Pergamon

Tetrahedron, Vol. 52, No. 38, pp. 12613-12622, 1996

0040-4020/96 $15.00 + 0.00

PII: S0040-4020(96)00749-1

How Good is Fluorine as a Hydrogen Bond Acceptor?

Judith A. K. Howard,* Vanessa J. Hoy, David O'Hagan* and Garry T. Smith

Department of Chemistry, University of Durham, Science Laboratories, South Road, Durham. DH1 3LE.

Abstract: This study is aimed at evaluating organic fluorine as a hydrogen bonding acceptor. A review of short F···H contacts from all of the organofluorine compounds deposited in the Cambridge Structural Database System (CSDS) was carried out and in parallel a theoretical estimate of the energy of such contacts with inter nuclear distance was executed. A total of 548 structures emerged which contained 1163 unique C-F bonds and only 166 of these fluorine atoms posessed short C-F···H-X contacts of ≤ 2.35Å. Contacts between fluorine and hydrogen bound to carbon (C-F···H-C) represent the major category of short contacts however these were not judged to be hydrogen bonds as they are weak with energies similar to those of van der Waals complexes. Short contacts between F and the acidic hydrogens of HO or HN are rare in the CSDS with only 12 and 28 occurring respectively. There was only one contact below 2.0Å. *Ab initio* calculations have evaluated the relative stability and optimum distance of C-F···H-O bonds between water and fluoromethane and fluoroethene. It emerges that the $C(sp^3)$-F fluorine in fluoromethane can enter into stronger hydrogen bonds than $C(sp^2)$-F of fluoroethene. The X-ray data reinforces the conclusion that $C(sp^3)$-F fluorine is a better hydrogen bond acceptor than $C(sp^2)$-F fluorine. The $C(sp^3)$-F···H-O bond is less than half the strength (2.38 kcal mol^{-1}) of a C-O···H-O and the $C(sp^2)$-F···H-O bond (1.48 kcal mol^{-1}) is about half as weak again. Overall however short contacts in the Database which are consistent with an optimal F···H bond are extremely rare.

INTRODUCTION

It has become a common practise in bio-organic chemistry to replace a hydrogen atom or hydroxyl group for fluorine to generate a fluorinated enzyme substrate analogue, which may act as a substrate or inhibitor in a given enzymatic process[1-3]. The rationale for such a strategy is that the size of the fluorine atom is intermediate between that of hydrogen and oxygen. The van der Waals radii of fluorine (1.47Å) can be compared to that of hydrogen (1.2Å) or oxygen (1.57Å) and it emerges that fluorine has a close isosteric relationship to oxygen[4]. To be a successful hydroxyl mimic in bio-organic chemistry the fluorine atom must replace the hydrogen bond acceptor ability of the hydroxyl oxygen. Clearly fluorine cannot replace the hydrogen bonding donor role as it is devoid of the acidic hydrogen (Figure 1).

Figure 1 The OH can act as a hydrogen bonding donor or acceptor whereas fluorine can only act as an acceptor.

Theoretical calculations variously estimate[5] the strength of a F···H bond to be between 2 to 3.2 kcal mol^{-1}. This can be compared[6] to an O···H hydrogen bond which is typically between 5 - 10 kcal mol^{-1}. Consistent with this the electrostatic influence of fluorine is approximately half that of oxygen[7]. Thus the greater electronegativity and lower polarisability of fluorine over oxygen, suppresses its electrostatic influence and renders it a poorer hydrogen bond acceptor.

X-ray structural data offers an arena in which to assess fluorine as a hydrogen bonding acceptor. In a recent survey Shimoni and Glusker,[8] building on an earlier study,[9] of organo-fluorine compounds deposited in the Cambridge Structural Database System (CSDS), revealed relatively few situations where fluorine was involved in short contacts to acidic hydrogens (HO or HN). The authors concluded that the weakness of the F···H interaction results in it being overridden by acidic hydrogens finding O and N acceptors to pair with in preference to fluorine. This study was particularly wide ranging and embraced CF, CF_2 and CF_3 containing structures including non-bonded F···H interactions up to 3Å in length. The mean H···F distances in the study emerged between 2.5-2.6Å, which is close to the sum of the van der Waals radii of hydrogen and fluorine.[4] Contacts of this length will constitute weak interactions in energy terms. If fluorine is to replace the oxygen atom directly in a highly preorganised binding situation eg. in an enzyme-substrate complex, then ideally it is required to replace oxygen in O···H and make a short F···H contact of about 2.0 - 2.3Å. The present study therefore aimed to review the shorter F···H contacts equal to or smaller than 2.35Å and was restricted to C-F containing systems. CF_2 and CF_3 systems were ignored. This restriction was introduced as the focus of interest was to evaluate the hydrogen bonding acceptor ability of the fluorine atom in monofluorinated functional groups (ie. F for O). It was judged important not to make false comparisons with CF_2 and CF_3 systems where the hydrogen bonding ability of these fluorines may be perturbed.

A theoretical study was carried out in parallel with the Database analysis to evaluate the strength of the C-F···H-O and C-F···H-C interactions with distance. Both the Database survey and the theoretical analysis make the distinction between the hydrogen bonding acceptor ability of fluorine bound to both sp^2 and sp^3 hybridised carbon. To make a quantitative assessment of the relative strengths of $C(sp^3)$-F and $C(sp^2)$-F bound fluorine atoms, the change in interaction energies with bond length between (**a**) fluoromethane and water, and (**b**) fluoroethene and water, were studied. Also this and the previous studies[8,9] have revealed that the most common F···H contacts in the Database occur between fluorine and non acidic hydrogen atoms (ie. to HC rather than to HO or HN). These C-F···H-C interactions should be weak and to assess their stabilising influence both with $C(sp^3)$-F and $C(sp^2)$-F bound fluorine atoms, the interaction energies between (**c**) fluoromethane and methane and (**d**) fluoroethene and methane, were also evaluated.

Methods

Cambridge Structural Database Search

Version 5.10 of the Cambridge Structural Database System[10] (CSDS: October 1995) containing 146,272 entries was used for the study. Searches for bonded substructures and for inter- and intra- molecular non-bonded contacts were carried out using the program QUEST3D[11]. Subsequent statistical analyses were performed using VISTA[11]. The CCDB was searched for all C-F containing structures with an R factor lower than 0.075. All CF_2 and CF_3 containing compounds were deselected. The search was restricted to shorter ($\leq$ 2.35Å) H···F contacts and was subdivided on the basis of hybridisation at carbon ie. C(sp^3)-F and C(sp^2)-F and on the *intra* or *inter* nature of the F···H contacts. Hydrogen atom positions were normalised by extending the H···X bond along the X-ray derived bond vector to a neutron derived mean X-H bond length.[12]

Theoretical calculations

The geometries of the CH_2=CHF, CH_3F and H_2O molecules were optimised at the second order Möller-Plesset level utilising analytical gradients as implemented in the GAMESS program.[13] The basis set used was Dunning's TZV[14] supplemented by 3d and 1f polarization functions with Bearpark and Handy's"V" exponents[15] and also a diffuse sp shell for C, O and F. Hydrogen atoms had a single shell of p polarisation functions and a single diffuse s shell. Diffuse functions were assigned the literature exponents[16]. This basis we denote TZV++(3d1f,1p). The molecules were then paired up to generate the appropriate dimers. The geometry chosen was that which made C-F···H-X bond colinear. Basis set expansion to higher angular momentum functions is required to saturate the dispersion term within the MP2 formalism.[17] A more economical way of approaching this situation is the use of functions at mid-bond positions.[18] Thus a dimer basis set of the TZV++(3d1f,1p) type as above supplemented with a (1p1d1f) expansion midway between the F and H atoms.

In all cases, a potential energy scan was performed by varying the F···H distance while keeping the monomer geometries frozen. Relaxation of the monomer geometries will presumably lower the dimer energy further, however due to computational cost this was not feasable and therefore we may assume that the dimer energies may be slightly underestimated. The total dimer energies were calculated at both the HF and MP2 levels on a one dimensional grid at 0.1Å spacing. Once the lowest energy point was determined, two further single point energy calculations 0.05Å either side of this were also calculated to obtain the absolute minima.

When considering the energy stabilisation on dimerisation, suitable monomer energies must be subtracted from the full dimer energy. However, straight subtraction of the energies obtained from the isolated geometry optimisations may over-estimate the binding energy due to the well known basis set superposition error (BSSE) which results from effectively performing the monomer and dimer energies with different basis sets. Thus when calculating the monomer energies, full dimer basis sets at the equlibrium position were used following Boys and Bernardi[19]. These energies were then subtracted from the dimer energies along the potential energy surface which assumed the counterpoise correction to remain constant for all monomer separations.

RESULTS

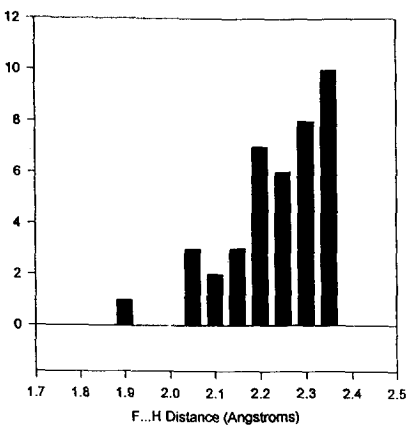
The results of the CSDS search are summarised Table 1 and in Figures 2-4.

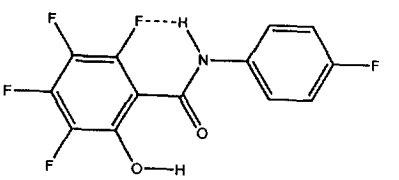
	TOTAL HITS		SHORT F···H CONTACTS			
			INTRA		INTER	
$C(sp^3)$-F	Compounds	C-F bonds	Compounds	Contacts	Compounds	Contacts
Total	177	237	28	29	20	22
O-H	71	89	5	5	1	1
N-H	69	85	8	8	3	3
C-H	177	237	15	16	18	18

			INTRA		INTER	
$C(sp^2)$-F	Compounds	C-F bonds	Compounds	Contacts	Compounds	Contacts
Total	371	926	48	65	45	50
O-H	89	177	3	3	3	3
N-H	113	214	7	9	8	8
C-H	360	929	41	53	35	39

Table 1 Summary statistics for the Total Hits and Short Contact searches showing numbers of compounds and C-F Bonds or contacts ≤ 2.35Å in each donor acceptor sub-set.

In Table 1 the search data is divided into short contacts between $C(sp^3)$-F and $C(sp^2)$-F and subdivided to distinguish *intra* and *inter* molecular F···H contacts. In the event 548 (177 + 371) C-F containing structures emerged with a total of 1163 (237 + 926) C-F bonds. Of all of the C-F bonds only 166 participated in non-bonded C-F···H-X contacts of 2.35Å or shorter. The majority of these were C-F···H-C contacts between fluorine and non acidic carbon bound hydrogen atoms as shown in Table 1. The weakness of these C-F···H-C is discussed later and it is more relevant to consider the stronger F···H contacts to more acidic hydrogens.





1

Figure 2 Histogram summarising the frequency and bond lengths of F···H-O/N contacts identified from the CSDS. There was only one instance (compound **1**, F···H = 1.86Å, Database Ref Code, YUYTOB[20]) of a contact shorter than 2.0Å.

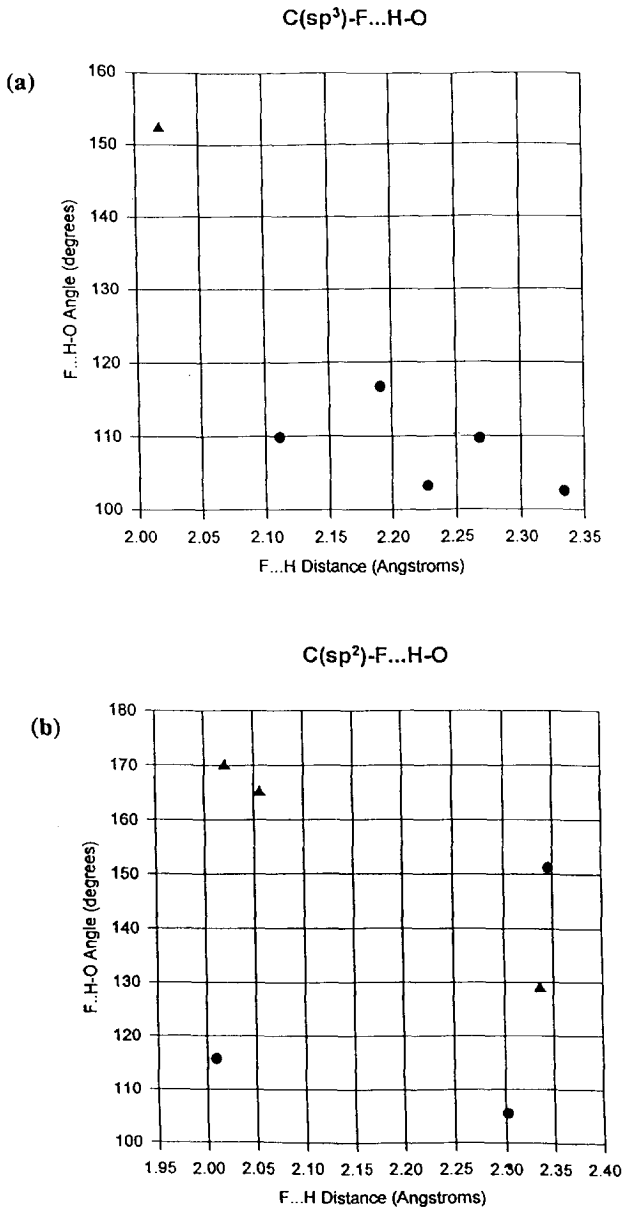


Figure 3 Scatter plots summarising the CSDS search showing the angles and lengths of *intra* (o) and *inter* (▲) molecular F···H-O contacts to (**a**) $C(sp^3)$-F bonded fluorine atoms and (**b**) to $C(sp^2)$-F fluorine bonded atoms.

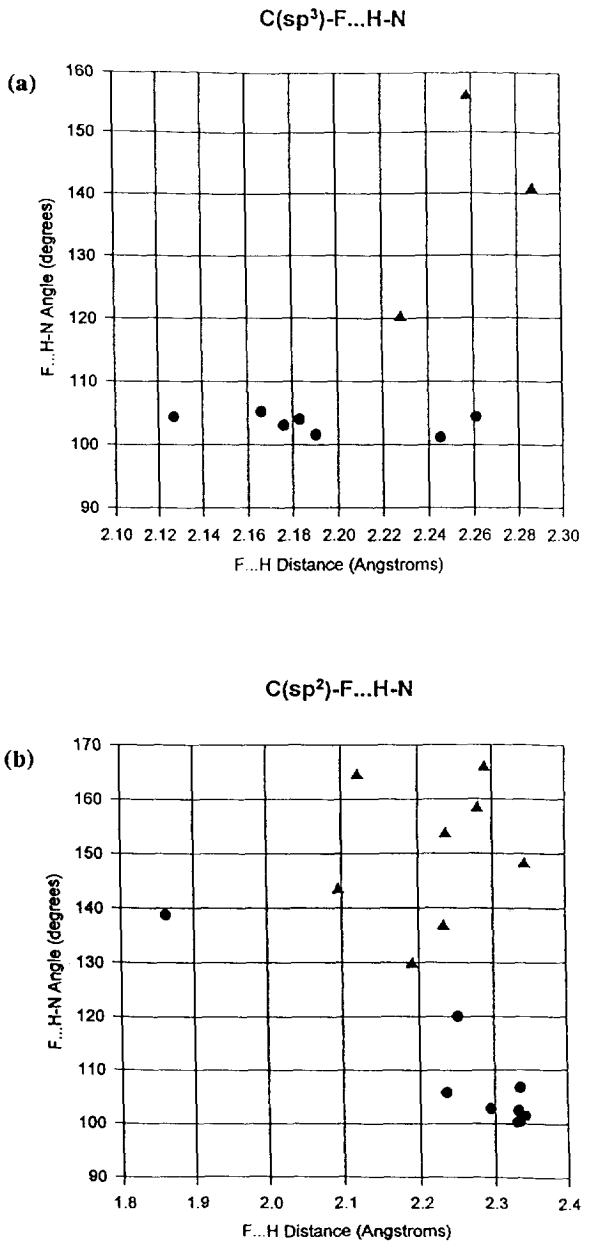


Figure 4 Scatter plots summarising the CSDS search showing the angles and lengths of *intra* (o) and *inter* (▲) molecular F···H-N contacts to (a) $C(sp^3)$-F bonded fluorine atoms and (b) to $C(sp^2)$-F fluorine bonded atoms.

Short C-F contacts to the acidic hydrogen atoms of HO (Figure 3) or HN (Figure 4) groups are surprisingly rare with only 12 and 28 occurring respectively. Only 7% of the total 565 C-F bonds in molecules also containing HO and HN groups are involved in these contacts. The F···HO and F···HN contact lengths are shown in Figures 3 and 4 and summarised in the histogram in Figure 2.

The shortest value in the Database is 1.86Å which is found in compound **1**.[20] It is clear from Figure 2 that there is no obvious clustering in bond lengths, but rather there is a steady increase in the bond length with the statistical increase in the number of contacts. In Figures 3 and 4 both the *intra* and *inter* molecular contacts to HO- and HN- groups are highlighted separately. The *intra* F···H-X angles are around 100-110° with few exceptions whereas the *inter* F···H-X angles are wider with no typical value. This lack of angular dependence with the *inter* molecular contacts suggests weak interactions, whereas the *intra* molecular contact angles cluster due to geometric constraints in forming a ring system.

The subdivision on the basis of hybridisation at carbon bound to fluorine is revealing. The $C(sp^3)$-F contacts to HO and HN hydrogens were statistically more frequent (9.8%, 17 from 174 C-F bonds) than the $C(sp^2)$-F contacts to hydrogen (5.88%, 23 from 391). Thus $C(sp^3)$-F bound fluorine atoms appear to enter into hydrogen bonding more frequently than $C(sp^2)$-F bound fluorine atoms. Anomalously the shortest contact in the Database found in compound **1** occurs to a $C(sp^2)$-F fluorine atom.

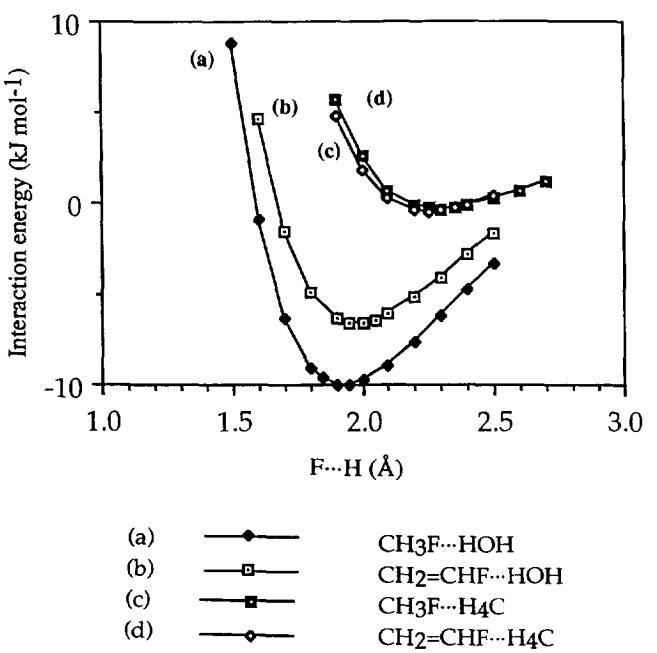


Figure 5 Plots derived from *ab initio* calculations of the interaction energy (kJ mol^{-1}) with F···H contact length for **(a)** fluoromethane and water **(b)** fluoroethene and water **(c)** fluoromethane and methane and **(d)** fluoroethene and methane.

The calculations of the interaction energy with bond distance of the four dimer interactions (**a**) - (**d**) are summarised in Figure 5. It is immediately apparent that the more stable interactions occur between C-F···H-O (**a** and **b** in Figure 5) rather than C-F···H-C (**c** and **d** in Figure 5). In the latter cases the equlibrium distance of ~2.2Å is extended and the energy minima is -0.85 kJ mol^{-1} (-0.2 kcal mol^{-1}), an energy consistent with that of a van der Waals complex rather than a hydrogen bond. On the other hand the complexes of water with fluoromethane and fluorothene do generate true hydrogen bonding dimers. That between water and fluoromethane (**a** in Figure 5) gave a minimum for the C(sp^3)-F···H-O interaction of -10 kJ mol^{-1} (-2.38 kcal/mol^{-1}) with an equilibrium distance of 1.9Å. This minimum value is consistent with other theoretical estimates[5] of the C-F···H-O bond. The C(sp^2)-F···H-O interaction between water and fluoroethene (**b** in Figure 5) has a similar equilibrium distance but is significantly weaker at 6.09 kJ mol^{-1} (1.48 kcal mol^{-1}).

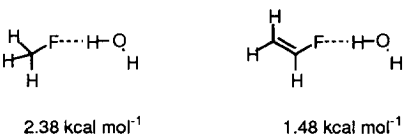


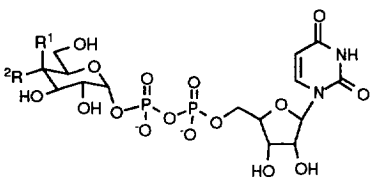
Figure 6 The F···H bond is stronger to an C(sp^3)-F rather than a C(sp^2)-F acceptor.

The increased donor ability of aliphatic over aryl bound fluorine atoms presumably arises as the fluorine lone pairs are in conjugation with the π-orbital system of the double bond and are less able to participate in H-bonding. Looking at the overall profiles in Figure 5 the data suggest that F···H-O and F···H-N contacts of 2.5Å and greater are very weak and come close in energy to van der Waals complexes, thus caution must be exercised in attributing a particular stabilising significance to interactions of this length and longer.

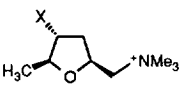
DISCUSSION

For all of the X-ray determined structures deposited in the CSDS which contain monofluorinated carbon atoms, there are very few instances where fluorine forms short contacts to acidic hydroxyl or amine protons. Thus fluorine is a poor hydrogen bond acceptor. Despite the dearth of F···H contacts to acidic protons there is a statistically significant increase in short contacts to C(sp^3)-F over C(sp^2)-F bound fluorine atoms. This observation implies that aliphatic fluorine atoms are better hydrogen bond acceptors than olefinic or aromatic fluorine atoms. This contention was reinforced by theoretical calculations which assessed the relative strengths of F···H bonds between (**a**) fluoromethane and water and (**b**) between fluoroethane and water. The former interaction emerged 0.8 kcal mol^{-1} more stable than the latter. A clear conclusion from this data is that vinyl (and aryl fluorines) are less effective than aliphatic fluorines as hydrogen bond acceptors. We conclude that enols and phenols will be poorly represented by their vinyl-fluorine and aryl-fluorine analogues respectively, when the oxygen atom acts as a hydrogen bond donor in interactions with proteins. For example it is anticipated that 4-fluorophenylalanine will be limited in this capacity when acting as a tyrosine mimic.
The predominant C-F···H-C contacts in the Database appear to have very little significance in energy terms and are essentially van der Waals complexes. The sum of many such contacts in a crystalline lattice[8], or in a polymer[5b] may add stabilisation to a macroscopic system and influence packing and properties, however for

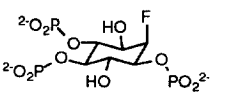
an enzyme-substrate binding interaction the energy of a single such contact will be insignificant. On the other hand F···H-O/N contacts, particularly to aliphatic fluorine atoms, are sufficiently stabilising in quantitative terms to influence binding energies. However in the context of fluorine replacing the hydroxyl oxygen as a hydrogen bond acceptor in substrate-protein interactions, then the structural data is not very supportive. The value of 10 kJ mol^{-1} (2.38 kcal mol^{-1}) for the optimal contact distance (1.9Å) between water and fluoromethane amounts to less than one half of the strength of a hydrogen bond between oxygen and an acidic proton (eg. O···HO) but it is striking that such optimal contacts are *extremely rare* in the Database. Encouraging for the bio-organic chemist is that substrate/protein interactions may offer an environment for optimal F···H bonding. In the pre-organised binding site of a receptor or enzyme, replacement of fluorine by hydroxyl in a given substrate should orientate the fluorine atom directly at the hydrogen bond donor, particularly if there are other peripheral stabilising interactions between the substrate and the protein. If such a situation is met then the F···H-X interaction may contribute to the overall binding energy, upto half of the strength of the original hydrogen bond to oxygen. However if the hydrogen bond donor can find stabilisation with an alternative acceptor to fluorine then it will do this and adversely influence the binding interaction.



R^1 = H, R^2 = F UDP-FGlc **2**
R^1 = F, R^2 = H UDP-FGal **3**



X = OH muscarine **4**
X = F **5**



6

The experimental evidence is mixed. Fluorodeoxy sugars often emerge as good substrate analogues for appropriate enzymes[21], and one recent study[22] provides convincing support for a F···H bonding controlling an enzymatic transformation. UDP-4-Deoxy-4-fluoroglucose (UDP-FGlc) **2** and UDP-4-deoxy-4-fluorogalactose (UDP-FGal) **3** were tested as substrates for UDP-D-glucose dehydrogenase, an enzyme which oxidises the C-6 hydroxyl group of UDP-D-glucose. UDP-FGlc **2** was an excellent substrate (Km of 30.2mM versus 9.6mM for natural substrate) for the enzyme whereas the diastereoisomeric C-4 epimer, UDP-FGal **3** was not a substrate but a competitive inhibitor (K_i = 20mM). Thus the configuration of the fluorine at C-4 appears crucial in securing the reactive conformation of the substrate on the enzyme surface for reaction at the remote C-6 centre. The correct stereoisomer **5** from various fluorodeoxy muscarine analogues was shown[23] to bind to the muscarinic receptors in heart tissue (guinea pig) by one order of magnitude *greater* than muscarine **4** itself in the heart receptors which control beat rate, and had a comparable effect with those receptors which control force. In another example[24] both D-*myo*-inositol-1,4,5-triphosphate and its fluorodeoxy analogue **6** were nearly equipotent [EC_{50} of **6** = 105nM, EC_{50} Ins(1,4,5)P_3 = 52nM] in their ability to mobilise sequestered Ca^{2+} ions, and this was judged consistent with the fluorine atom of **6** accepting a hydrogen bond from the receptor. These examples suggest that F is replacing OH, possibly in its role as a hydrogen bond acceptor, however such successful examples are few and more often than not[25,26,27] the substitution proves detrimental to the binding affinity or the kinetics of turnover.

References

1. Welch, J. T.; Eswarakrishnan, S. *Fluorine in bioorganic chemistry*, J. Wiley & Sons, New York, 1991; Welch, J. T.*Tetrahedron*, **1987**, *43*, 3123.
2. Mann, J.*Chem. Soc. Rev.*, **1987**, *16*, 318.
3. Seebach, D. *Angew Chem., Int. Ed. Engl.*, **1990**, *29*, 1320.
4. Bondi, A. *J. Phys. Chem.*, **1964**, *68*, 441.
5. (a) Smart, B. E. *"Characteristics of C-F Systems"* in *"Organofluorine Chemistry: Principals and Commercial Applications"* Ed Banks, R. E. Plenum Press, New York, 1994. (b) Dixon, D. A.; Smart, B. E. *Selective fluorination in organic and bioorganic chemistry*, ACS Symposium Series 456 (Welch, J. T. Ed), Washington D.C., 1991; (c) Dixon, D. A.; Smart, B. E. *J. Phys. Chem.*, **1991**, *95*, 1602.
6. Warshel, A.; Papazyan, A.; Kollman, P. A. *Science*, **1995**, *269*, 102; Fersht, A. *Enzyme structure and mechanism*, W. H. Freeman and Co, New York, 2nd Ed, p298, 1984.
7. O'Hagan, D.; Rzepa, H. S.*Chem. Commun.*, **1996**, in press.
8. Shimoni, L.; Glusker, J. P. *Structural Chem.*, **1994**, *5*, 383.
9. Murray-Rust, P.; Stalling, W. C.; Monti, C. T.; Preston, R. K.; Glusker, J. P. *J. Am. Chem. Soc.*, **1983**, *105*, 3206.
10. Allen, F. H.; Davies, J. E.; Galloy, J. J.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. *J. Chem. Inf. Comp. Sci.*, **1991**, *31*, 187.
11. Cambridge Structural Database System Users Manual (1994), Cambridge Crystallographic Data Centre, Cambridge, UK.
12. Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R. *J. Chem. Soc., Perkin Trans 2.*, **1987**, S1-S19.
13. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comp. Chem.*, **1993**, *14*, 1347.
14. Dunning, T. H. *J. Chem. Phys.*, **1971**, *55*, 716.
15. Bearpark, M. J.; Handy, N. C. *Theor. Chim. Acta.*, **1992**, *84*, 115.
16. Clark, T.; Chandresekhar, J.; Spitznagel, G. W.; Schleyer, P. von R. *J. Comp. Chem.*, **1983**, *4*, 294.
17. Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.*, **1994**, *94*, 1723.
18. Hobza, P.; Selze, H. L.; Schlag, E. W. *Chem. Rev.*, **1994**, *94*, 1767.
19. Boys, S. F.; Bernardi, F. *Mol. Phys.*, **1970**, *19*, 553.
20. Banks, R. E.; DuBoisson, R. A.; Pritchard, R. G.; Tipping, A. E. *Acta Crystallogr.*, **1995**, *C51*, 1427.
21. Phelps, M. E.; Hoffman, E. J.; Sterlin, C.; Huang, S. C.; Robinson, G.; McDonald, N.;. Schelbert, H.; Kuhl, D. E. *J. Nuc. Med.*, **1978**, *19*, 1311.
22. Chapeau, M-C.; Frey, P. A. *J. Org. Chem.*, **1994**, *59*, 6994.
23. Bravo, P.; Resnati, G.; Angeli, P.; Frigerio, M.; Viani, F.; Arone, A.; Marucci, G.; Cantalamessa, F. *J. Med. Chem.*, **1992**, *35*, 3102.
24. Lampe, D., Liu, C., Mahon, M. F., Potter, B. V. L. *J. Chem. Soc., Perkin Trans 1.*, 1996, 1717.
25. McCarter, J. D.; Adams, M. J.; Withers, S. G. *Biochem. J.*, **1992**, *286*, 721.
26. White, A.; Tull, D.; Johns, K.; Withers, S. G.; Rose, D. R. *Nature Struct. Biol.*, **1996**, *3*, 149.
27. Guédat, P.; Poitras, M.; Spiess, B.; Guillemette, G.; Schlewer, G. *Biorg. Med. Chem. Letts.*, **1996**, *6*, 1175.

(*Received in UK* 5 *July* 1996; *revised* 15 *July* 1996; *accepted* 15 *August* 1996)