DISSERTATION

# RELATIONSHIPS BETWEEN HYDROGEN BONDS AND HALOGEN BONDS IN BIOMOLECULAR ENGINEERING

Submitted by

Rhianon Kay Rowe Hartje

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 2019

Doctoral Committee:

Advisor: P. Shing Ho

Melissa Reynolds
Christopher Snow
Tim Stasevich
Robert Woody

ABSTRACT

RELATIONSHIPS BETWEEN HYDROGEN BONDS AND HALOGEN BONDS IN BIOMOLECULAR ENGINEERING

In this dissertation, we will explore the interconnectedness between halogen bonds (X-bonds) and hydrogen bonds in rational biomolecular engineering efforts. As X-bonds are not readily designed into biomolecules, we aim to show how they can be advantageous for molecular design. We will begin by considering how X-bonds compare to H-bonds and show how the two can work in harmony to provide enhanced stabilizing potential. In two unique protein engineering efforts we will show 1) how the X-bond can be just as specifying in terms of molecular assembly as compared to the H-bond, and 2) how it can coordinate with the H-bond to increase protein stability. One study shows the specifying potential the X-bond possesses in terms of coiled-coil assembly. While the study points to a direct application of a sensing probe, the scope of the work will aid others using coiled-coils for materials purpose, designing protein interfaces or potential ligand binding sites. In the other protein engineering study, we will survey how a protein with an intrinsically disordered region responds to hydrogen enhanced halogen bond engineering. We show how we can drastically increase the thermal stability of the protein through minimal change to its primary sequence. This study lends itself to exploring bigger structure-function questions and how the stabilizing capacity of halogen bonds fits into this. Through this work we aspire to show how useful X-bonds can be for biological engineering efforts by exhibiting their specifying and stabilizing characteristics in these settings.

# ACKNOWLEDGEMENTS

While my initial passion and naïve curiosity for science dates back to my childhood, as I was encouraged by my parents, some other individuals who ignited a deeper zeal in me and believed in me throughout the years deserve a huge thank you as well. First off, my chemistry teacher in high school, Dr. Stribling. He really piqued my interest in chemistry in 11$^{th}$ grade, and unbeknownst to him is one of the main reasons I pursued a bachelor's degree in chemistry. He always pushed me through thoughtful questions and was encouraging through trying times. I owe much of my "love" for chemistry to him. Dr. Stanley Smith has also always been a light for me to look up to as well. He was my mentor at the SURE program for 2 years and throughout that time I discovered my fondness for working at the bench. I've kept in contact with him throughout graduate school, and his kind and encouraging words have always kept me going through hard times. I'd also like to thank many previous members of the Ho lab, Crystal Vander Zanden, Melissa Ford, Matthew Scholfield, Anna-Carin Carlsson, and Grace Heaslip in the Peersen lab. They will always be people I admire and look up to as not only friends, but scientific role models as well. Each of them has made a huge impact on my life in the most positive way, and I thank them dearly for that.

I'd like to thank my student advisory committee- Tim Stasevich, Melissa Reynolds, Christopher Snow, and Robert Woody. They too are amazing scientific role models who have pushed me and made me a better scientist.

I would especially like to thank my advisor Shing Ho. His enthusiasm for science in general, and love for teaching is truly contagious. I have enjoyed his hands-off approach as it has made me think through and solve problems as they arise. This alone has made me a more confident

and capable scientist. I am so very grateful that he has let me explore alternate paths of my projects and take risks that may not have panned out – let me indulge in the creative side of science. He has given me so many opportunities that not every graduate student receives, which has been a huge blessing during graduate school. He has truly been a wonderful mentor that I can't even begin to explain how much I have learned from him. Thank you, Shing.

My family has been a huge part of my success. My mom and dad especially have always pushed me to do the best I can at whatever I am doing. Often, I tell my mom about problems at work and though she doesn't understand she always says, "You will figure it out, you always do." And most of the time, she has been right. I am so thankful for my parents continued support throughout my entire life.

Lastly, I would like to thank my amazing husband, Luke Hartje. I am so thankful I met the love of my life during grad school, because it has made one of the most stressful times in my life a little bit easier. It's nice to have a partner who understands what you are going through, so I will be forever grateful to him for always being my support. He is one of the most intelligent and hardest working people I know, and I'm so thankful he always has my back through every aspect of life.

TABLE OF CONTENTS

ABSTRACT……ii

ACKNOWLEDGEMENTS……iii

CHAPTER 1: INTRODUCTION- HYDROGEN AND HALOGEN BONDS ROLE IN BIOMOLECULES……..…….1

REFERENCES……....6

CHAPTER 2: RELATIONSHIPS BETWEEN HYDROGEN AND HALOGEN BONDS IN BIOLOGICAL SYSTEMS…….….8

2.1 Summary…….…...8

2.2 Introduction……8

2.3 A Brief History: Hydrogen Bonds and Halogen Bonds…….9

2.4 Competitive Relationships: HB against XB……15

2.5 Substitutions: XB for HB…….19

2.6 Orthogonal Relationships: XBs perpendicular to HBs……26

2.7 Perspectives……..33

REFERENCES……..37

CHAPTER 3: HYDROGEN BOND ENHANCED HALOGEN BONDS: A SYNERGISTIC INTERACTION IN CHEMISTRY AND BIOCHEMISTRY……...48

3.1 Summary……..48

3.2 Introduction……..49

3.3 Experimental Characterization of the HBeXB……49

3.3.1 HBeXB Increases Anion Binding…….50

3.3.2 HBeXB Increases Enzyme Stability and Function……...54

3.3.3 Survey of HBeXB in Cambridge Structure Database and PDB……...58

3.4 Conclusions and Perspectives……..62

REFERENCES……..66

CHAPTER 4: ENGINEERING HYDROGEN AND HALOGEN BONDS TO ENGINEER A SPECIFIC COILED-COIL SENSING PEPTIDE…….69

4.1 Summary……..……69

4.2 Introduction…….……69

4.2.1 Engineering Protein-Protein Interactions…….…….69

4.2.2 Halogen Bonds in Biomolecules……..……70

4.2.3 Coiled-Coils as Model Engineering Platform……...……71

4.3 Experimental Design……74

4.3.1. Peptide Design and Purification……...74

4.3.2. Crystallization, Data Collection, and Structure Determination……...75

4.3.3. Melting Profiles Determined with Differential Scanning Calorimetry……75

4.3.4 Circular Dichroism Spectroscopy Titration……..77

4.3.5 Quantum Mechanical Calculations……...77

4.4 Results……..78

4.4.1. Crystal Structures of $(GCN4\text{-}N16A)_2$/GCN4-$N16^{ZF}$ Heterotrimers……81

4.4.2. Thermal Melting Parameters from Differential Scanning Calorimetry…...85

4.4.3. Heterotrimer Formation Determined by CD Spectroscopy Titrations…..…89

4.5 Discussion……93

REFERENCES…….……98

CHAPTER 5: HYDROGEN BOND ENHANCED HALOGEN BONDS TO INCREASE YEAST KIX DOMAIN STABILITY……..……...103

5.1 Summary……103

5.2 Introduction……103

5.2.1 KIX Domain……....104

5.2.2 Engineering with HeXBs……106

5.2.3 Pdr1p Binding Partner…….108

5.2.4 Summary of Study……..108

5.3 Experimental Section…….……109

5.3.1 Protein Expression……..109

5.3.2 Protein Purification…….114

5.3.3 Pdr1p-12mer Peptide Purification……...114

5.3.4 Differential Scanning Calorimetry……..115

5.3.5 Circular Dichroism Studies…….116

5.4 Results……117

5.4.1 Protein Expression with Non-Canonical Amino Acids……..117

5.4.2 Circular Dichroism of Apo KIX Proteins……...117

5.4.3 Differential Scanning Calorimetry to Assess KIX Stability……...117

5.4.4. CD Titration of KIX with Binding Partner Pdr1p-12mer……..122

5.5 Discussion and Conclusions…….…….122

REFERENCES……129

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS ……..…131

6.1 Summary……..……..131

6.2. A Complex Interplay Exists Between Hydrogen and Halogen Bonds…….……131

6.3 Hydrogen Bond Enhanced Halogen Bonds Exist in Not Only Biomolecular Systems……..…..132

6.4 Halogen Bonds Can Behave as a Specifying Protein-Protein Interaction……..…132

6.5 Hydrogen Bond Enhanced Halogen Bonds Can Increase Stability in KIX……..133

6.6 Future Directions……...133

APPENDIX I……..…….135

# CHAPTER 1: INTRODUCTION- HYDROGEN AND HALOGEN BONDS ROLE IN BIOMOLECULES

It has long been established that hydrogen bonds (HB) are a crucial part of biology. From the structural biology viewpoint, they are involved in protein folding and tertiary structures, ligand binding and affinity, DNA base recognition, protein-protein interactions, protein-DNA interactions, as well as many others [1]. Their strength depends on many factors including environment, and the respective donor and acceptor atoms involved in the bond [2]. While hydrogen bonds are critical for function in naturally occurring systems, one such bond that is not as naturally abundant, but behaves quite similarly, is the halogen bond (XB).

Halogen bonds date back to the 1800's, but more recently have influenced crystal engineering, metal organic frameworks, and organocatalysis fields [3–5]. They are similar in nature to hydrogen bonds as they share the same electron-rich Lewis base acceptors, and typically form at a distance shorter than the sum of their respective van der Waal radii [6]. However, XB donors must be halogens (Cl, Br, or I). These electronegative atoms can interact with other electronegative atoms (O,N,S) due to the formation of the σ-hole on the halogen. This σ-hole or electropositive crown forms on the halogen as it undergoes σ-bond formation with a substituent group. During σ-bond formation, a valence electron on the halogen gets pulled back for the covalent bond, leaving an anisotropic charge distribution around the halogen. The partially positive σ-hole forms diametrically opposed to the σ-bond, while a negative ring forms perpendicular to this around the belt of the halogen (figure 1.1). As the σ-hole is ultimately what drives XB formation, XBs prefer a linear geometric dependence, leading to higher specificity. The various halogens and choice of substituent group the halogen is covalently bound to leads to an energetically tunable aspect

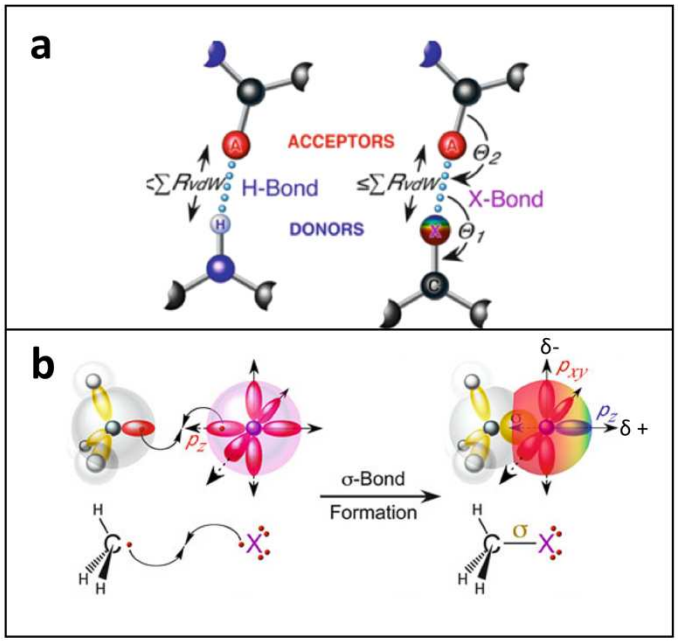



Figure 1.1
Halogen bonds compared to hydrogen bonds and formation of the σ-hole. a) The hydrogen bond (HB) as it relates to the halogen bond (XB). Hydrogens and halogens form interactions with similar electron-rich acceptor atoms (O, N, S). Both of their interaction distances are shorter than the sum of the interacting atoms van der Waals radii ($\Sigma R_{vdW}$), and the bonds have an angular dependence. The XB forms in a linearly dependent manner as dictated by the σ-hole (partially positive region) on the crown of the halogen. b) Formation of the σ-hole on the halogen. As the unpaired valence electron gets pulled back to form the σ-bond, a depopulation of electrons occurs coaxially with the σ-bond, which creates a partially positive region. A partial negative region exists as a belt around the halogen perpendicular to the σ-bond. This anisotropic charge distribution makes the halogen more versatile in terms of potential interactions compared to hydrogen. Figures adapted from Ho (2015) [7].

associated with the interaction. And like the HB, depending on environment and choice of donor and acceptor atoms, halogen bonds can vary in strength. However, they are often stronger than HB's, imparting to their use in molecular and biomolecular engineering efforts. One of the few naturally occurring halogen bonds (XB) is between the human thyroid receptor and 3,5,3-triiodothyroxine (T3) [8]. The short I --- O distance seen in the structure is evidence that XBs can play an important role in terms of ligand recognition.

This notion of recognition, specificity, and stability that ensues in XB potential is what has drawn many medicinal chemists and pharmaceutical companies to utilize this interaction in small-molecule therapeutics. Many of the therapeutics that make it through clinical trials and eventually get launched contain halogens [9,10]. The halogens' hydrophobic nature allows for easier access across lipid bilayers, and the geometrically defined bonds halogens make increase the drug's binding affinity. Both of these aspects lead to more potent therapeutics. As many small-molecule medicines bear halogens, pharmaceutical companies are beginning to think about larger macromolecules (e.g. biologics - proteins and peptides) as therapeutics. Because they are constructed from biological materials, they will have potentially lower toxicity in the cell compared to small molecules, and their larger size/surface area allows for more selective binding and less off-target effects. A few peptide biologics that have made it to the market include calcitonin (32 aa) for osteoporosis and hypercalcemia; teriparatide (34 aa), a parathyroid hormone analog; Fuzeon (36 aa) an antiretroviral; and a 41-aa corticotropin releasing hormone [11]. The market is growing rapidly as the success rate of biologics in the market is about double that of small molecules at this point[11]. As much work still needs to be done in this field, one concept that is lacking and could be hugely beneficial to not only pharmaceutical companies, but also biological engineers, is how halogens and halogen bonds behave in bigger biological molecules. While much

is known about how halogens behave in small-molecule settings, aid in directed organic synthesis, and help assemble supramolecular assemblies and metal organic frameworks [9,12,13], the lack of quantitative information about them in proteins in particular has prevented their widespread use for biological protein engineering efforts (e.g. designing protein interfaces, more stable enzymes to carry out reactions, or constructing multivalent proteins).

From a basic research angle, asking simple questions as to how halogen bonds effect protein structure, function, stability, and specificity will be a beneficial initial characterization for their exploitation in the challenging realm of protein engineering. A detailed understanding of their structure-energy relationship, and especially how the XB compares to its cousin interaction, the HB, will be advantageous for future rational design of biomolecules bearing halogens. With this in mind, we have elected to study XBs in a peptide-based system (30 aa) and a small partially stable protein system (90 aa). These two studies show how XBs can act as specifying and stabilizing interactions in protein contexts.

As mentioned above, the heart of this work is the investigation of the structure-energy relationship, particularly between XBs and HBs engineered into biological systems from an *inter-* and *intramolecular* perspective. We will specifically look at how XBs compare to more traditional HBs as an *intermolecular* protein-protein interaction and driver of specificity and, how the two bonds work together in an *intramolecular* fashion to increase protein stability. To thoroughly do so, the aspects we will be addressing are: (1) How can the HB and XB compete against each other, be substituted for each other, and act orthogonal to each other in biological and non-biological frameworks? (2) How can the HB enhance the strength of the XB (HBeXB) from an intramolecular orthogonal approach? (3) Can we design a sensor probe peptide capable of specifying assembly of

a higher order structure through an engineered HB or XB? (4) Can we stabilize an unstable protein through HBeXB engineering efforts?

We will address the above questions throughout the dissertation in the following manner. Chapter 2 will give a more detailed description of the HB and XB, and through examples from the PDB and literature, the interplay found between the bonds will be dissected as points of interest moving forward in terms of future engineering efforts. Chapter 3 describes the serendipitous discovery of the first two purposefully designed systems bearing the Hydrogen Enhanced Halogen Bond. Quantum mechanical calculations were done to better understand the inductive effects of the hydroxyl group as an *ortho, meta,* or *para* director, and how it can behave as an electron donating or an electron withdrawing group. Chapter 4 discusses the design and structural and thermodynamic characterization of an engineered GCN4 coiled-coil heterotrimer capable of assembling through the specific HB or XB electrostatic interaction. Through crystallography, differential scanning calorimetry, circular dichroism, and quantum mechanical calculations, it delves into the enthalpy-entropy compensation seen between the H and XB and how they compare to less specifying hydrophobic interactions. And finally, Chapter 5 addresses the question of can the HBeXB increase the stability of a partially stable protein without effecting its overall fold or function. Through non-canonical amino acid incorporation of meta-halogenated tyrosines into yeast KIX (Kinase Inducible Domain), the *in vitro* characterization of these engineered proteins is carried out via differential scanning calorimetry and circular dichroism. This *in vitro* study will pave the way for these proteins to be explored *in vivo*.

This dissertation describes the interplay that exists between hydrogen and halogen bonds and how the two primarily electrostatic interactions can be used as a tool for protein engineering efforts.

REFERENCES


1. McClellan, A. L. The significance of hydrogen bonds in biological structures. *J. Chem. Educ.* **44**, 547 (1967).
2. Kollman, P. A. & Allen, L. C. The theory of the hydrogen bond. *Chem. Rev.* **72**, 283–303 (1972).
3. Brammer, L., Bruton, E. A. & Sherwood, P. Understanding the Behavior of Halogens as Hydrogen Bond Acceptors. *Cryst. Growth Des.* **1**, 277–290 (2001).
4. 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**, 6721–6732 (2014).
5. Aakeröy, C. B., Wijethunga, T. K., Benton, J. & Desper, J. Stabilizing volatile liquid chemicals using co-crystallization. *Chem. Commun.* **51**, 2425–2428 (2015).
6. Desiraju, G. R. *et al.* Definition of the halogen bond (IUPAC Recommendations 2013). *Pure Appl. Chem.* **85**, 1711–1713 (2013).
7. Ho, P. S. Biomolecular Halogen Bonds. *Top. Curr. Chem* **358**, 241–276 (2015).
8. 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).
9. 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).

10. Wilcken, R. & Zimmermann, M. Principles and applications of halogen bonding in medicinal chemistry and chemical biology. *J. Med.* (2013).

11. Craik, D. J., Fairlie, D. P., Liras, S. & Price, D. The Future of Peptide-based Drugs. *Chem. Biol. Drug Des.* **81**, 136–147 (2013).

12. Bulfield, D. & Huber, S. M. Halogen Bonding in Organic Synthesis and Organocatalysis. *Chem. - A Eur. J.* **22**, 14434–14450 (2016).

13. Bertani, R. *et al.* Halogen bonding in metal-organic-supramolecular networks. *Coord. Chem. Rev.* **254**, 677–695 (2010).

# CHAPTER 2: RELATIONSHIPS BETWEEN HYDROGEN AND HALOGEN BONDS IN BIOLOGICAL SYSTEMS[1]

## ○ 2.1 Summary

Non-covalent interactions are some of the most essential interactions when it comes to defining how proteins and other macromolecules fold into their tertiary structures. Hydrogen bonds are a key component of this group of interactions. Another non-covalent interaction that has been recently "re-discovered" is the halogen bond. Similar to the hydrogen bond, it is an electrostatic interaction between a partially negative acceptor atom such as O, N, or S in biological systems, and a partially positive halogen atom - Cl, Br, and I instead of H. Due to the nature of this bond, it has the ability to be stronger and more geometrically defined than a traditional H-bond, thus making it potentially useful in biological engineering efforts. While the two bonds are distinct, they overlap in many regards giving way for useful interplay. In this slightly adapted review, we will discuss these two bonds (H and X-bonds) and how they can substitute for one another, compete against each other, and be orthogonal to each other, making for a dynamic interplay.

## ○ 2.2 Introduction

Hydrogen bonds (HBs) are ubiquitous in biology[1,2] as they are the noncovalent interactions responsible for controlling, for example, the assembly of the DNA and RNA double-helices [3-5] and maintaining higher-order protein conformations[6]. Interest in halogen bonds (XBs) has seen a revival in the past couple of decades, as the potential to exploit their unique properties has become better recognized [7-9] XBs are widely applied to the chemical engineering of crystals and supramolecular assemblies, [10-11] including designing liquid crystals [12-15], and

[1] Adapted from published article "Relationships Between Hydrogen Bonds and Halogen Bonds in Biological Systems." by Rowe, R. K.; Ho, P. S (2017)

organometallic frameworks[16], stabilizing volatile liquids [17], synthesizing organic catalysts [18-20] designing anion receptors [21-28] and in host–guest assemblies [29-33]. In biology [34-35] XBs are seen to direct DNA macromolecular conformations[36], to increase the affinity of agonists[37] and antagonists to protein targets [35,38-42], and more recently in rational drug design [23,41-43].

XBs are understood to be analogous and comparable to HBs in terms of their similarities in geometries and energetics [44]; thus, it would be conceptually simple to replace an HB with an XB for applications in both medicinal chemistry and biomolecular engineering. In this review, we explore the relationship between the two molecular interactions, in terms of how they compete against each other, are substituted for one another, and are orthogonal and thus independent of each other. We focus specifically on examples in biological systems and in particular those relationships that are supported by structures from X-ray crystallography. First, however, we start with some background on the individual 'bonding' interactions to provide the reader with some context for these relationships.

- **2.3 A Brief History: Hydrogen Bonds and Halogen Bonds**

Linus Pauling first recognized the significance of the HB in his models for the α-helix and β-sheet [45-46] (Fig. 2.1), describing the interaction as '… an atom of hydrogen is attracted by rather strong forces to two atoms, instead of only one, so that it may be considered to be acting as a bond between them'[47]. The International Union of Pure and Applied Chemistry (IUPAC) published a more modern definition, which has its roots in the Pauling definition, but expands upon it to include a list of criteria by which an HB can be identified [48]. The HB is now recognized as a weak, noncovalent interaction with some degree of directionality that falls between the nondirectional van der Waals interaction and the highly directional covalent bond[49]. The strength of an HB depends on the electronegativity of the HB donor (D) and the

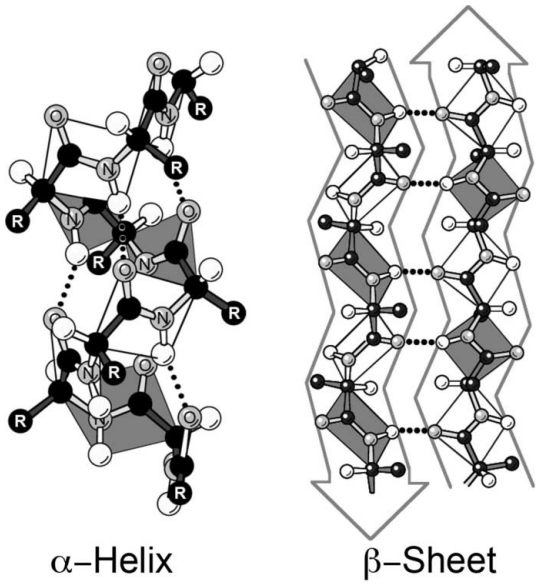



Figure 2.1
Structures of the α-helix and, β-sheet, with intra-strand HBs in the helix and inter-strand HBs in the sheet. Adapted from van Holde et al. (2006) [57].

acceptor (in the D—H. . .A interaction, where D is the heavy atom to which the hydrogen is covalently bound and A is the electron-rich acceptor atom). The resulting broad range of stabilization energies (accepted to be from 2 to 272 kJ $mol^{-1}$ [50]) has led to an artificial distinction between strong HBs (including the Pauling HBs N—H. . .O, N—H. . .N, O—H. . .O, O—H. . .N and O— H. . .$O^{-}$ interactions) and weak HBs (C—H. . .O, C—H. . .N and C—H. . .π). The fundamental physical roots of the interaction are recognized to be primarily electrostatic, with contributions from polarization and dispersion, although the role of covalency, as originally espoused by Pauling [47], is still debated.

In biology, there is no doubt that the HB plays a large role in defining molecular structure and function [1,52]. The hydrophobic effect in protein folding, for example, has been attributed to the loss of entropy in forming strongly hydrogen-bonded water clathrates around alkyl and aromatic C atoms[53,54]. Although the HB was key to Pauling's models for helices in proteins [45,46], the connection was lost in his triple- helical model of DNA [55]. The significance of HBs in conferring specificity in base pairing, however, was not lost on Watson & Crick [56] in their double-helical model of DNA. The complementary nature of Watson–Crick base pairing, and its variants in DNA and RNA, lay the foundation for modern molecular genetics[58] and the technological advances in genome manipulation that followed [59].

The halogen bond traces its roots as far back, if not further, as the HB. Interactions between halogens and electron-rich atoms to form Lewis acid–base complexes were known as early as the 1800s [60,61]. The recognition that such complexes are 'bridging' or 'bonding'-type interactions came from the crystallographic studies of Hassel & Strømme[62], later called 'charge-transfer bonds' [63] as an extension of Mulliken's charge-transfer theory[64,65].

The interaction was seen to be analogous to the HB by Bertrán and Rodríguez [66], leading to the more widely accepted term 'halogen bond' [67,68].

XBs are understood to be primarily an electrostatic interaction between a halogen bond donor and an electron-rich acceptor atom [9,69](O, N, S; Fig. 2.2). As with the HB, however, there remains considerable debate as to whether the XB is entirely electrostatic. High-level computational studies suggest that dispersion[70] and steric effects may be more important than electrostatics in conferring directionality[71]. The role of charge transfer in defining the interaction has been resurrected[72-74] , while several competing electrostatic models have also been put forward [75,76].

Without delving deeply into the merits of all the competing models, perhaps the most accessible explanation for how halogen substituents of covalent compounds interact with electron-rich acceptors is the σ-hole theory (Fig. 2.3), as articulated by Clark, Murray & Politzer [77]. In this model, a so-called 'σ-hole' develops as the valence electron in the half-filled $p_z$ atomic orbital is pulled into a covalent σ-bond between the halogen and bonded partner. The depopulation of the $p_z$ orbital results in a depletion of the electron density, which accounts for the electropositive charge and flattening of the atomic radius diametrically opposed to the σ-bond – this is the σ-hole. However, the $p_{x,y}$ atomic orbitals that are perpendicular to the σ-bond retain their full complement of electrons, leaving an electronegative belt around the halogen waist. Thus, halogen substituents are amphipathic[78-79], potentially serving simultaneously as an XB donor in the direction of the σ-hole and HB or XB acceptor in the perpendicular direction.

The size of the σ-hole increases with the size and polarizability of the halogen; therefore, the strength of XBs follow the series F << Cl < Br < I. Furthermore, the halogen's covalent

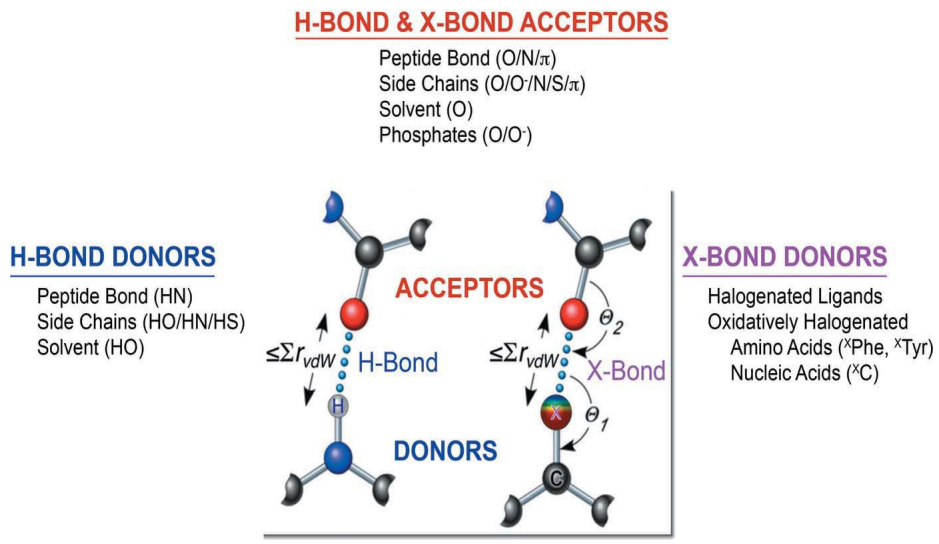



Figure 2.2

Comparison of HBs to XBs. The HB donor is a hydrogen (H) attached to an electronegative atom, which withdraws electron density from the H, allowing an electron-rich acceptor to interact electrostatically at a distance that is shorter than the sum of their respective van der Waals radii ($\Sigma r_{vdW}$). In biological systems, HB donors include any H—N, H—O or H—S group on a protein, nucleic acid or other molecule, including the solvent. XB donors in biology include halogenated ligands, including inhibitors and drugs, along with oxidatively halogenated amino acids and nucleic acid bases. HBs and XBs share a common set of acceptors, formally neutral or anionic O atoms (O or $O^-$), nitrogens, sulfurs and aromatic electron systems ($\pi$). Adapted from Scholfield et al. [35].

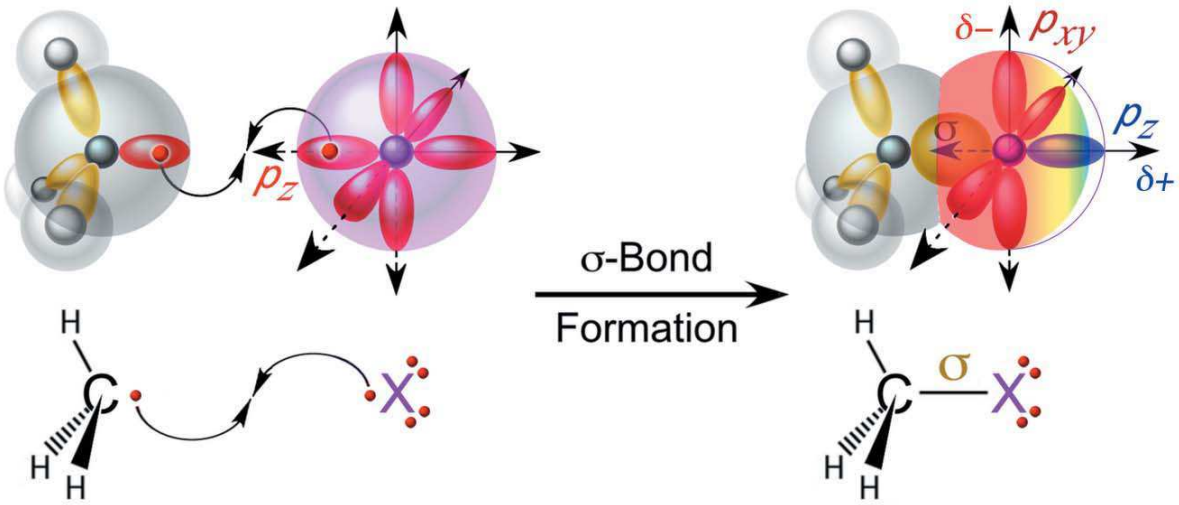



Figure 2.3

The σ-hole model for XB. The valence electron from the pz-atomic orbital of a halogen, when paired with that of a carbon, results in a σ-bond (C—X bond), which depopulates the lobe of the pz-orbital opposite this covalent bond. As a result, the halogen is flattened relative to its standard rvdW, and the charge distribution is anisotropic, with the halogen holding a partial positive charge (δ+) opposite the σ-bond (this is termed the σ-hole) and partial negative charge (δ-) perpendicular to the σ-bond. Adapted from Scholfield, Ford, Vander Zanden et al. [104].

partner also plays a role in determining the size of the σ-hole, with electron-withdrawing groups enhancing and electron-donating groups diminishing the electropositive crown. Taken together, the XB is seen to be highly 'tunable' [35,80-85].

The σ-hole model has been proposed as a unifying concept to explain a broad range of electrostatically driven interactions (including halogen, chalcogen and pnictogen bonds [87-89]). The HB from this perspective is seen as an extreme example, where the σ-hole is manifest as a near-naked proton [90]. This concept thus provides the backdrop for the similarities between XBs and HBs, including commonalities in geometric constraints and acceptors.

When viewed from the perspective of the approach of the acceptor towards the donor (the $\theta_1$ angle), both HBs and XBs are directional, with a tendency of being linear [69] ($\theta_1 = 180^\circ$). XBs, however, as a consequence of the anisotropic distribution of charge, are more strongly directional [67,91-93] and are thus more geometrically constrained than HBs.

In this review, the relationships between HBs and XBs are explored in greater detail from a biomolecular perspective (Fig. 2.4). By competing or substituting an XB for an HB, the geometries and energies of the two interactions are compared and contrasted. Alternatively, the amphipathic nature of halogens and the commonality in acceptors with HBs can result in the interactions being orthogonal to each other from an intra and intermolecular perspective.

- **2.4 Competitive Relationships: HB against XB**

HBs and XBs share identical sets of acceptors, which naturally leads to the two interactions competing against each other, with several studies taking advantage of this competitive relationship to compare the stabilizing potentials of the two interactions. In a small molecule example, Corradi et al. [94] had developed a competition assay in which a bipyridine derivative is crystallized from a solution containing both potential HB and XB donors, and found the XB to

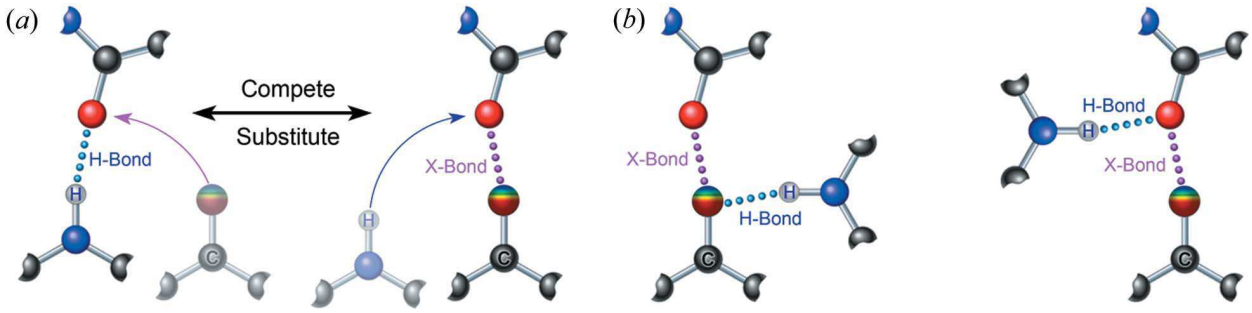



Figure 2.4
Relationships between HBs and XBs in biology. (a) Competing and substitution relationships. (b) Two classes of orthogonal relationships. The relationship on the left takes advantage of the amphipathic nature of the charge distribution of the halogen, allowing it to serve as both an XB donor and HB acceptor. On the right, an XB and HB that share a common acceptor can be orthogonally related to each other.

win out. In a similar study, Aakeröy et al. [95] showed that the two interactions competed in the co-crystallization of 4,4' -azabi- pyridines, with XBs showing distinct differences in donor type (Br versus I), while the various HB donors behaved very similarly to each other.

The only truly analogous crystallographic study that competes an XB directly against an HB in a biological system centers around a model DNA assembly called the Holliday junction, in which XBs have been engineered to 'direct' the conformation of the junction [36,96] (Fig. 2.5). The structure of the Holliday junction in an inverted repeat decanucleotide sequence had been shown to be stabilized by a set of HBs from the cytosine base of a C•G base pair to the phosphate oxygen at the positions where the DNA strands kink and thus cross over to link two near-continuous double-helices [97-100]. In this system, one set of C•G base pairs was replaced by a $^{Br}$U•A base pair (where $^{Br}$U is a 5-bromouracil), with the corresponding HB replaced by an XB [36]. In order to determine whether the XB or HB was the stronger interaction, two complementary strands of DNA (one with the native C•G base pair, and one with the halogenated $^{X}$U•A base pair) were designed to assemble into a four-stranded complex. If the HB is favored over the XB, then the C•G base pair would be seen at the cross-over strand of the resulting junction, and vice versa. By determining the ratio of Br at the crossing and non- crossing strands in the single-crystal structure, the Br XB energy was estimated to be ~17 kJ mol$^{-1}$ more stabilizing than the competing HB in this DNA system.

The work of Carter et al. [96] extended this crystallographic assay to F, Cl and I, and applied differential scanning calorimetry (DSC) to determine the explicit stabilizing potentials in solution[101]. From the crystallographic studies, the ability of XBs to effectively compete against the HB followed the expected series F < Cl < Br < I.

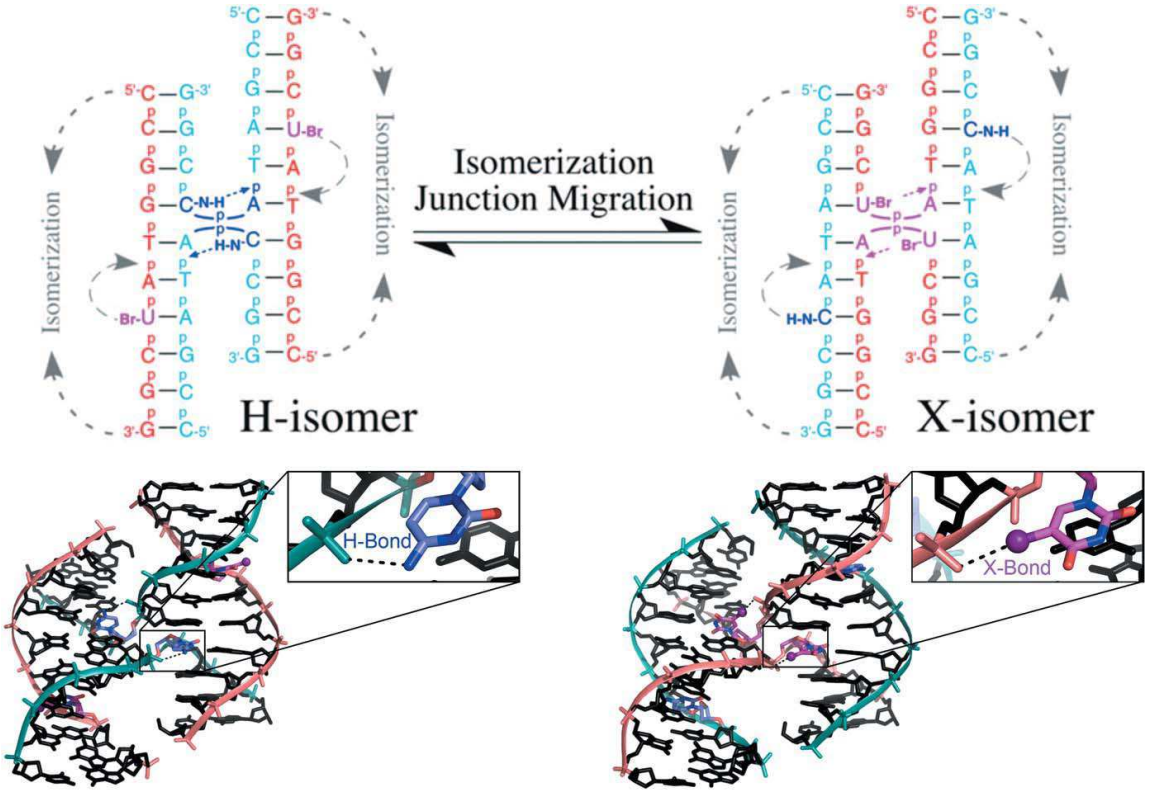



Figure 2.5
Competing HBs and XBs in a four-stranded DNA Holliday junction. A DNA assembly is constructed in which one set of strands (cyan) has cytosines that can form HBs to the DNA backbone to stabilize the junction, paired with a complementary set of strands (red) that places halogenated uracils at the analogous positions that can form XBs to the backbone. The HB stabilized junction is in the H-isomer, which must compete against the XB-stabilized X-isomer, with the isomer that is seen to be dependent on whether the HB or XB is more stabilizing. Adapted from Voth, Hays & Ho [36], PDB code 2org.

In addition, the geometries of interaction become more ideal (shorter interaction distances and more linear $\theta_1$ angles of approach) as the XB becomes more favorable. The series was recapitulated in $\Delta H_M$ from DSC studies, becoming more positive moving from F to I [96]. However, an enthalpy–entropy compensation effect was observed, where the very strong enthalpic iodine XB was also associated with a loss in dynamics of the system and a concomitant loss in $\Delta G$ of stabilization. Consequently, the Br XB was seen to be the most stable interaction overall in the DNA junction system, leading to the obvious conclusion that an enthalpically stable interaction may come at an entropic cost. Thus, when trying to apply XB concepts to molecular engineering, particularly in biological systems, it is important to consider not only the obvious enthalpic effects, but also how introducing such a stabilizing interaction affects the dynamics of the molecular system (including the solvent).

- **2.5 Substitutions: XB for HB**

While the competition assay in the DNA Holliday junction is fairly straightforward and easily interpretable, developing an analogous direct competition between an HB and XB proves to be much more difficult in a protein system. An alternative approach is to determine how replacing an HB donor with a halogen affects structure and function in a protein. In one such study, Kraut et al. [102] asked how replacing the HB with an XB would affect the catalytic function of ketosteroid isomerase (Fig. 2.6). In this enzyme, replacing a Tyr residue (Y16), which forms an HB to the oxyanion hole in the active site, with a non-interacting Phe dramatically reduces the catalytic activity. The expectation was that replacing this Tyr with para-halogenated Phe (F, Cl or Br) would introduce an XB that could restore or possibly enhance the catalytic activity. None of the halogenated constructs, however, had $k_{cat}$ or $k_{cat}/K_M$ values that were comparable to the wild-type. Unfortunately, since the halogenated substrates were models rather

than crystal

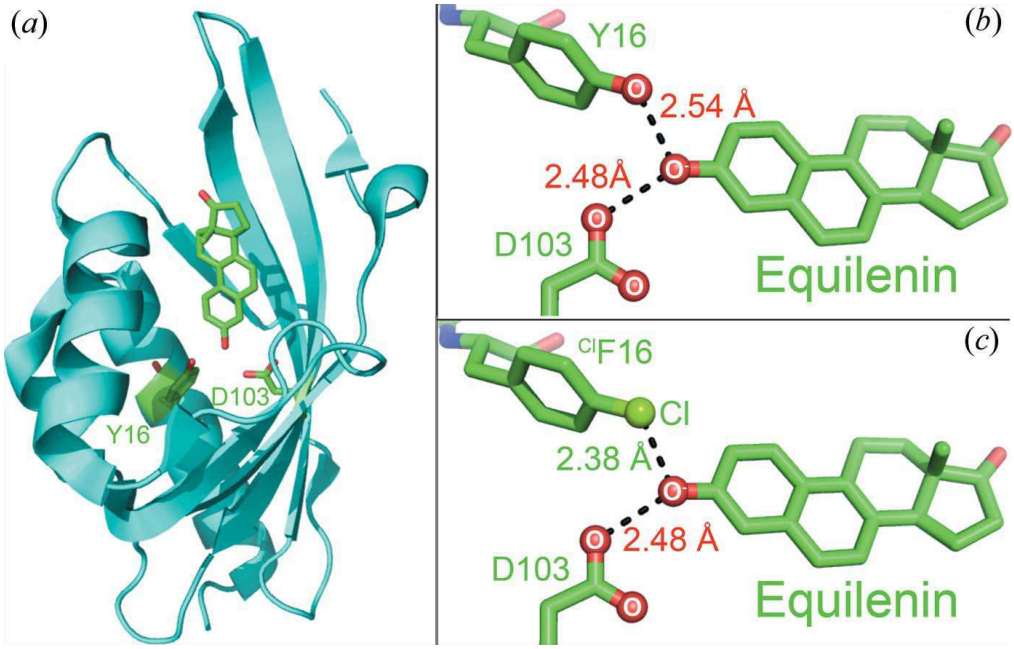



Figure 2.6
Structure of ketosteroid isomerase bound to the estrogenic steroid hormone equilenin (a). (b) Detail of the interactions from Y16 and D103 to the oxyanion hole ($O^-$) of equilenin. (c) Model of the construct in which Y16 is replaced by a chlorinated Phe residue ($^{Cl}F$). The distance from the chlorine to the oxyanion hole was estimated to be ~72% of the ΣrvdW, suggesting the formation of an XB. Data from Kraut et al. [102], PDB code 5g2g.

structures, it is difficult to assess whether any XB was actually formed, leaving the question open as to how comparable an XB is to an HB in this enzyme system.

Our laboratory has started to construct a system to directly determine the effect that replacing an HB with an XB has on protein stability by introducing unnatural halogenated amino acids[103] site specifically into T4 lysozyme [104] (Fig. 2.7a), a classic model for studying effects of non-covalent interactions on protein stability [105]. As with the isomerase studies, we have replaced a tyrosine with para-halogenated phenylalanines, but we have studied the effects of the HB to XB substitution on the structure by X-ray crystallography, and on its thermodynamic stability. In this case, the crystal structures demonstrated that the Br and I constructs formed XBs (Figs. 7b and c). Although these constructs were overall less stable than the wild-type, it was clear that the XB helped to partially rescue the general destabilizing effect that halogenation has on protein stability at a control site that could not form an XB.

Short of substitutions within an intramolecular system, examples from ligand binding[35, 38, 40, 42, 106-109] provide insight into how replacing an HB with an XB affects stability, not in a protein per se, but in a protein– ligand complex. The caveat, of course, is that it is difficult to decipher contributions from single isolated interactions to the overall binding affinity [110]. Well-designed studies, however, could allow a semiqualitative assessment of the stabilizing potential between HBs and XBs. For example, there have been extensive studies on protein kinase inhibitors as potential anticancer drugs, with many inhibitors being ATP analogs that are halogenated and involve XBs for recognition and specificity[106]. An interesting study from Johnson's group [111] showed that XBs substituting for HBs from ATP define the specificity of the inhibitor 5,6-dichlorobenzimidazone-1-,β-d-ribofuranoside (DRB) for Cdk9 over the structurally similar Cdk2.

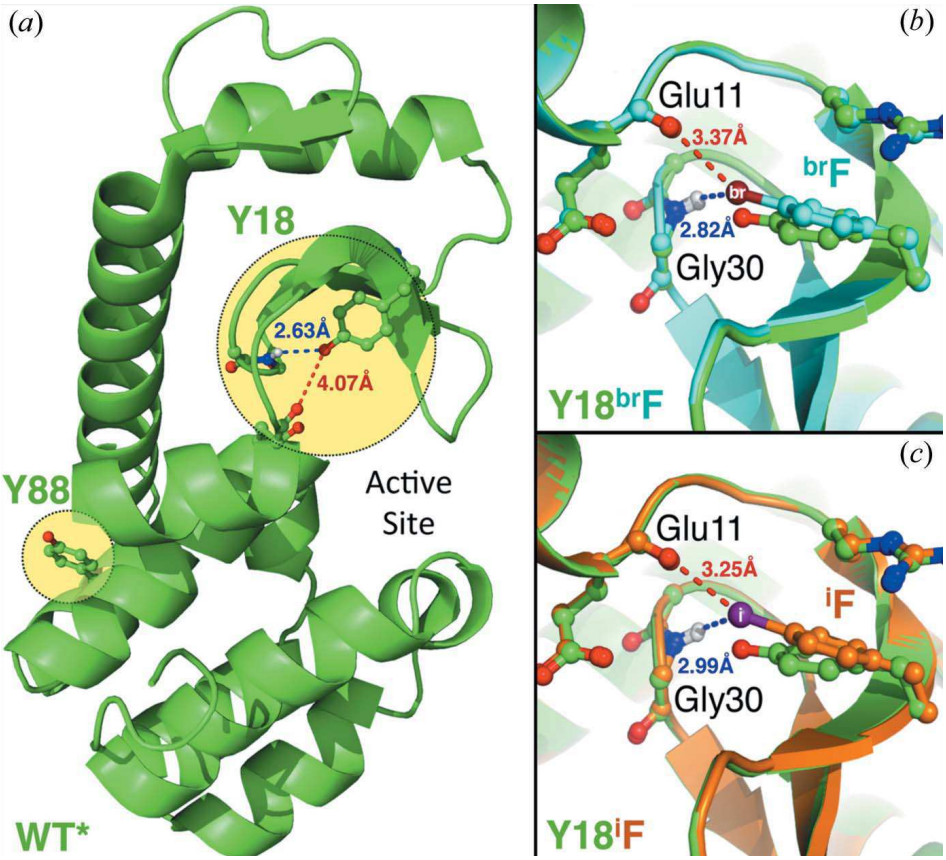



Figure 2.7
XBs engineered into T4 lysozyme. (a) Two aromatic sites (Y18 and Y88) where standard Tyr residues were replaced by bromophenylalanine (brF), (b) or iodophenylalanine (iF), (c). (b) The brF construct at Y18 shows the bromine pulled towards the peptide oxygen to within 100% of the $\Sigma r_{vdW}$. (c) The iF engineered protein at Y18 shows the iodine pulled even closer (93% of $\Sigma r_{vdW}$) to the oxygen to form a stronger XB. Substitutions at Y88 show no distortions to the protein structure, and are highly destabilizing. Adapted from Scholfield et al. [104].

Cdk9 is the kinase subunit of the transcription elongation factor b, while Cdk2 promotes the formation of the nuclear pore complex during cell division; thus, the specificity of DRB is important in defining which cellular pathway and the type of cancer the inhibitor will target. The crystal structure of the Cdk9-inhibitor complex showed that the purine ring of DRB is in the anti-conformation, allowing two XBs from the chlorines to substitute for the two HBs that would normally form with the ATP substrate [112] in the binding pocket (Fig. 2.8). In contrast, DRB in the ATP pocket of Cdk2 has a predominant form in which the chlorinated purine ring is in the syn-conformation, which allows for the formation of only a single Cl XB. This difference in binding modes accounts for the near 300-fold difference in $IC_{50}$ values of the inhibitor against CDK9 versus CDK2.

Fanfrlík et al. [113] took the opposite approach by substituting an amine ($NH_2$, an HB donor) for the XB donor (Br or I) found in an inhibitor against aldose reductase (Fig. 2.9). The crystal structure of this non-halogenated inhibitor in the ligand binding site of the enzyme showed that indeed an HB was formed, as proven by the short donor-to-acceptor distance. A comparison of combined quantum mechanical/ scaled QM/molecular mechanical (QM/SQM/MM) calculations (in the gas phase) on the amine to the two XB inhibitor– enzyme complexes showed that the HB inhibitor had the most negative energy of interaction. However, the amine inhibitor pays a hefty desolvation penalty compared with the halogenated inhibitors, resulting in an overall score and free energy that is least favorable in this comparison. These trends were borne out experimentally, with the iodinated inhibitor having the lowest $IC_{50}$ value. Thus, we must take into account the solvent effects when trying to design inhibitors against their protein targets [114].

Collectively, these studies show that we cannot simply substitute a halogen for an HB donor with the expectation that a stabilizing XB will form in replacement.

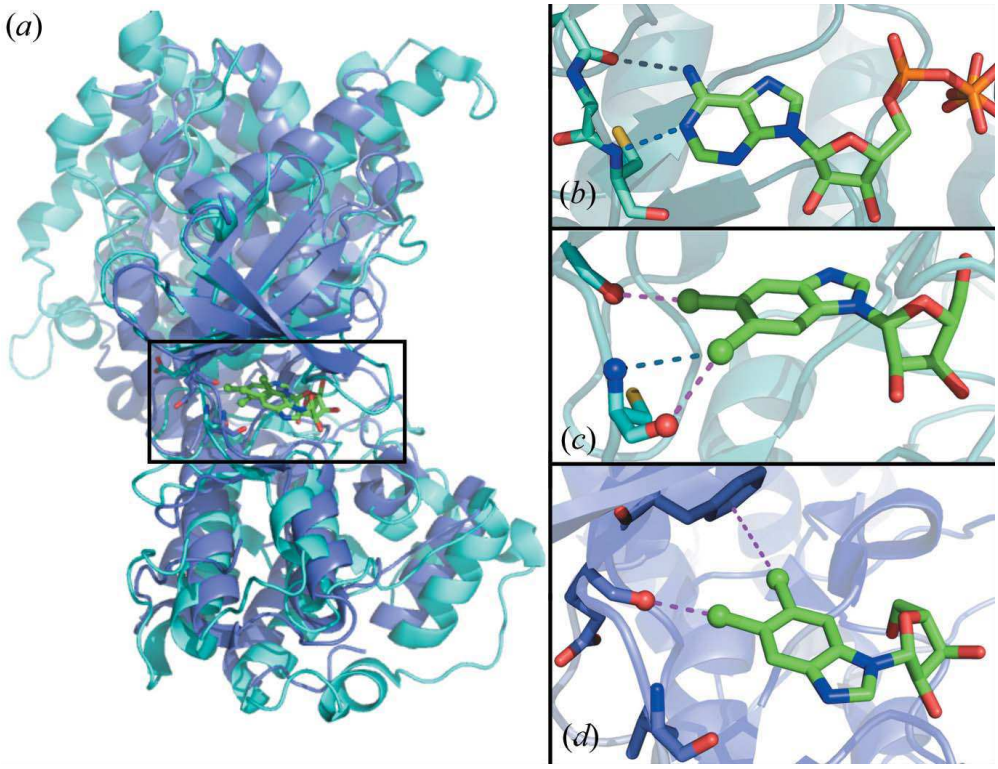



Figure 2.8
The inhibitor complex of Cdk9 or Cdk2. (a) The structure of Cdk9 (traced in cyan, PDB 3my1) with the DRB is overlaid on that of Ckd2 (traced in purple, DPB 3my5) with DRB in their respective ATP binding sites [111]. (b) Structural details of Cdk9 bound to ATP (in the anti-conformation) show hydrogen bonds (blue dots) between the adenine base and the protein backbone in the hinge region of the kinase [112] (PDB code 3blq). (c) Chlorines (green spheres) of the DRB inhibitor (in the anti-conformation) form two XBs (magenta dots) and an HB to the peptide backbone of Cdk9. (d) The major conformation of the DRB inhibitor is rotated to the syn-conformation, allowing XBs (magenta dots) only to form one XB to the peptide bond, and an XB exists with the aromatic ring of a phenylalanine side chain.

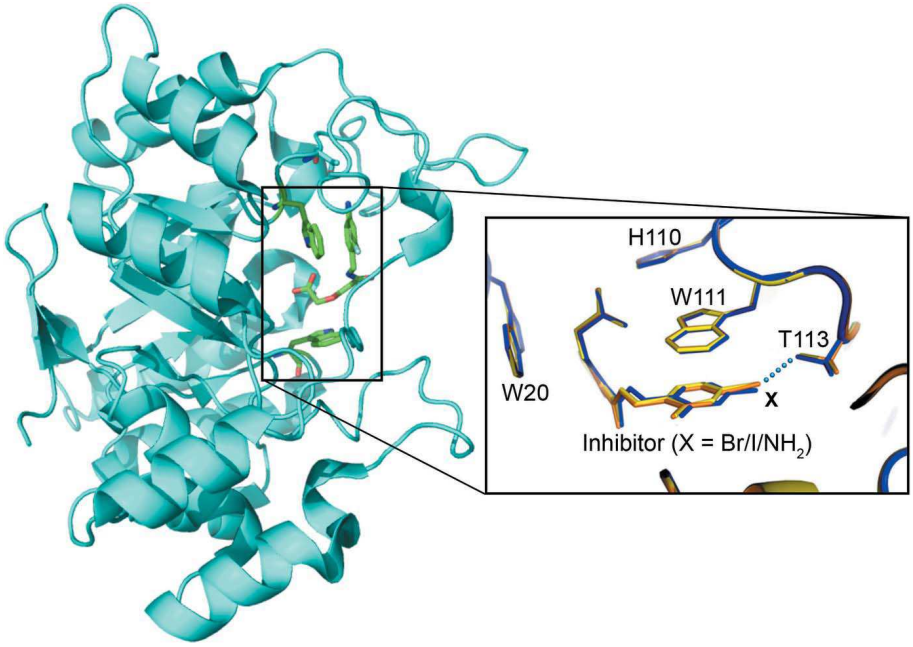



Figure 2.9
Crystal and modeled QM/SQM/MM structures of aldose reductase in complexes with the inhibitor {2-[(4-amino-2 fluorobenzyl)carbamoyl]-5-chlorophenoxy}acetic acid. The inset shows the crystal structures of the inhibitor, with the substituent X = Br (yellow), I (orange) or NH2 (blue). The QM/SQM/MM modeled structure aligned well with the respective crystal structures, and both indicate that replacing the halogen with NH2 did not alter the binding mode. Adapted from Fanfrlík et al. [113], PDB code 4qxi.

As with the competition assays, any efforts to rationally engineer such substitutions must be aware of the more restrictive geometry constraints for the XB and, therefore, be ready to accept that what appears initially to be a logical replacement may end with less than satisfying results. When trying to substitute an HB with a halogen on a ligand such as an inhibitor, we must also take into account solvation effects on the unbound and bound molecules – recognizing that halogens are generally hydrophobic atoms [115].

### ○ 2.6 Orthogonal Relationships: XBs perpendicular to HBs

Orthogonality as a concept extends to both geometric as well as functional properties when considering HBs and XBs. We will see that the geometric definition – that the interactions can be perpendicular to each other – will be evident from two properties of halogens: (i) the amphipathic nature of the halogen and (ii) the ability to share a common electronegative acceptor with an HB donor.

Recall that the charge across the surface of a halogen substituent is anisotropically distributed [35], with a partial positive charge opposite and an electro- negative belt perpendicular to the covalent σ-bond (Fig. 2.3), allowing it to act as both an XB donor and an HB acceptor (Fig. 2.4). This charge distribution allows for the halogen to partake in two simultaneous interactions. We have seen intermolecular X-bonds through the σ-hole and intermolecular H-bonds to the negative belt. More recently though, we have seen *intramolecular* H-bonds. That is, an H-bond donor on the same molecule forming an H-Bond to the halogen and increasing the size of the σ-hole on the halogen through polarization[116,117]. This unique interaction we note to be a Hydrogen Enhanced Halogen bond (HBeXB Bond), and will be reviewed in more detail in Chapter 3.

Lu, Wang et al. [118] surveyed the Protein Data Bank (PDB [119]) and found that the most frequent HBs to halogens (X = Cl, Br or I) in crystal structures of protein–inhibitor complexes were from C—H donors, followed by N—H, then O—H. As expected, the most probable HB approach to the halogen was ~100$^{o}$ (close to the 90$^{o}$ predicted from charge distributions) and the H. . .X distances shortened when going from weaker to stronger HB donors (C—H. . .X > N—H. . .X > O—H. . .X in terms of the H. . .X distance). The trends in the interaction distances, as reflected in the HB energies, were consistent with quantum chemical calculations (at the MP2 level) on model HB complexes between halogenated benzene and $CH_4$, $NH_3$ or $H_2O$ donors.

Lu, Wang et al. [118] found only three PDB entries (one crystal structure and two NMR models) with a halogen serving as both HB acceptor and XB donor. The one crystal structure (PDB 1gjd) was of a trypsin-like serine protease in a complex with an iodinated analog of APC-8696, an amidine inhibitor [120]. The authors had shown that replacing a single hydrogen with a halogen increased the specificity of APC-8696 by up to 220-fold for the class of proteases with a Ser at position 190 (which includes the urokinase-type plasminogen activator and factor VIIa) over a similar class of inhibitors with an alanine in place of Ser190. The structure with the iodo-analog shows an XB to the carbonyl oxygen of Val41 (Fig. 2.10), but also two sets of C—H. . .I HBs (from the C$\alpha$ of G193 and C$\beta$ of S195) orthogonal to the XB [118]. In addition, the amino NH of G193 and the OH of Ser195 are also within HB distance to the iodine, thus potentially expanding the sphere of interactions that are orthogonal to the XB.

The ability of an electron-rich atom to serve as an acceptor for both HBs and XBs sets up a different class of orthogonal relationship, in which simultaneous bonds to a common acceptor are geometrically perpendicular and thermodynamically independent of each other (Fig. 2.4). The

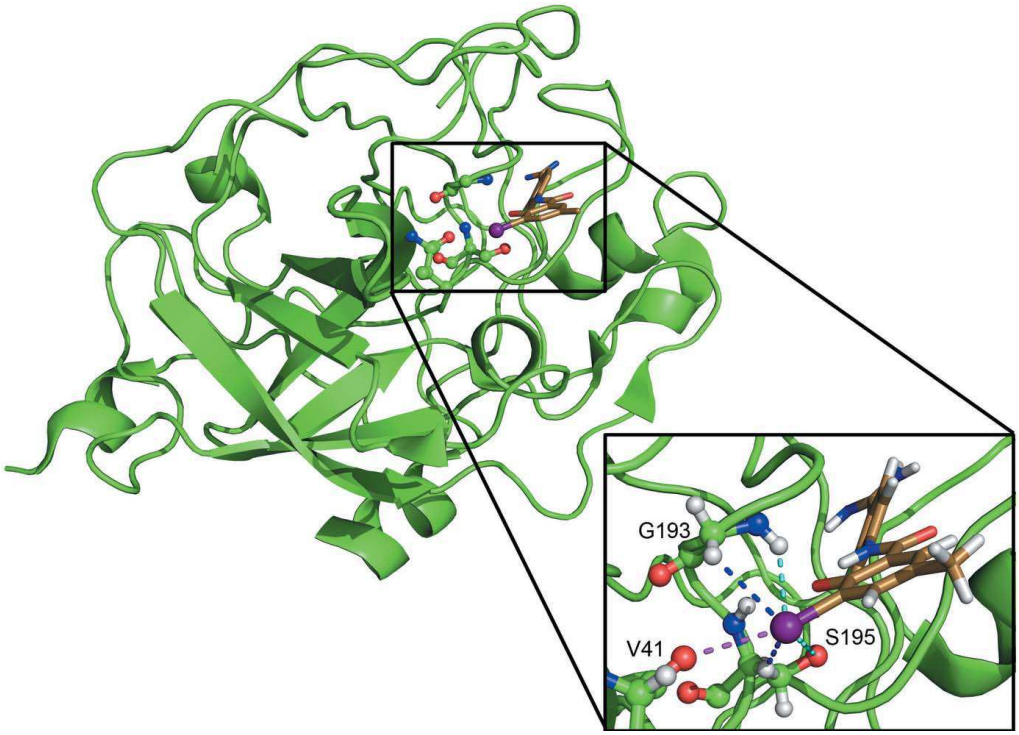



Figure 2.10
Structure of the iodinated APC-8696 inhibitor in complexes with trypsin, a Ser190 trypsin-like serine protease[120]. The inset shows the iodine of the inhibitor forming an XB (magenta dashes) to the backbone oxygen of Val41, along with two HBs (blue dashes) from Gly193 and Ser195 that were identified by Lu, Wang et al. [118]. In addition, the iodine is within HB distance of the backbone amino group of Gly193 and the hydroxyl group of Ser195 (cyan dashes).

evidence for this relationship came initially from a survey of the PDB by Voth et al. [121], which showed that an XB from a halogen to an acceptor (typically the halogen of an inhibitor to the carbonyl oxygen of the polypeptide backbone) is geometrically perpendicular to and HB that shares this same acceptor (with an average —X…O…H— angle of ~88°; Fig. 2.11a). An analysis of the secondary structures (α-helices and β-sheet conformations) shows that this geometric orthogonality is inherent in the relationship between the two bond types. The HBs between strands in a β-sheet are approximately aligned along the peptide planes, placing the shared XB perpendicular to the plane and the β-strands. In contrast, the HBs within an α-helix are aligned ~40° out of the peptide plane. Consequently, XBs approach the peptide planes in an α-helix at an angle of ~50° to maintain the orthogonal geometry. In addition to their geometric relationship, the two interactions were shown to be energetically orthogonal. Quantum chemical calculations on a set of model compounds that mimic interactions from halogenated ligands to the hydrogen-bonded peptide backbone indicated that adding or removing an XB to the carbonyl oxygen acceptor does not affect the energy of the HB peptides (Figs. 11b and c). Thus, by analyzing the HB pattern in a ligand binding site of a protein, we can predict the optimum geometry of a stabilizing XB and that adding the XB will not significantly disrupt the structural integrity of the folded protein.

Vasylyeva et al. [122] applied the orthogonality concept to drive the self-assembly of N-methylacetamide (NMA, a peptide backbone mimic) with a series of halogenated aromatic compounds (as XB donors), with the goal of developing the concept for future crystal engineering projects. In such cocrystals, the NMA molecules packed in a way that mimicked the polypeptide backbone in β-sheets, with HB donor/acceptor pairs aligned in the peptide plane. The dihalotetrafluorobenzene molecules were oriented approximately perpendicular to the NMA

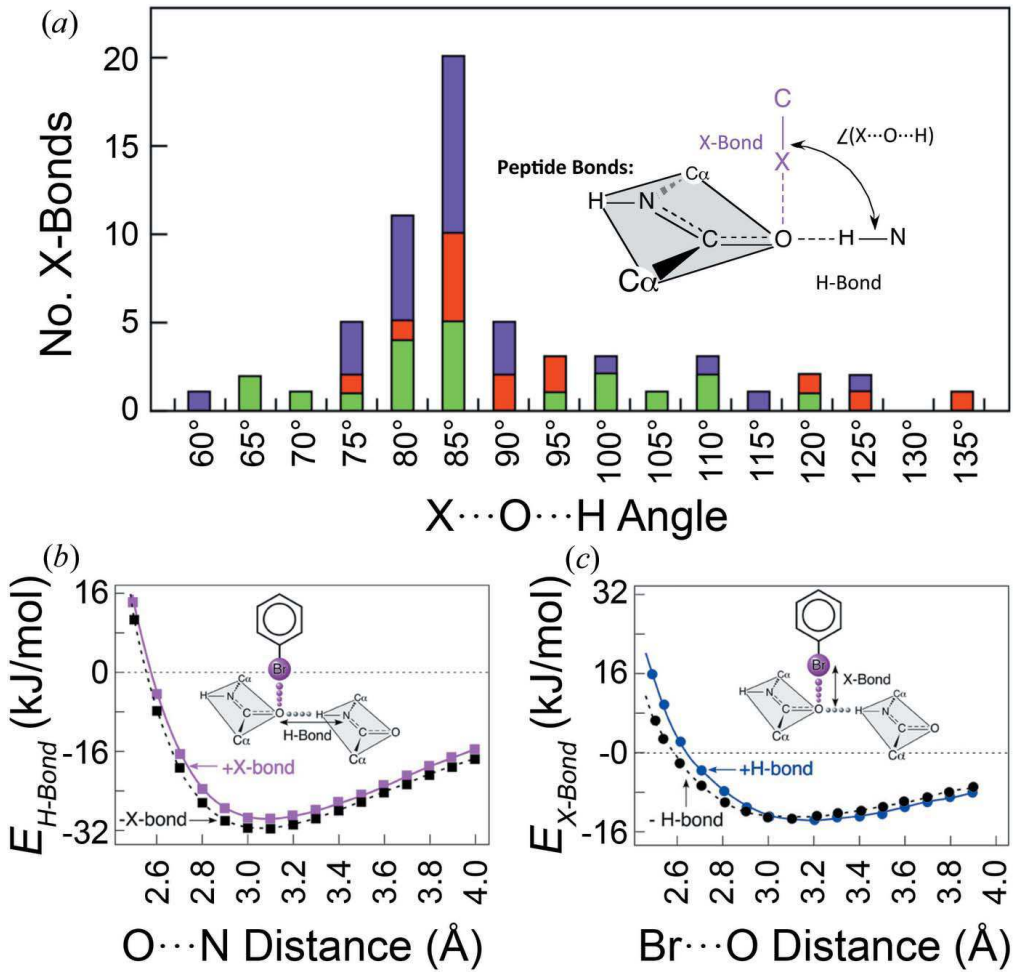



Figure 2.11
Orthogonal HBs and XBs that share a common acceptor. (a) Distribution of X---O-H angles seen in the PDB (Cl in green bars, Br in red bars and I in purple bars). (b) Energy of HB peptide bonds (EHB) in the presence (magenta squares) or absence (black squares) of an orthogonal XB. (c) Energy of an XB (EXB) in the presence (blue circles) or absence (black circles) of an orthogonal HB. Adapted from Voth et al. [121].

sheets, with their halogens halogen-bonding to the carbonyl O atoms. As expected, the —X. . .O… H— angles varied between 77.9 and 98.5$^{o}$, in agreement with Voth et al. [121].

The orthogonality concept has now been seen to extend to assemblies of non-protein/peptide molecules as well. Takemura et al. [123] observed orthogonal XBs and HBs in co- crystals of 1,4-diiodotetrafluorobenzene with acetamide or with N-methylbenzamide. In these crystals, the respective amides form HB sheets, with the opposing iodines of each 1,4-diiodotetrafluorobenzene forming XBs that bridge the carbonyl O atoms across two neighboring amide sheets. The average —X. . .O. . .H— angle in each case (~80$^{o}$) is geometrically orthogonal.

While much has been carried out on the halogen atom acting as a site for two intermolecular interactions, our lab and the Berryman lab simultaneously uncovered the ability for an intramolecular H-bond to form to the halogen. As the H-bond forms to the negative belt around the halogen, an increase in the σ-hole size is noted (Fig. 2.12). This enhancement is thought to come from the polarization of the electron density by the H-bond toward the $p_{XY}$ oribitals. We term this variation of an orthogonal X-bond the Hydrogen Enhanced Halogen bond (HBeXB bond). In engineering meta-halogenated tyrosine into T4 lysozyme, Carlsson et al [117] found through DSC, crystallography, and quantum mechanical calculations that the hydroxyl group on the tyrosine (when oriented toward the halogen), could increase the thermal stability, enthalpy, and activity of the enzyme. Due to the pre-existing structural rigidity of the protein, only meta-chloro-tyrosine formed a stabilizing HBeXB bond. Riel et al. [116] was working on an organic small molecule example of this at the same time. They were able to increase the size of the σ-hole of two iodines on a planar aromatic compound with the presence of an $NH_2$ group.

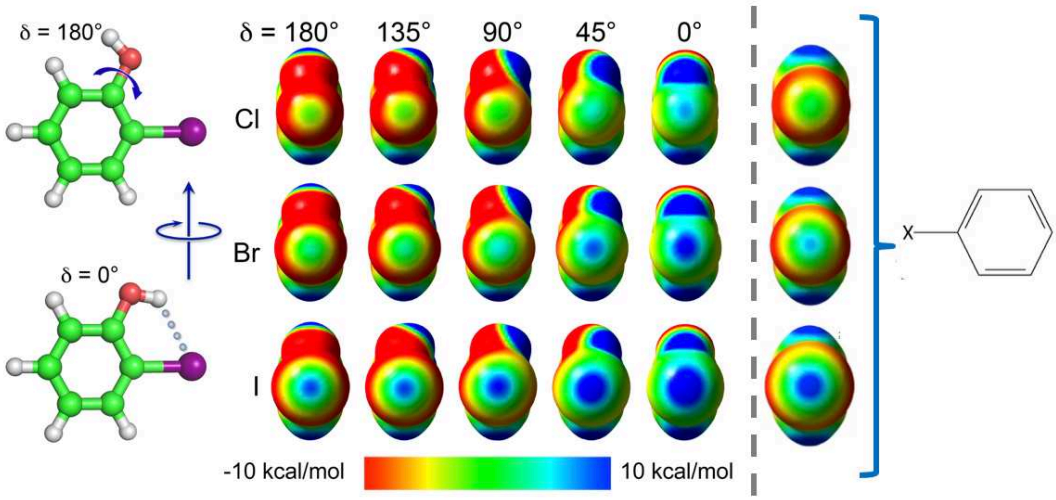



Figure 2.12
Electrostatic potential maps of 2-chlorophenol, 2-bromophenol, and 2-iodophenol showing how the size of the σ-hole increase as the hydrogen rotates from the trans non-HB orientation to the cis HB orientation. ESP of halobenzene molecules are to the right for σ-hole size comparison.

This bidentate compound was able to bind halides at a much higher affinity compared to a similarly designed compound with hydrogens instead of iodines. Herein is another example of a H for X substation that increases binding affinity, while also demonstrating the utility and power of the HBeXB bond that has only recently been identified.

The X—H orthogonality concept is now finding utility in the design of inhibitors against therapeutic targets. In one interesting example, Huang et al. [124] designed a peptide inhibitor against the oncoprotein encoded by the mouse double minute-2 (MDM2) gene (Fig. 2.13), a ubiquitin ligase E3 specific for p53 tumor suppressor protein. The authors based their design on a 12-amino acid peptide from p53, which included metabromophenylalanine at the third residue. Quantum chemical and QM/MM hybrid calculations suggested that the halogenated residue forms an XB to the peptide oxygen of Gly58 of the MDM2 protein, which is further perpendicular to an HB from the amino group of the Met62 peptide bond. The orthogonal interaction was estimated to increase the affinity of the halogenated ligand by fourfold. Thus, the orthogonal XB/HB serves as a starting point in designing new halogenated inhibitors against clinically important targets [109].

- **2.7 Perspectives**

The HB is ubiquitous in biology[1,2], responsible for defining the conformational details of proteins and nucleic acids, the interactions that determine the specificity between the two, and the affinity of various ligands for each. With the growing recognition that XBs are important in biology [34, 35, 108] particularly in agonists and antagonists that bind clinically important targets, the relationships between the two interactions are seen to be strongly related to each other, often times in relationships that are more complex than the simple competition that we would expect for such similar interactions. In this review, we have shown that XBs can substitute for,

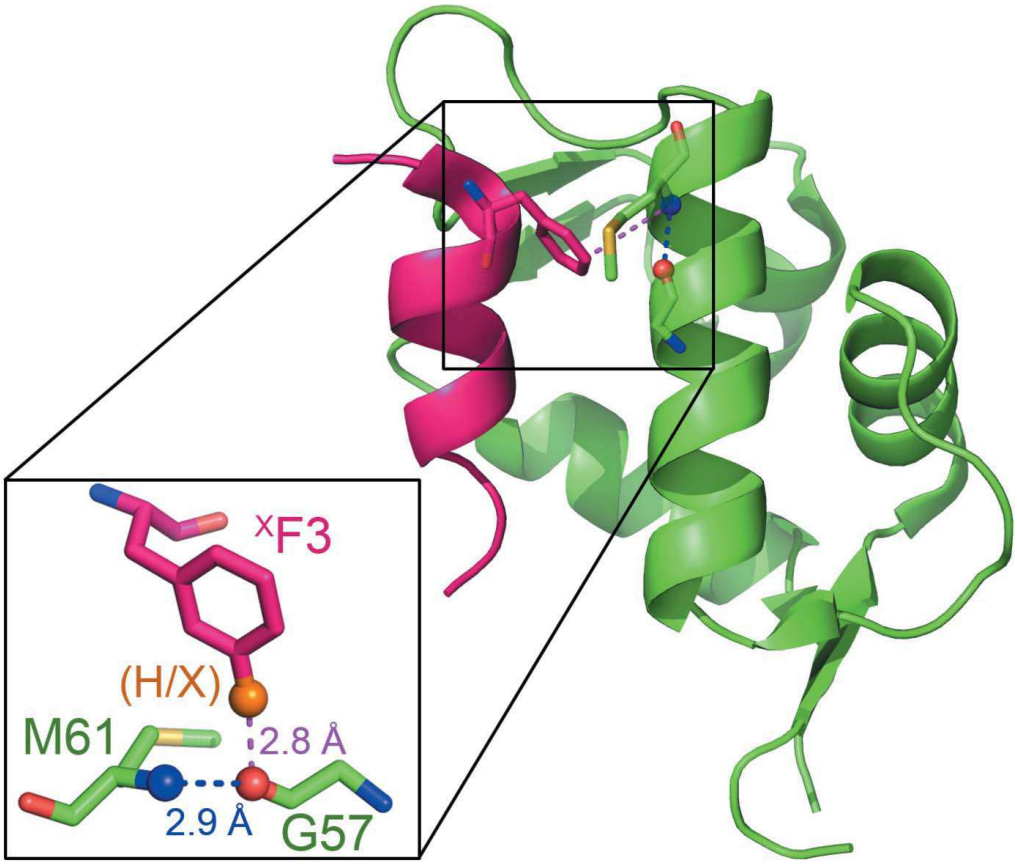



Figure 2.13
MDM2 in complexes with a peptide derived from the p53 tumor suppressor. AC—H ----O HB is seen from the aromatic side chain of F3 to the peptide oxygen of G57. An orthogonal HB to this same oxygen originates from the peptide amide of M61. When the hydrogen at position 3 of F3 is replaced by a halogen (H/X), an energetically independent orthogonal molecular interaction could form to help stabilize the peptide–MDM2 complex. Adapted from Huang et al. [124] , PDB code 3jzo.

compete against, or be orthogonal to HBs. However, these interactions and their respective contributions to molecular structure have primarily been seen in hindsight. In order to take advantage of these relationships in medicinal chemistry or biomolecular engineering, the individual interactions and their more complex relationships need to be more accurately modeled and thus be rationally designed. Quantum mechanical (QM) calculations can now accurately model complex interactions, including XBs and HBs; however, most biological systems remain recalcitrant to high-level QM methods[109]. Molecular mechanics approaches to modeling the XB are being developed and refined [86, 125-130], but the remaining challenge is in how to treat the competitive and orthogonal relationships.

Throughout this review, water is seen to play an important role in affecting relationships between HBs and XBs. In addition to their electrostatic interactions, halogens are known to be hydrophobic. Introducing halogens, either in replacement of or in addition to OH or NH groups, will therefore change the solvation pattern around a protein or nucleic acid structure, and these must be taken into account when trying to understand the energetics of XBs in biology. Furthermore, it has been shown that waters can mediate between halogens and both HB donor and acceptor groups in protein–ligand complexes [131].

As the molecular interactions seen in crystal structures of proteins and nucleic acids become better studied and methods to identify such interactions become better developed [132-136] we will invariably find additional, perhaps more complex relationships that may not fit neatly into any of the ones discussed here. The substitution of a hydroxyl or amine group with a halogen replaces not only an HB donor with an XB donor, but the HB acceptor functions are also affected. Although a halogen can also serve an HB acceptor, the angle of approach of the donor differs among each of these substituents. XBs were found to be important in biology

not from any preconception that they should be, but from finding that what was already present in the PDB [119] could not be explained from a biologist's simple understanding of halogens and their chemical properties. In short, there remains much to be learned about molecular interactions from single-crystal structures, not only about how proteins and nucleic acids function in the cell, but also about fundamental chemical properties that may be surprising not only to the biologist, but perhaps even to the chemist.

REFERENCES


1. Baker, E. N. (2001). International Tables for Crystallography, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 546–552. Dordrecht: Kluwer Academic Publishers.
2. Kuriyan, J., Konforti, B. & Wemmer, D. E. (2013). The Molecules of Life. New York: Garland Science.
3. Saenger, W. (1984). Principles of Nucleic Acid Structure. New York: Springer-Verlag.
4. Neidle, S. (1999). Oxford Handbook of Nucleic Acid Structure. Oxford University Press
5. Carter, M. & Ho, P. S. (2011). DNA Replication – Current Advances, edited by H. Seligmann, pp. 3–28. Rijecka, Croatia: InTech.
6. Fasman, G. (1989). Prediction of Protein Structure and the Principles of Protein Conformation. New York: Plenum Press.
7. Erdélyi, M. (2012). Chem. Soc. Rev. 41, 3547–3557.
8. Gilday, L. C., Robinson, S. W., Barendt, T. A., Langton, M. J., Mullaney, B. R. & Beer, P. D. (2015). Chem. Rev. 115, 7118–7195.
9. Cavallo, G., Metrangolo, P., Milani, R., Pilati, T., Priimagi, A., Resnati, G. & Terraneo, G. (2016). Chem. Rev. 116, 2478–2601.
10. Metrangolo, P. & Resnati, G. (2001). Chemistry, 7, 2511–2519.
11. Metrangolo, P., Neukirch, H., Pilati, T. & Resnati, G. (2005). Acc. Chem. Res. 38, 386–395.
12. Nguyen, H. L., Horton, P. N., Hursthouse, M. B., Legon, A. C. & Bruce, D. W. (2004). J. Am. Chem. Soc. 126, 16–17.

13. Bruce, D. W. (2008). Struct. Bond. 126, 161–180.

14. Chen, Y. J., Yu, H. F., Zhang, L. Y., Yang, H. A. & Lu, Y. F. (2014). Chem. Commun. 50, 9647–9649.

15. Fernandez-Palacio, F., Poutanen, M., Saccone, M., Siiskonen, A., Terraneo, G., Resnati, G., Ikkala, O., Metrangolo, P. & Priimagi, A. (2016). Chem. Mater. 28, 8314–8321.

16. Maharramov, A. M., Shixaliyev, N. Q., Gurbanov, A. V., Mahmudov, K. T., Nenajdenko, V. G., Pombeiro, A. J. L. & Kopylovich, M. N. (2016). Non-covalent Interactions in the Synthesis and Design of New Compounds, edited by A. M. Maharramov, K. T. Mahmudov, M. N. Kopylovich & A. J. L. Pombeiro. Hoboken, NJ: John Wiley and Sons, Inc

17. Aakeröy, C. B., Wijethunga, T. K., Benton, J. & Desper, J. (2015). Chem. Commun. 51, 2425–2428.

18. Coulembier, O., Meyer, F. & Dubois, P. (2010). Polym. Chem. 1, 434–437.

19. Jungbauer, S. H. & Huber, S. M. (2015). J. Am. Chem. Soc. 137,12110–12120.

20. Matsuzawa, A., Takeuchi, S. & Sugita, K. (2016). Chem. Asian J. 11, 2863–2866.

21. Sarwar, M. G., Dragisic, B., Sagoo, S. & Taylor, M. S. (2010). Angew. Chem. Int. Ed. 49, 1674–1677..

22. Kilah, N. L., Wise, M. D. & Beer, P. D. (2011). Cryst. Growth Des. 11,4565–4571.

23. Lu, Y. X., Li, H. Y., Zhu, X., Liu, H. L. & Zhu, W. L. (2012). J. Mol. Model. 18, 3311–3320.

24. Tepper, R., Schulze, B., Jäger, M., Friebe, C., Scharf, D. H., Görls, H. & Schubert, U. S. (2015). J. Org. Chem. 80, 3139–3150.

25. Amendola, V., Bergamaschi, G., Boiocchi, M., Fusco, N., La Rocca, M. V., Linati, L., Lo Presti, E., Mella, M., Metrangolo, P. & Miljkovic, A. (2016). RSC Adv. 6, 67540–67549.

26. Wageling, N. B., Neuhaus, G. F., Rose, A. M., Decato, D. A. & Berryman, O. B. (2016). Supramol. Chem. 28, 665–672.

27. Mullaney, B. R., Cunningham, M. J., Davis, J. J. & Beer, P. D. (2016). Polyhedron, 116, 20–25.

28. Hoque, M. N. & Das, G. (2016). J. Mol. Struct. 1108, 298–306.

29. Dumele, O., Trapp, N. & Diederich, F. (2015). Angew. Chem. Int. Ed. 54, 12339–12344.

30. Amombo Noa, F. M., Bourne, S. A. & Nassimbeni, L. R. (2015). Cryst. Growth Des. 15, 3271–3279.

31. Amombo Noa, F. M., Bourne, S. A., Su, H., Weber, E. & Nassimbeni, L. R. (2016). Cryst. Growth Des. 16, 4765–4771.

32. González, L., Zapata, F., Caballero, A., Molina, P., Ramírez de Arellano, C., Alkorta, I. & Elguero, J. (2016). Chem. Eur. J. 22, 7533–7544.

33. Puttreddy, R., Beyeh, N. K. & Rissanen, K. (2016). CrystEngComm, 18, 793–799.

34. Auffinger, P., Hays, F. A., Westhof, E. & Ho, P. S. (2004). Proc. Natl Acad. Sci. USA, 101, 16789–16794.

35. Scholfield, M. R., Vander Zanden, C. M., Carter, M. & Ho, P. S. (2013). Protein Sci. 22, 139–152.

36. Voth, A. R., Hays, F. A. & Ho, P. S. (2007). Proc. Natl Acad. Sci. USA, 104, 6188–6193.

37. Rohde, L. A. H., Ahring, P. K., Jensen, M. L., Nielsen, E. O., Peters, D., Helgstrand, C., Krintel, C., Harpsoe, K., Gajhede, M., Kastrup, J. S. & Balle, T. (2012). J. Biol. Chem. 287, 4248–4259.

38. Lam, P. Y. S., Clark, C. G., Smallwood, A. M. & Alexander, R. S. (2009). Abstracts of the 238th Am. Chem. Soc. National Meeting, edited by P. Metrangolo & G. Resnati, p. 58. Washington, DC: American Chemical Society.

39. Lu, Y., Shi, T., Wang, Y., Yang, H., Yan, X., Luo, X., Jiang, H. & Zhu, W. (2009). J. Med. Chem. 52, 2854–2862.

40. Hardegger, L. A., Kuhn, B., Spinnler, B., Anselm, L., Ecabert, R., Stihle, M., Gsell, B., Thoma, R., Diez, J., Benz, J., Plancher, J. M., Hartmann, G., Banner, D. W., Haap, W. & Diederich, F. (2011). Angew. Chem. Int. Ed. 50, 314–318.

41. Xu, Z., Yang, Z., Liu, Y., Lu, Y., Chen, K. & Zhu, W. (2014). J. Chem. Inf. Model. 54, 69–78.

42. Wilcken, R., Zimmermann, M. O., Lange, A., Joerger, A. C. & Boeckler, F. M. (2013). J. Med. Chem. 56, 1363–1388.

43. Sirimulla, S., Bailey, J. B., Vegesna, R. & Narayan, M. (2013). J. Chem. Inf. Model. 53, 2781–2791.

44. Metrangolo, P. & Resnati, G. (2008). Science, 321, 918–919.

45. Pauling, L., Corey, R. B. & Branson, H. R. (1951). Proc. Natl Acad. Sci. USA, 37, 205–211.

46. Eisenberg, D. (2003). Proc. Natl Acad. Sci. USA, 100, 11207–11210.

47. Pauling, L. (1960). Nature of the Chemical Bond. An Introduction to Modern Structural Chemistry, 3d ed. Ithaca: Cornell University Press.

48. Arunan, E., Desiraju, G. R., Klein, R. A., Sadlej, J., Scheiner, S., Alkorta, I., Clary, D. C., Crabtree, R. H., Dannenberg, J. J., Hobza, P., Kjaergaard, H. G., Legon, A. C., Mennucci, B. & Nesbitt, D. J. (2011). Pure Appl. Chem. 83, 1637–1641.

49. Desiraju, G. R. (2011). Angew. Chem. Int. Ed. 50, 52–59.

50. Głowacki, E. D., Irimia-Vladu, M., Bauer, S. & Sariciftci, N. S. (2013). J. Mater. Chem. B, 1, 3742–3753.

51. Grabowski, S. J. (2011). Chem. Rev. 111, 2597–2625.

52. McClellan, A. L. (1967). J. Chem. Educ. 44, 547–551.

53. Tanford, C. (1978). Science, 200, 1012–1018.

54. Tanford, C. (1979). Proc. Natl Acad. Sci. USA, 76, 4175–4176.

55. Pauling, L. & Corey, R. B. (1953). Proc. Natl Acad. Sci. USA, 39, 84–97.

56. Watson, J. D. & Crick, F. H. (1953). Cold Harb. Symp. Quant. Biol. 18, 123–131

57. Holde, K. van, Johnson, W. C. & Ho, P. S. (2006). Principles of Physical Biochemistry, 2nd Ed. New Jersey: Pearson Prentice Hall.

58. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K. & Walter, P. (2015). Molecular Biology of the Cell, 6th ed., pp. 1–1342. New York: Garland Science.

59. Zhang, S. G. & Wittig, B. (2015). Nat. Biotechnol. 33, 593–598.

60. Colin, M. (1814). Ann. Chim. 90, 252–272.

61. Guthrie, F. (1863). J. Chem. Soc. 16, 239–244.

62. Hassel, O. & Strømme, K. O. (1958). Nature, 182, 1155–1156.

63. Hassel, O. (1972). Nobel Lectures, Chemistry 1963–1970. Amsterdam: Elsevier.

64. Mulliken, R. S. (1950). J. Am. Chem. Soc. 72, 4493–4503.

65. Mulliken, R. S. & Person, W. B. (1969). J. Am. Chem. Soc. 91, 3409–3413.

66. Bertrán, J. F. & Rodríguez, M. (1979). Org. Magn. Reson. 12, 92–94.

67. Lommerse, J. P. M., Stone, A. J., Taylor, R. & Allen, F. H. (1996). J. Am. Chem. Soc. 118, 3108–3116.

68. Legon, A. C. (1999). Angew. Chem. Int. Ed. 38, 2686–2714.

69. Desiraju, G. R., Ho, P. S., Kloo, L., Legon, A. C., Marquardt, R., Metrangolo, P., Politzer, P., Resnati, G. & Rissanen, K. (2013). Pure Appl. Chem. 85, 1711–1713.

70. Riley, K. E. & Hobza, P. (2011). Cryst. Growth Des. 11, 4272–4278.

71. Riley, K. E., Vazquez, M., Umemura, C., Miller, C. & Tran, K. A. (2016). Chem. Eur. J. 22, 17690–17695.

72. Wang, C. W., Danovich, D., Mo, Y. R. & Shaik, S. (2014). J. Chem. Theory Comput. 10, 3726–3737.

73. Rosokha, S. V., Neretin, I. S., Rosokha, T. Y., Hecht, J. & Kochi, J. K. (2006). Heteroat. Chem. 17, 449–459.

74. Řezáč, J. & de la Lande, A. (2017). Phys. Chem. Chem. Phys. 19, 791–803.

75. Jahromi, H. J. & Eskandari, K. (2013). Struct. Chem. 24, 1281–1287.

76. Eskandari, K. & Zariny, H. (2010). Chem. Phys. Lett. 492, 9–13.

77. Clark, T., Hennemann, M., Murray, J. S. & Politzer, P. (2007). J. Mol. Model. 13, 291–296.

78. Brammer, L., Bruton, E. A. & Sherwood, P. (2001). Cryst. Growth Des. 1, 277–290.

79. Zhou, P. P., Qiu, W. Y., Liu, S. & Jin, N. Z. (2011). Phys. Chem. Chem. Phys. 13, 7408–7418.

80. Sarwar, M. G., Dragisic, B., Salsberg, L. J., Gouliaras, C. & Taylor, M. S. (2010). J. Am. Chem. Soc. 132, 1646–1653.

81. Riley, K. E., Murray, J. S., Fanfrlík, J., Řezáč, J., Solá, R. J., Concha, M. C., Ramos, F. M. & Politzer, P. (2011). J. Mol. Model. 17, 3309–3318.

82. Boeckler, F. M., Lange, A., Heidrich, J. & Zimmermann, M. O. (2014). Abstr. Pap. Am. Chem. Soc. p. 248.

83. Dumele, O., Wu, D. N., Trapp, N., Goroff, N. & Diederich, F. (2014). Org. Lett. 16, 4722–4725.

84. Koskinen, L., Jääskeläinen, S., Hirva, P. & Haukka, M. (2015). Cryst. Growth Des. 15, 1160–1167.

85. Carlsson, A. C. C., Mehmeti, K., Uhrbom, M., Karim, A., Bedin, M., Puttreddy, R., Kleinmaier, R., Neverov, A. A., Nekoueishahraki, B., Gräfenstein, J., Rissanen, K. & Erdélyi, M. (2016). J. Am. Chem. Soc. 138, 9853–9863.

86. Scholfield, M. R., Ford, M. C., Vander Zanden, C. M., Billman, M. M., Ho, P. S. & Rappé, A. K. (2015). J. Phys. Chem. B, 119, 9140–9149.

87. Murray, J. S., Lane, P., Clark, T., Riley, K. E. & Politzer, P. (2012). J. Mol. Model. 18, 541–548.

88. Bundhun, A., Ramasami, P., Murray, J. S. & Politzer, P. (2013). J. Mol. Model. 19, 2739–2746.

89. Politzer, P., Murray, J. S. & Clark, T. (2015). Halogen Bonding I: Impact on Materials Chemistry and Life Sciences, Vol. 358, pp. 19–42. New York: Springer.

90. Hennemann, M., Murray, J. S., Politzer, P., Riley, K. E. & Clark, T. (2012). J. Mol. Model. 18, 2461–2469.

91. Ramasubbu, N., Parthasarathy, R. & Murray-Rust, P. (1986). J. Am. Chem. Soc. 108, 4308–4314.

92. Shields, Z. P., Murray, J. S. & Politzer, P. (2010). Int. J. Quantum Chem. 110, 2823–2832.

93. Huber, S. M., Scanlon, J. D., Jimenez-Izal, E., Ugalde, J. M. & Infante, I. (2013). Phys. Chem. Chem. Phys. 15, 10350–10357.

94. Corradi, E., Meille, S. V., Messina, M. T., Metrangolo, P. & Resnati, G. (2000). Angew. Chem. Int. Ed. 39, 1782–1786.

95. Aakeröy, C. B., Panikkattu, S., Chopade, P. D. & Desper, J. (2013). CrystEngComm, 15, 3125–3136.

96. Carter, M., Voth, A. R., Scholfield, M. R., Rummel, B., Sowers, L. C. & Ho, P. S. (2013). Biochemistry, 52, 4891–4903.

97. Ortiz-Lombardía, M., González, A., Eritja, R., Aymamí, J., Azorín, F. & Collect, M. (1999). Nat. Struct. Biol. 6, 913–917.

98. Eichman, B. F., Vargason, J. M., Mooers, B. H. M. & Ho, P. S. (2000). Proc. Natl Acad. Sci. USA, 97, 3971–3976.

99. Eichman, B. F., Ortiz-Lombardia, M., Aymami, J., Collect, M. & Ho, P. S. (2002). J. Mol. Biol. 320, 1037–1051.

100. Hays, F. A., Teegarden, A., Jones, Z. J., Harms, M., Raup, D., Watson, J., Cavaliere, E. & Ho, P. S. (2005). Proc. Natl Acad. Sci. USA, 102, 7157–7162.

101. Vander Zanden, C. M., Carter, M. & Ho, P. S. (2013). Methods, 64, 12–18.

102. Kraut, D. A., Churchill, M. J., Dawson, P. E. & Herschlag, D. (2009). ACS Chem. Biol. 4, 269–273.

103. Wang, L., Xie, J. & Schultz, P. G. (2006). Annu. Rev. Biophys. Biomol. Struct. 35, 225–249

104. Scholfield, M. R., Ford, M. C., Carlsson, A. C., Butta, H., Mehl, R. A. & Ho, P. S. (2017). Biochemistry. 56, 2794-2802

105. Matthews, B. W. (1996). FASEB J. 10, 35–41.

106. Voth, A. R. & Ho, P. S. (2007). Curr. Top. Med. Chem. 7, 1336–1348.

107. Zhou, P., Tian, F., Zou, J. & Shang, Z. (2010). Mini Rev. Med. Chem. 10, 309–314

108. Parisini, E., Metrangolo, P., Pilati, T., Resnati, G. & Terraneo, G. (2011). Chem. Soc. Rev. 40, 2267–2278.

109. Ford, M. C. & Ho, P. S. (2016). J. Med. Chem. 59, 1655–1670.

110. Li, L. & Meroueh, S. O. (2008). Wiley Encyclopedia of Chemical Biology, edited by T. Begley, pp. 1–13. New York: John Wiley and Sons, Inc.

111. Baumli, S., Endicott, J. A. & Johnson, L. N. (2010). Chem. Biol. 17, 931–936.

112. Baumli, S., Lolli, G., Lowe, E. D., Troiani, S., Rusconi, L., Bullock, A. N., Debreczeni, J. E., Knapp, S. & Johnson, L. N. (2008). EMBO J. 27, 1907–1918

113. Fanfrlík, J., Ruiz, F. X., Kadlcikova, A., Rezac, J., Cousido-Siah, A., Mitschler, A., Haldar, S., Lepsik, M., Kolar, M. H., Majer, P., Podjarny, A. D. & Hobza, P. (2015). ACS Chem. Biol. 10, 1637–1642.

114. Xu, Z. J. et al. (2011). J. Med. Chem. 54, 5607–5611.

115. Vallejos, M., Auffinger, P. & Ho, P. (2012). International Tables for Crystallography, Vol. F, edited by D. M. Himmel, Ch 23.26. Heidelberg: Springer.

116. Riel, A.S.M., Decato, D.A., Sun, J., Massena, C.J., Jessop, M.J., Berryman, O.B. (2018). Chem. Sci. 9, 5828-5836.

117. Carlsson, A-C. C., Scholfield, M.R., Rowe, R.K., Ford, M.C., Alexander, A.T., Mehl, R.A., and Ho, P.S. (2018) Biochemistry. 57, 4135-4147.

118. Lu, Y., Wang, Y., Xu, Z., Yan, X., Luo, X., Jiang, H. & Zhu, W. (2009). J. Phys. Chem. B, 113, 12615–12621.

119. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). Nucleic Acids Res. 28, 235–242.

120. Katz, B. A., Sprengeler, P. A., Luong, C., Verner, E., Elrod, K., Kirtley, M., Janc, J., Spencer, J. R., Breitenbucher, J. G., Hui, H., McGee, D., Allen, D., Martelli, A. & Mackman, R. L. (2001). Chem. Biol. 8, 1107–1121.

121. Voth, A. R., Khuu, P., Oishi, K. & Ho, P. S. (2009). Nat. Chem. 1, 74–79

122. Vasylyeva, V., Nayak, S. K., Terraneo, G., Cavallo, G., Metrangolo, P. & Resnati, G. (2014). CrystEngComm, 16, 8102-8105.

123. Takemura, A., McAllister, L. J., Hart, S., Pridmore, N. E., Karadakov, P. B., Whitwood, A. C. & Bruce, D. W. (2014). Chem. Eur. J. 20, 6721–6732.

124. Huang, A. Z., Zhou, L., Zhang, D. W., Yao, J. L. & Zhang, Y. (2016). J. Mol. Graphics Modell. 70, 40–44.

125. Ibrahim, M. A. A. (2011a). J. Chem. Inf. Model. 51, 2549–2559

126. Ibrahim, M. A. A. (2011b). J. Comput. Chem. 32, 2564–2574.

127. Ibrahim, M. A. (2012). J. Phys. Chem. B, 116, 3659–3669.

128. Kolář, M. & Hobza, P. (2012). J. Chem. Theory Comput. 8, 1325–1333.

129. Kolář, M. H. & Hobza, P. (2016). Chem. Rev. 116, 5155–5187.

130. Carter, M., Rappé, A. K. & Ho, P. S. (2012). J. Chem. Theory Comput. 8, 2461–2473.

131. Zhou, P., Lv, J., Zou, J., Tian, F. & Shang, Z. (2010). J. Struct. Biol. 169, 172–182.

132. Hunter, C. A. (2004). Angew. Chem. Int. Ed. 43, 5310–5324.

133. Imai, Y. N., Inoue, Y. & Yamamoto, Y. (2007). J. Med. Chem. 50, 1189–1196.

134. Schreyer, A. & Blundell, T. (2009). Chem. Biol. Drug Des. 73, 157–167.

135. Bissantz, C., Kuhn, B. & Stahl, M. (2010). J. Med. Chem. 53, 5061–5084.

136. Boeckler, F. M., Zimmermann, M. O., Lange, A., Veyisoglu, E., Heidrich, J., Onila, I. & Exner, T. E. (2014). Abstr. Pap. Am. Chem. Soc. p. 248

# CHAPTER 3: HYDROGEN BOND ENHANCED HALOGEN BOND: A SYNGISTIC INTERACTION IN CHEMISTRY AND BIOCHEMISTRY[1]

## ○ 3.1 Summary

The halogen bond (XB) has become an important tool for molecular design in all areas of chemistry, including crystal and materials engineering and medicinal chemistry. Its similarity to the hydrogen bond (HB) makes the relationship between these interactions complex, at times competing against and other times orthogonal to each other. Recently, our two laboratories have independently reported and characterized a synergistic relationship, in which the XB is enhanced through direct intramolecular HBing to the electron-rich belt of the halogen.

In one study, intramolecular HBing from an amine polarizes the iodopyridinium XB donors of a bidentate anion receptor. The resulting hydrogen bond enhanced halogen bond (or HBeXB) preorganizes and further augments the XB donors. Consequently, the affinity of the receptor for halogen anions was significantly increased. In a parallel study, a meta-chlorotyrosine was engineered into T4 lysozyme, resulting in a HBeXB that increased the thermal stability and activity of the enzyme at elevated temperatures. The crystal structure showed that the chlorine of the noncanonical amino acid formed an XB to the protein backbone, which augmented the HB of the wild-type enzyme.

In sum, the two systems described here show that the HBeXB concept extends the range of interaction energies and geometries to be significantly greater than that of the XB alone.

[1] Adapted from published article "Hydrogen Bond Enhanced Halogen Bonds: A Synergistic Interaction in Chemistry and Biochemistry" by Asia Marie S. Riel, Rhianon K. Rowe, Ethan N. Ho, Anna-Carin C. Carlsson, Anthony K. Rappe, Orion B. Berryman, and Pui Shing Ho (2019)

Additionally, surveys of structural databases indicate that the components for this interaction are already present in many existing molecular systems. The confluence of the independent studies from our two laboratories demonstrates the reach of the HBeXB across both chemistry and biochemistry and that intentional engineering of this enhanced interaction will extend the applications of XBs beyond these two initial examples.

- **3.2 Introduction**

The hydrogen bond (HB) has become a central topic in chemistry, since it was first described in water nearly a century ago[1-4]. In structural biochemistry, HBs are the primary noncovalent interactions that define the functional conformations of nucleic acids and proteins[5-7]. More recently, the halogen bond (XB)[8] and its cousins (e.g., the chalcogen, pnictogen, and tetrel bonds) are becoming increasingly recognized as important contributors to molecular assembly and recognition across diverse fields of chemistry, chemical engineering, and biology[9]. The relationships among these various noncovalent interactions can be complex, particularly when two or more are present in the same system[10-12]. Here, we highlight a synergistic relationship, recently described separately in a chemical and a biochemical system, in which a HB greatly enhances the XB potential of a halogen substituent. The principle behind this HB enhanced XB (HBeXB for short) can potentially be applied to other pairs of noncovalent interactions, thereby extending their range of energies and, consequently, applications as design tools for molecular engineering.

- **3.3 Experimental Characterization of the HBeXB**

We present here rare quantification of intramolecular HBeXBing in solution with supporting computational and crystallographic evidence, in which a HB to a halogen substituent increases the XB donating potential. This HBeXB was quantified in a bidentate halide receptor [19] and, independently, with a meta-halotyrosine-modified enzyme [20]. The manifestation of HBeXBs

in both small molecules and a protein suggests that this synergistic interaction will be widely relevant across the fields of chemistry. In this Account, we summarize the studies characterizing the HBeXBs in these two experimental systems. In addition, we present results from surveys of structural databases indicating that HBeXBs are highly prevalent across a broad range of chemical compounds and complexes.

- ***3.3.1 HBeXB Increases Anion Binding***

The Berryman laboratory recently developed bisethynylpyridinium XBing receptors that bind anions and neutral Lewis bases in a bidentate fashion [21,22]. The alkynes promote rigidity and directionality; however, their low rotational barrier allowed the scaffolds to adopt three planar binding conformations. After considering ways to preorganize the structure, we determined that macrocyclization and external intramolecular HBs (away from the binding site) were not synthetically tractable. Instead, we introduced an electron-deficient aniline to HB to the electron-rich belt of the XB donors. This internal intramolecular HB was a unique departure from traditional preorganization techniques in that it also directly enhances XB donor strength.

First generation 1,3-bisethynyl iodopyridinium **1** (no intramolecular HB donor) and second generation **2** (intramolecular HB and fluorine) receptors were recently synthesized (Figure 3.1). The HB's role in preorganization and enhanced XBing (in **2**) as compared to our first-generation receptor (**1**) was quantified by $^{1}$HNMR titrations with chloride, bromide, and iodide. Intramolecular HBeXBing increased halide binding by nearly 9-fold over **1** (in 40% $CDCl_3$ /60% $CD_3NO_2$ ), which lacked the HBeXB. The halide $K_{11}$ values for **2** are 23 700 $M^{-1}$ for $Cl^-$ , 32 900 $M^{-1}$ for $Br^-$ , and 36 900 $M^{-1}$ for $I^-$ . However, **1** binds halides much more weakly, with association constants of 2630 $M^{-1}$ for $Cl^-$ , 4690 $M^{-1}$ for $Br^-$ , and 4380 $M^{-1}$ for $I^-$ . The second binding event ($K_{12}$ ) for all receptors is quite weak and presumably represents nonspecific

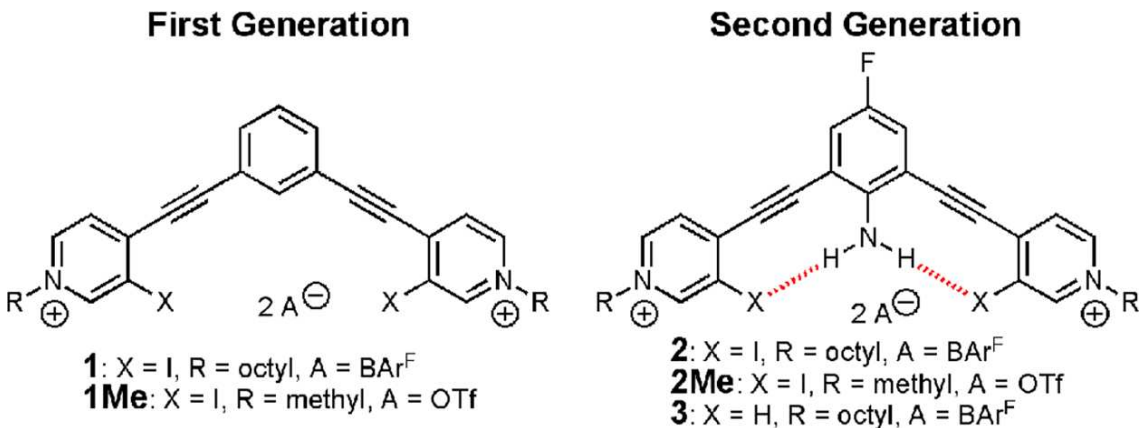



Figure 3.1
Schematics of first generation XB receptor (1 and 1Me) and second generation XB and HB receptors (2, 2Me, and 3). Syntheses can be found in the original publications [20,21]. Adapted with permission from Riel et al.[19] Copyright 2018 Royal Society of Chemistry.

ion pairing to balance charge. To further assess the HBeXB and verify that the amine is not the primary contributor to the increased binding strength, we compared **2** to **3** , which lacks the XB donors. We observed nearly an order of magnitude stronger binding for **2** compared to **3**, concluding that the amine HB donor of **3** does not significantly contribute to the anion affinity in this system. These first solution studies of HBeXBing demonstrate that intramolecular preorganization and enhanced XBing is operable and contributes to the improved halide recognition.

Simultaneous preorganization and enhancement of the XB was further confirmed by gas-phase computations (B3LYP, 6-31+G(d,p), aug-cc-pVTZ and LANDL2DZdp ECP). DFT single point energy calculations demonstrate that the bidentate conformation, with intramolecular HBs, is more stable than the conformation without HBs by 1.29 kcal/mol. Additionally, electrostatic potential (ESP) maps illustrate that **2Me** , with the intramolecular HBeHBs, contains a larger more electrophilic σ-hole (Figure 3.2b) compared to **1Me**, which lacks the HBeXBs (Figure 3.2a). Additional ESP maps of **2Me** with no amine (Figure 3.2c) and **2Me** with no fluorine (Figure 3.2d) verify that the enhanced polarization is caused by the intramolecular HBs from the amine. The magnitude by which HBing enhances bidentate XBing was calculated through interaction energies of **2Me** and **2Me-no NH2** (with no amine therefore no intramolecular HBs) with $Br^-$ . These energies with bromide highlight that the bidentate intramolecular HBeXBs in **2Me** are over 3.2 kcal/mol stronger than solely the XBs in **2Me-no NH2**, which lacks the HBeXBs. These calculations suggest that a single HBeXB interaction in this system provides approximately 1.6 kcal/mol of stabilization. Together, these calculations corroborate the solution data and dual role of the intramolecular HBeXB to enhance the σ -hole and promote preorganization.

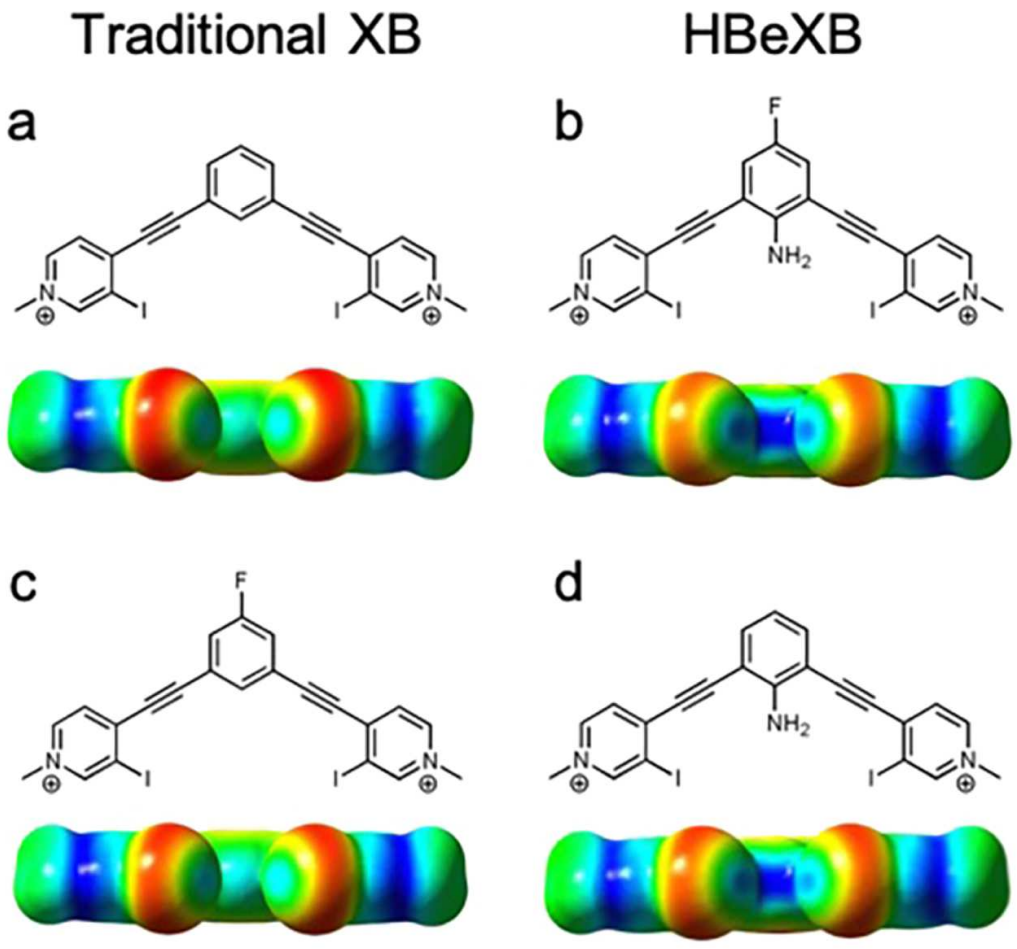



Figure 3.2
Schematics and associated electrostatic potential (ESP) maps of 1Me (a), 2Me (b), 2Me-no amine (c), and 2Me-no fluorine (d) showing HBeXB enhancement of the electropositive σ-holes. ESP maps drawn at a 0.004 au isodensity. Adapted with permission from Riel et al.[19] Copyright 2018 Royal Society of Chemistry.

Crystallographic data with halide counteranions provided detailed structural evidence of HBeXBing (Figure 3.3). When the intramolecular HB is present, we observed a 5% contraction of the XB distances between methyl derivatives of **2** (**2Me**) and both Br− and I− compared to **1Me**. Additionally, the intramolecular HBeXB preorganizes the complexes of **2** and **2Me** with Br− and I− , which promotes planarity in the receptor backbone. The pyridinium rings of **1Me** twist out of planarity up to 15° ; however the addition of the HBing amine decreases ring twist by over half, with the smallest angle at 2.4° . The crystals of **2** and **2Me** confirm that the intramolecular HBeXB can preorganize a receptor while simultaneously improving XB strength.

- ***3.3.2 HBeXB Increases Enzyme Stability and Function***

The growth in the development of polypeptide-based therapeutics spurred the Ho laboratory to determine whether XBs can be engineered to stabilize protein structures, using the enzyme T4 lysozyme (T4L) as the model system (Figure 3.4)[15]. Within the active site of T4L, tyrosine residue (Y18) forms a water-mediated HB to the carbonyl oxygen of a neighboring glutamate (E11) that is essential for the enzyme's structure and function. In order to determine whether an XB can replace this critical HB, we made T4L constructs in which Y18 was replaced by a halogenated phenylalanine ($^{X}$F18, where X = Cl or I) [15]. As a control, we made analogous $^{X}$F replacements at position Y88, a solvent-exposed residue that cannot form XBs, and found these constructs to be destabilizing to the protein. The $^{X}$F18-T4L constructs, however, formed XBs that replaced the essential HB of Y18, thus rescuing the stability of the protein (with Cl < Br < I) relative to the Y88 controls. The rescue, however, was incomplete, in that the engineered XB could not entirely compensate for the loss of stability and function afforded by the essential HB from the hydroxyl of Y18.

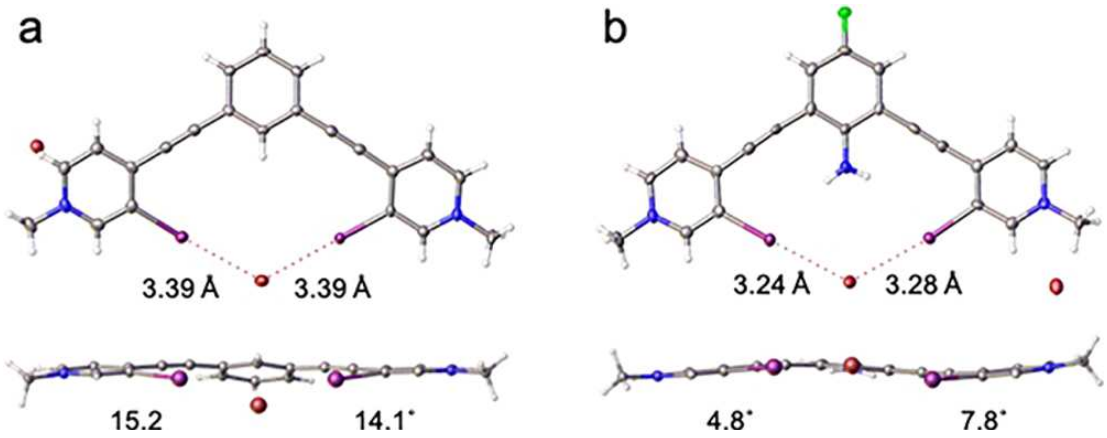



Figure 3.3
Crystal structures of 1Me with bromide top view (a, top) and planar view (a, bottom) comparing distances with 2Me and bromide (b). The planar views include the degrees that the pyridinium rings twist out of coplanarity with the benzene (a, bottom) or fluoroaniline (b, bottom) core. Adapted with permission from Riel et al.[19] Copyright 2018 Royal Society of Chemistry.

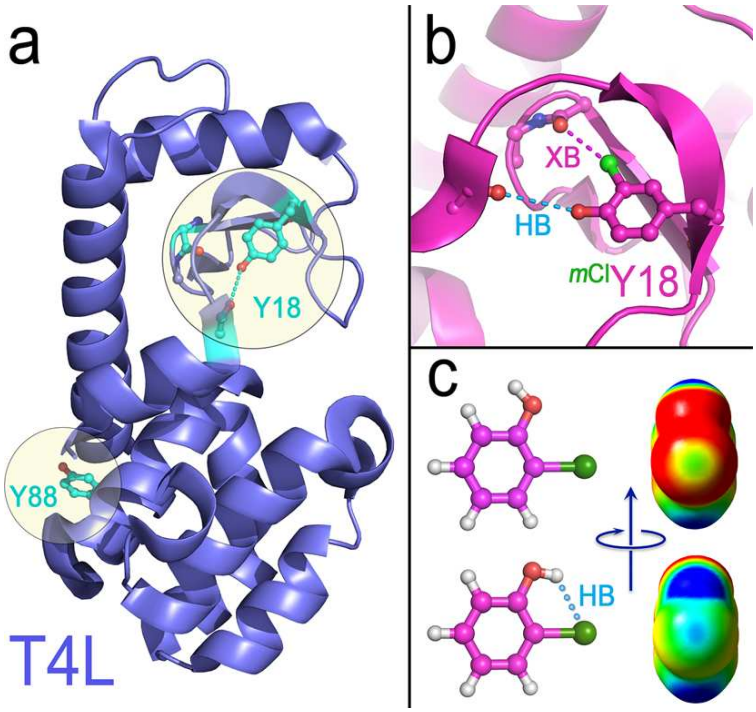



Figure 3.4
T4 lysozyme (T4L) model system for XB studies. (a) The hydroxyl of the tyrosine amino acid at position 18 (Y18) forms a HB to the polypeptide backbone of glutamate E11 (dashes). The side chain of tyrosine at Y88, however, is solvent exposed and does not interact with the remainder of the protein. (b) Replacing Y18 with a meta-chlorotyrosine ($^{Cl}$Y18) maintains the essential HB to E11, with the addition of a XB from the Cl to the peptide oxygen of glycine G28 (with a Cl···O distance of 3.11 Å or 95% of the van der Waals radii). (c) Electrostatic potentials of a chlorophenol model of the $^{Cl}$Y18 side chain. The Cl substituent shows a weak σ-hole when the hydrogen of the OH is rotated away from the halogen (top) but becomes significantly enhanced when rotated to form a HB.

We next attempted to augment, rather than replace, the critical hydroxyl HB of Y18 by introducing a halogen that can form a XB to a different, nearby carbonyl oxygen acceptor (at G28)[20]. This T4L variant was constructed by replacing Y18 with a meta-halotyrosine ($^{X}$Y18). The engineered $^{Cl}$Y18-T4L indeed showed that the chlorine formed a XB to the peptide G28 backbone, resulting in a protein that was more thermally stable than the wild-type one. The melting temperature (Tm) and melting enthalpy (**Δ**Hm) were both elevated (1 ° C and ∼3 kcal/mol, respectively), as determined by differential scanning calorimetry (DSC). In addition, this chlorinated construct showed 15% greater enzymatic activity over the wild-type one at 40 ° C.

The surprising aspect of these results was that the increased stabilization and elevated activity came from adding a single chlorinated substituent, while the brominated and iodinated variants (both expected to have larger σ -holes and therefore stronger XBs) had no effect or were destabilizing. The iodine of the $^{I}$Y18 was too large to fit into the tight and rigid loop into which the halogen must sit and, thus, was seen to be rotated to expose the halogen to solvent, thereby destabilizing the protein. The $^{Br}$Y18 placed the intermediate sized halogen partially exposed and partially XBed within the protein loop (with a very short Br⋯ O distance of 2.88 Å or 85% of the van der Waals radii), with the stabilizing and destabilizing effects essentially neutralizing each other. Only the small chlorine fits into this loop to form a stabilizing XB, with Cl⋯ O distance at 95% of the van der Waals radii, near the statistical mode for biological XBs [13].

The question, however, is why the Cl-XB of the $^{Cl}$Y18 construct has such a significant stabilizing influence on this protein. We had previously shown that a Cl-XB to a very strong anionic oxygen acceptor only provides 0 to 0.5 kcal/mol of stabilizing potential in a DNA system, while Br- and I-XBs contributed 2 to >6 kcal/mol of enthalpic stability [14,16]. Furthermore, quantum mechanical (QM) calculations suggest that a hydroxyl group should be electron donating to ortho

substituents (inset Figure 3.5) and, therefore, the chlorine of the $^{Cl}$Y18 should be a weaker XB donor even compared to a $^{Cl}$F.

The solution to this conundrum came from considering not simply the standard substituent effects of the hydroxyl group but also its ability to serve as a HB acceptor to the OH of the Y18 side chain. QM analyses on chlorophenol models indicate that when the OH is rotated to form an intramolecular HB to the electronegative annulus of the chlorine, the σ -hole becomes enhanced, resulting in a XB-donor that is comparable to that of bromo- or iodobenzene in stabilizing potential. The significantly stronger XB interaction observed in $^{Cl}$Y18-T4L can thus be attributed to this HBeXB. Such an intramolecular O−H⋯ X HB is supported by calculations and experiments on halophenols in nonaqueous environments [23]. In addition to its enhanced stabilizing potential, the QM calculations also indicate that the σ-hole encompasses a larger area of the atomic surface and, therefore, the HBeXB also should show a broader range of angles ($\theta_1$) for the approach of acceptors to the halogen XB donor [20]. The resulting enhanced XB in the $^{Cl}$Y18 T4 lysozyme is thus the first recognition that a HBeXB can increase the stability and function of a biomolecule.

- ***3.3.3 Survey of HBeXB in Cambridge Structure Database and Protein Data Bank***

Interest in the XB has dramatically increased since the turn of the century, with the number of annual publications on the topic growing from <10 prior to 2000 to >450 in 2017. This dramatic increase parallels the application of XBs as a molecular design element in nearly all fields of chemistry. The XB is very similar to the HB in terms of competing acceptors and interaction energies, but the more directional nature of the XB has been seen as a limitation, particularly in biomolecular engineering. However, the HBeXB has the potential to extend the application of XBs by not only increasing the strength of the interaction to be comparable to a traditional HB but also

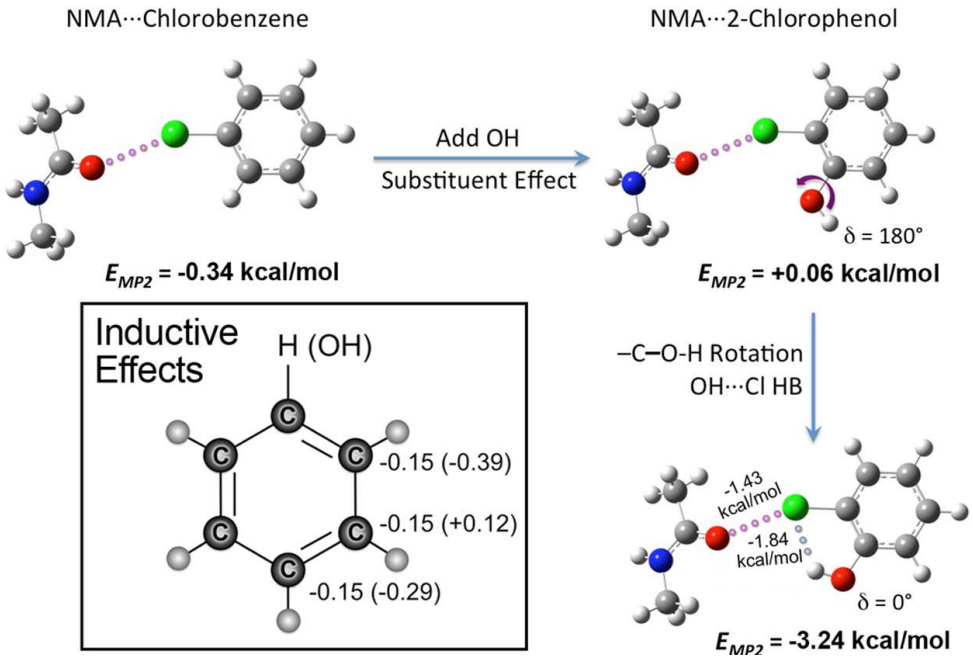



Figure 3.5
Quantum mechanics (MP2) calculated energies (*E*MP2) of XBs from chlorobenzene to the carbonyl oxygen of *N*-methylacetamide (NMA, a model for a peptide bond) and effects from adjacent hydroxyl groups. The Cl-XB is fairly weak, and addition of a hydroxyl to an adjacent (*ortho*) carbon weakens the interaction further. Rotation of the OH to form a HB to the Cl, however, significantly increases the stabilizing potential of the Cl-XB (with *E*MP2 becoming more negative by ∼1.5 kcal/mol). The inset shows the MP2 calculated inductive effects of a hydroxyl (OH) substituent on charges at the carbons of benzene (phenol). The carbons of benzene carry a charge of −0.15*e*, determined through an MP2 calculation. The charges at the *ortho*- and *para*-carbons of the phenols become more negative, reflecting the electron donating effect, while that of the *meta*-carbon become more positive, indicative of the electron withdrawing effect of the hydroxyl group to these positions. The Hammett constants[24] for hydroxyl substituents are −0.37 for the *para*- and +0.12 for the *meta*-position, consistent with the quantum calculated effects on the carbon charges.

expanding the atomic surface encompassed by the electropositive σ -hole and, consequently, extending the angles of approach by the acceptor.

The experimental observations of HBeXBs in a small molecule anion receptor and in an engineered protein, involving various XB donors and HB donors, suggest that this interaction is applicable across a wide range of chemical systems. We thus addressed the question of whether HBeXBs could be present in other chemical systems by surveying the Cambridge Structural Database (CSD[25]) for structures in which the basic elements of this interaction are present. Our initial survey searched for aromatic compounds with Cl, Br, or I that are ortho to OH or NH2 substituents and within short distance (≤ 105% of the sum of the van der Waals radii, **Σ**RvdW ) of a XB acceptor (O or N) of an interacting compound. This analysis identified 772 complexes, indicating that the potential for HBeXBs is very high, even with these very limited criteria.

A radial distribution plot of the XB donor approach to the acceptor (the $\theta_1$ -angle) showed that these interactions cluster around the σ -hole of the halogen ($\theta_1 \approx 180°$ , Figure 3.6), as expected for XBs, but extend to the electronegative annulus ($\theta_1 \approx 90°$ ). However, detailed analysis of the distance from the acceptor to the halogen ($R_{X\text{-}A}$) versus the acceptor to the HB donor atoms ($R_{O/N-A}$) shows a significant number of these contacts are primarily HB interactions to the *ortho* − OH or − NH substituents, instead of XBs to the XB donors (Figure 3.6b). With these HBs removed from the data set, the resulting radial analysis is even more highly clustered around the σ -hole but more broadly distributed across $\theta_1$ -angles (with $\hat{R}_{X\cdots(O/N)} \leq 1.0$ and $\theta_1 \geq 135°$ ) than seen in previous surveys of XBs (Figure 3.6c). Thus, surveys that attempt to identify XBs or HBeXBs in systems with neighboring HB donors using only the $R_{X-A}$ as a measure must be cognizant of other classes of noncovalent interactions that may confound the results. It is interesting that at $\theta_1 \approx 180°$, the normalized XB distance ($\hat{R}_{X\cdots(O/N)}$ ) does not extend beyond 100% of $\Sigma R_{vdW}$, even though the

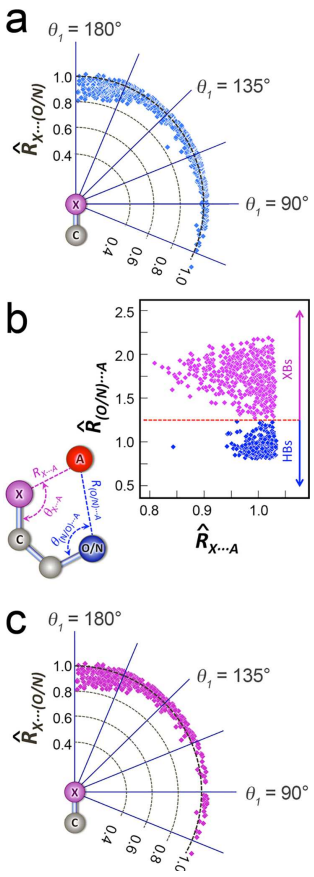



Figure 3.6

Results from survey of structures from the Cambridge Structural Database (CSD) for potential HBeXBs. (a) Radial distribution of potential HBeXBs. The CSD was surveyed for structures of halogenated aromatic compounds (Cl, Br, or I) with HB donors (OH or NH2) at the ortho-position that form complexes with a XB/HB acceptor (O or N). The distance from the halogen to the acceptor atom, normalized to the sum of the respective van der Waals radii ($\hat{R}_{X\cdots(O/N)} \leq 1.05$) are plotted radially relative to the angle of approach of the acceptor to the C−X bond ($\theta_1$). (b) Plot of normalized distances from the acceptor (A) to the halogen ($\hat{R}_{X\cdots A}$) versus the distance to the HB donor ($\hat{R}_{(O/N)\cdots A}$). HB interactions from the acceptor to the O/N HB donor are distinguished from XBs from the acceptor to the XB donor by $\hat{R}_{(O/N)\cdots A} \leq 1.25$. (c) Radial plot of XBs from panel a, with HBs to the O/N HB donors removed according to the criteria in panel b. This distribution analysis indicates that HBeXBs fall in the range of $\theta_1$ angles from ~135° to 180°.

survey extends beyond that distance, consistent with the HBeXBs being particularly strong interactions (assuming the strength of a noncovalent interaction is reflected in the contact distance). An initial survey of the Protein Data Bank with these same limiting criteria identified over 1000 structures with the components required to form HBeXBs, consistent with the previous survey by Lin and MacKerell [18]. The difference, however, is that the results from our CSD survey now allow us to distinguish between HBs and potential HBeXBs in biomolecular structures. Thus, the two examples of HBeXBs found in the halide receptor and the model protein, as described here, are most likely not singular exceptions but simply the first experimental recognition of a potentially prevalent molecular interaction.

- **3.4 Conclusions and Perspectives**

This Account highlights the concept that a HB directly to the electron-rich region of a halogen augments its potential as a XB donor. The resulting HBeXB extends the stability and the geometry of an XB interaction, rendering it comparable and potentially stronger than a classical HB. While the concept of the HBeXB had previously been suggested, our studies are highlighted as the first direct characterization of this synergistic relationship. Our initial surveys of the CSD and PDB highlight the strong likelihood that HBeXBs are common in both chemical and biochemical molecular systems.

Although we have now sampled HBeXBs at the two extremes of chemical complexity and with different pairs of XB and HB donors, there remain many aspects of polarization effects that are yet to be explored. For example, we expect that a stronger HB donor in these coordinated systems will strengthen the XB donor potential of the halogen. Similarly, the geometry of the HB⋯ X (the distance and angle of approach of the HB interaction) would affect the XB acceptor to donor geometry. A shorter HB, for example, would be expected to have a stronger polarizing

effect on the σ -hole and, thus, produce a stronger attractive force with the acceptor. Alternatively, the angle of approach of the HB to the halogen could affect the position of the σ -hole at the halogen surface, thereby affecting the approach $\theta_1$ -angle of the acceptor. Quantum calculations on simple model systems show that when the HB deviates from being perpendicular to the C− X bond (near the optimum 90° of the electronegative center), the most electropositive point of the σ-hole deviates from the ideal 180° along the C− X bond. We should note, however, that the two experimental systems presented here are in essentially nonaqueous environments, in an organic solvent or a solvent-excluded pocket of a protein. As with other noncovalent interactions, accessibility to water could reduce the overall potential of the HBeXB through either direct competition for the acceptor or, bulk solvent, by increasing the dielectric constant of the environment. Our research groups are studying these and other physical properties and effects on the HBeXB to better understand how we can rationally design the interaction for molecular and biomolecular engineering.

Finally, we highlight that the HBeXB is a type of polarization-enhanced noncovalent cooperativity (Figure 3.7, green box, σ-bond cooperativity). As a subclass of noncovalent cooperativity [26], polarization-enhanced XBs are unique: they can be polarized either directly through noncovalent interaction with the donor (e.g., HBeXB) or indirectly by noncovalent interaction with an adjacent atom that shares a σ-bond with the donor (Figure 3.7, red box). An example of the latter was seen in the adducts of dihalogens with heterocyclic pentatomic chalcogen donors, where a HB at one end of the dihalogen enhances the XB donor potential at the other end [17]. The generality of this approach is foreshadowed by a recent computational study showing that intramolecular HBs also can enhance tetrel bonds (HBeTtB) in fluorosilyl and fluorogermanium complexes [27]. The HBeXB and HBeTtB can be considered as two related subcategories of

polarization-enhanced noncovalent interactions (Figure 3.7 , pink and orange boxes), where HBs and other noncovalent interactions can cooperatively strengthen or weaken the noncovalent bonding of polarizable atoms (such as halogen or tetrel substituents). While this strategy is only now being explored, it has the potential to extend the utility of these interactions in chemistry and biochemistry, providing powerful alternatives to the classic HB in molecular engineering.

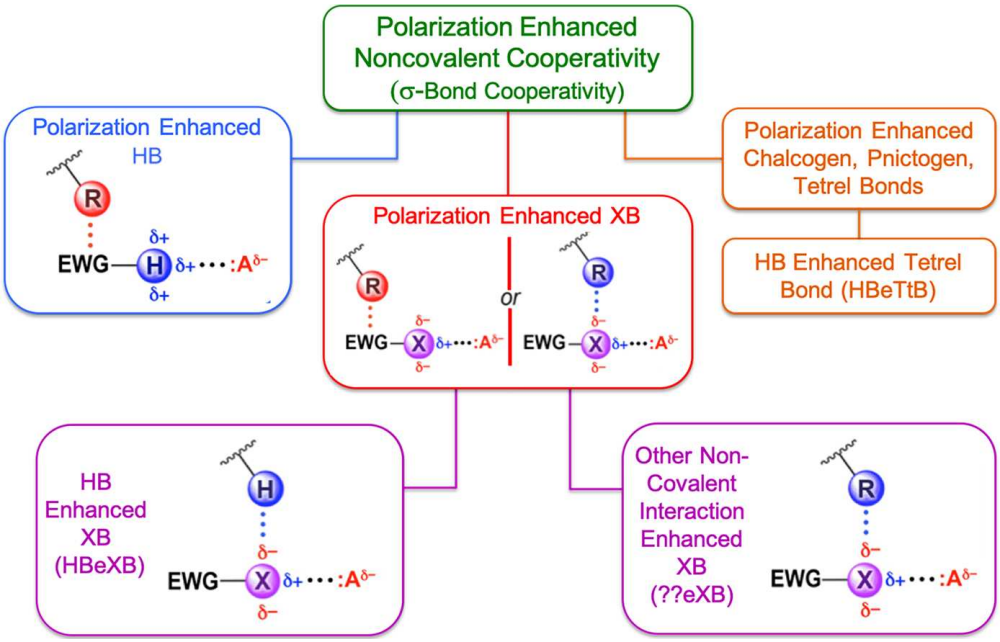



Figure 3.7
Different types of polarization enhanced noncovalent cooperativity. The HBeXB is a subclass of polarization enhanced XBs where HBing directly to the XB donor enhances the XB interaction. EWG is an electron-withdrawing group adjacent to a HB or XB donor, while A refers to electron-rich acceptors of HBs or XBs.

# REFERENCES


1. Latimer, W. M.; Rodebush, W. H. Polarity and Ionization from the Standpoint of the Lewis Theory of Valence. J. Am. Chem. Soc. 1920, 42, 1419−1433.
2. Pauling, L. The Nature of the Chemical Bond and the Structure of Molecules and Crystals; An Introduction to Modern Structural Chemistry, 3rd ed.; Cornell University Press: Ithaca, N.Y., 1960.
3. Smith, D. A. A Brief History of the Hydrogen Bond. In Modeling the Hydrogen Bond; Smith, D. A., Ed.; ACS Symposium Series; American Chemical Society: Washington DC, 1994; pp 1−4.
4. Pimentel, G. C.; McClellan, A. L. The Hydrogen Bond; W. H. Freeman and Co.: San Francisco, 1960.
5. Baker, E. N. Hydrogen Bonding in Biological Macromolecules. In International Tables for Crystallography Vol. F, Crystallography of biological macromolecules; Rossmann, M. G., Arnold, E., Eds.; Springer Netherlands: Dordrecht, 2001; pp 546−552.
6. Saenger, W. In Principles of Nucleic Acid Structure; Cantor, C. R., Ed.; Springer-Verlag: New York, 1984.
7. Kuriyan, J.; Konforti, B.; Wemmer, D. E. The Molecules of Life; Garland Science, Taylor & Francis Group: New York, NY, 2013.
8. Desiraju, G. R.; Ho, P. S.; Kloo, L.; Legon, A. C.; Marquardt, R.; Metrangolo, P.; Politzer, P.; Resnati, G.; Rissanen, K. Definition of the Halogen Bond (IUPAC Recommendations 2013). Pure Appl. Chem. 2013, 85, 1711−1713.
9. Cavallo, G.; Metrangolo, P.; Milani, R.; Pilati, T.; Priimagi, A.; Resnati, G.; Terraneo, G. The Halogen Bond. Chem. Rev. 2016, 116, 2478−2601.

10. Metrangolo, P.; Resnati, G. Chemistry: Halogen versus Hydrogen. Science 2008, 321, 918−919.

11. Aakeröy, C. B.; Panikkattu, S.; Chopade, P. D.; Desper, J. Competing Hydrogen-Bond and Halogen-Bond Donors in Crystal Engineering. CrystEngComm 2013, 15, 3125−3136.

12. Rowe, R. K.; Ho, P. S. Relationships between Hydrogen Bonds and Halogen Bonds in Biological Systems. Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. 2017, 73, 255−264.

13. Scholfield, M. R.; Zanden, C. M.; Carter, M.; Ho, P. S. Halogen Bonding (X-Bonding): A Biological Perspective. Protein Sci. 2013, 22, 139−152.

14. Voth, A. R.; Khuu, P.; Oishi, K.; Ho, P. S. Halogen bonds as orthogonal interactions to hydrogen bonds. Nat. Chem. 2009, 1, 74−79.

15. Scholfield, M. R.; Ford, M. C.; Carlsson, A.-C. C.; Butta, H.; Mehl, R. A.; Ho, P. S. Structure-Energy Relationships of Halogen Bonds in Proteins. Biochemistry 2017, 56, 2794−2802.

16. Carter, M.; Voth, A. R.; Scholfield, M. R.; Rummel, B.; Sowers, L. C.; Ho, P. S. Enthalpy-Entropy Compensation in Biomolecular Halogen Bonds Measured in DNA Junctions. Biochemistry 2013, 52, 4891−4903.

17. Montis, R.; Arca, M.; Aragoni, M. C.; Bauzá, A.; Demartin, F.; Frontera, A.; Isaia, F.; Lippolis, V. Hydrogen-and halogen-bond cooperativity in determining the crystal packing of dihalogen charge- transfer adducts: a study case from heterocyclic pentatomic chalcogenone donors. CrystEngComm 2017, 19, 4401−4412.

18. Lin, F. Y.; MacKerell, A. D., Jr. Do Halogen-Hydrogen Bond Donor Interactions Dominate the Favorable Contribution of Halogens to Ligand-Protein Binding? J. Phys. Chem. B 2017, 121, 6813−6821.

19. Riel, A. M.; Decato, D. A.; Sun, J.; Massena, C. J.; Jessop, M. J.; Berryman, O. B. The Intramolecular Hydrogen Bonded-Halogen Bond: A New Strategy for Preorganization and Enhanced Binding. Chem. Sci. 2018, 9, 5828−5836.

20. Carlsson, A.-C. C.; Scholfield, M. R.; Rowe, R. K.; Ford, M. C.; Alexander, A. T.; Mehl, R. A.; Ho, P. S. Increasing Enzyme Stability and Activity through Hydrogen Bond-Enhanced Halogen Bonds. Biochemistry 2018, 57, 4135−4147.

21. Massena, C. J.; Riel, A. M. S.; Neuhaus, G. F.; Decato, D. A.; Berryman, O. B. Solution and Solid-Phase Halogen and C-H Hydrogen Bonding to Perrhenate. Chem. Commun. 2015, 51, 1417−1420.

22. Riel, A. M. S.; Jessop, M. J.; Decato, D. A.; Massena, C. J.; Nascimento, V. R.; Berryman, O. B. Experimental Investigation of Halogen-Bond Hard-Soft Acid-Base Complementarity. Acta Crystal- logr., Sect. B: Struct. Sci., Cryst. Eng. Mater. 2017, 73, 203−209.

23. Abraham, M.H.; Abraham, R.J.; Aliev, A.E.; Tormena, C.F.; Is There an Intramolecular Hydrogen Bond in 2-Halophenols? A Theoretical and Spectroscopic Investigation. Phys. Chem. Chem. Phys. 2015, 17, 25151-25159

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

25. Groom, C.R.; Bruno, I.J.; Lightfoot, M.P.; Ward, S.C. The Cambridge Structural Database. Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. 2016, 72, 171-179.

26. Jeffrey, G.A. Hydrogen-Bonding; An Update. Crystallogr. Rev. 1995, 4, 213-254

27. Trujillo, C.; Alkorta, I.; Elguero, J.; Sanchez-Sanz, G. Cooperative Effects in Weak Interactions: Enhancement of Tetrel Bonds by Intramolecular Hydrogen Bonds.

# CHAPTER 4: ENGINEERING HYDROGEN AND HALOGEN BONDS TO ENGINEER A SPECIFIC COILED-COIL SENSING PEPTIDE[1]

## ○ 4.1 Summary

While the X-bond has previously been engineered into a DNA junction system and into a protein as an intramolecular interaction[1,2], it has not yet been characterized as a protein-protein interaction or an intermolecular interaction. Due to the X-bond's geometrically specific nature this initial engineering effort will be important to the field of biological engineering. Here we will address whether X-bonds can confer binding specificity and sensing ability. We will use a model coiled-coil system (dimerization domain of GCN4 transcription factor) as a simplified protein complex to address how the X-bond donor on a single α-helix can sense and specifically bind to a pre-existing coiled-coil dimer through the formation of an X-bond, to create a unique heterotrimer. By comparing our results to similarly designed heterotrimers assembling with non-specific hydrophobic interactions and an alternative electrostatic interaction, the H-bond, we will be able to determine how the X-bond compares as an intermolecular protein-protein interaction capable of sensing for the first time.

## ○ 4.2 Introduction

### ○ *4.2.1 Engineering Protein-Protein Interactions*

A major challenge in biomolecular engineering is to design artificial protein-protein interactions that show the same degree of binding affinity and specificity as oligomeric complexes that have evolved through natural selection[3]. A common scaffold for many such engineering

[1] Adapted from manuscript in process for submission by Rhianon Kay Rowe Hartje, Marco Ferrero, Gabriella Cavallo, Alessandro Gori, Pierangelo Metrangolo, and P. Shing Ho

efforts are the coiled-coils, naturally occurring α-helical assemblies that have been coopted for a variety of biomedical and material science applications [4,5]. We show here that introducing either a hydrogen bond (H-bond) or halogen bond (X-bond) interaction into an α-helix results in a peptide that binds with high specificity and affinity, serving as a sensor for the dimerization domain of the GCN4 transcription factor.

Recent studies have revealed the fundamental steps of the central dogma—replication, transcription, translation—in dramatic fashion in live cells, owing in part to the development of advanced tools, including ultra-resolution microscopy[6]. At the cellular level, molecular tools such as nanobodies[7] that can recognize and bind to proteins or nucleic acid targets are crucial for labeling specific components of the cellular machinery. As such studies expand beyond visualizing basic steps to regulatory mechanisms [8–10], there is a need to engineer new sensing molecules, which in turn will require the design of new molecular (*e.g.*, protein-protein) interactions and interfaces.

Although artificial protein-protein interfaces have been designed computationally [11], and through directed evolution[12], there have been only a few examples to date of success in the de novo engineering of new artificial protein assemblies [13]. The landscape available for such engineering challenges, however, can be greatly expanded by extending the tool box of molecular interactions beyond those found in the naturally occurring amino acids to those available from non-canonical amino acids[14]. Although H-bonds are commonly seen as an interaction that provides specificity in biological structures and ligand binding, X-bonds have recently gained significant attention for molecular recognition and biomolecular engineering[15,16].

- ***4.2.2 Halogen Bonds in Biomolecules***

X-bonds are analogous to H-bonds in that both present an electropositive atomic surface as a donor function to form a favorable non-covalent interaction with an electron-rich acceptor

atom (such as an O, N, or S) [17]. A simple, though incomplete, theory for the root of X-bonds is the electrostatic σ-hole model. This model posits that in forming a covalent bond with another atom, the $p_Z$-orbital of a halogen substituent becomes depopulated, leaving an electron-depleted crown (called the σ-hole) diametrically opposed to the resulting σ-molecular orbital. It is this σ-hole that provides the electropositive potential to interact with electron-rich acceptors. X-bonds are now recognized to play important roles in biology,[18,19] including defining the recognition of halogenated inhibitors by their protein targets [20–24] and the design of new inhibitors against clinically important targets [22,25,26]. We had previously engineered X-bonds into DNA junctions to control the molecular conformation [1] and into the classic T4 lysozyme enzyme [2] to increase its thermal stability and activity at elevated temperatures [16]. In this study, X-bonds are introduced through non-canonical amino acids to engineer a specific protein-protein interaction in coiled-coil complexes.

- ***4.2.3 Coiled-Coils as Model Protein Engineering Platform***

We asked here whether X-bonds and H-bonds could be engineered to give specificity to complexes of coiled-coils—two or more α-helices that self-assemble into superhelical bundles. These assemblies have been broadly adapted to tackle a range of problems, including serving as model systems to map protein folding pathways[27,28] and sequence-structure-function relationships[29], and in the design of cavities for enzyme-substrate complexes[30] and viral fusion inhibitors as potential therapeutics[31–33]. This adaptability is due in part to their predictable α-helix structure, which further stems from a simple repeated heptad motif of $(\mathbf{a}bc\mathbf{d}efg)_n$ [34]. The amino acids at positions **a** and **d** are typically hydrophobic residues, while *e* and *g* are typically charged. The helical repeat of ~3.6 amino acids/turn aligns the characteristic hydrophobic residues along

one face of the α-helix, and it is the association of the hydrophobic faces among strands that ultimately drives the assembly of higher-order coiled-coil bundles.

One coiled-coil helix that has been well-studied as a design scaffold is the "leucine zipper", so-called because of the regular placement of Leu at position **d** of the heptad repeat[32,35–37]. The ~30 amino acid dimerization domain of the GCN4 transcription factor is a naturally occurring leucine zipper that assembles as a parallel-aligned homodimer. Changing the amino acids within the GCN4 heptad motif changes the topology of coiled-coil assemblies. Harbury *et al*., showed that varying the hydrophobic residue at position **a** while maintaining a Leu at **d** resulted in GCN4 peptides that assemble as dimers, trimers, or tetramers and in parallel or antiparallel alignments[36,38]. It has been shown that the oligomerization state of coiled coil helices are controlled by "trigger sequences" [39]. One such trigger sequence that converts the GCN4 dimer into a trimer involves a series of H-bonds and salt bridges among side chains that line the exterior of the coil-coil assembly.

For the current study, we wanted to exploit X-bonds and H-bonds for the design of a heptad repeat peptide that would specifically recognize and bind to a GCN4 dimer and, thus, serve potentially as a cellular sensor for this genetic regulator. For this challenge, we started with the study by Gonzalez *et al.,* which showed that mutating an asparagine at position 16 to an alanine still allowed the formation of a stable GCN4 dimer[40]. The loss of the amide substituent of the side chain, however, created a ligand binding cavity that upon the addition of benzene to the system, the GCN4-N16A sequence assembled as a parallel trimer. In examining the structure of the modeled dimeric and benzene-induced trimeric states of GCN4-N16A, we observed that a carbonyl oxygen of the peptide backbone also became more accessible (figure 4.1), leading us to consider a design in which an X-bonding or H-bonding function on an aromatic side chain could

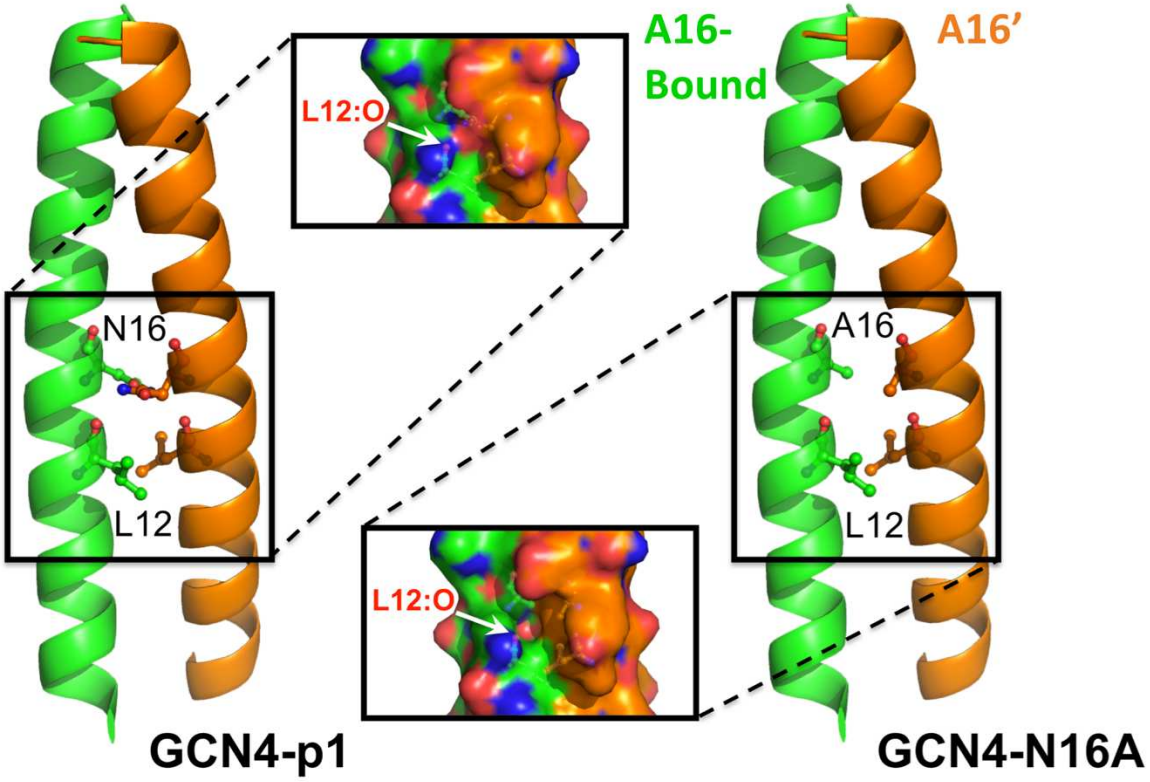



Figure 4.1
The starting model for our engineering efforts. GCN4-p1 is WT with N16 to promote parallel dimer formation. GCN4-N16A is a less stable dimer than GCN4-p1, but the mutation creates a slight hydrophobic pocket and makes the carbonyl oxygen in L12 more accessible. The green will denote the A16-bound strand throughout the paper, and the orange will denote the A16' strand.

not only substitute for the benzene in filling the hydrophobic cavity, but also provide specificity through a non-covalent interaction. Herein we describe the successful efforts in engineering and characterizing heterotrimeric GCN4 coiled-coils formed by the association of this GCN4-N16A dimer with an α-helix monomer engineered with either a Tyr (to form an H-bond) or halogenated-Phe (to form an X-bond) to fill the ligand cavity. We will call this third engineered peptide thesensing strand ($^{Z}$F) of the heterotrimer. Although the peptide oxygens of both GCN4-N16A strands are accessible, the chirality of amino acids positions the Z substituent of the $^{Z}$F strand only towards one of the two GCN4-N16A chains. We will distinguish this "sensed" strand as A16-bound and the non-sensed strand as A16' for the $(GCN4\text{-}N16A)_2$ dimer (figure 4.1).

- **4.3 Experimental Design**

To carry out this study we utilized a few different techniques to better understand how these engineered electrostatic and hydrophobic interactions influenced the structure of the coiled-coil. We crystallized four unique heterotrimers to glean insight in how the non-canonical amino acid packs in the hydrophobic core or designed "ligand binding" site of the GCN4-N16A coiled-coil. We carried out differential scanning calorimetry (DSC) and circular dichroism (CD) studies to assess the thermodynamic stability and solution state of the coiled-coils. We also performed quantum mechanical calculations on coordinates from the crystal structure to address the energy change we noted from the DSC studies.

- ***4.3.1 Peptide Design and Purification***

30mer peptides (RMKQLEDKVEELLSK**Z**YHLENEVARLKKLV) with position 16 being A (for A16-bound and A16' sequences), or $^{Z}$F with Z being a hydroxyl group ($^{OH}$F), a methyl group ($^{Me}$F16), or iodine ($^{I}$F16) were ordered crude from Biomatik and came 40-60% pure. The 30mer sequence with 2,3,5,6-fluoro, 4-iodo-phenylalanine ($^{TFI}$F16) at position 16 was synthesized

according to Bergamaschi et al. in the Metrangalo lab in Milan, Italy[41]. Lyophilized peptides were resuspended in 0.1% trifluoroacetic acid (TFA) in water and HPLC purified on a C18 semi-preparative column with a gradient between 0.1% TFA in water and 0.1% TFA in acetonitrile. HPLC fractions containing the full-length pure peptide were dried down in a speed vacuum and resuspended in the appropriate buffer for the set of experiments.

- ***4.3.2 Crystallization, Data Collection, Structure Determination***

Crystallization was carried out via the hanging drop vapor diffusion method. Drops were prepared by mixing 2 μL of 1.46 mM peptide with 2 μL of mother liquor and allowed to equilibrate at 25 °C over a well containing 500 μL of mother liquor. The 1.46 mM peptide was prepared in a 2:1 ratio of GCN4-N16A:GCN4N16-$^{Z}$F16, with $^{Z}$F being 2,3,5,6-fluoro, 4-iodo-phenylalnine ($^{TFI}$F16), 4-methyl-phenylalnine ($^{Me}$F16), 4-iodo-phenylalanine ($^{I}$F16), or tyrosine ($^{OH}$F16) in 10 mM potassium phosphate, 100 mM potassium chloride pH 7.0. Crystals were obtained for all of the constructs in mother liquors ranging from 0.9-1.1 M sodium citrate and 90-110 mM HEPES pH 7.0. Crystals were looped, flash frozen in liquid nitrogen and mounted, so data could be collected from the crystal on the Rigaku Compact Home Lab with a PILATUS detector (figure 4.2). HKL2000 was used to index, integrate, and scale the data. The phase of the structures was determined with molecular replacement [using the GCN4-N16A trimer as the starting search model (PDB 1SWI)] and refined using Phenix.

- ***4.3.3 Melting Profiles Determined with Differential Scanning Calorimetry***

DSC samples were prepared by mixing GCN4-N16A with GCN4-N16-$^{Z}$F16 in a 2:1 ratio at 0.54 mM in 100 mM sodium chloride, 50 mM HEPES pH 7.0. The data were collected on a TA instruments Nano DSC. Thermal scans were set-up from 5°C to 105°C at a rate of 1°C /min at a constant pressure of 3.0 atm with 600 s equilibration. Melting temperatures ($T_M$) and enthalpies of

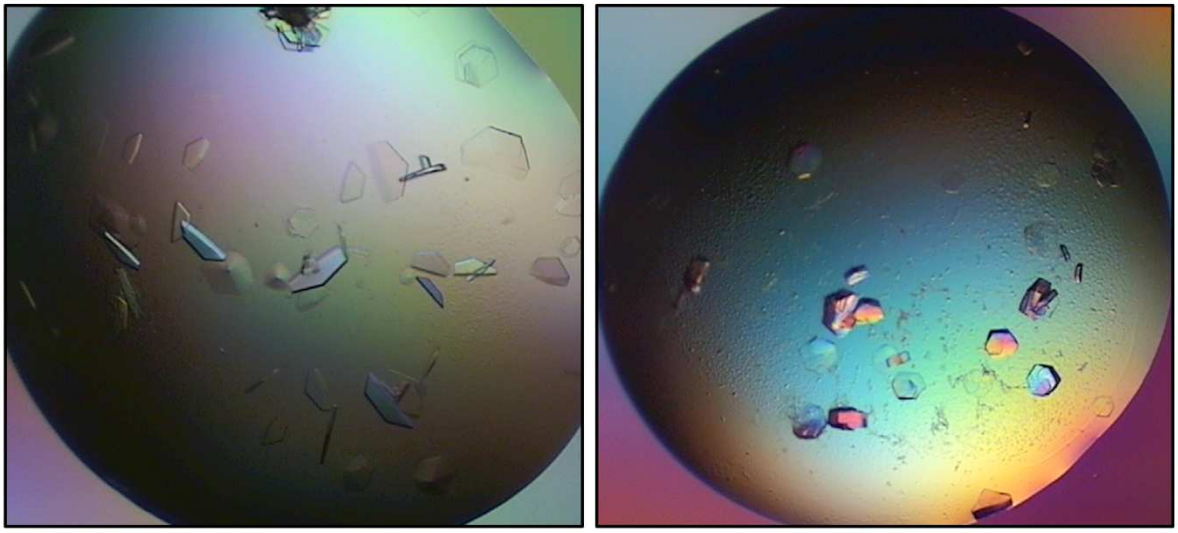


Figure 4.2
Representative crystals that were looped and data collected from for each of the four crystal structures determined. All crystals indexed to the same space group (C121) with similar unit cell dimensions and angles.

melting ($\Delta H_M$) were determined by fitting the data with TA Nano Analyze software, using a two-state scaled model. At least 4 replicate measurements were conducted for each construct. The average and standard deviation were calculated and reported from the 4 replicates. An entropy of melting ($\Delta S_M$) for each scan was calculated from the $T_M$ and $\Delta H_M$, as $\Delta S_M = \Delta H_M / T_M$.

- ***4.3.4 Circular Dichroism Spectroscopy Titration***

CD spectra were obtained on a Jasco spectrometer J-1000 and analyzed with Spectrum Manager. A titration of increasing GCN4-N16$^Z$F16 to GCN4-N16A was carried out in order to assess the ratio of the two sequences that formed a stable coiled-coil in solution. Titration points included ratios of GCN4-N16A: GCN4 N16-$^Z$F16 of 1:0, 10:1, 7:1, 5:1, 4:1, 3:1, 5:2, 2:1, 3:2, 1:1, 1:2 and 1:3. The $\Delta\varepsilon$ at 208 and 222 nm were monitored and their ratio (222/208) was calculated from the spectra to indicate higher ordered coiled-coil formation. 10 scans were carried out and averaged for each titration point, and each titration was done in triplicate. Scans were taken from 190 nm to 300 nm over 50 nm/min. Samples were prepared by mixing the appropriate ratio of GCN4-N16A to GCN4-N16-$^Z$F16 to a volume of 500 μL and 50 μM in 5 mM HEPES pH 7.48 and placed in a 1mL cuvette for data collection.

- ***4.3.5 Quantum Mechanical (QM) Calculations***

The atomic coordinates from the crystal structures were taken from the interacting residues (A16 and L12 from one chain, and $^Z$F16 from alternate chain) and were simplified down to 2 N-methyl-acetamide molecules (A16 and L12) and a benzene mimicking the $^Z$F aromatic residue. An AM1 geometry optimization on the hydrogen atoms was performed on the complex prior to energy calculations. For energy calculations, Møller-Plesset second-order (MP2) level of theory with the aug-cc-PVTZ (aug-cc-PVTZ-PP from EMSL basis set exchange for $^I$F16 and $^{TFI}$F16) basis set was used in cyclohexane (D=2 compared to vacuum) to mimic the hydrophobic interior of the coiled-

coil. Basis set superposition error (BSSEs) calculations were also carried out and summed into the solvent phase energy.

- **4.4 Results**

The objective of the current study is to engineer a peptide that specifically recognizes and binds to, or senses, a variant of the GCN4 dimerization domain to form a heterotrimeric coiled-coil assembly. For the current study, we have designed a sensing peptide strand that is based on the GCN4 leucine zipper, but with a substituted phenylalanine ($^{Z}$F, where *Z* is an -OH, -I, or methyl (Me)) incorporated to provide binding affinity and specificity. The $^{Z}$F is expected to fill the hydrophobic cavity created in a mutant of the GCN4 dimerization domain in which the Asn residue at position 16 is replaced by an Ala (GCN4-N16A) and provide specificity through an H-bond or an X-bond to the now accessible carbonyl oxygen of the polypeptide backbone (figure 4.1). The GCN4-N16A is combined with a GCN4-N16 $^{Z}$F peptide in a 2:1 ratio in order to form a potential heterotrimeric complex (table 4.1). The assembly of the predicted heterotrimeric coiled-coils and the role of the H- or X-bonding interaction are characterized in the single-crystal structures of the complexes (figure 4.3). In addition, the structures provide the detailed geometries of the $^{Z}$F substituent interaction with the carbonyl oxygen of the L12 peptide backbone, along with other near-neighbor interactions, in order to assess the potential formation of X- or H-bonds. Of particular interest are the distances ($d_{Z\cdots O}$) and the angle of approach ($\theta_1$) of the *Z* substituent to the carbonyl oxygen of L12. The oligomeric states of the assemblies were characterized in solution from the thermal melting parameters determined by differential scanning calorimetry (DSC) and through titration studies that monitor the superhelical state by circular dichroism (CD) spectroscopy.

Table 4.1
The abbreviations of the strands referred to throughout and the sequences of the respective strands used for all the experiments carried out. The 30mer sequence is derived from the dimerization domain of the GCN4 transcription factor. All of the sequences are identical expect for position 16 which is bolded.

| STRANDS | SEQUENCES |
|---|---|
| A16' – GCN4 | N-RMKQLEDKVEELLSK**A**YHLENEVARLKKLV-C |
| A16 – BOUND – GCN4 | N-RMKQLEDKVEELLSK**A**YHLENEVARLKKLV-C |
| $^{OH}$F – GCN4 | N-RMKQLEDKVEELLSK$^{\mathbf{OH}}$**F**YHLENEVARLKKLV-C |
| $^{I}$F – GCN4 | N-RMKQLEDKVEELLSK$^{\mathbf{I}}$**F**YHLENEVARLKKLV-C |
| $^{TFI}$F – GCN4 | N-RMKQLEDKVEELLSK$^{\mathbf{TFI}}$**F**YHLENEVARLKKLV-C |
| $^{Me}$F – GCN4 | N-RMKQLEDKVEELLSK$^{\mathbf{Me}}$**F**YHLENEVARLKKLV-C |

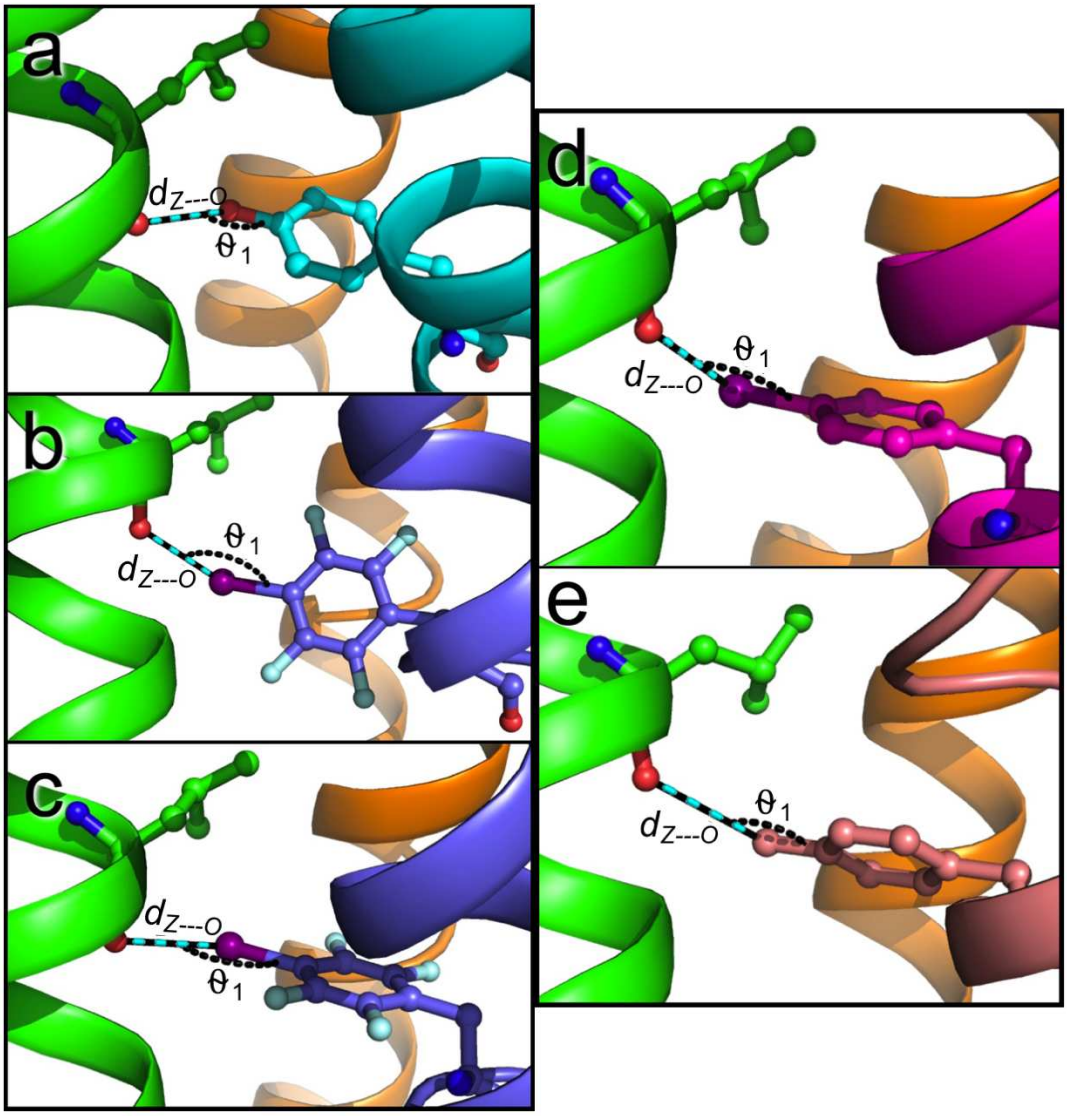



Figure 4.3
A zoom in on the electrostatic or hydrophobic interaction taking place between the $^{Z}$F strand and the A16-bound strand. a) The H-bond that forms between the OH on the tyrosine and the L12 carbonyl oxygen in the A16-bound chain. b) Rotamer A in the $^{TFI}$F16 structure (73% occupancy) with the X-bond to the carbonyl oxygen. c) Rotamer B in the $^{TFI}$F16 structure (27% occupancy) with the X-bond shown. d) Shows the shallow angle of approach the iodine in $^{I}$F16 has to the carbonyl oxygen while e) shows the similar interaction the methyl group in $^{Me}$F16 has.

The specificity and affinity afforded by a classical H-bond was evaluated through an engineered tyrosyl (GCN4-N16$^{OH}$F) sensing strand. The ability of an X-bond to substitute for an H-bond will be compared for an iodinated-Phe (GCN4-N16$^{I}$F) construct, which has a strong but standard σ-hole, and a unique *para*-iodotetrafluoro-Phe (GCN4-N16$^{F4I}$F) unnatural amino acid, where the electron-withdrawing fluorines greatly enhance the σ-hole of the iodine to create an unusually strong X-bond donor [41]. Finally, the role of steric interactions and hydrophobic effects on the sensing peptide will be assessed using a methyl-Phe construct (GCN4-N16$^{Me}$F).

### ○ *4.4.1 Crystal Structures of (GCN4-N16A)$_2$/GCN4-N16$^{Z}$F Heterotrimers*

Mixtures of GCN4-N16A with GCN4-N16$^{Z}$F in 2:1 ratios all formed isomorphous crystals and their X-ray diffraction structures determined to 1.8Å to 2.3 Å resolution. All of the structures of these complexes were seen to be (GCN4-N16A)$_2$/GCN4-N16$^{Z}$F heterotrimers, with the peptides in parallel alignment (figure 4.3). In each structure, the cavity at the A16 position is filled by the aromatic side chain of the $^{Z}$F amino acid from the GCN4-N16$^{Z}$F peptide. The specific or non-specific interaction of the side chain with GCN4-N16A depends on the identity of substituent Z.

The structure of the GCN4-N16$^{OH}$F construct showed the formation of an H-bond from the hydroxyl of the tyrosine to the carbonyl oxygen of the peptide backbone at position L12 of the A16-bound strand of the dimer. The O···O distance and $\theta_1$ angle is near the standard geometries for an H-bond (table 4.2). The expectation, therefore, is that this would be a fairly strong interaction that can provide specificity of the engineered $^{Z}$F-strand for the (GCN4-N16A)$_2$ dimer.

We next asked whether an X-bond could substitute for the OH H-bond of the $^{Z}$F-strand with a GCN4-N16$^{I}$F construct. The structure of this (GCN4-N16A)$_2$/GCN4-N16$^{I}$F construct showed that the iodine was also positioned towards the carbonyl oxygen of L12 of the A16-bound strand (figure 4.3). Although the $d_{Z\cdots O}$ distance was short (≤89% of the sum of the sum of the respective

Table 4.2
Bond geometries from the crystal structures. The Z --- O distance with the calculated percent of the sum of van der Waal radii and the angle of approach the Z substituent takes to the carbonyl oxygen ($\theta_1$) is noted.

| **CONSTRUCT ($^{Z}$F STRAND)** | $R_{Z\cdots O}$ (%$\Sigma R_{VDW}$) | $\theta_1$ |
|---|---|---|
| $^{OH}$F16 – GCN4 | 3.1 Å (101%) | 132° |
| $^{I}$F16 – GCN4 | 3.1 Å (88%) | 112° |
| $^{TFI}$F16 – GCN4 (ROT A) | 2.9 Å (83%) | 135° |
| $^{TFI}$F16 – GCN4 (ROT B) | 3.2 Å (91%) | 136° |
| $^{ME}$F16 – GCN4 | 3.3 Å (94%) | 128° |

van der Waals radii, $\sum R_{vdW}$), the shallow $\theta_1$ angle of approach was not consistent with an X-bond (table 4.2). A quantum calculated electrostatic surface potential (ESP) map of iodobenzene (a model for the $^{I}$F side chain) indicates that a $\theta_1$ angle < 145° would point the oxygen acceptor towards the electronegative surface of the iodine (figure 4.4), which would not be a stabilizing interaction. The additional interactions of the iodine with carbons of various side chains in the cavity suggests that the halogen is behaving as a non-specific hydrophobic substituent and less as a donor in a specific electrostatic-type interaction. To test this hypothesis, we characterized the crystal structure of the GCN4-N16$^{Me}$F16 construct, where a hydrophobic methyl group replaces the iodine ($R_{vdW}$ ~2.0 Å for methyl compared to 1.98 Å for iodine). In comparing the $^{Me}$F16 and $^{I}$F16 structures, we noticed the rings and the methyl and iodine of $^{I}$F and $^{me}$F nearly perfectly superimposed. The similarity among these structures suggests that the behavior of the iodine is structurally similar to that of the methyl group in this system, and that an iodine X-bond could not substitute for the OH H-bond of the tyrosine in the GCN4-$^{OH}$F construct.

The X-bonding potential of a halogen substituent can be enhanced through inductive effects from electron-withdrawing groups, such as fluorine. We thus engineered an $^{Z}$F peptide with a 2,3,5,6-fluoro-4-iodo-phenylalanine ($^{TFI}$F) in place of the N16 residue, creating a GCN4-$^{TFI}$F construct. The ESP surface map of a 2,3,5,6-fluoro-4-iodobenzene model of the $^{TFI}$F side chain shows that the charge neutral point of the iodine is now extended to a $\theta_1$ angle ≈ 125°, which can accommodate a much shallower approach of the oxygen acceptor to the σ-hole of the halogen (figure 4.4). The crystal structure of the GCN4-$^{TFI}$F construct was again a heterotrimer, with the $^{TFI}$F seen to occupy two rotamer positions. The major rotamer (A) shows a 73% partial occupancy, while the minor rotamer B is occupied 27% of the time (figure 4.3 and table 4.2). The rotation of

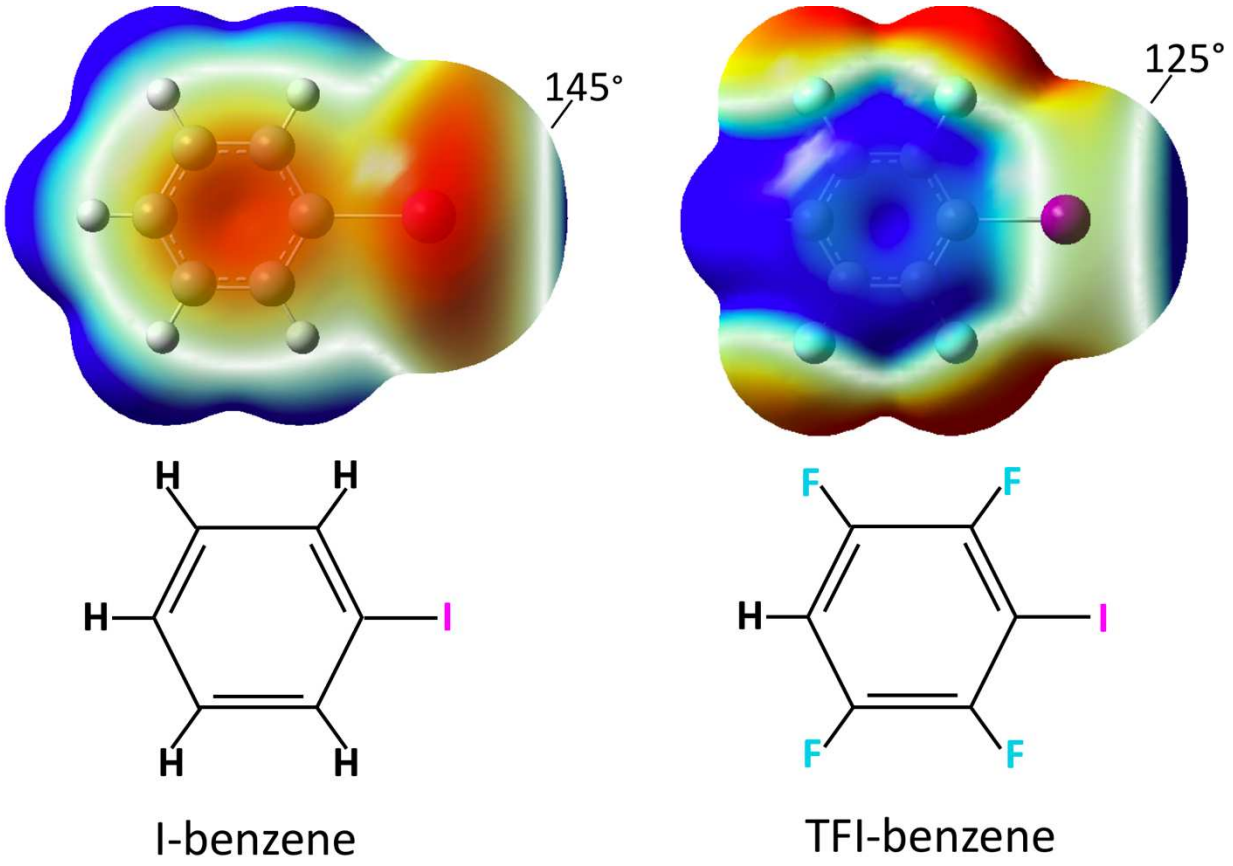



Figure 4.4
Electrostatic potential maps of small molecule mimics of $^{I}F$ (I-Benzene) and $^{TFI}F$ (TFI-benzene). With the replacement of hydrogens for 4 fluorines on the aromatic ring, the σ-hole of the halogen gets much bigger. With this enlargement, the angle of approach an acceptor atom can take widens from 145° to 125°.

the plane of the B-rotamer's aromatic ring is similar to that of the $^{OH}$F , $^{I}$F , and $^{me}$F side chains, but that of rotamer A is rotated ~75°. The iodines in both positions are in geometries compatible with the formation of X-bonds to the carbonyl oxygen of L12 of the A16-bound strand. The B-rotamer shows a slightly longer $d_{Z\cdots O}$ relative to the A-rotamer, while the $\theta_1$ angles are about the same (table 4.2). It is not possible from the structure alone to determine which interaction is stronger, and how either compares to the OH H-bond seen in the GCN4-$^{OH}$F construct. To better understand how the specific H-bond and X-bond and non-specific hydrophobic interactions at this site 16 influence the coiled-coil's overall stability, we carried out differential scanning calorimetry studies.

One structural difference in the (GCN4-N16A)$_2$/GCN4-N16$^{TFI}$F trimer and that of the other complexes is that the intramolecular *i*-*i*+4 H-bonds along the α-helix backbone is elongated from the average 1.5-3.1 Å to 3.5 Å (which is no longer an H-bond) between residues 14 and 18 of the A16-bound peptide. Thus, there is a local unwinding of the helical backbone in order to accommodate this larger $^{TFI}$F side chain and its two rotamers.

### ○ *4.4.2 Thermal Melting Parameters from Differential Scanning Calorimetry*

The crystallographic studies suggest that the (GCN4-N16A)$_2$/GCN4-N16$^{Z}$F heterotrimers fall into two classes. The *Z* = OH or TFI involve electrostatic H- or X-bonds and are expected to provide specificity to the A16-bound peptide for the GCN4-N16A dimer. The *Z* = I or Me are involved in non-specific hydrophobic interactions and, therefore expected to be non-specific. The question at this point is how these predictions from the crystal structures are manifest in solution. We first attacked this question by comparing the DSC-determined thermal melting behavior for each set of complexes, without any preconceived idea as to their oligomeric state.

We started the DSC studies by determining the melting temperature ($T_M$) and melting enthalpy ($\Delta H_M$) of the GCN4-N16A construct, which was previously shown to form a homodimer in solution[40]. The resulting DSC thermal melting curve was fitted with a simple two-state melting process (figure 4.5). The $T_M$ measured by DSC in this study is nearly identical to the previously published value of 62.2 °C [40] and, therefore, we are confident that the GCN4-N16A construct is also a dimer.

Addition of the GCN4-N16$^{OH}$F strand resulted in a thermal melting curve that was very similar to the GCN4-N16A dimer, and similarly fitted with a two-state model with random residuals. Thus, the 2(GCN4-N16A) + GCN4-N16$^{OH}$F mixture appears to be homogeneous in forming a single species. The $T_M$ is increased by 16.3° C, while the $\Delta H_M$ and resulting $\Delta S_M$ are 52% and 38% higher, respectively, than the GCN4-N16A dimer (figure 4.5, table 4.3). These results are consistent with this mixture forming a (GCN4-N16A)$_2$/GCN4-N16$^{OH}$F heterotrimer.

The 2(GCN4-N16A) + GCN4-N16$^{TFI}$F mixture shows the same melting profile as that with the GCN4-N16$^{OH}$F strand, with the $T_M$ increased by 16.1° C, while the $\Delta H_M$ and $\Delta S_M$ are both ~44% higher than the GCN4-N16A dimer (table 4.3). The DSC analysis are, thus, also indicative of a homogeneous species that can be assigned as a (GCN4-N16A)$_2$/GCN4-N16$^{TFI}$F heterotrimer.

The $T_M$s for the mixtures of 2(GCN4-N16A) + GCN4-N16$^{Z}$F (where $Z$ = I or Me) were shifted to even higher temperatures but could not be fitted to a simple two-state model without significant residual errors (supplemental figure 1 in Appendix I). These data were interpreted as being heterogeneous mixtures of oligomers and, therefore, the melting parameters were not analyzed in detail.

The DSC results show that the $Z$ = OH and TFI constructs both form single heterotrimeric species while those of the $Z$ = I and Me constructs could not be assigned to a single oligomeric

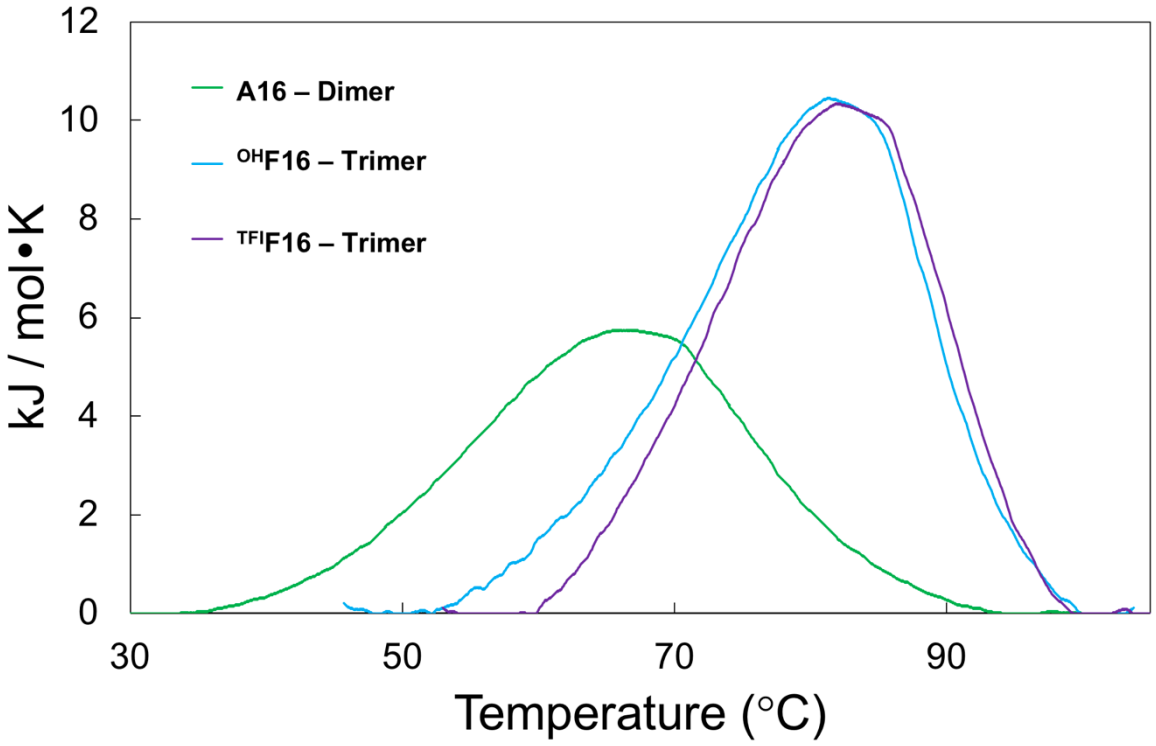



Figure 4.5
Buffer-subtracted and baseline-subtracted representative melting profiles obtained for the A16 dimer, $^{OH}$F16 heterotrimer and $^{TFI}$F16 heterotrimer. A clear shift is seen between the dimer and trimer species with a significant increase in $T_M$. All of the data for these species were fit to a simple two-state scaled model with reasonable residuals around the fit.

Table 4.3. Thermodynamic data obtained from DSC experiments. The $T_M$ and enthalpy ($\Delta H_M$) are found with the two-state fit, while the $\Delta S_M$ is calculated from those two values using the equation $\Delta G = \Delta H - T\Delta S$. The averages and their respective standard deviations from the 4 replicate runs are reported in the table.

| CONSTRUCT | $T_M$ (°C) | $\Delta H_M$ (KCAL/MOL) | $\Delta S_M$ (KCAL/MOL • K) |
|---|---|---|---|
| A16 | 64.4 ± 1.1 | 32.9 ± 1.8 | 97.7 ± 5.3 |
| 2:1 A16: $^{OH}$F16 – GCN4 | 80.7 ± 0.7 | 50.0 ± 1.9 | 134.4 ± 4.8 |
| 2:1 A16:$^{TFI}$F16 – GCN4 | 80.5 ± 0.9 | 47.5 ± 1.7 | 141.1 ± 5.4 |

state. We interpret these results as indicating that hydrophobic interactions in themselves are not sufficient to result in a single unique assembly, but that the electrostatic H- and X-bonds provide the specificity required to assemble a unique trimeric complex. The ~2 kcal/mol higher $\Delta H_M$ for the GCN4-N16$^{OH}$F complex over GCN4-N16$^{TFI}$F complex suggests either that the H-bond is slightly more enthalpically stabilizing than even the highly polarized X-bond, or that the loss of the H-bond in the local unwinding of the GCN4-N16$^{TFI}$F helix has a slight destabilizing effect on the complex, or both.

The ~7 cal/mol·K higher $\Delta S_M$ for the iodo construct suggests that the difference in enthalpic energy is compensated by a gain in entropy, resulting in nearly identical $T_M$ values for the two heterotrimers. We interpret the increased entropy contribution to the TFI construct as coming from burying the hydrophobic halogen into the cavity of the trimer interface, rather than from conformational entropy. An analysis of the normalized crystallographic temperature factors (*B*-factors) for the GCN4-N16A, and the *Z* = OH and TFI trimers shows that the X-bonded complex shows a higher *B*-factor Ser14, associated with the slight unwinding of the local helix to accommodate the larger $^{TFI}$F16 side chain and its two rotamers (figure 4.6). If the entropic difference were conformational, we would have predicted that $\Delta S_M$ would be lower for the X-bonded trimer.

### ○ *4.4.3 Heterotrimer Formation Determined by CD Spectroscopy Titrations*

Although the DSC studies indicated the Y16 and $^{TFI}$F16 coiled-coils assembled exclusively as heterotrimers in solution, the melting parameters did not provide information on the specificity or the stability of this complex at room temperature. In order to estimate the affinities of the $^{Z}$F strands for the GCN4-N16A dimer, we titrated the $^{Z}$F strand into solutions of the A16 dimer. The change in the ratio of ellipticity at 222 nm versus 208 nm ($\Delta(\Delta\varepsilon_{222nm}/\Delta\varepsilon_{208nm})$) determined in the

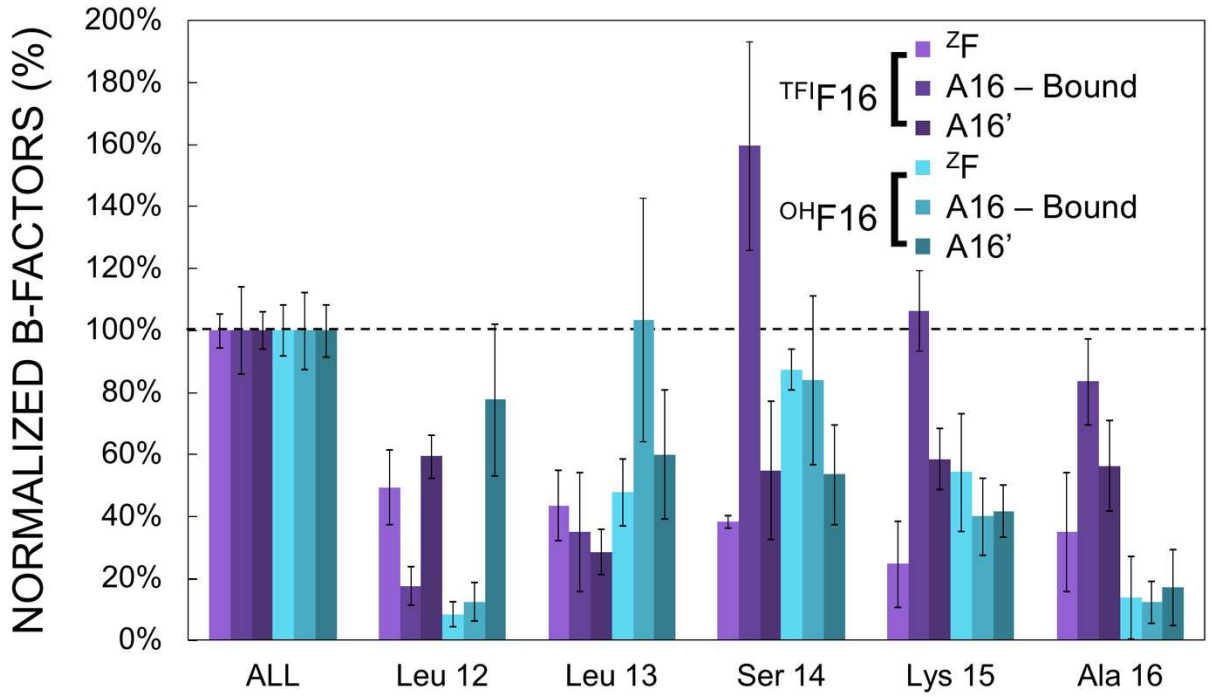



Figure 4.6

B-factor analysis on the $^{OH}$F16 and $^{TFI}$F16 structures. The B-factors for all of the backbone atoms (N, Ca, C, O) in Leu 12 – Ala16 /$^{Z}$F16 in the three chains were averaged and set to 100%. The backbone atoms in the individual residues were then normalized to that average value, and the average of the four normalized values is reported. The error bars are the standard deviation of the mean. While most of the values are below the average value (100%), the A16-bound strand in the $^{TFI}$F16 structure has a very high B-factor associated with Ser14 associated with the local unwinding (loss of i—i+4 H-bond) in the helix to accommodate the large $^{TFI}$F ring.

CD spectra at each titration point is indicative of formation of superhelical oligomers, without preconception of the number of strands in the assemblies. At each titration step, the total concentrations of peptides were held constant at 50μM (10-fold lower than the DSC studies), while varying the ratio of the $^{Z}$F strand to the A16 dimer.

The CD titrations with the $^{Me}$F16 peptide resulted in data that asymptotically approached a maximum at 0% A16 peptide. This binding curve is consistent with a low affinity binding process (figure 4.7d) and was interpreted as evidence the $^{Me}$F16 peptide forms a superhelical homodimer with a $K_D \approx 43$ μM. The CD titration curve for the $^{I}$F16 strand showed the same homodimerization behavior, with a $K_D \approx 14$ μM (figure 4.7c).

The CD titration curve for the $^{TFI}$F peptide, however, was more complex, with the $\Delta(\Delta\varepsilon_{220nm}/\Delta\varepsilon_{208nm})$ increasing nearly linearly from a $^{TFI}$F16 to A16 ratio of 0:2 to 1:2, then falling and then rising again (figure 4.7b). Our interpretation is that the first component is a high affinity binding of the $^{TFI}$F16 to A16 dimer to form the heterodimer at a 1:2 ratio, with $^{TFI}$F16 being limiting. The drop in the superhelicity after the 1:2 ratio comes from loss of the heterotrimer as the A16 strand becomes limiting, leaving excess $^{TFI}$F16 strand in solution. As the $^{TFI}$F16 concentration continues to increase, this peptide forms its own superhelical homodimer. This behavior can be modeled as two competitive binding events, with the formation of the heterotrimer having an apparent $K_D \leq 10$ nM, and the $^{TFI}$F16 homodimer having a $K_D \approx 5.5$ mM.

The titration of the $^{OH}$F peptide into the A16 dimer resulted in a curve that is nearly identical to that of $^{TFI}$F16 (figure 4.7a), with apparent $K_D \leq 10$ nM, and the $^{OH}$F16 homodimer having a $K_D \approx 6$ mM. The $K_D$ for the heterotrimer is considered to be an apparent value, since although the simplest model assumes that the $^{TFI}$F16 or $^{OH}$F binds to a preformed A16 dimer, it is possible that the dimer forms concertedly with $^{TFI}$F16 or $^{OH}$F binding.

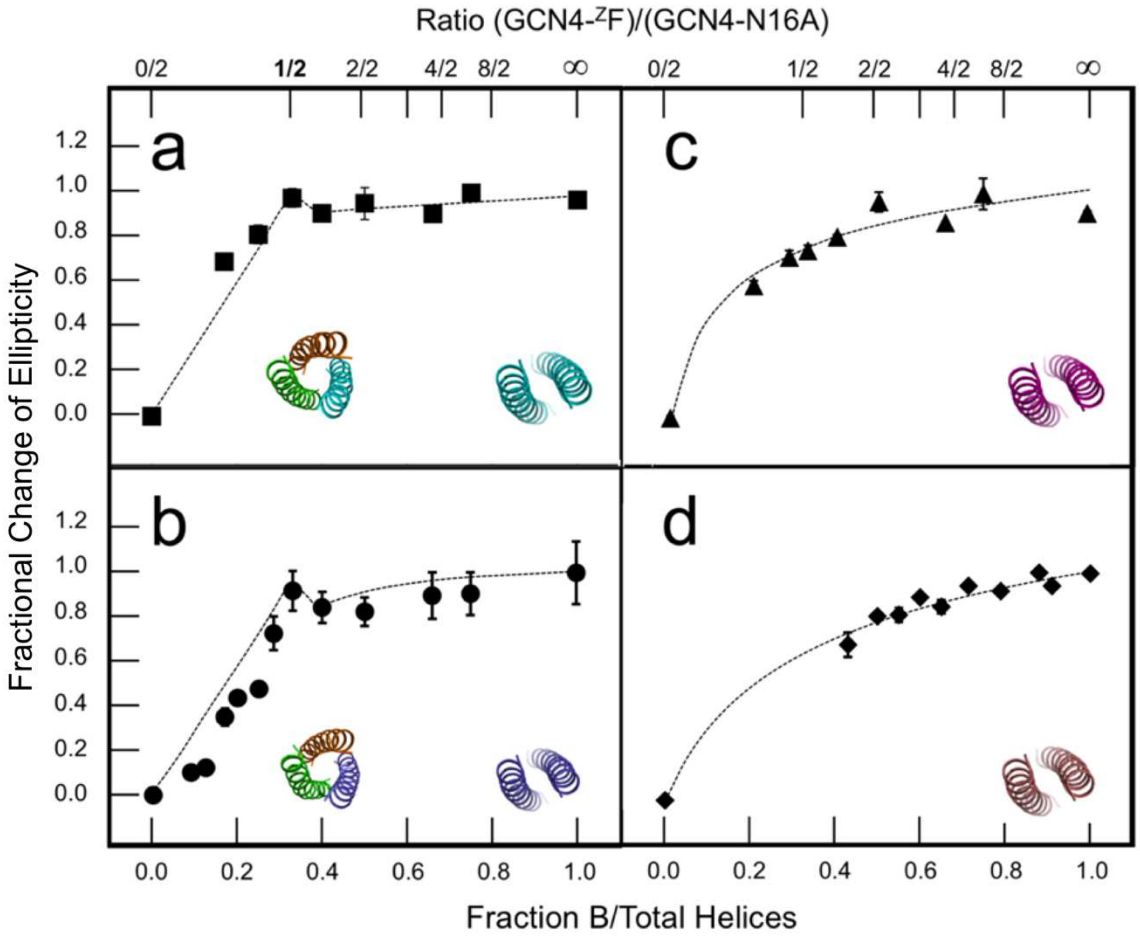



Figure 4.7
CD titration of $^{Z}$F strand into A16. The fractional change of ellipticity at 222 nm versus 208 nm calculated from the CD spectra is plotted against fraction of helices formed or the ratio of $^{Z}$F16 to A16. While A16 is kept constant, it acts as a "limiting reagent" in solution. Once it is all bound up, in the case of $^{OH}$F16 and $^{TFI}$F16, those strands begin to be in excess and start dimerizing with themselves. As $^{Me}$F and $^{I}$F do not bind A16 at this low concentration, they just dimerize with themselves throughout the titration. a) Titration results for $^{OH}$F16. The ratio ellipticity climbs quickly and once a ratio of 1:2 $^{OH}$F16:A16 is met (heterotrimer formation), the slight dip down in the ratio ellipticity and subtle increase indicates the homodimer formation. Similar results are seen in b) with the $^{TFI}$F16 titration. c) shows the titration results for $^{I}$F16, while d) shows similar results when $^{Me}$F16 is titrated into A16.

## 4.5 Discussion

We have shown with this study that a highly stable heterotrimer coiled-coil complex can be designed with an intermolecular electrostatic interaction to the peptide backbone, providing specificity to the complex. Single crystal structures show that the iodine of the unnatural amino-acid $^{TFI}$F forms a short X-bond, while the hydroxyl of the analogous $^{OH}$F construct forms a near ideal H-bond to the peptide oxygen of a neighboring strand. The substituent groups of $^{I}$F and $^{Me}$F sit in the same pocket formed in the A16 dimer; however, these structures suggest that the interactions are primarily van der Waals interactions and, therefore, may not provide the specificity afforded by the X- and H-bonding interactions.

CD titration studies support the predictions for specificity from the crystal structures. Both the $^{TFI}$F and $^{OH}$F constructs were seen to form complexes at ratios of 1:2 $^{Z}$F:A16, consistent with stable heterotrimers, while the $^{I}$F and $^{Me}$F peptides formed only homodimers at these low concentrations. The specificity comes from both the high affinity of the X- and H-bonding peptides with the A16 dimer and low affinity for themselves, thereby, lowering the probability of forming homodimers compared to the $^{I}$F and $^{Me}$F peptides.

In the DSC studies, it is clear that all constructs mixed in 1:2 $^{Z}$F:A16 formed heterotrimers, which is not surprising since these studies were performed at ~10-fold higher concentrations for all species than for the CD studies. These results suggest that the $K_D$ for the binding of $^{I}$F and $^{Me}$F strands to the A16 dimer is between 50μM and 500 μM. Although explicit thermodynamic melting parameters could not be determined for these two constructs, we show that a melting profile for the $^{I}$F homodimers could be derived as the residual after subtracting the $^{TFI}$F:A16 heterotrimer and A16 homodimer profiles (supplemental figure 2 in Appendix I). Thus, it appears that the shift in the $^{I}$F:A16-trimer melting profile to higher temperatures compared to the $^{TFI}$F and $^{OH}$F constructs

is likely due to the presence of significant quantities of $^{I}$F homodimer and not a result of a more thermally stable trimer or presence of higher order structures.

The crystallization of each construct was at millimolar concentrations and, therefore, follows the trend of increasing the probability of heterotrimer formation as the concentrations of material increases. Thus, the oligomeric states of the $^{I}$F and $^{Me}$F constructs are concentration dependent, forming homodimers at the lower concentrations and heterotrimers at high concentrations of the peptides, while the $^{TFI}$F and $^{OH}$F constructs favor the heterotrimers at all concentrations, reflecting the specificity provided by the X- and H-bonds, respectively.

Although the $^{TFI}$F and $^{OH}$F heterotrimer complexes show similar melting profiles and $T_M$s, they reach that point from different thermodynamic routes. The $^{OH}$F complex is stabilized primarily from enthalpic contributions, while the $^{TFI}$F complex shows additional contributions from entropic stabilization of the heterotrimer. The difference of 2.5 kcal/mol in enthalpic stabilization for $^{OH}$F over $^{TFI}$F can be attributed to the electrostatic X- and H-bonds. This interpretation is supported by quantum mechanical (QM) calculations on the interacting components of these assemblies. For these calculations, we constructed ternary complexes of small-molecule mimics, consisting of a hydroxybenzene or tetrafluoroiodobenzene (representing the $^{OH}$F and $^{TFI}$F side chains, respectively), an N-methylacetamide (NMA) to represent the L12 peptide backbone that provides the carbonyl oxygen acceptor for the H- or X-bond, and a second NMA that is H-bonded to the L12 peptide. The components were placed in positions as defined by the crystal structures, with hydrogens added and positioned through QM optimization. The overall energies of the resulting complexes showed that the $^{OH}$F complex is 2.4 kcal/mol more stable than the $^{TFI}$F assembly, with the energy of the latter weighted according to the proportion of the two conformations seen in the crystal structure (figure 4.8). The contribution of each interaction within the ternary complex can

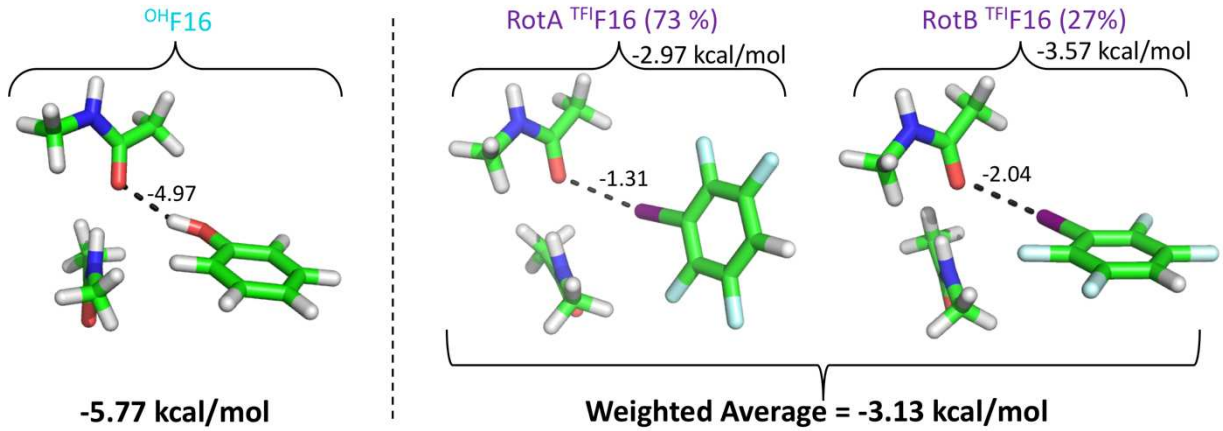



Figure 4.8
Quantum mechanical calculation set-up. Coordinates from L12 and A16 in the A16-bound strand were pulled from the crystal structure and adapted to N-methyl-acetamide (NMA) molecules. Coordinates from ZF were also pulled and adapted to a benzene small-molecule mimic. Energy calculations on each individual piece were carried out and then subtracted from the energy calculation on the ternary structure to get the overall energy of the complex. The energy of the H-bond and X-bonds alone were calculated in a similar fashion.

be estimated by systematically calculating the pair-wise energy. The pair-wise energies, again, show that the H-bond of $^{OH}$F is ~3.6 kcal/mol more stable than the weighted X-bond of $^{TFI}$F.

The compensatory entropic stabilization of the $^{TFI}$F heterotrimer can come from higher conformational or solvent entropy compared to the $^{OH}$F complex. The conformational entropy could, for example, reflect the two conformations seen for the $^{TFI}$F side chain, while the solvent entropy would be associated with burying its more hydrophobic side chain. In order to determine the contribution of conformational entropy, we normalized the B-factors of the two structures and compared the thermal factors for each of the interacting amino acids (figure 4.5). From this analysis, it is clear that the B-factors of the $^{TFI}$F complex are much higher than the comparable residues in $^{OH}$F, suggesting that the interactions of the larger $^{TFI}$F amino acid actually makes this region of the complex more dynamic overall. Thus, we would expect the change in entropy during unfolding to be lower for $^{TFI}$F versus $^{OH}$F, which is the opposite of what is observed from the DSC studies.

The $^{I}$F and $^{Me}$F constructs, however, show that hydrophobicity without the H- or X-bond interaction is not sufficient to provide the same level of stability to the heterodimer, particularly over their homodimeric forms.

This is the first study, to our knowledge, that engineers electrostatic interactions in the interior of a hydrophobic coiled-coil to test their ability to assemble hetero-oligomeric complexes. More specifically, this is the first time an X-bond has been purposefully designed to create a new protein-protein interaction. Coiled-coil assemblies have been the prototype for demonstrating the power of the "knobs-in-holes" concept for the design of intermolecular protein-protein interfaces[42,43].

In this study, we demonstrate that designing a cavity for an aromatic amino acid that also exposes the peptide backbone in a hole provides unique specificity that goes beyond classic hydrophobic effects.

The GCN4 coiled-coil that serves as the framework system for this study provides clear potential applications for the H- or X-bonding peptides characterized here. It would be interesting, for example, to apply these peptides as specific inhibitors to control GCN4 activated gene expression. As a "master" regulator, GCN4 controls ~10% of all genes in yeast, including those associated with protein synthesis and lifespan [44]. An alternative application would be as a cellular probe for localizing GCN4 regulation sites along the DNA and transcription of the genes associated with these sites.

The ability to assemble different helical peptides into heterotrimeric complexes allows us to design systems that bring multiple components together in a specific manner for bioengineering applications. Finally, the concepts developed here of constructing highly specific "knobs-in-holes" interfaces can be applied to control homo- or hetero-oligomerization states in a wide variety of proteins, with the potential to create new allosteric interactions.

REFERENCES


1. Voth, A. R., Hays, F. a & Ho, P. S. Directing macromolecular conformation through halogen bonds. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 6188–6193 (2007).
2. Scholfield, M. R. *et al.* Structure-Energy Relationships of Halogen Bonds in Proteins. *Biochemistry* **56**, 2794–2802 (2017).
3. Schreiber, G. & Fleishman, S. J. ScienceDirect Computational design of protein – protein interactions. *Curr. Opin. Struct. Biol.* **23**, 903–910 (2013).
4. Wu, Y. & Collier, J. H. α-Helical coiled-coil peptide materials for biomedical applications. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology* **9**, 1–17 (2017).
5. Lapenta, F., Aupič, J., Strmšek, Ž. & Jerala, R. Coiled coil protein origami: From modular design principles towards biotechnological applications. *Chem. Soc. Rev.* **47**, 3530–3542 (2018).
6. Godin, A. G., Lounis, B. & Cognet, L. Super-resolution microscopy approaches for live cell imaging. *Biophys. J.* **107**, 1777–1784 (2014).
7. Pauw, D. Nanobodies and their potential applications P erspective. 1013–1026 (2013). doi:10.2217/nnm.13.86
8. Kuroyanagi, H., Ohno, G., Sakane, H., Maruoka, H. & Hagiwara, M. Visualization and genetic analysis of alternative splicing regulation in vivo using fluorescence reporters in transgenic Caenorhabditis elegans. *Nat. Protoc.* **5**, 1495–1517 (2010).
9. Chen, W. *et al.* The Development of Fluorescent Probes for Visualizing Intracellular Hydrogen Polysulfides. *Angew. Chemie - Int. Ed.* **54**, 13961–13965 (2015).
10. Liu, Z. & Tjian, R. Visualizing transcription factor dynamics in living cells. *J. Cell Biol.*

**217**, 1181–1191 (2018).

11. Liu, S. *et al.* Nonnatural protein-protein interaction-pair design by key residues grafting. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5330–5 (2007).
12. Jäckel, C., Kast, P. & Hilvert, D. Protein Design by Directed Evolution. *Annu. Rev. Biophys.* **37**, 153–173 (2008).
13. Yagi, S., Akanuma, S. & Yamagishi, A. Creation of artificial protein–protein interactions using α-helices as interfaces. *Biophys. Rev.* **10**, 411–420 (2018).
14. Coin, I. Application of non-canonical crosslinking amino acids to study protein–protein interactions in live cells. *Curr. Opin. Chem. Biol.* **46**, 156–163 (2018).
15. Lange, A. *et al.* Targeting the Gatekeeper MET146 of C-Jun N-Terminal Kinase 3 Induces a Bivalent Halogen/Chalcogen Bond. *J. Am. Chem. Soc.* **137**, 14640–14652 (2015).
16. Carlsson, A. C. C. *et al.* Increasing Enzyme Stability and Activity through Hydrogen Bond-Enhanced Halogen Bonds. *Biochemistry* **57**, 4135–4147 (2018).
17. Desiraju, G. R. *et al.* Definition of the halogen bond (IUPAC Recommendations 2013). *Pure Appl. Chem.* **85**, 1711–1713 (2013).
18. Scholfield, M. R., Vander Zanden, C. M., Carter, M. & Ho, P. S. Halogen bonding (X-bonding): A biological perspective. *Protein Sci.* **22**, 139–152 (2013).
19. Ho, P. S. Biomolecular Halogen Bonds. *Top. Curr. Chem* **358**, 241–276 (2015).
20. Auffinger, P., Hays, F. a, Westhof, E. & Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16789–16794 (2004).
21. Lange, A., Zimmermann, M. O., Wilcken, R., Zahn, S. & Boeckler, F. M. Targeting Histidine Side Chains in Molecular Design through Nitrogen − Halogen Bonds. (2013).
22. Li, Y., Yu, B. X., Lou, A. Y. & A, T. W. Rational Design of an Orthogonal Molecular

Interaction System at the Complex Interface of Lung Cancer-Related MDM2 Protein with p53 Peptide. 1167–1171 (2016).

23. Lu, Y. *et al.* Halogen bonding--a novel interaction for rational drug design? *J. Med. Chem.* **52**, 2854–2862 (2009).

24. Fanfrílk, J. *et al.* The Effect of Halogen-to-Hydrogen Bond Substitution on Human Aldose Reductase Inhibition. *ACS Chem. Biol.* **10**, 1637–1642 (2015).

25. Wilcken, R. *et al.* Halogen-enriched fragment libraries as leads for drug rescue of mutant p53. *J. Am. Chem. Soc.* **134**, 6810–6818 (2012).

26. 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).

27. Jelesarov, I. & Bosshard, H. R. Thermodynamic characterization of the coupled folding and association of heterodimeric coiled coils (leucine zippers). *J. Mol. Biol.* **263**, 344–358 (1996).

28. Boice, J. A., Dieckmann, G. R., DeGrado, W. F. & Fairman, R. Thermodynamic analysis of a designed three-stranded coiled coil. *Biochemistry* **35**, 14480–14485 (1996).

29. Grigoryan, G. & Keating, A. A. Structural specificity in coiled-coil interactions. *Curr. Opin. Struct. Biol.* **18**, 477–483 (2008).

30. Yadav, M. K. *et al.* Structure-based engineering of internal cavities in coiled-coil peptides. *Biochemistry* **44**, 9723–9732 (2005).

31. Watanabe, S. *et al.* Functional importance of the coiled-coil of the Ebola virus glycoprotein. *J Virol* **74**, 10194–10201 (2000).

32. Liu, N., Caderas, G., Gutte, B. & Thomas, R. M. An artificial HIV enhancer-binding

peptide is dimerized by the addition of a leucine zipper. *Eur Biophys J* **25**, 399-403. (1997).

33. Bianchi, E. *et al.* Covalent stabilization of coiled coils of the HIV gp41 N region yields extremely potent and broad inhibitors of viral infection. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12903–8 (2005).

34. O'Shea, E. K., Rutkowski, R. & Kim, P. S. Evidence that the leucine zipper is a coiled coil. *Science* **243**, 538–542 (1989).

35. Gonzalez, L., Woolfson, D. N. & Alber, T. Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat. Struct. Biol.* **3**, 1011–1018 (1996).

36. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**, 1401–1407 (1993).

37. Ibarra-Molero, B., Makhatadze, G. I. & Matthews, C. R. Mapping the energy surface for the folding reaction of the coiled-coil peptide GCN4-p1. *Biochemistry* **40**, 719–731 (2001).

38. Woolfson, B. D. N. The Design of Coiled-Coil Structures and Assemblies. *Advances in Protein Chem.* **70**, 79–112 (2005).

39. Lupas, A. N. & Gruber, M. The structure of alpha-helical coiled coils. *Adv. Protein Chem.* **70**, 37–78 (2005).

40. Gonzalez, L. An engineered allosteric switch in leucine zipper oligomerization. *Nature* **3**, 510–515 (1996).

41. Bergamaschi, G. *et al.* organocatalysis in water †. *Chem. Commun.* **1**, 4–7 (2018).

42. Schnarr, N. A. & Kennan, A. J. Specific Control of Peptide Assembly with Combined Hydrophilic and Hydrophobic Interfaces. *J. A. Chem. Soc.* 125, 667–671 (2003).

43. Schnarr, N. A. & Kennan, A. J. Peptide Tic-Tac-Toe : Heterotrimeric Coiled-Coil Specificity from Steric Matching of Multiple Hydrophobic Side Chains Thermodynamic Characterization . *J. A. Chem. Soc.* 124, 9779–9783 (2002).

44. Mittal, N. *et al.* The Gcn4 transcription factor reduces protein synthesis capacity and extends yeast lifespan. *Nat. Commun.* **8**, (2017).

# CHAPTER 5: HYDROGEN BOND ENHANCED HALOGEN BONDS TO INCREASE YEAST KIX DOMAIN STABILITY

## ○ 5.1 Summary

In this study, we have site-specifically incorporated unnatural amino acids (meta-halo-tyrosines) into yeast KIX, a protein with an intrinsically disordered region, to better understand how the formation of a Hydrogen Bond enhanced Halogen bond (HBeXB) can affect this class of proteins stability and function. This *in vitro* characterization of engineered yeast KIX will show how the HBeXB can provide up to ~4 kcal/mol of stabilizing potential while minimizing structural perturbations. As structure determines function in the cell, probing how slight structural and energetic changes affect these proteins' (with intrinsically disordered regions) function will be beneficial as these proteins can be associated with cellular dysfunction, so stabilization could be a means to combat some of the unwanted side effects that complement their structure.

## ○ 5.2 Introduction

Proteins that participate in many functions and bind a plethora of factors in the cell often have intrinsically disordered regions, as this state allows them to adopt an assortment of conformations to carry out their various tasks [1]. While this feature is crucial for their role, the instability inherent in this lends these proteins prone to misfolding or aggregating, having a shorter half-life as disordered regions are susceptible to proteasome degradation [2], and if their expression levels are not tightly regulated, the abundance of the low-affinity transient interactions they make could lead to off-target interactions which may lead to disease states [3]. More so, cancer cells benefit from the capacity of these disordered proteins to form assemblages leading to cellular dysregulation[4].

While stabilization of such proteins seems like an "easy" fix to prevent these damaging effects to the cell, on the contrary, locking these proteins into a particular conformation may force them into a certain cellular pathway or prevent them from operating in another crucial pathway. The increased stability may increase binding to certain targets or eliminate binding to others. The structural flexibility innate to these proteins is what defines their cellular function. So, how might their function alter upon stabilization? On a broad scale, how will cells respond to a stabilized version of a protein? In thinking about these proteins as potential therapeutic targets, due to some of the detrimental effects that accompany their structure, could stabilization be a beneficial strategy to combat some of the effects or will other cellular issues ensue from the presence of the stabilized protein?

- ***5.2.1 KIX Domain***

Before tackling these larger questions, we must first choose an appropriate protein to stabilize and carry out *in vitro* experiments to evaluate its engineered stability and function. As such, we elected to work with the KIX domain of the yeast Gal11p complex, a mediator of RNA polymerase II transcription subunit 15 (MED15) involved in transcription of specific genes [5]. The KIX domain is thought to be one of the most important molecular recognition sites for gene regulation as it's required in the assembly of the transcriptional apparatus in not only yeast, but mammals as well. In yeast, KIX interacts with a range of transcription factors including Gal4p, Gcn4, Pdr1/3, and Oaf1, to name a few [5]. It is a key part of the pleiotropic drug response pathway in yeast as it activates transcription of genes that work to pump toxic compounds/antifungals out of the cell (figure 5.1). As such, it is a clinical target in combatting antifungal resistance [6]. While a solution NMR structure of KIX exists [7], no crystal structure of it in its apo state has been solved. However, several crystal structures of the domain bound to certain factors do exists[8–10]. While the

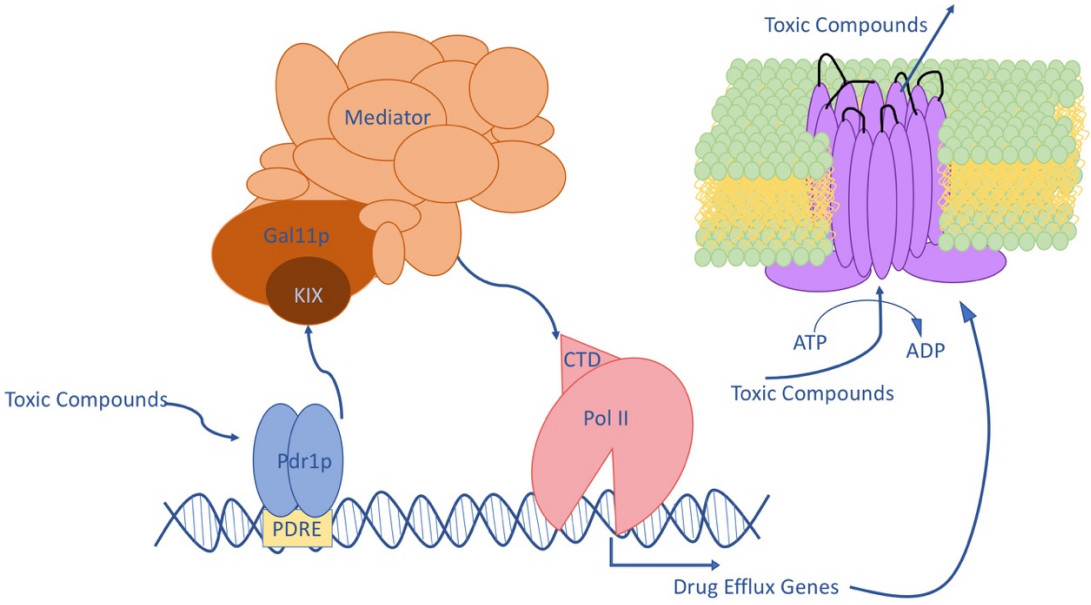



Figure 5.1
Pleiotropic drug efflux pathway KIX is involved in in yeast. Pdr1p (Pleiotropic Drug Response Protein 1) binds its enhancer (PDRE), and antifungals/toxic compounds come and bind Pdr1p, which then binds KIX in the mediator complex. The mediator can then bind the C-terminal domain of Polymerase II to activate transcription of drug efflux genes. These genes then work to pump the toxic compounds/antifungal treatments out of the cell.

yeast KIX structure, according to the NMR results, closely resembles the mammalian KIX structure (figure 5.2), a three-helix bundle, the sequences are only ~25% homologous.

- ***5.2.2 Engineering with HBeXBs***

In regard to how to stabilize this protein, our goal is to maximize the increase in stability with minimal change to the protein sequence. By not adjusting the protein's primary sequence so much, we hope to minimally disturb the overall fold of the protein, so its entire function is not obliterated.

We previously engineered a hydrogen bond enhanced halogen bond (HBeXB) into T4 lysozyme and found with a single mutation (Y18 to $^{mCl}$Y18) the enzyme had increased thermal stability, activity, and an increase in enthalpy of unfolding of ~3 kcal/mol compared to WT [11]. These increases came from the addition of a single atom (Cl) to the sequence that participated in a very strong HBeXB in the structure. As this study showed the amount of energy you can harness with the incorporation of a single atom oriented properly to make a stabilizing HBeXB, we sought to apply this same engineering strategy to yeast KIX. There is a tyrosine conserved across multiple KIX domains (yeast, mouse, and human) that appears to be making an H-bond with a carbonyl oxygen in α1, which is thought to be an intrinsically disordered region[12]. The tyrosine is in the hydrophobic core of the helix bundle near the N-terminus of the protein in α3. Due to its conservation and placement (figure 5.2), this is the tyrosine we have chosen to replace with a meta-halogenated-tyrosine ($^{mX}$Y). With this study on an engineered yeast KIX protein, we also aim to show the progression of enthalpic stability derived from the incorporation of meta-chloro and meta-iodo-tyrosine into yeast KIX. This will allow us to better understand the tunability of the HBeXB in a protein framework as well.


Figure 5.2
Structural alignment of Yeast KIX (orange PDB: 2k0n) and Mammalian KIX (cyan PDB: 2agh). While the sequences are only ~25% homologous, they both have the same three-helix bundle core. Mammalian KIX has two additional $3_{10}$ helices. One near the N-terminus, and the other between α1 and α2. The tyrosine at 66 lies in α3 and is one of the few residues conserved across species.

In considering our overall goal of understanding how stabilized proteins may affect overall cellular function, the first step is engineering a more stable protein. We expect the HBeXB engineering route will allow us to enthalpically stabilize yeast KIX while not altering its overall fold significantly. We anticipate the minimal change in primary sequence to be reflected as a slight structural change. However, even slight structural changes can make a big impact in terms of function. To glean insight on how the engineered KIX protein's function or activity may alter *in vitro*, we have opted to work with one of its binding partners, Pdr1p.

### ○ *5.2.3 Pdr1p Binding Partner*

Pdr1p is a transcription factor that regulates the pleiotropic drug response[13]. It is a zinc cluster protein involved in regulation of multidrug resistance genes. It is thought that xenobiotics bind a ligand-binding domain of Pdr1p, which then allows the activation domain of Pdr1p to bind KIX, a part of the mediator complex. The mediator then interacts with the C-terminal domain of Pol II to recruit Pol II to transcribe genes encoding drug efflux pumps [13] (figure 5.1). This is a mechanism in yeast responsible for multidrug resistance as the drug efflux pumps propel antifungal drugs out of the yeast, rendering them useless. As the binding of KIX to Pdr1p is a crucial part of this pathway, examining how the engineered KIX mutants bind this partner will be an informative first step for evaluating if and how one of the protein's functions has been affected *in vitro*. We will be using a 12mer peptide from Pdr1p for this study (Pdr1p-12mer) that was previously found to bind KIX and induce structural changes from an NMR titration [13].

### ○ *5.2.4 Summary of Study*

In this study, we have engineered a meta-chloro-tyrosine and meta-iodo-tyrosine into yeast KIX, a protein with an intrinsically disordered region, and carried out initial biophysical experiments (DSC and CD) characterizing the engineered proteins *in vitro*. In applying HBeXB

engineering efforts to this protein we will be able to address 1) how the meta-halo-tyrosines affect yeast KIX's structure, 2) if the meta-halo-tyrosine can stabilize the protein, and 3) how meta-halo-tyrosine influences one of KIX's functions – binding to Pdr1p-12mer. Answering these questions through an *in vitro* analysis will be an important first step for future *in vivo* experiments.

### ○ 5.3 Experimental Section

To carry out the *in vitro* characterization, we first engineered the KIX proteins containing meta-chloro and meta-iodo-tyrosine through non-canonical amino acid incorporation. We then carried out circular dichroism and differential scanning calorimetry studies to test the mutant's secondary structural changes, stability, and binding potential compared to WT KIX.

#### ○ *5.3.1 Protein expression*

All KIX constructs started from the yeast KIX gene, with a 6-His tag appended to the C-terminus for purification ease (figure 5.3). The WT KIX gene was ordered from Geneblock codon optimized in a pET26 vector, which bears a kanamycin resistance gene. The plasmid used for non-canonical amino acid incorporation was generated from site-directed mutagenesis of the WT plasmid. The codon from Y66 (TAC) was changed to the amber stop codon (TAG) (so a C to G mutagenesis), which is the codon most typically utilized for non-canonical amino acid incorporation[14] .

WT KIX plasmid was transformed into BL21 codon plus *E. coli.* cells via heat shock at 42°C for 20 seconds. 1 µL of 5 µg/µL of WT KIX plasmid was added to 75 µL of thawed BL21 codon plus cells. The cells sat on ice for 20 minutes after addition of the plasmid and then underwent heat shock to open up the cell walls to allow the plasmid into the cell. After the heat shock, the cells were brought up to 1 mL with LB media and were placed in an incubator to recover


Figure 5.3
Gene construct design for KIX expression. Restriction enzyme sites (NdeI and XhoI) were incorporated for cloning purposes. The 6 HIS tag was purposefully placed on the C-terminus for purification ease of $^{mCl}$Y66 and $^{ml}$Y66 KIX proteins. As codon 66 was changed to the Amber stop codon (TAG) for expression of these proteins, only full length protein with the non-canonical amino acid properly incorporated would have the 6 HIS tag. This allows us to purify away the truncated protein (not containing the HIS tag) from the full length in the initial nickel column purification step. The GSSS is a linker that could be cleaved for removal of the HIS tag.

at 37°C for 1 hour with shaking at 220 rpm. After one hour, approximately 200 µL of the cells were plated onto agar plates with appropriate antibiotics (kanamycin for WT KIX and chloramphenicol for the codon plus cells), and the plates were placed in a 37°C incubator to grow overnight. Colonies were picked from the plates the next day and put in liquid LB (with appropriate antibiotics) to confirm plasmid incorporation, and glycerol stocks were made from these starters.

Several experiments were carried out to optimize the conditions associated with incorporation of the non-canonical amino acids. Initially, control experiments with a TAG-GFP construct were carried out to test the tRNA-synthetase's ability to incorporate the meta-halogenated tyrosine in the given media and temperature conditions (figure 5.4a), and evaluate the incorporation efficiency through SDS-PAGE. Once we verified the incorporation, several cell lines were tested including pLysS and codon plus, before settling on DE3 which showed the highest expression. Throughout this process, various induction OD values were tested (cultures were induced at $OD_{600}$ values ranging from 0.5 – 1.5), as well as expression times. An $OD_{600}$ ~0.75 and a 17-hour expression at 23 °C was found to yield the highest amount of protein.

For protein expression, WT KIX cells from the glycerol stocks were grown in 2xYT with appropriate antibiotics (kanamycin and chloramphenicol) at 37°C while being shaken at 220 rpm until the culture reached an $OD_{600}$ of 0.75. IPTG was then added (to a concentration of 1mM) to induce protein expression in the cultures, and the temperature was dropped to 23°C for expression over 17 hours. The temperature was dropped to reduce the likelihood of the protein unfolding or adopting a non-native structure. After expression, cells were harvested by centrifugation at 7000 rpm for 30 minutes, and the supernatant was decanted, and the bacterial pellet stored at -80°C until purification.

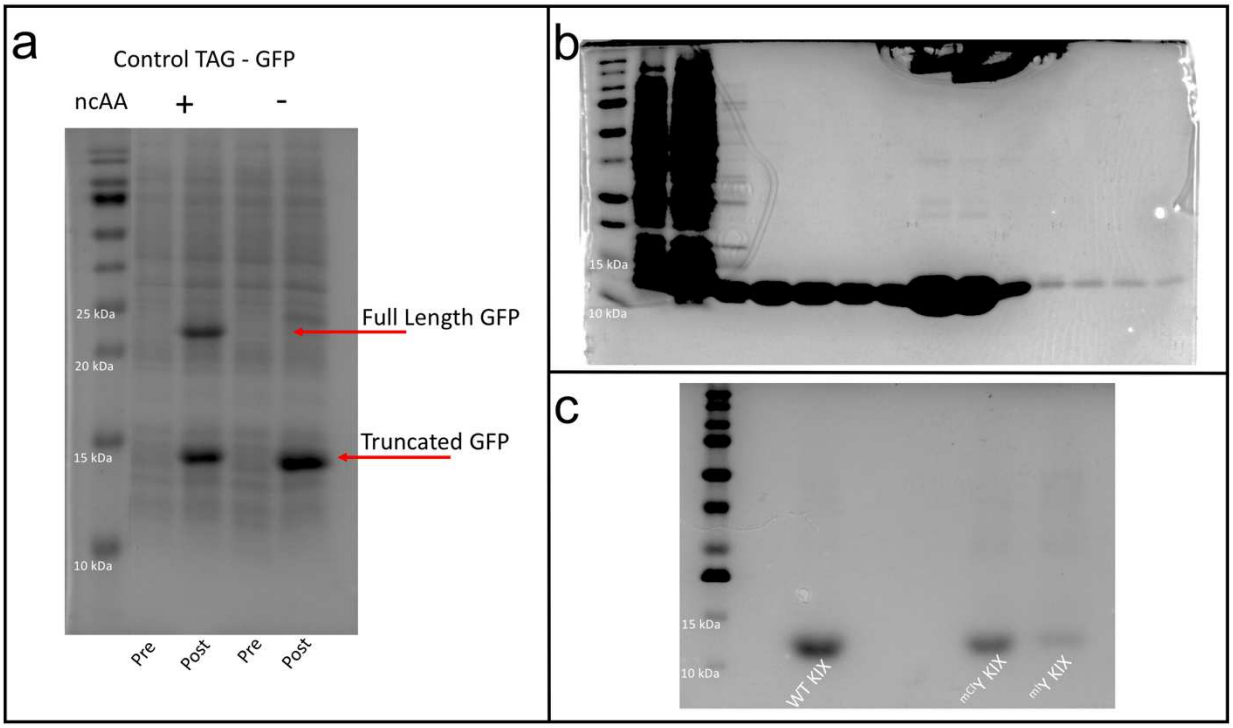



Figure 5.4

Example protein expression and purification gels. a) Control GFP study done to evaluate efficiency of the tRNA-synthetase pair incorporating meta-halo-tyrosine into GFP. Without the non-canonical amino acid (ncAA) present no full length protein is produced, however expression with the ncAA revealed about a 50% incorporation yield. b) Fractions of WT KIX after Ni column purification step. Fractions were run on a 15% polyacrylamide gel in order to determine which fractions to pool and concentrate for further purification/buffer exchange. c) Final purified WT, $^{mCl}$Y66, and $^{mI}$Y66 KIX proteins after gel filtration column showing purity and relative expression yields with ncAA.

The pET26 plasmid bearing the TAG KIX sequence was transformed into BL21 DE3 cells, and this cell line was made competent to allow for a subsequent transformation of the *pDule2-Mb-ClTyrRSC6* plasmid. Competent cell prep started by making Inoue transformation buffer (55 mM manganese chloride, 15 mM calcium chloride, 250 mM potassium chloride, 10 mM PIPES pH 6.7). A starter culture of the DE3 cells containing the TAG KIX plasmid and kanamycin was added to 250 mL of LB. When an $OD_{600}$ of 0.6 for the cells was reached, the cells were spun down at 3900 rpm at 4°C for 10 minutes. The supernatant was subsequently poured off and the pellet resuspended in 80 mL of Inoue transformation buffer. This 80 mL of cells in Inoue buffer was centrifuged at 3900 rpm at 4°C for 20 minutes and the supernatant poured off again. The pellet was resuspended in another 20 mL of Inoue transformation buffer and 1.5 mL of DMSO was added. The cells were aliquoted and stored in the -80°C freezer. These cells were then ready to be transformed with the plasmid containing the unnatural amino acid incorporation machinery. The *pDule2-Mb-ClTyrRSC6* plasmid contains the orthogonal *Mb* $tRNA_{CUA}$ and 3-halo-Tyr amino acyl-tRNA synthetase to allow for incorporation of specifically meta-halogenated tyrosine amino acids. The BL21 DE3 cells bearing both the KIX and tRNA/synthetase machinery plasmids were grown in LB media with the appropriate antibiotics (kanamycin for the TAG KIX plasmid and spectinomycin for the tRNA/synthetase plasmid) for expression at 37°C while being shaken at 220 rpm. The required non-canonical amino acid ($^{mCl}$Y or $^{mI}$Y) was added to a final concentration of 1mM before beginning expression. When an $OD_{600}$ of 0.75 was reached, IPTG to 1mM was added to induce expression of $^{mCl}$Y66 or $^{mI}$Y66 KIX protein. After induction, the cultures were allowed to grow for an additional 17 hours at 23°C. After expression, the cells were harvested by centrifugation at 7000 rpm for 30 minutes. The supernatant was subsequently decanted, and the bacterial pellets stored at -80°C until purification.

- ***5.3.2 Protein Purification***

The frozen bacterial pellets were resuspended in ~15-20 mL of buffer A (500 mM sodium chloride, 50 mM TRIS, and 0.02% sodium azide pH 7.7) and allowed to gently thaw on ice. The thawed pellet was lysed by sonication on ice for three 30s intervals on a Branson Sonifier 450 sonicator (duty cycle of 70%, output control of 7). After cell lysis, the suspension was centrifuged at 17000 rpm and 4°C for 45 minutes. The supernatant was then filtered through a 0.45 μm pore syringe filter. The filtered lysate was loaded on to the AKTA start FPLC system applying 10% buffer B (500 mM sodium chloride, 50 mM TRIS, 0.02% sodium azide, and 500 mM imidazole pH 7.7) onto a 5 mL HisTrap HP column. Non-bound protein was washed out with 17% buffer B over 3 column volumes. The His-tagged KIX was eluted over a 75-100% buffer B gradient over 5 column volumes. Fractions thought to contain the protein, according to the chromatogram, were run on a 15% polyacrylamide gel (figure 5.4b) and then combined and concentrated at 4°C in a 10000 MWCO Amicon Ultra Millipore device. The concentrated protein was then loaded onto a GE PD MiniTrap G-25 gravity column equilibrated with appropriate buffer specific for DSC or CD experiments (figure 5.4c). The DSC buffer was 100 mM Na/K phosphate and 200 mM sodium chloride pH 6.2, and the CD buffer was 50 mM TRIS and 150 mM sodium perchlorate pH 6.2. After the buffer exchange, the protein was aliquoted and stored in -80°C until needed.

- **5.3.3 Pdr1p-12mer Peptide Purification**

A 12mer peptide with sequence N-EDLYSILWSDVY-C was ordered crude from Biomatik and came 57% pure. The lyophilized peptide was resuspended in HPLC Buffer A (0.1% trifluoroacetic acid (TFA) and 99.9% water) and purified on a C18 semi-preparative column with a gradient between 0.1% TFA in water and 0.1% TFA in acetonitrile. HPLC fractions containing

the full-length pure peptide were dried down in a speed vacuum and resuspended in the appropriate buffer for the set of experiments.

- ***5.3.4 Differential Scanning Calorimetry***

Protein samples at 1 mg/mL in 100 mM Na/K Phosphate and 200 mM sodium chloride pH 6.2 were used for DSC experiments. Melting profiles were collected on a TA Instruments Nano DSC model 602001 under 3 atm constant pressure. The concentration of at least 1 mg/mL was required to obtain sufficient signal-to-noise from the instrument. Samples were set to equilibrate for 600s before a heating cycle from 0 – 90°C at a scan rate of 1°C/min followed by a cooling cycle to confirm refolding reversibility. A minimum of four replicate experiments were performed on each protein. The $T_M$, $\Delta H_M$ and $\Delta S_M$ values reported are averages from the four scans with a standard deviation calculated from the replicates as well.

Melting data were analyzed and thermodynamic properties determined with NanoAnalyze Data Analysis, version 3.6.0, from TA Instruments. All of the scans were background subtracted to eliminate heat changes from the buffer over the temperature range. The melting temperatures ($T_M$) and melting enthalpies ($\Delta H_{fit}$) were found by fitting the buffer subtracted and baseline corrected data to the two-state scaled model, as the folding/re-folding process is reversible. The $\Delta H_{cal}$ was found from the baseline subtracted raw data and compared to the $\Delta H_{fit}$. One measure of "goodness of fit" for the DSC data is found in the $\Delta H_{fit}/\Delta H_{cal}$ ratio. This ratio should be as close to 1 as possible, and all of the ratios calculated were in the 0.97-1.01 range. Another value to consider is the Aw value. It is a value that scales the data from the $\Delta H_{cal}$ according to the user's concentration and molecular weight input. This value should also be 1, indicating the proper values are being used to fit the data. The Aw's ranged from 0.99-1.07 for all the data. The software

automatically calculates a $\Delta S_M$ from the $\Delta H_{cal}$ due to the property of $\Delta G = 0$ at the $T_M$, so $\Delta H = T\Delta S$ from the equation $\Delta G = \Delta H - T\Delta S$. A $\Delta S_M$ from the $\Delta H_{fit}$ can be calculated in the same way.

- ***5.3.5 Circular Dichroism Studies***

CD spectra were collected on just the KIX proteins (apo), as well as the KIX protein with ligand peptide present to understand the effects the $^{mX}$Y has on the fold of the protein as well secondary structural changes that ensue upon binding the peptide. Protein samples at 5 µM were prepared in 50 mM TRIS and 150 mM sodium perchlorate pH 6.2. Samples were placed in a 1 mm cuvette and at least three scans were collected on each sample. Scans from 190 to 285 nm were taken at a rate of 1 nm/sec. The spectra from the buffer was subtracted out of each of these. The mean residue ellipticity was calculated for each scan and the averages graphed. The spectra were put through the algorithm BeStSel[15] to calculate the % helicity. These results correlate with the 222/208 ratios calculated from the spectra as well. CD experiments of the KIX protein with the ligand peptide were also performed in triplicate. Samples of different KIX protein: Pdr1p-12mer peptide in molar ratios of 10:1 5:1 2:1 1:1 1:5 and 1:10 were tested to understand how the secondary structure of the protein changes as it binds the peptide. The KIX protein was held at a constant 5 µM throughout the titration with increasing molar amounts of peptide added. As far as analyzing this data, the CD signal from the buffer was subtracted out and the signal from the peptide at its concentration was also subtracted out. The Pdr1p-12mer peptide alone looked like a classic example of a "random coil." The 222/208 spectral shifts were monitored throughout the titration indicating how the structure of the protein changes in relation to the amount of peptide present. All of these spectra were also put into BeStSel[15] to calculate the % helicity at each titration point as well.

- **5.4 Results**
- ***5.4.1 Protein Expression with Non-Canonical Amino Acids***

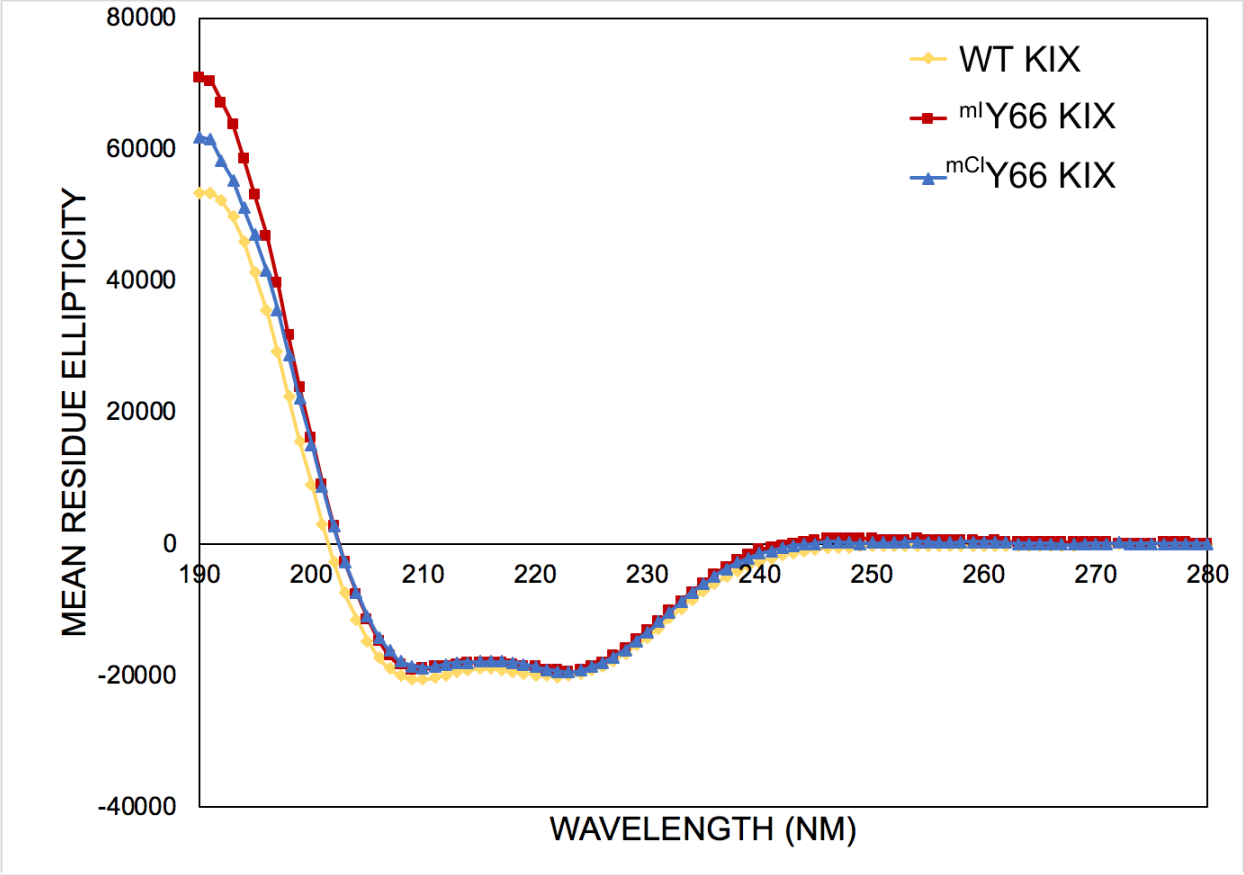
Expression of WT KIX in 1 L of media typically yields around 2 mL of 20 mg/mL. The $^{mCl}$Y66 KIX expression normally generates around 1 mL of 10 mg/mL, or about 25% of the WT expression amount. The $^{ml}$Y66 KIX expression has a slightly lower yield than the $^{mCl}$Y66 KIX at about 1 mL of 8 mg/mL (figure 4c). It is fairly common in this field to get around 10% of the WT expression level when expressing a protein with a non-canonical amino acid, so the current incorporation strategy is surpassing this mark.

- ***5.4.2 Circular Dichroism of Apo KIX Proteins***

These experiments were done to address if all of the proteins were folding properly with the incorporation of the $^{mX}$Y amino acid and maintaining their expected helical rich structure, as suggested by NMR and crystallography. The scans of just the purified proteins alone (WT, $^{mCl}$Y66, and $^{ml}$Y66) (figure 5.5) indicate that all of the proteins maintain their secondary structure as the dominating secondary structure is α-helical. The calculated MRE from the raw data was put into BeStSel[15] and the % helicity calculated with the algorithm revealed the engineered KIX proteins to have as much as 10% higher helicity. This may indicate that the 3-helix bundle of the protein is tighter, or the partially unfolded helix (α1)[16] in the structure has folded with $^{mX}$Y present.

- ***5.4.3 Differential Scanning Calorimetry to Assess KIX Stability***

Once we determined the $^{mX}$Y were not disrupting the proteins fold significantly, DSC studies were carried out to assess the engineered proteins thermal stability and understand if the halogen was providing any enthalpic stability. Prior to carrying out the data analysis described above (5.3.4), we needed to address whether the yeast KIX domain adopts a simple two state unfolding – folding mechanism, in order to fit the data to the simple "two-state-scaled" model. To

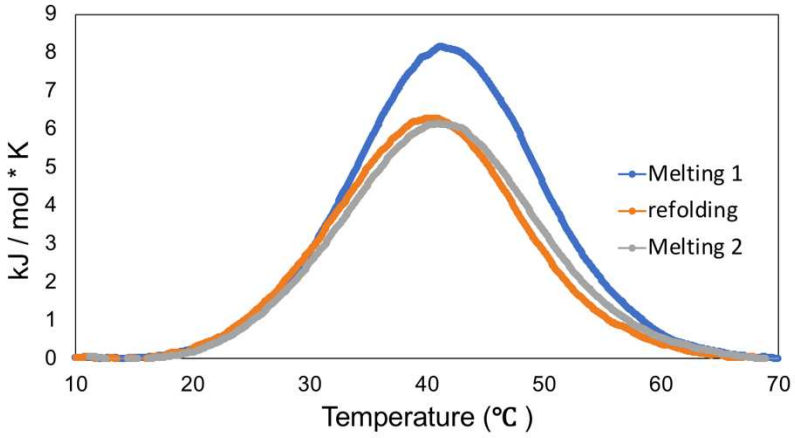



| Contruct | BeStSel % Helicity |
|---|---|
| WT KIX | 66.9 |
| $^{mCl}$Y66 KIX | 72.6 |
| $^{ml}$Y66 KIX | 80.8 |

Figure 5.5
CD spectra of Apo WT, $^{mCl}$Y66 and $^{ml}$Y66 KIX proteins. When the Mean Residue Ellipticity calculated from the spectra was put through BeStSel, the % helicity calculated increased from the WT to the $^{ml}$Y66 KIX protein.

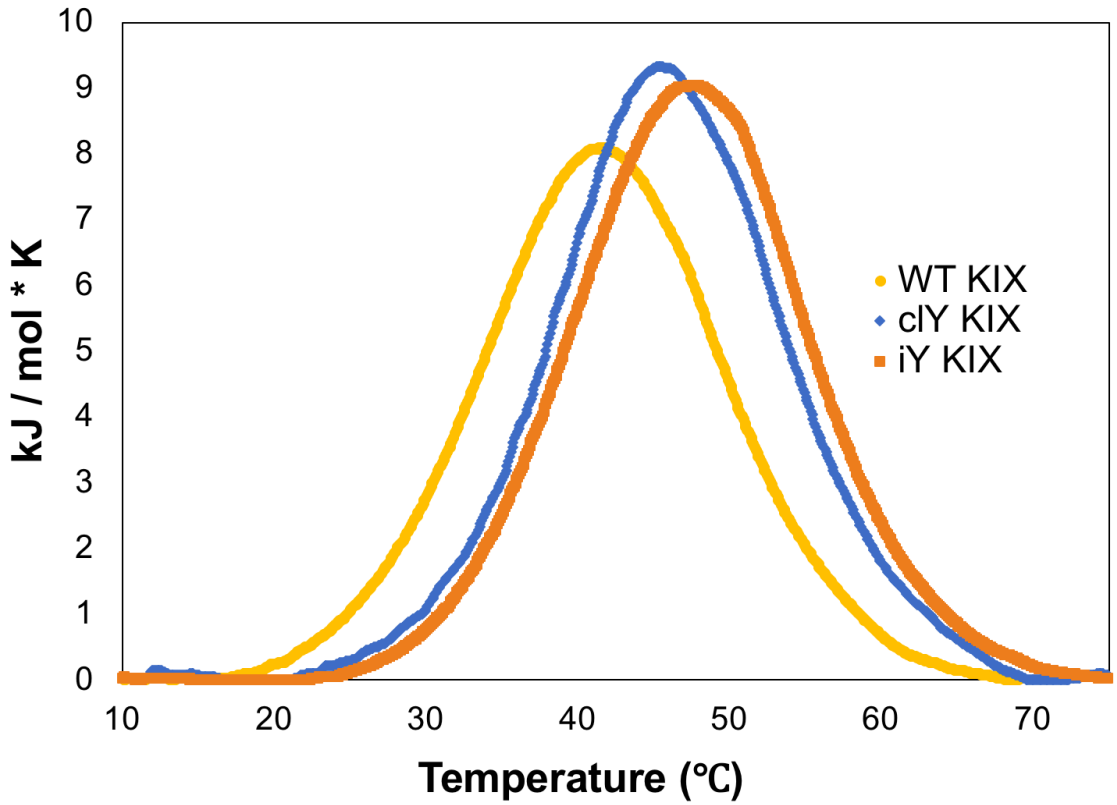
do this, we carried out the normal heating scan on the WT protein, and then performed a cooling scan from 90°C - 0°C at 1°C/minute where the protein was allowed to cool and refold. Following this cooling scan, we did another heating scan from 0°C - 90°C at 1°C/minute. The area under each heating (2 of them) and cooling scan (1) was calculated as it represents the relative enthalpy of unfolding or folding. To assume a simple two-state unfolding-folding mechanism, the enthalpy of folding and refolding should be the same [17]. However, because not 100% of the protein refolds properly, the relative enthalpies need to be scaled according to the amount of protein that refolded. In looking at the two heating scans, the enthalpy of the second heating scan is ~77% of the initial heating scan, implying ~77% of the protein refolded properly. The folding enthalpy was then scaled to the amount of protein that refolded (77%), and the difference between the unfolding and folding enthalpies differ by only 0.3 kJ/mol, suggesting the yeast KIX follows a simple two-state folding-refolding mechanism (figure 5.6). Troilo *et al.* also found KIX to adopt a simple two-state fold-refolding mechanism through stopped-flow and equilibrium experiments [16].

The melting profiles from all three proteins revealed an increase in $T_M$ with the addition of the unnatural amino acid (figure 5.7). The KIX domain became ~3°C more stable than WT with the addition of $^{mCl}$Y66 and gained another ~2°C (so 5°C more stable than WT) with the addition of $^{mI}$Y66 (table in figure 5.7). This alone tells us that the protein has become more thermally stable, but taking a closer look at the enthalpy (ΔH) and entropy (ΔS) will help us discern if the amino acid incorporated is forming a stabilizing bond, or just acting as a hydrophobic-space filling molecule in this context. In looking at the $\Delta H_M$, both engineered proteins have gained a couple kcal/mol of energy at the $T_M$, with the $^{mCl}$Y66 KIX gaining ~2 kcal/mol, and the $^{mI}$Y66 KIX gaining ~3.5 kcal/mol. The $\Delta S_M$ can be calculated from the $T_M$ and $\Delta H_M$ though the equation $\Delta G = \Delta H$ -




| Heating Cycle | Area/Enthalpy (kJ/mol *K) | % Enthaly relative Initial Heating | $T_{max}$ (°C) |
|---|---|---|---|
| Melting 1 | 161.66 | 100 | 41.6 |
| Refolding | 124.70 | 77 | 40.0 |
| Melting 2 | 126.15 | 78 | 41.1 |

Figure 5.6
DSC analysis to verify the KIX domain has a simple two-state unfolding-refolding behavior. The $T_{max}$'s for the two melting profiles are almost identical, and the calculated % refolded protein from the cooling scan almost exactly matches the relative enthalpy calculated from the 2$^{nd}$ melting profile.




| Construct | $T_M$ (°C) | $\Delta H_M$ (kcal/mol) | $\Delta S_M$ (cal/mol * K) |
|---|---|---|---|
| WT KIX | 41.9 ± 0.1 | 39.9 ± 0.6 | 126.8 ± 2.0 |
| $^{mCl}$Y66 KIX | 44.9 ± 0.23 | 41.9 ± 0.9 | 131.5 ± 3.0 |
| $^{mI}$Y66 KIX | 46.6 ± 0.1 | 43.6 ± 0.6 | 136.5 ± 2.0 |

Figure 5.7
The graph shows representative DSC melting profile for WT, $^{mCl}$Y and $^{mI}$Y KIX proteins. A clear shift in the $T_M$ can be seen from these profiles. The table displays the average and standard deviation of the thermodynamic information calculated from the four replicate DSC scans done on each protein.

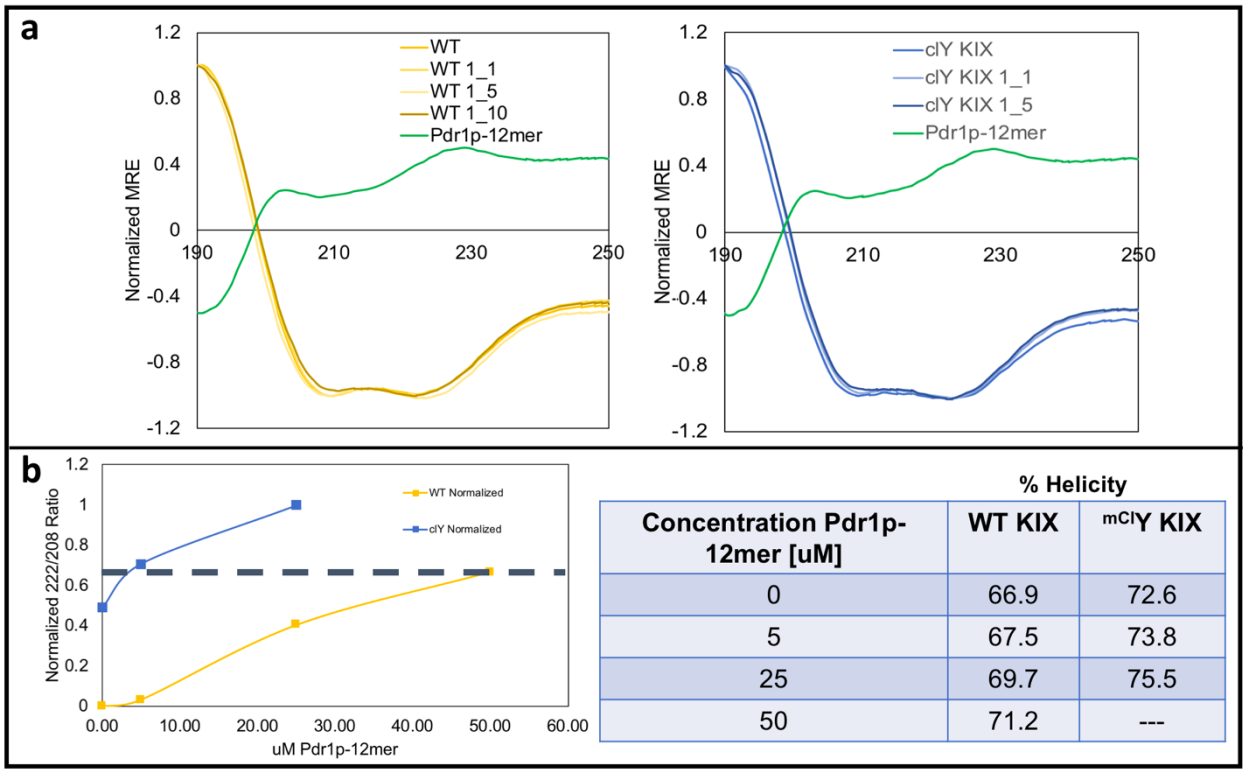
$T\Delta S$ as $\Delta G$ at the $T_M$ is 0. This increase in $\Delta H_M$ corresponds to an increase in $\Delta S_M$ of ~4.5 cal/mol * K for the $^{mCl}$Y66 KIX, and an increase of ~9.2 cal/mol * K for the $^{mI}$Y66 KIX.

- ***5.4.4. CD Titration of KIX with Binding Partner Pdr1p-12mer***

A CD titration was also carried out on the WT and $^{mCl}$Y66 KIX proteins to ask if the engineered proteins can bind Pdr1p-12mer, and what (if any) structural changes ensue upon binding. Over the course of the titration an increase in helicity is noted by an increase in the ellipticity at 222 nm /208 nm ratio. The 222/208 changes from a ratio of 0.97 to 1.03 for the WT KIX and 1.02 to 1.06 for the $^{mCl}$Y66 KIX. Furthermore, this ~0.04 increase in 222/208 ratio occurs at a lower concentration of Pdr1p-12mer present for $^{mCl}$Y66 KIX compared to WT (figure 5.8).

- **5.5 Discussion and Conclusions**

In this preliminary *in vitro* work on yeast KIX, we see through non-canonical amino acid incorporation, we have not significantly affected the fold of the protein with the incorporation of $^{mX}$Y at position 66, however, we have significantly increased the thermal stability of yeast KIX with this engineering strategy. The two engineered proteins $^{mCl}$Y66 and $^{mI}$Y66 KIX both have higher significantly higher melting temperatures ($T_M$) and melting enthalpies ($\Delta H_M$) compared to WT. We have also seen with the CD titration with Pdr1p-12mer, that $^{mCl}$Y66 KIX potentially binds the peptide at a "higher affinity" compared to WT KIX as noted by a larger shift in the 222 nm/208 nm ratio at lower concentrations of Pdr1p-12mer. In our previous engineering efforts of T4L with meta- chlorotyrosine, we only slightly increased the melting temperature (~1°C) of the protein. Herein, we have shown how impactful HBeXB engineering efforts are in terms of stabilizing a protein with an intrinsically disordered region. The addition of the single chlorine or iodine atom has increased the $T_M$ by 2 and 3 °C respectively and allowed the protein to gain an additional 2-4 kcal/mol of enthalpy.




| Concentration Pdr1p-12mer [uM] | % Helicity WT KIX | % Helicity $^{mCl}$Y KIX |
|---|---|---|
| 0 | 66.9 | 72.6 |
| 5 | 67.5 | 73.8 |
| 25 | 69.7 | 75.5 |
| 50 | 71.2 | --- |

Figure 5.8

a) Shows graphs of the CD spectra titration study. A slight change can be seen as more Pdr1p-12mer is added to the protein. b) A graph showing the change in 222/208 ratio vs. the concentration of Pdr1p present throughout the titration, and a table showing the % Helicity as calculated from BeStSel. The $^{mCl}$Y KIX approaches a 222/208 ratio of 1 at a much lower concentration of Pdr1p present compared to WT KIX. The $^{mCl}$Y KIX also has a higher helicity throughout the titration compared to WT KIX.

While these results indicate the increased stability and binding activity of the protein are due to the HBeXB bond, without a definitive crystal structure, it is hard to say if the halogen on the meta-halotyrosine is making a halogen bond and, moreover, if it is being enhanced by the hydroxyl group. In looking at the ESP maps of 2-chlorophenol and 2-iodophenol, we see the neutral point on the surface of the molecule extends to 150° and 125° respectively when the hydrogen is pointed towards the negative annulus of the halogen, which suggests acceptor atoms can approach the halogen at angles as low as those, and still be participating in an X-bond (figure 5.9). As the protein has a slightly unfolded region, the incorporation of the meta-halotyrosines could slightly rearrange atoms in the helical bundle making it hard to predict where the halogen aligns. However, we can model meta-halotyrosines into existing KIX structures to perhaps get an idea of how the halogen could be behaving.

In modeling meta-chlorotyrosine into an NMR structure of yeast KIX (PDB: 2k0n), we see the chlorine could have a $\theta_1$ as high as 162° with an interaction distance to the oxygen of 3.0 Å, which is ~92% of the van der Waal radii (figure 5.10). We have carried out some theoretical quantum mechanical calculations on small-molecule mimics, 2-halophenols interacting with N-methylacetamide (NMA), at various distances and angles to better understand the energy profile. These simplified calculations show when 2-chlorophenol interacts with NMA at a $\theta_1$ of 160° and a distance of 3.0 Å to the acceptor atom, the energy of the interaction is estimated to be 1.3 kcal/mol. The energy of the intramolecular H-bond to the halogen is estimated to be ~1.85 kcal/mol from our previous T4L engineering efforts[11]. Together, these bonds would provide ~3.2 kcal/mol

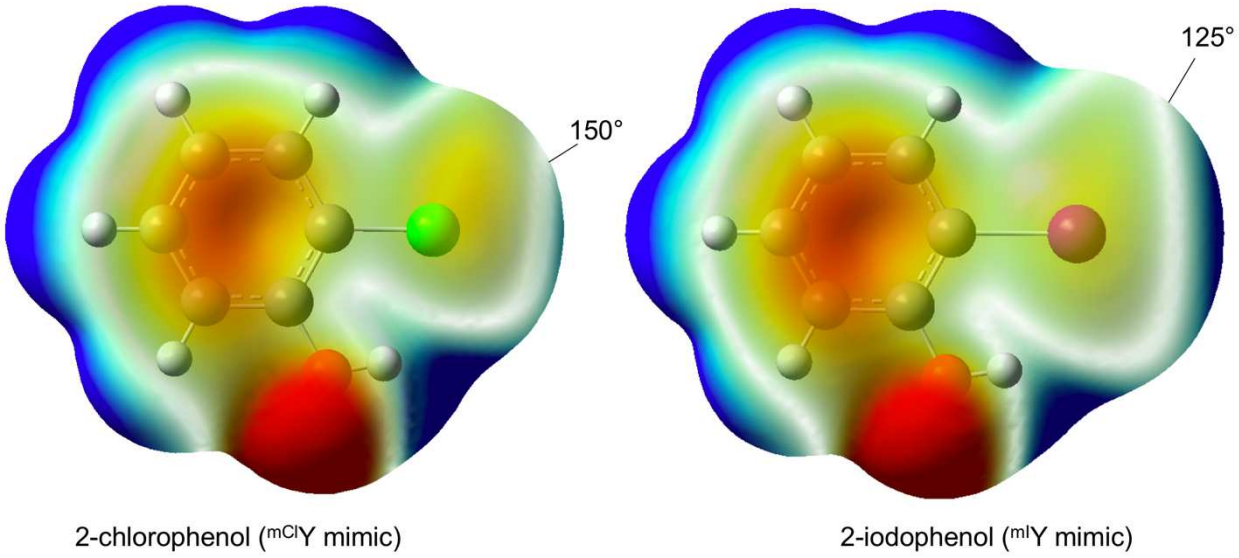



Figure 5.9
Electrostatic potential maps showing the theoretical neutral point of the σ-hole on 2-chlorophenol and 2-iodophenol. These angles represent the maximum angle of approach an acceptor molecule could take to interact with the positive σ-hole.

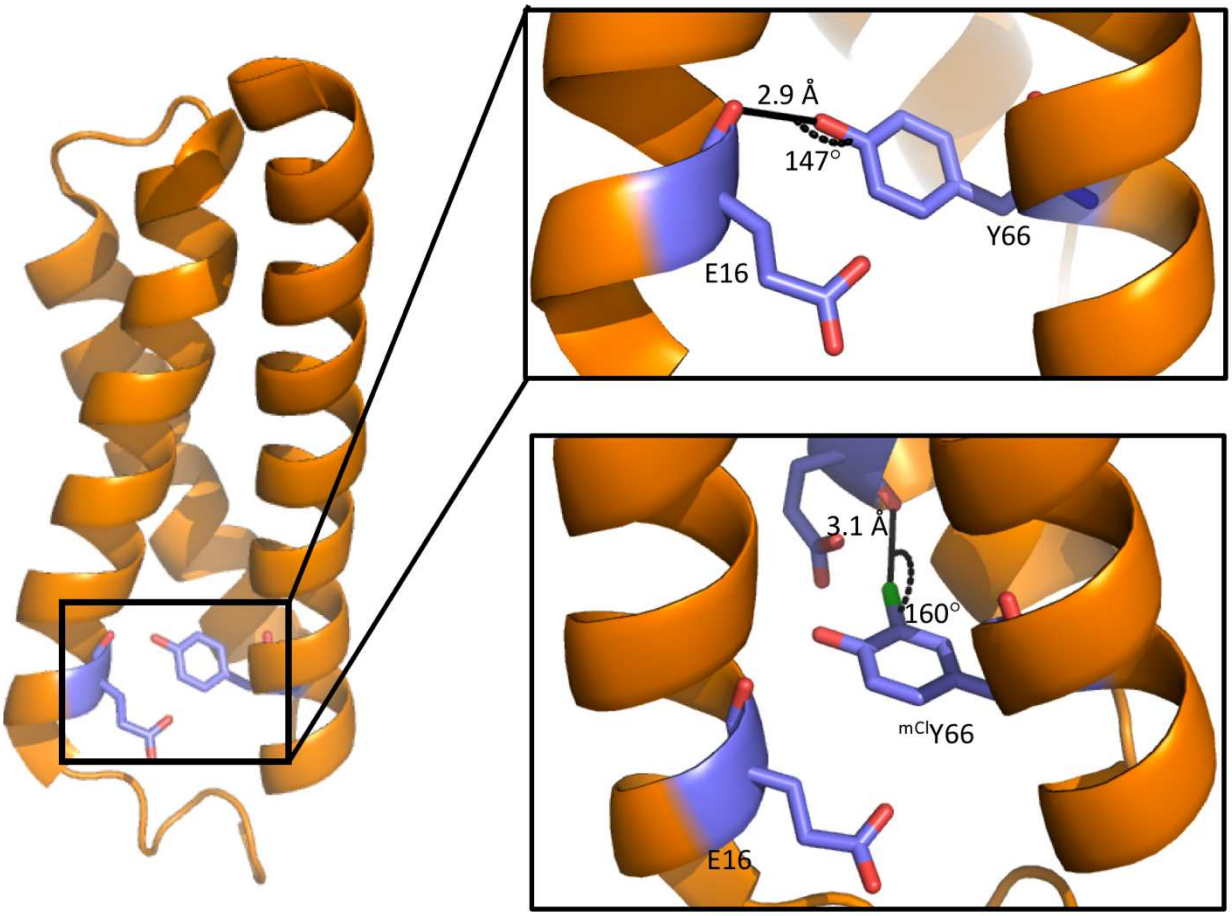



Figure 5.10
NMR structure of yeast KIX (PDB 2k0n) with conserved Y66 and a modeled in $^{mCl}$Y66. The chlorine could be making a very strong X-bond with a distance of 3.1 Å and a $\theta_1$ of 160°. The quantum mechanical calculations were carried out based on this modeled geometry.

in enthalpy. As this is slightly higher than we see in the DSC results (2 kcal/mol), perhaps the chlorine is positioned at a lower $\theta_1$ approach angle or is interacting with an acceptor atom at a longer distance. An alternative interpretation could be that the chlorine is at this modeled geometry (160° and 3.0 Å), but, the hydrogen is not pointing towards the chlorine, and a true HBeXB is not forming. In this scenario, the modest increase of 1.3 kcal/mol aligns well with a simple chlorine XB. Carrying out the same calculations on 2-iodophenol at a $\theta_1$ of 160° and a distance of 2.6 Å to the oxygen in NMA reveals an interaction energy of ~1.45 kcal/mol. Adding in the energy of the intramolecular H-bond would make the overall energy ~3.3 kcal/mol, which aligns well with our DSC results. These hypotheses seem to support the enthalpic terms obtained from the DSC measurements, however, getting crystal structures of the proteins will resolve any speculation about the placement of the halogen and the existence of the HBeXB.

The CD titration showed us that the $^{mCl}$Y66 KIX increases its structural helicity in the presence of a lower concentration of Pdr1p-12mer peptide compared to WT KIX. While we were not able to calculated a Kd from this titration, the shift in the spectra is preliminary evidence of an increased binding affinity[13]. Perhaps the stabilized KIX is locked in a conformation more conducive to binding this partner. Or, perhaps, the meta-halo-tyrosine helped fold α1 in the structure in a way that allows for a tighter interaction to take place. This assay at the very least shows $^{mCl}$Y66 KIX has not lost its ability to interact with Pdr1—12mer. The result will need to be confirmed and compared to a titration with $^{mI}$Y66 KIX to further evaluate how the different halogens are affecting potential binding activity. Applying this increase in affinity in an *in vivo* context, we start to think about how this could be altering the protein's function. More often than not in the cell, proteins have particular binding affinities to ligands to carry out tasks accordingly. Low binding affinities might indicate the need for high turnover on a particular reaction as the $k_{off}$

could be high. Contrary, high binding affinities are usually associated with very low $k_{off}$ rates. Since the KIX protein binds the Pdr1p protein as a means to help turn on transcription of drug efflux genes[5] (figure 5.1) altering this behavior could have vast effects on these genes' expression levels. The increased binding affinity could prevent KIX from releasing Pdr1p, possibly augmenting transcription of the drug efflux genes. Alternatively, the increased affinity could prevent KIX's release from Pdr1p potentially inhibiting or delaying the transcription of the genes, thereby affecting the genes' downstream influences. Perhaps the increased binding affinity of KIX to Pdr1p will prevent the export of "toxic drugs" allowing the anti-fungals or "toxic drugs" to be useful again.

The very preliminary results from this study are promising and exciting and have raised new questions. One key aspect of this study that is most exciting for the field is the impact a single atom can have in terms of stabilizing a protein with an intrinsically disordered region, and how the slight structural change can impact function. These initial findings suggest the incorporation of meta-halotyrosine into a protein with an intrinsically disordered region can aid its enthalpic stability, with a tunable aspect in terms of halogen choice. This should aid those interested in stabilizing aggregation-prone proteins that are associated with disease states [18,19], and perhaps be a valuable strategy in terms of designing more stable biologic therapeutics . It will be interesting to discover with the *in vivo* studies how the engineered KIX proteins binding partners vary, and how that plays out in terms of gene regulation, half-life, and proteasome degradation.

REFERENCES


1. Latysheva, N. S., Flock, T., Weatheritt, R. J., Chavali, S. & Babu, M. M. How do disordered regions achieve comparable functions to structured domains? *Protein Sci.* **24**, 909–922 (2015).
2. van der Lee, R. *et al.* Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep.* **8**, 1832–1844 (2014).
3. Madan Babu, M. *et al.* Intrinsically disordered proteins: regulation and disease This review comes from a themed issue on Sequences and topology Edited. *Curr. Opin. Struct. Biol.* **21**, 1–9 (2011).
4. Pajkos, M., Mészáros, B., Simon, I. & Dosztányi, Z. Is there a biological cost of protein disorder? Analysis of cancer-associated mutations. *Mol. Biosyst.* **8**, 296–307 (2012).
5. Thakur, J. K., Yadav, A. & Yadav, G. Molecular recognition by the KIX domain and its role in gene regulation. **42**, 2112–2125 (2014).
6. Goffeau, A. The fight against fungi. **452**, (2008).
7. Thakur, J. K. *et al.* ARTICLES A nuclear receptor-like pathway regulating multidrug resistance in fungi. **452**, (2008).
8. Palazzesi, F., Barducci, A., Tollinger, M. & Parrinello, M. The allosteric communication pathways in KIX domain of CBP. (2013). doi:10.1073/pnas.1313548110
9. Guzman, R. N. De, Goto, N. K., Dyson, H. J. & Wright, P. E. Structural Basis for Cooperative Transcription Factor Binding to the CBP Coactivator. 1005–1013 (2006). doi:10.1016/j.jmb.2005.09.059
10. Toto, A., Giri, R., Brunori, M. & Gianni, S. The mechanism of binding of the KIX domain

to the mixed lineage leukemia protein and its allosteric role in the recognition of c-Myb. **23**, 962–969 (2014).

11. Carlsson, A. C. C. *et al.* Increasing Enzyme Stability and Activity through Hydrogen Bond-Enhanced Halogen Bonds. *Biochemistry* **57**, 4135–4147 (2018).
12. Troilo, F. *et al.* The Folding Pathway of the KIX Domain. (2017). doi:10.1021/acschembio.7b00289
13. Thakur, J. K. *et al.* A nuclear receptor-like pathway regulating multidrug resistance in fungi. *Nature* **452**, 604–609 (2008).
14. Wals, K. & Ovaa, H. Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. *Front. Chem.* **2**, 1–12 (2014).
15. Micsonai, A. *et al.* BeStSel: A web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **46**, W315–W322 (2018).
16. Troilo, F. *et al.* The Folding Pathway of the KIX Domain. *ACS Chem. Biol.* **12**, 1683–1690 (2017).
17. Freire, E. Statistical thermodynamic analysis of differential scanning calorimetry data: Structural deconvolution of heat capacity function of proteins. *Methods Enzymol.* **240**, 502–530 (1994).
18. Koo, E. H., Lansbury, P. T. & Kelly, J. W. Amyloid diseases: abnormal protein aggregation in neurodegeneration. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9989–90 (1999).
19. Blokhuis, A. M., Groen, E. J. N., Koppers, M., Van Den Berg, L. H. & Pasterkamp, R. J. Protein aggregation in amyotrophic lateral sclerosis. *Acta Neuropathol.* **125**, 777–794 (2013).

# CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS

## ○ 6.1 Summary

The work herein has looked at the relationship between hydrogen and halogen bonds from a *inter* and *intra*molecular level. Through protein engineering efforts, we have shown how the bonds are similar in terms of specifying capabilities. We have shown how they can act synergistically and enhance one another in an intramolecular environment. And, we have shown how the bonds can differ from an enthalpy – entropy standpoint. The work will add to the growing fundamental knowledge of the structure-energy relationship X-bonds possess in biomolecular systems. This will be valuable to pharmaceutical companies looking to incorporate halogens into peptide or protein therapeutics. But also, the peptide (GCN4) and protein (yeast KIX) systems we selected to work with are biologically relevant. The detailed work on the GCN4 peptide could become a probe for transcription as the designed H- or X-bond-bearing peptide could act as a transcriptional regulator when binding endogenous GCN4. And, the HBeXB stabilization efforts of yeast KIX will aid those studying how to stabilize aggregation-prone proteins.

The next step for both of these projects is to study them in an *in vivo* environment. This would allow for the designed peptide from GCN4 to act as potential transcription sensing probe, and will allow us to investigate the change in cellular regulation pathways derived from stabilized KIX.

## ○ 6.2 A Complex Interplay Exists Between Hydrogen and Halogen Bonds

In Chapter 2 we delved into the relationship between hydrogen and halogen bonds seen from a biomolecular perspective. We broke their relationship into 3 main categories 1) competition between H and X-bonds in biological systems, 2) substitution of an X-bond for an H-bond and

how that affects the system, and 3) the orthogonal or synergistic relationship seen between the bonds. It's no trivial task predicting how the bonds may complement or compete against one another, but chemists and biochemists alike should consider these relations in molecular design. As more X-bond engineering work is carried out, more defined directives can be established for molecular design strategies. We kept the concepts learned from this review in mind while carrying out our own protein engineering endeavors discussed in Chapter 4 and 5.

- **6.3 Hydrogen Bond Enhanced Halogen Bonds Exist in Not Only Biomolecular Systems**

Chapter 3 describes the opportune discovery of the hydrogen bond enhanced halogen bond in two unique systems. This study shows the wide application of the HBeXB as it was discovered in an organic small-molecule system as well as a protein system. The HBeXB theory suggests the hydroxyl is acting as an electron-withdrawing group as it enhances the size of the σ-hole, so quantum mechanical calculations were carried out to better understand this. As OH is typically thought of as an electron-donating group, the charges on the carbon atoms in phenol reflect this for the *ortho-* and *para-* carbons (negatively charged), but, the *meta-* position carbon is positively charged reflecting the electron-withdrawing ability of the hydroxyl. A survey of the CSD and PDB was carried out looking for the existence of HBeXB's in small molecule and biological systems. As many hits appeared, it seems the Berryman and Ho labs are just the first to experimentally recognize the interaction and its potential impact on the halogen bonding field.

- **6.4 Halogen Bonds Can Behave as a Specifying Protein-Protein Interaction**

Chapter 4 discusses how engineering an H-bond or X-bond donor into the GCN4 N16A coiled coil sequence aids in the assembly of a specific heterotrimer complex. We showed the engineered electrostatic interactions are capable of assembling the specific heterotrimer complex across a range of concentrations (signifying their high affinity), unlike the non-specific

hydrophobic molecules also tested. This study points at the potential the H- and X-bond peptide possesses as a transcriptional inhibitor of GCN4, while also revealing how powerful the designed electrostatic interaction could be for other peptide assemblies. This study opens a up a new "knobs-in-holes" realm for coiled-coils as the concepts could be applied to finely tune oligomerization rate and behavior.

- **6.5 Hydrogen Bond Enhanced Halogen Bonds Can Increase Stability in KIX**

In Chapter 5 we incorporated $^{mX}$Y into yeast KIX to better understand how the potential HBeXB could impact this protein's structure and function *in vitro.* We found the $^{mX}$Y did not significantly affect KIX's secondary structure from CD, but it did signifigantly increase its thermal stability. We do not know for sure whether a true HBeXB bond is forming in the structure. We are working on crystallizing KIX, which will give us more information about the placement of the halogen. The CD titration revealed a slight increase in helicity at lower Pdr1p-12mer concentrations for the $^{mCl}$Y66 KIX as compared to WT, perhaps indicting the $^{mCl}$Y66 binds this peptide at a higher affinity. As KIX's intrinsically disordered region partially defines its function, investigating how this stabilized version of KIX will impact its function and influence overall cellular regulation from an *in vivo* standpoint will be interesting to look at next.

- **6.6 Future Directions**

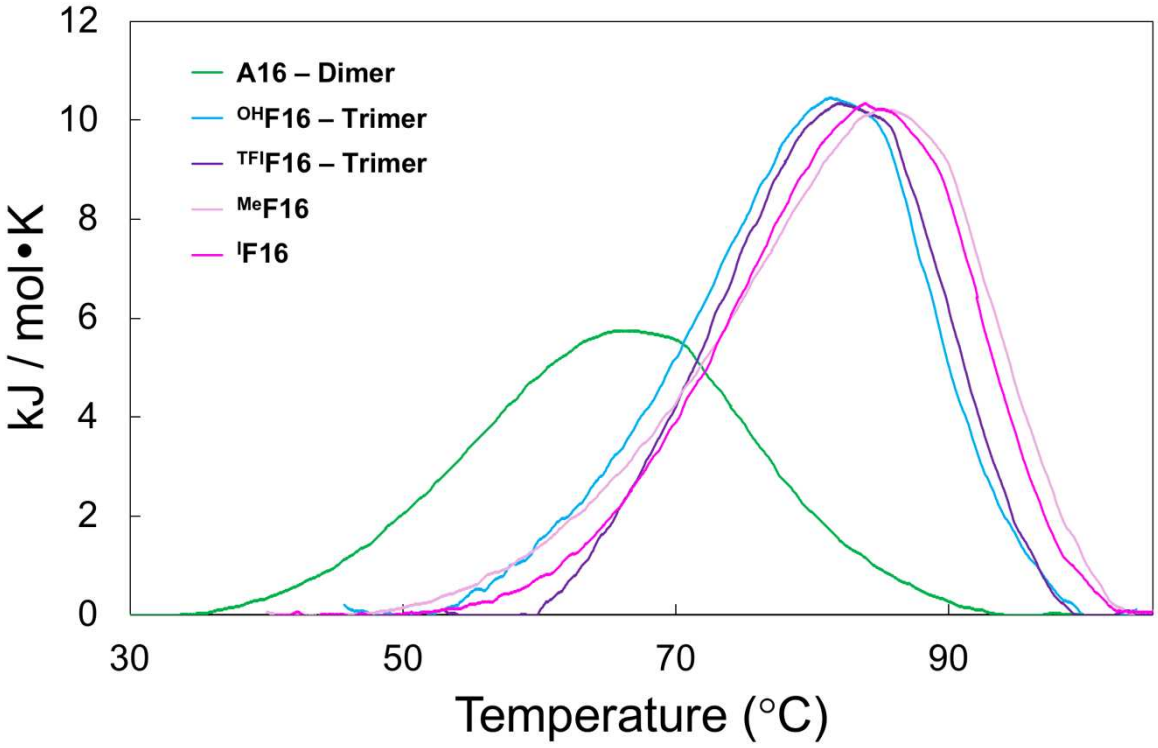
For GCN4's use as a sensing probe, or understanding how stabilized KIX modifies cellular function, the next steps for the work depicted in Chapters 4 and 5 would be to evaluate the proteins *in vivo.* The *in vivo* characterization will allow for the application of the designed GCN4 peptide as a sensing probe. We could tag endogenous GCN4 and load the designed GCN4 peptide into yeast cells looking for colocalization among the proteins. We could also set-up a reporter gene (lacZ and X-Gal) assay in yeast to monitor how the designed GCN4 peptide could act as a

transcriptional switch. We hypothesize in the absence of the designed peptide, the yeasts cells would express lacZ and turn blue. Upon loading in the peptide, endogenous GCN4 may be inhibited by the peptide and not be able to properly promote expression of lacZ, leaving the cells opaque. These *in vivo* studies could show how useful the designed H-bond or X-bond-specifying peptide could be as a transcriptional probe.

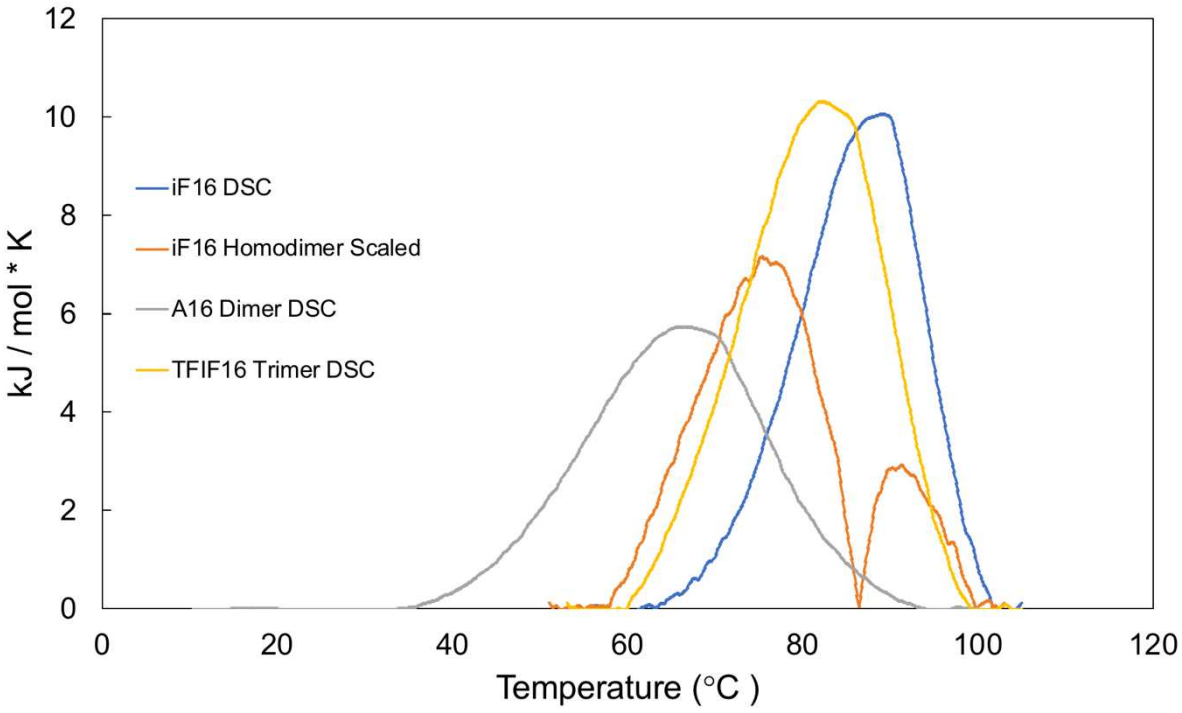
Expressing the KIX mutants in yeast cells (which could involve eukaryotic unnatural amino acid incorporation) and carrying out RNA-seq studies would allow us to tease out which genes get up/down-regulated as a result of the presence of $^{mX}$Y KIX. To gauge how KIX is modifying cellular pathways, we must take a look at the genomic level, and this type of experiment will help us determine which cellular pathways are potentially distressed or unaffected. As yeast KIX is part of the pleiotropic drug response pathway, we could look at those genes specifically to glean insights on how KIX is binding with partners in the pathway. A more detailed understanding of these interactions could be useful, as KIX is a clinical target.

Overall, this dissertation has looked at how useful X-bonds can be for biological engineering efforts. Compared to H-bonds, they have unique properties that will be advantageous for exploitation in biologically derived therapeutics, increasing enzyme activity, designing high affinity protein sensing molecules, and stabilizing aggregation-prone proteins.

APPENDIX I




Supplemental Figure 1.
DSC melting profiles from all of the constructs. The $^{Me}$F16 and $^{I}$F16 profiles are wider than the trimer melting profile and shifted over another 4 °C (FWHM is ~25 °C). When the data were fit to one simple two-state model (as all of the other constructs were), the residual around the fit was high and thermodynamic values could not be determined.




Supplemental Figure 2.
Homodimer presence in $^{I}$F16 melting profile. The buffer and baseline subtracted A16 and $^{TFI}$F16 data were subtracted from the $^{I}$F16 to indicate the presence of an $^{I}$F16 homodimer.