Fluorine Bonding − How Does It Work In Protein−Ligand Interactions?

Peng Zhou,[†,‡] Jianwei Zou,[‡] Feifei Tian,[§,#] and Zhicai Shang*[,†]

Department of Chemistry, Zhejiang University, Hangzhou 310027, China, Key Laboratory for Molecular Design and Nutrition Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China, College of Bioengineering, Chongqing University, Chongqing 400044, China, and Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, Florida 32611

Received July 7, 2009

Although fluorination of pharmacologically active compounds has long been a common strategy to increase their metabolic stability and membrane permeation, the functionality of protein−ligand interactions involving fluorine atoms (fluorine bonding) was only recently recognized in the chemistry and biology communities. In this study, the geometric characteristics and the energetic behaviors of fluorine bonding were systematically investigated by combining two quite disparate but complementary approaches: X-ray structural analysis and theoretical calculations. We found that the short contacts involving fluorine atoms (generalized fluorine bonding) between proteins and fluorinated ligands are very frequent, and these contacts, compared to those routine hydrogen/halogen bonding, are more similar to sulfur-involved hydrogen bonding observed in proteins. ONIOM-based quantum mechanics/molecular mechanics analysis further revealed that fluorine bonding does play an essential role in protein−ligand binding, albeit the strength of isolated fluorine bonding is quite modest. Furthermore, 14 quantum mechanics (QM) and molecular mechanics (MM) methods were performed to reproduce fluorine bonding energies obtained at the rigorous MP2/aug-cc-pVDZ level of theory, and the results showed that most QM and very few MM methods perform well in the reproducibility; the MPWLYP functional and MMFF94 force field are recommended to study moderate and large fluorine bonding systems, respectively.

INTRODUCTION

Inter- and intramolecular interactions involving halogens in supra- and biomolecular systems have been studied extensively over the past several years. The well-known halogen bonding, which was shown to be parallel with classical hydrogen bonding, plays an important role in molecular recognition and assembly.[1] Fluorine is an unique halogen element which does not participate in routine halogen bonding and hardly acts as a hydrogen-bond acceptor,[2,3] however, Cambridge structural database surveys uncovered that there are a considerable number of short contacts between covalent fluorine and other atoms including H, C, O, and even F.[4−6] Theoretical and experimental studies also revealed a versatility of fluorine in the interaction with various types of atoms and groups,[7−9] this peculiarity was ingeniously utilized by chemists and biologists to control crystal aggregations[10] and to design functional peptides and proteins.[11]

Numerous works have been addressed for ascertaining the nature of noncovalent interactions involving fluorine atoms, and various mechanisms were proposed by different investigators to elucidate such interaction behaviors (Table 1). Borisenko et al. early suggested the electrostatic nature of F···Se interaction observed for an organic selenide in the gas phase,[12] and, on the other hand, Iwaoka et al. characterized orbital interactions associated with F approaching to Se by using NMR and orbital analysis.[13] Matta et al. employed electronic topology theory to analyze bond path linking to saturated fluorine atoms and concluded that the F···F bonding is shown to exhibit all the hallmarks of a closed-shell weak interaction.[14] Lee et al. highlighted the importance of head-to-tail and antiparallel dipolar interactions involving fluorine in crystal packing.[15] Olsen et al. described an orthogonal multipolar interaction of the ligand C−F bond with the backbone amide of thrombin at the active site.[16] Plenio pointed out that the organofluorine, like oxygen, is capable of participating in coordination interactions with metal ions, such as Na^+, K^+, Rb^+, and Ca^{2+}.[17] In addition, the hydrogen bonding accepted by fluorine was also investigated thoroughly, from geometric characteristics to energetic profile.[3,18−21] The results showed that the fluorine is a

Table 1. Diverse Fluorine Bonding

interaction type	example	interaction energy, kcal/mol
orbital interaction[13]	C−F···Se	∼ −1
closed-shell interaction[14]	C−F···F−C	∼ −14
antiparallel dipolar interaction[15]	C−F−C−F	NR[a]
orthogonal multipolar interaction[16]	C−F···C(N)═O	−0.2 to −0.3
coordination interaction[17]	C−F···K^+	NR[a]
hydrogen bonding[19]	C−F···H−O	∼ −3
halogen bonding[23]	CH_2F^+···NH_3	−0.13 to −9.30

[a] NR: not reported.

* Corresponding author. Telephone: +86-0571-87952379. Fax: +86-0571-87951895. E-mail: shangzc@zju.edu.cn.
† Department of Chemistry, Zhejiang University.
‡ Ningbo Institute of Technology, Zhejiang University.
§ College of Bioengineering, Chongqing University.
# Department of Chemistry, University of Florida.

10.1021/ci9002393 CCC: $40.75
Published on Web 09/29/2009

bad hydrogen-bond acceptor with interaction energies at most 3 kcal/mol.[20,21] Recently, Politzer et al. provided computational evidence that the covalent fluorine atoms can form typical halogen bonds with electron donors, if their bonded atoms/groups are sufficiently electron withdrawing.[22] This finding was further confirmed in our study.[23] This phenomenon can be explained by the σ-hole theory, which claims that many covalently bonded atoms of Groups V−VII possess an electron-deficient outer lobe of a half-filled p (or nearly p) orbital, giving rise to an electropositive region on the outermost surfaces of these atoms.[24−28] Considering the diverse appearances that fluorine-involved nonbonding interactions present, here we use the term fluorine bonding to underline the peculiarity of the fluorine atoms imposing on these bonding behaviors.

Fluorination of pharmacologically active compounds has long been a common strategy to increase their metabolic stability and membrane permeation.[29,30] Since the 1950s, over 150 fluorinated drugs have come to market and now make up ∼20% of all pharmaceuticals, with even higher figures for agrochemicals (up to 30%).[31,32] Traditionally, pharmacologists are mainly concerned on fluorine's effects on pharmacokinetic properties of drugs. However, recent structure determinations of protein complexes with bound fluorinated ligands have unraveled that the fluorine atoms can also directly participate in nonbonding protein−ligand interactions. For example, a representative example of a very short orthogonal C−F···C═O interaction in the complex of the potent serine protease inhibitor ZK-807834 bound to the active site of its biological target, factor Xα (PDB entry: 1FJS), was described by Paulini et al.[33] Moreover, fluorine substitution can significantly alter the electronic distribution of the aromatic ring and, hence, indirectly influences the binding affinity of receptor−ligand complexes. This conjecture was experimentally proved by Kim and co-workers by systematically manipulating fluorine atoms on the benzyl ring of *N*-(4-sulfamylbenzoyl)benzylamine, a selective carbonic anhydrase II inhibitor.[34]

When compared to small molecules, much less is known about the roles that fluorine-involved nonbonding interactions play in biological assemblies. To better understand the significance of fluorine bonding in drug affinity and selectivity to biomolecules, in the current study, we launched a comprehensive investigation on the short contacts involving fluorine atoms between proteins and their ligands by surveying all high-resolution crystal structures of protein−ligand complexes deposited in the Protein Data Bank (PDB)[35] and by performing systematic calculations on fluorine-containing model and real systems at quantum mechanics (QM), molecular mechanics (MM), and combination (QM/MM) levels. We also evaluated the performance of several sophisticated MM and QM methods in comparison with the results obtained from the accurate but expensive correlation ab initio level. This study would help to clarify how fluorine bonds participate in ligand binding and how they can potentially be used in rational drug design.

MATERIALS AND METHODS

Database Survey. On May 6, 2009, there were 6 466 ligand types and 190 928 models recorded in PDB-Ligand database, version 1.4.[36] The structures of these ligands in complex with receptors were extracted from the PDB from which we only considered the X-ray crystal structures of protein−small ligand complexes solved at 3 Å or better. Using the 3 Å as a cutoff was due to Auffinger et al. demonstrating that this resolution level is enough to provide a statistically significant analysis on the geometric characteristics of halogen-involved nonbonding interactions in biomolecules.[37] Further, all selected complex structures were subjected to a pretreatment procedure, that is, (i) removing water molecules and metal ions, (ii) repairing missing side chains of protein residues, using the newly released SCWRL4 program,[38] and (iii) adding hydrogen atoms for all protein and ligand heavy atoms, using the REDUCE program.[39] REDUCE was adopted here because this program was tested in our previous study to be capable of precisely reproducing the hydrogen positions measured by neutron diffraction[40] and also because a HET dictionary was released with this program to assist adding hydrogen atoms to nonstandard amino acid atoms, such as small ligands or modified residues.

From these prepared structures, we culled those with F···X (where the F belongs to ligands, and the X is the H, C, N, O, or S atom of proteins) distances that are shorter than the sum (R) of their respective van der Waals radii (r) [with $r_F = 1.47$, $r_H = 1.20$, $r_C = 1.70$, $r_N = 1.55$, $r_O = 1.52$, and $r_S = 1.80$ Å, leading to R(F···H) = 2.67, R(F···C) = 3.17, R(F···N) = 3.02, R(F···O) = 2.99, and R(F···S) = 3.27 Å][41] to define a distinct data set of generalized fluorine bonding.

Theoretical Calculations on Model Systems. Møller−Plesset second-order perturbation theory (MP2)[42] in conjunction with the Dunning's augmented correlation consistent basis set, aug-cc-pVDZ,[43] were employed to accurately determine fluorine-bonding energies of model systems. This level of theory was thought to be adequate for reasonably weak nonbonding interactions and was used in recent theoretical studies of hydrogen- and halogen-bonded complexes formed between fluorides and their binding partners.[22,23,44] In order to comprehensively investigate the performance of lower-level theories in reproducing fluorine-bonding energies obtained at the expensive MP2/aug-cc-pVDZ level, other methods, including two density functional theories (B3LYP and MPWLYP), one noncorrelation ab initio theory (HF), two semiempirical methods (AM1 and PM3), and nine MM methods (UFF, Dreiding, MM2, MM3, MM+, CFF, MMFF94, OPLS-AA, and CHARMm) were tested by performing calculations on 14 types of short fluorine-bonding contacts observed in protein−ligand interactions. The contact energies (ΔE_{ce}) calculated using QM methods were obtained under the indirect supermolecule approach,[45] this method considers the difference between the total energy of the complex and the sum energies of isolated monomers, viz. $\Delta E_{ce} = E_{complex} - (E_{monomer1} + E_{monomer2})$, whereas the MM-based ΔE_{ce} was determined directly by nonbonding terms (i.e., Coulomb and van der Waals terms) of force field functions. Basis set superposition error (BSSE) associated with the ΔE_{ce} calculated by MP2, B3LYP, MPWLYP, and HF were corrected by means of the counterpoise strategy.[46]

ONIOM-Based QM/MM Calculations on Real Systems. To reduce computational cost, two-layered ONIOM-based QM/MM calculations were carried out for the real fluorine bonding systems, i.e. whole protein−ligand complex structures extracted from PDB.[47] The ligands and the correspond-

ing protein residues that possibly form fluorine bonds with the ligands were included in the QM layer and treated at a high level of DFT calculation (MPWLYP/6-31G(d)),[48] while the rest atoms were included in the MM layer and treated at a low level of MM calculation (AMBER parm96).[49] The MPWLYP functional was proved to provide accuracies close to a high-level MP2 method and has been successfully used in a very recent QM/MM analysis of halogen bonds formed between proteins and their ligands.[50] In the MM layer, water molecules were described by the TIP3P model (unlike database survey, water molecules were kept in the QM/MM calculation procedure),[51] and the restricted electrostatic potential (RESP) fitting procedure was employed to obtain partial atomic charges of the ligands and the nonstandard amino acid atoms.[52] Parameters that are not found in standard AMBER force field are defined using the generalized amber force field (GAFF).[53]

First, structures of the whole protein−ligand complex systems were fully optimized based on the two-layer ONIOM protocol described above without any constraints. Then the interaction energy (ΔE_{int}) between protein and ligand in the optimized complex was calculated according to the strategy proposed by Alzate-Morales et al.,[54] this was accomplished by performing single-point energy computations twice, one on the complex system ($E_{complex}$) and the other on the same system, but its members (protein and ligand) were separated by several angstroms of distance from each other ($E_{complex'}$). In this way, we derived that the $\Delta E_{int} = E_{complex} - E_{complex'}$. For the single-point energy computation procedure, the QM layer was extended to the residues within 6 Å distance from ligand and treated at a higher level.

RESULTS AND DISCUSSION

General Statistics of Short Contacts Involving Fluorine Atoms. The PDB database contains 655 entries of crystal protein−ligand complexes (with resolutions of 3 Å or better) showing 2 476 C−F bonds in the ligands (~70% are alkyl fluorine and other are aromatic fluorine). These covalent fluorine atoms were further observed to be in short contact (distance is shorter than the sum of respective van der Waals radii) with protein atoms at 2 966 times, as listed in the Supporting Information, Table S1. Here we classified protein atoms into 14 types in terms of their chemical properties and hybridization states to ascertain the propensity of ligand fluorine atoms approaching to different protein atoms. Atomic classifications and the frequency of their participation in short contact with fluorine are summarized in Table 2. This table also lists the contact rate (the ratio of those in short contact with fluorine to all presented in our data set) for each protein atomic type. As can be seen, some ubiquitous atomic types in proteins, such as sp3 carbon (CT) and sp2 nitrogen in amides (N), show a lower preference to interact with fluorine. This is expected because these atoms are shielded compactly by their bonded atoms or/and possess a large electron density to repulse the approaching electronegative fluorine atoms. However, there are also few exceptions exhibiting an abnormally high propensity in short contact with fluorine; a typical case is the sp2 oxygen in carboxyl group (OC), a negatively charged atom in native proteins. According to our statistics, 1.455‰ of OC are in short contact with fluorine, this rate is even higher than that of positively charged nitrogen (NC) − a considerably fluorophilic atom described by Müller et al.[31] It can be explained by the fact that the carboxyl groups frequently occur in the ligand binding sites, and the naked OC is completely exposed to those vicinal fluorine atoms (this is unlike NC which bonded with two hydrogen atoms). Not surprisingly, most ligand C−F bonds do not approach carboxyl O−C bonds in a head-to-head manner, alternatively, these contacts are mostly orthogonal − by this way the electrostatic repulsive effect between them can be largely eliminated. In addition, the short C−F···C═O contacts (where the C in C═O is sp2 carbon in amides) highlighted by Paulini et al.[33,55] were not found to be particularly frequent in protein−ligand interactions. Conversely, the usually neglected nonpolar hydrogen (H) seems to be quite active in the interaction with the fluorine atoms.

Table 2. Classification of Protein Atoms and the Times and Contact Rates of Them in Short Contact with the Ligand Fluorine Atoms

type[a]	description	times in short contact	contact rate (‰)
CT	sp3 carbon	484	0.403
C	sp2 carbon in amides	98	0.233
CC	sp2 carbon in carboxyl group	23	0.499
CA	sp2 carbon in aromatic ring	186	0.671
CZ	sp2 carbon in guanidinium group of arginine	21	1.100
N	sp2 nitrogen in amides	119	0.282
NA	sp2 nitrogen in aromatic rings	36	1.275
NC	positively charged nitrogen	70	0.898
O	sp2 oxygen in amides	76	0.180
OC	sp2 oxygen in carboxyl group	134	1.455
OH	sp3 oxygen in hydroxyl group	108	1.738
S	sp3 sulfur	32	2.088
H	nonpolar hydrogen	1 287	0.728
HP	polar hydrogen	292	0.511

[a] Atomic type symbols are modified from the definition of AMBER.[49]

Geometric Characteristics of Short Contacts Involving Fluorine Atoms. The geometric characteristics of fluorine-involved short contacts (C−F···X, where X = C, N, O, S, and H) observed in our data set are characterized using two parameters: one is contact distance (d) indicating the separation between F and X, and other is contact angle (θ) reflecting the degree of C−F···X deviating from linear arrangement. The polar scatter plots of θ vs d for the 2 966 short contacts are separately shown in Figure 1a−d in terms of their acceptor types. The θ distribution presents a clear propensity with most samples >80°, indicating that the fluorine-involved interactions, like hydrogen bonds, have strong directionality. In addition, except for hydrogen as acceptors, the θ of some short contacts are remarkably small and fall into the region ranging from 30 to 70°. By visually examining these small-angle contacts, we noticed that most of them are typical antiparallel dipolar interactions, this interaction mode was also observed in organic complexes.[15] Figure 2a illustrates the histogram distributions of the θ respecting different acceptor types. It can be seen that the θ distribution of fluorine-involved short contacts exhibits pronounced Gaussian modality, with fitting center and deviation falling into the ranges of 118.4−130.4° and 38.0−91.7°. Compared to C−F···H and C−F···C, the θ of C−F···O/S and C−F···N are distributed wider and have no significant preferences;

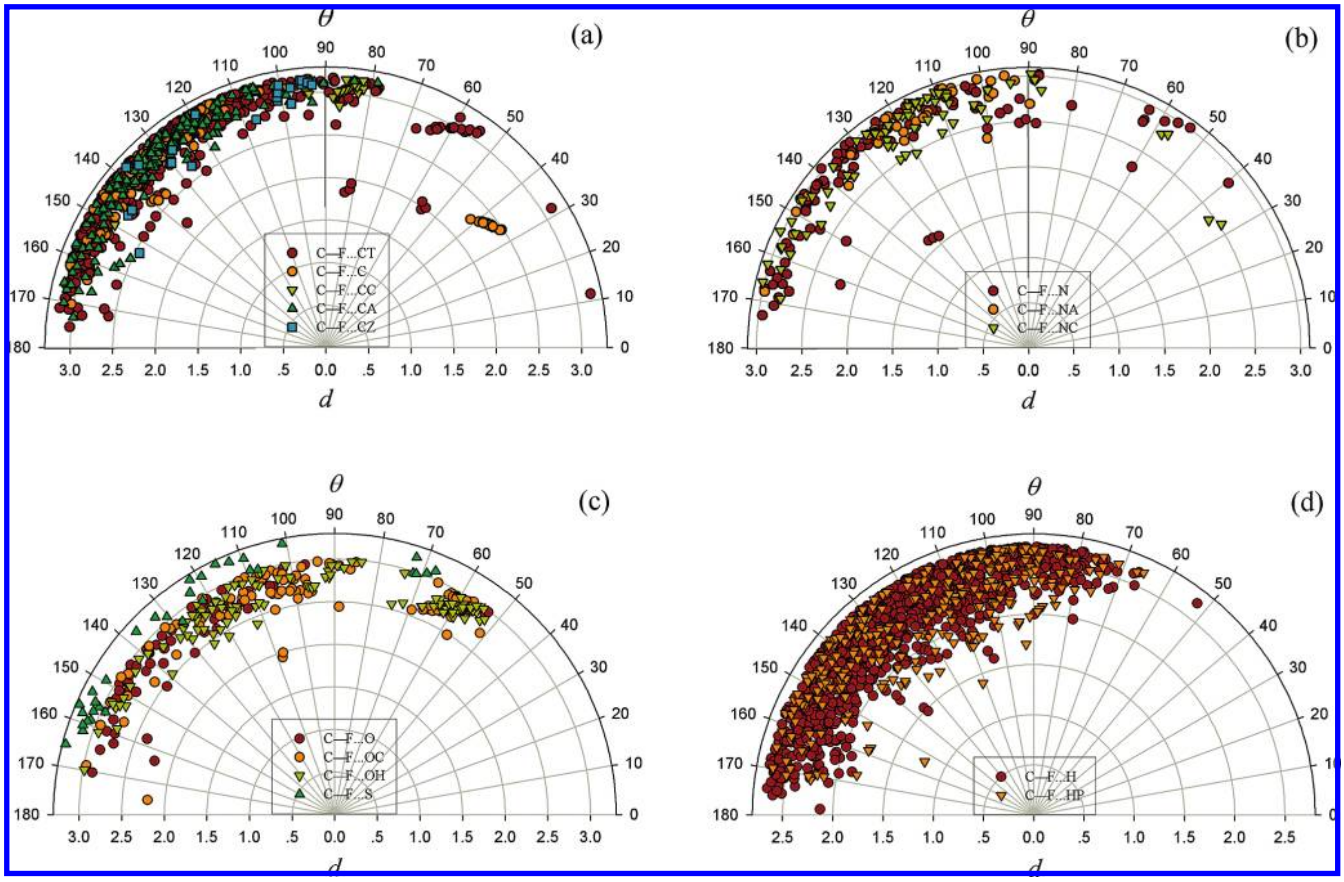

Figure 1. Polar scatter plots of θ vs d for fluorine-involved short contacts with different acceptors. (a) Fluorine−carbon contacts (C−F···C). (b) Fluorine−nitrogen contacts (C−F···N). (c) Fluorine−oxygen or sulfur contacts (C−F···O/S). (d) Fluorine−hydrogen contacts (C−F···H).

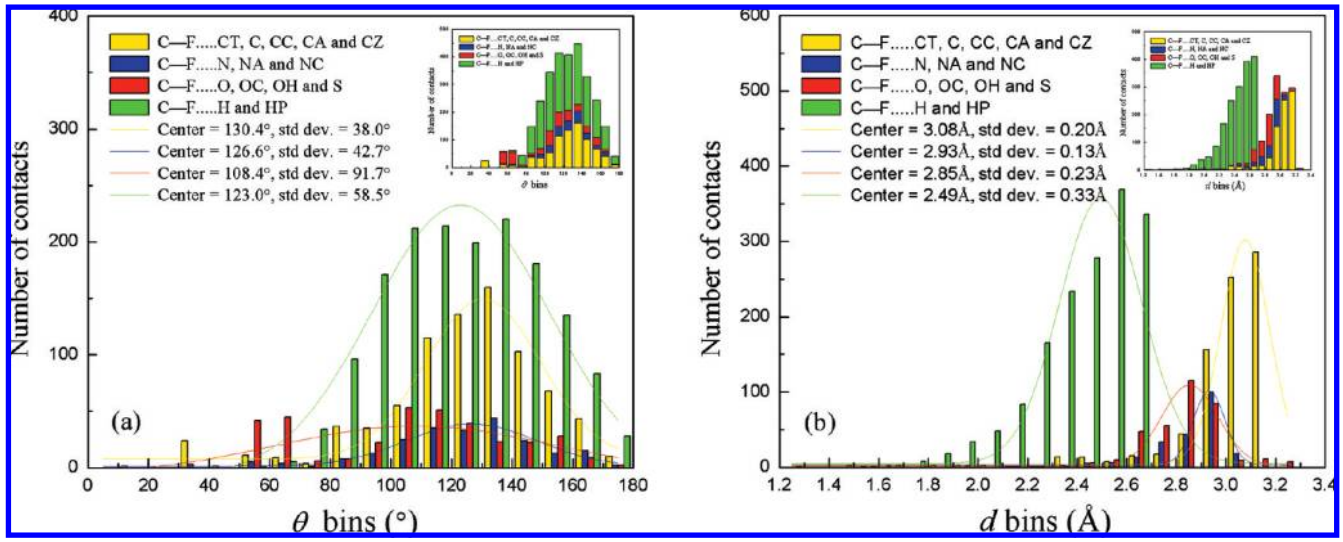

Figure 2. Histogram distributions of (a) θ (in 10° bins) and (b) d (in 0.1 Å bins) for fluorine-involved short contacts with different acceptors. The distributions are fitted using standard Gaussian function, yielding the distributional center and standard deviation for θ and d. Corresponding cumulative distributions are shown in the build-in plots.

strong electronegativity attached to the oxygen/sulfur and nitrogen atoms gives rise to a repulsive effect (rather than an attractive effect) for the C−F···O/S and C−F···N interactions, so their contact orientations are more random than those of stable nonbonding attractions.

For the d, as seen in Figure 1, the most contacts are longer than 2.5 Å (for C−F···C, C−F···N, C−F···O/S) or 1.5 Å (for C−F···H), however, there are also few samples with an abnormally short d, only accounting for half of the sum of their van der Waals radii. We found that these abnormal samples can be divided into two cases to explain: one results from the compromise between global stability and local collision-free, and another is lead by experimental measurement errors (notice that not every atom in the crystal structures is located at a high precision). Figure 2b demonstrates that the d values also obey the Gaussian distribution,

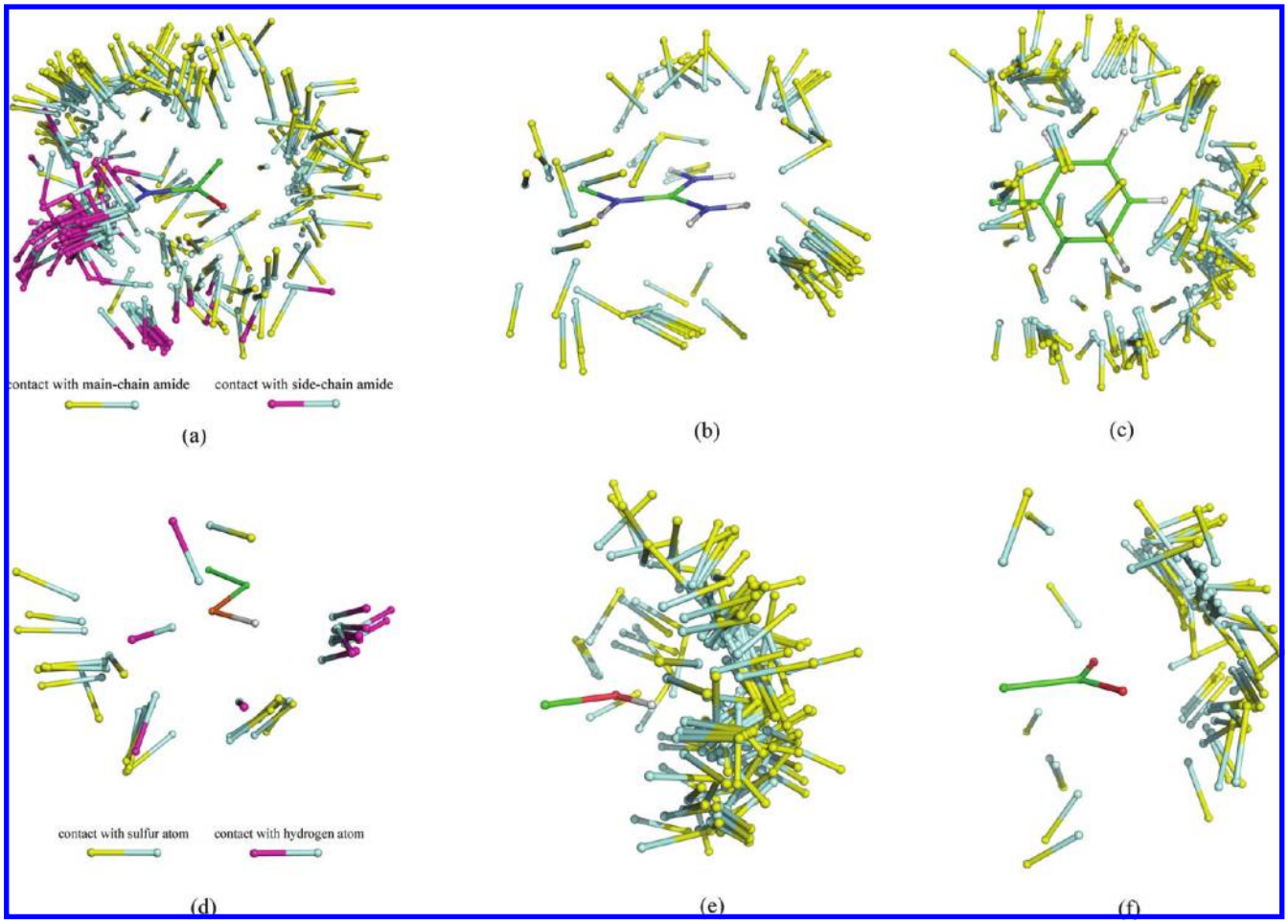

Figure 3. Superposition of ligand C−F bonds around different protein moieties. Fluorine and carbon atoms are colored cyan and yellow/pink, respectively. (a) C−F bonds around both main and side chain amides. (b) C−F bonds around the guanidinium group of arginine. (c) C−F bonds around the phenyl moiety of phenylalanine and tyrosine. (d) C−F bonds around the sulphydryl group of cysteine. (e) C−F bonds around the hydroxyl group of serine, threonine, and tyrosine. (f) C−F bonds around the carboxyl group of both glutamic and aspartic acid.

if ignoring the effect of an arbitrary cutoff on d. The centers and deviations of different contact types range from 2.49 to 3.08 Å and from 0.13 to 0.33 Å, respectively.

Generally speaking, geometric characteristics of short contacts involving fluorine atoms are very similar to that of sulfur-involved hydrogen bonds described in our previous report;[40] their θ can reach to a smaller value, and d varies more significantly than those routine hydrogen/halogen bonds observed in biomolecules.[37,56] These features suggest that the fluorine-involved interactions are relatively modest and may impose some subtle effects on protein−ligand recognition and association.

Ligand C−F Bonds around Protein Moieties. To investigate the interaction behaviors of ligand fluorine atoms with some important functional groups of proteins, we separately superposed the C−F bonds interacting with protein amides and phenyl moieties and with guanidinium, sulphydryl, hydroxyl, and carboxyl groups into Figure 3a−f (interaction is considered as the fluorine in short contact with any atom of a given group).

As seen in Figure 3a, the C−F bonds roughly isotropically surround the main-chain amides, only lack in the regions near to carbonyl O and Cα atoms (due to electrostatic repulsive and steric hindrance effects). In contrast, most C−F prefer to point frontally onto the side-chain $H_2N-C{=}O$ moiety, and a few others are more prone to laterally approach the N, C, and H atoms of the side-chain amides. Despite a moderate number of orthogonal C−F···C═O interactions observed in the survey, this interaction mode, according to our statistics, is not more preponderant than other short contacts between the ligand C−F and the protein amides. Our finding is distinct to that concluded by Paulini et al.,[33,55] we think this is due to the relative values that were emphasized in our survey, whereas the absolute magnitudes attracted more attention in Paulini et al.'s study.

The guanidinium group of arginine is a considerably fluorophilic group and can strongly attract negatively polarized fluorine atoms. As shown in Figure 3b, most C−F were observed to orient either parallel to or more orthogonally to the guanidinium plane with its delocalized positive charge, which is consistent with a previous report.[31] In addition, we also found that some C−F point toward the N of the $-NH_2$ moiety and form nonlinear hydrogen bonds with the H bonded to the N.

An aromatic ring is a π-electron-rich entity that hinders the C−F approaching to it in an orthogonal manner. In Figure 3c, only a few C−F are parallelly laid on the arene plane, while most of the others prefer to surround the phenyl group and interact with its C−H moieties. This observation agrees

Table 3. Contact Energies (ΔE_{ce}, in kcal/mol) Calculated Using Different QM MM Methods for the 14 Fluorine-Involved Contact Types Observed in Protein−ligand Interactions

	QM[a]						MM								
contact type	MP2[b]	B3LYP[c]	MPWLYP[c]	HF[c]	PM3	AM1	UFF	Dreiding	MM2	MM3	MM+	CFF	MMFF94	OPLS−AA	CHARMm
C−F···CT	−0.122	0.299	−0.055	0.548	0.069	0.262	−0.351	−0.116	−0.432	−0.164	−0.195	−0.044	−0.242	−0.191	−0.298
C−F···C	−0.155	0.338	−0.203	0.190	0.447	0.497	−0.028	−0.386	−0.312	0.165	−0.457	−0.033	−0.005	−0.080	−0.350
C−F···CC	5.016	6.134	5.381	6.976	4.903	5.267	−0.352	−0.319	−1.167	−0.793	−0.386	4.093	7.005	−0.524	−1.892
C−F···CA	0.096	1.434	0.762	1.995	1.088	1.639	−0.884	−0.777	−0.843	−0.091	−0.873	0.089	1.129	0.032	0.757
C−F···CZ	−6.538	−6.148	−6.998	−7.140	−4.648	−4.823	−0.662	−0.639	0.354	1.115	−0.688	−3.340	−5.721	0.939	3.271
C−F···N	−0.021	0.372	−0.340	0.565	0.296	0.246	−0.433	−0.300	−0.582	−0.056	−0.400	0.036	−0.044	−0.244	−0.283
C−F···NA	−0.652	0.307	−0.241	0.671	0.468	0.630	−0.732	−0.559	−0.443	0.056	−0.688	−0.018	−0.280	0.098	0.686
C−F···NC	−10.101	−10.300	−10.893	−11.135	−7.199	−8.708	−0.283	0.104	−6.411	−5.413	−0.071	−5.538	−10.010	−1.945	−12.694
C−F···O	1.779	2.554	1.872	3.119	1.555	1.850	−0.162	−0.154	1.178	1.121	−0.202	1.206	3.148	1.235	3.037
C−F···OC	5.513	6.790	5.979	8.098	4.799	7.680	−0.143	−0.138	2.535	3.622	−0.185	4.020	8.904	0.829	8.397
C−F···OH	1.535	2.267	1.791	2.994	0.963	1.421	−0.257	−0.234	0.283	0.928	−0.285	0.910	2.420	1.242	2.306
C−F···S	0.943	1.630	0.913	2.339	2.880	2.046	−0.273	−0.190	0.121	0.479	−0.270	0.390	1.662	0.351	1.180
C−F···H	−1.602	−1.700	−1.842	−1.697	−0.907	−1.027	−0.204	−0.058	−1.206	−1.183	−0.163	−1.276	−1.747	−1.336	−1.877
C−F···HP	−0.355	−0.035	−0.455	−0.022	−0.275	−0.290	−0.188	−0.010	−0.248	−0.118	−0.113	−0.096	−0.152	−0.217	−0.263

[a] Basis set superposition error (BSSE) associated with the contact energies calculated at the MP2, B3LYP, MPWLYP, and HF levels were corrected with the standard counterpoise method of Boys and Bernard.[46] [b] In conjugation with the Dunning's augmented basis set, aug-cc-pVDZ. [c] In conjugation with the Pople's standard basis set, 6-311++G(d,p).

with the repulsive nature of C−F···arene−π interaction with a slight energetic penalty of <1 kcal/mol.[57]

According to our statistics, sulfur has the largest contact rate (2.088 ‰) with the ligand fluorine in all protein atoms. In fact, all of sulfur atoms in short contact with fluorine belong to the sulphydryl group of cysteine (none of methionine). In Figure 3d, most C−F contacting with sulphydryl S are also in an interaction with the sulphydryl H. Apparently, by this way systemic destabilization caused by the electrostatic repulsion between the negatively polarized F and the lone pair of S can obtain an extra compensation from the F···H attractions. Moreover, there are also some C−F bonds pointing toward H−S and forming a weak hydrogen-bonding interaction with the H atoms.

The protein hydroxyl group is a typical fluorophilic moiety that attracts a considerable number of C−F to interact with it (Figure 3e). Although the nature of C−F···HO− interaction, like C−F···HS−, is hydrogen bonding, the interaction energy of C−F···HO− is moderately strong and can achieve an energy of −1.601 kcal/mol − nearly half of hydrogen-bonding energy in a standard water dimer (−3.2 kcal/mol).[58] Therefore, according to our observation, most C−F···HO− are close to a linear arrangement with an interacting distance $d \approx 2.1$ Å (slightly longer than the routine hydrogen bond =O···HN− of about 1.9 Å).[59]

The ionized carboxyl group possesses a negative formal charge and, thus, in theory, refuses the fluorine close to it. However, by the above discussion, the oxygen atoms of carboxyl group have an abnormally high propensity in short contact with fluorine. In Figure 3f, the C−F does not point to the carboxyl group by a head-to-head manner. Instead, they are prone to orthogonally interact with the carboxyl group, by this way, the electrostatic repulsion raised from C−F···O can be partially compensated by electrostatic attraction of F−C···O.

Contact Energy Analysis. Of course, we are aware that "atoms in crystals need to go somewhere" as stated by Dunitz, and that not every short contact may be an attractive one.[33] Although, generalized fluorine bonding can be regarded as having all interaction types, including attractive and repulsive ones, between the organofluorine and its vicinities, those thermodynamically stable interactions involving fluorine are more interesting for chemistry, biological, and pharmacological applications. Therefore, here we employed the rigorous MP2/aug-cc-pVDZ theory and the 14 lower-level methods to investigate the contact energies (ΔE_{ce}) of covalent fluorine with the 14 protein atomic types. ΔE_{ce} is defined as the minimum interaction energy of donor atom (F) separating from acceptor atom (X) at the van der Waals contact distance. Since ∼70% of organofluorine in ligands are alkyl fluorine, we used fluoromethane as the fluorine provider in the model systems. On the other hand, methane, formamide, ionized formic acid, benzene, methanol, methanethiol, and guanidinium cations were considered as the fluorine receivers for the model systems. All model dimers were arranged in low-energy conformations with the donor atom in the van der Waals contact with acceptor atom (see Supporting Information, Figure S1). ΔE_{ce} calculated using MP2/aug-cc-pVDZ as well as the other 14 methods are tabulated in Table 3.

The ΔE_{ce} based on MP2/aug-cc-pVDZ shows that most individual fluorine-involved interactions, whether repulsively or attractively, are quite modest, with their ΔE_{ce} falling into the range of −1.602−1.779 kcal/mol. In contrast, fluorine contacting with charged molecules, such as ionized formic acid and guanidinium cation, gives rise to a noticeably strong interaction where the ΔE_{ce} magnitudes are even larger than 10 kcal/mol. These phenomena indicate that fluorine bonding in protein−ligand interactions are mainly derived from electrostatic force, which is more like the halogen bonding than the hydrogen bonding (hydrogen bonding is imparted more covalent and polar component).[60]

Fluorine interacting with sp2 carbon in amides (C−F···C) has received a slight attraction (−0.155 kcal/mol by MP2), this interaction mode was called orthogonal multipolar interaction by Paulini et al.[33] and highlighted in contributing to the binding affinity of thrombin with its tricyclic inhibitors.[61,62] Experimental investigations of a model system in chemical double mutant cycles also revealed the attractive nature of C−F···C, with a contribution to binding free energy in nonpolar environments of −0.2 to −0.3 kcal/mol (note this value is very close to our computed result).[63]

Fluorine interacting with sp2 carbon in aromatic rings (C−F···CA) has been systematically investigated by both

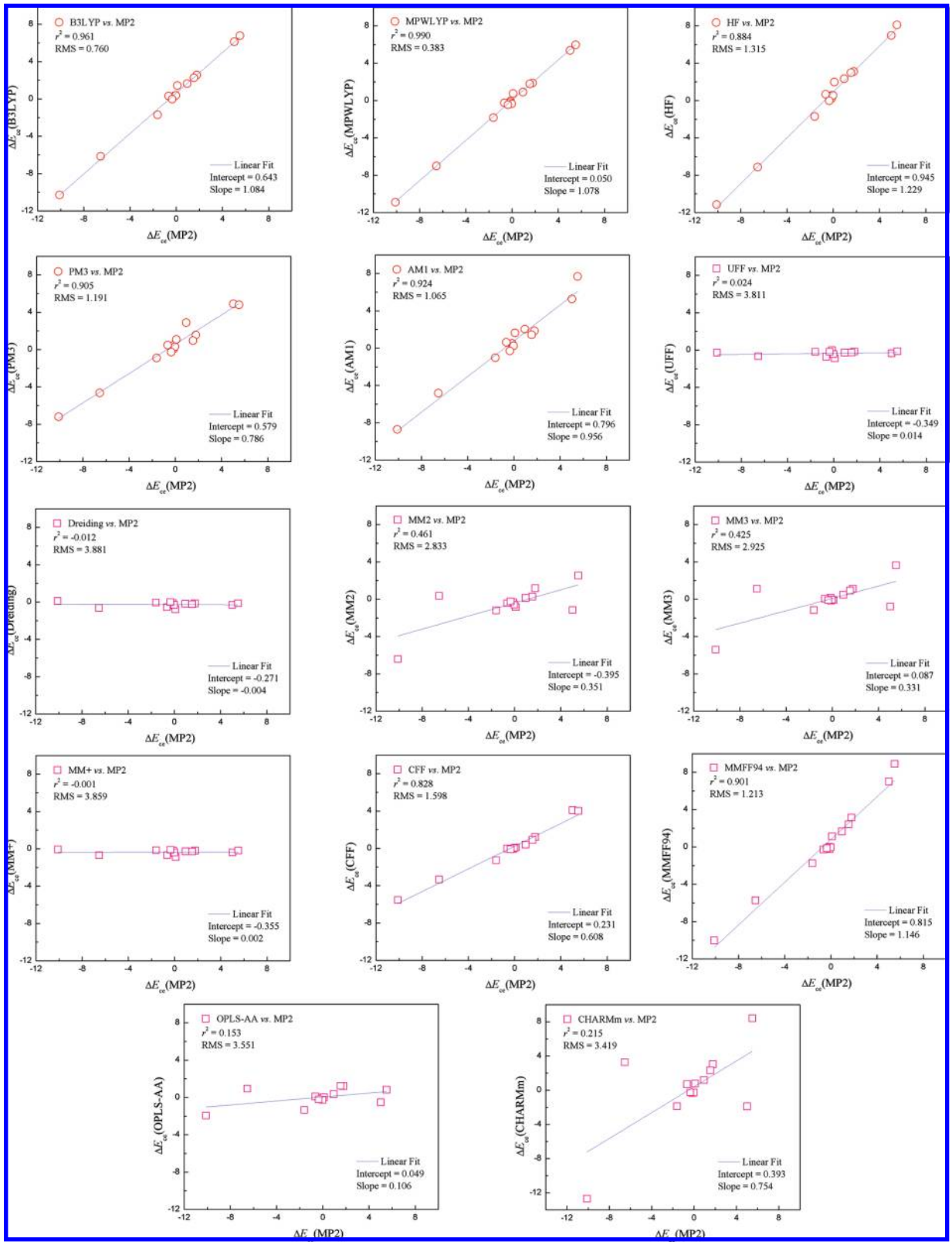

Figure 4. Plots of the contact energies (ΔE_{ce}) calculated using different lower-level methods against that derived from the rigorous MP2/aug-cc-pVDZ level of theory. The scatters of QM- and MM-based plots are represented in circle and square, respectively.

theoretical and experimental approaches,[57] results indicated the substantial repulsion of this interaction with an energy penalty of <1 kcal/mol. Our calculations confirmed this conclusion and further quantified the repulsive energy of 0.096 kcal/mol — this value is relatively small since the attractive effect between the fluorine and the adjacent hydrogen was also accounted into it.

Fluorine interacting with polar hydrogens (C−F···HP) is traditionally called hydrogen bonding.[3,18−21] MP2 calculations showed that this interaction is weaker than routine hydrogen bonding with its ΔE_{ce} of −1.602 kcal/mol. By natural bond orbital (NBO)[64] analysis of charge transfers (CTs) between the fluorine and the hydrogen, we found that only a very few electrons are transferred to fluorine from

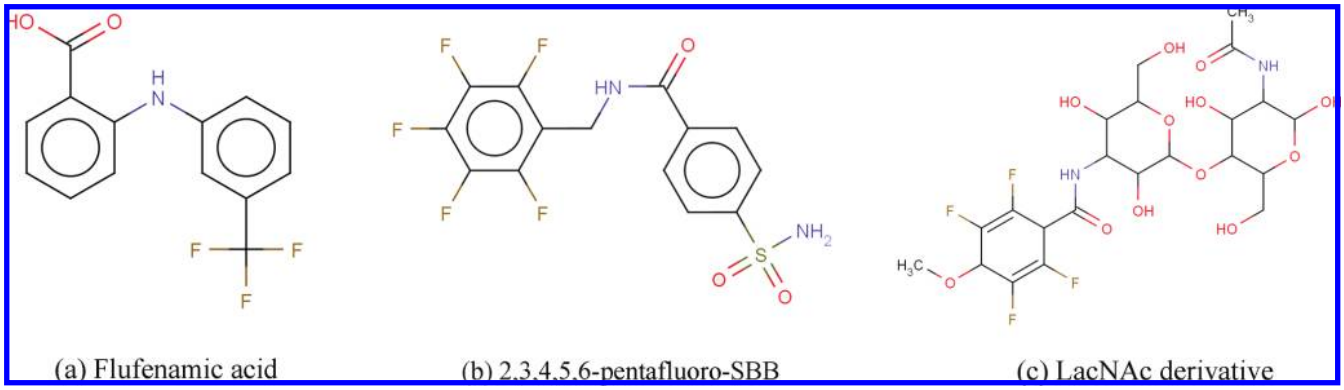

Figure 5. Chemical structures of the investigated ligands. (a) 2-[[3(trifluoromethyl)phenyl]amino]benzoic acid (flufenamic acid). (b) 4-(aminosulfonyl)-*N*-[(2,3difluorophenyl)methyl]benzamide (2,3,4,5,6-pentafluoro- SBB). (c) 2,3,5,6-tetrafluoro-4-methoxybenzamide-D-galactose-*N*-acetyl-D-glucosamine (LacNAc derivative).

hydrogen upon the bonding. Therefore, the term of dipolar interaction may be more suitable for the C−F···HP interactions. In addition, the contribution of fluorine interacting with nonpolar hydrogen (C−F···H) to protein−ligand binding was usually neglected in previous studies. However, a large number of such interactions observed in our data set (about half of ligand C−F bonds participate in C−F···H interactions) implied that the C−F···H may play some subtle roles in protein−ligand recognition, such as adjusting the ligand arrangement in an active pocket or improving the binding specificity of ligand to receptor, although its ΔE_{ce} is quite moderate at −0.355 kcal/mol.

The plots of ΔE_{ce} calculated using the 14 lower-level methods against that derived from the rigorous MP2/aug-cc-pVDZ level of theory are shown in Figure 4. As seen, the QM methods, especially DFT, perform very well in reproducing MP2-based ΔE_{ce}; the greatest root-mean-square error (rms) is only 0.383 kcal/mol, as obtained by the MPWLYP functional, which was tested also to be excellent in determining various halogen-bonding energies.[65] Not surprisingly, the ab initio HF theory performs worse than the semiempirical AM1 and PM3 methods, since the electron correlation neglected by HF is implicitly involved in the experimental data utilized for parametrizing semiempirical methods. In contrast, most MM methods are incapable of reproducing well the MP2-based ΔE_{ce}, and, worst, the correlation coefficients r^2 between the ΔE_{ce} obtained by general-purpose force fields UFF, Dreiding, and MM+, and those by MP2 are even near to 0. Exceptionally, the MMFF94 force field seems to be comparable in reproducibility to the QM methods, with corresponding r^2 and rms of 0.901 and 1.213 kcal/mol. These results suggest that the MPWLYP is a good choice for accurately determining the fluorine-bonding energy of moderate systems, while the PM3, AM1, and MMFF94 could be suitable candidates for large or quite large fluorine-bonding systems, such as fluorinated biomolecules and supramolecules.

QM/MM Analysis of Fluorine Bonds in Protein−ligand Complexes. To better understand the effects of fluorine bonding casting to real protein−ligand complex systems, X-ray structures of three fluorinated ligands, i.e. flufenamic acid,[66] 2,3,4,5,6-pentafluoro-SBB,[67] and LacNAc derivative[68] (their chemical structures are sketched in Figure 5), in complexes with their cognate receptors (PDB entries: 1BM7, 1G54 and 1KJR), were selected to perform ONIOM-based QM/MM calculations. For the purpose of comparison, the corresponding defluorinated complexes were also investigated herein, that is, the fluorine atoms of the native ligands were artificially mutated to hydrogen atoms.

The complex structures of these studied systems as well as their defluorinated ones were fully optimized using the two-layered QM/MM scheme, as described in the Materials and Methods section. The electrostatic interactions between QM and MM layers were treated in terms of the mechanical embedding strategy to save computational cost. The optimized results are illustrated in Figure 6, where the native ligands are shown in superposition with their defluorinated ones. It can be seen that the binding modes of different ligands within proteins are not ultimately changed by the fluorinations, but it should also be noted that the fluorine indeed plays a functional role in these complex systems due to their participation in nonbonding interactions with vicinal protein atoms, and these interactions, i.e., fluorine bonding, give rise to significant conformational perturbations on the local structures of ligands. For example, the trifluoromethyl moiety of flufenamic acid is twisted and the phenyl moieties of 2,3,4,5,6-pentafluoro-SBB and LacNAc derivatives are drawn by forming fluorine bonds.

To quantitatively measure the contribution of fluorine atoms to ligand binding, experimentally determined dissociation constants (K_d), QM/MM-calculated interaction energies (ΔE_{int}), and root mean-square derivation (rmsd) values of the native ligands, relative to their defluorinated ones, are summarized in Table 4. As can be seen, there is a strongly linear correlation between the ΔE_{int} and the negative logarithm K_d (pK_d), with correlation coefficient R of 0.954 (Figure 7), albeit the absolute magnitudes of ΔE_{int} are overestimated by QM/MM (according to $\Delta E_{int} \approx \Delta G = -RT \ln(1/K_d)$). The differences in the ΔE_{int} between the three native ligands and their defluorinated ones are −2.54, −8.26, and −4.85 kcal/mol. Note, these values are significantly larger than the interaction energy of isolated fluorine bonding, indicating that the changes in ligand affinity upon fluorination are not only contributed by the form of fluorine bonds but also a result of the indirect effects of fluorine altering the electron distribution of ligand molecules. It is worth noting that the fluorinated sites at the aromatic ring have a larger indirect effect than those at the alkyl group. The deviation of the structures of the optimized native ligands relative to defluorinated ones can be quantified by the computations of rmsd values. It can be seen from Table 4 that fluorine substitutions have only a slight influence on the global

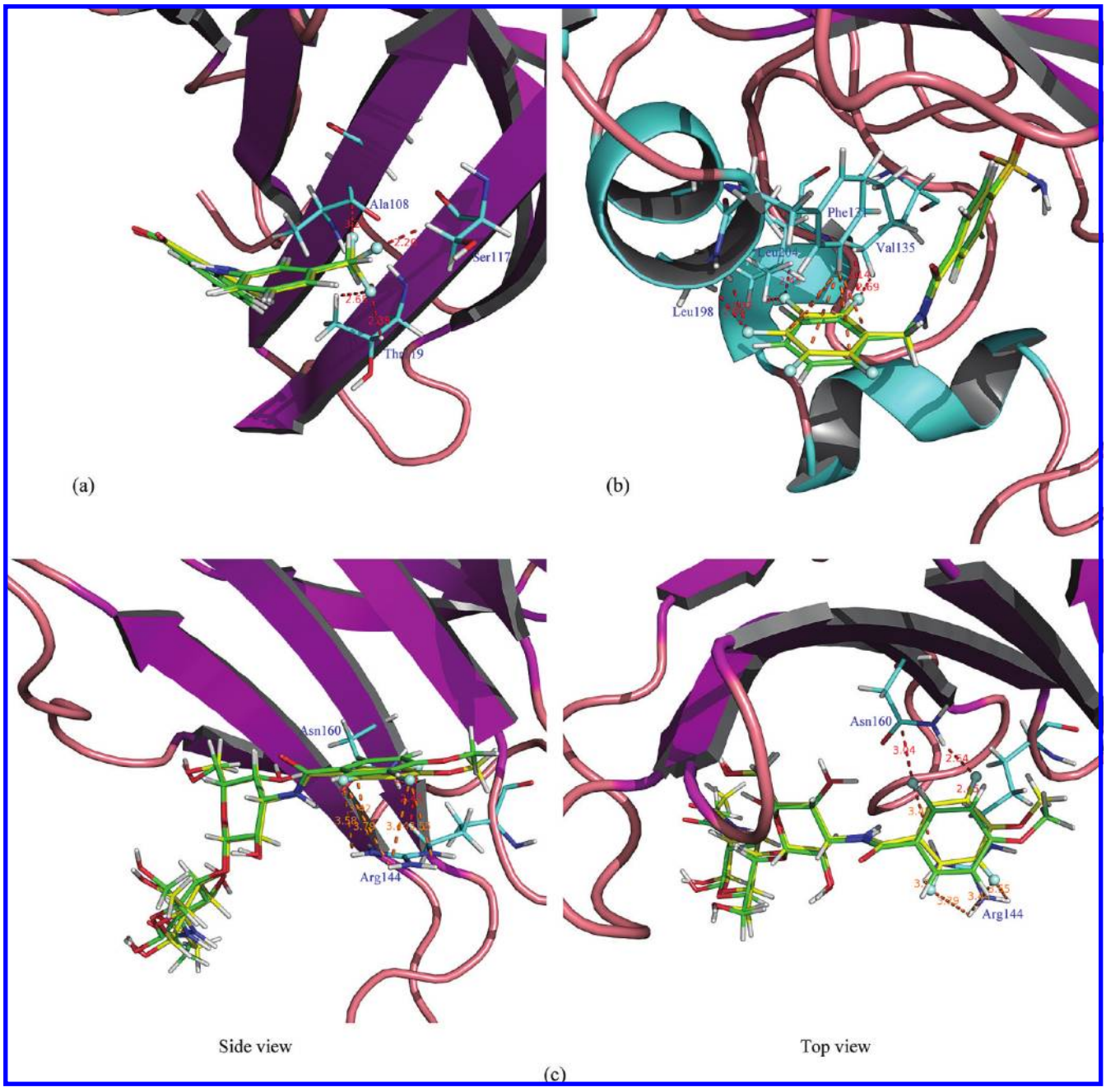

Figure 6. Superposition of the QM/MM-optimized native ligand structures (yellow) with their defluorinated ones (green) in protein active pockets (only the proteins in complex with native ligands are shown). In which, ligands and important residues are shown in stick style, fluorine atoms are represented as ball, and fluorine bonds and other key interactions are shown in broken line. (a) Flufenamic acid in complex with transthyretin (PDB entry: 1BM7). (b) 2,3,4,5,6-pentafluoro-SBB in complex with carbonic anhydrase II (PDB entry: 1G54). (c) LacNAc derivative in complex with human galectin-3 (PDB entry: 1KJR).

conformations of ligand molecules, with average atomic movements less than 1 Å. Nevertheless, these movements mainly concentrate on the fluorinated portions of the ligands (Figure 6), in this way the fluorine-bonding geometries are adjusted into a more favorable arrangement.

Transthyretin (TTR)−Flufenamic Acid Complex. Several human amyloid diseases, including familial amyloid polyneuropathy and senile systemic amyloidosis, are associated with the TTR amyloid fibrils. Flufenamic acid bound to TTR can inhibit the conformational changes of TTR and, thus, block the process of amyloid fibril formation.[66] According to QM/MM calculations, defluorination of flufenamic acid will produce an interaction energy loss of about 2.54 kcal/mol, which results from the breaking of four fluorine bonds: an orthogonal multipolar interaction and three dipolar interactions (Figure 6a). In Figure 6a, the trifluoromethyl moiety is remarkably rotated about the axis along with C−C bond by forming the fluorine bonds, indicating a stabilization contribution to the TTR−flufenamic acid complex. In addition, the dissociation constant K_d of defluorinated flufenamic acid was predicted to be 78.5 nM (no experimentally determined value available), which should be reliable since this point is a typical interpolation in the regression model (Figure 7).

Carbonic Anhydrase II (CAII)−2,3,4,5,6-Pentafluoro-SBB Complex. CAII is a zinc metalloenzyme that catalyzes the hydration of CO_2 to yield a bicarbonate and a proton. *N*-(4-sulfamylbenzoyl)benzylamine (SBB) is a tight-binding in-

Table 4. Summary of Energetic and Geometric Properties for the Investigated Ligands

ligand name	K_d[a]	ΔE_{int}[b]	rmsd[c]
Flufenamic acid	30.0 nM (78.5 nM[e])[d]	−22.01 (−19.47)[d]	0.512
2,3,4,5,6-pentafluoro-SBB	1.5 nM (2.1 nM)[d]	−32.76 (−24.50)[d]	0.664
LacNAc derivative	0.88 μM (3.6 μM)[d]	−14.48 (−9.63)[d]	0.838

[a] Experimentally measured dissociation constant (extracted from MOAD database).[69] [b] QM/MM-calculated interaction energy between the ligands and the proteins in kcal/mol. [c] Rmsd of the QM/MM-optimized ligand structures relative to the corresponding defluorinated ones in Å. [d] The values in parentheses are of corresponding defluorinated ones. [e] This value was predicted using the regression model.

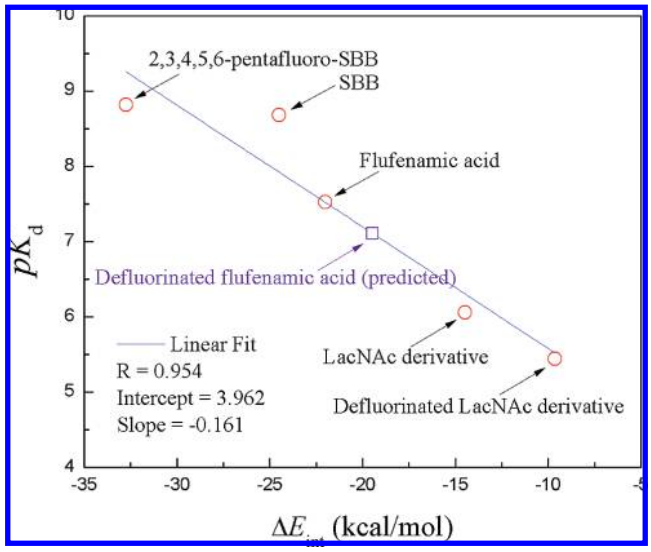

Figure 7. Correlation scatters between the experimentally measured affinity (pK_d) and the QM/MM-calculated interaction energy (ΔEint) for the investigated ligands binding to their cognate receptors.

hibitor of human CAII and can form a edge-to-face (or T-shape) interaction with CAII Phe131's side chain through its benzyl moiety. Kim and co-workers experimentally demonstrated that fluorination of the benzyl ring can slightly increase the binding affinity for ligands.[67] This observation contradicts the expected behavior of fluoronated aromatic systems, since fluorine substitution on π-donor aromatic rings reduces the overall quadrupole moment and, thus, reduces the strength of T-shape aromatic interactions.[8,57] By QM/MM analysis of the CAII-2,3,4,5,6−pentafluoro-SBB complex, we found that the ligand fluorination results in the establishment of six weak fluorine-bonding interactions with the nonpolar hydrogen atoms of Val135, Leu198, and Leu204 (Figure 6b), leading to a slight motion of the ligand's benzyl moiety toward Phe131 (this phenomenon was also observed by crystallographic experiment).[67] Therefore, we conclude that the fluorine bonding is mainly responsible for the increase of binding affinity, this effect is partially counteracted by weakening quadrupole−quadrupole interactions between the ligand's benzyl moiety and the CAII Phe131's side chain upon the fluorination of ligand. This conclusion was further confirmed by the fact that only a little change in experimental K_d value occurs from the fluorination (from 2.1 to 1.5 nM), if considering two contrary effects coexist in this process.

Human Galectin-3−LacNAc Derivative Complex. The galectins are a family of proteins defined by a carbohydrate recognition domain (CRD) of about 135 amino acids and have been clearly implicated in physiological and pathological mechanisms in inflammation, immunity, and cancer. The crystal structures of human galectin-3 were solved in complex with a series of fluorinated *N*-acetyllactosamine (LacNAc) derivatives and unraveled so that the ligands bound to galectin-3 were characterized by a cation−π interaction between the ligand's aromatic substituent and Arg144's guanidinium.[68] QM/MM analysis revealed a substantial motion of the aromatic substituent toward guanidinium upon fluorination (Figure 6c [side view]). This may be caused by the strong electrostatic interactions (charge-assisted fluorine bonding) between the negatively polarized fluorine and the positively charged hydrogen. Moreover, two fluorine atoms close to an active pocket form dipolar and multipolar interactions with Asn160, as a result, the aromatic moiety of ligand is drawn toward the protein in some degree (Figure 6c [top view]). According to calculations and experiments, fluorination can contribute −4.85 kcal/mol of interaction energy to the human galectin-3−LacNAc derivative binding and gives rise to 2.72 μM improvement in the affinity value (from 3.6 to 0.88 μM).

CONCLUDING REMARKS

The main goal of this study is to comprehensively understand the functionality of fluorine bonding in protein−ligand interactions. To achieve this, we have combined two quite disparate but complementary approaches, i.e., X-ray structural analysis and theoretical calculations, to systematically investigate the geometric characteristics and the energetic behaviors of fluorine bonding in crystal structures and in model/real systems. Here we conclude with the following remarks to close this article:

(i) Fluorine bonding behaves more similarly to hydrogen bonding involving sulfur atoms than routine hydrogen/halogen bonding from both geometry and energy points of view.

(ii) In comparison with other fluorine-involved short contacts, the orthogonal multipolar interactions of C−F···C═O, highlighted in previous studies, were not found to be particularly frequent in protein−ligand interactions.

(iii) A large number and moderate strength of C−F interacting with nonpolar hydrogen imparted that this previously neglected type of fluorine bonding may play some subtle roles in protein−ligand interactions.

(iv) Most QM methods perform well in determining fluorine-bonding energy, whereas most MM methods are incapable of properly treating fluorine bonding. The MPW-LYP functional and the MMFF94 force field are recommended to study moderate and quite large fluorine-bonding systems.

(v) Albeit the strength of isolated fluorine bonding is quite modest, fluorination of ligand molecules can contribute significantly to binding affinity and to specificity from two aspects as direct fluorine-bonding interactions and indirect electronic effects.

ACKNOWLEDGMENT

This work was supported by the State Key Laboratory of Trauma, Burns, and Combined Injury Foundation (no.

SKLKF200904) and the Natural Science Foundation of Ningbo (no. 2008A610068).

Supporting Information Available: The conformations of model dimers arranged to low-energy conformations with the fluorine atom in van der Waals contact with acceptor atom. Summary of the short contacts between the ligand's fluorine atoms and the protein atoms extracted from the PDB-Ligand database and Protein Data Bank (May 2009 release) are tabulated. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Metrangolo, P.; Neukirch, H.; Pilati, T.; Resnati, G. Halogen bonding based on recognition processes: a world parallel to hydrogen bonding. *Acc. Chem. Res.* **2005**, *38*, 386–395.

(2) Politzer, P.; Lane, P.; Concha, M. C.; Ma, Y.; Murray, J. S. An overview of halogen bonding. *J. Mol. Model.* **2007**, *13*, 305–311.

(3) Dunitz, J. D.; Taylor, R. Organic fluorine hardly ever accepts hydrogen bonds. *Chem.—Eur. J.* **1997**, *3*, 89–98.

(4) Alkortaa, I.; Rozasa, I.; Elgueroa, J.; Foces-Foces, C.; Canob, F. H. A statistical survey of the Cambridge structural database concerning density and packing. *J. Mol. Struct.* **1996**, *382*, 205–213.

(5) Ahuja, R.; Samuelson, A. G. Non-bonding interactions of anions with nitrogen heterocycles and phenyl rings: a critical Cambridge structural database analysis. *CrystEngComm* **2003**, *5*, 395–399.

(6) Awwadi, F. F.; Willett, R. D.; Peterson, K. A.; Twamley, B. The nature of halogen···halogen synthons: crystallographic and theoretical studies. *Chem.—Eur. J.* **2006**, *12*, 8952–8960.

(7) Kawahara, S.; Tsuzuki, S.; Uchimaru, T. Theoretical study of the C-F/π interaction: attractive tnteraction between fluorinated alkane and an electron-deficient π-system. *J. Phys. Chem. A* **2004**, *108*, 6744–6749.

(8) Riley, K. E.; Merz, K. M. Effects of fluorine substitution on the edge-to-face interaction of the benzene dimer. *J. Phys. Chem. B* **2005**, *109*, 17752–17756.

(9) Chopra, D.; Nagarajan, K.; Row, T. N. G. Analysis of weak interactions involving organic fluorine: insights from packing features in substituted 4-keto-tetrahydroindoles. *J. Mol. Struct.* **2008**, *888*, 70–83.

(10) Vishnumurthy, K.; Row, T. N.; Venkatesan, K. Fluorine in crystal engineering: photodimerization of (1E,3E)-1-phenyl-4-pentafluorophenylbuta-1,3-dienes in the crystalline state. *Photochem. Photobiol. Sci.* **2002**, *1*, 427–430.

(11) Jäckel, C.; Koksch, B. Fluorine in peptide design and protein engineering. *Eur. J. Org. Chem.* **2005**, *21*, 4483–4503.

(12) Borisenko, K. B.; Broschag, M.; Hargittai, I.; Klapö tke, T. M.; Schröder, D.; Schulz, A.; Schwarz, H.; Tornieporth-Oetting, I. C.; White, P. S. Preparation of $N(SeCl)_2^+X^-$ (X = $SbCl_6$ or $FeCl_4$), $F_3CCSeNSeCCF_3^+SbCl_6^-$, $F_3CCSeNSeCCF_3$, $F_3CCSeNSeCCF_3$ and $F_3CCSeSeC(CF_3)C(CF_3)SeSeCCF_3$. Electron diffraction study of $F_3CCSeSeCCF_3$ and crystal structure of the eight-membered heterocycle $F_3CCSeSeC(CF_3)C(CF_3)SeSeCCF_3$. *J. Chem. Soc., Dalton Trans.* **1994**, 2705–2712.

(13) Iwaoka, M.; Komatsu, H.; Katsuda, T.; Tomoda, S. Quantitative evaluation of weak nonbonded Se···F interactions and their remarkable nature as orbital interactions. *J. Am. Chem. Soc.* **2002**, *124*, 1902–1909.

(14) Matta, C. F.; Castillo, N.; Boyd, R. J. Characterization of a closed-shell fluorine-fluorine bonding interaction in aromatic compounds on the basis of the electron density. *J. Phys. Chem. A* **2005**, *109*, 3669–3681.

(15) Lee, S.; Mallik, A. B.; Fredrickson, D. C. Dipolar-dipolar interactions and the crystal packing of nitriles, ketones, aldehdyes, and C(sp2)-F groups. *Cryst. Growth Des.* **2004**, *4*, 279–290.

(16) Olsen, J. A.; Banner, D. W.; Seiler, P.; Sander, U. O.; D'Arcy, A.; Stihle, M.; Müller, K.; Diederich, F. A fluorine scan of thrombin inhibitors to map the fluorophilicity/fluorophobicity of an enzyme active site: evidence for C-F···C═O interactions. *Angew. Chem., Int. Ed.* **2003**, *42*, 2507–2511.

(17) Plenio, H. The Coordination chemistry of fluorine in fluorocarbons. *ChemBioChem* **2004**, *5*, 650–655.

(18) Kovács, A.; Varga, Z. Halogen acceptors in hydrogen bonding. *Coord. Chem. Rev.* **2006**, *250*, 710–727.

(19) Bettinger, H. F. How good is fluorine as a hydrogen-bond acceptor in fluorinated single-walled carbon nanotubes. *ChemPhysChem* **2005**, *6*, 1169–1174.

(20) Parsch, J.; Engels, J. W. C-F···H-C hydrogen bonds in ribonucleic acids. *J. Am. Chem. Soc.* **2002**, *124*, 5664–5672.

(21) Carosati, E.; Sciabola, S.; Crucia, G. Hydrogen bonding interactions of covalently bonded fluorine atoms: from crystallographic data to a new angular function in the GRID force field. *J. Med. Chem.* **2004**, *47*, 5114–5125.

(22) Politzer, P.; Murray, J. S.; Concha, M. C. Halogen bonding and the design of new materials: organic bromides, chlorides and perhaps even fluorides as donors. *J. Mol. Model.* **2007**, *13*, 643–650.

(23) Lu, Y.; Zou, J.; Yu, Q.; Jiang, Y.; Zhao, W. Ab initio investigation of halogen bonding interactions involving fluorine as an electron acceptor. *Chem. Phys. Lett.* **2007**, *449*, 6–10.

(24) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. Halogen bonding: the σ-hole. *J. Mol. Model.* **2007**, *13*, 291–296.

(25) Murray, J. S.; Lane, P.; Clark, T.; Politzer, P. σ-hole bonding: molecules containing group VI atoms. *J. Mol. Model.* **2007**, *13*, 1033–1038.

(26) Politzer, P.; Murray, J. S.; Concha, M. C. σ-hole bonding between like atoms; a fallacy of atomic charges. *J. Mol. Model.* **2008**, *14*, 659–665.

(27) Murray, J. S.; Concha, M. C.; Lane, P.; Hobza, P.; Politzer, P. Blue shifts *vs* red shifts in σ-hole bonding. *J. Mol. Model.* **2008**, *14*, 699–704.

(28) Murray, J. S.; Lane, P.; Politzer, P. Expansion of the σ-hole concept. *J. Mol. Model.* **2009**, *15*, 723–729.

(29) Böhm, H.; Banner, D.; Bendels, S.; Kansy, M.; Kuhn, B.; Müller, Obst-Sander, U.; Stahl, M. Fluorine in medicinal chemistry. *ChemBioChem* **2004**, *5*, 637–643.

(30) Gerebtzoff, G.; Li-Blatter, X.; Fischer, H.; Frentzel, A.; Seelig, A. Halogenation of drugs enhances membrane binding and permeation. *ChemBioChem* **2004**, *5*, 676–684.

(31) Müller, K.; Faeh, C.; Diederich, F. Fluorine in pharmaceuticals: looking beyond intuition. *Science* **2007**, *317*, 1881–1886.

(32) Jeschke, P. The unique role of fluorine in the design of active ingredients for modern crop protection. *ChemBioChem* **2004**, *5*, 570–589.

(33) Paulini, R.; Müller, K.; Diederich, F. Orthogonal multipolar interactions in structural chemistry and biology. *Angew. Chem., Int. Ed.* **2005**, *44*, 1788–1805.

(34) Kim, C.; Chandra, P. P.; Jain, A.; Christianson, D. W. Fluoroaromatic-fluoroaromatic interactions between inhibitors bound in the crystal lattice of human carbonic anhydrase II. *J. Am. Chem. Soc.* **2001**, *123*, 9620–9627.

(35) 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.

(36) Shin, J.; Cho, D. PDB-Ligand: a ligand database based on PDB for the automated and customized classification of ligand-binding structures. *Nucleic Acids Res.* **2005**, *33*, D238–D241.

(37) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci., U.S.A.* **2004**, *101*, 16789–16794.

(38) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, Jr. R. L. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009** [Epub ahead of print].

(39) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735–1747.

(40) Zhou, P.; Tian, F.; Lv, F.; Shang, Z. Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins* **2009**, *76*, 151–163.

(41) Bondi, A. van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 441–451.

(42) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618–622.

(43) Dunning, T. H., Jr. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(44) Sahu, P. K.; Lee, S. Hydrogen-bond interactions in THF-H_2O-HF: a theoretical study. *Int. J. Quantum Chem.* **2007**, *107*, 2015–2023.

(45) Chesnut, D. B.; Moseley, R. W. The geometries of molecular complexes: an extended Hückel approach. *Theor. Chim. Acta.* **1969**, *13*, 230–248.

(46) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.

(47) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: a multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363.

(48) Adamo, C.; Barone, V. Exchange functionals with improved longrange behavior and adiabatic connection methods without adjustable parameters: the mPW and mPW1PW models. *J. Chem. Phys.* **1998**, *108*, 664–675.

(49) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, Jr, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(50) Lu, Y.; Shi, T.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. Halogen bonding−a novel interaction for rational drug design. *J. Med. Chem.* **2009**, *52*, 2854–2862.
(51) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
(52) Comell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.
(53) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Developing and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(54) Alzate-Morales, J. H.; Caballero, J.; Jague, A. V.; Nilo, F. D. G. Insights into the structural basis of N2 and O6 substituted guanine derivatives as cyclin-dependent kinase 2 (CDK2) inhibitors: prediction of the binding modes and potency of the inhibitors by docking and ONIOM calculation. *J. Chem. Inf. Model.* **2009**, *49*, 886–899.
(55) Olsen, J. A.; Banner, D. W.; Seiler, P.; Wagner, B.; Tschopp, T.; Obst-Sander, U.; Kansy, M.; Müller, K.; Diederich, F. Fluorine interactions at the thrombin active site: protein backbone fragments H-Cα-C═O comprise a favorable C-F environment and Interactions of C-F with electrophiles. *ChemBioChem* **2004**, *5*, 666–675.
(56) Morozov, A. V.; Kortemme, T.; Tsemekhman, K.; Baker, D. Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc. Natl. Acad. Sci., U.S.A.* **2004**, *101*, 6946–6951.
(57) DiMagno, S.; Sun, H. The strength of weak interactions: aromatic fluorine in drug design. *Curr. Top. Med. Chem.* **2006**, *6*, 1473–1482.
(58) Feyereisen, M. W.; Feller, D.; Dixon, D. A. Hydrogen bond energy of the water dimer. *J. Phys. Chem.* **1996**, *100*, 2993–2997.
(59) Taylor, R.; Kennard, O.; Versichel, W. Geometry of the N-H···O═C hydrogen bond. 1. Lone-pair directionality. *J. Am. Chem. Soc.* **1983**, *105*, 5761–5766.
(60) Yerushalmi, R.; Brandis, A.; Rosenbach-Belkin, V.; Baldridge, K. K.; Scherz, A. Modulation of fragmental charge transfer via hydrogen bonds. Direct measurement of electronic contributions. *J. Phys. Chem. A* **2006**, *110*, 412–421.
(61) Obst, U.; Gramlich, V.; Diederich, F.; Weber, L.; Banner, D. W. Design of novel, nonpeptidic thrombin inhibitors and structure of a thrombin-inhibitor complex. *Angew. Chem., Int. Ed.* **1995**, *34*, 1739–1742.
(62) Obst, U.; Banner, D. W.; Weber, L.; Diederich, F. Molecular recognition at the thrombin active site: structure-based design and synthesis of potent and selective thrombin inhibitors and the X-ray crystal structures of two thrombin-inhibitor complexes. *Chem. Biol.* **1997**, *4*, 287–295.
(63) Hof, F.; Scofield, D. M.; Schweizer, W. B.; Diederich, F. A weak attractive interaction between organic fluorine and an amide group. *Angew. Chem., Int. Ed.* **2004**, *43*, 5056–5059.
(64) Foster, J. P.; Weinhold, F. Natural atomic orbitals and natural population analysis. *J. Am. Chem. Soc.* **1980**, *102*, 7211–7218.
(65) Lu, Y.; Zou, J.; Fan, J.; Zhao, W.; Jiang, Y.; Yu, Q. Ab initio calculations on halogen-bonded complexes and comparison with density functional methods. *J. Comput. Chem.* **2009**, *30*, 725–732.
(66) Peterson, S. A.; Klabunde, T.; Lashuel, H. A.; Purkey, H.; Sacchettini, J. C.; Kelly, J. W. Inhibiting transthyretin conformational changes that lead to amyloid fibril formation. *Proc. Natl. Acad. Sci., U.S.A.* **1998**, *95*, 12956–12960.
(67) Kim, C.; Chang, J. S.; Doyon, J. B.; Baird, T. T.; Fierke, C. A.; Jain, A.; Christianson, D. W. Contribution of fluorine to protein-ligand affinity in the binding of fluoroaromatic inhibitors to carbonic anhydrase II. *J. Am. Chem. Soc.* **2000**, *122*, 12125–12134.
(68) Sörme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. M.; Nilsson, U. J. Structural and thermodynamic studies on cation-π interactions in lectin-ligand complexes: high-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction. *J. Am. Chem. Soc.* **2005**, *127*, 1737–1743.
(69) Benson, M. L.; Smith, R. D.; Khazanov, N. A.; Dimcheff, B.; Beaver, J.; Dresslar, P.; Nerothin, J.; Carlson, H. A. Binding MOAD, a high-quality protein-ligand database. *Nucleic Acids Res.* **2008**, *36*, D674–D678.

CI9002393