Fluorine in Pharmaceuticals: Looking Beyond Intuition

Klaus Müller,[1]* Christoph Faeh,[2] François Diederich[2]*

Fluorine substituents have become a widespread and important drug component, their introduction facilitated by the development of safe and selective fluorinating agents. Organofluorine affects nearly all physical and adsorption, distribution, metabolism, and excretion properties of a lead compound. Its inductive effects are relatively well understood, enhancing bioavailability, for example, by reducing the basicity of neighboring amines. In contrast, exploration of the specific influence of carbon-fluorine single bonds on docking interactions, whether through direct contact with the protein or through stereoelectronic effects on molecular conformation of the drug, has only recently begun. Here, we review experimental progress in this vein and add complementary analysis based on comprehensive searches in the Cambridge Structural Database and the Protein Data Bank.

Fluorinated compounds are the least abundant natural organohalides (*1*). Most terrestrial F is bound in insoluble form, hindering uptake by bioorganisms. Until 1957, no F-containing drug had been developed. Since then, over 150 fluorinated drugs have come to market and now make up ~20% of all pharmaceuticals (*2*–*5*), with even higher figures for agrochemicals (up to 30%) (*4*). Top-selling fluorinated pharmaceuticals include the antidepressant fluoxetine (Prozac) (*6*), the cholesterol-lowering drug atorvastatin (Lipitor) (*7*), and the antibacterial ciprofloxacin (Ciprobay) (*8*) (Fig. 1).

Chemists have known about F's inductive effects for decades from small molecule studies (such as Hammett linear free-energy relationships). Also, F's capacity to enhance metabolic stability (mainly by lowering the susceptibility of nearby moieties to cytochrome P450 enzymatic oxidation) has become increasingly clear recently (*9*). In contrast, an understanding of how F affects binding affinity and selectivity at the molecular level is just starting to develop. Here, we highlight recent findings of F···protein interactions and complement the discussion with the analysis of structures in the Cambridge Structural Database (CSD) and the Protein Data Bank (PDB).

Structural information is essential for rationalizing F contributions to protein binding affinity. Taking the three pharmaceuticals mentioned above as examples, the CF_3 group in fluoxetine (table S1, entry 1) and the F substituents in atorvastatin and ciprofloxacin enhance potency, but this gain can be explained with confidence only for atorvastatin, which has been structurally characterized bound to its target. Atorvastatin (median inhibitory concentration = 8 nM) inhibits 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, an essential enzyme in the biosynthesis of cholesterol. In an early stage of development, a series of inhibitors featuring a pyrrole core similar to the one in atorvastatin was screened. The 4-fluorophenyl derivative was found to be superior to ligands with hydroxy (by a factor of 2), hydrogen (factor of 5), or methoxy (factor of 10) groups in this position; only the chlorinated ligand was of similar potency (*10*). The x-ray crystal structure of atorvastatin bound to HMG-CoA reductase (table S1, entry 2) revealed that the aromatic C–F of the ligand approaches the guanidinium side chain of Arg590, [distance d(F···C(N3) = 2.9 Å)], hinting at a favorable polar interaction (*11*) discussed in detail below. Although published structure-activity relationships for quinolone antibiotics (*12*) such as ciprofloxacin might again hint at a favorable polar interaction of the essential F substituent at position 6, the absence of structural information limits confirmation of any such hypothesis (for the structure of ciprofloxacin bound to the AcrB multidrug efflux pump, see table S1, entry 3).

Synthetic Advances

Long after Moisson's preparation of elemental F_2 in 1886, its extreme reactivity still limited widespread laboratory fluorinations. This situation changed around 1970 with the introduction of safe and selective fluorinating agents that were compatible with ordinary laboratory equipment and therefore amenable to elaboration of lead compounds (*13*–*15*).

Today, an increasing number of such agents are directly available to researchers from commercial suppliers. Examples of nucleophilic reagents used to form C–F bonds (Fig. 2A) include diethylaminosulfur trifluoride (DAST) (*16*), 2,2-difluoro-1,3-dimethylimidazolidine (DFI) (*17*), and bis(2-methoxyethyl) aminosulfur trifluoride (Deoxofluor) (*18*); these reagents transform alcohols into monofluorides and carbonyls into gem-difluorides. A wide range of electrophilic reagents bearing a R_2N–F or R_3N^+–F unit has also been developed and commercialized, elaborated from the first such agent, pyridinium poly(hydrogen fluoride) (Olah's

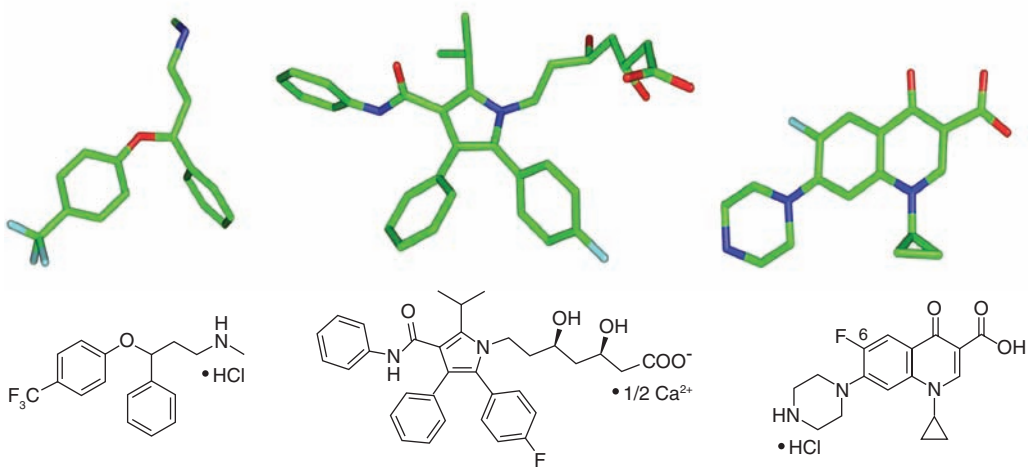


Fig. 1. Major fluorinated drugs: the antidepressant Prozac (table S1, entry 1), cholesterol-lowering drug Lipitor (table S1, entry 2), and quinolone antibiotic Ciprobay (table S1, entry 3). The molecular-model conformations are from crystal structures. Ligand Cs, green; O atoms, red; N atoms, dark blue; and F atoms, light blue. Unless otherwise stated, this color code also applies to the images in Figs. 3 and 5 and the supporting online material (SOM). Images generated with MacPyMol (*68*).

[1]Pharmaceuticals Division, Discovery Chemistry, F. Hoffmann–La Roche, CH-4070 Basel, Switzerland. [2]Laboratorium für Organische Chemie, ETH Zürich, Hönggerberg, CH-8093 Zürich, Switzerland.

*To whom correspondence should be addressed. E-mail: klaus.mueller@roche.com (K.M.); diederich@org.chem.ethz.ch (F.D.)

reagent) (*13*). Examples (Fig. 2B) include 1-chloromethyl-4-fluorodiazoniabicyclo [2.2.2]octane bis(tetrafluoroborate) (Selectfluor) (*13*) and *N*-fluorobenzene-sulfonimide (NFSI) (*13*). Trifluoromethyl groups are conveniently introduced with trimethyl(trifluoromethyl)silane (Ruppert-Prakash reagent) (*13*) or the more recently developed trifluoroacetamides (Fig. 2C) (*19*). Protocols for asymmetric fluorinations are also expanding (*20*). Nevertheless, most of these reagents are too expensive for plant-scale production, where traditional methods such as the use of elemental F prevail.

Physical and Pharmacokinetic Properties

Fluorine is the most electronegative element. The C–F bond is one of the strongest known (table S2), and adjacent C–C single bonds are also strengthened, whereas allylic C=C double bonds are weakened by F substitution (*21*). The very low polarizability of organofluorine substituents also impacts intermolecular interactions (*22*). Further, the nuclear magnetic resonance (NMR) activity of F's sole natural isotope, ^{19}F, is convenient for characterization (*23*).

Fluorine often replaces H in organic molecules but the size and stereoelectronic influences of the two atoms are quite different (for bond lengths and van der Waals radii and volumes, see table S2). The bond length of C–F (1.41 Å) is actually more similar to C–O (1.43 Å) than to C–H (1.09 Å), although packing-radius comparisons are a subject of ongoing research (*24*). The van der Waals volume of the trifluoromethyl (CF_3) group (as in fluoxetine) is similar to that of the ethyl group (CH_3CH_2) but the shapes of the two groups are very different. Despite suggestions that CF_3 and isopropyl [$(CH_3)_2CH$] are interchangeable (*21*), isopropyl has a larger volume and is axially anisotropic (*25*).

Bioisosterism is an important concept in lead optimization. It refers to the capacity of atoms or functional groups with similar sizes or shapes to be interchanged without substantially altering biological behavior such as binding affinity (*26*). Thus, the fluorovinyl group (C=CHF) has been used as a replacement for the peptide bond (*27*). Fluorine takes the position of the carbonyl O, and the planarity of the vinyl unit makes it quite a good match in size and geometry, as shown in the inhibition of dipeptidyl peptidase IV. The C–F bond length and the total extension of a C–F unit are similar to the values for the C=O group (table S2). The C–CF_3 fragment has also been introduced as a substitute for the C=O group, providing a substantial gain in potency of cathepsin K inhibitors (*28*). The C–F and C–O dipoles can undergo similar multipolar interactions with neighboring dipoles.

Bioisosterism of C–F, C–OH (*29*), and C–OMe (where Me is methyl) was also observed for a series of tricyclic inhibitors of thrombin, a Ser protease from the blood-coagulation cascade (*30*). Similar potency was found for inhibitors in which any of the three groups were bound at specific positions on the central scaffold so as to point into the region of the catalytic triad and the oxyanion hole of the enzyme. C–F, C–OH, and C–OMe thus appear to be bioisosteric (in terms of binding efficacy) if the negative poles (O, F) interact with the positive pole of another dipole or a positively charged center, provided that the Me group of C–OMe can be accommodated without strain and that the OH group finds a H-bond acceptor. A striking example of this analogy between C–OH and C–F has also been noted (*11*) in a crystal structure of a HIV protease complex with a bound peptidic inhibitor containing a central α-difluoroketone hydrate unit (table S1, entry 4).

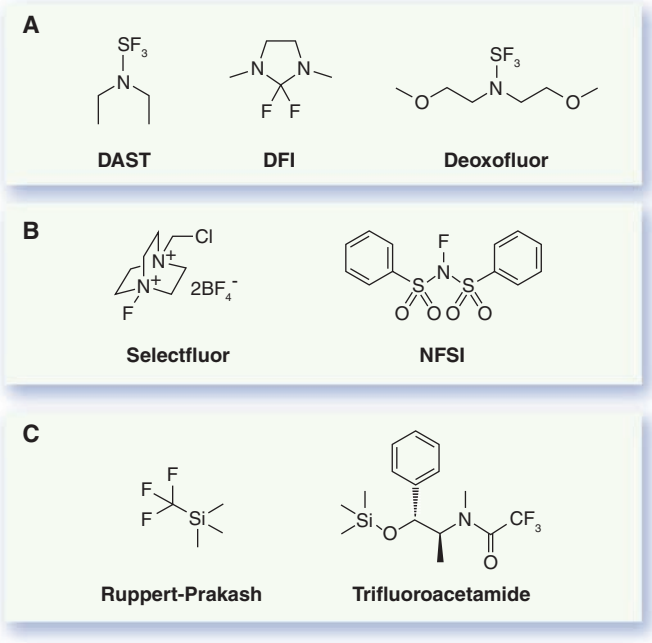


Fig. 2. Examples of safe and selective fluorination agents. (**A**) Nucleophilic agents, (**B**) electrophilic agents, and (**C**) reagents to introduce CF_3 groups.

Conformational and Stereoelectronic Influences

Knowledge about the energetically most favorable conformation of a ligand is essential for optimizing the binding efficacy, which increases with the ligand's degree of preorganization: The more closely the geometry of the bound ligand resembles the lowest-energy conformation of the free ligand, the stronger the gain in binding free energy. Ligand preorganization usually translates into enforcing the most favorable conformation through rigidification. When binding a preorganized ligand, no enthalpy loss occurs to reach the favorable binding geometry, and there is no need to freeze out the desirable binding conformation in an entropically unfavorable way. Conformations of free ligands are usually determined experimentally in solution by ^1H NMR measurements, and this information is frequently complemented by gas-phase theoretical calculations and conformational searches in the CSD (*31*, *32*). Information on the conformation of bound ligands is usually extracted from protein-ligand cocrystal structures.

Substitution of H by F can profoundly change the conformational preferences of a small molecule because of size and stereoelectronic effects. A comparison between methoxyphenyl and trifluoromethoxyphenyl groups illustrates the influence of F on conformation. Methoxy groups lie in the plane of the phenyl ring because the p orbital of the sp^2-hybridized O is in π conjugation with the aromatic π system. This conformation is preferred by ~3.0 kcal/mol (*33*, *34*). In contrast, trifluoromethoxy groups tend to turn out of plane because of their larger size and, presumably more substantially, stereoelectronic effects. Orienting C–F bonds antiperiplanar to the lone pairs of the now sp^3-hybridized O results in an anomeric n_O–σ^*_{CF} conjugation with concomitant lengthening of the C–F bonds (*35*). This effect reduces the conjugation between O and the aromatic π system and eliminates the energetic preference of a planar, in-plane conformation.

Computational studies and searches in the CSD (fig. S1) (*31*, *32*) and the PDB (*36*) were carried out to investigate the conformational preferences of aromatic OCH_3, OCH_2F, $OCHF_2$, and OCF_3 groups. Only cases in which at least one ortho position of the aromatic ring was unsubstituted were considered, as two non-H ortho groups lead to orthogonal orientations of all alkoxy groups for steric reasons. Neither CSD nor PDB searches yielded any structures with aryl–OCH_2F motifs; computational studies revealed (*33*, *34*) two conformational energy minima with dihedral angles θ [C_{aryl}–C_{aryl}–O–C(H_2F)] of 24° (ΔE_{rel} = 0.0 kcal/mol) and 0° (ΔE_{rel} = 4.2 kcal/mol). Whereas the high-energy conformation is predicted to be planar without an anomeric effect, the low-energy twisted conformation has the C–F unit in an anomeric arrangement. Overall, the change from OCH_3 to OCH_2F reduces the in-plane conformational preference.

The CSD search provided seven aryl–$OCHF_2$ fragments, with torsional angles [C_{aryl}–C_{aryl}–O–C(HF_2)] between 1° and 90°. Thus, there seems

to be no orientational preference, as was also confirmed by calculations that afforded equal-energy minima at θ = 33° and 90°. The orthogonal conformation has both C–F bonds in an anomeric (endo) orientation similar to the trifluoromethoxy group, whereas the twisted conformation is predicted to have only one C–F bond in an anomeric position (with a slightly longer C–F bond length than the nonanomeric C–F bond). This prediction is particularly nicely borne out by a crystal structure (table S1, entry 5) in which the asymmetric unit contains two independent molecules, one with an approximate orthogonal conformation (θ = 86°) with two anomeric C–F bonds, the other with a twisted arrangement (θ = 58°) and only one anomeric C–F bond. This structure provides experimental evidence that the two conformations cannot be largely different in energy as they coexist within the same crystal. Seven structures in the PDB feature ligands with aryl–$OCHF_2$ fragments (table S1, entries 6 to 11). The measured dihedral angles vary from θ = 0° to 50°. Three of the seven x-ray crystal structures were complexes of phosphodiesterase 4 with roflumilast (DAXAS), a drug against respiratory diseases that was withdrawn from the market in 2005 (Fig. 3) (*37*).

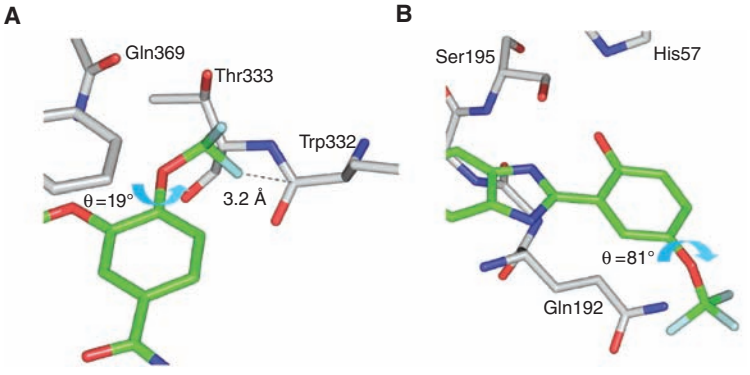


Fig. 3. (**A**) Aryl–$OCHF_2$ fragment (θ = 19°) of roflumilast bound to phosphordiesterase 4 (table S1, entry 7) shows a multipolar C–F···C=O contact with the backbone amide of Trp332. (**B**) Aryl–OCF_3 fragment (θ = 81°) of an inhibitor bound to the Ser protease trypsin (table S1, entry 13). Protein Cs are shown in gray (also applies to Fig. 5 and the SOM).

Thirteen structures with aryl–OCF_3 fragments were found in the CSD. The majority show a preference for a dihedral angle C_{aryl}–C_{aryl}–O–C(F_3) of θ ≈ 90°. Thus the O–CF_3 bond seems to prefer an orthogonal orientation to the aromatic plane, although the calculations suggest that the energetic differences in the dihedral-angle range of 0° to 90° are small (~1 kcal/mol). The PDB contains eight structures with aryl–OCF_3 fragments (table S1, entries 12 to 19); six have a dihedral angle θ between 81° and 86°, so here too we see a preference for orthogonal alignment of the O–CF_3 bond (Fig. 3). The one structure in the CSD (table S1, entry 20) with an aryl–SCF_3 bond also shows a 90° dihedral angle.

The introduction of F into piperidine rings decreases the basicity of the N center (*38*), thereby improving oral bioavailability, as has been shown for ligands of the human 5-HT_{1D} receptor, a target in migraine therapy (*39*), and for antagonists of the h5-HT_{2A} receptor, a target in schizophrenia therapy (*40*). Fluorine in protonated 3-fluoro- and 3,5-difluoropiperidines strongly prefers the axial position in aqueous solution, whereas after deprotonation, the F substituent adopts an equatorial position (*41*, *42*). In the axial orientation, the polar C–F and N^+–H fragments undergo favorable antiparallel dipolar interactions. These intramolecular interactions are quite effective, and the axial preference of 3-fluoro- substituents is maintained in protonated 3-fluoro-*N*-alkylpiperidinium salts and even in quaternary 3-fluoro-*N*,*N*-dialkylpiperidinium salts, despite steric congestion.

Influence of logD and pK_a Effects

Biological absorption and distribution are largely controlled by the ionization state and balance of lipophilicity and hydrophilicity in a drug molecule. Enhanced lipophilicity can increase the measured binding free energy through more favorable partitioning between the polar aqueous solution and the less polar receptor site. A convenient measure of lipophilicity is the logarithmic coefficient (logD) for distribution (D) of a compound between octanol and water at pH 7.4. In general, H/F exchange leads to a more lipophilic molecule. The logD values of nearly 300 compounds have been measured, and they followed this trend (*43*). A single H/F exchange raises the logD value by approximately 0.25. A large increase in logD is usually seen when F is introduced nearby a basic N. Amine basicity is decreased because of the σ-inductive effect of F, and thus the ratio of neutral to protonated molecules increases.

However, there are exceptions wherein the introduction of F, in particular into aliphatic chains and rings, leads to a reduction in the logD value. LogD values sometimes decrease when F is introduced near O (*43*) or N atoms (*30*). Such a decrease was observed when the O···F distance of at least one low-energy conformer in a molecule was smaller than 3.1 Å, and this finding has been tentatively explained by solvation effects. On the other hand, replacement of a methoxy group by a trifluoromethoxy group attached to an aryl unit may result in a marked increase in lipophilicity as seen, for example, in the ΔlogP of 1.1 between trifluoromethoxy- and methoxybenzene [the logarithmic partition coefficient logP (octanol/water) and logD are identical for nonionizable solutes] (*44*). This finding can be explained by a primary H/F effect augmented by a depolarization and desolvation effect by the orthogonally oriented OCF_3 group, partially shielding one π face of the aromatic ring.

Also somewhat surprisingly, (3-fluoropropyl)-benzene and (3,3,3-trifluoropropyl)benzene are markedly less lipophilic than nonfluorinated propylbenzene, with ΔlogP = –0.7 and –0.4, respectively (fig. S2). This result can be explained by the introduction of polarity to a highly hydrophobic domain. In this example, the polarity effect is stronger for the monofluorinated analog than for the trifluorinated one.

Fluorine introduction also strongly reduces amine basicity, impacting membrane permeability (*45*), the potential liability for phospholipidosis (*46*), and interference with the *hERG* (*human ether a-go-go–related gene*) K^+ channel associated with cardiovascular toxicity (*47*–*49*). Useful predictive rules have been developed for tuning the pK_a values (where K_a is the acid dissociation constant) of basic amine centers through σ-transmission effects of F, O, N, and S functionalities (*38*, *50*). Thus, the pK_a value steadily decreases upon F introduction in the series $CH_3CH_2NH_2$ (10.7) → $FCH_2CH_2NH_2$ (9.0) → $F_2CHCH_2NH_2$ (7.3) → $F_3CCH_2NH_2$ (5.7) (*38*). A sufficient number of nearby F atoms can leave an amine unprotonated at physiological pH, resulting in higher bioavailability, as shown for inhibitors of the human 5-HT_{1D} receptor (*39*). In alicyclic systems, substantial conformational and stereoelectronic effects are only beginning to be identified and understood (*38*, *51*). For example, σ-transmission effects in five-membered rings (such as pyrrolidines) are only 70 to 80% as efficient as those in six-membered rings (such as piperidines), in which perfectly staggered conformations, similar to those of aliphatic chains, can be adopted (*38*).

Over the past 5 years, we have conducted a "fluorine scan" of tricyclic inhibitors of thrombin to map the fluorophilicity and/or fluorophobicity of the enzyme active site (*30*, *52*–*56*). Fluorine was systematically introduced at various positions of the inhibitor skeleton to explore specific interactions of the halogen with active-site amino acid residues of the enzyme. The binding mode of the tricyclic inhibitors at the thrombin active site was confirmed by several crystal structures of protein-ligand complexes and is schematically shown in Fig. 4 (*30*, *52*, *56*, *57*). Remarkably, the pK_a value of the tertiary-amine center in the inhibitors can be tuned from the usual value near 10 to less than 2, through σ-transmission effects of remote Fs. The pK_a value of the tertiary-amine center in tricyclic (±)**1** is 7.0, 3 units below the

approximate value for a simple trialkylamine (~10.2) and some 4 units below the pK_a of heliotridane, a methylated hexahydropyrrolizidine that is the closest comparison compound for the bicyclic fragment that incorporates the tertiary-amine center (*38*). This pK_a lowering results from the electron-withdrawing effects of the two imide carbonyl groups. The phenylamidinium moiety in (±)**2** reduces the pK_a value further by 2.5 units. Adding one F to the terminal five-membered ring [(+)**3** to (+)**6**] lowers the pK_a value to ~3.3. In each of these compounds, F is in both β and γ positions of the tertiary amine and thus draws electron density through two σ pathways. The pK_a reduction by OH [(+)**7**] and MeO [(+)**8**] groups is much less effective. The protonated difluorinated compounds (+)**9** and (±)**10** are even moderately strong acids (pK_a < 2).

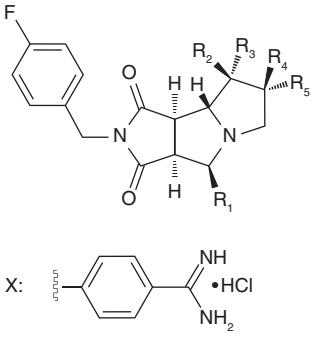
Carbonic anhydrase II is another enzyme for which organofluorine effects on physical properties and binding efficacies of aliphatic and benzenesulfonamide-based inhibitors have been extensively studied (*58*, *59*). The introduction of F near the sulfonamide (RSO_2NH_2) moiety increases the acidity of the N–H bond, facilitating deprotonation and stronger binding of the resulting anion to the Zn(II) ion at the active site of the enzyme. This point is nicely illustrated by comparing the inhibitory potency of weakly acidic methylsulfonamide ($CH_3SO_2NH_2$) [pK_a ≅ 10.5, inhibition constant (K_i) = 10^{-4} M] to that of the much more acidic trifluoromethylated counterpart $CF_3SO_2NH_2$ (pK_a = 5.8, K_i = 2 × 10^{-9} M) (*60*).

Selective Protein-Ligand Interactions

Electronegativity considerations would suggest that C–F behaves similarly to C–O and C–N fragments and acts as a good H-bond acceptor. However, an extensive search of the CSD and the PDB revealed this not to be the case (*61*). The C–F unit is a poor H-bond acceptor: Organic F has a very low proton affinity and is weakly polarizable. Nevertheless, the large number of C–F···H–X (where X = O, N, S) as well as C–F···H–C_α (C_α carbon of α amino acids) contacts points to the fact that it is favorable for the C–F dipole to undergo multipolar interactions (*11*). C–F···H–N (backbone amide) interactions are abundant in the PDB (*36*). Out of 788 C–F containing structures, constraining the F···N separation below the van der Waals contact distance of d_1 = 3.1 Å and the angles $\alpha_1 \geq 150°$ and $90° \leq \alpha_2 \leq 150°$ gave 11 structures in which the C–F moiety of the ligand points toward the H–N bond (Fig. 5A). In the case of two thrombin inhibitors that differ by only one H/F substitution (*43*), the F-containing inhibitor is more potent by a factor of 5 and shows a dramatic conformational change in its bound state when compared to the nonfluorinated ligand (Fig. 5D and fig. S3). The crystal structures (table S1, entries 21 and 22) reveal for the fluorinated analog a dipolar C–F···H–N interaction with a distance of 3.5 Å, which could be responsible for the observed change in conformation. This distance is well beyond H-bonded contact distances, but the conformational change and the resulting gain in potency provide particularly strong evidence for energetically favorable dipolar interactions.

Orthogonal multipolar C–F···C=O interactions were nicely revealed during the fluorine scan of tricyclic thrombin inhibitors (*52*). Introduction of F in the para position of the benzyl ring occupying the D pocket of thrombin (Fig. 4) enhanced the binding affinity by a factor of 6 ($\Delta\Delta G$ = –1.1 kcal mol^{-1}, where $\Delta\Delta G$ is the difference in binding free enthalpy between para-F-substituted and -unsubstituted ligands) (*52*). X-ray crystallography revealed (table S1, entry 23) that the C–F residue interacts not only at short distance with H–C_α [d(F···C 3.1 Å)] but also in an orthogonal fashion, with the backbone C=O group of Asn98 (Fig. 5C). Such orthogonal multipolar interactions were subsequently shown to be abundant in both small molecule x-ray crystal structures and in protein-ligand complexes (*11*, *54*), although they had not been recognized as such [for an early example, see fig. S4 (*62*)]. Investigations of a model system in chemical double-mutant cycles subsequently confirmed the attractive nature of the orthogonal C–F···C=O interaction, with a contribution in binding free enthalpy in apolar environments of $\Delta\Delta G$ = –0.2 to –0.3 kcal mol^{-1}, about a third of the gain from a neutral H bond (*63*). Such contacts are observed for both aliphatic and

A



Compound	R^1	R^2	R^3	R^4	R^5	pK_a
(±)-**1**	H	H	H	H	H	7.0
(±)-**2**	X	H	H	H	H	4.5
(+)-**3**	X	F	H	H	H	3.4
(+)-**4**	X	H	F	H	H	3.3
(+)-**5**	X	H	H	F	H	3.3
(+)-**6**	X	H	H	H	F	3.3
(+)-**7**	X	OH	H	H	H	4.1
(+)-**8**	X	OMe	H	H	H	3.7
(+)-**9**	X	F	F	H	H	<2
(±)-**10**	X	H	H	F	F	<2

B

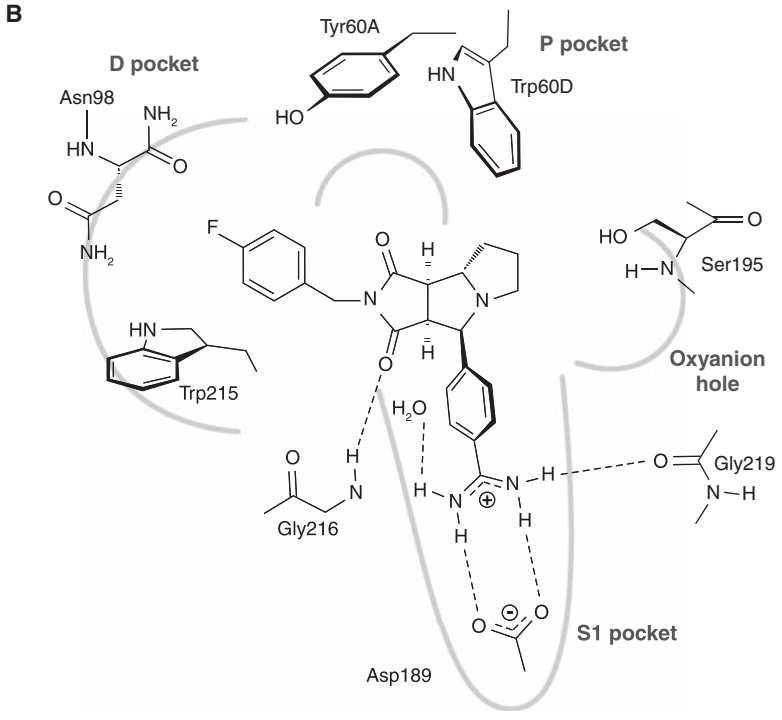


Fig. 4. (**A**) pK_a values for the tertiary-amine center in tricyclic inhibitors of the Ser protease thrombin. (**B**) Binding mode of the inhibitors as confirmed by x-ray crystal structures of protein-ligand complexes.

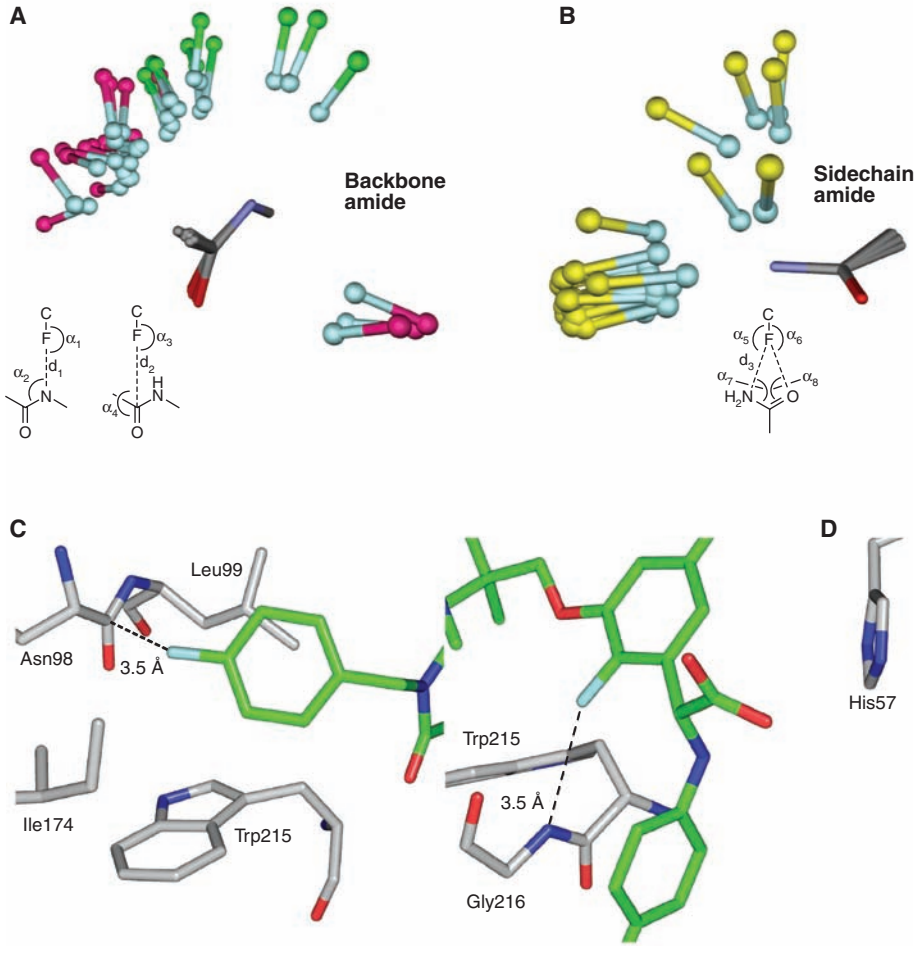


Fig. 5. (**A**) Fluorine interacts favorably with peptidic N–H (ligand Cs in green) and C=O (ligand Cs in purple) moieties. (**B**) Fluorine undergoes dipolar interactions with side-chain amides of Gln and Asp (ligand Cs in yellow). (**C**) The C–F residue of a tricyclic inhibitor undergoes a multipolar interaction with the backbone C=O of Asn98 in the D pocket of thrombin (table S1, entry 23). (**D**) A dipolar N–H···F–C interaction induces the shown conformation of a thrombin inhibitor within the enzyme (fig. S3 and table S1, entry 22).

aromatic C–F and are also seen for CF_3 groups. In the structure of an inhibitor of the trifluoroacetyl peptide class bound to porcine pancreatic elastase (table S1, entry 24), all three F atoms of the CF_3CO group interact in an orthogonal fashion with three backbone C=O groups of the protein. At short contact distance, orthogonal C–F···C=O interactions are much more frequent than the energetically more favorable antiparallel dipolar alignments (*11*), which can be explained by reduced steric hindrance.

An earlier search for orthogonal C–F···C=O interactions in the PDB (*11*, *54*) was repeated with an upper cutoff limit for the F···C distance of d_2 = 3.3 Å (Fig. 5A) and angles set to values of $\alpha_3 \geq 140°$ (C–F···$C_{C=O}$) and $70° \leq \alpha_4 \leq 110°$ (F···$C_{C=O}$=O). This search yielded 20 hits shown in Fig. 5A. Most of these hits also feature additional favorable C–F···H–C_α interactions below the van der Waals distance. An overlay of both C–F···H–N and C–F···C=O interactions illustrates how F organizes around backbone amides, which clearly provide a fluorophilic environment.

The new search also revealed a specific geometric preference for the interaction between C–F and side-chain amide residues of Gln and Asn. In 17 out of 23 hits, C–F points frontally onto the H_2N–C=O moiety, with the F···N distance as the shortest ($d_3 \leq 3.1$ Å). The O···F–C and N···F–C angles amount to $120° \leq \alpha_{5/6}$, whereas the $C_{C=O}$–N···F and $C_{C=O}$–O···F angles have values of $60° \leq \alpha_{7/8} \leq 150°$ (Fig. 5B). An example for this preferred interaction is found in the complex of endoglucanase Cel5A bound to a fluorinated inhibitor (table S1, entry 25); here, the F···N distance is 2.9 Å.

Beyond highlighting the favorable character of orthogonal C–F···C=O interactions, the fluorine scan of the thrombin inhibitors leads to some general conclusions about fluorophobic environments. C–F bonds pointing into highly polar environments such as the oxyanion hole of thrombin were found to reduce binding affinity. Also, C–F bonds avoid pointing directly at the O atom of C=O groups. Thus, the behavior of F contrasts with that of the larger, more polarizable halogens Br and in particular I, for which such linear C–X···O=C alignment is observed ("halogen bonding") (*11*). Tricyclic thrombin inhibitors directing fluoroalkyl and alkyl residues of appropriate and similar size into the narrow P pocket of thrombin, which is lined by the side chains of Trp60D, Tyr60A, and His57 (Fig. 4), gave similar binding affinities. However, overlays of crystal structures suggest that a CHF_2 group in the P pocket points away from the electron-rich π surface of the indole ring of Trp60D lining the pocket (*55*). A preference for positively polarized environments is consistent with F's high electronegativity.

This conclusion also finds support in a PDB search for C–F interactions with the guanidinium group of Arg, inspired by observations for the complex of atorvastatin bound to HMG-CoA reductase (table S1, entry 2). In many protein-ligand complexes, C–F bonds point toward the guanidinium moiety of Arg (fig. S5). However, no linear C–F···H–N interactions were observed, in agreement with the poor H-bond-accepting capacity of F. Instead, C–F bonds were found to orient either parallel to or more orthogonally to the guanidinium plane with its delocalized positive charge. A high total of 32 structures showed distances below 3.8 Å (the distance cutoff takes into account the longer range of interactions involving charges) between the F atom and the central C atom of the guanidinium residue, which clearly highlights the fluorophilic character of the Arg side chain.

Fluorine also affects the aromatic interactions of the phenyl ring to which it is attached. The positive polarization of neighboring ortho H atoms is increased, strengthening C–H···X (where X = O, N) and C–H···π interactions. Upon moving from benzene to hexafluorobenzene, the quadrupole moment changes with increasing fluorination from a large negative to a large positive value (*64*–*66*). In benzene, the negative poles are on the π surfaces and the positive poles on the C–H residues; in hexafluorobenzene, the charge distribution is exactly opposite. This distinction has strong implications for aromatic interactions. Benzene and hexafluorobenzene undergo efficient eclipsed face-to-face stacking interactions, which have been used as a construction principle in supramolecular chemistry (*67*). In medicinal chemistry, it was shown that face-to-face stacking interactions between the pentafluorophenyl ring of a 1,3,4-thiadiazole-2-thione-based inhibitor and Tyr155 of the metalloprotease stromelysin (table S1, entry 26) strongly contribute to the protein-ligand binding affinity. Nevertheless, a variety of electrostatic forces such as dipole-dipole, dipole-induced dipole, or dipole-quadrupole interactions may dominate quadrupole-moment interactions and induce other orientational preferences between aromatic and fluoroaromatic rings. Thus, a rare edge-to-face interaction between Phe131 and a perfluorophenyl ring was

seen in the crystal structure of an inhibitor bound to carbonic anhydrase II (table S1, entry 27).

Outlook

It is becoming clear that F can enhance binding efficacy and selectivity in pharmaceuticals. As small atoms of high electronegativity, F substituents on ligands prefer to orient toward electropositive regions of receptor sites. Distinct fluorophilic environments in proteins include the ubiquitous peptide bonds (particularly those in hydrophobic environments), which undergo multipolar C–F···H–N, C–F···C=O, and C–F···H–C_α interactions, as well as the side-chain amide residues of Asn and Glu and the positively charged guanidinium side chain of Arg. Correspondingly, F introduction into regions of high electron density can adversely affect the binding affinity. The introduction of fluoroalkyl substituents into tight lipophilic pockets lined by electron-rich aromatic rings neither increases nor decreases binding affinity substantially, as compared with similarly sized alkyl residues. However, taking into account advantageous effects on physicochemical properties, an overall benefit may well result from the decoration of ligands with fluoroalkyl residues to occupy apolar aromatic pockets. On the basis of these conclusions, we suggest systematic fluorine scans of ligands as a promising strategy in lead optimization, not only to enhance physicochemical and adsorption, distribution, metabolism, and excretion properties, but also to strengthen protein-ligand binding interactions.

References and Notes

1. D. B. Harper, D. O'Hagan, C. D. Murphy, in *The Handbook of Environmental Chemistry*, vol. 3P, G. W. Gribble, Ed. (Springer, Heidelberg, Germany, 2003), pp. 141–169.
2. Integrity (Prous Science, Barcelona, Spain, database analysis done on 18 August 2006, www.prous.com).
3. J. P. Bégué, D. Bonnet-Delpon, *J. Fluorine Chem.* **127**, 992 (2006).
4. C. Isanbor, D. O'Hagan, *J. Fluorine Chem.* **127**, 303 (2006).
5. K. L. Kirk, *J. Fluorine Chem.* **127**, 1013 (2006).
6. D. T. Wong, F. P. Bymaster, E. A. Engleman, *Life Sci.* **57**, 411 (1995).
7. B. D. Roth, in *Progress in Medicinal Chemistry*, vol. 40, F. D. King, A. W. Oxford, Eds. (Elsevier, Amsterdam, 2002), pp. 1–22.
8. K. Drlica, M. Malik, *Curr. Top. Med. Chem.* **3**, 249 (2003).
9. B. K. Park, N. R. Kitteringham, P. M. O'Neill, *Annu. Rev. Pharmacol. Toxicol.* **41**, 443 (2001).
10. B. D. Roth *et al.*, *J. Med. Chem.* **33**, 21 (1990).
11. R. Paulini, K. Müller, F. Diederich, *Angew. Chem. Int. Ed.* **44**, 1788 (2005).
12. J. Matsumoto *et al.*, *J. Med. Chem.* **27**, 292 (1984).
13. P. Kirsch, *Modern Fluoroorganic Chemistry* (Wiley-VCH, Weinheim, Germany, 2004).
14. M. Schlosser, *Angew. Chem. Int. Ed.* **45**, 5432 (2006).
15. W. R. Dolbier Jr., *J. Fluorine Chem.* **126**, 157 (2005).
16. W. J. Middleton, *J. Org. Chem.* **40**, 574 (1975).
17. H. Hayashi, H. Sonoda, K. Fukumura, T. Nagata, *Chem. Commun.* **2002**, 1618 (2002).
18. G. S. Lal, G. P. Pez, R. J. Pesaresi, F. M. Prozonic, H. Cheng, *J. Org. Chem.* **64**, 7048 (1999).
19. J. Joubert *et al.*, *Angew. Chem. Int. Ed.* **42**, 3133 (2003).
20. J. A. Ma, D. Cahard, *Chem. Rev.* **104**, 6119 (2004).
21. B. E. Smart, in *Organofluorine Chemistry*, R. E. Banks, B. E. Smart, J. C. Tatlow, Eds. (Plenum, New York, 1994), pp. 57–88.
22. B. E. Smart, *J. Fluorine Chem.* **109**, 3 (2001).
23. R. Martino, V. Gilard, F. Desmoulin, M. Malet-Martino, *Chemotherapy* **52**, 215 (2006).
24. J. D. Dunitz, W. B. Schweizer, *Chem. Eur. J.* **12**, 6804 (2006).
25. F. Leroux, *ChemBioChem* **5**, 644 (2004).
26. G. A. Patani, E. J. LaVoie, *Chem. Rev.* **96**, 3147 (1996).
27. K. Zhao, D. S. Lim, T. Funaki, J. T. Welch, *Bioorg. Med. Chem.* **11**, 207 (2003).
28. W. C. Black *et al.*, *Bioorg. Med. Chem. Lett.* **15**, 4741 (2005).
29. J. C. Biffinger, H. W. Kim, S. G. DiMagno, *ChemBioChem* **5**, 622 (2004).
30. E. Schweizer *et al.*, *ChemMedChem* **1**, 611 (2006).
31. CSD Version 5.28 (Cambridge Crystallographic Data Centre, Cambridge, UK, updated January 2007, www.ccdc.cam.ac.uk/).
32. All searches have been performed on structures with an *R* factor ≤ 0.1 (no errors, no partially disordered structures), excluding polymeric, metalloorganic, and powder structures.
33. Computational studies were done using density functional theory [B3LYP (Becke 3-Parameter, Lee, Yang, Parr)] with the basis set 6-31G* within Spartan06 (*34*).
34. Spartan06 (Wavefunction, Irvine, CA, 2006).
35. F. Leroux, P. Jeschke, M. Schlosser, *Chem. Rev.* **105**, 827 (2005).
36. PDB searches were performed using the Web-based version of Relibase, http://relibase.ccdc.cam.ac.uk [Version 1.3.2 (Cambridge Crystallographic Data Centre, Cambridge, UK), August 2005].
37. L. Wollin, D. S. Bundschuh, A. Wohlsen, D. Marx, R. Beume, *Pulm. Pharmacol. Ther.* **19**, 343 (2006).
38. M. Morgenthaler et al., *ChemMedChem* **2**, 1100 (2007).
39. M. B. van Niel *et al.*, *J. Med. Chem.* **42**, 2087 (1999).
40. M. Rowley *et al.*, *J. Med. Chem.* **44**, 1603 (2001).
41. J. P. Snyder, N. S. Chandrakumar, H. Sato, D. C. Lankin, *J. Am. Chem. Soc.* **122**, 544 (2000).
42. A. M. Sun, D. C. Lankin, K. Hardcastle, J. P. Snyder, *Chem. Eur. J.* **11**, 1579 (2005).
43. H. J. Böhm *et al.*, *ChemBioChem* **5**, 637 (2004).
44. MedChem database version 06 (Daylight Chemical Information Systems, Aliso Viejo, CA, www.daylight.com/).
45. A. Avdeef, *Curr. Top. Med. Chem.* **1**, 277 (2001).
46. H. Fischer, M. Kansy, D. Bur, *Chimia* **54**, 640 (2000).
47. D. Alberati *et al.*, *Bioorg. Med. Chem. Lett.* **16**, 4311 (2006).
48. B. Fermini, A. A. Fossa, *Nat. Rev. Drug Discov.* **2**, 439 (2003).
49. C. Jamieson, E. M. Moir, Z. Rankovic, G. Wishart, *J. Med. Chem.* **49**, 5029 (2006).
50. D. D. Perrin, B. Dempsey, E. P. Serjeant, *pKa Prediction for Organic Acids and Bases* (Chapman & Hall, London, 1981).
51. H. H. Jensen, L. Lyngbye, A. Jensen, M. Bols, *Chem. Eur. J.* **8**, 1218 (2002).
52. J. A. Olsen *et al.*, *Angew. Chem. Int. Ed.* **42**, 2507 (2003).
53. J. Olsen *et al.*, *Org. Biomol. Chem.* **2**, 1339 (2004).
54. J. A. Olsen *et al.*, *ChemBioChem* **5**, 666 (2004).
55. A. Hoffmann-Röder *et al.*, *ChemMedChem* **1**, 1205 (2006).
56. E. Schweizer *et al.*, *Org. Biomol. Chem.* **4**, 2364 (2006).
57. U. Obst, D. W. Banner, L. Weber, F. Diederich, *Chem. Biol.* **4**, 287 (1997).
58. C.-Y. Kim *et al.*, *J. Am. Chem. Soc.* **122**, 12125 (2000).
59. V. M. Krishnamurthy *et al.*, *Chem. Asian J.* **2**, 94 (2007).
60. T. H. Maren, C. W. Conroy, *J. Biol. Chem.* **268**, 26233 (1993).
61. J. D. Dunitz, R. Taylor, *Chem. Eur. J.* **3**, 89 (1997).
62. Z. Wang *et al.*, *Structure* **6**, 1117 (1998).
63. F. Hof, D. M. Scofield, W. B. Schweizer, F. Diederich, *Angew. Chem. Int. Ed.* **43**, 5056 (2004).
64. J. Hernandez-Trujillo, A. Vela, *J. Phys. Chem.* **100**, 6524 (1996).
65. A. Matsushima, T. Fujita, T. Nose, Y. Shimohigashi, *J. Biochem. (Tokyo)* **128**, 225 (2000).
66. C.-Y. Kim, P. P. Chandra, A. Jain, D. W. Christianson, *J. Am. Chem. Soc.* **123**, 9620 (2001).
67. E. A. Meyer, R. K. Castellano, F. Diederich, *Angew. Chem. Int. Ed.* **42**, 1210 (2003).
68. MacPyMol v0.99 [W. L. DeLano, PyMOL Molecular Graphics System (DeLano Scientific, Palo Alto, CA, 2006, www.pymol.org)].

69. This work was supported by the ETH Research Council and F. Hoffmann–La Roche.

Supporting Online Material

www.sciencemag.org/cgi/content/full/317/5846/1881/DC1
Figs. S1 to S5
Tables S1 and S2
References

10.1126/science.1131943

Fluorine in Pharmaceuticals: Looking Beyond Intuition

Klaus Müller, Christoph Faeh and François Diederich

Science **317** (5846), 1881-1886.
DOI: 10.1126/science.1131943