

Plugging the explicit σ -holes into molecular docking

Michal Kolář^{ab}

Pavel Hobza^{ac}

Agnieszka K. Bronowska^{de*}

^a*Institute of Organic Chemistry and Biochemistry and Gilead Science Research Center, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 16610 Prague 6, Czech Republic*

^b*Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University in Prague, Albertov 6, 12843 Prague 2, Czech Republic*

^c*Department of Physical Chemistry, Palacký University, Olomouc, 77146 Olomouc, Czech Republic*

^d*Faculty of Chemistry, University of Heidelberg, Im Neuenheimer Feld, D-69115 Heidelberg, Germany*

^e*Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnengasse 35, D-69118 Heidelberg, Germany.
E-mail: agnieszka.bronowska@h-its.org; Fax: +49 6221 533298; Tel: +49 6221 533510*

A novel approach in molecular docking was successfully used to reproduce protein-ligand experimental geometries. When dealing with halogenated compounds the correct description of halogen bonds between the ligand and the protein is shown to be essential. Applying a simple molecular mechanistic model for halogen bonds improved the protein-ligand geometries as well as halogen bond features, which makes it a promising tool for future computer-aided drug development.

A promising tool for computer-aided molecular design is presented. By means of molecular docking we calculated the binding poses of a series of 92 halogenated inhibitors and successfully reproduced the crystallographic data in 90 out of 92 instances. For the first time we incorporate into a docking program suite a molecular-mechanical approach that correctly describes halogen bonding. The approach is based on a massless positive point charge included in addition to the halogen atoms, which mimics the anisotropy of the charge density around the halogen atom, known as the σ -hole. We show that this description of halogen bonding considerably improves the reliability of the protein-ligand geometries determined by a docking process, especially in those cases where more than one halogen bond is established between the ligand and the active site of the protein.

The cost of a drug being developed by a major pharmaceutical company is at least \$4 billion, and it can be as much as \$11 billion.¹ The time required for the drug development may vary, but

typically it takes from 7 to 12 years.² Considering that the major reason for drug failing is lack of efficacy,³ the new methods for describing the drug-target binding are actively being sought.

Many drugs available on the market and new bioactive chemical entities are halogenated compounds. The halogen atoms are introduced to increase the membrane permeability hence improving oral absorption, to fill hydrophobic cavities at the protein binding site, to facilitate the blood-brain barrier crossing, and to prolong the lifetime of the drug.³ Apart from those non-specific effects, the halogens were recognized as being able to participate in a highly specific, directional, non-covalent interaction, known as halogen bond.^{4,5} According to the most recent “provisional recommendation” by IUPAC, it is an attractive interaction occurring between an electrophilic region of a halogen atom and a nucleophilic region of another atom or a molecular fragment such as a carbonyl oxygen. The strength of the interaction increases with the atomic number of the halogen

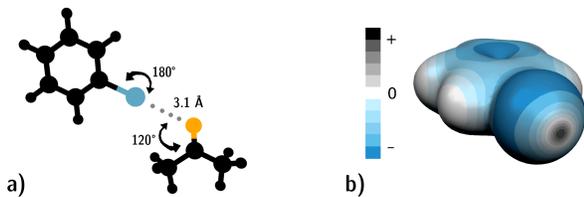


Figure 1: (a) Geometrical features of a typical halogen bond between bromobenzene and acetone. (b) The charge distribution around the bromobenzene molecule. The regions of negative electrostatic potential are in blue, positive regions in grey. The grey disc in the foreground is called σ -hole.

reaching several kcal mol^{-1} .⁶ Typical binding geometry is depicted in Fig. 1a. The nature of the attraction lies, in large part, in a so-called σ -hole.^{7–9} Quantum chemical calculations revealed that the charge distribution around the halogen atom is highly anisotropic, creating a region with positive electrostatic potential located on top of the halogen atom. This positive region, the σ -hole, attracts the negative lone-pair of the Lewis base (Fig. 1b). This poses a serious challenge to current modelling approaches, which treat halogen atoms as having all-negative electrostatic potential, thus failing to correctly describe the halogen-bonded systems, such as protein-ligand complexes. It should be emphasized, though, that the role of halogen bonding in tuning the intermolecular interactions is not limited to medicinal and pharmaceutical chemistry. There is a growing recognition of this type of interactions among inorganic and supramolecular chemists in the applications of halogen-bonding in liquid crystals, light-induced surface patterning of supramolecular polymers and crystal engineering^{10,11}

Halogen bonding is described well at Hartree-Fock or DFT levels of theory, providing at least a double-zeta basis set is used. Semi-empirical methods fail to describe halogen bonds as well as standard molecular mechanics. In the case of computer-aided drug design the use of computationally cheap methods is inevitable. Since a correct description of halogen bonding is of such a fundamental importance, molecular mechanical approaches correctly describing σ -holes were intro-

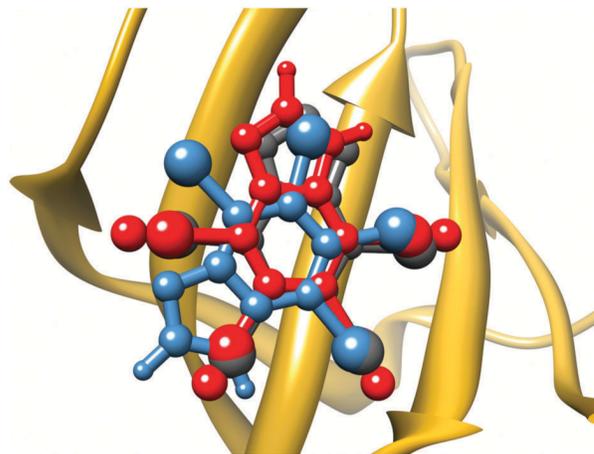


Figure 2: The overlay of the predicted binding poses of the K17 inhibitor of casein kinase 2 (PDB code 2OXY) with (red) and without (blue) explicit σ -holes (ESH) and comparison with the crystal structure (grey).

duced by several laboratories.^{12–15} The essential component of all these approaches was a positively charged, optionally massless, dummy-atom, representing the σ -hole. However, all these corrections were applied to the molecular-mechanical force fields, which require at least preliminary structural data and which are designed to study the dynamic behaviour of systems of known structures. So far, no improvements have been implemented in the suites, which are designed to predict the structure of macromolecular complexes.

Herein we applied such a concept for the first time to the molecular docking scheme. The entity, denoted explicit σ -hole (ESH), was used in conjunction with the UCSF DOCK molecular docking suite.¹⁶ The performance of the improved docking has been tested on 92 protein-ligand complexes for which the crystallographic data are available. The geometries calculated with and without the ESH concept were compared with the experimental geometries (Fig. 2). It should be emphasised that the faithful geometrical representation of the protein-ligand complexes is the essential prerequisite for any further computational investigation not only in drug design studies.

Four pharmaceutically attractive protein targets were chosen, namely aldose reductase (ALDR), cyclin-dependent kinase 2 (CDK2), casein kinase 2

(CK2), and human immunodeficiency virus 1 reverse transcriptase (HIVRT), since they are known to be effectively inhibited by halogenated ligands. From the Protein Data Bank¹⁷ a set of protein-ligand X-ray geometries was collected (see ESI) and their analysis revealed the following facts: the set contained 55 chlorinated, 38 brominated and one iodinated ligand and about 55% of ligands contained more than one possible halogen-bond donor (i.e. Cl, Br or I). The set comprised both halogen bond complexes (about 57%) as well as the complexes without any significant halogen–Lewis base contact (43%). In some instances, mostly in the CK2 case, two or three halogen bonds were identified. About 85% of halogen bonds were established with protein backbone carbonyl oxygens. No nitrogen was involved in halogen bonds which reflects the low abundance of nitrogen acceptors in the protein structures contrary to e.g. advanced crystalline materials.¹¹ All the ligands were subject to the docking procedure, which is described in detail in the ESI section. The outcome of the docking was a set of geometries of the protein-ligand complexes. For each ligand 25 highest-ranked geometries were analysed. Although the ranking is based on electrostatic and van der Waals interactions, and therefore quite simplistic, it can consistently filter out non-physical ligand orientations. The root-mean-square deviation (RMSD) of the heavy atoms of the ligand was calculated with respect to the X-ray geometry.

The RMSDs are summarised in Table 1. The lowest RMSD, the highest RMSD and the average RMSD over all ligands and all their docked orientations were calculated. By visual inspection also the correct binding poses (i.e. “native orientations”) were distinguished. Evidently, all the RMSD descriptors are improved by inclusion of ESH. In other words, the geometries predicted by including ESH overcome those without ESH. The most striking distinctions appear in the case of CK2, where more than one halogen bond contributes to the binding arrangement.

The analysis of the lengths of the halogen bonds in protein-ligand complexes is presented in Table 2. The effect of ESH is emphasised by the halogen-acceptor distances, where ESH typically provides

	Lowest RMSD [Å]	Highest RMSD [Å]	Average RMSD [Å]	Natives detected
ALDR				
No ESH	0.11	3.83	1.74	7/7
ESH	0.08	3.67	1.21	7/7
CDK2				
No ESH	0.41	11.46	6.25	26/32
ESH	0.32	8.89	4.16	32/32
CK2				
No ESH	0.83	10.17	5.76	11/16
ESH	0.09	5.22	3.21	16/16
HIVRT				
No ESH	0.45	15.86	7.63	29/37
ESH	0.17	9.52	3.59	35/37

Table 1: Root mean square deviations (RMSDs) of heavy atoms of the ligand calculated with respect to the X-ray experimental geometries. Natives stands for the number of correctly identified binding poses

shorter halogen-acceptor contacts than those predicted in the absence of ESH, the shorter distances agreeing better with the experimental geometries. Also the number of halogen bonds established between the pose and protein is affected by the ESH presence. In CK2 complexes, which contain more than one halogen bond, the docking with ESH was able to reproduce 7 of 10 complexes with the correct halogen bonds pattern (i.e. all amino acids involved agreed with the X-ray data) compared to 3 of 10 without ESH.

To summarise, we performed a docking study of halogenated enzyme inhibitors. By including a molecular mechanistic model for σ -hole description we obtained generally better protein-ligand geometries than those accessed so far. It has to be noted that due to the simplicity of the ESH model, the improvement was reached without significant additional computational cost which makes it promising for all future docking studies involving halogenated compounds.

This work was supported by the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic [RVO: 61388963], and the Czech Science Foundation [P208/12/G016]. This work was also supported by the Operational Program Research and Development for Innovations – European Science Fund (CZ.1.05/2.1.00/03.0058). The support of

	No ESH	ESH	X-ray
ALDR			
V47(O)	3.53	3.38	3.04
T113(OG1)	3.31	3.32	3.32
CDK2			
I10(O)	4.17	3.67	4.15
E12(N)	5.59	3.58	3.56
F80(ring COM)	3.91	3.85	3.43
L83(O)	7.96	3.01	2.90
Q131(O)	4.93	3.64	3.62
D145(O)	7.55	3.61	4.06
CK2			
E108(O)	3.46	3.39	3.41
V110(O)	3.36	3.18	2.97
HIVRT			
K101(O)	6.99	3.88	4.06
Y188(O)	4.51	4.09	3.18
F227(ring COM)	6.39	4.76	4.40
L234(O)	5.48	4.00	4.12

Table 2: The lengths of halogen-acceptor contacts averaged over all protein-ligand complexes. Items in the first column, e.g. V47(O), stand for the average distance between valine 47 oxygen (acceptor atom) and the closest ligand halogen. All distances are in Å.

Praemium Academiae of the Academy of Sciences of the Czech Republic awarded to P. H. in 2007 is acknowledged. Gratitude is also expressed towards Interdisciplinary Centre for Scientific Computing (IWR) and HGS MathComp Graduate School, University of Heidelberg, Germany, for support of this study.

References

- [1] B. H. Munos, InnoThink Center For Research In Biomedical Innovation, Thomson Reuters Fundamentals via FactSet Research Systems, 2012.
- [2] A. Merino, A. K. Bronowska, D. B. Jackson and D. J. Cahill, *Drug Discovery Today*, 2010, 15, 749.
- [3] M. Z. Hernandez, S. M. Cavalcanti, D. R. Moreira, W. F. de Azevedo Jr and A. C. Leite, *Curr. Drug Targets*, 2010, 11, 303.
- [4] Y. Lu, T. Shi, Y. Wang, H. Yang, X. Yan, X. Luo, H. Jiang and W. Zhu, *J. Med. Chem.*, 1999, 52, 2854.
- [5] P. Auffinger, F. A. Hays, E. Westhof and P. S. Ho, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, 101, 16489.
- [6] P. Politzer, K. E. Riley, F. A. Bulat and J. S. Murray, *Comput. Theor. Chem.*, 2012, 998, 2.
- [7] P. Politzer, P. Lane, M. Concha, Y. Ma and J. Murray, *J. Mol. Model.*, 2007, 13, 305.
- [8] T. Clark, M. Hennemann, J. S. Murray and P. Politzer, *J. Mol. Model.*, 2007, 13, 291.
- [9] P. Politzer, J. S. Murray and M. Concha, *J. Mol. Model.*, 2008, 14, 659.
- [10] E. Corradi, S. V. Meille, M. T. Messina, P. Metrangolo and G. Resnati, *Angew. Chem., Int. Ed.*, 2000, 39, 1782.
- [11] P. Metrangolo, H. Neukirch, T. Pilati and G. Resnati, *Acc. Chem. Res.*, 2005, 38, 386.
- [12] M. A. A. Ibrahim, *J. Comput. Chem.*, 2011, 32, 2564.
- [13] S. Pieraccini, A. Forni and M. Sironi, *Phys. Chem. Chem. Phys.*, 2011, 13, 19508.
- [14] M. Kolář and P. Hobza, *J. Chem. Theory Comput.*, 2012, 8, 1325.
- [15] W. L. Jorgensen and P. Schyman, *J. Chem. Theory Comput.*, 2012, 8(10), 3895.
- [16] P. T. Lang, S. R. Brozell, S. Mukherjee, E. F. Pettersen, E. C. Meng, V. Thomas, R. C. Rizzo, D. A. Case, T. L. James and I. D. Kuntz, *RNA*, 2009, 15, 1219.
- [17] www.pdb.org.

Electronic Supplementary Information (ESI) for Chemical Communications

In this work, the following protein-ligand complexes were studied: 1IEI, 1US0, 1Z89, 1Z8A, 2IKI, 2IKJ, 2PFH, 2R3F, 1H07, 2VU3, 2R3K, 1WCC, 1PXI, 1FVT, 1H1R, 1H08, 1H01, 2R3J, 2J9M, 2V22, 2R3Q, 2I40, 2C68, 1P5E, 3MY5, 2VTJ, 2R3R, 2R3P, 1Y8Y, 1YKR, 3UNK, 2R3L, 2IW6, 2C69, 3LFS, 2VTR, 2B54, 2BHE, 3LE6, 3KXH, 2OXX, 1ZOH, 3PVG, 3KXG, 1ZOG, 3KXN, 1J91, 2PVK, 2OXY, 3KXM, 2QC6, 2OXD, 1ZOE, 3NGA, 3RPS, 1HNV, 1HNI, 3DYA, 3DLE, 3C6U, 2VG6, 1TKZ, 3MEC, 2VG5, 1FK9, 1TL1, 1RT5, 3C6T, 1VRU, 1RT6, 3I0R, 3FFI, 3DI6, 2RKI, 1TL3, 1RT7, 3E01, 2RF2, 1EP4, 3T19, 3DRP, 2VG7, 1TKT, 3I0S, 3DLG, 1DTT, 3R8D, 3QIN, 3HYF, 1JLG, 2YKM, 1IKX.

The charges of the ligands were assigned by the UCSF Chimera program suite [S1] at the AM1-BCC level in a standard manner [S2]. Then, the ESH was constructed as the nF model [S3] as described in Kolář and Hobza [S4]: the dummy atom with a desired positive charge was added to the halogen and the charge of the halogen was lowered by the same value. Hence, the net charge of the ESH-halogen pair remained identical as the initial halogen atom charge. None of the other atoms was modified. This model is well suited for high-throughput calculations since it does not require any additional quantum chemical calculation, once the atomic partial charges are known. On the other hand the effect of σ -hole is reduced only to the vicinity of the halogen. Nevertheless, its performance on interaction energies was proven to be sufficient [S4].

The ESH was added to all halogen atoms except fluorine, which is known not to create halogen bonds in organic drug-like molecules [S5]. The ESH parameters (charge, ESH-halogen distance) were chosen as follows and were not subject of any further optimization: (0.1 e, 1.0 Å) for chlorine, (0.2 e, 1.3 Å) for bromine, and (0.3 e, 1.6 Å) for iodine. These parameters follow the recommendation in Ref. 14 and also the known features of halogens, where iodine exhibits the largest -hole and chlorine the smallest. The ESH-halogen distance was, however slightly shortened when compared with Ref. 14 (i.e. 1.3 Å vs. 1.5 Å for bromine). Large ESH-halogen distance caused problems with the docking algorithm. Consequently, the change in improved scoring arises mainly from the improved electrostatics and also from the shape complementarity between ligand and receptor (distances being corrected for the presence of dummy atom mimicking the σ -hole). Molecular docking was performed using UCSF DOCK6.5 suite [S6], using a grid scoring, in an implicit solvent. The grid spacing was 0.25 Å, and the grid box included 12 Å beyond the ligand binding site. The energy score has been regarded as a sum of electrostatic and Van der Waals contributions. In the course of the docking procedure, the ligand was subjected to 2500 cycles of molecular- mechanical energy minimization. The number of maximum orientations was 5000.

[S1] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.*, 2004, 25, 1605.

[S2] A. Jakalian, D. B. Jack, C. I. Bayly *J. Comput. Chem.*, 2002, 23, 1623.

[S3] The script used to introduce ESH into MOL2 files is available upon request.

[S4] M. Kolář, P. Hobza, *J. Chem. Theory Comp.*, 2012, 8, 1325.

[S5] P. Politzer, K. E. Riley, F. A. Bulat, J. S. Murray, *Comput. Theor. Chem.*, 2012, 998, 2.

[S6] P. T. Lang, S. R. Brozell, S. Mukherjee, E. F. Pettersen, E. C. Meng, V. Thomas, R. C. Rizzo, D. A. Case, T. L. James, I. D. Kuntz, *RNA*, 2009, 15, 1219.