Acta Cryst B
STRUCTURAL SCIENCE
CRYSTAL ENGINEERING
MATERIALS
ISSN 2052-5206

Crystallographic searches for weak interactions – the limitations of data mining

Hans-Jörg Schneider*

FR Organische Chemie, Universität des Saarlandes, Stadtwald, Saarbrücken, D-66041, Germany. *Correspondence e-mail: ch12hs@rz.uni-sb.de

Received 20 April 2018
Accepted 24 May 2018

Edited by A. Nangia, CSIR–National Chemical Laboratory, India

Keywords: weak non-covalent interactions; data mining; database searches; crystallography.

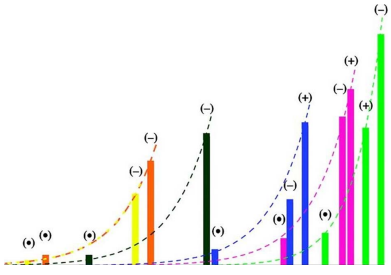
1. Introduction

Weak non-covalent interactions can exert a major influence on molecular structures and properties, due to their great number particularly in organic compounds. Weak interactions have been aptly reviewed, mainly with respect to the often used method of crystallography (Desiraju & Steiner, 1999, 2001; Desiraju, 2002, and references cited therein). Typically one observes large scatter and broad histograms, *e.g.* with the very weak C—H$\cdots\pi$ bonds (Takahashi *et al.*, 2010). Even stronger hydrogen bonds with ionic partners show up as rather broad profiles in histograms of interaction energies, *E*, as functions of donor–acceptor distances (Fig. 1) (Gilli *et al.*, 2009).

2. Organic fluorine as hydrogen-bond acceptor

The possible hydrogen bond with organic fluorine is one of the most debated weak non-covalent interactions and it has been reviewed several times (Schneider, 2012; Champagne *et al.*, 2015; Dalvit & Vulpetti, 2016); including a recent paper in this journal (Taylor, 2017). There the often-cited controversy is blamed on a misunderstanding, mainly due to the neglect of competing stronger interactions. However, the most often-cited 1997 paper by Dunitz and Taylor (1997) not only carries in the title *Organic Fluorine Hardly Ever Accepts Hydrogen Bonds*, based on their finding of only 0.6% relevant hits in the CSD, but claimed that the weakness of this hydrogen bond was backed by molecular orbital calculations, was in accord with physicochemical studies and with physical properties of fluorinated organic compounds.

The problem with many crystallographic studies is that all kinds of fluorine-containing compounds are present in the Cambridge Structural Database (CSD), and that most often fluorine has not been introduced in there with the aim to allow hydrogen bonds with fluorine. Fluorine is used very often in view of its well known strong substituent effects (Smart, 2001), also on electrostatic potentials; it can in steroids, for example, lead to a twofold decrease of reaction rates at distances as large as 10 Å, or corresponding ^{13}C NMR shift changes by 0.25 p.p.m. (Schneider & Becker, 1989). That statistical analyses of the Protein Data Bank (PDB) exhibit more contacts with fluorine as acceptor is not really surprising (Taylor, 2017), as in contrast to databases such as the CSD, which include all kinds of synthetic derivatives, many



© 2018 International Union of Crystallography

compounds in the PDB have been included knowing that fluorine produces special effects in reaction with proteins. A recent publication describes solid solutions of two drugs which differ only by a single fluorine atom in which F···F interactions play a decisive role (de Castro Fonseca *et al.*, 2018).

With necessary precautions, all new crystallographic and computational analyses, as well as equilibrium measurements in solution, now support the existence of weak hydrogen bonds with organic fluorine.

3. The limitations of data mining

Data mining has become a powerful tool to extract chemical information from a multitude of data sets, including crystal structures (Hautier, 2014). In particular drug discovery is widely supported by data mining (Wassermann *et al.*, 2015; Yang *et al.*, 2009) and is closely related to the identification of non-covalent forces for drug binding. Related approaches for the prediction of host–guest complexes are based on the use of training sets from data of known complexes and/or molecular similarity-based screening methods. Usually many descriptors are necessary for stability prediction, *e.g.* seven descriptors for cyclodextrin complexes (Steffen *et al.*, 2017).

Chemistry and physics are sciences where compounds or systems can be designed which allow properties such as weak interactions to be analysed experimentally with intelligent approaches, in contrast to social sciences, economics or biology which essentially rely on the analysis of already existing systems. Thus, systematic analyses of supramolecular complexes in solution have provided consistent experimental data for all kinds of intermolecular forces, which can be used also to reliably predict stabilities of biological associations and

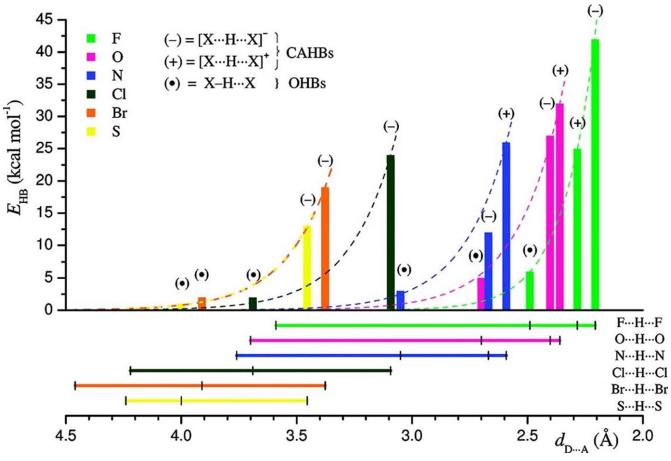


Figure 1
Large variations of donor···acceptor distances in hydrogen bonds. Hydrogen-bond energies E as a function of $D\cdots A$ distances; negative or positive charge-assisted hydrogen bonds marked as (−) or (+); ordinary hydrogen-bond marked as (●); colored horizontal lines on the bottom show the ranges of variation of the $D\cdots A$ distances for each type of bond from $d_D{\cdot}A$ (vdW) to the shortest value d_D. The dashed curve connecting the (●) points represents the particularly weak interactions with organic halogen acceptors. Reprinted with permission from Gilli *et al.* (2009). Copyright with permission 2009 American Chemical Society.

can be compared to interactions in crystals (Schneider, 2009; Biedermann & Schneider, 2016; Hunter, 2004).

With respect to crystallographic methods, one should not forget that they measure structures, not energies, and that crystal packing is determined by a multitude of interactions (Dunitz & Gavezzotti, 2005, 2009). The use of van der Waals radii as the cut-off criterion for weak interactions has been criticized (Aakeroy *et al.*, 1999). Empirical distance and angle relationships between large numbers of interacting aggregates have been proposed for the distinction between van der Waals and hydrogen-bonded interactions (van den Berg & Seddon, 2003). It also has been stated that the lowest-energy structure need not be present in a crystal (Dunitz & Gavezzotti, 2009). For example, a solid-state structure of a complex between a cyclophane and benzene showed the benzene ring outside the cavity (Hilgenfeld & Saenger, 1982), while an NMR study in solution showed the benzene ring inside in the cavity (Wald & Schneider, 2009).

As far as the frequency of intermolecular contacts in crystals is used as the indicator for non-covalent interactions, the indiscriminate use of all available data in the CSD can be misleading, particularly when it comes to weak forces. Instead one can, for example, limit the search to compounds which contain only weak donors and acceptors. Desiraju *et al.* have studied many structures where fluorine is the only heteroatom and always found strong evidence for its propensity as an acceptor, even with the C—H bond as a donor (Desiraju, 2005; Thakur *et al.*, 2010). Solid solutions of benzoic acid and fluorobenzoic acids show evidence of C—H···F hydrogen bonds, with their strength increasing with higher F content in the crystal structures (Chakraborty & Desiraju, 2018). Alternatively, one could exclude not only metal-containing structures but also those with short distances to known stronger acceptors, such as oxygen, nitrogen, *etc*. Such a screening would also limit the often observed scatter and large variation of distances (Fig. 1) in crystals.

References

Castro Fonseca, J. de, Clavijo, J. C. T., Alvarez, N., Ellena, J. & Ayala, A. P. (2018). *Cryst. Growth Des.* **18**, 3441–3448.

Aakeroy, C. B., Evans, T. A., Seddon, K. R. & Pálinkó, I. (1999). *New J. Chem.* **23**, 145–152.

Berg, J. van den & Seddon, K. R. (2003). *Cryst. Growth Des.* **3**, 643–661.

Biedermann, F. & Schneider, H.-J. (2016). *Chem. Rev.* **116**, 5216–5300.

Chakraborty, S. & Desiraju, G. R. (2018). *Cryst. Growth Des.* **18**, 3607–3615.

Champagne, P. A., Desroches, J. & Paquin, J.-F. (2015). *Synthesis*, **47**, 306–322.

Dalvit, C. & Vulpetti, A. (2016). *Chem. Eur. J.* **22**, 7592–7601.

Desiraju, G. R. (2002). *Acc. Chem. Res.* **35**, 565–573.

Desiraju, G. R. (2005). *Chem. Commun.* **24**, 2995–3001.

Desiraju, G. R. & Steiner, T. (1999). In *The Weak Hydrogen Bond in Structural Chemistry and Biology*. Oxford University Press.

Desiraju, G. R. & Steiner, T. (2001). In *The Weak Hydrogen Bond in Structural Chemistry and Biology*. 2nd edition. Oxford University Press.

Dunitz, J. D. & Gavezzotti, A. (2005). *Angew. Chem. Int. Ed.* **44**, 1766–1787.

Dunitz, J. D. & Gavezzotti, A. (2009). *Chem. Soc. Rev*. **38**, 2622–2633.
Dunitz, J. D. & Taylor, R. (1997). *Chem. Eur. J.* **3**, 89–98.
Gilli, P., Pretto, L., Bertolasi, V. & Gilli, G. (2009). *Acc. Chem. Res.* **42**, 33–44.
Hautier, G. (2014). *Top. Curr. Chem.* **345**, 139–179.
Hilgenfeld, R. & Saenger, W. (1982). *Angew. Chem. Int. Ed. Engl.* **21**, 1690–1701.
Hunter, C. A. (2004). *Angew. Chem. Int. Ed.* **43**, 5310–5324.
Schneider, H.-J. (2009). *Angew. Chem. Int. Ed.* **48**, 3924–3977.
Schneider, H.-J. (2012). *Chem. Sci.* **3**, 1381–1394.
Schneider, H.-J. & Becker, N. (1989). *J. Phys. Org. Chem.* **2**, 214–224.
Smart, B. E. (2001). *J. Fluor. Chem.* **109**, 3–11.
Steffen, A., Karasz, M., Thiele, C., Lengauer, T., Kämper, A., Wenz, G. & Apostolakis, J. (2017). *New J. Chem.* **31**, 1941–1949.
Takahashi, O., Kohno, Y. & Nishio, N. (2010). *Chem. Rev*. **110**, 6049–6076.
Taylor, R. (2017). *Acta Cryst.* B**73**, 474–488.
Thakur, T. S., Kirchner, M. T., Bläser, D., Boese, R. & Desiraju, G. R. (2010). *CrystEngComm*, **12**, 2079–2085.
Wald, P. & Schneider, H.-J. (2009). *Eur. J. Org. Chem.* **2009**, 3450–3453.
Wassermann, A. M., Lounkine, E., Davies, J. W., Glick, M. & Camargo, L. M. (2015). *Drug Discovery Today*, **20**, 422–434.
Yang, Y., Adelstein, S. J. & Kassis, A. I. (2009). *Drug Discovery Today*, **14**, 147–154.