From Holliday Junctions

to Biomolecular Halogen Bonds:

How Biology Informs Us

about Complex Chemistry

P. Shing Ho ’79

Department of Biochemistry & Molecular Biology

1870 Campus Delivery

Colorado State University, Fort Collins, CO 80523-1870

Shing.Ho@Colostate.edu

Abstract

The greatest joy for a scientist is to discover something that is totally new and unexpected—the “eureka moment.” This essay serves as a travel log of a journey that led to the unexpected discovery of halogen bonding in biological molecules, or biomolecules. Halogens are not among the primary elements in biology, but they are found in a large number of natural and synthetic compounds that bind to and, in many instances, inhibit the functions of proteins and enzymes. Outlined here is a circuitous path that started with studies on an unusual four-stranded DNA structure (known as a Holliday junction) that led to a set of studies to map the occurrence of electrostatic interactions between halogens and various electron-rich atoms in proteins and nucleic acids (these interactions are now called halogen bonds). With this map, we have now charted a strategy to exploit halogen bonds as a molecular tool to design new compounds for potential clinical use. Along the way, we (and here, I mean my students) have established that halogen bonds have geometries that are unique to biomolecules, experimentally measured their stabilizing potentials, and developed mathematical equations that relate their geometries with their energies. It is perhaps even more satisfying to see how this particular
Introduction

The official definition of the “halogen bond” (or X-bond) for the International Union of Pure and Applied Chemistry (IUPAC, the arbiters of definitions and naming conventions in chemistry) was just published (1). Your first reaction to this announcement may be “what the heck is a halogen bond?” The fact that most readers of this article have never heard of this interaction (although most have heard of hydrogen bonds, or H-bonds (2)) may lead you to think that this must be something new and exciting in chemistry. You would be partly correct; X-bonds have become very exciting, but they are hardly new—only our understanding of their occurrence and use. Halogens, as you may recall from freshman chemistry, are the Group VII elements (fluorine, F; chlorine, Cl; bromine Br; iodine, I; and astatine, At), residing on the penultimate column of the periodic table. With this essay, I will introduce the reader to X-bonds, focusing primarily on those involved in structures and interactions with biological molecules (I will call these biomolecular halogen bonds, or BXBs for short); however, this is not meant to be a comprehensive review. There have already been several such tomes published describing how X-bonds are being applied in both material science (3, 4) and in biology (5–7). Instead, I will present to the readers of F&M Scientist a personal historical perspective on how a set of lucky scientific accidents through a series of seemingly unrelated experiments led us to stumble onto the X-bond, eventually leading to our contribution to writing its official definition for the IUPAC. I intend this narrative to emphasize the role that serendipity plays in the process of scientific discovery. As Isaac Asimov put it, “The most exciting phrase…in science, the one that heralds new discoveries, is not ‘Eureka!’ but ‘That’s funny…’”

Holliday junctions and the “stupid” experiment

Just at the turn of the 21st century, the research in my laboratory took a turn, literally, toward an unusual structure of DNA called the Holliday junction. The Holliday junction is a four-stranded DNA structure that was proposed by Dr. Robin Holliday as an intermediate during the exchange of genetic information.

eureka moment is becoming accepted as not only a new paradigm, but also as a practical concept in the biological and medicinal chemical fields.
in cells, known as recombination (8, 9) (for a review of recombination and its role in human sexual development, see http://www.hhmi.org/biointeractive/meiosis). Up to that point, I had established a research program in structural biology to study how the three-dimensional structure of double-helical DNA affects genetic function (a quick review of DNA structure and base pairing can be found at http://www.hhmi.org/biointeractive/paired-dna-strands and http://www.hhmi.org/biointeractive/watson-constructing-base-pair-models). We had, for example, computationally mapped the occurrence of Z-DNA across the human genome, suggesting that this left-handed double-helical form accumulated at transcriptional start sites for genes (10–12), and had published a movie showing in atomic-level detail how the B-form of DNA (the conformation proposed by Watson ad Crick (13)) converted to the A-form (the structure determined by Rosalind Franklin (14)) (15).

In the mid-1990s, we started working in collaboration with Dr. John Hearst (professor emeritus at the University of California, Berkeley), to determine how the drug psoralen affects the structure of DNA. Psoralens are a class of heterocyclic coumarins, which interact with DNA by slipping (intercalating) between the stacked base pairs and, upon photoactivation by ultraviolet light (16), cross-link the double-strands and eventually lead to programmed cell death (17). They are one of the oldest drugs still in clinical use—PUVA (psoralens plus ultraviolet light) is used today as a treatment for psoriasis, eczema, and vitiligo. An entry in Eber’s papyrus (a medical text written in ~2,000 BC) described how a species of plants found along the banks of the Nile River could be used to treat various skin ailments (there are similar descriptions for the use of such plants found along rivers in India and China, reviewed in (18)).

We had grown crystals of a DNA where the two central thymine bases were photocross-linked with psoralen, and collected X-ray diffraction data that lead to the calculation of electron density maps for the molecules in the crystals (19). The maps that my graduate student (Dr. Brandt Eichman, currently an Associate Professor at Vanderbilt University) drew showed breaks in the electron density along the phosphoribose backbone near the cross-linked Ts. It took several more crystals, however, before we were finally convinced the DNA backbone had crossed-over from one DNA duplex to an adjacent duplex in the crystal (persistence is needed to convince an advisor of something new and
unexpected). Thus, the DNA formed a four-stranded structure, the Holliday junction, with two extended strands that looked like standard B-DNA, and two strands that form tight U-turns connecting between them (Figure 1). The Holliday junction in solution had been studied for years (20), but an atomic level structure had remained elusive—the Holy Grail of DNA structure.

Brandt then ran the control experiment to determine whether the junction was drug induced by determining the structure of the DNA alone. To our surprise, the DNA itself formed a Holliday junction (21). By combining this information with that from a paper that had just been published on the structure of a junction that contained GA mispairs (22), we proposed that H-bonding interactions within the ACC nucleotides at the 6th, 7th, and 8th

Figure 1. Structure of the four-stranded DNA Holliday junction. A. Structure of psoralenated CCGT*ACCGG (19), where T* indicates the thymine bases that are cross-linked across the DNA duplex by 4’-hydroxymethyl-4,5',8-trimethylpsoralen (see inset for molecular structure). The experimental electron densities for two unique strands (backbone traced in green and in gold) of the four-stranded DNA junction are rendered as green chicken wire mesh. The psoralen molecules, in magenta, are shown cross-linking the thymines T5 and T15 (labeled). B. Structure of CCGGTACCGG DNA (21). The DNA sequence, in the absence of any cross-linking, adopts a compact stacked-X form of the junction (20). In this structure, the arms stack as two pairs, forming near continuous B-type duplexes, interrupted only by the crossing strands (in yellow and green) that connect the duplexes. Two sets of H-bonds from cytosines to the oxygens of the DNA phosphoribose backbone help stabilize the crossing strands and, thus, the junction: the H-bond from C7 (blue box) is important, while that from C8 (red box) is considered to be essential, but not sufficient for junction formation.
positions, common to both sequences, was the trinucleotide core responsible for the formation of the junction.

To test this hypothesis, I asked the next graduate student to enter the lab, Frank Hays (now an assistant professor at the University of Oklahoma Health Sciences Center), to determine the crystal structures of all 64 combinations of the sequence CCnnnN$_6$N$_7$N$_8$GG, where the nucleotides at the N$_6$N$_7$N$_8$ junction core contained any of the four common bases and nnn are nucleotides that maintain the self-complementary nature of the overall sequence. The idea was that by solving the structures of all possible combinations, we would be able to determine how individual nucleotides help to stabilize junctions and, consequently, how they affect recombination. I called this the “stupid” experiment, because it took no additional thought on my part, just persistence and ingenuity from Frank and an army of undergraduate and high school students. They were able to crystallize 63 combinations and solve the structures of over half of these. The resulting crystal structures generated a phase map that related the DNA sequence to structure (23). In particular, a cytosine was seen to be essential at the N$_8$ position of the N$_6$N$_7$N$_8$ core, associated with an H-bond from the N4 amino group of the cytosine back to the phosphate oxygen of the preceding nucleotide. A cytosine is also preferred at the N$_7$ position, again, showing an H-bond forming to the base to the backbone. The resulting sequence preference in junction stabilization led to a new model for how recombination can be site specific even if the recombination enzymes show no specificity for binding a particular DNA sequence (24).

You may be asking, “why all this detail about Holliday junctions?” We will see how the experiment to fish out junction forming sequences led to halogen bonds, and how the stabilizing interactions in junctions can be used to measure the stabilizing energies of a halogen bond.

**From Holliday junctions to halogen bonds**

In the course of the stupid experiment, Frank Hays ran across a sequence, CCAGTACTGG, the structure of which could not be easily determined from the X-ray diffraction data. A standard trick in such situations is to introduce an electron rich atom, such as a halogen, into the crystal to help phase the data (X-rays are diffracted by electrons, and an electron-rich halogen serves as a beacon to help guide the calculation of the electron density map).
Consequently, he crystallized the sequence CCAGTACBrUGG, where BrU is the 5-bromouracil base (a thymine with the methyl group replaced by a Br atom). This brominated structure was seen to be a Holliday junction (25), which was surprising since the sequence did not contain a cytosine at the essential N8 position, as we had previously discussed (21, 26) The structure of the native thymine-containing sequence, once solved, was a standard B-DNA duplex, suggesting that the bromine was responsible for stabilizing the junction.

A careful examination of the brominated junction structure (Figure 2) showed that the halogen of the BrU base was ~3 Å away from the same phosphate oxygen that is H-bonded to the C8 that stabilizes the junction (or 0.4 Å shorter than the sum of the van der Waals radii, \( \Sigma R_{vdW} \), of the two atoms). This made no sense from the perspective of our understanding of the standard chemical properties of bromine (5). Halogens are electron rich atoms and, therefore, considered to carry an over electronegative charge—they should repel and not be attracted to a formally negatively charged phosphate oxygen of the DNA backbone, and certainly should not replace a stabilizing H-bond. Since we had

---

**Figure 2.** Bromine stabilized junction. **A.** Single crystal structure of CCAGTACBrUGG as a four-stranded junction (backbone traced by a yellow ribbon) (25). The electron density map (inset, 2Fo-Fc, at 2\( \sigma \) and 3\( \sigma \) levels are rendered as chicken wires, blue and red, respectively) shows a bromine (Br) to phosphate oxygen distance of 3.0 Å (van der Waals radius for Br = 1.9 Å and for O = 1.5 Å, for a sum = 3.4 Å). **B.** Single crystal structure of CCAGTACTGG is a standard B-DNA duplex.
no reasonable explanation for the short, stabilizing Br···O interaction, we did not discuss it when the structures were published in 2003 (25).

In 2004, a Fulbright grant to study in Strasbourg, France, provided me the luxury to pursue a number of different research projects, including time to think about the Br···O conundrum. My “thinking” led me to try a Google search for the phrase “short bromine oxygen interactions”, which further led me to the 1970 Nobel lecture by Odd Hassel on charge transfer bonds (27). The charge transfer bond was one in which the formally neutral atoms of two compounds (for example, Br₂ and the oxygen of dioxane) can come together by transferring charges (28), such that one becomes formally positive and the other formally negative, consequently, allowing the two atoms to interact electrostatically at very short distances (20% or more shorter than ∑RᵥvdW). In short, nihil novi sub sole (“there is nothing new under the sun.” [Ecclesiastes 1:9]).

A more comprehensive search of the literature, beyond Google, showed that the topic of charge transfer bonding went through about two decades of near hibernation, with only a few publications discussing the interaction each year. In the late 1990s, interests in halogen interactions re-emerged in the material chemistry literature, but were now called “halogen bonds,” or X-bonds, in analogy to the better recognized H-bond (29). The old dog was resurrected, not only with a new name, but it ends up that it was not really even a dog—charge transfer was not the root of short halogen to oxygen interactions.

The emerging understanding of X-bonds is that they are electrostatic in nature, but that their ability to attract electron-rich electronegative atoms, such as oxygen and nitrogen, is best described as a quantum chemical effect (Figure 3). To understand the current model, we start with the basic chemistry of a covalent carbon-halogen (C—X) single bond (or σ-bond). To form this σ-bond, one electron from the sp-hybrid orbitals of the carbon would combine with one from a p-atomic orbital of the halogen. Now, recall that the electronic configuration of the outer-most orbitals of halogens is s²p⁵, or that there is a single valence electron available to form a covalent bond (the standard convention is that this electron is in the pₓ-orbital, which must point in the direction of resulting covalent bond). By pairing this electron to form the σ-bond, we have effectively depopulated the pₓ-orbital. This depletion of electrons is called the σ-hole (30), which partially exposes the positive charge of the nucleus in the direction diametrically opposed to the covalent bond (an
Electropositive crown) that can attract an electron-rich atom. X-bonds are now being referred to by some as an extreme example of the broader class of σ-hole bonding (31, 32).

This σ-hole model makes several predictions: 1) atoms with more polarizable electrons (i.e., larger halogens whose electrons are more readily redistributed) will form larger σ-holes; 2) when electrons are drawn into the σ-bond and away from the halogen, it will exaggerate the σ-hole; and 3) electron-rich atoms will interact primarily with the electron-depleted $p_z$-orbital. Indeed, the strength of an X-bond increases as we go down the Group VII column of the periodic table (F < Cl < Br < I) (33); as we add electron withdrawing substituent groups, particularly in aromatic systems, opposite the halogen (34); and electron-rich oxygens, nitrogens, and sulfurs in X-bonds tend to be linear relative to the covalent C—X bond ($\theta_1$-angle = 180°).

This is not to imply that there is a complete physical understanding of X-bonding; there remains debate as to how other forces including dispersion

Figure 3. Halogen bonds. In a halomethane (X-Me), the C—X bond is pulled from the atomic $p_z$-orbital of the halogen, exposing the nuclear charge in the direction opposite this σ-bond (36). The resulting electropositive crown (the σ-hole) is seen mapped on the halogen surface of the halomethane series (top row), looking down the halogen along the C—X bond (blue surface as an electropositive potential, red as electronegative potential, as calculated at the B3LYP level, energies in kcal/mol). The σ-hole is seen to grow as the halogen becomes larger and, thus, more polarizable. Attaching the halogen to a strong electron withdrawing aromatic group, such as a uracil base (bottom row) further exaggerates the polarization and the σ-hole.
and even charge transfer contribute to the interaction (4, 33). However, there is agreement that X-bonds are largely electrostatic in nature, much like H-bonds, and the physical characteristics are analogous between the two (Figure 4). In analogy to the partial positive hydrogen in an H-bond, the electropositive σ-hole is considered to be the X-bond donor, and any atom or functional group that can serve as an H-bond acceptor is recognized as being a potential X-bond acceptor (1). The general principles have allowed chemists to apply the concept of X-bonding to control the formation of crystals and liquid crystals to engineer new materials, including organic conductors, molecular sensors, and so forth (4).

**The biomolecular X-bond**

When we fell upon the concept of X-bonds, it was not clear how important they would be in biology; halogens are not typically found in proteins, nucleic acids, lipids, etc. Thus, the first question we asked after our Google “experiment” was, are X-bonds present in biomolecules? To address this question, I worked
with Dr. Pascal Auffinger in Professor Eric Westhof’s lab (my Fulbright host) to perform a structural survey for short halogen interactions in the Protein Data Bank (PDB) (35), a repository of crystal structures of all biomolecular structures, similar to surveys of X-bonds in the Cambridge Structural Database of small molecules. To our surprise, there were over 100 examples of X-bonds in the PDB (36)—we were not alone in ignoring something we did not understand. Perhaps not surprisingly, most of these structures were complexes between proteins and halogenated ligands, many of which were inhibitors. In one example, the bromine of the inhibitor to the enzyme aldose reductase was shown to account for a greater than 1,000-fold selectivity against the very similar enzyme, aldehyde reductase (37).

The properties of these biomolecular X-bonds, or BXBs (36) were seen to be very similar to those seen in small molecules (38). The $\theta_1$-angle is again fairly linear to the C—X bond (Figure 5A), indicating that the positive crown of the \( \sigma \)-hole is the primary X-bond donor. The approach of the halogen to the acceptor atom ($\theta_2$-angle) in small molecules is generally $\sim$120°, implicating the non-bonding electrons of the acceptor. In BXBs, this is also the case for Cl, but as we move toward the larger halogens (Br and I), this angle approaches 90° (Figure 5B). The implications are that the X-bond is to the delocalized \( \pi \)-system of the peptide bond of proteins (36). A more detailed analysis showed that this geometry is dictated by the H-bond to the carbonyl oxygen of the peptide bond (Figure 5C), which are present to help stabilize structures such as \( \alpha \)-helices and \( \beta \)-sheets in proteins, leading us to the unique proposal that X-bonds are orthogonal (both geometrically and energetically) to H-bonds when they share a common acceptor atom (39). This concept of orthogonal interactions provides a framework for designing new halogenated inhibitors against protein targets: the atomic level structure of an enzyme in complex with an inhibitor candidate can inform us as to where to place a halogen to engineer an X-bond with optimum geometry and, thus, improve the efficacy of the potential inhibitor.

The principle of X-bonds to \( \pi \)-systems is now seen as being general, with aromatic rings being recognized as potential acceptors (40, 41). In particular, X-bonding to the \( \pi \)-electrons of carbonyl oxygen has recently moved out of the realm of biomolecules, and has been applied in the material chemistry field to design a cage compound that recognizes molecular halogens (42).
How stable is a biomolecular X-bond

Although the geometry and chemistry of an X-bond are very similar to the H-bond, we wanted to determine whether their stabilizing potentials were comparable. A calorimetric study of a bipyridine derivative mixed with an X- and H-bond donor demonstrated that the two interactions do compete against...
each other in forming a crystal (43). We wanted to determine if this was also true for biomolecular interactions, since we would expect the hydrophobic halogens to potentially behave differently in aqueous solution. For these studies, Andrea Regier Voth (a former graduate student now at the NIH) designed a four-stranded DNA junction to directly compete a stabilizing H-bond against an engineered X-bond (Figure 6). The idea was very simple: a junction was assembled from two different sequences—one with a cytosine at the \( N_7 \) nucleotide position that can form the standard H-bond, and a second with a brominated uracil (BrU) at the comparable nucleotide position. We

![Figure 6](image-url). Competition of a hydrogen bond against a halogen bond in a DNA junction construct (45). A. Assay to compare the stabilizing potentials of an X-bond to an H-bond was developed by constructing a DNA junction in which one set of strands (blue) contains G\(_4\)-C\(_7\) base pairs (green) that can form H-bonds (blue box) to the backbone to stabilize the junction. The complementary strands (red) contain halogenated uracil bases (xU) at the A\(_4\)-xU\(_7\) base pair positions (magenta) that can form X-bonds (magenta box) to the backbone. The two interactions compete to define an H-bond stabilized conformer (H-isomer) or an X-bonded conformer (X-isomer). B. Residual electron density of brominated DNA junctions. The difference electron density maps (2Fo-2Fc) are shown at the cross-over region of the construct in A, with two BrU (Br\(_2\)J) or one BrU (Br\(_1\)J) incorporated into the potential X-bonding strands. The residual density (rendered as magenta chicken wire) in both cases shows the bromine residing primarily at the junction cross-over, indicating that the X-bond is more stabilizing than the competing H-bond. The X-bond in the Br\(_2\)J construct is shorter than that in Br\(_1\)J.
knew that junctions can isomerize (44) and, therefore, predicted that the junction could be stabilized by the H-bond (the H-isomer), or adopt an isomer form that it is stabilized by the competing X-bond (the X-isomer) (45). If the bromine is observed on the outside strand, then we had the H-isomer and the H-bond is more stabilizing. However, if the bromine is on the inside cross-over strand, then the junction is in the X-isomer, and the X-bond is more stabilizing. The assumption in this assay is that what is observed in the crystal structure reflects the distribution of isomers in solution.

The junction constructed with two brominated strands, which competed two X-bonds against two H-bonds (called the Br2J construct), showed the bromine only on the inside strand, indicating that the X-bond out competed the H-bond one-on-one (45) (Figure 6B). Next, Andrea assembled a junction in which only one strand of the DNA junction was brominated (the Br1J construct), competing one X-bond competed against two H-bonds. In this case, although the junction was predominantly in the X-isomer form, there was a small fraction of H-isomer observed. This mixture suggested that in the 1:2 competition, the X-bond still won, but the observed ratio allowed us to estimate an actual difference in energy, with the bromine X-bond (Br-bond) being ~ 3 kcal/mol more stable than the H-bond in this system.

Megan Carter (a recently completed Ph.D., now at Stockholm University), used differential scanning calorimetry (DSC) to measure a stabilizing enthalpy of -3.6 kcal/mol for the Br-bond relative to the competing H-bond, thereby validating the crystallographic assay (the energy of this H-bond is ~0.6 kcal/mol) (46). She, along with Brittany Rummel, an undergraduate student who has since completed an M.S., extended this study to determine the energies of X-bonds for F and Cl using the crystallographic assay, and Br and I using the DSC assay (Table 1) (47). The results showed that the enthalpies of stabilization follow the predictions from the σ-hole model; as the halogen becomes larger, the X-bonding strength increases (F < Cl < Br < I), because of increased polarization. However, there is an associated entropic cost when trying to fit the largest halogen (I) into the pocket of the junction, resulting in the most stabilizing X-bond, according to the Gibb’s free energy, being associated with bromine rather than iodine in this system.
Now that we have all the energies and associated geometries of BXBs from the DNA junction assay, what do we do with them? In order that our studies not be simply an academic exercise, we must be able to apply what we have learned to something useful. Recall that our initial survey of BXBs found that most were involved in protein-inhibitor complexes and, therefore, it is natural to believe that we can start to exploit the BXB to facilitate the structure-based design of new halogenated compounds with potential therapeutic applications.

Rational drug design is based on the ability to accurately compute the energy of protein binding to an inhibitor; unfortunately, all current programs model halogens classically as negatively charged atoms. What we need to do, therefore, is to develop a set of mathematical equations that describe the σ-hole that allows halogens to form X-bonds—a force field for BXBs (which we abbreviate as $ff_{BXB}$) that can be incorporated into classical molecular mechanics (MM) based molecular modeling programs.

There had already been several computational approaches at the time to model X-bonds. The most accurate methods are quantum mechanics (QM) calculations on the ground-state energies of X-bonded pairs (48, 49), but these

---

**Table 1.** Stabilizing potentials of halogen bonds in a DNA junction (47). The energies associated with halogens interacting with the formally negative phosphate oxygen ($\text{O}^{-\frac{1}{2}}$) of the DNA backbone serving as the acceptor. For each X-bond, the method by which the energies were determined is indicated: X-ray crystallographic titration (X-ray) or differential scanning calorimetry (DSC). Listed are the differences in enthalpy, entropy, and Gibbs free energy of the X-bond relative to the competing H-Bond ($\Delta H_{(X-H)}$, $\Delta S_{(X-H)}$, and $\Delta G_{(X-H)}$, respectively, see Figure 6), where the $\Delta G$ of the H-bond is estimated to be 0.64 ± 0.07 kcal/mol.

<table>
<thead>
<tr>
<th>Halogen</th>
<th>Method</th>
<th>Distance</th>
<th>$\Delta H_{(X-H)}$</th>
<th>$\Delta S_{(X-H)}$</th>
<th>$\Delta G_{(X-H)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>X-Ray</td>
<td>3.2 Å</td>
<td>ND</td>
<td>ND</td>
<td>0.12 kcal/mol</td>
</tr>
<tr>
<td>Cl</td>
<td>X-Ray</td>
<td>2.92 Å</td>
<td>ND</td>
<td>ND</td>
<td>-0.15 kcal/mol</td>
</tr>
<tr>
<td>Br</td>
<td>X-Ray</td>
<td>3.32 Å</td>
<td>ND</td>
<td>ND</td>
<td>-1.6 kcal/mol</td>
</tr>
<tr>
<td>Br</td>
<td>DSC</td>
<td>ND</td>
<td>-3.6 kcal/mol</td>
<td>+4.2 cal/mol K</td>
<td>-4.8 kcal/mol</td>
</tr>
<tr>
<td>I</td>
<td>X-Ray</td>
<td>2.97 Å</td>
<td>ND</td>
<td>ND</td>
<td>≤-1.46</td>
</tr>
<tr>
<td>I</td>
<td>DSC</td>
<td>ND</td>
<td>-5.9 kcal/mol</td>
<td>-12.0 cal/mol K</td>
<td>-2.3 kcal/mol</td>
</tr>
</tbody>
</table>

Computational models of biomolecular X-bonds

Now that we have all the energies and associated geometries of BXBs from the DNA junction assay, what do we do with them? In order that our studies not be simply an academic exercise, we must be able to apply what we have learned to something useful. Recall that our initial survey of BXBs found that most were involved in protein-inhibitor complexes and, therefore, it is natural to believe that we can start to exploit the BXB to facilitate the structure-based design of new halogenated compounds with potential therapeutic applications. Rational drug design is based on the ability to accurately compute the energy of protein binding to an inhibitor; unfortunately, all current programs model halogens classically as negatively charged atoms. What we need to do, therefore, is to develop a set of mathematical equations that describe the σ-hole that allows halogens to form X-bonds—a force field for BXBs (which we abbreviate as $ff_{BXB}$) that can be incorporated into classical molecular mechanics (MM) based molecular modeling programs.

There had already been several computational approaches at the time to model X-bonds. The most accurate methods are quantum mechanics (QM) calculations on the ground-state energies of X-bonded pairs (48, 49), but these
are extremely cumbersome, computationally costly, and fraught with errors for such large biological molecules. Another approach is to apply the QM calculations in and around the ligand-binding site, leaving the modeling of the remainder of the protein to the more efficient MM calculations. This hybrid QM/MM approach (50) has seen some limited success in predicting inhibitor affinities to some enzymes (51). A completely classical MM method has been developed by adding a positive extra point (PEP) on the surface of the halogen, diametrically opposed to the C—X bond in order to model the electropositive crown of the σ-hole (52, 53). The PEP method was capable of putting in correct order a list of inhibitors against a protein target, but the magnitude of the resulting energies were much larger (by an order of magnitude) than observed for BXBs. This problem suggested that BXBs, and perhaps X-bonds in general, are not entirely electrostatic interactions and, therefore, cannot be entirely accounted for by simple point charge models. The situation is improved when dispersion forces are added to the PEP model (54).

We decided that the most robust approach would be to derive an ffBXB based entirely on well-defined and measurable physical properties inherent in halogens—the physical shape and the non-uniform distribution of charge at the surface. Our strategy for deriving the ffBXB is quite simple in concept: 1) develop equations that describe the effects of distance and angle (θi) on the strength of the X-bond; 2) build molecular models of the BXBs in various geometries; 3) calculate the energies of the BXBs in their various geometries to assign values to the parameters in the equations (parameterize the force field); and validate the ffBXB by comparing the calculated with X-bonding enthalpies measure from the DNA junction studies. This task was given to Megan Carter, who worked collaboratively with Professor Anthony Rappé from the Chemistry Department to perform accurate QM calculations.

For the first step in this derivation, we considered the Coulombic potential that defines the effect of charge and the van der Waals potential of the atomic size on molecular interactions. The standard Coulombic potential \( V_C \) for the interaction energy (Equation 1) is dependent on the charges \( Z_1 \) and \( Z_2 \) of the two interacting atoms, the charge of an electron \( e \), the dielectric constant of the medium separating the two charges \( D \), and the distance between the charges \( r \). The van der Waals potential \( V_{vdw} \), Equation 2) includes a short-range attractive term that is dependent on the 6th-power of the distance.
separating two atoms ($r^6$) and an even shorter-range repulsive term that is dependent on the 12th-power of the distance ($r^{12}$), along with the effective size of the atoms, as reflected in their van der Waals radii ($R_{vdW}$).

**Equation 1**

$$V_c = \frac{Z_1 Z_2 e^2}{Dr}$$

**Equation 2**

$$V_{vdW} = 4\varepsilon \left[ \left( \frac{\sum R_{vdW}}{r} \right)^{12} - \left( \frac{\sum R_{vdW}}{r} \right)^6 \right]$$

In this model for X-bonds, the effective charge of the X-bond donor is dependent on the angle in which the acceptor approached the halogen: at a 180° $\theta_1$-angle approach, the acceptor will see a positively charged halogen, while the halogen will appear to be negatively charged from a 90° $\theta_1$-angle approach. To model this angle-dependence, we defined the charge of the halogen ($Z_X$) as a cosine function of $\alpha = (180° - \theta_1)$, adding a scaling ($A$) and baseline offset ($B$) for this function, and allowing the power of the distance relationship to float as a variable $n$ (**Equation 3**).

**Equation 3**

$$V_c = \frac{[A \cos(\nu \alpha) + B] Z_2 e^2}{Dr^n}$$

**Equation 4**

$$V_{vdW} = 4\varepsilon \left[ \left( \frac{R_{vdW(A)} + <R_{vdW(X)}> - \Delta R \cos(\nu \alpha)}{r} \right)^{12} - \left( \frac{\sum R_{vdW}}{r} \right)^6 \right]$$

We already knew from the PEP models that X-bonds could not be modeled entirely from an electrostatic interaction perspective and considered whether the effective size of halogens were also angle-dependent. Indeed, when the surface of one bromine in Br$_2$ is probed with a helium atom, using a high level QM calculation, the zero-point energy for interaction was seen to be shorter at $\theta_1 = 180°$ than for $\theta_1 = 90°$, suggesting that the $R_{vdW}$ for Br is flatter in the direction of the $\sigma$-hole than perpendicular to the hole. This prediction is consistent with the flattening of halogens in the direction opposite its covalent bond as observed in the high-resolution crystal structures of a series of small organic molecules. Once again, we modeled this polar flattening by allowing the deviation ($\Delta R$) from the average value ($<R_{vdW}>$) to be a cosine function of $\alpha$ (**Equation 4**). The resulting mathematical \textit{ff}BXB, once parameterized
Figure 7. Energy landscape of bromine X-bonds (67). A. Map of the force field for biomolecular halogen bonds (ffBXB) calculated energies (y-axis) as a function of distance from the center of the bromine (along the horizontal plane) in a bonded C—Br atom pair (carbon in grey, and bromine colored according to the calculated electrostatic potential, Figure 3). The three-dimensional energy surface is shown looking along the horizontal plane for a formally negatively charged oxygen serving as the X-bond acceptor, with energies from white (≤-10 kcal/mol), to green (0 kcal/mol), to blue (≥ +10 kcal/mol). B. The ffBXB calculated energy landscape for Br X-bond, with a formally neutral oxygen (-0.4e charge), as a model for a carbonyl oxygen in a peptide bond. The energies of interaction are mapped relative to the radial distance from the center of the bromine (see A for color scheme). C. The ffBXB calculated energy landscape for Br interactions with a formally positive hydrogen (+0.4e charge), as a model for an H-bond.
The ffBXB potentials were extended to map the structure-energy landscapes for other atom types by applying different charges to these atoms (Figure 7B). For example, by defining \( Z = -0.4 \), we can model the X-bond to the carbonyl oxygen of a peptide bond, as seen in most interactions of inhibitors against their protein targets (36, 55, 56), or a \( Z = +0.4 \) allows us to predict the geometry preferences of H-bonding interactions (57) to the bromine surface. Thus, the ffBXB, derived based on the intrinsic physicochemical properties of halogens, is applicable to modeling the electrostatic type interactions available to halogens, including X- and H-bonding.

Table 2. Parameters for the angle-dependent ffBXB functions describing the anisotropic shape and electrostatic potential energy functions for the bromine of BrU (Equations 3 and 4)

<table>
<thead>
<tr>
<th>Angle</th>
<th>Electrostatic Parameters (Equation 3)</th>
<th>Shape Parameters (Equation 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nu )</td>
<td>( A )</td>
<td>( B )</td>
</tr>
<tr>
<td>2.31</td>
<td>2.84</td>
<td>1.53</td>
</tr>
</tbody>
</table>

Table 3. Comparison of experimental and calculated enthalpies (kcal/mol) for bromine X-bonds in the Br1J and Br2J conformations of DNA junctions (45). Interaction enthalpies for X- minus H-bond (\( \Delta H_{X-H} \)) determined from a crystallographic competition assay (45) and by differential scanning calorimetry in solution (65) are compared to X-bond energies from QM and ffBXB calculations applied to X-bonds of the BrU···H2PO2⁻¹ or BrU···Dimethylphosphate (DMP⁻¹) model systems.

<table>
<thead>
<tr>
<th>X-Bond Geometry</th>
<th>Experimental ( \Delta H_{X-H} )</th>
<th>Calculated Energies (H₂PO₂⁻¹/DMP⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>( \theta_{1,\text{Angle}} )</td>
<td>Crystal Assay</td>
</tr>
<tr>
<td>3.32 Å</td>
<td>167.2°</td>
<td>-2.0 ± 0.5</td>
</tr>
<tr>
<td>2.87 Å</td>
<td>163.2°</td>
<td>-</td>
</tr>
</tbody>
</table>


**Perspectives**

In the decade or so since we started studying biomolecular X-bonds (36), there has been a number of variations seen for the molecular interaction, including interactions in which waters can intervene between the donor and acceptor atoms (58), those with aromatic rings and even anionic halides (59) as acceptors. More importantly, we have seen the acceptance of the interaction as being important in biomolecular systems and as being important for the rational design of new inhibitors against therapeutically important targets (6, 7, 60–62). There are those, however, who question why anyone should put so much effort and devote some much creative thought into studying an interaction with only 2–6 kcal/mol of stabilizing energy. Let’s consider the following scenario: you found a lead compound from a standard drug screen that seems to be promising, but shows only marginally affinity (a dissociation constant $K_D$ of ~10 µM). How do you redesign this to be more effective? If you can add a single interaction that contributes an additional 5 kcal/mol stabilization to the complex, that would buy you >4,000 fold increase in affinity, or that the $K_D$ will now be ~2 nM. Why not do this with a classic H-bond instead of an X-bond? The primary reasons are that 1) H-bonds are ubiquitous and, therefore, constantly compete against other H-bonds and 2) halogens are hydrophobic and, thus, prefer to sequester into environments such as protein binding pockets and away from the aqueous solvent.

I should note that we were not the first to recognize the potential to apply X-bonding for inhibitor and drug design. In 1996, a group at Merck had started to consider the use of halogens and X-bonding for the design of inhibitors against the blood clotting factor Xa (63). The effort had lead to the development of a higher affinity inhibitor; however, the research did not result in a marketed product. Such are the limitations of industrial research, where novel approaches are not readily disseminated to the scientific community at large.

In considering our own research, we propose a strategy for applying X-bonding as a concept to help improve the affinity and potential efficacy of an inhibitor (Figure 8). This strategy starts with the structure of a protein in complex with an unhalogenated lead compound. Using the concept of orthogonality, we can predict from the H-bonding pattern of the protein where to place halogens that would be geometrically optimum for X-bonding and in
a way that does not compete against the established H-bonds that stabilize the protein structure. This concept then provides the medicinal chemist with options as to where to add a halogen substituent onto the lead compound. Finally, the ffBXB is used to optimize the geometry and simulate the dynamics of the complex between the new inhibitor and its protein target, providing a calculation for the enthalpic and entropic energy terms required to accurately estimate the effect of the halogen on the affinity of the inhibitor. The field is starting to move towards applying X-bonding to optimizing inhibitor and drug efficacy (62, 63); more accurate descriptions of how halogens contribute energetically to ligand recognition and specificity, through X- and H-bonds, will greatly enhance their use in biomolecular engineering.

We have made significant strides toward the goal of developing this design strategy, but there remains a significant amount of work to be done before this strategy becomes fully functional. The most immediate task is to extend the ffBXB beyond bromine and onto the other halogens (F, Cl, and I) seen in drugs and drug candidates. We also need to address the question of why halogens are inherently hydrophobic, when they can form stabilizing electrostatic interactions (both H- and X-bonds) to water molecules. The answer to this question will require that we consider the entropic costs of such interactions. We have already seen how the entropy of a halogenated system is reduced

![Figure 8](image_url)

**Figure 8.** Strategy for rational design of X-bonding inhibitors. Starting with the structure of a lead compound in complex with its therapeutic protein target (A), the concept of orthogonality predicts where a halogen or halogens can form optimal X-bond(s) relative to the structural H-bonds in the protein (B). The predicted halogen positions will be useful to inform where a halogen can be introduced as a substituent on the lead compound (C). Finally, the ffBXB is used to optimize the geometry of complex and to predict the effect of the halogen X-bond on the affinity of the proposed halogenated inhibitor (D).
as a result of crowding effects (47)—we suspect that there will be additional conformational constraints imposed on waters when they must accommodate the electrostatic interactions with halogens (47, 64).

To close, the research in our laboratory has strayed from studying the detailed molecular interactions that could account for sequence-specific recombination events to the effects of electron distributions on the specificity of halogenated drugs. Essentially, we have taken a fairly simple biomolecular system (the Holliday junction) to learn new things about some relatively complex quantum chemistry of atoms. And this is how a structural biologist, by pursuing serendipitous observations that could easily have been ignored, helped to author the official definition of halogen bonds for the IUPAC, that most chemical of organizations.

**Acknowledgments**

The success of any research laboratory is dependent on the quality of the students who do the actual work. I am very grateful for the large number of outstanding students, past and present, graduates and undergraduates, who have contributed to these and other studies at Oregon State University and Colorado State University. The studies on halogen bonds from this laboratory have been supported by grants from the Franco-American Fulbright Commission, the National Science Foundation, and Colorado State University.
References


About the author

It is hard to imagine that the middle child from a family of five children, born in the slums of Hong Kong, would now find himself as the leader of an academically vibrant department at a Carnegie Research Intensive University—public education and access to higher education are the great equalizers. P. Shing Ho received his bachelor’s of science degree in chemistry from Franklin & Marshall College in 1979 and earned his Ph.D. in 1984 from the Department of Biochemistry, Molecular, and Cellular Biology at Northwestern University under Professors Brian M. Hoffman and Emanuel Margoliash. He continued his scientific training as an American Cancer Society Postdoctoral Fellow with Dr. Alexander Rich at the Massachusetts Institute of Technology in Cambridge, Mass. Dr. Ho launched his academic career in 1987 in the Department of Biochemistry & Biophysics at Oregon State University (OSU), was promoted to associate professor with tenure in 1992 and to professor in 1997. He became chair of the Department at OSU in 2002 and was recruited to serve as chair of the Department of Biochemistry & Molecular Biology at Colorado State University (CSU) in 2007 (the position he currently holds). Throughout his career, he has been recognized for contributions in research (including the Milton Harris Award in Basic Research from the College of Science at OSU, the Researcher of the Year Award from the OSU chapter of Sigma Xi, and the Discovery Award from the Medical Research Foundation of Oregon) and in teaching (including the L. Carter Award in Undergraduate Teaching and nomination for the L. Carter Award in Graduate Teaching from the College of Science at OSU, and the Provost N. Preston Davis Award for Innovation in Teaching at CSU). He has received fellowships to pursue sabbatical studies abroad, including a Wellcome Travel Fellowship at the University of Dundee in Scotland and Fulbright Award to study at the Université de Strasbourg in France.
Looking back on where I am now, it seems unfathomable that I had essentially failed Algebra I in middle school. At that stage in my life, a career in science never came to mind; my focus was in the fine arts (drawing, painting, and photography), where I had found initial success and recognition. The most enduring impact of a liberal arts education from F&M on my life was in the realization that I could marry the free form creativity of the fine arts with the discipline of science—the point of convergence is in research and discovery. My interest in chemistry started in high school in the general, organic, and biological chemistry courses I took from Dr. John Jacobs, but only fully blossomed into a passion in the classrooms at F&M, particularly that of Dr. Bonnie Sandel and later in Professor Claude Yoder’s research laboratory. I remember an evening during my sophomore year as I considered what was due the next day—a midterm in Dr. Sandel’s Organic Chemistry class and an art project for Dr. James Peterson’s painting class. The only solution was to take my Org Chem notes with me to the art studio, alternating between mapping synthetic pathways using the Grignard reaction and mastering the complex brush strokes of Édouard Manet in a study of his famous painting, *Un Bar aux Folies-Bergère*. During my senior year, I had the good fortune to work as a research student with Dr. Yoder, synthesizing and characterizing new silicon-based organic compounds. Between the eureka moments were the inevitable periods of frustration—those lapses would have been unbearable, except for the outlet of the F&M drawing studio. I continue to take advantage of the unique breadth of education that I received at F&M and, in this way, strive to find creative solutions to complex questions in biology, but always from a strong chemical perspective. Perhaps more importantly, I learned how to effectively communicate the excitement of science, both in printed and spoken words, and on digital canvases. Despite much of the evidence to the contrary, the general populace is hungry for new scientific discoveries; our task is to communicate these discoveries clearly and concisely so that the relevance of science to society as a whole remains unquestioned.