Electronic Supporting Information

Unravelling the electronic nature of C-F…O-C non-covalent interaction in proteins and small molecules in the solid state

Abhijit Rana^[ab], Bartomeu Galmés^[c], Antonio Frontera^[c]*, Himansu S. Biswal^[ab]*, Deepak Chopra^[d]*

Computational Methods

For small molecules (from CSD): The Cambridge Structure Database (CSD)¹ version 5.41 with updates until November 2019 was inspected using ConQuest version 2.0.4 (build 270014). All queries were limited to single crystal structures (no polymers) that contained 3D coordinates and had no errors. Only those CIFs with F…O distances \leq than the sum of van der Waals radii and C–F…O angle $\geq 160^{\circ}$ were selected and further inspected manually. The statistical plots showing the distribution of F…O distances and C-F…O angles for fluorine atoms connected to sp^2 and sp^3 hybridized carbon atoms is given below (Figure S1).



Figure S1: Distribution of C-F covalent bond lengths, F…O distances and C-F…O angles in the crystal structures of small molecules.

The energies of all complexes included in this study were computed at the PBE0-D3/def2-TZVP level of theory by using the program TURBOMOLE $v7.0.^2$ For the calculations we have used the Weigend def2-TZVP^{3,4} basis set and the PBE0⁵ DFT functional. The TURBOMOLE default cut-offs have been used for the optimizations. The interaction energy (or binding energy in this work) E_{int} , is defined as the energy difference between the complex and the sum of the energies of the monomers. The molecular electrostatic potential surfaces have been plotted using the 0.001 a.u. isosurface using Gaussian-16 software at the PBE1PBE-D3/def2-TZVP level of theory.⁶ The PBE1PBE is the notation used by Gaussian-16 for the hybrid functional PBE0.

The QTAIM method has been used to obtain the critical points and bond paths via analysis of the topology of the electron density,^{7,8} using the PBE1PBE-D3/def2-TZVP wavefuntion and the AIMALL program.⁹ The NCIPLOT is a computational index that is convenient for the easy visualization and identification of NCIs. The noncovalent contacts are identified with the peaks that emerge in the reduced density gradient (RDG) at low densities.¹⁰ These are plotted by mapping an isosurface of s (s = $|\nabla p|/p^{4/3}$) for a low value of RDG. When a supramolecular dimer is formed, the RDG changes at the CPs in between the monomers due to the annihilation of the density gradient at these CPs. Therefore, this index visualizes the extent to which NCIs stabilize a supramolecular assembly qualitatively and reveals which molecular regions interact. The color code is red-yellow-green-blue where the red and yellow colors are used for strong and weak repulsive (ρ_{cut}^{-}), respectively. The Natural Bond Orbital analysis of orbital donor-acceptor interactions has been performed at the same level using the NBO 3.1 program.¹¹

For macromolecules (from Protein data bank): A data set of 11,332 protein X-ray structures with a high resolution (less than 1.5) Å was retrieved from the RCSB protein data bank.¹² The geometric parameters used to search C-F...O interactions were (a) 2.6 Å \leq distance between F and O \leq 3.3 Å , and (b) $120^{\circ} \leq (\angle C - F^{\circ\circ\circ}O) \leq 180^{\circ}$. These interactions were found using an in-home program written in Python, which was designed to execute batch processing of the PDB files. We found 212 interactions satisfying these structural criteria. Stride program¹³ was used to assign secondary structures to the amino acid residues, which contain the interacting oxygen atom.

The geometry optimization of protein-ligand complexes wasperformed at the B97-D3/aug-cc-pVDZ level of theory by using the program TURBOMOLE v6.5.² This

level of theory has been benchmarked with experimental data¹⁴. We performed only freeze-coordinate geometry optimization i.e. the co-ordinates of non-hydrogen atoms were freezed at their atomic positions in the crystal structure of the protein. The procedure and level of theory used for AIM, NBO, MEP and NCI were same as described for thesmall molecule structures from the CSD.

QTAIM (Small Molecules):

In Figures S2 and S3 (below) the distribution of critical points (CPs) and bond paths obtained for the structures analyzed from the CSD. It can be observed that in all cases there is a bond critical point (green sphere) and bond path interconnecting the O and F-atoms, which is an unambiguous evidence of contact. Moreover, in most of the structures there is an intricate combination of interactions (several bond CP and bond paths connecting both monomers).



Figure S2: QTAIM distribution of bond, ring and cage CPs in structures CEMYEY, (a), DAZHEU (b), IMOPIL (c) and AXELIA (d) and QECCUW (e).



Figure S3: QTAIM distribution of bond, ring and cage CPs in structures (a)AFUBIO, (b)JIQTUB, (c) JIYWOF, (d) MATSEG, (e)BEWDIS, (f) RABBUN, (g) POKJUU and (h) ISASAY.



Figure S4: Relevant NBO donor-acceptor interactions in (a)PALDOW and (b)DUSREQ (H-atoms omitted). Distances in Å.



Figure S5: Relevant NBO donor-acceptor interactions in (a)AKUDIV and (b) BAQGAE. Distances in Å.

Table S1:CSDrefcodesof X-ray structures exhibiting C–F···O contacts. For each structure, the MEP values (kcal/mol) at the F atom opposite to C–F, and at the O-atom of the electron donor group is given (isosurface 0.001 a.u.) at the PBE1PBE-D3/def2-TZVP level of theory.

REFCODE	V _{s,C-F}	V _{s,O}	REFCODE	V _{s,C-F}	V _{s,O}
C(sp2)FO			C(_{sp3})FO		
CEMYEV	-18.8	-36.4	AFUBIO	-7.5	-24.5
DAZHEU	-11.9	-20.7	JIQTUB	-5.1	-10.0
IMOPIL	-24.5	-35.8	JIYWOF	-6.3	-9.4
AXELIA	-17.6	-28.8	MATSEG	-9.4	-30.7
QECQUW	-6.9	-21.3	BEWDIS	-20.7	-15.7
			RABBUN	-7.5	+3.1
			POKJUU	-20.0	-25.1
			ISASAY	-12.5	-2.6

REFCODE	ΔE_{int}	Distance (Å)	Angle (°)			
C(_{sp2})-FO						
PALDOW	-2.36	3.124	175.1			
AKUDIV	-2.81	3.010	167.7			
CEMYEY	-0.65	3.165	166.1			
DAZHEU	-6.08	2.966	175.7			
IMOPIL	-6.66	3.048	167.3			
AXELIA	-2.16	3.043	171.8			
QECCUW	-8.16	3.066	164.6			
C(_{sp3})-FO						
DUSREQ	-7.73	3.041	168.7			
BAQGAE	-9.88	3.138	162.0			
AFUBIO	-11.02	3.093	163.9			
JIQTUB	-3.68	3.184	177.5			
JIYWOF	-11.69	2.932	175.8			
MATSEG	-2.37	3.125	163.5			
BEWDIS	-2.02	3.154	166.2			
RABBUN	-2.53	2.882	173.4			
POKJUU	-9.98	2.955	171.0			
ISASAY	-8.68	3.092	162.0			
PDB Code						
5JID	-1.75	2.820	142.0			
5HVE	-3.48	3.057	140.9			
5LWU	+0.38	2.919	150.0			
6FEX	-5.80	2.940	166.0			
5UQA	-0.73	2.624	147.7			
4QZS	-0.14	2.883	174.1			

Table S2. Interaction energies (kcal/mol) at the PBE1PBE-D3/def2-TZVP level of theory and geometric features of the dimers exhibiting $C-F\cdots O$ contacts.

It is particularly strange the case of BEWDIS, since the MEP value opposite to the C–F bond is large and negative, however the binding energy is attractive and there is only

one bond CP connecting both dimers. Perhaps there is some long range attraction between the F-atom and the H-atoms.

Some structures present clear C–H \cdots F bonds and other C–H \cdots O/N interactions that clearly dominate the interaction and explain the magnitudes of the obtained large binding energies of Table S2.

References

1 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, ActaCryst B, 2016, B72, 171-179.

2 R. Ahlrichs, M. Bär, M. Hacer, H. Horn, and C. Kömel, Chem. Phys. Lett. 1989, 162, 165–169.

F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. 2005, 7, 3297–3305.

4 F. Weigend, Phys. Chem. Chem. Phys. 2006, 8, 1057–1065.

5 C. Adamo and V. Barone, J. Chem. Phys., 1999, 110, 6158–6169.

Gaussian 16, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E.
Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H.
Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts,
B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D.
Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T.
Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M.
Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y.
Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E.
Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N.
Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J.
C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R.
Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D.
J. Fox, Gaussian, Inc., Wallingford CT, 2016.

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

8 R. F. W. Bader, Atoms in Molecules, A Quantum Theory, Clarendon, Oxford, 1990.

9 AIMAll (Version 17.11.14), T. A. Keith, TK Gristmill Software, Overland Park KS, USA, (aim.tkgristmill.com), 2013.

10 E. R. Johnson, S. Keinan, P. Mori-Sanchez, J. Contreras-Garcia, A. J. Cohen and W. Yang, J. Am. Chem. Soc., 2010, 132, 6498.

11 F. Weinhold and C. R. Landis, Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective, Cambridge University Press, Cambridge, UK, 2005.

12URL <u>www.rcsb.org</u>: H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. The Protein Data Bank Nucleic Acids Research, 2000,28, 235-242.

13 M. Heinig, D. Frishman, STRIDE: a Web server for secondary structure assignment from known atomic coordinates of proteins. Nucl. Acids. Res., 2004, 32, W500-2.

14 Thioamide, a Hydrogen Bond Acceptor in Proteins and Nucleic Acids. V. R.Mundlapati, S. Gautam, D. K. Sahoo, A. Ghosh, and H. S. Biswal; J. Phys. Chem. Lett. 2017, 8, 18, 4573–4579.