JOURNAL OF CHEMICAL INFORMATION AND MODELING

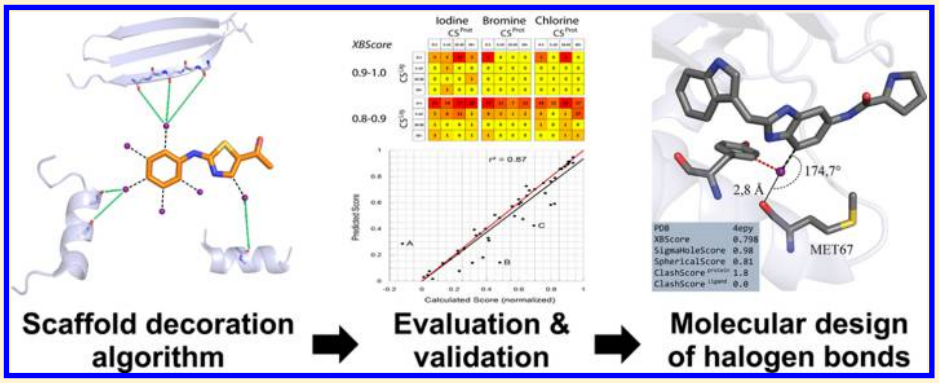
Article

pubs.acs.org/jcim

Evaluating the Potential of Halogen Bonding in Molecular Design: Automated Scaffold Decoration Using the New Scoring Function XBScore

Markus O. Zimmermann, Andreas Lange, and Frank M. Boeckler*

Laboratory for Molecular Design and Pharmaceutical Biophysics, Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmacy, Eberhard-Karls-University Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany



ABSTRACT: We present a QM-derived empirical scoring function for the interaction between aromatic halogenated ligands and the carbonyl oxygen atom of the protein backbone. Applying this scoring function, we developed an algorithm that evaluates the potential of protein-bound ligands to form favorable halogen-bonding contacts upon scaffold decoration with chlorine, bromine, or iodine. Full recovery of all existing halogen bonds in the PDB involving the protein backbone was achieved with our protocol. Interestingly, the potential for introducing halogen bonds through scaffold decoration of unsubstituted aromatic carbon atoms appears to easily match the number of previously known halogen bonds. Our approach can thus be used as a blueprint for integration of halogen bonding into general empirical scoring functions, which at present ignore this interaction. Most importantly, we were able to identify a substantial number of protein−ligand complexes where the benefits and challenges of introducing a halogen bond by molecular design can be studied experimentally.

INTRODUCTION

Halogen bonding is increasingly gaining recognition in the fields of medicinal chemistry and molecular design.[1,2] Various examples of the application of halogen bonding in biological systems have recently been reported.[3−6] However, integration into the drug design process is hampered by the complex nature of this interaction. Halogen bonding can be described as a favorable but rather directional interaction between an electropositive region on a halogen atom, the σ-hole,[7,8] and a nucleophilic interaction partner such as an oxygen atom or the π system of an aromatic moiety.[9] It should be noted that the σ-hole interaction (Figure 1) is typically recognized as the signature aspect of halogen bonding. However, interactions with the negative belt of the anisotropic electron distribution (δ^- interactions) may additionally contribute to or even synergistically enhance the overall effect.

Our recent analysis of the Protein Data Bank (PDB)[11] (as of Feb 2, 2014) revealed that good halogen-bonding interactions between ligands and protein-backbone carbonyl oxygen atoms occur about twice as often as interactions with any other side-chain atoms combined (π systems not included). This certainly correlates to the more frequent occurrence of carbonyl oxygen atoms compared with all other interaction partners in a protein binding site.[12,13] Accessibility and flexibility issues might also play a role in this statistical assessment.[1]

In this work, we strive to compare the potential of introducing halogen bonds into existing protein−ligand complexes by molecular design to the number of halogen bonds found in the PDB. We focus on targeting the carbonyl oxygen atom by halogen bonding and on evaluating the quality of the resulting halogen bonds in relation to a reference system.

For the drug discovery process, halogenated aromatic or heteroaromatic scaffolds are the most useful halogen-bond donors. Alkyl halides and other reactive species can form unspecific, irreversible covalent bonds to the target protein, causing severe problems with drug safety aspects. Still, concerns about adverse ADMET properties as well as chemical and metabolic stability, particularly of the heavier aryl halides, have also been raised and discussed, possibly limiting the applicability in the drug discovery process.[1,2] Using quantum-chemical calculations, we have evaluated the interaction

Received: December 1, 2014
Published: February 5, 2015

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699

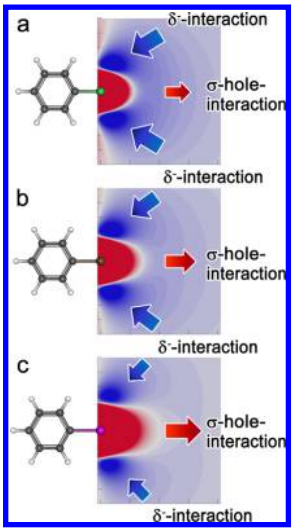


Figure 1. 2D plots of the electrostatic potentials of (a) chlorobenzene, (b) bromobenzene, and (c) iodobenzene calculated at the MP2/TZVPP level of theory, highlighting halogen-bonding contributions from the σ-hole interaction and potential additive or synergistic δ^- interactions with the negative electrostatic belt. The shielding of the σ-hole is larger for chlorine than for bromine and iodine. Thus, its σ-hole interactions may be weaker and its δ^- interactions can be stronger. In contrast, σ-hole interactions will be stronger and δ^- interactions can be weaker for iodine. The figure was generated using MOLCAD.[10]

energies between halobenzenes (serving as simple model ligands for halogen-bond donors) and the oxygen of *N*-methylacetamide (representing a carbonyl function of a protein backbone). For this purpose, we have extended previously reported QM studies[14] quite significantly.

In a protein−ligand complex, truly optimal interaction geometries are very rarely observed. Often, attractive interactions can compete with each other, causing the binding mode to deviate from the optimal interaction geometry. Thus, usually a blend of good, moderate, and suboptimal geometries are to be expected. In order to investigate all spatial dependences of the halogen bond, we systematically altered the distance, the σ-hole angle, and the spatial orientation of our model systems in small steps and calculated the interaction energies as MP2/TZVPP single-point values. On the basis of these calculations, we developed a method for scaffold decoration of aromatic moieties that makes use of a new scoring function for halogen bonding that can be applied to any protein-bound ligand in a crystal structure or to any docking pose. Hence, the scoring function can be used either to evaluate existing halogen bonds or to assess the quality of putative halogen bonds arising from scaffold decoration approaches in molecular design. With this scoring function, we aim to foster the applicability of halogen bonding in applied molecular design projects in academia and the pharmaceutical industry.

FROM QM TO SCORING FUNCTIONS

Apart from computational resources, the success of using a specific program for molecular design purposes mainly relies on the application of a suitable force field or scoring function.[15−18] Generally, there are three types of scoring functions: knowledge-based scoring functions, empirical scoring functions, and force-field-based scoring functions. Knowledge-based scoring functions are derived from deep statistical analyses of experimental data using atom pair potentials. These analyses, however, are always dependent on the quality and composition of the respective data sets. On the basis of the currently available crystallographic data on halogen bonds in the PDB, parts of the required parameter space for establishing a knowledge-based scoring function may still be underrepresented. The frequencies of occurrence of the halogens iodine, bromine, and chlorine in a protein-bound ligand differ substantially. More importantly, target-dependent bias can lead to overestimation of the binding strengths of certain interaction geometries simply due to their large abundance. For example, by analysis of all chlorine halogen-bonding contacts in the PDB, one certain recurring interaction stands out. In a large number of crystal structures of Factor Xa, the ligand invariably forms a chlorine halogen bond with the protein's Tyr228 residue in exactly the same orientation.[19]

Common force fields aim to describe covalent and noncovalent interactions of and between molecules by calculating distinct energy terms such as bond stretching or van der Waals interactions of atom pairs that are derived from simplified physical models. Since each atom is described by an atom type and a partial charge, this method fails to describe the characteristic anisotropic electron distribution around a halogen atom correctly.

Jorgensen and Schyman[20] solved this problem by adding an additional positive partial charge in the region of the σ-hole to their force field implementation of OPLS-AA. In contrast, Ho and colleagues[21] created a force field (*ff*BXB) that provides an anisotropic treatment of both the shape and electrostatic charge parameters of bromine. They suggested that electrostatics alone cannot account for the very short distances of bromine halogen bonds and that a flattening of the effective van der Waals radius is required. This approach was recently also extended to chlorine and iodine.[22] With EMPIRE,[23] Clark and Hennemann generated a massively parallel implementation of restricted Hartree−Fock (RHF) neglect of diatomic differential overlap (NDDO)-based self-consistent field (SCF) calculations, which has been demonstrated to reproduce halogen bonds in crystal structures known from crystal engineering studies.[24]

Řezáč and Hobza improved the semiempirical QM method PM6, which in its unaugmented form fails to handle halogen bonding correctly, leading to underestimated interaction distances and overestimated interaction energies due to a lack of repulsion. In PM6-D3H4X, an empirical energy correction term is applied to the method to compensate for this problem and other issues.[25−27]

Empirical scoring functions try to estimate the binding affinity of a protein−ligand complex by approximating different energy terms representing the contributions of hydrogen bonds, van der Waals interactions, and other interactions to overall binding. Kuhn et al.[28] recently introduced the empirical scoring function ScorpionScore that recognizes halogen bonds by applying distance and angle cutoffs.

Because halogen bonds have been argued to consist of varying contributions from electrostatics, dispersion, polarization, and charge transfer,[29−35] it is difficult to find a simplified form to accurately describe the geometry-dependent behavior within a reasonable margin of error. At the currently available level of simplification, it is difficult to make a suitable assessment of the quality of a halogen bond or whether the term *halogen bonding* is even applicable. Undisputedly, quantum-chemical calculations at the MP2 level of theory can be a valuable source for the description of the interactions of aryl halides with nucleophiles in the protein binding site. We

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699

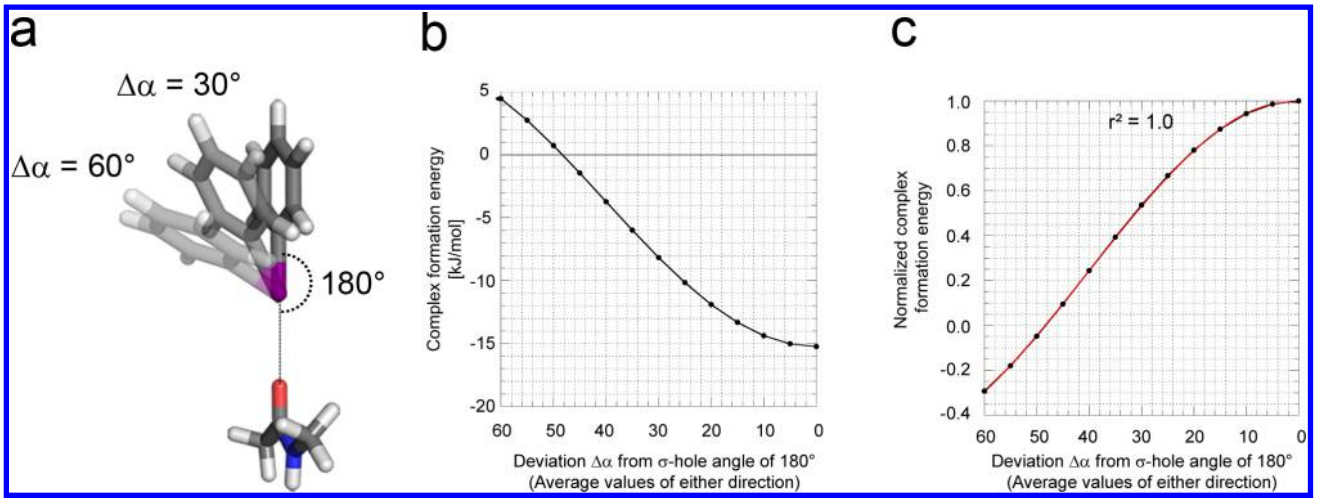


Figure 2. σ-Hole scans and curve fitting for iodobenzene and *N*-methylacetamide. (a) Illustration of the σ-hole scan. (b) Impact of nonideal σ-hole angle on the interaction energy. (c) Normalized energy curve (black dots) and fitted function with a correlation coefficient of 1.0 (red curve).

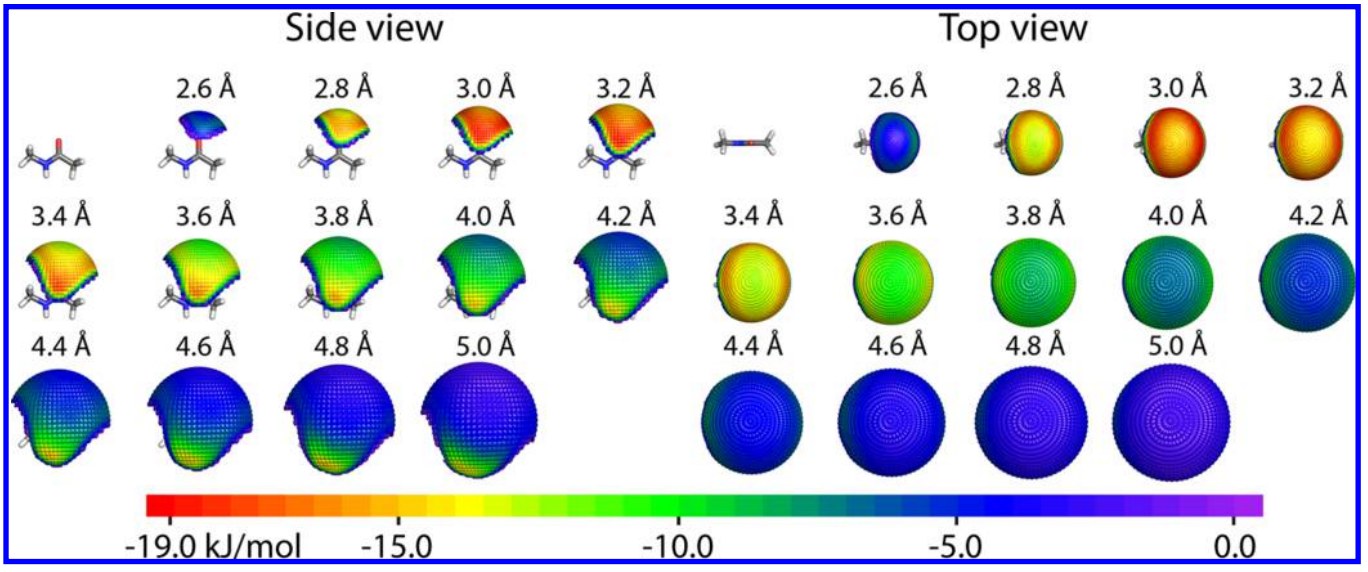


Figure 3. Spherical scans of iodobenzene and *N*-methylacetamide. With a fixed σ-hole angle of 180°, the spherical scan was performed at 13 different distances. Each depicted point represents the coordinate of the iodine atom in one of 2664 MP2/TZVPP single-point calculations colored by the interaction energy. For the sake of clarity, only favorable interaction energies are shown.

have found in recent studies that MP2/TZVPP-calculated energies are in reasonable agreement with CCSD(T) reference calculations using a complete basis set (CBS) extrapolation scheme.

On the basis of approximately 35 000 MP2 single-point calculations for each of the halogens chlorine, bromine, and iodine, we derived two separate components of the final scoring function for the interaction between a halobenzene (iodobenzene, bromobenzene, or chlorobenzene) and the oxygen atom of *N*-methylacetamide (representing the protein backbone): a σ-hole score (termed *SigmaHoleScore*) to address the directionality of the halogen bond and a spherical score (denoted as *SphericalScore*) for the spatial position of the ligand system with respect to the oxygen atom. The two scores are combined in one halogen-bonding score, called the *XBScore*.

The σ-Hole Score. Halogen-bonding strength is affected by a number of different factors such as interaction geometry, tuning effects, the interaction partner, and the type of the halogen. The most important element of halogen-bond directionality is the σ-hole angle. Deviations from the optimum of close to 180° diminish the overall interaction strength quite significantly.

Subsequently, we will focus on discussing the interaction between iodobenzene and *N*-methylacetamide. Similar results were obtained for bromobenzene and chlorobenzene.

The starting geometry of iodobenzene and the backbone model system was set to a σ-hole angle of 180°. Iodobenzene was placed at an optimal distance from the oxygen (around 3.0 Å) on the elongation of the carbonyl carbon−carbonyl oxygen bond (Figure 2a). In this linear geometry, the interaction energy amounts to −15.2 kJ. Deviations of 20° from the optimal angle in either direction result in an energy loss of around 22%, decreasing the interaction energy to −11.9 kJ. An additional deviation of 20° leads to a loss of more than 75%, with a total energy of only −3.7 kJ. At σ-hole angles below 130°, the interaction becomes unfavorable and the interaction energy turns positive (Figure 2b).

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699

In order to describe the quality of the interaction geometry with respect to the σ-hole angle as a relative quantity, all of the values were normalized to the best obtained energy. After normalization, the scores ranged between 0 (worst value) and 1 (best value). Negative values (representing areas of repulsive behavior) were set to zero. These normalized values were used to perform a curve fit with a cubic polynomial of the form $f(x) = ax^3 + bx^2 + cx + d$. The following polynomial, with a coefficient of correlation (r^2) of 1.0, was obtained using MATLAB:

$$SigmaHoleScore_{Iodine}(x) = 6.556\cdot10^{-6}x^3 + 2.747\cdot10^{-3}x^2 - 0.3542x + 13.98$$

The polynomial is shown in Figure 2c as the red curve overlaid with the normalized values. This equation represents the first part of the *XBScore*. The respective polynomial equations fitted to the calculated values for bromine and chlorine are the following:

$$SigmaHoleScore_{Bromine}(x) = 8.347\cdot10^{-6}x^3 + 3.564\cdot10^{-3}x^2 - 0.4747x + 19.64$$

$$SigmaHoleScore_{Chlorine}(x) = 1.022\cdot10^{-5}x^3 + 4.459\cdot10^{-3}x^2 - 0.6157x + 26.93$$

The Spherical Score. The second part of the *XBScore* is derived from interaction spheres between the halobenzene and *N*-methylacetamide with a fixed σ-hole angle of 180° at 13 different distances ranging from 2.6 to 5.0 Å in steps of 0.2 Å. These interaction spheres combine the two remaining spatial variables of halogen bonding: the interaction distance and the spatial position of the halobenzene with respect to the carbonyl function. Each interaction sphere was obtained by calculating MP2/TZVPP single-point energies of 2664 different interaction geometries. Figure 3 presents all 13 spheres for iodobenzene and *N*-methylacetamide, viewed from the side and from the top.

The dissimilar color gradient profiles at different distances indicate that the spatial variables distance and spherical orientation are in some areas rather interdependent and hence should not be treated as separate contributions but rather as one. It is not possible to interpolate the energies of one interaction sphere to its adjacent spheres using a simple mathematical operation. Visualization of clusters of favorable spatial interaction geometries for chlorine, bromine, and iodine grouped by interaction energies (Figure 4) reveals a significant transition from n-electron to π-electron interactions for the weaker σ-holes of bromine and chlorine compared with iodine. Because of this transition behavior, separate approaches are required in order to describe each halobenzene system correctly.

The second part of the *XBScore* is implemented as a table lookup, as no reasonable three-dimensional mathematical fit was possible without loss of information. In all of the interaction spheres for chlorobenzene, bromobenzene, and iodobenzene, the orientation of *N*-methylacetamide in the coordinate system is identical. Hence, the data can be reduced to the coordinates of the halogen atom and the corresponding interaction energy. As before, the interaction energy is normalized to a value between 0 and 1 using the best observed interaction energy of any sphere. All positive energies (unfavorable interactions) are set to a score of 0. The adduct

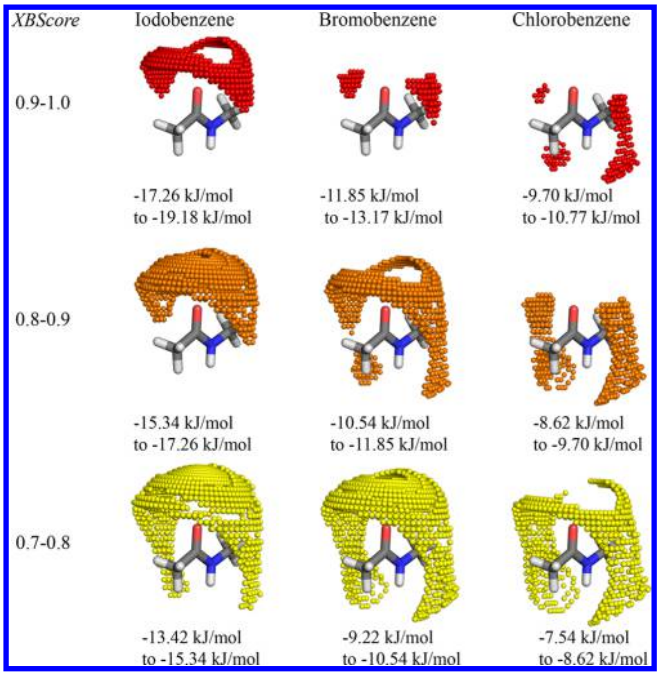


Figure 4. Depiction of interaction geometries for the top three levels of *XBScore* for iodobenzene, bromobenzene, and chlorobenzene. Each point represents the position of the respective halogen atom. The halobenzene ligand is not depicted explicitly. The interaction energies presented for each class are based on a linear orientation of the halobenzene (σ-hole angle of 180°) and MP2/TZVPP calculations.

formation energies of the top three *XBScore* classes (1.0−0.9, 0.9−0.8, and 0.8−0.7, as depicted in Figure 4) range from −19.2 to −13.4 kJ/mol for iodobenzene, −13.2 to −9.2 kJ/mol for bromobenzene, and −10.8 to −7.5 kJ/mol for chlorobenzene, respectively. For iodobenzene they include 3617 interaction geometries. The number of interaction geometries with an *XBScore* of >0.7 was 2348 for bromobenzene and only 1170 for chlorobenzene.

Application to PDB Files. In order to allocate the correct spherical score to a halogen atom in the ligand of a PDB file, the coordinates of the protein-backbone carbonyl function are matched onto the backbone model system as accurately as possible by minimizing the root-mean-square deviation (RMSD). The halogen is translated accordingly, and the nearest neighbor within all 13 spherical scans of the respective halogen is identified. If there is more than one match within 0.07 Å, the average of all normalized energies is taken.

To illustrate the usefulness of this *XBScore* for drug discovery purposes, we have implemented it in a PDB-mining protocol in order to assess the potential of introducing halogen bonds into protein−ligand crystal structures by molecular design.

The Automated Scaffold Decoration Protocol. First, a protein preparation step is conducted that removes all hydrogen atoms, metal cations, and water molecules from the PDB file. Then the protein is checked for ligand molecules that are not covalently bound to the protein and contain five or more heavy atoms (Figure 5a). For each of these ligands, all of the ring systems are identified and checked for planarity to select aromatic rings only (Figure 5b). Aromaticity assessment is done by calculating the average of all possible dihedral angles within the ring. The dihedral angle of four atoms in a plane should be close to 0°, but here an average deviation of 5° is tolerated. This step is done to compensate for artifacts

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699

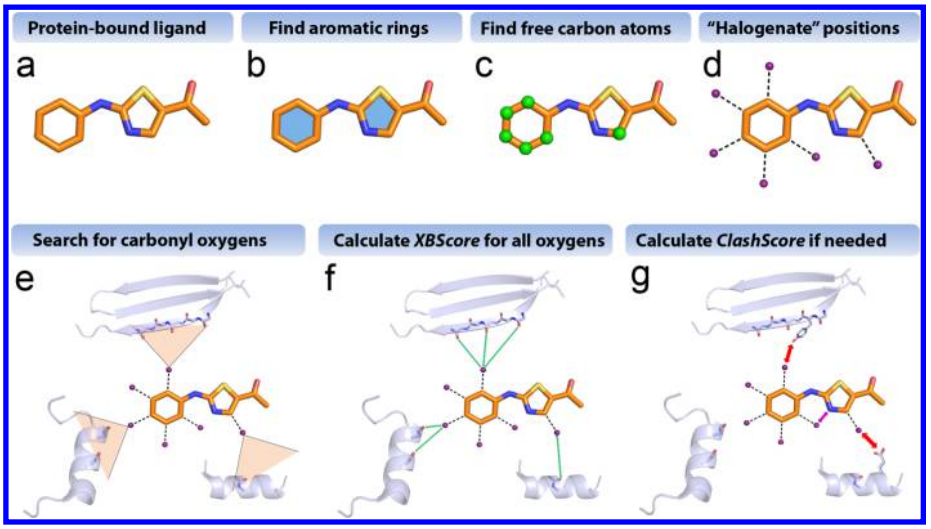


Figure 5. Explanation of the halogen-bonding-based scaffold decoration algorithm. When a protein-bound ligand is found (a), the algorithm searches for aromatic ring systems in the ligand (b). Next, it checks for unsubstituted aromatic carbon atoms (c). If such an atom is found, the respective position is halogenated with chlorine, bromine, or iodine using bond lengths acquired through QM calculations (d). Then the protein surroundings are searched for carbonyl oxygen atoms that could act as halogen-bond acceptors (e). For each halogen−oxygen pair, an *XBScore* is calculated to assess the quality of the putative halogen bond (f). Additionally, a *ClashScore* is calculated to check for van der Waals clashes of the newly introduced halogen with protein atoms (red arrows) or the ligand (purple arrow) (g).

generated by mapping atoms onto electron densities, which can lead to inaccurate atom placement in the crystal structure. In addition, we do not have to rely on the aromaticity definitions of other programs or file formats, which are often incomplete.

Next, all of the aromatic carbon atoms are checked for substituents (Figure 5c), and each unsubstituted carbon atom is subsequently used for halogen-bonding-based scaffold decoration through putative halogenation using chlorine, bromine, or iodine (Figure 5d). It has to be noted that this step does not take into account whether this halogenation is feasible or how reactive or stable the resulting molecule would be. Unsubstituted aromatic heteroatoms such as nitrogen and oxygen are always directly omitted from halogenation.

In order to obtain the direction vector for the placement of the new halogen, we apply the *h_add* feature within PyMol to the position of interest. The halogen atom is translated along this vector until a halogen−carbon bond distance from the respective carbon atom corresponding to that in the MP2/TZVPP-optimized structure of iodobenzene (2.08 Å), bromobenzene (1.89 Å), or chlorobenzene (1.73 Å) is reached. For each newly introduced halogen atom, the surrounding binding site is searched for carbonyl oxygen atoms that could possibly act as halogen-bond acceptors (Figure 5e). For each halogen···$O_{carbonyl}$ contact within 5.0 Å, the *XBScore* is calculated (Figure 5f).

As the newly introduced halogen atom can cause perturbations in the ligand and the protein, e.g., through van der Waals clashes, we have introduced a score, denoted as *ClashScore*, that evaluates all of the halogen contacts with protein atoms (red arrows) and ligand atoms (purple arrow) that are located closer than the sum of their van der Waals radii (Figure 5g). However, we also need to take into consideration the fact that proteins are typically not rigid structures and sometimes can be highly flexible. From small adaptive movements of flexible side chains in the binding site to whole induced-fit rearrangements of loop regions forming the binding site, various degrees of protein flexibility have been observed. Hence, it is impossible to precisely predict, whether a "clash" in the perturbed system will be a "clash" in a relaxed system. As a consequence, *ClashScore* is intended to be used as a "miner's canary" reporting possible issues with the halogen placement by detecting putative sources for generating conformational strain in the ligand or the binding site upon halogenation of the respective unsubstituted carbon position.

The Clash Score. *ClashScore* is a crude measure of the severity of the van der Waals clashes that might occur between heavy atoms in the protein or the ligand and the suggested halogen. Each individual contribution to *ClashScore* is calculated as follows: The ratio of the actual distance to the sum of the van der Waals radii of the two atoms is subtracted from 1. The resulting value represents the relative violation caused by the electron density overlap of the intersecting van der Waals spheres. This value is then squared to account for the steep increase in repulsion caused by increasingly overlapping electron densities. These squared values are then summed up over all individual contributions and multiplied by a factor of 100 to give the final *ClashScore* (eq 1):

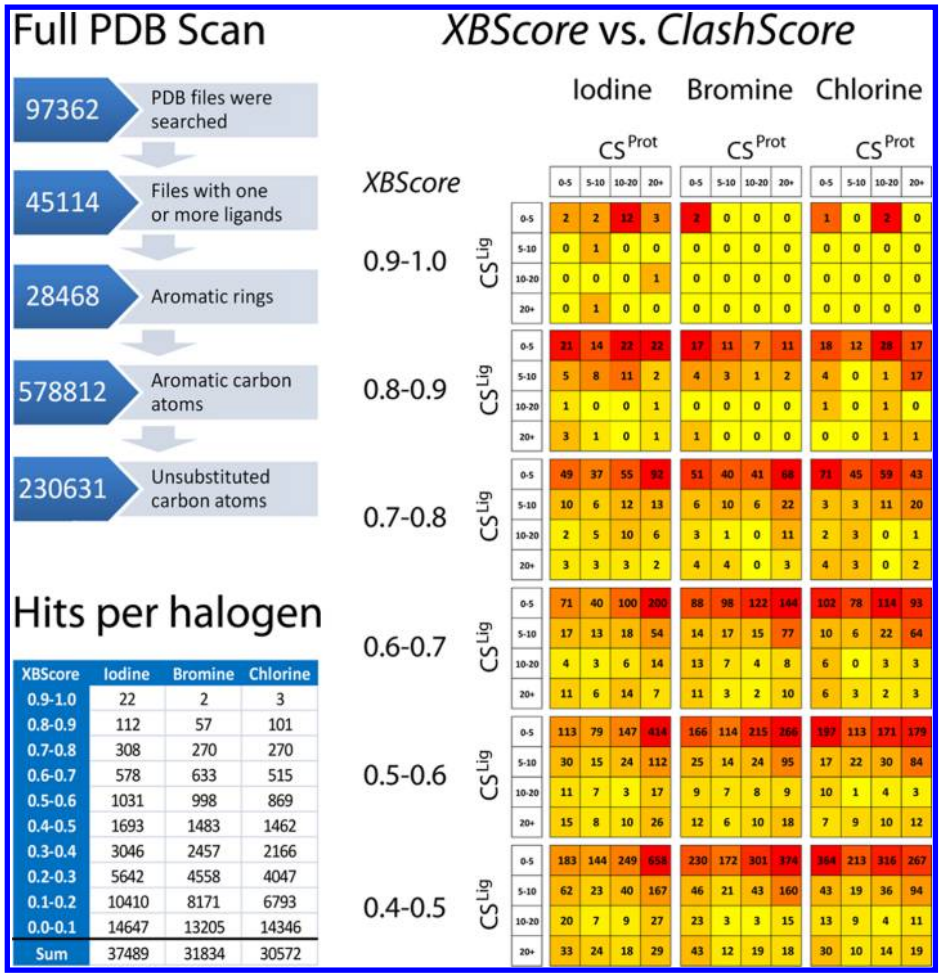
$$ClashScore = 100 \cdot \sum_{i=1}^{n} \left(1 - \frac{\text{actual distance}}{\text{sum of vdW radii}}\right)_i^2 \quad (1)$$

where n is the number of clashes for the respective halogen. This quadratic behavior results in a *ClashScore* contribution of 1 for a distance corresponding to 90% of the sum of the van der Waals radii, a contribution of 4 at 80%, and a contribution of 9 at 70%. Thus, a total *ClashScore* of 5 could be caused by one significant clash at ∼78% of the sum of the van der Waals radii or by five small contributions at 90% of the sum of the van der Waals radii. The protein clash score (CS^{Prot}) and the ligand clash score (CS^{Lig}) were calculated using eq 1 with n set equal to the number of clashes with protein atoms and ligand atoms, respectively.

RESULTS AND DISCUSSION

The scaffold decoration algorithm was applied to the PDB (as of Feb 4, 2014), comprising 97 362 protein structures (Figure

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699



XBScore	Iodine	Bromine	Chlorine
0.9-1.0	22	2	3
0.8-0.9	112	57	101
0.7-0.8	308	270	270
0.6-0.7	578	633	515
0.5-0.6	1031	998	869
0.4-0.5	1693	1483	1462
0.3-0.4	3046	2457	2166
0.2-0.3	5642	4558	4047
0.1-0.2	10410	8171	6793
0.0-0.1	14647	13205	14346
Sum	37489	31834	30572

Figure 6. Summary of the PDB mining protocol and results. The numbers of results obtained for iodine, bromine, and chlorine are reported for different levels of halogen-bonding quality (*XBScore*). These results are further differentiated by their various levels of intermolecular protein or intramolecular ligand clashes (CS^{Prot}, CS^{Lig}).

6). Of these, 45 114 structures possessed a ligand of any kind (metals and water molecules are not considered as ligands). A total of 28 468 aromatic moieties consisting of 578 812 aromatic carbon atoms were found, of which 230 631 were unsubstituted carbon atoms, to which the algorithm was applied. Using *XBScore* $\geq$ 0.4 as a threshold without initially taking possible clashes into account resulted in 3744 hits for iodine. The numbers of hits found for bromine and chlorine were of similar magnitude (3443 and 3220, respectively).

Differences in the spatial distributions of hits for the three halogens can be attributed predominantly to the shorter C−Br and C−Cl bond lengths and the smaller van der Waals radii of bromine and chorine. When hydrogen is substituted by a halogen, bromine and chlorine are more easily accommodated in most binding sites, leading to fewer clashes. This allows the brominated or chlorinated ligand to address the carbonyl oxygen atom from different spatial positions that are less accessible for iodobenzenes or would lead to substantial van der Waals clashes with the protein. These variations in the spatial orientation can be easily observed by plotting the position of the halogen (colored by *XBScore* bin) onto the *N*-methylacetamide model structure (Figure 7). While preferential halogen bonds (*XBScore* > 0.8) involving iodine are clustered around the highest n-electron densities of the carbonyl oxygen, for bromine and chlorine a progressive trend to bind from the opposite hemisphere is visible as well. The basis for this effect is

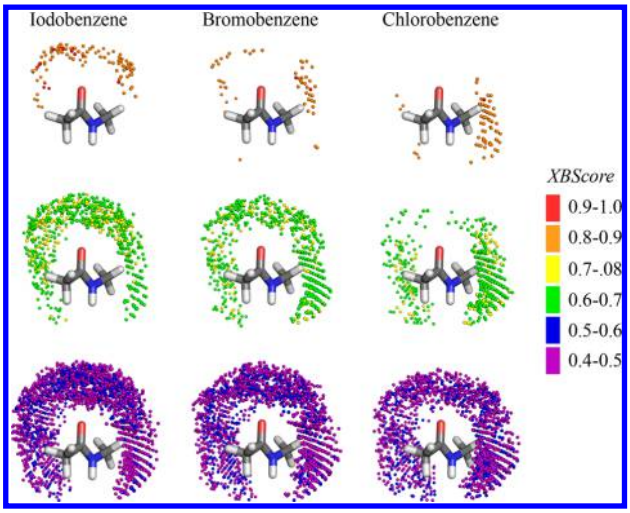


Figure 7. Spatial distributions of hits with *XBScore* $\geq$ 0.4. In contrast to Figure 4, the presentation of the halogen atom coordinates as points does not necessarily imply that the σ-hole angle is always close to the optimal ~180°. Therefore, in various cases the dominant interaction is a δ^- contact rather than a σ-hole contact.

a transition of the preferred nucleophile in the backbone amide system from the n electrons to the π electrons for the smaller

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699

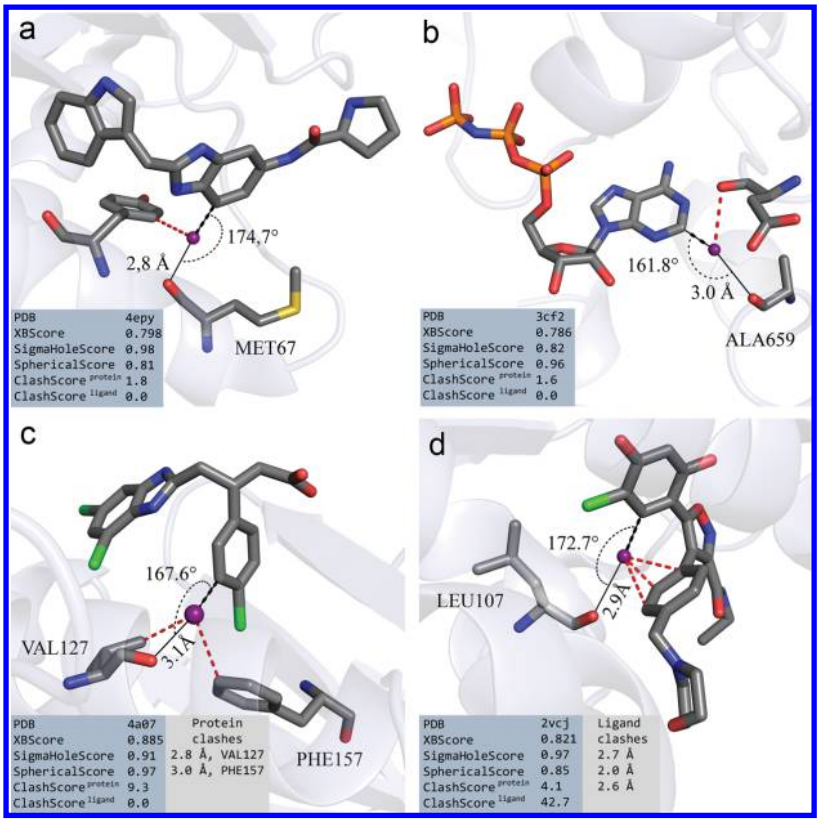


Figure 8. High-scoring examples with varying *ClashScore* levels: (a) PDB entry 4epy;[36] (b) PDB entry 3cf2;[37] (c) PDB entry 4a07;[38] (d) PDB entry 2vcj.[39]

halogens. Additionally, δ^- interactions (see Figure 1) instead of σ-hole interactions can play a more dominant role for the smaller halogens on the basis of overlaps of the negative electrostatic potential in the equatorial position of the halogen with the positive electrostatic potential of the amide hydrogen or the weaker potential of the hydrogen atom of the Cα atom (represented by CH_3 groups in the model). It further has to be kept in mind that the *XBScore* values are normalized to the best possible energy of the respective halogen. This means that an *XBScore* of 1.0 is equivalent to an MP2/TZVPP complex formation energy of −19.2 kJ/mol for iodine, −13.2 kJ/mol for bromine, or −10.8 kJ/mol for chlorine (for comparison, see Figure 4).

Subsequently, we will discuss four high-scoring examples found by the algorithm, which characterize different levels of the two clash scores (CS^{Prot} and CS^{Lig}).

Figure 8a shows a ligand bound to the hydrolase K-Ras.[36] The benzimidazole scaffold was decorated at position 7, leading to an almost optimal halogen-bonding geometry with the carbonyl oxygen of MET67. With a σ-hole angle of 174.7° and an interaction distance of 2.8 Å, the *XBScore* amounts to 0.798. With no *ClashScore* for the ligand ($CS^{Lig} = 0$) and a more than tolerable *ClashScore* for the protein ($CS^{Prot} = 1.8$), this is one of the best examples obtained through the scaffold decoration algorithm. Upon decomposition of CS^{Prot} into single-clash contributions, there is only one significant clash (contribution to $CS^{Prot} = 1.64$; distance ≈ 87% of the sum of the vdW radii), which occurs between the halogen and one of the Cδ atoms of TYR71. Because TYR71 seems not to be packed tightly or engaged in close favorable contacts, it is reasonable to assume that it will be able to adapt to the introduction of a halogen atom by rotation around the χ_1 and/or χ_2 angle.

Quite a few hits were obtained for cofactors such as NAD, FAD, ADP, and ATP and their stabilized analogues. Figure 8b shows AMP-PMP bound to the ATPase p97.[37] The ATP derivative was decorated at position 2 of the adenine ring, leading to a halogen-bonding interaction with ALA659. This interaction is rated with an *XBScore* of 0.786 and causes only one relevant clash (CS^{Prot} contribution of 1.2), involving the carbonyl oxygen of ASP478. This aspartate is located within a loop forming a part of the binding site for AMP-PMP. On the basis of the flexibility of the loop, the clash could be easily resolved by small structural changes.

Figure 8c shows an allosteric activator of human PDK1 kinase binding to the PIF pocket of the protein.[38] With a distance of 3.1 Å and a σ-hole angle of 167.7°, the interaction geometry is almost optimal. However, CS^{Prot} is increased to a moderate level of 9.3, resulting predominantly from two overlaps of van der Waals radii between the suggested iodine and carbon atoms of the side chains of VAL127 and PHE157 (red dotted lines in Figure 8c). However, both overlaps occur with atoms of flexible side chains that could possibly avoid the clash by transitioning into different rotameric states. For VAL127 (distance to iodine ≈ 76% of the sum of the vdW radii; CS^{Prot} contribution = 5.7), rotation around the χ_1 angle could resolve the steric clash. However, the transition into another preferred rotameric state might cause other clashes. In the case of PHE157 (distance to iodine ≈ 82% of the sum of the vdW radii; CS^{Prot} contribution = 3.3), which is rather tightly packed by surrounding residues and interacts with the ligand through a halogen−π contact, there are only a few possibilities to resolve the clash.

As a last example, Figure 8d shows a 4,5-diarylisoxazole ligand bound to HSP90.[39] Here CS^{Lig} = 42.7, resulting from

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699

severe overlaps of the van der Waals spheres of the suggested iodine with three aromatic carbon atoms of the para-substituted phenyl moiety (red dotted lines). With distances of only 2.0, 2.6, and 2.7 Å (46% to 27% below the sum of the van der Waals radii), these contacts contribute 22.0, 9.3, and 7.7 to CS^{Lig}, respectively. Molecular-mechanics-based optimization of the iodinated complex in MOE[40] with restraints on the halogen bond surprisingly indicated that even these highly unfavorable clashes can possibly be overcome through ligand flexibility. However, loss of affinity as a result of changes in the binding mode is possible, rendering this example less favorable. Another aspect that certainly needs to be considered is synthetic accessibility. Hence, a multisubstituted aromatic scaffold, such as the 4-chloro-6-heteroarylbenzene-1,3-diol in 2vcj, appears to be less suitable for further scaffold decoration.

Validation of the Scoring Function. To assess the prediction accuracy of our scoring function, we applied the algorithm to the PDB (as of Feb 2, 2014), comprising more than 97 000 structures. The full range of *XBScore* from 0 (no interaction) to 1 (ideal halogen bond) was split into 10 classes with a fixed class range of 0.1 to facilitate the selection of 50 equally distributed examples for further validation. For each file, the respective geometry was recreated (Figure 9a) using iodobenzene to represent the ligand (matching the model system onto the aromatic or heteroaromatic five- or six-membered ring and the suggested iodine pseudoatom) and *N*-methylacetamide as the protein's backbone structure. The halogen-bonding geometry of the example was transferred as accurately as possible to the small model system (by minimizing the RMSD), and the interaction energy for the model system was obtained from an MP2 single-point calculation. The resulting energy was normalized to obtain a "QM-Calculated Score" ($Score^{QM}$) and plotted against the *XBScore* (Figure 9b). With only a few exceptions, most of the deviations between the $Score^{QM}$ and the predicted *XBScore* were below 10%, leading to a good overall coefficient of determination of $r^2 = 0.87$. The average deviation for all of the data points was 0.063. For 10 of the 50 examples, the deviation between calculated and predicted scores was higher than 0.1, while 28 examples deviated by less than 0.02. Interestingly, when it was assumed that the predicted *XBScore* should ideally match the $Score^{QM}$ by fitting to the equation $y = x$, the r^2 value remained at 0.85 (Figure 9b, red line). The majority of data points are even better fitted by this equation, while a few outliers are responsible for the slight decrease in r^2. Particularly these outliers require more attention: only one example (point A in Figure 9b) shows a high overestimation of the halogen-bond quality, while a few show a significant underestimation (e.g., points B and C in Figure 9b). These examples bearing the largest deviations (0.407 for A, 0.34 for B, and 0.269 for C) are evaluated and discussed below.

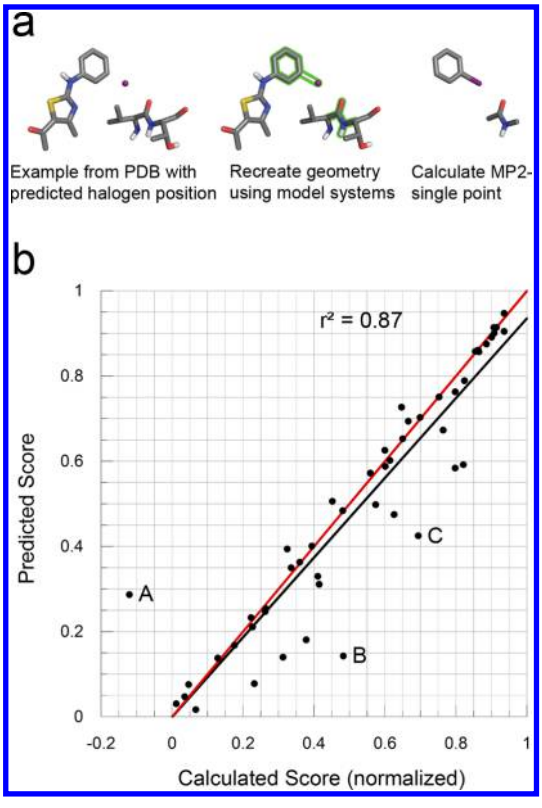


Figure 9. Validation of the scoring function. (a) Crystal structure geometries and predicted position of the halogen are recreated using iodobenzene and *N*-methylacetamide as model systems. MP2/TZVPP single points are calculated. (b) Predicted score vs calculated score. The biggest outliers are highlighted using the letters A–C. The correlation coefficient of the best linear fit (black line) was $r^2 = 0.87$. A fit to the equation $y = x$ (red line) had $r^2 = 0.85$.

While the geometry for point A (Figure 10a) has a predicted score of 0.287, the calculated score reveals an unfavorable interaction with a score of −0.12. With an interaction distance of 2.8 Å and a σ-hole angle of 146.3°, the equatorial, negative electrostatic potential of the iodine most probably clashes with the electron density of the backbone oxygen atom. Such a combination of a very short distance and a rather poor σ-hole angle is currently not well represented in the *XBScore* because the calculation of the product of the normalized *SphericalScore* and *SigmaHoleScore* assumes that the σ-hole dependence is identical at each distance. This artifact should be restricted to very weak halogen bonds at particularly short distances and can be overcome by the introduction of a penalty term.

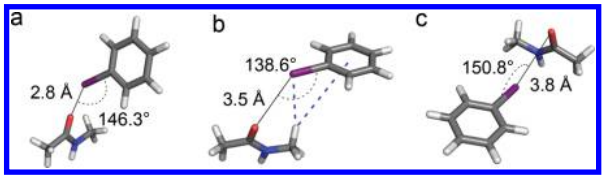


Figure 10. Explanation of outliers. (a) Geometry of the highest deviation from *XBScore*. As opposed to the prediction, this interaction is unfavorable because of repulsion of the iodine atom and the carbonyl oxygen atom with a poor σ-hole angle. (b) Geometry of the second-worst outlier from the prediction. Failure to recognize additional CH···I and CH···π contacts (blue lines) leads to an underestimation of the halogen-bonding interaction. (c) Geometry of the third-biggest outlier from the validation. The interaction between the halogen atom and the electron-rich peptide bond of the carbonyl function compensates for the moderate σ-hole angle toward the carbonyl oxygen atom.

For point B, with a predicted *XBScore* of 0.143 and a calculated $Score^{QM}$ of 0.483, the quality of the interaction seems to be notably underrated. Examination of the geometry of the model system (Figure 10b) indicates that additional secondary interactions (CH···I and CH···π contacts) improved the QM-calculated energy significantly. Thus, quite a few outliers are generated because the *XBScore* (in contrast to the QM

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687–699

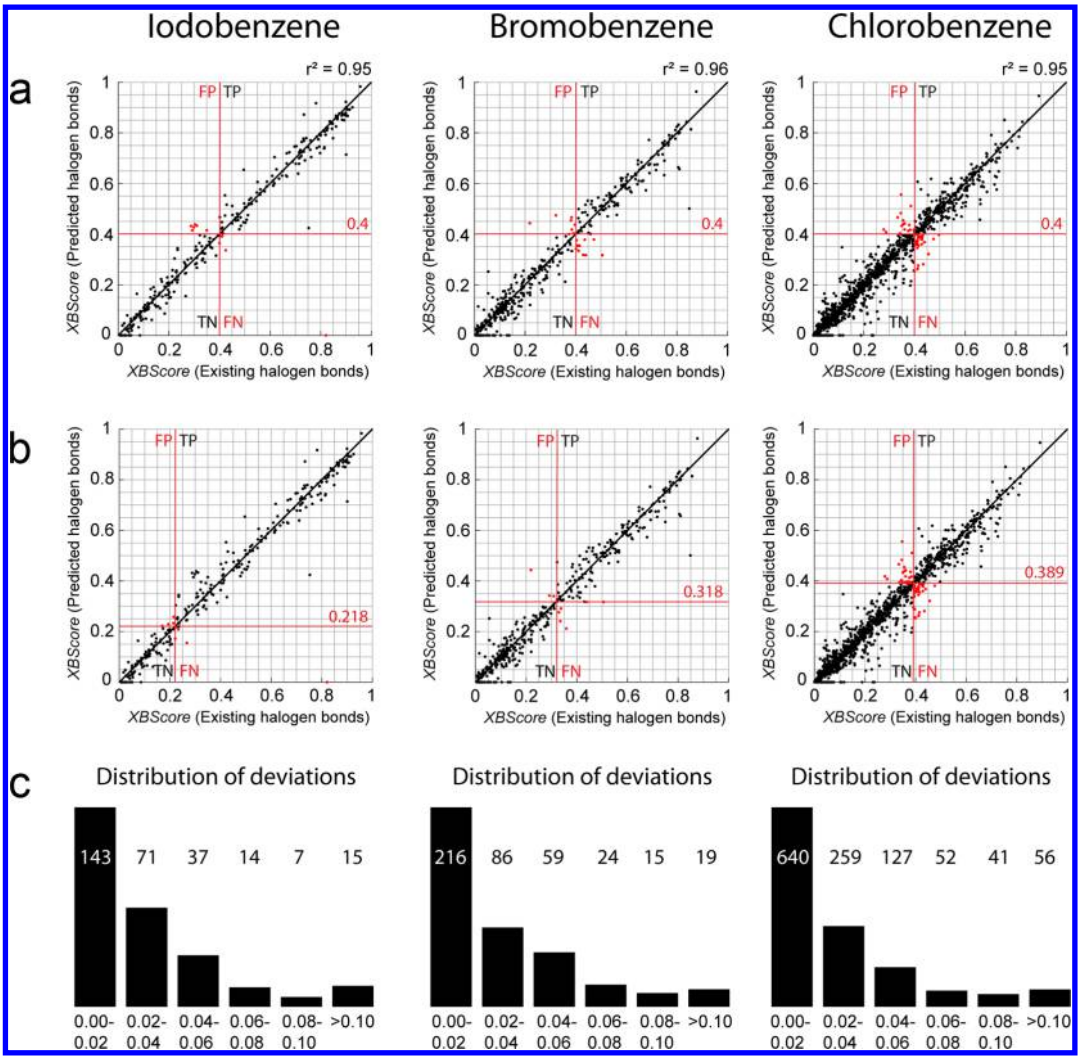


Figure 11. Validation of the scoring function by prediction of existing halogen bonds. Existing halogen bonds were scored by *XBScore*. Subsequently, all halogen atoms were substituted by hydrogen atoms, and these dehalogenated compounds were submitted to the automated scaffold decoration protocol. The predicted *XBScore* was calculated and plotted against the original *XBScore* of the crystal structure. The correlation coefficients (r^2) for iodine, bromine, and chlorine were found to be 0.95, 0.96, and 0.95, respectively. In (a), the classification for halogen bonds versus weak interactions was defined as a relative criterion (*XBScore* = 0.4), while in (b) an absolute energy value (ΔE = 1 kcal/mol) was used. The distribution of deviations (c) shows a vast dominance of very small differences between the predicted and original *XBScores*. The majority of scores deviate by only 2% of the maximal interaction energy of the respective halogen (*XBScore* = 0.02).

calculation) cannot recognize the contribution of additional attractive interactions.

For point C, with a rather moderate σ-hole angle of 150.8°, the *XBScore* of 0.425 is underestimated by 0.269. This is due to the fact that the iodine is directed toward the peptide bond (between the carbonyl carbon and the nitrogen atom), a very electron-rich area (see Figure 10c). However, the transition of the interaction partner from n to π electrons has not yet been implemented into the oxygen-centered construction of the *XBScore*.

Prediction of Existing Halogen Bonds. As an alternative validation, we checked whether we could rediscover existing halogen bonds with reasonable interaction qualities in the PDB. Therefore, we identified all halogen···carbonyl oxygen contacts in protein−ligand complexes that show a halogen···oxygen distance between 2.6 and 5.0 Å. All of these contacts were first scored on the basis of their interaction geometry in the crystal structure. We propose that an *XBScore* of 0.4 (equivalent to an MP2/TZVPP complex formation energy of −7.7 kJ/mol for iodine, −5.3 kJ/mol for bromine, or −4.3 kJ/mol for chlorine) is sufficient to classify a contact as a halogen bond.

All of the relevant halogen atoms from each ligand were then substituted by hydrogen atoms, and the automated scaffold decoration protocol was applied as before. The hereby-retrieved predicted *XBScore* of each configuration is plotted against the *XBScore* obtained for the original crystal structure in Figure 11a. On the basis of our proposed definition of a halogen bond (*XBScore* = 0.4), in the case of iodine 129 halogen bonds (45%) were correctly predicted as halogen bonds (true positives = TP), while 147 weaker contacts (51%) were correctly predicted as weaker contacts (true negatives = TN). For only 11 (4%) was the classification incorrect (false positive (FP) or false negative (FN)). The overall correlation was r^2 = 0.95. For the true positives only, r^2 slightly decreased to 0.89. For the true negatives only, r^2 amounted to 0.92. For bromine and chlorine, very similar rates of incorrect predictions (FP + FN = 4%) were found. A significant shift from true positives to true negatives can be observed when comparing chlorine to bromine and

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699

Table 1. Comparison of the Existing Halogen Bonds Found in the PDB to the Potential Halogen Bonds Suggested by the Scaffold Decoration Algorithm Described Herein

		existing XBs		designed XBs	
selection criteria	halogen	occurences	unique structures	occurences	unique structures
$XBScore \geq 0.4$	iodine	134	85	3744	2570
$XBScore \geq 0.4 + CS^{Lig} \geq 5$	iodine	121	78	2751	1948
$XBScore \geq 0.4 + CS^{Lig} \leq 5 + CS^{Prot} \leq 10$	iodine	120	77	763	600
$XBScore \geq 0.218^a$	iodine	190	99	11114	6206
$XBScore \geq 0.218^a + CS^{Lig} \geq 5$	iodine	172	93	8427	5033
$XBScore \geq 0.218^a + CS^{Lig} \leq 5 + CS^{Prot} \leq 10$	iodine	171	92	2010	1415
$XBScore \geq 0.4$	bromine	133	63	3443	2403
$XBScore \geq 0.4 + CS^{Lig} \geq 5$	bromine	129	60	2577	1854
$XBScore \geq 0.4 + CS^{Lig} \leq 5 + CS^{Prot} \leq 10$	bromine	127	59	996	788
$XBScore \geq 0.318^a$	bromine	162	71	5334	3510
$XBScore \geq 0.318^a + CS^{Lig} \geq 5$	bromine	154	68	4064	2788
$XBScore \geq 0.318^a + CS^{Lig} \leq 5 + CS^{Prot} \leq 10$	bromine	152	67	1562	1178
$XBScore \geq 0.4$	chlorine	332	215	3220	2222
$XBScore \geq 0.4 + CS^{Lig} \geq 5$	chlorine	328	211	2519	1816
$XBScore \geq 0.4 + CS^{Lig} \leq 5 + CS^{Prot} \leq 10$	chlorine	326	209	1215	943
$XBScore \geq 0.389^a$	chlorine	346	224	3395	2338
$XBScore \geq 0.389^a + CS^{Lig} \geq 5$	chlorine	342	220	2657	1905
$XBScore \geq 0.389^a + CS^{Lig} \leq 5 + CS^{Prot} \leq 10$	chlorine	340	218	1290	997

[a]The *XBScore* corresponds to an attractive energy contribution of 1 kcal/mol.

iodine. For chlorine only 26% were true positives. For bromine this rate increased to 29%, whereas for iodine almost half of the evaluated contacts (45%) were true positives. This highlights that whenever iodine is found in a reasonable proximity to a carbonyl oxygen of the protein backbone, it considerably more often forms a decent to good halogen bond than bromine or chlorine. This trend is even more pronounced when the definition of a proper halogen bond is based on absolute energies instead of relative values. For comparison, we evaluated the difference when 1 kcal/mol (4.1868 kJ/mol) was used as the classification criterion for a suitable halogen bond (Figure 11b). With an *XBScore* of 0.389, the border for the classification shifted only marginally in the case of chlorine. Thus, the true positive rate was still 26%. With the 1 kcal/mol criterion, the *XBScore* decreased to 0.318 in the case of bromine, yielding a true positive rate of 37%. In the case of iodine, the *XBScore* was even more reduced to 0.218, yielding a true positive rate of 65%.

In general, the correlations for bromine and chlorine are of similar quality as the correlation found for iodine. Bromine yields an overall r^2 of 0.96 and chlorine an overall r^2 of 0.95. Most of the deviations (Figure 11c) are only marginal (<2%). Only very few deviations are problematic (>10%). Most of these issues can be attributed to a somewhat distorted geometry of the ligand. In some cases, this may be explained by conformational strain imposed on the ligand by competing strong interactions. In some cases, the distortion is most likely an artifact of the crystal structure (e.g., substituents of aromatic rings are strongly bent out of plane).

METHODS

Optimizations and Single-Point Calculations. Geometry optimizations and single-point calculations were carried out at the MP2/TZVPP[41] level of theory using the TURBOMOLE 6.4 suite of programs.[42,43] Relativistic effects for iodine were considered by using an effective core potential (ECP).[44] The calculations were done in combination with the resolution of identity (RI) technique[45−47] and the frozen core approximation. The frozen core orbitals were attributed by the default setting in TURBOMOLE, by which all orbitals possessing energies below 3.0 au are considered core orbitals. Using the frozen core approximation to treat relativity in iodine may have effects on the polarizability of the outer electrons and consequently might not represent dispersion correctly. The SCF convergence criterion was increased to 10^{-8} Hartree for all of the calculations.

σ-Hole Scans. All of the σ-hole scans were performed using the optimized MP2/TZVPP geometries as starting points. The C−X group was placed on the vector defined by the C═O group, with both the angle $\alpha_{C=O\cdots X}$ and the σ-hole angle $\alpha_{O\cdots X-C}$ constrained to 180°; the bond distance for each halobenzene was set to the most favorable interaction distance (iodine, 3.03 Å; bromine, 3.06 Å; chlorine, 3.12 Å). The σ-hole angle $\alpha_{O\cdots X-C}$ was then successively modified from 180° by an amount ranging from −90° to +90° in steps of 5° (see Figure 2a), and the MP2/TZVPP single-point energy was calculated for each angle. To avoid having to differentiate between the energies for angles of 180° + x° and 180° − x°, we used the average of the two calculations for the subsequently developed *XBScore*. It should be noted that the deviation between the energies for angles of 180° + x° and 180° − x° was marginal in all cases.

Spherical Scans. Input files were generated from the optimized MP2/TZVPP geometries subject to certain constraints. The C−X group was placed on the vector defined by the C═O group, with both the angle $\alpha_{C=O\cdots X}$ and the σ-hole angle $\alpha_{O\cdots X-C}$ constrained to 180°; the bond distance was freely optimized. In order to generate a full sphere of input geometries for subsequent calculations, the optimized structure was transformed as follows: The oxygen atom was placed on the origin of the coordinate system, and the entire complex was rotated until the halogen atom was positioned on the positive x axis. Let α denote the angle of rotation counterclockwise around the z axis and β the angle of rotation counterclockwise around the x axis. The value of α was gradually increased from 0° to 180° in steps of 5°. For each α value, β was varied from 0°

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699

to 355° in steps of 5°, leading to a total number of 2664 halogen positions distributed on a sphere. The structure of the ligand was not altered during this transformation process. Single-point (SCF) calculations at the MP2/TZVPP level were then done.

For calculation and visualization purposes, several Python scripts were written and executed in PyMol version 1.2.[48] MATLAB (release 2012a)[49] was used for curve fitting.

CONCLUSIONS AND OUTLOOK

Characterizing the potential and facilitating the applicability of halogen bonding in molecular design are key objectives of this study. We have strived to make a thorough comparison of the current level of halogen bonding in ligand−protein complexes in the PDB with the potential to introduce halogen bonds by molecular design into existing crystal structures. We have focused on halogen contacts with the ubiquitously present backbone carbonyl moieties on the basis of previous systematic quantum-chemical studies,[14] which we have extended herein. We have found 780 crystal structures containing iodine, 1056 crystal structures containing bromine, and 9154 crystal structures containing chlorine in the PDB. Applying relative energy cutoffs for the identification of halogen bonds with reasonable qualities ($XBScore \geq 0.4$), we have found that 10.9% of the iodine-containing structures exhibit a halogen bond (see Table 1). For bromine, this percentage decreases to 6.0%, while for chlorine only 2.3% of the crystal structures show a halogen bond toward a backbone carbonyl. On the basis of an absolute energy cutoff (1 kcal/mol), only for iodine ($XBScore \geq 0.218$) was a moderate increase in the number of existing halogen bonds observed (10.9% to 12.7%). For bromine and chlorine, the numbers of crystal structures showing halogen bonds hardly changed. When additionally constraints for possible clashes with the ligand ($CS^{Lig} \leq 5$) or the protein ($CS^{Prot} \leq 10$) were applied, the percentage of observed halogen bonds remains almost unchanged.

In contrast to 134 existing halogen bonds involving iodine in 85 different crystal structures, in this study we have identified 3744 possible halogen bonds of reasonable quality ($XBScore \geq 0.4$) involving iodine that could be introduced into existing protein−ligand complexes by halogenation of aromatic scaffolds. Because the introduction of an iodine atom can lead to clashes with the protein and/or the ligand, we have proposed discarding every suggested halogen bond exceeding certain cutoffs ($CS^{Lig} > 5$ and $CS^{Prot} > 10$). Even so, the number of potential halogen bonds (763) is still ∼6-fold higher than the number of existing ones. Using an absolute energy cutoff (1 kcal/mol), we initially retrieved 11 114 suggestions, of which 2010 pass the standard clash threshold. This indicates that the potential to find new halogen bonds is even much stronger (>11-fold) when weak to moderate strengths are taken into consideration as well. For bromine, we have found a similar level of existing halogen bonds in comparison to iodine but a slightly larger number of halogen bonds by design when applying a relative cutoff ($XBScore \geq 0.4$). Thus, the ratio of designed to existing halogen bonds increases to 7.8, whereas in the case of an absolute energy cutoff (1 kcal/mol), the ratio of 10.3 is slightly worse than for iodine. This behavior can be easily explained by the size of bromine. The introduction of bromine causes fewer clashes, therefore leading to a smaller number of suggestions that do not pass the clash cutoffs. However, when the quality of the halogen bond is judged using an absolute energy criterion, the inferior strength of the bromine−carbonyl interaction causes a significantly lower number of suggested halogen bonds to be found initially (11 114 for iodine vs 5334 for bromine).

Similar trends can be observed for chlorine. The absolute number of putative halogen bonds by design increases for a relative quality definition ($XBScore \geq 0.4$) from 763 for iodine and 996 for bromine to 1215 for chlorine. However, on the basis of a roughly 10-fold larger amount of crystal structures containing chlorine, also the number of existing halogen bonds is higher for chlorine (326) than for the other halogens. As a result, the proportion of designed to existing halogen bonds (3.7) is considerably lower than for the previously discussed halogens. Because the relative energy cutoff ($XBScore \geq 0.4$) and the absolute energy cutoff (1 kcal/mol) are almost identical, the values and ratios based on the absolute energy cutoff are closely related.

In summary, it can be concluded that iodine certainly holds the greatest potential to form halogen bonds that are able to contribute reasonably to the free energy of binding. However, it also can cause significantly more disturbances in the existing binding modes of the decorated ligands, while bromine and particularly chlorine are more easily accommodated by the binding site. For systems where new crystal structures can be easily obtained, using iodine in molecular design may provide the greatest benefit, while changes in the binding mode can be easily monitored. In contrast, when structural insights about the changes of the binding mode cannot be readily obtained, decorations with chlorine and bromine appear more reasonable in order to scout for improved affinities through halogen bonding.

Using two separate approaches, we have validated both the newly generated scoring function and the molecular design algorithm. We have demonstrated that we can reproduce the QM-calculated energies with a very reasonable r^2 of 0.87 for iodine using a diverse set of 50 predicted structures. In addition, the almost complete retrieval of all existing halogen bonds by our scaffold decoration algorithm (after initial deletion of the halogen from the scaffold) is a useful proof of the exhaustive search and thorough prediction of the entire approach. Hence, we have created a robust and reliable method for integrating halogen bonds into molecular design, yielding a vast number of interesting test cases in a broad variety of targets. Currently, some of the more promising suggestions are under experimental evaluation.

On the basis of this first scoring function for chlorine, bromine, or iodine contacts to backbone carbonyls, we aim to implement this and additional scoring terms for halogen bonds into an empirical scoring function to facilitate its application in protein−ligand docking experiments. Scoring and visualization of halogen bonds toward backbone carbonyls will also be made available via a Web interface at www.halogenbonding.com.

AUTHOR INFORMATION

Corresponding Author

*E-mail: frank.boeckler@uni-tuebingen.de.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) 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.

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699

(2) Zimmermann, M. O.; Lange, A.; Wilcken, R.; Cieslik, M. B.; Exner, T. E.; Joerger, A. C.; Koch, P.; Boeckler, F. M. Halogen-enriched fragment libraries as chemical probes for harnessing halogen bonding in fragment-based lead discovery. *Future Med. Chem.* **2014**, *6*, 617−639.
(3) Wilcken, R.; Liu, X.; Zimmermann, M. O.; Rutherford, T. J.; Fersht, A. R.; Joerger, A. C.; Boeckler, F. M. Halogen-enriched fragment libraries as leads for drug rescue of mutant p53. *J. Am. Chem. Soc.* **2012**, *134*, 6810−6818.
(4) Fanfrlik, J.; Kolar, M.; Kamlar, M.; Hurny, D.; Ruiz, F. X.; Cousido-Siah, A.; Mitschler, A.; Rezac, J.; Munusamy, E.; Lepsik, M.; Matejicek, P.; Vesely, J.; Podjarny, A.; Hobza, P. Modulation of Aldose Reductase Inhibition by Halogen Bond Tuning. *ACS Chem. Biol.* **2013**, *8*, 2484−2492.
(5) Scholfield, M. R.; Vander Zanden, C. M.; Carter, M.; Ho, P. S. Halogen bonding (X-bonding): A biological perspective. *Protein Sci.* **2013**, *22*, 139−152.
(6) Carter, M.; Voth, A. R.; Scholfield, M. R.; Rummel, B.; Sowers, L. C.; Ho, P. S. Enthalpy−Entropy Compensation in Biomolecular Halogen Bonds Measured in DNA Junctions. *Biochemistry* **2013**, *52*, 4891−4903.
(7) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. Halogen bonding: The σ-hole. *J. Mol. Model.* **2007**, *13*, 291−296.
(8) Politzer, P.; Murray, J. S.; Clark, T. Halogen bonding and other σ-hole interactions: A perspective. *Phys. Chem. Chem. Phys.* **2013**, *15*, 11178−11189.
(9) Desiraju, G. R.; Ho, P. S.; Kloo, L.; Legon, A. C.; Marquardt, R.; Metrangolo, P.; Politzer, P.; Resnati, G.; Rissanen, K. Definition of the halogen bond (IUPAC Recommendations 2013). *Pure Appl. Chem.* **2013**, *85*, 1711−1713.
(10) Brickmann, J.; Exner, T. E.; Gimmler, J.; Lautenschläger, P.; Heiden, W.; Moeckel, G.; Zahn, D. *MOLCAD II*, version 1.4; MOLCAD GmbH: Darmstadt, Germany; http://www.molcad.de.
(11) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(12) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Zahn, S.; Kirchner, B.; Boeckler, F. M. Addressing Methionine in Molecular Design through Directed Sulfur−Halogen Bonds. *J. Chem. Theory Comput.* **2011**, *7*, 2307−2315.
(13) Lange, A.; Zimmermann, M. O.; Wilcken, R.; Zahn, S.; Boeckler, F. M. Targeting Histidine Side Chains in Molecular Design through Nitrogen−Halogen Bonds. *J. Chem. Inf. Model.* **2013**, *53*, 3178−3189.
(14) Wilcken, R.; Zimmermann, M.; Lange, A.; Zahn, S.; Boeckler, F. Using halogen bonds to address the protein backbone: A systematic evaluation. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 935−945.
(15) Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discovery* **2005**, *4*, 649−663.
(16) Gohlke, H.; Klebe, G. Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644−2676.
(17) Schneider, G.; Böhm, H.-J. Virtual screening and fast automated docking methods. *Drug Discovery Today* **2002**, *7*, 64−70.
(18) Huang, S.-Y.; Grinter, S. Z.; Zou, X. Scoring functions and their evaluation methods for protein−ligand docking: Recent advances and future directions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12899−12908.
(19) Roehrig, S.; Straub, A.; Pohlmann, J.; Lampe, T.; Pernerstorfer, J.; Schlemmer, K. H.; Reinemer, P.; Perzborn, E. Discovery of the novel antithrombotic agent 5-chloro-*N*-({(5*S*)-2-oxo-3-[4-(3-oxomorpholin-4-yl)phenyl]-1,3-oxazolidin-5-yl}methyl)thiophene-2-carboxamide (BAY 59-7939): An oral, direct Factor Xa inhibitor. *J. Med. Chem.* **2005**, *48*, 5900−5908.
(20) 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.
(21) Carter, M.; Rappé, A. K.; Ho, P. S. Scalable Anisotropic Shape and Electrostatic Models for Biological Bromine Halogen Bonds. *J. Chem. Theory Comput.* **2012**, *8*, 2461−2473.
(22) Scholfield, M. R.; Ford, M. C.; Vander Zanden, C. M.; Billman, M. M.; Ho, P. S.; Rappé, A. K. Force Field Model of Periodic Trends in Biomolecular Halogen Bonds. *J. Phys. Chem. B* **2014**, DOI: 10.1021/jp509003r.
(23) Hennemann, M.; Clark, T. EMPIRE: A highly parallel semiempirical molecular orbital program: 1: Self-consistent field calculations. *J. Mol. Model.* **2014**, *20*, No. 2331.
(24) Metrangolo, P.; Meyer, F.; Pilati, T.; Proserpio, D. M.; Resnati, G. Dendrimeric Tectons in Halogen Bonding-Based Crystal Engineering. *Cryst. Growth Des.* **2007**, *8*, 654−659.
(25) Kolar, M.; Hobza, P. On Extension of the Current Biomolecular Empirical Force Field for the Description of Halogen Bonds. *J. Chem. Theory Comput.* **2012**, *8*, 1325−1333.
(26) Řezáč, J.; Hobza, P. A halogen-bonding correction for the semiempirical PM6 method. *Chem. Phys. Lett.* **2011**, *506*, 286−289.
(27) Brahmkshatriya, P. S.; Dobes, P.; Fanfrlik, J.; Rezac, J.; Paruch, K.; Bronowska, A.; Lepsik, M.; Hobza, P. Quantum Mechanical Scoring: Structural and Energetic Insights into Cyclin-Dependent Kinase 2 Inhibition by Pyrazolo[1,5-*a*]pyrimidines. *Curr. Comput.-Aided Drug Des.* **2013**, *9*, 118−129.
(28) Kuhn, B.; Fuchs, J. E.; Reutlinger, M.; Stahl, M.; Taylor, N. R. Rationalizing Tight Ligand Binding through Cooperative Interaction Networks. *J. Chem. Inf. Model.* **2011**, *51*, 3180−3198.
(29) Murray, J. S.; Riley, K. E.; Politzer, P.; Clark, T. Directional Weak Intermolecular Interactions: σ-Hole Bonding. *Aust. J. Chem.* **2010**, *63*, 1598−1607.
(30) Riley, K. E.; Hobza, P. Investigations into the Nature of Halogen Bonding Including Symmetry Adapted Perturbation Theory Analyses. *J. Chem. Theory Comput.* **2008**, *4*, 232−242.
(31) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(32) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H−Pu. *J. Chem. Phys.* **2010**, *132*, No. 154104.
(33) Hennemann, M.; Murray, J. S.; Politzer, P.; Riley, K. E.; Clark, T. Polarization-induced σ-holes and hydrogen bonding. *J. Mol. Model.* **2012**, *18*, 2461−2469.
(34) Riley, K. E.; Hobza, P. The relative roles of electrostatics and dispersion in the stabilization of halogen bonds. *Phys. Chem. Chem. Phys.* **2013**, *15*, 17742−17751.
(35) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918−1931.
(36) Sun, Q.; Burke, J. P.; Phan, J.; Burns, M. C.; Olejniczak, E. T.; Waterson, A. G.; Lee, T.; Rossanese, O. W.; Fesik, S. W. Discovery of Small Molecules That Bind to K-Ras and Inhibit Sos-Mediated Activation. *Angew. Chem., Int. Ed.* **2012**, *51*, 6140−6143.
(37) Davies, J. M.; Brunger, A. T.; Weis, W. I. Improved Structures of Full-Length p97, an AAA ATPase: Implications for Mechanisms of Nucleotide-Dependent Conformational Change. *Structure* **2008**, *16*, 715−726.
(38) Lopez-Garcia, L. A.; Schulze, J. O.; Fröhner, W.; Zhang, H.; Süß, E.; Weber, N.; Navratil, J.; Amon, S.; Hindie, V.; Zeuzem, S.; Jørgensen, T. J. D.; Alzari, P. M.; Neimanis, S.; Engel, M.; Biondi, R. M. Allosteric Regulation of Protein Kinase PKCζ by the N-Terminal C1 Domain and Small Compounds to the PIF-Pocket. *Chem. Biol.* **2011**, *18*, 1463−1473.
(39) Brough, P. A.; Aherne, W.; Barril, X.; Borgognoni, J.; Boxall, K.; Cansfield, J. E.; Cheung, K.-M. J.; Collins, I.; Davies, N. G. M.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Finch, H.; Fink, A.; Hayes, A.; Howes, R.; Hubbard, R. E.; James, K.; Jordan, A. M.; Lockie, A.; Martins, V.; Massey, A.; Matthews, T. P.; McDonald, E.; Northfield, C. J.; Pearl, L. H.; Prodromou, C.; Ray, S.; Raynaud, F. I.; Roughley, S. D.; Sharp, S. Y.; Surgenor, A.; Walmsley, D. L.; Webb, P.; Wood, M.; Workman, P.; Wright, L. 4,5-Diarylisoxazole Hsp90 Chaperone

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699

Inhibitors: Potential Therapeutic Agents for the Treatment of Cancer. *J. Med. Chem.* **2007**, *51*, 196−218.
(40) *Molecular Operating Environment (MOE)*, version 2009.10; Chemical Computing Group: Montreal, QC, Canada, 2009.
(41) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(42) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. Electronic-Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165−169.
(43) *TURBOMOLE*, version 6.4; TURBOMOLE GmbH: Karlsruhe, Germany, 2010.
(44) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16−18 elements. *J. Chem. Phys.* **2003**, *119*, 11113−11123.
(45) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of Approximate Integrals in ab Initio Theory. An Application in MP2 Energy Calculations. *Chem. Phys. Lett.* **1993**, *208*, 359−363.
(46) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: Optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143−152.
(47) Hattig, C. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core−valence and quintuple-ζ basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59−66.
(48) *The PyMOL Molecular Graphics System*, version 1.2; Schrödinger, LLC: New York.
(49) *MATLAB*, release 2012a; The MathWorks: Natick, MA, 2012.

DOI: 10.1021/ci5007118
J. Chem. Inf. Model. 2015, 55, 687−699