# PCCP

## Accepted Manuscript

# Journal Name

## COMMUNICATION

# *Statistical Analysis of σ-Holes: A Novel Complementary View on Halogen Bonding*

Michal H. Kolář,[abc*] Paolo Carloni[bcd] and Pavel Hobza[ae]

**To contribute to the understanding of noncovalent binding of halogenated molecules with a biological activity, electrostatic potential (ESP) maps of more than 2,500 compounds were thoroughly analysed. A peculiar region of positive ESP, called σ-hole, is a concept of central importance for halogen bonding. We aim at simplifying the view on σ-holes and provide general trends in organic drug-like molecules. The results are in fair agreement with crystallographic surveys of small molecules as well as of biomolecular complexes and attempt to improve the intuition of chemists when dealing with halogenated compounds.**

Halogen bonding (XB) is an attractive interaction between a halogen atom and a Lewis base. It has received appreciable attention in the last decade. It has found a distinguished place in drug[1-3] and material design.[4] The electrostatic and London dispersion interactions are the most prominent effects contributing to the stabilization of XB complexes.[5-7]

The anisotropy of electrostatic potential (ESP) (i.e. the σ-hole) has been recognised by quantum chemical calculations on halogen atoms,[8] and later also on the atoms of IV, V and VI groups.[9,10] The properties of halogen σ-holes have recently been abstracted into two characteristics: i) σ-hole *magnitude*, which stands for the value of the most positive electrostatic potential on the 0.001 e/bohr$^3$ electron density isosurface, and ii) σ-hole *size*, i.e. the spatial extent of the positive region.[11] Although the σ-hole magnitude and size were initially introduced only for molecules with a $C_{2v}$ symmetry point group, it has been shown that such characteristics provide a simplified but relevant view on ESP anisotropy and XB. The

magnitude correlates with the strength of XB[6] while the size is related to the directionality of XB, understood as an angular channel, where the interaction between the molecules is attractive.[11]

Here, we propose a generalization of the concept of σ-hole magnitude and size for non-symmetric organic molecules. Further, we apply such descriptors to characterize halogenated molecules in ZINC database.[12,13] This database is a keystone of virtual screening,[14] which is a starting point for numerous drug-design studies. Thus, we aim at elucidating the typical σ-hole of a drug candidate. So far, we have calculated and analysed ESPs of 331 molecules containing chlorine, 1,267 molecules containing bromine and 1,003 molecules containing iodine. All of these molecules have at least one C-X (X=Cl, Br, I) covalent bond. The actual number of σ-holes analysed was slightly higher, because some of the molecules comprised more than one halogen. From ZINC, we selected neutral purchasable molecules. The upper bound for molecular weight was 250, 300 and 350 Da for chlorinated, brominated and iodinated compounds, respectively.

All of the compounds were subjected to full energy minimization using a hybrid B3LYP functional with a 6-31G* basis set for all atoms except for iodine, for which a LANL2DZ basis with a pseudopotential for the inner-core electrons was used. The default convergence criteria of the Gaussian09 program package[15] were adopted (maximum force < 4.5e-4 a.u., root-mean-square force (RMS) < 3.0e-4 a.u., maximum displacement < 1.8e-3 a.u., RMS displacement < 1.2e-3 a.u.). The electrostatic potential maps were calculated on a three dimensional grid from the electron densities at the B3LYP/def2-QZVP level.

The σ-hole *magnitude* was defined, like in Ref. 10, as the value of the most positive (or the least negative) ESP lying on the halogen
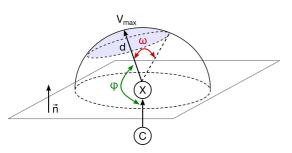
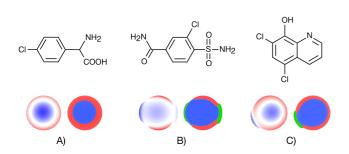Figure 1: The definitions of σ-hole characteristics.



Figure 2: Top views on three ESP maps of chlorine and their respective simplifications used for analysis. In the maps, the scale goes from the negative red trough white zero to the positive blue. In the simplifications, the σ-hole region is in blue, the negative ESP in red and the positive ESP excluded by clustering algorithm from the σ-hole region is in green; A) Baclofen (ZINC00000061), B) 4-chloro-3-sulfamoylbenzamide (ZINC00002088), C) Chloroxine (ZINC00001131), (ESP of the chlorine in para- position wrt the hydroxyl).



Figure 3: Probability density functions (pdf) of A) the magnitude, B) the size, C) angular deviation $\varphi$ D) threshold angle $\omega$. E) The dependence of the ESP on angle $\omega$ and F) the pdf of the σ-hole range as the boundary of the positive and negative ESP in the X-$V_{max}$ direction. The pdf was summed for distances larger than 9.0 Å.

boundary, arbitrarily defined as the 0.001 e/bohr³ electron density isosurface.[16] Only the isosurface lying in a certain half-space was considered.

The half-space was defined by the position of the halogen and by a plane whose norm *n* was collinear with the vector of the halogen and its closest atom (carbon) (see Figure 1). We further defined distance *d* between the halogen and the point with the $V_{max}$, C-X-$V_{max}$ angular deviation $\varphi$ and threshold angle $\omega$ between the σ-hole boundary and the vector *d* (Figure 1). Since other atoms may induce the most positive ESP with $\varphi$ of about 90, the magnitude is found as the local maximum of the ESP closest to the elongation of the C-X bond.

The σ-hole *size* for non-symmetrical molecules was defined as the area of the 0.001 e/bohr³ electron density isosurface where the ESP was positive. As shown in Figure 2, there may be molecules with a complicated shape of the positive ESP making the *total* area of positive ESP useless for σ-hole characterisation. Thus, we applied a clustering algorithm to select only such a part of the positive surface that reflects the topology of the ESP. It means that the area of positive ESP, which was induced by the surrounding chemical entities, was discarded from the total positive ESP area.

In order to obtain an approximately rounded boundary of the area of discrete set of positive ESP points, the *size* of the σ-hole was obtained as follows:

1) The point with the most positive ESP closest to the elongation of the C-X bond was localized.

2) For each *grid point* with a positive ESP, the angle $V_{max}$-X-point was calculated. A histogram of these angles was constructed.
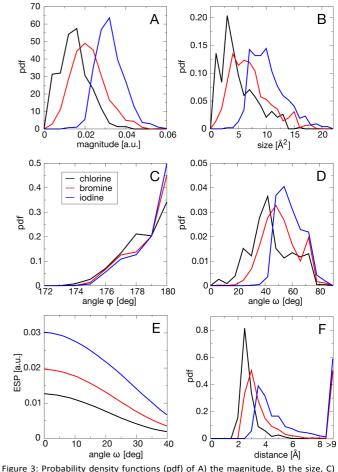
3) We counted all the points up to the first local minimum of the histogram and multiplied the sum by the area per point. The minimum was denominated as *threshold angle $\omega$*.

Such an approach selects only those positive ESP points, which deviate from the ESP maximum by less than the value of *threshold angle $\omega$*.

Obviously by this approach, no σ-hole size could be assigned to halogens with overall negative ESP, even though such negative regions have also been denominated as σ-holes before.[9] However here, only eight instances out of the entire set of drug-like organic compounds have been identified with no positive ESP region. Thus, a positive σ-hole is a general feature of such compounds.

The probability density functions (pdfs) are summarized in Figure 3. The magnitudes of the σ-holes (Figure 3A) show the previously known trend[5,17,18] that the maximum of the ESP increases with the increasing atomic number of the halogen. The pdfs are gaussian-like with the mean values of 0.0127, 0.0198 and 0.0304 a.u. for chlorinated, brominated and iodinated molecules. On average, the iodinated molecules thus have more than a twice more positive σ-hole than the chlorinated ones.

The deviation angle $\varphi$ (Figure 1) is higher than 178° for about 50% of the set and higher than 172° for almost the entire set. Surprisingly enough, the spatial position of $V_{max}$ in organic drug-like molecules is

hence only little affected by the nature of the molecule or, in other words, by the chemical environment of the halogen. It implies that the electrostatic interaction of σ-holes supports the strictly linear arrangement of halogen bonds (see below). Finally, the high values of angle φ provide the justification of placing a fixed positively charged particle on top of halogen atoms in molecular mechanistic studies.[19,20]

The sizes of σ-holes (Figure 3B) increase with the atomic number of the halogen. Obviously, the size correlates well with the angle ω ($R^2$=0.87). Moreover with the increasing atomic number of the halogen, the ω pdfs shift to higher values and their width decreases. Consequently, the size variability of the chlorine σ-holes is higher than in the case of the iodine ones. The size correlates with the magnitude ($R^2$=0.91), which was observed previously for symmetric molecules.[11]

The ESP projections on the 0.001 e/bohr$^3$ isodensities provide information on the electrostatics in the vicinity of a compound. We calculated the average distance of the $V_{max}$ point from the halogen, which was 1.93, 2.04 and 2.19 Å for chlorinated, brominated and iodinated compounds, respectively. However, it is often overlooked that the ESP changes with the distance from the halogen and consequently that the σ-hole as a region of the positive ESP may vanish at longer distances. We calculated the *range* of the σ-hole as the distance from the halogen, where the ESP changes its sign from positive to negative. The direction of X-$V_{max}$ was evaluated (Figure 3F). The maximums of the pdfs of the σ-hole ranges were 2.5, 3.0 and 3.5 Å from the chlorine, bromine and iodine, respectively, which is 1.25-, 1.50- and 1.75-multiple of their vdW radii. Importantly, the σ-holes of more than 24% of chlorinated, 25% of brominated and 30% of iodinated compounds are preserved beyond 9 Å from the halogen, which likely contributes to the middle-range stabilization of XB complexes.

As stated in our previous work,[11] "the channel allowing the approach of hydrogen fluoride to halobenzenes is the narrowest for Cl and the broadest for I," which is supported here by the pdfs of angle ω (Figure 3D). Thus the angular freedom of electron donors to establish an attractive interaction (i.e. with positive stabilization energy) with iodinated compounds should be larger than with chlorinated or brominated ones. This is, however, not reflected in crystallographic studies, from which it has resulted that the XBs involving iodine tend to be linear more than those with chlorine and bromine[21-24] This merely agrees with the pdfs of angle φ (Figure 3C), although the differences are rather small. Furthermore, Figure 3E shows the angular profile of the ESP averaged over the set of molecules. The profile is the steepest for iodine compounds. Hence, the force acting on the electron donor is the largest when iodine is involved, favouring the linear arrangement of XB more than in the cases of bromine and chlorine. In other words, the electrostatic energy penalty needed to deform the linear arrangement is higher for iodine than for chlorine.

We have to distinguish between directionality, previously determined as a solid-angular channel allowing attractive interaction, and tendency to linearity. Directionality seems to be well captured only by the size of the σ-hole. This means that directionality increases on passing e.g. from chlorobenzene to iodobenzene. The tendency to linearity is expressed by the angular variations of electric field and

thus it depends on both the magnitude and size of the σ-hole. Both, the magnitude and the size of the σ-hole, increase with the atomic number of halogen, but the increase of the magnitude is steeper.

In biomolecular complexes, it was observed[24] that the angular variations of C-X...O are significantly lower for iodine when compared to bromine and chlorine. This agrees well with the statistical analysis of σ-hole sizes and angles ω while it remains slightly puzzling for halogen with such large polarisability.

## Conclusions

To summarize, we have presented an abstraction of the σ-hole – a three-dimensional object pertaining to halogen atoms. The statistical analysis of σ-hole magnitude, size, angular deviation φ and range has revealed a surprisingly low effect of intramolecular polarization on the spatial position of $V_{max}$. The other properties are much more affected. Our findings bring a novel, refined view on halogenated molecules and their ability to participate in noncovalent interactions and help approach them since the analysed molecules represent a realistic set for general use. Finally, we admit that because the σ-holes of atoms of IV, V and VI group are significantly more complicated in shape, similar analysis of them seems not to be as straightforward as for halogens.

## Notes and references

$^a$ Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v. v. i., Flemingovo nam. 2, 166 10 Prague 6, Czech Republic. Email: michal.kolar@uochb.cas.cz

$^b$ Computational Biophysics, German Research School for Simulation Sciences GmbH, 52425 Jülich, Germany

$^c$ Institute for Advanced Simulations (IAS-5), Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany

$^d$ Institute of Neuroscience and Medicine (IMN-9), Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
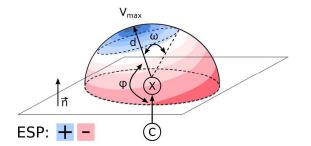
$^e$ Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, Olomouc, 771 46 Olomouc, Czech Republic

1   A. R Voth and P. S. Ho, *Curr. Top. Med. Chem.* 2007, **13**, 1336-1348.

2   Y. Lu, T. Shi, Y. Wang, H. Yang, X. Yan, X. Luo, H. Jiang and W. Zhu, *J. Med. Chem.* 2009, **52**, 2854-2862.

3   J. Fanfrlík, M. Kolář, M. Kamlar, D. Hurný, F. X. Ruiz, A. Cousido-Siah, A. Mitschler, J. Řezáč, E. Munusamy, M. Lepšík, P. Matějíček, J. Veselý, A. Podjarny and P. Hobza, *ACS Chem. Biol.* 2013, **8**, 2484-2492.

4   P. Metrangolo, H. Neukirch, T. Pilati and G. Resnati, *Acc. of Chem. Res.* 2005, **38**, 386-395.

5   P. Politzer, P. Lane, M. Concha, Y. Ma and J. S. Murray, *J. Mol. Model.* 2007, **13**, 305-311.

6   K. E. Riley and P. Hobza, *J. Chem. Theory Comput.* 2008, **4**, 232-242.

7   P. Politzer, J. S. Murray and T. Clark, *Phys. Chem. Chem. Phys.* 2010, **12**, 7748-7757.

8   T. Clark, M. Hennemann, J. S. Murray and P. Politzer, *J. Mol. Model.* 2007, **13**, 291-296.

9   J. S. Murray, P. Lane and P. Politzer, *J. Mol. Model.* 2009, **15**, 723-729.

10  P. Politzer, J. S. Murray, T. Clark, *Phys. Chem. Chem. Phys.* 2013, **15**, 11178.

11  M. Kolář, J. Hostaš and P. Hobza, *Phys. Chem. Chem. Phys.* 2014, **16**, 9987-9996.

12  J. J. Irwin and B. K. Shoichet, *J. Chem. Inf. Model.* 2004, **45**, 177-182.

13  J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, *J. Chem. Inf. Model.* 2012, 52, 1757-1768.

14  G. Klebe, *Drug Discov. Today* 2006, **11**, 580-594.

15  M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson et al., Gaussian09, Revision D.01, Gauss-ian, Inc., Wallingford CT, 2009.

16  R. F. W. Bader, M. T. Carroll, J. R. Cheeseman and C. Chang, *J. Am. Chem. Soc.* 1987, **109**, 7968-7979.

17  F. H. Allen, J. P. M. Lommerse, V. J Hoy, J. A. K. Howard and G. R. Desiraju, *Acta Crystallogr. B* 1997, **53**, 1006-1016.

18  J. Řezáč, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.* 2012, **8**, 4285-4292.

19  M. Kolář and P. Hobza, *J. Chem. Theory Comput.* 2012, **8**, 1325-1333.

20  M. Kolář, P. Hobza and A. K. Bronowska, *Chem. Commun.* 2013, **49**, 981-983.

21  F. H. Allen, *Acta Crystallogr. B* 2002, 58, 380-388.

22  J. P. M. Lommerse, A. J. Stone, R. Taylor and F. H. Allen, *J. Am. Chem. Soc.* 1996, **118**, 3108-3116.

23  A. Bauzá, D. Quiñonero, P. M. Deyà and A. Frontera, *CrystEngComm* 2013, **15**, 3137-3144.

24  P. Auffinger, F. A. Hays, E. Westhof, and P. S. Ho, *Proc. Natl. Acad. Sci. U. S. A.* 2004, **101**, 16789-16794.

**Physical Chemistry Chemical Physics Accepted Manuscript**

Graphical abstract:



Halogen σ-hole characteristics are defined and their typical values are provided based on an extensive analysis of more than 2500 molecules form ZINC database.