bonding interaction was observed in the H-isomers for all constructs, indicating that the H-isomers are all nearly identical and that the H2J junction is a good model for this conformer in all of the halogenated DNA constructs of this study. The halogenated constructs are analyzed for the presence of potential X-bonds in their X-isomer forms in relation to these intrastrand interactions of H2J structure.

The F2J construct was observed to be predominantly in the H-isomer form (Fig. 5.6A). The fluorine to oxygen distance in the X-isomer is slightly longer than the accepted ΣR_{vdW}. A superposition of the X-isomer structure over the H-isomer of the F2J construct showed the two conformers to be nearly identical, with only a slight shift of the fluorinated uracil base away from the phosphate and towards the minor groove of the $xU_7 \cdot A_{14}$ base pair relative to the cytosine of the H-isomer. Thus, although the directional approach of the acceptor phosphate oxygen is towards the halogen σ-hole, this particular fluorine interaction was not strictly defined as X-bonding.

X-bonds were observed in X-isomers forms of the Cl1J, Cl2J, Br1J, Br2J, I1J, (Fig. 5.6) and I2J constructs, as supported by halogen to phosphate oxygen distances that are closer than ΣR_{vdW} and the directional approach of the acceptor towards the halogen (Table 5.3) (Metrangolo and Resnati 2012). The X-bond in the Br1J X-isomer was previously identified as being 2% shorter that the ΣR_{vdW} (3.34 Å), while the interaction in the Br2J X-isomer is 16% shorter than the ΣR_{vdW}. This is equivalent to an ~0.5 Å difference in the Br···O distances, resulting from a rotation about the β-angle of the preceding base that shifts the position of the phosphate relative to xU_7. In contrast, the Cl···O distances in the X-isomer structures of Cl1J and Cl2J are reduced by 11% and 13% in the ΣR_{vdW}, respectively. This difference is seen as a slight shift of less than 0.1Å between the Cl donor and O acceptor atoms, with a difference in Θ_1 of only 6° between the two constructs; thus, we consider the X-conformers to be nearly identical in structure and associated energies of interaction for Cl_1J and Cl_2J. The X-isomer forms are nearly identical in I1J and I2J, with the I···O distances being shorter than the ΣR_{vdW} by 17% and 15%, respectively, similar to that of the Br2J X-isomer. The angular approach of the



C D

Fig. 5.6. Analysis of junction stabilizing interactions. X-isomers were superimposed on the H-isomer of F2J for direct structural comparison. (A) F2J X-isomer (cyan) is most similar to the H-isomer (olive) with slight movement of the halogenated uracil. (B) Cl1J (gray) and Cl2J (yellow) X-isomers are similar in structure while the Br1J (magenta) and Br2J (blue) have differing geometry base on rotation of the phosphate about the β-angle resulting in a longer Br1J X···O interaction distance. I1J (green) and I2J (orange) are similar in both phosphate and halogen location.

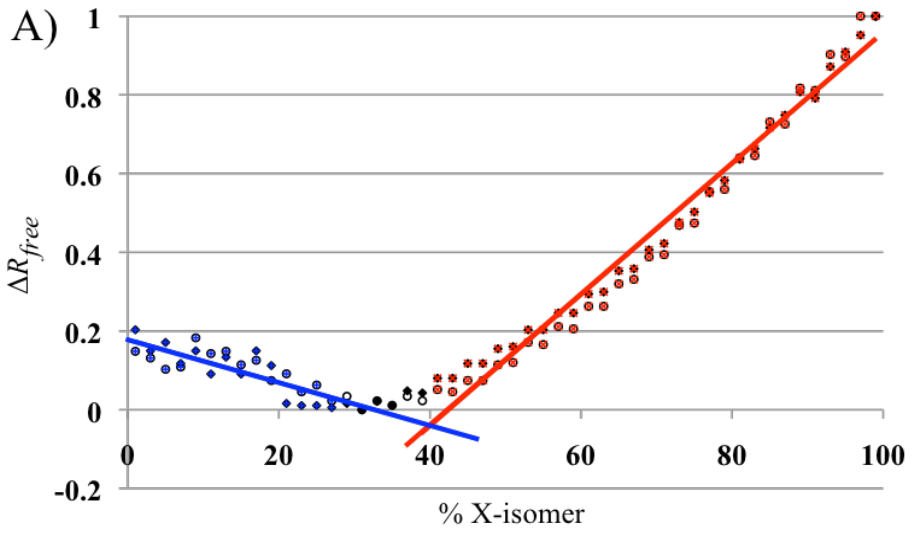
Table 5.3. $X\cdots O^-$ Interaction Geometry. Halogen (X) to O1P phosphate oxygen (O) X-isomer geometries of each constructs were analyzed. The angular approach of the acceptor towards the halogen is reported as Θ_1. The % Reduction in ΣR_{vdW} is calculated as the difference in the predicted interatomic distance and observed interatomic distance divided by the predicted interatomic distance.

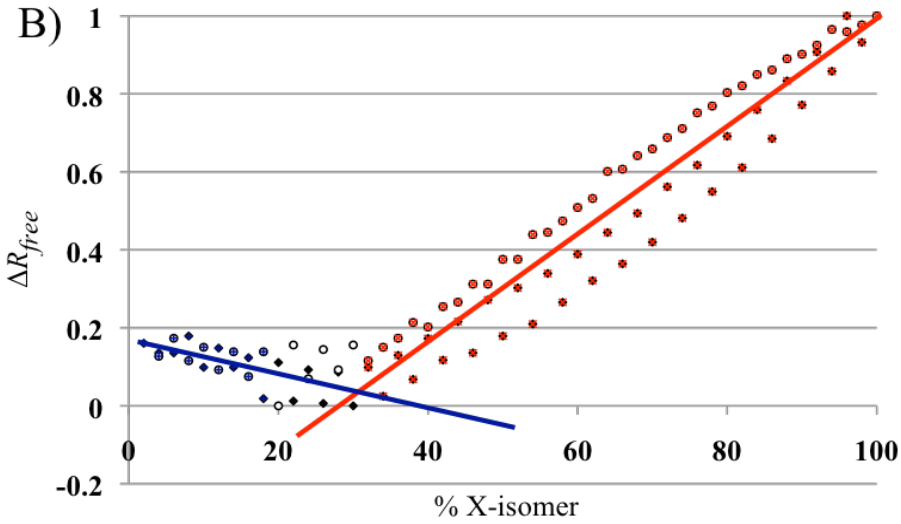
Construct	$X\cdots O^-$ Distance	% Reduction in ΣR_{vdW}	Θ_1
F2J	3.20 Å	-5%	153.5°
Cl1J	2.95 Å	11%	152°
Cl2J	2.88 Å	13%	146°
Br1J	3.32 Å	2%	167.2°
Br2J	2.87 Å	16%	163.2°
I1J	2.92 Å	17%	164.4°
I2J	3.01 Å	15%	170.7°

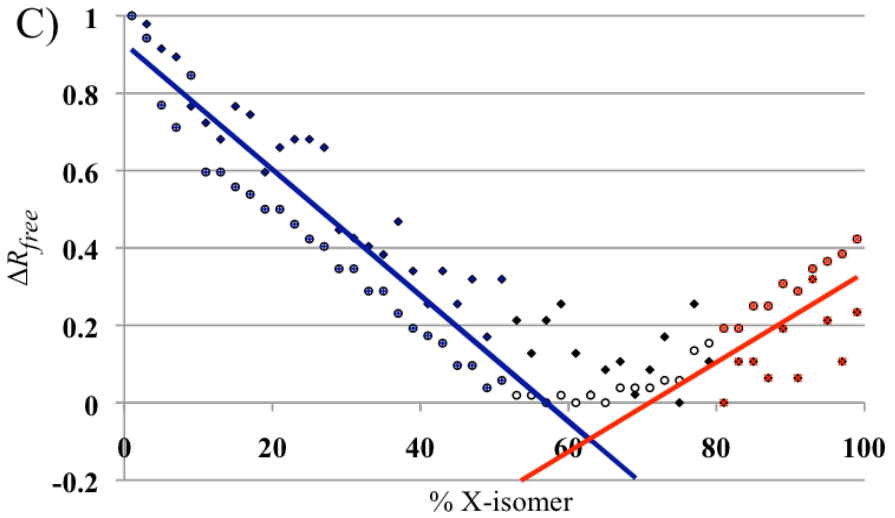
acceptor phosphate oxygen to the C-X bond (the Θ_1-angle) becomes progressively more linear (approaching 180°) as the polarizability and the size of the halogen donor increase. This suggests that as the σ-hole becomes more pronounced, it has a greater influence on the geometry of the X-bond.

4.2 Crystallographic occupancy titrations to determine X- and H-bonding energies

The ratio of X- to H-isomers present in crystal formation of each construct was quantified from the single-crystal studies using an occupancy titration method that had previously been shown to correlate well with solution studies (Carter and Ho 2011). In each case, the structures were initially refined as a single model without specifying an isomer form. As the refinement neared convergence, an overlapping model with equal contribution of X- and H-isomers was generated to perform occupancy titration with. In the titration, the X-isomer contribution was increased from 0 to 100% while the H-isomer occupancy was decreased from 100 to 0% percent (Fig. 5.7). This was repeated in the opposite direction, increasing the H-isomer and decreasing the X-isomer contributions to control for any hysteretic effects. The resulting percentages were confirmed by comparison to the electron densities calculated from models refined at 0%, 100%, and estimated effective X-isomer contributions. From these titrations, it was clear that the fluorinated and chlorinated junctions behaved similarly to the previously reported Br1J construct in that they did not adopt entirely the X- or the H-isomers, but were mixtures of the two. By monitoring the minimum in the change in R_{free} (ΔR_{free}) as the contributions of the isomers were varied, we observed that the F2J and Cl1J junctions were predominantly H-isomer, while Cl2J was predominantly X-isomer (Table 5.4). The I1J and I2J titrations


A)
1
0.8
0.6
0.4
0.2
0
-0.2
ΔR_{free}
0
20
40
60
80
100
% X-isomer


B)
1
0.8
0.6
0.4
0.2
0
-0.2
ΔR_{free}
0
20
40
60
80
100
% X-isomer


C)
1
0.8
0.6
0.4
0.2
0
-0.2
ΔR_{free}
0
20
40
60
80
100
% X-isomer

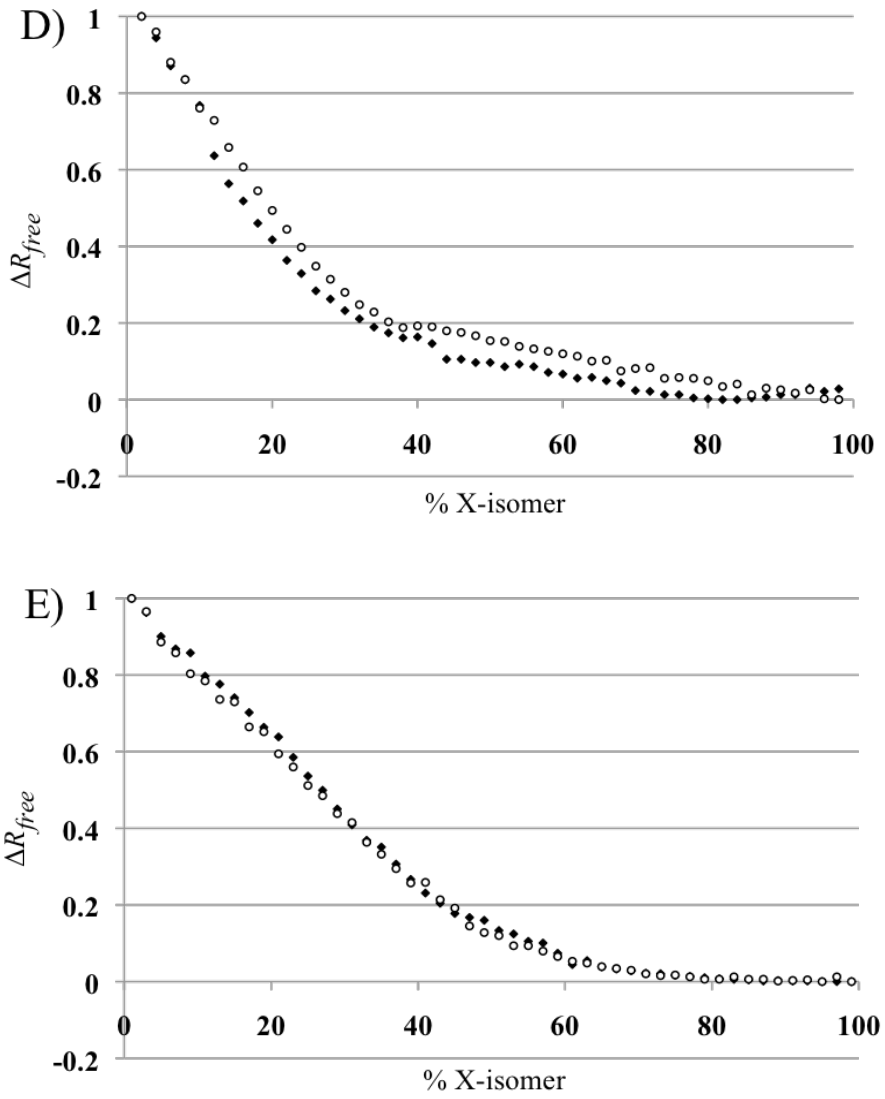


Fig. 5.7. Occupancy Titration. Distribution of H- and X-isomers are estimated by systematically increasing (black diamonds) or decreasing (open circles) X-isomer occupancy, correlated with occupancy of the H-isomer. Titration data with clear minima, A) F2J, B) Cl1J, and C) Cl2J, are fit with linear models for the fitting (blue) and overfitting (red) portions. The minima is taken as the intersection of the two models with the error of the minima determined form the x-axis intersection of each model. The minima in normalized ΔR_{free} values indicates the most likely ratio of X-isomer indicated by the crystallographic data. Titration data for D) I1J and E) I2J lack clear overfitting and therefore are estimated at greater than 80% X-isomer and greater than 100% X-isomer respectively.

Table 5.4. Crystallographic Competition Assay Results. The %X-isomer as determined from occupancy titration minima in ΔR_{free} is used to determine the stabilization energy of the X-isomer vs. the H-isomer. The individual $X\cdots O^-$ stabilization energy is determined after correcting for the number of halogens present and the H-isomer stabilization energy.

Construct	% X-isomer	$\Delta E_{IsoX - IsoH}$ (kcal/mol)	$\Delta E_{X\cdots O}$ (kcal/mol)
F2J	40% ± 4	0.24 ± 0.05	-0.54 ± 0.13
Cl1J	30% ± 6	0.50 ± 0.13	-0.82 ± 0.28
Cl2J	63% ± 7	-0.32 ± 0.24	-0.82 ± 0.17
Br1J	84% ± 4	-0.98 ± 0.15	-2.30 ± 0.29
Br2J	100% ± 5	ND	ND
I1J	≥80%	≥-0.82	≥-2.14
I2J	100% ± 5	ND	ND

indicated that both iodinated junctions were predominantly in the X-isomer form, as expected from the very high polarizability and positive electrostatic potential of the σ-hole. The ΔR_{free} of the I2J titration data asymtoped toward 100% X-isomeric form and did not show any evidence for the H-isomer. The I1J titration, however, shows a linear reduction in ΔR_{free} that intersects at 80% X-isomer, with a slight indication of H-isomer at ~95%. Therefore the I1J construct was determined to be ≥ 80% X-isomer, while the I_2J construct is 100% X-isomer.

The resulting ratio of X- to H-isomer could further be used to calculate the difference in energy between the two isomers according to the relationship $\Delta E_{IsoX-IsoH} = -RT \ln\left(\frac{\%IsoX}{\%IsoH}\right)$. Since we had determined that the X-bonds in the Cl1J and Cl2J constructs were essentially identical in structure, we can further assume that they are similar in energy. With this assumption allowed, we could construct a set of equations that allowed us to estimate the absolute energies of interactions of both the Cl···O X-bond ($E_{Cl\cdot O}$) and the competing H-bond ($E_{H\text{-}bond}$) from the $\Delta E_{IsoX\text{-}IsoH}$ of the chlorinated junctions.

$$\Delta E_{IsoX-IsoH(Cl_1J)} = \Delta E_{Cl\cdots O} - 2\Delta E_{H-bond}$$

Eq. 1

$$\Delta E_{IsoX-IsoH(Cl_2J)} = 2\Delta E_{Cl\cdots O} - 2\Delta E_{H-bond}$$

Eq. 2

Direct comparison of $X\cdots O^{-1}$ X-bonding energies show a trend of stabilization increasing from F2J < Cl1J = Cl2J < Br1J ≤ I1J . These results follow the trend of polarization and are, therefore, in agreement with predictions from the s-hole model for

X-bonding. The Br2J and I2J junctions were entirely X-isomer, thereby precluding the determination of their X-bonding energies from this crystallographic assay.

4.3 X-Bonding energies from differential scanning calorimetry

The H-bonding energies in H2J (Carter and Ho 2011), and the bromine and iodine X-bonding energies of I2J and Br2J (Carter and Ho 2011) were determined by solution-state melting studies through differential scanning calorimetry (DSC). The energies of interaction that contribute to stabilization of the junction core were segregated from the standard DNA base pairing and base stacking energies by exploiting the concentration dependent transition from duplex to junction (Hays, Schirf et al. 2006). Constructs annealed at low concentration are predominantly in the duplex form, evident from the single melting component DSC profile that is indicative of a duplex to single strand DNA transition. At high concentrations, the DSC profiles showed two melting components with melting temperatures (T_m). The lower temperature component had the same T_m and melting enthalpies (ΔH_m) as those of the low concentration scans and, therefore, could be attributed to the melting of the duplex form. The higher component with the higher T_m was associated with a higher ΔH_m and, thus, was attributed to melting of the junction to single strand DNA (Fig. 5.8). Consequently, the stabilizing energy of the junction core interactions can be determined by subtracting the ΔH_m of the high T_m component (at high DNA concentrations) from that of the low T_m component (at low and high DNA concentrations) after appropriate extrapolation of each to a reference temperature (T_{ref}) of 25°C (Eq. 3). The entropic contributions can then be calculated from the ΔH_m and T_m (Eq. 4), and extrapolated to a reference temperature for direct comparison (Eq. 5). The

total free energy of interaction at the reference temperature for the X- vs H- isomer is the negative of the resulting melting free energy ($\Delta G_{m(Tref)}$), Eq. 6) .

$$\Delta H_{m(T_{ref})} = \Delta H_{m(T_m)} + \Delta C_p (T_{ref} - T_m) \qquad \text{Eq. 3}$$

$$\Delta S_{m(T_m)} = \frac{\Delta H_{m(T_m)}}{T_m} \qquad \text{Eq. 4}$$

$$\Delta S_{m(T_{ref})} = \Delta S_{m(T_{ref})} + \Delta C_p \ln\left(\frac{T_{ref}}{T_m}\right) \qquad \text{Eq. 5}$$

$$\Delta G_{m(T_{ref})} = \Delta H_{m(T_{ref})} - T_{ref} \Delta S_{m(T_{ref})} \qquad \text{Eq. 6}$$

The energy of the X-bond relative to the H-bond in solution is calculated by subtracting components of the junction stabilizing energy of the H-isomer (H2J) from that of the X-isomers (Br2J and I2J). This removes the contribution of accessory stabilizing interactions including, for example, the H-bond from the base at N8 to the preceding phosphate oxygen that is essential for formation of the junction in inverted repeat sequences (Hays, Teegarden et al. 2005). In this manner, the relative X- vs H-bond enthalpy, entropy, and total free energy of stabilization can be determined for the Br2J bromine and the I2J iodine X-bonds (Table 5.5). The results indicate that while the iodine X-bond is more enthalpically favorable, in agreement with increased polarizability of this larger halogen, it is entropically less favorable; resulting in a total free energy of stabilization less favorable than that of the bromine X-bond.

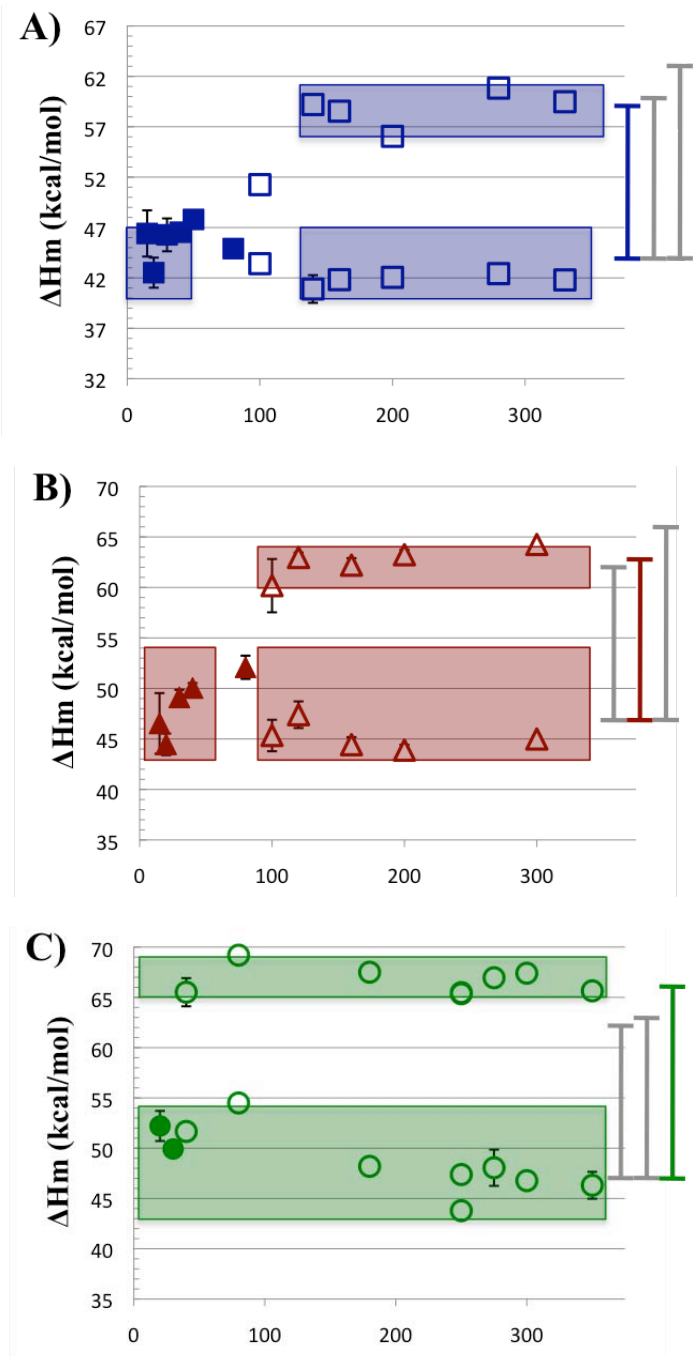


Fig. 5.8. Differential Scanning Calorimetry. Enthalpies of melting (ΔH_m) from duplex or junctions to single-stranded DNA measured by DSC results for H2J (A), Br2J (B), and I2J (C) constructs at increasing DNA concentrations. Solid symbols are DSC data analyzed with a single component (duplex) two-state model, while open symbols are data analyzed by a two-component (duplex and junction) two-state melting model. Boxes represent data for the duplex and junction used to calculate the averages and standard deviations of ΔH_m.

Table 5.5. DSC Results. Individual junction stabilizing interactions, including ΔH ,ΔS, and ΔG, of the Br2J, I2J, and H2J constructs at 25°C are determined by subtracting the stabilization energy of duplex arms from the junction after extrapolating to 25°C. The difference in stabilization between the X- and H-isomer junction stabilizing interactions including, $\Delta\Delta H_{X\text{-}H}$, $\Delta\Delta S_{X\text{-}H}$, $\Delta\Delta G_{X\text{-}H}$, are reported as the difference in the X-isomer junction stabilizing energy and H-isomer junction stabilization energy.

Construct	$\Delta H_{25°C}$ (kcal/mol)	$\Delta S_{25°C}$ (cal/mol·K)	$\Delta G_{25°C}$ (kcal/mol)
Br2J Junction-Duplex	-16.7 ± 0.9	-39 ± 3	-5.3 ± 1.3
I2J Junction-Duplex	-19.0 ± 0.6	-55 ± 2	-2.6 ± 0.9
H2J Junction-Duplex	-13.1 ± 0.9	-43 ± 3	-0.4 ± 1.3

Construct	$\Delta\Delta H_{X\text{-}H}$ (kcal/mol)	$\Delta\Delta S_{X\text{-}H}$ (cal/mol·K)	$\Delta\Delta G_{X\text{-}H}$ (kcal/mol)
Br2J X-bond – H-bond	-3.6 ± 1.3	4.2 ± 4	- 4.8 ± 1.8
I2J X-bond – H-bond	-5.9 ± 1.1	-12.0 ± 4	- 2.3 ± 1.5

5. Conclusions

We present here a comprehensive analysis of the structure-energy relationships of X-bonds (or X-bond like interaction in the case of flourine) involving the halogens F, Cl, Br (Voth 2007), and I as donors in the context of a DNA junction stabilized by this interaction. The crystallographic structures indicate that all of these halogens are capable of forming X-bonds, with their geometries and energies of stabilization in the crystal being highly correlated with the associated polarizability of the halogen as we go down the Group 17 column of atoms. We see from these structures that going through the halogen series from least to most polarizable (F < Cl < Br < I), the distance between the halogen and the oxygen acceptor becomes generally shorter, in terms of the percent reduction in the respective van der Waals radii, and the angle of approach of the acceptor to the halogen becomes more linear. These structural properties are consistent with a progressively stronger X-bond, which is reflected in the stabilization of the X-bonded stabilized X-isomeric form relative to the competing H-bond stabilized H-isomer of the junction. The relative stability of the isomer forms allowed us to tease out the absolute energies of interactions for the X-bonds as well as the competing H-bond in this system.

The crystal structures of the DNA junctions, however, do not tell the entire story. The calorimetric energies of the Br and I containing DNA constructs in solution showed that the X-bonding energies in the crystals largely reflect the enthalpic contributions of the interaction, which draw primarily from the σ-hole formed as a result of polarization of the halogen. It is clear, however, that these two large halogens have different effects on the dynamics, with the Br placed in the center of the junction having very little effect while the larger I in this position greatly reduces the entropy of the system. We can ask at

this point whether this effect on the entropy arises primarily from conformational or solvent effects. It is easy to see how placing the larger, more hydrophobic iodine atom into this relatively small space can result in a more tightly packed junction interior and/or differences in exposure of the halogen surface to the bulk solvent might significantly effect the dynamics of the system.

To address the solvation component, we calculated the solvent accessible surface (SAS) of the X-bonding halogen and the amino group of guanine complementary to the competing H-bond forming cytosine in the X- versus the H-isomeric form (ΔSAS_X and ΔSAS_{NH2}, respectively) (Table 5.6). These were translated to equivalent solvent free energy (SFE) for solvating each group, again in the X- versus the H-isomers (ΔSFE_X and ΔSFE_{NH2}, respectively). We can then estimate the effect of each halogen and amino group on the solvation of the X- and the H-isomer and, consequently, the difference in free energy of solvation for the two conformers ($\Delta G_{IsoX\text{-}IsoH}$). The results of this analysis show that the overall contributions from solvation are slightly stabilizing and fairly consistent across the halogen types ($<\Delta G_{IsoX\text{-}IsoH}>$ = -0.48 kcal/mol, S.D. = 0.32 kcal/mol). This stabilization is associated with increased disorder of solvent molecules as the hydrophobic halogen is moved into the junction core during X-isomer formation. This leaves the entropic difference between the Br and I junctions to come primarily from differences in the conformational entropies. This model is supported by comparing the average crystallographic temperature- or B-factors for the Br2J and I2J structures (Fig. 5.9). From this analysis, we see that both the outside and crossing strands of the junction follow a sinusoidal pattern with peaks at nucleotides $N_3 - N_4$ and valleys at nucleotides $N_6 - N_7$.

Table 5.6. Contribution of solvation to the free energy difference between the X- and H-isomeric forms ($\Delta G_{IsoX\text{-}IsoH}$) of the X-bonded DNA junction. $\Delta G_{IsoX\text{-}IsoH}$ for each halogenated construct of the junction was determined from the solvent accessible surfaces (SAS, calculated with a probe with a 1.4 Å radius using the Discovery Studio program, Accelrys, San Diego). The SAS for the respective halogen in the X- versus the H-isomer were scaled according the atomic solvation parameter (ASP_X for each halogen and ASP_{NH2} for the amino group = kcal/mol/Å^2) derived from the partitioning of these atom types from water to octanol. This served as a model for burying the atomic surfaces into a closed DNA environment (Kagawa, Howell et al. 1993) in order to determine the solvent free energy for the halogen and the NH2 group in the X- and H-isomers (ΔSFE_X and ΔSFE_{NH2}). The $\Delta G_{IsoX\text{-}IsoH}$ for each construct is weigthed according to the stoichiometric ratio of halogens per DNA strand pair for one continuous and one cross-over strand.

Group	ASP_X (kcal/mol/Å^2)
F	0.041
Cl	0.033
Br	0.047
I	0.048
NH_2	-0.043

Junction	ΔSAS_X	ΔSFE_X (kcal/mol)	ΔSAS_{NH2}	ΔSFE_{NH2} (kcal/mol)	$\Delta G_{IsoX\text{-}IsoH}$ (kcal/mol)
F2J	-7.0 Å^2	-0.28	-2.2 Å^2	0.10	-0.18
Cl1J	-11.1 Å^2	-0.37	-0.9 Å^2	0.04	-0.15
Cl2J	-8.7 Å^2	-0.29	3.2 Å^2	-0.14	-0.86
Br1J	-6.1 Å^2	-0.32	1.7 Å^2	-0.07	-0.23
Br2J	-11.4 Å^2	-0.54	1.1 Å^2	-0.05	-0.59
I1J	-17.1 Å^2	-0.82	1.0 Å^2	-0.04	-0.45
I2J	-15.7 Å^2	-0.75	3.3 Å^2	-0.14	-0.89

For the Br2J junction, the two strands show the same pattern with very little variation from the average, suggesting that the variations in B-factor in this structure likely reflect the effect of the sequence. The I2J structure, however, shows that the B-factors of the central nucleotides N_3 to N_7 of the crossing strand are significantly lower than the outside strand, with the lowest average value at the N_6 and N_7 positions. These are the two nucleotides that are engaged in the iodo-X-bond, with the xU_7 uracil forming a strong X-bond to the preceding phosphate oxygen, suggesting that the xU_7 base and, more significantly, the phosphate become highly constrained in forming the iodine X-bond. Thus, analysis of the crystal structure along with the entropic component of the X-bonding energy supports a model that the center of the junction cross-over is just about the right size to accommodate the bromine, but an iodine fits a bit too tightly.

In the context of using X-bonding as a tool for biomolecular engineering, the current study indicates that polarization effects, which can be readily modeled by quantum mechanical calculations (Clark, Hennemann et al. 2007), define the enthalpy of the X-bonding interaction. However, the relationship between the size of the halogen and the space into which it fits, and probably the strength of the interaction (through enthalpy-entropy compensation (Schmidt 2005)) will affect the dynamics of the system, which, together with the enthalpy, determines the overall free energy of interaction. In the current DNA system, the interplay between enthalpic and entropic terms apparently defines bromine as the optimal halogen to form an X-bond to stabilize the junction, even though iodine is more polarizable and, thus, is expected to form a stronger interaction. In other systems, however, one must determine the relationship between these two

competing thermodynamic components separately, which in turn begs for a method that can properly model halogen interactions in molecular dynamics simulations.

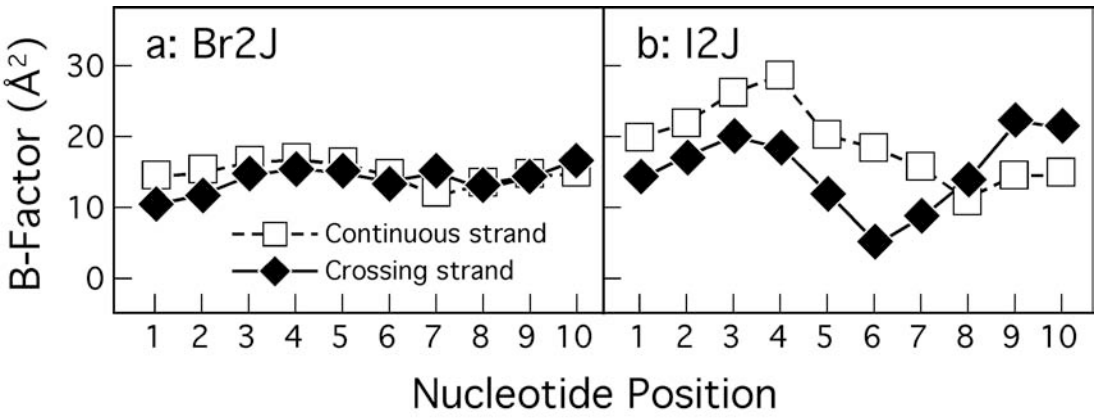


Fig. 5.9. Nucleotide B-factors of the continuous and crossing DNA strands in the X-bonded Br2J (a) and I2J (b) junctions. The B-factor averaged over all atoms at each nucleotide (from 1 to 10) are shown for the crossing strand (solid diamonds) and outside continuous strand (open squares) of each junction construct.

6. References

Brünger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr 54 (Pt 5): 905-921.

Carter, M. and P. S. Ho (2011). "Assaying the Energies of Biological Halogen Bonds." Crystal Growth & Design 11(11): 5087-5095.

Carter, M., Rappé, A., and P. S. Ho (2012). "Scalable Anisotropic Shape and Electrostatic Models for Biological Bromine Halogen Bonds." J. Chem. Theory Comput. Articles ASAP.

Clark, T., M. Hennemann, et al. (2007). "Halogen bonding: the sigma-hole. Proceedings of "Modeling interactions in biomolecules II", Prague, September 5th-9th, 2005." J. Mol. Model. 13(2): 291-296.

Duckett, D. R., A. I. H. Murchie, et al. (1988). "The structure of the Holliday junction, and its resolution." Cell 55: 79-89.

Eichman, B. F., J. M. Vargason, et al. (2000). "The Holliday junction in an inverted repeat sequence: sequence effects on the structure of four-way junctions." Proc. Natl. Acad. Sci., USA 97: 3971-3976.

Fourmigué, M. (2009). "Halogen bonding: Recent advances." Curr Opin Solid State Mat Sci.

Hays, F. A., V. Schirf, et al. (2006). "Solution formation of Holliday junctions in inverted-repeat DNA sequences." Biochemistry 45(8): 2467-2471.

Hays, F. A., A. Teegarden, et al. (2005). "How sequence defines structure: a crystallographic map of DNA structure and conformation." Proc Natl Acad Sci, USA 102(20): 7157-7162.

Hays, F. A., J. M. Vargason, et al. (2003). "Effect of sequence on the conformation of DNA holliday junctions." Biochemistry 42(32): 9586-9597.

Ho, P. S., and Eichman, B. F. (2001). "The crystal structures of DNA Holliday junctions." Current Opin. Struct. Biol. 11: 302-308.

Ho, P. S. a. C., Megan (2011). DNA structure: Alphabet soup for the cellular soul. DNA Replication-Current Advances. D. H. Seligmann, InTech.

Lu, Y., Shi, T., Yong, W. Huaiyu, Y., Xiuhua, Y., Xiaoming, L., Hualiang, J., and W., Zhu (2009). "Halogen Bonding- A Novel Interaction for Rational Drug Design?" J. Med. Chem. 52: 2854-2862.

Matter, H., M. Nazare, et al. (2009). "Evidence for C-Cl/C-Br...pi interactions as an important contribution to protein-ligand binding affinity." Angew Chem Int Ed Engl 48(16): 2911-2916.

Metrangolo, P., Neukirch, H., Pilati, T., and Resnati, G. (2005). "Halogen bonding based recognition processes: A world parallel to hydrogen bonding." Acct. Chem. Res. 38(5): 386-395.

Metrangolo, P. and G. Resnati (2001). "Halogen bonding: a paradigm in supramolecular chemistry." Chemistry 7(12): 2511-2519.

Metrangolo, P. and G. Resnati (2012). "Recommendation from IUPAC Task Group 2009-032-1-100: Definition of the Halogen Bond." Personal Cummunication, In Review.

Murray, J. S., K. E. Riley, et al. (2010). "Directional Weak Intermolecular Interactions: sigma-Hole Bonding." Australian Journal of Chemistry 63(12): 1598-1607.
Kagawa, T. F., M. L. Howell, et al. (1993). "Effects of base substituents on the hydration of B- and Z-DNA: correlations to the B- to Z-DNA transition." Nucleic Acids Res 21(25): 5978-5986.
Otwinowski, Z. and W. Minor (1997). "Processing of x-ray diffraction data collected in oscillation mode." Methods Enzymol. 276: 307-326.
Politzer, P., J. S. Murray, et al. (2007). "Sigma-hole bonding and hydrogen bonding: Competitive interactions." Int. J. Quantum Chem. 107: 3046-3052.
Riley, K. E., J. S. Murray, et al. (2011). "Halogen bond tunability I: the effects of aromatic fluorine substitution on the strengths of halogen-bonding interactions involving chlorine, bromine, and iodine." J Mol Model 17(12): 3309-3318.
Riley, K. E., J. S. Murray, et al. (2009). "Br···O Complexes as probes of factors affecting halogen bonding: Interactions of bromobenzenes and bromopyrimidines with acetone." J Chem Theory Comput 5: 155-163.
Ritter, S. (2009). "Halogen bonding begins to fly." Chemical and Engineering News 87(38): 39-42.
Schmidt, J. R., and Polik, W.F. (2005). "WebMO." 6.0. from http://www.webmo.net.
Shields, Z. P., J. S. Murray, et al. (2010). "Directional Tendencies of Halogen and Hydrogen Bonds." International Journal of Quantum Chemistry 110(15): 2823-2832.
Voth, A. R., Hays, F.A., and Ho, P.S. (2007). "Directing macromolecular conformation by halogen bonds." Proc. Natl. Acad. Sci., USA 104: 6188-6193.
Voth, A. R., Khuu, P., Oishi, K., and Ho, P.S. (2009). "Halogen bonds as orthogonal molecular interactions to hydrogen bonds." Nature Chemistry 1: 74-79.
Xu, Z., Z. Liu, et al. (2011). "Utilization of halogen bond in lead optimization: a case study of rational design of potent phosphodiesterase type 5 (PDE5) inhibitors." J Med Chem 54(15): 5607-5611.
Zhou, P., J. Huang, et al. (2012). "Specific noncovalent interactions at protein-ligand interface: implications for rational drug design." Curr Med Chem 19(2): 226-238.

CHAPTER 6

CONCLUSIONS

1. Enthalpic stabilization of X-bonds increases with increasing polarizability of the halogen involved.

The crystallographic structures and interaction energies of X-bonds observed in DNA Holliday junctions indicate that the enthalpic stabilization of observed X-bonds increases from F<Cl<Br<I with increasing polarizability. The observed X-bond geometries have reduced interatomic distances, as assessed from the sum of the van der Waals radii, and an angle of approach of the acceptor towards the halogen σ-bond (Θ1) that approaches 180°. These trends are consistent with progressively stronger X-bonds, as they maximize the interaction between the halogen σ -hole and the acceptor. The stabilization energy of X-bonds, determined from crystallographic competition assays, increases from F<Cl<Br as indicated by increasing populations of X-bond stabilized X-isomers. The stabilizing energy of these interactions is primarily derived from enthalpic contributions with minimal entropic influence. This is supported by DSC measurment of enthalpic and entropic stabilization energies for bromine X-bond. In addition, movement from solution state to the crystallographic environment should serve to reduce the entropic contribution, as relative movement and temperature are both reduced in crystal conditions. The trend of enthalpic stabilization also increases from bromine to iodine X-bonds as determined from solution state DSC experiments. The culmination of these

results provide experimental evidence that the enthalpic strength of X-bonds increases with increasing polarizability from F<Cl<Br<I in the DNA Holliday junction system. These findings are in agreement with trends in interaction energies as determined by quantum mechanical calculations for multiple donor-acceptor pairs.

2. Entropic effects can provide an opposing force to enthalpic stabilization and must be considered to determine the overall free energy of stabilization of X-bonds.

We have demonstrated that the enthalpic strength of X-bonds increases from F<Cl<Br<I. However, the overall effect of halogen interactions as determined by the total free energy is also dependent on entropic contributions. Entropic stabilization associated with solvent interactions was estimated from solvent accessible surface calculations to be relatively minimal and slightly stabilizing for X-isomer formation involving all four halogens. This stabilization is associated with increased disorder of solvent molecules as the hydrophobic halogen is moved into the junction core. However, our results from solution state DSC experiments demonstrate that in our DNA Holliday junction system the increased enthalpic stabilization of the iodine X-bond was countered by reduced entropy, thereby resulting in an overall decrease in stabilization energy as compared to the analogous bromine halogen interaction. There is evidence for this specific interaction that the loss of stabilization is primarily a result of reduced conformational entropy in the DNA junction. Clearly the balance of entropic and enthalpic contributions must be assessed for halogen interactions in order to engineer the most favorable interaction. Currently, enthalpic energy contributions can be estimated with quantum mechanical calculations for relatively small systems. To estimate entropic contributions and enthalpic

stabilization of larger molecular systems, accurate force fields must be developed such that molecular mechanic and molecular dynamic methods are accessible for halogen modeling. Thus, development of accurate mathematical representation for halogen interactions has been a major goal of research presented in this dissertation.

3. Both anisotropic shape and electrostatic surface potential must be depicted in order to accurately model halogen interactions.

Accurate mathematical representation of halogen interactions is required for use of molecular mechanic and molecular dynamic calculations in the prediction of interaction energy and molecular structures. The development of mathematical representation has also shed light on physical properties of halogens and the influence they have on interaction energies and geometries. The anisotropic electrostatic surface potential of polarized halogens and formation of a positive electrostatic potential at the σ-hole is accepted as the distinguishing characteristic responsible for X-bond formation. Therefore, in order to model halogen interactions it was essential that this property be depicted in mathematical representation. This was achieved by introducing an angular dependence to the electrostatic potential for halogen atoms. However, this modification to classical force fields was not sufficient to accurately reproduce observed interactions for covalently bound bromine. The representation of an anisotropic shape, or effective van der Waals radius, was also necessary. Specifically, a reduction in the van der Waals radius along the σ-bond, at the σ-hole, was required in order to model the close interatomic distances of the halogen and the acceptor atom observed experimentally. Modeling an anisotropic shape of polarized halogens is supported by the σ-hole model in which electron density is

depleted at the σ-hole as a consequence of p_z orbital electron redistribution into the σ-bond. The anisotropic shape distribution of bromine, often referred to as polar flattening, characterized and parameterized for here is in agreement with that observed previously in both crystallographic and ab initio calculations (Nyburg 1979; Peebles, Fowler et al. 1995). Our results suggest that polar flattening, in addition to anisotropic electrostatic distribution, is responsible for the close interatomic distances, often closer than the sum of the respective van der Waals radii, between the acceptor and donor that are a hallmark of X-bonding (Metrangolo and Resnati 2012). Furthermore, we demonstrated that the dispersive interactions of polarized bromine remained largely angular independent while the steric repulsive interactions were significantly affected by polar flattening and require angular dependent representation of the van der Waals radius for accurate mathematical modeling. Our resulting force fields modeled both an anisotropic van der Waals radius and electrostatic surface charge for steric repulsive potentials and electrostatic interactions respectively. Multiple parameters were introduced that allowed control of the details of shape and charge distribution. Parameterization to our model system not only allowed us to recreate quantum mechanically and experimentally determined interaction profiles for polarized bromine but also shed light on halogen properties previously uncharacterized. These properties, described below, are as of yet only supported from parameterization of bromo-uracil and should be verified for the remaining halogens and with additional external validation. However, there is already some support for our conclusions from outside sources.

3.1 The distance dependence of halogen electrostatic interactions is between that of charge-dipole and dipole-dipole interactions.

The distance dependence of electrostatic interactions for polarized bromine as represented by $1/r^n$ was determined to be best fit for n = 2.5. This distances dependence was determined from the parameterization of bromo-uracil to hypophosphite interactions after isolating the electrostatic interactions energies. This was achieved by subtracting out bromo-uracil interactions to a neutral $H_2PO(OH)$ acceptor. A distances dependence represented by n = 2.5 is slightly longer than a classic charge-dipole (n=2) interaction and slightly shorter than a dipole-dipole (n=3) interaction. It is logical that electrostatic interactions with the polarized halogen are reduced as a function of distance in this manner. For example, while the X-bond acceptor is at short distance to the σ-hole it is minimally affected by the ring of negative electrostatic potential maintained by the p_x and p_y orbitals. However, as the acceptor moves further away from the σ-hole the influence of this negative electrostatic potential increases and reduces the electrostatic interaction potential energy. The manner in which the electrostatic potential energy falls off as a function of distance for halogen interactions is characteristic of an interaction between a charge-dipole and dipole-dipole interaction.

3.2 The p_x and p_y orbitals of polarized covalently bound bromine are tipped away from 90° relative to the σ-bond.

The period of the cosine function for the angular dependence is determined by the parameter ν. This parameter was fit to quantum mechanical calculations of two different model systems. The parameterization to the bromo-uracil hypophosphite interaction

resulted in a period of 2.29 while the parameterization of a molecular bromine and helium interaction were best fit with a period of 2.53. A period of 2.0 would indicate a maximum bulge of the effective van der Waals radius and negative electrostatic potential exactly 90° relative to the σ-bond while the increase to 2.3-2.5 indicates these maxima are located closer to 101-109° relative to the σ-bond. This suggests that the p_x and p_y orbitals do not remain perpendicular to the σ-bond but become canted toward the depopulated p_z orbital. For the two models investigated, the extent of tipping appears to be dependent on the extent of polarization. Tipping increases from the bromo-uracil to molecular bromine models correlated with the effective polarization of the bromine. The potential energy map for the Br···H interaction shows the potential effect of p_x and p_y orbital tipping on the geometry of H-bonds formed when bromine acts as an H-bond acceptor. These observations are in agreement with recent calculations on the multipolar electron densities showing the maximum charge concentration for bromine from 95° to 110° relative to the C-Br bond (personal communication Prof. E. Espinosa, Université de Nancy, France).

4. How to use halogens in drug design: a recipe

The culmination of the work in this thesis, in addition to previously published studies, provides a growing set of principles to guide knowledge-based application of halogens in drug design. The halogen class of elements, including fluorine (F), chlorine (Cl), bromine (Br), and iodine (I), show a tendency to become polarized when covalently bound to another atom. Polarization of halogens results in both anisotropic, or non-uniform, electrostatic potential and effective van der Waals radius of the atom. These

characteristics have a direct effect on non-covalent halogen interactions and result in a unique class of interactions called halogen bonds or X-bonds.

Polarization of Group 16 and 17 elements is effectively described by the σ-hole model developed by Politzer et al. (Clark, Hennemann et al. 2007; Politzer, Murray et al. 2007). The σ-hole model details the depopulation of the p_z orbital electrons into the σ-bond, resulting in a reduced electron density along the σ-bond. This exposes the nuclear charge, creating a region of positive electrostatic potential called the σ-hole at the crown of the halogen along the σ-bond (Fig. 2). The remaining electrons in the p_x and p_y orbitals maintain a ring of negative charge roughly perpendicular to the σ-bond (Carter, Rappé et al., 2012). The extent of σ-hole formation is dependent on the polarizability of the halogen, which increases from F<Cl<Br<I (Clark, Hennemann et al. 2007). Increasing the electron withdrawing ability of the covalently bound substituent can also increase relative polarization and σ-hole formation (Riley, Murray et al. 2011). The electropositive σ-hole region is able to form a distinct class of electrostatic interactions termed halogen-bonds or X-bonds. X-bonds are a non-covalent electrostatic attractive interactions between the σ-hole of a polarized halogen, the X-bond donor, and an electron rich Lewis base, the X-bond acceptor, that typically result in interatomic distances closer than the sum of the respective van der Waals radii. X-bonds are similar to H-bonds in strength and share a common class of acceptor atoms. However, X-bonds display a strong directionality preference. The dependence of polarization upon both the involved halogen and the covalently bound substituent proved a broad range of tunable interactions. (Metrangolo and Resnati 2012).

In the past decade there has been a substantial increase into the theoretical understanding and application of X-bonds in material science, crystal engineering, and drug design (Ritter 2009; Metrangolo and Resnati 2001; Fourmigué 2009; Metrangolo 2005; Zhou, Huang et al. 2012; Xu, Liu et al. 2011). X-bonds have been implicated as practical and effective tools for the use in rational, or bottom up, drug design and have been used successfully in a few preliminary cases (Xu, Liu et al. 2011; Matter, Nazare et al. 2009). The recent advances in X-bond characterization can be used to direct the knowledge based development of halogens in drug design. Important considerations include, but are not limited to, the strong directionality preferences of halogen interactions, the ability of halogens to act as H-bond acceptors, the preference of halogen bond acceptors, the orthogonality principal of H- and X-bonds sharing the same acceptor, the effect of polarization on X-bond strength, polarized halogen solvent interactions, and the effect of halogen size on steric interactions. These principles are applied in the following discussion to the selection of X-bond acceptors in a protein binding pocket, optimal placement of the halogen on the lead compound, and selecting which halogen is best suited for a particular interaction.

The process of structure-based drug or ligand design often begins with identification of a small molecule that binds, with variable affinity, to the target of choice by high-throughput screening. In addition to traditional large-scale screening of small molecule libraries approaches in vitro, a large number of potential ligands can be screened for fitting into the binding pocket of a receptor through in-silico, or computational, database searching. Once a lead compound is identified, the process of lead optimization begins in which the small molecule is modified in attempts to increase

binding affinity and/or efficacy of the drug (Ilag, Ng et al. 2002). An alternative to the process of lead identification and optimization is the approach of "building" ligands from scratch, commonly referred to as de novo design. In this process the ligands are designed within the constraints of a known binding pocket by assembling structural fragments in a stepwise manner (Wang, Gao et al. 2000; Yuan, Pei et al. 2011).

Halogens share a common set of acceptors to H-bonds, including both charged and uncharged oxygens, amino and imino nitrogens, sulfurs, and aromatic rings in biological systems. Thus, the number of possible acceptors is large and the challenge often becomes narrowing down the number of acceptors to pursue. One approach to identifying preferred acceptors is to consider structures of known X-bonds and identify specific trends. One of the first surveys carried out in 2004 identified 113 X-bonds in protein and nucleic acid structures, summarized in (Auffinger, Hays et al. 2004). The majority of the structures were of protein-ligand complexes. The most common acceptors were carbonyl oxygen (69%) and hydroxy oxygen of protein side chains (15%). Of the X-bonds observed with carbonyl oxygens, a majority of these were directed toward the π-system of the O=C bond rather that the lone pairs. A more recent survey in 2009 found similar results identifying 397 structures in the PDB with X-bond type interactions involving Cl, Br, and I. A majority of these, 53%, were also found to involve oxygen as an acceptor. Of these interactions it was reported that ~75% of them were to the backbone carbonyl oxygen. The prevalence of X-bonds to carbonyl oxygen was attributed to the availability of both their lone pair and π-system electrons for interaction. The second most prevalent X-bonds found in this survey, at 33% a total of 146 contacts, were to π-orbital system of aromatic amino acid side chains. This was followed by

contacts to nitrogen, where roughly half of these were to the backbone amine nitrogen. As a whole, these surveys call attention to the preference of halogenated small molecules to bind to the carbonyl oxygen followed by interactions with the π-system of aromatic side chains. Of the X-bonds to carbonyl oxygens, an astonishing 80% or more surveyed in 2009 showed associated H-bonds occurring to the carbonyl oxygen at the same time (Voth 2009). Simultaneous H- and X-bonds to common carbonyl oxygen were found to have an orthogonal relationship in respect to both energetics and prefer geometry. In other words, the two interactions are energetically independent and a preferred geometry of 90° in relation to each other. Therefore, X-bonds can be engineered toward backbone carbonyl oxygen participating in existing H-bonds, including those essential to secondary structure formation, without perturbing the established H-bonding interactions. By targeting X-bonds engineered in this fashion, one can feasibly avoid unnatural perturbation of the native protein structure and help prevent the possibility of drug resistant mutations. It should be noted that these observations are only from existing structures in the PDB and do not represent the only possibilities for stabilizing interactions. However, they do give us an indication of what can work when selecting X-bond acceptors in a protein system.

Placement of the halogen on a lead compound should optimize the strong directional preferences of X-bonding interactions. Primarily, the increased enthalpic stabilization of X-bonds when moving toward a linear approach of the acceptor towards the σ-hole must be considered. This has been observed in countless surveys and is supported by interpretation of X-bond stabilization via the σ-hole theory. In addition, the enthalpic strength of X-bonds is well correlated with the positive electrostatic potential of

the σ-hole and demonstrated experimentally and in quantum mechanical calculations to be related to the extent of polarization. Up until recently these two significant characteristics of covalently bond halogens and associated X-bonds were not accurately modeled in computational prediction methods. The development of anisotropic electrostatic and steric force fields has made modeling these influences possible in molecular mechanic, drug docking, and by extension molecular dynamic software (Chapter 4). These force fields are able to model not only the anisotropic properties of halogens, but also the extent of polarization, which increases with polarizability of the halogen (F<Cl<Br<I) and electron withdrawing ability of the substituent.

Entropic contributions, namely solvent and conformational entropies, must also be considered to determine the overall free energy of interaction of the ligand and associated protein-binding pocket. These contributions can be estimated using molecular dynamic simulations based on anisotropic force fields; these methods are currently under development. Important to these considerations is the increase in size of the halogens from F<Cl<Br<I. In addition to determining the polarizability, the size of the halogen may also affect the conformational entropy of a system (Chapter 5). Therefore, halogen selection for a given interaction should optimize the balance between both enthalpic and entropic stabilization.

The above considerations will help guide the use of halogens in drug development and lead optimization. Carbonyl oxygen, π-systems of aromatic side chains, and backbone nitrogen atoms represent the overwhelming majority of X-bond acceptors observed in known protein-ligand complexes. Multiple surveys have demonstrated the prevalence of X-bonds to the carbonyl oxygen and the orthogonal relationship of these

interactions to existing H-bonds sharing the same acceptor. Once a set of X-bond acceptors has been selected the placement of a halogen on the lead compound should optimize the directional preference of acceptor approach toward the σ-hole of the polarized halogen. The optimal halogen for a given interaction and geometry should be selected to balance the enthalpic stabilization, which increases with increasing polarization, and entropic stabilization, influenced by the size and solvent interaction of a given halogen. The above considerations are an attempt to synthesize the current knowledge of halogen properties and X-bond interaction profiles for use in knowledge base drug design. Some of these considerations, such as enthalpic dependence on directionality and polarizability of the halogen, can be accounted for in computer simulators implementing anisotropic force fields developed specifically for halogen interactions. However, many of these considerations, such as the orthogonality principal and entropic effects are not yet modeled by drug docking and dynamic programs and must be implemented directly by the scientist or engineer involved.

5. References

Auffinger, P., F. A. Hays, et al. (2004). "Halogen bonds in biological molecules." Proc. Natl. Acad. Sci., USA 101(48): 16789-16794.

Carter, M., Rappé, A., and P. S. Ho (2012). "Scalable Anisotropic Shape and Electrostatic Models for Biological Bromine Halogen Bonds." J. Chem. Theory Comput. Articles ASAP.

Flanagan, A. e. a. (2002). "A Revolution in R&D: how genomics and genetics are transforming the biopharmaceutical industry." Boston Consulting Group Report 26(November, 1-64).

Ilag, L. L., J. H. Ng, et al. (2002). "Emerging high-throughput drug target validation technologies." Drug Discovery Today 7(18): S136-S142.

Metrangolo, P. and G. Resnati (2012). "Recommendation from IUPAC Task Group 2009-032-1-100: Definition of the Halogen Bond." Personal Cummunication, In Review.

Nyburg, S. (1979). "Polar flattening: Non-spherical effective shapes of atoms in crystals." Acta Cryst A35: 641-645.

Peebles, S. A., P. W. Fowler, et al. (1995). "Anisotropic Repulsion in Complexes B-Center-Dot-Center-Dot-Center-Dot-Cl-2 and B-Center-Dot-Center-Dot-Center-Dot-Hcl - the Shape of the Chlorine Atom-in-a-Molecule." Chemical Physics Letters 240(1-3): 130-134.

Riley, K. E., J. S. Murray, et al. (2009). "Br···O Complexes as probes of factors affecting halogen bonding: Interactions of bromobenzenes and bromopyrimidines with acetone." J Chem Theory Comput 5: 155-163.

Shields, Z. P., J. S. Murray, et al. (2010). "Directional Tendencies of Halogen and Hydrogen Bonds." International Journal of Quantum Chemistry 110(15): 2823-2832.

Voth, A. R., Khuu, P., Oishi, K., and Ho, P.S. (2009). "Halogen bonds as orthogonal molecular interactions to hydrogen bonds." Nature Chemistry 1: 74-79.

Wang, R. X., Y. Gao, et al. (2000). "LigBuilder: A multi-purpose program for structure-based drug design." J Mol Model 6(7-8): 498-516.

Yuan, Y. X., J. F. Pei, et al. (2011). "LigBuilder 2: A Practical de Novo Drug Design Approach." Journal of Chemical Information and Modeling 51(5): 1083-1091.