

Halogens in Protein–Ligand Binding Mechanism: A Structural Perspective

Nicolas K. Shinada,^{*,†,‡,§,||} Alexandre G. de Brevern,^{‡,§,||} and Peter Schmidtke[†]

[†]Discngine S.A.S., 79 Avenue Ledru Rollin, 75012 Paris, France

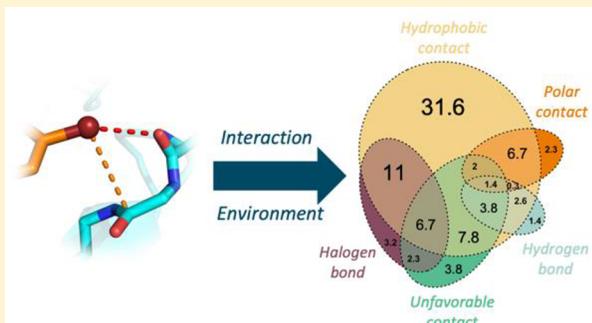
[‡]Biologie Intégrée du Globule Rouge UMR S1134, Inserm, Université Paris Diderot, Sorbonne Paris Cité, Université de la Réunion, Université des Antilles, 75739 Paris, France

[§]Laboratoire d'Excellence GR-Ex, 75739 Paris, France

^{||}Institut National de la Transfusion Sanguine (INTS), 75739 Paris, France

Supporting Information

ABSTRACT: Halogen atoms have been at the center of many rational medicinal chemistry applications in drug design. While fluorine and chlorine atoms are often added to enhance physicochemical properties, bromine and iodine elements are generally inserted to improve selectivity. Favorable halogen interactions such as halogen bond have been thoroughly studied through quantum mechanics and statistical analyses. Although most of the studies focus on halogen interaction through its σ -hole, hydrogen bonding also has a significant impact. Here, we present an analysis describing the interacting environment of halogen atoms in protein–ligand context. With consideration of structural redundancy in the PDB, tendencies toward specific molecular interactions consideration have been refined and implications for rational drug design with halogens further discussed. Finally, we highlight the moderate occurrence of halogen bonding and present the other roles of halogen in protein–ligand complexes, completing the medicinal chemistry guide to rational halogen interactions.



INTRODUCTION

In recent years, the amount of publicly available structural data on the Protein Data Bank¹ (PDB) structures increased from 47 000 to 134 000 in only 10 years. A field that benefits from this wealth of structural information is computer-aided drug design by increasing the understanding of small molecule binding. As such, a multitude of methods have been developed for protein–ligand interaction visualization,^{2,3} characterization,^{4,5} and comparison.⁶ Over the years, numerous studies have described specific interactions in detail such as hydrogen bonds⁷ and sulfur–oxygen interactions.⁸

Halogen elements, i.e., fluorine (F), chlorine (Cl), bromine (Br), and iodine (I), have been incorporated in designed drugs for distinct reasons: (i) to improve their selectivity with the addition of bromine or iodine,⁹ (ii) to increase their ADME properties by including chlorine and fluorine,¹⁰ or (iii) to reduce undesired reactions such as ring hydroxylation.¹¹ While their multiple functions in medicinal chemistry have been explored,¹² their ability to form molecular interactions have also been studied to understand their contribution in binding affinity improvements.¹³

σ -Hole and Halogen Bonding. Heavy halogen elements, defined as chlorine, bromine, and iodine, have an anisotropic electron distribution on their equatorial sides resulting in a positive outer region along their covalent bond called the σ -

hole¹⁴ (Figure 1). Its size depends on multiple factors: a heavier halogen or electron-withdrawing scaffold contribute to a more positive σ -hole.^{14–17} Fluorine displays such a region in exceptional conditions,¹⁸ but is generally considered as σ -hole deficient due to its high electronegativity.¹⁴

Molecular interactions involving halogens have been mostly studied through its σ -region. First evidence of noncovalent halogen interaction was described in the 1950s^{20,21} with the name “halogen bonding” emerging from the late 1990s.²² Multiple nucleophiles act as halogen bond acceptors such as oxygen, sp³-hybridized nitrogen, aromatic ring, or sulfur.

Halogen bond stabilization energy is frequently compared to a strong hydrogen bond (5.8 kcal/mol) as suggested by Kolář.¹⁹ Quantum mechanics (QM) evaluations of halogen bonding indicate an optimal interaction at a distance slightly shorter than the sum of van der Waals radii, and C–X…D angle (θ_1) close to linearity²³ (Figure 2). Meanwhile, a distance increase of 1.0 Å or an θ_1 angle deviation of 25–30° result in a 50% energy loss.²⁴ Furthermore, evaluations using the force field model from Scholfield indicate a zero-energy interaction occurring at 130° for Cl and 140° for Br and I elements.²⁵ Statistical analysis across crystallographic databases, such as the Auffinger et al. study,

Received: September 19, 2018

Published: May 22, 2019

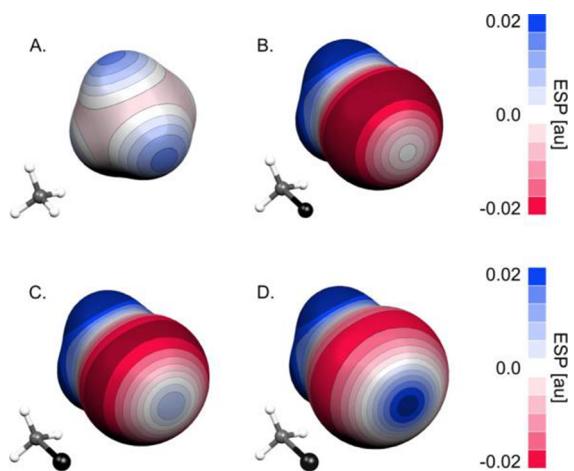


Figure 1. Electrostatic potential surface of (A) methane, (B) chloromethane, (C) bromomethane, and (D) iodomethane. Image from Kolář et al.,¹⁹ reproduced with permission from *Chemical Reviews*.¹⁹ Copyright 2016 American Chemical Society.

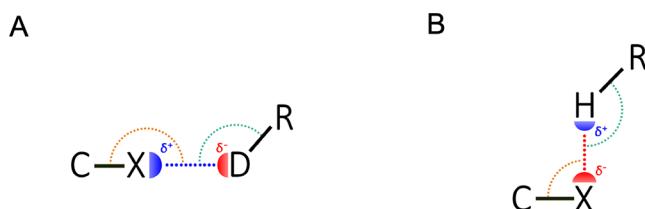


Figure 2. 2D representation of interaction involving halogen X with (A) halogen bond donor D, with δ^+ being the σ -hole region and δ^- being the electronegative region, and (B) hydrogen bond donor H, with δ^- being the electronegative belt. θ_1 angle is represented in orange, while the angle defined by the residue is designated as θ_2 angle.

underlined extensive instances of θ_1 angle of 165° on 964 C–X...O halogen interactions observed in the PDB using sum of van der Waals radii as a distance threshold.²⁶ Moreover, a Cambridge Structural Database (CSD) survey in 2013 by Bauzá et al. highlighted a higher propensity of linear configurations for iodine atoms compared to bromine halogen bonds.²⁷

Experimentally, halogen bonds have largely contributed to affinity improvement on specific drugs. Anticoagulant protein family such as factor X or prothrombin has been optimized by adding a chlorine interacting with the aromatic group of tyrosine 228 in the S1 pocket.²⁸ Interestingly, the inclusion of bromine and iodine, e.g., better halogen bond donors, decreases the binding affinity. Similarly, Rowlinson et al.²⁹ and Koch and co-workers³⁰ underlined the benefit of halogen addition to molecules targeting cyclooxygenase-2 and aldose reductase proteins.

Halogens and Hydrogen Bonding Properties. Due to its anisotropic electron distribution, halogens can act as both a halogen bond donor and hydrogen bond acceptor. In addition to halogen bond depiction, Murray-Rust et al. have underlined in 1979 high propensity of C–I...O interactions in θ_1 angle range of 90° .³¹ Hydrogen bond donors were highlighted in the vicinity of both fluorine atoms (θ_1 angle from 120° to 160°) and heavy halogens (90 – 130° range) in the 2001 CSD data set by Brammer et al.³² More recently, Lin et al. have measured interaction energy for hydrogen bond ranging from 2 kcal/mol for water–chlorine interaction to 14 kcal/mol with a positively

charged guanidine group.³³ Experimentally, the contribution of such an interaction resulted in an affinity increase by a factor of 250 in the design of hepatitis C virus 5B polymerase inhibitors.³⁴

Fluorine element as a weak hydrogen bond acceptor has been debated recurrently in biological conditions. Studies across the CSD database have highlighted only rare occurrences of such interactions with only 0.6% of fluorine atoms involved.^{35,36} On the contrary, Dalvit and Vulpetti have performed extensive experimental analyses to assess fluorine's hydrogen bond aptitude.^{37–39} Recent large-scale assessment from Sirimulla et al. has shown a strong presence of backbone nitrogen and oxygen at a θ_1 angle of 120° around fluorine.⁴⁰

Yet, addition of fluorine groups has contributed to an affinity increase in multiple drug design applications. The inclusion of a trifluoroethyl group to the PhiKan083 drug, a stabilizer of mutant p53, resulted in an affinity increase by a 5-fold factor.⁴¹ Similarly, a 6.7-fold affinity increase was observed in the design of dipeptidyl peptidase IV inhibitor induced by a fluorine addition.⁴²

Complementary Role of Halogen. The halogenation of an aromatic ring is a chemical reaction frequently used to moderate its hydroxylation;¹¹ furthermore it contributes to the ring electron delocalization toward the halogen elements. As a result, electron-depleted surfaces located on each plane of the aromatic ring, i.e., π -hole (Figure S1 in Supporting Information), interact with electronegative groups, e.g., lone pair (lp)– π interaction.⁴³ Therefore, the π -hole intensity depends on the number and nature of electron-withdrawing moieties present on the aromatic group such as fluorine and chlorine atoms.⁴⁴ Wang and co-workers have highlighted a competition between the σ -hole and π -hole in C_6F_5X aromatic ring, with iodine favoring halogen bonding, while aromatic stacking prevails in the presence of bromine.^{45,46}

Description of the Environment. Large-scale studies of halogen interactions in protein–ligand focused mostly on one interaction type, specifically on halogen bonding through the σ -hole. However, a ligand atom is generally surrounded by multiple receptor atoms in a binding context. Hence, potential co-occurring interactions patterns on a single halogen atom are generally overlooked in these analyses. This amphoteric property of heavy halogens has been explored experimentally by Paolo et al.⁴⁷ and more recently by using QM suggesting a strong cooperativity.^{48,49} Lu et al. have twice mentioned this specific interaction pattern without exploring their proper contribution.^{50,51}

Here, we decided to explore through intermolecular contacts, including unfavorable ones, the entire interaction environment of fragments holding halogens in the protein–ligand context. With the consideration of structural redundancy in the PDB, we reassess the occurrences of multiple types of halogen interactions such as halogen and hydrogen bonding, with the consideration of key interactions in redundant structures. A detailed analysis of protein environment around halogen atoms is for the first time analyzed and discussed. Results presented herein update behaviors related to known interactions but also description of unfavorable interactions and hydrophobic contacts. These differences are thoroughly discussed. Last, we propose a catalog of the most commonly observed geometries and interactions involving halogens in macromolecular structures.

RESULTS

Halogen Frequency in the PDB. Using our relational database 3decision (<https://3decision.discngine.com/>), 136 318 PDB entries were stored and processed to identify protein–ligand complexes using heteroatom labeling (March 2018 release). One out of four nonredundant ligands (5950 out of 24 733 using SMILES fingerprint) contains at least one halogen atom. Halogen presence per molecule in PDB ligand is unbalanced; more than half of our molecules contains only one halogen (53.8%), while several instances are composed of three or more halogens (24%, Figure 3). Halogen-rich ligands are

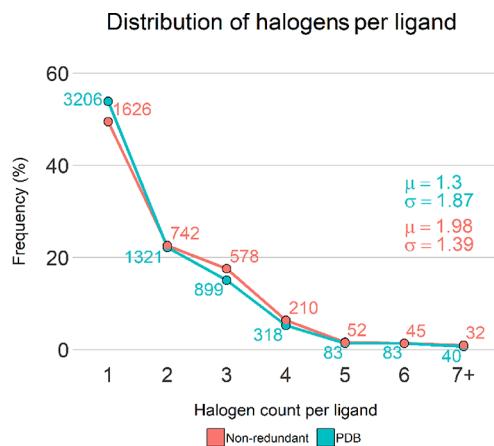


Figure 3. Frequency of halogen atom per ligand in the PDB and after a filtering step preserving nonredundant ligand using SMILES fingerprint (blue and red, respectively). Number of ligands appears as label on distribution.

mostly constituted of fluorine elements, e.g., tricosakis contains for instance 23 fluorine atoms (fluorenlydodecanoylamino-propanoic acid, PDB heteroatom code W1Z).

In comparison to the work of Sirimulla et al.⁴⁰ in 2013, a 3-fold increase of detected contacts involving halogen and protein atoms is observed using sum of van der Waals radii as distance threshold, highlighting their expanding interest in drug design. As multiple evidences of weak interactions occurring at a distance greater than 1.0 Å over the sum of van der Waals radii was previously described,²⁴ this updated distance threshold was used to identify potential interactions. A total of 139 850 detected contacts across 12 200 different complexes, including carbon atoms as receptor, were retrieved.

To avoid the inherent redundancy of complexes involving identical ligands and binding sites in the PDB, a merging step was performed (see **Material and Methods**). Redundancy filtering provided a data set where 26.9% of the original 12 200 halogen complexes were retained, i.e., 3285 complexes. While distribution of halogen elements per molecule remains unchanged (49.5% ligands containing one halogen), the average

number of halogens per ligand increases from 1.3 to 1.9 in our nonredundant data set (Figure 3).

Discarding protein–ligand complexes redundancy allows us to identify multiple occurrences of one ligand bound to different proteins. As such, 11 distinct kinase complexes involving the molecule dasatinib and nine complexes with indomethacin (PDB identifiers 1N1 and IMN) are present in our data set (Figure S2A and Figure S2B in Supporting Information). Overall, 192 out of 3001 unique ligands are bound to multiple proteins in our data set. Similarly, some proteins are also repeatedly represented bound to different ligands, generally due to significant research interest. For instance, β -secretase 1 (named BACE 1), carbonic anhydrase, and capsid protein V1 are the most recurrent receptors within our data set representing respectively 146, 74, and 64 complexes (Table S1 in Supporting Information).

Similar to Hernandes' observations on FDA-approved drugs,¹² halogen elements distribution in our data set features a similar imbalance: ligands in the PDB are mostly composed of fluorine (3807 atoms, 60.4%) involved in 21 423 close contacts and chlorine atoms (2000 occurrences, 31.8%) with 15 551 short-distance contacts. Bromine and iodine are largely underrepresented with only 383 (6.1%) and 115 occurrences (1.8%) accounting for 2915 and 947 contacts, respectively (Table 1). This trend is not specific to the PDB only, as it was also observed for ChEMBL. The recent release of ChEMBL23 database⁵² encompasses similar proportions compared to our data set with 660 174 fluorine elements (55.1%), 430 052 chlorine elements (35.9%), 90 418 bromine elements (7.6%), and 16 666 iodine atoms (1.4%). Close contacts with residue not associated with UniProt assignment were discarded.

Strong fluorine propensity is partly due to electron-withdrawing substituent groups such as trifluoromethyl ($-\text{CF}_3$) and difluoromethylene ($-\text{CHF}_2$), commonly used to lower ligand basicity.⁵³ Combined, they represent 42% of fluorinated fragments in our data set (Figure S3 in Supporting Information), the other half being mostly fluorine atoms directly bound to an aromatic ring, added to protect a reactive methyl group and enhance metabolic stability.¹¹

The prevalence of chlorine atom is directly related to the risk of adding bromine and iodine elements. Indeed, their consequent molecular weights, influential in the drug design ADME properties, prevent the potential selectivity gain induced by their addition.¹² Fragment distribution highlights a 95% representation of chlorine bound to aromatic carbons (Figure S2 in Supporting Information), the remaining fragments $-\text{RCH}_2\text{Cl}$ are mostly present for alkylating activity.⁵⁴ More generally, 2 out of 3 ligands contain a halogenated aromatic ring in our data set.

Molecular Interaction Involving Halogens. In addition to optimizing ligand ADME properties, halogen elements play a significant role in the selectivity and binding affinity properties. Hence, a distance-based contact detection was achieved around

Table 1. Statistical Summary of Halogen Atom Counts and Their Corresponding Interaction Frequency in Our Data Set

halogen	atom count	halogen bond	hydrogen bond	interaction with amide	hydrophobic contacts	polar contacts	unfavorable contacts
aromatic fluorine	1606	NA	211	220	1124	NA	678
aliphatic fluorine	2201	NA	228	280	1630	NA	867
chlorine	2000	438	188	124	1562	463	595
bromine	383	112	33	33	278	61	121
iodine	115	53	14	11	82	16	27
total	6305	603	674	668	4676	540	2188

every halogen atom in our data set. Halogen atoms are surrounded by an average of 6.5 protein atoms at close distance (with 4.7% halogen in contact with only one amino acid atom, **Figure S4 in Supporting Information**). While not every contiguous protein atom interacts directly with the halogen, this value contrasts significantly with the usual interaction description of two atoms facing each other. A description of favorable and unfavorable groups facing each electrostatic region of heavy halogens and fluorine atoms is underlined in the subsequent sections emphasizing the diversity of its binding pattern.

σ -Hole Interaction Partners. Out of the 19 413 distance-based contacts detected involving heavy halogens, 567 contacts result in proper halogen bonds with a Lewis base. 46.1% of iodine atoms, i.e., the strongest σ -hole, are involved in a halogen bond compared to 29.0% of bromine and 21.9% of chlorine as illustrated in **Table 1**. This interaction type is observed in 1 out of 4 complexes involving a heavy halogen.

Similar to recent large-scale PDB observations from Wilcken et al., backbone carbonyl oxygen was identified as the most prominent partner for each heavy halogen.²⁴ As Voth and Ho revealed,⁵⁵ aromatic interactions also play a prominent role as a halogen bond acceptor in our data set, most notably between chlorine and tyrosine side chains (106 occurrences out of 567, i.e., 19.6%). This prevalence is noticeably due to numerous complexes involving anticoagulant inhibitors in our data set such as factor Xa with 48 complexes featuring a halogen bond. In this instance, halogen bond between chlorine and tyrosine 228 is prominent to obtain nanomolar selectivity.²⁸ **Figure 4** illustrates

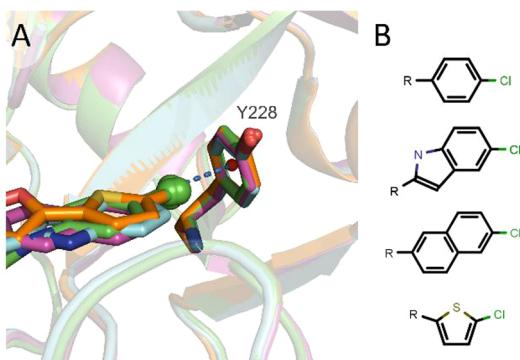


Figure 4. 3D representation of halogen bond (defined by dashes) between prominent chlorine (green sphere) and tyrosine 228 in anticoagulant factor Xa inhibitors with their corresponding scaffold (from top to bottom: PDB codes 2pr3, 3q3k, 4y71, 2boh).

this prominent halogen bond across four different aromatic scaffolds in factor Xa ligands. Halogen bond as a key interaction in the binding mechanism is also observed in E3 ubiquitin-protein ligase mdm2 with the peptide carbonyl oxygen of histidine 96 in 26 instances (**Table S2 in Supporting Information**). Repetitive occurrences were also detected to a lesser extent in heavier halogens as nine bromine elements from distinct structures interact with the aromatic group of CDK1 phenylalanine 80. MAP kinase 1 also features a key halogen bond with nine complexes identified as redundant with valine 127 peptide carbonyl oxygen interacting with iodine inhibitors.

Such examples illustrate the inherent structure redundancy present in the RCSB PDB and therefore should be considered carefully to avoid over-representation and infer biased

tendencies. As such, by use of UniProt sequence residue assignment, cases of multiple instances of identical protein and halogen elements, e.g., factor Xa and chlorine, were merged into only one occurrence. Those initial 567 observations can be differentiated into 327 unique halogen bonds through redundancy consideration. While the carbonyl group of the residue backbone remains the most frequent halogen bond donor despite a 2-fold decrease, tyrosine aromatic ring propensity is greatly reduced (**Figure 5**). Those observations lead to balanced distributions of halogen bond donors across every side chain with a moderate edge for aromatic moieties. Iodine and bromine, due to their rareness, are barely affected by redundancy filtering.

Due to its relatively small range, the σ -hole surface limits the possibility to interact with multiple halogen bond donor. We consider distinct residues interacting with the same halogen at a distance less than 3.8 Å, e.g., sum of sulfur and iodine van der Waals radii. Rare instances of bifurcated halogen bonds are identified in five complexes. **Figure 6** illustrates bifurcated bonds involving backbone carbonyl of valine 41 and sulfur atom of cysteine 42 in a trypsin-like serine protease protein. Interestingly, this cysteine is already involved in a disulfide bond with cysteine 58, highlighting a complex interaction network. Co-occurring interactions involving disulfide bond have been previously described in the literature coupled with hydrogen bonding capabilities.^{56,57} This bifurcated interaction pattern is reminiscent of analysis by Voth et al. as orthogonal interactions between hydrogen and halogen bonding.⁵⁸ Results from this study indicated that interaction energy of both noncovalent bonds are noncompetitive and remained stable in those specific cases.

Interestingly, expanding our distance detection threshold highlights an interesting pattern: a majority of chlorine elements are separated from their halogen bond acceptors by an additional 0.3 Å compared to the sum of van der Waals radii (**Figure 7**). On the contrary, bromine and iodine are more likely to be at their theoretical distance. Halogen bond energy is supposedly to be at its maximum with a distance slightly shorter than the sum of van der Waals (3.27 Å for chlorine and oxygen) and linear configuration (θ_1 angle of 180°).²⁴ θ_1 angle distribution indicates that proper linear arrangement ($\theta_1 > 165^\circ$) is frequently observed for chlorine and iodine atom while most of the bromine interactions are realized in the 155° range. However, high frequencies of weak chlorine halogen bonds are observed within the 140° range and should be taken into consideration in molecular interaction detection descriptions (**Figure 7**). Those results correlate with previous observations from Auffinger and Wilcken.^{24,26}

Few occurrences of potentially unfavorable hydrogen bond donors facing the σ -hole were observed in our data set. Twenty-five cases are mainly constituted of arginine and asparagine side-chain nitrogen atoms facing the positively charged region. Few cases have to be acknowledged properly as 56 interactions involve the hydroxyl group of tyrosine, serine, and threonine side chains where either a halogen bond or unfavorable interaction are potentially occurring regarding the hydrogen position.

Negative Belt Interaction Partners. Anisotropic electron distribution in halogen results in the σ -hole presence but also an electron-rich belt perpendicular to the covalent bond. The hydrogen bond acceptor role of heavy halogens through this electronegative belt was also further explored. While Lin et al. suggest that its contribution to binding could be more influential than halogen bonding, only 9.4% of chlorine, 8.6% of bromine,

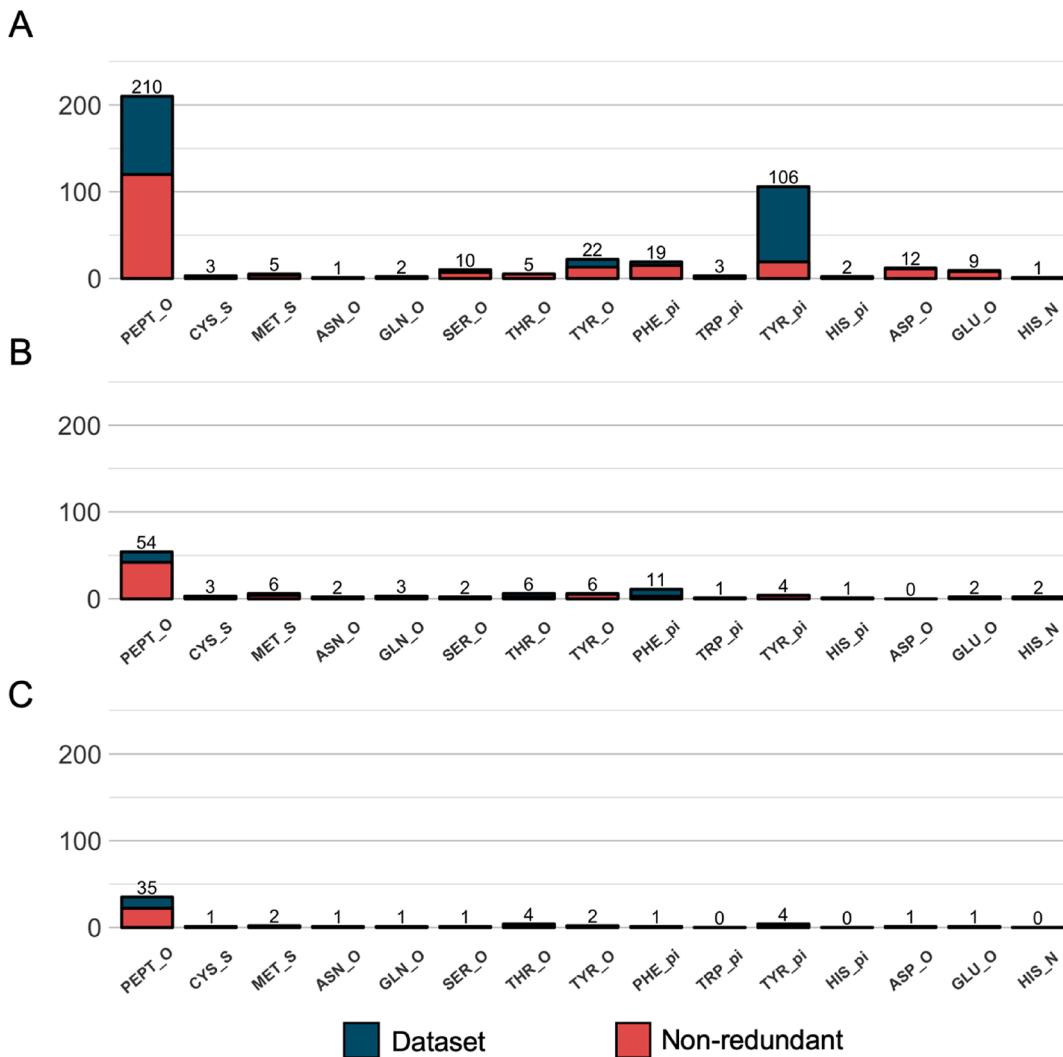


Figure 5. Halogen bond donor distribution observed initially in the PDB data set (opal) and with the consideration of one interaction per repeated complex (red) for each halogen type. Label “PEPT_O” refers to peptide carbonyl oxygen atom regardless of the residues, while other moieties correspond to side chain groups. (A) Chlorine atoms, (B) bromine atoms, and (C) iodine atoms.

and 12.2% iodine atoms are involved in our data set out of 2498 heavy halogens (**Table 1**).

Their occurrences are limited; as such, it is difficult to extract significant amino acid propensities. Hydroxyl group of tyrosine side chain is a recurrent hydrogen bond donor, with 17 bromine and 22 chlorine identified interactions (**Figure 8A** and **Figure 8B**). Threonine hydroxyl is also one of the most preponderant hydrogen bond donors facing the chlorine electronegative belt with 38 potential hydrogen bonds identified, behind the 74 peptide nitrogen interactions (**Figure 8A**). Nonetheless, instances involving hydroxyl groups have to be considered carefully due to the lack of hydrogen in our data set resulting in either a hydrogen bond (favorable) or an unfavorable interaction. Occurrences are too rare for extensive analysis in iodine case.

A significant number of these hydrogen bonds are complex-specific; nonredundant interactions tally 151 unique occurrences from the 241 initial observations. Thus, coagulation factors are largely detected as involved in side hydrogen bonding with 21 and 12 occurrences of peptide nitrogen of serine 214 in factor XI and factor Xa, respectively. Four instances of capsid protein complexes also featured a recurrent hydrogen bonding

with hydroxyl group of tyrosine close to the electronegative belt of brominated ligands in our data set.

Amide groups, due to their dipole moment, offer a positive moment on the carbon atom that can be considered as a potential attractive partner with the heavy halogens’ electron-rich region. This interaction type has been identified and named by Bissantz et al. as “side-on carbon interaction”.⁵⁹ Often studied in carbonyl–carbonyl interaction,⁶⁰ these weak attractive polar interactions are similar to hydrogen bonding with the negatively charged belt. As a result, the covalently bonded halogen has to be parallel with the amide carbon plane and located above the carbon atom to interact.

While these positively charged groups interact with a moderate frequency of heavy halogen, they appear to be very consistent to specific complexes. As displayed by **Table 1**, 7.0% of heavy halogens are involved in this interaction type (171 atoms). β -Secretase 1 inhibitors are recurrently interacting with amide groups as 12 chlorine atoms and 2 bromine atoms are in contact with glycine 291 (**Figure 9B**). Bromine presence and a modest backbone rotation result in a shorter distance between the backbone oxygen and the bromine electron-rich region, leading to a potential unfavorable interaction interfering with the

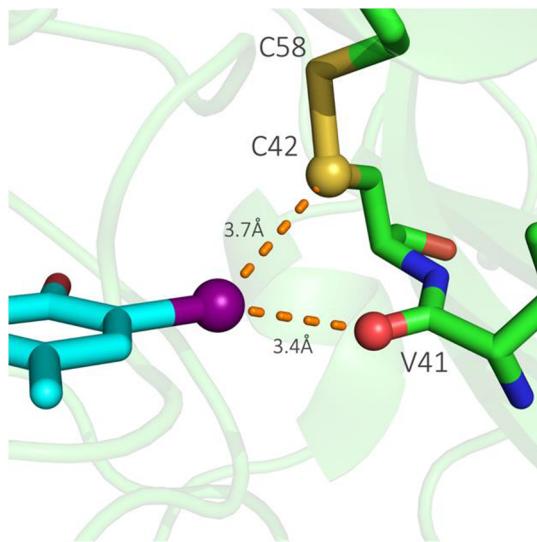


Figure 6. 3D representation of bifurcated halogen bonding involving iodine (purple sphere) in trypsin-like serine protease complex (dashed lines, PDB code 1gjd). Geometric descriptors of both halogen bonds indicate a θ_1 angle of 164° with the valine 41 residue (V41) and 143° with the cysteine 42, respectively.

Geometric descriptors distribution of halogen bond

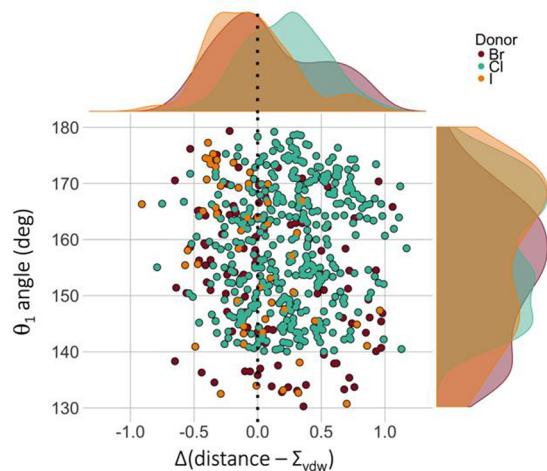


Figure 7. Distribution of halogen bond acceptors geometric descriptors facing the σ -hole region of heavy halogen (bromine in purple, chlorine in green, and iodine in orange). Distance is expressed as the difference between measured distances and sum of van der Waals radii. Densities are highlighted as 2D distributions on x and y axes.

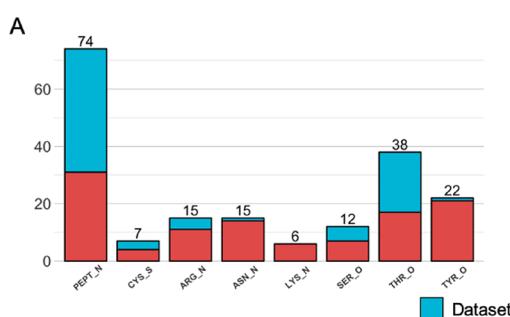


Figure 8. Distribution of hydrogen bond donor in the vicinity of electron-rich belt of heavy halogens in data set (blue) and with the consideration of nonredundant interaction (red) for (A) chlorine and (B) bromine (iodine not shown here due to a maximum number of three observations).

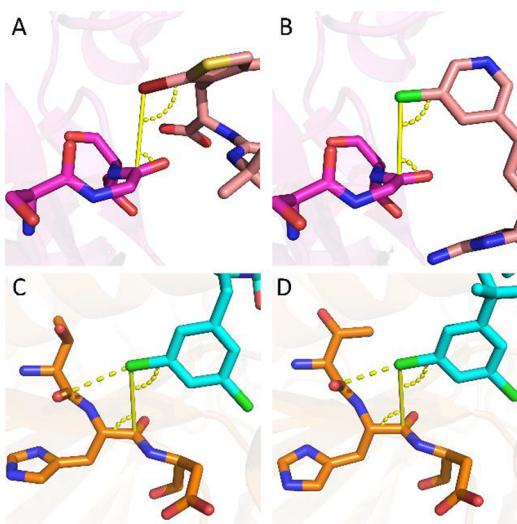
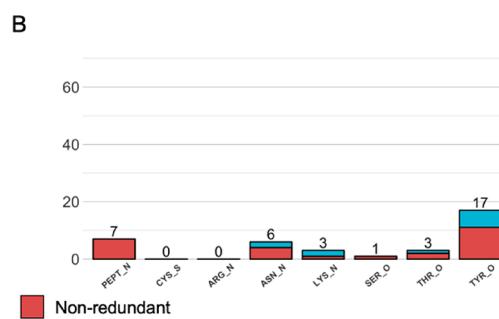


Figure 9. 3D representation of attractive interaction between heavy halogen electronegative belt and amide positively charged carbon. (A, B) β -Secretase 1 inhibitors with bromine and chlorine interacting with glycine 291 amide plane (PDB codes 4i0e and 4pxz). (C, D) Adenovirus 8 protease inhibitors with identical scaffold where chlorine interacts with histidine 25 amide plane and through fuzzy amide interaction or halogen bonding with threonine 24 (PDB codes 4piq and 4wx4).

amide carbon–bromine atom (Figure 9A). This interaction is also observed multiple times in adenovirus 8 protease inhibitors with 5 amide–chlorine interactions detected in our data set with the σ -hole facing the backbone oxygen of the contiguous residue (see Figure 9C,D).

Fluorine Attractive Interactions. Fluorine atoms are considered separately due to their lack of electropositive region resulting in their inability to form classical halogen bonds in biological conditions.^{61,62} Here, fluorine elements mediated interactions are analyzed in two different axes based on the fragment nature: (i) fluorine atoms covalently bound to an aromatic moiety and (ii) fluorine atoms bound to aliphatic carbon such as the recurrent trifluoromethyl. A total of 21 423 fluorine-mediated contacts were found in our data set.

Only 13.0% of aromatic fluorine and 10.0% of aliphatic fluorine have been identified as involved in hydrogen bonding despite its wide θ_1 angle threshold used in our detection method (Table 1). Peptide nitrogen group was the most frequent hydrogen bond donor for both aromatic and aliphatic fluorine atoms accounting for 30.0% and 28.0% of those interactions, respectively. Residue hydrogen bond donor tendencies are



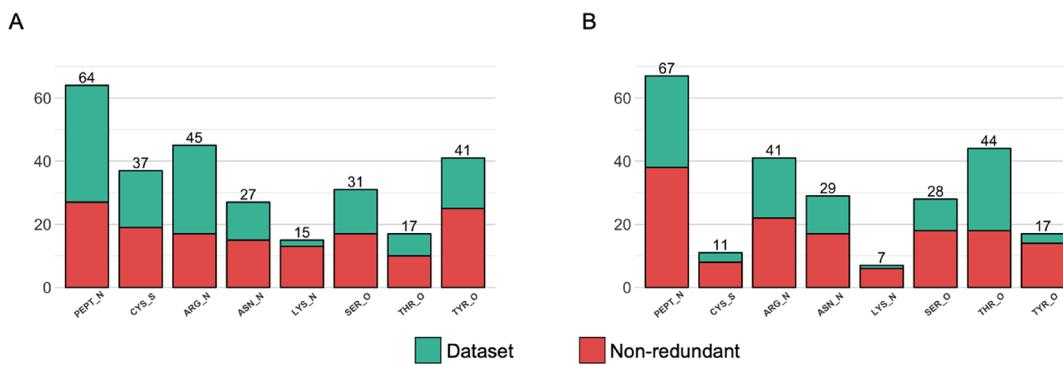


Figure 10. Hydrogen bond donor propensity interacting with fluorine atoms in our data set. Distinct colors are displayed to highlight interaction redundancy due to overrepresentation of specific target and ligand scaffold (unfiltered in green and nonredundant in red): (A) residue propensity for aliphatic fluorine element; (B) residue propensity for aromatic fluorine element.

different regarding the fluorine type. Hence, aliphatic fluorine atoms are involved in weak hydrogen bonding with cysteine's sulfur, hydroxyl group of tyrosine (Figure 10A). Conversely, high tendencies toward hydroxyl group of threonine are observed more frequently for aromatic fluorine atoms than aliphatic ones (Figure 10B).

Looking into details, redundancy of interactions is highly present in both fluorine cases resulting in unique hydrogen bond counts of 143 for aliphatic fluorine and 141 for aromatic ones (from 277 and 244 initial observations, respectively). For instance, cysteine 443 sulfur atom of coagulation factor X is detected in 15 distinct complexes as hydrogen bond donor in contact with aliphatic fluorinated inhibitors. Aromatic fluorine elements are also subject to similar repetitions; inhibitors of carbonic anhydrase 2 are close to threonine 200 hydroxyl side chain in 13 instances. Distribution of hydrogen bond donors is more balanced with the consideration of interaction redundancy, especially in aliphatic fluorine where peptide nitrogen propensity is similar to the tyrosine side chain (Figure 10).

Interestingly, these results are in contradiction with observations from Zhou et al.⁶³ as they described backbone nitrogen atoms as the least occurring partners. These updated results underline the necessity to regularly revise the molecular interaction analyses with resources like the PDB.

Multiple hydrogen bond donors in the vicinity of fluorine atoms are only found in a limited number of instances. A recurrent pattern among carbonic anhydrase involves two inhibitors with seven identified occurrences (Figure 11). Otherwise, multiple attractive partners are very rare and are not recurrent to specific protein.

Attractive interactions involving amide group dipole moment and the electronegative fluorine elements have been explored in specific cases such as thrombin, menin-LL, and p53 inhibitors leading to 5-fold improvement in affinity.^{64,65} Large scale studies have been performed in the past on the CSD and PDB being referenced as orthogonal multipolar interactions.^{59,66} Pollock et al. have developed FMAP, an algorithm highlighting fluorophilic sites near peptide carbonyl group.

While those interactions are mostly described for the trifluoromethyl group, we have detected a similar percentage of fluorine elements involved in this interaction type for aromatic and aliphatic moieties with 13.7% and 12.8%, respectively (Table 1). Receptors such as MAP-kinase 14, cathepsin S, menin-LL or receptor to androgen ligands all contain a recurrent, i.e., more than 10 times, orthogonal polar interaction in our data set. These repeated interaction patterns

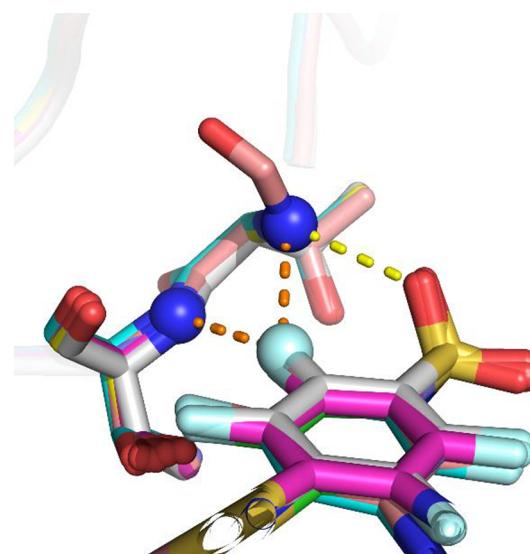


Figure 11. 3D representation of carbonic anhydrase 2 recurrent weak hydrogen bonding though interaction between fluorine (light blue sphere) interaction and both threonine 199 and threonine 200 nitrogen atoms (dark blue spheres) across multiple different ligands (PDB codes 4pzr, 5llc, 5drs, 4dz7, 4qjm, 4ht0, 4dz9).

apply for both aliphatic and aromatic fluorine atoms, tallying for 162 and 148 unique interactions out of 583 initial interactions.

Fluorine atoms are more inclined to be in a tilted conformation relative to the amide plane as geometric descriptors indicate. T-shape conformations, where the fluorine atom faces directly the carbon, i.e., θ_1 angle greater than 160°, are rarely observed. Aromatic fluorine atoms tend to be more frequently in parallel arrangement relative to the plane compared to the tilted conformation of aliphatic halogen with a peak θ_1 angle at 110° and 130°, respectively (Figure 12).

Few cases involve multiple amide planes forming a cage-like subpocket around the fluorine atom, similar to orthogonal interaction patterns mentioned by Voth around carbonyl oxygen with halogen and hydrogen bonding.⁵⁸ MAP kinase 13 protein is repeatedly observed as displaying such patterns with 14 instances between leucine 104 and valine 105 (Figure S5A in Supporting Information). In a similar fashion, alanine 213 and phenylalanine 215 backbone surround six recurrent fluorine atoms in neutrophil elastase ligands and entrectinib is close to amide of arginine 1254 and glycine 1269 in ALK tyrosine kinase

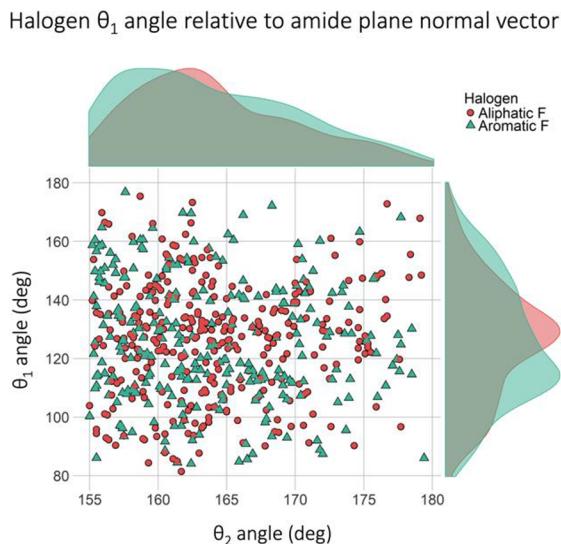


Figure 12. Distribution of geometric descriptors for both aromatic and aliphatic fluorine polar interactions with peptide amide carbon plane (θ_1 and θ_2 angles).

receptor (Figure S5B and Figure S5C in Supporting Information).

Hydrophobic Environment. Known attractive interactions relative to halogen atoms have been assessed in the PDB in the past years.^{24,26,40} However, a recurrent hydrophobic environment surrounding halogen atoms in protein–ligand context has been implied recently.⁶⁷ Therefore, hydrophobic contacts with carbon moieties were also computed in our data set. A total of 8273 contacts have been identified as proper hydrophobic interactions involving 4676 halogen atoms.

Fluorine, bromine, and iodine atoms all range around the same propensity of atoms involved in carbon interactions with 72.3%, 72.6%, and 71.3%, respectively. Chlorine is found to be involved more frequently in this type of contact, 78.1% of all chlorine atoms (Table 1).

The number of hydrophobic partners surrounding halogen atom was also measured. Despite the significant size difference between iodine and fluorine, 0.5 Å in their respective van der Waals radii, and hence the larger detection threshold for contact detection, halogens display homogeneous hydrophobic contact frequencies (Figure 13A). Over 45% of the halogens interact with one hydrophobic moiety. Heavier halogens appear to be more frequently located in hydrophobic-rich environments, e.g.,

at least four hydrophobic partners, with 6.8% for chlorine to 13.4% for iodine in those cases. This observation results from the larger detection distance relative to the bigger van der Waals radius of iodine. However, some packed environments are specific to some proteins and have been detected multiple times: 17 instances of chlorinated inhibitors in ubiquitin ligase mdm2 complexes (Figure S6 in Supporting Information) and 15 complexes involving carbonic anhydrase 2 for instance.

The methyl group is the most frequently hydrophobic moiety in contact with halogen atom (more than 50% for fluorine and chlorine elements) followed by $-\text{CH}_2$ moiety with around than 35% (Figure 13B). Propensity toward a hydrophobic partner is mostly similar across every halogen with the exception of heavier ones where iodine and bromine favor interaction with methylene moieties over methyl groups.

Since no specific θ_1 angle filter was applied on the hydrophobic contact detection process, geometric descriptors were studied accordingly. The geometry of these interactions differs regarding the interacting halogen element (Figure 14). Interestingly, we can corroborate that the presence of hydrophobic atoms facing the σ -region of heavy halogens is limited. Hydrophobic moieties are located either in front of the negative belt for chlorine and bromine, e.g., θ_1 angle of 90°, or with a slight inclination of 110–120° for iodine and fluorine halogens.

Halogen Interaction Environment in Protein–Ligand. Molecular interactions between a protein and a ligand are generally viewed as one or maximum two interactions considered for each atom. However, atoms are generally in contact with numerous atoms (as described in Figure S4); hence it is noteworthy to describe this environment. Wilcken et al. have suggested for instance the presence of primary, secondary, and even tertiary halogen bond in some complexes.²⁴ Here, we will describe protein environment by the interaction nature of the surrounding atoms directed toward the halogen. Furthermore, geometric thresholds defined previously for halogen and hydrogen bonding omit polar contacts. As a result, interactions with a θ_1 angle between 110° and 130–140° for heavy halogens are usually neglected in interaction analyses due to the absence of a polar force in this region. Interacting with polar moiety, they will be referred to as polar contacts in the following analysis. 5724 out of 6305 halogen atoms from our initial data set were defined with at least one protein moiety directed toward these atoms.

The Venn diagram in Figure 15 shows the co-occurrence between different interaction types on the same halogen atom for each protein ligand complex. First and foremost, halogen

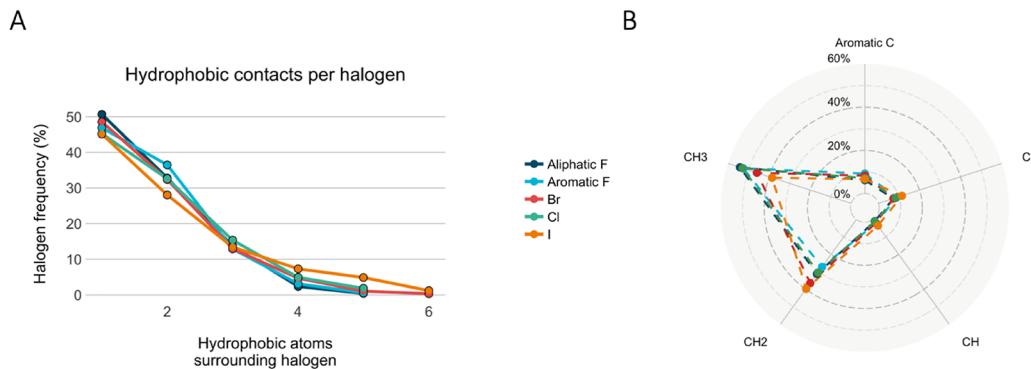


Figure 13. Statistical distribution of hydrophobic environment around halogen atoms: (A) quantitative distribution of hydrophobic moiety surrounding for each halogen atom; (B) frequency of hydrophobic partner involved in hydrophobic contact for each halogen type.

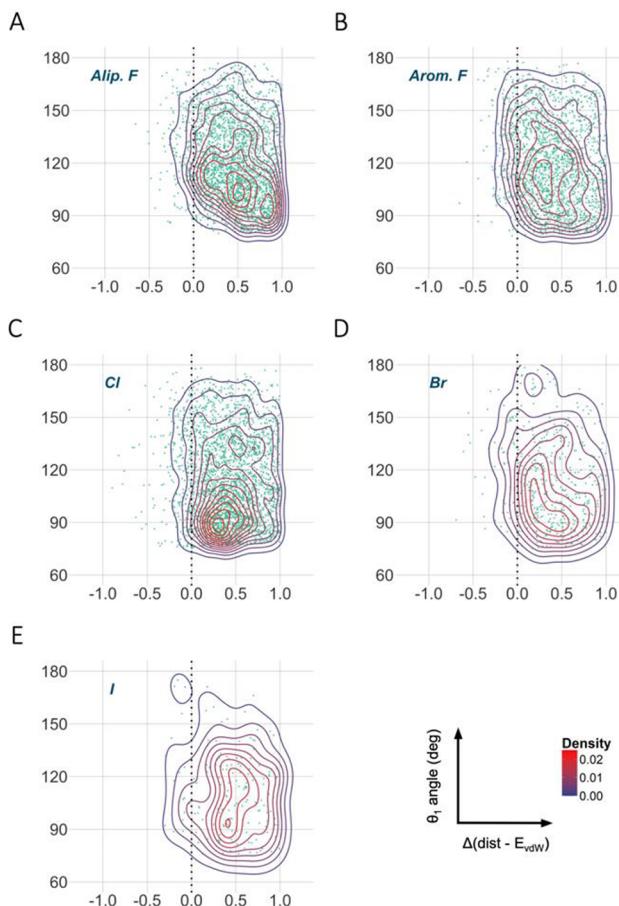


Figure 14. Distribution of geometric descriptors, distance and θ_1 angle, of hydrophobic contact partner surrounding each halogen type (observations plotted in green). Density of distribution is highlighted with blue corresponding to low density of points and red to high density for (A) aliphatic fluorine, (B) aromatic fluorine, (C) chlorine, (D) bromine, and (E) iodine.

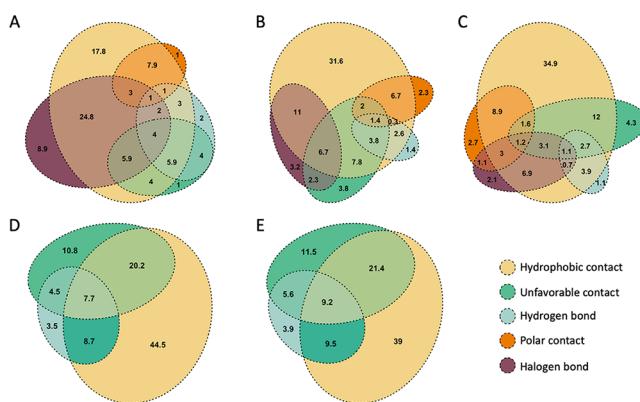


Figure 15. Frequency distribution of electrostatic interactions and interatomic contacts surrounding each halogen element in our data set through Venn diagrams: (A) iodine; (B) bromine; (C) chlorine; (D) aliphatic fluorine; (E) aromatic fluorine. For visual clarity, Venn diagrams were optimized and might miss some values, explaining why the sum of percentage can be below 100%. Interactions were computed as hydrophobic contact (yellow), unfavorable contact (green), hydrogen bond (blue), polar contact (orange), and halogen bond (purple).

atoms are hardly observed in an environment composed of only attractive moieties. Less than 4% of fluorine atoms are surrounded by only hydrogen bond donors or the partially charged carbon from amide groups. Interestingly, iodine elements are less frequently surrounded by unfavorable moieties, e.g., hydrogen bond donor facing the positive region, with 26.7% of halogen in such conditions compared to 35.1% of bromine and 32.2% of chlorine atoms.

The nature of protein environment is dependent on the halogen involved. Both fluorine types are equally surrounded by homogeneous environment, 58.8% for aliphatic fluorine atoms and 54.4% for aromatic ones. Hydrophobic-only donors are more frequently observed in those homogeneous pockets, representing more than 40% of fluorine atoms. Heavier halogen elements tend to be surrounded more frequently by heterogeneous environment: 45.1% of chlorine atoms are in such configuration in opposition to 30.7% of iodine due to large contribution from hydrophobic groups. Noticeably, the nature of the homogeneous environment is greatly element-dependent. 34.9% of chlorine atoms are enclosed in a hydrophobic-only surrounding compared to 17.8% of iodine in similar condition. Halogen bond is generally located in nonpolar or favorable subpockets with only few instances of co-occurring unfavorable contacts.

The amphoteric role of heavy halogens as both a Lewis base and acid occurs rarely since 128 instances were detected in our data set (not shown for bromine on Figure 15). These amphoteric halogens are characteristic of anticoagulant factors, detected in 21, 15, and 12 complexes of factor XI, factor Xa, and prothrombin inhibitors. Figure 16A illustrates both interactions in anticoagulant inhibitors where peptide nitrogen of serine 114 is directed between peptide carbonyl group and halogen side. Four cases of estrogen receptor inhibitors also indicate a dual role, more specifically with the amide plane pointing toward the side of bromine (Figure 16B).

A significant percentage of unfavorable elements, e.g., hydrogen bond acceptor near negative belt, are located in the vicinity of halogen atoms. Interestingly, the heavier halogens display a higher propensity to both attractive and unfavorable elements within the same environment, with 21.8% of iodine and 13.3% of chlorine in such situation (20.6% of bromine).

Polar contacts, defined as polar atoms interacting with the nonpolar region of the heavy halogen, i.e., around θ_1 angle of 120° , are also highly present in the protein–ligand context involving 25.0% of iodine, 17.7% of bromine, and 15.8% of chlorine atoms. The presence of a proper interaction or its electrostatic nature is up to debate; this region constitutes a fuzzy boundary between the electron-depleted σ -hole and the electron-rich belt, usually considered as noninteracting. Statistically, the amount of halogen atoms involved in polar contacts is close to halogen bond frequency with a total of 540 interactions. Peptide oxygen is the most frequent contact partner with 254, 23, and 5 occurrences for chlorine, bromine, and iodine, respectively. Nitrogen of the residue peptide bond and aromatic side chains are also regularly observed in the vicinity of this neutral region of chlorine (56 and 74 cases) where θ_1 angle distribution indicates a large count of polar partners at 120° (Figure S7 in Supporting Information). Their presence in a protein environment is largely coupled with a hydrophobic moiety around halogen atoms.

Recurrent patterns of halogen environment are observed such as three hydrophobic moieties in contact with 14 distinct estrogen receptor ligands containing chlorine, bromine, or

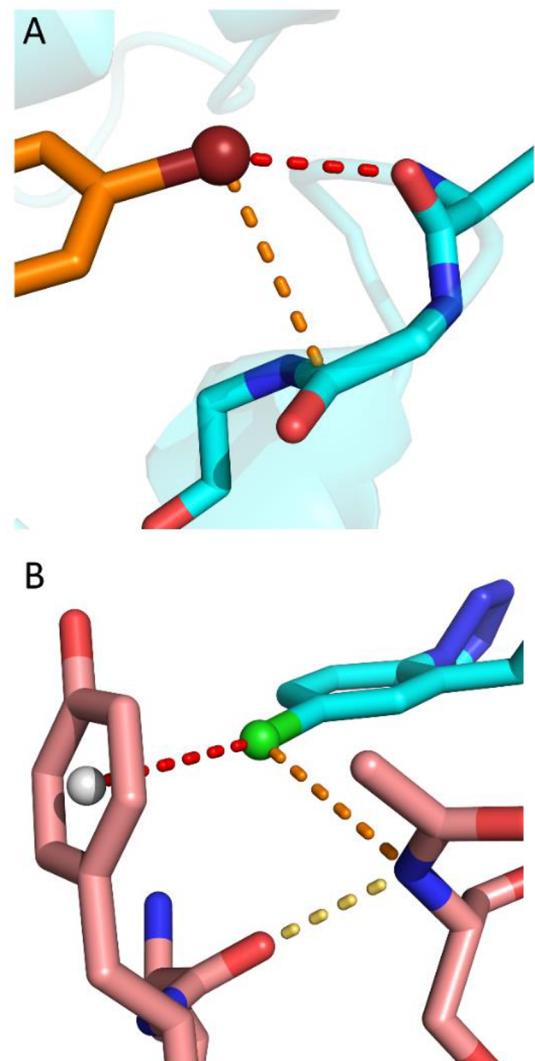


Figure 16. Amphoteric role of bromine and chlorine acting as halogen bond donor and attractive polar interaction on respective negative belt. (A) Estrogen receptor ligand bromine (red sphere) interacting with peptide carbonyl oxygen of glycine 419 (red) and amide positive carbon of glycine 420 (orange, PDB code 5tm8). (B) Factor Xa inhibitor chlorine (green sphere) interacting with serine 214 (orange) and tyrosine 228 (red) (PDB code 4y8x).

iodine. In a similar fashion, anticoagulant factor XI subpocket comprises a halogen bond acceptor, a hydrogen bond donor, unfavorable element, and hydrophobic moiety in eight occurrences; seven cases of prothrombin subpockets are also displaying similar pattern. Interestingly, cyclin-dependent kinase 2 ligands have a recurrent environment pattern within the same binding sites where chlorine atoms interact in similar fashion as bromine but in a distinct subpocket ([Figure 17A](#)).

Last, the interaction environment of dasatinib, a chlorinated nonselective inhibitor, was analyzed in more detail. Different interaction patterns around the chlorine have been observed depending on the protein targeted. Targeting for instance ephrin type-A receptor a halogen bond with serine 756 is observed ([Figure 17B](#)). However, this halogen bond is not detected on other tyrosine kinases. Instead, interaction patterns are very diverse. Hydrogen bonding is protein-specific in two instances: sterile 20-like kinase and Abelson tyrosine-protein kinase. Interestingly, [Figure 17B](#) displays recurrent hydrophobic moieties across

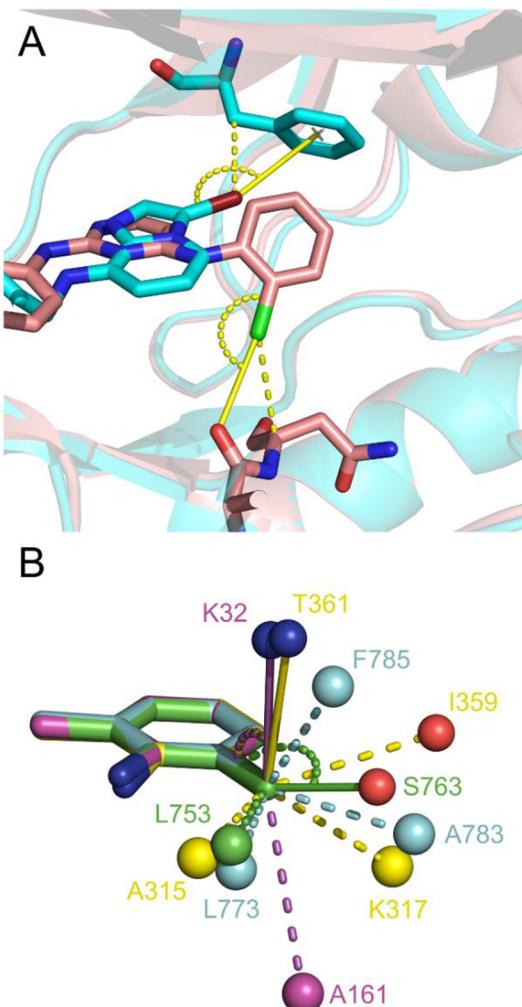


Figure 17. Superposition of similar interaction pattern in identical receptor or identical ligand. (A) Similar interaction pattern on CDK2 inhibitor observed on two distinct subpockets involving halogen bond and hydrophobic contact (halogen bond in line, hydrophobic contact in dashes, PDB code 2r31 in blue and PDB code 3unk in pink). (B) Interaction environment around chlorine of dasatinib in mammalian sterile 20-like kinase 3 (purple, PDB code 4qms), ephrin type-A receptor 4 (green, PDB code 2y6o), Abelson tyrosine-protein kinase (yellow, PDB code 4xli), and epithelial discoidin domain-containing receptor 1 (light blue, PDB code 5bvw) with hydrophobic and polar contacts displayed in dashes and attractive interactions in straight lines (red sphere, oxygen; blue sphere, nitrogen; others spheres, carbon).

different protein such as discoidin-domain receptor leucine 773 and alanine 783 and Abelson tyrosine kinase alanine 315 and lysine 317.

Aromatic Stacking Involving Halogen atom. Aromatic hydrocarbon rings display specific interacting properties. Electrons can move along the covalent bond in the aromatic ring system and thus build up two negatively charged regions referred to as π -systems. Such regions are commonly admitted for natural aromatic amino acids for example.

Prominent work from Hunter et al.^{68,69} has highlighted the preferential arrangement of aromatic fragments by studying phenylalanine residue in protein. T-shape, edge-to-face, and parallel displaced configurations are usually considered as favorable, while sandwich arrangement is most likely to be unfavorable.

However, these configurations are supposed to be suited only for aromatic groups where lone pairs extend in the ring's plane; electron-withdrawing groups have a profound effect on the electron delocalization. As a consequence, fluorine elements and to a lesser extent chlorine atoms attract electrons toward their position on the ring leading to two electron-depleted regions on both sides of the aromatic ring plane called π -hole.⁴⁴

Interestingly, in current molecular modeling tools and molecular mechanics force fields, interactions involving a π -hole with other aromatic fragment are generally considered as π -stacking interactions, even though one electron-deficient region is involved. This type of interaction was referred to more suitably as aromatic donor–acceptor interaction by Martinez et al.⁷⁰ According to Wang et al.,⁴⁴ a benzene with at least three fluorines can yield such an electropositive moiety. Here, we decided to analyze interactions with the consideration of the number of halogen substituents in our aromatic groups.

In theory, π -hole reverses the electrostatic properties of the aromatic rings with electron-depleted region located on each plane surface instead of electronegative regions. Parallel aromatic arrangement referred to as sandwich configuration, unfavorable in non-halogen aromatic stacking, is attractive in the presence of halogens. Moreover, a T-shape arrangement with the edge of a benzene facing the π -hole highlights two identical polarities facing each other, i.e., repellent force.

Histidine was arbitrarily not considered in aromatic contact analysis of halogen containing molecules. A distance-based contact detects a large number of aromatic contacts in our data set. 10% of our aromatic interactions involve at least one halogen in our data set (Figure 18). While the large majority contains only one halogen, the presence of two should not be neglected with 294 unique interactions.

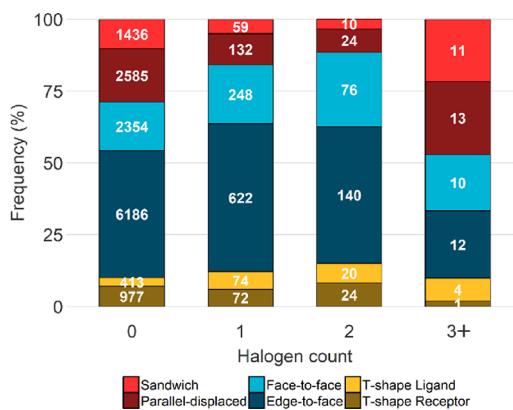


Figure 18. Distribution of aromatic- π -hole spatial arrangement relative to the number of halogens present on the aromatic ring.

In both our nonhalogen and halogen data set, edge-to-face is clearly the most recurrent configuration with observed frequencies of 44.3% and around 50%, respectively (Figure 18). Favorable interactions between the electron-depleted region of π -hole and electron-rich surface of protein aromatic moiety, face-to-face and sandwich, are surprisingly under-represented with a combined 21.6% and 4.6% for both configurations.

Two-proportions Z-test between nonhalogen and halogen aromatic data sets (except three or more) indicates different results. First, proportions are deemed as similar in one-halogen and two-halogen data set with *p*-values ranging from 0.06

(25.8% and 20.5% of face-to-face arrangement) to a maximum of 0.35 (3.4% and 4.9% of sandwich conformation). Observations between nonhalogenated and one halogen within the aromatic fragment are considered as divergent with a maximum *p*-value of 0.001 for face-to-face conformation (16.9% and 20.5%, respectively). Aromatic groups containing two halogens and without halogens were also tested with the only *p*-value above 0.05 observed for edge-to-face geometry (*p*-value = 0.289; 47.6% and 44.3%, respectively). Those comparisons indicate differences in arrangement distribution in the presence of halogen in the aromatic ring despite the similar behaviors observed.

The increase in T-shape configuration with the receptor aromatic ring edge facing the π -hole is unexpected due to its unfavorable positive–positive interaction. On the contrary, the electron-withdrawing group of π -hole aromatic group allows for the remaining edge to be more electron-depleted, favoring T-shape interaction with the trunk formed by the ligand. Increase in face-to-face configuration for halogen-rich aromatic ring is expected because of the presence of the positive region of π -hole.

Aromatic fragments containing three or more halogens, i.e., proper π -hole, are rarely detected in the PDB with 51 occurrences. Therefore, the interpretation of the results should be considered carefully due to the low occurrence. Except for both T-shape configurations (9.8%), other configurations are similarly represented with a minimum of 12 observations for edge-to-face arrangement to a maximum of 13 occurrences for parallel-displaced.

DISCUSSION

Data Set Relevance and Current Pitfalls. For more than 20 years, a constant increase in halogenated ligand registration is observed in the PDB, with the exception of iodine which remains steady (Figure S8 in Supporting Information). Its lower rate is due to multiple factors such as its high atomic weight unsuitable for substantial ligands and surely its expensive synthesis.¹² Multiple studies have analyzed the propensity and applicability of halogen bonds in protein–ligand complexes and their interest in drug discovery.⁷¹

However, statistical analyses of molecular interaction have to be considered carefully due to the prominent redundancy of structures in the PDB and must not be neglected. It is obvious that identical complexes are found across one or multiple PDB entries; not considering these repetitions during analysis contributes to misleading tendencies. A critical example is the large number of chlorinated anticoagulant inhibitors. The high number of chlorine–aromatic interactions through the σ -hole can lead us to think this interaction as very common and preferential.

The parameters and descriptors used in the interaction definition of interaction also play a key role in its study. For instance, aromatic interactions were considered individually between each aromatic carbon atoms and a halogen atom when separated by a short distance in the Sirimulla study.⁴⁰ Using similar rules instead of considering one interaction per aromatic ring moiety led to a 2-fold overestimation of halogen–aromatic interactions in our data set, biasing residue propensity interpretation (Table S3 in Supporting Information).

Interacting Partners. Molecular interactions involving halogen atoms are still largely misunderstood. For instance, the automatic interaction detection program Protein-Ligand Interaction Profiler⁷² considers only halogen bonding for this element type, with fluorine characterized as a halogen bond

donor. Using similar criteria, only 2158 interactions out of our 40 000 contacts would have been considered in this context.

Quantum mechanics studies have highlighted the presence of the σ region on halogens, and multiple analyses have enumerated Lewis bases facing this region in the PDB.²⁴ As described in our previous sections, halogen bonding is involved in one halogen out of five and is very specific to some protein targets such as anticoagulation factors or kinase. Rare instances of unfavorable elements facing the electropositive region consolidate the existence of σ -hole.

Still, the remaining halogen atoms also play a role in binding mechanism such as hydrogen bonding or attractive polar interactions underlining the prerequisite to explore different interaction patterns. While looking thoroughly at halogen interactions involving the σ -hole in a similar fashion as Wilcken et al.,²⁴ we also have included analyses of potential hydrogen bonding but also interactions that are usually overlooked in halogen biomolecular interactions. Recent studies have highlighted the hydrogen bonding role of heavy halogen in binding mechanism. However, unfavorable partners are also frequently observed in the electronegative belt vicinity, underlining the specificity of hydrogen bonding aptitude of heavy halogen to particular complexes. Furthermore, a significant number of polar atoms located in the neutral region of heavy halogens were observed, generally ignored in molecular interaction detection software. These polar contacts can play a key role in the binding mechanism by contributing to the hydrophobic component. Moreover, with our data coming from crystallographic data and therefore representing a static representation of a dynamic mechanism, those polar atoms facing the neutral region can potentially become an attractive halogen bond with a small motion.

Fluorine has seen an increased interest in its potential role as an interacting moiety in recent years. Normally considered as a weak hydrogen bond acceptor, we have found that more fluorine atoms are involved in attractive interaction with the positively charged carbon of amide group. This observation correlates to recent studies on a specific target such as p53 kinase where the trifluoromethyl group acts as a prominent moiety in the selectivity.⁴¹ Nonetheless, attractive interacting property of fluorine atom is still rarely observed with more unfavorable elements than attractive ones in its vicinity. Overall, fluorine atoms are still more recurrently detected in hydrophobic environment and increase affinity through entropy gain.

The usual description of molecular interaction in protein–ligand relies on one or two attractive interactions per atom. However, ligand and protein moieties are usually surrounded by a multitude of other neighboring atoms in their vicinity. Characterization of the entire environment using both attractive and unfavorable interactions results in an appropriate description of the binding pattern of halogen atom in our context. These patterns could be used as a signature to specific binding site and therefore guiding the medicinal chemist in halogen favorable environment because of the identification of similar pattern surrounding one interacting atom.

Alternative Interacting Roles. Here, we also analyzed the arrangement between aromatic fragments in protein ligand binding. Aromatic ring spatial arrangements were compared between nonhalogenated and π -hole rings. Our results suggest changes in 3D configuration depending on the presence of a halogen or not in one of the aromatic fragments (Figure 18). Although preferences can be seen in nonhalogenated aromatic ring toward face-to-face and parallel-displaced arrangement

tendencies, they are widely different in the presence of halogens. Therefore, observing a decrease in the favorable parallel-displaced conformation in the presence of halogen is expected due to the shift in dipole moment on aromatic moieties. The same can be said about the T-shape conformation where the ligand plays the role of the trunk of the T; its remaining edges, nonhalogenated, will display a more positive dipole moment compared to benzene, favorable to an interaction with the π negative region of a halogen-depleted aromatic ring. Meanwhile, face-to-face arrangement, because of matching electrostatic regions, is recurrent in the presence of halogen and can be considered as a favorable interaction.

In the protein–ligand binding mechanism, some ligand atoms have a bigger contribution in the binding affinity. Those can be identified either by coupling experimental data with crystal structure and when a high number of structures of the same targets are available.

Chlorinated aromatic elements for instance greatly enhance the binding on factor X, Xa or prothrombin and appear as an interacting conserved fragment in the drug design process against those targets. Therefore, the binding mode remains similar across multiple complexes. This type of moiety has to be clearly distinguished from the rest of the ligand, more mobile and less influential in the ligand selectivity (Figure S9 in Supporting Information). However, the detection and characterization of conserved and flexible ligand atoms require either a large amount of data or a suitable molecular dynamic to discriminate those two states.

Caveats. The lack of protonation states in the PDB complicates the analysis of each interaction type. The dual interacting role of hydroxyl in serine, threonine, and tyrosine as a hydrogen bond donor and a Lewis base can generate either an unfavorable contact or a proper halogen bond when positioned in front of the σ region.

Moreover, it is substantial to consider a bigger intermolecular context while studying protein–ligand binding mechanism. Molecular interactions between protein atoms, especially hydrogen bonding, also play a key role and can improve the description of complex molecular interaction pattern. It is particularly striking in the case of aequorin (Figure 6) where geometrically, a bifurcated interaction through the σ -hole seems to be possible at first, but an α -helix hydrogen bond is prioritized.

While the high quantity of chlorine and fluorine allows for a suitable study, making conclusions on iodine and bromine has to be put into perspective due to their scarcity in the PDB. It will be interesting to perform the same type of analysis with a bigger data set for these two atoms in the future. Finally, crystal structures are a snapshot of a dynamic process; therefore the conformational state that we observe in our data set is limited. While these structures are supposed to be at an average state, it is possible that other appropriate binding conformations favoring proper halogen bonding instead of hydrogen bonds exist. Interpretation is narrowed by the conformational states that we possess in each of the protein–ligand binding “picture”. Development of halogen bond parameters in force field, e.g., CHARMM,⁷³ is still in its infancy.

■ CONCLUSION

Although theoretical studies around halogen interactions have been performed, only a few have explored different interaction types outside halogen bonding. Here we highlighted an updated and more appropriate description of halogen environment in the

Table 2. Description of Considered Aromatic Arrangement and Their Respective Geometry Descriptors

Illustration	Arrangement name	Geometric descriptors
	Sandwich	$\theta_1 > 160^\circ$ (and) $\theta_2 > 160^\circ$ (and) Arrangement angle > 160°
	Parallel-displaced	$135^\circ > \theta_1 > 160^\circ$ (and) $135^\circ > \theta_2 > 160^\circ$ (and) Arrangement angle > 160°
	Face-to-face	$\theta_1 > 160^\circ$ (or) $\theta_2 > 160^\circ$ $105^\circ > \text{Arrangement angle} > 160^\circ$
	Edge-to-face	$135^\circ > \theta_1 > 160^\circ$ (or) $135^\circ > \theta_2 > 160^\circ$ (and) $105^\circ > \text{Arrangement angle} > 160^\circ$
	T-shape ligand	$75^\circ > \theta_1 > 105^\circ$ (and) $\theta_2 > 150^\circ$ (and) $75^\circ > \text{Arrangement angle} > 105^\circ$
	T-shape receptor	$\theta_1 > 150^\circ$ (and) $75^\circ > \theta_2 > 105^\circ$ (and) $75^\circ > \text{Arrangement angle} > 105^\circ$

protein–ligand context. Although known interactions such as halogen or hydrogen bonding are present, numerous protein atoms directed toward halogen have been referenced and described. Overlooked interactions, such as unfavorable electrostatic moieties, are largely represented and could improve both understanding and comparison of binding mechanism. While these interactions and their arrangements still have to be explained from a quantum view, their large numbers and conserved patterns in specific protein family indicate a significant role in the binding process. Those results paired with bioisosteres experimental studies show the prominent role that halogen have in the drug discovery field.

MATERIAL AND METHODS

Following IUPAC recommendations,⁷⁴ an acceptor such as halogen bond acceptor is defined by the presence of a lone-pair or an electronegative moment. On the contrary, a donor must exhibit a positive electrostatic region to be considered as such. Moreover, a clear distinction between interactions and contacts must be done. A contact is defined when the distance is less than the defined threshold, while an interaction involves surface accessibility and electrostatic components. An interaction can either be attractive or unfavorable.

Data Set and Contact Detection. The entire Protein Data Bank¹ (PDB, release March 2018) was analyzed. Structures containing small molecular entities identified by heteroatom tags (HETATM labels) were considered for further analyses. In the case of alternates, the atom associated with the maximal occupancy is conserved or the first alternate conformation if the occupancies are equal. In addition to NMR structures, X-ray structures with resolution lower than 2.5 Å were considered. Ligands were also filtered based on their molecular weight (between 250 and 850 Da), their size (more than five heavy atoms), at

least one ring, and 90% of single bond fraction to keep only relevant ligands in drug discovery and filter out porphyrins, lipids, and other nonrelevant ligands.

Contacts between two atoms (one from the protein, one from the ligand) were defined by a fixed threshold corresponding to van der Waals radii of each atom + 1.0 Å. Structures and contacts were then processed within Pipeline Pilot 2017R2⁷⁵ for characterization purpose such as atom hybridization or ligand fragmentation. Molecular fragmentation was performed using an in-house script where ring assemblies were separated, long chains are split but functional, and terminal groups are conserved. Contacts corresponding to θ_1 angle of approximatively 70° were dismissed due to the logical fact that they are mostly interacting with the covalently bonded carbon following the isosceles rule (Figure S10 in Supporting Information).

Protein ligand contact data was stored in Discngine's 3decision relational knowledge-based database (<https://3decision.discngine.com/>), which is augmented with data sources such as the PDB,¹ UniProt,⁷⁶ and ChEMBL.⁵² Further meta-data are added to each PDB entry as automatic pocket detection using the fpocket algorithm.⁷⁷

Geometric Descriptors. Distances and angles between halogens and protein residues were computed using the methods described by Sirimulla et al.⁴⁰ and Wilcken et al.²⁴ For the angles, vectors were created alongside the C–X group, the Z–D group and in both directions of X–D atoms as defined in Figures 2 and S11 in Supporting Information. If the halogen bond donor has two covalent bonded atoms, then the geometric center was computed to define the vector. If more than two, then a tetrahedron using the normal vector created with the three covalently bonded atoms was processed to build the vector.

Interaction Characterization. Halogen bonding was defined if an θ_1 angle is greater than 140° for chlorine and 130° for bromine and iodine element and facing a halogen bond acceptor defined as oxygen, nitrogen, sulfur displaying a lone pair or an aromatic center. Hydrogen bonding was characterized if a hydrogen bond donor, i.e., a polar atom

with an implicit hydrogen, was facing an electronegative surface: fluorine atom with a θ_1 angle greater than 70° or a heavy halogen facing the electronegative belt (θ_1 angle between 70 and 110°). Positively charged carbon in amide group was treated as an interaction moiety if the angle between halogen and normal vector of amide group is greater than 155° . Unfavorable contacts are detected when similar charged regions face each other, e.g., hydrogen bond donor facing a σ -hole.

Aromatic Ring Arrangement. Normal vector starting from the aromatic ring center of mass was computed. Angles defined by each vector and the opposing center of mass were called θ_1 angle, e.g., position of protein center of mass relative to halogen aromatic ring, and θ_2 angle, e.g., position of halogen center of mass relative to protein aromatic group. Relative spatial arrangement between two aromatic rings was assessed using the angle between the two normal vectors named arrangement angle.

The labeling of aromatic rings arrangement is described in Table 2 and was inspired from works of Aravinda et al.⁷⁸

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jmedchem.8b01453](https://doi.org/10.1021/acs.jmedchem.8b01453).

Additional data sets (ZIP)

Figures illustrating protein chain redundancy, fragment diversity, definition of geometric values, and additional 3D examples (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: nicolas.shinada@inserm.fr.

ORCID

Nicolas K. Shinada: [0000-0003-2060-9676](https://orcid.org/0000-0003-2060-9676)

Alexandre G. de Brevern: [0000-0001-7112-5626](https://orcid.org/0000-0001-7112-5626)

Peter Schmidtke: [0000-0003-3172-3669](https://orcid.org/0000-0003-3172-3669)

Notes

The authors declare no competing financial interest.

Biographies

Nicolas K. Shinada graduated from Paris VII Diderot University (France) with an M.S. degree in Bioinformatics in 2015. After finishing his Master's degree focused on structural biology, he recently finished his Ph.D. studies in collaboration between Discngine company and DSIMB unit (INSERM, Paris VII University, National Institute of Blood Transfusion), supervised by Peter Schmidtke and Alexandre G. de Brevern, respectively. His research is centered on detection and characterization of molecular interaction in protein–ligand context and protein structure.

Alexandre G. de Brevern, trained as a cell biologist, has been a structural bioinformatician for 20 years. As Senior Researcher at the French National Institute for Health and Medical Research (INSERM), he is the head of the team 2 of INSERM UMR_S 1134 located at the National Institute for Blood Transfusion (INTS). He provided near 20 tools, Web servers, and databases dealing with protein behavior prediction and analyses such as flexibility. Concrete applications to red blood cells and platelets were also extensively studied. He also extended his work to drug design with collaborations with private industries and next-generation sequencing groups. He had authored more than 115 papers, is an editor in five peer-reviewed journals, and is implicated in numerous scientific societies and institutes.

Peter Schmidtke received his M.S. in Bioinformatics (2008) from the University Paris Diderot, France, and Ph.D. (2011) from the University of Barcelona, Spain, under the supervision of Prof. Xavier Barril. He

received the Extraordinary Thesis Award from the Faculty of Pharmacy of the University of Barcelona in 2011. Following postdoctoral studies in Pierre Ducrot's (Ph.D.) CADD group at Servier he joined Discngine in 2013. At Discngine he is now product owner of Discngine 3decision. His scientific contributions include several highly cited papers and open source software such as fpocket, rdock, and dUCK.

■ ACKNOWLEDGMENTS

N.K.S. acknowledges support from ANRT and Discngine S.A.S. This work was supported by grants from the Ministry of Research (France), Discngine S.A.S., University Paris Diderot, Sorbonne, Paris Cite (France), University of La Reunion, Reunion Island, National Institute for Blood Transfusion (INTS, France), National Institute for Health and Medical Research (INSERM, France), and Labex GR-Ex. The Labex GR-Ex, reference ANR-11-LABX-0051, is funded by the program "Investissements d'Avenir" of the French National Research Agency, reference ANR-11-IDEX-0005-02. A.G.d.B. acknowledges Indo-French Centre for the Promotion of Advanced Research/CEFIPRA for Collaborative Grant 5302-2.

■ ABBREVIATIONS USED

Br, bromine; Cl, chlorine; CDK, cyclin dependent kinase; CSD, Cambridge Structural Database; ESP, electrostatic surface potential; F, fluorine; FMAP, fluorine mapping; I, iodine; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; PLIP, Protein-Ligand Interaction Profiler; RCSB, Research Collaboratory for Structural Bioinformatics; SMILES, simplified molecular input line entry specification; UniProt, Universal Protein Resource

■ REFERENCES

- (1) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (2) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: A Program to Generate Schematic Diagrams of Protein-Ligand Interactions. *Protein Eng., Des. Sel.* **1995**, *8*, 127–134.
- (3) Jubb, H. C.; Higuero, A. P.; Ochoa-Montaña, B.; Pitt, W. R.; Ascher, D. B.; Blundell, T. L. Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J. Mol. Biol.* **2017**, *429*, 365–371.
- (4) Deng, Z.; Chuaqui, C.; Singh, J. Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein-Ligand Binding Interactions. *J. Med. Chem.* **2004**, *47*, 337–344.
- (5) Da, C.; Kireev, D. Structural Protein–Ligand Interaction Fingerprints (SPLIF) for Structure-Based Virtual Screening: Method and Benchmark Study. *J. Chem. Inf. Model.* **2014**, *S4*, 2555–2561.
- (6) Lenselink, E. B.; Jespers, W.; van Vlijmen, H. W. T.; IJzerman, A. P.; van Westen, G. J. P. Interacting with GPCRs: Using Interaction Fingerprints for Virtual Screening. *J. Chem. Inf. Model.* **2016**, *S6*, 2053–2060.
- (7) Nittinger, E.; Inhester, T.; Bietz, S.; Meyder, A.; Schomburg, K. T.; Lange, G.; Klein, R.; Rarey, M. Large-Scale Analysis of Hydrogen Bond Interaction Patterns in Protein–Ligand Interfaces. *J. Med. Chem.* **2017**, *60*, 4245–4257.
- (8) Zhang, X.; Gong, Z.; Li, J.; Lu, T. Intermolecular Sulfur...Oxygen Interactions: Theoretical and Statistical Investigations. *J. Chem. Inf. Model.* **2015**, *S5*, 2138–2153.
- (9) Gerebtzoff, G.; Li-Blatter, X.; Fischer, H.; Frentzel, A.; Seelig, A. Halogenation of Drugs Enhances Membrane Binding and Permeation. *ChemBioChem* **2004**, *S*, 676–684.
- (10) Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A. Applications of Fluorine in Medicinal Chemistry. *J. Med. Chem.* **2015**, *S8*, 8315–8359.

- (11) Purser, S.; Moore, P. R.; Swallow, S.; Gouverneur, V. Fluorine in Medicinal Chemistry. *Chem. Soc. Rev.* **2008**, *37*, 320–330.
- (12) Hernandes, M.; Cavalcanti, S. M.; Moreira, D. R.; de Azevedo Junior, W.; Leite, A. C. Halogen Atoms in the Modern Medicinal Chemistry: Hints for the Drug Design. *Curr. Drug Targets* **2010**, *11*, 303–314.
- (13) Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; et al. Systematic Investigation of Halogen Bonding in Protein-Ligand Interactions. *Angew. Chem., Int. Ed.* **2011**, *50*, 314–318.
- (14) Politzer, P.; Lane, P.; Concha, M. C.; Ma, Y.; Murray, J. S. An Overview of Halogen Bonding. *J. Mol. Model.* **2007**, *13*, 305–311.
- (15) Valerio, G.; Raos, G.; Meille, S. V.; Metrangolo, P.; Resnati, G. Halogen Bonding in Fluoroalkylhalides: A Quantum Chemical Study of Increasing Fluorine Substitution. *J. Phys. Chem. A* **2000**, *104*, 1617–1620.
- (16) Riley, K. E.; Murray, J. S.; Fanfrlík, J.; Rezáč, J.; Solá, R. J.; Concha, M. C.; Ramos, F. M.; Politzer, P. Halogen Bond Tunability I: The Effects of Aromatic Fluorine Substitution on the Strengths of Halogen-Bonding Interactions Involving Chlorine, Bromine, and Iodine. *J. Mol. Model.* **2011**, *17*, 3309–3318.
- (17) Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; Plancher, J. M.; Hartmann, G.; Isshiki, Y.; Morikami, K.; Shimma, N.; Haap, W.; Banner, D. W.; Diederich, F. Halogen Bonding at the Active Sites of Human Cathepsin L and MEK1 Kinase: Efficient Interactions in Different Environments. *ChemMedChem* **2011**, *6*, 2048–2054.
- (18) Metrangolo, P.; Murray, J. S.; Pilati, T.; Politzer, P.; Resnati, G.; Terraneo, G. The Fluorine Atom as a Halogen Bond Donor, Viz. a Positive Site. *CrystEngComm* **2011**, *13*, 6593.
- (19) Kolár, M. H.; Hobza, P. Computer Modeling of Halogen Bonds and Other σ -Hole Interactions. *Chem. Rev.* **2016**, *116*, 5155–5187.
- (20) Mulliken, R. S. Structures of Complexes Formed by Halogen Molecules with Aromatic and with Oxygenated Solvents. *J. Am. Chem. Soc.* **1950**, *72*, 600–608.
- (21) Hassel, O.; Hvoslef, J.; Vihovde, E. H.; Sørensen, N. A. The Structure of Bromine 1,4-Dioxanate. *Acta Chem. Scand.* **1954**, *8*, 873–873.
- (22) Legon, A. C. Prereactive Complexes of Dihalogens XY with Lewis Bases B in the Gas Phase: A Systematic Case for the Halogen Analogue B \cdots XY of the Hydrogen Bond B \cdots HX. *Angew. Chem., Int. Ed.* **1999**, *38*, 2686–2714.
- (23) Jorgensen, W. L.; Schyman, P. Treatment of Halogen Bonding in the OPLS-AA Force Field: Application to Potent Anti-HIV Agents. *J. Chem. Theory Comput.* **2012**, *8*, 3895–3901.
- (24) 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*, 1363–1388.
- (25) Scholfield, M. R.; Ford, M. C.; Vander Zanden, C. M.; Billman, M. M.; Ho, P. S.; Rappe, A. K. Force Field Model of Periodic Trends in Biomolecular Halogen Bonds. *J. Phys. Chem. B* **2015**, *119*, 9140–9149.
- (26) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen Bonds in Biological Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16789–16794.
- (27) Bauzá, A.; Quiñonero, D.; Deyà, P. M.; Frontera, A. Halogen Bonding Versus Chalcogen and Pnicogen Bonding: A Combined Cambridge Structural Database and Theoretical Study. *CrystEngComm* **2013**, *15* (16), 3137–3144.
- (28) Maignan, S.; Guilloteau, J.-P.; Choi-Sledeski, Y. M.; Becker, M. R.; Ewing, W. R.; Pauls, H. W.; Spada, A. P.; Mikol, V. Molecular Structures of Human Factor Xa Complexed with Ketopiperazine Inhibitors: Preference for a Neutral Group in the S1 Pocket. *J. Med. Chem.* **2003**, *46*, 685–690.
- (29) Rowlinson, S. W.; Kiefer, J. R.; Prusakiewicz, J. J.; Pawlitz, J. L.; Kozak, K. R.; Kalgutkar, A. S.; Stallings, W. C.; Kurumbail, R. G.; Marnett, L. J. A Novel Mechanism of Cyclooxygenase-2 Inhibition Involving Interactions with Ser-530 and Tyr-385. *J. Biol. Chem.* **2003**, *278*, 45763–45769.
- (30) Koch, C.; Heine, A.; Klebe, G. Tracing the Detail: How Mutations Affect Binding Modes and Thermodynamic Signatures of Closely Related Aldose Reductase Inhibitors. *J. Mol. Biol.* **2011**, *406*, 700–712.
- (31) Murray-Rust, P.; Motherwell, W. D. S. Computer Retrieval and Analysis of Molecular Geometry. 4. Intermolecular Interactions. *J. Am. Chem. Soc.* **1979**, *101*, 4374–4376.
- (32) Brammer, L.; Bruton, E. A.; Sherwood, P. Understanding the Behavior of Halogens as Hydrogen Bond Acceptors. *Cryst. Growth & Des.* **2001**, *1* (4), 277–290.
- (33) Lin, F.-Y.; MacKerell, A. D. Do Halogen–Hydrogen Bond Donor Interactions Dominate the Favorable Contribution of Halogens to Ligand–Protein Binding? *J. Phys. Chem. B* **2017**, *121*, 6813–6821.
- (34) Ontoria, J. M.; Rydberg, E. H.; Di Marco, S.; Tomei, L.; Attenni, B.; Malancona, S.; Martin Hernando, J. I.; Gennari, N.; Koch, U.; Narjes, F.; Rowley, M.; Summa, V.; Carroll, S. S.; Olsen, D. B.; De Francesco, R.; Altamura, S.; Migliaccio, G.; Carfi, A. Identification and Biological Evaluation of a Series of 1H-Benz[de]Isoquinoline-1,3(2H)-Diones as Hepatitis C Virus NSSB Polymerase Inhibitors. *J. Med. Chem.* **2009**, *52*, 5217–5227.
- (35) Howard, J. A. K.; Hoy, V. J.; O'Hagan, D.; Smith, G. T. How Good Is Fluorine as a Hydrogen Bond Acceptor? *Tetrahedron* **1996**, *52*, 12613–12622.
- (36) Dunitz, J. D.; Taylor, R. Organic Fluorine Hardly Ever Accepts Hydrogen Bonds. *Chem. - Eur. J.* **1997**, *3*, 89–98.
- (37) Dalvit, C.; Vulpetti, A. Intermolecular and Intramolecular Hydrogen Bonds Involving Fluorine Atoms: Implications for Recognition, Selectivity, and Chemical Properties. *ChemMedChem* **2012**, *7*, 262–272.
- (38) Dalvit, C.; Invernizzi, C.; Vulpetti, A. Fluorine as a Hydrogen-Bond Acceptor: Experimental Evidence and Computational Calculations. *Chem. - Eur. J.* **2014**, *20*, 11058–11068.
- (39) Dalvit, C.; Vulpetti, A. Weak Intermolecular Hydrogen Bonds with Fluorine: Detection and Implications for Enzymatic/Chemical Reactions, Chemical Properties, and Ligand/Protein Fluorine NMR Screening. *Chem. - Eur. J.* **2016**, *22*, 7592–7601.
- (40) Sirimulla, S.; Bailey, J. B.; Vegeasna, R.; Narayan, M. Halogen Interactions in Protein–Ligand Complexes: Implications of Halogen Bonding for Rational Drug Design. *J. Chem. Inf. Model.* **2013**, *53*, 2781–2791.
- (41) Bauer, M. R.; Jones, R. N.; Baud, M. G. J.; Wilcken, R.; Boeckler, F. M.; Fersht, A. R.; Joerger, A. C.; Spencer, J. Harnessing Fluorine–Sulfur Contacts and Multipolar Interactions for the Design of P53 Mutant Y220C Rescue Drugs. *ACS Chem. Biol.* **2016**, *11*, 2265–2274.
- (42) Kim, D.; Wang, L.; Beconi, M.; Eiermann, G. J.; Fisher, M. H.; He, H.; Hickey, G. J.; Kowalchick, J. E.; Leiting, B.; Lyons, K.; Marsilio, F.; McCann, M. E.; Patel, R. A.; Petrov, A.; Scapin, G.; Patel, S. B.; Sinha Roy, R.; Wu, J. K.; Wyvratt, M. J.; Zhang, B. B.; Zhu, L.; Thornberry, N. A.; Weber, A. E. (2R)-4-Oxo-4-[3-(Trifluoromethyl)-5,6-Dihydro-[1,2,4]Triazolo[4,3-a]Pyrazin-7(8H)-YL]-1-(2,4,5-Trifluorophenyl)-Butan-2-Amine: A Potent, Orally Active Dipeptidyl Peptidase IV Inhibitor for the Treatment of Type 2 Diabetes. *J. Med. Chem.* **2005**, *48* (1), 141–151.
- (43) Garau, C.; Frontera, A.; Quiñonero, D.; Ballester, P.; Costa, A.; Deyà, P. M. A Topological Analysis of the Electron Density in Anion-Pi Interactions. *ChemPhysChem* **2003**, *4* (12), 1344–1348.
- (44) Wang, H.; Wang, W.; Jin, W. J. σ -Hole Bond vs π -Hole Bond: A Comparison Based on Halogen Bond. *Chem. Rev.* **2016**, *116*, 5072–5104.
- (45) Zhang, Y.; Ji, B.; Tian, A.; Wang, W. Communication: Competition Between $\pi\cdots\pi$ Interaction and Halogen Bond in Solution: A Combined ^{13}C NMR and Density Functional Theory Study. *J. Chem. Phys.* **2012**, *136* (14), 141101.
- (46) Ma, N.; Zhang, Y.; Ji, B.; Tian, A.; Wang, W. Structural Competition Between Halogen Bonds and Lone-Pair $\cdots\pi$ Interactions in Solution. *ChemPhysChem* **2012**, *13* (6), 1411–1414.
- (47) Paolo, T. D.; Sandorf, C. On the Hydrogen Bond Breaking Ability of Fluorocarbons Containing Higher Halogens. *Can. J. Chem.* **1974**, *S2*, 3612–3622.

- (48) Zhou, P.-P.; Qiu, W.-Y.; Liu, S.; Jin, N.-Z. Halogen as Halogen-Bonding Donor and Hydrogen-Bonding Acceptor Simultaneously in Ring-Shaped H₃N-X(Y)-HF (X = Cl, Br and Y = F, Cl, Br) Complexes. *Phys. Chem. Chem. Phys.* **2011**, *13* (16), 7408–7418.
- (49) Yang, X.; Zhou, P.-P.; Zhou, D.-G.; Zheng, P.-J.; Dai, Y.; Yan, C.-X.; Yang, Z. Simultaneous Interactions of Amphoteric Halogen in XY (X = Cl, Br and Y = F, Cl, Br) with C and O Atoms of CO₂ in Ring-Shaped CO₂ ·X(Y)·CO₂ Complexes. *Comput. Theor. Chem.* **2016**, *1076*, 32–41.
- (50) Lu, Y.; Wang, Y.; Xu, Z.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. C-X··H Contacts in Biomolecular Systems: How They Contribute to Protein-Ligand Binding Affinity. *J. Phys. Chem. B* **2009**, *113* (37), 12615–12621.
- (51) Lu, Y.; Wang, Y.; Zhu, W. Nonbonding Interactions of Organic Halogens in Biological Systems: Implications for Drug Discovery and Biomolecular Design. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4543.
- (52) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42*, D1083–D1090.
- (53) Shah, P.; Westwell, A. D. The Role of Fluorine in Medicinal Chemistry: Review Article. *J. Enzyme Inhib. Med. Chem.* **2007**, *22* (5), S27–S40.
- (54) Holland, J.; Frei, E. *Holland-Frei Cancer Medicine*, 6th ed.; Kufe, D. W., Pollack, R. E., Weichselbaum, R. R., Bast, R. C., Jr., Gansler, T. S., Holland, J. F., Emil Frei, I. I. I., Eds.; BC Decker Inc., 2003.
- (55) Voth, A. R.; Ho, P. S. The Role of Halogen Bonding in Inhibitor Recognition and Binding by Protein Kinases. *Curr. Top. Med. Chem.* **2007**, *7* (14), 1336–1348.
- (56) Gregoret, L. M.; Rader, S. D.; Fletterick, R. J.; Cohen, F. E. Hydrogen Bonds Involving Sulfur Atoms in Proteins. *Proteins: Struct., Funct., Genet.* **1991**, *9*, 99–107.
- (57) Zhou, P.; Tian, F.; Lv, F.; Shang, Z. Geometric Characteristics of Hydrogen Bonds Involving Sulfur Atoms in Proteins. *Proteins: Struct., Funct., Genet.* **2009**, *76*, 151–163.
- (58) Voth, A. R.; Khuu, P.; Oishi, K.; Ho, P. S. Halogen Bonds as Orthogonal Molecular Interactions to Hydrogen Bonds. *Nat. Chem.* **2009**, *1*, 74–79.
- (59) Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010**, *53*, 5061–5084.
- (60) Choudhary, A.; Gandla, D.; Krow, G. R.; Raines, R. T. Nature of Amide Carbonyl-Carbonyl Interactions in Proteins. *J. Am. Chem. Soc.* **2009**, *131*, 7244–7246.
- (61) Carosati, E.; Sciabola, S.; Cruciani, G. Hydrogen Bonding Interactions of Covalently Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field. *J. Med. Chem.* **2004**, *47*, 5114–5125.
- (62) Kovacs, A.; Varga, Z. Halogen Acceptors in Hydrogen Bonding. *Coord. Chem. Rev.* **2006**, *250*, 710–727.
- (63) Zhou, P.; Zou, J.; Tian, F.; Shang, Z. Fluorine Bonding — How Does It Work In Protein-Ligand Interactions? *J. Chem. Inf. Model.* **2009**, *49*, 2344–2355.
- (64) Olsen, J. A.; Banner, D. W.; Seiler, P.; Obst Sander, U.; D'Arcy, A.; Stihle, M.; Müller, K.; Diederich, F. A Fluorine Scan of Thrombin Inhibitors to Map the Fluorophilicity/Fluorophobicity of an Enzyme Active Site: Evidence for C—F···C=O Interactions. *Angew. Chem., Int. Ed.* **2003**, *42* (22), 2507–2511.
- (65) Pollock, J.; Borkin, D.; Lund, G.; Purohit, T.; Dyguda-Kazimierowicz, E.; Grembecka, J.; Cierpicki, T. Rational Design of Orthogonal Multipolar Interactions with Fluorine in Protein–Ligand Complexes. *J. Med. Chem.* **2015**, *58*, 7465–7474.
- (66) Paulini, R.; Müller, K.; Diederich, F. Orthogonal Multipolar Interactions in Structural Chemistry and Biology. *Angew. Chem., Int. Ed.* **2005**, *44*, 1788–1805.
- (67) Mendez, L.; Henriquez, G.; Sirimulla, S.; Narayan, M. Looking Back, Looking Forward at Halogen Bonding in Drug Discovery. *Molecules* **2017**, *22*, 1397.
- (68) Hunter, C. A.; Singh, J.; Thornton, J. M. π-π Interactions: The Geometry and Energetics of Phenylalanine-Phenylalanine Interactions in Proteins. *J. Mol. Biol.* **1991**, *218*, 837–846.
- (69) Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. Aromatic Interactions. *J. Chem. Soc. Perkin Trans. 2* **2001**, 651–669.
- (70) Martinez, C. R.; Iverson, B. L. Rethinking the Term “Pi-Stacking”. *Chem. Sci.* **2012**, *3*, 2191.
- (71) 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*, 69–78.
- (72) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: Fully Automated Protein–Ligand Interaction Profiler. *Nucleic Acids Res.* **2015**, *43*, W443–W447.
- (73) Gutierrez, I. S.; Lin, F.-Y.; Vanommeslaeghe, K.; Lemkul, J. A.; Armacost, K. A.; Brooks, C. L.; MacKerell, A. D. Parametrization of Halogen Bonds in the CHARMM General Force Field: Improved Treatment of Ligand–Protein Interactions. *Bioorg. Med. Chem.* **2016**, *24*, 4812–4825.
- (74) McNaught, A. D.; Wilkinson, A. *Compendium of Chemical Terminology (IUPAC Chemical Data)*; Wiley, 1997.
- (75) Dassault Systemes BIOVIA, *BIOVIA Pipeline Pilot*, release 2017; Dassault Systemes, 2017.
- (76) The UniProt Consortium.. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* **2016**, *45*, D158–D169.
- (77) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinf.* **2009**, *10*, 168.
- (78) Aravinda, S.; Shamala, N.; Das, C.; Sriranjini, A.; Karle, I. L.; Balaram, P. Aromatic-Aromatic Interactions in Crystal Structures of Helical Peptide Scaffolds Containing Projecting Phenylalanine Residues. *J. Am. Chem. Soc.* **2003**, *125*, 5308–5315.