DISSERTATION

EXPLORING THE HYDROGEN BOND ENHANCED HALOGEN BOND AS A TOOL FOR BIOMOLECULAR DESIGN AND ENGINEERING

Submitted by

Margaret Grace Walker

Department of Biochemistry and Molecular Biology

In partial fulfilment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2026

Doctoral Committee:

Advisor: P. Shing Ho

Chris Snow
Marcelo Melo
Debbie Crans

ABSTRACT

EXPLORING THE HYDROGEN BOND ENHANCED HALOGEN BOND AS A TOOL FOR BIOMOLECULAR DESIGN AND ENGINEERING

The field of therapeutic drug discovery continues to embrace computational design, development and optimization methods, and has been further invigorated by the development of artificial intelligence (AI) and machine learning (ML) techniques for biomolecular modeling, prediction, and docking purposes. Halogens have long been accepted as meaningful and beneficial additions to biological drug design. Due to their unique properties, halogens not only contribute lipophilicity and stability, but also binding affinity and specificity by forming noncovalent interactions, called halogen bonds (XBs), with electron rich atoms. In addition to their roles as small molecule ligand constituents, XBs have emerged as interesting and unique design tools for protein engineering purposes for augmenting function and stability. This thesis aims to demonstrate the value of the hydrogen bond enhanced halogen bond (HBeXB), a cooperative noncovalent interaction network, as a tool for biomolecular design. To accomplish this goal, we approach protein design applications as a whole by exploring nontraditional noncovalent interactions currently in protein engineering, the HBeXB in catalytic enzyme design, and the HBeXB impact in protein-ligand binding. Investigating the influence of HBeXBs on enzymatic catalysis and protein-ligand binding provides the foundation for accurately and effectively incorporating these cooperative noncovalent interactions in protein and drug design.

AWKNOWLEDGEMENTS

I could not have undertaken this journey without my advisor, Professor P. Shing Ho. Thank you for always encouraging me to pursue new learning and training opportunities, and for your advocacy, patience, and mentorship throughout my doctoral studies. I am also deeply grateful to my student advisory committee—Jeff Hansen, Chris Snow, Debbie Crans, and Marcelo Melo—for their guidance, mentorship and support throughout my PhD. I am very thankful to the Ho Lab, especially Ryan and Alex for their training and encouragement, and Gustavo for being my teammate throughout this process.

To my parents, Steve and Carrie, thank you for the opportunities and blessings you have worked so hard to provide. You have both embodied virtue and compassion, integrity and diligence, and an unwavering faith. I will always be grateful for the examples you have set and the life you have given me. Your love, support, and encouragement have meant so much to me throughout this journey. Chris, Hannah, Joe, Sam, and JoAnna, you are my long distance, built-in best friends. Each of you brings something unique and invaluable to my life. The last five years have been quite a journey for each of us, and I know that I could not have gotten through it without each of you. To my sisters, our relationship means everything to me. You hold me accountable to my goals, lift me up when I am down, and always have my back. You are two of the strongest and most compassionate people I know, and I aspire to be more like you every day.

Finally, thank you, Alex, for being my best friend and for listening to me talk about this project and this experience incessantly over the past four years. You encouraged me, believed in me, and were so patient with me every time I panicked about things that now seem so small. I will always be grateful for the support and love you have shown me.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Protein design has been rapidly revolutionized with the introduction of artificial intelligence (AI) and machine learning (ML) methods to predict their structures and molecular interactions.[1, 2, 3, 4] The field has bloomed with new and evolving techniques boost knowledge of the structural and dynamic features of macromolecules. It is truly an exciting time to be a structural biologist.

Even so, it remains necessary and beneficial to build upon our understanding and explore creatively the protein engineering tools at our disposal. This includes nontraditional molecular interactions that are infrequently observed in nature and, thus, less likely to appear in AI/ML training databases. This thesis establishes foundational knowledge of the hydrogen bond enhanced halogen bond (HBeXB), a recently discovered nonclassical noncovalent (ncNC) interaction, to inform structure-based design strategies and how we can exploit and tune these nontraditional interactions for creative and unconventional biological applications.

1. The Halogen Bond

The halogen bond (XB), a ncNC interaction, is classified as primarily an electrostatic interaction, with some controversy as to contributions from forces such as charge transfer and dispersion.[5,6] The XB is most commonly described by the σ-hole model as a quantum anisotropic effect when a halogen atom covalently binds another atom (C, P, S).[7] Due to the electronic configuration of halogens, the single unpaired valence electron is used to form a σ molecular orbital, resulting in a depleted region of electron density (commonly referred to as the σ-hole)

directly opposed to the newly formed molecular orbital. The majority of the electron density remains perpendicular to the σ-hole. This anisotropic redistribution of electrons results in an amphoteric atom, with both an electronegative region and an electropositive region; the latter with a capacity for interacting with electron rich atoms such as oxygen, nitrogen and sulfur (Figure 1). Additionally, this redistribution is highly influenced by polarizability and so the strength of the interaction also follows with I > Br > Cl >>F (Fluorine being so small and electronegative that it rarely forms a σ-hole).

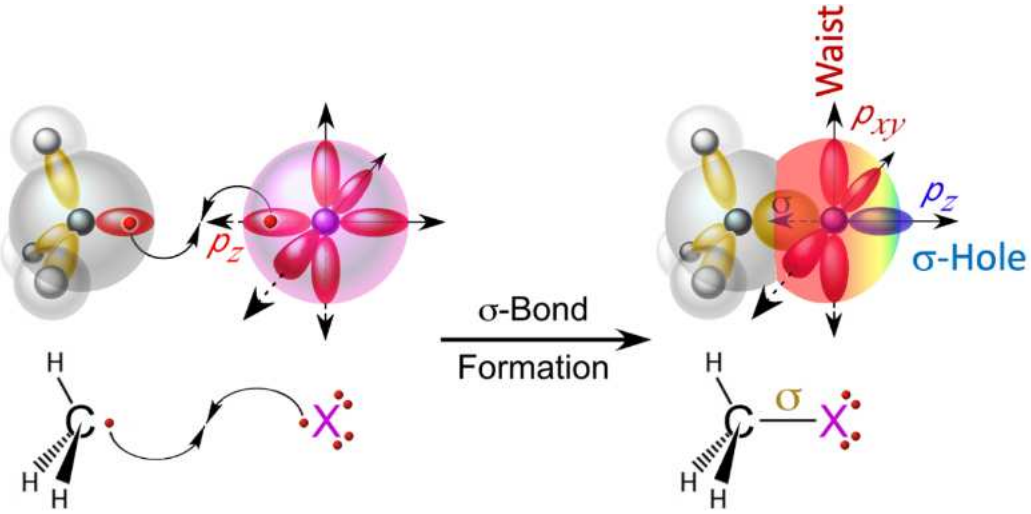


Figure 1.1: Visual of the σ-hole model of XB. [8,9]

The first reported evidence of halogens interacting with electronegative atoms occurred in 1814, when J.J. Colin first noticed that combining iodine and amylose produced a blue-black color, forming $I_2 \cdots NH_3$. [10, 11] However, the exact molecular makeup of the liquid was not confirmed for another 50 years by F. Guthrie.[12]

Throughout the next two centuries, pieces and fragments of evidence leading the scientific community to understanding the XB emerged irregularly. This includes observing and

identifying XB roles in self-assembling products, UV-VIS spectra shifts and the shift in color upon reactions with iodine (such as what J.J. Colin first observed).[13]

In the 1950s, Odd Hassel's research involving x-ray crystallography revealed the structural features of what he called "halogen molecule bridges" between dioxane molecules.[14] These structures revealed Br···O interactions with distances less than the sums of their van der Waals radii (Σ_{VDWR}) and Br···Br···O angles approaching 180°.[15] In his Nobel lecture in 1970, Hassel recognized these short "halogen molecule bridges" as charge transfer bonds.[16]

The computational work P. Politzer and J. S. Murrey in the 1990s described covalently bound halogens having an anisotropic charge distribution, which would create the foundation for later defining the σ-hole.[17, 18, 19]

In 2013, the XB was officially defined as "A halogen bond occurs when there is evidence of a net attractive interaction between an electrophilic region associated with a halogen atom in a molecular entity and a nucleophilic region in another, or the same, molecular entity" by the International Union of Pure and Applied Chemistry's (IUPAC) recommendation, a culmination of 200 years of research into short halogen interactions.[20]

Over the last two centuries, the XB has made significant impacts on material sciences, crystal engineering, and small molecule catalysis.[13,21] It is at times characterized alongside other ncNC interactions (tetrel, pnictogen and chalcogen bonds) that are also explained as σ-hole interactions.[22, 23, 24, 25] However, understanding the role of XB interactions in biomolecules is a rapidly growing field of scientific inquiry.

1.2 Biological Halogen bonds

XBs emerged unexpectedly in biological contexts in 2003 when usage of bromouracil (BrU) to help solve the crystal structure of a DNA sequence resulted in the formation and stabilization of 4-stranded DNA Holliday junction instead of a duplex B-DNA.[26] The resulting junction showed the Br ~3 Å from a phosphate oxygen (~ 89% Σ_{VDWR}), in place of a HB that had been previously demonstrated to stabilize the junction formation of other nonhalogenated sequences.

To further explore this unique and unexpected relationship in biomolecules, Auffinger *et al.,* conducted a structural bioinformatics survey of the Protein Data Bank (PDB) in search of short halogen interactions.[27] Subsequent analysis of the patterns of interactions from this survey suggested that XBs and HBs are orthogonal when sharing a common acceptor atom, giving XBs unique geometries when interacting with peptide carbonyl oxygens.

Biological XBs can be found in small molecule ligands and inhibitors in complex with their protein targets,[28,29] and as contributors to the recognition and specificity of thyroid hormone receptors.[30] The first intentional use of XBs in structure-based drug design (SBDD) was reported by Patrick Lam at the 238th ACS National Meeting in 2009 on factor Xa inhibitors, forever altering how halogens are perceived in rational drug design to include added specificity and affinity.[31]

ncNC interactions, such as XBs, also present chemical and biological engineers alike with the opportunity for building networks of noncovalent interactions. Theoretical and experimental evidence of cooperative XBs has been demonstrated for purposes such as self-assembly of molecular structures and membrane-ligand binding.[32, 33, 34] Recently, Baykov *et al.*, demonstrated

a 3-part cooperative ncNC network to enhance XB donation via π-π stacking, π-chalcogen bonding, and σ-hole chalcogen bonding interactions.[35]

Another example includes the hydrogen bond enhanced halogen bond (HBeXB). Due to its electronegative waist perpendicular to the σ-hole, the XB donor can also act as a hydrogen bond (HB) acceptor.[36] Accepting an HB further polarizes the X atom, drawing on its electron density, and results in enhancing the size of the σ-hole. This enhancement, the HBeXB, adds another layer of tunability to XBs as ncNC interactions by widening the angle of approach for XB acceptors as well as potentially influencing XB interaction strength.[37, 38, 9]

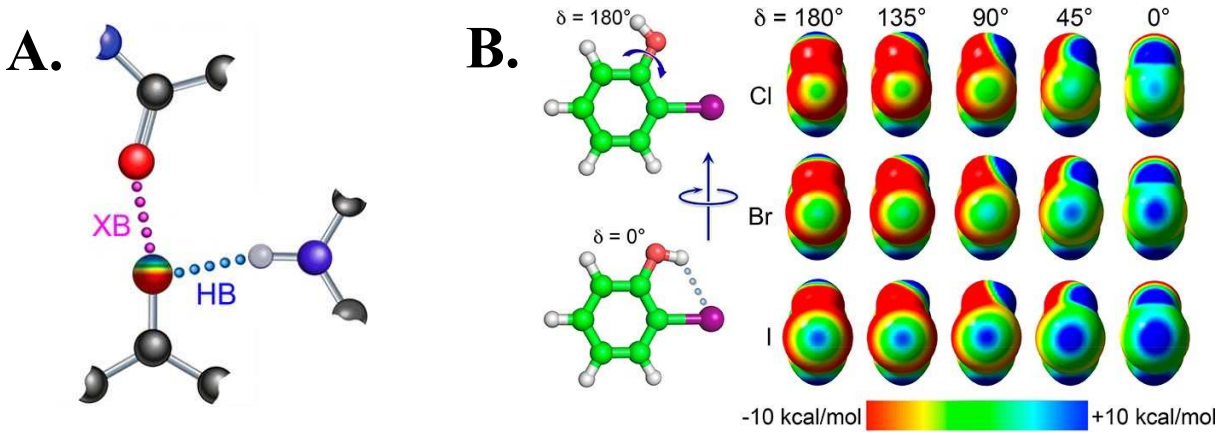


Figure 1.2: The hydrogen bond enhanced halogen bond (HBeXB). A. Visual description of the HBeXB with the X atom donating a XB and accepting a HB.[39] B. Impact of OH rotations on the surface electrostatic potential (ESP) of 2-halophenol. The scale shows -10 kcal/mol to +10 kcal/mol of ESP to a positive point charge.[38]

1.3 Research Goals

To demonstrate how nontraditional noncovalent interactions can be utilized in protein engineering, I will explore the HBeXB in previous protein engineering applications, as an enzymatic catalyst, and in protein-ligand binding interactions, aiming to answer the following questions:

How can ncNC interactions be utilized to enhance, support, or substitute for enzymatic catalytic function? How can incorporation of these interactions contribute favorably to lowering the activation energy needed to catalyze enzymatic reactions? How can ncNC interactions optimize intermolecular binding? How does the complex environment of proteins impact ncNC capabilities?

Chapter 2 of this thesis discusses how ncNC interactions have generally been utilized for protein engineering applications, highlighting XBs and HBeXBs, when this study first began. This chapter will further explain the previous studies involving the HBeXB in protein design, demonstrating the tunability, variety, and stability that it can offer to structural biologists. However, the direct application to biological protein function remains an area for continued investigation.

In this dissertation, I will address these questions by investigating how the HBeXB can substitute for enzymatic catalytic function and how protein structural environments impact the HBeXB in return. I will also explore the tunability of the HBeXB in protein systems by examining the impact of various amino acid residue side chains on the XB energies, aiming to parse out the interaction energy of the XB itself while enhanced by a HB.

In chapter 3, we will demonstrate an engineered halogen bond catalyzed DNA-cleaving endonuclease and explore using the HBeXB to widen functional opportunities in protein engineering. The purpose of this study is to further grasp the potential applications of using XBs in design of novel enzymes but also to question whether enzymatic active sites can be harnessed to complete XB dependent catalysis. Additionally, this study probes the question of whether the XB is capable of replacing the function of metal cofactors in enzymatic active sites and therefore, challenges how we understand metalloenzymes. The reported DNA-cleaving

endonuclease activity was attributed to a bridging HBeXB interaction functionally substituting for the wildtype's magnesium orchestrated DNA hydrolysis reaction. I will provide evidence for the HBeXB mechanism through a series of mutagenesis studies, pH titrations and σ-hole electrostatic potential calculations that validate the key characteristics of the novel XB dependent enzyme.

In chapter 4, I will examine existing biological XBs and HBeXBs, and the impact of their geometric and energetic properties as they apply to protein-ligand binding. A comprehensive survey of XBs and HBeXBs in the Protein DataBank (PDB) reveal geometric trends in protein-ligand binding interactions. Utilizing these findings to build computational models for energy computation resulted in an examination of HB donor impacts on electrostatic potential of the σ-hole and the impact of positively charged HB donors on overall binding energy of the cooperative noncovalent interaction. These findings demonstrate the structural context dependence and complexity of the HBeXB, and how the protein pocket directs the strength and geometry of such interactions, exemplifying how ncNC interactions can contribute to protein-substrate binding.

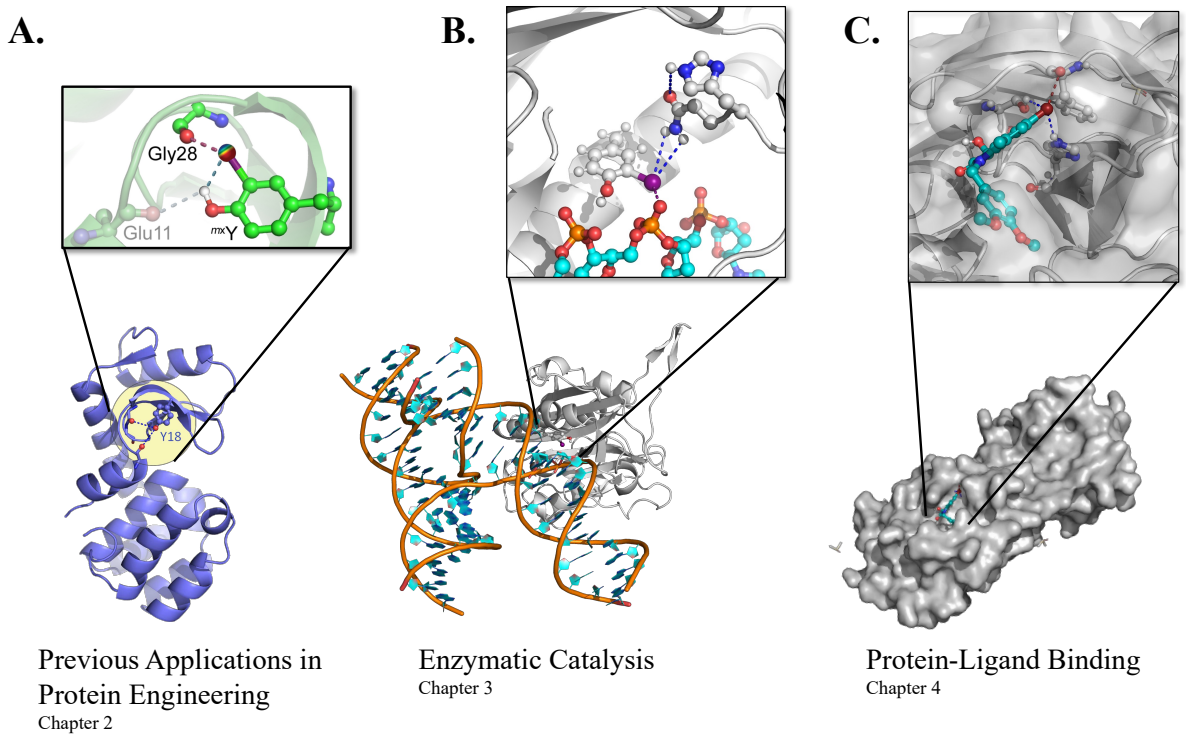


Figure 1.3: Demonstrating the HBeXB as a valuable tool for structure-based design. A. Structure of halogenated T4 Lysozyme (PDB: 5V7D).[38] B. Halogenated Mouse endonuclease G (136^{mI}Y) bound to 4-stranded Holliday junction (structure prediction with Chai1 Discovery.[40] C. Structure *of SARS*-CoV-2 main protease in complex with N1U (PDB: 7GDX).[41]

Overall, this study of HBeXB interactions in protein systems will further inform the opportunities, limitations and complexities of implementing this network of noncovalent interactions in design of proteins and protein-ligand interactions. The significance of these findings lies in the biomolecular design applications, both experimentally and computationally. To fully model and apply nonadditive ncNC networks in protein engineering, we must better understand their capabilities under influence of protein environments. Through exploring the HBeXB in previous protein design applications, as a catalyst for enzymatic reactions, and as a contributor to protein-ligand binding, I will demonstrate the HBeXB as a valuable tool for structure-based design.

REFERENCES

(1) S. Borkotoky, A. Joshi, V. Kaushik, and A. Nath Jha, "Machine Learning and Artificial Intelligence in Therapeutics and Drug Development Life Cycle," in *Drug Development Life Cycle*, ed. Juber Akhtar et al. (IntechOpen, 2022), https://doi.org/10.5772/intechopen.104753.

(2) J. Villalobos-Alva, L. Ochoa-Toledo, M.J. Villalobos-Alva, A. Aliseda, F. Pérez-Escamirosa, N.F. Altamirano-Bustamante, F. Ochoa-Fernández, R. Zamora-Solís, S. Villalobos-Alva, C. Revilla-Monsalve, N. Kemper-Valverde, M.M. Altamirano-Bustamante, "Protein Science Meets Artificial Intelligence: A Systematic Review and a Biochemical Meta-Analysis of an Inter-Field," *Frontiers in Bioengineering and Biotechnology* 10 (July 2022): 788300, https://doi.org/10.3389/fbioe.2022.788300.

(3) M. Suleiman Jada, H. Stephen, M.M. Usman, A.U. Wurochekke, "Application of Artificial Intelligence in Biochemistry Research: A Review," *Journal of Biochemistry, Microbiology and Biotechnology* 11, no. 2 (2023): 1–10, https://doi.org/10.54987/jobimb.v11i2.835.

(4) Mohammed Abdul Lateef Junaid, "Artificial Intelligence Driven Innovations in Biochemistry: A Review of Emerging Research Frontiers," *Biomolecules and Biomedicine* 25, no. 4 (2025): 739–50, https://doi.org/10.17305/bb.2024.11537.

(5) Anthony C. Legon, "The Halogen Bond: An Interim Perspective," *Physical Chemistry Chemical Physics* 12, no. 28 (2010): 7736, https://doi.org/10.1039/c002129f.

(6) Marc Fourmigué, "Halogen Bonding: Recent Advances," *Current Opinion in Solid State and Materials Science* 13, nos. 3–4 (2009): 36–45, https://doi.org/10.1016/j.cossms.2009.05.001.

(7) T. Clark, M. Hennemann, J.S. Murray, and P. Politzer, "Halogen Bonding: The σ-Hole: Proceedings of 'Modeling Interactions in Biomolecules II', Prague, September 5th–9th, 2005," *Journal of Molecular Modeling* 13, no. 2 (2007): 291–96, https://doi.org/10.1007/s00894-006-0130-2.

(8) M.R. Scholfield, M. Coates Ford, A.C. Carlsson, H. Butta, R.A. Mehl, and P.S. Ho, "Structure–Energy Relationships of Halogen Bonds in Proteins," *Biochemistry* 56, no. 22 (2017): 2794–802, https://doi.org/10.1021/acs.biochem.7b00022.

(9) A.M.S. Riel, R.K. Rowe, E.N. Ho, A. C. Carlsson, A.K. Rappé, O.B. Berryman, and P.S. Ho, "Hydrogen Bond Enhanced Halogen Bonds: A Synergistic Interaction in Chemistry and Biochemistry," *Accounts of Chemical Research* 52, no. 10 (2019): 2870–80, https://doi.org/10.1021/acs.accounts.9b00189.

(10) Colin, M. M.; Gaultier de Claubry, H. Sur Le Combinaisons de L'iode Avec Les Substances Végétales et Animales. Ann. Chim. 1814, 90, 87−100.

(11) Colin, M. Note Sur Quelques Combinaisons de L'iode. Ann. Chim. 1814, 91, 252−272.

(12) Guthrie, F. On the Iodide of Iodammonium. J. Chem. Soc. 1863, 16, 239−244.

(13) G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati, and G. Terraneo, "The Halogen Bond," *Chemical Reviews* 116, no. 4 (2016): 2478–601, https://doi.org/10.1021/acs.chemrev.5b00484.

(14) O. Hassel, J. Hvoslef, The Structure of Bromine 1,4-Dioxanate. Acta Chem. Scand. 1954, 8, 873.

(15) O. Hassel, C. Romming, Direct Structural Evidence for Weak Charge-Transfer Bonds in Solids Containing Chemically Saturated Molecules. Q. Rev., Chem. Soc. 1962, 16, 1−18.

(16) O. Hassel, "Structural Aspects of Interatomic Charge-Transfer Bonding," *Science* 170, no. 3957 (1970): 497–502, https://doi.org/10.1126/science.170.3957.497.

(17) T. Brinck, J.S. Murray, P. Politzer, Surface Electrostatic Potentials of Halogenated Methanes as Indicators of Directional Intermolecular Interactions. Int. J. Quantum Chem. 1992, 44, 57−64.

(18) T. Brinck, J.S. Murray, and P. Politzer, Molecular Surface Electrostatic Potentials and Local Ionization Energies of Group V−VII Hydrides and Their Anions: Relationships for Aqueous and Gas-Phase Acidities. Int. J. Quantum Chem. 1993, 48, 73−88.

(19) J.S. Murray, K. Paulsen, and P. Politzer, Molecular Surface Electrostatic Potentials in the Analysis of Non-Hydrogen-Bonding Noncovalent Interactions. Proc. Indian Acad. Sci., Chem. Sci. 1994, 106, 267−275.

(20) G. R. Desiraju, P. S. Ho, L. Kloo, A. C. Legon, R. Marquardt, P. Metrangolo, P. Politzer, G. Resnati, and K. Rissanen, (2013) Definition of the halogen bond (IUPAC Recommendations 2013). Pure Appl. Chem. 85, 1711−1713.

(21) Revannath L. Sutar and Stefan M. Huber, "Catalysis of Organic Reactions through Halogen Bonding," *ACS Catalysis* 9, no. 10 (2019): 9622–39, https://doi.org/10.1021/acscatal.9b02894.

(22) I. Alkorta, J. Elguero, and A. Frontera, "Not Only Hydrogen Bonds: Other Noncovalent Interactions," *Crystals* 10, no. 3 (2020): 180, https://doi.org/10.3390/cryst10030180.

(23) F. De Vleeschouwer, F. De Proft, **Ö**. Ergü**n**, W. Herrebout, and P. Geerlings, "A Combined Experimental/Quantum-Chemical Study of Tetrel, Pnictogen, and Chalcogen Bonds of Linear Triatomic Molecules," *Molecules* 26, no. 22 (2021): 6767, https://doi.org/10.3390/molecules26226767.

(24) L. Brammer, A. Peuronen, and T.M. Roseveare "Halogen Bonds, Chalcogen Bonds, Pnictogen Bonds, Tetrel Bonds and Other σ-Hole Interactions: A Snapshot of Current Progress," *Acta Crystallographica Section C Structural Chemistry* 79, no. 6 (2023): 204–16, https://doi.org/10.1107/S2053229623004072.

(25) Robin Taylor, "Aerogen Bond, Halogen Bond, Chalcogen Bond, Pnictogen Bond, Tetrel Bond, Triel Bond ... Why So Many Names?," *Crystal Growth & Design* 24, no. 10 (2024): 4003–12, https://doi.org/10.1021/acs.cgd.4c00303.

(26) F.A. Hays, J.M. Vargason, and P.S. Ho, "Effect of Sequence on the Conformation of DNA Holliday Junctions," *Biochemistry* 42, no. 32 (2003): 9586–97, https://doi.org/10.1021/bi0346603.

(27) P. Auffinger, F.A. Hays, E. Westhof, and P.S. Ho, "Halogen Bonds in Biological Molecules," *Proceedings of the National Academy of Sciences* 101, no. 48 (2004): 16789–94, https://doi.org/10.1073/pnas.0407607101.

(28) N. Muzet, B. Guillot, C. Jelsch, E. Howard, and C. Lecomte, "Electrostatic Complementarity in an Aldose Reductase Complex from Ultra-High-Resolution Crystallography and First-Principles Calculations," *Proceedings of the National Academy of Sciences* 100, no. 15 (2003): 8742–47, https://doi.org/10.1073/pnas.1432955100.

(29) A. Voth and P.S. Ho, "The Role of Halogen Bonding in Inhibitor Recognition and Binding by Protein Kinases," *Current Topics in Medicinal Chemistry* 7, no. 14 (2007): 1336–48, https://doi.org/10.2174/156802607781696846.

(30) N.F. Valadares, L.B. Salum, I. Polikarpov, A. D. Andricopulo, and R.C. Garratt, "Role of Halogen Bonds in Thyroid Hormone Receptor Selectivity: Pharmacophore-Based 3D-QSSR Studies," *Journal of Chemical Information and Modeling* 49, no. 11 (2009): 2606–16, https://doi.org/10.1021/ci900316e.

(31) Lam, P. Y. S.; Clark, C. G.; Smallwood, A. M.; Alexander, R. S. Structure-Based Drug Design Utilizing Halogen Bonding: Factor Xa Inhibitors. In The 238th ACS National Meeting; Metrangolo, P.; Resnati, G., Eds.; ACS: Washington, DC, **2009**; Vol. ORGN, p 58.

(32) X.C. Yan, P. Schyman, and W.L. Jorgensen, "Cooperative Effects and Optimal Halogen Bonding Motifs for Self-Assembling Systems," *The Journal of Physical Chemistry A* 118, no. 15 (2014): 2820–26, https://doi.org/10.1021/jp501553j.

(33) M. Kaasik, S. Kaabel, K. Kriis, I. J **ä**rving, R. Aav, K. Rissanen, and T. Kanger, "Synthesis and Characterisation of Chiral Triazole-Based Halogen-Bond Donors: Halogen Bonds in the Solid State and in Solution," *Chemistry – A European Journal* 23, no. 30 (2017): 7337–44, https://doi.org/10.1002/chem.201700618.

(34) R. Santana Nunes, D. Vila-Viçosa, and P.J. Costa, "Halogen Bonding: An Underestimated Player in Membrane–Ligand Interactions," *Journal of the American Chemical Society* 143, no. 11 (2021): 4253–67, https://doi.org/10.1021/jacs.0c12470.

(35) S.V. Baykov, I.I. Fedorova, D.M. Ivanov, A.V. Semenov, L.E. Zelenkov, S.O. Baykova, and V.Y. Kukushkin, "Dual σ- and π-Hole Chalcogen Bonding with Cooperative π-Stacking: Noncovalent Forces That Amplify Halogen Bond Donors," *Crystal Growth & Design*, February 25, 2026, acs.cgd.5c01745, https://doi.org/10.1021/acs.cgd.5c01745.

(36) Y.V. Nelyubina, M.Y. Antipin, D.S. Dunin, V.Y. Kotov, and K.A. Lyssenko, "Unexpected 'Amphoteric' Character of the Halogen Bond: The Charge Density Study of the Co-Crystal of N-Methylpyrazine Iodide with I2," *Chemical Communications* 46, no. 29 (2010): 5325, https://doi.org/10.1039/c0cc01094d.

(37) A.M.S. Riel, D.A. Decato, J. Sun, C.J. Massena, M.J. Jessop, and O.B. Berryman, "The Intramolecular Hydrogen Bonded–Halogen Bond: A New Strategy for Preorganization and Enhanced Binding," *Chemical Science* 9, no. 26 (2018): 5828–36, https://doi.org/10.1039/C8SC01973H.

(38) A.C. Carlsson, M.R. Scholfield, R.K. Rowe, M Coates Ford, A.T. Alexander, R.A. Mehl, and P.S. Ho, "Increasing Enzyme Stability and Activity through Hydrogen Bond-Enhanced Halogen Bonds," *Biochemistry* 57, no. 28 (2018): 4135–47, https://doi.org/10.1021/acs.biochem.8b00603.

(39) Rhianon K. Rowe and P. Shing Ho, "Relationships between Hydrogen Bonds and Halogen Bonds in Biological Systems," *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials* 73, no. 2 (2017): 255–64, https://doi.org/10.1107/S2052520617003109.

(40) Chai Discovery, J. Boitreaud, J. Dent, M. McPartlon, J. Meier, V. Reis, A. Rogozhnikov, and K. Wu, "Chai-1: Decoding the Molecular Interactions of Life," preprint, Synthetic Biology, October 11, 2024, https://doi.org/10.1101/2024.10.10.615955.

(41) M.L. Boby, D. Fearon, M. Ferla, M. Filep, L. Koekemoer, M.C. Robinson, The Covid Moonshot Consortium, J.D. Chodera, A.A. Lee, N. London, A. Von Delft, and F. Von Delft, "Open Science Discovery of Potent Noncovalent SARS-CoV-2 Main Protease Inhibitors," *Science* 382, no. 6671 (2023): eabo7201, https://doi.org/10.1126/science.abo7201.

CHAPTER 2

NONCLASSICAL NONCOVALENT σ-HOLE INTERACTIONS IN PROTEIN STRUCTURE AND FUNCTION: CONCEPTS FOR POTENTIAL PROTEIN ENGINEERING APPLICATIONS[1]

2.1 Summary

The structures and associated functions of biological molecules are driven by noncovalent interactions, which have classically been dominated by the hydrogen bond (H-bond). Introduction of the s-hole concept to describe the anisotropic distribution of electrostatic potential of covalently bonded elements from across the periodic table has opened a broad range of nonclassical noncovalent (*nc*NC) interactions for applications in chemistry and biochemistry. Here, we review how halogen bonds, chalcogen bonds and tetrel bonds, as they are found naturally or introduced synthetically, affect the structures, assemblies, and potential functions of peptides and proteins. This review intentionally focuses on examples that introduce or support principles of stability, assembly and catalysis that can potentially guide the design of new functional proteins. These three types of *nc*NC interactions have energies that are comparable to the H-bond and, therefore, are now significant concepts in molecular recognition and design. However, the recently described H-bond enhanced X-bond shows how synergism among *nc*NC interactions can be exploited as potential means to broaden the range of their applications to affect protein structures and functions.

[1] The work in this chapter was written, formatted, and submitted as a review article in 2023 for *Chemistry, an Asian Journal*. I cowrote this paper, specifically the section on halogen bonds and hydrogen bond enhanced halogen bonds in protein engineering.

Walker MG, Mendez CG, Ho PS. Non-classical Non-covalent σ-Hole Interactions in Protein Structure and Function: Concepts for Potential Protein Engineering Applications. Chem Asian J. 2023 Apr 3;18(7)

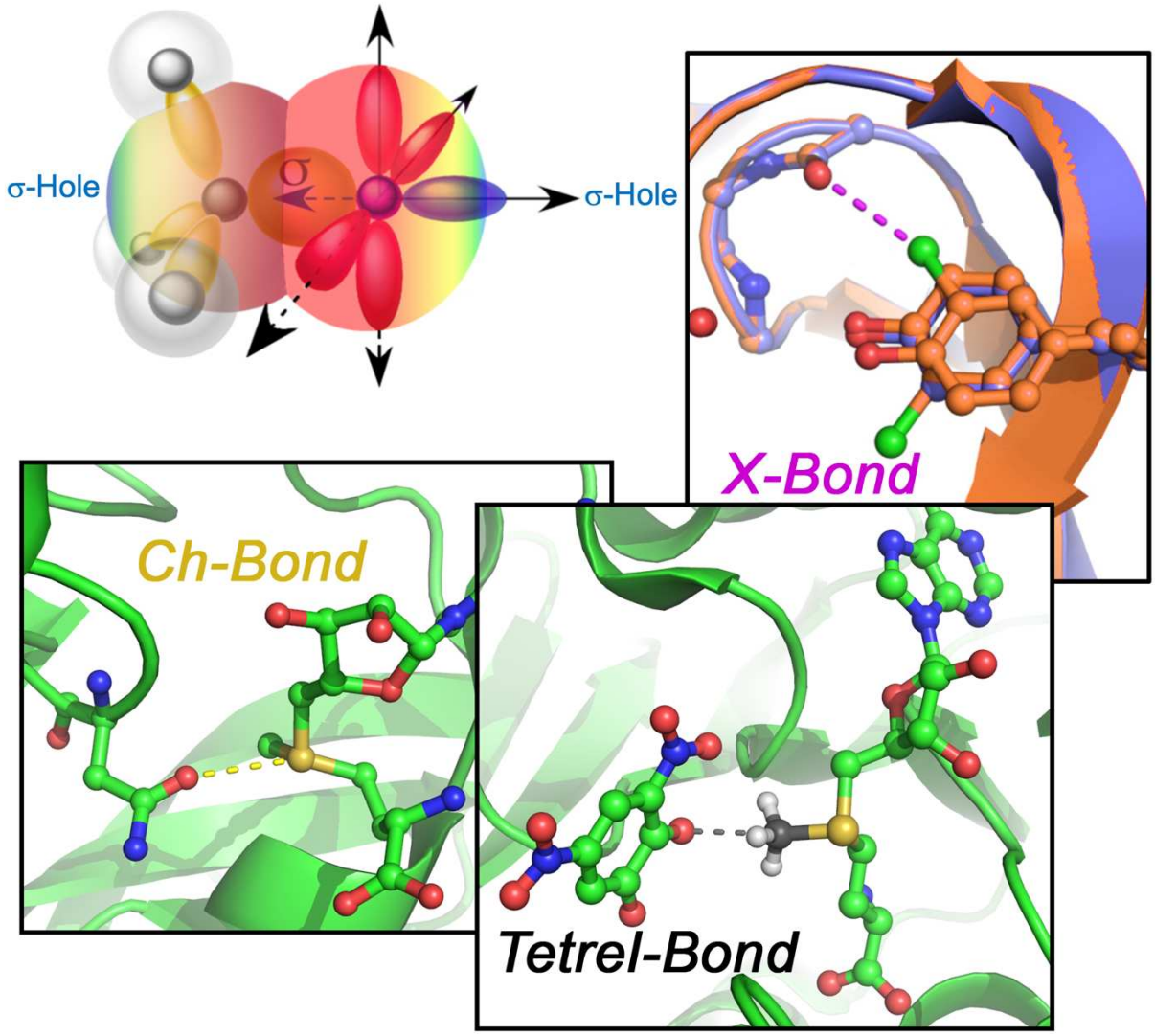


Figure 2.0: Graphical Abstract of Chapter 2.

2.2. Introduction

The folding of a protein is driven by noncovalent (NC) interactions, with the hydrogen bond (H-bond) being classically the dominant determinant of its functional three-dimensional conformation[1]. However, there is now a greater appreciation that nonclassical NC (*nc*NC) interactions involving elements from groups IV to VII of the periodic table can play significant roles in defining the structures and functions of molecular systems at various levels in chemistry[2], which can by extension be applied to the future design of new functional proteins.

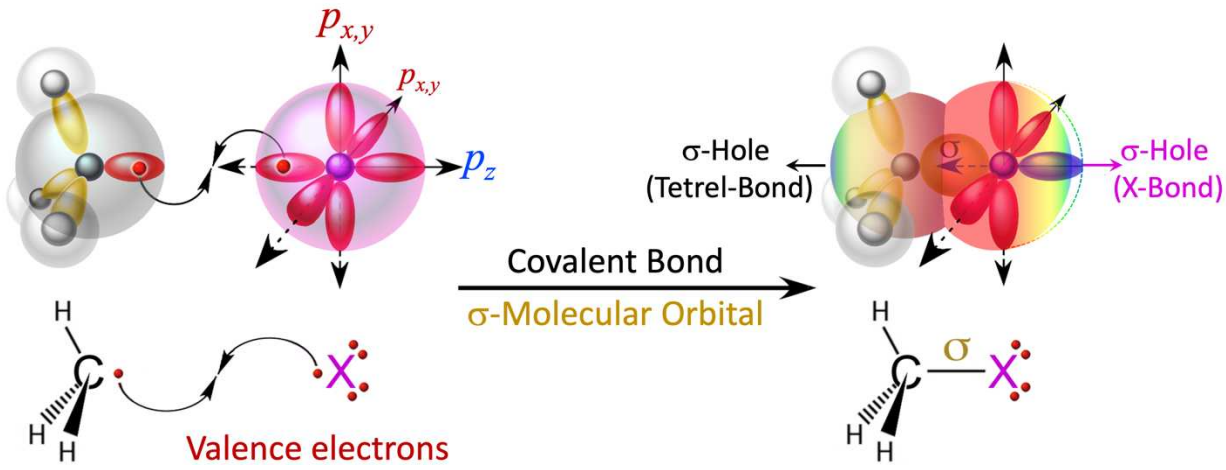


Figure 2.1. Schematic of the s-hole model to describe halogen and tetrel bonding. In this model, pairing an electron from the valence p_z-orbital of a halogen (X) with that of, for example, the carbon of a methyl group to form a single covalent bond results a s-molecular orbital (s-MO). We note that only the p-orbital valence electrons are shown in this figure. This s-MO subsumes the electron from the halogen p_z-orbital, leading to an electropositive crown and flattening of the van der Waals radius (the s-hole) but leaving an electronegative waist around the halogen substituent. This s-hole provides an electrostatic basis for the halogen bond. In similar fashion, a s-hole is created on the surface of the carbon of the methyl group opposite the s-MO, which in turn forms the electrostatic basis for a tetrel bond. Adapted from Ho[3].

A general model describing the class of *nc*NC interactions is the s-hole theory (Fig. 2.1)[4]. This model posits that when the valence electrons of an atom's p-orbital forms a covalent bond

with another atom, this orbital becomes depopulated in the direction diametrically opposed to the associated s-molecular orbital. The result is an electropositive crown called the s-hole, which sits directly opposite the covalent bond. The electropositive potential of the s-hole is accentuated by electron withdrawing groups. Such *nc*NC interactions have now been given bonding names specific to their groupings in the periodic table: tetrel bonds, pnictogen bonds, chalcogen bonds, and halogen bonds (Fig. 2.2) but have more generally been referred to as "s-hole bonds" [5]. Even the H-bond has at times been placed under this broader classification, if we consider its 1s orbital to be similarly polarized when a hydrogen is chemically bonded to any other atom[6]. Each of these *nc*NC interactions pair its s-hole (analogous to the H-bond donor) with an electron-rich acceptor, typically an oxygen, nitrogen, or delocalized electron system (such as an aromatic ring). We should note here, however, that the s-hole describes only the electrostatic component of such interactions. For at least the halogen bond, additional components, including charge-transfer (its initial conception[7]), and dispersion[8], as well as electrostatics[9] have been supported by various levels of computational theory. It is likely that these concepts will be consistent for describing the various physical contributions to *nc*NC interactions across the group IV to VII elements[10].

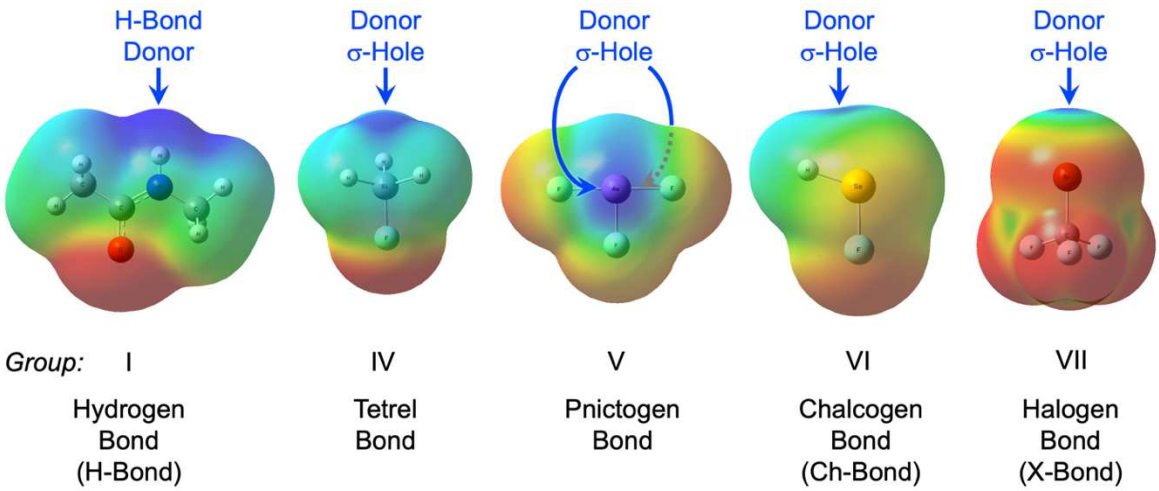


Figure 2.2. Examples of s-hole leading to electrostatic "bonding" interactions for various elemental groups. Electrostatic potential (ESP) maps were calculated for a H-bond donor in *N*-methylacetimide (NMA) as a model peptide bonds of proteins, for a tetrel bond donor in $FSiH_3$ from group IV, for a pnictogen bond donor in F_3As from group V, for a chalcogen bond (Ch-bond) donor in FSeH from group VI, and for halogen bond (X-bond) donor in F_3CBr from group VII. Positive electrostatic potentials are shown as blue surfaces while negative electrostatic potentials are in red. Adapted from Czarny, *et al.*[2f].

In chemistry, these *nc*NC interactions have been shown to facilitate the assembly of supramolecular complexes (including crystals and liquid crystals), the recognition of ions and ligands, and the catalysis of organic reactions[2e]. The study of *nc*NC interactions in biology have not been as extensive as in small organic and inorganic molecular systems but have been found to primarily be important as recognition elements in ligands/inhibitors against protein targets[2f, 11].

As the field begins to mature, there is a growing mountain of studies on how various *nc*NC interactions are found in a multitude of biomolecular and biomimetic systems. The intent of this review, however, is not to be a comprehensive summation of all occurrences of *nc*NC interactions (see Jena, *et al.*,[12] for a more complete summary of various *nc*NC interactions found in proteins system) but explicitly on exploring the applications of *nc*NC interactions that are emerging as being potential design elements to engineer functional proteins. In particular, we

will focus on how halogen, chalcogen, and tetrel bonding are applied to understand and/or manipulate the assembly, stability, and function of peptide and protein systems, along with their role in the recognition of functional cofactors. For each of these *nc*NC interactions, we start by discussing studies that survey the Protein Data Bank (PDB[13]) for their occurrence, followed by a discussion of the energies of such interactions, and conclude with examples of where such interactions have been found or engineered to control the structure, stability, and/or function of a peptide/protein system. These discussions initiate with the halogen bond, since its role in biology has been most extensively studied.

2.3. The Halogen Bond

The halogen bond (X-bond, for short) is now well recognized[14] and is finding utility across a broad range of chemical applications[15]. The X-bond was first recognized as an important *nc*NC interaction in biology through a survey of the PDB[11a]. This initial PDB survey showed that X-bonds were primarily important in recognition of halogenated ligands and inhibitors to protein targets, including the hormone thyroxine to its receptors[16] and agonists and antagonists against cancer related protein kinases[17] (Fig. 2.3). The primary X-bond acceptor in such cases, similar to H-bonds, is the carbonyl oxygen of the peptide backbone of the protein, likely because of its ubiquity in such systems. The s-hole model predicts X-bonding potentials to be dependent on the polarizability of the donor halogen and, consequently, correlated with the size of the halogen. The predicted order of X-bond strength would thus be I > Br > Cl >> F (F is seldom seen as X-bond donors in biological systems). The energies of biological X-bonds have been estimated through a unique assay system[18] that takes advantage of the similarity between H- and X-bonds to compete one against the other in a four-stranded DNA Holliday junction[19]. The resulting X-

bond energies range from +0.8 kcal/mol (F) to -0.15 kcal/mol (Cl) to -4.2 kcal/mol (Br) and to -6.5 kcal/mol (I)[20]. These fundamental concepts of biological X-bonds will be foundational for any design element in protein engineering.

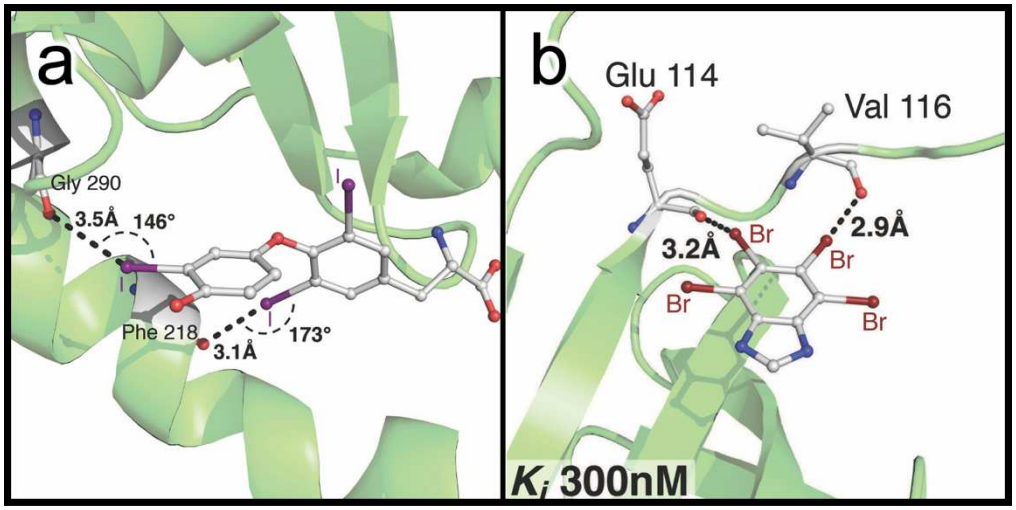


Figure 2.3. Examples of X-bonds in protein-ligand complexes. a. The hormone 3,5,3'-triiodothyroxine in the binding site of human thyroid hormone receptor. The distances and angles that define the geometries for two X-bonds involved in recognition of the iodinated hormone are labeled. b. Two X-bonds that facilitate binding of the tetrabromobenzimidazole inhibitor to cyclin kinase CK2 are shown, which contribute to the near 600-fold reduction in its inhibitory effect (as reflected in the K_i in the nanomolar range) on the enzyme target. Adapted from Scholfield, *et al.*[17b].

Subsequent surveys further revealed the amphipathic nature of the halogen in serving as both X-bond donor and H-bond acceptor[21], the orthogonality between H- and X-bond donors when both are paired to the same acceptor[22] (Fig. 2.4b), and how water can serve as a bridge between X-bond donors and receptors[23]. Thus, the X-bond plays an important function in understanding and designing therapeutic agents against clinically important protein targets. The biomedical role of X-bonds has been reviewed extensively elsewhere[17b, 24] and, therefore, will not be discussed further here. In this review, we will focus on studies that explore the effects of

*nc*NC interactions that of halogenated systems that affect the structure, stability, and/or function of peptides and proteins.

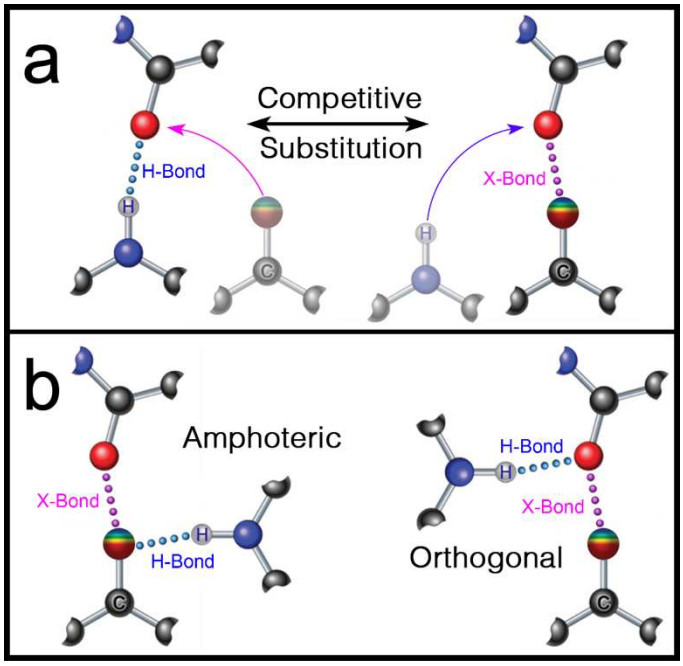


Figure 2.4. Relationships between X- and H-bonds in protein systems. a. The commonality in acceptors between H- and X-bonds indicates that they can compete and/or substitute for each other. b. The amphipathic nature of a halogen substituent (left)[25] allows it to serve as both an X-bond donor and H-bond acceptor simultaneously. In contrast, when an H-bond and a X-bond donor share the same acceptor atoms (right), the interaction is an orthogonal relationship[22] in respect to their perpendicular geometry and the independence of their interaction energies. Adapted from Rowe and Ho[21b].

2.3.1 Halogen bonds in the assembly of amino acids and peptides

The fundamental building block of proteins is the amino acid. Naturally halogenated amino acids have been found to play significant roles in affecting protein structure and function across a broad range of organisms. For example, halogenated tyrosine (Y) and tryptophan (W) in marine organisms have been found to facilitate protein crosslinking[26], while chloro- and bromotyrosines resulting from oxidative halogenation are markers for asthma in humans[27].

The assembly of amino acids into proteins requires formation of peptide bonds, which is mimicked, in its simplest form, as *N*-methylacetamide (NMA). Vasylyeva, *et al.*,[28] showed that cocrystals of NMA with dihalotetrafluorobenzene forms H-bonded layers (reminiscent of interactions that stabilize regular structures in proteins), with each NMA layer bridged by a pair of perpendicular X-bonds from individual aromatic ligands (Fig. 2.5a). The concept of orthogonality relating X- and H-bonds, initially conceptualized from surveys of proteins[22], thus extends even within simple peptide mimics.

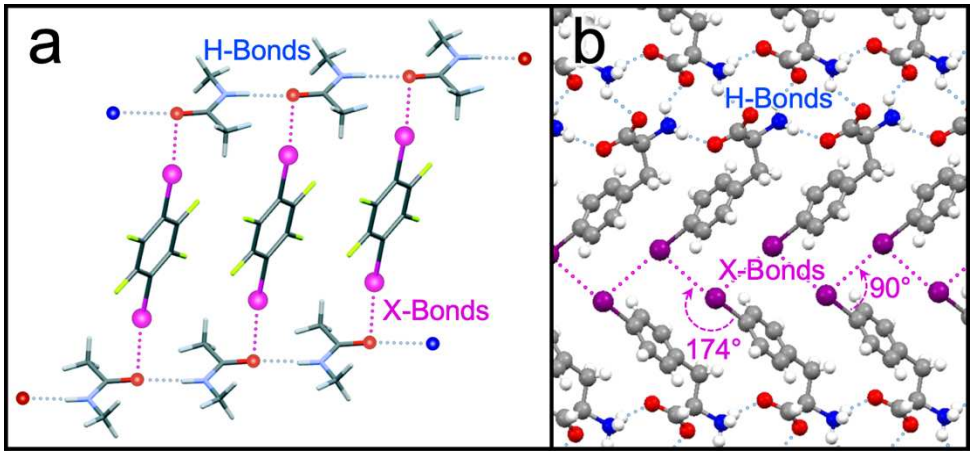


Figure 2.5. X-bonds bridging assemblies of an H-bonded peptide mimic (**a**) and H-bonded iodophenylalanine (**b**). **a**. X-bonds (magenta dots) from diiodotetrafluorobenzene bridge the antiparallel arrays of H-bonded (blue dots) antiparallel *N*-methylacetimide (NMA) chains. Adapted from Vasyleva, *et al.,*[28]. **b**. Crystal structure of *p*-iodophenylalanine shows chains assembled through N–H···O H-bonds between the amino and carboxylate groups, with each set of arrays linked through I···I X-bonds. Labeled are the near linear C–I···I angle aligning the electropositive s-hole of the donor to acceptor and the nearly perpendicular C–I···I angle for the electronegative waist of the acceptor to the donor. Data from Bertolani, *et al.*,[29].

Studies by Bertolani, *et al.*[29], show that isolated iodophenylalanine (^{I}F) amino acids crystallize with the backbone amides forming H-bonded layers (Fig. 2.5b). These layers are connected through a series of I···I X-bonds in which the electropositive s-hole of one iodine points towards the electronegative waist of the second. Similarly, *para*-iodophenylalanine-phenylalanine (^{I}F-F) dipeptide was found to crystallize as b-sheet like complexes, with I···I X-

bonds bridging across the H-bonded sheets[30]. Both the mono and di-iodinated phenylalanyl dipeptides were seen to form nanotube-like macrostructures in solutions of DMSO, with red-shifts in the infrared spectra in DMSO solvated solids or films supporting formation of weak X-bonds in the solid state. Thus, even in the simplest cases, X-bonds facilitate assembly of halogenated amino acids and dipeptides into extended complexes.

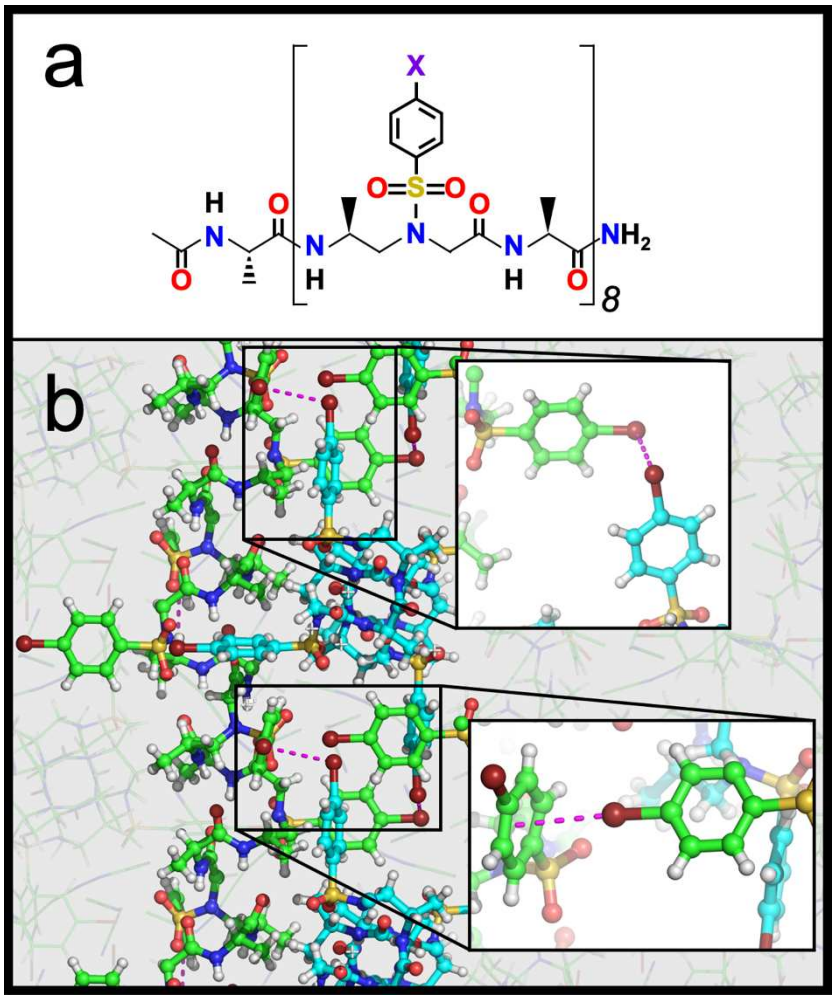


Figure 2.6. Peptide foldamers assembled from *para*-halo-a/L-sulfono-g-amino acids. a. Sequence of the synthetic peptide containing eight repeats of the halogenated unnatural amino acid. b. Foldamer structure showing one set of 4_{13} helices packed against neighboring parallel (green) and perpendicular (cyan) oriented helices. A type I I···I contact between parallel and perpendicular helices is shown in the top inset, while an I···p X-bond between parallel helices is shown in the bottom inset. Data from Teng, *et al.*,[31], adapted from Ho and Anderson[32].

A synthetic peptide construct containing eight unnatural α/L-sulfono-γ-amino acids (halogenated with Cl, Br, or I at the *para*-position of the phenyl ring, Fig 2.6a) was seen by Teng, *et al.,*[31] to self-assemble into right-handed 4_{13}-type helical (4_1 symmetry) "foldamer" structures (Fig. 2.6b), with sets of helices stacked in near continuous alternating parallel and perpendicular arrangements. X-bonds from the halogen of the sulfono side chain to the center of the phenyl ring of the sulfono side chain across pairs of adjacent parallel-aligned helices help hold the assemblies together (Fig. 2.6b inset), with QM calculated energies ranging from 2.5 kcal/mol to 2.8 kcal/mol. Close halogen-halogen interactions were also seen linking perpendicularly packed helices but these synthons did not appear to be consistent with standard X-bond geometries in proteins[14, 17b].

X-bonding can play a role in the assembly of amyloid-related peptides. The DFNKF pentapeptide is a core motif that triggers fibrillation of human calcitonin polypeptides. Bertolani, *et al*., saw that halogenation of the phenylalanine (F) residues in this pentapeptide promotes the formation of hydrogels, which are seen to result from formation of fibrillar networks[29]. The dependence of fibrillation of the halopeptides follows the trend I > Br > Cl, which supports the involvement of X-bonds in the process (the unhalogenated peptide did not form hydrogels under similar experimental conditions). A follow-up crystallographic study by this group found that waters were the X-bond acceptors in these fibers, thereby suggesting that the role of the halogen is to structure the solvent at the wet interface of the hydrogels and potentially promote other NC (including C–H··· p and p-p stacking) interactions[33].

2.3.2 Halogen bonds affecting structure and function of peptides and proteins

In humans, oxidative halogenation of tyrosines in proteins by myeloperoxidase or eosinophile peroxidase[34] have been associated with aging, neurodegenerative disease and cancer[35]. A recent study showed that a single halogenated tyrosine (^{X}Y) at the interface between the *N*- and *C*-termini can affect the ability of tubulin-like proteins to undergo GTP-dependent self-organize and polymerize[36], although X-bonding was apparently not a factor. This raises the questions of how halogenation affects the structure, stability, and function of proteins and whether X-bonds can affect these properties.

Danelius, *et al.*[37] studied the structures of a series of cyclic peptide and found that an intramolecular Cl···O X-bond across the peptide (from a chloroalanine at the third residue to a methylates serine at position 8) helped stabilize a b-hairpin foldamer over conformation (Fig. 2.7). The cyclic peptide was seen by solution-state NMR to be an intrinsically dynamic system. In the presence of the Cl···O X-bond, the peptide was seen to have a 74% probability of folding into the b-hairpin foldamer, while that probability was reduced to 58% when the Cl is replaced by an H-bonding OH and to only 29% when replaced by a non-interacting CH_3 group. A QM calculation indicated that the Cl··O X-bond was weak, with the stabilizing potential being only ~1.5 kcal/mol. This cyclic peptide system, therefore, was presented as a means to evaluate the effect and potential energy of weak *nc*NC interactions such as the Cl···O X-bond in a peptide/protein system.

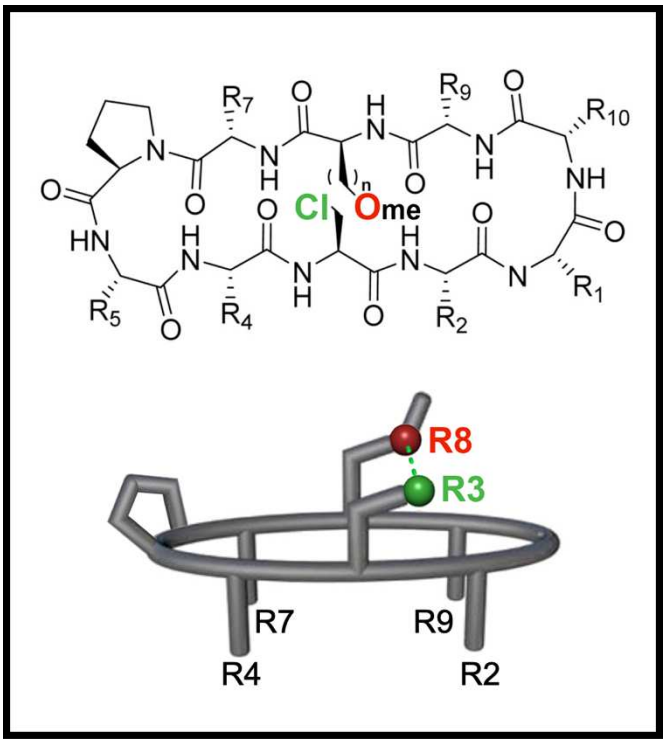
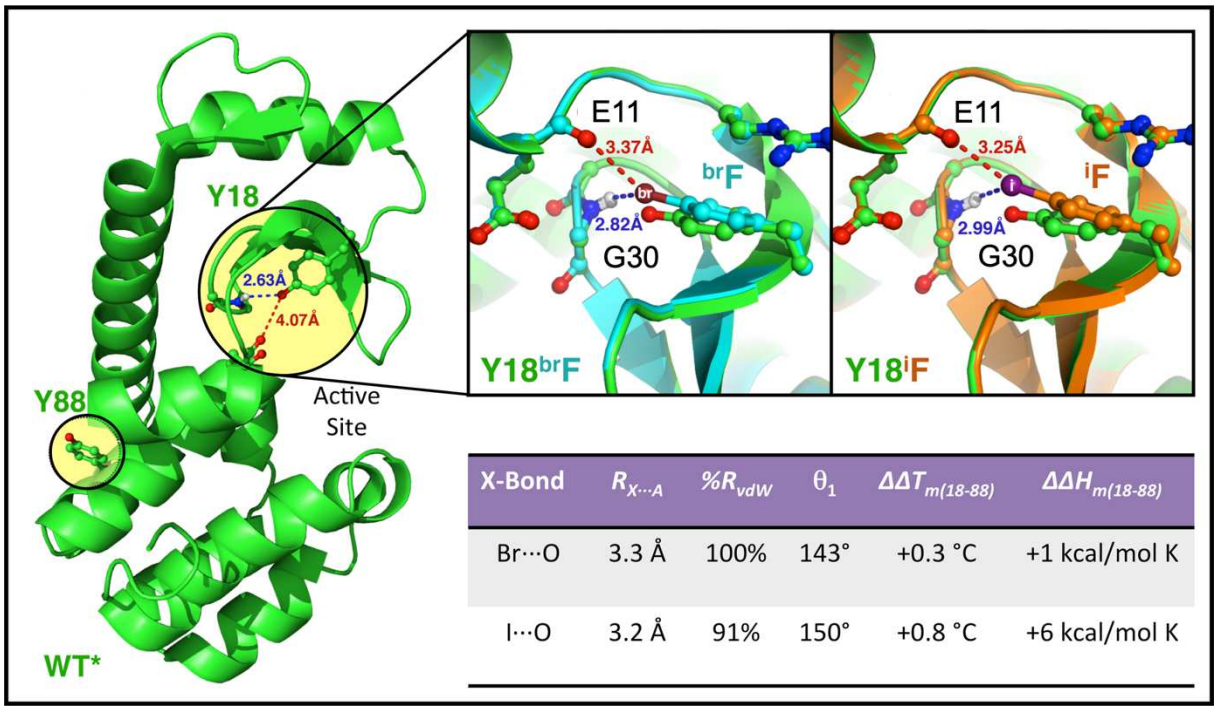


Figure 2.7. NMR structure of a cyclic decapeptide held in a b-hairpin conformation by an X-bond. The chemical schematic of the cyclic decapeptide is shown on top, while a cartoon of its b-hairpin structure determined by solution-state NMR is shown on the bottom. The chlorine (green sphere) of residue 3 sidechain (R3) is shown X-bonded (green dotted line) to the oxygen (red sphere) of the methylserine residue at position 8 (R8). Adapted from Danelius, *et al.*[37].

Halogenated amino acids can be incorporated into proteins through site-specific incorporation of unnatural amino acids[38]. One application of halogenation is to facilitate phasing of X-ray diffraction data to help solve the single crystal structures of proteins, with the argument that such modifications would have little effect on the protein itself. Indeed, the first example of such a strategy was seen with the incorporation of an iodophenylalanine (^{I}F) a hydrophobic pocket of T4 lysozyme (a classic model enzyme to study protein folding), with no X-bond acceptors within 4 Å of the iodine[39]. More recently, the halogenated amino acids have been incorporated in order to introduce X-bonding as a design element to affect the stability and activity of proteins and enzymes.



X-Bond	$R_{X\cdots A}$	$\%R_{vdW}$	θ_1	$\Delta\Delta T_{m(18\text{-}88)}$	$\Delta\Delta H_{m(18\text{-}88)}$
Br···O	3.3 Å	100%	143°	+0.3 °C	+1 kcal/mol K
I···O	3.2 Å	91%	150°	+0.8 °C	+6 kcal/mol K

Figure 2.8. T4 lysozyme constructs with tyrosines at residue 18 (Y18) or 88 (Y88) replaced by halophenylalanines (^{X}F). The structure of the wildtype T4 lysozyme (WT*) is shown as a cartoon, with the active site labeled and tyrosine to be mutated (Y18 and Y88) shown in ball and stick models and highlighted with yellow circles. The left inset compares details of the crystal structures of the brominated Y18^{br}F construct (cyan) against WT* (green). The Br···O X-bond to glutamate11 (E11) is shown as red dots and its distance labeled, while the H-bond from the amino group of glycine 30 (G30) is shown as blue dots and its distance labeled. The right inset compares details of the crystal structures of the iodinated Y18^{ir}F construct (orange) against WT* (green). The I···O X-bond to E11 is shown as red dots and its distance labeled, while the H-bond from the amino group of G30 is shown as blue dots and its distance labeled. The table lists for the brominated and iodinated constructs the distance between the halogen X-bond donor and oxygen acceptor ($R_{X\cdots A}$), the associated percent of the sum of the van der Waals radii for the two interacting atoms ($\%R_{vdW}$), the X-bond angle from donor to acceptor (q_1), the difference in DSC measured melting temperature between the mutant at the Y18 *versus* Y88 position ($\Delta\Delta T_{m(18\text{-}88)}$), and the difference in DSC measured melting enthalpies between the mutant at the Y18 *versus* Y88 position ($\Delta\Delta H_{m(18\text{-}88)}$). Data from Scholfield, *et al.*[40].

In our own laboratory, we asked the question of whether an X-bond can be introduced to replace an H-bond to stabilize the structure of T4 lysozyme[40]. Our first set of studies focused on tyrosine at position 18 (Y18, Fig 8), which sits on a b-strand adjacent to the active-site of the enzyme. Y18 accepts an H-bond from the NH of the peptide bond at glycine (G30) and is

bridged to the peptide oxygen of the catalytic glutamic acid E11[41] by a series of water molecules. The structure of the protein in which Y18 is replaced to either a bromo- or iodophenylalanine (Y18^{br}F or Y18^{i}F mutant constructs) showed that the halogens formed X-bonds to the backbone of E11, while maintaining the H-bond from G30 (Fig. 2.8 inset). The solvent exposed amino acid Y88 served as a control for the effect of substituting the hydrophilic OH group with a either Br or I substituent atoms. In all cases, substitution of the OH with a halogen (whether at Y18 or Y88) reduced the overall stability of the protein as reflected in the lower melting temperature (T_m) and enthalpy of melting (ΔH_m) determined by differential scanning calorimetry (DSC), indicating that the OH of Y18 is critical to the stability of T4 lysozyme (Fig. 8). However, the in comparing these DSC thermal parameters among the mutants at the two positions, the halogens in the X-bonding site of Y18 increased the T_m by 0.3° to 0.8°C and ΔH_m by +1 to +6 kcal/mol relative to the analogous mutants at the solvent exposed Y88 position. Thus, we showed that introducing a halogen into a solvent exposed position is destabilizing (consistent with the effect of exposing a hydrophobic group, in this case a halogen atom) but the ability to form an intramolecular X-bond rescues this loss in stability by upwards of 6 kcal/mol.

Recognizing the critical nature of the interactions of Y18, we next asked whether an X-bond can be introduced to augment the stability of T4 lysozyme[42]. For this study, we replaced Y18 with a *meta*-halotyrosine (Y18$^{m\mathrm{X}}$Y mutants, where X is Cl, Br, or I). The crystal structures of these constructs showed that this pocket did not fully accommodate the halogens (Fig 9). The chlorine of the $^{m\mathrm{Cl}}$Y18 side chain was seen to be 54% sitting inside the pocket and X-bonded to the carbonyl oxygen of glycine G28, while 46% of this side chain was rotated outside and not forming any *nc*NC intramolecular interactions (Fig. 2.9a). The halogen of the $^{m\mathrm{Br}}$Y18 construct was 22% X-bonded inside and 78% outside, while that of $^{m\mathrm{I}}$Y18 was 100% out (Fig. 2.9b),

indicating that this pocket in T4 lysozyme is very rigid and can only partially accommodate the smaller Cl halogen substituent. Even then, however, $^{m\mathrm{Cl}}$Y18 was seen to increase the T_m by 1°C and the ΔH_m by 3 kcal/mol relative to the wildtype enzyme, indicating that this partial X-bond seen in the structure resulted in a more stable protein. Furthermore, the chlorinated mutant showed ~15% increased enzymatic activity compared to the wildtype at elevated temperatures. A similar but smaller increase in activity was also observed for the brominated construct, suggesting that as the temperature increased, the protein pocket in which $^{m\mathrm{X}}$Y18 resides becomes more accommodating to the halogen substituents, thereby allowing more of the side chain to rotate to form a stabilizing X-bond.

We noticed, however, in the $^{m\mathrm{Cl}}$Y18 T4 lysozyme studies that the energy associated with the Cl···O X-bond was significantly larger than the chlorine X-bonding energies seen in any biomolecular system so far. A Cl···O energy of ~3 kcal/mol is greater than the 0.15 kcal/mol determined by competition with H-bonds in our prior DNA junction systems[18, 20] or the ~1.5 kcal/mol from the cyclic peptide system described above[37] and is more similar to what is expected for a Br···O X-bond. This anomalously large stabilizing energy led us to propose that the X-bonding potential of the chlorine was enhanced by the adjacent OH substituent of the tyrosyl side chain (Fig. 2.9c). Hydroxyl groups are known to be electron donating substituents and we would expect that the X-bonding potential of a halogen adjacent to an OH would be reduced. However, QM calculations indicate that when the hydroxyl is rotated such that the OH can form an H-bond to the electronegative waist of the halogen, the electropositive s-hole becomes significantly enlarged due to polarization effects, thereby increasing both the stabilizing potential and the angular range afforded the halogen as an X-bond donor. We called this an H-bond enhanced X-bond, or HB*e*XB for short.

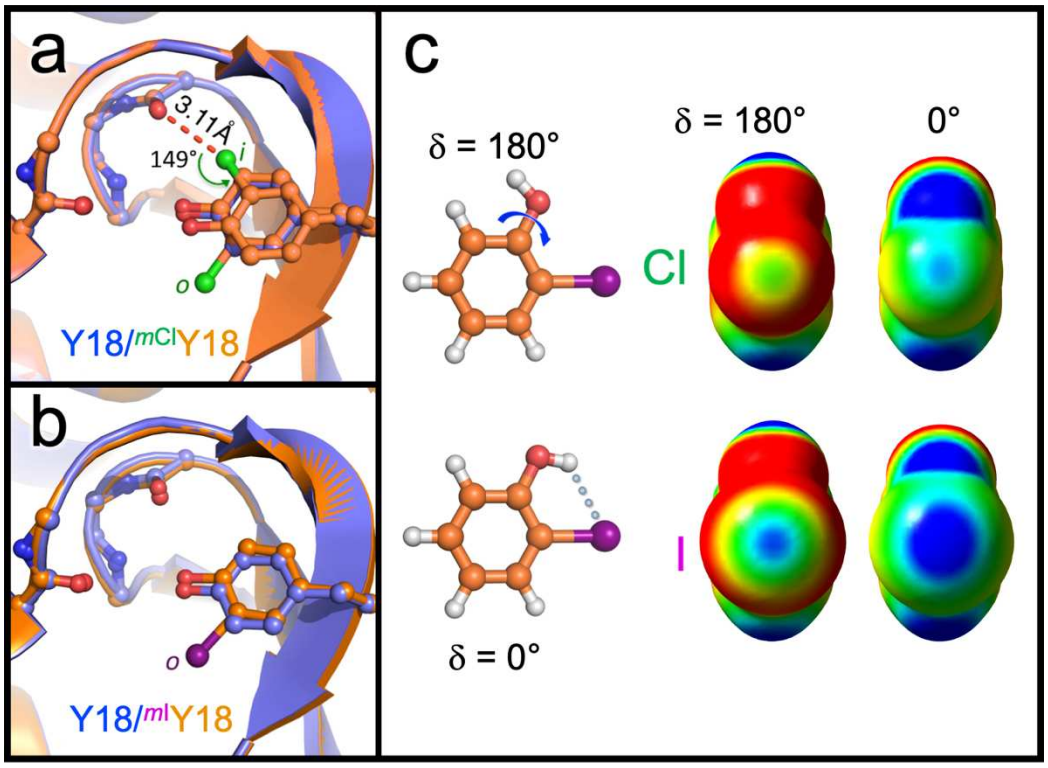


Figure 2.9. Comparison of the details of T4 lysozyme structure with tyrosine at position 18 (Y18) replaced by *meta*-halotyrosine (^{mX}Y18). a. Structure of the chlorinated ^{mCl}Y18 mutant (orange) compared to WT* (blue). The sidechain of ^{mCl}Y18 can be oriented to point the chlorine into the pocket (*i*) to form an X-bond (distance and q_1-angle as labeled) or out of the pocket (*o*) with no X-bond interaction with the protein. b. Structure of the iodinated ^{mI}Y18 mutant (orange) compared to WT* (blue). The sidechain of ^{mI}Y18 was seen to be oriented only to place the iodine out of the pocket (*o*) with no X-bond interaction with the protein. c. Effect of rotating the hydrogen of the hydroxyl group in the ^{mX}Y18 sidechain either away (d-angle = 180°) or towards (d-angle = 0°) the halogen substituent. In the d= 0° orientation, the OH serves as an H-bond donor to the electronegative waist of the halogen, which in turn enhances the s-hole of the halogen (Cl top and I bottom) and consequently its potential to serve as an X-bond donor. Adapted from Carlsson, *et al.*[42]

A similar effect was observed at nearly the same time by Berryman's group in a small molecule sensing system. In their studies, the affinity for halide ions of a bidentate anion receptor increased when an NH_2 substituent is added to form H-bonds to the two iodine X-bond donors of the iodopyridinium groups[43]. Surveys of the Cambridge Structural Database (CSD) and the PDB

indicates that the requirements for HB*e*XB interactions are prevalent in both small and large molecule systems[44]. There is some indication from QM studies that with multiple enhancing H-bonds, even fluorine can serve as a potent X-bond donor[45]. Thus, the concept of the synergistic HB*e*XB seems to be general across chemistry and biochemistry and greatly expands the range of energies and geometries available for molecular design applications compared to X-bonds alone.

2.4. The Chalcogen Bond

In biology, chalcogen bonds (Ch-bonds) primarily involve either S or Se as donors and O, N, or S as acceptors[11c]. Unlike the X-bond, the atoms of Ch-bond donors have typically two covalent bonds, which means that there will potentially be two s-holes, each aligned opposite of their associated covalent bonds[2d, 46]. Surveys for Ch-bonds would thus be complicated by potential interactions of acceptors with an additional electropositive surface, but also possible steric clashes with adjacent substituent groups. A survey of protein-ligand complexes in the PDB by Kriz *et al.,* showed that nearly a quarter (23%) of ligands containing S or Se showed geometries consistent with Ch-bonding[11c]. Biswal, *et al.*'s comprehensive survey of S/Se donors to O acceptors in the PDB found that Ch-bonds are dominated by S···O interactions in turn structural element of proteins (although all secondary structures of proteins were represented)[47]. The survey had identified primarily intermolecular Ch-bonds in protein-ligand complexes, but one interesting intramolecular S···O was found within the Cys558 residue (Fig. 2.10a). Quantum mechanical calculations indicated that the interaction energies of Ch-bonds selected from the survey could range from -0.7 to -14.1 kcal/mol. As with the X-bond, the focus of this discussion will not be on Ch-bonds associated with the design of protein inhibitors (of which there are several examples) but on one that is integral to the protein's natural function.

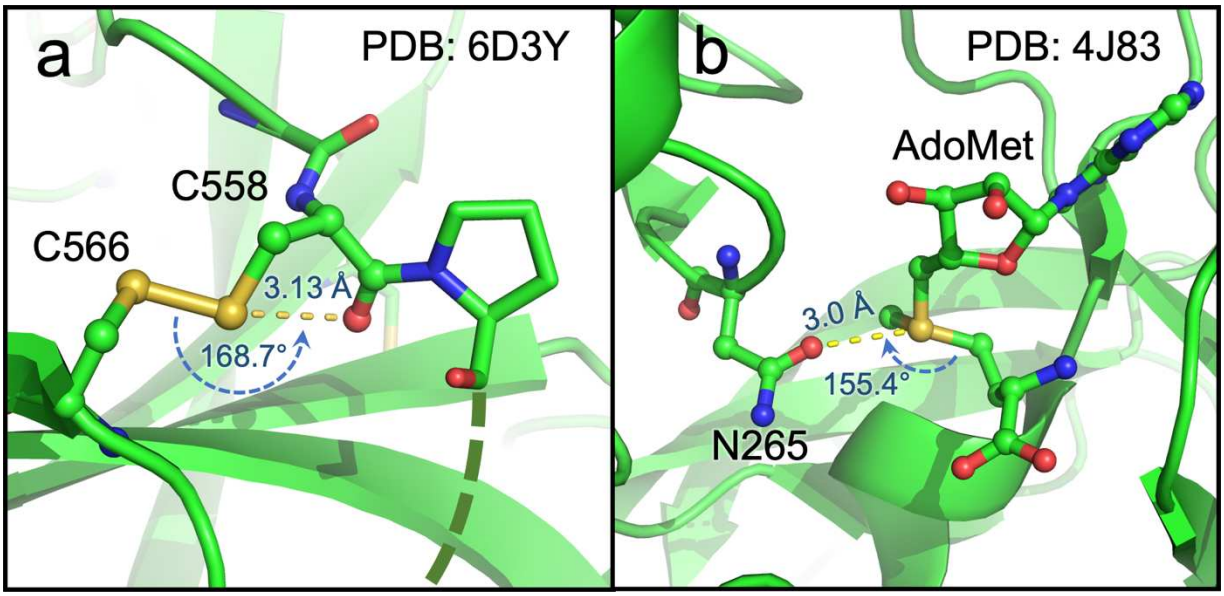


Figure 2.10. Chalcogen (Ch-bonds) in proteins. a. An intramolecular chalcogen bond (yellow dots) was observed from the sidechain sulfur to the peptide carbonyl oxygen of cysteine Cys588. The s-hole on this sulfur is established by the covalent S–S disulfide bond to cysteine C566. Data from Swedberg, *et al.*[48] (PDB code; 6D3Y). b. A chalcogen bond (yellow dots) is highlighted between the sulfur of AdoMet and carbocyclic oxygen of Asparagine at position 265 (N265) in histone-lysine N-methyltransferase Set 7/9. Data from Horowitz, *et al.,*[49] (PDB code 4J83).

A particularly interesting class of proteins where *nc*NC interaction, including Ch-bonds may play an important role in function are the *S*-adenosylmethionine (AdoMet)-dependent methyltransferases (MTases). AdoMet-dependent MTases are enzymes (ubiquitous across all phyla of life) that add methyl groups from AdoMet to small molecules in the biosynthesis of secondary metabolites, and2 to nucleic acids and proteins to control cellular functions. The SET domain containing lysine methyltransferase (SET7/9) specifically methylates a lysine on the histone 3 protein, which is a marker for transcriptional activation[50]. Fick, *et al.*[51] found that the oxygen of the asparagine265 (N265) side chain of SET7/9 is positioned 3.0 Å from the with the sulfur of the bound AdoMet cofactor, consistent with an S···O Ch-bond (Fig. 2.10b). QM calculation on models that mimic this interaction resulted in an overall (–S-CH_3)···O interaction

energy of ~ 16 kcal/mol, although it was not possible to disentangle the contributions of the S···O Ch-bond from the C–H···O H-bond. In order to determine the role of this Ch-bond, the authors replaced N265 with an alanine (N265A mutant). Removing the amide of the side chain resulted in ~10-fold reduction in the affinity of the enzyme for AdoMet and ~7-fold reduction in the catalytic rate but, again, it would be difficult to distinguish the contributions to the binding of each of the two classes of *nc*NC interactions. One possible approach may be to study the system with a selenium analog of AdoMet, which should significantly increase its potential as a Ch-bond donor but not the H-bonding potential of the methyl group. We will see in the next section how tetrel bonds may play a catalytic role in this same class of enzymes.

2.5. The Tetrel Bond

Although tetrel bonding in chemistry can involve any of the group IV elements (C, Si, Ge, Sn, or P), the primary donor for this class of interactions in proteins are the carbons of methyl groups. The identification of tetrel bonds to methyl groups would be complicated by potential weak C–H···Acceptor H-bonds and, therefore, care must be taken to distinguish among the two competing interactions. Mooibroek's 2019 survey of tetrel bonds published found nearly 2,800 such *nc*NC interactions in the PDB[52]. The survey showed that oxygen dominated over sulfur and aromatic acceptors, with water O representing over half the interactions (57%), followed by amide O (~24%), carboxylic O (~15%), thiol S (4%) and aromatic rings (<1%). QM calculations on interaction energies representing tetrel bond type interactions are weak (ranging from -0.8 kcal/mol to -3.7 kcal/mol) but not insignificant.

As with the Ch-bond, proteins where tetrel bonds may be particularly important are the AdoMet-dependent MTases. In this case, however, it is not the sulfur of AdoMet involved in the *nc*NC interaction but the methyl group that is being transferred, with the tetrel bond proposed as stabilizing a state that precedes the S_N2 transition state of the transfer reaction[53]. Although carbons of methyl groups would not typically form strong tetrel bonds, the positively charged sulfonium of the AdoMet cation could strengthen the *nc*NC interaction through polarization[49, 54]. A 2018 survey by Trievel and Schiener[55] of the PDB for tetrel bonds (Fig. 2.11) specifically in structures of AdoMet MTases found 20 unique examples of the *nc*NC interaction between the CH_3 of the AdoMet cofactor and the oxygen of the water solvent or a ligand in the active site (one example was to a chlorine of a ligand). The tetrel bond lengths ranged from 78% to 106% of the sum of the van der Waals radii of the carbon tetrel bond donor to O/Cl acceptor (average = 90.1 ± 7.5%) for AdoMet to ligands and 95.7% to 101% (average = 98.1 ± 1.8%) for interactions to water. The S–C···(O/Cl) angles ranged from 160° to 176° (average = 169.0° ± 5.0°). QM calculations on model systems mimicking four structures from survey showed that interaction energies CH_3···(O/Cl) ranges from -5.2 kcal/mol to -9.0 kcal/mol for neutral acceptors and potentially to -65.7 kcal/mol for an anionic oxygen acceptor. A nonbonding orbital analysis of these energies indicate that the contribution from the tetrel bond 1.6 to 8-times greater than that from the competing C–H···(O/Cl) H-bonds of the methyl group.

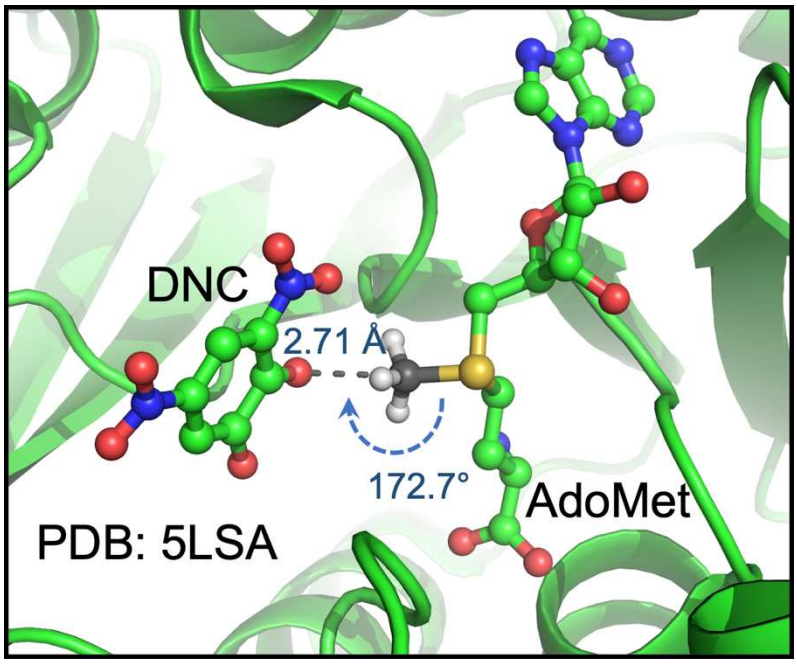


Figure 2.11. Tetrel bond in methyltransferase. A tetrel bond interaction (black dots) is shown between the methyl group of the AdoMet cofactor in human catechol O-methyltransferase and the ligand 3,5-dinitrocatechol (DNC). (PDB code 5LSA).

2.6. Conclusions and Perspectives

We had previously argued that to be considered biologically relevant, an *nc*NC interaction must 1) affect a biological function, 2) affect the structures/stabilities of the biological molecule that defines the function, and 3) have energies that are biologically relevant (*i.e.*, similar to that of H-bonds)[2f]. The discussion in this review demonstrates that X-bonds, Ch-bonds, and tetrel bonds all fit these criteria and all are fairly well represented in biological systems, as seen from surveys of the PDB. These surveys further point us in multiple directions for how and where such *nc*NC interactions can play significant roles in biology. The most obvious applications of these interactions would be towards inhibitor and drug design against therapeutical protein targets. This could include the design of small molecule ligands as well as peptides that bind specifically to target proteins. The intentional focus of the current review, however, is in *nc*NC

interactions that could serve as design elements applicable to engineering proteins that are more stable and/or with additional functions. An exciting area of development of *nc*NC interactions in chemistry is towards the design of new catalysts for organic reactions[56]. We see here that tetrel bonds may play a significant role already in catalyzing SAM dependent methyl transfer[49, 53, 55]. Thus, it we can now imagine how various *nc*NC interactions can be used to design new enzymatic catalysts.

Finally, we currently have a good understanding for how the energies of each type of *nc*NC interaction is defined directly by the element type (with the s-hole being more electropositive as we go down the column of each elemental group) as well as the electron withdrawing potential of whatever our donor atom is attached. Such effects are now readily estimated from higher level QM calculations. In addition, classical molecular mechanics/dynamics simulations can now start to model at least the X-bond in proteins and other biological systems by incorporating corrections to the typical spherically uniform forcefield parameters[57], including the positive extra point approach[58] and the forcefield for biological halogen bonds[59]. The concept that the synergistic HB*e*XB can as much as double the potential energy of an X-bond[42, 44], however, indicates that *nc*NC interactions can be strongly affected by its environment, potentially through synergistic coupling among different types of noncovalent interactions, not just H-bonds. Thus, to fully model and apply X-, Ch-, and tetrel bonds in protein engineering applications, we must better understand how each are affected by the ubiquitous H-bond and by each of the other *nc*NC interactions in the environments of proteins.

2.7. Funding

The studies from P. S. Ho's laboratory were supported by grants CHE-1905328 and CHE-2203161 from the National Science Foundation of the United States of America.

REFERENCES

(1) a. L. Pauling, R. B. Corey, H. R. Branson, *Proc Natl Acad Sci, USA* **1951**, *37*, 205-211; b. G. Fasman, *Prediction of Protein Structure and the Principles of Protein Conformation*, Plenum Press, New York and London, **1989**; c. E. N. Baker, in *International Tables for Crystallography Volume F: Crystallography of biological macromolecules* (Eds.: M. G. Rossmann, E. Arnold), Springer Netherlands, Dordrecht, **2001**, pp. 546-552; d. J. Kuriyan, B. Konforti, D. E. Wemmer, *The Molecules of Life*, Garland Science, Taylor & Francis Group, New York, NY, **2013**.

(2) a. A. Bauza, T. J. Mooibroek, A. Frontera, *Angew Chem Int Edit* **2013**, *52*, 12317-12321; b. A. Priimagi, G. Cavallo, P. Metrangolo, G. Resnati, *Acc Chem Res* **2013**, *46*, 2686-2695; c. S. Scheiner, *Acc Chem Res* **2013**, *46*, 280-288; d. L. Vogel, P. Wonner, S. M. Huber, *Angew Chem Int Edit* **2018**, *58*, 1880-1891; e. L. Brammer, *Faraday Discuss* **2017**, *203*, 485-507; f. R. S. Czarny, A. N. Ho, P. S. Ho, *Chem Rec* **2021**, *21*, 1240-1251.

(3) P. S. Ho, *Halogen Bonding I: Impact on Materials Chemistry and Life Sciences* **2015**, *358*, 241-276.

(4) a. T. Clark, M. Hennemann, J. S. Murray, P. Politzer, *J Mol Model* **2007**, *13*, 291-296; b. P. Politzer, K. E. Riley, F. A. Bulat, J. S. Murray, *Comput Theor Chem* **2012**, *998*, 2-8.

(5) a. P. Politzer, J. S. Murray, T. Clark, *Halogen Bonding I: Impact on Materials Chemistry and Life Sciences* **2015**, *358*, 19-42; b. J. S. Murray, P. Politzer, *Wires Comput Mol Sci* **2017**, *7*.

(6) J. S. Murray, K. E. Riley, P. Politzer, T. Clark, *Aust J Chem* **2010**, *63*, 1598-1607.

(7) O. Hassel, in *Nobel lectures, Chemistry 1963-1970.*, Elsevier publishing company, Amsterdam, **1972**.

(8) K. E. Riley, M. Vazquez, C. Umemura, C. Miller, K. A. Tran, *Chem-Eur J* **2016**, *22*, 17690-17695.

(9) a. K. Eskandari, H. Zariny, *Chem Phys Lett* **2010**, *492*, 9-13; b. H. J. Jahromi, K. Eskandari, *Struct Chem* **2013**, *24*, 1281-1287.

(10) a. G. Haberhauer, R. Gleiter, *Angew Chem Int Edit* **2020**, *59*, 21236-21243; b. D. J. Pascoe, K. B. Ling, S. L. Cockroft, *J Am Chem Soc* **2017**, *139*, 15160-15167; c. L. D. Santos, T.

A. Hamlin, T. C. Ramalho, F. M. Bickelhaupt, *Phys Chem Chem Phys* **2021**, *23*, 13842-13852; d. S. Scheiner, *Phys Chem Chem Phys* **2021**, *23*, 5702-5717.

(11) a. P. Auffinger, F. A. Hays, E. Westhof, P. S. Ho, *Proc. Natl. Acad. Sci., USA* **2004**, *101*, 16789-16794; b. A. Bauza, A. Frontera, *Crystals* **2016**, *6*; c. K. Kriz, J. Fanfrlik, M. Lepsik, *ChemPhysChem* **2018**, *19*, 2540-2548.

(12) S. Jena, J. Dutta, K. D. Tulsiyan, A. K. Sahu, S. S. Choudhury, H. S. Biswal, *Chem Soc Rev* **2022**, *51*, 4261-4286.

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

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

(15) G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati, G. Terraneo, *Chem Rev* **2016**, *116*, 2478-2601.

(16) N. F. Valadares, L. B. Salum, I. Polikarpov, A. D. Andricopulo, R. C. Garratt, *J Chem Inf Model* **2009**, *49*, 2606-2616.

(17) a. A. R. Voth, and Ho, P.S., *Curr. Topics Med. Chem.* **2007**, *7*, 1336-1348; b. M. R. Scholfield, C. M. Zanden, M. Carter, P. S. Ho, *Protein Sci* **2013**, *22*, 139-152.

(18) A. R. Voth, Hays, F.A., and Ho, P.S., *Proc. Natl. Acad. Sci., USA* **2007**, *104*, 6188-6193.

(19) a. B. F. Eichman, J. M. Vargason, B. H. M. Mooers, P. S. Ho, *Proc. Natl. Acad. Sci., USA* **2000**, *97*, 3971-3976; b. F. A. Hays, J. Watson, P. S. Ho, *J Biol Chem* **2003**, *278*, 49663-49666.

(20) a. M. Carter, P. S. Ho, *Cryst Growth Des* **2011**, *11*, 5087-5095; b. M. Carter, A. R. Voth, M. R. Scholfield, B. Rummel, L. C. Sowers, P. S. Ho, *Biochemistry* **2013**, *52*, 4891-4903; c. C. M. Vander Zanden, M. Carter, P. S. Ho, *Methods* **2013**, *64*, 12-18.

(21) a. Y. Lu, Y. Wang, Z. Xu, X. Yan, X. Luo, H. Jiang, W. Zhu, *J Phys Chem B* **2009**, *113*, 12615-12621; b. R. K. Rowe, P. S. Ho, *Acta Crystallogr B Struct Sci Cryst Eng Mater* **2017**, *73*, 255-264.

(22) A. R. Voth, Khuu, P., Oishi, K., and Ho, P.S., *Nature Chemistry* **2009**, *1*, 74-79.

(23) P. Zhou, J. Lv, J. Zou, F. Tian, Z. Shang, *J Struct Biol* **2010**, *169*, 172-182.

(24) a. Y. Lu, T. Shi, Y. Wang, H. Yang, X. Yan, X. Luo, H. Jiang, W. Zhu, *J. Med. Chem.* **2009**, *52*, 2854-2862; b. C. Bissantz, B. Kuhn, M. Stahl, *J Med Chem* **2010**, *53*, 5061-

5084; c. M. Z. Hernandes, S. M. Cavalcanti, D. R. Moreira, W. F. de Azevedo Junior, A. C. Leite, *Curr Drug Targets* **2010**, *11*, 303-314; d. R. Wilcken, M. O. Zimmermann, A. Lange, A. C. Joerger, F. M. Boeckler, *J Med Chem* **2013**, *56*, 1363-1388.

(25) C. B. Aakeröy, S. Panikkattu, P. D. Chopade, J. Desper, *Crystengcomm* **2013**, *15*, 3125-3136.

(26) A. Pizzi, C. Pigliacelli, G. Bergamaschi, A. Gori, P. Metrangolo, *Coordin Chem Rev* **2020**, *411*.

(27) U. Egner, J. Kratzschmar, B. Kreft, H. D. Pohlenz, M. Schneider, *Chembiochem* **2005**, *6*, 468-479.

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

(29) A. Bertolani, L. Pirrie, L. Stefan, N. Houbenov, J. S. Haataja, L. Catalano, G. Terraneo, G. Giancane, L. Valli, R. Milani, O. Ikkala, G. Resnati, P. Metrangolo, *Nat Commun* **2015**, *6*, 7574.

(30) A. Pizzi, L. Catalano, N. Demitri, V. Dichiarante, G. Terraneo, P. Metrangolo, *Peptide Sci* **2020**, *112*.

(31) P. Teng, G. M. Gray, M. Zheng, S. Singh, X. Li, L. Wojtas, A. van der Vaart, J. Cai, *Angew Chem Int Ed Engl* **2019**, *58*, 7778-7782.

(32) P. S. Ho, D. M. Anderson, in *Halogen Bonding in Solution* (Ed.: S. Huber), Wiley-VCH Verlag GmbH & Co, Weinheim, Germany, **2021**, pp. 335-362.

(33) A. Pizzi, N. Demitri, G. Terraneo, P. Metrangolo, *Crystengcomm* **2018**, *20*, 5321-5326.

(34) A. Bachi, I. Dalle-Donne, A. Scaloni, *Chem Rev* **2013**, *113*, 596-698.

(35) a. A. O. Olaniran, E. O. Igbinosa, *Chemosphere* **2011**, *83*, 1297-1306; b. M. B. Feeney, C. Schoneich, *Antioxid Redox Sign* **2012**, *17*, 1571-1579; c. A. L. Santos, A. B. Lindner, *Ox Med Cell Longev* **2017**, *2017*; d. R. Tinikul, N. Lawan, N. Akeratchatapan, P. Pimviriyakul, W. Chinantuya, C. Suadee, J. Sucharitakul, P. Chenprakhon, D. P. Ballou, B. Entsch, P. Chaiyen, *FEBS J* **2021**, *288*, 3246-3260.

(36) H. Sun, H. Jia, O. Kendall, J. Dragelj, V. Kubyshkin, T. Baumann, M.-A. Mroginski, P. Schwille, N. Budisa, *Nat Commun* **2022**, *13*.

(37) E. Danelius, H. Andersson, P. Jarvoll, K. Lood, J. Grafenstein, M. Erdelyi, *Biochemistry* **2017**, *56*, 3265-3272.

(38) a. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182-188; b. L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498-500.

(39) J. M. Xie, L. Wang, N. Wu, A. Brock, G. Spraggon, P. G. Schultz, *Nat Biotech* **2004**, *22*, 1297-1301.

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

(41) B. W. Matthews, *FASEB J* **1996**, *10*, 35-41.

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

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

(44) A. M. S. Riel, R. K. Rowe, E. N. Ho, A. C. Carlsson, A. K. Rappe, O. B. Berryman, P. S. Ho, *Acc Chem Res* **2019**, *52*, 2870-2880.

(45) S. Scheiner, *J Phys Chem* **2022**, *126*, 9691-9698.

(46) J. S. Murray, P. Lane, P. Politzer, *J Mol Model* **2009**, *15*, 723-729.

(47) H. S. Biswal, A. K. Sahu, B. Galmes, A. Frontera, D. Chopra, *ChemBiochem* **2022**, *23*.

(48) J. E. Swedberg, G. J. Wu, T. Mahatmanto, T. Durek, T. T. Caradoc-Davies, J. C. Whisstock, R. H. P. Law, D. J. Craik, *J Med Chem* **2019**, *62*, 552-560.

(49) S. Horowitz, L. M. A. Dirk, J. D. Yesselman, J. S. Nimtz, U. Adhikari, R. A. Mehl, S. Scheiner, R. L. Houtz, H. M. Al-Hashimi, R. C. Trievel, *J Am Chem Soc* **2013**, *135*, 15536-15548.

(50) a. H. B. Wang, R. Cao, L. Xia, H. Erdjument-Bromage, C. Borchers, P. Tempst, Y. Zhang, *Mol Cell* **2001**, *8*, 1207-1217; b. J. P. Mariot, J. Y. K'nevez, *P I Mech Eng K-J Mul* **2002**, *216*, 203-211.

(51) R. J. Fick, G. M. Kroner, B. Nepal, R. Magnani, S. Horowitz, R. L. Houtz, S. Scheiner, R. C. Trievel, *Acs Chemical Biology* **2016**, *11*, 748-754.

(52) T. J. Mooibroek, *Molecules* **2019**, *24*.

(53) R. W. Woodard, M. D. Tsai, H. G. Floss, P. A. Crooks, J. K. Coward, *J Biol Chem* **1980**, *255*, 9124-9127.

(54) a. U. Adhikari, S. Scheiner, *J Phys Chem A* **2013**, *117*, 10551-10562; b. S. Horowitz, J. D. Yesselman, H. M. Al-Hashimi, R. C. Trievel, *J Biol Chem* **2011**, *286*, 18658-18663.

(55) R. C. Trievel, S. Scheiner, *Molecules* **2018**, *23*.

(56) a. R. L. Sutar, S. M. Huber, *ACS Catal* **2019**, *9*, 9622-9639; b. S. Benz, A. I. Poblador-Bahamonde, N. Low-Ders, S. Matile, *Angew Chem Int Edit* **2018**, *57*, 5408-5412.

(57) M. C. Ford, P. S. Ho, *J Med Chem* **2016**, *59*, 1655-1670.

(58) a. M. A. A. Ibrahim, *J. Comput. Chem.* **2011**, *32*, 2564-2574; b. M. A. Ibrahim, *J Phys Chem B* **2012**, *116*, 3659-3669; c. M. Kolar, P. Hobza, *J. Chem Theory Comput* **2012**, *8*, 1325-1333.

(59) a. M. Carter, A. K. Rappe, P. S. Ho, *J. Chem Theory Comput* **2012**, *8*, 2461-2473; b. M. C. Ford, A. K. Rappe, P. S. Ho, *J. Chem Theory Comput* **2021**, *17*, 5369-5378.

CHAPTER 3

DESIGN OF A HALOGEN BOND CATALYZED ENDONUCLEASE[2]

3.1. Summary

In this study, we expand the repertoire of biological catalysts by showing that a halogen bond (X-bond) can functionally replace the magnesium (Mg^{2+}) cofactor in mouse endonuclease G (*m*EndoG). We mutated the metal coordinating glutamate E136 in *m*EndoG to a *meta*-halotyrosine (^{mX}Y, X = chlorine or iodine) to form a ^{mX}Y-*m*EndoG construct that is both acid and base catalyzed. Under basic conditions, the enzyme is inactivated by the metal chelator ethylene diamine tetraacetic acid (EDTA), indicating that the halogen substituent facilitates deprotonation of tyrosyl hydroxyl group, allowing recruitment of Mg^{2+} to restore the metal dependent catalytic center. At low pHs, we observe that the ^{mX}Y-*m*EndoG is resistant to EDTA inactivation and that the iodinated constructed is significantly more active than the chlorinated analogue. These results implicate a hydrogen bond (H-bond) enhanced X-bond as the catalyst in the ^{mX}Y-*m*EndoG, with asparagine N103 serving as the H-bond donor that communicates the protonation state of histidine H104 to the halogen. This model is supported by mutation studies and electrostatic potential calculations on models for the protonated and unprotonated ^{mX}Y···N103···H104 system compared to the Mg^{2+} coordination complex of wildtype. Thus, we have designed and engineered an enzyme that utilizes a previously unexplored unnatural catalyst in its active site—

[2] The work in this chapter was written, formatted, and submitted as a research article in *Proceedings of the National Academy of Sciences, USA*. I contributed to the EDTA and pH titration assays, HBeXB Bridge mutagenesis and activity assays, and quantum energy calculations.

M.G. Walker, C.G. Mendez, A.N. Ho, R.S. Czarny, A.K. Rappé, & P.S. Ho, Design of a halogen bond catalyzed DNA endonuclease, *Proc. Natl. Acad. Sci. U.S.A.* 122 (14) e2500099122, https://doi.org/10.1073/pnas.2500099122 (2025).

a catalytic X-bonding enzyme, or *cX*-Zyme—by controverting what constitutes a metal catalyst in biochemistry.

3.2. Significance

A fundamental concept in chemistry is that the properties of elements are periodic. Magnesium is a ubiquitous alkaline Earth (group II element) metal cofactor in enzymes that catalyzes many biochemical reactions, including DNA processing. We show here that the catalytic Mg^{2+} in a DNA endonuclease can be functionally replaced by the iodine or chlorine (group VII elements) of a halogenated tyrosine. This unnatural amino acid forms a hydrogen bond enhanced halogen bond to render the DNA backbone susceptible to hydrolysis, mechanistically analogous to but different from Mg^{2+}. This unique catalytic center opens the field of synthetic molecular biology to the design of novel enzyme active sites and potentially challenges the fundamental concept of what defines a metal catalyst in biology.

3.3. Introduction

Nature is adept at exploiting a limited number of chemical elements and mechanisms to evolve all the functional biological molecules that define life on Earth. Recently, synthetic biologists have started to expand the molecular tools to design functional biomolecules beyond what has naturally evolved. For example, non-canonical amino acids (ncAA) have been introduced to create new types of proteins[1], while unnatural nucleic acids can be incorporated into DNA to expand the genetic code[2]. Directed evolution has become an important approach to extending natural selection to engineer biomolecules with new biological and chemical functions[3], including the site-specific incorporation of ncAAs.

There has not been as much progress in exploiting the breadth of chemical catalytic mechanisms to design new enzymes with novel mechanisms and/or functions. Unnatural amino acids have been successfully engineered into enzymes to improve their catalytic functions and/or substrate selectivities but their mechanisms of action remain largely "natural"[4]. For example, an unnatural amino acid has been introduced into an enzyme as a copper chelator to catalyze the cleavage of small non-coding RNAs[5] but metal catalysis is a well-established mechanism for chemical cleavage of nucleic acids.

Nonclassical noncovalent interactions (halogen, chalcogen, pnictogen bonds) are now emerging as new catalyst for classic chemical reactions, including halogen abstraction and Diels-Alder reactions[6, 7]. Recently, unnatural enzyme cofactors have been designed to utilize the pnictogen bond of a chelated antimony V metalloid center to catalyze the reduction of a fluorogenic quinoline substrate[8]. However, since this cofactor sits at the protein surface rather than the active site, the catalyzed reaction is not stereospecific. In the current study, we have designed an unnatural enzyme active site in which the catalytic magnesium dication (Mg^{2+}) of

mouse endonuclease G (*m*EndoG, Fig. 3.1a) is functionally replaced by a halotyrosine ncAA. A hydrogen bond enhanced halogen bond is implicated in the catalytic mechanism, thereby introducing a new type of chemistry for designing catalysts for synthetic molecular biology.

Halogen bonds (X-bonds) are non-classical noncovalent interactions[9] that are now commonly used in chemistry to design novel materials[10]. The X-bond is most readily described by the σ-hole model[11], where the σ-molecular orbital of the covalent bonded halogen subsumes the electron from the atomic p_Z-orbital, resulting in an electropositive "σ-hole" at the tip and leaving an electronegative waist around the halogen substituent. X-bonds are analogous to and, thus, interchangeable with hydrogen bonds (H-bonds)[12, 13] and are seen to primarily replace the functions of H-bonds in recently engineered chemical catalysts[6].

X-bonds in biology[14] contribute to recognition of halogenated inhibitors and drugs[15-17] and have been engineered to control the conformations of DNA[18, 19] and of peptides and proteins[20-22]. In this study, we address the question of whether a metal cofactor can be replaced by an X-bond by displacing the Mg^{2+} cofactor in *m*EndoG with a *meta*-halogenated tyrosine ($^{m\mathrm{X}}$Y) (Fig. 1c-d). *m*EndoG specifically cleaves 5-hydroxymethylcytosine ($^{5\mathrm{hm}}$C) modified DNAs[23] in the context of four-stranded DNA junctions[24]. DNA cleavage by *m*EndoG follows a classic endonuclease mechanism where an Mg^{2+} cofactor coordinates to and renders the DNA phosphodiester backbone susceptible to cleavage (Fig. 3.1b). Mutagenesis studies on the human homolog of EndoG supports this overall mechanism[25]. In addition, Robertson, *et al.*, had shown that a histidine to asparagine mutation at position 97 (H97N) inactivates *m*EndoG[23], supporting the requirement of a general base to provide the nucleophilic water. Finally, we will show here that replacing glutamate E136 that helps coordinate the Mg^{2+} cofactor with a bulky tyrosine also inactivates the mouse enzyme, thereby supporting the role of the metal in the cleavage

mechanism. In this study, we demonstrate that the ^{mX}Y-*m*EndoG constructs show the same specificity as the wild type (WT) enzyme but functions through a unique catalytic mechanism in which the halogen substituent mimics the "metallic" electrostatic properties of the Mg^{2+} cofactor through a hydrogen bond enhanced halogen bond (HB*e*XB).

3.4. Results

Replacing E136 (Fig. 3.1c) in *m*EndoG with an ^{mX}Y (X = Cl or I, Fig. 1d and e) creates ^{mX}Y-*m*EndoG constructs that cleave ^{5hm}C-modified DNA junction and duplex substrates (Fig. 3.1f), resulting in products that are identical in size to those from the WT-*m*EndoG. In contrast, the unhalogenated tyrosyl E136Y mutant is significantly less active across all conditions, indicating that the halogen substituent of the ^{mX}Y ncAA is the essential catalytic element in the ^{mX}Y-*m*EndoGs. The *para*-phenylalanine construct (^{pX}F) is similarly or less inactive (Supplemental Figure S3.1), meaning that simply adding a halogen to any aromatic amino acid is insufficient to induce catalysis.

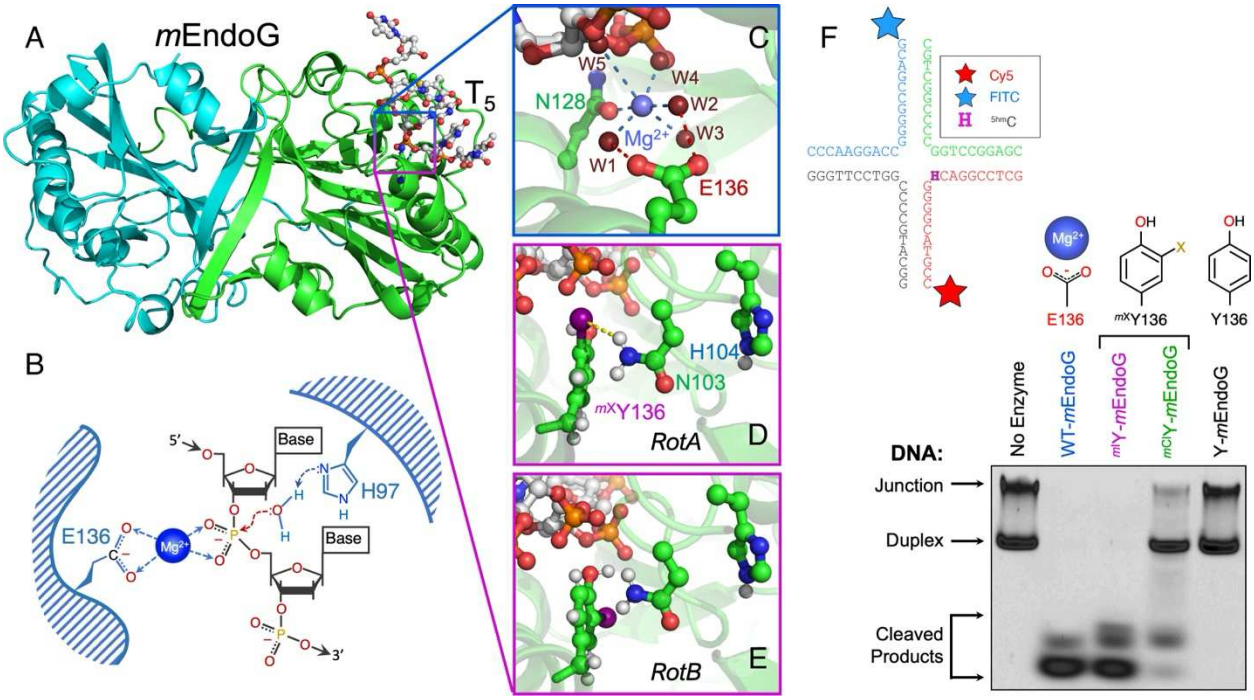


Fig. 3.1. Structure and catalytic activities of *m*EndoG constructs. a. Crystal structure of the *m*EndoG homodimer (separate monomers colored cyan and green), with the T_5 DNA from the *C. elegans* CPS6 ortholog of EndoG docked into the active site of the green monomer (see Materials and Methods). **b.** Schematic for the mechanism of *m*EndoG, showing glutamate E136 coordinated to the catalytic Mg^{2+} and histidine H97 serving as a general base to deprotonate the nucleophilic water. **c.** Active site in WT-*m*EndoG shown with Mg^{2+} coordinated (dashed lines) to the carboxylate oxygens of E136 through bridging waters and the carbonyl oxygen of asparagine N128. **d.** Model of the E136 replaced by ^{mI}Y136, generating the ^{mI}Y-*m*EndoG construct. This *RotA* model positions the iodine to potentially form X-bonds to the scissile phosphate oxygens of the DNA. Asparagine N103 is shown as an H-bond donor to bridge ^{mI}Y136 to the protonated histidine H104. **e.** The *RotB* model of ^{mI}Y-*m*EndoG construct positions the iodine away from the DNA phosphate oxygens and N103. **f.** Junction substrate and DNA cutting activities of non-halogenated and halogenated *m*EndoG constructs. Schematic of the DNA junction substrate (top left), showing the 5-hydroxymethylcytosine base (^{5hm}C, bold, outlined text) position and the strands labeled at their 5'-ends with cyan (Cy5, red star) or fluorescein (FITC, blue star) fluorophores. Duplex DNA substrates result from incomplete formation of junctions from the sequences. Polyacrylamide gel electrophoresis (PAGE, bottom right) analysis of the DNA substrates (four-stranded junction and duplex, lane 1) and products after treatment with WT-*m*EndoG (lane 2), iodinated ^{mI}Y-*m*EndoG (lane 3), chlorinated ^{mCl}Y-*m*EndoG (lane 4), or the nonhalogenated tyrosyl Y-*m*EndoG (lane 5).

The apparent rate constants (k_{APP}) for catalytic cleavage of ^{5hm}C-junctions by WT-*m*EndoG under pseudo-first order reaction conditions, where the enzyme is in excess, had previously been determined[24] to be ~0.2 to 0.5 min^{-1} and are reproduced here. The equivalent rate for $^{m\mathrm{I}}$Y-*m*EndoG ($k_{APP} \approx 0.06$ min^{-1}) is ~10-fold lower than that of WT-*m*EndoG at *p*H 7.0, while the chlorinated $^{m\mathrm{Cl}}$Y-*m*EndoG construct ($k \approx 0.03$ min^{-1}) is ~2-fold lower than $^{m\mathrm{I}}$Y-*m*EndoG and ~20-fold lower than WT-*m*EndoG. The higher activity of $^{m\mathrm{I}}$Y-*m*EndoG over $^{m\mathrm{Cl}}$Y-*m*EndoG mirrors the dependence of the σ-hole on the size and the polarizability of halogen substituents (26), thereby implicating X-bonding as responsible for facilitating DNA cleavage. The $^{m\mathrm{X}}$Y-*m*EndoGs are significantly more active than the Mg^{2+} dependent EndoG ortholog from *Caenorhabditis elegans* [27], which is >200-fold less active ($k_{APP} \approx 8 \times 10^{-4}$ min^{-1}) than WT-*m*EndoG. Thus, the $^{m\mathrm{X}}$Y-*m*EndoGs catalyzed reactions fall well within the range of rates for this class of Mg^{2+} dependent DNA nucleases. Furthermore, the $^{m\mathrm{X}}$Y-*m*EndoGs show the same specificity for four-stranded junctions over duplex DNA as WT-*m*EndoG, as evident in $^{m\mathrm{Cl}}$Y-*m*EndoG (Fig. 3.1f), where nearly all the junction and very little of the duplex DNAs are cleaved.

The differences in activities among the $^{m\mathrm{X}}$Y-*m*EndoG constructs and relative to WT-*m*EndoG are not defined by their binding affinities to the ^{5hm}C-junction substrate (Supplemental Figure S3.2 and Supplemental Table S3.1). While the iodinated $^{m\mathrm{I}}$Y-*m*EndoG shows an ~17-fold lower affinity for the substrate compared to WT, that of the chlorinated $^{m\mathrm{Cl}}$Y-*m*EndoG for the DNA substrate is higher than both WT-*m*EndoG and $^{m\mathrm{I}}$Y-*m*EndoG, indicating that the activities of the haloenzymes are dependent on the intrinsic catalytic potential of the halogen. We will primarily focus the remainder of our discussion on the more active iodinated $^{m\mathrm{I}}$Y-*m*EndoG.

The *p*H dependence of $^{m\mathrm{I}}$Y-*m*EndoG and WT-*m*EndoG were assayed to probe the differences in their catalytic mechanisms. WT-*m*EndoG showed a *p*H profile (Fig. 3.2a) with the

maximum activity seen between *p*H 7.0 and 7.5, associated with two *pKa* values (pKa_1 = 6.8 and pKa_2 = 7.7). The increase in activity going from pH 6 to 7.5 is consistent with a mechanism in which His97 acts as a general base to increase the nucleophilicity of the hydrolytic water (Fig. 3.1b). We interpret the reduction in activity at *p*H ≥ 7.7 as the enzyme switching from overall positively to negatively charged, thereby reducing its affinity for the negative DNA substrate. This interpretation is supported by an isoelectric point *p*I = 8.1 calculated for *m*EndoG using the Isoelectric Point Calculator 2.0[28].

In contrast, the *p*H profile for $^{m\mathrm{I}}$Y-*m*EndoG is inverted relative to that of the wildtype, showing high activities both above and below *p*Hs 8.0 (Fig. 3.2b). We probed the two *p*H effects separately. The high *p*H active $^{m\mathrm{I}}$Y-*m*EndoG has a *p*Ka ≈ 8.7 (Fig. 3.2c) and is as or more sensitive than the wildtype to inactivation by the metal chelator ethylenediaminetetraacetic acid (EDTA, Supplemental Fig. S3.3). We interpret these results as the halogen lowering the *p*Ka for deprotonation of the tyrosyl hydroxyl group through standard inductive effects (the *p*Ka of the OH of $^{m\mathrm{X}}$Y is reported to be ~8.3 for both the chloro- and iodo-forms[29, 30]. This mechanism may account for the nominal activity observed for the non-halogenated tyrosyl mutant. The sensitivity to EDTA inactivation suggests that the deprotonated $^{m\mathrm{X}}$Y recruits an Mg^{2+} to the active site, recapitulating the metal-dependent mechanism of the wildtype enzyme.

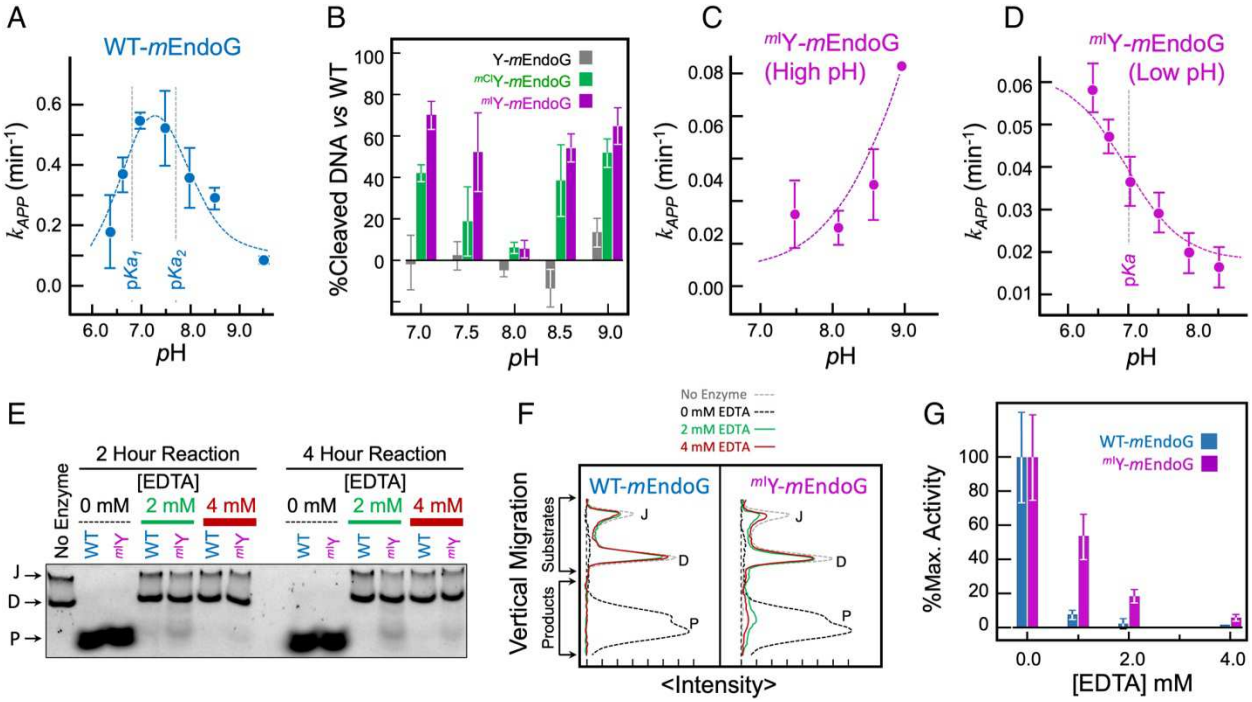


Fig. 3.2. Effects of *p*H and EDTA on the DNA cutting activity of WT- and halogenated ^{mX}Y-*m*EndoG. a. Apparent pseudo-first order kinetic rate constants (k_{APP}) for WT-*m*EndoG. The *p*H titration curve (dotted line) was modeled using standard Henderson-Hasselbach equations[36, 37]. The titration curve was modeled with two *p*Ka values (pKa$_1$ = 6.78, pKa$_2$ = 7.69), with R^2 = 0.91. **b.** Quantitation of cleaved DNA products after 30 minutes reaction from halogenated (^{mCl}Y-*m*EndoG, green bars, and ^{mI}Y-*m*EndoG, magenta bars) and unhalogenated (grey bars) tyrosine mutant constructs relative to those from WT-*m*EndoG at *p*Hs from 7.0 to 9.0. **c.** Effect of *p*H on k_{APP} for the base (High *p*H) facilitated ^{mI}Y-*m*EndoG activity, analyzed similarly to WT-*m*EndoG in **a**. The Henderson-Hasselbach curve was fit with a *p*Ka = 8.7 (R^2 = 0.9). **d.** Effect of *p*H on k_{APP} for the acid (Low *p*H) facilitated ^{mI}Y-*m*EndoG activity. The Henderson-Hasselbach curve was fit with a *p*Ka = 6.98 (R^2 = 0.98). **e.** PAGE analysis of junction (J) and duplex (D) substrates and products (P) after treatment with WT-*m*EndoG or ^{mI}Y-*m*EndoG in 0 mM, 2 mM, or 4 mM concentrations of EDTA at pH 7.0 for 2- or 4-hours. **f.** Normalized intensity (<Intensity>) profiles of the scanned gel lanes from **e** using the NIH ImageJ program[38] for WT-*m*EndoG and ^{mI}Y-*m*EndoG after 4 hours catalyzed reactions in 0 mM, 2 mM and 4 mM EDTA. **g.** Percent activity for WT-*m*EndoG (blue) and ^{mI}Y-*m*EndoG (magenta) from **f** and Supplemental Figure S5 relative to their respective maximum activities at 0 mM EDTA. Error bars are the standard errors of the mean from at least 3 independent measurements.

The higher activity under acidic conditions is consistent with general acid facilitated catalysis, with a single *p*Ka value of ~6.95 (Fig 3.2d and Supplemental Figure S3.4). Under these conditions, $^{m\mathrm{I}}$Y-*m*EndoG is significantly less sensitive to EDTA inactivation than the WT-*m*EndoG (Fig. 3.2e-f), with the wildtype losing over 90% of its activity at 1 mM EDTA (Supplemental Fig. S3.5) while the iodoenzyme shows residual activity up to 4 mM EDTA. Thus, the $^{m\mathrm{I}}$Y-*m*EndoG construct is shown to be significantly less metal dependent than wildtype. The loss of activity of $^{m\mathrm{I}}$Y-*m*EndoG at higher EDTA concentrations can be attributed to the change in conformation of the junction substrate as the chelator strips Mg^{2+} from the DNA[31], an effect seen with other DNA junction processing enzymes[32].

We interpret the acid facilitated activity of $^{m\mathrm{I}}$Y-*m*EndoG to an X-bond being augmented by an H-bond to the electronegative waist of the halogen in a synergistic interaction that we had previously identified as an H-bond *enhanced* X-bond (an HB*e*XB)[22, 33]. This leaves the question as to what the H-bond donor is for the HB*e*XB. Our prior studies on T4 lysozyme had shown that the hydroxyl group of a $^{m\mathrm{X}}$Y amino acid can serve as the H-bond donor to an HB*e*XB interaction[22]. However, the *p*Ka ≈ 7 for this reaction is ~1.3 *p*H units lower than any singly halogenated tyrosine found in the literature[29, 30], thereby precluding the tyrosyl OH and pointing to a histidine being the more likely H-bond donor to the HB*e*XB.

In surveying the WT-*m*EndoG structure[24], we found H104 to be the closest histidine to the $^{m\mathrm{I}}$Y residue but is too distant to be a direct H-bond donor. However, an intervening asparagine (N103) is present that could accept an H-bond from H104 and donate an H-bond to $^{m\mathrm{I}}$Y (Fig. 3.1c). This arrangement suggested that the *p*H-dependent protonation state of H104 could be communicated to the halogen through the H-bonded N103 bridge (Fig. 3.3a). To test this model, we mutated N103 or H104 separately to alanine residues (N103A and H104A mutants,

respectively) in the ^{ml}Y-*m*EndoG and WT-*m*EndoG constructs (Fig. 3.3b). Both the N103A and H104A mutants of ^{ml}Y-*m*EndoG were found to be inactive. In contrast, the N103A-WT-*m*EndoG showed diminished but substantial activity, while the H104A-WT-*m*EndoG mutant showed near wildtype activity. Thus, both N103 and H104 are essential to the catalytic activity of ^{ml}Y-*m*EndoG but neither are required in the catalytic mechanism of WT-*m*EndoG.

To further probe this mechanism, we calculated the electrostatic potential (ESP) surfaces of simplified molecular models that represent the various protonated and unprotonated H104···N103···X networks and compared these ESP values to those of models for the coordinated Mg^{2+} environment (Fig. 3.3c). The ESP were calculated for the ^{mX}Y136 models for both iodine and chlorine as the X substituent. The models used in the ESP calculations for the coordination complex with Mg^{2+} and of the H104···N103···X networks are described in greater detail in the Materials and Methods section.

We note that in the published crystal structure[24], the NH_2 of the amide side chain points towards H104, which, at first would preclude it from serving as an H-bond donor to the halogen of ^{ml}Y136. However, X-ray diffraction cannot distinguish between an oxygen and nitrogen at the reported resolution and, therefore, it is not clear what the actual orientation of the amide is in respect to H104 and ^{ml}Y136 and whether this H104···N103···X network model would require rotation of N103.

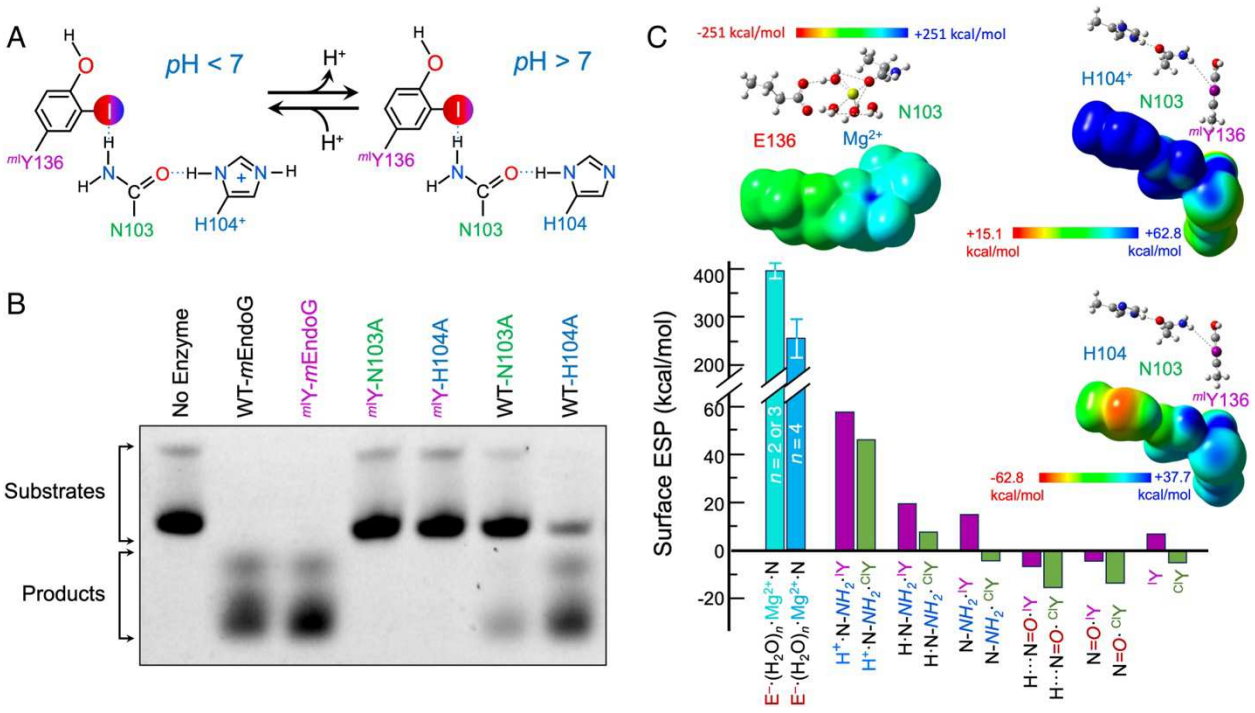


Fig. 3.3. H-bond network to enhance the ^{mI}Y X-bond and quantum mechanical (QM) analysis of electrostatic potentials (ESP) of models for WT- and ^{mX}Y-*m*EndoG active site residues. a. Model showing the proposed H104···N103···^{mI}Y136 network. In this model, protonation of H104 under acidic conditions creates the H-bonded H104-H$^+$···O=C-N103-NH$_2$···I network that significantly enhances the X-bonding potential of the ^{mI}Y136 iodine substituent. **b.** PAGE assay of DNA substrates and products generated by WT-*m*EndoG, ^{mI}Y-*m*EndoG and their respective N103A and H104A mutants. **c.** Comparison of ESP values for the Mg^{2+} cofactor in WT-*m*EndoG to those of the halogens in models for the ^{mI}Y-*m*EndoG active site. The model of Mg^{2+} coordinated to N128 and E136 in WT-*m*EndoG was constructed as complexes with acetamide and acetate molecules, respectively. All atoms, along with waters were placed according to their respective positions as seen in the crystal structure of WT-*m*EndoG[24]. The complex included either all four waters from the crystal structure (cyan bar), or with waters removed (leaving 2 or 3 coordinated solvent molecules, blue bar) that would allow coordination of the Mg^{2+} to the phosphate oxygens of the DNA substrate. ESP were calculated for models of ^{mCl}Y (green bars) and ^{mI}Y (magenta bars) residues alone, in complex with the carbonyl oxygen (N=O) or the amino group (NH$_2$) of acetamide (model for the N103 side chain) pointed towards the halogen, with neutral imidazole (model for the H104 side chain, labeled H) or protonated imidazole (H$^+$) in the network. Values were plotted for the most positive point on the respective ESP surfaces.

The ESP calculations consider both orientations of the amide with either the carbonyl O or the amino NH_2 pointing towards the halogen. However, the difference in overall energies calculated for these complexes with the carbonyl oxygen pointed towards the halogen (as in the H104···N103=O···X model) is +132 kcal/mol (for iodine) to +138 kcal/mol (for chlorine) compared to models with the amino group towards the halogen. This large positive energy difference strongly supports the H104···N103-NH_2···X model as being most likely for these studies.

The maximum ESP for Mg^{2+} in its coordination complex is highly positive (>200 kcal/mol), as expected for a dicationic metal. We modeled this complex with the Mg^{2+} coordinated to four waters, as seen in the crystal structure[24] or 1 or 2 waters (labeled W4 and W5 in Fig. 1c) removed, leaving 2 or 3 solvent molecules coordinated to mimic the metal complex when DNA sits in the active site (there was very little difference in the calculated ESP between these two water depleted complexes). The neutral halogens in all the $^{m\mathrm{X}}$Y136 models are significantly less positive (≤ 60kcal/mol). The ESP values of models calculated with the carbonyl oxygen of the N103 amide pointed towards either the $^{m\mathrm{X}}$Y136 is either negative or only slightly positive and overall less positive than that of a halophenol model alone. In contrast, all models with the amide NH_2 pointed towards the halogen have significantly positive or near neutral maximum ESP surfaces. The observation that $^{m\mathrm{I}}$Y-*m*EndoG shows at least weak activity under neutral *p*H conditions suggests that the amide is oriented to form an NH_2···I H-bond. Addition of the imidazole side chain of H104 to the model, whether protonated or not, renders all the ESPs to be positive. When this H104 imidazole is protonated, the maximum ESP values increase significantly.

We also compared the ESP of each H104···N103···X model with the tyrosyl hydroxyl pointed either towards or away from the halogen to determine how much a potential OH···X HB*e*XB could contribute to the electropositive σ-hole (Supplemental Figure S3.6). In the models absent the protonated $H104^+$, this OH···X interaction contributed significantly (upwards of 18 kcal/mol) but the overall ESP is less than that of the respective $H104^+$···N103···X models. With the H104 protonated and positively charged, the OH···X H-bond contributed less than 10 kcal/mol (<10%) of the electropositive charge for either the iodinated or chlorinated ESP values. Thus, above neutral pHs, the tyrosyl OH could create a significantly positive σ-hole but that this is dwarfed by the effects of the protonated $H104^+$ under acidic conditions where the $^{m\mathrm{X}}$Y-*m*EndoG constructs are fond to be most active.

We can now compare the various ESP values to determine how large a role electrostatics plays in the mechanism of $^{m\mathrm{I}}$Y-*m*EndoG versus WT-*m*EndoG. We see that the ESP of the protonated H104···N103···I model network is ~2-3-fold more positive than the neutral network with the unprotonated H104, which corresponds well with the ~3-fold higher k_{APP} at pH 6.5 versus 8.5 (Fig. 3.2d). In addition, the iodinated enzyme shows ~60% higher activity at pH 6.5 compared to the chlorinated analogue, while their ESPs differ by ~40%. Thus, electrostatics can account generally for the acid facilitated catalysis and the difference in reactivity between the two halogens. Finally, we can compare the ESPs of the $^{m\mathrm{X}}$Y-*m*EndoG models to those of the Mg^{2+} coordination complex. The k_{APP} for $^{m\mathrm{I}}$Y-*m*EndoG at *p*H 6.5 is ~ 9-fold lower than that of WT-*m*EndoG, which is comparable to the ~7-fold difference in ESP values for the protonated $H104H^+$···N103···I network and the 2/3 water coordinated complexes of Mg^{2+}. Thus, the ESP calculations correlate well with the relative activities of the *m*EndoG constructs studied here. These findings indicate that the enhancement of the σ-hole by the $H104H^+$···N103···I network

mimics the electrostatic effects of the Mg^{2+} cofactor in the *m*EndoG active site and generally supports our supposition that the DNA cutting activity of the ^{mX}Y-*m*EndoG enzyme under acidic conditions is catalyzed by the resulting HB*e*XB interaction.

3.5. Conclusions and Discussion

This study demonstrates that the essential catalytic Mg^{2+} cofactor in the *m*EndoG endonuclease can be functionally replaced by an unnatural halogenated ^{mX}Y amino acid, creating for the first time a catalytic X-bond dependent enzyme—a *cX*-Zyme. The inversion of the *p*H profile is the first evidence that ^{mX}Y-*m*EndoG follows a catalytic mechanism that is distinct from the wildtype. Could the acid facilitated catalysis by ^{mX}Y-*m*EndoG be attributed to residual metal in the enzyme's active site? Although halogens can form complexes with various transition metals[34], the finding that the ^{pI}F construct is essentially inactive supports our contention that a halogen substituent in itself does not recruit metals to catalyze DNA cutting. Furthermore, ^{mI}Y-*m*EndoG at low *p*Hs and in the presence of the metal chelator EDTA cuts DNA while the WT-enzyme is nearly completely inactive. One would expect any catalysis by residual Mg^{2+} bound to ^{mI}Y-*m*EndoG would be as or even more sensitive to EDTA as WT-*m*EndoG. Indeed, the high *p*H activity serves as an internal control, where the Mg^{2+} is more readily extracted by EDTA from the deprotonated halotyrosine of ^{mI}Y-*m*EndoG than from wildtype. Therefore, we are confident that ^{mI}Y-*m*EndoG cuts DNA without the need for Mg^{2+} in the active site, even though the metal may be necessary to define the conformation of the DNA junction substrate[32].

The evidence that ^{mX}Y-*m*EndoG is dependent on X-bonding, more specifically an HB*e*XB, comes initially from the observation that the nonhalogenated tyrosyl construct of *m*EndoG is largely inactive, indicating that the halogen substituent of the ^{mX}Y is the critical element for

catalytic activity. The finding that the iodinated $^{m\mathrm{I}}$Y-*m*EndoG is more active than the analogous chlorinated construct follows the trend expected for the polarizability of halogen substituents and their X-bonding potentials according to the σ-hole model[11], implicating X-bonding in the catalytic mechanism. The requirement of both H104 and N103 for catalysis further indicates that this is an HB*e*XB, with the N103 serving as the H-bond donor and as a bridge to communicate the *p*H-dependent protonation/charge state of H104. This model is supported by the correspondence between the ESP calculations on models of the H104H$^+$···N103···X network and Mg^{2+} coordination complex in their various states with the experimental k_{APP} values. Therefore, the experimental and computational evidence points to $^{m\mathrm{X}}$Y-*m*EndoG being a catalytic X-bonding enzyme (a *cX*-Zyme), whose DNA processing activity can be readily tuned according to the type of halogen incorporated into the unnatural amino acid.

The results from this study challenges what defines a metal catalyst in biology. We learned early in our chemistry classes that the properties of the elements are dictated by the periodic table, with metals occupying the left-hand columns of the table and non-metals relegated to the right-hand columns. In this study, we show that a group II alkaline earth metal catalyst (Mg^{2+}, an element on the left-hand side of the table) can be functionally replaced by a group XVII halogen on the far right-hand side. There is a growing appreciation that these basic assumptions about how nature works are challenged by quantum mechanics. In the simplest interpretation, the electropositive σ-hole of the halogen substituent in $^{m\mathrm{X}}$Y-*m*EndoG is mimicking the charge of Mg^{2+}, rendering the DNA's phosphodiester bond more susceptible to nucleophilic attack by a hydrolytic water. We have not, however, explored the possibility that additional components of X-bonds, including covalency, may contribute to the catalytic activity of the *cX*-Zyme. The X-

bond and indeed the plethora of so-called σ-hole interactions are quantum effects that now blur the line between metals and non-metals, at least in terms of their electrostatic properties.

The ^{m}XY-*m*EndoG is the first in what could potentially be a larger class of *cX*-Zymes. We note that *m*EndoG is a member of the HNH endonuclease, which have been shown in some cases to be tolerant of multiple amino acid substitutions in their active sites[35]. It would be interesting and important to determine whether an X-bond can be as effective as a catalyst in DNA processing enzymes that are either more or less accommodating to mutations of their catalytic residues. The HNH superfamily of enzymes is fairly large and, therefore, the *cX*-Zyme concept would still find broad application, even if restricted to this class of endonucleases.

This proof of concept that an X-bond can serve as an unnatural enzyme catalytic center, thus, presents a unique tool for the design of new enzymes in synthetic biology. X-bonds in small organic catalysts have been shown to accelerate an expanding range of important synthetic reactions[6, 7]. Incorporating analogous X-bonding facilitated reactions into proteins could allow the *cX*-Zyme concept to be extended beyond DNA processing to develop novel biological-based catalysts for a much broader range of chemical reactions besides simple hydrolysis.

3.6. Methods

Protein Constructs and Expression

Expression vectors for various forms of mouse endonuclease G (*m*EndoG) were constructed starting with the *pET-28a* plasmid containing the gene for the maltose binding protein (MBP)/*m*EndoG fusion protein (truncated at the N-terminus by 41 amino acids), as previously described[23, 24]. The MBP-*m*EndoG fusion is followed downstream by the sequence for a T7 promoter and the inhibitor dEndoGI from *Drosophila melanogaster*[35]. This starting construct expresses what we refer to as the wildtype (WT) form of *m*EndoG. An inactive form of the enzyme was constructed by replacing the Mg^{2+} chelating glutamate E136 with a tyrosine (the E136Y mutant). Expression vectors for haloenzymes containing unnatural *meta*-halogenated tyrosine (^{mX}Y) amino acids were constructed by replacing E136 in WT-*m*EndoG with the AMBER stop codon TAG[39-42].

Nonhalogenated forms of *m*EndoG were expressed in BL21 Codon+ cells. Cells were grown at 37°C in 1 liter of NZCYM media in the presence of the appropriate antibiotics (kanamycin and chloramphenicol) and 1% (w/v) dextrose. Protein expression was induced with 1 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) when cells reached growth measured as 0.6 OD_{600}. The expression cultures continued to grow with shaking at 37°C for 90 minutes or 48 hours at 16°C. Cells were centrifuged into pellets and stored at -20°C.

Constructs for *m*EndoG containing the unnatural ^{mX}Y amino acids were expressed in BL21 Codon+ cells by co-transforming the ^{mX}Y-*m*EndoG vector with the *pdule-Mb-haloTyrRS-C6* plasmid, which encodes *Mb*$tRNA_{CUA}$ and 3-halo-L-Tyr amino acyl-tRNA synthetase, and that was altered to include ampicillin resistance (amp^R). Cells were grown at 37°C in 1 liter of

NZCYM media with kanamycin (52 mg), chloramphenicol (36 mg) and ampicillin (36 mg). Protein expression was induced with addition of 1 mM IPTG and 1 mM ^{mX}Y (^{mI}Y or ^{mCl}Y) at 0.6 OD_{600}. The culture solution then continued to shake for 4.5 hours at 25°C before centrifugation and storing the pellet at -20°C.

Protein Purification

Cell pellets after protein expression were resuspended in ~25 mLs of 50 mM Tris buffer at pH 8.0, 50 mM NaCl, 1 mM MgCl2, 700 uL/L b-mercaptoethanol (BME) and 0.5% w/v glycerol, then incubated with 3 mL of 0.5 M EDTA and 10 µL of the detergent nonyl phenoxypolyethoxylethanol prior to sonication. The sonicated solution was then centrifuged at 15,000 rpm and the supernatant filtered through a 0.22 um filter. The filtered solution was passed through an MBP-Trap column and eluted with 20 mM maltose. The eluate was collected and incubated with tobacco etch virus (TEV) protease for 16 hours at 4°C to remove the MBP tag. The cleaved protein solution was passed through a Heparin column and the *m*EndoG construct eluted with 1 M NaCl. The eluate was concentrated to 1 mL then passed through a Sephracyl S-100 HR gel filtration column. The final sample was concentrated and stored at -80°C until used.

DNA Substrate Preparation

Holliday junctions were assembled by annealing four complementary DNA oligonucleotides[24, 43], one strand modified at the 5'-end with the cyan Cy5 and a second with the fluorescein FITC fluorophores (sequences listed below). Oligonucleotides, purchased in HPLC purified form from Integrated DNA technologies (Coralville, IA), were dissolved in nano-pure water and junction buffer (2x TBE buffer at pH 7.8 plus 0.1M $MgCl_2$, 0.2M NaCl) to a

concentration of 7μM. The DNA heated to 90°C for 20 minutes, followed by a 2-3 hour slow cooldown to room temperature. The annealed DNA was separated using a native polyacrylamide gel electrophoresis. Gel slices containing the junction DNAs were excised, crushed, and resuspended in 1X junction buffer. Duplex forms of the DNA result from mis-assembly or self-disassembly of the junction sequences.

The DNA sequences annealed as junctions are the following (^{5hm}C: 5-hydroxymethylcytosine):

Cy5-CCGTACGGGG(^{5hm}C)CAGGCCTCG

GGGTTCCTGGCCCCGTACGG

CGAGGCCTGGCCCCGGCTGC

FITC-GCAGCCGGGGCCAGGAACCC

Nuclease Assays for Activity

DNA substrates for nucleases assays (prepared as described above) were diluted to 45-50 nM in reaction buffer (20 mM Tris buffer at *p*H 7.5, 0.4% glycerol, 10 mM BME, 0.1% Triton X-100) and incubated with 0.5 μM enzyme for varying time points (the buffer was altered for *p*H and EDTA content as necessary for the respective assays). Reactions were quenched with 1.5 uM proteinase k and 0.02% sodium dodecyl sulfate (SDS). The quenched reaction mixtures were separated by native or denaturing polyacrylamide gel electrophoresis (PAGE) and imaged using a Typhoon FLA 9500 gel imager (GE Healthcare). The PAGE images were scanned and quantified using the NIH ImageJ program[38].

Kinetics of the enzyme catalyzed reactions were treated as pseudo-first order reactions and fitted as single-exponential functions. The resulting apparent rate constants (k_{APP}) were measured at least twice for each condition through a minimum of two, typically three independent replicate experiments.

Fluorescence Polarization Assays for Affinity

The affinity of various *m*EndoG constructs for the Cy5-labeled DNA junctions were assayed by determining their dissociation constants (K_Ds) from fluorescence polarization (FP) measurements using a PerkinElmer 1420 Multilabel counter plate reader and FP curves analyzed using the program KaleidaGraph. Assays were performed in solutions containing 20 nM FITC-labeled DNA plus enzyme concentrations ranging from 0.1 to 10 μM in 50 mM TRIS buffer at *p*H 8.0, plus 50 mM NaCl and 1 mM $MgCl_2$. Under these solution conditions, the enzymes do not show significant cleavage of the DNA (WT-*m*EndoG cleaves <2% of the DNA after 2 hours incubation).

Molecular Models of *m*EndoG Constructs

Initial models for the various halogenated and nonhalogenated constructs of the *m*EndoG enzyme were constructing starting with the crystal structure of the WT enzyme (PDB ID 6NJU)[24] using the BUILD functions in the program PYMOL[44]. The PYMOL SCULPTING function was used to remove steric clashes but no additional higher-level modeling was performed on the structures in order to eliminate any bias in the modeling.

Models for Mg^{2+} coordination complex in WT-mEndoG: Models for the *m*EndoG metal site (Mg^{2+}-N128-E136) was constructed by placing the atoms of acetamide and acetate around the

Mg^{2+} cation at equivalent positions seen in the WT-*m*EndoG crystal structure. In addition, waters W1 to W4 in Fig. 1c were placed according to their positions in the crystal structure to construct the model with four solvent molecules (W5 was not included, since it would have been replaced by an oxygen from the DNA phosphoribose backbone). The model for the three-water complex included W1-3, while the two-water complex included only W1 and W2. Hydrogens were added to all molecules including the waters. The positions of the hydrogen atoms were geometry optimized using density functional theory (DFT) optimization, while holding the position of all heavy atoms fixed in their respective positions. ESP surface calculations were performed on these hydrogen optimized models.

Models for H-bonded network in ^{mX}Y-mEndoG: Ortho-halophenol molecules were used as minimal models for the aromatic side chains of the respective iodo- and chloro- *meta*-halotyrosine amino acids. As described above, hydrogens were added to models and their geometries optimized by DFT calculations, while holding all heavy atoms fixed. The hydrogens of the hydroxyl substituents in these models were oriented away from the halogens, since it was determined that this OH is likely not the H-bond donor of the HB*e*XB. Calculations with the OH oriented towards the halogens increased the ESP values by +1 to +8 kcal/mol of each model but did not affect the overall trends observed from the calculations. The simplest models were that of the *o*-halophenol to mimic the halogenated tyrosyl side chain with no enhancement of its X-bonding potential from an H-bond donor.

To build the H-bonded HB*e*XB network, we started with the *o*-halophenol and an acetamide molecule to mimic N103, with all atoms placed in positions as seen in the crystal structure of *m*EndoG[24], with the carbonyl oxygen is oriented towards the halogen. In this and all subsequent models, the atomic positions of the carbon equivalent to the Cα of the enzyme

backbone were held fixed. The positions of the equivalent Cβ atom of the acetamide was fixed in place, while the coordinates of the remainder of the molecule were geometry optimized by DFT calculations. The H-bonded *o*-halophenol was constructed by rotating the amide to point the NH2 group towards the halogen and, again, the acetamide component geometry optimized with the equivalent Cβ atom fixed.

Next, we added an imidazole molecule positioned to model the neutral H104···N103···^{mX}Y136 network in the crystal structure, initially with the Nδ equivalent nitrogen protonated and the Nε unprotonated. Again, the positions of atoms of the N103 and H104 mimics were geometry optimized by DFT while holding the Cβ of the acetamide and the imidazole components fixed. The remainder of the atoms were allowed to rotate around the Cα-Cβ bonds.

Quantum Mechanical (QM) Calculations of Electrostatic Potentials (ESPs)

Minimum molecular models were constructed from small molecules that mimic the catalytic components of the WT-mEndoG and the ^{mX}Y-mEndoG constructs, as described in detail below. Models did not include any atoms from the docked DNA substrate, since these positions are uncertain. Electrostatic potentials for each model system were calculated at the Møller-Plesset second-order (MP2) level of theory using the program GAUSSIAN[45]. Calculations on non-iodine models were performed with the 6-311+g(d) basis set, with an implicit solvent model having a dielectric constant $D = 2$, shown to be appropriate for electrostatic type interactions in active sites of proteins, including those that bind nucleic acids[46-49]. Calculations on the iodinated models applied an EMSL basis set exchange[42].

3.7. Funding

This work was supported by the following grants from the National Science Foundation of the United States of America: CHE-220316 (to PSH and AKR) and MCB-2124202 (to PSH).

REFERENCES

(1) L. Wang, J. Xie, P. G. Schultz, Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225-249 (2006).

(2) J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, Enzymatic Incorporation of a New Base Pair into DNA and Rna Extends the Genetic Alphabet. *Nature* **343**, 33-37 (1990).

(3) F. Arnold, Advances in directed protein evolution. *Abstr. Pap. Am. Chem. S.* **257** (2019).

(4) D. Mondal, H. M. Snodgrass, C. A. Gomez, J. C. Lewis, Non-Native Site-Selective Enzyme Catalysis. *Chem. Rev.*, **123**, 10381-10431 (2023).

(5) N. Ahmed, N. Ahmed, D. A. Bilodeau, J. P. Pezacki, An unnatural enzyme with endonuclease activity towards small non-coding RNAs (vol 14, 3777, 2023). *Nat. Commun.* **14** (2023).

(6) R. L. Sutar, S. M. Huber, Catalysis of Organic Reactions through Halogen Bonding. *ACS Catal.* **9**, 9622-9639 (2019).

(7) N. Tarannam, M. H. H. Voelkel, S. M. Huber, S. Kozuch, Chalcogen vs Halogen Bonding Catalysis in a Water-Bridge Cocatalyzed Nitro-Michael Reaction. *J. Org. Chem.* **87**, 1661-1668 (2022).

(8) G. Renno *et al.*, Pnictogen-Bonding Enzymes. *Angew. Chem. Int. Edit.* **63** (2024).

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

(10) G. Cavallo *et al.*, The Halogen Bond. *Chem. Rev.* **116**, 2478-2601 (2016).

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

(12) P. Metrangolo, Resnati, G. , Chemistry: Halogen versus hydrogen. *Science* **321**, 918-919 (2008).

(13) R. K. Rowe, P. S. Ho, Relationships between hydrogen bonds and halogen bonds in biological systems. *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater.* **73**, 255-264 (2017).

(14) P. Auffinger, F. A. Hays, E. Westhof, P. S. Ho, Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci., USA* **101**, 16789-16794 (2004).

(15) L. A. Hardegger *et al.*, Systematic Investigation of Halogen Bonding in Protein-Ligand Interactions. *Angew. Chem. Int. Edit.* **50**, 314-318 (2011).

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

(17) P. J. Costa, R. Nunes, D. Vila-Viçosa, Halogen bonding in halocarbon-protein complexes and computational tools for rational drug design. *Expert Opin. Drug Dis.* **14**, 805-820 (2019).

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

(19) M. Carter *et al.*, Enthalpy-Entropy Compensation in Biomolecular Halogen Bonds Measured in DNA Junctions. *Biochemistry* **52**, 4891-4903 (2013).

(20) E. Danelius *et al.*, Halogen Bonding: A Powerful Tool for Modulation of Peptide Conformation. *Biochemistry* **56**, 3265-3272 (2017).

(21) D. Vila-Viçosa, R. Nunes, P. J. Costa, Folding of cyclic peptides stabilized by halogen bonds. *Eur. Biophys. J. Biophy.* **48**, S197-S197 (2019).

(22) A. C. Carlsson *et al.*, Increasing Enzyme Stability and Activity through Hydrogen Bond-Enhanced Halogen Bonds. *Biochemistry* **57**, 4135-4147 (2018).

(23) A. B. Robertson, J. Robertson, M. Fusser, A. Klungland, Endonuclease G preferentially cleaves 5-hydroxymethylcytosine-modified DNA creating a substrate for recombination. *Nucleic Acids Res.* **42**, 13280-13293 (2014).

(24) C. M. Vander Zanden, R. S. Czarny, E. N. Ho, A. B. Robertson, P. S. Ho, Structural adaptation of vertebrate endonuclease G for 5-hydroxymethylcytosine recognition and function. *Nucleic Acids Res.* **48**, 3962-3974 (2020).

(25) S. L. Wu *et al.*, Mutagenesis identifies the critical amino acid residues of human endonuclease G involved in catalysis, magnesium coordination, and substrate specificity. *J. Biomed. Sci.* **16**, 6 (2009).

(26) J. S. Murray, L. Macaveiu, P. Politzer, Factors affecting the strengths of sigma-hole electrostatic potentials. *J. Comput. Sci.-Neth.* **5**, 590-596 (2014).

(27) J. L. Lin, C. C. Wu, W. Z. Yang, H. S. Yuan, Crystal structure of endonuclease G in complex with DNA reveals how it nonspecifically degrades DNA as a homodimer. *Nucleic Acids Res.* **44**, 10480-10490 (2016).

(28) L. P. Kozlowski, IPC 2.0: prediction of isoelectric point and dissociation constants. *Nucleic Acids Res.* **49**, W285-W292 (2021).

(29) N. G. Faleev, O. V. Axenova, T. V. Demidkina, R. S. Phillips, The role of acidic dissociation of substrate's phenol group in the mechanism of tyrosine phenol-lyase. *Biochim. Biophys. Acta* **1647**, 260-265 (2003).

(30) T. N. Das, Redox Chemistry of 3-Iodotyrosine in Aqueous Medium. *J. Phys. Chem. A* **102**, 426-433 (1998).

(31) D. R. Duckett, A. I. Murchie, D. M. Lilley, The role of metal ions in the conformation of the four-way DNA junction. *EMBO J.* **9**, 583-590 (1990).

(32) Y. Y. Sun, J. Y. Yang, G. Z. Xu, K. Y. Cheng, Biochemical and Structural Study of RuvC and YqgF from Deinococcus radiodurans. *mBio* **13:**e01834-22 (2022).

(33) A. M. S. Riel *et al.*, Hydrogen Bond Enhanced Halogen Bonds: A Synergistic Interaction in Chemistry and Biochemistry. *Acc. Chem. Res.* **52**, 2870-2880 (2019).

(34) I. Benito, R. M. Gomila, A. Frontera, On the energetic stability of halogen bonds involving metals: implications in crystal engineering. *CrystEngComm.* **24**, 4440-4446 (2022).

(35) J. H. Eastberg, J. Eklund, R. Monnat, B. L. Stoddard, Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism. *Biochemistry* **46**, 7215-7225 (2007).

(36) L. J. Henderson, Concerning the relationship between the strength of acids and their capacity to preserve neutrality. *Am. J. Physiol.* **21**, 173-179 (1908).

(37) K. A. Hasselbalch, Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebundenen Kohlensäure desselben, und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl. *Biochem. Z.* **78**, 112-144 (1917).

(38) C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671-675 (2012).

(39) L. Wang, A. Brock, B. Herberich, P. G. Schultz, Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500 (2001).

(40) J. M. Xie *et al.*, The site-specific incorporation of *p*-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **22**, 1297-1301 (2004).

(41) H. S. Jang *et al.*, Efficient Site-Specific Prokaryotic and Eukaryotic Incorporation of Halotyrosine Amino Acids into Proteins. *ACS Chem. Biol.* **15**, 562-574 (2020).

(42) M. R. Scholfield *et al.*, Structure-Energy Relationships of Halogen Bonds in Proteins. *Biochemistry* **56**, 2794-2802 (2017).

(43) C. M. Vander Zanden, R. K. Rowe, A. J. Broad, A. B. Robertson, P. S. Ho, Effect of Hydroxymethylcytosine on the Structure and Stability of Holliday Junctions. *Biochemistry* **55**, 5781-5789 (2016).

(44) G. Janson, C. Zhang, M. G. Prado, A. Paiardini, PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. *Bioinform. (Oxford, England)* **33**, 444-446 (2017).

(45) G. W. T. M. J. Frisch, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, (2009) Gaussian 09. (Gaussian, Inc.).

(46) B. Honig, A. Nicholls, Classical Electrostatics in Biology and Chemistry. *Science* **268**, 1144-1149 (1995).

(47) R. Rohs *et al.*, The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248-1253 (2009).

(48) Y. C. Li *et al.*, The impact of interior dielectric constant and entropic change on HIV-1 complex binding free energy prediction. *Struct. Dynam.-US* **5** (2018).

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

CHAPTER 4

THE EFFECTS OF PROTEIN ENVIRONMENTS ON THE GEOMETRIES AND ENERGIES OF HYDROGEN BOND ENHANCED HALOGEN BONDS[3]

4.1. Summary

Halogen bonds (XBs) are significant interactions in protein-inhibitor complexes. The XB potentials of halogens are greatly influenced by hydrogen bonds (HBs) through synergistic hydrogen bond enhanced halogen bond (HB*e*XB) interactions. In this survey of the Protein Data Bank, we find that HB*e*XBs constitute ~45% of all XBs, although their overall geometries are similar to those of XBs alone. However, shorter HBs are seen to correlate with more linear XBs. Uncharged HB donors increase the electrostatic surface potentials (ESPs) of HB*e*XB models by 2 -10 kcal/mol, while ESPs with charged donors increase by 60-70 kcal/mol. This charge effect on ESPs translates to an increase in interaction energy of 5-6 kcal/mol. Finally, IC_{50}s for a limited set of halogenated inhibitors appear to be associated with HB angles of approach. Thus, HBs significantly influence the geometries and energies of HB*e*XBs, providing insights useful for the design of more effective therapeutic inhibitors.

[3] The work in this chapter was written, formatted, and submitted as a research article for the *Journal of Medicinal Chemistry*. I contributed by analysis of the dataset collected from the survey, quantum energy calculations, and co-writing. I am responsible for conceptualization of the pairwise binding energy calculations.

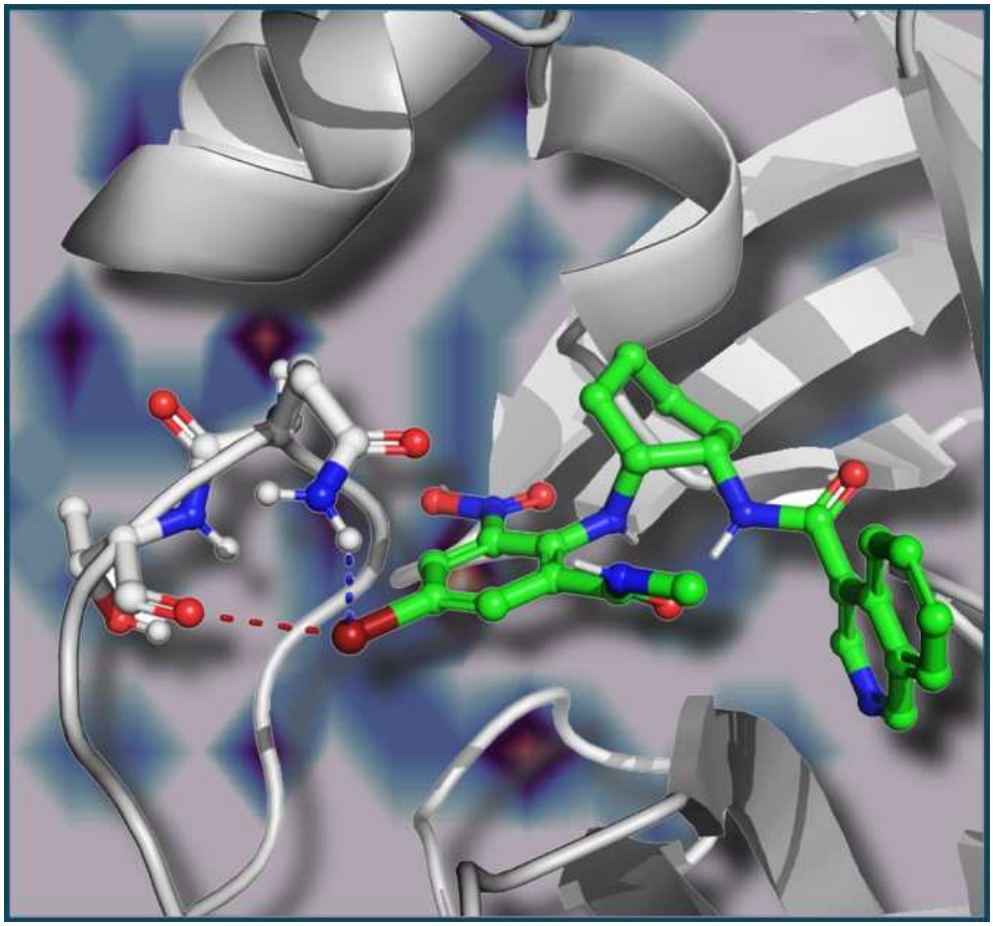


Figure 4.0: Graphical abstract of Chapter 4.

4.2. Introduction

Nonclassical noncovalent (*nc*NC) interactions are now well established as important tools for molecular design and engineering across multiple areas of chemistry[1]. Among this class of molecular interactions, the halogen bond (XB) stands out for its broad applications in the biological sciences[2], particularly in medicinal chemistry[3, 4] and, more recently, in designing peptides and proteins with enhanced physical properties[5, 6] and new catalytic mechanisms[7]. The molecular contexts for such biological XBs are complex in that this type of *nc*NC interaction resides in an environment that is replete with many other, more classical electrostatic interactions, not the least of which is the hydrogen bond (HB). The relationship between HBs and XBs in the biological context is also complex[8], being competitive[9] or orthogonal[10] when both are vying for the same acceptor atom. We recently proposed that an HB can also serve to enhance an XB, resulting in what we call the hydrogen bond enhanced halogen bond, or HB*e*XB[5, 11]. The HB*e*XB concept had previously been invoked to help rationalize the ability of halogenated tyrosine (^{X}Tyr, an unnatural amino acid) to affect the physical properties of proteins [5] and to function as catalysts in DNA processing enzymes[7]. In this study, we surveyed the Structural Bioinformatics Protein Data Bank (the PDB for short)[12] for potential HB*e*XBs to determine their prevalence and geometries (Figure 1a) of interaction, and apply quantum chemical analyses of models for average and idealized HB*e*XBs to better understand how the protein environment affects their energies of interaction in protein-inhibitor complexes.

Halogens are rarely seen as covalent substituents in natural biological macromolecules but, for proteins, can result from oxidative halogenation by peroxidases as inflammatory responses in respiratory diseases[13] such as asthma. Bromination of DNA bases by eosinophil peroxidase may serve as markers for DNA damage[14], while halogenated cytosines may induce spontaneous

transition mutations[15]. However, the most common occurrence of halogens in biology are as substituents in inhibitors and drugs that bind protein targets[2, 16, 17]. Indeed, ~20% of all therapeutics and drugs, including ~50% of the highest grossing drugs, are halogenated, with the heaviest of the halogens (I, Br and particularly Cl) persisting through the drug discovery pipeline[18]. Recent studies have shown that nearly one-third of drugs approved by the FDA in 2024 are halogenated (primarily F and Cl)[19].

The classical reasons for the incorporation and persistence of halogens in drug discovery pipelines include the elements' lipophilicity and their electron withdrawing abilities, which affects the reactivities and properties of other substituents, such as those involved in HBs. For example, Poznański, *et al.*[20] showed from a survey of PDB structures that distances of HB donors from inhibitors to their acceptor atoms in the target protein kinases and acyltransferases were shorter when the ligands are halogenated *versus* those that are unhalogenated. This effect can be attributed to the increase in the acidity of the HB donor from the electron withdrawing inductive property of the halogen, with the most electronegative fluorine showing the strongest effects on most types of HBs, while nonfluorine halogens show the most significant shortening only of NH···O interactions.

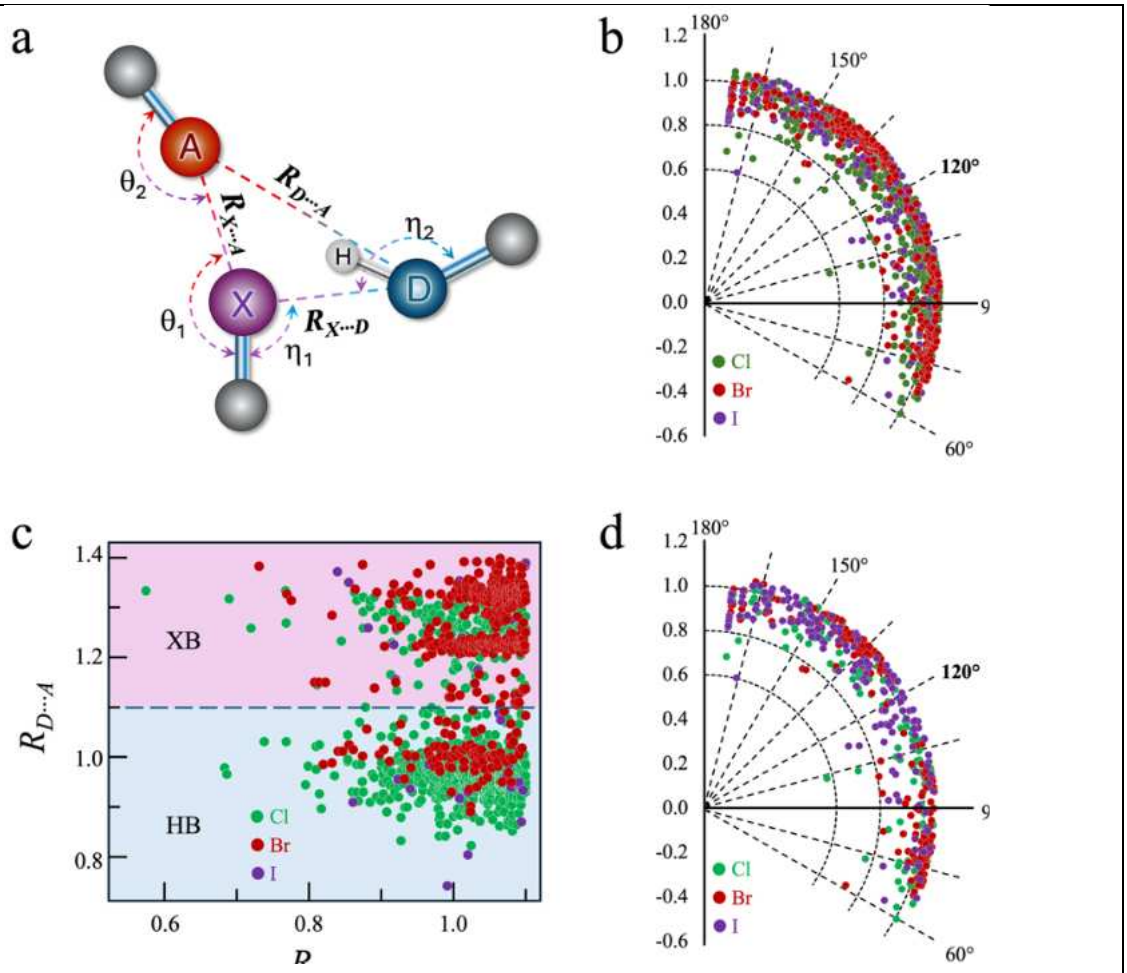


Figure 1: PDB survey of biological halogen bonds (XBs) and hydrogen bond enhanced halogen bonds (HB*e*XBs). **a.** Geometric parameters for XBs and HBeXBs. Distances are labeled from the halogen (X) to the XB acceptor (A, $R_{X\cdots A}$), from X to the HB donor (D, $R_{X\cdots D}$) and from the HB donor to the XB acceptor ($R_{D\cdots A}$). The approach of A to the σ-hole of X is the θ_1-angle, approach of X to the A electrons is the θ_2-angle, approach of the HB donor to X is the η_1-angle and D to A is the η_2-angle. **b.** Polar radial plot representing the HB*e*XB θ_1 angles and $R_{X\cdots A}$ distances (calculated as the respective $f\sum R_{vdW}$) for X= Cl (green), Br (red) or I (violet). **c.** HBeXB populations plotted as $R_{D\cdots A}$ versus $R_{X\cdots A}$ distances. Populations were distinguished as primarily HB or primarily XB interactions by the $R_{D\cdots A}$ distance, with demarcation at $R_{D\cdots A} \approx 1.1$ of $f\sum R_{vdW}$. **d.** Polar radial plot representing the θ_1 angle and $R_{X\cdots A}$ distance HB*e*XB population after filtering the data to remove from **b** interactions with $R_{D\cdots A} < 1.1$ of $f\sum R_{vdW}$.

The significance of halogens in biological systems has become greatly expanded with the discovery that XBs are major components in biomolecular recognition, particularly in inhibitor and drug recognition and binding[2]. However, XBs in these protein-ligand systems were

primarily identified retroactively. Among the first intentional application of XBs in therapeutics was a report in 2009 on the design of factor Xa inhibitors[3]. XBs are now seen as significant contributors to occurrence and the persistence of larger halogens through the drug discovery pipeline[17].

XBs are members of a larger class of what are now known as s-hole *nc*NC interactions[1, 21]. The s-hole concept posits that an atom participating in a covalent bond will have its valence electron subsumed by the s-molecular orbit (s-MO), resulting in an electropositive crown (the s-hole) diametrically opposed to the orbital[22]. For XBs, this polarization leaves an electronegative waist perpendicular to the s-MO that can serve as an HB acceptor. The size and electropositive charge of the s-hole increases as the atomic radius and associated polarizability of the halogen increases (with Cl < Br < I). We should note that various levels of computational theory support the inclusion of charge-transfer (its initial conception[23]), and dispersion [24], as well as electrostatics[25] in XB interactions. However, for convenience, we will focus primarily on the electrostatic s-hole model for the remainder of this discussion.

XBs are analogous to the more ubiquitous HB[21, 26]. As such, the relationship between XBs and HBs in protein environments have been of great interest, with surveys of structures in the PDB providing important insights into their relationships. These two molecular interactions share common electronegative acceptors, meaning that XBs could substitute for or compete against HBs[8, 9]. When sharing the same acceptor atoms, such as the carbonyl oxygen of a peptide bond, XBs and HBs have been seen from such surveys to be geometrically and thermodynamically orthogonal and independent of each other[10].

The supposition of the HB*e*XB concept is that an HB to the electronegative waist of a halogen further polarizes its electron distribution, thereby enhancing the s-hole

and consequently its potential to serve as an XB donor[11]. We had first recognized the enhancing effects of the HB*e*XB when characterizing the thermal stability and function of T4 lysozyme (T4L) construct engineered with an unnatural ^{X}Tyr. The chlorinated ^{Cl}Tyr-T4L was found to have an ~3 kcal/mol higher enthalpic stability and 15% greater enzymatic activity at elevated temperatures compared to the wildtype (WT) enzyme[5]. All prior studies on chlorine as a XB donor in a biological context indicated that the effects of this smaller halogen should be minimal (no more than ~1 kcal/mol in interaction energy) [27]. For example, the ability of a chlorine XB (Cl-XB) to stabilize a four-stranded DNA junction was found to be 0 – 0.8 kcal/mol[28], while Br- and I-XBs provided as much as 5 kcal/mol thermal stabilization[29]. This apparently anomalously large Cl-XB effect on ^{Cl}Tyr-T4L stability was rationalized through a proposal that the adjacent hydroxyl group of the Tyr residue could form an intramolecular HB to the halogen and, in this way, enhance its s-hole. Calculations on the electrostatic surface potential (ESP) in support of this proposal indicated that the XB potential of the Cl substituent could be increased by 3 kcal/mol[5]. This effect of an internal HB enhancing the XB potential of a halogen substituent is supported by quantum mechanical (QM) calculations applied to various molecular models in which HB donors are placed ortho to a halogen XB donor[30].

More recently, we asked the question of whether XBs can serve as a biochemical catalyst in enzymes. This idea derived from reports that this *nc*NC interaction can catalyze several classical chemical processes, including halogen abstraction and Diels-Alder reactions[31]. In our biochemical study, we replaced a glutamic acid (Glu) amino acid that chelates the essential catalytic magnesium dication (Mg^{2+}) in a mammalian endonuclease with an ^{X}Tyr residue[7]. This halo-enzyme construct was found to actively cleave a DNA substrate. The observation that the iodinated enzyme was approximately twice as active as the chlorinated analog and less sensitive

to deactivation by the metal chelator ethylenediamine tetra-acetic acid (EDTA) than wildtype implicated the XB as replacing the catalytic function of the essential Mg^{2+}. The XB potential shown to be increased by an asparagine (Asn103) that served as an HB bridge to communicate the protonation and charge state of a histidine (His104) to the ^{X}Tyr. Thus, this catalytic XB enzyme (or *cX*-Zyme) is dependent on an HB*e*XB as an intermediary to transmit the properties of the protein environment to the XB catalyst.

The objective of the current study is to identify distinctive properties of HB donors to halogen substituents that may affect the geometries and potential energies of XBs by surveying the PDB for HB*e*XB interactions. A prior survey of the Cambridge Structural Database (CSD) had shown HB*e*XBs are prevalent in small molecule complexes and that this type of *nc*NC can be distinguished from purely HB interactions[11]. The protein environment is dominated by HB donors and acceptors, making it much more complicated than that of small molecule complexes. A study by Han, *et al.* [32], had shown that HBs to fluorine substituents of ligands can induce an electropositive s-hole, which in analogous fashion to the HB*e*XB, generate a hydrogen bond-induced fluorine bonding (HBiFB) interaction. Their survey of the PDB showed that ~11% of fluorine atoms of ligands with "XB-like" geometries have the potential to form an HBiFB and that such interactions could provide up to 1 kcal/mol of interaction energy with its protein target. Our current study focuses on the heavier halogens (Cl, Br and I), with the goal of determining whether and how this synergistic interaction potentially affects the geometries and energies of XBs in the protein environment as initial steps towards applying HB*e*XBs for the molecular design of inhibitors against therapeutic targets and of biomolecular systems with novel functions.

4.3. Materials and Methods

Construction of the biological halogen bond (XB) data set: The dataset of XBs was constructed by surveying the 2023 release of the RCSB-PDB[12] using a locally written program that opens PDB files, searches for the occurrence of halogen atom codifiers (Cl, Br or I) in the ATOMS fields, identifies the atom(s) that the halogen is covalently attached to (to ensure that it is a halogen substituent), identifies all electron-rich atoms within 140% of the sum of their respective van der Waals radii ($\sum R_{vdW}$) from the halogen as potential XB acceptors and identifies all oxygen, nitrogen, or sulfur atoms within 140% of the sum of their respective van der Waals radii ($\sum R_{vdW}$) from the halogen as potential HB donors. A filter is applied to retain only crystal structures at resolutions reported to be at or better than 3Å. The dataset catalogs the halogen type as XB donors, potential XB acceptors (atom name, atom number, residue name, residue number), XB donor to acceptor distances ($R_{(X\cdots A)}$ in Å and as the fraction of the sum of the respective van der Waals radii, $f\sum R_{vdW}$) and angles (q_1 and q_2, see Figure 1a) and potential HB donors (atom name, atom number, residue name, residue number), HB donor to acceptor distances ($R_{(X\cdots D)}$) and angles (h_1 and h_2, see Figure 1a). The data was subsequently filtered to define interactions as XBs using the standard criteria of $R_{(X\cdots A)} \leq 1.05$ of $f\sum R_{vdW}$. Subsequent analyses of the dataset further restrict XBs to $q_1 \geq 120°$. To minimize overrepresentation of any specific type of interaction, the dataset was manually filtered for redundant representations of identical or near identical XB interactions. In cases where a ligand is present multiple times within a PDB structure (*e.g.*, as symmetry related subunits), the geometries of interactions are averaged and only the averaged parameters are retained. In cases where there are multiple entries of identical structures at different resolutions, only the parameters from the highest resolution structure are retained.

Construction of the hydrogen bond enhanced halogen bond (HBeXB) dataset: Construction of the HB*e*XBs dataset started with the XB data and identifying all interactions with potential HB donors within $R_{(X\cdots D)} \leq 1.05$ of $f\sum R_{vdW}$ are included. Analysis of the $R_{(X\cdots A)}$ *versus* HB donor to acceptor distances ($R_{(D\cdots A)}$) showed that there are two distinct populations of potential HB*e*XBs (Figure 1c). Prior analysis of structures from a survey of HB*e*XBs in the Cambridge Structural Database (CSD) [11] had found a similar pattern, with the distribution at lower $R_{(D\cdots A)}$ values being more properly classified as HB interactions with a halogen incidentally nearby but not a significant contributor to the interaction. We interpreted the $R_{(X\cdots A)}$ *versus* $R_{(D\cdots A)}$ results similarly and set a threshold of $R_{(D\cdots A)} \geq 1.1\, f\sum R_{vdW}$ as the criterion for potential HB*e*XBs (Figure 1d). Finally, we needed to ensure that the potential HB donor for the HB*e*XB has its hydrogen aligned towards the halogen. Hydrogens are not typically observed in crystal structures in the PDB but the h_2-angle can serve as a proxy for this H–X alignment. Although HBs are not as directional as XBs, we considered a limit of $h_2 \geq 90°$ to be reasonable.

Construction of the XB-Only dataset: The XB-Only dataset was constructed from the overall XB data after interactions that fit the criteria for HB*e*XBs were removed.

Distribution analyses of XB and HBeXB geometries: The interactions in each of the individual datasets were grouped into bins of $0.2(f\sum R_{vdW})$ for all distance measurements and bins of 5° for all angle measurements to construct distribution maps. MATPLOTLIB[33] was used to construct heatmaps for interactions clustered as geometric pairs using the binned data.

Quantum mechanical (QM) electrostatic surface potential (ESP) calculations for hydrogen bonded halogen substituents: Models for ESP calculations were constructed starting with halogenated benzene as a model for the halogenated ligand. Amino acid side chains of HB

donors identified from the PDB survey were modeled as methanethiol for Cys, methanol for Ser and Thr, phenol for Tyr, methyl amide for Gln, methylimidazole for His and methylguanidine for Arg. Each HB complex was constructed using the average geometries for each type of interaction observed in the PDB survey.

The ESP for each complex was calculated using GAUSSIAN09 program[34] and applying Møller-Plesset second-order (MP2) theory. Calculations for chlorine and bromine interactions were conducted with the 6-311+g(d) basis set and with a dielectric constant of 2 (scrf=cyclohexane), the accepted solvent model for the environment of a ligand binding pocket in proteins[35]. Iodine interactions applied the aug-cc–PVTZ-PP basis set from the EMSL Basis Set Exchange[36]. The ESP value of the s-hole was determined at the surface point from the halogen at $q_1 = 180°$.

QM calculations of charged HB donors on energies for X···O (carbonyl) interactions. Trimolecular models to probe the influence of charged HB donors on XBs centered on effects that His (unprotonated and uncharged *versus* protonated and charged) as an HB donor has on Br as an XB donor to the carbonyl oxygen (the most frequent XB acceptor found from the PDB survey). The His sidechain in these complexes is modeled by 4-methylimidazole (labeled "D" as the HB donor), the XB donor by a bromobenzene (labeled "X") and the carbonyl oxygen XB acceptor by formaldehyde (labeled "A").

The conformations for these model complexes started with "ideal" geometries, with the carbonyl oxygen placed at 0.95 of $f\sum R_{vdW}$ from and aligned at q_1 and $q_2 = 180°$ from the halogen of bromobenzene. The initial geometries of the HB donors relative to the bromobenzene were similarly set at ideal distances (0.95 of $f\sum R_{vdW}$) and angles ($h_1 = 180°$ and $h_2 = 129°$). These starting models were then geometry optimized using GAUSSIAN09 program[34] and applying

MP2 theory, with the 6-311+g(d) basis set and a dielectric constant set $D = 2$. To ensure that the components do not deviate significantly from the expected geometries for HBs and XBs, the specific carbon atoms were held fixed (the methyl of the 4-methylimidazole, C4 carbon of the benzene ring of the bromobenzene and the carbon of the formaldehyde models) during the optimization calculations. Optimizations were performed separately with the unprotonated and uncharged His (labeled optHis) and with the protonated and positively charged His^+ (labeled optHis+). In addition, a model for the protonated His+ in the unprotonated conformation of the HB*e*XB complex was constructed by adding a proton to the optHis model (unoptHis+) and a model for the unprotonated His in the protonated conformation of the HB*e*XB complex was constructed by removing the proton to the optHis+ model (unoptHis).

A thermodynamic cycle for the changes in the interaction energies for optHis+ to unoptHis (deprotonation) to optHis (conformational rearrangement) to unoptHis+ (protonation) to optHis (conformational rearrangement) was constructed by calculating the MP2 energies of each model, with basis set superposition error corrected using counterpoise of three[37] and subtracting the energies of the individual components from each model (Eq. 1). These calculations were replicated using density functional theory (DFT) wB97XD method with cc-pVTZ basis set and corrected using counterpoise=3.

$$E_{int} = E_{XDA} - E_X - E_D - E_A \qquad \text{Eq. 1}$$

where E_{int} is the net interaction energy of interest, E_{XDA} is the QM calculated energy of the three-component complex (see Fig. 1a for the definitions X, D and A), E_X is the QM energy of the XB donor component (X), E_D is the QM energy of the HB donor component (D) and E_A is the QM energy of the XB acceptor component (A).

To estimate the impact of the positively charged HB to the HB*e*XB for individual interactions, the interaction energy (E_{int}) was calculated by treating two components as a single paired assembly (e.g., X and D as a preassembled X-D HBed complex), leaving the third (in this example, the component A) as a monomer component (Eq. 2).

$$E_{int} = E_{ij} - E_i - E_j \quad \text{Eq. 2}$$

where i = D and A, and j = X,i = A and X, and j = D or i = X and D, and j = A

These calculations were replicated using DFT wB97XD method with cc-pVTZ basis set.

4.4. Results

4.4.1. Datasets from the survey of the PDB

In this study, we surveyed the 2023 release of the RCSB-PDB[12] to construct a list of biomolecular structures that included covalent halogen molecular substituents (Cl, Br or I) as potential XB donors that were positioned close (initially within 140% of the sum of their respective van der Waals radii, $\sum R_{vdW}$) to an electronegative atom (O, N or S) as XB acceptors. This master list of over 19,000 halogen structures ignored F as potential XB donors, since fluorine very seldom is found in this class of *nc*NC interactions in biological systems. The dataset was further filtered by applying both the distance (as a fractional ratio to the sum of the respective van der Waals radii, or $f\sum R_{vdW} = 1.05$) and angle of acceptor to halogen approach ($q_1 \geq 120°$), criteria that are recognized as characteristic of XBs (Figure 1b). Finally, redundant entries were manually identified and removed to minimize biases resulting from multiple identical or near identical structures in the PDB.

The final dataset (Supplemental Table 1) that was used for all subsequent analyses included 821 XB interactions and is nearly entirely structures with halogenated ligands interacting with

proteins. Thus, the remainder of the discussions will focus only on ligand-protein interactions. As consistent with all previous surveys of XBs, the set is dominated by chlorine as the donor and carbonyl oxygens (labeled "=O") as the acceptor. The average $f\sum R_{vdW}$ distances and q_1-angles trend towards shorter and more linearly aligned XB interactions going from Cl to Br to I, as predicted by the s-hole theory. However, we note that the distributions of these two geometric parameters are skewed as a result of the distance and angle cutoffs. Consequently, the standard deviations reported here should be interpreted only as an indication of the size of the spread in the data and not a true measure of population statistics.

In addition, we must recognize that there are limitations associated with geometries derived from crystal structures. First, although the crystal structures in the dataset would be considered to be at or near "atomic resolution" [38], the uncertainty in the atomic coordinates could range from 0.1 Å for the highest resolution structures to upwards of 0.3 Å (or near 10% of the van der Waals radius of chlorine and oxygen) for the lowest resolution structures [39]. In addition to the inherent uncertainties associated with resolution, we further note that the models reported in the PDB were constructed with additional constraints outside of the X-ray data, including restraints to relative atomic positions to libraries of geometric parameters, particularly of covalently bound atoms. Finally, we note that any structure that has been solved and refined using energy forcefields as restraints would necessarily be biased against XB-type interactions, given the limitations of classical forcefields to model such interactions. The standard deviations listed in these and subsequent tables do not reflect these experimental coordinate errors.

The distribution of atom types that serve as XB acceptors follow generally their occurrences in proteins, with the exception of the higher representation than expected occurrence of sulfur in the Br-XB dataset. The approach of the halogen towards the XB acceptor

(q_2-angle) indicates that these interactions are towards the nonbonding electrons of the oxygen and sulfur atoms. However, the halogens approach the nitrogen acceptors at a sharper q_2-angle, suggesting these XBs are pointed perpendicular and primarily to the π-electron system of peptide bonds in proteins. This overall XB dataset was then segregated into a group consisting of interactions in which an oxygen, nitrogen or sulfur potential HB donor is localized within 110% of the $f\sum R_{vdW}$ of the halogen substituent (the starting HB*e*XB dataset) and a separate group with interactions that lack these potential HBs (the XB-Only set).

We repeated the $R_{D\cdots A}$ versus $R_{X\cdots A}$ analysis from the CSD survey[11] on the PDB HB*e*XB data and found again that there are two distinct populations, the first with $R_{D\cdots A} \leq$ 1.1 of $f\sum R_{vdW}$ and the second with $R_{D\cdots A} \geq 1.1$ of $f\sum R_{vdW}$ (Figure 1c). As with the CSD analysis, this first group was interpreted as being primarily HB interactions, with the halogen being not the primary but perhaps an ancillary interaction partner. This set of interactions were removed from further analysis in all datasets. The remaining interactions we interpret as being primarily driven by the XB and were used for further analysis. We next reassigned to the XB-Only group any interaction where the angle of approach of the HB donor to the halogen (h_1-angle, Figure 1d) is outside the range of 60° – 120°, thereby removing HBs that are not pointed towards the electronegative waist of the halogen. The resulting filtered data constitutes the final HB*e*XB dataset for further analyses.

4.4.2. Analyses of the HBeXB dataset

The HB*e*XB dataset represents ~45% of the total XBs in this study, reflecting the prevalence and significance of this class of interactions in protein-ligand interactions. The geometric properties and distribution of XB acceptor types in both the XB-Only

and HBeXB data mirror those of the overall XB set (Supplemental Tables S2 and S3). Thus, unlike HBeXB in small molecules[11], the HB apparently does not generally affect the geometries of the XBs in protein environments. This lack of effect is likely due to the XB being only one of several noncovalent interactions, including HBs, π-π stacking and hydrophobic interactions that define how ligands bind in protein pockets, regardless of the strength of the halogen's s-hole.

The HB*e*XB dataset was further dissected according to the halogen serving as the XB donor, the acceptor for each donor class (O, N, or S) and finally the type of HB donor to those acceptors (Supplemental Tables S4-S6). We distinguished between nitrogenous HB donors that are neutral (amides of the peptide bond and sidechains of Asn, Gln and Trp) from those that likely carry a formal positive charge in the binding pocket (side chains of Arg, His and Lys). Overall, the majority of HB donors to HB*e*XBs are –OH groups (sidechains of Ser, Thr, and Tyr), followed by the uncharged –NH, the positively charged $–NH^+$ and finally –SH (from Cys). However, we find that this distribution of HB donors is dependent on the XB halogen type, with the Cl-XBs dominating the statistics (constituting >61% of the HB*e*XBs). For Br-XBs and even more so for I-XBs, $–NH^+$ donors are as prevalent as –OH, indicating that positively charged nitrogens need to be considered when analyzing the properties of HB*e*XBs.

The angle at which the HB donor aligns with the C–X bond of the halogen is ideally h_1 = 90° (towards the most electronegative potential of the halogen substituent's waist). The actual h_1-values for all donors are indeed centered at ~100° (Figure 2a). However, the h_1-distribution for –OH and –SH donors appear to peak at ~75° and ~130° (Figure 2b). This 25°-30° deviation away from the overall data and from the ideal value can be attributed to the analysis of h_1 being to the heavy atom of the donors, since hydrogens are not observed in most crystal structures. Thus, although the hydrogen may indeed be pointed to the electronegative waist of the halogen,

the h_1 analysis in this manner may not reflect that. The h_1-distribution for nitrogen HB donors (Figure 2c) are centered at ~100° but there are two wings both below and above this value that are similar to those in the O/S distribution. After more detailed analyses of the actual HB donors in each population, we conclude that the amino acids contributing to the central $h_1 \approx 100°$ population are primarily sp^2-hybrid type nitrogens (Arg, Asn, Gln and His), while those with h_1 at ~75° and ~130° are primarily sp^3-hybrid type donors (Lys, Ser, Thr and Tyr).

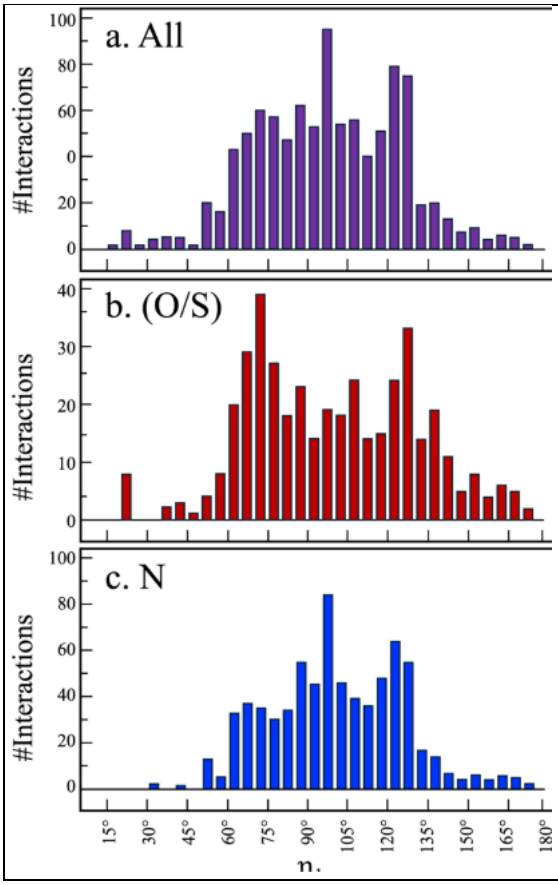


Figure 2: Distribution of HB*e*XB interactions according to their η_1-angles for HB donors. **a.** Distribution of η_1-angle for all HB donors. **b.** Distribution of η_1-angles for oxygen (Ser, Thr, and Tyr) and sulfur (Cys) HB donors. **c.** Distribution of η_1-angles for nitrogenous HB donors (from Asn, Arg, Gln, His, Lys, Trp amino acid side chains and the amino group of the peptide backbone).

Surveys of XBs typically start with an analysis of the relationship between q_1 and $R_{X\cdots A}$. In this study, these angle-distance analyses are presented as contour heat maps showing the occurrence of interactions distributed across the two primary geometric measures (Figure 3). The data presented in this manner, as opposed to a radial plot, allows us to visualize and analyze geometric trends across the entire population of interactions. For the dominant X···O interactions

in the full XB datasets (Figure 3a), the most populated clusters trend towards slightly shorter and more linear (q_1 approaching 180°) geometries as the halogen size increases (Cl < Br < I for $R_{X\cdots A}$). This trend generally continues in the HB*e*XB dataset (Figure 3b), although not as pronounced for the smaller I-HB*e*XB dataset.

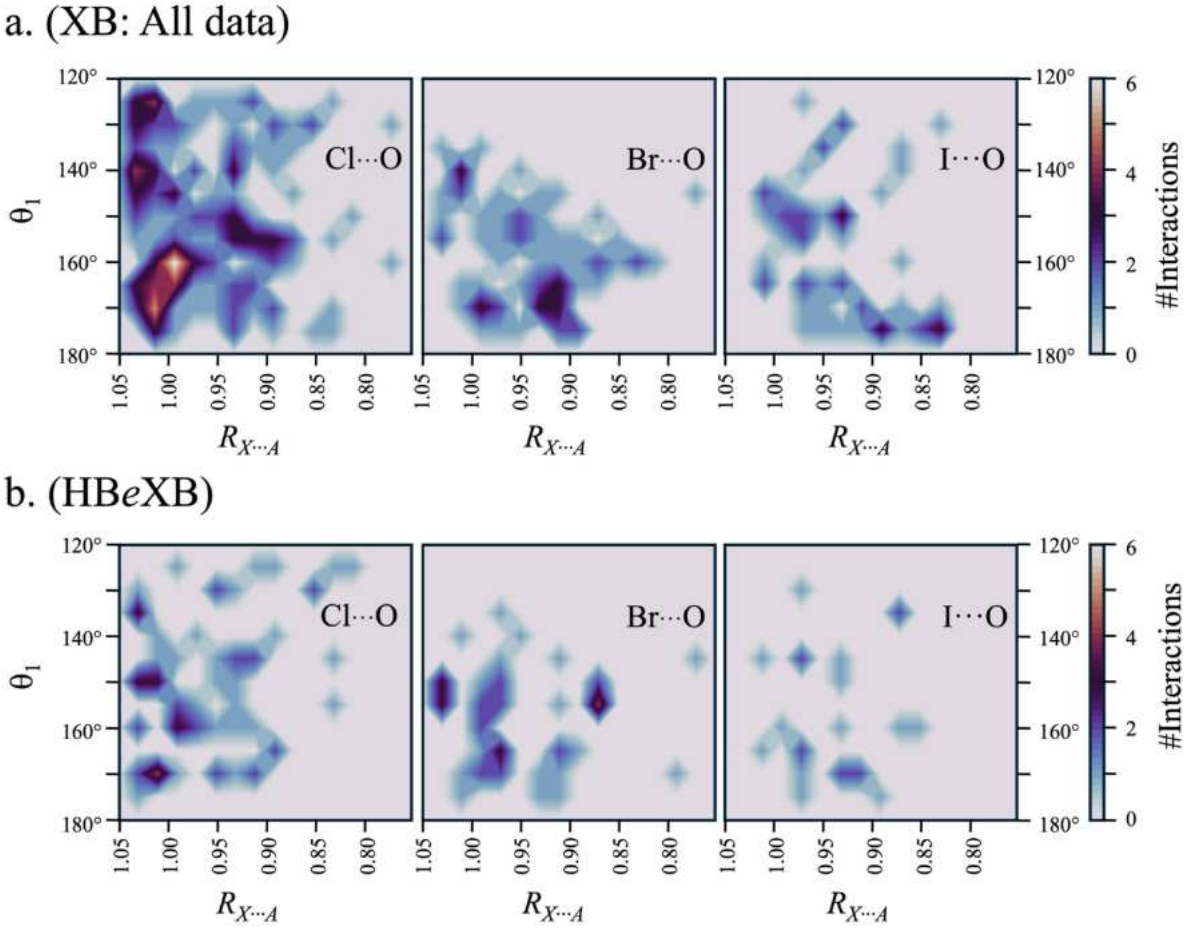


Figure 3: Contour heat maps of θ_1-angles and $R_{X\cdots A}$-distances for distributions for numbers of X···O XB (**a**) and HB*e*XB (**b**) interactions. The color scales for the numbers of interactions are shown on the right.

We next addressed the question of how the HB donor interaction affects the geometry of the XB in the HB*e*XB data. One trend for the Cl-HBeXBs that stood out was that the q_1-angle becomes more linear as the halogen to HB donor distance ($R_{X\cdots D}$) becomes shorter for the most highly populated clusters of interactions (#Interactions > 2, Figure 4a). This inverse relationship

appears also for the Br-HB*e*XBs but not as prominently. To determine whether this relationship is statistically significant, we started by plotting out the occurrence of Br-HB*e*XB interactions as a function of q_1 and observed that there were two distinct populations (colored as blue and red in the top graph Figure 4b, with *p*-values < 10^{-4} for the population means). We then replotted the interactions now according to their $R_{X\cdots D}$ values, keeping track of each interaction's initial q_1-grouping (bottom graph Figure 4b). The *p*-values for the blue versus red categories had a *p*-value < 0.01 based on the $R_{X\cdots D}$ value means, indicating that they remained statistically distinct populations. This analysis was repeated starting by grouping interactions according to $R_{X\cdots D}$ values, then replotting against q_1 (Figure 4c). Again, replotting the groupings resulted in a *p*-value < 0.01, indicating that they remained statistically distinctive populations. Thus, for both Cl- and Br-HB*e*XB interactions, shorter and presumably stronger HBs result in more linear interactions of the halogen XB donors to their respective acceptors. Unfortunately, the data for I- HB*e*XBs is too limited for a similar analysis.

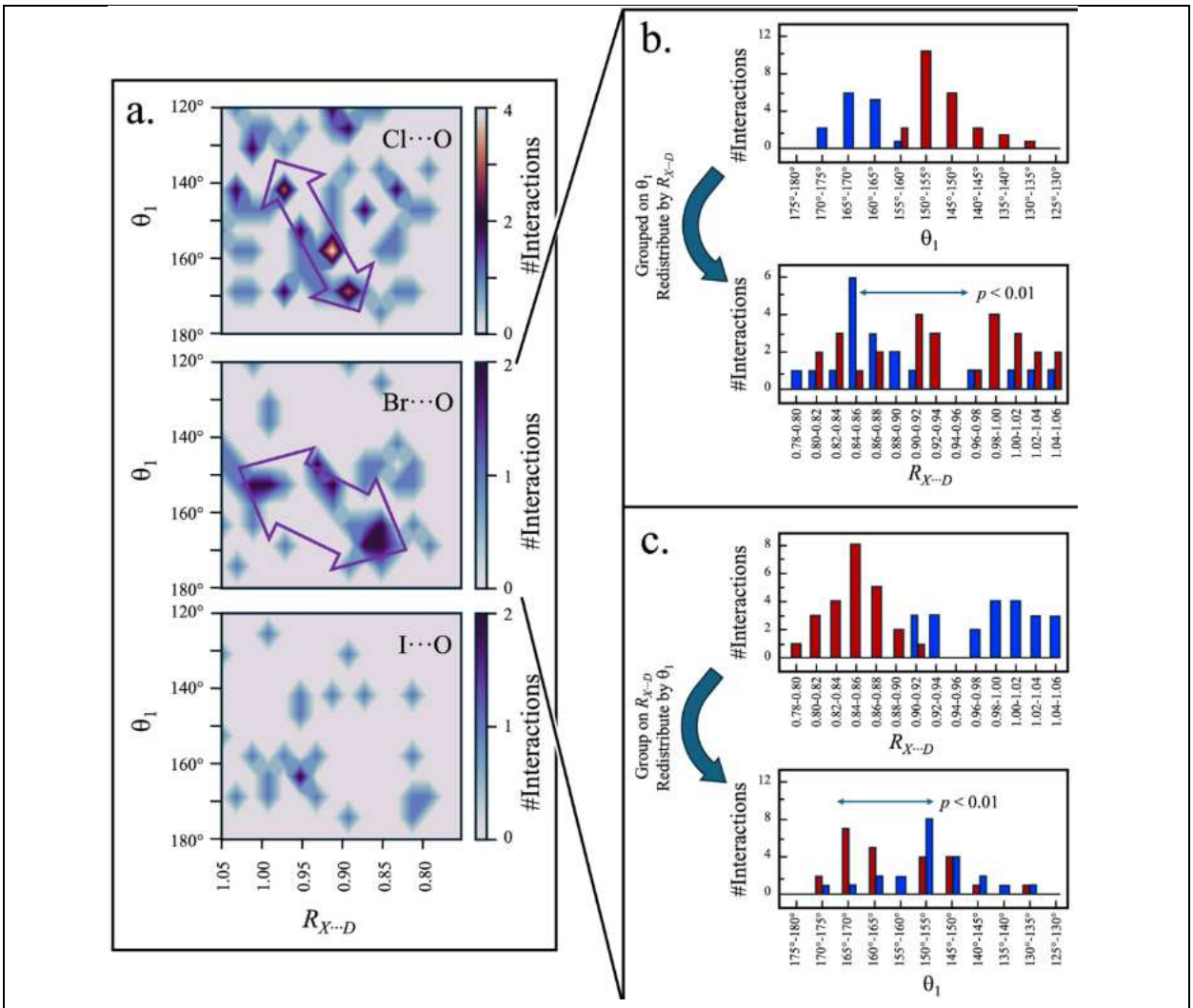


Figure 4: Relationship between θ_1-angles and $R_{X\cdots D}$ HB distances for numbers of X···O HB*e*XB interactions. **a.** Contour heatmaps for Cl···O, Br···O, and I···O. The color scales for the numbers of interactions (#Interactions) are shown on the right. Open arrows connect high occurrence populations in the Cl···O (#Interactions ≥ 3) and Br···O (#Interactions = 2). **b.** Distributions of Br···O HBeXBs grouped by θ_1 (top panel) and redistributed by their $R_{X\cdots D}$ (lower panel). The two distinct distributions according to θ_1-angles are colored blue and red. These colors for the θ_1-angles groups are maintained in the bottom panel after redistribution according to $R_{X\cdots D}$. The redistributed blue and red interactions in the lower panel are distinctive, with *p*-value < 0.01. **c.** Distributions of Br···O HBeXBs grouped by $R_{X\cdots D}$ (top panel) and redistributed by their θ_1-angles (lower panel). The two distinct distributions according to $R_{X\cdots D}$ are colored blue and red. These colors for the $R_{X\cdots D}$ groups are maintained in the bottom panel after redistribution according to θ_1-angles. The redistributed blue and red in the lower panel are distinctive, with *p*-value < 0.01.

4.4.3. Analyses of HB effects on XB energies

We next address the question of whether the effects of HBs on the XB geometries have any bearing on the HB*e*XB interaction energies. For these analyses, we apply quantum mechanical (QM) calculations to determine how various HB donors contribute to the electrostatic surface potentials (ESPs) of the halogen s-hole and how these ESP values may affect the overall interaction energies of halogenated ligands into protein pockets with available XB acceptor and HB donor functions (Figure 5a).

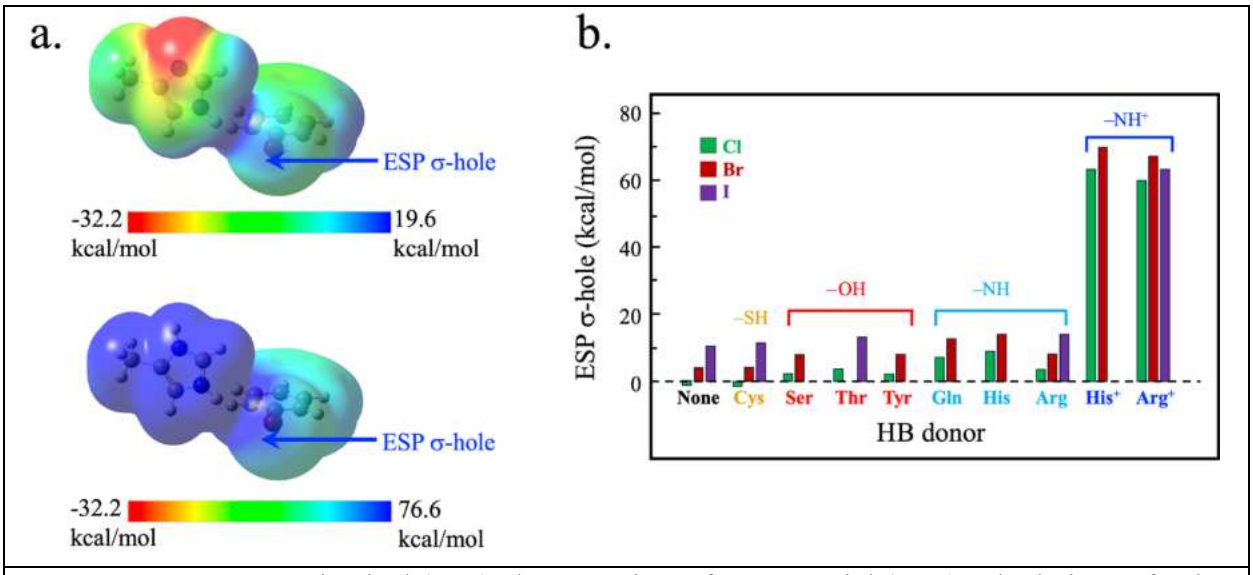


Figure 5: Quantum mechanical (QM) electrostatic surface potential (ESP) calculations of X in models for HB*e*XB interactions. **A.** ESP of bromobenzene model for a XB donor with neutral unprotonated (top) and charged protonated (bottom) 4-methylimidazole models for His and His^+, respectively. The energy scales for each ESP calculation are shown under each surface. Arrows indicate the surface points where the ESPs for the σ-holes of each halogen are calculated. **B.** ESP calculated at the σ-holes for chlorobenzene (green bars), bromobenzene (red bars) or iodobenzene (violet bars), with various amino acid side chains as HB donors as found in the HB*e*XB dataset. The geometries for the models for HB*e*XB interactions.

Models for the ESP calculations were constructed starting with a halogenated benzene representing chlorinated, brominated or iodinated ligands (Figure 5b). These model ligands were then placed in geometries with average h_1, h_2 and $R_{X\cdots D}$ values for the HB components as determined from the PDB survey of –OH, –SH, neutral –NH and positively

charged $–NH^+$ donors for each halogen type. Results from these calculations show that –OH and –NH HB donors significantly increase the electropositive ESP of the s-hole (by 2 to 10 kcal/mol) for all halogens over having no HB or –SH as the HB donor. The most dramatic effects come from the protonated and positively charged $–NH^+$ donors, with the His^+ and Arg^+ sidechains increasing the electropositive potential by 50 – 60 kcal/mol.

At this point, we can ask how this very large increase in the electropositive potential of the s-hole from the positively charged HB donors could affect the interaction energies of the ligand to its protein target (E_{int}). We address this question by constructing trimolecular complexes that model a ligand sitting in a protein binding pocket (Figure 6). Each complex contains a bromobenzene molecule serving as a model halogenated ligand, formaldehyde as a model XB acceptor and 4-methylimidazole as a model for a His side chain as an HB donor that can be unprotonated (neutral His model) or protonated (positively charge His^+ model), with each component placed in "ideal" geometries for XB ($q_1 \approx 180°$, $q_2 \approx 120°$ and $R_{X\cdots A}$ at $0.95\sum R_{vdW}$) and HB ($h_1 \approx 90°$ and $R_{X\cdots D}$ at $0.9\sum R_{vdW}$). The trimolecular complexes were subjected to QM geometry optimization with either a neutral His or protonated His^+ as an HB donor.

The difference in E_{int} between the optimized His^+ ($E_{opt\text{-}His+}$) and optimized neutral His complexes ($E_{opt\text{-}His}$) is $\Delta E_{His+–His}$ or ΔE_{int} = -5.6 to -5.9 kcal/mol. We note that the absolute energy values depend on the method used in their calculation, with DFT energies being about 1 kcal/mol more negative than their MP2 calculated counterparts. However, the differences when comparing among different protonation and/or conformational states are very small. With these optimized charged and neutral complexes, we constructed a thermodynamic cycle to determine how this difference in E_{int} arises from protonation of the HB donor in different conformations of the binding pocket (Fig. 6). We start with the energetically most favorable

model, which is the optimized His^+ complex. Deprotonating the His^+ without additional geometry optimization results in a loss of 6.3 kcal/mol of interaction energy ($E_{unopt\text{-}His}$). Optimizing the geometry of this unprotonated complex results in the optimized neutral His complex ($E_{opt\text{-}His+}$) that regains a small 0.6 kcal/mol energy. Thus, this leg of the cycle indicates that the proton of His^+ with the ligand in the pocket is greatly stabilizing within the His^+ conformation, but that once deprotonated the XB and HB geometries will readjust to a slightly more stable uncharged conformation.

Starting with the optimized neutral His complex, the second leg of the cycle indicates that protonation of this neutral configuration ($E_{unopt\text{-}His+}$) results in reduction of the E_{int} by ~5.5 kcal/mol. Optimization of this protonated complex provides an additional 0.1 kcal/mol of stabilizing energy to E_{int}. The overall result is that there is ~1.8 kcal/mol difference in interaction energy between the optimized uncharged His and optimized His^+.

In the calculations for this thermodynamic cycle, the XB and HB donors and acceptors were each treated as separate components, resulting in interaction energies that reflect an artificial situation where these components starting at infinite distances are brought together in concerted fashion into completed assemblies (Figure 7, panels labeled X:A:D). The issue with this type of analysis is that ligands do not bind to proteins in this manner. In addition, we cannot from this type of analysis disentangle the effects of charge on the XB from that on the HB, because of the non-additive effects that the HB and XB have on the polarization of the halogen, on that of the HB donor and on the XB acceptor.

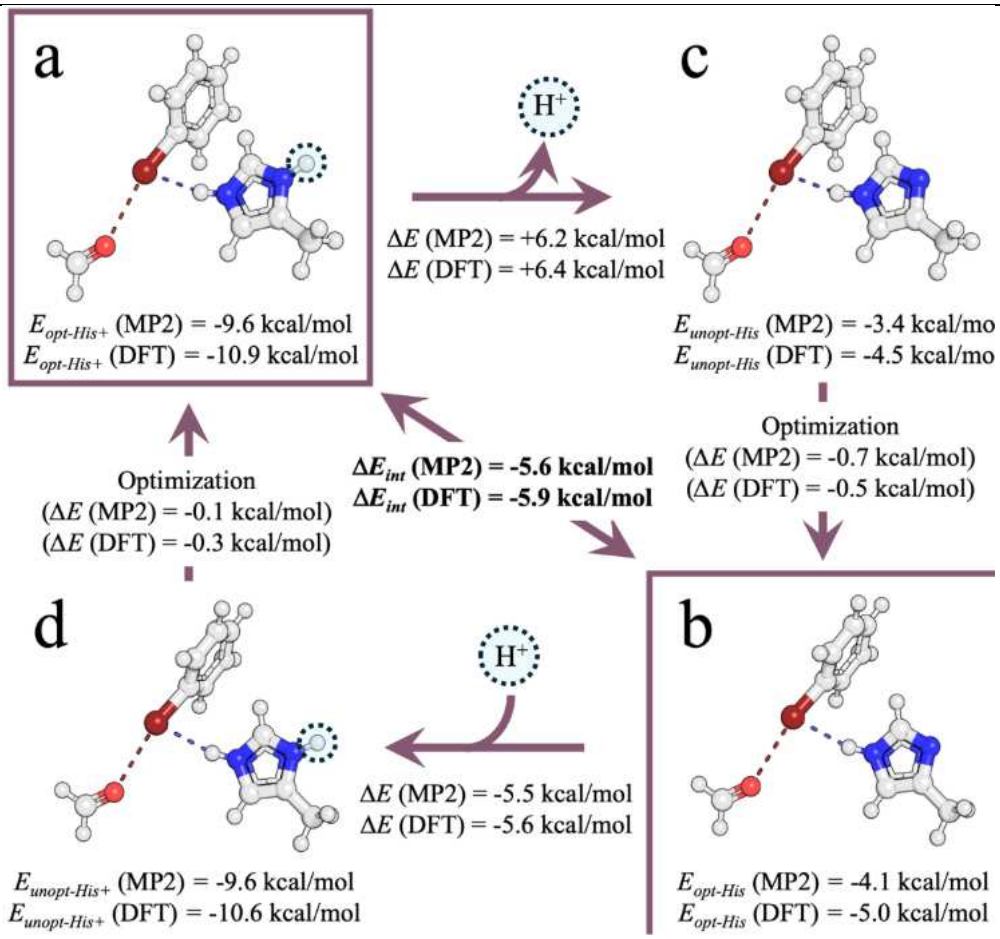


Figure 6: QM thermodynamic cycle analysis focusing on effects of charge and conformation of HB donors on interaction energies of an HB*e*XB. Trimolecular complexes composed of bromobenzene as a XB donor and HB acceptor (mimicking a brominated ligand), formaldehyde as a XB acceptor (mimicking a carbonyl oxygen of the peptide backbone) and imidazole as an HB donor (mimicking the sidechain of a His amino acid) was assembled as model systems of HB*e*XB interaction systems. The geometry optimized system with a protonated and positively charged 4-methylimidazole (leading to the $E_{opt\text{-}His+}$ energy) and geometry optimized system with an unprotonated and uncharged imidazole (leading to the $E_{opt\text{-}His}$ energy) are shown in boxes **a** and **b**, respectively. Removing the proton and charge from the imidazole in **a** results in an unprotonated complex in the charged optimized conformation (**c**, the $E_{unopt\text{-}His}$ energy), with the difference between $E_{opt\text{-}His+}$ and $E_{unopt\text{-}His}$ terms representing the effect of the charge on the energy of the complex in the optimized charged conformation. The difference in energy between $E_{unopt\text{-}His}$ and $E_{opt\text{-}His}$ represents the difference in energy between the charge optimized and uncharged optimized conformations for an uncharged complex. Adding a proton and charge to the imidazole in **b** results in an protonated complex in the uncharged optimized conformation (**d** and the $E_{unopt\text{-}His+}$ energy), with the difference between $E_{opt\text{-}His}$ and $E_{unopt\text{-}His+}$ terms representing the effect of the charge on the energy of the complex in the optimized uncharged conformation. The difference in energy between $E_{unopt\text{-}His+}$ and $E_{opt\text{-}His+}$ represents the difference in energy between the charge optimized and uncharged optimized conformations for a charged complex. Energies calculated from both MP2 and DFT methods are listed for comparison.

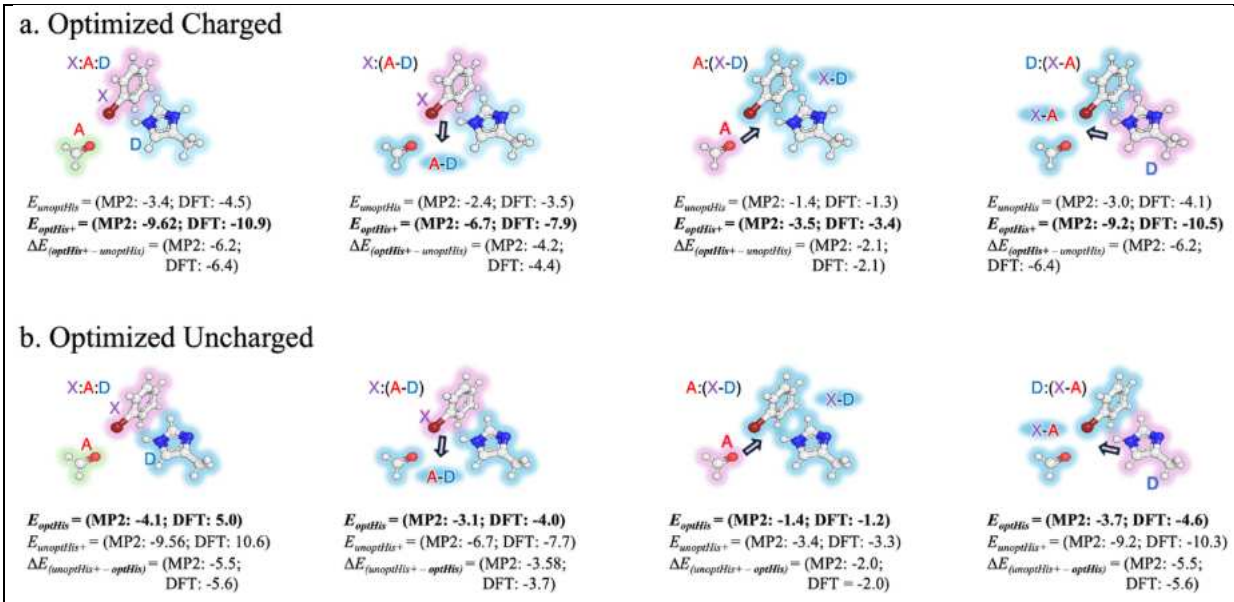


Figure 7: Analysis of energies for HB*e*XB models with protonated and charged (**a**) or unprotonated and uncharged (**b**) HB donors, with various components treated as associated pairs. The X:A:D assemblies and energies in **a** are the same as those in Figure 6**a** and **c**, while X:A:D assemblies and energies in **b** are those in Figure 6**b** and **d**. X is the XB donor and HB acceptor, A is the XB acceptor and D is the HB donor component of the HB*e*XB models. The X:(A-D) complexes treat A and D as an associated pair (highlighted by the blue "glow" around the components) in the energy calculations, with X as the "binding" independent ligand (highlighted by the red "glow" around that molecule). The A:(X-D) complexes treat X and D as an associated pair, with A as the "binding" independent ligand. The D:(X-A) complexes treat X and A as an associated pair, with D as the "binding" independent ligand. The differences in the calculated energies (ΔE) reflect the effect of adding a proton and positive charge to the HB donor (D) in either the charge optimized conformations (upper assemblies **a**) or the uncharged optimized geometries (lower assemblies **b**). Energies calculated from both MP2 and DFT methods are listed for comparison.

To address these issues, we used the same molecular models but reanalyzed the interaction energies by pairing two of the three components into a single assembly (allowing the calculation to account for the properties such as basis set correction of the paired assembly[31]), leaving the third as its own entity. In this manner, we can recalculate and interpret the resulting E_{int} as the binding of this third individual component to a preassembled complex. For example, in Fig 7 panel labeled X:(A-D), the XB acceptor (A) and HB donor (D) are treated as an assembled pair and the XB donor (X) treated individually. The interpretation is that the resulting interaction

energy is for X binding to a pre-established A-D complex, which is a more relevant model for how a halogenated ligand binds to the environment of a protein pocket. For these pair-wise calculations, we can again compare the E_{int} energies of conformations that have been geometry optimized with the protonated and charged His^+ (Figure 7a) and that of the unprotonated and uncharged His (Figure 7b). The differences in these energies allow us to estimate the effect of placing a charge on the HB donor for the conformation that has been geometry optimized for the charged state ($\Delta E_{(\mathbf{optHis}+ \text{ - } unoptHis)}$) or the uncharged state and ($\Delta E_{(unoptHis+ \text{ - } \mathbf{optHis})}$).

The resulting analyses of the X:(A-D) complexes indicate that the halogenated ligand placed in a charged pocket has an interaction energy that is 3.6 kcal/mol ($\Delta E_{(unoptHis+ \text{ - } \mathbf{optHis})}$) to 4.2 kcal/mol ($\Delta E_{(\mathbf{optHis}+ \text{ - } unoptHis)}$) more favorable than in an uncharged pocket and that the difference in this charge effect depends on whether the A-D pair is optimized for the charged or uncharged conformation (similar to the conclusion drawn from the thermodynamic cycle in Fig. 6). However, this set of pair-wise interaction energies still does not provide any insights into how the charge affects the XB versus the HB.

The A:(X-D) complex is seen as bringing an XB acceptor towards the XB donor into a pre-established HBed halogen. There is a surprising consistency in the interaction energy in terms of the uncharged and charged HB donor in the (X-D) pair, regardless of how its geometry was optimized (charged or uncharged). The effect of the charge is interpreted as reflecting ~2 kcal/mol energy attributed primarily to the stabilization of the XB interaction.

Finally, analyses of the D:(X-A) complexes shows that bringing the HB donor towards the halogen that is already involved in a XB interaction is highly dependent on the relative conformation of the HB donor. In this case, the effect of the positively charged HB donor on the

interaction energies are seen to be ~5.5 kcal/mol ($\Delta E_{(unoptHis+ - \mathbf{optHis})}$) or 6.3 kcal/mol ($\Delta E_{(\mathbf{optHis}+ - unoptHis)}$), indicating that the primary conformational effects are harbored in the HB to the halogen.

4.5. Discussion

Both HB and XB interactions are recognized as strong potential contributors to molecular recognition but there is now a greater appreciation that how each relates to the other is important, particularly in the design of inhibitors against clinically important protein targets. Their interrelationships can range from competitive to mere indifference[8]. However, it is more interesting to consider how XBs and HBs work cooperatively, such as in HB*e*XBs[5, 11], to enhance ligand selectivity and/or affinity. The environment of a binding site within a protein is replete with both HB donors and XB acceptors available to establish these types of synergistic interactions. In this study, we surveyed the PDB to better understand how HBs would affect the geometric and energetic properties of XBs. We note that this study applied very stringent criteria for what constitutes an HB*e*XB. As with the prior CSD survey, it was clear that not all halogen substituents that were in close proximity to both XB acceptors and HB donors would or should be categorized in this way. A large number of interactions from the initial screen were eventually found to be primarily HBs to HB acceptors (as defined by their relative geometries), with the halogen being close by but likely not the primary driver of the interactions. Still, nearly half (~45%) of halogenated ligands that participate in XB interactions were identified as legitimate HB*e*XBs. Thus, the HB*e*XB has a significant presence in protein-ligand complexes.

Taken in aggregate, the XB-Only and HB*e*XB datasets look very similar, with the number of interactions both following the trend Cl > Br > I. In addition, the primary XB acceptors are the

carbonyl oxygens of the polypeptide backbone in both sets. Further analyses of the two types of interactions showed no significant differences in the overall geometries of the XB interactions. The XB-Only interactions tend towards slightly shorter distances ($R_{X\cdots A}$) relative to their respective van der Waals radii and very slightly more linear approach of the XB acceptors to the halogen s-holes (q_1-angles) going from Cl to Br to I, following the trend of a deeper s-hole as the halogen becomes larger and more polarizable. The XB donor-acceptor distances tend to be indistinguishable within the error and the angle of approach becomes less linear in the HB*e*XB set. This lack of general effect of HBs on XBs is likely due to the large number of molecular interactions in a protein pocket, leaving XBs as one of several contributing rather than the dominant factor that defines the geometry of a bound ligand. However, distinct relationships and trends become apparent when delving into the details of specific sets of parameters.

We can ask what are the amino acids that serve as HB donors to the HB*e*XB. Looking at the dataset as a whole, it is clear that the most common HB donors are hydroxyl groups from Ser, Thr and Tyr residues. In general, finding that –OH (50%) > $–NH^+$ (25%) > –NH (9%) > –SH (6%) as HB donors generally follows the relative occurrence of each of these classes of amino acid in proteins (15.6% –OH, 13.2% $–NH^+$, 9% –NH and 1.4% –SH). It should be noted that the representation of –NH as HB donors should be larger, since the polypeptide backbone has one NH for each amino acid. However, the HB*e*XB data found that backbone NHs very rarely form HBs to ligands probably because backbone amino groups are typically already involved in HBs to peptide carbonyl oxygens of the backbone. Unlike HB acceptors such as the carbonyl oxygen, –NH as a donor typically makes a single HB, precluding it from bifurcated interactions with the peptide backbone and a halogen substituent.

Diving into the HB*e*XB data more deeply, it is clear that the prevalence of –OH HB donors is skewed by the larger Cl data. For the Br and I data, the positively charged $–NH^+$ donors from Arg, His and Lys are as common as the –OH donors. It is unclear from this analysis why the Cl-HB*e*XB data is skewed in this way. The higher occurrence of $–NH^+$ in the Br and I sets could reflect the preference for Arg and His, along with Trp and Tyr in protein binding pockets[40].

One immediate observation from a more detailed consideration of the HB*e*XB geometries is that the q_1-angles for both Cl and Br become more linear (approaching 180°) as the HB donor to halogen acceptor distance becomes shorter ($R_{X\cdots D}$), suggesting that a stronger HB results in the halogen becoming a more linearly aligned and possibly stronger XB interaction. This trend is less apparent with iodine probably because of the small size of the HB*e*XB dataset for this halogen. Thus, there is an apparent effect of the HB geometry on that of the associated XB interaction.

Does the HB in the protein environment actually affect the XB energy? To address this question, we first analyzed the ESP of the halogen s-hole in the absence and presence of each type of HB donor found in the HB*e*XB dataset. The models for these calculations were constructed around the average conformational geometries observed in the data. Not surprisingly, the uncharged –OH and –NH HBs increased the ESP values by 3 to 8 kcal/mol. However, the –SH donor had very little effect. Apparently, the type of HB donor significantly affects the XB potential of halogen substituents.

Protonation and adding a positive charge to an Arg or His residue increases the ESP values by 60 to 65 kcal/mol, but does this translate to significantly higher ligand interaction energies? From this thermodynamic cycle, we can conclude first that the neutral His complex is

conformationally more stable when ignoring the charge but that protonating the HB donor provides sufficient electrostatic potential to pull the complex towards the less favorable His^+ conformation. Thus, if the ligand enters a protein pocket with a preprotonated HB donor, deprotonation results in a large loss in electrostatic potential, followed by a small structural rearrangement to a slightly more favorable conformation. Similarly, if the ligand enters a protein pocket that starts with a neutral HB donor, protonation of the donor provides near equal favorable electrostatic potential and the subsequent conformational rearrangements to account for the positive charge results in a near negligible change in conformational energy. The overall effect, taking both electrostatic and conformational changes into account is ~5.6 to 5.9 kcal/mol stabilization of the complex resulting from protonation and associated positive charge of the HB donor.

Finally, we addressed the question of whether the effect of charge is primarily on the XB or the HB terms of the interaction energies by analyzing the three-component system as series of pairwise interactions between one component (either X, A or D) and a bimolecular complex of the remaining two components (A-D, X-D or X-A). The result of this analysis showed that the positive charge has a fairly consistent ~2 kcal/mol effect in increasing the strengths of the XB regardless of the conformation of the X-D pair, while the effect on the HB is strongly dependent on the conformation of the HB donor relative to the halogen. The halogenated ligand entering a protein pocket that is already predisposed to a charged HB donor that forms an HB*e*XB will experience greater than 5 kcal/mol increased interaction energy (equivalent to increasing the affinity by upwards of three orders of magnitude) than one entering a pocket with an HB donor uncharged.

We temper these discussions for how *nc*NC interaction energies may translate to overall binding affinities by recognizing that ligand binding cannot be considered in such simple, isolated constructs. Forming a ligand-protein complex requires desolvation of both the ligand and the protein pocket. The protein pocket in its "empty" form will be filled with solvent[41] that will partner with potential XB acceptors and HB donors for the halogen substituent. Furthermore, the halogens of the ligand in the unbound state will interact with the bulk solvent[42]. Finally, although we tend to think about a protein's binding pocket as being fairly rigid, the absence of a ligand may allow the HB donating sidechain to seek other acceptors. Thus, for the HB*e*XB to form upon binding, there may also need to be some conformational readjustment to set the HB donor and the XB acceptor to accommodate these *nc*NC interactions. How the energies and entropies associated with the halogen/protein interactions are balanced with that of desolvation and conformational rearrangement of both components will have the final say on how any HB*e*XB contributes to overall affinity of a halogenated ligand.

How could these structure and energy analyses facilitate the design of inhibitors to a protein target? It is clear that the protein environment, particularly the HB donors immediately around the halogen substituent affect both the geometries and, specifically for the positively charged donors, interaction energies of the XBs. However, XBs are only one of several intermolecular interactions that define how a ligand binds in the pocket and is unlikely to significantly affect the overall conformation of the complex.

The second observation is that the HB to the halogen has some effect on both the geometry and interaction energy of the XB. Figure 8 compares the XB and HB geometries of five different inhibitor/protein complexes with Br-HB*e*XB*s* to the carbonyl oxygens of the peptide backbone and different HB donors, selected on the basis that their respective distances and angles are close

to the ideal HBeXB geometries. For this set of inhibitors across ~10-fold differences in IC_{50} values, we observe a trend that lower IC_{50}'s are associated with h_1 values that are closer to the ideal 90° from the s-hole. Thus, as with the energy calculations in Figures 6 and 7, the geometry of the HB donor to the halogen has the most significant influence on the HB*e*XB and its effects on ligand binding. The sample size for this IC_{50} trend is small and these associations may not hold for tightly bound inhibitors, where the influence of the HB*e*XB may be diminished by presence of many more and potentially stronger classical noncovalent interactions (other HBs, π-stacking, etc.).

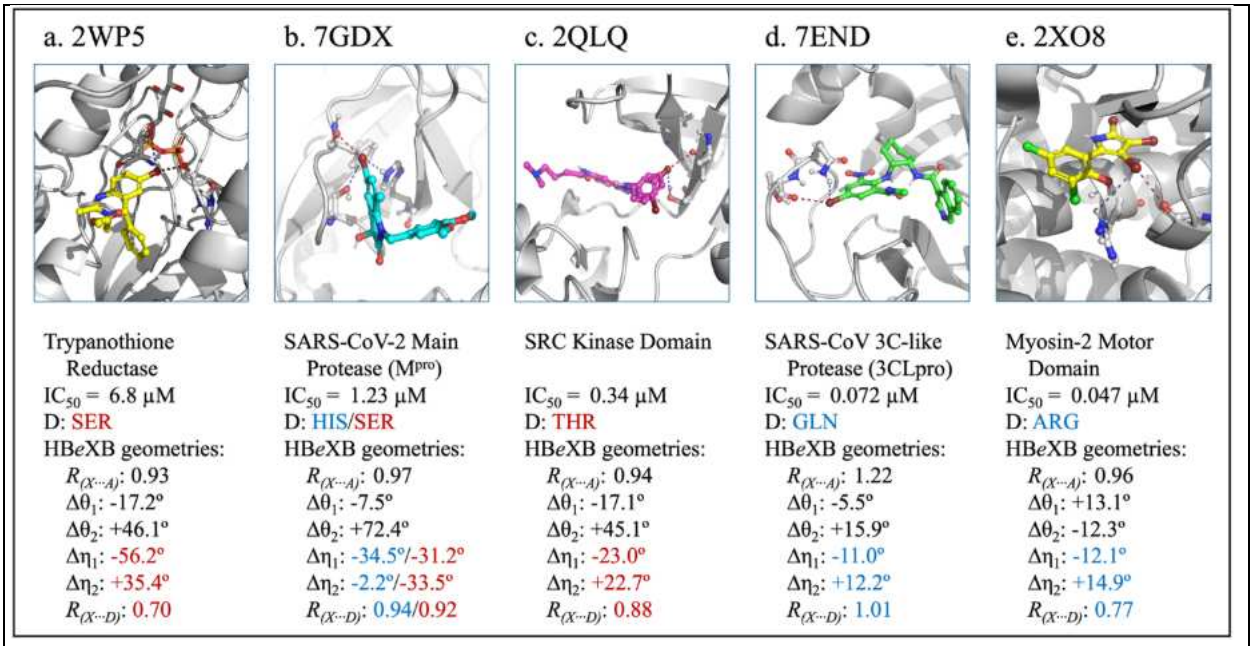


Figure 8: Selected examples of Br···O HBeXB interactions from the PDB survey representing a broad range of IC_{50} values for their respective inhibitors. In each example, the XB acceptor is the carbonyl oxygen of the peptide backbone. The PDB codes (a. 2WP5[43], b. 7GDX[44], c. 2QLQ[45], d. 7END[46] and e. 2XO8[47]), associated protein targets and amino acids serving as HB donors are listed, with hydroxyl type donors in red and nitrogenous donors in blue fonts. The distance parameters for the XB and HB interactions ($R_{X\cdots A}$ and $R_{X\cdots D}$) are listed as the fraction of the sums of the van der Waals radii ($f\sum R_{vdW}$) for the respective heavy atoms. Angle parameters are listed deviations from their ideal values for XBs ($\theta_1 = 180°$, $\theta_2 = 120°$) and HBs ($\eta_1 = 90°$ and $\eta_2 = 109.5°$ for OH, 129° for His and 120° for GLN and ARG HB donors). Figures for the protein-inhibitor complexes were created using PYMOL[48].

4.6. Conclusion

Since a halogenated ligand designed *de novo* would not have any predetermined geometry for the components of the HB*e*XB interaction in protein binding pocket, it would be difficult to *a priori* predict what the conformation and energies of the interaction could be. However, if the structure of ligand in its target binding pocket is available, the results from this study could be used to help identify potential sites for adding a halogen substituent to optimize its potential as a XB donor and just as important, as an HB acceptor to maximize its ability to serve as an inhibitor. Artificial intelligence is helping to predict biomolecular interactions, including XBs[49] and the type of data presented here will further facilitate the application of these methods for structure-based inhibitor and drug design.

One limitation of this study is that we have surveyed and analyzed the charge effects of only the first shell HB donors to halogen substituents. We had already seen with the *cX*-Zyme construct with *m*EndoG that the charge of an HB donor from the second shell can have profound effects, in this case, the catalytic activities of a XB[7]. In addition, it has been shown that waters can serve as bridges between an XB donor and acceptor[50]. Thus, a more comprehensive study of second shell HBs, particularly in regard to charged species, would provide a more complete view of how the complex environment of a binding pocket in a protein structure affects the *nc*NC interaction that is so dominant in drug design and emergent in biomolecular engineering

4.7. Funding

This study was supported by a grant (CHE-2203161) from the National Science Foundation.

REFERENCES

(1) (a) Brammer, L.; Peuronen, A.; Roseveare, T. M. Halogen bonds, chalcogen bonds, pnictogen bonds, tetrel bonds and other σ-hole interactions: a snapshot of current progress. *Acta Crystallogr C-Struct Chem,* **2023**, *79*, 204-216. DOI: 10.1107/S2053229623004072. (b) Walker, M. G.; Mendez, C. G.; Ho, P. S. Non-classical Non-covalent σ-Hole Interactions in Protein Structure and Function: Concepts for Potential Protein Engineering Applications. *Chem-Asian J* **2023**, *18* (7), e202300026 DOI: 10.1002/asia.202300026.

(2) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci., USA* **2004**, *101* (48), 16789-16794. DOI: 10.1073/pnas.0407607101.

(3) Lam, P. Y. S.; Clark, C. G.; Smallwood, A. M.; Alexander, R. S. Structure-based drug design utilizing halogen bonding: Factor Xa inhibitors. In The 238th ACS National Meeting, Washington, DC; 2009.

(4) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J Med Chem* **2013**, *56* (4), 1363-1388. DOI: 10.1021/jm3012068.

(5) Carlsson, A. 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* (28), 4135-4147. DOI: 10.1021/acs.biochem.8b00603.

(6) (a) Danelius, E.; Andersson, H.; Jarvoll, P.; Lood, K.; Grafenstein, J.; Erdelyi, M. Halogen Bonding: A Powerful Tool for Modulation of Peptide Conformation. *Biochemistry* **2017**, *56* (25), 3265-3272. DOI: 10.1021/acs.biochem.7b00429. (b) Pizzi, A.; Demitri, N.; Terraneo, G.; Metrangolo, P. Halogen bonding at the wet interfaces of an amyloid peptide structure. *CrystEngComm* **2018**, *20* (36), 5321-5326.

(7) Walker, M. G.; Mendez, C. G.; Ho, A. N.; Czarny, R. S.; Rappe, A. K.; Ho, P. S. Design of a halogen bond catalyzed DNA endonuclease. *Proc Natl Acad Sci, USA* **2025**, *122* (14), e2500099122. DOI: 10.1073/pnas.2500099122.

(8) Rowe, R. K.; Ho, P. S. Relationships between hydrogen bonds and halogen bonds in biological systems. *Acta Crystallogr B Struct Sci Cryst Eng Mater* **2017**, *73* (Pt 2), 255-264. DOI: 10.1107/S2052520617003109.

(9) Aakeröy, C. B.; Panikkattu, S.; Chopade, P. D.; Desper, J. Competing hydrogen-bond and halogen-bond donors in crystal engineering. *CrystEngComm* **2013**, *15* (16), 3125-3136. DOI: 10.1039/c2ce26747k.

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

(11) Riel, A. M. S.; Rowe, R. K.; Ho, E. N.; Carlsson, A. C.; Rappe, A. K.; Berryman, O. B.; Ho, P. S. Hydrogen Bond Enhanced Halogen Bonds: A Synergistic Interaction in Chemistry and Biochemistry. *Acc Chem Res* **2019**, *52* (10), 2870-2880. DOI: 10.1021/acs.accounts.9b00189.

(12) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28* (1), 235-242. DOI: 10.1093/nar/28.1.235.

(13) (a) Wu, W., Samoszuk, M. K., Comhair, S. A. A., Thomassen, M. J., Farver, C. F., Dweik, R. A., Kavuru, M. S., Erzurum, S. C., and Hazen, S. L. Eosinophils generate brominating oxidants in allergen-induced asthma. *J. Clin. Invest.* **2000**, *105*, 1455-1463. (b) Buss, I., H., Senthilmohan, R., Darlow, B.A., Mogridge, N., Kettle, A. J., and Winterbourn, C.C. 3-Chlorotyrosine as a Marker of Protein Damage by Myeloperoxidase in Tracheal Aspirates From Preterm Infants: Association With Adverse Respiratory Outcome. *Pediatric Res,* **2003**, *53*, 455-462.

(14) Shen, Z.; Mitra, S. N.; Wu, W.; Chen, Y.; Yang, Y.; Qin, J.; Hazen, S. L. Eosinophil peroxidase catalyzes bromination of free nucleosides and double-stranded DNA. *Biochemistry* **2001**, *40* (7), 2041-2051. DOI: 10.1021/bi001961t.

(15) Vargason, J. M.; Eichman, B. F.; Ho, P. S. The extended and eccentric E-DNA structure induced by cytosine methylation or bromination. *Nat Struct Biol* **2000**, *7*, 758-761.

(16) (a) Voth, A. R., and Ho, P.S. The Role of Halogen Bonding in Inhibitor Recognition and Binding by Protein Kinases. *Curr Topics Med Chem* **2007**, *7*, 1336-1348. (b) Lu, Y.; Shi, T.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. Halogen bonding--a novel interaction for rational drug design? *J Med Chem* **2009**, *52* (9), 2854-2862. DOI: 10.1021/jm9000133. (c) Hernandes, M. Z.; Cavalcanti, S. M.; Moreira, D. R.; de Azevedo Junior, W. F.; Leite, A. C. Halogen atoms in the modern medicinal chemistry: hints for the drug design. *Curr Drug Targets* **2010**, *11* (3), 303-314. DOI: 10.2174/138945010790711996. (d) 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* **2013**, *53* (11), 2781-2791. DOI: 10.1021/ci400257k.

(17) Xu, Z.; Liu, Z.; Chen, T.; Chen, T.; Wang, Z.; Tian, G.; Shi, J.; Wang, X.; Lu, Y.; Yan, X.; et al. Utilization of halogen bond in lead optimization: a case study of rational design of potent phosphodiesterase type 5 (PDE5) inhibitors. *J Med Chem* **2011**, *54* (15), 5607-5611. DOI: 10.1021/jm200644r.

(18) Xu, Z.; Yang, Z.; Liu, Y.; Lu, Y.; Chen, K.; Zhu, W. Halogen bond: its role beyond drug-target binding affinity for drug discovery and development. *J Chem Inf Model* **2014**, *54* (1), 69-78. DOI: 10.1021/ci400539q.

(19) (a) Ali, S.; Tian, X. C.; Meccia, S. A.; Zhou, J. Highlights on US FDA-approved halogen-containing drugs in 2024. *Eur J Med Chem* **2025**, *287*. DOI: 10.1016/j.ejmech.2025.117380. (b) Ali, S.; Zhang, C.; Zhou, J. Prevalence of Halogens in 2025 FDA-Approved Drugs: Highlight on Suzetrigine, the First Nonopioid Painkiller in Decades. *J Med Chem* **2026**, *69* (4), 3649-3655. DOI: 10.1021/acs.jmedchem.6c00417.

(20) Poznanski, J.; Poznanska, A.; Shugar, D. A Protein Data Bank Survey Reveals Shortening of Intermolecular Hydrogen Bonds in Ligand-Protein Complexes When a Halogenated Ligand Is an H-Bond Donor. *Plos One* **2014**, *9,* e99984. DOI: 10.1371/journal.pone.0099984.

(21) Murray, J. S.; Lane, P.; Politzer, P. Expansion of the sigma-hole concept. *J Mol Model* **2009**, *15* (6), 723-729. DOI: 10.1007/s00894-008-0386-9.

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

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

(24) Riley, K. E.; Vazquez, M.; Umemura, C.; Miller, C.; Tran, K. A. Exploring the (Very Flat) Potential Energy Landscape of R-Br center dot center dot center dot pi Interactions with Accurate CCSD(T) and SAPT Techniques. *Chem-Eur J* **2016**, *22* (49), 17690-17695. DOI: 10.1002/chem.201603674.

(25) (a) Eskandari, K.; Zariny, H. Halogen bonding: A lump-hole interaction. *Chem Phys Lett* **2010**, *492* (1-3), 9-13. DOI: 10.1016/j.cplett.2010.04.021. (b) Jahromi, H. J.; Eskandari, K. Halogen bonding: a theoretical study based on atomic multipoles derived from quantum theory of atoms in molecules. *Struct Chem* **2013**, *24* (4), 1281-1287. DOI: 10.1007/s11224-012-0156-2.

(26) 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* (8), 1711-1713. DOI: 10.1351/Pac-Rec-12-05-10.

(27) Ford, M. C.; Rappe, A. K.; Ho, P. S. A Reduced Generalized Force Field for Biological Halogen Bonds. *J Chem Theory Comput* **2021**, *17* (8), 5369-5378. DOI: 10.1021/acs.jctc.1c00362.

(28) 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* (29), 4891-4903. DOI: 10.1021/bi400590h.

(29) Scholfield, M. R.; Ford, M. C.; Carlsson, A. C.; Butta, H.; Mehl, R. A.; Ho, P. S. Structure-Energy Relationships of Halogen Bonds in Proteins. *Biochemistry* **2017**, *56* (22), 2794-2802. DOI: 10.1021/acs.biochem.7b00022.

(30) Scheiner, S. Enhancement of Halogen Bond Strength by Intramolecular H-Bonds. *J Phys Chem A* **2023**, *127* (21), 4695-4703. DOI: 10.1021/acs.jpca.3c02049.

(31) Sutar, R. L.; Huber, S. M. Catalysis of Organic Reactions through Halogen Bonding. *ACS Catal* **2019**, *9* (10), 9622-9639. DOI: 10.1021/acscatal.9b02894.

(32) Han, Z.; Li, J.; Liao, Q.; Luo, R.; Wu, L.; Shi, Y.; Zhou, L.; Zhang, Q.; Erdélyi, M.; Zhu, W.; Xu, Z. Hydrogen Bond-Induced Binding between Organofluorine and Protein via Fluorine

Atoms: A Database Survey and Quantum Chemistry Calculation Study. *J Med Chem* **2025**, *68* (14), 14882–14894. DOI: 10.1021/acs.jmedchem.5c01043.

(33) Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput Sci Eng* **2007**, *9* (3), 90-95. DOI: Doi 10.1109/Mcse.2007.55.

(34) *Gaussian 09*; Gaussian, Inc.: 2009.

(35) Zhou, H. X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem Rev* **2018**, *118* (4), 1691-1741. DOI: 10.1021/acs.chemrev.7b00305.

(36) Peterson, K. A.; Shepler, B. C.; Figgen, D.; Stoll, H. On the spectroscopic and thermochemical properties of ClO, BrO, IO, and their anions. *J Phys Chem A* **2006**, *110* (51), 13877-13883. DOI: 10.1021/jp065887l.

(37) Boys, S. F.; Bernardi, F. Calculation of Small Molecular Interactions by Differences of Separate Total Energies - Some Procedures with Reduced Errors. *Mol Phys* **1970**, *19* (4), 553-566. DOI: Doi 10.1080/00268977000101561.

(38) van Holde, K.; Johnson, W. C.; Ho, P. S. *Principles of Physical Biochemistry, 2nd Ed.*; Pearson Prentice Hall, 2006.

(39) Hawkins, P. C. D.; Warren, G. L.; Skillman, A. G.; Nicholls, A. How to do an evaluation: pitfalls and traps. *J Comput Aid Mol Des* **2008**, *22* (3-4), 179-190. DOI: 10.1007/s10822-007-9166-3.

(40) Villar, H. O.; Kauvar, L. M. Amino-Acid Preferences at Protein-Binding Sites. *FEBS Let* **1994**, *349* (1), 125-130. DOI: Doi 10.1016/0014-5793(94)00648-2.

(41) Lai, J. K.; Ambia, J.; Wang, Y. M.; Barth, P. Enhancing Structure Prediction and Design of Soluble and Membrane Proteins with Explicit Solvent-Protein Interactions. *Structure* **2017**, *25* (11), 1758-+. DOI: 10.1016/j.str.2017.09.002.

(42) Verteramo, M. L.; Ignjatovic, M. M.; Kumar, R.; Wernersson, S.; Ekberg, V.; Wallerstein, J.; Carlstrom, G.; Chadimova, V.; Leffler, H.; Zetterberg, F.; et al. Interplay of halogen bonding and solvation in protein-ligand binding. *Iscience* **2024**, *27* (4). DOI: 10.1016/j.isci.2024.109636.

(43) Patterson, S.; Alphey, M. S.; Jones, D. C.; Shanks, E. J.; Street, I. P.; Frearson, J. A.; Wyatt, P. G.; Gilbert, I. H.; Fairlamb, A. H. Dihydroquinazolines as a Novel Class of Trypanothione Reductase Inhibitors: Discovery, Synthesis, and Characterization of their Binding Mode by Protein Crystallography. *J Med Chem* **2011**, *54* (19), 6514-6530. DOI: 10.1021/jm200312v.

(44) Boby, M. L.; Fearon, D.; Ferla, M.; Filep, M.; Koekemoer, L.; Robinson, M. C.; Chodera, J. D.; Lee, A. A.; London, N.; von Delft, A.; et al. Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors. *Science* **2023**, *382* (6671), eabo7201. DOI: 10.1126/science.abo7201.

(45) Michalczyk, A.; Klüter, S.; Rode, H. B.; Simard, J. R.; Grütter, C.; Rabiller, M.; Rauh, D. Structural insights into how irreversible inhibitors can overcome drug resistance in EGFR. *Bioorgan Med Chem* **2008**, *16* (7), 3482-3488. DOI: 10.1016/j.bmc.2008.02.053.

(46) Hou, N. K.; Shuai, L.; Zhang, L. J.; Xie, X. P.; Tang, K. M.; Zhu, Y. K.; Yu, Y.; Zhang, W. Y.; Tan, Q. Z.; Zhong, G. X.; et al. Development of Highly Potent Noncovalent Inhibitors of SARS- CoV-2 3CLpro. *ACS Central Sci* **2023,** *9* (2), 217-227. DOI: 10.1021/acscentsci.2c01359.

(47) Preller, M.; Chinthalapudi, K.; Martin, R.; Knölker, H. J.; Manstein, D. J. Inhibition of Myosin ATPase Activity by Halogenated Pseudilins: A Structure-Activity Study. *J Med Chem* **2011**, *54* (11), 3675-3685. DOI: 10.1021/jm200259f.

(48) Janson, G.; Zhang, C.; Prado, M. G.; Paiardini, A. PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. Bioinformatics 2017, 33 (3), 444-446. DOI: 10.1093/bioinformatics/btw638.

(49) Engelhardt, M. U.; Mier, F.; Zimmermann, M. O.; Boeckler, F. M. Expanding and Enhancing Neural Network-Based QM-AI Models for the Accurate Prediction of Halogen-π Interaction Energies in Protein Contexts. *J Chem Inf Modeling* **2026**. DOI: 10.1021/acs.jcim.5c03249.

(50) (a) Lu, S. Y.; Jiang, Y. J.; Zhou, P.; Zou, J. W.; Wu, T. X. Geometric characteristics and energy landscapes of halogen-water-hydrogen bridges at protein-ligand interfaces. *Chem Phys Lett* **2010**, *485* (4-6), 348-353. DOI: 10.1016/j.cplett.2009.12.077. (b) Zhou, P.; Lv, J.; Zou, J.; Tian, F.; Shang, Z. Halogen-water-hydrogen bridges in biomolecules. *J Struct Biol* **2010**, *169* (2), 172-182. DOI: 10.1016/j.jsb.2009.10.006.

CHAPTER 5

DISCUSSION

The studies in this thesis aimed to explore HBeXB interactions, from engineering XB dependent enzymatic catalysts to examining protein-ligand binding energies. The state of the field of nonclassical noncovalent interactions in protein engineering, as described in chapter 2, had just begun to explore the HBeXB as a tool for enhancing protein stability. In this work, we demonstrate a closer look at the importance of protein environmental context, specifically HB donors, when incorporating XBs for functional and structural purposes. Studying the HBeXB as an enzymatic catalyst exhibits the functional opportunities and limitations of using this interaction in the active site. Surveying the PDB for existing HBeXBs in protein-ligand binding interactions provides a comprehensive look at conformational and energetic impacts of HB enhancement. In examining HBeXBs in both the catalytic active site and the ligand binding site, we have illustrated the utility and advantages with incorporating networks of ncNC interactions in structure-guided design.

5.1. HBeXB dependent catalysis of DNA-cleavage

By incorporating a halo-tyrosine in the metal-chelating position of mouse endonuclease G, we have demonstrated how a XB can functionally substitute for magnesium in DNA-cleavage catalysis. Evidence for XB dependence in the cX-Zyme includes reduced metal dependence and an inverse pH dependence to that of the wildtype mEndoG. The bridging HBeXB interaction from a bridging asparagine-histidine network demonstrates the importance of protein

environmental context and exploring primary but also secondary influences on the XB σ-hole. The hypothesis describing the bridging HBeXB as responsible for the XB-dependent DNA-cleavage catalysis was validated by the σ-hole analysis with a protonated HBeXB bridge, further emphasizing the importance of finding or designing the optimal environment for the engineered XB, especially for purposes as ambitious as enzymatic catalysis.

5.2. HBeXB impact on XB binding energies and σ-hole in protein ligand binding

From a survey of existing XBs and HBeXBs, we have observed the prevalence and geometric parameters for these interactions in protein-ligand binding. The presence of a HB can impact the XB, potentially contributing to inhibitor potency, by increasing the σ-hole ESP, stabilizing conformation, and enhancing binding energies.

A unique aspect of this study was the exploration of the impact on the XB σ-hole with various amino acid HB donors, demonstrating the variability and tunability of HBeXB interactions using nature's building blocks. Utilizing models built from the geometric parameters collected from the PDB survey, the result of accepting a HB on the X atom's σ-hole for various HB donor types (-SH, -OH, -NH, -NH+) is calculated to be from no change (-SH) to up 60-65 kcal/mol increase in ESP (-NH+).

While σ-hole size is difficult to translate *exactly* to XB strength, it can impact angle of approach for XB interactions. Furthermore, we do observe how HB donation from -NH+ donors can contribute to σ-hole size as well as HBeXB binding energy compared to -NH donors. Calculating the impact of positively charged 4-methylimidazole on the HBeXB's binding energies, we can approximate ~2 kcal/mol increase in favorable energy with protonation,

ultimately demonstrating that HB enhancement does impact XB interaction strength. Additionally, the conformation of the binding pocket plays a large role with up to ~4 kcal/mol of favorable energy if the binding pocket is adapted for protonation of the HBeXB.

5.3. HBeXB limitations in protein engineering

These findings further establish the HBeXB as a tool for structure-based design. While demonstrating the HBeXB's value in enzymes as well as protein-ligand binding, we have also examined its complexity.

In the cX-Zyme, the HBeXB is crucial to XB-dependent catalysis but also renders the enzyme more pH dependent than its wildtype. Additionally, it was not originally considered that the HB donor would come from an indirect interaction. This XB-dependent endonuclease is a prime example that secondary interactions, such as a bridging HB donor across multiple residues, should be considered in engineering nontraditional noncovalent interactions in protein systems. The exceptional difference the secondary interaction made for XB-dependent catalysis appears to be a rate contributing factor which strengthens the claim that the HB enhancement increases XB strength.

Through the survey analysis and computation, the HBeXB is shown to be impacted greatly by the conformation of the binding pocket, a complication that is not present in small molecule catalytic HBeXBs in solution. It is difficult to predict the exact conformation when engineering proteins and adds a measure of uncertainty if a network like the HBeXB will function as envisioned. Additionally, protein environments are complex, offering a variety of other interactions that contribute to both catalysis and protein-ligand binding which could conflict with or support the HBeXB. These are considerations needed for including the HBeXB

and other nontraditional noncovalent interactions in design applications. Another limitation for using this interaction in design of biomolecules is the small dataset of existing interactions. As drug development grows in *in silico* AI/ML methods, the accuracy of those calculations does heavily rely on the training sets available.

5.4. Final remarks

Besides its novelty, why might XB dependent proteins be useful? With the creation of the cX-Zyme, the next question becomes whether this functional capability extends past DNA hydrolysis. Given the proposed mechanism, that the HBeXB stabilizes the electrophilic DNA phosphate backbone for nucleophilic attack, it is reasonable to inquire whether the HBeXB can participate in catalyzing similar reactions.

One future application would be to harness the enzymatic active site for chemical synthesis. A unique property of enzymes is that their active site is built to promote the reaction's transition state, lowering the activation energy and producing stereospecific products. Given the XB's use in small molecule catalysis, utilizing this interaction in a protein pocket built for stereospecific products could be a rational design choice for biomanufacturing industry. Upon further optimizing the cX-Zyme for catalyzing DNA cleavage, is metal-free DNA cleavage catalysis possible? Perhaps the cX-Zyme could be utilized for *de novo* design of enzymes, especially if the intent is to engineer an enzyme that does not entirely depend on metal cofactors. Essentially, the creation of the cX-Zyme demonstrates that our knowledge of XB applications in enzyme function is not complete. There is more we can use this ncNC interaction for.

While many noncovalent interactions are involved in protein-ligand binding, the understanding of how HB enhancement can impact biological XBs better equips us for intentional engineering of the HBeXB for improved binding affinity and stability in protein-binding pockets. This thesis has thoroughly explored the variety of ways to tune XBs with different HB donors, protonation and geometric constraints involved in binding sites. What else can be tuned or enhanced or strengthened? Intentional engineering of the HBeXB in protein pockets could inspire other cooperative noncovalent networks worth building, expanding the design toolkit using ncNC interactions such as σ-hole interactions, π-interactions, etc.

How do we realistically incorporate this knowledge in design applications for proteins or small molecule drugs alike? One thought would be to integrate these interactions in computational ligand docking and structure predictions methods for structure-based or ligand-based drug design. Given their geometric constraints, it would be beneficial to have a program that could identify and calculate potential networks of noncovalent interactions in biomolecular structure, in this case, HBeXBs, especially including secondary and tertiary HB donors. Having identified possible HB donors for enhancing the target of interest, the impact on binding energy could be approximated and further inform how this design choice might influence the structure activity relationship of the molecule of interest.

Using a combination of existing techniques, such as molecular dynamics (MD) simulations, Quantum Theory of Atoms in Molecules (QTAIM)/Natural Bond Order (NBO) and Symmetry Adapted Perturbation Theory (SAPT) calculations, one could also identify ncNC interactions and calculate their impact on binding energies. However, these are computationally expensive methods and require rigorous quantum mechanical calculations.

Within the last year, there has been movement towards utilizing QM/MM binding energy calculations for training machine learning systems for protein-ligand binding purposes[1]. For nontraditional noncovalent interactions specifically, recent work has demonstrated integration of QM calculations with machine learning training to assess halogen-π interactions and predict interaction energies for use in molecular recognition and drug development[2]. The anisotropic nature of these interactions can be a complication when using classical molecular mechanics (MM) techniques[3-4]. Therefore, the QM/AI approach makes major steps towards more efficient methods of studying σ-hole and other ncNC interactions in biomolecular engineering by integrating computational efficiency and speed with biochemical accuracy.

In theory, training a neural network to recognize ncNC networks, especially indirect, based on their geometric and energetic constraints, like those found in these studies, to accomplish the same goal as the techniques described above could potentially streamline efforts for understanding mechanism of action. Having observed just how impactful an indirect HB donor was in the cX-Zyme and how much influence HB donation has on XB binding energies, it is apparent that indirect (secondary, tertiary, etc.) and network interactions are important considerations for structure-based design.

In the case of drug design, with protein targets, the smallest differences in protein environment can impact activity amongst tissue or species isoforms. It is the little differences, such as noncovalent interactions, direct or indirect, that can impact structure activity relationships and help determine potency of these developed drugs. A method for detecting or predicting how networks, like the HBeXB, contribute to binding could help piece together the puzzle that is structure activity relationship, especially amongst enantiomers and isoforms with

varying resulting activities. Based on these studies, we are now better informed for tuning existing HBeXB protein-ligand interactions to achieve more favorable binding energies.

Ultimately, this examination of HBeXBs provides structural biologists with a foundational understanding of the potential value of incorporating noncovalent interaction networks by demonstrating this unique ncNC network in two crucial steps of biological reactions: binding and catalysis. With this knowledge, we can further explore engineering cooperative nontraditional noncovalent networks in biomolecules efficiently and advantageously.

REFERENCES

(1) S. Borkotoky, A. Joshi, V. Kaushik, and A. Nath Jha. “Machine Learning and Artificial Intelligence in Therapeutics and Drug Development Life Cycle.” In *Drug Development Life Cycle*, edited by Juber Akhtar, Badruddeen, Mohammad Ahmad, and Mohammad Irfan Khan. IntechOpen, 2022. https://doi.org/10.5772/intechopen.104753.

(2) M.U. Engelhardt, F. Mier, M.O. Zimmermann, and F.M. Boeckler. “A QM-AI Approach for the Acceleration of Accurate Assessments of Halogen-π Interactions by Training Neural Networks.” *Journal of Chemical Information and Modeling* 65, no. 24 (2025): 13132–44. https://doi.org/10.1021/acs.jcim.5c02136.

(3) K. Raha, M.B. Peters, B. Wang, N. Yu, A.M. Wollacott, L.M. Westerhoff, and K.M. Merz, The role of quantum mechanics in structure-based drug design. *Drug Discov Today*. 2007 Sep;12(17-18):725-31. doi: 10.1016/j.drudis.2007.07.006.

(4) A. Anighoro, (2020). Underappreciated Chemical Interactions in Protein–Ligand Complexes. In: Heifetz, A. (eds) Quantum Mechanics in Drug Discovery. Methods in Molecular Biology, vol 2114. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-0282-9_5

APPENDIX

Supplemental Material (Walker, M.G., Mendez, C.G., Ho, A.N., Czarny, R.S., and Ho, P.S.)[4]

1. **Supplemental Figure S3.1.** DNA cutting activity of the nonhalogenated E136Y mEndoG construct compared to the enzyme with E136 replaced by a *para*-iodophenylalanine (^{I}F-*m*EndoG).
2. **Supplemental Figure S3.2.** Fluorescence polarization titration to determine dissociation constants (K_D) of the *meta*-iodotyrosyl mouse endonuclease G (^{mI}Y-*m*EndoG) construct.
3. **Supplemental Table S3.1.** Dissociation constants (K_D) of wild type (WT) and various halogenated and unhalogenated mouse endonuclease G (*m*EndoG) constructs with DNA Holliday junctions, as determined by fluorescence polarization assays.
4. **Supplemental Figure S3.3.** Effect of the metal chelator ethylenediaminetetraacetic acid (EDTA) on DNA cutting activity of ^{mI}Y-*m*EndoG *versus* WT-mEndoG under basic conditions.
5. **Supplemental Figure S3.4.** *p*H Dependence of DNA cleavage rates for ^{mI}Y-*m*EndoG.
6. **Supplemental Figure S3.5.** Denaturing PAGE gel analysis on effects of EDTA (0 mM, 1 mM and 2 mM concentrations) on DNA cutting activity of WT-*m*EndoG and ^{mX}Y-*m*EndoG (X = I or Cl) under neutral (*p*H = 7.0) conditions
7. **Supplemental Figure S3.6.** Quantum mechanical (QM) analysis of electrostatic potentials (ESP) of models for WT- and ^{mX}Y-*m*EndoG active site residues.

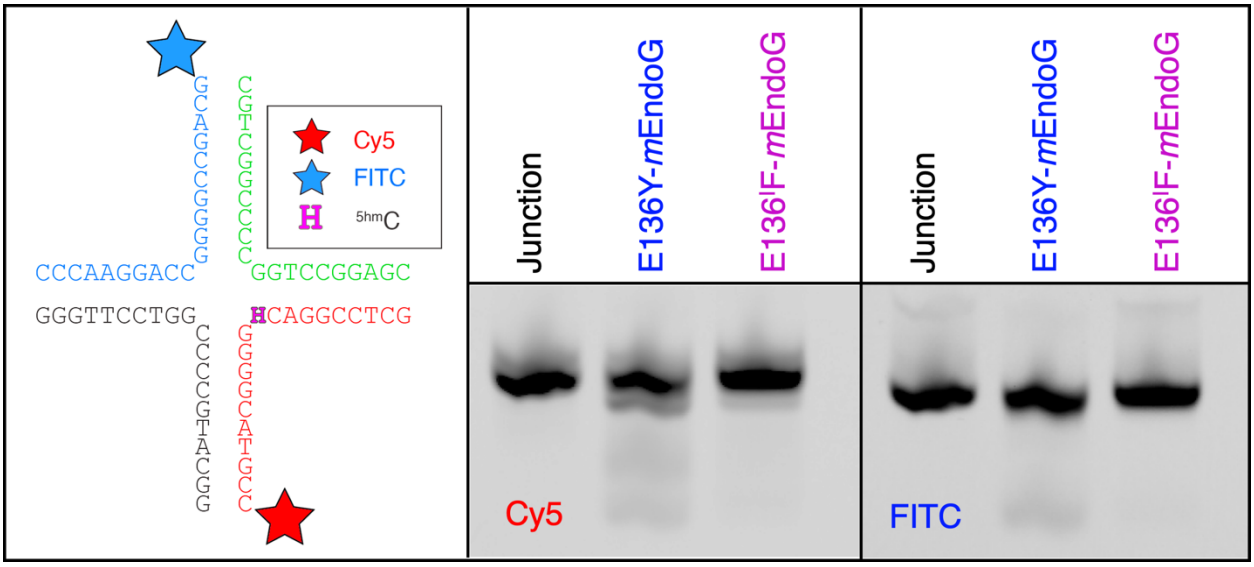
[4] The supplementary figures in this section are from the research article published in PNAS.
M.G. Walker, C.G. Mendez, A.N. Ho, R.S. Czarny, A.K. Rappé, & P.S. Ho, Design of a halogen bond catalyzed DNA endonuclease, *Proc. Natl. Acad. Sci. U.S.A.* 122 (14) e2500099122, https://doi.org/10.1073/pnas.2500099122 (2025).

Supplemental Material (Walker, M.G., Ho, E.N., Govindarajan, V., Rappé, A.K., and Ho, P.S.)[5]

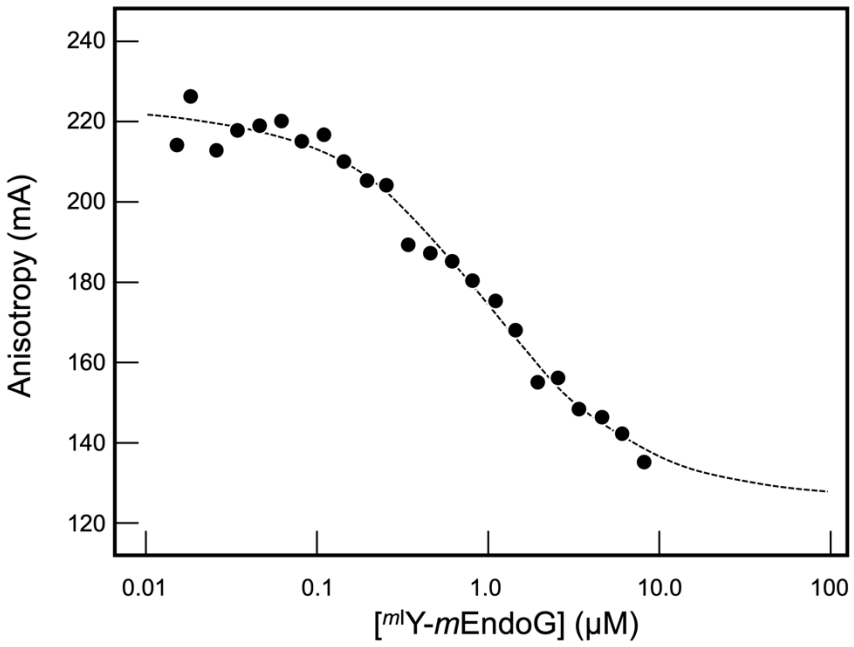
1. **Supplemental Table S4.1.** Statistics of all bXB interactions, including number of interactions, and averages and standard deviations (in parentheses) for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.
2. **Supplemental Table S4.2.** Statistics for bXB acceptors in the bXB-Only dataset, including number of interactions, and averages and standard deviations (in parentheses) for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.
3. **Supplemental Table S4.3.** Statistics for XB acceptors in the HB*e*XB dataset, including number of interactions, and averages and standard deviations (in parentheses) for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.
4. **Supplemental Table S4.4.** Statistics for HB donors in the Cl-HB*e*XB dataset , including number of interactions, and averages and standard deviations for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.
5. **Supplemental Table S4.5.** Statistics for HB donors in the Br-HB*e*XB dataset, including number of interactions, and averages and standard deviations for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.

[5] The tables in this section were written, formatted, and submitted as a research article from Chapter 4.

6. **Supplemental Table S4.6.** Statistics for HB donors in the I-HB*e*XB dataset, including number of interactions, and averages and standard deviations for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.



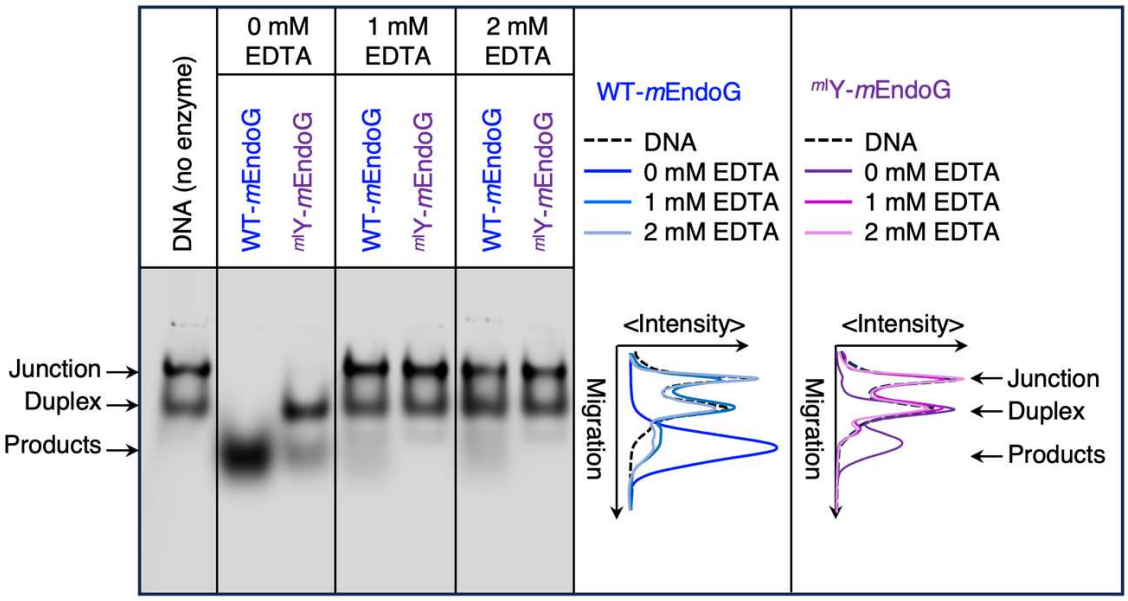
Supplemental Figure S3.1. DNA cutting activity of the nonhalogenated E136Y-*m*EndoG construct compared to the enzyme with E136 replaced by a *para*-iodophenylalanine (^{I}F-*m*EndoG). Left-hand panel shows junction construct for the assay, showing the DNA strands labeled with the Cy5 (red star) or fluorescene (FITC, blue star). The DNA cleavage products were analyzed by a denaturing polyacrylamide gel after 90 minutes of reaction at *p*H6.5.



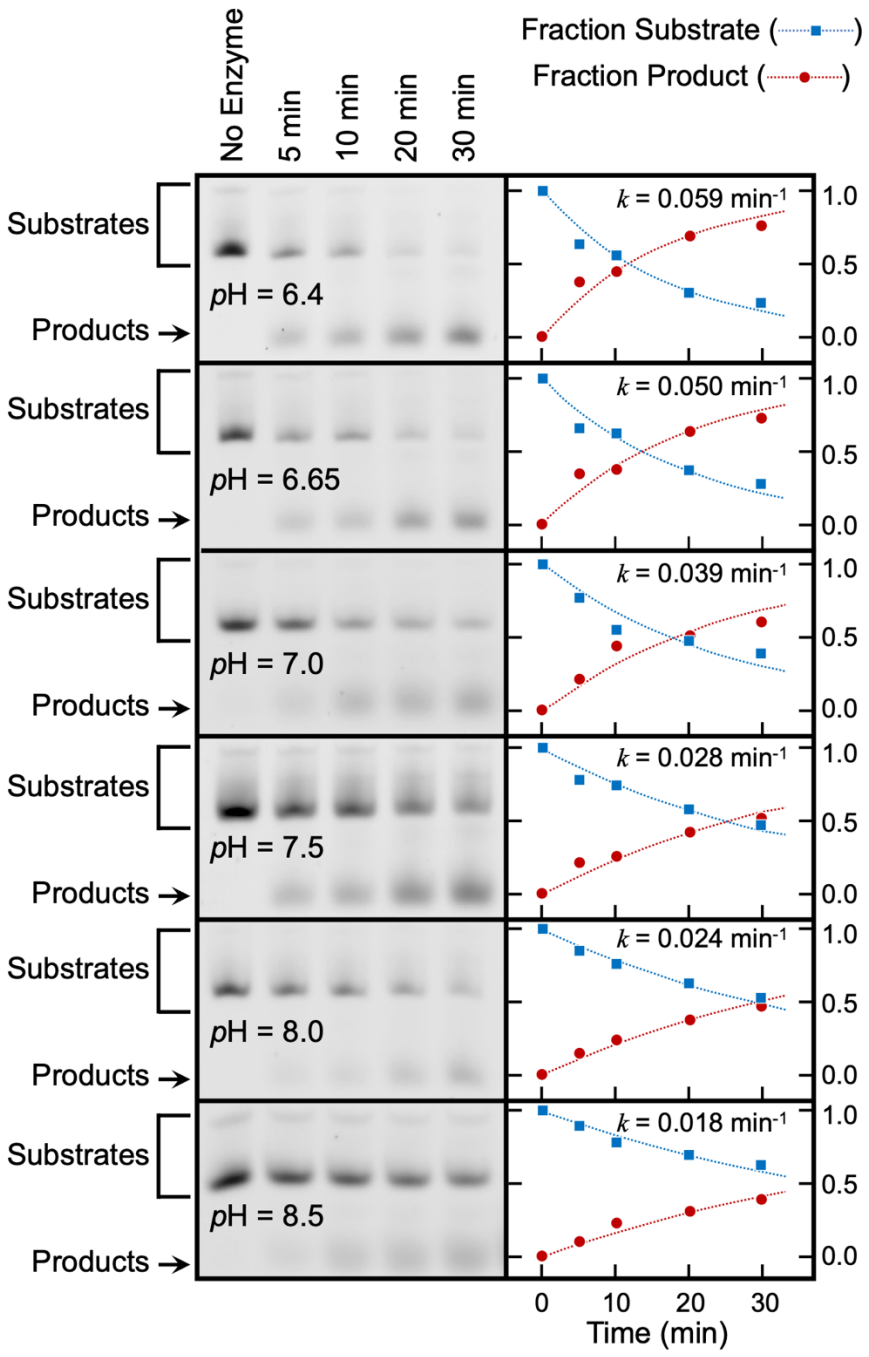
Supplemental Figure S3.2. Fluorescence polarization titration to determine dissociation constants (K_D) of the *meta*-iodotyrosyl mouse endonuclease G (^{mI}Y-*m*EndoG) construct. See Materials and Methods section of main text for conditions.

Supplemental Table S3.1. Dissociation constants (K_D) of wild type (WT) and various halogenated and unhalogenated mouse endonuclease G (*m*EndoG) constructs with DNA Holliday junctions, as determined by fluorescence polarization assays.

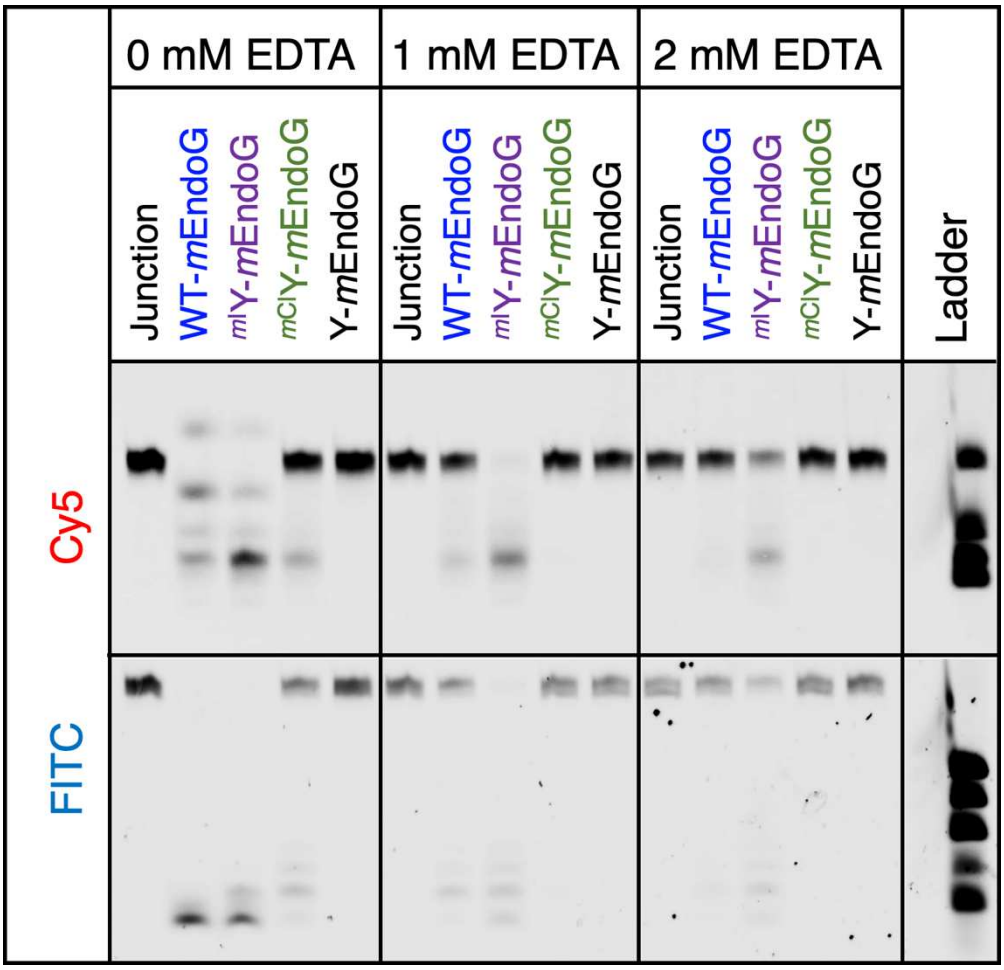
Construct	$\boldsymbol{K_D}$ **(nM)**
WT-*m*EndoG	57 ± 9
Y-*m*EndoG	490 ± 75
^{mI}Y-*m*EndoG	980 ± 191
^{mCl}Y-*m*EndoG	390 ± 124



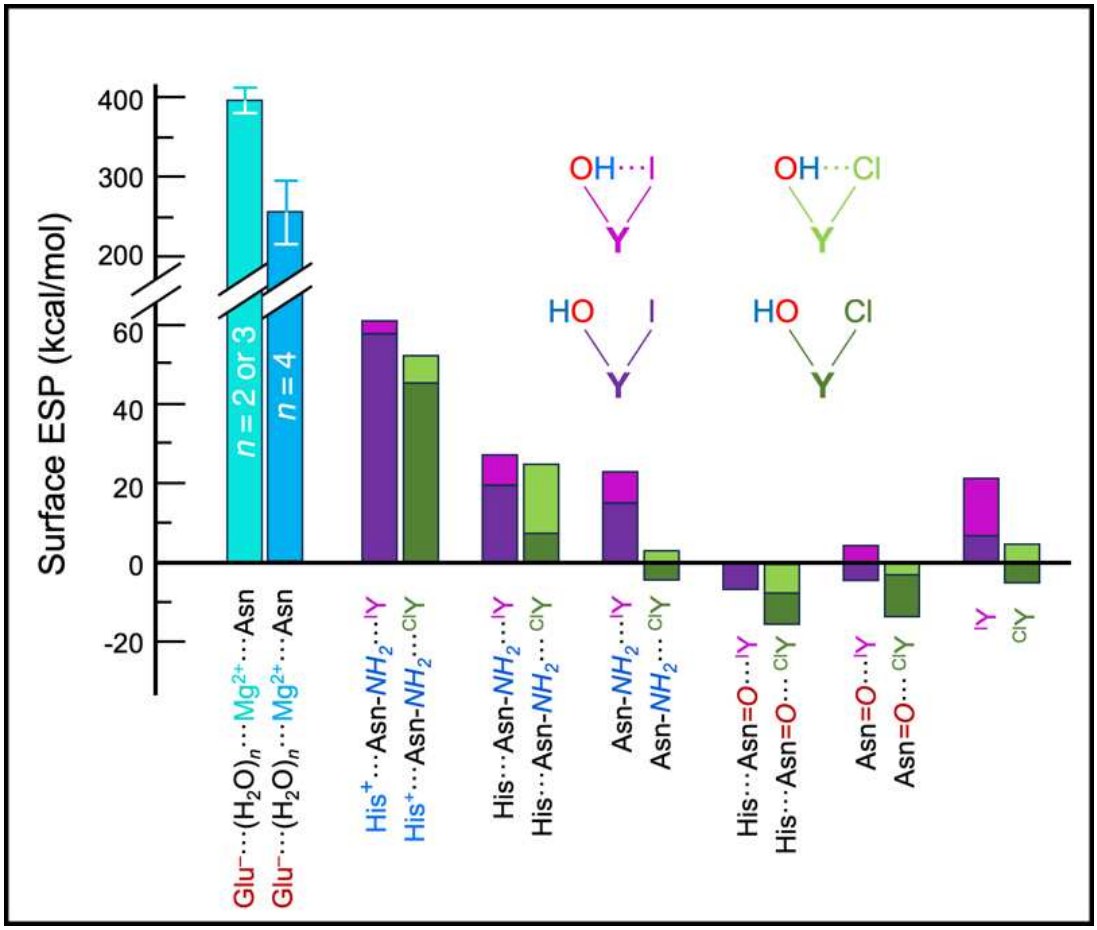
Supplemental Figure S3.3. Effect of the metal chelator ethylenediaminetetraacetic acid (EDTA at 0 mM, 1 mM and 2 mM concentrations) on DNA cutting activity after 2 hours incubation with WT-*m*EndoG or ^{mI}Y-*m*EndoG under basic (*p*H 8.5) conditions. Normalized intensity (<Intensity>) profiles of the PAGE analyses for WT-*m*EndoG (blue lines) and ^{mI}Y-*m*EndoG (magenta lines) are shown in the two right-hand panels. Black dashed lines are the profiles from the DNA only (far left, no enzyme) lane of the gel.



Supplemental Figure S3.4. *p*H Dependence of DNA cleavage rates for ^{ml}Y-*m*EndoG. Data from one example titration experiment to determine k_{APP} for ^{ml}Y-*m*EndoG. Scans of intensities from gels with no enzyme (representing the 0 time point) and after 5, 10, 20 and 30 minutes of reaction (left panels) are normalized and plotted as fractions as substrate (red circles) and products (blue squares) for each time period (right panels). The fractions for each time period were treated as pseudo-first-order reactions to determine the k_{APP} values at each *p*H.



Supplemental Figure S3.5. Denaturing PAGE gel analysis on effects of EDTA (0 mM, 1 mM and 2 mM concentrations) on DNA cutting activity of WT-*m*EndoG and ^{mX}Y-*m*EndoG (X = I or Cl) under neutral (*p*H = 7.0) conditions. The slow migrating bands in the 0 mM WT-*m*EndoG and ^{mX}Y-*m*EndoG Cy5 lanes of the PAGE gel are small fragments that migrate anomalously due to the positive charge of the Cy5 label (Killelea, *et al.*, Anomalous electrophoretic migration of short oligodeoxynucleotides labelled with 5′-terminal Cy5 dyes, *Electrophoresis*, **35**(14): 1938–1946, 2014).



Supplemental Figure S3.6. Quantum mechanical (QM) analysis of electrostatic potentials (ESP) of models for WT- and ^{mX}Y-*m*EndoG active site residues. In this analysis, the top lighter colored bar associated with the ESP calculation for each ^{mX}Y-*m*EndoG model shows the additional energy associated with the tyrosyl OH rotated towards and H-bonded to the halogen. Comparison of ESP values for the Mg^{2+} cofactor in WT-*m*EndoG to those of the halogens in models for the ^{mI}Y-*m*EndoG active site (see main text for description for how each model was constructed). The complex included either all four waters from the crystal structure (cyan bar), or with waters removed (leaving 2 or 3 coordinated solvent molecules, blue bar) that would allow coordination of the Mg^{2+} to the phosphate oxygens of the DNA substrate. ESP were calculated for models of ^{mCl}Y (green bars) and ^{mI}Y (magenta bars) residues alone, in complex with the carbonyl oxygen (N=O) or the amino group (NH_2) of acetamide (model for the N103 side chain) pointed towards the halogen, with neutral imidazole (model for the H104 side chain, labeled H) or protonated imidazole (H^+) in the network. Values were plotted for the most positive point on the respective ESP surfaces.

Supplemental Table S4.1. Statistics of all bXB interactions, including number of interactions, and averages and standard deviations (in parentheses) for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.

XB Donor	XB Acceptor	# XB Interactions (821 Total)	Average XB $f(\sum R_{vdW})$ (0.96)	Average θ_1 (148.3°)	Average θ_2 (116.0°)
Cl	***Total:***	***521***	***0.97 (0.06)***	***147.6° (14.7°)***	***114.8° (20.1°)***
	=O	229	0.97 (0.06)	151.4° (14.4°)	114.0° (16.6°)
	O(H)	70	0.92 (0.08)	146.1° (15.4°)	126.4° (19.8°)
	N	154	0.99 (0.05)	143.3° (12.9°)	108.1° (21.2°)
	S	68	0.97 (0.07)	145.9° (15.7°)	120.6° (22.1°)
Br	***Total:***	***193***	***0.96 (0.06)***	***148.6° (15.3°)***	***117.3° (21.4°)***
	=O	101	0.96 (0.05)	153.3° (14.5°)	120.6° (18.1°)
	O(H)	21	0.97 (0.06)	152.9° (8.3°)	102.8° (14.1°)
	N	34	0.98 (0.04)	140.5° (13.6°)	104.5° (14.1°)
	S	38	0.94 (0.07)	141.2° (14.5°)	127.5° (28.2°)
I	***Total:***	***107***	***0.93 (0.06)***	***151.3° (16.0°)***	***119.8° (15.8°)***
	=O	69	0.92 (0.05)	154.5° (16.2°)	120.7° (15.4°)
	O(H)	17	0.93 (0.05)	149.6° (14.6°)	125.3° (13.7°)
	N	15	0.94 (0.06)	139.7° (13.9°)	106.9° (11.1°)
	S	6	0.98 (0.02)	148.7° (6.0°)	125.2° (19.6°)

XB acceptors labeled as "=O" are carbonyl oxygens, primarily from the peptide backbone and to a lesser extent from the side chains of Asp, Asn, Glu and Gln. The "O(H)" acceptors are the hydroxyl sidechains of Ser, Thr and Tyr, "N" are the amino and imino nitrogens of the peptide backbone and various nitrogen containing amino acid side chains, while "S" are to the sulfur atoms in the side chains of Cys and Met.

Supplemental Table S4.2. Statistics for bXB acceptors in the bXB-Only dataset, including number of interactions, and averages and standard deviations (in parentheses) for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.

XB-Only Donor	XB-Only Acceptor	#XB-Only Interactions (419)	Average XB $f(\sum R_{vdW})$ (0.97)	Average θ_1 (148.9°)	Average θ_2 (112.8°)
Cl	***Total:***	***277***	***0.97 (0.07)***	***147.9° (15.4°)***	***111.0° (21.2 °)***
	=O	103	0.99 (0.06)	154.2° (14.8°)	113.5° (18.8°)
	O(H)	35	0.92 (0.08)	146.5° (15.6°)	122.2° (22.8°)
	N	97	0.99 (0.05)	141.6° (13.0°)	102.7° (20.4°)
	S	39	0.97 (0.07)	148.6° (16.4°)	116.6° (20.4°)
Br	***Total:***	***64***	***0.98 (0.04)***	***149.0° (15.0°)***	***113.0° (20.6°)***
	=O	40	0.97 (0.04)	155.4° (13.5°)	119.0° (18.6°)
	O(H)	0	–	–	–
	N	14	1.00 (0.03)	142.8° (12.0°)	103.0° (10.5°)
	S	10	0.99 (0.03)	132.4° (6.2°)	102.9° (29.4°)
I	***Total:***	***78***	***0.94 (0.06)***	***152.4° (15.7°)***	***118.9° (16.4°)***
	=O	54	0.92 (0.05)	154.3° (15.9°)	120.1° (15.4°)
	O(H)	10	0.94 (0.04)	150.8° (17.6°)	125.3° (16.8°)
	N	10	0.93 (0.08)	143.2° (15.4°)	104.3° (11.1°)
	S	5	0.98 (0.02)	149.7° (6.2°)	123.6° (21.5°)

XB acceptors labeled as "=O" are carbonyl oxygens, primarily from the peptide backbone and to a lesser extent from the side chains of Asp, Asn, Glu and Gln. The "O(H)" acceptors are the hydroxyl sidechains of Ser, Thr and Tyr, "N" are the amino and imino nitrogens of the peptide backbone and various nitrogen containing amino acid side chains, while "S" are the sulfurs in the side chains of Cys and Met.

Supplemental Table S4.3. Statistics for XB acceptors in the HB*e*XB dataset, including number of interactions, and averages and standard deviations (in parentheses) for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.

bXB-Only Donor	bXB-Only Acceptor	#bXB-Only Interactions (402 total)	Average XB $f(\sum R_{vdW})$ (0.96)	Average θ_1 (147.7°)	Average θ_2 (119.4°)
Cl	***Total:***	***244***	***0.97 (0.06)***	***147.2° (13.8°)***	***119.1° (17.9 °)***
	=O	126	0.97 (0.05)	149.1° (13.6°)	114.3° (14.8°)
	O(H)	35	0.92 (0.07)	145.8° (15.4°)	130.6° (15.5°)
	N	57	0.99 (0.05)	146.2° (12.3°)	117.4° (19.4°)
	S	26	0.97 (0.06)	142.3° (14.8°)	130.5° (20.4°)
Br	***Total:***	***129***	***0.96 (0.06)***	***148.4° (15.1°)***	***119.4° (21.5°)***
	=O	80	0.96 (0.05)	152.1° (13.8°)	116.9° (19.0°)
	O(H)	19	0.89 (0.06)	151.3° (12.7°)	130.4° (19.2°)
	N	21	0.97 (0.05)	139.7° (14.9°)	106.3° (16.2°)
	S	10	0.98 (0.06)	131.6° (10.3°)	145.7° (24.5°)
I	***Total:***	***29***	***0.92 (0.05)***	***148.3° (16.4°)***	***122.0° (13.8°)***
	=O	16	0.92 (0.05)	153.5° (18.4°)	122.8° (15.1°)
	O(H)	7	0.91 (0.06)	148.0° (10.0°)	125.4° (8.9°)
	N	5	0.94 (0.04)	132.7° (7.2°)	112.1° (13.6°)
	S	1	0.97	143.7°	133.4°

XB acceptors labeled as "=O" are carbonyl oxygens, primarily from the peptide backbone and to a lesser extent from the side chains of Asp, Asn, Glu and Gln. The "O(H)" acceptors are the hydroxyl sidechains of Ser, Thr and Tyr, "N" are the amino and imino nitrogens of the peptide backbone and various nitrogen containing amino acid side chains, while "S" are the side chains of Cys and Met.

Supplemental Table S4.4. Statistics for HB donors in the Cl-HB*e*XB dataset , including number of interactions, and averages and standard deviations for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.

XB Donor	XB Acceptor	HB Donor	#HB*e*XB Interactions (339 total)	Average XB $f(\sum R_{vdW})$	Average θ_1	Average θ_2	Average HB $f(\sum R_{vdW})$	Average η_1	Average η_2
Cl	***Total***		***244***	***0.97 (0.06)***	***147.2° (13.8°)***	***119.1° (17.9 °)***	***0.92 (0.07)***	***86.8° (16.3°)***	***117.9° (23.0°)***
	O	*Total*	*161*	*0.96 (0.06)*	*148.4° (14.0°)*	*117.9° (16.4°)*	*0.91 (0.07)*	*88.6° (16.2°)*	*115.3° (20.8°)*
		–O(H)	94	0.95 (0.06)	148.1° (15.1°)	118.7° (18.3°)	0.92 (0.07)	89.4° (14.9°)	115.4° (16.8°)
		–S(H)	2 (to same acceptor)	1.02 -	166.8° -	125.1° -	0.85 (0.05)	98.2° (16.1°)	136.6° (10.2°)
		–N(H)	37	0.97 (0.06)	143.5° (11.8°)	116.4° (13.9°)	0.88 (0.07)	82.4° (16.2°)	105.5° (23.3°)
		–$N(H)^+$	28	1.00 (0.04)	154.6° (9.7°)	116.4° (12.3°)	0.90 (0.07)	93.3° (15.5°)	126.5° (23.3°)
	N	*Total*	*58*	*0.99 (0.05)*	*146.2° (12.4°)*	*117.4° (19.4°)*	*0.93 (0.08)*	*78.1° (15.0°)*	*127.5° (24.1°)*
		–O(H)	37	1.00 (0.03)	148.8° (13.7°)	118.5° (18.6°)	0.94 (0.08)	76.0° (13.8°)	131.4° (15.2°)
		–S(H)	1	0.96 -	167.1° -	115.4° -	0.98 -	75.7° -	143.2° -
		–N(H)	14	0.95 (0.06)	142.5° (6.7°)	113.0° (20.4°)	0.95 (0.08)	72.0° (6.5°)	125.3° (31.8°)
		–$N(H)^+$	6	0.98 (0.06)	138.4° (9.0°)	120.7° (23.3°)	0.89 (0.04)	105.2° (4.0°)	108.5° (40.7°)
	S	*Total*	*26*	*0.97 (0.06)*	*142.3° (14.8°)*	*130.5° (20.4°)*	*0.92 (0.07)*	*94.7° (16.1°)*	*112.6° (28.3°)*
		–O(H)	7	0.99 (0.03)	154.7° (8.5°)	119.3° (11.0)	0.94 (0.05)	105.9° (14.7°)	98.9° (15.0°)
		–S(H)	5	0.91 (0.04)	129.1° (10.2°)	143.0° (13.3°)	0.90 (0.02)	99.6° (22.6°)	118.7° (21.0°)
		–N(H)	5	0.97 (0.07)	153.2° (5.7°)	123.4° (7.0°)	0.83 (0.03)	83.6° (5.1°)	149.9° 38.0°)
		–$N(H)^+$	9	0.98 (0.07)	133.7° (13.0°)	136.2° (28.5°)	0.98 (0.06)	89.5° (12.0°)	99.2° (11.6°)

HB donating amino acids include some or all of the following in for each type of halogen XB donor: -O(H): Ser, Thr, Tyr; -S(H): Cys; -N(H): peptide backbone, Asn, Gln, Trp; -$N(H)^+$: Arg, His, Lys.

Supplemental Table S4.5. Statistics for HB donors in the Br-HB*e*XB dataset, including number of interactions, and averages and standard deviations for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.

XB Donor	XB Acceptor	HB Donor	#HB*e*XB Interactions	Average XB $f(\sum R_{vdW})$	Average θ_1	Average θ_2	Average HB $f(\sum R_{vdW})$	Average η_1	Average η_2
Br	***Total***		***66***	***0.957 (0.067)***	***150.8° (12.8°)***	***115.9° (21.1°)***	***0.879 (0.135)***	***94.6° (16.6°)***	***120.4° (25.4°)***
	O	*Total*	*100*	*0.96 (0.06)*	*153.0° (13.5°)*	*119.3° (19.3°)*	*0.867 (0.148)*	*94.5° (16.0°)*	*119.7° (25.3°)*
		–O(H)	39	0.95 (0.05)	154.7° (15.9°)	119.7° (19.0°)	0.93 (0.09)	90.8° (19.5°)	113.0° (24.1°)
		–S(H)	12	0.90 (0.06)	150.9° (9.6°)	127.2° (9.0°)	0.98 (0.05)	98.8° (7.1°)	130.0° (17.7°)
		–N(H)	16	0.96 (0.06)	145.0° (8.7°)	122.9° (21.7°)	0.95 (0.11)	97.5° (18.2°)	112.1° (26.4°)
		$-N(H)^+$	33	0.97 (0.05)	151.8° (12.8°)	114.9° (21.5°)	0.91 (0.09)	93.4° (16.5°)	118.2° (22.7°)
	N	*Total*	*16*	*0.97 (0.05)*	*138.9° (14.7°)*	*105.6° (16.3°)*	*0.93 (0.07)*	*91.4° (17.8°)*	*116.9° (26.5°)*
		–O(H)	4	0.97 (0.05)	130.7° (4.6°)	109.5° (13.2°)	0.88 (0.09)	86.2° (18.3°)	116.5° (19.4°)
		–S(H)	0	–	–	–	–	–	–
		–N(H)	2	0.90 (0.02)	144.7° (26.4°)	116.1° (7.8°)	0.96 -	117.3 -	146.0° -
		$-N(H)^+$	10	0.97 (0.06)	144.9° (16.3°)	100.4° (18.8°)	0.95 (0.05)	90.4° (15.3°)	109.3° (30.6°)
	S	*Total*	9	0.99 (0.06)	129.7° (8.9°)	148.75° (24.0°)	1.00 (0.09)	75.8° (19.3°)	120.4° (28.2°)
		–O(H)	8	1.00 (0.04)	127.3° (5.5°)	150.9° (24.7°)	1.02 (0.08)	72.9° (19.0°)	124.5° (27.2°)
		–S(H)	0	–	–	–	–	–	–
		–N(H)	0	–	–	–	–	–	–
		$-N(H)^+$	1	0.84	149.1°	131.0°	0.85	98.8°	87.7°

HB donating amino acids include some or all of the following in for each type of halogen XB donor: -O(H): Ser, Thr, Tyr; -S(H): Cys; -N(H): peptide backbone, Asn, Gln, Trp; $-N(H)^+$: Arg, His, Lys.

Supplemental Table S4.6. Statistics for HB donors in the I-HB*e*XB dataset, including number of interactions, and averages and standard deviations for fraction of the sum of the van der Waals radii of the interacting atoms [$f(\sum R_{vdW})$] and interaction angles (see Figure 1) from the survey of the PDB.

XB Donor	XB Acceptor	HB Donor	#HB*e*XB Interactions	Average XB $f(\sum R_{vdW})$	Average θ_1	Average θ_2	Average HB $f(\sum R_{vdW})$	Average η_1	Average η_2
I	***Total***		***29***	***0.92 (0.05)***	***148.3° (16.4°)***	***122.0° (13.8°)***	***0.93 (0.08)***	***89.0° (17.5°)***	***125.3° (21.2°)***
	O	*Total*	23	0.92 (0.05)	151.8° (16.3°)	123.6° (13.4°)	0.92 (0.08)	86.9° (17.3°)	128.5° (19.5°)
		–O(H)	8	0.90 (0.05)	136.3° (15.2°)	119.0° (8.3°)	0.91 (0.08)	87.5° (20.6°)	123.6° (14.3°)
		–S(H)	5	0.88 (0.06)	163.0° (9.0°)	127.7° (5.6°)	0.93 (0.08)	88.9° (21.5°)	120.3° (9.2°)
		–N(H)	0	–	–	–	–	–	–
		$-N(H)^+$	10	0.95 (0.04)	158.7° (9.9°)	125.3° (18.4°)	0.92 (0.09)	85.3° (13.9°)	136.4° (24.6°)
	N	*Total*	*5*	*0.94 (0.04)*	*132.7° (7.2°)*	*112.1° (13.6°)*	*0.95 (0.05)*	*96.4° (18.1°)*	*114.2° (27.5°)*
		–O(H)	3	0.94 (0.04)	134.9° (9.2°)	108.5° (16.8°)	0.97 (0.06)	84.0° (7.6°)	114.2° (20.9°)
		–S(H)	0	–	–	–	–	–	–
		–N(H)	0	–	–	–	–	–	–
		$-N(H)^+$	2	0.95 (0.04)	129.5° (1.1°)	117.5° (9.0°)	0.93 (0.06)	115.1° (5.2°)	114.4° (46.3°)
	S	*Total*	*0*	–	–	–	–	–	–
		–O(H)	0	–	–	–	–	–	–
		–S(H)	0	–	–	–	–	–	–
		–N(H)	0	–	–	–	–	–	–
		$-N(H)^+$	0	–	–	–	–	–	–

HB donating amino acids include some or all of the following in for each type of halogen XB donor: -O(H): Ser, Thr, Tyr; -S(H): Cys; -N(H): peptide backbone, Asn, Gln, Trp; -N(H)+: Arg, His, Lys.