Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 8, No.2, April 2026, pp. 02602010-01 ~ 02602010-06
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v8i2.2757

Analysis for Non-covalent Bonds in Protein Structures Using the DV-Xα Method

Takeru Kajiyama[1], Kota Aramaki[1], Kohei Kanazaki[1], Ayami Matsushima[2*]

[1]Department of Chemistry, Graduate School of Science, Kyushu University, Fukuoka, 819-0395 Japan

[2]Department of Chemistry, Faculty of Science, Kyushu University, Fukuoka, 819-0395 Japan

ayami@chem.kyushu-univ.jp

Abstract. Halogen atoms are increasingly recognized for their ability to form non-covalent halogen bonds in protein–ligand complexesy. To systematically evaluate these interactions, we employed the DV-Xα method to calculate bond overlap populations (BOPs) in protein structures containing halogenated ligands. Using the Protein Data Bank, we identified thousands of entries with fluorine, chlorine, bromine, or iodine atoms. Structural coordinates were extracted around the chlorine atoms of the ligand and subjected to *ab initio* calculations under multiple truncation schemes. The findings highlight the utility of DV-Xα calculations in characterizing weak interactions and provide insights for future computational drug design employing halogen-containing compounds.

Keywords: computational chemistry, DV-Xα method, non-covalent bond, molecular orbital calculation

(Received 2025-09-15, Revised 2025-11-11, Accepted 2026-01-30, Available Online by 2026-03-17)

1. Introduction

Many biologically active ligands, such as hormones and pharmaceuticals, contain halogen atoms. In biological systems, hydrogen bonds are crucial for protein conformation and interactions. Halogen atoms also exhibit a weaker interaction known as halogen bonds, which share similarities with hydrogen bonds.

Halogen bonds have garnered significant attention as an interaction between receptor proteins and ligands [1][2]. A halogen bond is an interaction between a Lewis acid (the halogen atom) and a Lewis base, and it's particularly important when the halogen atom acts electrophilically. Conversely, halogen

atoms can also act nucleophilically. This diverse range of interactions stems from the charge bias induced by atoms covalently bonded to the halogen. Since halogen atoms aren't naturally found in the amino acids composing in vivo proteins, halogen bonds in proteins and other biomolecules often form between the oxygen atoms of protein-constituent amino acids and the halogen atoms of small molecule ligands. Recently, the Protein Data Bank (PDB, https://www.rcsb.org) has accumulated over 230,000 protein X-ray crystallographic and NMR structural analysis datasets, along with more than 990,000 protein structure prediction data from AlphaFold [3]. By leveraging these structural data to rationally incorporate halogen atoms into compounds, it becomes possible to effectively design drugs that utilize non-covalent bond to improve selectivity and binding affinity [4-14].

We previously conducted binding assays of environmental chemicals with proteins and evaluated its crystal structures. A large number of environmental chemicals contain halogen atoms [15]. While numerous biologically active substances, including anesthetics, psychotropic drugs, and even poisons like dioxin, contain fluorine or chlorine, the only naturally occurring halogen-containing compound in living organisms is thyroid hormone, which contains iodine [16-19]. Therefore, developing halogen (*i.e*, fluorine, chlorine, bromine, and iodine)-containing drugs could enable the design of compounds exhibiting novel and beneficial in vivo interactions.

2. **Methods**

2.1 Searching Structures

We searched for crystal structures registered in the Protein Data Bank (PDB) that contain a halogen atom in the ligand. We used the SQL search function provided by the Protein Data Bank Japan (PDBj) to perform the structural search [20-22].

*First, we searched for structures where the ligand's composition included a fluorine atom using the SQL query: "SELECT * FROM chem_comp WHERE formula LIKE '%F%'". Because the condition `"formula LIKE '%F%'"$ unintentionally includes ligands with iron (Fe) in their composition, we removed structures whose ligands contained only iron and no fluorine from the search results. Next, we searched for structures where the ligand's composition included a chlorine atom using the SQL query: "SELECT * FROM chem_comp WHERE formula LIKE '%Cl%'".*

*Then, we searched for structures where the ligand's composition included a bromine atom using the SQL query: "SELECT * FROM chem_comp WHERE formula LIKE '%Br%'". Finally, we searched for structures where the ligand's composition included an iodine atom using the SQL query: "SELECT * FROM chem_comp WHERE formula LIKE '%I%'". Because the condition `"formula LIKE '%I%'"$ unintentionally includes ligands with indium (In) or iridium (Ir) in their composition, we removed structures whose ligands contained only indium or iridium and no iodine from the search results.*

2.2 Preparation of Crystal Structures for Calculation

We performed preprocessing on the crystal structures for use in subsequent calculations. First, ligands that did not contain halogen atoms and water molecules—which were not the subjects of the calculation—were removed from the structures. Since PDB crystal structures often lack hydrogen atoms, we created structures with added hydrogen atoms. We used the Molecular Operating Environment (MOE), developed by Chemical Computing Group ULC., for hydrogen addition [23]. We used MOE 2024.0601 for this study. Hydrogen atoms were added using MOE's Protonate3D function. MOE's Protonate3D function explores the most stable state from the possible hydrogen atom addition states of the three-dimensional structure using the Unary Quadratic Optimization algorithm. To prepare the structures for calculation, we extracted the amino acid residues within a 4.0 Å radius of the halogen atom in the ligand, ensuring that the termini of the peptide backbone were capped as amides. We performed this same operation at 0.5 Å intervals.

2.3 Evaluation of Non-covalent bond

Bond overlap populations (BOPs) were calculated by DV-Xα method. The BOP is an index used in quantum chemistry calculations to quantitatively assess the nature and strength of a specific interatomic bond within a molecule. This value is calculated based on the overlap integral of the Atomic Orbitals (AO) that constitute the Molecular Orbitals (MO), as well as the contribution (coefficients) of each AO to the MOs. The total bond overlap population between atoms A and B, is obtained by summing these values over all pairs of AOs on atom A and AOs on atom B. We performed a first-principles calculation based on the DV-Xα method on the prepared structures to calculate the BOP [24].

3. Results and Discussion

3.1. Detabase Search

We utilized the deposited protein structures to PDB database(https://www.rcsb.org) for *ab initio* calculation [3]. As of September 15, 2025, the PDB database (URL) contains 197,747 X-ray crystallographic structure data entries and 999,251 AlphaFold-predicted structural data entries. For our calculations, we decided to use only the X-ray crystallographic structures. Among these structures, 11,661 contained fluorine atoms, 28,036 contained chlorine atoms, 3,077 contained bromine atoms, and 1,926 contained iodine atoms. Of these, the number of structures bound to DNA was 196 for fluorine-containing, 20 for chlorine-containing, 231 for bromine-containing, and 57 for iodine-containing cases. In addition, there were 100 structural data entries in which fluorine atoms, 20,114 entries in which chlorine atoms, 706 entries in which bromine atoms, and 1,222 entries in which iodine atoms were considered to be present as counterions (Table 1). In this study, we focused on the structure of the estrogen receptor, a female hormone receptor that has been attracting attention as a target of environmental chemicals. Among environmental chemicals, bisphenol A is widely used as a raw material for plastics, and its safety has raised considerable concern. Consequently, various derivatives of bisphenol A have been employed, one of which is bisphenol C, a compound containing chlorine atoms. The estrogen receptor (ER), which is a receptor for the female hormone estrogen, has in two isoforms: the ERα and the ERβ. We identified the crystal structure of ERα bound to bisphenol C, which is deposited under PDB ID: 3UUC [25][26]. This structure was selected for further use in our calculations.

Table 1. The number of protein crystal structures containing halogen atoms

Halogen	All	As counterions	Bound to DNA	Bound to Peptide	Bound to Saccharide
Fluorine	11,661	100	196	331	183
Chlorine	28,036	20,114	20	346	41
Bromine	3,077	706	231	87	27
Iodine	1,926	1,222	57	118	2

3.2. Extraction of Structural Coordinates

For the *ab initio* calculations, we first determined the coordinates to be used. Because employing the coordinates of the entire protein structure would be computationally expensive, we chose to use only the coordinates surrounding the chlorine atom. We have previously examined which range of atoms should be extracted, and in this study, we followed the same procedure [27]. We utilized the crystal structure of the bisphenol C-bound ERα ligand-binding domain to conduct ab initio calculations aimed at uncovering the mechanisms linked to bisphenol C 's strong binding affinity for the ERα, which features two chlorine atoms. Amino acid residues in proximity to these chlorine atoms were chosen from the deposited structure for *ab initio* calculations. To enhance the accuracy of our results, we employed three compensatory methods for calculations involving the terminal regions. In type 1, hydrogen atoms were added, type 2 involved reconstructing each truncated amino acid residue as individual amino acids,

and type 3 expanded the analyzed region to include the nitrogen atom at the i-1 position with standard protonation (Figure 1). We performed ab initio calculations using coordinate data ranging from 4 Å in 0.5 Å increments for the chlorine atoms of bisphenol C in the protein complex under these three conditions. The calculated bond overlap population of covalent bonds between chlorines and the carbon atom in the complex is summarized in table 2. The bond overlap population is used to assess the covalency contribution in target bonds. The findings indicated that the compensatory method in type 3 was effective for *ab initio* calculations, as the calculated bond overlap populations were relatively stable and converged when the target regions for calculation were expanded.

Table 2. Bond overlap population between the chlorine and carbon atoms.

Range from chlorine (Å)	Type 1 Cl (1)	Type 1 Cl (2)	Type 2 Cl (1)	Type 2 Cl (2)	Type 3 Cl (1)	Type 3 Cl (2)
4.0	0.720	0.710	0.730	0.726	0.72	0.745
4.5	0.725	0.722	0.731	0.727	0.718	0.740
5.0	0.682	0.722	0.736	0.724	0.719	0.745
5.5	0.682	0.725	0.727	0.724	0.745	0.727
6.0	0.700	0.725	0.727	0.685	0.745	0.727
6.5	0.710	0.722	0.731	0.684	0.726	0.723
7.0	0.690	0.721	0.732	0.712	0.725	0.725
7.5	0.708	0.722	0.737	0.695	0.73	0.723
8.0	0.706	0.724	0.718	0.725	0.724	0.723
8.5	0.735	0.722	0.736	0.722	0.72	0.730
9.0	0.722	0.727	0.731	0.726	0.725	0.727
9.5	0.723	0.720	0.736	0.725	0.726	0.725

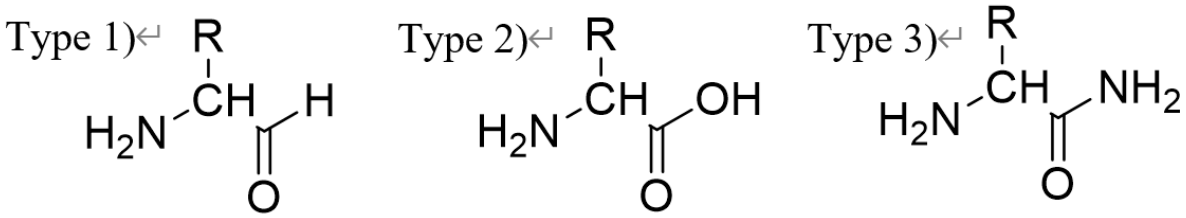

Figure 1. Three types of extraction of the structural coordinates.

3.3. Evaluation of noncovalent interactions

Because there is a possibility of a noncovalent interaction between the chlorine atom and neighboring carbon atoms, we measured the interatomic distances. As a result, the distances between the chlorine atom and the surrounding carbon atoms in the protein were 6.1, 3.8, 4.8, 5.3, 6.7 Å. The van der Waals radius of chlorine is approximately 1.82 Å, while that of carbon is 1.77 Å [28]. Although it is evident that these atoms are not covalently bonded, the possibility of an interaction cannot be excluded. Typically, in the field of structural biochemistry, it is customary to regard atoms as interacting when the distance is less than 5 Å. Considering that the resolution of the structure used for the calculations is 2.1 Å, it is reasonable to assume that some of the surrounding atoms are engaged in noncovalent interactions with the chlorine atom.

4. Conclusion

The protein crystal structure database contains numerous structures with halogen atoms, many of which are considered to interact with proteins. Future calculations utilizing these structures are expected to be

highly valuable, extending DV-Xα to larger systems or integrating it with molecular dynamics for sustainability-focused modeling.

Acknowledgments
This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant No JP23H00521 to A.M., in part by a research grant from the Mitsubishi Foundation.

References

[1] M. L. Verteramo *et al.*, "Interplay of halogen bonding and solvation in protein–ligand binding," *iScience*, vol. 27, no. 4, p. 109636, Apr. 2024, doi: 10.1016/j.isci.2024.109636.
[2] B. Inscoe, H. Rathnayake, and Y. Mo, "Role of Charge Transfer in Halogen Bonding," *J. Phys. Chem. A*, vol. 125, no. 14, pp. 2944–2953, Apr. 2021, doi: 10.1021/acs.jpca.1c01412.
[3] H. M. Berman, "The Protein Data Bank," Nucleic Acids Research, vol. 28, no. 1, pp. 235–242, Jan. 2000, doi: 10.1093/nar/28.1.235.
[4] R. Sharma, S. Sarkar, S. Chattopadhayay, J. Mondal, and P. Talukdar, "A Halogen-Bond-Driven Artificial Chloride-Selective Channel Constructed from 5-Iodoisophthalamide-based Molecules," *Angew Chem Int Ed*, vol. 63, no. 19, p. e202319919, May 2024, doi: 10.1002/anie.202319919.
[5] M. G. Walker, C. G. Mendez, A. N. Ho, R. S. Czarny, A. K. Rappé, and P. S. Ho, "Design of a halogen bond catalyzed DNA endonuclease," Proc. Natl. Acad. Sci. U.S.A., vol. 122, no. 14, p. e2500099122, Apr. 2025, doi: 10.1073/pnas.2500099122.
[6] K. Tsuji *et al.*, "Exploratory Studies of Effective Inhibitors against the SARS-CoV-2 Main Protease by Halogen Incorporation and Amide Bond Replacement," *Chem. Pharm. Bull.*, vol. 71, no. 12, pp. 879–886, Dec. 2023, doi: 10.1248/cpb.c23-00562.
[7] W. Liu *et al.*, "New Cocrystals of Antipsychotic Drug Aripiprazole: Decreasing the Dissolution through Cocrystallization," *Molecules*, vol. 26, no. 9, p. 2414, Apr. 2021, doi: 10.3390/molecules26092414.
[8] J. O'Brien, N. Melnyk, R. S. Lee, M. James, and C. Trujillo, "Computational Design of Bidentate Hypervalent Iodine Catalysts in Halogen Bond-Mediated Organocatalysis," *ChemPhysChem*, vol. 25, no. 22, p. e202400515, Nov. 2024, doi: 10.1002/cphc.202400515.
[9] K. Mu, Z. Zhu, A. Abula, C. Peng, W. Zhu, and Z. Xu, "Halogen Bonds Exist between Noncovalent Ligands and Natural Nucleic Acids," *J. Med. Chem.*, vol. 65, no. 6, pp. 4424–4435, Mar. 2022, doi: 10.1021/acs.jmedchem.1c01854.
[10] A. Docker, Y. C. Tse, H. M. Tay, A. J. Taylor, Z. Zhang, and P. D. Beer, "Anti-Hofmeister Anion Selectivity via a Mechanical Bond Effect in Neutral Halogen-Bonding [2]Rotaxanes," *Angew Chem Int Ed*, vol. 61, no. 50, p. e202214523, Dec. 2022, doi: 10.1002/anie.202214523.
[11] N. Gutiérrez-Sánchez, J. Kästner, F. Mendizábal, and S. Miranda-Rojas, "Insights into the Catalytic Mechanism and Selectivity of *S* -Adenosyl Methionine (SAM)-Dependent Fluorinase toward Carbon–Halogen Bond Formation," *J. Chem. Inf. Model.*, vol. 65, no. 13, pp. 6998–7012, July 2025, doi: 10.1021/acs.jcim.5c00544.
[12] M. U. Engelhardt et al., "Halogen Bonding on Water—A Drop in the Ocean?," *J. Chem. Theory Comput.*, vol. 20, no. 23, pp. 10716‑10730, Dec. 2024, doi: 10.1021/acs.jctc.4c00834.
[13] J. Sun, D. A. Decato, V. S. Bryantsev, E. A. John, and O. B. Berryman, "The interplay between hydrogen and halogen bonding: substituent effects and their role in the hydrogen bond enhanced halogen bond," *Chem. Sci.*, vol. 14, no. 33, pp. 8924–8935, 2023, doi: 10.1039/D3SC02348F.
[14] M. H. Daniels *et al.*, "Discovery of ATX968: An Orally Available Allosteric Inhibitor of DHX9," *J. Med. Chem.*, vol. 68, no. 9, pp. 9537–9554, May 2025, doi: 10.1021/acs.jmedchem.5c00252.
[15] M. Iwamoto *et al.*, "Bisphenol A derivatives act as novel coactivator-binding inhibitors for estrogen receptor β," *Journal of Biological Chemistry*, vol. 297, no. 5, p. 101173, Nov. 2021, doi: 10.1016/j.jbc.2021.101173.

[16] X. Yao, S. Geng, L. Zhu, H. Jiang, and J. Wen, "Environmental pollutants exposure and gestational diabetes mellitus: Evidence from epidemiological and experimental studies," *Chemosphere*, vol. 332, p. 138866, Aug. 2023, doi: 10.1016/j.chemosphere.2023.138866.

[17] H. M. Dusza, K. E. Manz, K. D. Pennell, R. Kanda, and J. Legler, "Identification of known and novel nonpolar endocrine disruptors in human amniotic fluid," *Environment International*, vol. 158, p. 106904, Jan. 2022, doi: 10.1016/j.envint.2021.106904.

[18] V. Suteau *et al.*, "Effects and risk assessment of halogenated bisphenol A derivatives on human follicle stimulating hormone receptor: An interdisciplinary study," *Journal of Hazardous Materials*, vol. 479, p. 135619, Nov. 2024, doi: 10.1016/j.jhazmat.2024.135619.

[19] K. Nowak and Ž. Jakopin, "In silico profiling of endocrine-disrupting potential of bisphenol analogues and their halogenated transformation products," *Food and Chemical Toxicology*, vol. 173, p. 113623, Mar. 2023, doi: 10.1016/j.fct.2023.113623.

[20] A. R. Kinjo *et al.*, "Protein Data Bank Japan (PDBj): updated user interfaces, resource description framework, analysis tools for large structures," *Nucleic Acids Res*, vol. 45, no. D1, pp. D282–D288, Jan. 2017, doi: 10.1093/nar/gkw962.

[21] A. R. Kinjo *et al.*, "New tools and functions in data-out activities at Protein Data Bank Japan (PDBj)," *Protein Science*, vol. 27, no. 1, pp. 95–102, Jan. 2018, doi: 10.1002/pro.3273.

[22] G. Bekker *et al.*, "Protein Data Bank Japan: Celebrating our 20th anniversary during a global pandemic as the Asian hub of three dimensional macromolecular structural data," *Protein Science*, vol. 31, no. 1, pp. 173–186, Jan. 2022, doi: 10.1002/pro.4211.

[23] *Molecular Operating Environment (MOE)*, 2024.0601 Chemical Computing Group ULC, 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, **2025**.

[24] H. Adachi, M. Tsukada, and C. Satoko, "Discrete Variational Xα Cluster Calculations. I. Application to Metal Clusters," *J. Phys. Soc. Jpn.*, vol. 45, no. 3, pp. 875–883, Sept. 1978, doi: 10.1143/JPSJ.45.875.

[25] Delfosse, V., Grimaldi, M., Bourguet, W., 2012, Crystal structure of hERa-LBD (wt) in complex with bisphenol-C, doi: 10.2210/pdb3UUC/pdb

[26] V. Delfosse *et al.*, "Structural and mechanistic insights into bisphenols action provide guidelines for risk assessment and discovery of bisphenol A substitutes," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 37, pp. 14930–14935, Sept. 2012, doi: 10.1073/pnas.1203574109.

[27] T. Masuya, M. Iwamoto, X. Liu, and A. Matsushima, "Discovery of novel oestrogen receptor α agonists and antagonists by screening a revisited privileged structure moiety for nuclear receptors," *Sci Rep*, vol. 9, no. 1, p. 9954, July 2019, doi: 10.1038/s41598-019-46272-y.

[28] S. Alvarez, "A cartography of the van der Waals territories," *Dalton Trans.*, vol. 42, no. 24, p. 8617, 2013, doi: 10.1039/c3dt50599e.