ISSN 0012-5008, Doklady Chemistry, 2016, Vol. 471, Part 1, pp. 338–342. © Pleiades Publishing, Ltd., 2016.
Original Russian Text © O.I. Titov, D.A. Shulga, V.A. Palyulin, N.S. Zefirov, 2016, published in Doklady Akademii Nauk, 2016, Vol. 471, No. 3, pp. 307–311.

CHEMISTRY

Quadrupole Correction for Halogen Bonding Description in Virtual Screening and Molecular Docking

O. I. Titov, D. A. Shulga, V. A. Palyulin*, and Academician N. S. Zefirov

Received April 29, 2016

Abstract—Halogen bonding is electrostatic attraction between halogen atoms in an organic molecule and Lewis bases. It is important to consider halogen bonding during molecular docking and virtual screening, in particular, at early stages of drug development. A new scoring function AutoDock-XB, which takes into account halogen bonding by means of the quadrupole correction, has been constructed. The function has been tested for a series of phosphodiesterase-5 inhibitors.

DOI: 10.1134/S0012500816110100

Halogen substitution has long been used in medicinal chemistry for improving the activity as well as pharmacokinetic and pharmacodynamic properties of compounds. Attention to the directional stabilizing interaction between Lewis bases and halogen atoms in organic molecules—halogen bonding (XB)—has been paid only recently. XB is explained by pronounced anisotropy of the molecular electrostatic potential (MEP) of heavy halogen atoms (Cl, Br, I). This anisotropy is conceptually explained by the presence of a σ-hole, a region of reduced electron density located on the extension of the carbon–halogen σ-bond [1].

Despite the fact that explicit XB modeling is promising for optimization of ligand–receptor interactions [2] and noticeable progress in theoretical description of XB at different levels applicable to modeling in drug development [3], practical and reliable tools for everyday work are still unavailable to researchers.

Virtual screening and molecular docking applied at early stages of the search for active compounds, as well as for subsequent optimization of a lead compound, use scoring functions (SFs) for fast evaluation of the ligand–protein interaction energy. To date, despite a significant number of available SFs, only few of them take into account halogen bonding, in particular, ScorpionScore [4], several modifications of XBPMF [5], and XBScore [6]; however, they are still not implemented in non-commercial software. In commercial program packages Glide [7], Gold [8], and some other, halogen bonding has already been considered; however, the corresponding mechanism has not been described in detail. Kolář et al. [9] have used a "dipole" correction to consider the anisotropy of the MEP of heavy halogen atoms in the framework of the DOCK6 program package [10]. This correction was modeled through the introduction of a positive extra-point charge (EP).

Previously, we have shown that the MEP anisotropy (σ-hole) is adequately described when placing a symmetric quadrupole on the halogen atom, with the quadrupole principal axis being oriented along the C–Hal bond [11]. It has also been demonstrated that, to a first approximation, the quadrupole magnitude weakly depends on the substituents in a molecule and is determined by the type of halogen atom [12]. Thus, the multipole description is a theoretically reasonable and practical way to create a new SF for XB.

We proceeded from AutoDock SF (v. 4.2 [13]), where the free energy of binding is evaluated by a weighted sum of molecular-mechanical components (electrostatic, van der Waals interactions, and hydrogen bonds) which also includes an empirical account for desolvation and the entropy penalty for freely rotating bonds. In this work, we introduced the sixth component into the AutoDock SF (E(AutoDock)), namely, the quadrupole halogen correction:

$$E(\text{AutoDock-XB}) = E(\text{AutoDock}) + \text{ScaleFactor} \times \text{XBCorrection}.$$

The correction was calculated as the energy of electrostatic interaction of symmetric atomic quadrupoles centered at the halogen atom with atomic charges of the protein target (XBCorrection) multiplied by a scaling factor (ScaleFactor). The quadrupole moments were taken to be 1.15, 1.59, and 2.76 au for the Cl, Br, and I atoms, respectively, which gave an acceptable description of the MEP of halogen-containing aromatic molecules [12]. The quadrupole moment itself was presented as a group of three

Moscow State University, Moscow, 119991 Russia
**e-mail: vap@qsar.chem.msu.ru*

Table 1. Results of the scoring power test of PDE5 inhibitors

Halogen (PDB ID)	ΔG_{bind}, kJ/mol			
	exp (IC_{50})	exp (ITC)	AutoDock	AutoDock-XB
H (4OEX)	−41.55	−33.35	−33.81	−33.81
F (3SHY)	−40.17	−32.89	−31.21	−31.21
Cl (3SHZ)	−42.43	−34.89	−32.68	−32.72
Br (3SIE)	−44.94	−36.40	−39.08	−39.04
I (4OEW)	−46.44	−38.91	−38.24	−37.99
R_P (IC_{50})			0.927	0.920
RMSE (IC_{50})			8.20	8.24
R_P (ITC)			0.840	0.829
RMSE (ITC)			1.76	1.76

Here and in Table 2, R_P is the Pearson coefficient, and RMSE is the root-mean-square error of the binding energy.

charges: the charge of the halogen atom and two EPs located on the carbon–halogen bond axis on both sides of the halogen atom at a distance of 0.1 Å (a multipole charge cluster model [11]). This allowed us to adequately describe XB with minor modification of the AutoDock program code. The extra charges and the halogen charge were calculated by the equations:

$$q_{outer} = \frac{Q_{zz}}{2r^2},$$

$$q_{Hal} = q_{partial} - 2q_{outer},$$

where q_{outer} is the extra-point charge, Q_{zz} is the quadrupole moment, q_{Hal} is the ultimate charge of halogen, $q_{partial}$ is the partial charge of the halogen atom, and r is

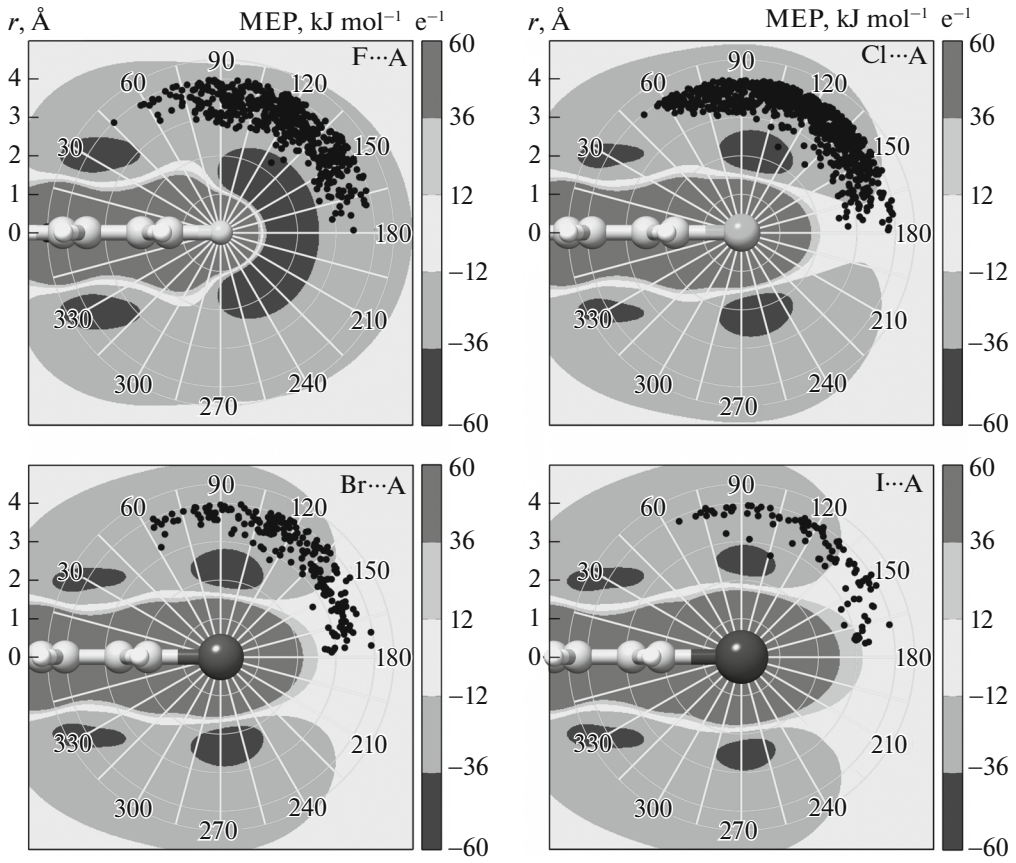

Fig. 1. Distribution of halogen bond acceptors around halogen atoms (rightward and downward: F, Cl, Br, I) against the background of the MEP of the corresponding phenyl halide.

Table 2. Results of the docking power test of PDE5 inhibitors

Halogen (PDB ID)	ΔG_{bind}, kJ/mol				RMSD, Å	
	exp (IC_{50})	exp (ITC)	AutoDock	AutoDock-XB	AutoDock	AutoDock-XB
H (4OEX)	–41.55	–33.35	–43.64	–43.76	1.38	1.4
F (3SHY)	–40.17	–32.89	–41.00	–40.88	1.92	1.94
Cl (3SHZ)	–42.43	–34.89	–45.31	–45.73	1.22	1.26
Br (3SIE)	–44.94	–36.40	–46.78	–48.99	2.76	2.65
I (4OEW)	–46.44	–38.91	–46.11	–47.91	2.90	2.85
R_P (IC_{50})			0.874	0.920		
RMSE (IC_{50})			1.84	2.64		
R_P (ITC)			0.799	0.845		
RMSE (ITC)			9.37	10.29		

the distance between the EP and the halogen. As partial charges of both the ligands and the proteins, we used Gasteiger charges calculated with AutoDock Tools [13].

To calibrate the scoring function, 560 complexes were selected from the PDBBind-CN database (version 2014) [14], in which 1609 contacts between the halogens in ligands (F, Cl, Br, I) and potential halogen bond acceptors in proteins (O, N, S) were found. The distributions (Fig. 1) show that the set contains all possible combinations of halogen contact distances and angles.

For the final test, a series of phosphodiesterase-5 inhibitors (PDE5) was selected. For this protein, the activities of five ligands that differ only by the substitution of one atom are available (H, F, Cl, Br, and I; PDB ID: 4OEX, 3SHY, 3SHZ, 3SIE, and 4OEW, respectively). The structures were prepared in a standard way: all water molecules and metal ions were removed, hydrogen atoms were added, and atomic charges were calculated. Special attention was paid to the retention of the amide tautomeric form of the ligand upon the saturation of valences with hydrogen atoms.

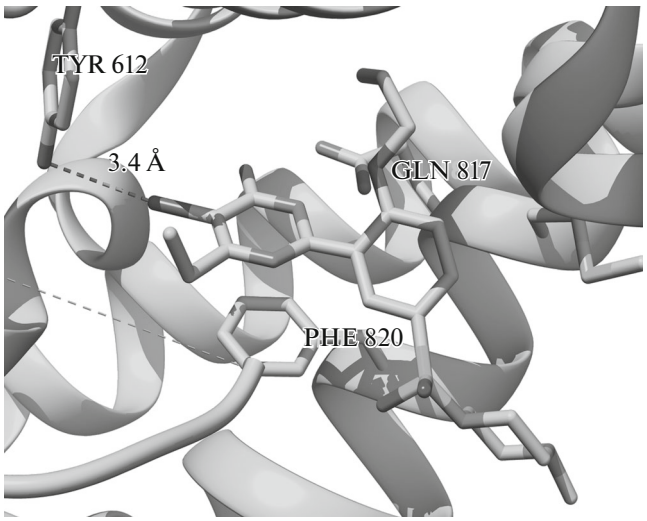

Fig. 2. A Br-containing ligand in the binding pocket of the PDE5 protein (PDB ID: 3SIE).

The SF was calibrated using the prepared set of 560 ligand–protein complexes. Minimization of the root-mean-square deviation gave ScaleFactor close to unity, whereas the factor for the electrostatic contribution of AutoDock SF was 0.1406.

The resulting SF (AutoDock-XB, ScaleFactor is 1) gives results close to the AutoDock SF (the correlation coefficient between the estimate and the experimental value is 0.53 for both functions, the root-mean-square deviation is 10.334 and 10.318 kJ/mol for AutoDock and AutoDock-XB, $N = 527$). A slight improvement of description is explained, first of all, by significant residual errors in the description of complex energies by means of modern SFs (standard deviation on the order of 7–10 kJ/mol [15]), comparable with the XB effect or exceeding it.

For most of the studied complexes, the contribution of the halogen correction with account for the protein environment of the ligand turned out to be insignificant, which corresponds to the absence of XB or to a combination of favorable and unfavorable contacts. However, in some complexes, the correction turned out to be rather significant and led to a noticeable improvement of the description of binding energy, for example, –6.11 kJ/mol in 3DP2 and +1.51 kJ/mol in 4F8H (the error was reduced from +8.33 to

+2.22 kJ/mol and from −1.26 to +0.25 kJ/mol for 3DP2 and 4F8H, respectively).

After calibration of the SF, it was tested for inhibitor-PDE5 complexes. The inhibitors under consideration are structural analogues of the commercially successful drug sildenafil and differ only by one halogen atom as a substituent in the heterocyclic ring. This halogen atom has a short contact with the oxygen atom of tyrosine, a typical halogen bond acceptor (Fig. 2). For these inhibitors, the data on activity (IC_{50}) and thermochemical estimates of free energy of binding (ITC) are available; therefore, such complexes are a good model for testing the XB description.

First, the binding energies for the complexes with the experimental geometry (taken from the PDB database) were evaluated without optimization (scoring power test). It follows from Table 1 that, for both experimental series (IC_{50} and ITC), the introduction of the anisotropic quadrupole halogen correction reduces the error of energy evaluation, even if insignificantly. This test shows that the developed SF can be used for estimating the binding positions revealed in the course of docking with the use of other SFs. This should lead to a better enrichment of the sets with promising structures during virtual screening.

In the second step, the resulting SF was tested in the docking problem, searching for a set of optimal ligand binding modes in the receptor pocket on the bases of the SF (docking power test). Typically, docking is performed into the receptor structure taken from the complex with another ligand or constructed by homology. Docking is used at later stages of drug development for optimization of lead compounds; therefore, it is equally essential in this experiment to reveal the binding mode and correctly evaluate the relative binding energy in a series of structural analogues of the ligands.

As a protein structure, the PDE5 complex with a Br-containing ligand (PDB ID 3SIE) was used since it has the best resolution. Docking was successful (Table 2). The root-mean-square deviation of the calculated ligand position from the experimental one in the complexes containing H, F, and Cl is within 2 Å, which is a very good result; for Br- and I-containing complexes, the standard deviation is within 3 Å, which is also accepted as successful reproduction of the complex geometry. For the description of the binding energy, both approaches—AutoDock and AutoDock-XB showed themselves equally well: although AutoDock-XB leads to a slightly higher error, it correctly describes the trend in activity (which is reflected in correlation coefficients between the calculated and experimental values), which is important for the ranking of ligands in the problems of optimization of the lead compounds when selecting the compounds for synthesis and tests.

It is worth noting that, in the case of the scoring power test, the binding energy estimate turned out to be close to the thermochemical data (ITC), whereas in the case of the docking power test, this estimate was close to the data on activity (IC_{50}) used for calibration of AutoDock SF, which are eventually more valuable for medicinal chemistry since they implicitly take into account important factors that have an effect on the experimental activity.

Thus, on the basis of the AutoDock SF, we constructed the new AutoDock-XB SF that uses the quadrupole correction for the description of the anisotropy of heavy halogen atoms and for more adequate description of XB.

The new SF was calibrated using a set of more than 500 ligand–protein complexes; it demonstrated performance comparable with original AutoDock. For the test complexes of the PDE5 protein with halogen-substituted inhibitors, it was shown that AutoDock-XB can be successfully used for estimating the binding energy and searching for the binding mode.

Improvement in accuracy, detailed description, and interpretability of intermolecular interactions is a current trend in the development of modeling tools used in drug design. Halogen bonding, rarely making the major contribution to the ligand–receptor interaction energy, is nevertheless an important component that allows one to optimize both the affinity of new ligands for a specified target and the selectivity to other targets as well as physicochemical and pharmacodynamic properties. The use of the quadrupole term for the description of XB has a clear theoretical basis and significantly simplifies further improvement of the methods considering XB due to concerted generation of atomic charges and the quadrupole correction for arbitrary organic molecules.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 14–03–00851-a).

REFERENCES

1. Clark, T., Hennemann, M., Murray, J.S., and Politzer, P., *J. Mol. Model.*, 2007, vol. 13, no. 2, pp. 291–296.
2. Lu, Y., Liu, Y., Xu, Z., Li, H., Liu, H., and Zhu, W., *Expert Opin. Drug Discov.*, 2012, vol. 7, no. 5, pp. 375–383.
3. Kolář, M.H. and Hobza, P., *Chem. Rev.*, 2016, vol. 117, no. 9, pp. 5155–5187.
4. Kuhn, B., Fuchs, J.E., Reutlinger, M., Stahl, M., and Taylor, N.R., *J. Chem. Inf. Model.*, 2011, vol. 51, pp. 3180–3198.
5. Liu, Y., Xu, Z., Yang, Z., Chen, K., and Zhu, W., *J. Mol. Model.*, 2013, vol. 19, no. 11, pp. 5015–5030.
6. Zimmermann, M.O., Lange, A., and Boeckler, F.M., *J. Chem. Inf. Model.*, 2015, vol. 55, no. 3, pp. 687–699.
7. Friesner, R.A., Murphy, R.B., Repasky, M.P., Frye, L.L., Greenwood, J.R., Halgren, T.A., Sanschagrin, P.C., and

Mainz, D.T., *J. Med. Chem*., 2006, vol. 49, pp. 6177–6196.

8. Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., and Taylor, R.D., *Proteins: Struct., Funct., Genet.*, 2003, vol. 52, pp. 609–623.
9. Kolář, M.H., Hobza, P., and Bronowska, A.K., *Chem. Commun*., 2013, vol. 49, pp. 981–983.
10. Lang, P.T., Brozell, S.R., Mukherjee, S., Pettersen, E.F., Meng, E.C., Thomas, V., Rizzo, R.C., Case, D.A., James, T.L., and Kuntz, I.D., *RNA*, 2009, vol. 15, no. 6, pp. 1219–1230.
11. Titov, O.I., Shulga, D.A., Palyulin, V.A., and Zefirov, N.S., *Dokl. Chem*., 2013, vol. 450, part 1, pp. 139–143.
12. Titov, O.I., Shulga, D.A., Palyulin, V.A., and Zefirov, N.S., *Mol. Inf.*, 2015, vol. 34, nos. 6/7, pp. 404–416.
13. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., and Olson, A.J., *J Comput. Chem*., 2009, vol. 30, pp. 2785–2791.
14. Liu, Z., Li, Y., Han, L., Li, J., Liu, J., Zhao, Z., Nie, W., Liu, Y., and Wang, R., *Bioinformatics*, 2015, vol. 31, no. 3, pp. 405–412.
15. Cheng, T., Li, X., Li, Y., Liu, Z., and Wang, R., *J. Chem. Inf. Model*., 2009, vol. 49, no. 4, pp. 1079–1093.

Translated by G. Kirakosyan