DISSERTATION

HALOGEN BONDS IN BIOLOGICAL MACROMOLECULES

Submitted by

Matthew Robert Scholfield

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

Fall 2016

Doctoral Committee:

Advisor: P. Shing Ho

John Fisk
Olve Peersen
Santiago Di Pietro

ABSTRACT

HALOGEN BONDS IN BIOLOGICAL MACROMOLECULES

The purpose of this dissertation is to study how halogen bonds (X-bonds) affect the stability of biological macromolecules and to develop a set of empirical mathematical equations that can provide insight into the anisotropic nature of covalently bound halogens. To achieve this end, we first conducted a detailed analysis of the Protein Data Bank (PDB) to determine the prevalence of X-bonding in biological macromolecules, which allowed us to study the geometrical trends associated with X-bonding. Quantum mechanical (QM) calculations were also applied to determine how the strength of X-bonds interaction could be "tuned." The next chapter used QM calculations to help parameterize an equation that can model the anisotropic size and charge of covalently bound chlorine, bromine and iodine. The energies obtained from this equation were validated on experimentally determined X-bond data by differential scanning calorimetry (DSC) in DNA holiday junctions and were found to nearly duplicate the energies obtained in the solution state experiments. In the final chapter, we engineer X-bonds into the structure of T4 lysozyme to studying structural and thermodynamic effects of X-bonds on protein. X-bonds were introduced into the enzyme via site-specific non-canonical amino acid incorporation and then the structure and stability of the protein were assayed via X-ray crystallography and DSC, respectively. The culmination of this work has elucidated many concepts that need to be considered when trying to engineer new biologically based materials with halogens.

ACKNOWLEGMENTS

It is hard for me to know who to acknowledge for this achievement as this journey started long before I came to Colorado. It has been much more difficult than I could have imagined and I can say with confidence that many of my grade school teachers would never have expected me to have earned a PhD, let alone graduate high school. They believed this because I was unable to read and, many teachers gave up on me. Perhaps it was their doubt in me that caused me to strive to the highest academic achievement. I was very fortunate to have a mother and farther that never gave up on me and they continued to fight for my right to get an education. They found me the best teachers, Betty Holle, Karen Toledo, Jan Secunda, and Kathlyn Grahm. These teachers believed in me and worked very hard to identify, and develop strategies I could use to work around my learning disability proving to me that it is could also be successful. Without them, I would have never made it. I owe much of my success to them, thank you.

There were many wonderful instructors at my undergraduate university and I want to thank them all for furthering my development and encouraging my scientific curiosity. I truly believe that New Mexico Tech has brilliant talent. I want to especially thank Dr. Snezna Rogeli who set up independent study classes for me so I could get more exposure to lab work. She helped me find my passion for research and encouraged me to keep striving for answers.

I want to thank everyone at Colorado State University. I cannot name everyone who contributed to my success but know that my fondest memories are of spending late nights in lab and talking with other graduate students. Those conversations of philosophy, and sciences are some of the most rewarding and fulfilling experiences that I will take with me. Your friendship and support is greatly appreciated. I want to especially thank my committee, Dr. Olve Peersen,

Dr. Santiago Di Pietro, and Dr. Nick Fisk for their mentorship and suggestions that contributed to my development as a scientist. A special thank you goes to Dr. Norman Curthoys, who retired before I was able to defend my dissertation. He is a great mentor and scientist.

I cannot fully express the gratitude that I hold for my advisor Dr. Shing Ho. I specifically chose him because he has incredibly high standards for his students and I believed that he would help me develop into the best possible scientist. I really appreciate the patience he gave me to learn and grow. Thank you for all you have done.

I must thank my family and friends for supporting me. Crystal is a great source of inspiration for me and helped me develop as a scientist. She also was there to help me through the more difficult challenges of graduate school. I'm most grateful for her friendship. My two brothers, Daniel and Gregory, for being good role models for me to live up to and for providing me insight on all topics scientific, philosophical, or personal. Bonnie, who is an invisible force for me and whose silent affection kept me grounded. Again, I must thank my mother and father for their love and support and without whom I would have never started down this road.

The biggest thank you goes to my wife Kristy for always being there for me. For understanding that when I said I would be home in 15 minutes, that I was really saying I would be home in 3-5 hours. She has always loved me and has been more patient than anyone should have to be. Thank you for loving me, thank you for being here, and for seeing greatness in me even when I cannot see it myself.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

I. THE SIGNIGICANCE OF STUDYING HALOGEN BONDS IN BIOMOLECULAR SYSTEMS

Halogens are an important molecular design tool, and their application in the life sciences have been increasing dramatically to aid in the development of a variety of new materials. In the chemical sciences, halogens have been used to developed novel nanomaterials[1], supramolecular materials[2], molecular conductors[3], molecular receptors[4], and liquid crystal lattices[5]. A field of study where halogens can have a substantial impact is in the biochemical and biomedical sciences, specifically where halogens can be used to design new therapeutics against clinically significant targets[6–8]. The benefits of halogens in these sciences can be further enhanced with the aid of computer modeling to rationally design new medicines and materials[9]. It is evident that as more information is learned about halogens and how they behave in biochemical systems, their applications will become even more effective and impactful. Despite halogens being used more frequently to develop new materials, the specific properties of halogens and how they behave in these materials is still poorly understood. To fully unlock the potential of halogens, we must take a comprehensive approach to study them in order to understand how to use them more effectively.

In the 1990's, a new term appeared in the scientific literature which was used to described a molecular interaction called the halogen bond (X-bond)[10–12]. This interaction described the ability of a covalently bound halogens to form short, stabilizing non-covalent

interaction with electron-rich Lewis bases such as oxygen or nitrogen (Figure 1.1A). Since then, there has been an abundance of research seeking to elucidate the physiochemical properties and the significance of this non-covalent interaction in biochemical systems. In 2004 Auffinger[13] and colleagues conducted a survey on the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PBD) for the existence of X-bonds in protein and nucleic acid structures and found evidence of 116 distinct X-bond interactions. With this study, they determined that halogens, via the X-bond, can drastically affect the binding affinity of halogenated inhibitors and can alter the structure of biomolecular systems. They learned that the introduction of halogens into biological systems have the potential to lead to unexpected and unexplainable effects from what would be expected. Indeed, not understanding how X-bonds contribute to the stability of such systems can result in unsuccessful and improperly designed experiments and possibly the failure of these developing sciences, which is why it is critical for the physiochemical properties of X-bonds to be understood. This impact does not only affect the basic sciences, it could potentially have negative economic impacts with wasted efforts and money used to develop these failed materials and medicines[6,8]. If we understand how halogens contribute to the stability of these systems, we can develop more efficacious, potent medicines and more effective biologically-based materials. In order to understand how a particular X-bond interaction may contribute to a system, we need to first understand and describe the structure-energy relationship of X-bonds in a quantitative way.

One of the primary questions that must be answered to describe X-bonding is “How can a short, stabilizing non-covalent interaction between a halogen, which is considered electron-rich, form with another electron-rich Lewis base such as an oxygen atom?” Our classical understanding of physicochemistry would predict that two negatively charged atoms would repel

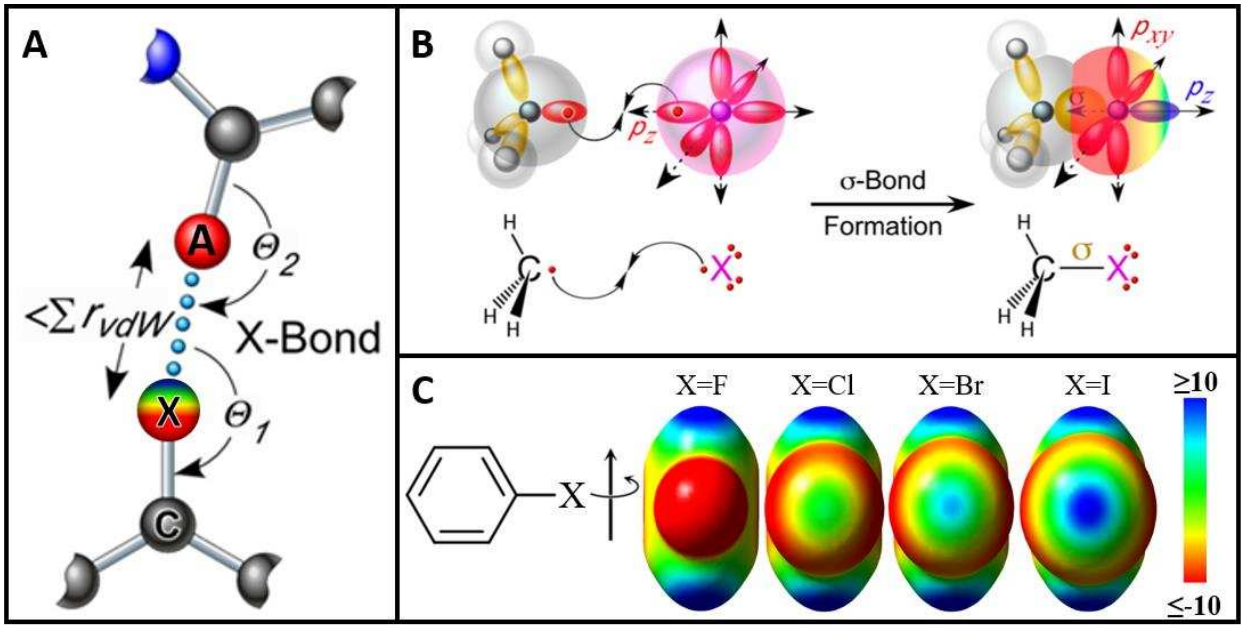

Figure 1.1. Halogen bond formation and polarization of halogens. (A) The geometry of a halogen bond (X-bond). The interaction is characterized as shorter than the sum of the van der Waals radii ($\sum R_{vdW}$) of the halogen (X) to an acceptor atom (A) like oxygen, nitrogen, or sulfur. The halogen of the X-bond is shown with the electropotential polarized from positive (blue) to neutral (green) to negative (red). The approach of the acceptor to the halogen and halogen to the acceptor are labeled as Θ_1 and Θ_2, respectively. (B) The σ-hole model of halogen bonds. The formation of a single C—X covalent bond (σ -bond, yellow) pairs an electron from the carbon with one from the valence p_z-orbital of the halogen. The depopulation of the p_z-orbital creates an electropositive crown (blue) and flattening of the atomic radius opposite the σ -bond, while the $p_{x,y}$-orbitals remain fully occupied, resulting in an electronegative annulus perpendicular to the covalent bond. (C) Electrostatic potential of halogenated benzene at 4-position. The DFT calculated electrostatic potentials (from >10 kcal/mol, in blue, to <-10 kcal/mol, in red) show σ -holes that increase in size and intensity as the size of the halogen increases from F to Cl to Br to I. Figure adopted from previous publication[13].

one another, but this is clearly not the case with X-bonding. To address this seemingly paradoxical behavior of X-bonds, one only needs to examine the distribution of charge across the surface of the halogen when it is a covalent substituent. An example is C—X, where the halogen (X) has formed a single covalent bond (—), known as a σ-bond, to a carbon atom (C). Covalently bound halogens experience polarization as a result of forming a σ-bond[14,15]. The polarization occurs because an electron from the valence of the halogen must participate in the formation of the molecular orbital, which results in a depopulation of elections opposite the bond (Figure 1.1B). The ultimate result is an anisotropic distribution of charge across the surface of the halogen with a positive electrostatic crown diametrically opposed to the C—X σ-bond and a negatively charged ring around the equator of the halogen. The shape of the halogen is also distorted in what is called "polar flattening." This change in shape results in shortening of the halogen's effective radius opposite of the C—X bond and lengthening perpendicular to the bond. Furthermore, halogens polarize to different extents with increasing polarization associated with increasing atomic mass. This means that fluorine is the least polarized, iodine being the most polarized, with chlorine and bromine falling in-between the two for the common stable halogens (Figure 1.1C). Astatine and element 117, the last two halogens in group 17, are intentionally left out left out of this dissertation because it is currently unclear if they follow this polarization trend. In addition, astatine isn't normally considered because it is radioactive and has a short half-life. For these reasons, the main halogens that are considered useful in the biological context are F, Cl, Br, and I. There will be a more comprehensive description of the polarization of halogens in subsequent chapters, but this brief explanation provides the basis for how an electron-rich halogen can have a short, stabilizing non-covalent interaction with an election-rich Lewis base.

The unique properties that allow halogens to polarize places geometrical constraints on the orientations that an X-bond can adopt. For example, if a Lewis base, such as oxygen, approaches a covalently bound halogen, the most favorable approach angle would be linear[13,16,17] with respect to the σ-bond (C—X···O), because the negatively charged oxygen interacts with the positive electrostatic crown of the halogen. The most unfavorable approach is perpendicular to the σ-bond, because the negatively charged oxygen will be repelled by the electronegative ring around the equator of the halogen. Additionally, we now know that the halogen's effective radius is shortened at a linear approach angle[17,18], allowing for interaction distances between the halogen and oxygen to be shorter than what would be predicted by their standard radii. Understanding the geometrical constraints of X-bonds allows for a more detailed definition of the interaction; X-bonding is primarily a stabilizing, electrostatic, non-covalent interaction between a covalently bound halogen and Lewis base with the interaction distance between the two atoms being shorter than the sum of their van der Waals radii, and near linear approach angle from the covalently bound halogen to the Lewis base[19]. This deeper insight into halogens should be considered when using them to develop new biologically-based materials and therapeutics.

Understanding that covalently bound halogens are not simply electron-rich atoms has a large impact in how halogens should be used, for example, in modeling halogens computationally to develop new pharmaceutical drugs. Rationally designed medicines are developed with the aid of computer modeling to predict how the addition or removal of an atom to a new drug candidate will increase or decrease its binding affinity to the target. Most atoms are modeled as spherical objects, with an isotropic distribution of charge across their surface. We now know that these assumptions are incorrect and modeling halogens this way would incorrectly predict unfavorable interaction between the halogen and the Lewis base, and can lead

to the failure to identify more potent compounds. Not properly predicting how X-bonds behave in biochemical systems further stresses the point that, to understand how a particular X-bond may contribute to the stability of a system, we need to first understand the structure-energy relationship of X-bonds.

The work presented in this dissertation is primarily concerned with how X-bonds contribute to the stability of biological systems, and specifically the structure-energy relationship of X-bonds in these systems. To address this research goal, this dissertation will strive to answer the following questions: (1) How is the strength of an X-bond to a biological Lewis base affected when the polarity of the environment is changed, the Lewis base becomes more electron donating, the halogen is changed, or the electron withdrawing ability of the halogen's covalently bound substituent is altered? (2) Is there an optimum X-bond interaction distance, and is that distance dependent on the halogen? (3) Can all of the biologically important halogens be modeled computationally with the force field for biological halogen bonds (*ff*BXB), and can we learn about the physicochemical properties of covalently bound halogens from the parameterized *ff*BXB? (4) How does the strength of X-bonds compare to the strength of H-bonds and steric interactions in protein systems?

This dissertation addresses the above questions by first discussing in Chapter 2 how the strength of X-bonds can be "tuned." This work was done with the aid of quantum mechanical calculations, which is used to investigation how the strength of an X-bond is affected when the electron donating ability of the Lewis base is altered, the electron withdrawing ability of the halogen's covalently bound substituent is changed, or the polarization of the solvent increases. Chapter 2 then conducts an updated survey of the PDB, in which biological X-bonds are identified and the optimal X-bond interaction distance is determined. Furthermore, the general

concepts that are important for X-bonding are discussed and explained from the perspective of a biologist. Chapter 3 seeks to extend the *ff*BXB to all of the biological halogens and presents the testing which demonstrates that this molecular mechanics approach can properly model X-bonds. This work was the first demonstration of a molecular mechanics approach directly modeling the experimentally determined energies of X-bonds with explicit geometries with a nearly 1 to 1 relationship. The parameters from the *ff*BXB can be readily interpreted in terms of the physicochemical properties of the halogen atoms which follow periodic trends. These results provide insight on the average radii and polar flattening of the halogens, and they show that the character of the X-bond interaction is between a dipole-charge and dipole-dipole interaction. Finally, Chapter 4 discusses how the strength of X-bonds compares to the strength of H-bonds and steric interactions in proteins. This was done by engineering halogenated amino acids into T4 lysozyme such that X-bonds were likely to form. The formation of the bonds were validated with X-ray crystallography and the thermostability was measured with differential scanning calorimetry (DSC). Taken together, this dissertation describes many concepts that must be taken into consideration when using X-bonds and it provides a basis for how X-bonds can contribute to the stability of biological systems. It also provides clarity on how X-bonds can be used in biological systems to make more efficient and effective therapeutics, biologically based materials, and novel recognition motifs for protein and nucleic acids engineering.

REFERENCES

1. Shirman, T., Kaminker, R., Freeman, D. & van der Boom, M. E. Halogen-bonding mediated stepwise assembly of gold nanoparticles onto planar surfaces. *ACS Nano* **5,** 6553–63 (2011).
2. Priimagi, A., Cavallo, G., Metrangolo, P. & Resnati, G. The halogen bond in the design of functional supramolecular materials: recent advances. *Acc. Chem. Res.* **46,** 2686–95 (2013).
3. Lieffrig, J., Pennec, R. Le, Jeannin, O., Auban-Senzier, P. & Fourmigué, M. Toward chiral conductors: combining halogen bonding ability and chirality within a single tetrathiafulvalene molecule. *CrystEngComm* **15,** 4408 (2013).
4. Chudzinski, M. G., McClary, C. A. & Taylor, M. S. Anion receptors composed of hydrogen- and halogen-bond donor groups: modulating selectivity with combinations of distinct noncovalent interactions. *J. Am. Chem. Soc.* **133,** 10559–67 (2011).
5. Nguyen, H. L., Horton, P. N., Hursthouse, M. B., Legon, A. C. & Bruce, D. W. Halogen bonding: a new interaction for liquid crystal formation. *J. Am. Chem. Soc.* **126,** 16–7 (2004).
6. Hernandes, M., Cavalcanti, S. M., Moreira, D. R., de Azevedo Junior, W. & Leite, A. C. Halogen Atoms in the Modern Medicinal Chemistry: Hints for the Drug Design. *Curr. Drug Targets* **11,** 303–314 (2010).
7. Roughley, S. D. & Jordan, A. M. The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* **54,** 3451–79 (2011).
8. Wilcken, R., Zimmermann, M. O., Lange, A., Joerger, A. C. & Boeckler, F. M. Principles

and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **56,** 1363–1388 (2013).

9. Ibrahim, M. A. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* **32,** 2564–74 (2011).

10. Lommerse, J. P. M., Stone, A. J., Taylor, R. & Allen, F. H. The Nature and Geometry of Intermolecular Interactions between Halogens and Oxygen or Nitrogen. *J. Am. Chem. Soc.* **118,** 3108–3116 (1996).

11. Legon, A. Prereactive Complexes of Dihalogens XY with Lewis Bases B in the Gas Phase: A Systematic Case for the Halogen Analogue B small middle dot small middle dot small middle dotXY of the Hydrogen Bond B small middle dot small middle dot small middle dotHX. *Angew. Chem. Int. Ed. Engl.* **38,** 2686–2714 (1999).

12. Metrangolo, P., Neukirch, H., Pilati, T. & Resnati, G. Halogen bonding based recognition processes: A world parallel to hydrogen bonding. *Acc. Chem. Res.* **38,** 386–395 (2005).

13. Auffinger, P., Hays, F. A., Westhof, E. & Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 16789–94 (2004).

14. Clark, T., Hennemann, M., Murray, J. S. & Politzer, P. Halogen bonding: the sigma-hole. Proceedings of 'Modeling interactions in biomolecules II', Prague, September 5th-9th, 2005. *J. Mol. Model.* **13,** 291–6 (2007).

15. Politzer, P., Murray, J. S. & Lane, P. σ-Hole bonding and hydrogen bonding: Competitive interactions. *Int. J. Quantum Chem.* **107,** 3046–3052 (2007).

16. Ouvrard, C., Le Questel, J.-Y., Berthelot, M. & Laurence, C. Halogen-bond geometry: a crystallographic database investigation of dihalogen complexes. *Acta Crystallogr. Sect. B Struct. Sci.* **59,** 512–526 (2003).

17. Politzer, P., Lane, P., Concha, M. C., Ma, Y. & Murray, J. S. An overview of halogen bonding. *J. Mol. Model.* **13,** 305–311 (2007).

18. Politzer, P., Riley, K. E., Bulat, F. A. & Murray, J. S. Perspectives on halogen bonding and other σ-hole interactions: Lex parsimoniae (Occam's Razor). *Comput. Theor. Chem.* **998,** 2–8 (2012).

19. Desiraju, G. R. *et al.* Definition of the halogen bond (IUPAC Recommendations 2013). *Pure Appl. Chem.* **85,** 1711–1713 (2013).

CHAPTER 2

HALOGEN BONDING (X-BONDING): A BIOLOGICAL PERSPECTIVE[†]

SUMMARY

The concept of the halogen bond (or X-bond) has become recognized as contributing significantly to the specificity in recognition of a large class of halogenated compounds. The interaction is most easily understood as primarily an electrostatically driven molecular interaction, where an electropositive crown, or σ-hole, serves as a Lewis acid to attract a variety of electronrich Lewis bases, in analogous fashion to a classic hydrogen bonding (H-bond) interaction. We present here a broad overview of X-bonds from the perspective of a biologist who may not be familiar with this recently rediscovered class of interactions and, consequently, may be interested in how they can be applied as a highly directional and specific component of the molecular toolbox. This overview includes a discussion for where X-bonds are found in biomolecular structures, and how their structure–energy relationships are studied experimentally and modeled computationally. In total, our understanding of these basic concepts will allow X-bonds to be incorporated into strategies for the rational design of new halogenated inhibitors against biomolecular targets or toward molecular engineering of new biological-based materials.

[†]Matthew R. Scholfield, Crystal M. Vander Zanden, Megan Carter, and P. Shing Ho
Department of Biochemistry and Molecular Biology, Colorado State University
Fort Collins, CO 80523

Reproduced with permission from Protein Science

I. INTRODUCTION

Nature is highly adept at taking advantage of the laws of chemistry and physics to evolve highly complex biological systems. For the most part, however, terrestrial biology makes rather restrictive use of the elements from the periodic table, being based primarily on six elements (C, H, O, P, N, and S), along with a smattering of Group I and II and transition metals. Halogens (the Group VII elements) are not widely discussed in biology, except in terms of the effects of their anionic (fluoride, chloride, bromide, and iodide) forms on properties such as the osmolarity and ionic strengths of solutions. In chemistry, however, molecular halogens are important for their high reactivity; consequently, halogenation is seen as an important step in synthetic organic chemistry. The prevalence of halogenated compounds has made them widely used as inhibitors against biomedically important targets, and with the halogens often providing several orders of magnitude in specificity for these targets. For most of the history of biochemistry and medicinal chemistry, however, halogens have been treated primarily as electron-rich, lipophilic atoms that do not, in themselves, participate in specific molecular interactions that contribute to the recognition of ligands by proteins.

The recent ‘‘rediscovery’’ of halogen bonds (or X-bonds, Figure 2.1) as highly directional, short-range electrostatic interactions with electron-rich atoms (oxygens, nitrogens, and sulfurs) provides us with a renewed appreciation for the role that this class of elements plays in recognition and, potentially, as a new tool for biomolecular design and engineering. In this review, we will discuss the physicochemical basis for X-bonds and how their energies are estimated from both theoretical and experimental bases. We will then explore the types of X-bonds seen in biomolecular complexes and how they are being applied in the design of new

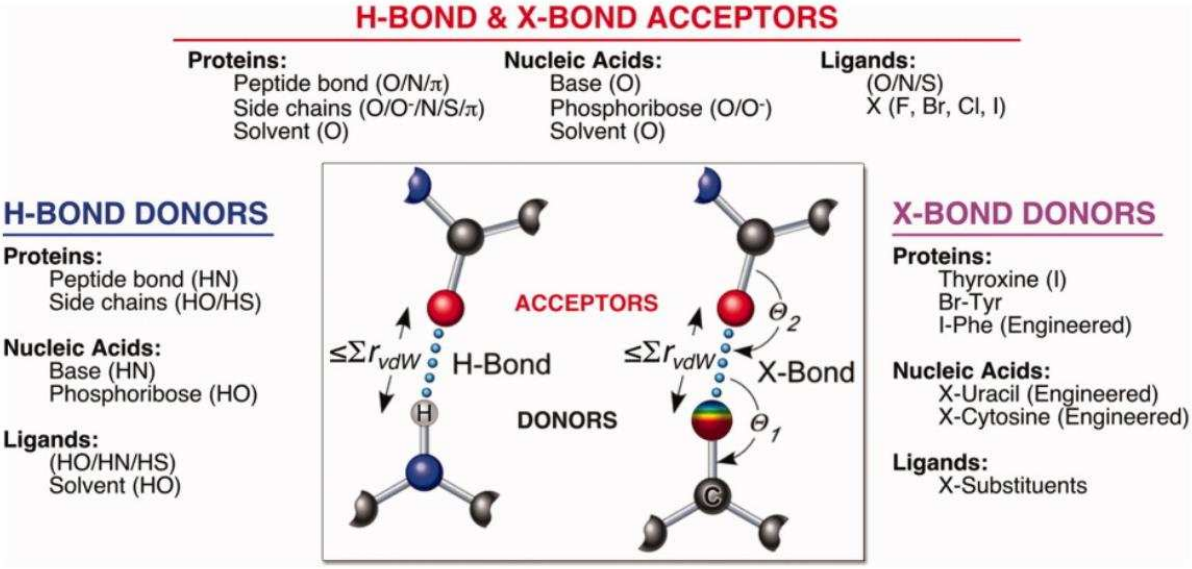

Figure 2.1. Hydrogen and halogen bonds. The geometries and types of donor and acceptor atoms are compared for classic hydrogen bonds (H-bonds) and halogen bonds (X-bonds) seen in biomolecular systems. Each interaction is characterized as being shorter than the sum of the van der Waals radii ($\sum R_{vdW}$). of the respective atoms. The halogen of the X-bond is shown with the electropotential polarized from positive (blue) to neutral (green) to negative (red). The approach of the acceptor to the halogen and halogen to the acceptor are labeled as Θ_1 and Θ_2, respectively. Acceptors that include the delocalized electrons of the amide peptide bond or the ring of an aromatic amino acid are listed as π.

inhibitors and to control molecular structures. We will start, however, with a brief history of this unique interaction and why it has become interesting from a biological perspective.

The first reports of halogens potentially serving as Lewis acids came in the mid to late 1800s with the description of complexes formed between molecular halogens (I_2, Br_2, and Cl_2) and ammonia and methylamines.[1] The detailed physical descriptions of such interactions came from the studies of Odd Hassel[2] in the mid-20th century on the crystal structures of molecular halogens in complex with organic Lewis bases, where it was observed that the interatomic distance from, for example, the bromine of Br_2 to the oxygen of dioxane could be as short as 2.7 Å , or >20% shorter than the sum of their respective van der Waals radii ($\sum R_{vdW}$). At that time, the interactions were called ‘‘charge transfer bonds,’’ referring to a bonding model in which the charge from the lone pairs of an electron-rich atom, such as an oxygen or nitrogen, is transferred to a Lewis acid, in this case the halogen, in a manner similar to what is commonly observed with transition metal complexes. From that point on, however, the field appeared to be relatively quiet until about 1990, when a new term (halogen bond) started to appear in the chemical literature.[3–5] These short-range interactions were being used to control the assembly of organic molecules in crystals and in solution, among other things. This new name now reflects the more electrostatic character of the interaction, similar to classic hydrogen bonding, rather than their charge-transfer nature.

The re-emergence of X-bonds in chemistry has, until recently, been largely invisible to the biological community. Indeed, when short-range interactions between halogens and Lewis bases were first noted in complexes of proteins and nucleic acids,[6,7] they were unexplainable from a simplistic understanding of periodic chemistry. A survey of the Protein Data Bank (PDB[8]) by Auffinger et al.[9] in 2004, however, showed that such interactions were common in the

crystal structures of biomolecular systems, particularly in complexes of halogenated ligands with their protein targets, but had remained ignored as significant contributors to specificity in molecular recognition. There was, at the time, at least one group trying to apply the concept of X-bonds for the rational design of new inhibitors against Factor Xa to serve as anticoagulants—unfortunately, this work had not been published.[10] Since then, the literature to characterize biological X-bonds, particularly as potential tools for rational drug design and molecular engineering has grown exponentially. Still, the contribution of X-bonds to any particular biomolecular structure is typically realized only in hindsight, and the biological studies to engineer X-bonds for such applications have lagged significantly behind the chemical and material sciences fields. This can be attributed to, until recently, a dearth of accurate and accessible methods to model the interaction in macromolecular systems.

For this review, we will focus on recent advances in the study of X-bonds from a perspective that helps to inform the biological community of their relevance and potential in conferring specificity. We will not attempt to summarize the advances made in the areas of the chemical, theoretical, and material sciences—there are already a number of recent reviews that focus on these particular areas of study.[5,11–13] We will, instead, discuss the current understanding of X-bonding, where the interactions are found in biology, their structure–energy relationships, how these relationships are modeled, and how this is now starting to come together in a manner that allows X-bonds to potentially become a powerful tool for molecular design.

II. FUNDAMENTALS OF HALOGEN BONDING

The basic principles that underlie the X-bonding concept come from quantum mechanical (QM) analyses of complexes of halogenated organic molecules with various types of Lewis

bases. The first question that comes to mind when we try to describe the concept of X-bonding is: How can halogens, considered to be electron rich in themselves, form short-range, stabilizing interactions with electron-rich Lewis bases? Fluorine, for example, is often used as a substitute for hydrogen-bond acceptors in carbohydrate chemistry,[14] which would appear to contradict the ability of halogens to serve as Lewis acids. To address this problem, we need to consider a more detailed description of the shape and charge distributions of halogens that are covalently bonded to other atoms (typically carbon in a C—X bond, where X is a polarizable halogen), as predicted by current QM models.

The original concept that charge transfer is the primary physicochemical basis for X-bonding has largely been replaced by electrostatic models based on the polarization of halogens that participate in covalent bonds. QM calculations applied at various levels, from Hartree–Fock method, to density functional theory (DFT), to Møller–Plesset perturbation (MP2) theory on simple halogenated organic compounds indicate that the distribution of electrostatic potential across the surface of the halogen is nonuniform. This anisotropic distribution of charge results in a crown of positive electrostatic potential directly opposite a covalent C—X bond, whereas the expected electronegative potential is manifested as a ring that encircles the halogen's girth perpendicular to the covalent bond.[15–19] Accompanying this anisotropic charge distribution is distortion to the halogen's shape, referred to as ''polar flattening,'' where the effective radius of the halogen is shorter in the direction of the covalent bond[20,21] by as much as 16% relative to the standard rvdW.[22] Together, the effects of flattening and charge depletion opposite the C—X bond provides a rationale for how halogens, such as bromine and iodine, can attract electron-rich oxygens and nitrogens to form ''bonds'' that are similar to classic H-bonds.

The σ-hole model

This leads to the next question, which is why are halogens polarized to such an extent that they form X-bonds? Perhaps the most accessible description, because of its simplicity, for how polarization affects the charge and shape properties of halogens is the σ-hole model (Figure 2.2) as formulated by Politzer, Murray, and Clark.[17,18]. In this model, the apparently unique nature of halogens has its roots in the fundamental properties of the covalent σ-bond between atoms. To understand this model, recall that Group VII atoms have five electrons residing in the p-atomic orbitals of the valence shell and that, according to molecular orbital theory, it is the single valence electron of the p_z orbital that participates in forming a covalent σ-bond to a carbon atom. Consequently, the depopulation of this orbital opposite the C—X σ-bond leaves a hole that partially exposes the positive nuclear charge. This σ-hole accounts for the electropositive crown and polar flattening associated with the polarization effects predicted from the QM calculations, whereas the four electrons remaining in the p_x, p_y orbitals account for the electronegative ring lying perpendicular to the σ-bond. In analogy to H-bonds, we can consider the σ-hole to be the donor to the electron-rich Lewis base acceptor in an X-bond.

One would expect that σ-bond formation to any atom would have this same type of polarization effect and, indeed, this has been predicted for other atoms, including those in the Group VI atoms of the periodic table. Thus, the X-bond is simply one example of a larger class of interactions that are now referred to as σ-hole bonding.[23] Halogens, however, are unique in that the electropositive crown is not masked by other covalently bonded groups or by lone pair electrons in nonbonding orbitals that extend in approximately the same orientation as the σ-hole. We should note, however, that this relatively straightforward electrostatic explanation of X-bonding may not tell the entire story, as there remains debate concerning the relative

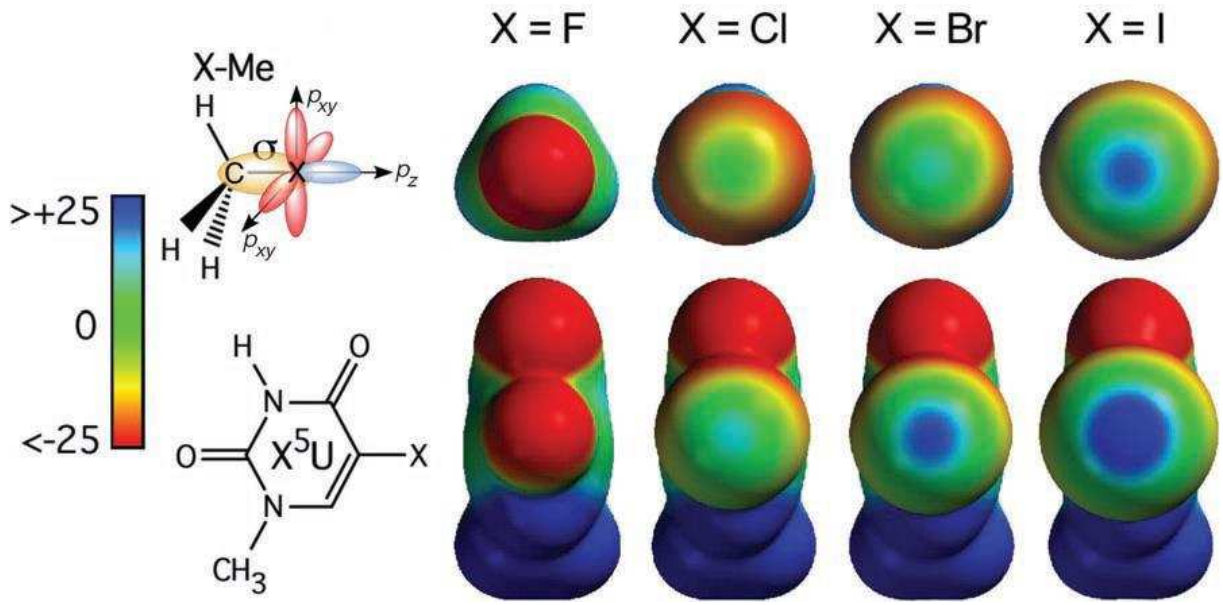

Figure 2.2. The σ -hole model and polarization of the electrostatic surface potential. The σ -hole resulting from redistribution of the valence electron in the p_z-atomic orbital (blue) to form the covalent C—X σ-bond (yellow) of a halomethane (X-Me) molecule results in depopulation of the p_z orbital, but maintaining the electrons of the p_x and p_y orbitals (red). The resulting polarization of the electrostatic potential of the halogen surfaces increases as the size of the halogen increases, from F to Cl to Br to I (viewed down the X—C bond). The halogen attached to a more electronegative molecule (e.g., a uracil base, X^5U) exaggerates the polarization effects. Electrostatic potential surfaces were calculated by DFT calculations at the 3-21G* level.[9]

contributions of dispersion and even the original charge transfer concept to the interaction. Attempts have been made to deconvolute the various components of the interaction using symmetry-adapted perturbation theory and natural bond orbital (NBO) analyses;[24–27] however, the conclusions from such studies are highly dependent on the model system the method of analysis.[24,28]

Geometry of X-bonds

The basic concept of the σ-hole makes the X-bond a highly directional interaction, as reflected in the angle of approach of the X-bond acceptor to the halogen relative to the direction of the σ-bond (Θ_1, Figure 2.1). Surveys of Θ_1 angles for small molecule structures in the Cambridge Database[29] as well as biomolecular structures in the PDB[9] indicate a strong preference for a near linear approach of the acceptor toward the electropositive crown of the σ-hole, with a significant drop-off as the acceptor approaches the crossing point between the positive and negative electrostatic potentials ($\Theta_1 \approx 140°$). The balance between the maximum positive electrostatic potential at $\Theta_1 = 180°$ with the increase in available surface area of the halogen atom as Θ_1 approaches 90° accounts for the preference for $\Theta_1 \approx 160°–165°$.

The geometry of the X-bond in terms of the angle of approach of the halogen toward the acceptor atom (Θ_2, Figure 2.1) shows that, for the most part, the σ-hole is attracted to the nonbonding electrons of the acceptor,[9] with $\Theta_2 \approx 120°$ and consistent with the geometries seen in small molecule structures.[29] The exceptions for biological X-bonds are when delocalized π-electrons are available, for example, from the side chain of an aromatic amino acid or the peptide bond of a protein backbone. Once again, such π-X-bonds were first identified from surveys of small molecule crystal structures in the Cambridge Database,[30] showing the halogen to be

directed perpendicular to the aromatic ring. Examples of π-X-bonds in biomolecules were first observed to the aromatic ring of Phe residues in the complexes of inhibitors to the protein kinases CDK2 and CK2.[31] This was further extended to a more comprehensive survey demonstrating the broad range of aromatic π-X-bonds in both proteins and small molecules structures, which, when coupled with MP2 calculations, could be attributed to upward of 2.5 kcal/mol toward the energy of interaction.[32] Once again, we see an obvious analogy between X- and H-bonds, in this case the π-X-bonds to π-H-bonds[33] to aromatic side chains of proteins.

The peptide bonds of most amino acids participate in H-bonds that help to define the secondary and tertiary structures of a folded protein. In these cases, the nonbonding orbitals of the carbonyl oxygen of the peptide bonds are occupied by H-bonds; thus, the only electronegative potential available for X-bonding comes from the π orbitals.[34] This type of interaction is the predominant type of X-bond seen in the crystal structures of protein–ligand complexes, and was shown to be oriented perpendicular to and energetically independent of the accompanying H-bond.[34] These observations lead to the hypothesis that biological X-bonds are orthogonal interactions to H-bonds when both share the carbonyl oxygen of a peptide bond as the common acceptor, suggesting that an X-bond can be introduced as an interaction for recognition without disrupting the H-bond stabilized structure of the protein target.

Tunability of X-bonds

The energy associated with any particular X-bond is dependent on several intrinsic properties of the interaction, including the polarizability of the donor halogen, the electron-withdrawing ability of the molecule that the halogen is covalently bound to, and the basicity of the acceptor atom. The most direct effect on not only the energy, but also the directionality of X-

bonds is the polarizability of the halogen, which follows the series F <Cl< Br < I, for common halogens. The electrons of fluorine are not so easily redirected to the σ-bond;[17,28] there is not a significant depopulation of the p_z-atomic orbital opposite the σ-bond and, consequently, F shows only a minimum σ-hole.[17,28] A study using NBO analysis on the molecule CF_3X (where X = F, Cl, Br, I) found that 71.4% of the σ-bond electron density was shifted toward the fluorine, whereas Cl, Br, and I tended to split the electrons equally with the carbon.[17] Thus, fluorine is generally considered to be a poor X-bond donor, except in cases where there is a very strong electron-withdrawing ability of the molecule that it is bonded to.[35]

For Cl, Br, and I, the electron-withdrawing ability of the molecule they are bound to helps tune the size of the σ-hole, which then affects the interaction energy (Figure 2.3A, B) as well as the size of the electropositive crown. A series of QM studies on various fluorinated benzene in complex with acetone demonstrated the degree of "tunability" of X-bonds.[36] We have repeated this type of calculation here for a bromobenzene interacting with *N*-methylacetamide (NMA), and their derivatives, to demonstrate the various effects on the energies of X-bonding interactions as we would expect to see them in most biological complexes (Figure 2.3C). In this case, pentafluorobromobenzene is seen to have over twice the stabilizing potential (−4 kcal/mol) of the nonfluorinated compound (−2 kcal/mol). A uracil base (Figure 2.2) was seen to approximate the electron-withdrawing ability of a tetrafluorobenzene, which would account for the very strong X-bond measured for a bromouracil to a phosphate oxygen acceptor in DNA.[37,38]

As a predominantly electrostatic interaction, we would expect the interaction energy of X-bonds to increase with the basicity of the acceptor group, as reflected in their partial charge. For any given Lewis base, however, substituent effects could greatly affect its basicity (Figure

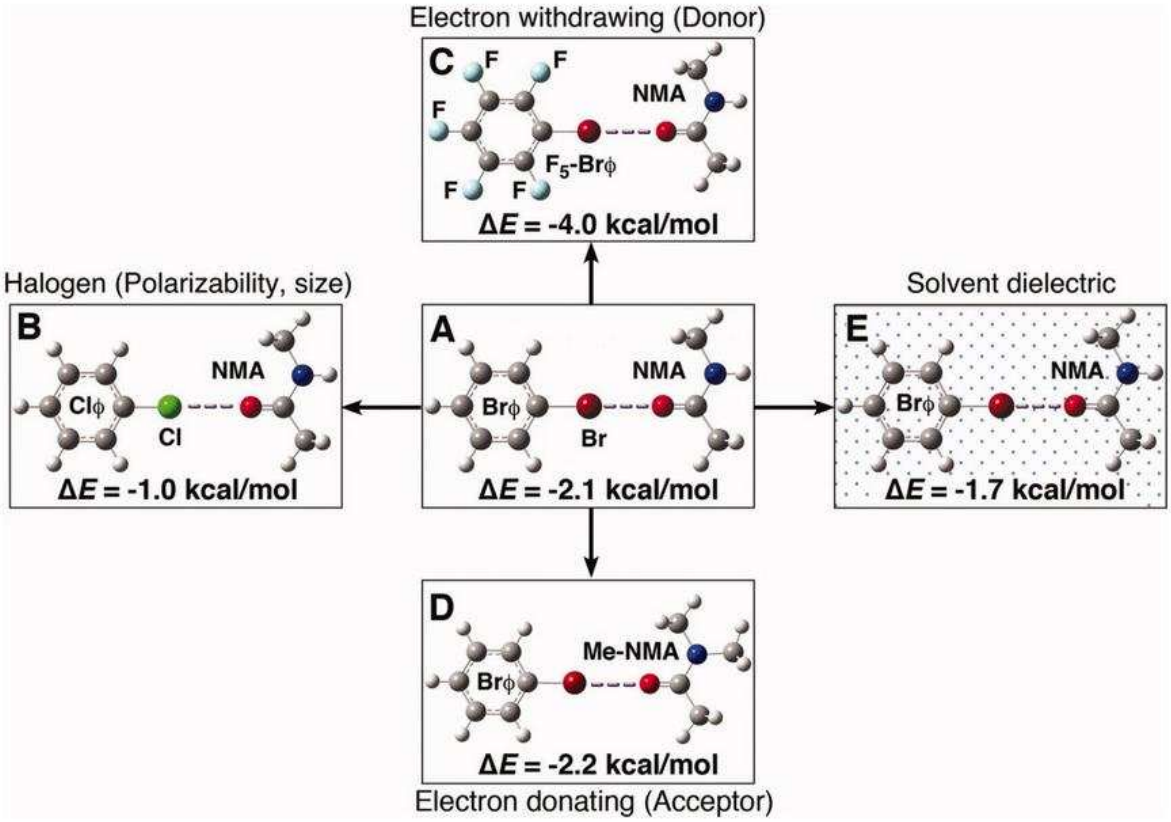

Figure 2.3. Tunability of halogen bonding energies. MP2 calculations at the 6-31G(d) level compares the effects on the energies ($\triangle E$) for bromobenzene (Brϕ) interacting with the carbonyl oxygen of *N*-methyl acetamide (NMA) in the gas phase (A) as the bromine is replaced by a less polarizable chlorine (B), with electron-withdrawing fluorine substituents added to the Brϕ donor (C), an electron donating methyl added to the NMA acceptor (D), or as it is transferred to solvent (cyclohexane, with a dielectric constant = 2.023, panel E).

2.3D), with electron donating groups predicted to increase the negative potential and withdrawing groups to reduce the potential of the acceptor. As discussed previously, π-electrons can serve as X-bonds acceptors and, although they are not as strong as the nonbonding electrons in terms of their basicity, are seen with aromatic amino acids or with the carbonyl oxygen of amides. Thus, the acidity of the donor and basicity of the acceptor combine to define the overall stabilizing potential of X-bonds. Finally, we would expect the polarizability of the solvent to affect the interaction energy, with an increase in the dielectric constant associated with a less favorable interaction (Figure 2.3E).

Relationship between hydrogen and halogen bonds

The interplay between X- and H-bonds can be very complicated. In addition to forming the σ-hole, polarization creates a negatively charged annulus perpendicular to the σ-bond. Halogens, therefore, serve not only as X-bond donors in the direction of the σ-hole, but also as acceptors to H- or X-bond donors. A survey of interactions with H-bond donors shows a much broader distribution across the Θ_1 angles than with X-bond acceptors,[19] which might be expected, as donors such as hydroxyl groups can interact in an H-bond or X-bond to the hydrogen, depending on their angle of approach. In addition, it has been shown that the X-bond-donating potential of the σ-hole can be extended to become an H-bond donor through a water bridge.[35] On the acceptor side, we have already seen that a carbonyl oxygen of a peptide bond can form an X-bond that is geometric and thermodynamically orthogonal to a pre-established H-bond. The relationship between X- and H-bonds, therefore, appears to be schizophrenic, being competing, complementary, or orthogonal, depending on the situation. Thus, the nature of any particular X-bond is best understood from a detailed analysis for that system.

III. WHERE ARE HALOGEN BONDS SEEN IN BIOLOGY?

With the increasing recognition of X-bonding as a significant contributor to specificity, the number of X-bonds observed in a variety of biomolecular systems has increased dramatically. For the most part, X-bonds continue to be recognized in crystal structures only in hindsight; however, there are increasing efforts to design the interaction into complexes to control specificity or the folding in biomolecules.

Survey of biological halogen bonds

The most comprehensive view of the variety of X-bonds seen in biology comes from surveying the PDB. Starting from the very first of such studies, which introduced X-bonds to the biological community,[9] detailed analysis of the crystal structures in the PDB have defined the geometries of and expanded the range of acceptors available for the interaction, and have delineated the complimentary and orthogonal relationships between H- and X-bonds.

The number of biological X-bonds identified in the PDB has increased significantly from 116 in 2004[9] to well over 600 in the current study (Figure 2.4), reflecting at some level the growth of structures in the PDB, along with a growing recognition of the interaction. In the current survey, we see acceptor interactions with halogens extending throughout the entire range of Θ_1 from 90° to 180°. As expected, the distributions peak at $\Theta_1 \approx 160°$ for the total of the X-bonds (the three bins from 140° to 150° appear to be unusually high, which we attribute to van der Waals contacts at the point of neutral electrostatic potential as well as contributions from potential H-bonding-type interactions). The X-bonds tend to be ~7% shorter than the $\sum R_{vdW}$ (Figure 2.4B). As the types of acceptors become expanded to include all possible H-bond

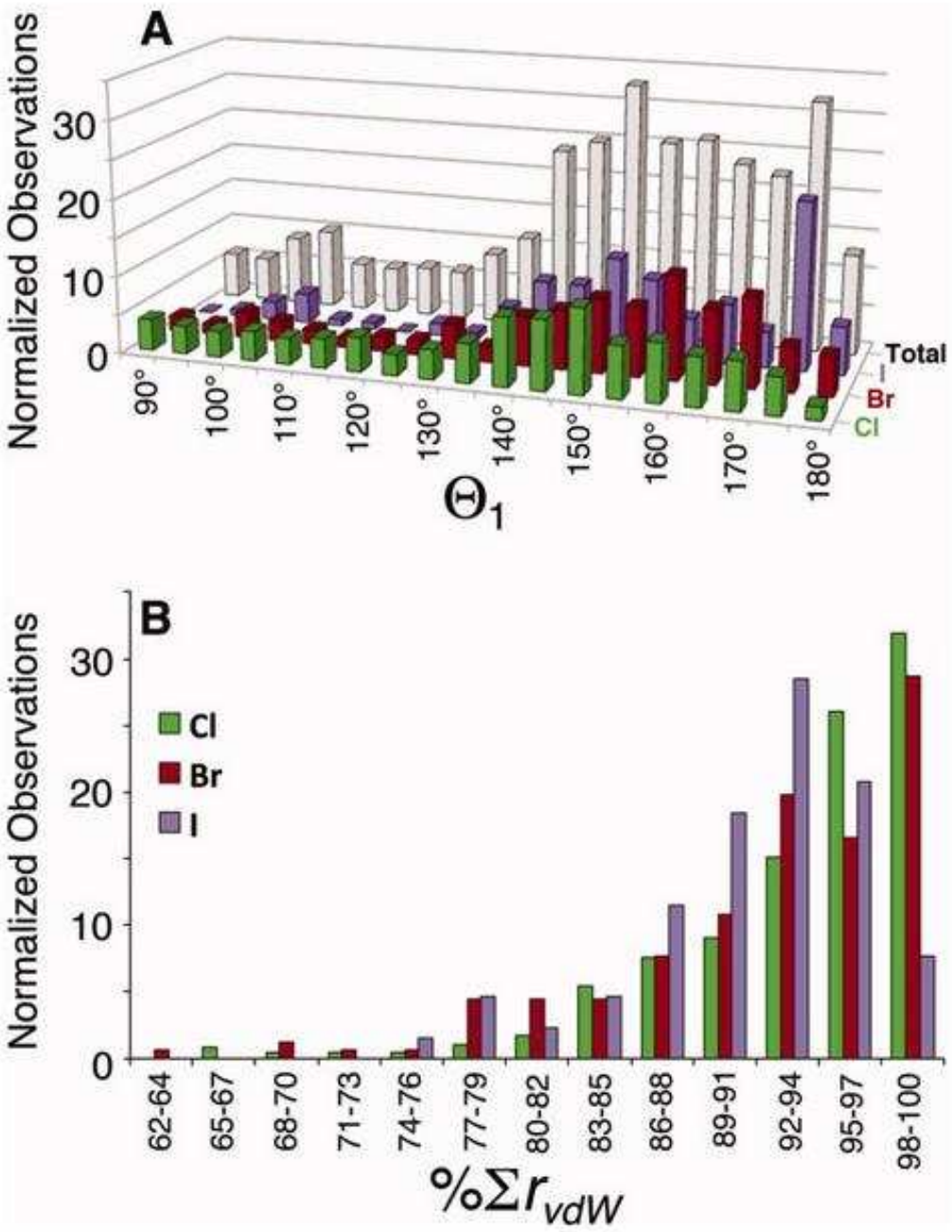

Figure 2.4. Survey of X-bonds in the PDB. Number of interactions at distances $\leq \Sigma R_{vdW}$ were tabulated for acceptor types that can only form X-bonds, including oxygens, nitrogens, and sulfurs and H-bond donors from Θ_1 = 140°–180° (up to the neutral point of the electrostatic potential). (A) Number of X-bonds to Cl (477 total), Br (157 total), and I (130 total), normalized for total number of observations for each type of halogen, and the total of these normalized observations. The X-bond distribution is centered at $\Theta_1 \approx 160°$ for all halogen types. (B) Distribution of distances between X-bond donors and acceptors as percentages of the $\sum R_{vdW}$ ($\%\sum R_{vdW}$).

acceptors, including aromatic side chains[31,32] and even anionic halides,[35] the list of biological X-bonds is expected to grow at an even faster rate.

Examples of biological halogen bonds

The variety of X-bonds depends on the variety of halogens seen in biology. There are very few examples of naturally halogenated proteins or nucleic acids, except as an oxidative response associated with, for example, asthma.[35] There is, however, an increasing number of halogenated proteins and nucleic acids used to help phase crystallographic data,[39] but these modifications are not entirely benign—it has been shown, for example, that X-bonds can facilitate formation of a number of multistranded DNA complexes,[40,41] including the four-stranded Holliday junction.[6]

X-bonds are seen predominantly in protein complexes with halogenated ligands. This is not surprising, given the prevalence of halogenated compounds found as secondary metabolites,[42] including a number that are antibiotics, and incorporated in screens to identify inhibitors against therapeutic targets.[35] At this point, we will consider in some detail the thyroid hormones as examples of naturally halogenated compounds and a set of halogenated inhibitors as anticancer drugs, to demonstrate how X-bonds can be useful in the design of therapeutics as treatments against human disease.

Thyroid hormones

The iodinated thyroid hormones represent a class of naturally occurring ligands where X-bonding plays a role in recognition.[26,43] Thyroxine and thyroid-like hormones are associated with a number of metabolic diseases such as obesity, hypercholesterolemia, diabetes, and

amyloidogenesis.[44,45] The role that X-bonds play in the recognition of thyroid hormones is evident in the short I···O interactions seen in the structure of tetraiodothyroxine with the transthyretin transport protein.[35] A brominated analog of the thyroid-like hormone (2-arylbenzoxazole) was found to bind at a 15-fold higher affinity to transthyretin than its nonhalogenated analog, and to inhibit the formation of transthyretin aggregates, which might lead to a treatment against transthyretin misfolding diseases.[45] In addition, the structures of the thyroid hormones with their receptors are seen to show interactions with geometries that are indicative of X-bonds (Figure 2.5), in both the brominated and iodinated forms.[44] Finally, it has been shown that iodination was a requirement in the recognition of thyroxine by RNA aptamers selected to bind to this hormone.[47]

The catabolism of thyroid hormones results in formation of iodotyrosine, which is then processed by iodotyrosine deiodinase to salvage the halogen. The enzyme's specificity for tyrosine analogs follows the order of polarizability of halogens (the series I>Br>Cl>F), suggesting the involvement of X-bonding in the recognition of the substrate. In a more recent study,[48] X-bonds are thought to slightly elongate the cleavable C—I bond, an important step in the enzymatic mechanism.[49] Thus, X-bonds appear to play crucial roles in the biology of thyroid hormones, from its recognition by receptors, to the salvage of iodine during its catabolism and as necessary for subsequent anabolism.

Inhibitors against cancer targets

Halogenated compounds are important inhibitors against proteins, including those that are involved in carcinogenesis. There have been extensive reviews on the role of X-bonds in the

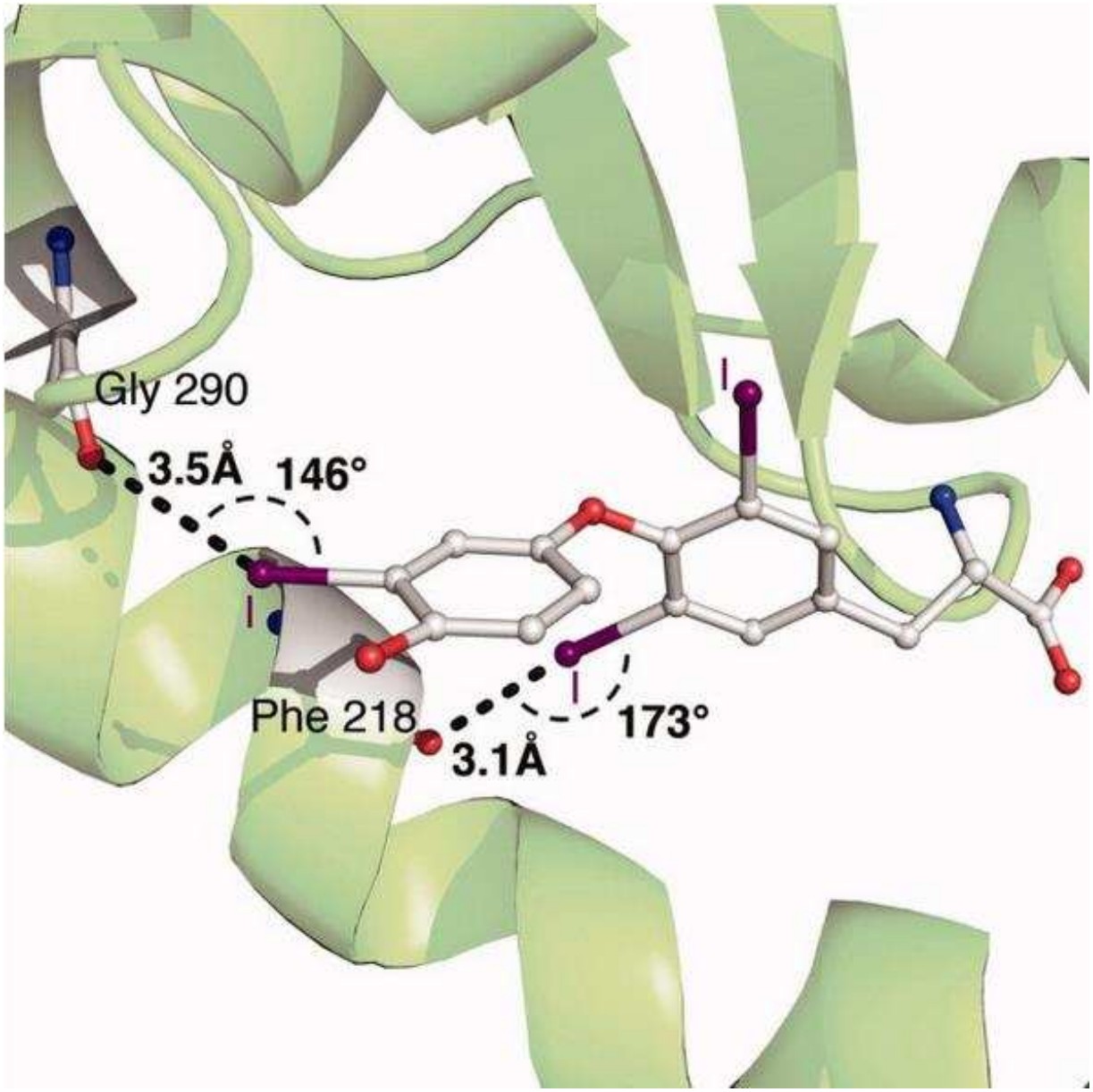

Figure 2.5. Recognition of 3,5,3′-triiodothyroxine (T3) by human thyroid hormone receptor. X-bonds (dotted lines) are shown from two iodines (purple) of T3 to the carbonyl oxygens (red) of the peptide bonds of the receptor, along with the distances and Θ_1 angles for each interaction (PDB-ID 2H79[46]).

recognition of various inhibitors against several classes of protein kinases.[31,50] Two recent examples include a new iodinated inhibitor designed to target the mitogen-activated protein kinase (MEK) and a chlorinated inhibitor to the CDC2-like kinase isoform 1 (CDK1), reinforcing the significance of X-bonds in conferring specificity of inhibitors against protein kinases (Table 2.1).

The structures of halogenated inhibitors in complex with epidermal growth factor receptor and maltripase[51] show that X-bonding can be generalized to other antitumor targets (Table 2.1). Finally, the five-order of magnitude lower K_i of a brominated compared with nonbrominated inhibitor against the tumor suppressor protein aminopeptidases-N (APN) was suggested to be associated with X-bonds rather than general hydrophobic effects,[55] indicating that this concept is becoming invoked even in the absence of specific structural evidence.

Each of the examples discussed so far have implicated X-bonds, again, in hindsight from the structural geometry of interactions or when comparing the efficacies of halogenated to nonhalogenated compounds against protein targets. Two recent studies show that the X-bonding concept can be incorporated at the design stage to increase the affinity of ligands as potential anticancer drugs. Wilcken *et al.*[56] generated a fragment library that was halogen-enriched with the intent of exploiting X-bonding to screen for high-affinity inhibitors against p53. The results of the studies showed that compounds containing iodine had significantly lower affinities compared with similar compounds containing the other halogens. The geometry of the I···O(pCO) interaction (where O(pCO) refers to the carbonyl oxygen of the peptide bond distance of $0.87\sum R_{vdW}$, Θ_1 angle = 172°) was evidence that an X-bond accounts for the halogen selectivity. Finally, Carpenter *et al.*[57]showed that a halogenated benzimidazole carboxamide inhibitor had a 1000-fold higher affinity against integrin $\alpha_4\beta_1$, a newly identified target to fight

Table 2.1. Role of X-Bonds in Recognition Specificity of Halogenated Inhibitors Against Anticancer Targets. The geometries of interaction of the halogen to the acceptor atom, including the carbonyl oxygen of the peptide backbone (pCO), in terms of the fractional distance relative to the $\sum R_{vdW}$ and Θ_1 angle (if a structure is available).

Protein target/inhibitor	**Geometry**		**Comments: types of X-bond (PDB-ID), affinity data**
	X···O Distance	Θ_1 **Angle**	
Mitogen-activated protein kinase (MEK)/G-894	$0.94\sum R_{vdW}$	176°	I···O (PDB: 3V04[52])
CDC2-like kinase isoform 1 (CDK1)/KH-CB19	$0.88\sum R_{vdW}$	171°	Cl···O (PDB: 2VAG[53])
Epidermal growth factor receptor (EGFR)/GEFITINB	$1.01\sum R_{vdW}$	163°	Cl···O (PDB: 2ITO[51])
Aminopeptidases N (APN)/1D	NA	NA	Ki = 60 pM for brominated; 1 μM for nonhalogenated inhibitor55
Matriptase/0NW	$0.98\sum R_{vdW}$	158°	Cl···O (PDB: 4E7R[54])

T- and B-cell lymphomas, than the nonhalogenated analog. We expect that as more is learned about the energy–structure relationship of X-bonds, the rational incorporation of the interaction at the initial stages of inhibitor and drug design will become more commonplace.

IV. STRUCTURE-ENERGY RELATIONSHIPS

We now have a good understanding for the geometry of X-bonds in various types of protein–ligand interactions, and see some initial successes with intentional design of the interaction into such complexes. To accelerate the incorporation of X-bonds at the initial design stage for engineering new or better inhibitors, however, we need to understand how these geometries define the energies of specific X-bonding interactions. The most accurate computational approach to defining the structure–energy of X-bonds is to perform high-level QM calculations on the structures of these complexes; however, these are cumbersome and, in the absence of ultrahigh-resolution crystal structures, are fraught with errors. What we really need is a computational approach that incorporates X-bonding into current molecular mechanics (MM) and their associated docking algorithms, and a set of experimental data to help validate both the QM and MM approaches.

Direct experimental measures of energies of biological X-bonds

There are very few direct ways to measure the energies of X-bonds with specific geometries in biological systems. With small molecules, the interactions between complexes of known geometries can be determined by measuring the melting thermodynamics of the crystalline complex.[58] A similar approach has been applied to directly correlate the energies of X-bonds relative to H-bonds with specific geometries in single crystals of a four-stranded DNA

junction.[37,38] In this system, however, the energies were initially estimated through a crystallographic competition assay, in which the stabilizing potential of a bromine X-bond is directly competed against that of an H-bond in 1:1 and 1:2 X- to H-bond ratios (Figure 2.6). In this assay, two specific geometries were observed, with a shorter Br···$O^{-1/2}$ interaction being ~5 kcal/mol and the longer interaction being ~2kcal/mol more stabilizing than the competing H-bond in this system.[37] A study applying differential scanning calorimetry to measure the melting thermodynamics of the interaction in this DNA system showed a similar energy in solution, thereby validating the results from the crystallographic assay.[38]

An X-bond, however, is not always stabilizing. An I···S X-bond to the side chain of a Met designed into a T4-lysozyme/ligand complex was found to have an interaction energy that was not significantly different from what is expected for a simple van der Waals attraction.[59] The contrast to the energies from the DNA studies may be associated with the much weaker basicity of the sulfur in a thiol ether in the protein as opposed to the formally anionic oxygen of the DNA backbone.

Indirect experimental measures of X-bonding energies in protein-ligand complexes

The large number of single crystal structures along with affinity measurements should provide a database of indirect measures of X-bonding energies of ligands in various protein environments. We consider these to be indirect measurements of X-bonding energies, because there is no measure of the energies of the components separately, particularly with the liganded protein in identical conformation as the unliganded form; however, they serve as reasonable estimates.

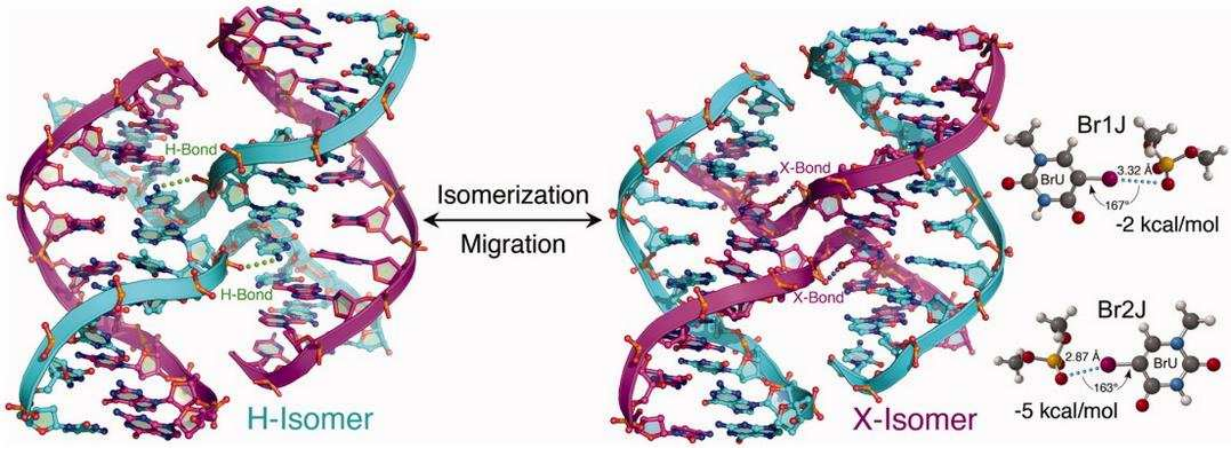

Figure 2.6. Four-stranded DNA junction as a competitive assay for H-bonding versus X-bonding energies. A four-stranded junction composed of unique DNA strands can isomerize to place either cytosine bases to form stabilizing H-bond (cyan strands) or 5-bromouracil (BrU) bases to form stabilizing X-bonds (magenta strands) to the sharp U-turn of the junction cross-over, adopting the H-isomer and X-isomer forms, respectively. In the X-isomer, the two possible $BrU\cdots OPO_3^{-1}$ interactions have energies that are 2–5 kcal/mol more stabilizing than the competing H-bonds.

There are some general trends seen in an analysis of ligand–protein structures and measures of their affinities as reflected in their K_i or IC50 values (Figure 2.7), particularly by comparing the affinities of halogenated and the unhalogenated inhibitors toward identical proteins or similar protein domains. The protein cathepsin L is a eukaryotic lysosomal endopeptidase that is associated with antigen processing, tumor invasion and metastasis, bone resorption, and turnover of protein involved in growth regulation. A set of structural studies show that X-bonding was an important contributor to the binding of a series of substituted nitrile inhibitors to a deep pocket in the active site of this enzyme (Figure 2.7A,B), with a methyl substituent[60] showing a 20-fold increase in the IC50 (equivalent to 1.8 kcal/mol reduction in affinity) relative to the iodinated analog.[61] Similarly, a comparison of casein kinase II (CK2) Ser/Thr kinase structures (Figure 2.7C,D) shows the imino nitrogens, an indazole inhibitor points toward the loop of the active site.[62] A similar tetrabromobenzimidazole inhibitor is rotated to point its halogens to form two X-bonds to this same loop,[63] which could account for the over 600-fold difference in the IC50 versus K_i for the indazole and tetrabromobenzimidazole, respectively.

Computational approaches to structure-energy relationships of X-bonds

Experimental assays allow us to now develop and validate computational methods at various levels to model specific X-bonding geometries and predict their associated energies. For biologist and medicinal chemists, the goal is to develop algorithms that can be readily applied to the design of new inhibitors against therapeutically important protein targets or new supramolecular complexes from biological systems.

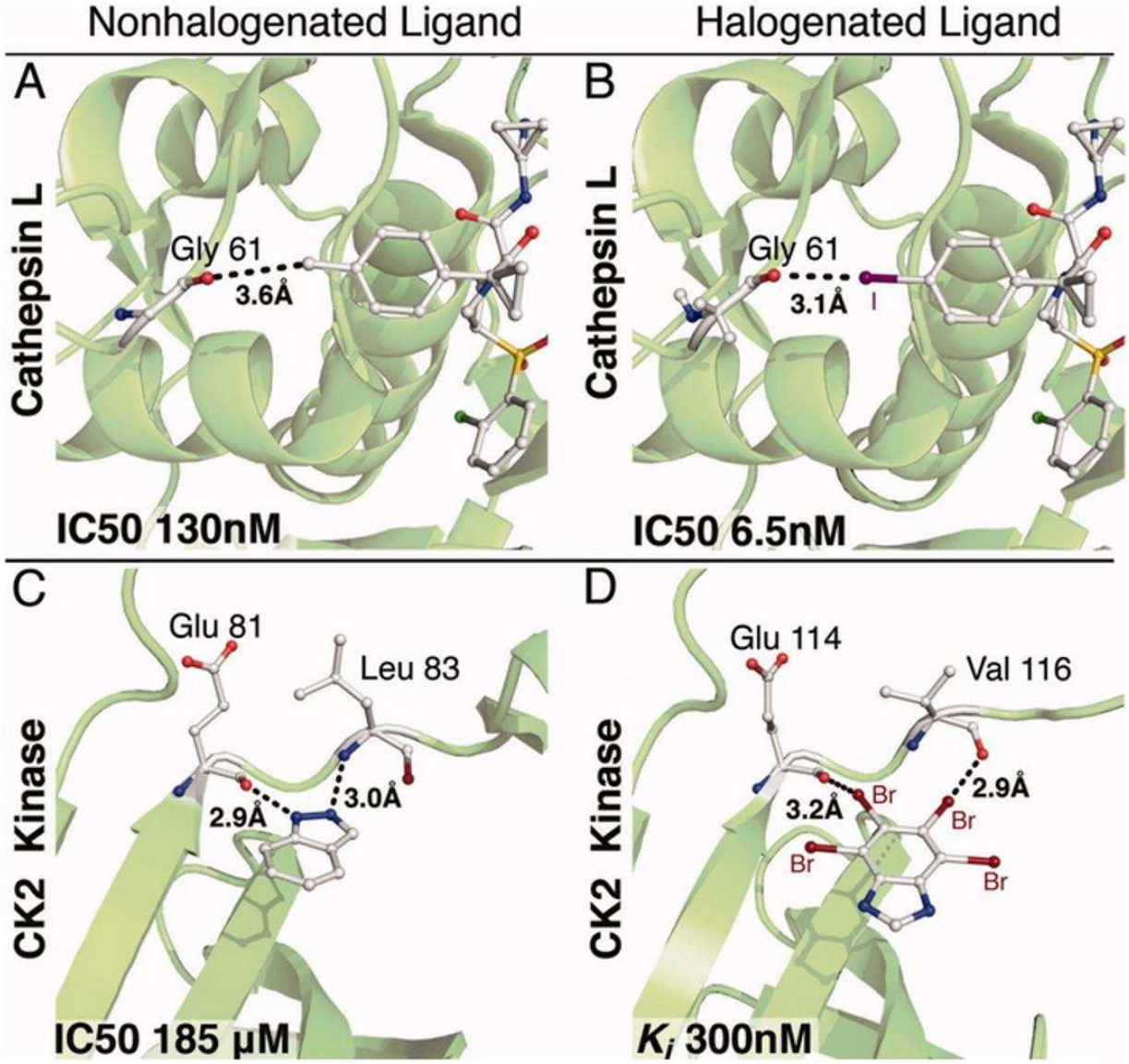

Figure 2.7. Comparison of halogenated and nonhalogenated inhibitors to protein targets. Cathepsin L in complex with the nonhalogenated ligand (2S,4R)-4-(2-chlorophenyl)sulfonyl-*N*-[1-(iminomethyl)cyclopropyl]-1-[1-(4-methylphenyl)cyclopropyl]carbonyl-pyrrolidine-2-carboxamide (A, PDB-ID 2XU5[60]) and its iodinated analog (B, PDB-ID 2YJ8[61]). Complex of cyclin kinase CK2 with 1H-indazole (C, PDB-ID 2VTA[62]) is compared with casein kinase CK2 bound to tetrabromobenzimidazole (D, 2OXY[63]). The polypeptides backbone are traced as a green ribbon, with amino acids and ligands involved in X-bonding interactions (black dashes) shown as ball stick models with carbons (gray), oxygen (red), nitrogen (blue), iodine (purple), and bromine (brown). The affinities of each ligand are labeled in terms of their IC50 or K_i values.

The most accurate means to modeling X-bonds is through QM calculations. For the theoretical chemist, high-level QM calculations have been and are currently used to better understand the foundational principles of X-bonding. The first application of QM showing X-bonding in a biological system (although not recognized as such by the authors) was a DFT calculation[64] based on the 0.66 Å structure of aldose reductase in complex with the brominated IDD594 inhibitor[7]—this ultrahigh-resolution structure provided the accuracy in the atomic coordinates required to minimize errors in the calculation. The crystal structure showed a short 3.0 Å Br···O interaction to the hydroxyl oxygen of the Thr113 side chain, whereas the DFT calculation on the entire complex indicated that this was primarily due to an electrostatic-type interaction. Although the energy of the Br···O interaction was not explicitly estimated in this study, this could account for the 1000-fold specificity of the brominated inhibitor for the aldose reductase over the similar aldehyde reductase.

More direct QM calculations have been performed, however, on models of the $Br\cdots O^{-1/2}$ interaction in the competition assay using DNA junctions.[35] In this case, higher level MP2 calculations were applied to the complex of bromouracil with hypophosphite to model the X-bonding interaction to the phosphate backbone of the DNA.[65] The MP2 calculations yielded energies that accurately mirrored those of the specific geometries seen in the experimental system (Table 2.2), indicating that the QM models were appropriate for this DNA system.

One approach to incorporating the accuracy of QM calculations with MM approaches that are more generally applicable to macromolecules is a hybrid QM/MM method, such as that implemented in the program ONIOM.[66] In this approach, the molecular system is segregated into those parts for which MM calculations can be accurately applied (most of the macromolecule plus the solvent) and those parts which requires a QM calculation (groups immediately around

Table 2.2. Experimental and Theoretical X-Bonding Energies of the Br1J and Br2J Conformations (Figure. 2.6) in the DNA Junction Competitive Assay. Experimental X-bond versus H-bond energies as determined crystallographically[37] ($\Delta E_{X\text{-}H(\mathrm{Xtal})}$) or by differential scanning calorimetry[38] ($\Delta E_{X\text{-}H(\mathrm{DSC})}$) are compared with calculated X-bonding energies[22] from quantum mechanical MP2 (E_{MP2}) and by the force field for biological X-bonds ($E_{ff\mathrm{BXB}}$)—all energies are in kcal/mol. The competing H-bond energy is estimated to be ∼ −1 kcal/mol.

Conformation	$\Delta E_{X\text{-}H(\mathrm{Xtal})}$	$\Delta E_{X\text{-}H(\mathrm{DSC})}$	E_{MP2}	$E_{ff\mathrm{BXB}}$
Br1J	−2.0 ± 0.5	ND	−3.1	−3.2
Br2J	−4.8 ± 0.5	−3.9 ± 1.3	−5.8	−5.5

the X-bond donor and acceptor). The application of this QM/MM approach to interactions between halogenated ligands and carbonyl oxygens found in the PDB was able to accurately reproduce the interaction geometry compared with data from crystallographic analysis.[67] The interaction energies, however, were variable depending on the approach implemented for the QM component of the calculation, but they were qualitatively in agreement with general halogen bonding trends.

Perhaps the approach that is most readily accessible to biological and medicinal chemists is the pure MM calculation used to determine both static and dynamic properties of biomolecular systems. There have recently been some significant efforts toward implementing X-bonding into programs such as AMBER[65,68] and OPLS-AA,[69] including the application of a positive extra point (PEP) approach and attempts to derive a set of potential energy force field functions that are specific for X-bonds.

The most straightforward approach to modeling X-bonds in AMBER, which utilizes all of the current functions of the force field, is the PEP. In this method, all of the standard MM parameters are assigned to the center of the halogen being modeled, whereas a pseudoatom having no mass or van der Waals energy, but a defined positive charge, is placed at some distance from the halogen center and diametrically opposed to the σ-bond. The first application by Ibrahim,[70,71] which placed the extra point charge at the halogen surface, showed that the energies calculated by this PEP method correlated well with the affinity of various halogenated benzimidazole inhibitors against CK2 kinase. We note, however, that the distances between the halogen donor and acceptor atoms tend to be on average ∼0.3 Å longer than seen in the crystal structures and the absolute $\Delta G°$ for binding are considerably more negative than expected from their dissociation constants.

A recent refinement of the PEP approach from Hobza's group[72] is to place the pseudoatom closer to the halogen center (1.5 Å from the bromine center) and with a compensatory charge of +0.2*e*. This allows a closer approach of the X-bond acceptor to donor, yielding donor–acceptor distances that better mirrored those seen in the crystal structures and energies that are well matched with gas-phase QM calculations.

Our own group took the approach of deriving a halogen-specific set of empirical potential energy functions, based on QM analyses of the DNA junction system, that constitute a force field for biological X-bonds (an *ff*BXB).[22] The *ff*BXB includes a directional model for the aspherical shape of a bromine (consistent with the polar flattening seen in high-resolution crystal structures of complexes of small halogenated compounds) and for the distribution of charges across halogen surface (from positive to negative going from Θ_1 = 180°–90°). This method very accurately reproduces the QM energies for the $BrU\cdots H_2PO_2^-$ model system as well as the experiment energies from the DNA junction competition assay (Table 2.2). When extended to an uncharged oxygen-type acceptor, the *ff*BXB accurately reproduces the QM-calculated energies and geometries for the X-bond interaction between acetone and bromobenzene, and its various fluorinated derivatives. As the *ff*BXB method models the surface potential of the halogen, its energies can be coupled with calculations of the surface area available at each θ_1 angle to predict a probability for interaction distributed across the angle of approach of the acceptor to the donor halogen relative to the σ-hole. When applied to the acetone–bromobenzene system, these *ff*BXB-calculated probabilities are seen here to predict accurately the distribution of X-bonds to the pCO acceptors found in the PDB. It also shows how H-bonds can interact with the negative annulus that is approximately perpendicular to these X-bonds, thereby, providing a model for the amphimorphic properties of halogens in terms of their electrostatic interactions.

V. CONCLUSIONS AND PERSPECTIVES

Although short stabilizing interactions involving halogenated complexes have been known for several decades, we have only recently established a good fundamental understanding for why they occur through accurate QM modeling of how electrons distribute between the atomic and molecular orbitals of the covalent bond. What we now call a halogen bond or X-bond has been increasingly recognized as not only being important in hindsight, but also as a potentially powerful tool to engineer specificity in molecular complexes with foresight. Chemists have taken advantage of this concept in designing new halogenated materials with unique properties.[35] Of course, not all halogenated compounds are considered to be beneficial, as evident from the heated debate concerning the potential health risks of polybrominated flame retardants such as decabromodiphenyl ether.[73,74] The perspective for the biological and medicinal chemists, however, is that we can exploit the X-bond to design new or better inhibitors against therapeutic targets, or biologically based materials with definable structural and biochemical properties. The biological applications, however, lag behind those of the small molecule chemist, first, because we were not as quick to recognize the existence of such an interaction and second, because we have not had the computational tools to accurately model the interaction in the large and complex biological molecular systems. However, there has been significant progress in both fronts.

A quick survey of publication and citation databases shows a dramatic rise in the number of halogen bond publications found after 2004 (Figure 2.8). This reflects the recognition and acceptance of X-bonds as a distinct molecular interaction that is relevant in biological systems, even if halogens are not commonly found in many naturally occurring proteins or nucleic acids, but are in the ligands that bind to them. The computational tools required to identify and model

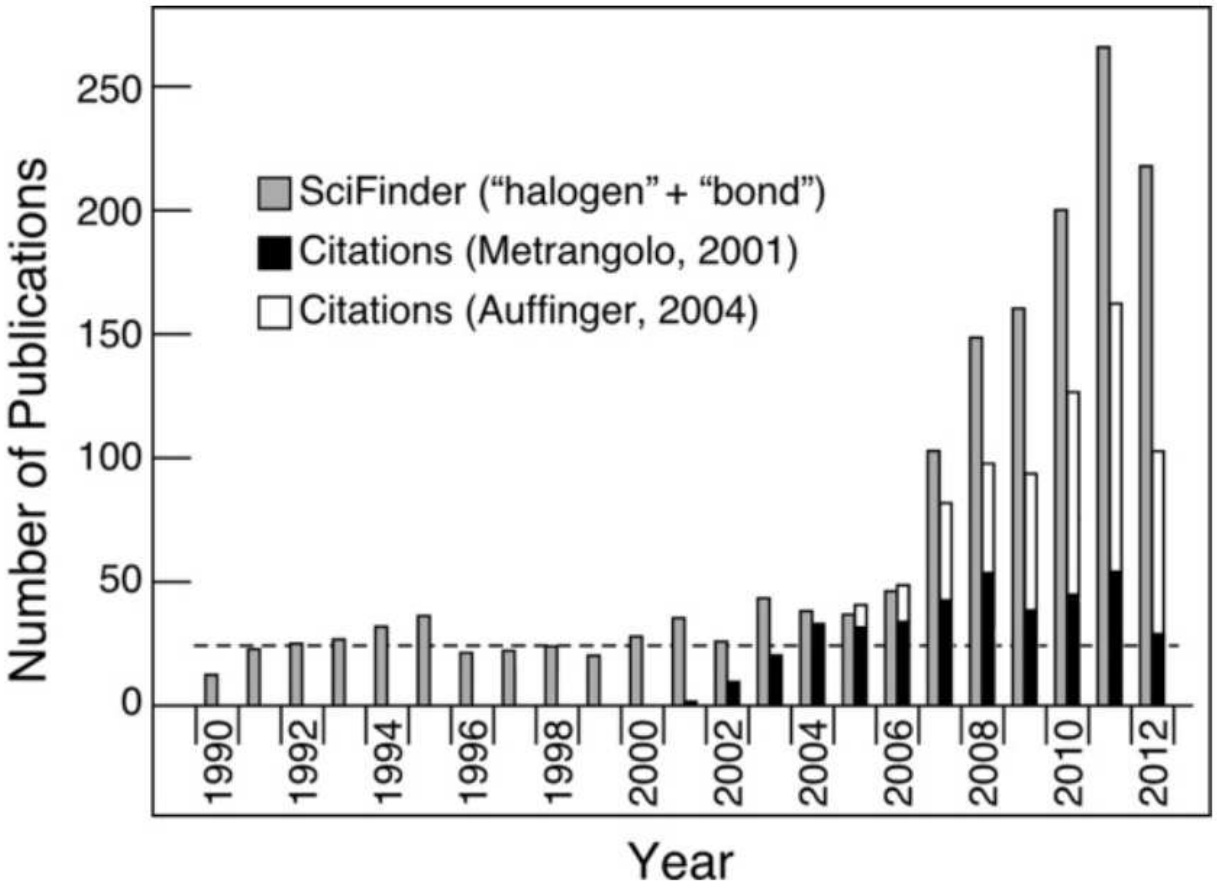

Figure 2.8. Number of X-bond publications from 1990 to 2012. The number of publications with the words "halogen" and "bond" in the title as found in SciFinder (gray bars) is compared with the number of citations to the publication of Metrangolo *et al.*[1111] (as a measure of interest in the material chemistry literature, black bars) and of citations to the publication of Auffinger *et al.*[9] (as a measure of interest in the biological literature, white bars). The horizontal dashed line indicates the average number of publications per year in which "halogen" and "bond" appear in SciFinder, but are not related to X-bonds. The biological literature has accounted for at least half the publications on X-bonds since 2007.

X-bonds are now being developed and refined. A tool to identify X-bonds in crystal structures of biomolecular systems, based on their geometries, has been incorporated the program HBAT,[75] which will allow structural biologist to better recognize the interaction if it exists.

Both the PEP and *ff*BXB approaches are being incorporated into molecular modeling algorithms to allow us to calculate the energies of the biomolecular complexes. When fully implemented, these computational tools will allow us to study and better understand the contribution of entropy, both in terms of the conformational entropy within the molecular system and the solvent entropy associated with the hydrophobic effect that together help to define the overall free energy of X-bonds. It is only then that there is the prospect to fully exploit the X-bonding interaction for biomolecular design and engineering using the Group VII atoms.

REFERENCES

1. Guthrie, F. XXVIII.?On the iodide of iodammonium. *J. Chem. Soc.* **16,** 239 (1863).
2. Hassel, O. Structural aspects of interatomic charge-transfer bonding. *Science* **170,** 497–502 (1970).
3. Lommerse, J. P. M., Stone, A. J., Taylor, R. & Allen, F. H. The Nature and Geometry of Intermolecular Interactions between Halogens and Oxygen or Nitrogen. *J. Am. Chem. Soc.* **118,** 3108–3116 (1996).
4. Legon, A. Prereactive Complexes of Dihalogens XY with Lewis Bases B in the Gas Phase: A Systematic Case for the Halogen Analogue B small middle dot small middle dot small middle dotXY of the Hydrogen Bond B small middle dot small middle dot small middle dotHX. *Angew. Chem. Int. Ed. Engl.* **38,** 2686–2714 (1999).
5. Metrangolo, P., Neukirch, H., Pilati, T. & Resnati, G. Halogen bonding based recognition processes: A world parallel to hydrogen bonding. *Acc. Chem. Res.* **38,** 386–395 (2005).
6. Hays, F. A., Vargason, J. M. & Ho, P. S. Effect of sequence on the conformation of DNA holliday junctions. *Biochemistry* **42,** 9586–97 (2003).
7. Howard, E. I. *et al.* Ultrahigh resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 A. *Proteins* **55,** 792–804 (2004).
8. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–42 (2000).
9. Auffinger, P., Hays, F. A., Westhof, E. & Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 16789–94 (2004).
10. Lam, P., Clark, C., Smallwood, A. & Alexander, R. Structure-based drug design utilizing halogen bonding: factor Xa inhibitors. in *Proceedings at the 238th ACS National Meeting*

59 (2009).

11. Metrangolo, P. & Resnati, G. Halogen bonding: a paradigm in supramolecular chemistry. *Chemistry* **7,** 2511–9 (2001).

12. Politzer, P., Murray, J. S. & Concha, M. C. Halogen bonding and the design of new materials: organic bromides, chlorides and perhaps even fluorides as donors. *J. Mol. Model.* **13,** 643–50 (2007).

13. Fourmigué, M. Halogen bonding: Recent advances. *Curr. Opin. Solid State Mater. Sci.* **13,** 36–45 (2009).

14. Scott, M. E. & Viola, R. E. The use of fluoro- and deoxy-substrate analogs to examine binding specificity and catalysis in the enzymes of the sorbitol pathway. *Carbohydr. Res.* **313,** 247–53 (1998).

15. Brinck, T., Murray, J. S. & Politzer, P. Surface electrostatic potentials of halogenated methanes as indicators of directional intermolecular interactions. *Int. J. Quantum Chem.* **44,** 57–64 (1992).

16. Murray, J. S., Lane, P., Clark, T. & Politzer, P. σ-hole bonding: molecules containing group VI atoms. *J. Mol. Model.* **13,** 1033–1038 (2007).

17. Clark, T., Hennemann, M., Murray, J. S. & Politzer, P. Halogen bonding: the σ-hole. *J. Mol. Model.* **13,** 291–296 (2007).

18. Politzer, P., Murray, J. S. & Lane, P. σ-Hole bonding and hydrogen bonding: Competitive interactions. *Int. J. Quantum Chem.* **107,** 3046–3052 (2007).

19. Lu, Y. *et al.* C−X···H Contacts in Biomolecular Systems: How They Contribute to Protein−Ligand Binding Affinity. *J. Phys. Chem. B* **113,** 12615–12621 (2009).

20. Nyburg, S. C. `Polar flattening': non-spherical effective shapes of atoms in crystals. *Acta*

Crystallogr. Sect. A **35,** 641–645 (1979).

21. Peebles, S. A., Fowler, P. W. & Legon, A. C. Anisotropic repulsion in complexes B.Cl2 and B.HCl: The shape of the chlorine atom-in-a-molecule. *Chem. Phys. Lett.* **240,** 130–134 (1995).

22. Carter, M., Rappé, A. K. & Ho, P. S. Scalable Anisotropic Shape and Electrostatic Models for Biological Bromine Halogen Bonds. *J. Chem. Theory Comput.* **8,** 2461–2473 (2012).

23. Riley, K. E., Murray, J. S., Politzer, P., Concha, M. C. & Hobza, P. Br···O Complexes as Probes of Factors Affecting Halogen Bonding: Interactions of Bromobenzenes and Bromopyrimidines with Acetone. *J. Chem. Theory Comput.* **5,** 155–63 (2009).

24. Riley, K. E. & Hobza, P. Investigations into the Nature of Halogen Bonding Including Symmetry Adapted Perturbation Theory Analyses. *J. Chem. Theory Comput.* **4,** 232–242 (2008).

25. Wu, J., Zhang, J., Wang, Z. & Cao, W. Theoretical study on intermolecular interactions in BrF/HnX adducts. *Chem. Phys.* **338,** 69–74 (2007).

26. Riley, K. E. & Hobza, P. Strength and Character of Halogen Bonds in Protein–Ligand Complexes. *Cryst. Growth Des.* **11,** 4272–4278 (2011).

27. Lu, Y.-X., Zou, J.-W., Wang, Y.-H., Jiang, Y.-J. & Yu, Q.-S. Ab initio investigation of the complexes between bromobenzene and several electron donors: some insights into the magnitude and nature of halogen bonding interactions. *J. Phys. Chem. A* **111,** 10781–8 (2007).

28. Murray, J. S., Riley, K. E., Politzer, P. & Clark, T. Directional Weak Intermolecular Interactions: σ-Hole Bonding. *Aust. J. Chem.* **63,** 1598 (2010).

29. Ouvrard, C., Le Questel, J.-Y., Berthelot, M. & Laurence, C. Halogen-bond geometry: a

crystallographic database investigation of dihalogen complexes. *Acta Crystallogr. Sect. B Struct. Sci.* **59,** 512–526 (2003).

30. Prasanna, M. . & Guru Row, T. . C–halogen···π interactions and their influence on molecular conformation and crystal packing: a database study. *Cryst. Eng.* **3,** 135–154 (2000).

31. Andrea, V. & P., H. The Role of Halogen Bonding in Inhibitor Recognition and Binding by Protein Kinases. *Curr. Top. Med. Chem.* **7,** 1336–1348 (2007).

32. Matter, H. *et al.* Evidence for C-Cl/C-Br...pi interactions as an important contribution to protein-ligand binding affinity. *Angew. Chem. Int. Ed. Engl.* **48,** 2911–6 (2009).

33. Steiner, T. & Koellner, G. Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.* **305,** 535–57 (2001).

34. Voth, A. R., Khuu, P., Oishi, K. & Ho, P. S. Halogen bonds as orthogonal molecular interactions to hydrogen bonds. *Nat. Chem.* **1,** 74–79 (2009).

35. Egner, U., Krätzschmar, J., Kreft, B., Pohlenz, H.-D. & Schneider, M. The target discovery process. *Chembiochem* **6,** 468–79 (2005).

36. Riley, K. E. *et al.* 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,** 3309–3318 (2011).

37. Voth, A. R., Hays, F. a & Ho, P. S. Directing macromolecular conformation through halogen bonds. *Proc. Natl. Acad. Sci.* **104,** 6188–6193 (2007).

38. Carter, M. & Ho, P. S. Assaying the Energies of Biological Halogen Bonds. *Cryst. Growth Des.* **11,** 5087–5095 (2011).

39. Vallejos, M. J., Auffinger, P. & Ho, P. S. in 821–826 (2012).

doi:10.1107/97809553602060000895

40. Sunami, T. *et al.* 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,** 422–31 (2004).

41. Sunami, T. *et al.* Structure of d(GCGAAAGC) (hexagonal form): a base-intercalated duplex as a stable structure. *Acta Crystallogr. D. Biol. Crystallogr.* **60,** 90–6 (2004).

42. Gribble, G. W. The diversity of naturally produced organohalogens. *Chemosphere* **52,** 289–97 (2003).

43. Eneqvist, T. *et al.* High Resolution Crystal Structures of Piscine Transthyretin Reveal Different Binding Modes for Triiodothyronine and Thyroxine. *J. Biol. Chem.* **279,** 26411–26416 (2004).

44. Valadares, N. F., Salum, L. B., Polikarpov, I., Andricopulo, A. D. & Garratt, R. C. Role of halogen bonds in thyroid hormone receptor selectivity: pharmacophore-based 3D-QSSR studies. *J. Chem. Inf. Model.* **49,** 2606–16 (2009).

45. Johnson, S. M., Connelly, S., Wilson, I. a. & Kelly, J. W. Biochemical and Structural Evaluation of Highly Selective 2-Arylbenzoxazole-Based Transthyretin Amyloidogenesis Inhibitors. *J. Med. Chem.* **51,** 260–270 (2008).

46. Nascimento, A. S. *et al.* Structural rearrangements in the thyroid hormone receptor hinge domain and their putative role in the receptor function. *J. Mol. Biol.* **360,** 586–98 (2006).

47. Lévesque, D., Beaudoin, J.-D., Roy, S. & Perreault, J.-P. In vitro selection and characterization of RNA aptamers binding thyroxine hormone. *Biochem. J.* **403,** 129–38 (2007).

48. McTamney, P. M. & Rokita, S. E. A mammalian reductive deiodinase has broad power to

dehalogenate chlorinated and brominated substrates. *J. Am. Chem. Soc.* **131,** 14212–3 (2009).

49. Manna, D. & Mugesh, G. Regioselective deiodination of thyroxine by iodothyronine deiodinase mimics: an unusual mechanistic pathway involving cooperative chalcogen and halogen bonding. *J. Am. Chem. Soc.* **134,** 4269–79 (2012).

50. Liao, J. J.-L. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J. Med. Chem.* **50,** 409–24 (2007).

51. Yun, C.-H. *et al.* Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **11,** 217–27 (2007).

52. Heald, R. A. *et al.* Discovery of novel allosteric mitogen-activated protein kinase kinase (MEK) 1,2 inhibitors possessing bidentate Ser212 interactions. *J. Med. Chem.* **55,** 4594–604 (2012).

53. Fedorov, O. *et al.* Specific CLK Inhibitors from a Novel Chemotype for Regulation of Alternative Splicing. *Chem. Biol.* **18,** 67–76 (2011).

54. Hammami, M. *et al.* New 3-amidinophenylalanine-derived inhibitors of matriptase. *Medchemcomm* **3,** 807 (2012).

55. Maiereanu, C. *et al.* A novel amino-benzosuberone derivative is a picomolar inhibitor of mammalian aminopeptidase N/CD13. *Bioorg. Med. Chem.* **19,** 5716–33 (2011).

56. Zimmermann, M. O. *et al.* Halogen-enriched fragment libraries as chemical probes for harnessing halogen bonding in fragment-based lead discovery. *Future Med. Chem.* **6,** 617–39 (2014).

57. Carpenter, R. D. *et al.* Halogenated benzimidazole carboxamides target integrin

alpha4beta1 on T-cell and B-cell lymphomas. *Cancer Res.* **70,** 5448–56 (2010).

58. Corradi, E., Meille, S., Messina, M., Metrangolo, P. & Resnati, G. Halogen Bonding versus Hydrogen Bonding in Driving Self-Assembly Processes Perfluorocarbon-hydrocarbon self-assembly, part IX. This work was supported by MURST (Cofinanziamento '99) and EU (COST-D12-0012). We thank Dr. A. Lunghi and Dr. P. Cardillo (Stazi. *Angew. Chem. Int. Ed. Engl.* **39,** 1782–1786 (2000).

59. Liu, L., Baase, W. A. & Matthews, B. W. 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,** 595–605 (2009).

60. Hardegger, L. a *et al.* Systematic Investigation of Halogen Bonding in Protein-Ligand Interactions. *Angew. Chemie Int. Ed.* **50,** 314–318 (2011).

61. Hardegger, L. A. *et al.* Halogen bonding at the active sites of human cathepsin L and MEK1 kinase: efficient interactions in different environments. *ChemMedChem* **6,** 2048–54 (2011).

62. Wyatt, P. G. *et al.* Identification of N-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *J. Med. Chem.* **51,** 4986–99 (2008).

63. Battistutta, R. *et al.* The ATP-binding site of protein kinase CK2 holds a positive electrostatic area and conserved water molecules. *Chembiochem* **8,** 1804–9 (2007).

64. Muzet, N., Guillot, B., Jelsch, C., Howard, E. & Lecomte, C. Electrostatic complementarity in an aldose reductase complex from ultra-high-resolution crystallography and first-principles calculations. *Proc. Natl. Acad. Sci. U. S. A.* **100,**

8742–7 (2003).

65. Kollman, P. A. *et al.* Modeling complex molecular interactions involving proteins and DNA. *Ann. N. Y. Acad. Sci.* **482,** 234–44 (1986).

66. Vreven, T., Morokuma, K., Farkas, O., Schlegel, H. B. & Frisch, M. J. Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *J. Comput. Chem.* **24,** 760–9 (2003).

67. Lu, Y. *et al.* Halogen bonding--a novel interaction for rational drug design? *J. Med. Chem.* **52,** 2854–62 (2009).

68. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26,** 1668–88 (2005).

69. Jorgensen, W. L. & Schyman, P. Treatment of Halogen Bonding in the OPLS-AA Force Field; Application to Potent Anti-HIV Agents. *J. Chem. Theory Comput.* **8,** 3895–3801 (2012).

70. Ibrahim, M. A. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* **32,** 2564–74 (2011).

71. Ibrahim, M. A. A. 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,** 3659–69 (2012).

72. Kolář, M. & Hobza, P. On Extension of the Current Biomolecular Empirical Force Field for the Description of Halogen Bonds. *J. Chem. Theory Comput.* **8,** 1325–1333 (2012).

73. Schulz, W. G. Ablaze Over Furniture Fires. *Chem. Eng. News Arch.* **90,** 28–33 (2012).

74. Hogue, C. EPA Targets Flame Retardants. *Chem. Eng. News Arch.* **90,** 34–37 (2012).

75. Tiwari, A. & Panigrahi, S. K. HBAT: a complete package for analysing strong and weak hydrogen bonds in macromolecular crystal structures. *In Silico Biol.* **7,** 651–61 (2007).

CHAPTER 3

FORCE FIELD MODEL OF PERIODIC TRENDS IN BIOMOLECULAR HALOGEN BONDS[†]

SUMMARY

The study of the noncovalent interaction now defined as a halogen bond (X-bond) has become one of the fastest growing areas in experimental and theoretical chemistry—its applications as a design tool are highly extensive. The significance of the interaction in biology has only recently been recognized, but has now become important in medicinal chemistry. We had previously derived a set of empirical potential energy functions to model the structure-energy relationships for bromines in biomolecular X-bonds (BXBs). Here, we have extended this force field for BXBs (*ff*BXB) to the halogens (Cl, Br, and I) that are commonly seen to form stable X-bonds. The *ff*BXB calculated energies show a remarkable one-to-one linear relationship to explicit BXB energies determined from an experimental DNA junction system, thereby validating the approach and the model. The resulting parameters allow us to interpret the stabilizing effects of BXBs in terms of well-defined physical properties of the halogen atoms, including their size, shape, and charge, showing periodic trends that are predictable along the Group VII column of elements. Consequently, we have established the *ff*BXB as accurate

---

[†]Matthew R. Scholfield*, Melissa Coates Ford*, Crystal M. Vander Zanden*, M. Marie Billman‡, P. Shing Ho*, And Anthony K. Rappé‡
*Department of Biochemistry and Molecular Biology, Colorado State University
‡Department of Chemistry, Colorado State University
Fort Collins, CO 80523

computational tool that can be applied to, for example, for the design of new therapeutic compounds against clinically important targets and new biomolecular based materials.

I. INTRODUCTION

The halogen bond (X-bond) has become an important molecular design tool in nearly all areas of the chemical sciences,[1] from liquid crystals[2] and molecular receptors[3] to nanomaterials[4,5] and molecular conductors.[6] Biology and biochemistry are areas where X-bonds have the potential to make a great impact, particularly in the computer-aided design of new halogenated inhibitors against clinically important targets.[7–10] This potential can only be fully realized with the development of computational tools to accurately model X-bonds in biological molecules (BXBs). We describe here the complete parameterization of a force field for biomolecular X-bonds (an *ff*BXB[11]) that is based on elemental properties of size, shape, and electrostatic potential of the halogens (Cl, Br, and I) commonly seen in biology,[8,12–14] and show for the first time that such a molecular mechanics (MM) approach can directly model experimentally determined energies of BXBs with explicitly defined geometries.[15–17]

The study of halogen bonds and their applications in material and biological chemistry has become one of the fastest growing research areas in chemistry (as seen in the explosive growth in numbers of articles and citations) in the past decade.[5,8] The concept that halogenated compounds can form stable noncovalent complexes was first recognized two centuries ago by Colin when the mixture of molecular iodine with ammonia was seen to produce a shiny liquid that differed from either starting component.[18] The extension of Mulliken's charge transfer concept[19] by Hassel to molecular complexes helped explain the short stabilizing interactions seen between molecular halogens and electron-rich atoms, *e.g.*, between Br_2 and dioxane.[20,21] The

charge transfer bond remained in scientific hibernation for two decades, until it reemerged in the 1990's as the halogen bond. The change in name reflected a change in the current understanding that the interaction is primarily electrostatic in nature,[22] analogous to the more widely recognized hydrogen bond (H-bond).[23] The recently published IUPAC definition[24] provides guidelines for how we describe and study X-bonds, and places the interaction well within the accepted chemical tools available for molecular design.

Perhaps the most readily accessible description for the origins of the X-bond is the σ-hole model,[25] which treats the interaction as primarily electrostatic (Figure 3.1b),[22] although there remains significant controversy over the contribution of dispersion[26] and even charge-transfer[27] to the interaction. The model describes X-bonding as a consequence of a halogen forming a σ -bond, such as a C—X bond. Covalently bonding a halogen to, for example, an aromatic uracil base results in the depopulation of the valence p_z-orbital, which creates an electropositive crown (the σ-hole) at the halogen surface diametrically opposed to the σ-bond. The σ-hole serves as the X-bond donor to an electron-rich acceptor. The size and electropositive potential of the σ-hole increases as the size of the halogen increases and as the electron-withdrawing ability of the substituent bound to the halogen increases; thus, the energies of X-bonds can be tuned and fine-tuned,[8,28] which make the interaction particularly useful as a design tool.

X-bonds in biomolecular systems (BXBs) are predominantly found in complexes of proteins with halogenated inhibitors,[8,12] with the carbonyl oxygen of the peptide bonds found to be the primary acceptor in these complexes[8,12–14,29]. For the most part, the geometries relating the donors to acceptors in BXBs follow those seen in simpler organic complexes, with distances shorter than the sums of the respective van der Waals radii ($\sum R_{vdW}$), an angle of approach of the acceptor to the σ-hole along C—X bond (Θ_1) that is nearly linear, and an angle of approach of

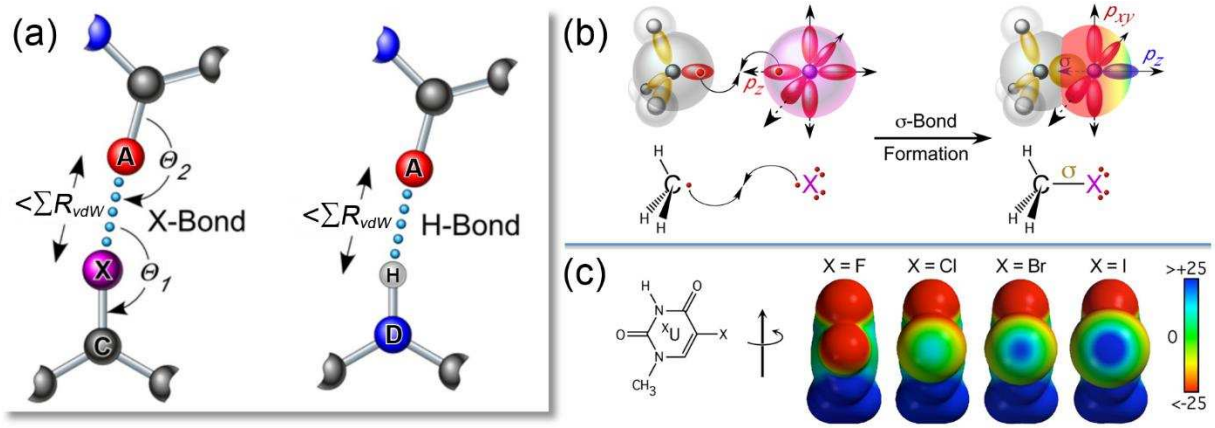

Figure 3.1. The halogen bond. (a) The geometry of a halogen bond (X-bond) relative to a hydrogen bond (H-bond), adapted from ref 12. The X-bond is similar to an H-bond, both defined by distances between the halogen donor (X) and the electron-rich acceptor (A) that is shorter than the sum of their standard van der Waals radii ($\sum R_{vdW}$). The X-bond is more directional, with the approach of the acceptor to C—X bond defined by the angle Θ_1 and the halogen to acceptor by Θ_2. (b) The σ-hole model of halogen bonds. The formation of a single C—X covalent bond (σ -bond, yellow) pairs an electron from the carbon with one from the valence p_z-orbital of the halogen. The depopulation of the p_z-orbital creates an electropositive crown (blue) and flattening of the atomic radius opposite the σ -bond, while the $p_{x,y}$-orbitals remain fully occupied, resulting in an electronegative annulus perpendicular to the covalent bond. (c) Electrostatic potential of halogen substituents at the 5-position of a uracil base (adopted from ref 12). The DFT calculated electrostatic potentials (from >+25 kcal/mol, in blue, to <-25 kcal/mol, in red) for the various 5-halouracils (^{X}U), show σ -holes that increase in size and intensity as the size of the halogen increases from F to Cl to Br to I.

the halogen to the acceptor covalent bond (Θ_2) either towards the non-bonding electrons or π-electrons of the acceptor.[8,12,29–31] One exception has been that most peptide bonds in proteins are hydrogen bonded (H-bonded) in secondary structures (α-helices and β-sheets), leading us to further propose that BXBs are geometrically and energetically orthogonal to H-bonds when they share the same acceptor atoms[32]—the relationship now also seen in small molecule complexes.[33,34] Not surprisingly, the most direct application of BXBs has been in medicinal chemistry for the computer-aided design of halogenated inhibitors and drugs.[7–10,35,36] Recently, a library of halogen-enriched fragments to help identify potential BXB sites to facilitate drug discovery,[37] in one example to design inhibitors that induce apoptosis in human cell lines carrying the cancer-related Y220C mutation of p53.[38]

The various geometries of BXBs are associated with different energies of interaction. Thus, in order to understand how a particular X-bonding interaction may contribute, for example, to the affinity of a protein for a halogenated ligand, we need to understand and describe the structure-energy relationships in a rigorously quantitative manner. There is now a fairly large library of crystal structures of proteins bound to halogenated ligands, along with their affinities as reflected in dissociation constants (K_Ds) or IC50s[29]. The energies for a specific geometry of any particular BXB in such complexes, however, cannot be easily isolated from all other competing or contributing interactions.

To overcome this problem, we developed a DNA model system (Figure 3.2)[15] in which the structures of various BXBs can be determined crystallographically[17] and their energies directly measured either through a crystallographic[15,16] or calorimetric assay.[17,40] The system competes a BXB against analogous H-bonds to stabilize a four-stranded DNA junction[15] and, with the energy of the H-bond now determined, the actual energy of each BXB geometry can be

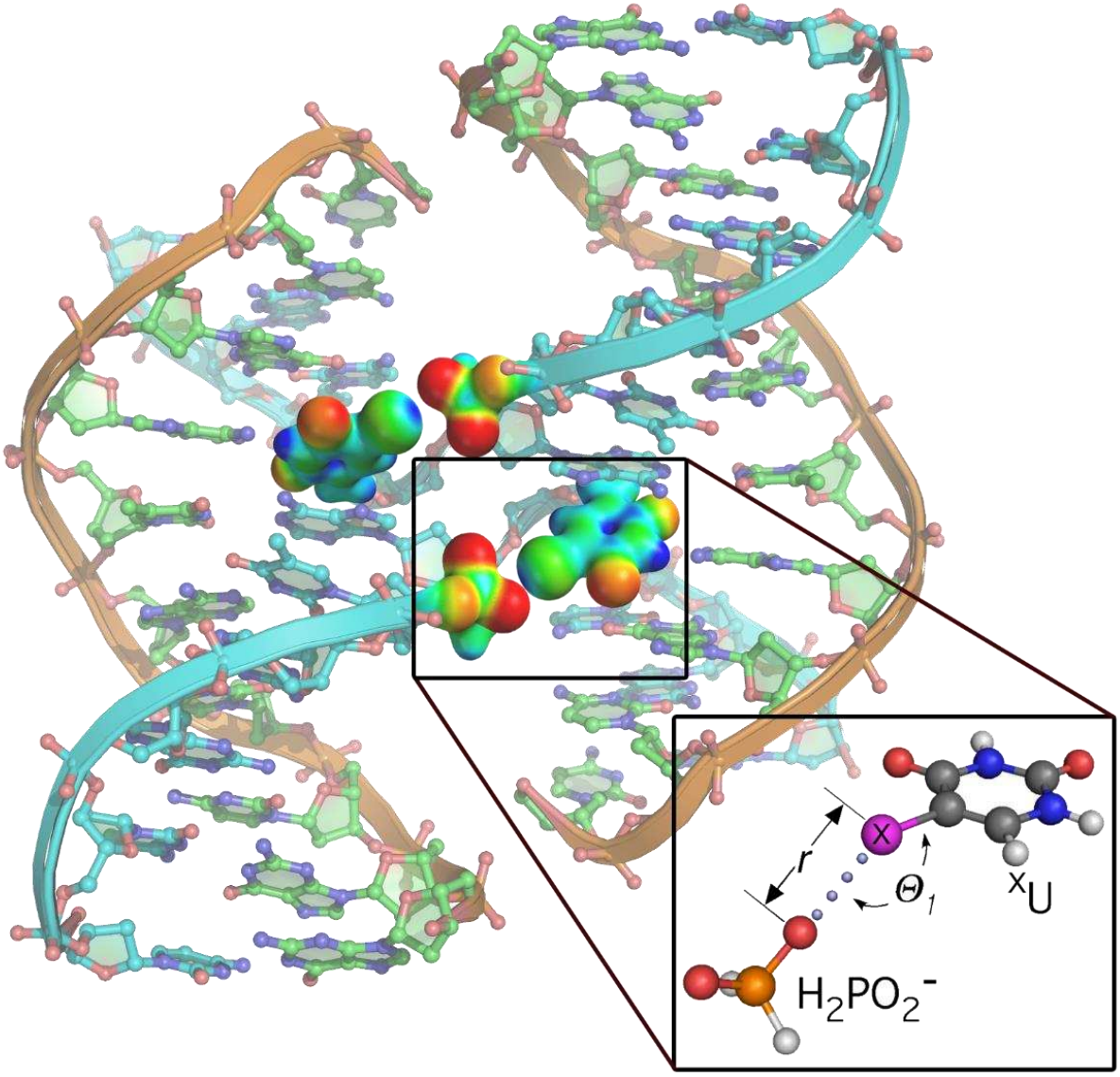

Figure 3.2. DNA junction system to determine experimental BXB energies. A four-stranded DNA junction[39] construct was designed to compete a stabilizing H-bond from the amino group of a cytosine base to the phosphate oxygen of the junction cross-over (orange strands) against X-bonds from an ^{X}U base to the analogous phosphate (cyan strands).[15] Experimental energies for the chlorinated Cl1J (one Cl BXB) and Cl2J (two Cl BXBs), brominated Br1J (one Br BXB) and Br2J (two Br BXBs), and the iodinated I2J (two I BXBs) constructs of the four-stranded DNA junction system were determined by a crystallographic assay or by differential scanning calorimetry (DSC) in solution.[17,40] The geometries of each construct were determined by X-ray crystallography. The specific X-bonded interacting groups are modeled by an $^{X}U\cdots H_2PO_2^{-1}$ (hypophosphite) pair[11] for high-level QM calculations (MP2 calculations, applying the aug-cc-PVTZ basis set for F, Cl, and Br; and aug-cc-PVTZ-PP for I, with BSSE correction).

determined in isolation.[17,40] Using this system, we have shown that the strength of halogen bonds increases according to the series F < Cl < Br < I, with energies ranging from -0.52 to -6 kcal/mol.[17] In addition, as the X-bonding energy becomes more negative and stabilizing, the geometry becomes more ideal (the $\sum R_{vdW}$ becomes shorter relative to the standard atomic radii and the approach of the acceptor, Θ_1, becomes more linear towards the σ-hole).[17]

The experimentally determined structure-energy relationships can now be uniquely applied to validate any type of computational model. Quantum mechanics (QM) is the most accurate method to computationally model X-bonds; however, the most accessible approach for modeling biomolecular systems is molecular mechanics (MM), which in its standard form improperly treats halogens as isotropic charges, thereby ignoring the σ-hole. It is essential, therefore, to develop a computational method that can accurately model BXBs, which can be incorporated into a standard MM force field to simulate the structure-energy relationships of X-bonds in biomolecular systems, including inhibitor-protein complexes.

The "extra point" or "explicit σ-hole" extensions[41–43] to the AMBER force field[44] was one of the first applications of MM to model BXBs. In this approach, the halogen is described by the standard MM parameters for size, shape and atomic charge, with the positive σ-hole mimicked by the addition of a massless positive charge placed at or near the surface of the halogen, opposite the σ-bond (Figure 3.3a). Such MM approaches have been successful in simulating the geometries and trends in IC50s (to reflect the affinities) of inhibitors and drugs to various proteins, including Clapthesin L, p53, and, in particular, protein kinases (reviewed in refs 9 and 30).

We have approached the problem of modeling BXBs from a different perspective, by deriving an empirical force field for BXBs[11] that, unlike the pseudo-atom approach, is based on

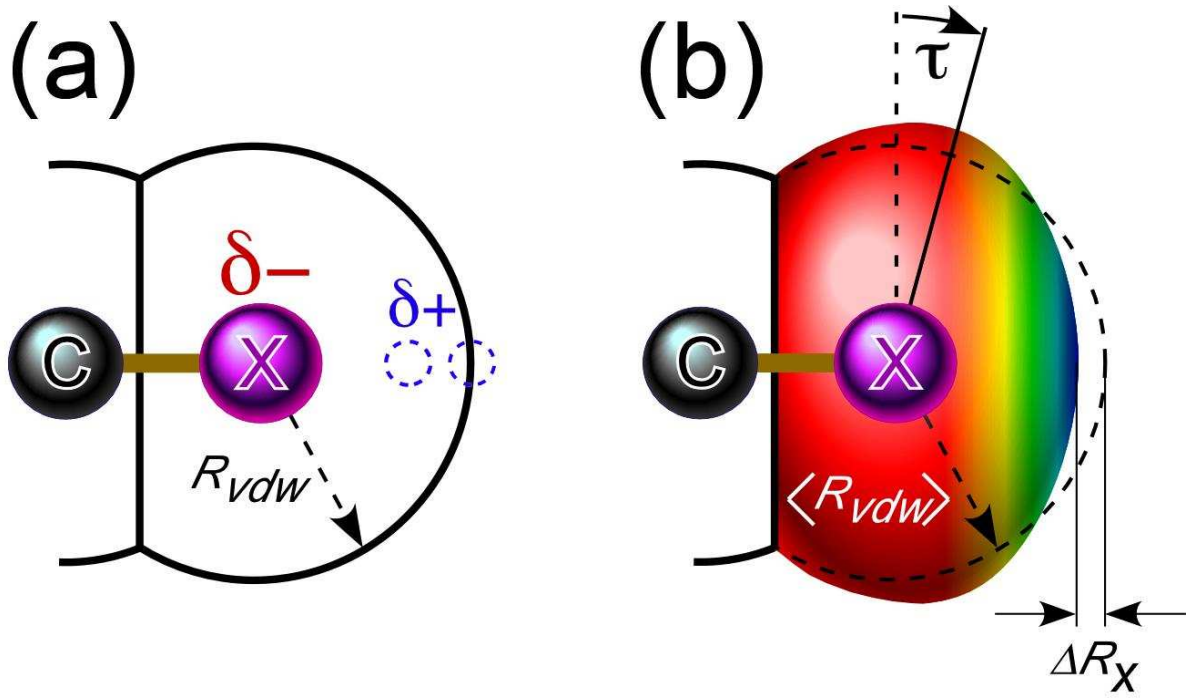

Figure 3.3. Molecular mechanics (MM) approaches to model BXBs. (a) positive extra point or explicit σ-hole models incorporate a massless positive pseudoatom at[41,42] or near the surface[43] of a standard electronegative and spherical halogen (with a standard R_{vdW}) to mimic the positive crown of the σ-hole. The force field for biological X-bonds (*ff*BXB[11]) model assumes that the size of the halogen varies by ΔR_X from an average radius ($<R_{vdW}>$) in an angle dependent manner (with the bulge tipped by an angle τ away from 90° relative to the C—X bond), and the charge of the halogen similarly dependent on the angle of approach, varying from positive opposite the C—X bond (blue) to neutral (green) to negative perpendicular to the C—X bond (red).

the fundamental atomic properties of size, shape, and electrostatic charge (Figure 3.3b). The resulting *ff*BXB assigns an anisotropic distribution of charge across the atom surface to model the directional electrostatic potential[45] and an aspherical atomic radius to model the polar flattening[46] (the reduction in the R_{vdW} in the direction of the σ-hole, as observed in high resolution crystal structures) of a halogen. The initial application of the *ff*BXB showed that the relative energies of Br-BXBs for two different geometries could be accurately modeled based on the AMBER force field.[11] In the current study, we fully parameterize the *ff*BXB, and show for the first time that a molecular mechanics model correlates in a direct one-to-one manner with the experimentally determined structure-energy relationships for all halogens commonly seen in biological systems (Cl, Br, and I). Furthermore, the parameters of the model can be readily interpreted in terms of the physicochemical properties of the halogen atom in covalent compounds, when X-bonded to an electron-rich acceptor.

II. THEORY AND METHODS

We had previously defined a set of potential energy force field functions to describe the structure-energy relationships of biomolecular halogen bonds—the *ff*BXB—that was specific for Br.[11] In the current study, we extend and parameterize the *ff*BXB for the common biological halogens Cl, Br, and I.

The *ff*BXB functions were derived based on well defined physical properties of covalently bonded halogen atoms, including the anisotropic distribution of charge across the atomic surface[45] and the aspherical shape of the atomic radius, seen as polar flattening;[46] both aspects are consistent with predictions from the σ-hole model.[25] The anisotropic charge distribution (positive for the linear approach towards the σ-hole, Θ_1 = 180°, and negative

perpendicular to the σ-hole, Θ_1 = 90°) was modeled simply by defining the effective charge of the halogen (Z_x) as the angle-dependent function $Z_X = A\cos(\upsilon\alpha) + B$ (where A is the amplitude of the charge difference between the linear and perpendicular approaches, *B* defines the baseline of the cosine function, $\alpha = 180° - \Theta_1$, and υ is the period of the cosine function).[11] The electrostatic potential for the interaction (V_{elect}) of an acceptor atom (with charge Z_A) at any angle of approach to the halogen is thus described by a standard potential (Eq. 1), where the exponent *n* defines the distance relationship (*r*) for the interaction, *e* is the standard charge of an electron, and *D* is the dielectric of the medium separating the atoms.

$$V_{elect} = \frac{Z_X Z_A e^2}{D r^n} \qquad \text{Eq. 1}$$

In addition to the electrostatics, the *ff*BXB incorporated polar flattening by defining the shape of covalently bonded halogens as being directional (shorter for a linear approach and larger for a perpendicular approach). The aspherical shape was manifested in an angle-dependent function that described the very short-range repulsive term of the standard Lennard-Jones (V_{LJ}) potential.[11] The X-bond specific V_{LJ}, in a form consistent with that used in the common AMBER force field, defines an average effective van der Waals radius for the halogen ($\langle R_{vdW(X)} \rangle$) and a perturbation to the radius that is dependent on the angle of approach, in a similar fashion as the charge of the halogen ($\Delta R_X \cos(\upsilon\alpha)$). The polar flattening is applied only to the repulsive term of the standard V_{LJ}, resulting in Eq. 2 for the overall interaction potential (where ε_X and ε_A are the energy contributions from the X-bond donor and acceptor, respectively, $R_{vdW(A)}$ is the standard van der Waals radius of the acceptor, and *r* is the distance between donor and acceptor atoms).

$$V_{LJ} = \sqrt{\varepsilon_X \varepsilon_A}\left[\left(\frac{R_{vdW(A)} + \langle R_{vdw(X)} \rangle - \Delta R_X \cos(\nu\alpha)}{r}\right)^{12} - 2\left(\frac{R_{vdW(A)} + \langle R_{vdw(X)} \rangle}{r}\right)^{6}\right] \qquad \text{Eq. 2}$$

In this work, we derive a set of parameters for the V_{elec} and V_{LJ} functions of the *ff*BXB that are consistent and applicable across the common halogens (Cl, Br, and I) that are seen to

form stabilizing X-bonds in the DNA model system. The derivation of these parameters follow a general strategy, where we: 1) apply high-level quantum mechanical (QM) calculations on a pair of simple compounds to model the energies at various distances and angles of approach between the X-bond donor and acceptor atoms in the experimental DNA junction system; 2) determine the shape parameters for covalent Cl, Br, and I atoms; 3) determine the size and charge parameters for Cl, Br, and I atoms; and 4) validate the model system and force field parameters by comparing to BXB structure-energy relationships experimentally determined in the DNA system.

QM Calculations for BXB Interaction Energies in Models of the Experimental DNA System

We had previously shown that the BXBs in the DNA system could be mimicked by a an isolated halogenated uracil base and a hypophosphate group (an $^{X}U\cdots H_2PO_2^{-1}$ interacting pair, Figure 3.2 inset)—a simple model that is amenable to high-level QM calculation.[11] In order to map the structure-energy relationships, we constructed $^{X}U\cdots H_2PO_2^{-1}$ pairs for X = Cl, Br, and I, maintaining an X···½O—P alignment of 180°. For each set, the BXB geometries were defined by varying the angle of approach from Θ_1 = 90° to 180°, and the interacting distances varying from 70% to 100% of the standard radius of the halogen of ^{X}U to the oxygen of the hypophosphite.

QM energies for each $^{X}U\cdots H_2PO_2^{-1}$ geometry were calculated using the GAUSSIAN 09, applying the Møller-Plesset second order (MP2) calculations with cyclohexane as the solvent (*D* = 2, relative to a vacuum). Appropriate polarizable basis sets that include dispersion were applied to the calculations, according to the halogen (aug-cc-PVTZ for F, Cl, and Br; aug-cc-PVTZ-PP for I from the EMSL Basis Set Exchange[47]). Basis set superposition errors (BSSE) were

determined from a separate counterpoise gas phase calculation and directly summed into the calculated solvent phase energy.

Determining ffBXB Parameters from QM calculated energies

The *ff*BXB has seven parameters that need to be determined for the V_{elect} and V_{LJ} potential functions. The advantage of the *ff*BXB is that these parameters describe well defined physical characteristics of each halogen and, theoretically, can be determined through global fitting of Eq. 1 and 2 to the QM energies calculated for each geometry of the $^{X}U\cdots H_2PO_2^{-1}$ pairs for Cl, Br, and I. Unfortunately, it is clear from attempts at such global fitting that some parameters are coupled, resulting in fitting errors of >100% in some cases. We therefore, needed to decouple some of the parameters from the *ff*BXB functions. Our previous studies indicated that the two parameters ΔR_X and ε_X in the V_{LJ} function could be determined independently from the inherent properties of the halogens themselves.[11] Once values for ΔR_X and ε_X were defined for a particular halogen, the remaining parameters could be robustly determined by the combined *ff*BXB functions applied to the QM energies of the $^{X}U\cdots H_2PO_2^{-1}$ pairs. The first step in parameterization, therefore, is to determine ΔR_X and ε_X for Cl, Br, and I. These parameters were derived using very high level QM calculations of the energies of the isolate halogen atoms interacting with a helium atom, with the He serving as a small neutral, non-polarizable probe.[11,48] We can then determine the size and shape parameters by fitting the V_{LJ} function against the QM calculated energies.

The ΔR_X and ε_X for Cl, Br, and I were obtained from counterpoise-corrected CCSD(T) and Hartree Fock potential curves for X···He, the helium either approaching the singly occupied σ p_z orbital or one of the doubly occupied π orbitals, p_x. The halogen basis sets, aug-cc-PVTZ for

F, Cl, and Br, and aug-cc-PVTZ-PP for I, were augmented by single d, f, and g functions, with exponents selected to maximize the X···He interaction energy at a distance near the respective van der Waals minima. The aug-cc-PVTZ basis set, augmented by single p, d, and f functions, was used for helium (augmentation exponents obtained from I···He). The selected exponents are collected in Table 3.1.

Once the exponents were obtained, the Hartree Fock and CCSD(T) potential curves were combined to obtain estimates for the correlation energy ($V_{corr} = V_{CCSD(T)} - V_{HF}$). A spherical estimate for correlation was obtained by forming a weighted average of the σ and π correlation curves (1/3 σ + 2/3 π). This spherical correlation term was added to the separate σ and π Hartree Fock curves to obtain interaction potential curve estimates. The resulting X···He potential curves were fit to Eq. 2, cos($\upsilon\alpha$) = -1(180˚) for σ approach and +1(90˚) for π approach. The He $\varepsilon_{vdW(A)}$ and $R_{vdW(A)}$ of 0.039 kcal/mol and 1.42 Å were obtained from a He potential curve obtained with the same augmented basis (Figure 3.4).

With ΔR_X and ε_X defined for each halogen, the remaining parameters for the *ff*BXB were determined by fitting the combined V_{elect} and V_{LJ} functions to the QM energies for the $^{X}U\cdots H_2PO_2^{-1}$ pairs for each halogen type (Cl, Br, and I). A program was written in Mathematica[49] to apply a nonlinear least squares fit of the *ff*BXB parameters (ΔR_X and ε_X fixed). The initial fit to all geometries of the X-bonded pair yielded parameters with very high error and large residuals around the minimum of the potential well (for all angles and halogens); it was clear that the very large steric repulsion energies were dominating the fitting routine. Our primary interest is to accurately model the energies at and near the potential wells, where X-bonds would form; thus, we applied a weighting factor (κ) that is biased toward the negative potential energy domains of the *ff*BXB functions. To determine how much weight to apply to the

Table 3.1. Dispersion augmentation exponents for Cl, Br, I, and He.

Halogen	p	d	f	g
Cl		0.0300	0.1200	0.1700
Br		0.1200	0.2000	0.1400
I		0.1600	0.3700	0.2800
He	0.2400	0.2900	0.3300	

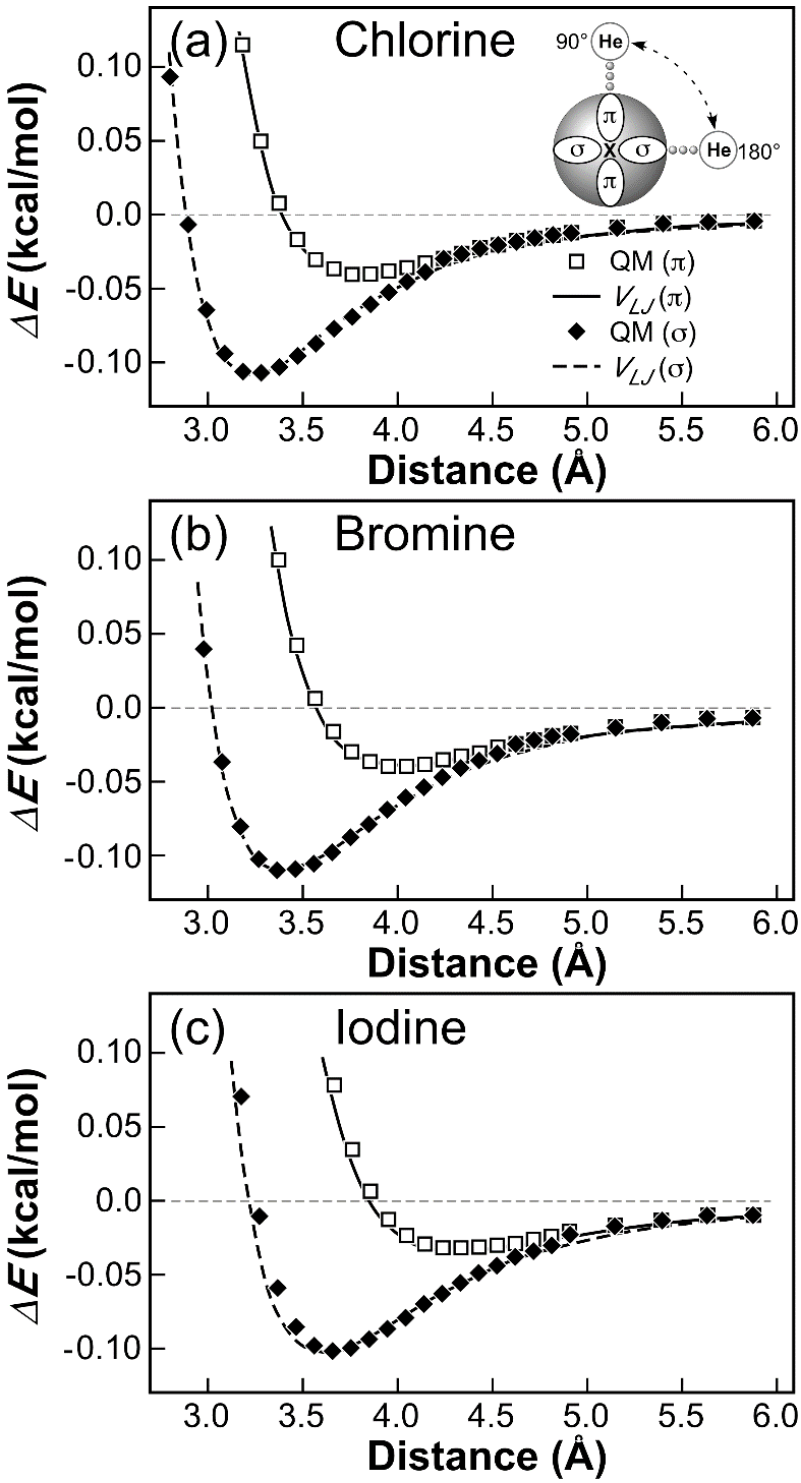

Figure 3.4. Size and shape of Cl (a), Br (b), and I (c). QM energies for each halogen (X), probed with a helium (He) atom, were calculated along the σ- (solid diamonds, 180°) and π- (open squares, 90°) directions. The QM calculated energies at various distances were used to determine the $<\Delta R_{vdW(X)}>$, ΔR_X, and ε_X parameters using the V_{LJ} function (Eq. 2) for the σ- (dashed curves) and π- (solid curves) directions.

fitting of each halogen, we systematically varied κ to minimize the residuals for the negative interactions energies, averaged between Θ_1 = 180˚ and 90˚. Once optimized for each halogen, we applied κ to define a weight, Weight = $1/(\kappa - E_{QM})^2$, to be applied to each. QM energy (E_{QM}) to bias the fitting of the *ff*BXB functions to determine the $<\Delta R_{vdW(X)}>$, ΔR_X and ε_X parameters (Figure 3.5).

III. RESULTS AND DISCUSSION

We had previously derived an empirical force field for biomolecular halogen bonds (*ff*BXB), which correlated well with the relative energies of bromine X-bonds in our DNA model system. In the current study, we have expanded the *ff*BXB to describe the anisotropic shape and surface charge distribution for all the common halogens seen in BXBs (Cl, Br, and I) and show that this molecular mechanics approach directly correlates with the absolute energies measured for specific geometries of these interactions. We have studied the interactions of F within the context of the biological system, but do not observe a stabilizing potential that can be attributed to an X-bond[17] and, therefore, have not included this smallest of halogens in the current force field. The resulting *ff*BXB parameters can be interpreted in terms of periodic trends of the atomic orbitals and their interaction energies for the common elements along the Group VII column that form stabilizing X-bonds in our system.

Our current extension of the *ff*BXB defines parameters for Cl, Br, and I that describe electrostatic interactions as modeled by an angle-dependent Coulombic-type potential (V_{elec}, Eq. 1) and van der Waals type interactions as modeled by an angle-dependent Lennard-Jones potential (V_{LJ}, Eq. 2). The parameters were derived to be consistent with the AMBER force field (one of the most commonly used in the simulation of biomolecular structures[44]). Parameters for

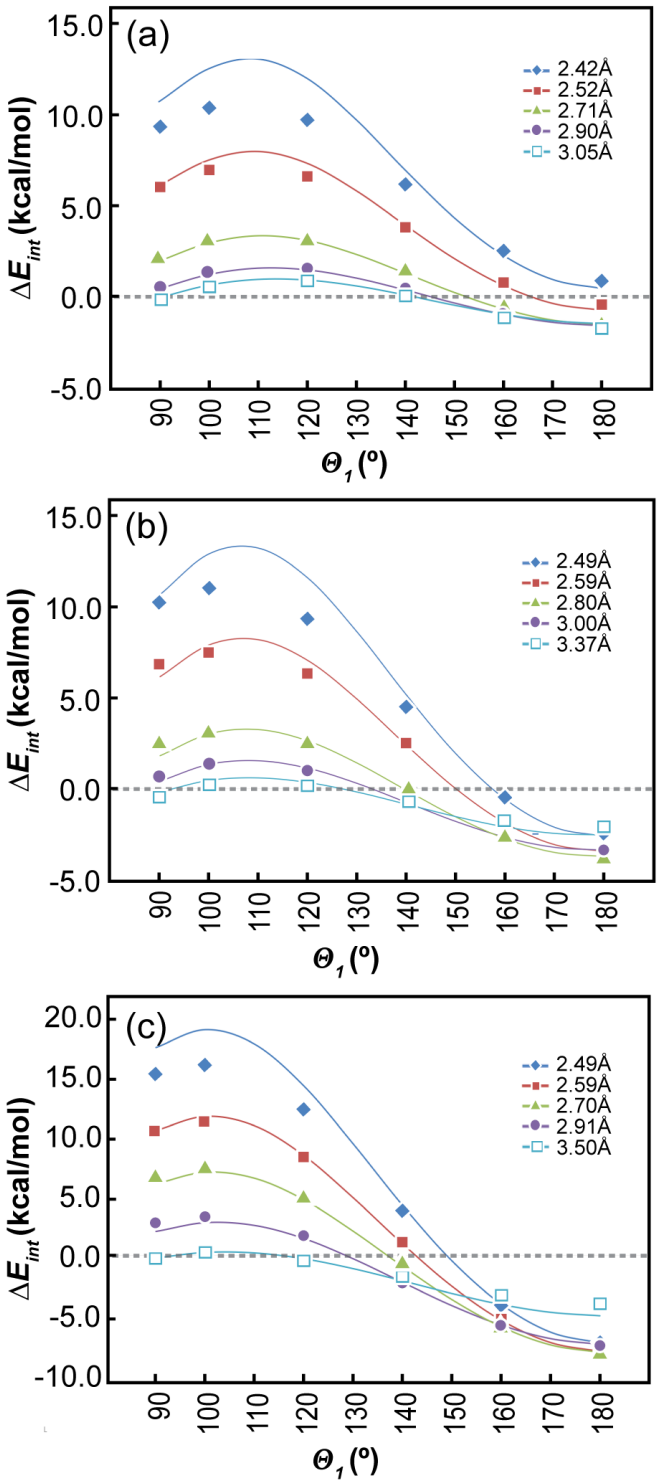

Figure 3.5. Parameterizing the *ff*BXB against QM calculated interaction energies (E_{int}) for Cl (a) Br (b), and (c). The QM E_{int} were calculated for $^{X}U\cdots H_2PO_2^{-1}$ pairs, with various distances separating donors and acceptors (data points) and, for each distance, at angles of approach (Θ_1) from 90° to 180°. The E_{int} were calculated using the parameterized *ff*BXB are shown as curves.

these two classes of interactions were derived by first performing high level QM calculations on model systems that probe the size/shape of halogens using an uncharged, nonpolarizable He atom, and the charge distribution across the halogen surface using a simple model for the $^{X}U\cdots H_2PO_2^{-1}$ pair that mimics the interactions in the experimental DNA junction system. The QM model calculations and the ability of the parameterized force field to model the structure-energy relationships of BXBs were validated by comparison to the absolute energies determined for various geometries of Cl, Br, and I X-bonds measured in the DNA system.

ΔR_X and ε_X ffBXB parameters

The atomic size and shape was initially determined through MP2 calculations of each halogen probed with a He atom at various distances of approach, and at angles of approach of Θ_1 = 180° (along the halogen σ-hole) and Θ_1 =90° (perpendicular to the halogen σ-hole) (Figure 3.3). The distances of the average energy minima at 180° and 90° are, as expected, consistent with the standard R_{vdW} defined in the current AMBER parameters for each halogen. It is clear that all the halogens show polar flattening of their R_{vdW}, with the 180° approach being ~0.15 Å and ~0.2 Å shorter than the 90° approach (equivalent to 15% to 31% flattening relative to their respective AMBER radii).

By fitting the V_{LJ} potential of Eq. 1 to the QM calculated interaction energies, we derived the size and shape parameters for the *ff*BXB for Cl, Br, and I (Table 3.2). We see that, as expected, the effective size ($<R_{vdW(X)}>$) and the polar flattening (ΔR_X) become progressively larger, but the effective depth of the energy well (ε_X) remains fairly similar from Cl to I. The $<R_{vdW(X)}>$ values are slightly larger than the respective AMBER radii (R_{AMBER}), while the ε_X

Table 3.2. Size and shape parameters for Cl, Br, and I. The average van der Waals radius ($<R_{vdW}>$), perturbation to the radius (ΔR_X), and depth of the energy well (ε_X) were calculated by fitting the V_{LJ} to the QM calculated energies for a He probe of each halogen (X). The resulting $<R_{vdW}>$ and ε_X parameters are compared to the radius and energy well (R_{AMBER} and ε_{AMBER}, respectively) assigned to each halogen in the AMBER99[44] parameter file.

X	$<R_{vdW(X)}>$ (Å)	R_{AMBER} (Å)	ΔR_X (Å)	ε_X (kcal/mol)	ε_{AMBER} (kcal/mol)
Cl	2.116	1.95	0.149 ± 0.001	0.107 ± 0.002	0.265
Br	2.369	2.22	0.160 ± 0.018	0.109 ± 0.038	0.320
I	2.658	2.35	0.185 ± 0.006	0.087 ± 0.008	0.400

values are significantly smaller and follow the opposite periodic trend compared to the AMBER energies (ε_{AMBER}). The discrepancy may result from the ε_{AMBER} values not accounting for the contribution of X-bonds. Of these parameters, we carried ΔR_X and ε_X to the next step of deriving the *ff*BXB charge parameters. The $<R_{vdW(X)}>$ will be re-determined in conjunction with the effective charges, because they appear to be codependent.[11]

Size, Shape, and Charge ffBXB Parameters

The model system to parameterize the size and charges of each halogen is the $^{X}U\cdots H_2PO_2^{-1}$ pair, in which the distance between the interacting X···O pair and the Θ_1 angles of approach are varied to sample a broad range of possible BXB geometries for each halogen. The size and shape parameters of $<R_{vdW(X)}>$ and υ (holding the values for ΔR_X and ε_X fixed to those determined by He probing study, Table 3.2), and the charge parameters *n*, *A*, and *B* were fit to best model the QM calculated energies of interaction of the $^{X}U\cdots H_2PO_2^{-1}$ pairs (Table 3.3).

The resulting parameters for the *ff*BXB fit the QM energies for the $^{X}U\cdots H_2PO_2^{-1}$ geometries very well (Figure 3.5). Applying the parameters to V_{elect} (Eq. 1) and V_{LJ} (Eq. 2), we can calculate the BXB energy landscapes for the various halogens for the $^{X}U\cdots H_2PO_2^{-1}$ pair as topological maps (Figure 3.6). These topological geometry-energy maps show that the energy wells become deeper and broader going from Cl to Br to I, as expected. Although the breadth of the wells increase with the size of the halogen, as expected, this is not a linear increase. For example, the -1 kcal/mol stabilizing isotherm extends to ~3.5 Å for Cl, to ~4 Å for Br, but well beyond 6 Å for I. Similarly, the 1 kcal/mol isotherm spans ~30° of Θ_1 for Cl, ~70° for Br, and 90° for I at a distance of 3.0 Å from the halogen. The maps, therefore, indicate that a BXB

Table 3.3. The *ff*BXB parameters for Cl, Br, and I. The parameters for size and shape ($<R_{vdW(X)}>$, υ, and τ) and charge (n, A, and B) were determined by fitting the V_{LJ} (Eq. 2) and V_{elect} (Eq. 1) functions to QM calculated energies for various geometries of the $^{X}U\cdots H_2PO_2^{-1}$ pairs (see text for definition of parameters).

	Size and Shape Parameters			Charge Parameters		
X	$<R_{vdW(X)}>$ (Å)	υ	τ	n	A (e^-)	B (e^-)
Cl	1.719 ± 0.010	2.46 ± 0.05	16.9°	2.40 ± 0.19	0.14 ± 0.03	-0.016 ± 0.005
Br	1.817 ± 0.014	2.42 ± 0.04	15.5°	2.79 ± 0.21	0.23 ± 0.05	0.15 ± 0.04
I	1.922 ±0.015	2.23 ± 0.04	9.2°	2.44 ± 0.15	0.46 ± 0.06	0.29 ± 0.06

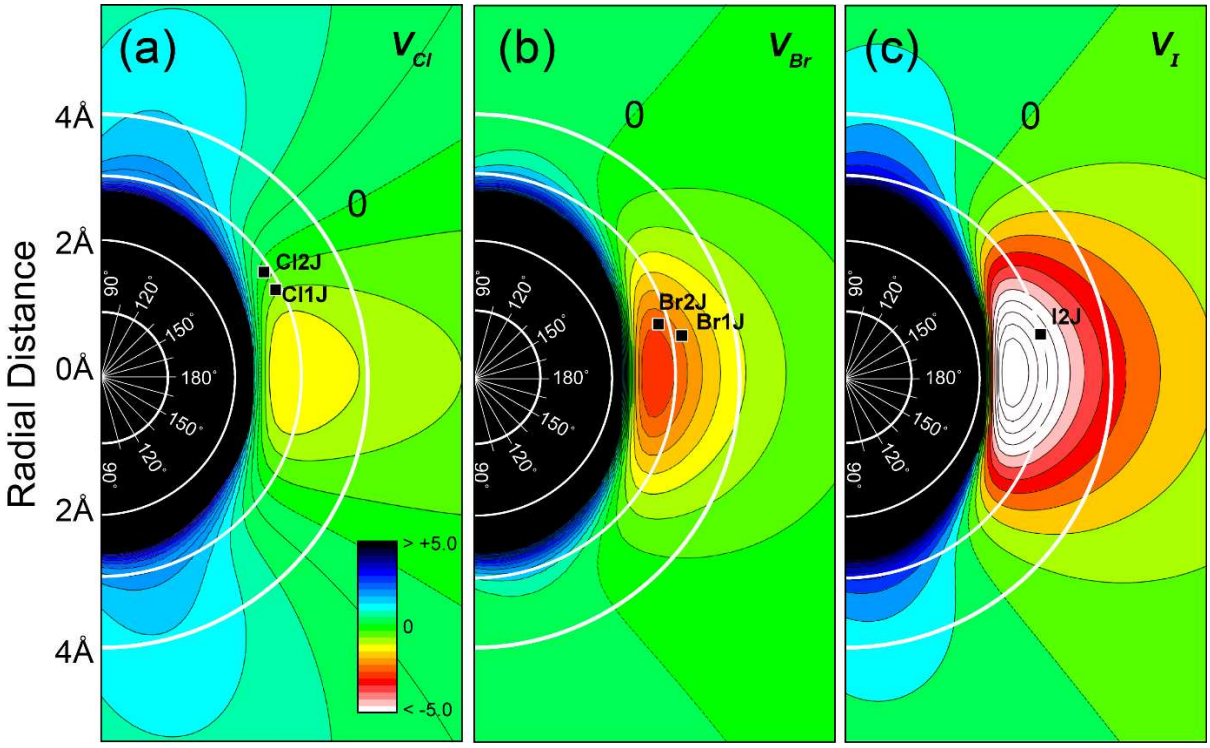

Figure 3.6. Potential energy landscapes calculated with the parameterized *ff*BXB. Energies were calculated for an acceptor atom (carrying a -0.4*e*- charge) to Cl (a), Br (b), and I (c). Energies range from < -5.0 kcal/mol to > +5.0 kcal/mol, from white to black (see color scale insert). The radial distances are shown as semicircular arcs spanning angles of approach from 90° to 180°. The Cl1J, Cl2J, Br1J, Br2J, and I2J BXBs geometries determined from the DNA junction system[15,17] are mapped onto the landscapes. The zero-energy isoenergetic curve is labeled.

acceptor will see a stabilizing interaction energy over a broader range of distance and angle geometries for an I donor compared to Cl. In addition, the greater depth (and accompanying steeper gradient of the well) will draw the acceptor more towards the optimum BXB geometry for I compared to Cl, as seen in the experimental DNA system[17].

The parameter *n* reflects the distance dependence of the V_{elec} potential, which in turn informs us of the type of electrostatic interaction involved in the BXB. For the three halogens parameterized here, the average $n = 2.54$ (S.D. = 0.11). The value of $n \approx 2.5$ suggests that X-bonds lie somewhere between a dipole-charge and dipole-dipole interaction, which accounts for its strong directionality, even compared to the H-bond. In the case of X-bonds, the dipole can be interpreted as being defined by the anisotropic charge distribution across the atomic surface of the halogen.

We can ask how well this simple $^{X}U\cdots H_2PO_2^{-1}$ model system, the associated QM calculations, and the resulting *ff*BXB parameters simulate the experimentally measured structure-energy relationships. The experimental DNA junction had uniquely allowed us to determine the explicit energies of BXBs for Cl, Br, and I,[17] either from a crystallographic assay or in solution by differential scanning calorimetry (DSC).[40] The DSC energies are measured as enthalpies (ΔH_{DSC}), while the crystallographic energies (ΔE_{cryst}) can be interpreted as equivalent to the enthalpies of interaction (assuming the entropies of the competing molecular interactions in the crystals are nearly identical). The experimental energies (ΔH_{DSC} and ΔE_{cryst}) for the various BXBs and their geometries show a linear relationship to the QM energies (ΔE_{QM}) calculated for the $^{X}U\cdots H_2PO_2^{-1}$ pair, that is remarkably close to one-to-one (Table 3.4, Figure 3.7).

The experimental *vs* calculated energies are related by a slope ~0.9, with a *y*-intercept ~ at -0.5 kcal/mol (R^2 correlation = 0.97). Thus, the $^{X}U\cdots H_2PO_2^{-1}$ pair and the QM calculations on

Table 3.4. Comparison of experimental BXB energies to QM and *ff*BXB calculated energies. Experimental energies for the various constructs in the four-stranded DNA junction (Figure 3.2) system were determined by from a crystallographic assay ($\Delta \boldsymbol{E}_{exp}$) or by DSC in solution ($\Delta \boldsymbol{H}_{exp}$). The QM ($\Delta \boldsymbol{E}_{QM}$) and *ff*BXB ($\Delta \boldsymbol{E}_{ffBXB}$) energies were calculated based on the $^{X}U\cdots H_2PO_2^{-1}$ pair fixed at the crystallographically determined geometries of each construct, defined by the distance between the donors and acceptors ($\boldsymbol{r}_{(X\cdots O)}$) and the angle of approach of the acceptor to the donor (Θ_1).

Structure (PDB code)	$\boldsymbol{r}_{(X\cdots O)}$	Θ_1	$\Delta \boldsymbol{E}_{exp}/\Delta \boldsymbol{H}_{exp}$	$\Delta \boldsymbol{E}_{QM}$	$\Delta \boldsymbol{E}_{ffBXB}$
Cl1J (4GSG)	2.88 Å	157.0°	-0.79 ± 0.12	-0.79	-0.76
Cl2J (4GQD)	2.88 Å	146.0°	-0.79 ± 0.12	0.02	-0.01
Br1J (2ORF)	3.32 Å	167.2°	-2.28 ± 0.11	-2.07	-2.41
Br2J (2ORG)	2.87 Å	163.2°	-3.6 ± 1.3	-2.93	-2.94
I2J (4GRE)	3.01 Å	170.7°	-5.9 ± 1.1	-5.93	-5.96

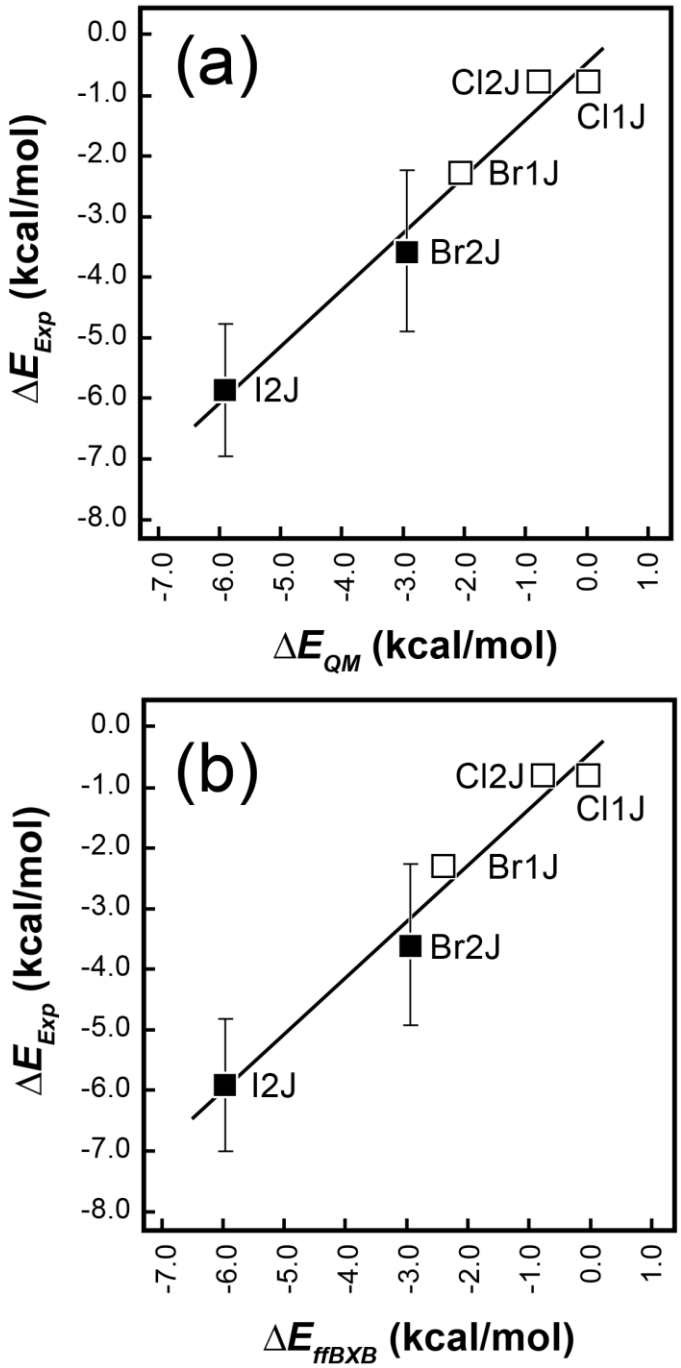

Figure 3.7. Correlation of experimental energies (ΔE_{Exp}) to QM (ΔE_{QM}, **a**) and *ff*BXB (ΔE_{ffBXB}, **b**) for BXBs. Experimental energies determined from crystallographic competition assays (open squares) or DSC measurements (filled squares) on DNA junctions are compared to energies calculated by QM or the *ff*BXB functions for the model $^{X}U\cdots H_2PO_2^{-1}$ pairs. Geometries for the $^{X}U\cdots H_2PO_2^{-1}$ were taken from the crystal structures, see Table 3.4. The linear relationships show $\Delta E_{Exp} = 0.921\Delta E_{QM} - 0.52$ kcal/mol for the QM energies ($R^2 = 0.97$), and $\Delta E_{Exp} = 0.916\Delta E_{ffBXB} - 0.46$ kcal/mol for the QM energies ($R^2 = 0.97$).

interaction energies between the components of the molecular pair are appropriate to model BXBs. The experimental energies also showed a strong linear one-to-one relationship to the *ff*BXB calculated interaction energies (near identical slope and *y*-intercept, and R^2 correlation =0.97), attesting to the ability of the empirical model to reproduce the experimental structure-energy relationships.

Periodic Trends and Interpretation of ffBXB Parameters

The *ff*BXB parameters show periodic tends for the Group VII elements, which allow QM calculations to be interpreted in terms of the physical properties of the halogens and their contributions to the BXB interactions (Figure 3.8). As expected, the average effective van der Waals radii $\langle R_{vdW(X)} \rangle$ become larger progressing down the column from Cl to I. The $\langle R_{vdW(X)} \rangle$ derived from the *ff*BXB, however, are smaller than the standard AMBER values[44] or the He probe values. Indeed, they are closer to the Bondi radii,[50] which uses an alternative definition for R_{vdW} and, thus, a V_{LJ} equation that differs from the AMBER equation. The shorter *ff*BXB $\langle R_{vdW(X)} \rangle$ distances may reflect the coupling of the shape and size of the halogen with the charge parameters.[11]

The parameter υ defines the angle (relative to the C—X bond) at which the halogen shows the largest effective radius and most electronegative charge. Relative to the σ-hole model, υ can be interpreted as the angle at which the $p_{x,y}$-atomic orbitals are tipped away from the *x,y*-plane (a tip angle τ), assuming that the σ-hole lies along the p_z-orbital. Ideally, $\tau = 0°$ (for υ = 2.0), that is, the $p_{x,y}$-orbitals should lie in the *x,y*-plane, and the halogen should be largest and most negative perpendicular to the σ-hole. We see that in fact τ is largest for Cl and approaches the ideal value for I, or that the orientation of the $p_{x,y}$-orbitals are most tipped for the smallest

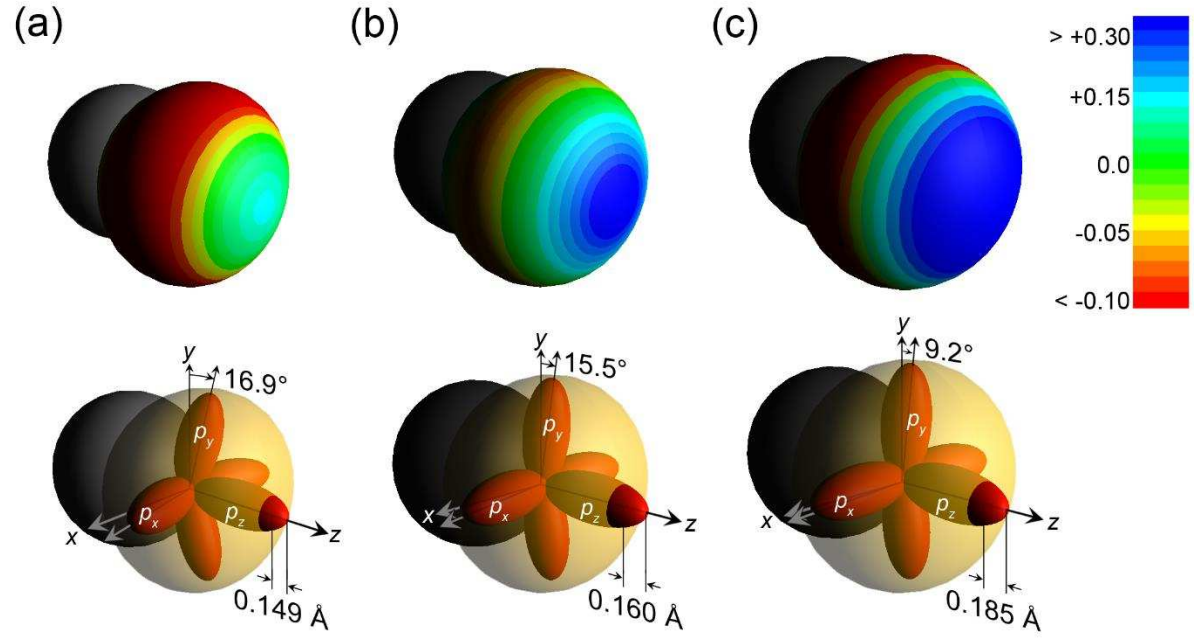

Figure 3.8. Atomic interpretation of the charge and shape *ff*BXB parameters for Cl (a), Br (b), and I (c). The electrostatic charge parameters *A* and *B* define the distribution of charges (top row, scaled from < -0.1 e^- in red to > +0.3 e^-). The bottom row shows the effect of the shape parameter ΔR_X in flattening of the effective $R_{vdW(X)}$ relative to the expected size of the p_z-atomic orbital, and the effect of υ on the tipping (τ angle) of the $p_{x,y}$-orbitals relative to the standard Cartesian coordinate plane.

atom. A near linear relationship is observed between τ and $\langle R_{vdW(X)} \rangle$ (τ = [-38.1 deg/Å] $\langle R_{vdW(X)} \rangle$ + 83.2°, R^2 = 0.90), suggesting that the further the $p_{x,y}$-orbitals are from the σ-hole (even when the hole is more electropositive) the less they are perturbed. The τ -angle has a number of effects on the non-covalent interactions of halogens with other atoms, including the optimum angle for potential H-bonds to the negative annulus of the halogen[51] and the electroneutral point across the halogen surface. Covalently bonded halogens are amphipathic, capable of serving as both X-bond donors and H-bond acceptors, H-bonds generally ascribed as being orthogonal to X-bonds when they share the same halogen atom. The τ -angle derived here suggests that this is true primarily for I, but the angular relationship deviates significantly going to Br and Cl, and may be even more shallow with a more electropositive σ-hole.

A and *B* define the charge of the atomic surface at any angle Θ_1, with *A* specifying the amplitude of the difference between the most positive and most negative partial charge, and *B* specifying the charge at the midpoint of the cosine function. Not surprisingly, *A* increases with the size of the halogen, which is consistent with the larger halogens being more polarized. Similarly, *B* is most negative for Cl and becomes more positive towards I. Indeed, the parameters *A* and *B* are highly correlated with $\langle R_{vdW(X)} \rangle$ (Figure 3.8). Thus, the σ-hole is seen to be least positive (+0.12 e^-) for Cl and becomes more positive proceeding down the column to Br (+0.38 e^-) and to I (+0.75 e^-).

The parameters *A*, *B*, and υ allow the electroneutral angle (Θ_0) to be calculated according to the relationship $\Theta_0 = \frac{\cos^{-1}(A/B)}{\nu}$. From the *ff*BXB parameters, Θ_0 = 146° for Cl, 126° for Br, and 122° for I. The Θ_0 angles reflect the angular size of the σ-hole for each halogen and, as predicted from the potential energy landscape, the range at which a BXB is electrostatically stable is broader for the larger halogens. In addition, *A*, *B*, and υ can be used to calculate the

overall charge across the entire accessible surface for each halogen. The resulting surface charges (-0.095*e* for Cl, +0.008*e* for Br, and 0.026*e* for I) follow closely the charges calculated by the restrained electrostatic potential (RESP) approach[52] (-0.046*e* for Cl, -0.0072*e* for Br, and 0.035*e* for I), a standard approach to defining atomic charges for an unknown ligand for the AMBER force field. Both methods indicate that Cl carries an overall negative charge, Br essentially neutral, and I is positive.

IV. CONCLUSIONS

In this study, we have extended the force field previously derived for Br X-bonds to be applicable to Cl, Br, and I. The empirical models were derived from QM calculations, and this general approach (both the QM models and *ff*BXB approach) is shown here to be valid by comparing the calculated structure-energies relationships to the explicit energies measured in a DNA system. The parameters derived for the *ff*BXB now allow us to interpret the periodic trends in terms of physical properties of the individual halogen atoms, including the size, shape, and anisotropy of the charge across the atomic surface. The X-bond is now interpreted as being more than a dipole-dipole and not quite a charge-dipole interaction, giving it a significantly shorter distance relationship and more directionality than expected for a purely Coulombic electrostatic interaction. As a shorter distance interaction, its influence would not be expected to extend much beyond 6 Å for any halogen type (being longest for I, which has the strongest □-hole). The overall charge across the halogen surface calculated from the *ff*BXB parameters closely matches the overall RESP charges. The strong linear correlation between the charges from the two approaches (R = 0.92) reflects the ability of the *ff*BXB to model the QM calculated electrostatic potentials of BXB donors.

That being said, the shorter $<R_{vdW(X)}>$ from the *ff*BXB compared to the AMBER radii may reflect an inability of a purely electrostatic model to account for the short BXB distances without significantly compressing the effective atomic radius. This shortfall suggests that the purely electrostatic definition of the σ-hole model, even with dispersion taken into account, may not fully account for the structure-energy relationships of BXBs. For example, charge-transfer is not an effect that is directly modeled by the electrostatic σ-hole model. If there is a significant contribution of charge transfer to X-bonds,[27] we would expect the transfer of an electron from the HOMO of the electron donor to the LUMO of acceptor to affect the atomic shape of both atoms, which may be reflected in the change in their size and shape and, for an MM model, the $<R_{vdW(X)}>$ and minimum radius in the V_{LJ} potential for an X-bonded pair.

Taken in total, we have described a set of empirical potential energy functions and parameters for these functions that perform well in modeling BXBs engineered into our experimental DNA junction system. To be generally useful, the *ff*BXB must be implemented into a workable computational method, in this case those that apply the AMBER force field for simulation—currently a project in progress. We can see, however, how the form of the *ff*BXB functions and the periodic trends help us in this implementation. For example, the short distance relationship ($r^{2.5}$) for V_{elec} indicates that the *ff*BXB can be cutoff at very short distances, limiting the number of possible interacting atoms that need to be considered in these simulations.

One question that can be reasonably raised is how can these parameters be derived for various halogenated compounds without having to perform extensive QM calculations on every type of BXB? Of the seven *ff*BXB parameters, we expect the shape and size parameters ($<R_{vdW(X)}>$, ΔR_X, ε_X, and υ) of the V_{LJ} function to remain fairly consistent for each halogen in any system, and n for V_{elec} to be 2.5 for all BXB systems. Thus, the only parameters that need to be

derived are the electrostatic variables *A* and *B*. The expectation is that one or both will depend greatly on the electron withdrawing or donating ability of the molecule that the halogen is immediately attached to. We have already seen that the overall charge is strongly correlated with the RESP charges (given that we currently have only three points, it is not clear whether the relationship is strictly linear or not). An RESP calculation[52], therefore, which is typically performed for any new molecule or compound prior to running an AMBER simulation, may provide relationships that will allow *A* and *B* to be estimated for the halogen atoms in these molecules.

We conclude, therefore that the functions and various parameters derived for these functions will be generally applicable to model the structure-energy relationships of BXBs. In particular, we see direct application of the *ff*BXB to predict how halogenated bases can be used to control the conformation of DNA constructs, particularly the geometric relationship of the arms in DNA junctions. DNA junctions have become an important scaffolding molecule for constructing various two and three-dimensional structures,[53] including DNA origami[54] and controllable crystal lattices.[55] The ability to use BXBs now in a computationally predictable way opens additional dimensions for such molecular engineering applications. We consider the approach, however, to be general to protein systems as well, including the design of inhibitors against enzymes. Thus, the *ff*BXB will be an important tool for many biomolecular engineering applications.

V. RECOGNITIONS

This work was funded in part by a grant from the National Science Foundation (CHE-1152494) and from Colorado State University. The authors declare no conflicts of interest in this study.

REFERENCES

1. Erdelyi, M. Scientific conferences: A big hello to halogen bonding. *Nat. Chem.* **6,** 762–4 (2014).
2. Nguyen, H. L., Horton, P. N., Hursthouse, M. B., Legon, A. C. & Bruce, D. W. Halogen bonding: a new interaction for liquid crystal formation. *J. Am. Chem. Soc.* **126,** 16–7 (2004).
3. Chudzinski, M. G., McClary, C. A. & Taylor, M. S. Anion receptors composed of hydrogen- and halogen-bond donor groups: modulating selectivity with combinations of distinct noncovalent interactions. *J. Am. Chem. Soc.* **133,** 10559–67 (2011).
4. Shirman, T., Kaminker, R., Freeman, D. & van der Boom, M. E. Halogen-bonding mediated stepwise assembly of gold nanoparticles onto planar surfaces. *ACS Nano* **5,** 6553–63 (2011).
5. Priimagi, A., Cavallo, G., Metrangolo, P. & Resnati, G. The halogen bond in the design of functional supramolecular materials: recent advances. *Acc. Chem. Res.* **46,** 2686–95 (2013).
6. Lieffrig, J., Pennec, R. Le, Jeannin, O., Auban-Senzier, P. & Fourmigué, M. Toward chiral conductors: combining halogen bonding ability and chirality within a single tetrathiafulvalene molecule. *CrystEngComm* **15,** 4408 (2013).
7. Lu, Y. *et al.* Halogen bonding for rational drug design and new drug discovery. *Expert Opin. Drug Discov.* **7,** 375–383 (2012).
8. Scholfield, M. R., Vander Zanden, C. M., Carter, M. & Ho, P. S. Halogen bonding (X-bonding): A biological perspective. *Protein Science* **22,** 139–152 (2013).

9. Wilcken, R., Zimmermann, M. O., Lange, A., Joerger, A. C. & Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **56,** 1363–1388 (2013).

10. Xu, Z. *et al.* Halogen bond: its role beyond drug-target binding affinity for drug discovery and development. *J. Chem. Inf. Model.* **54,** 69–78 (2014).

11. Carter, M., Rappé, A. K. & Ho, P. S. Scalable Anisotropic Shape and Electrostatic Models for Biological Bromine Halogen Bonds. *J. Chem. Theory Comput.* **8,** 2461–2473 (2012).

12. Auffinger, P., Hays, F. A., Westhof, E. & Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 16789–94 (2004).

13. Zhou, P., Tian, F., Zou, J. & Shang, Z. Rediscovery of halogen bonds in protein-ligand complexes. *Mini Rev. Med. Chem.* **10,** 309–314 (2010).

14. Hardegger, L. a *et al.* Systematic Investigation of Halogen Bonding in Protein-Ligand Interactions. *Angew. Chemie Int. Ed.* **50,** 314–318 (2011).

15. Voth, A. R., Hays, F. a & Ho, P. S. Directing macromolecular conformation through halogen bonds. *Proc. Natl. Acad. Sci.* **104,** 6188–6193 (2007).

16. Carter, M. & Ho, P. S. Assaying the Energies of Biological Halogen Bonds. *Cryst. Growth Des.* **11,** 5087–5095 (2011).

17. Carter, M. *et al.* Enthalpy-entropy compensation in biomolecular halogen bonds measured in DNA junctions. *Biochemistry* **52,** 4891–4903 (2013).

18. Colin, J.-J. Sur l'iode. *Ann Chim* **91,** 252–272 (1814).

19. Mulliken, R. S. Molecular Compounds and their Spectra. II. *J. Am. Chem. Soc.* **74,** 811–824 (1952).

20. Hassel, O. Structural aspects of interatomic charge-transfer bonding. *Science* **170,** 497–

502 (1970).

21. Hassel, O. Structural Aspects of Interatomic Charge-Transfer Bonding. *Nobel Lect. Chem. 1963-1970, Elseveir Publ. Company, AMsterdam* (1972).

22. Politzer, P., Murray, J. S. & Clark, T. Halogen bonding: an electrostatically-driven highly directional noncovalent interaction. *Phys. Chem. Chem. Phys.* **12,** 7748–57 (2010).

23. Arunan, E. *et al.* Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl. Chem.* **83,** 1637–1641 (2011).

24. Desiraju, G. R. *et al.* Definition of the halogen bond (IUPAC Recommendations 2013). *Pure Appl. Chem.* **85,** 1711–1713 (2013).

25. Clark, T., Hennemann, M., Murray, J. S. & Politzer, P. Halogen bonding: the sigma-hole. Proceedings of 'Modeling interactions in biomolecules II', Prague, September 5th-9th, 2005. *J. Mol. Model.* **13,** 291–6 (2007).

26. Riley, K. E. & Hobza, P. Investigations into the Nature of Halogen Bonding Including Symmetry Adapted Perturbation Theory Analyses. *J. Chem. Theory Comput.* **4,** 232–242 (2008).

27. Wang, C., Danovich, D., Mo, Y. & Shaik, S. On The Nature of the Halogen Bond. *J. Chem. Theory Comput.* **10,** 3726–37 (2014).

28. Riley, K. E. *et al.* 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,** 3309–3318 (2011).

29. Wilcken, R., Zimmermann, M. O., Lange, A., Zahn, S. & Boeckler, F. M. Using halogen bonds to address the protein backbone: a systematic evaluation. *J. Comput. Aided. Mol. Des.* **26,** 935–45 (2012).

30. Andrea, V. & P., H. The Role of Halogen Bonding in Inhibitor Recognition and Binding by Protein Kinases. *Curr. Top. Med. Chem.* **7,** 1336–1348 (2007).

31. Matter, H. *et al.* Evidence for C-Cl/C-Br...pi interactions as an important contribution to protein-ligand binding affinity. *Angew. Chem. Int. Ed. Engl.* **48,** 2911–6 (2009).

32. Voth, A. R., Khuu, P., Oishi, K. & Ho, P. S. Halogen bonds as orthogonal molecular interactions to hydrogen bonds. *Nat. Chem.* **1,** 74–79 (2009).

33. El-Sheshtawy, H. S., Bassil, B. S., Assaf, K. I., Kortz, U. & Nau, W. M. Halogen bonding inside a molecular container. *J. Am. Chem. Soc.* **134,** 19935–41 (2012).

34. Takemura, A. *et al.* Halogen- and hydrogen-bonded salts and co-crystals formed from 4-halo-2,3,5,6-tetrafluorophenol and cyclic secondary and tertiary amines: orthogonal and non-orthogonal halogen and hydrogen bonding, and synthetic analogues of halogen-bonded biological syst. *Chemistry* **20,** 6721–32 (2014).

35. Lam, P., Clark, C., Smallwood, A. & Alexander, R. Structure-based drug design utilizing halogen bonding: factor Xa inhibitors. in *Proceedings at the 238th ACS National Meeting* 59 (2009).

36. Hernandes, M., Cavalcanti, S. M., Moreira, D. R., de Azevedo Junior, W. & Leite, A. C. Halogen Atoms in the Modern Medicinal Chemistry: Hints for the Drug Design. *Curr. Drug Targets* **11,** 303–314 (2010).

37. Zimmermann, M. O. *et al.* Halogen-enriched fragment libraries as chemical probes for harnessing halogen bonding in fragment-based lead discovery. *Future Med. Chem.* **6,** 617–39 (2014).

38. Wilcken, R. *et al.* Halogen-enriched fragment libraries as leads for drug rescue of mutant p53. *J. Am. Chem. Soc.* **134,** 6810–8 (2012).

39. Eichman, B. F., Vargason, J. M., Mooers, B. H. & Ho, P. S. The Holliday junction in an inverted repeat DNA sequence: sequence effects on the structure of four-way junctions. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 3971–6 (2000).

40. Vander Zanden, C. M., Carter, M. & Ho, P. S. Determining thermodynamic properties of molecular interactions from single crystal studies. *Methods* **64,** 12–8 (2013).

41. Ibrahim, M. A. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* **32,** 2564–74 (2011).

42. Ibrahim, M. A. A. 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,** 3659–69 (2012).

43. Kolář, M. & Hobza, P. On Extension of the Current Biomolecular Empirical Force Field for the Description of Halogen Bonds. *J. Chem. Theory Comput.* **8,** 1325–1333 (2012).

44. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26,** 1668–88 (2005).

45. Murray, J. S., Riley, K. E., Politzer, P. & Clark, T. Directional Weak Intermolecular Interactions: σ-Hole Bonding. *Aust. J. Chem.* **63,** 1598 (2010).

46. Nyburg, S. C. `Polar flattening': non-spherical effective shapes of atoms in crystals. *Acta Crystallogr. Sect. A* **35,** 641–645 (1979).

47. Peterson, K. A., Shepler, B. C., Figgen, D. & Stoll, H. On the spectroscopic and thermochemical properties of ClO, BrO, IO, and their anions. *J. Phys. Chem. A* **110,** 13877–83 (2006).

48. Peebles, S. A., Fowler, P. W. & Legon, A. C. Anisotropic repulsion in complexes B.Cl2 and B.HCl: The shape of the chlorine atom-in-a-molecule. *Chem. Phys. Lett.* **240,** 130–

134 (1995).

49. Wolfram Reseach, I. Mathematica. (2012).

50. Bondi, a. van der Waals Volumes and Radii. *J. Phys. Chem.* **68,** 441–451 (1964).

51. Aakeröy, C. B. *et al.* Combining halogen bonds and hydrogen bonds in the modular assembly of heteromeric infinite 1-D chains. *Chem. Commun.* 4236 (2007). doi:10.1039/b707458a

52. Wang, J., Cieplak, P. & Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21,** 1049–1074 (2000).

53. Seeman, N. C. Structural DNA nanotechnology: an overview. *Methods Mol. Biol.* **303,** 143–66 (2005).

54. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440,** 297–302 (2006).

55. Paukstelis, P. J., Nowakowski, J., Birktoft, J. J. & Seeman, N. C. Crystal Structure of a Continuous Three-Dimensional DNA Lattice. *Chem. Biol.* **11,** 1119–1126 (2004).

CHAPTER 4

ENGINEERING HALOGEN BONDS TO AFFECT PROTEIN STABILITY[†]

SUMMARY

The structures and stabilities of proteins are defined by a series of weak non-covalent electrostatic, van der Waals, and hydrogen bond (HB) interactions. We have now engineered halogen bonds (XBs) site-specifically in order to affect the stability of a model protein, T4 lysozyme. The evidence for XBs is the displacement of the aromatic side chain towards an oxygen acceptor, at distances that are at or less than the sums of their respective van der Waals radii, when the hydroxyl substituent of the wildtype tyrosine is replaced by either a bromine or iodine. In addition, the iodine XB rescues the stabilization energy from an otherwise destabilizing substitution. Thus, XBs are shown to be a powerful new component in the toolbox of non-covalent interactions available for biomolecular engineering, including applications not only to affect protein stability, but also in the design of new therapeutics, protein assemblies, and enzymes.

[†]Matthew R. Scholfield[1], Melissa Coates Ford[1], Anna-Carin C. Carlsson[1], Hawera Butta[1], Ryan A. Mehl[2], and P. Shing Ho[1*]
[1]Department of Biochemistry & Molecular Biology, Colorado State University,
Fort Collins, CO 80523-1870, USA
[2]Department of Biochemistry and Biophysics, Oregon State University
Corvallis, OR 97331, USA

Submitted manuscript

I. INTRODUCTION

Nature is adept at engineering broad spectra of functional nano-architectures from biological molecules. Our efforts to expand on what has naturally evolved are limited by the molecular tools provided by nature. The advent of methods to incorporate non-canonical building blocks into proteins[1–3] and nucleic acids[4] have helped to overcome some of these limitations. We remain, however, constrained by the library of non-covalent interactions that dictate molecular folding. In this study, we introduce the halogen bond (XB) as a new component in the toolbox of molecular interactions available to the biomolecular engineer to affect protein stability and, by inference, in designing new folds, recognition motifs, or assemblies.

Biomolecular engineers rely on a menu of standard non-covalent interactions[5,6] to design more stable molecules, or novel assemblies[7]. DNA, though a highly versatile engineering platform[8–10], still relies primarily on the specific pairing of nucleotide bases through the classical hydrogen bond (HB). The variety of non-canonical amino acids (ncAAs) available for designing polypeptides is much more expansive, but the toolbox for protein engineering remains limited to this same standard menu of molecular interactions.

Recently, the XB[11], which is analogous to the HB, has become widely applied as a design element in supramolecular chemistry[12]. In biology, XBs are critical in defining the specificity and affinity of halogenated inhibitors against protein targets[13], making them important new tools in medicinal chemistry[14–16]. We present here a proof of concept that XBs can be engineered to affect protein stability, specifically of T4 lysozyme, resulting in the first detailed characterization of XB geometry-energy relationships in a protein. Although the physicochemical basis for a halogen as a covalent substituent, interacting with an electron-rich oxygen, nitrogen, or sulfur

remains debated, a readily accessible explanation for the XB is the σ-hole model[17] (Figure 4.1a). When a halogen is covalently bonded to a carbon atom, its valence electron is pulled into the σ-molecular orbital, resulting in a "σ-hole"—an electropositive crown that serves as the donor in the XB and accounts for the strong directionality of the interaction (Figure 4.1b). Since, the atomic orbitals perpendicular to the σ-hole retain their full complement of electrons, the halogen is amphoteric[16,18], with the halogen serving as electropositive XB donor along the σ-hole and as an electronegative HB acceptor in the perpendicular direction[19,20].

The typical XB donor in protein systems is a halogen substituent of an enzyme inhibitor, while the acceptor is most often the peptide carbonyl oxygen of the protein's backbone[13,21–24]. The strength of the XB depends on the size of the σ-hole, which in turn follows the polarizability of the halogen (F << Cl < Br < I, Figure 4.1c). The XB can be further tuned according to the electron-withdrawing ability of the compound that the halogen is covalently bonded to[16,25], with aromatic groups accentuating XBs[26]. XB energies to anionic oxygen acceptors measured in a model DNA system are equal to or greater than that of a competing HB[27–29]. Furthermore, their geometries become more ideal (with shorter distances and more linear alignment to the X—C bond, Θ_{XB}) as their strength increases[29]. The structure-energy relationships of XBs to a neutral carbonyl oxygen acceptor in a protein[13,21–23], however, have not been previously determined.

To study the impact of XBs on protein stability, we started with T4 lysozyme—a classic model system to study the effects of molecular interactions on protein structure and stability in crystals and in solution[30,31]. An iodinated phenylalanine had previously been incorporated into lysozyme to demonstrate the utility of ncAAs to solve the phase problem in protein crystallography; however, this iodine was not positioned to form an XB[32]. For the current study, we have replaced tyrosines (Y) with a chemically modified phenylalanines (^{Z}F) at two different

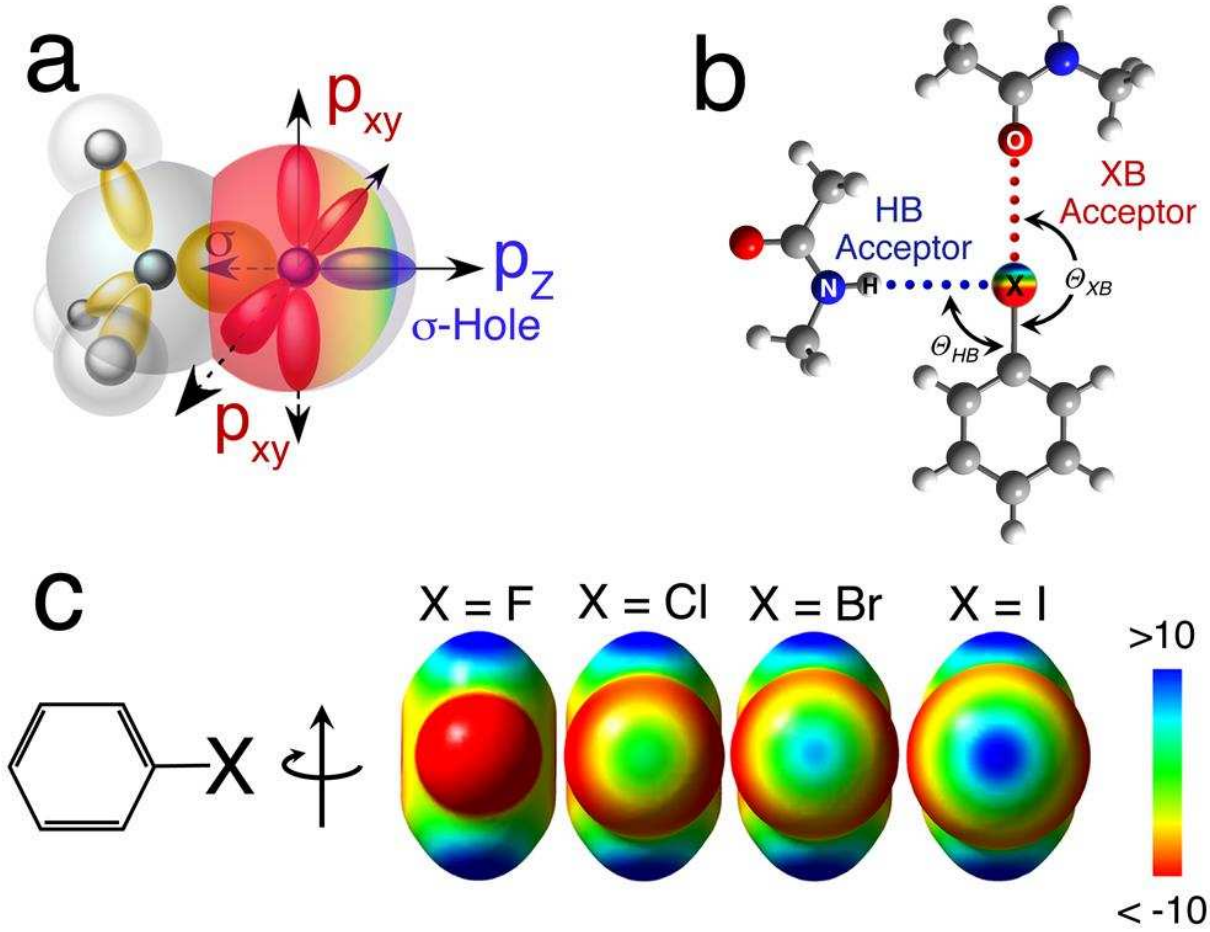

Figure 4.1. The halogen bond (XB). (a) The σ-hole model for X-bonding[17]. The electropositive σ-hole that serves as the XB donor is created as a result of pairing the valence electron from a halogen (typically assigned to the p_z-atomic orbital) with, in this example, an electron from a carbon to form a covalent C—X single bond. The resulting σ-molecular orbital depopulates the outside lobe of the p_z-orbital, leaving an electropositive crown (blue surface) and flattening of the atomic radius—the σ-hole. The $p_{x,y}$ -orbitals remain fully occupied, resulting in an electronegative (red surface) ring perpendicular to the σ-molecular orbital. (b) Relationship between XBs and hydrogen bonds (HBs). The XB (red dotted line) is defined by contact distances that are less than the sum of the standard van der Waals radii for the halogen donor and the acceptor, and an approximate linear approach of the acceptor to the halogen ($\Theta_{XB} \approx 180^{\circ}$). The $p_{x,y}$-orbitals of the halogen can also serve as an acceptor to an HB (blue dotted line), an approach that is approximately orthogonal to the XB ($\Theta_{HB} \approx 90^{\circ}$). (c) Electrostatic potential of halobenzenes. The DFT calculated electrostatic potentials at the 6-311+g(d,p) level (from >10 kcal·mol^{-1}, in blue, to <-10 kcal·mol^{-1}, in red) show the σ-hole increases in size and intensity as the size of the halogen increases from F to Cl to Br to I.

sites: the first position (Y18) has the potential to interact with the protein backbone through HBs and/or XBs, while the second (Y88) serves as a control for the effects of each substituent at a non-interacting site. The non-interacting site mutations are suitable controls because they share many conditions in common with that of the interacting (Y18) mutations. Both mutation locations have cation-pi interactions with a nearby arginine, which was unique between the interacting and non-interacting locations. In addition, both sites have approximately the same solvent accessibility when only comparing their phenyl ring structures, the only difference is in the solvent accessibility of the Z substituent with the Y88 mutations being more exposed. Finally, mutations at the non-interacting site allow us to assay if there are any effects of theses mutations on the denatured form of the protein as the Y88 mutations are exposed in both the native and denatured structures. Comparing the structure and energy effects separately at the two positions reveals how the engineered XB interactions affect the structure and stability of the protein.

II. MATERIALS AND METHODS

Site directed mutagenesis and protein expression

All T4 lysozyme constructs started with the gene of the pseudo wild type (WT*) protein,[33] with the DNA sequence coding for a 6 His-tag appended at the C-terminus to facilitate protein purification. The modified DNA sequences were inserted into the pBAD vector for expression in DH5α *E. coli*.

Expression vectors for WT* containing canonical amino acids were transformed into BL21 (DE3) pLysS *E. coli*. Transformed cells were grown in 2xYT media with appropriate antibiotic (ampicillin and chloramphenicol) and incubated at 37°C with shaking at 250 rpm until

an $O.D._{600}$ of 0.4-0.6 was reached. The cells were induced with arabinose added directly to the cultures to a final concentration of 0.2%(w/v) and allowed to grow for 3 additional hours. Subsequently, the cells were harvested by centrifugation at 3.9K RCF, the supernatant was decanted, and the bacterial pellets were stored at -80 °C.

WT* constructs that incorporate non-canonical amino acids (ncAAs) were expressed in *E. coli* cells, as previously described[34]. Briefly, for the constructs with halogenated or methylated Phe, the codons for Y18 or Y88 were replaced by an AMBER (TAG) codon. The expression vectors were co-transformed with pBAD (gene) and pDule2-pCNF (containing the orthogonal aminoacyl-tRNA synthetase and tRNA pair) into BH10β *E. coli.* The cells were grown overnight in 5 mL of non-inducing media with appropriate antibiotics. These cultures were used to inoculate 1 L of autoinduction medium without arabinose, and with appropriate antibiotics at 37 °C with shaking at 250 rpm. At an $O.D._{600}$ of 0.8, ncAAs were added to the cultures to a final concentration of 1 mM. Once cells reached an of $O.D._{600}$ of 3.0, arabinose was added to a final concentration of 0.1%(w/v) and cell growth continued for 3 additional hours. The cells were harvested by centrifugation at 3.9K RCF, the supernatant was decanted, and bacterial pellets were stored at -80°C.

Protein purification

Bacterial pellets were suspended in 35 mL of binding buffer (40 mM potassium phosphate pH 7.4, 500 mM sodium chloride, 50 mM imidazole, and 0.02%(w/v) sodium azide), thawed in a 37°C water bath for 5 minutes, and all subsequent steps were performed at 4°C. Thawed pellets were sonicated for 1.5 minutes, and then centrifuged for 30 minutes at 35.3K RCF. The supernatant was decanted and directly loaded on to a HisTrap HP column on an

ÄKTA FPLC, after which His-tagged protein was eluted with an increasing imidazole gradient with elution buffer (40 mM potassium phosphate pH 7.4, 500 mM sodium chloride, 500 mM imidazole, and 0.02%(w/v) sodium azide). Selected fractions were concentrated in an Amicon Ultra centrifugal concentrator (NMWL, 10Kda) to 1 mL, then loaded onto a gravity fed Sephadex G-50 fine column equilibrated in buffer specific for crystallization or differential scanning calorimetry (DSC) experiments (crystallization buffer: 500 mM sodium chloride pH 7.4, 50 mM sodium phosphate, and 0.02%(w/v) sodium azide; DSC buffer[31]: 20 mM glycine-HCl pH 3.5, 80 mM NaCl, and 1 mM EDTA). Selected fractions were combined and used for crystallization or DSC experiments.

Protein crystallization

Combined fractions in the crystallization buffer were concentrated to 18-20 $mg \cdot mL^{-1}$. T4 lysozyme crystals were grown at 18 °C using the hanging drop vapor diffusion method with a 1:1 ratio of protein (18-20 mg/mL) to precipitant solution (2.0-2.4 M potassium phosphate, pH 6.6-7.5, 50 mM 2-hydroxyethyldisulfide, 50 mM 2-mercaptoethanol), as previously described[35,36]. Diffraction quality crystals grew in 1-7 days. Crystals were harvested using a cryo-loop, flash frozen, and stored in liquid nitrogen until X-ray data collection.

X-ray data collection and structure determination

X-ray diffraction data were collected on crystals held in a cryogenic nitrogen stream (100K), on the home source (Rigaku copper anode X-ray generator, 1.54Å, Dectris Pilatus 200K detector) or the Advanced Light Source (ALS) Beamline 4.2.2 at Berkeley National Laboratory (1.00Å, Research Detectors Inc. complementary metal-oxide-semiconductor 8M detector).

Diffraction data from the home source were reduced using Denzo/Scalepack,[37] or for data from the ALS beamline using d*TREK and CCP4 suite[38]. X-ray data were phased by molecular replacement, applying the atomic coordinates of WT* (PDB code 1L63)[39] as the starting model, yielding initial models with R_{work} that ranged from 31.8% to 35.5% and R_{free} that ranged from 31.0% to 36.1%. The PHENIX suite of crystallographic software[40] was used for subsequent refinement, which resulted in final structures with R_{work} that ranged from 16.0% to 20.3% and R_{free} that ranged from 19.6% to 24.7%.

Differential scanning calorimetry

Combined fractions of T4 lysozyme, after the gel filtration purification, were diluted to a concentration of 0.1 mg·mL^{-1} using the DSC buffer, and stored at -80°C. A low pH was used to help promote reversible folding[41]. Melting curves were collected on a TA Instruments Nano DSC under constant pressure (3.0 atm) with all samples matched against identical buffers in the reference cell. Samples were equilibrated for 600 seconds, followed by melting data collected through heating cycles from 10 °C to 90 °C (at scan rates of 1 °C·min^{-1}). Reversibility was confirmed for all constructs by performing a cooling scan from 90 °C to 10 °C (at scan rates of 0.5 °C·min^{-1}) and a subsequent heating cycle. A minimum of 5 replicate experiments were conducted for each mutant. Melting data were analyzed using NanoAnalyze Data Analysis, version 3.5.0 software from TA Instruments to extract the melting temperatures (T_m), and enthalpies ($\Delta H°_{Tm}$).

Quantum mechanical (QM) calculations

The atomic coordinates for the interacting residues (11, 18, 30) were taken directly from the refined structures of each construct. Residues 11 and 30 were reduced to N-methylacetamides and residue 18 was reduced to Z-benzene to decrease the computational time (Figure 4.2). QM energies were calculated using Gaussian 09e[42], with the Møller−Plesset second-order (MP2) calculations in cyclohexane as the solvent (D = 2, relative to a vacuum). Geometry optimization of the hydrogen atoms were performed with the Hartree-Fock method prior to the energy calculations. Polarizable basis sets including dispersion were applied to the calculations (aug-cc-PVTZ for W.T.*, Y18F, Y18^{m}F, Y18^{br}F, and aug-cc-PVTZ-PP[43] from EMSL Basis Set Exchange for Y18^{i}F). Basis set superposition errors (BSSE)[44,45] were determined from a separate counterpoise gas phase calculation and directly summed into the calculated solvent phase energy.

Turbidity assay

The activity of T4 lysozyme was monitored via a standard cell clearing assay[46,47]. *M. lysodeikitcus* bacteria was grown in 2xYT media overnight and then diluted in 50 mM sodium phosphate buffer until an O.D.$_{450}$ of 1.0 was reached. Purified T4 lysozyme was added to the solution at room temperature to reach a final concentration of 0.1 mg·mL^{-1} and the absorbance change over time was measured.

III. RESULTS

The constructs to engineer XBs into a model protein are based on the modified WT* form of T4 lysozyme, which has its two disulfide forming Cys residues replaced and, thus, follows a classical two-state reversible folding/unfolding pathway[31]. To identify positions where

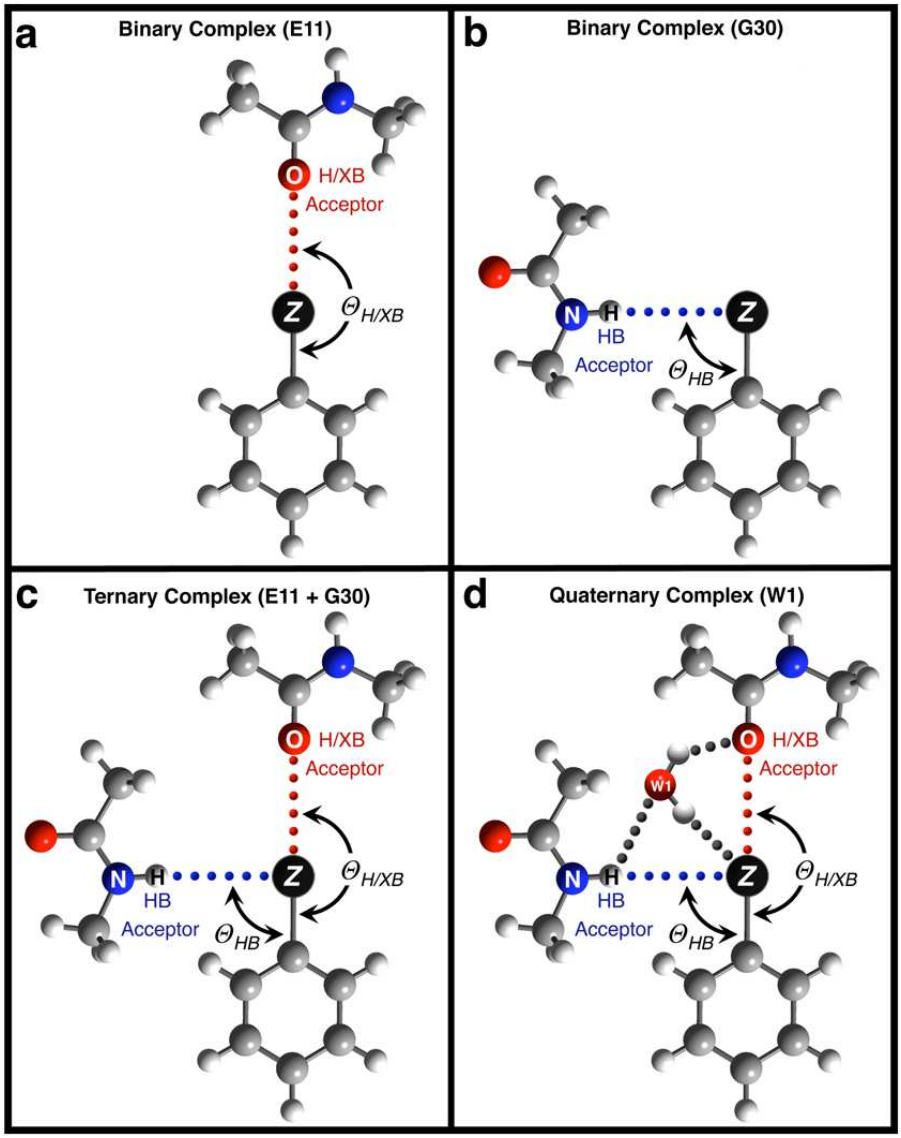

Figure 4.2. Schematic of the molecular models used in the quantum mechanical calculations of HB/XB interactions at the Y18 site of T4 lysozyme constructs. The modified side chain at the Y18 position was modeled as a benzene molecule with a single substituent (*Z* = H, OH, CH_3, Br, or I), while *N*-methylacetamides were used to model the HB donor from Gly30 and HB/XB acceptor of Glu11 (atomic coordinates were taken directly from the refined crystal structures of the interacting residues (11, 18, 30) of each Y18 construct). QM energies were calculated using Gaussian 09e[42], with the Møller−Plesset second-order (MP2) calculations performed in cyclohexane. The positions of the hydrogen atoms were optimized prior to the energy calculations as their positions are dependent on Z. Interaction energies were calculated for the binary complexes of two residues (a) ([Glu11(O) + Y18(C–Z)] and (b) [Gly30(N–H) + Y18(C–Z)]); for the ternary complex of three residues (c) ([Glu11(O) + Gly30(N–H) + Y18(C–Z)]); or for the quaternary complex with water W1 (d) ([Glu11(O) + Gly30(N–H) + Y18(C–Z) + (W1)]). The dashed lines in each panel indicate the interactions for which energies were calculated.

an optimal XB could be introduced into this system, we started by identifying all aromatic amino acids with side chains that are in close proximity to a peptide carbonyl oxygen. Tyrosine 18 (Y18) was selected as the site to engineer XB interactions; the hydroxyl group of Y18 serves as an HB acceptor to a backbone amino group at Gly30 and is close, but does not form a direct HB to the peptide oxygen of Glu11 (Figure 4.3a). Constructs of Y18 were engineered as Y18^{Z}F, where ^{Z}F is a phenylalanine (F) residue with a *Z*-substituent—hydrogen (F), bromine (^{br}F), iodine (^{i}F), or a methyl group (^{m}F). We expected both of the halogenated ^{Z}F analogues to form stabilizing XBs to the Glu11 oxygen. The amphoteric nature of halogens predicts that the Y18^{br}F and Y18^{i}F constructs would also maintain the orthogonal HBs to Gly30[48]. The Y18F mutant was designed as a control with no HB or XB capabilities, while the methylated Y18^{m}F was designed to mimic the size and hydrophobic properties of the halogens.

Constructs at Y88 (Y88^{Z}F) serve as controls to determine the nonspecific effects of each substituent on the structure and stability of the enzyme. All of the Y88^{Z}F constructs place the potential interacting (Z) substituent exposed to solvent and thus unable to form direct intramolecular interactions within the protein (Figure 4.3a). The effects of the engineered XB on the protein structure were characterized by comparing the single-crystal structures of the Y18^{Z}F to Y88^{Z}F constructs, while the effects on stability were determined by comparing the thermodynamics of melting by differential scanning calorimetry (DSC).

Single-Crystal Structures

The crystals from all of the Y18^{Z}F to Y88^{Z}F constructs were isomorphous and diffracted to sufficiently high resolution to provide highly accurate geometries (Table 4.1 – 4.3). The single crystal structures of Y18F and Y18^{m}F place the phenyl rings in near identical geometries relative

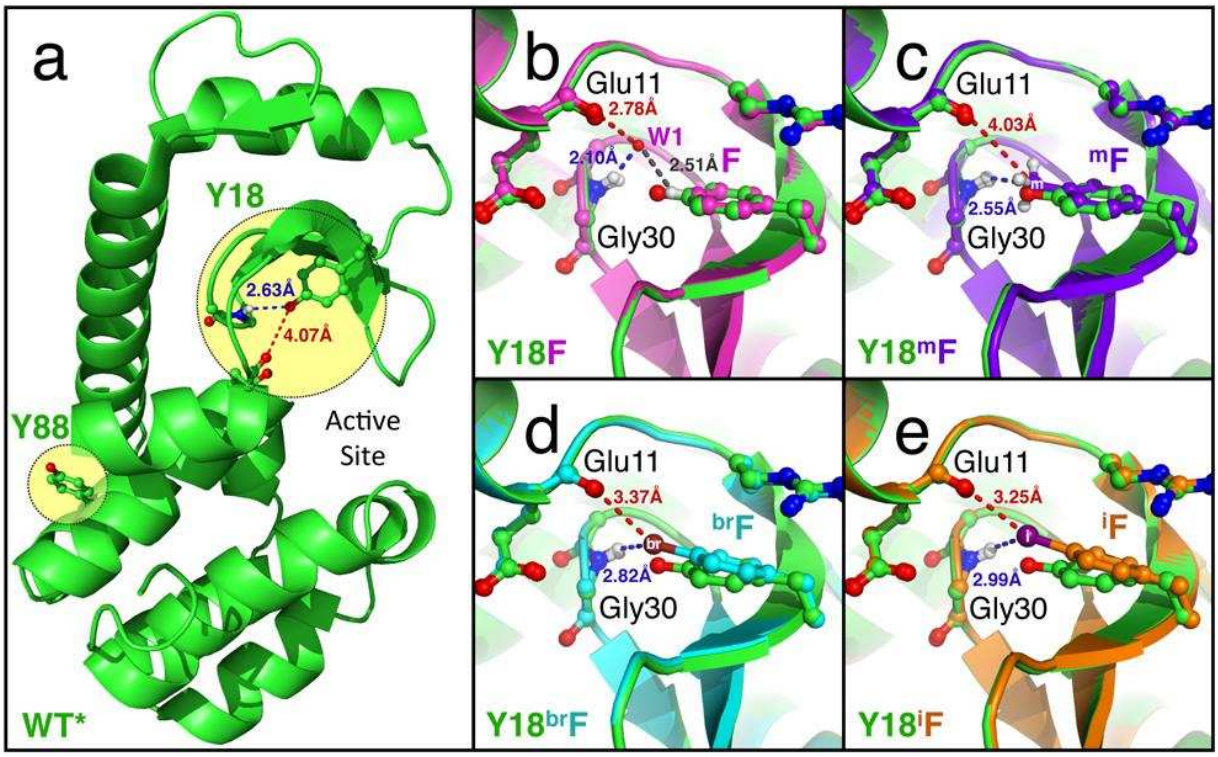

Figure 4.3. Structure of halogenated and non-halogenated T4 lysozyme constructs. (a) Cartoon representation of T4 lysozyme structure. The aromatic amino acid residues selected for modifications in the current study (tyrosine 18 and 88) are circled and labeled (Y18 and Y88, respectively). The side chain at Y18 makes specific intramolecular interactions with the carboxyl oxygen of Glu11 (red dash line) and the amine of Gly30 (blue dash line). (b-e) Crystal structures of halogenated and non-halogenated T4 lysozyme constructs at Y18. In each panel, the structures of the modified ^{Z}F residues at 18 are superimposed on the structure of WT* (carbons in green). The modified constructs are colored as (b) Y18F magenta, (c) Y18^{m}F purple, (d) Y18^{br}F cyan, (e) Y18^{i}F orange. The blue dashes indicate close distances ($R_{N-H\cdots Z}$) of each modified substituent (Z) to the HB donor of Gly30, while the red dashes indicate close distances ($R_{Z\cdots O}$) to the potential HB or XB acceptor oxygen of Glu11.

Table 4.1. Crystallographic parameters for non-halogenated T4 lysozyme constructs. All constructs crystallized in the space group $P3_221$ (angles for the unit cells: $\alpha = \beta = 90.0°$; $\gamma = 120.0°$). All data were collected on the Rigaku home source, as described in the Materials and Methods section.

Parameter	**WT***	**Y18F**	**Y18^{m}F**
		Crystal	
Unit Cell Lengths	$a = b$ = 60.248 Å; c = 96.454 Å	$a = b$ = 59.763 Å; c = 95.119 Å	$a = b$ = 60.197 Å; c = 95.969 Å
		Data Collection	
Resolution (Å)[1]	21.98 – 1.49 (1.54 – 1.49)	29.88 – 1.46 (1.51 – 1.46)	18.76 – 1.50 (1.55 – 1.50)
#Total Reflections	298,362	283,211	310,833
#Unique Reflections[1]	31,370 (3,322)	34,257 (3,097)	31,673 (2,670)
Multiplicity	9.5	8.3	9.8
Completeness[1]	92% (99%)	98% (85%)	96% (67%)
Mean I/σ(I)[1]	28.4 (2.9)	31.0 (2.1)	23.1 (1.8)
R_{merge}[1]	0.085 (0.592)	0.065 (0.594)	0.115 (0.757)
R_{meas}[1]	0.089 (0.623)	0.069 (0.727)	0.121 (0.825)
		Structure Refinement	
Molecular Replacement: Initial Model Statistics			
R_{work}	0.3418	0.3536	0.3384
R_{free}	0.3479	0.3438	0.3312
Final Model			
PDB Code	5KHZ	5KI1	5KI2
A.A. Residues	162	162	162
Non-Solvent atoms	1,374	1,313	1,383
Solvent atoms	383	325	367
R_{work}[1]	0.1755 (0.2164)	0.1800 (0.2919)	0.1879 (0.2760)
R_{free}[1]	0.1962 (0.2537)	0.2067 (0.343)	0.2216 (0.3188)

[1]Values in parentheses are for the highest resolution shell.

Table 4.2. Crystallographic parameters for halogenated T4 lysozyme construct and one control construct. All constructs crystallized in the space group $P3_221$ (angles for the unit cells: $\alpha = \beta = 90.0°$; $\gamma = 120.0°$).

Parameter	**Y18^{br}F**	**Y18^{i}F**	**Y88F**
		Crystal	
Unit Cell Lengths	$a = b$ = 59.589 Å; c = 95.150 Å	$a = b$ = 60.103 Å; c = 96.080 Å	$a = b$ = 60.368Å; c = 96.571Å
		Data Collection	
Data collect at:	ALS	Rigaku home source	Rigaku home source
Resolution (Å) [1]	34.98 – 1.65 (1.71 – 1.65)	32.03 – 1.63 (1.69 – 1.63)	22.99 – 1.49 (1.55 – 1.49)
#Total Reflections	236,904	218,989	269,948
#Unique Reflections[1]	23,668 (2,084)	25,701 (2,478)	33,480 (3,216)
Multiplicity	5.3	8.5	8.1
Completeness[1]	98% (92%)	99% (100%)	99% (94%)
Mean I/σ(I) [1]	4.6 (1.0)	31.0 (2.5)	42.2 (3.7)
R_{merge}[1]	0.210 (0.759)	0.098 (0.792)	0.061 (0.339)
R_{meas}[1]	0.233 (0.913)	0.103 (0.865)	0.064 (0.470)
		Structure Refinement	
Molecular Replacement: Initial Model Statistics			
R_{work}	0.3344	0.3396	0.3374
R_{free}	0.3379	0.3494	0.3272
Final Model			
PDB Code	5KI3	5KIO	5KIG
A.A. Residues	162	162	162
Non-Solvent atoms	1,365	1,367	1,408
Solvent atoms	248	264	343
R_{work}[1]	0.2028 (0.3547)	0.1858 (0.2232)	0.1628 (0.2336)
R_{free}[1]	0.2473 (0.3744)	0.2165 (0.2959)	0.1958 (0.2857)

[1]Values in parentheses are for the highest resolution shell.

Table 4.3. Crystallographic parameters for control T4 lysozyme constructs (Res. 88). All constructs crystallized in the space group $P3_221$ (angles for the unit cells: $\alpha = \beta = 90.0°$; $\gamma = 120.0°$).

Parameter	**Y88^{m}F**	**Y88^{br}F**	**Y88^{i}F**
Crystal			
Unit Cell Lengths	$a = b$ = 60.183 Å; c = 95.834 Å	$a = b$ = 60.013 Å; c = 95.945 Å	$a = b$ = 60.283 Å; c = 96.197 Å
Data Collection			
Data collect at:	Rigaku home source	ALS	Rigaku home source
Resolution (Å) [1]	45.79 – 1.56 (1.62 – 1.56)	45.70 – 1.55 (1.61 – 1.55)	28.76 – 1.50 (1.55 – 1.50)
#Total Reflections	361,463	163,055	431,722
#Unique Reflections[1]	29,011 (2836)	28,167 (2,068)	33,083 (3,250)
Multiplicity	12.4	1.9	13.0
Completeness[1]	99% (98%)	95% (71%)	100% (99%)
Mean I/σ(I) [1]	43.6 (2.4)	19.1 (6.3)	70.4 (6.5)
R_{merge}[1]	0.065 (0.783)	0.026 (0.310)	0.058 (0.341)
R_{meas}[1]	0.068 (0.870)	0.037 (0.438)	0.060 (0.376)
Structure Refinement			
Molecular Replacement: Initial Model Statistics			
R_{work}	0.3430	0.3178	0.3546
R_{free}	0.3290	0.3095	0.3613
Final Model			
PDB Code	5KII	5KI8	5KIM
A.A. Residues	162	162	162
Non-Solvent atoms	1,393	1,366	1,436
Solvent atoms	342	205	368
R_{work}[1]	0.1999 (0.2893)	0.1757 (0.2757)	0.1604 (0.2117)
R_{free}[1]	0.2244 (0.3198)	0.2101 (0.3147)	0.1844 (0.2326)

[1]Values in parentheses are for the highest resolution shell.

to WT* (Figure 4.3b and c), suggesting that the local loop structure is not perturbed by interactions to Gly30 or lack of interaction to Glu11. In contrast, the orientation of the aromatic rings in both the Y18^{br}F and Y18^{i}F constructs are significantly perturbed, showing the halogens pulled towards the Glu11 oxygen (Figure 4.3d – e, 4.4, and 4.5). This ring displacement was notably absent from any of the control Y88^{Z}F constructs (Figure 4.6). In Y18^{br}F, the bromine is within van der Waals distance of the Glu11 oxygen (~100% of the sum of the van der Waals radii, $\sum R_{vdW}$, Table 4.4), indicating the formation of an XB. This cannot be a steric effect, since the Y18^{m}F remains unperturbed. The distortion is even greater with Y18^{i}F, where the I···O distance is within the optimum distance of ~93% of $\sum R_{vdW}$ for XBs in proteins[16]. The angle of approach of the Glu11 oxygen towards the Br—C and I—C bonds (Θ_{E11} = 142.7° and 150.0°, respectively) places it within the electropositive σ-hole for the respective halogens[16]. Thus, the geometries indicate that XBs were formed in both the Y18^{br}F and Y18^{i}F constructs, with the iodine being more ideal and thus potentially a stronger interaction.

In addition to the XBs, the Y18^{br}F and Y18^{i}F constructs maintain the HB to Gly30 seen in WT*, but the angles of approach of the Gly30 N—H to each halogen (Θ_{G30} = 130.3° and 125.9° for Br and I, respectively) are significantly far from the optimum[29], suggesting weaker HBs compared to that in WT*.

Quantum mechanical (MP2) analyses on simple models for the molecular interactions to Glu11 and Gly30 (Table 4.5, Figure 4.2) predict Y18^{i}F to be the most stable ternary construct ($E_{MP2\ (ternary)}$). The XB from the iodine to the Glu11 oxygen was the strongest interaction of any construct, as expected from the short O···I distance. However, as predicted from the N–H···I angle, the iodine is not as strong an HB acceptor as the Tyr OH in WT* (Table 4.6). Indeed,

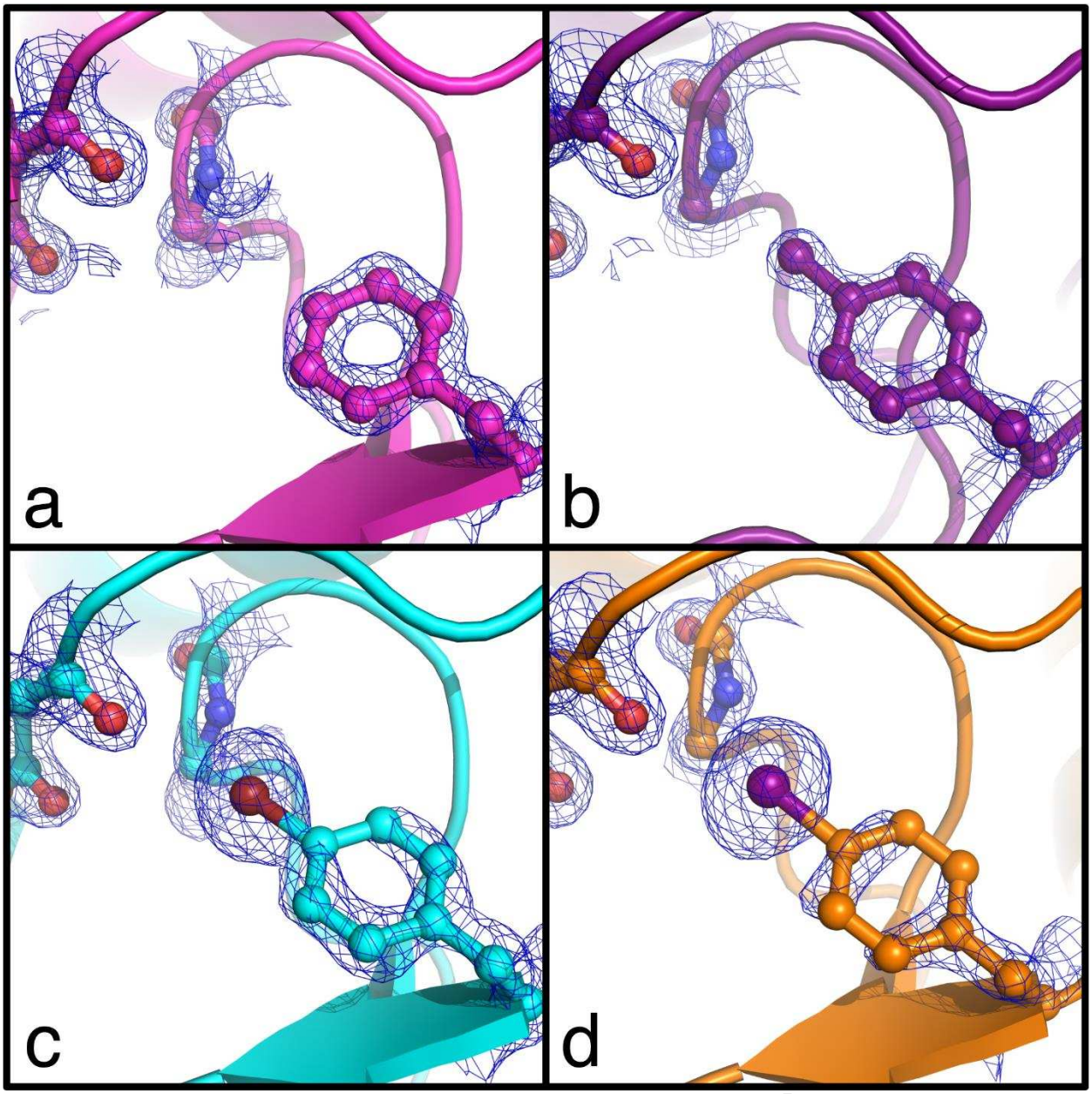

Figure 4.4. Omit electron density maps from crystal structures of Y18^{Z}F T4 lysozyme constructs. 2Fo-Fc electron densities are rendered at the 2σ level of contours with Y18F (a) in magenta, Y18^{m}F (b) in purple, Y18^{br}F (c) in cyan, and Y18^{i}F (d) in orange.

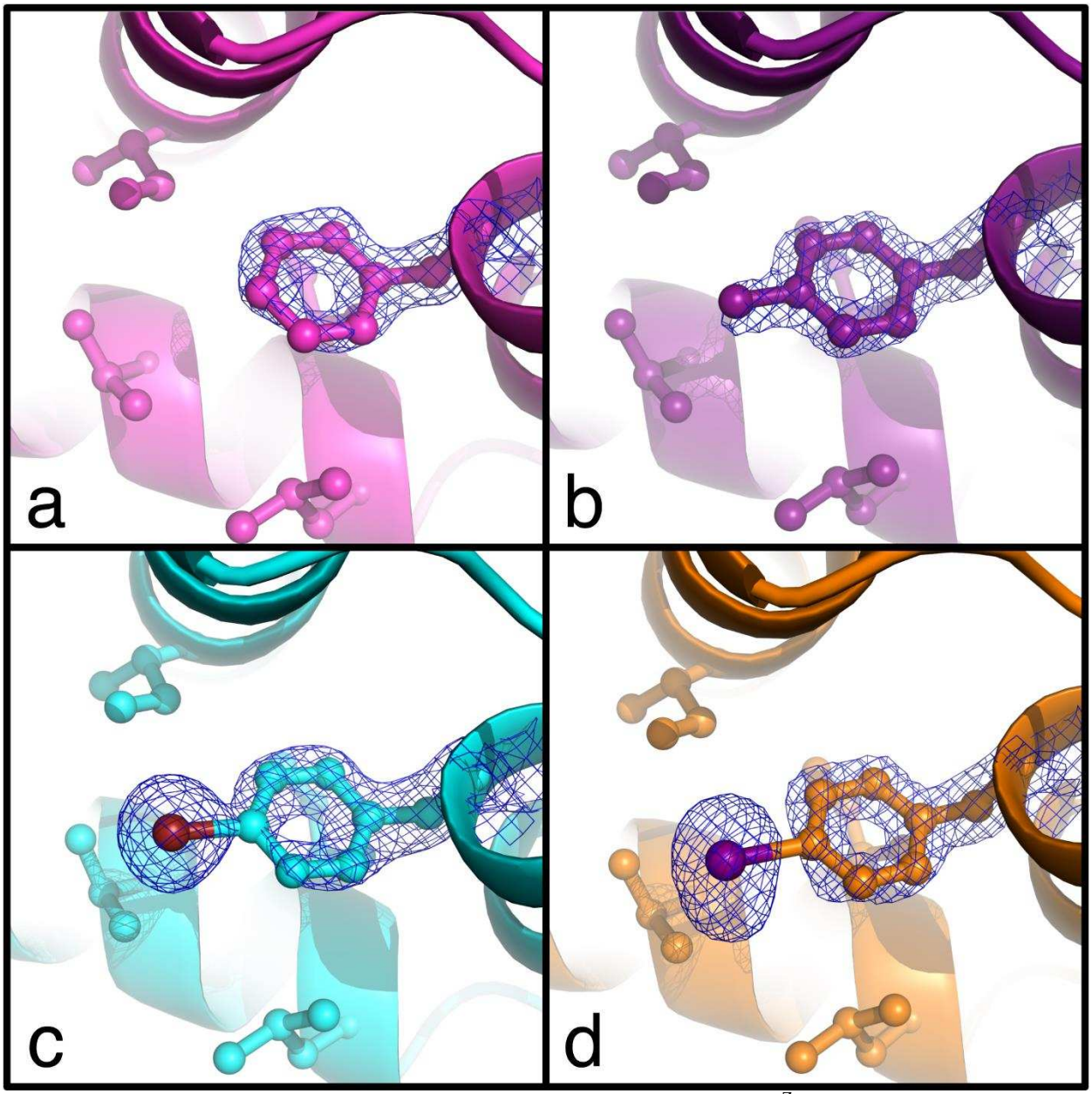

Figure 4.5. Omit electron density maps from crystal structures of Y88^{Z}F T4 lysozyme constructs. 2Fo-Fc electron densities are rendered at the 2σ level of contours with Y88F (a) in magenta, Y88^{m}F (b) in purple, Y88^{br}F (c) in cyan, and Y88^{i}F (d) in orange.

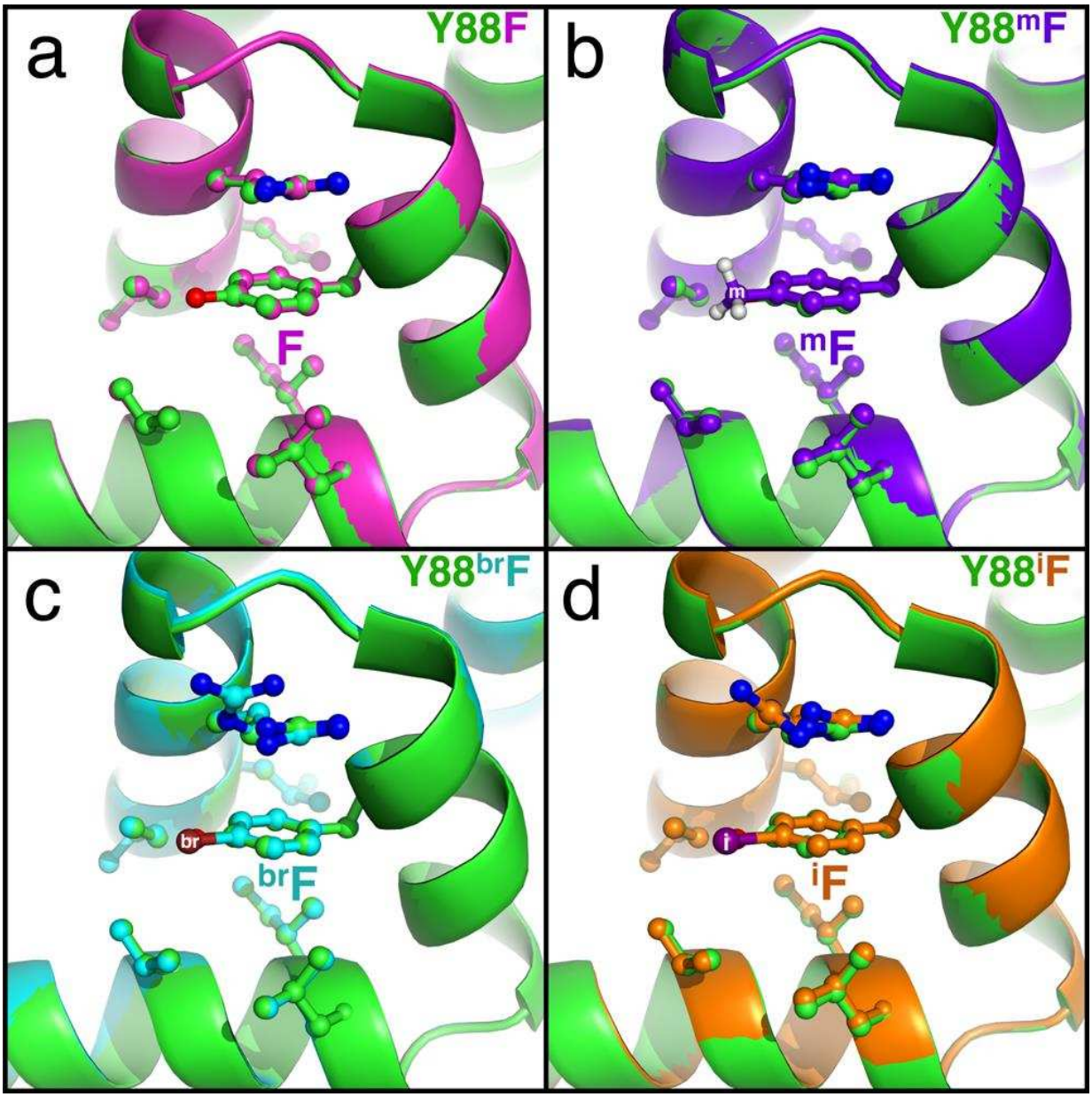

Figure 4.6. Details of the crystal structures of halogenated and non-halogenated T4 lysozyme constructs at Y88 position. In each panel, the structures of the modified residues are superimposed on the structure of the WT* enzyme (carbons and backbone trace in green). The carbons and backbone traces of the modified constructs are colored as (a) Y88F magenta, (b) Y88^{m}F purple, (c) Y88^{br}F cyan, (d) Y88^{i}F orange.

Table 4.4. Interaction geometries for T4 lysozyme constructs. The distances from the modified substituent (*Z*) to the potential H/XB acceptor oxygen atom of Glu11 ($R_{O\cdots Z}$) or the HB donor of Gly30 ($R_{N-H\cdots Z}$) identify HB or XB interactions, in terms of the percent of the sum of the standard van der Waals distances of the interacting atoms ($\%\Sigma R_{vdW}$). Approach of *Z* to the potential H/XB acceptor oxygen atom of Glu11 (Θ_{E11}) or to the HB donor of Gly30 (Θ_{G30}) determines whether the interaction is linear.

		Glu11$_{(O)}$			*Gly30*$_{(N)}$		
Construct	**Substituent (*Z*)**	$R_{O\cdots Z}$ **(Å)**	$\%\Sigma R_{vdW}$ **(%)**	Θ_{E11} **(°)**	$R_{N-H\cdots Z}$ **(Å)**	ΣR_{vdW} **(%)**	Θ_{G30} **(°)**
WT*	OH	4.07Å	133.0%	126.4°	2.63Å	96.7%	147.1°
Y18F	H	5.38Å	175.2%	142.1°	3.21Å	108.1%	155.0°
Y18^{m}F	CH_3	4.03Å	148.2%	133.4°	2.55Å	106.3%	156.0°
Y18^{br}F	Br	3.37Å	100.0%	142.7°	2.82Å	92.5%	130.3°
Y18^{i}F	I	3.25Å	92.9%	150.0°	2.99Å	94.0%	125.9°

Table 4.5. Experimental and calculated interaction energies for T4 lysozyme constructs. The experimental melting enthalpies ($\Delta H°_M$) and melting temperatures (T_M) are measured by DSC for each Y18 and Y88 construct. The differences in $\Delta H°_M$ ($\Delta\Delta H°_{M\ (18\text{-}88)}$) and T_M ($\Delta T_{M\ (18\text{-}88)}$) between the Y18 and Y88 constructs reflect the energies associated with specific interactions at the Y18 site relative to a non-interacting substitution at Y88. Errors are the standard deviations of mean for each measurement. The H/XB interaction energies for the ternary complex of the interacting residues of ^{Z}F18 to Glu11 and Gly30 ($E_{MP2\ (ternary)}$) were calculated by the Møller-Plesset 2 (MP2) method, applying the aug-cc-PVTZ basis set, with atomic coordinates for individual residues taken from the single crystal structures, in the absence of waters. The quaternary complex MP2 interaction energies of W1 (atomic coordinates taken from the single crystal structures seen in Figure 4.7) with residues Y18^{Z}F, Glu11, and Gly30 ($E_{MP2\ (quaternary)}$) shows how the position of this water contributes additionally to the stabilization of each Y18 construct ($E_{MP2\ (quaternary)}$) was not calculated for Y18^{br}F construct, since W1 was not observed in this structure). See Figure 4.2 for more detail.

Construct	**Substituent** ***(Z)***	$\Delta H°_M$ **(kcal·mol^{-1})**	$\Delta\Delta H°_{M\ (18\text{-}88)}$ **(kcal·mol^{-1})**	T_M **(°C)**	$\Delta T_{M\ (18\text{-}88)}$ **(°C)**	$E_{MP2\ (ternary)}$ **(kcal·mol^{-1})**	E_{MP2} (quaternary) **(kcal·mol^{-1})**
			Nonhalogenated T4 Lysozyme Constructs				
WT*$_{18}$	OH	130 ± 1	-	57.56 ± 0.02°C	-	-3.4	-12.5
WT*$_{88}$	OH	130 ± 1	-	57.56 ± 0.02°C	-	-	-
Y18F	H	122.6 ± 0.4	7 ± 1	56.65 ± 0.04°C	-0.12 ± 0.05°C	-0.3	-11.2
Y88F	H	115 ± 1	-	56.77 ± 0.04°C	-	-	-
Y18^{m}F	CH_3	118.7 ± 0.5	2 ± 1	55.32 ± 0.03°C	-1.12 ± 0.05°C	-1.4	-7.6
Y88^{m}F	CH_3	116.4 ± 0.7	-	56.53 ± 0.04°C	-	-	-
			Halogenated T4 Lysozyme Constructs				
Y18^{br}F	Br	115 ± 1	1 ± 1	55.21 ± 0.02°C	0.27 ± 0.07°C	-2.4	-
Y88^{br}F	Br	114 ± 1	-	54.94 ± 0.06°C	-	-	-
Y18^{i}F	I	119 ± 1	6 ± 1	56.21 ± 0.07°C	0.79 ± 0.08°C	-3.5	-7.5
Y88^{i}F	I	113 ± 1	-	55.43 ± 0.03°C	-	-	-

Table 4.6. Quantum mechanical energies of the binary and ternary complexes. MP2 level quantum mechanical energies were calculated for noncovalent interactions from the *Z*-substituent of $Y18^{Z}F$ constructs (where *Z* = H, OH, CH_3, Br, or I) to the HB/XB acceptor of Glu11 ($E_{MP2\ (binary,\ E11)}$), to the HB donor of Gly30 ($E_{MP2\ (binary,\ G30)}$), or to both ($E_{MP2\ (ternary)}$). Quantum mechanical energies were calculated by Møller-Plesset 2 method, applying the aug-cc-PVTZ basis set, in cyclohexane with BSSE corrections. Refer to Figure 4.2 for a reference of the schematic showing the geometries of the interacting residues. We note that the individual interaction energies ($E_{MP2\ (binary,\ E11)} + E_{MP2\ (binary,\ G30)}$) sums to approximately that calculated with both interaction groups in the model, indicating that the ternary model is fairly accurate, and can be segregated into its individual components.

Construct	**Substituent (*Z*)**	$E_{MP2\ (binary,\ E11)}$ **(kcal·mol^{-1})**	$E_{MP2\ (binary,\ G30)}$ **(kcal·mol^{-1})**	$E_{MP2\ (ternary)}$ **(kcal·mol^{-1})**
WT*	OH	-1.0	-2.7	-3.4
Y18F	H	-0.2	-0.1	-0.3
$Y18^{m}F$	CH_3	-0.6	-0.8	-1.4
$Y18^{br}F$	Br	-0.6	-2.4	-2.4
$Y18^{i}F$	I	-1.6	-1.9	-3.5

WT* was predicted to have an overall energy for direct interactions that is nearly identical to that of Y18^{i}F. Similarly, the weak XB in Y18^{br}F (a consequence of the less positive σ-hole and less ideal geometry of the bromine) is compensated by a stronger N–H···Br interaction. Not surprisingly, the Y18^{m}F construct shows only weak interactions to the amine group at this site.

An MP2 analysis of Y18F indicates that there are essentially no stabilizing interactions from the Phe to either Glu11 or Gly30. Overall, the MP2 analyses predict an order for the stability of the Y18 constructs to be Y18^{i}F > WT* > Y18^{br}F > Y18^{m}F > Y18F. However, a water (W1, Figure 4.3b) fills in the cavity in the Y18F, bridging Glu11, and Gly30 through HBs. We see that the positioning of W1 is relatively conserved for all of the Y18 constructs, except for Y18^{br}F where it is absent and Y18F where it moves significantly compared to the other Y18 constructs (Figure 4.7). Taking W1 into account, the MP2 energies ($E_{MP2\ (quaternary)}$) show that the Y18F construct becomes very stabilizing, much more so than the other Y18 halogenated or methylated constructs (Table 4.5, 4.7). Thus, the MP2 analysis including W1 predicts an order for the stability of WT* > Y18F > Y18^{i}F ≈ Y18^{m}F > Y18^{br}FF.

Thermal Melting Studies to Assess Protein Stability

The DSC determined melting temperatures (T_M) and melting enthalpies ($\Delta H°_M$) for each construct (Table 4.5) showed that any substitution made either at the Y18 or Y88 positions is destabilizing relative to WT*, reinforcing the general understanding that it is very difficult to engineer a more stable T4 lysozyme[49]. Indeed, the order of stability, as reflected by the T_Ms, is WT* > Y18F > Y18^{i}F > Y18^{m}F > Y18^{br}F, which initially did not follow our predictions from the MP2 analysis until W1 was added to the calculations. Analyses of the control Y88^{Z}F constructs

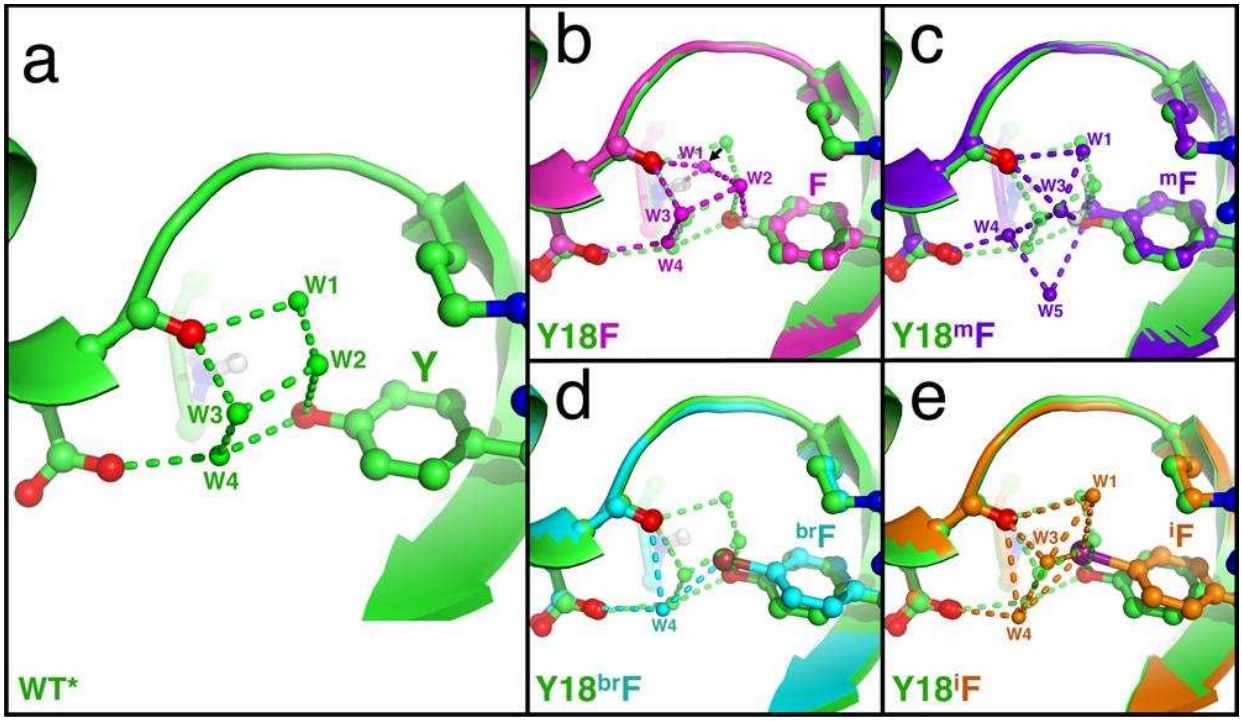

Figure 4.7. Structure of water molecules in the halogenated and non-halogenated T4 lysozyme constructs. (a-e) Details of the crystal structures of the halogenated and non-halogenated T4 lysozyme constructs around the Y18 residue. In each panel, the structures of the modified residues 18 are superimposed on the structure of WT* (a, in green). The modified constructs are colored as (b) Y18F magenta, (c) Y18^{m}F purple, (d)Y18^{br}F cyan, (e) Y18^{i}F orange. The non-bonded spheres are water molecules, which are colored with their associated structure. The waters in each modified construct are aligned and labeled relative to the closest corresponding water in WT*. The arrow in (b) shows the shift in the position of W1 in order to accommodate the loss of the Tyr hydroxyl group in the Y18F construct.

Table 4.7. Quantum mechanical energies for HB interactions of water W1 to the oxygen (O) of Glu11, amide group (N–H) of Gly30, and/or the ring hydrogen (C–H) of Phe18 in the Y18F construct. Interaction energies (E_{MP2}) were calculated by Møller-Plesset 2 method with the aug-cc-PVTZ basis set in cyclohexane, with BSSE corrections applied. E_{MP2} values were calculated for W1 paired with the individual amino acids (atomic coordinates taken from the refined crystal structures), with pairs of amino acids ([Glu11(O) + Gly30(N–H)], [Glu11(O) + Phe18(C–H)], or [Gly30(N–H) + Phe18(C–H)]), or with all three amino acids as a quaternary complex ([Glu11(O) + Gly30(N–H) + Phe18(C–H)]. The positions of the hydrogen atoms were optimized prior to the energy calculations. From this analysis, we see that the E_{MP2} values for pairs that are approximately the sums of the energies to the respective individual amino acids, and the quaternary complex the sums of the individual residues, or the pairs plus the individual residues. Thus, the overall energies of interactions within the quaternary complex is accurate, and can be accurately segregated into the individual components.

Interacting Residue (Functional Group)	E_{MP2} **(kcal·mol^{-1})**
Glu11(O)···W1	-5.3
Gly30(N–H)···W1	-4.0
Phe18(C–H)···W1	-0.9
[Glu11(O) + Gly30(N–H)]···W1	-10.0
[Glu11(O) + Phe18(C–H)]···W1	-4.9
[Gly30(N–H) + Phe18(C–H)]···W1	-6.3
[Glu11(O) + Gly30(N–H) + Phe18(C–H)]···W1	-10.9

indicate that all *Z*-substitutions destabilize the protein, with halogenation being the most destabilizing.

Comparing the thermodynamic parameters for the Y18^{Z}F to Y88^{Z}F sites ($\Delta T_{M\,(18\text{-}88)}$ and $\Delta\Delta H°_{M\,(18\text{-}88)}$) allows us to gauge the effects of the XB and HB interactions at the Y18 site on the stability of the protein relative to the general substituent effects. We see that the Y18^{i}F construct has the most significant effect on both the differences in $\Delta T_{M\,(18\text{-}88)}$ and $\Delta\Delta H°_{M\,(18\text{-}88)}$, reflecting the contribution of the iodine XB to the enthalpic stabilization of the protein and the associated increase in the relative melting temperature.

The Y18F shows the largest $\Delta\Delta H°_{M\,(18\text{-}88)}$, which is in agreement to the MP2 analysis of the direct intramolecular interactions at this site when W1 is included. As previously suggested W1 of Y18F shifts to compensate for the lost hydroxyl group in WT* (Figure 4.7b). Indeed, the MP2 analysis shows that W1 becomes energetically more favorable by ~2 kcal·mol^{-1} when shifted from its WT* position to that of the Y18F construct, which may help to account for the increased $\Delta\Delta H°_{M\,(18\text{-}88)}$ for this construct (Table 4.5). This enthalpic stabilization, however, does not translate into an increase in the $\Delta T_{M\,(18\text{-}88)}$, which could be interpreted as the entropic penalty resulting from positioning this water into a structurally fixed position.

Both the Y18^{m}F and Y18^{br}F constructs show small effects on $\Delta\Delta H°_{M\,(18\text{-}88)}$, but the methyl substituent results in significant thermal destabilization, while the bromine slightly increases the stability. Thus, the $\Delta\Delta H°_{M\,(18\text{-}88)}$ values reflect the contributions of each substituent group on the molecular interactions at the Y18 site, as quantified by the MP2 analysis of the crystal structures. The data indicates that the iodine of Y18^{i}F forms an XB that contributes significantly to the stability of the protein in solution, relative to iodination at site 88 that cannot

form an XB. The effect of these interactions on the overall thermal stability of the protein, however, remains a more complex relationship.

IV. CONCLUSIONS AND DISCUSSION

The geometry-energy relationships for XBs are determined for the first time in a protein. The direct correlation of MP2 analysis of crystal structures to DSC measured energies[29] for XBs engineered previously into a model DNA system allowed us to develop a force field that indeed is accurate in modeling XBs to anionic acceptors. The energetic contribution of XBs in the T4 lysozyme now allows us to refine and test this force field[29] to more accurately model XBs in proteins, including direct applications in the design of new halogenated inhibitors against clinically important cellular targets[14-16] by informing the development of new scoring functions for lead discovery[50].

In the current study, we see a similar correlation between the MP2 calculated molecular interactions and the melting enthalpies ($\Delta\Delta H°_{M\,(18\text{-}88)}$) in T4 lysozyme, but in the case where the tyrosine is replaced by a phenylalanine, we need to consider the solvent structure for the relationship to hold. This analysis shows that the hydroxyl group in WT* is too far to form a direct HB to the Glu11 peptide oxygen, yet WT* remains overall more stable than any of the Y18^{Z}F constructs. Again, we must consider the solvent structure. A detailed analysis of all the current crystal structures that maintain the tyrosine at the Y18 position (WT* and all the Y88 constructs) shows four very well defined water molecules help bridge the Y18 hydroxyl to the Glu11 peptide oxygen (Figure 4.7a), one of which also connects Y18 to the side chain of the Glu11 in the catalytic pocket of the enzyme (labeled W4). In all constructs where Y18 remains intact, these bridging waters remain in place. In all cases where Y18 is replaced, W4 remains

intact, while W1 or W2 is either repositioned or is displaced (Figure 4.7b – e). The inference is that W1 and W2 are particularly important in conferring stability to the entire protein. In the case of Y18F, W1 is repositioned to fill the void space left by the loss of the tyrosyl OH group, while in $Y18^{m}F$, $Y18^{br}F$, and $Y18^{i}F$, W2 are displaced. Our MP2 calculations show that W1 positioned at the Y18F position is ~2 kcal·mol^{-1} more favorable than its similar position in WT*, which explains why removing the hydroxyl group was not as detrimental to the protein's stability. Thus, when we consider how to engineer a more stable protein, it is perhaps not surprising that we must pay attention not only to the direct interactions within the protein, but also how they affect the solvent structure.

This same concept that water structure is important applies also when attempting to engineer enzyme function. The hydroxyl group of Y18 sits near the substrate-binding pocket and interacts with the side-chain of the catalytic Glu11 residue through a bridging water (W4 in Figure 4.7). We see that the activities, as monitored by a turbidity assay[46], of the substituted Y18 constructs are all significantly diminished when the OH of the Tyr residue is replaced by a halogen or non-halogen substituent, even though W4 remains intact. The activities of constructs in which Y18 and its constellation of waters remain intact are all at least 40% of the WT* (Figure 4.8). In contrast, the halogenated and methylated constructs, where this cluster of waters is disrupted show diminished activities. The Y18F construct, however, retains significant activity even though it has lost the direct interactions to Glu11, suggesting that indeed W1 helps to compensate for the missing OH.

The additive nature of non-covalent interactions suggests that introducing XBs into the system can result in a more thermally stable protein. Although we did not succeed in creating an overall more thermally stable T4 lysozyme, we have shown that XBs will help rescue effects that

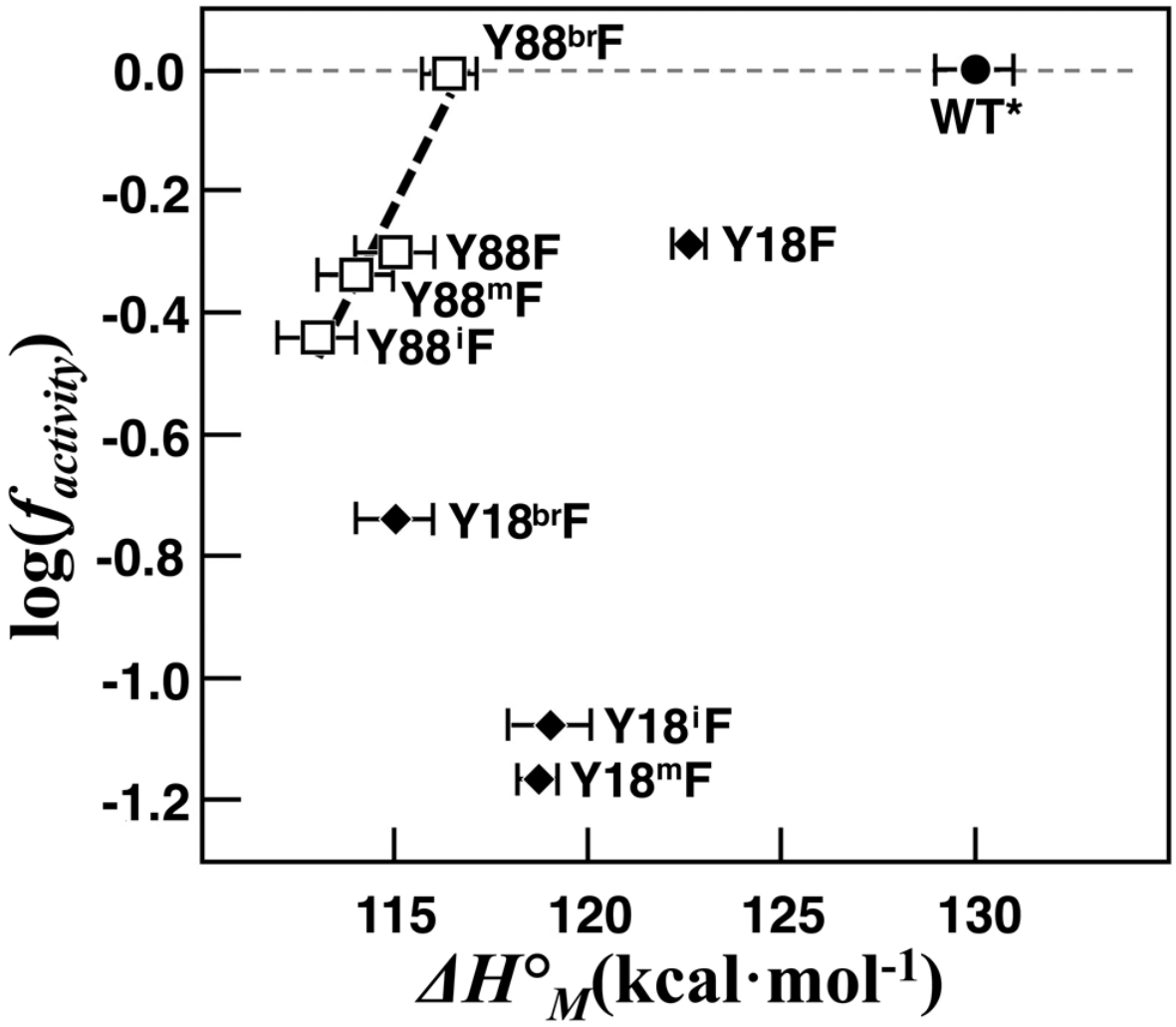

Figure 4.8. Cell clearing activity *vs* melting enthalpy. The base-10 logarithm of the fraction of the enzymatic activities for each construct relative to that of WT* (log($f_{activity}$)) are plotted against their melting enthalpies ($\Delta H°_M$), as determined from the DSC measured melting data (Table 4.5). Diamonds show the T4 lysozyme constructs that have been modified at the interacting Y18 residue, while squares are the control Y88 constructs. WT* is shown as a circle. The log($f_{activity}$) are linearly correlated with $\Delta H°_M$ for the control Y88 constructs (R^2=0.92). Error bars show standard deviations of the mean for the $\Delta H°_M$ measured for each construct.

generally causes instability in a protein. We saw that halogenation at Y88 has a generally destabilizing effect. By moving the halogen from this non-interacting position to one that can form a relatively strong XB (at Y18), we can restore nearly half the loss in thermal stability. The XB at this position, however, was not in an ideal geometry, with the Θ_{XB} = 150° (30° from the optimum linear 180° angle). Thus, there is significant room for improving the geometry and thus increasing the stabilizing potential of the engineered XB in proteins. As ncAAs, including halogenated residues, become more widely applied to engineer proteins with new functions, the XB can provide added stability to an otherwise destabilizing substitution.

The application of XBs in protein engineering can be easily expanded beyond simply affecting protein stability. We can envision that XBs can be introduced at interfaces to engineer new protein-protein interactions, recognition sites, and even XB dependent enzymatic catalysts. Organocatalysts have been designed in which XB donors help to accelerate halide abstraction[51,52], nucleophilic substitution[53], and aza-Diels-Alder reactions[54]. In addition, XBs are thought to facilitate iodination abstraction by iodothyronine deiodinase[55]. In such catalysts, the XB interaction helps to weaken the covalent bond to facilitate extraction of the leaving group. The introduction of amino acids as XB donors can thus provide new catalytic capabilities that take advantage of their tunability and high directionality.

V. RECOGNITIONS

We thank Crystal Vander Zanden for thoughtful discussion throughout this project and Rhianon Kay Rowe for help in editing the manuscript. This work was supported by a grant from the National Science Foundation (CHE-1608146). A.-C.C.C. acknowledges *the* Wenner-Gren

Foundation, Helge Ax:son Johnsons Stiftelse, Richard Dahlboms Stiftelse, and Stiftelsen *Olle Engkvist Byggmästare for funding support*.

REFERENCES

1. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292,** 498–500 (2001).
2. Xie, J. & Schultz, P. G. An expanding genetic code. *Methods* **36,** 227–238 (2005).
3. Zhang, W. H., Otting, G. & Jackson, C. J. Protein engineering with unnatural amino acids. *Curr. Opin. Struct. Biol.* **23,** 581–587 (2013).
4. Malyshev, D. A. *et al.* A semi-synthetic organism with an expanded genetic alphabet. *Nature* **509,** 385–388 (2014).
5. Bissantz, C., Kuhn, B. & Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **53,** 5061–5084 (2010).
6. Hunter, C. A. Quantifying Intermolecular Interactions: Guidelines for the Molecular Recognition Toolbox. *Angew. Chemie Int. Ed.* **43,** 5310–5324 (2004).
7. Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **6,** e19230 (2011).
8. Seeman, N. C. Structural DNA nanotechnology: an overview. *Methods Mol. Biol.* **303,** 143–66 (2005).
9. Tintoré, M., Eritja, R. & Fábrega, C. DNA Nanoarchitectures: Steps towards Biological Applications. *ChemBioChem* **15,** 1374–1390 (2014).
10. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440,** 297–302 (2006).
11. Desiraju, G. R. *et al.* Definition of the halogen bond (IUPAC Recommendations 2013). *Pure Appl. Chem.* **85,** 1711–1713 (2013).

12. Priimagi, A., Cavallo, G., Metrangolo, P. & Resnati, G. The halogen bond in the design of functional supramolecular materials: recent advances. *Acc. Chem. Res.* **46,** 2686–95 (2013).

13. Auffinger, P., Hays, F. A., Westhof, E. & Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 16789–94 (2004).

14. Zhou, P., Tian, F., Zou, J. & Shang, Z. Rediscovery of halogen bonds in protein-ligand complexes. *Mini Rev. Med. Chem.* **10,** 309–314 (2010).

15. Wilcken, R., Zimmermann, M. O., Lange, A., Joerger, A. C. & Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **56,** 1363–1388 (2013).

16. Scholfield, M. R., Vander Zanden, C. M., Carter, M. & Ho, P. S. Halogen bonding (X-bonding): A biological perspective. *Protein Science* **22,** 139–152 (2013).

17. Clark, T., Hennemann, M., Murray, J. S. & Politzer, P. Halogen bonding: the sigma-hole. Proceedings of 'Modeling interactions in biomolecules II', Prague, September 5th-9th, 2005. *J. Mol. Model.* **13,** 291–6 (2007).

18. Nelyubina, Y. V., Antipin, M. Y., Dunin, D. S., Kotov, V. Y. & Lyssenko, K. A. Unexpected 'amphoteric' character of the halogen bond: the charge density study of the co-crystal of N-methylpyrazine iodide with I2. *Chem. Commun.* **46,** 5325 (2010).

19. Brammer, L., Bruton, E. A. & Sherwood, P. Understanding the Behavior of Halogens as Hydrogen Bond Acceptors. *Cryst. Growth Des.* **1,** 277–290 (2001).

20. Aakeröy, C. B. *et al.* Combining halogen bonds and hydrogen bonds in the modular assembly of heteromeric infinite 1-D chains. *Chem. Commun.* 4236 (2007). doi:10.1039/b707458a

21. Lu, Y. *et al.* Halogen bonding--a novel interaction for rational drug design? *J. Med. Chem.* **52,** 2854–62 (2009).

22. Hardegger, L. a *et al.* Systematic Investigation of Halogen Bonding in Protein-Ligand Interactions. *Angew. Chemie Int. Ed.* **50,** 314–318 (2011).

23. Wilcken, R., Zimmermann, M. O., Lange, A., Zahn, S. & Boeckler, F. M. Using halogen bonds to address the protein backbone: a systematic evaluation. *J. Comput. Aided. Mol. Des.* **26,** 935–45 (2012).

24. Sirimulla, S., Bailey, J. B., Vegesna, R. & Narayan, M. Halogen Interactions in Protein–Ligand Complexes: Implications of Halogen Bonding for Rational Drug Design. *J. Chem. Inf. Model.* **53,** 2781–2791 (2013).

25. Riley, K. E. *et al.* 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,** 3309–3318 (2011).

26. Voth, A. R. & Ho, P. S. The Role of Halogen Bonding in Inhibitor Recognition and Binding by Protein Kinases. *Curr. Top. Med. Chem.* **7,** 1336–1348 (2007).

27. Carter, M. & Ho, P. S. Assaying the Energies of Biological Halogen Bonds. *Cryst. Growth Des.* **11,** 5087–5095 (2011).

28. Vander Zanden, C. M., Carter, M. & Ho, P. S. Determining thermodynamic properties of molecular interactions from single crystal studies. *Methods* **64,** 12–8 (2013).

29. Scholfield, M. R. *et al.* Force Field Model of Periodic Trends in Biomolecular Halogen Bonds. *J. Phys. Chem. B* **119,** 9140–9149 (2015).

30. Matthews, B. W. Structural and genetic analysis of the folding and function of T4 lysozyme. *FASEB J.* **10,** 35–41 (1996).

31. Carra, J. H., Murphy, E. C. & Privalov, P. L. Thermodynamic effects of mutations on the denaturation of T4 lysozyme. *Biophys. J.* **71,** 1994–2001 (1996).

32. Xie, J. *et al.* The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **22,** 1297–301 (2004).

33. Matsumura, M. & Matthews, B. Control of enzyme activity by an engineered disulfide bond. *Science (80-.).* **243,** 792–794 (1989).

34. Peeler, J. C. & Mehl, R. A. in *Methods Mol Biol.* **8,** 125–134 (2012).

35. Eriksson, A. E., Baase, W. A. & Matthews, B. W. Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J. Mol. Biol.* **229,** 747–69 (1993).

36. Liu, L., Baase, W. A. & Matthews, B. W. 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,** 595–605 (2009).

37. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* **276,** 307–326 (Elsevier, 1997).

38. Winn, M. D. *et al.* Overview of the CCP 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 235–242 (2011).

39. Nicholson, H., Anderson, D. E., Dao Pin, S. & Matthews, B. W. Analysis of the interaction between charged side chains and the .alpha.-helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry* **30,** 9816–9828 (1991).

40. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 213–21 (2010).

41. Heinz, D. W., Baase, W. A. & Matthews, B. W. Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino

acid sequence. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 3751–5 (1992).

42. Frisch, M. J. *et al.* Gaussian 09, Revision E.01. (2009).

43. Peterson, K. A., Shepler, B. C., Figgen, D. & Stoll, H. On the spectroscopic and thermochemical properties of ClO, BrO, IO, and their anions. *J. Phys. Chem. A* **110,** 13877–83 (2006).

44. Boys, S. F. & Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **19,** 553–566 (1970).

45. Simon, S., Duran, M. & Dannenberg, J. J. How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *J. Chem. Phys.* **105,** 11024 (1996).

46. Kuroki, R., Weaver, L. H. & Matthews, B. W. Structure-based design of a lysozyme with altered catalytic activity. *Nat. Struct. Biol.* **2,** 1007–1011 (1995).

47. Gorin, G., Wang, S.-F. & Papapavlou, L. Assay of lysozyme by its lytic action on M. lysodeikticus cells. *Anal. Biochem.* **39,** 113–127 (1971).

48. Yang, X. *et al.* Simultaneous interactions of amphoteric halogen in XY (X=Cl, Br and Y=F, Cl, Br) with C and O atoms of CO2 in ring-shaped CO2·X(Y)·CO2 complexes. *Comput. Theor. Chem.* **1076,** 32–41 (2016).

49. Baase, W. A., Liu, L., Tronrud, D. E. & Matthews, B. W. Lessons from the lysozyme of phage T4. *Protein Sci.* **19,** 631–641 (2010).

50. Boeckler, F. M. *et al.* Towards a QM-based scoring function for the recognition and assessment of halogen bonds. in *ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY* **248,** (2014).

51. Walter, S. M., Kniep, F., Herdtweck, E. & Huber, S. M. Halogen-Bond-Induced

Activation of a Carbon-Heteroatom Bond. *Angew. Chemie Int. Ed.* **50,** 7187–7191 (2011).

52. Kniep, F. *et al.* 5-Iodo-1,2,3-triazolium-based multidentate halogen-bond donors as activating reagents. *Chem. Commun.* **48,** 9299 (2012).

53. Kniep, F. *et al.* Organocatalysis by Neutral Multidentate Halogen-Bond Donors. *Angew. Chemie Int. Ed.* **52,** 7028–7032 (2013).

54. Takeda, Y., Hisakuni, D., Lin, C.-H. & Minakata, S. 2-Halogenoimidazolium Salt Catalyzed Aza-Diels–Alder Reaction through Halogen-Bond Formation. *Org. Lett.* **17,** 318–321 (2015).

55. Bayse, C. A. & Rafferty, E. R. Is Halogen Bonding the Basis for Iodothyronine Deiodinase Activity? *Inorg. Chem.* **49,** 5365–5367 (2010).

CHAPTER 5

CONCLUSIONS

I. INTRODUCTION

The work in this dissertation has helped established the significances of how halogen bonds (X-bonds) can be used more effectively in biomolecular systems, which until recently had been overlooked. It is clear that X-bonds are becoming an increasingly important molecular design tool that can be used to develop novel biologically based materials and the key to successful engineering of X-bonds into these systems is understanding how they are affected by X-bonds. This work has helped elucidate this concept by asking the question of what is the structure-energy relationship of X-bonds in biological macromolecules and how do halogen bonds contribute to the stability of these systems. This question was addressed by studying the occurrence of X-bonds in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PBD), parameterizing a force field that provides information about the physiochemical properties of halogens, and by engineering X-bonds in to a protein system to affect thermostability. From these studies, several conclusions can be drawn.

II. HALOGEN BONDS OCCUR IN A VARIETY OF GEOMETRIES AND STRENGHTS MAKING THEM A TUNABLE MOLECULAR INTERACTION

We demonstrated this in Chapter 2 with the aid of quantum mechanical (QM) calculations, which investigated how the strength of an X-bond is affected when the environment is altered. Specifically, we determined how the strength of an X-bond is affected when the

electron withdrawing ability of the halogen's covalently bound substituent is changed, the electron donating ability of the Lewis base is altered, or the polarization of the solvent increases. To further understand the structure-energy relationship of X-bonds in biological macromolecules we conducted a survey of the PDB in which over 600 structures were found to have the geometrical definition of an X-bond, e.g. short interaction distances and linear approach angles (Θ_1). The conclusion that X-bonds have short interactions distances seems obvious because there are numerous specific examples of this phenomenon in the literature. What was missing was a comprehensive study that determined if these individual occurrences had anything in common. From these 600 structures we were able to determined that the optimal interaction distance for an X-bond is 7% shorter than the sum of the van der Waal radii of participating atoms. Thus, both the specific environment and geometry of an X-bond must be considered when determine where to engineer the interaction in biological systems.

III. THE PHYSIOCHEMICAL PROPERTIES OF COVALENTLY BOUND HALOGENS CAN BE DESCRIBED BY A ELECTROSTATIC FORCE FIELD MODEL

In Chapter 3 we expanded the force field for biological halogen bonds (*ff*BXB) to include chlorine, bromine, and iodine. The strength of this model is that it is relatively straight forward in its approach to describe X-bonds as a primarily electrostatic interaction and this molecular mechnaics approach is able to directly model the experimentally determined energies of X-bonds in a nearly 1 to 1 relationship. Another advantage is that the physiochemical parameters, from which the *ff*BXB is based, can be readily interpreted to provide information about the size, shape, and anisotropic distribution of charge across the surface of a covalently bound halogen. The results from the parameterization of the *ff*BXB follows periodic trends, for example, it predicts

that the average radius (<R_{vdW}>) increase with Cl < Br < I. Polar flattening, or the change in the radius of the halogen (ΔR_x), is also described by this model and as expected the halogens that experience stronger polarization also experience the most dramatic polar flattening with I > Br > Cl. The charge parameters (*A*, *B*, and υ) of the *ff*BXB can be used to describe the anisotropic distribution of charge across the surface of each halogen and can be used to calculate the overall charge across the entire surface of the halogen. These parameters also allow for the electroneutral angle (Θ_0) to be calculated for each halogen, which reflects the angular size of the σ-hole and again this follows the expectation that the more polarizable halogens have larger σ-holes with I > Br > Cl. The ultimate impact of this model not only allows for the interaction strength of an X-bond in various geometries to be predicted, but also provides clarity to the relationship of the how physicochemical properties of covalently bound halogens affect the strength of X-bonds.

IV. ENGINEERING HALOGEN BONDS INTO PROTEIN AFFECTS STABILITY

The goal in Chapter 4 was to demonstrate that X-bonds could be used to increase the stability of biological macromolecules by engineering them into T4 lysozyme. In the current study, we did not succeed in designing a more thermally stable form of T4 lysozyme. In hindsight, we were trying to produce an even more stable protein which is known already to be incredibly stable. One explanation to why were weren't successful in developing a more stable protein is that the halogenated constructs did not achieve a linear approach angle (Θ_1), thus the relative strength of the X-bonds in this study are weaker than they would have been at 180°. None the less, that doesn't lessen the fact that we were, indeed, successful in engineering X-bonds into the enzyme with site-specific unnatural amino acid incorporation and determined that

X-bonds can reduce protein stability. The stability provided by the Y18^{x}F constructs were primarily due to the formation of an X-bond and not caused by a steric or hydrophobic effect which is evidenced by the phenol being unperturbed in the Y18F, Y18^{me}F and all of the Y88 constructs. Furthermore, we see that halogenation of T4 lysozyme on the solvent exposed surface where no intramolecular interactions can occur leads to a general destabilizing effect and that the X-bonds at position Y18^{x}F restores approximately half of the thermostability. Even though there is significant room for improvement in this study for engineering X-bonds that will increase thermostability, we successfully demonstrated that it is possible to intentionally incorporate X-bonds into the structure of proteins, but the rules that govern if the mutation is stabilizing may not be as straight forward as with canonical non-covalent interactions.

V. FUTURE DIRECTIONS AND CONCLUSIONS

This dissertation has elucidated some of the considerations necessary to understand how X-bonds can be used in biological macromolecules. Future directions for this field is to developed simple computational tools that scientists in other fields, such as biologists, can easily use to determine if the addition of a X-bond is viable. Developing these tools would allow for better understanding of how X-bonds contribute to the entropy of biological macromolecules, and increase the success rate when rationally designing new medicines. These tools can also help engineer novel recognition surfaces such as molecular receptors, nanomaterials, supramolecular materials, and liquid crystal lattices. Taken together, this dissertation describes how the structure-energy relationship of X-bonds can affect the stability of biological macromolecules. Specifically, it helps elucidate many of the concepts that must be taken into consideration when

engineering an X-bond into biological macromolecules, and provides a basis for how X-bonds affect the stability of biological macromolecules.