

Hydrogen Bonding Interactions of Covalently Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field

Emanuele Carosati, Simone Sciabola, and Gabriele Cruciani*

Laboratory for Chemometrics and Cheminformatics, Department of Chemistry, University of Perugia, Via Elce di Sotto, 10, I-06123, Perugia, Italy

Received February 27, 2004

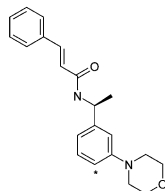
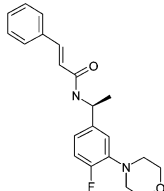
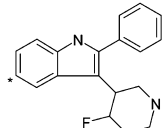
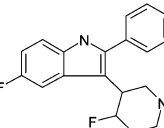
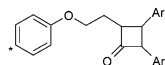
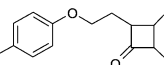
Through the years the GRID force field has been tuned to fit experimental observations in crystal structures. This paper describes the determination of the hydrogen bonding pattern for organic fluorines based on an exhaustive inspection of the Protein Data Bank. All the PDB complexes, whose protein structures have cocrystallized fluorine-containing ligands, were examined and geometrically inspected. By applying statistics, the hydrogen bonding geometry was described as a distribution function of the angle at the fluorine: a new specific angular function was consequently defined and inserted in the program GRID to estimate the effect of fluorine hydrogen bonds on the ligand–protein binding. All the fluorine-containing ligands collected from the PDB were docked within their corresponding protein binding sites: introducing the fluorine hydrogen bonding contribution improves the results of the docking experiments in terms of accuracy and ranking.

Introduction

Fluorine atoms are often present in drugs and drug-like molecules as proven by the relatively frequent occurrence (14.4%) in the refined MDDR database of fluorine-containing molecules.^{1,2} In fact, fluorine atoms are often introduced in drug skeletons to modify pharmacokinetic properties, such as oral absorption,³ and to occupy key positions⁴ where they will modulate metabolic reactions,⁵ blocking metabolic routes of oxidation^{6,7} as shown by the examples presented in Table 1. Thus, knowledge of the nonbonding behavior of covalently bonded fluorine is necessary to approach any kind of molecular modeling problem correctly, whether it is molecular recognition or pharmacokinetics and metabolism.

Evidence regarding fluorine hydrogen bonding has largely been published from the 1960s,⁸ and in the recent decades the hydrogen bonding acceptor capability of all the halogen atoms has often been investigated, in both their ionic and bonded forms. It is commonly accepted that fluorine is a stronger acceptor than the other halogens⁸ but is not as strong as oxygens and nitrogens,^{9–11} whereas in their ionic^{12–15} and metal-bonded¹⁶ forms all the halogens act as considerable proton acceptors. The acceptor capability of halogenated compounds was quantitatively measured by Laurence and Berthelot,¹⁷ who used FT-IR technique to produce a broad scale (more than one thousand values) of hydrogen bonding acceptor capability. In this scale fluorinated compounds exhibit substantially different values from those of their heavier halogen counterparts, as the latter are very weak acceptors comparable to π bases.^{17,18} The survey by Howard and co-workers¹⁹ of the X-ray data stored in the Cambridge Structural Database (CSD)²⁰ revealed short contacts of fluorine atoms to acidic hydrogens and was reinforced by theo-

Table 1. Reports of How Introducing a Fluorine Atom on the Molecular Skeleton of Different Drugs Overcomes a Metabolic Reaction, as Benzylic Hydroxylation

Drug classification	Site of metabolism (*)	Modified drug	Ref
Potassium channel opener			4
h5-HT _{2A} receptor antagonist			6
Cholesterol absorption inhibitor			7

retical calculations from which half of the binding energy of a hydrogen bond to oxygen was assigned to the fluorine. Furthermore, the large scale analysis conducted by Shimoni and Glusker²¹ in the CSD revealed that acidic hydrogens prefer to bind to the stronger acceptors such as oxygens and nitrogens compared to fluorine atoms. Brammer and co-workers¹¹ recently examined the CSD and confirmed the weakness of fluorine hydrogen bonding interactions, revealing shorter fluorine...donor contacts for C(sp³)-F structures compared to C(sp²)-F, as previously observed by Howard and co-workers.¹⁹ The geometry of hydrogen bonds involving fluorines was only investigated on small-molecule contacts in the CSD²² although the amount of

* Corresponding author. Phone: +39 075 5855550, Fax: +39 075 45646, e-mail: gabri@chemiome.chm.unipg.it.

deposited crystal structures on the Protein Data Bank archive²³ has increased significantly through the years (1000% of growth in the last 10 years).²⁴ The first and only attempt to define the fluorine hydrogen bonding behavior within the PDB seems to have been carried out 10 years ago by Dunitz and Taylor.²² Fourteen protein–ligand complexes were identified as geometrically possible hydrogen bonds; the contacts were investigated in detail by measuring the distance $F\cdots D$ and the angle at the donor, but without defining the angle at the acceptor (fluorine).

The weakness of fluorine hydrogen bonding interactions may be responsible for the lack of a detailed definition of the hydrogen bonding geometry, which in the opinion of the authors may only be found by statistical analysis of the angle made at the fluorine. In addition, the specific behavior of geminal fluorines has neither been investigated over the PDB nor over the CSD, and they were even excluded from the analysis whenever comparison between $C(sp^3)-F$ and $C(sp^2)-F$ structures was performed.^{11,19} Thus, to obtain an accurate definition of the behavior of the fluorine as hydrogen bonding acceptor and investigate its geometric pattern, an exhaustive and up-to-date investigation of the PDB archive was considered to be indispensable as this is actually the depository for more than 23 000 crystal structures. Consequently, a PDB search was carried out to extract and analyze those protein–ligand complexes with fluorines as acceptors and protein NH and OH groups as hydrogen donors. To avoid misleading contacts, the chemical environment of each protein–ligand complex was analyzed in detail and the resulting set was then further reduced, leaving few (but certain) interactions. Thus, statistical analysis was applied to these in order to design a new angular function which will describe the fluorine hydrogen bonding geometry on the GRID force field.^{25,26} It will be shown that the developed function provides a very good prediction of the fluorine hydrogen bonding properties, with a significant improvement in agreement between experimental and computational data.

Computational Methods

PDB Searches. The chemical behavior of fluorine atoms present in many drugs and drug-like molecules depends on their chemical surroundings: hydrogen bonding ability is likely to be affected by the presence of other geminal fluorines, as well as by the aromatic/aliphatic character of the molecular skeleton. Therefore, any kind of covalently bonded fluorine-containing molecules were sought within the PDB, and ligands were subdivided into two main groups according to the sp^2 or sp^3 hybridized form of the fluorine-bonded carbon. Consequently, the resulting fluorines referred to in this paper were classified as “aromatic” and “aliphatic”, respectively. Those bonded to olefin skeletons were included in the broader aromatic class, whereas those of CF_2 and CF_3 groups, which are characterized by two and three geminal fluorines respectively, were considered separately from the aliphatic class.

The geometry of more than 23 000 of the ligand–protein complexes extracted from the PDB (up to November 2003) was analyzed. The search was restricted to those structures with crystallographic *R*-factor under 50% and

resolution up to 3.0 Å. Only the complexes respecting geometrical constraints were stored, verified by using the coordinates of both the carbon–fluorine system from the ligand, and the potential hydrogen bond donors (N and O) from the protein, here briefly coded as D (Donor). The geometrical constraints will be described in detail later on in the manuscript. The processing of the PDB archive was carried out using a combination of PERL scripts which were programmed to perform the following operations:

(1) Fluorine atoms were searched within all the PDB entries, and only those entries in which the cocrystallized ligands contained the fluorine atom in its covalently bonded form were retained.

(2) The existence of any kind of oxygen or nitrogen within a fluorine-centered sphere (whose radius was defined by the distance constraints described in more detail later in this section) was then verified. Only those atoms which are likely to be hydrogen bonding donors were considered, thus excluding the oxygen atoms of peptide bonds (coded as O in PDB files) and retaining nitrogen and oxygen atoms coded as N, ND2, NE, NE1, NE2, NH1, NH2, NZ, OD1, OD2, OE1, OE2, OG, OG1, and OH.

(3) The geometry of the interaction was investigated in detail. Angles at acceptor atoms were calculated and compared with the angular constraints illustrated in Figure 1, as discussed below.

PDB Selection. The first step of the described procedure was applied to the complexes of the PDB that respect the constraints of resolution (3.0 Å) and *R*-factor (50%). The longer accepted distance between donor and acceptor (fluorine), which is referred to as distance constraint in the second point listed above, corresponds to the straight angle $D-H\cdots F$. Thus, the distance $D\cdots F$ is calculated as the sum of the distance $F\cdots H$ and the bond distance $D-H$. As suggested by several authors,^{27,28} the van der Waals radii sum (F, 1.47 Å; H, 1.20 Å) was assigned to the distance $F\cdots H$, whereas for both nitrogen and oxygen donors the value of 1.0 Å was assigned to the bond distance $D-H$, according to the GRID force field.²⁵ Finally, a tolerance of 15% was added to avoid loss of useful information in the second phase, and the value of the distance constraint was 4.0 Å. After that, in the third and final step of the automated refinement procedure, the contacts whose angles $C-F\cdots D$ lay between 90° and 270° were accepted as possible hydrogen bonds.^{28,29} In fact, bent angles may correspond to unstable situations due to the repulsion between the donor and the molecular skeleton, and consequently they were excluded. Nevertheless, the values of the angles $C-F\cdots D$ within the broader interval [60°; 300°] for geminal fluorines were also retained in order to avoid the exclusion of three-centered interactions, as illustrated in Figure 1.

All the crystal structures respecting these geometrical constraints were then analyzed in detail to avoid misinterpreting forced close contacts as hydrogen bonds. Thus, hydrogens of donor groups were placed where hydrogen bonding was most likely to occur. Essential hydrogens were added to the functional groups as follows. First, hydrogens of peptide bonds were placed according to the residue coordinates, as well as nitrogen-bonded hydrogens of side chain for asparagine (Asn),

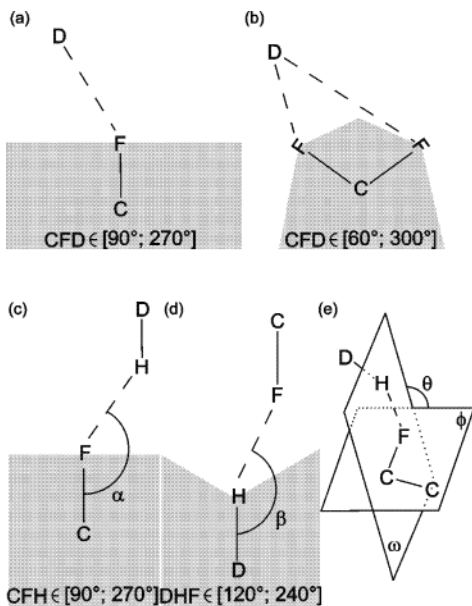


Figure 1. Constraints regarding the angle $C-F\cdots D$ were introduced to avoid the selection of PDB entries where steric hindrance would make the donor–ligand contact energetically disfavored. Inaccessible regions to the donor are represented as gray areas. In the case of aromatic and aliphatic fluorines (a) the angular range is between 90° and 270° . The angle is wider for geminal fluorines (b) to ensure three-centered interactions, which are generally characterized by more bent angles $C-F\cdots D$, to be included. In addition, two angles (α and β) are defined by each acceptor–donor system. (c) Fixing the acceptor, the hydrogen is free to move everywhere around it, with the exception of the inaccessible gray area. (d) Fixing the donor, the fluorine is free to move within a restricted area, due to the collinearity of the donor–hydrogen...acceptor system. (e) The torsional angle θ is made by the planes ϕ and ω , in which the systems $C-C-F$ and $H\cdots F-C$ lie, respectively. The angle θ may rotate over the whole range $[0^\circ; 360^\circ]$, not being involved in any constraint.

glutamine (Gln), and tryptophan (Trp). Similarly, five hydrogens were added to the guanidine group of the arginine (Arg), considered in its charged form. Furthermore, the lysine (Lys) was positively charged on the lateral chain, and the $C-N$ bond was rotated by aligning one hydrogen toward the nearest fluorine atom. The uncharged form was assigned to all the remaining residues, and hydrogens were carefully added to tautomeric residues if necessary. The two imidazole nitrogens (ND1 and NE2) could be protonated in the case of histidine (His), but only the nitrogen nearest to the fluorine of the ligand was considered to be the donor. Similarly, for the aspartic acid (Asp) and glutamic acid (Glu) the oxygens nearest to the acceptor were protonated: each added hydrogen atom was orientated toward the nearest fluorine by rotating the torsional angle of the corresponding $C-O$ bonds, keeping the angle $C-O-H$ constant. The hydroxyl hydrogens added to serine (Ser) and threonine (Thr) were orientated using the same procedure, whereas only the two torsional orientations coplanar to the aromatic ring were considered for the phenoxyl group of tyrosine (Tyr). Whenever the acceptor was a CF_2 or CF_3 group rather than aliphatic or aromatic fluorine, the aforementioned rotatable bonds $C-C-O-H$ (for any of the Ser, Thr, Asp, and Glu residues) and $C-C-N-H$ (for the Lys residues) were orientated toward the central carbon.

Essential hydrogens were also added wherever hydrogen bonds support the protein tertiary structure of binding sites.

The addition of hydrogens allowed the effective measurement by means of the SYBYL program³⁰ of several geometric properties such as distances $F\cdots H$, angles $C-F\cdots H$ (α) and $D-H\cdots F$ (β), and torsional angles $C-C-F\cdots H$ (θ), shown in Figure 1. The angle α was only acceptable within the interval $[90^\circ; 270^\circ]$, whereas for β the narrower interval $[120^\circ; 240^\circ]$ was considered due to the necessary collinearity of donor–hydrogen...acceptor complex. To facilitate any kind of statistical approach, the symmetry of α and β with respect to the corresponding axes $C-F$ and $D-H$ was considered. Whenever any value was larger than 180° , a physical reflection with respect to the corresponding symmetry axis was executed. Thus, the corresponding values obtained by the reflection substituted all the α and β values, whose resulting sets were restricted to the intervals $[90^\circ; 180^\circ]$ and $[120^\circ; 180^\circ]$, respectively. The torsional angle $C-C-F\cdots H$ (θ) was made at the $C-F$ axis by the planes $C-C-F$ (ϕ) and $C-F\cdots H$ (ω). Again, all of the $C-C-F\cdots H$ values were reported to the intervals $[0^\circ; 90^\circ]$ and $[0^\circ; 120^\circ]$ for those angles involving aromatic and aliphatic fluorines, respectively.

The GRID Program. The GRID procedure was developed for determining energetically favorable binding sites for small chemical groups, called probes, on macromolecular target(s).²⁵ The probe group is moved through a regular grid of points around the target molecule and the interaction energy is calculated by an empirical energy function (eq 1), where the index j ranges over all the target atoms.

$$E = \sum_j E_{lj} + \sum_j E_{el} + \sum_j E_{hb} + S \quad (1)$$

The energy function contains Lennard–Jones (E_{lj}), electrostatic (E_{el}), and hydrogen bonding (E_{hb}) terms, as well as an entropic contribution (S) in some special cases. The Lennard–Jones and electrostatic terms have been widely described elsewhere.²⁵ The hydrogen bonding term often provides a relevant contribution to the probe–target interaction and defines specificity for such interactions. It is the product of three terms (eq 2) and depends on the chemical nature of the hydrogen bonding atoms as well as on the length and orientation of the hydrogen bond.

$$E_{hb} = E_r \times E_t \times E_p \quad (2)$$

The term E_r is an 8-6 function depending on the optimum energy and length of the hydrogen bonding donor–acceptor pair, and on their distance r . The orientational dependence of the hydrogen bond is described by E_t and E_p , which are functions of the angles made by the hydrogen bond at the target and probe atoms, respectively. Their values lie between zero and one according to t and p angles illustrated in Figure 2. Whenever the geometry of the interaction is not optimally orientated at the target or at the probe, the values of the corresponding term E_t or E_p is lower than one, and the main term E_{hb} is consequently lowered.³¹ The E_p equation depends on the probe geometry and appropriate parameters are used to define each probe. A

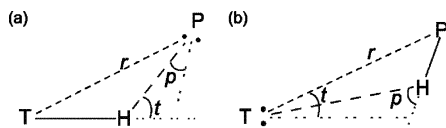


Figure 2. Generic target-probe interaction with the hydrogen held either by the target (a) or by the probe (b). The probe is not optimally orientated toward the target because of the stronger second hydrogen bond (not shown here) characterized by its value of ρ being equal to zero.

single-atom probe could be thought as a small spherical center with hydrogen atoms and lone-pair electrons condensed in it. Whenever a donor probe is able to donate two or more hydrogen bonds its orientation depends on their relative strengths: the strongest hydrogen bond occurs with linear geometry and $E_p = 1.0$, whereas the second and further additional hydrogen atoms could lie out from the corresponding target lone-pair directions, giving rise to E_p contributions differing from unity (Figure 2). The same happens at acceptor probes, with two or more lone-pairs available, or at probes with both donor and acceptor features. The probe orientates itself whenever it accepts or holds one only hydrogen bond: this is optimally aligned, the angle ρ is equal to zero, and E_p is consequently equal to one.

The angular functions E_t and E_p were chosen by fitting to experimental data on hydrogen bonding geometry in crystal structures of small organic molecules and proteins, observed by X-ray or neutron diffraction.^{31–33} The modeling of angular functions required the large amount of data presented in ref 31: the observations were grouped at 10° angular intervals and the probability distribution was calculated. The resulting probability values were then converted into relative energies assuming a Boltzmann distribution (eq 3), where P_i is the probability of formation of a hydrogen bond of energy E_i and P_0 is the probability of formation of an optimally orientated hydrogen bond of energy E_0 . The resulting values of energy were plotted against t angles. Finally, suitable analytical functions, continuous and simple to compute, were chosen to fit them.

$$E_i = E_0 - RT \ln(P_i/P_0) \quad (3)$$

In the present study, the authors used the same approach for treating the fluorine hydrogen bonds observed within the PDB in order to model an analytical function for such a class of atoms. Actually, in the previous versions of the program GRID the weak hydrogen bonding acceptor capability was approximated by a preferred frontal geometry, due to a lack of experimental information on crystal structures.

Results and Discussion

Crystallographic Approach. The PDB database, containing 23 549 deposited structures (November 2003) was downloaded and exhaustively analyzed: 275 different entries whose ligands present at least one covalently bonded fluorine were found. Overall, 998 fluorine atoms were present in the database, distributed between 359 aromatic or aliphatic C–F groups, 123 CF₂ groups, and 131 CF₃ groups. This dataset was then reduced according to the geometrical constraints described in the Methods section to only fluorine atoms

Table 2. van der Waals Radius, Mean Distance, and Standard Deviation for Each Class of Fluorine Atoms

fluorine	hydrogen bonding contacts	mean distance F...H (Å)	standard deviation (Å)
aliphatic	20 ^a	2.113	0.164
aromatic	25	2.698	0.176
geminal	60 ^b	2.350	0.285
overall	105	2.313	0.229

^a Two aliphatic fluorine atoms are involved in double interactions, all the other 18 in single interactions. ^b Forty four different donors generate 60 contacts with 60 different geminal fluorine atoms, which are involved in 28 single and 16 bifurcated hydrogen bonds.

interacting with donor groups of the surrounding protein environments. Then, further refinement was performed and only “certain” hydrogen bonds were selected. In fact, the agreement with geometrical constraints was considered a necessary but not sufficient condition for fluorine–donor contact to be caused by hydrogen bond. Thus, complexes were excluded whenever either a factor disfavoring the hydrogen bond formation such as steric hindrance or other stronger interaction responsible for the ligand orientation within the binding site was observed. At the end of the refinement process 105 contacts were observed and 49 of the 275 analyzed complexes were retained. Thus, a significant part (10%) of the overall amount of fluorine atoms from the PDB are involved in hydrogen bonding contacts, whereas there is an 18% of chance that a generic PDB entry with a cocrystallized fluorine-containing ligand presents a hydrogen bond occurring at the fluorine.

The hydrogen bonding interactions can be classified according to the mode of interaction as single, double, triple, or bifurcated. Single interactions involve only one fluorine atom as hydrogen bonding acceptor and one donor group from the protein, whereas double interactions involve two donor groups interacting with the same fluorine. Triple interactions, involving three donor groups and the same fluorine, were never observed. Last, bifurcated interactions involve the same donor group directed between two fluorines, which are usually geminal. Eighteen fluorine atoms of aliphatic ligands share single hydrogen bonds, whereas only twice the same fluorine atom is involved in a double interaction. Single hydrogen bonds involve 25 fluorine atoms of the aromatic ligands, and last, 28 single and 16 bifurcated interactions implicate 60 geminal fluorine atoms. The contacts collected are presented in Table 2, and a detailed summary is provided in the Supporting Information section (Tables S1–S3).

Most of the PDB entries selected were released in the last 10 years, and so this set is wider than the set selected by Dunitz and Taylor.²² Six out of fourteen entries are still present in our set (namely **1elb**, **1elc**, **2est**, **4est**, **6gch**, and **7gch**) whereas five other complexes (**1bcd**, **1hld**, **1rds**, **4gpb**, **7est**) do not respect the geometrical constraints of this study regarding angles and distances. Moreover, an oxygen atom clearly competes with the fluorine as acceptor in three other cases (**1apv**, **1apw**, **1ela**). Two of the mentioned PDB entries, **1apw** and **4est**, are shown in Figure 3 as examples of the discussed refinement. In Figure 3a the “apparent interaction” within the PDB entry **1apw** is presented: the oxygen OD2 of Asp213 respects the

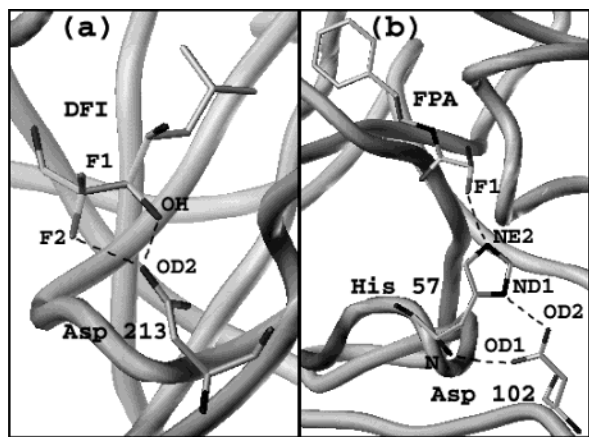


Figure 3. The PDB entries **1apw** and **4est** are here presented for illustrative purposes, and only one (**4est**) was included in the working set. (a) The hydrogen atom from the OD2 atom of Asp213 might bind two acceptor atoms, namely F2 and OH, but the first is geometrically favored. (b) Two hydrogen bonds linking Asp102 and His57 are evident from the close proximity of both the nitrogen ND1 of the imidazole ring and the backbone nitrogen N of His57 to the oxygens OD2 and OD1 of Asp102, respectively. In addition, the imidazole ring is likely to be protonated on its NE2 nitrogen, giving rise to the hydrogen bond with the fluorine F1 of the ligand.

mentioned geometrical constraints. In fact, the hydrogen from OD2 could point toward one fluorine atom of the ligand DFI: acceptor (F2) and donor (OD2) could be 3.0 Å apart and the angle O–H···F could be equal to 158°. Furthermore, the oxygen (OH) of the ligand could also accept the hydrogen held by OD2,³⁴ with even better donor-hydrogen···acceptor alignment (angle OD2–H···OH equal to 169°) and shorter donor–acceptor distance (2.7 Å). It is highly probable that the interaction involving the fluorine is energetically disfavored, and the complex was therefore excluded from the set. Figure 3b shows the hydrogen bonding linkage formed within the PDB entry **4est**, in which two hydrogen bonds link the residues His57 and Asp102. In addition, His57 is able to donate the acidic hydrogen in its imidazole ring, held by the nitrogen NE2, to the fluorine F1 of the ligand. Since the study was performed at pH 5.0,³⁵ the geometry of the system led to the supposition that both hydrogen bonds exist at the imidazole ring, and so the PDB entry **4est** was added to the selected set.

Statistic Survey. The collected PDB entries within each of the aliphatic, aromatic, and geminal classes were grouped and progressively numbered according to the release date in order to assign a unique code to each hydrogen bonding geometry. According to the mode of ligand-protein binding, 14 groups were distinguished within the aliphatic class, 6 groups within the aromatic class, and 19 groups within the geminal class, as listed in Table 3. Each group contains at least one PDB entry which might repeatedly bind the ligand in the same manner by means of several binding sites from one or more chains. Moreover, different PDB entries may belong to the same group whenever their binding modes are comparable and the hydrogen bonds occurring at fluorines involve the same kind of donor atoms.

Since the objective of this study was the classification of hydrogen bonding features of aliphatic, aromatic, and geminal fluorine atoms, only accurate use of statistics

would allow, first, any comparison between different classes, and second, the whole set to be considered in general. Thus, with the purpose of considering all the observed hydrogen bonding geometries equally, appropriate weights were assigned according to the population of each group and to the presence of similar binding sites. For a generic group containing w observed binding sites the contribution to the statistics of any observed binding geometry was weighted as $1/w$. For example, in the case of group R6 the weight was 1/8 (i.e. 0.125) both because of the similarity between the PDB entries **1o28** and **1o29**, and because each crystal structure is composed of four chains (A, B, C, and D). In addition, to equalize the importance of single and bifurcated interactions, the number of fluorine atoms sharing the same hydrogen donor was multiplied by w when calculating the final weights. For example, in the case of the PDB entry **1a08** (group G11), for two chains (A and B), a bifurcated interaction involves two fluorine atoms and the OG donor atom of Ser180: the resulting weight for each contact was 1/4 (i.e. 0.250). This treatment was applied to the calculation of the mean distances for the aromatic, aliphatic, and geminal classes as well as to the distribution count of the angle C–F···H discussed above.

The resulting values of the mean distances F···H calculated for aliphatic, aromatic, and geminal fluorines, were 2.113 Å, 2.698 Å, and 2.350 Å, respectively. These mean values led to the postulation that there is a significant greater number of short contacts in aliphatic systems compared to olefin/aromatic systems, which agrees with the statistical considerations carried out by Howard and co-workers.¹⁹ The increased acceptor ability of aliphatic over aryl-bonded fluorine atoms presumably arises because the fluorine lone-pairs are in conjugation with the π -orbital system and are less able to participate in hydrogen bonding. Nevertheless, introducing the standard deviation values obtained from the distribution of aliphatic, aromatic, and geminal classes, which are respectively 0.164 Å, 0.176 Å, and 0.285 Å, both the difference between geminal and aromatic classes and the difference between geminal and aliphatic classes were not statistically significant. Therefore, the unique value of mean distance 2.313 Å for the generic F···H interaction was calculated, as reported in Table 2. The scattering of the torsional angle C–C–F···H (θ) revealed no observed angular preference related both to aliphatic/aromatic/geminal atoms and to single/bifurcated interactions (Figure 4a). Consequently, a circular symmetry around the C–F axis was assumed. From the distribution of the angle C–F···H little difference was observed on the angular preferences (Figure 4b): hydrogens principally approach any aromatic fluorine laterally [90°; 120°], whereas aliphatic interactions are preferentially straight [120°; 180°], especially if single. On the contrary, the geometry of two hydrogens approaching the same fluorine is characterized by bent angles: it was generally observed for CF₂ and CF₃ groups, and it was generalized as being typical for three-centered interactions. Thus, the ability to accept two hydrogen bonds when bent geometry occurs is mainly an aliphatic and geminal peculiarity. The lateral approach is also favorite by aromatic fluorines, whereas the remaining two-centered interactions generally occur with straight

Table 3. PDB Entries with Geometrically Similar Binding Sites Belong to the Same Group, Progressively Named According to the Deposition Date with the Prefixes L, R, and G for Aliphatic, Aromatic, and Geminal Classes, Respectively

aliphatic fluorine			aromatic fluorine			geminal fluorine		
PDB code (chains)	deposition date	group code	PDB code (chains)	deposition date	group code	PDB code (chains)	deposition date	group code
1exp ^a	11-JAN-96	L1	1rds ^a	19-AUG-94	R1	2est (E)	24-MAR-86	G1
2xyl ^a	20-NOV-97		1aid (A,B) ^b	16-APR-97	R2	4est (E)	15-MAY-89	G2
1tsn ^a	02-DEC-96	L2	2aid (A)	17-APR-97	R3	6gch ^a	06-APR-90	G3
4a3h (A)	22-JUL-98	L3	1g4j (A)	27-OCT-00	R4	7gch ^a	06-APR-90	G4
6a3h (A)	22-JUL-98	L4	1hwi (A,B,C,D)	09-JAN-01	R5	1elb (A)	07-DEC-93	G5
5a3h (A)	23-JUL-98		1hwj (A,B,C,D)	09-JAN-01		1elc (A)	07-DEC-93	
1qi2 (A)	02-JUN-99		1hwk (A,B,C,D)	09-JAN-01		1dif (A+B) ^b	09-OCT-95	G6
1h11 (A)	01-JUL-02		1o28 (A,B,C,D)	18-FEB-03	R6	1cx2 (A,B,C,D)	17-DEC-96	G7
1bvv ^a	18-SEP-98	L5	1o29 (A,B,C,D)	18-FEB-03		6cox (A,B)	18-DEC-96	G8
1c5i (A)	24-NOV-99					1ah3 ^a	12-APR-97	G9
2nlr (A)	02-NOV-98	L6				6csc (B)	19-JUN-97	G10
1b99 (A)	22-FEB-99	L7				1a08 (A,B)	09-DEC-97	G11
1e0v (A)	10-APR-00	L8				1bwf (O,Y)	22-SEP-98	G12
1e0x (A,B)	10-APR-00	L9				1dy8 (B)	18-JAN-00	G13
1e73 (M)	16-AUG-00	L10				1e0y (B)	11-APR-00	G14
1e70 (M)	23-AUG-00	L11				1e7c (A) ^c	26-AUG-00	G15
1ga8 (A)	29-NOV-00	L12				1e8g (A,B)	20-SEP-00	G16
1iew (A)	11-APR-01	L13				1g4p (A,B)	27-OCT-00	G17
1qx1 (A)	04-SEP-03	L14				1g6c (A,B)	03-NOV-00	
						1mmj (N)	04-SEP-02	G18
						1qz0 (A,B)	15-SEP-03	G19

^a Any letter is not used in the PDB file for recognizing the chain. ^b The chains A and B define the binding site together. ^c Seven molecules of the ligand HLT are cocrystallized within the protein.

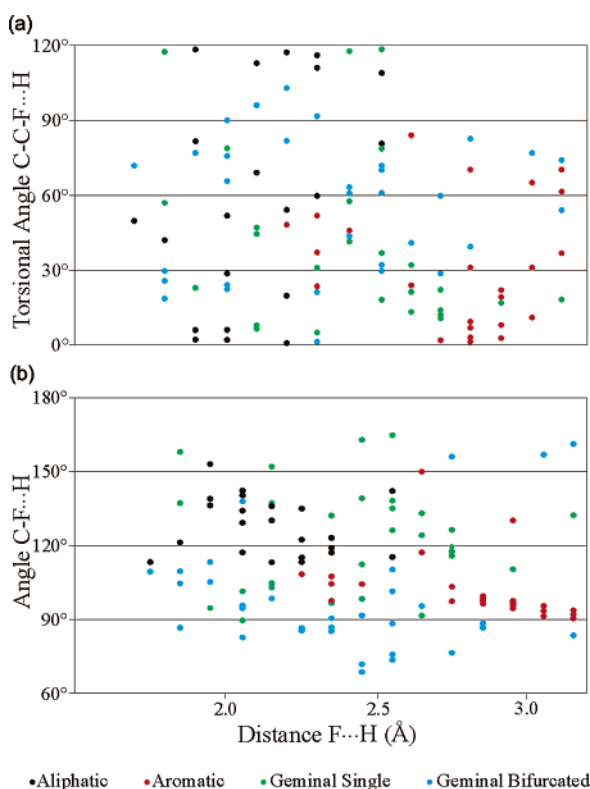


Figure 4. (a) The torsional angles C–C–F...H for hydrogen bonds involving aliphatic and geminal fluorines are within the range [0°; 120°] due to the trigonal symmetry of the fluorine hybridization form. On the contrary, aromatic fluorines, which have symmetry with respect to the aryl plane and to the plane perpendicular to it, lie within the narrower range [0°; 90°]. (b) The angles C–F...H for hydrogen bonds involving aliphatic and aromatic fluorines are within the range [90°; 180°]. Geminal fluorines lie within the wider range [60°; 180°] because three-centered interactions make the angle C–F...H more bent.

geometry. The resulting dispersion of fluorines over the whole range [60°; 180°] of C–F...H angles led to its

hydrogen bonding interactions being defined as not extremely directional.

The Probability for Hydrogen Bond Formation.

The features revealed for each class of fluorine atoms from the analysis of crystallographic data prompted to the definition of a unique new angular function E_b , able to describe the hydrogen bonding capability of aromatic and aliphatic as well as geminal fluorine atoms.

It was derived from the statistics applied to the whole set, which was treated as described below to ensure that any geometrically different hydrogen bonding contact would contribute equally. Thus, whenever very similar hydrogen bonding contacts were observed, appropriate weights were assigned in order to guarantee a unitary contribution to the statistics from any hydrogen bonding geometry.

With the aim of producing a description able to represent fluorine atoms in all their forms, both the float and trigonal geometries were hypothesized as potentially true. Moreover, four-centered hydrogen bonds were never observed, whereas three-centered interactions were frequently observed. Consequently, only two lone-pairs could be simultaneously available for protons. This led to the choice of the flat geometry for further calculations regarding the angular preference of hydrogen bonds occurring at fluorines. In fact, the alternative choice of the trigonal geometry would entail the assumption that up to three hydrogen bonds could occur simultaneously at the fluorine acceptor, which contrasts with the findings.

As previously described, only those complexes free from misleading contacts and with a reliable geometry were retained, i.e., angles D–H...F lying in the range [120°; 180°] and angles C–F...H lying in the range [90°; 180°], or [60°; 180°] for geminal fluorine atoms. The angle subtended at the fluorine by the hydrogen and the bisector of lone-pairs, namely C–F axis, is supplementary to the measured C–F...H, which will be referred to as t .

$$t = 180^\circ - \text{C}-\text{F}\cdots\text{H} \quad (4)$$

In the past the same approach was applied to other chemical groups: the values of measured angles were grouped in 10° intervals and each resulting number of observations was converted into a probability value by applying the geometrical correction.³⁶ The circular symmetry around the C–F axis is responsible for the different volumes of regions corresponding to intervals of the same breadth. Therefore, we assigned different weights by dividing each value representing the number of observations by the difference between the cosine trigonometric function of the two angles which are at the extremities of the corresponding 10° segment, as proposed in ref 31. In addition, we divided the value assumed by each bar of the histogram distribution by the value assumed by the highest bar in order to obtain a measure of unity scaling, with frequency values of hydrogen bond formation lying between 0 and 1, as presented in Figure 5.

In agreement with the expected chemical behavior of fluorine atoms, the interval $[60^\circ; 70^\circ]$ possesses the highest number of observations: the angles t made by the lone-pair for flat and trigonal geometries are 60° and 70.5° respectively, and the corresponding histogram bar is the most populated. On the two extremities of the distribution the frequency decreases due to the steric hindrance, whereas medium-frequency values in the middle prove the isotropic character of hydrogen bonding interactions (Figure 5).

The number of observed contacts is low when compared with those obtained by Boobbyer and co-workers for many kinds of nitrogen and oxygen atoms.³¹ For the interval $[0^\circ; 10^\circ]$ of angle t the amount of observed contacts was equal to zero. Hence, P_i/P_0 would assume a value of zero, and consequently its natural logarithm would not be defined and the energy expressed as in the eq 3 could not be calculated. Thus, the frequency values P_i/P_0 were plotted against the angle t as histogram distribution, as Figure 5 reports in graph form. Then, a trigonometric function was chosen to fit the frequency distribution. This trigonometric function is the combination of two functions of the angle t : the donor could be between the two lone-pairs or outside them, according to the different geometries illustrated in Figure 6. It could be plotted without using any mathematical transformation because the intrinsic nature of such trigonometric functions guarantees a unitary scaling system. The resulting function, defined as following, is overlaid to the frequency histogram bars in Figure 5.

$$E_t = |\sqrt{\cos(60^\circ - |t|)}| \text{for } [-60^\circ; 60^\circ] \quad (5a)$$

$$E_t = |\cos^3(|t| - 60^\circ)| \text{for } -120^\circ; -60^\circ \text{ [and] } 60^\circ; 120^\circ \quad (5b)$$

The extremes are included for the interval of 5a and excluded for the interval of 5b.

In correspondence with the directions of the lone-pairs the two functions assume the unit value, and both their derivatives approach zero. Along the whole existence field the E_t function is continuous and derivable, with the exception of the cusp point, which corresponds to the C–F axis: at t set to zero the function is not

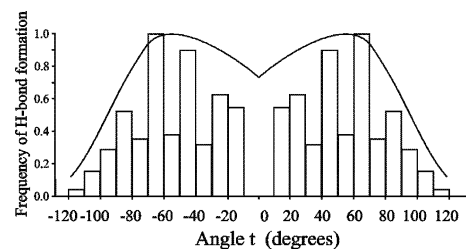


Figure 5. Histogram distribution of the frequency of hydrogen bond formation for fluorine atoms derived from crystallographic data, plotted together with the new angular energy function E_t , presented as the black curve fitting the data.

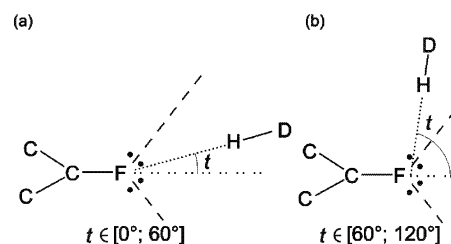


Figure 6. The probe D is able to donate only one hydrogen bond, optimally orientated toward the acceptor (fluorine). It may lie in the region between the two lone-pairs (a) or outside them (b).

derivable. Between the two lone-pairs (eq 5a) the energy is lowered due to the absence of collinearity between the lone-pair direction and the hydrogen, but the lowering is small because no steric hindrance occurs. Mathematically, the square root function allows this small decreasing of E_t when t lies between 0° and 60° and, obviously, between -60° and 0° . On the contrary, the energy is quickly reduced by a decreasing function out of the lone-pairs, where the interaction is disfavored by the steric hindrance (eq 5b). Again, the cube power is mathematically responsible for this rapid decrease for t values greater than 60° or lower than -60° .

In conclusion, the lowest energy (and strongest hydrogen bonding interaction) occurs along the lone-pair direction: the highest probability corresponds to the maximum value of the E_t function, that is to say, within the t intervals $[60^\circ; 70^\circ]$ and $[-70^\circ; -60^\circ]$, where the hydrogen approaches along the lone-pair direction of the fluorine. The lowering toward 120° is deep and rapid, caused by the steric hindrance of the substituents which are in α to fluorines, whereas the decrease toward 0° is small. Indeed, this new angular function is very good at describing the isotropic character of interactions occurring at fluorine atoms in all its forms, aliphatic and aromatic as well as geminal. In fact, frontal hydrogen bonds are still significant, whereas far bent approaches are allowed although energetically disfavored.

Superposition of GRID Maps and Experimental Findings. A probe able to donate a hydrogen bond interacts with the target generating energetically attractive regions, and GRID isocontour maps highlight those regions around the target where the heavy atom of the probe finds energetically favorable sites. To distinctly evaluate aromatic, aliphatic, and geminal fluorines of CF_2 and CF_3 groups, fluoromethane, fluorobenzene, difluoromethane, and trifluoromethane were considered as four different small targets. Each target was the input for a GRID run with the amide nitrogen

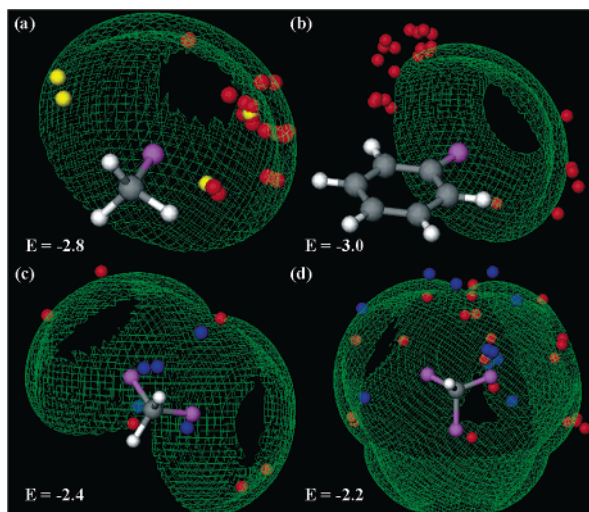


Figure 7. GRID isocontour maps generated by the amide nitrogen probe (N1), able to donate one hydrogen bond, over the fluorine-containing targets: fluoromethane (a), fluorobenzene (b), difluoromethane (c), and trifluoromethane (d). All the collected fluorine-containing ligands were superimposed over the corresponding targets with respect to the C–F bond; the heavy atoms (N and O) of donor groups involved in hydrogen bonds to fluorine were extracted from the corresponding protein structure and are shown as colored balls, according to their single (red), double (yellow), and bifurcated (blue) character. The GRID fields (green), representing the energetically favorable regions for the heavy atom of donor groups in proximity to the fluorine, were generated at specific energy values corresponding to 85% of the minimum energy value: the regions are large enough to cover the great majority of donors occurring at aromatic fluorines (b) as well as the totality of donors occurring at aliphatic (a) and geminal fluorines (c and d).

N1 as donor probe, and the grid spacing was set to 0.2 Å to guarantee excellent resolution.

For each PDB entry the fluorine-containing ligand was aligned to the corresponding target listed above by fixing the fluorine atom and the ligand skeleton, whereas the donor groups (N–H and O–H) which interact with the fluorine were extracted from the corresponding protein structure. For each class all the observed donors were grouped together, and their position was maintained with respect to the corresponding ligand. Figure 7 visualizes them in the same plot with GRID isocontour energy maps (green fields), colored according to the type. The heavy atoms oxygen and nitrogen were treated in the same manner: they are presented in yellow for donors of single interactions, whereas the color red is used for donors of double interactions and the color blue for donors of bifurcated ones. GRID allows the fluorine lone-pairs to rotate around the C–F axis, giving rise to toroidal regions of favorable energy where the heavy atoms of hydrogen donors could be placed to establish hydrogen bonds. The better the fit of experimental donors over these regions, the more the experimental data and the GRID hydrogen bonding angular function can be assumed to agree. To define significant energetically favorable regions, energy isocontour maps are presented in Figure 7: any selected energy value is 85% of the corresponding best minimum GRID energy value.

Hydrogen bond donors lie over the energy isocontour map, plotted at -2.8 kcal/mol for the aliphatic fluorine (7a): double interactions (yellow) have bent angles due

to simultaneous approach causing steric hindrance, whereas the occurrence of single interactions (red) is mainly frontal. Nevertheless, the region is large enough to cover all of them. In the case of aromatic fluorines the bent approach prevails, and some donors lie over the edge of the GRID field plotted at -3.0 kcal/mol (7b). The packed group on the left with respect to the fluorine atom of fluorobenzene belongs to the aromatic groups R5 and R6 (Table 3), and its abundance is consequently deceptive. Contacts occurring at geminal fluorines are well covered by the isocontour maps, plotted at higher energy values, namely -2.4 kcal/mol for difluoromethane (7c) and -2.2 kcal/mol for trifluoromethane (7d). In these cases a highly attractive region between two or three fluorine atoms is the effect of their synergy: many donors are grouped there, especially those corresponding to bifurcated interactions (blue). The remaining donor groups, which occur more frontally, are mainly involved in single interactions. Indeed, a good fit of overlaying isocontour maps, generated at low energy values (85% of the minimum), and the corresponding hydrogen bonding protein donors graphically confirmed the agreement with X-ray data of the new angular function of the GRID force field assigned to fluorines.

Application: Docking Experiments on Observed Ligand–Protein Complexes. The effect of the new function describing fluorine hydrogen bonding on the GRID force field was evaluated by using docking experiments. The focus of a generic docking program is to reproduce the bound conformation of a ligand in an active site. GRID has its own docking procedure called GLUE,²⁶ which uses the GRID force field to predict the ligand orientation within the protein active site.

The GLUE procedure mainly works in two steps. First, the protein cavity is mapped by several GRID runs and a view of how the cavity is energetically felt by the ligand is produced. In fact, any type of atom forming part of the ligand is used as a probe to find its own most favorable binding sites. For example, in the case of the ligand coded as FCR ($\text{CF}_3\text{--C}_6\text{H}_4\text{--OH}$), which is cocrystallized within the PDB entry **1e0y**, GRID maps are produced using the probes H, OH2, DRY, OH, and F. The first three probes (H, OH2, and DRY) are always used in order to define the regions of the site accessible to the ligand and the hydrophobic and hydrophilic binding sites of the cavity. The other two probes, OH and F, are used to find the favorite places for atoms of the ligand such as the phenoxyl oxygen, and the three fluorines; all the remaining hydrophobic atoms are already simulated by the hydrophobic probe DRY. The corresponding polar and hydrophobic chemical groups of the ligand are fitted on the resulting minima: for instance, the carbonyl oxygen of the ligand might be placed at a favorable energy minimum in the GRID map for the O probe. While keeping the first polar atom at the first minimum, the program would then search for a second favorable minimum of the target, at which one of the other polar atoms of the molecular probe could be placed. As soon as two polar atoms have been placed they define an axis round which the molecular probe might rotate until a third polar atom is close to a third appropriate energy minimum of the target. The same procedure is also used to fit hydrophobic atoms at hydrophobic locations of the target at the same time as

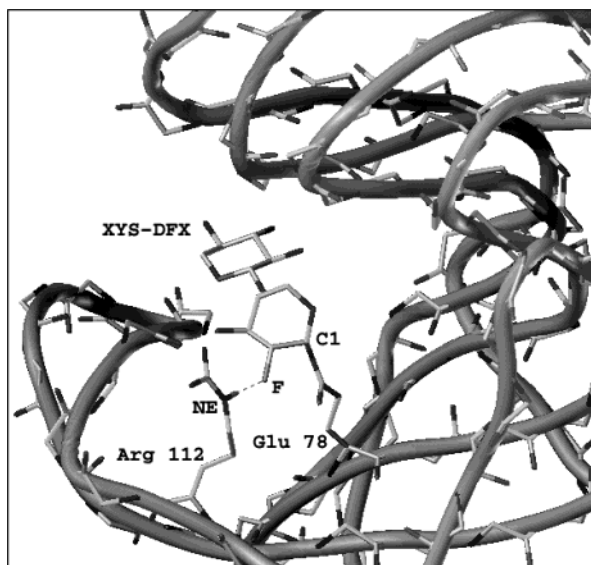


Figure 8. Since the fluorine of the ligand XYS-DFX hydrogen bonds with the Arg112, it is covalently bonded to the residue Glu78 of the protein **1c5i** through its C1 carbon.

it fits the polar groups of the molecular probe into their corresponding maps. At the end of this iterative procedure, a large number of solutions are temporally stored. In the second step, since all the ways in which three atoms of the molecular probe could bind with the target have been identified, many triplets of polar atoms are quickly eliminated from redundancy and steric hindrance constraints. Thus, the remaining binding modes of the substrate are finally reassessed and optimized toward its preferred positions within the cavity, driven by the GRID force field. The interaction energy between the whole ligand and the protein binding site is calculated by using the GRID-GLUE equation, taking Lennard–Jones, electrostatic, hydrogen bonding, and entropic energy contributions into account.

Gaining an understanding of the fluorine hydrogen bonding effect on ligand–protein binding was at the core of this application. Thus, docking experiments were carried out using the GRID force field with and without the new angular function for fluorines. Then, each fluorine atom was considered to be a polar atom accepting up to two hydrogen bonds in the first system, or a hydrophobic group in the second system. For each method all the ligand orientations proposed by the program were stored with the corresponding energies from the force field; each solution was compared by using the root-mean-square distance (rmsd) to its originating X-ray structure, calculated taking the eventual symmetry of groups into consideration. In addition to the rmsd and the energy values, the ranking of the best-placed solution was considered so that comparison of the two methods could be made.

The entire set of 49 complexes was unsuitable for the docking experiment because some ligands were covalently bonded to the corresponding protein. For 13 PDB entries (reported in detail in Supporting Information, Table S5) one of the carboxylic oxygens of aspartic or glutamic acid is linked to the carbon C1 of six-membered rings derived from carbohydrate skeleton, whereas the fluorine is always bonded to the carbon C2, as illustrated in Figure 8. On the contrary, another five

Table 4. Results over the 29 Docking Experiments

description	absence/presence of fluorine hydrogen bonding function on the GRID force field	
	absence	presence
success rate evaluated over the first solutions proposed by the GLUE docking method	48.3%	62.1%
success rate evaluated over the whole sets of solutions proposed by the GLUE docking method	86.2%	100.0%
Mean rmsd (SD)	0.72 (0.30) Å	0.71 (0.30) Å
positive effect over the docking energy ^a	24.0%	52.0%
positive effect over the docking ranking ^b	4.0%	32.0%

^a Number of complexes whose energy difference of docking solutions has either negative or positive values for, respectively, the absence or presence of fluorine hydrogen bonding function on the GRID force field. For detail see Figure 9. ^b Number of complexes whose ranking difference of docking solutions has either negative or positive values for, respectively, the absence or presence of fluorine hydrogen bonding function on the GRID force field. For detail see Figure 10.

ligands are linked in different manners to the oxygen OG of serine residues. Care was taken to treat each system in a uniform and consistent manner to avoid introduction of bias. Thus, the absence of a specific and defined binding site prompted the exclusion of the complexes **1e7c** and **1e70** from the docking experiments, the former for the abundance and low size of its ligand HLT, which presents seven different binding modes, and the latter for the inaccuracy of the coordinates of the residue Asp409, which is directly involved in binding the corresponding ligand G2F. The atoms CA, CB, CG, and CD of the side chain are defined twice, whereas OE1 is defined only once, as part of the chain B, and furthermore, OE2 is defined once, but as part of the chain A. Therefore, it was impossible to define a unique and acceptable coordinate system for the Asp409.

The remaining 29 complexes underwent the docking procedure. Data were treated uniformly, and the conformations of both proteins and ligands were taken from the X-ray crystal structures. Thus, the flexibility of ligands was not considered because the focus of this application was only on the evaluation of the effect of fluorine hydrogen bonding on ligand–protein binding rather than focusing on discussing the docking performances and drawbacks. The proteins were treated on a case-by-case basis: a specific 3D-cage for the GRID calculations was defined by using visual inspection in order to ensure the completeness of the active site and each tautomeric and ionization state was individually evaluated. Care was taken to add waters, metals, and cofactors to define the binding sites, as listed in the footnotes in Table S5 of the Supporting Information, and more rarely reaction product was also necessary to define the pocket more specifically.³⁷ Finally, only the best chain in terms of β -factor was selected for the docking experiment, when several chains were available for the same protein. After accurately producing the input for GRID, several probes were used to chemically mime the ligand. The resulting fields were converted to files containing the coordinates of the energy minima, which were used to define the starting positions for the

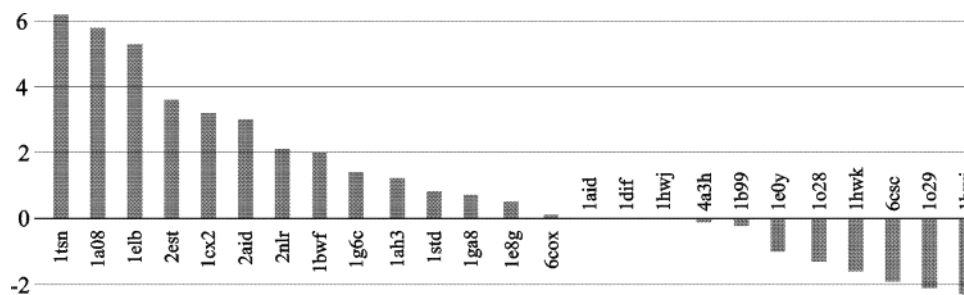


Figure 9. The bar values correspond to difference values obtained from the two systems, which differ for the absence or presence of fluorine hydrogen bonding character in the GRID force field. Therefore, a positive value for each complex, expressed in kcal/mol, represents a positive effect of the fluorine hydrogen bonding character on the overall docking energy. On the contrary, a negative value represents a negative effect.

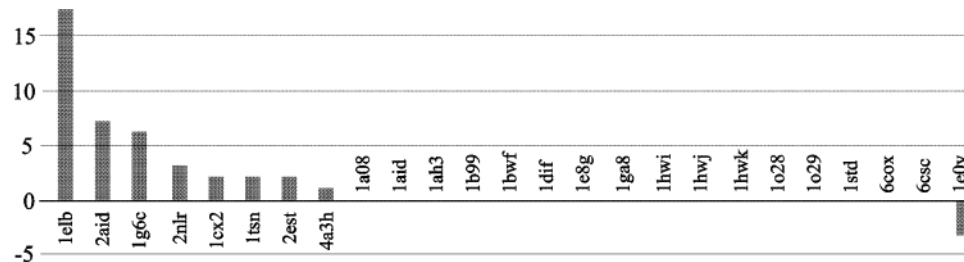


Figure 10. The bar values correspond to difference values obtained from the two systems, which differ for the absence or presence of fluorine hydrogen bonding character in the GRID force field. Therefore, a positive value for each complex, expressed in positions within the ranking scale obtained from the docking experiments, represents a positive effect of the fluorine hydrogen bonding character on the relative position of the well-docked solution within the set of solutions proposed. On the contrary, a negative value represents a negative effect.

minimization step. This was executed twice because each fluorine of the ligand was considered either as a polar atom accepting up to two hydrogen bonds or as a hydrophobic atom. This was carried out by manually manipulating the KOUT format of the ligands' files. All the docking experiments were executed uniformly by applying the default parameters. To compare the two systems, tabular results were collected and are reported in Table 4. For each ligand the orientation geometrically closest to its X-ray structure in terms of rmsd was considered as the best-docked solution, and acceptable deviations ranged from 0.0 to 2.0 Å. For each best-docked solution the energy value defined by the force field was stored in addition to the ranking position, which is its position within the set of solutions proposed by the program.

The success rate was defined as the percentage of docked ligand orientations whose rmsd to the X-ray structure is less than 2.0 Å. The result was 62.1% and 48.3% for the systems with and without the fluorine hydrogen bonding features, with accurately docked defined as only those ligands for which the best-docked solution is the first proposed by the program.

Furthermore, the presence of fluorine hydrogen bonding guarantees the complete accuracy (100%) when considering the ability of the method to reproduce the bound conformation of the ligand in the corresponding active site without considering the ranking, whereas this is less accurate with the other system, dropping to 86.2%. It is worth noting that 4 of the 29 complexes do not have any solution proposed by the system without the fluorine hydrogen bonding: these are **1elc**, **1g4j**, **1g4p**, and **1qz0**, which represent the remaining 13.8% of the set.

About half of the set (14 complexes) revealed better docking results in terms of solutions, ranking and

energy values. For **1e8g**, **1ga8**, **1std**, **1ah3**, and **1g6c** the energy difference is between 0.5 and 2.0 kcal/mol, for **1bwf** and **2nlr** between 2.0 and 3.0 kcal/mol and for **2aid**, **1cx2**, **2est**, **1elb**, **1a08**, and **1tsn** more than 3.0 kcal/mol in favor of fluorine hydrogen bonding effect. Opposite behavior (between -2.5 and -0.5 kcal/mol) was revealed for **1e0y**, **1o28**, **1hwk**, **6csc**, **1o29**, and **1hwi**. On these bases were calculated the mean values reported in Table 4. The values of energy difference for each complex herein mentioned are shown in Figure 9.

Figure 10 reports the difference in ranking caused by absence or presence of fluorine hydrogen bonding character in the GRID force field. For eight complexes (**4a3h**, **2est**, **1tsn**, **1cx2**, **2nlr**, **1g6c**, **2aid**, and **1elb**) the solution proposed by the method with the new fluorine function was closer to the corresponding crystallographic structure. For example, the PDB entry **1g6c** ranked ninth out of twenty solutions (without fluorine hydrogen bonding contribution) and ranked third out of twenty solutions with fluorine hydrogen bonding contribution. Opposite behavior was only observed for the complex **1e0y**.

The improvement of docking performances in terms of energy could be expected due to the greater contribution of hydrogen bonding energy with respect to the hydrophobic contribution, but this is only valid if a real hydrogen bond occurs. Therefore, the docking results confirmed the goodness of introducing the new angular function for describing the geometry of fluorine hydrogen bonding on the GRID force field.

A summary of the comparison of the docking results with and without the fluorine hydrogen bonding function is displayed in Table 4 and illustrated in Figures 9 and 10. A full table (S4) listing the rmsd and energy values, together with the ranking position, of both

methods for each complex is available as Supporting Information.

Conclusions

The GRID method is a fast and appropriate rational procedure for locating regions where ligands, as represented by probes, bind to molecular targets. The GRID method achieves this by determining the energy of interaction between the probe and the target. The hydrogen bonding term of the GRID energy function has been defined in the past years by investigating X-ray data for the most common chemical groups, and functions and parameters were assigned for many acceptor atoms.

The hydrogen bonding geometry of fluorine described in this paper was studied by an efficient and precise investigation of crystallographic data, and the resulting angular function was added to the GRID force field. An overview of the whole set of X-ray data depicts the observed hydrogen bonds as weak and not strongly directional. The weakness of such interactions was considered in the isotropic character given to the function, obtained by fitting the whole set of experimental observations. Nevertheless, differences were found on distances and angular preferences: hydrogen bonds of aliphatic fluorine atoms are prevalently straighter and shorter than those occurring at aromatic fluorine atoms. For the aliphatic fluorines the angular preference is reversed when two hydrogen atoms approach the same atom: the geometry is bent, due to steric hindrance between the two donors. Bifurcated interactions were not observed for C–F or Ar–F fluorines, whereas they are very common for CF₂ and CF₃ groups. In fact, a donor group may hydrogen bond to geminal fluorines either toward two fluorines atom or toward only one, with bent and straight geometries, respectively. On the contrary, four-centered interactions involving one fluorine atom and three donors, as well as four-centered interactions involving all the three fluorines of CF₃ group and the same donor, were never observed. Consequently, the flat geometry was preferred to the trigonal one.

Indeed, the angular function developed is capable of describing all the fluorine features, as highlighted by the superposition of protein donor groups, extracted from X-ray observations, on GRID isocontour maps, which graphically demonstrates the reliability of the angular function with X-ray data.

GRID maps were also used to dock each ligand within the corresponding protein binding site by using the docking procedure of GRID, called GLUE. As a consequence of the addition of the new angular function for fluorine, a significant improvement of docking performances was obtained, that is to say, in terms of success rate, energy and ranking.

Although they are weak as hydrogen bonding acceptors, fluorine atoms of ligands were observed in hydrogen bonding interactions on 18% of the complexes with fluorine-containing ligand, whereas 10% of the overall amount of fluorine atoms in the PDB is involved in hydrogen bonds. Since the use of fluorine to modify the bioavailability of drugs and lead compounds is common in pharmaceutical chemistry, and also to overcome their

metabolic instability, the correct treatment of fluorine will improve the approach to any molecular modeling query.

Acknowledgment. We gratefully acknowledge Prof. Peter Goodford for his suggestions and assistance, and Molecular Discovery staff for helpful discussions.

Supporting Information Available: Tables presenting crystallographic data for the fluorine classes, as well as docking results for all the analyzed complexes, are freely available via the Internet at <http://pubs.acs.org>.

References

- (1) MDL Drug Data Report is available from Molecular Design Limited Information Systems, Inc., 14600 Catalina St., San Leandro, CA 94577.
- (2) Muegge, I.; Heald, S. L.; Brittelli, D. Simple Selection Criteria for Drug-like Chemical Matter. *J. Med. Chem.* **2001**, *44*, 1841–1846.
- (3) Van Niel, M. B.; Collins, I.; Beer, M. S.; Broughton, H. B.; Cheng, S. K. F.; Goodacre, S. C.; Heald, A.; Locker, K. L.; MacLeod, A. M.; Morrison, D.; Moyes, C. R.; O'Connor, D.; Pike, A.; Rowley, M.; Russel, M. G. N.; Sohal, B.; Stanton J. A.; Thomas, S.; Verrier, H.; Watt, A. P.; Castro, J. L. Fluorination of 3-(3-(Piperidin-1-yl)propyl)indoles and 3-(3-(Piperazin-1-yl)propyl)indoles Gives Selective Human 5-HT_{1D} Receptor Ligands with Improved Pharmacokinetic Profiles. *J. Med. Chem.* **1999**, *42*, 2087–2104.
- (4) Wu, Y.-J.; Davis, C. D.; Dworetzky, S.; Fitzpatrick, W. C.; Harden, D.; He, H.; Knox, R. J.; Newton, A. E.; Philip, T.; Polson, C.; Sivarao, D. V.; Sun, L.-Q.; Tertyshnikova, S.; Weaver, D.; Yeola, S.; Zoeckler, M.; Sinz, M. W. Fluorine Substitution Can Block CYP3A4 Metabolism-Dependent Inhibition: Identification of (S)-N-[1-(4-Fluoro-3-morpholin-4-ylphenyl)ethyl]-3-(4-fluorophenyl)acrylamide as an Orally Bioavailable KCNQ2 Opener Devoid of CYP3A4 Metabolism-Dependent Inhibition. *J. Med. Chem.* **2003**, *46*, 3778–3781.
- (5) Zamora, I.; Afzelius, L.; Cruciani, G. Predicting Drug Metabolism: A Site of Metabolism Prediction Tool Applied to the Cytochrome P450 2C9. *J. Med. Chem.* **2003**, *46*, 2313–2324.
- (6) Rowley, M.; Hallett, D. J.; Goodacre, S.; Moyes, C.; Crawford J.; Sparey, T. J.; Patel, S.; Marwood, R.; Patel, S.; Thomas, S.; Hitzel, L.; O'Connor, D.; Szeto, N.; Castro, J. L.; Hutson, P. H.; MacLeod, A. M. 3-(4-Fluoropiperidin-3-yl)-2-phenylindoles as High Affinity, Selective, and Orally Bioavailable h5-HT_{2A} Receptor Antagonists. *J. Med. Chem.* **2001**, *44*, 1603–1614.
- (7) Dugar, S.; Yumibe, N.; Clader, J. W.; Vizziano, M.; Huie, K.; Van Heek, M.; Compton, D. S.; Davis Jr. H. R. Metabolism and structure activity data based drug design: discovery of (–) SCH 53079 an analog of the potent cholesterol absorption inhibitor (–) SCH 48461. *Biorg. Med. Chem. Lett.* **1996**, *6*, 1271–1274.
- (8) West, R.; Powell, D. L.; Whatley, L. S.; Lee, M. K. T.; Schleyer, P. von R. The relative strengths of alkyl halides as proton acceptor groups in hydrogen bonding. *J. Am. Chem. Soc.* **1962**, *84*, 3221–3222.
- (9) O'Hagan, D.; Rzepa, H. S. Some influences of fluorine in bioorganic chemistry. *Chem. Commun.* **1997**, 645–652.
- (10) Murray-Rust, P.; Stallings, W. C.; Monti, C. T.; Preston, R. K.; Glusker, J. P. Intermolecular Interactions of the C–F Bond: The Crystallographic Environment of Fluorinated Carboxylic Acids and Related Structures. *J. Am. Chem. Soc.* **1983**, *105*, 3206–3214.
- (11) Brammer, L.; Bruton, E. A.; Sherwood, P. Understanding the Behaviour of Halogens as Hydrogen Bond Acceptors. *Cryst. Growth Des.* **2001**, *1*, 277–290.
- (12) Stainer, T. Hydrogen-bond distances to halide ions and organometallic crystal structures: up-to-date database study. *Acta Crystallogr.* **1998**, *B54*, 456–563.
- (13) Mascal, M. A statistical analysis of halide...H–A (A = OR, NR₂, N⁺R₃) hydrogen bonding interactions in the solid state. *J. Chem. Soc., Perkin Trans. 2* **1997**, 1999–2001.
- (14) Allerhand, A.; Scheleyer, P. von R. Halide anions as proton acceptors in hydrogen bonding. *J. Am. Chem. Soc.* **1963**, *85*, 1233–1237.
- (15) Harbell, S. A.; McDaniel, D. H. Strong Hydrogen Bonds. II. The Hydrogen Difluoride Ion. *J. Am. Chem. Soc.* **1964**, *86*, 4497–4497.
- (16) Brammer, L.; Bruton, E. A.; Sherwood, P. Fluoride Ligands exhibit marked departures from the hydrogen bond acceptor of their heavier halogen congeners. *New J. Chem.* **1999**, *23*, 965–968.
- (17) Laurence, C.; Berthelot, M. Observations on the strength of hydrogen bonding. *Persp. Drug Discovery Des.* **2000**, *1*, 39–60.

- (18) Ouvrard, C.; Berthelot, M.; Laurence, C. The first basicity scale of fluoro-, chloro-, bromo- and iodo-alkanes: some cross-comparison with simple alkyl derivatives of other elements. *J. Chem. Soc., Perkin Trans. 2* **1999**, 1357–1362.
- (19) Howard, J. A. K.; Hoy, V. J.; O'Hagan, D.; Smith, G. T. How Good is Fluorine as a Hydrogen Bond Acceptor? *Tetrahedron* **1996**, *52*, 12613–12622.
- (20) Allen F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. The development of versions 3 and 4 of the Cambridge Structural Database System. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187–204.
- (21) Shimoni, L.; Glusker, J. P. The Geometry of Intermolecular Interactions in Some Crystalline Fluorine-Containing Organic Compounds. *Struct. Chem.* **1994**, *5*, 383–397.
- (22) Dunitz, J. D.; Taylor, R. Organic Fluorine Hardly Ever Accepts Hydrogens Bonds. *Chem. Eur. J.* **1997**, *3*, 89–98.
- (23) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (24) <http://www.rcsb.org/pdb/>
- (25) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (26) The version 22 of the GRID package is available from Molecular Discovery Ltd., 4, Chandos Street, W1A 3BQ, London, United Kingdom. <http://www.moldiscovery.com>
- (27) Thallypally, P. K.; Nangia, A. A Cambridge Structural Database analysis of the C–H...Cl interaction: C–H...Cl- and C–H...Cl-M often behave as hydrogen bonds but C–H...Cl–C is generally a van der Waals interaction. *Cryst. Eng. Commun.* **2001**, *3*, 114–119.
- (28) Aakeroy, C. B.; Evans, T. A.; Seddon, K. R.; Palinko, I. The C–H...Cl hydrogen bond: does it exist? *New. J. Chem.* **1999**, *23*, 145–152.
- (29) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441–451.
- (30) SYBYL 6.7, Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144.
- (31) Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New Hydrogen-Bond Potentials for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. *J. Med. Chem.* **1989**, *32*, 1083–1094.
- (32) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 1. Ligand Probe Groups with the Ability to Form Two Hydrogen Bonds. *J. Med. Chem.* **1993**, *36*, 140–147.
- (33) Wade, R. C.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 2. Ligand Probe Groups with the Ability to Form More Than Two Hydrogen Bonds. *J. Med. Chem.* **1993**, *36*, 148–156.
- (34) James, M. N. G.; Sielecki, A. R.; Hayakawa, K.; Gelb, M. H. Crystallographic Analysis of Transition State Mimics Bound to Penillopepsin: Difluorostatine- and Difluorostatone-Containing Peptides. *Biochemistry* **1992**, *31*, 3872–3886.
- (35) Takahashi, L. H.; Radhakrishnan, R.; Rosenfield, R. E.; Meyer, Jr. E. F. Crystallographic Analysis of the Inhibition of Porcine Pancreatic Elastase by a Peptidyl Boronic Acid: Structure of a Reaction Intermediate. *Biochemistry* **1989**, *28*, 7610–7617.
- (36) Kroon, J.; Kanters, J. A. Nonlinearity of hydrogen bonds in molecular crystals. *Nature* **1974**, *248*, 667–668.
- (37) Peapus, D. H.; Chiu, H.-J.; Campobasso, N.; Reddick, J. J.; Begley, T. P.; Ealick, S. E. Structural Characterization of the Enzyme–Substrate, Enzyme–Intermediate, and Enzyme–Product Complexes of Thiamin Phosphate Synthase. *Biochemistry* **2001**, *40*, 10103–10114.

JM0498349