Propensities of Polar and Aromatic Amino Acids in Noncanonical Interactions: Nonbonded Contacts Analysis of Protein−Ligand Complexes in Crystal Structures

Yumi N. Imai,*,† Yoshihisa Inoue,† and Yoshio Yamamoto‡

Discovery Research Center and Strategic Research Planning Dept., Pharmaceutical Research Division, Takeda Pharmaceutical Company Limited, 2-17-85, Juso-honmachi, Yodogawa-ku, Osaka 532-8686, Japan

Received August 30, 2006

The nonbonded contacts analysis of 14 polar and aromatic amino acid side chains was carried out for protein−ligand complexes derived from the crystal structures in the Protein Data Bank. Through the exhaustive analysis, several unusual contacts were observed as well as the well-known interactions. CH−S interactions were frequently found in Met-related contacts, which have not yet been the subject of systematic investigations. We have also described the propensity of each amino acid for nonbonded interactions. All amino acids studied in this work showed high frequencies for the canonical hydrogen-bonding NH−O, OH−N, and OH−O interactions, while the preferences in noncanonical interactions such as CH− π interactions were not always consistent among the side chains with similar characteristics. Understanding such amino acid side chain propensities is important for improving the accuracy of structure-based drug design, and this study will open new possibilities for developing unique compounds with high binding affinity.

Introduction

In recent years, detailed analysis of crystal structures of protein−ligand complexes has established the importance of weak interactions, which are directly relevant to rational drug design. A number of crystal structures have been reported regarding weak nonbonded contacts such as CH−O,[1] CH−π,[2−4] $\pi-\pi$,[5] and cation−π interactions[6] which greatly contribute to the molecular recognition in protein−ligand interactions. Though the importance of these weak interactions in biological systems is well established,[7] there has yet to be a systematic analysis of the roles and frequencies of these interactions in protein−ligand complexes. The Protein Data Bank (PDB)[8] is a database containing crystal structures of biological macromolecules, and the number of the registered structures has grown exponentially in recent years. Its potential as a useful resource for protein−ligand interactions has been increased significantly by the registration of a large number of high-resolution crystal structures determined by the use of synchrotron X-ray diffraction. An exhaustive analysis of this database will provide us with new insights for the patterns and the preferences of protein−ligand interactions and will aid structure-based drug design of more potent ligands. However, errors and inconsistencies in the contents of the PDB data have prevented us from handling the information effectively. We therefore employed the protein−ligand interaction database, Relibase+,[9] which possesses well-formatted data and python libraries, as an analysis tool to solve these problems. Utilizing Relibase+, we carried out systematic analysis of the nonbonded contacts of 14 polar and aromatic amino acid side chains in the crystal structures of protein−ligand complexes from the PDB. In this article, we report the preferences and propensities of the patterns and types of weak interactions for each amino acid side chain. We also describe unusual or unexpected interactions identified in this work.

* Corresponding author. Tel: +81-6-6308-9053; Fax: +81-6-6308-9010. E-mail: imai_yumi@takeda.co.jp.
† Discovery Research Center.
‡ Strategic Research Planning Dept.

Results and Discussion

More than 25 000 PDB structures in Relibase+ (August 2004) were exhaustively analyzed, and the propensities for nonbonded contacts of 14 amino acids were elucidated. Although the frequency of each interaction was obtained, only preferences are discussed, because the redundant data from certain protein families were not completely precluded in this work. The statistics of the search results are shown in Table 1. Relibase+ version 1.2 includes 151 416 ligand data derived from 26 538 PDB datasets. A total of 349 575 hits were retrieved by primary query searches, which were filtered and pruned to provide 6091 structures for this analysis. The number of hits for each amino acid shown in the second column of Table 1 means the number of ligands retrieved by primary searches of the whole database via the graphical user interface of Relibase+, and the actual hit number of interactions for each amino acid is larger than that because there were cases where a ligand contained two or more interactions with amino acid residues of the same kind. The third column provides the number of the "interaction" data per amino acid after processing and filtering by the criteria shown in the Experimental Section. The resulting number of the final interaction data per amino acid after removing duplicates is provided in the fourth column, and the percentages described for each amino acid in the following sections were calculated by using this number as a denominator.

Arg. For the contacts of nitrogen atoms of Arg side chains, N−O (e.g., NH−O or OH−N) interactions accounted for more than 80%, and a distinct preference for the salt bridge with a carboxylate (ca. 60%) was exhibited. About 5% of all the Arg-related noncanonical interactions were observed as cation−π interactions between the guanidinium group of the Arg side chain and an aromatic ring of the ligand. The guanidinium group of Arg can participate in cation−π interactions in two geometries, parallel and T-shaped (Figure 1(a)) in crystal protein structures.[10] In this work, we found additional types, parallel displaced (e.g., PDB 1EC3, Figure 1(c)) and herringbone (e.g., PDB 1ANK Figure 1(d)) geometries (Figure 1(b)) similar to $\pi-\pi$ interactions,[11] where a guanidinium group of Arg may play a role as a π donor. Computationally, the T-shaped

10.1021/jm061038a CCC: $37.00
Published on Web 02/22/2007

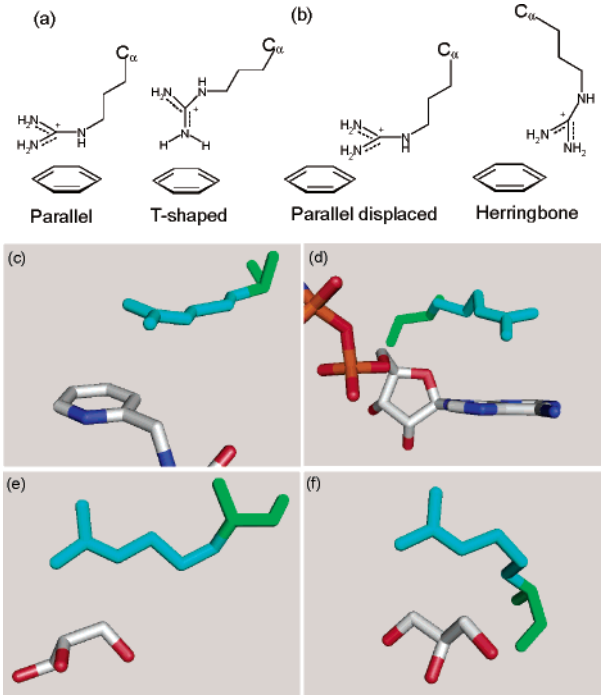


Figure 1. Noncanonical interactions of Arg side chain. Geometries in cation−π interactions involving Arg side chain: parallel ((a), left) and T-shaped ((a), right), parallel displaced ((b), left) and herringbone ((b), right). Additional geometries in cation-π interactions of Arg side chains found in this work are parallel displaced geometry (PDB 1EC3, (c)) and herringbone geometry (PDB 1ANK, (d)). CH−N interactions involving Arg side chain (PDB 1O6I, (e) and PDB 1KTB, (f)).

Table 1. Statistics of the Search Results and the Postprocessing Procedures in This Work

side chain	no. of hits by query search	no. of hits after the filtration	no. of hits after removing duplicants
Arg	26912	14346	1197
Asn	10457	3459	378
Asp	44246	8283	860
Cys	7733	784	99
Gln	11027	3310	295
Glu	39507	6800	794
His	51145	8328	559
Lys	15558	7005	423
Met	7622	523	95
Phe	31979	1456	137
Ser	24893	7389	337
Thr	18589	6122	231
Trp	15431	2797	209
Tyr	44476	6595	477
total	349575	77197	6091

geometry is favored in the gas phase, but, in solution, the parallel geometry is preferred in protein structures.[12] In agreement with previous studies,[11,13−15] we found that parallel geometry is preferred in protein−ligand interactions.

A number of unexpected contacts between a guanidinium nitrogen atom of Arg and a methylene group of the ligand were observed, where the distance between a nitrogen atom of the Arg side chain and a methylene carbon atom of the ligand was within 3.2 Å (e.g., PDB 1O6I, 1KTB, Figure 1(e)(f)). In particular, a moderately strong interaction was seen in the atom pair between a guanidinium nitrogen atom of Arg and a carbon atom adjacent to a nitrogen or an oxygen atom of the ligand. Taylor and Kennard reported that CH−N interactions as well as CH−O, CH−Cl, and CH−S interactions represent nonconventional hydrogen bonds based on 113 high quality organic crystal structures determined by neutron diffraction.[16] Steiner

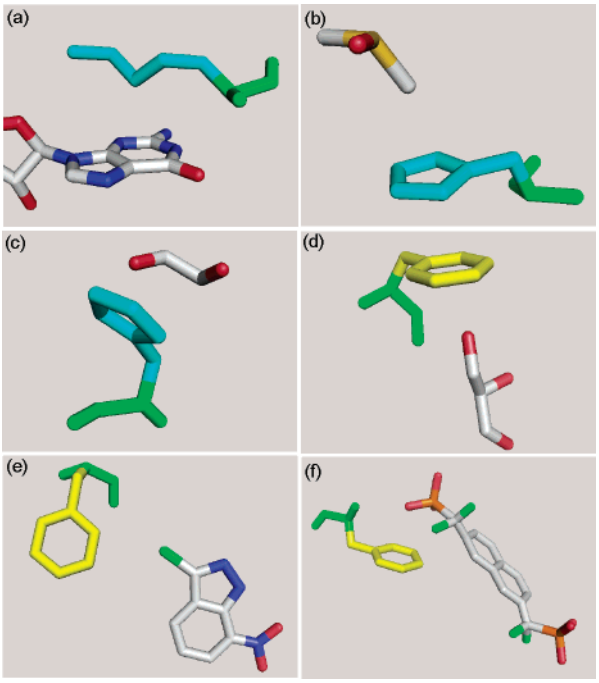


Figure 2. Lys, His, and Phe side chain-related interactions. CH−π interactions of methylene group of Lys side chain. (PDB 1AGP, (a)) CH−π interaction (PDB 1NO0, (b)) and OH−π interaction (PDB 1KJQ, (c)) involving His aromatic ring. OH−π interaction (PDB 1KFW, (d)), bromine−π interaction (PDB 1D0C, (e)) and fluorine−π interaction (PDB 1KAK, (f)) in Phe aromatic ring-related interactions.

further studied the nature of hydrogen bonds in CH−X (X = N, O, Cl) interactions, elucidating the correlation between the acidity of the CH group and the CH−N distance among the data extracted from the Cambridge Structural Database (CSD).[17] To form a CH−N hydrogen bond in this case, however, the CH group should act as an H-donor and the guanidinium nitrogen atom as an H-acceptor despite the fact that a guanidinium group is protonated at neutral pH and is unlikely to be an H-acceptor. We therefore cannot make any conclusion regarding the nature of this type of contact. Further analysis should be done in order to elucidate its mechanism. For the contacts of carbon atoms of the Arg side chain, those between a methylene group of Arg and an aromatic ring of the ligand were found and may be categorized as CH−π interactions.

Lys. For the contacts of nitrogen atoms of Lys side chains, N−O (e.g., NH−O or OH−N) interactions comprised the majority of the interactions, where either a phosphate group or a sugar molecule was involved in many cases besides a carboxylate group. The contacts between a nitrogen atom of the Lys side chain and a methylene group of the ligand were observed most frequently in the noncanonical interactions (ca. 15% of all interactions), as well as in the case of Arg. About 14% of all Lys-related noncanonical interactions were observed as cation−π interactions between the amine nitrogen of Lys and an aromatic ring of the ligand. This cation−π interactions have been well discussed by the Dougherty group.[10]

For the contacts of carbon atoms of Lys side chains, significant preference for CH−O interactions was observed (ca. 53% of all side chain carbon atom-related interactions), but this result may be biased by a large amount of data involving a phosphate group or a sugar group. A contact between a methylene group of Lys and an aromatic ring of the ligand was found and categorized as a CH−π interaction (e.g., PDB 1AGP, Figure 2(a)). This interaction was also found in the crystal structure of the v-src SH2 domain complexed with tyrosine-

phosphorylated peptides, where the aromatic ring of the phosphotyrosine was recognized by the carbon atoms of the Lys side chain, with C−C distances of approximately 3.7 Å between Lys and the aromatic ring.[18]

His. Because it is difficult to distinguish two flipping states and/or the protonation states of His side chains even at a moderate resolution, the preferences based on atom species may be insignificant. However, many cases were found where His side chains were located around phosphate groups of ligands, and nitrogen atoms of the His side chain may be preferentially involved in the N−O (e.g., NH−O or OH−N) interactions. Although His can participate in cation−π interaction as either a cation or π donor depending on its protonation states, cation−π interactions involving the His side chain were seldom observed. This may be explained by the reduced electrostatic potential of the His side chain compared to the Phe side chain that has been reported.[10]

Alternatively, $\pi-\pi$ interactions were observed more frequently, and three types of $\pi-\pi$ interactions, parallel, pseudo-T-shaped, and herringbone,[11] were found, of which the parallel type was predominant. The stacked His side chain against another π donor has been commonly observed in protein crystal structures. For instance, recently Henderson et al. reported that the stacking of the His side chain against the phenol group of the ligand was demonstrated to play a key role in the control of the fluorescence efficiency.[19] For other interactions categorized as a π donor, CH−π interactions (PDB 1NO0, Figure 2(b)) and OH−π interactions were observed (PDB 1KJQ, Figure 2(c)).

Phe. No salient propensity was observed in the contacts of Phe side chains, and $\pi-\pi$ interactions (ca. 30%), CH−π interactions (ca. 25%), and CH−O interactions (ca. 30%) accounted for most of the interactions. Among the CH−O interactions, cases where the CZ or CD atom of the Phe side chain was in close proximity to a bridging (C−O−C) oxygen of the ligand were predominant. A survey of the interactions of small molecules in the CSD revealed that the oxygen of furan rarely accepts a conventional hydrogen bond, but short CH−O interactions are sometimes found.[20] Theoretically calculated hydrogen-bond strengths have supported the observation that ether oxygens are quite common hydrogen bond acceptors, while aromatic oxygens are weaker acceptors than ether oxygens (e.g., calculated hydrogen bond energies: dimethyl ether−methanol, −21.2 kJ/mol; furan−methanol, −14.4kJ/mol).[21] For the $\pi-\pi$ interactions, the Phe side chain was mainly found in parallel, parallel displaced, or T-shaped geometries, and the T-shaped geometry was observed to be preferred over parallel in this work, which is consistent with previous surveys[20,22] and recent theoretical calculations.[23−25] For CH−π interactions, the interactions with CH groups derived from carbohydrate residues of the ligands were frequently found. It is suggested that these interactions play an important role in the ligand-recognition function of carbohydrate-binding proteins[26] and are examples of CH−π interactions involved in actual binding events. For OH−π interactions, a few examples such as PDB 1KFW were found (Figure 2(d)).

For halogen−aromatic interactions, only a very small number of cases could be found, for instance, in PDB 1D0C or 1KAK (Figure 2(e)(f)). Chlorine and Bromine were obviously involved in CH−halogen or halogen−π interactions as shown in Figure 2(e), while the interaction involving fluorine such as in PDB 1KAK may be a cooperative interaction or an artifact due to the low resolution, although in fact, the resolution of PDB 1KAK is 2.5 Å. Halogen substitution is an important approach for drug design, since halogenation may result in altered physicochemical properties and enhanced membrane solubilities. However, explanation of the structure−activity effects of halogens have been limited to considerations of membrane solubility and the steric effects of halogen substituents on aromatic rings, though much is known about the effect of halogens on crystal packing.[27] Recently, Codding et al.[28] has reported that (a) halogen atoms are potential H-bond acceptors able to interact with strong and weak H-bond donors,[29] although the evidence is equivocal for C−F as a H-bond acceptor;[30] (b) CH−halogen interactions are weakly attractive yet highly dependent on the molecular environment of the halogen,[31,32] and (c) halogen−aromatic ring and halogen−H interactions are stronger than halogen−halogen interactions.[33] Our findings are basically consistent with their results.

Tyr. For the nonbonded contacts of phenol OH, OH−O interactions were predominant, accounting for about 60% of the interactions, and only about 13% of the interactions were cases where Tyr was involved in CH−O interactions as a proton acceptor. As a π donor, no clear preference was observed as with Phe; $\pi-\pi$ interactions (ca. 28%), CH−π interactions (ca. 30%), and CH−O interactions (ca. 17%) were observed. For $\pi-\pi$ interactions, parallel geometry was predominant, and this observation is consistent with the previous theoretical calculations regarding the substituent effects on $\pi-\pi$ interactions, where phenol binds to benzene more favorably in parallel than T-shaped geometry, whereas the benzene dimer favors T-shaped geometry.[24]

Trp. The nitrogen atom of the Trp side chain was involved in N−O (e.g., NH−O or OH−N) interactions, which accounted for the overwhelming number (ca. 80%) of the all nitrogen-related nonbonded contacts. However, for the noncanonical interactions, the nitrogen atom did not seem to play an important role. As a π donor a clear preference was observed for OH−π interactions (ca. 45%), and to a lesser extent CH−π interactions (ca. 29%). Gallivan and Dougherty reported that 26% of all Trp residues were involved in energetically significant cation−π interactions among side chains within proteins from an analysis of structures in the PDB.[10] Although our observation cannot be compared directly with their result, the rate of cation−π interactions in the noncanonical interactions of Trp was as low as about 1.5%. For $\pi-\pi$ interactions, T-shaped geometry was frequently observed, as well as parallel or herringbone geometries. Also, strong contacts between the Trp aromatic ring and a methylene group of the ligand were observed, where the distance between the ring atom of Trp and a methylene carbon atom of the ligand was within 2.9 Å.

Thr and Ser. The majority of all contacts in each side chain were accounted for by OH−O interactions for canonical interactions and by CH−O interactions for noncanonical interactions, respectively.

Met. The contact where the sulfur atom of Met is in proximity to a π donor appeared with greater frequency than was expected (ca. 22%). This contact was seen not only with an aromatic ring but also with an amide group as a π donor; it was observed even when the π density is weaker as in the case of perfluorobenzene (e.g., PDB 1MGO, Figure 3(a)). Evidence for these S−π interactions have been reported in the statistical analysis of interactions of amino acids in protein crystal structures[34−36] and in small molecule crystal structures in the CSD,[37] as well as studies involving molecular mechanics[38] and quantum chemical calculations.[39,40] In divalent sulfur systems such as thioethers, the sulfur atom may even carry a small positive partial charge, and in this situation, the sulfur atom may be

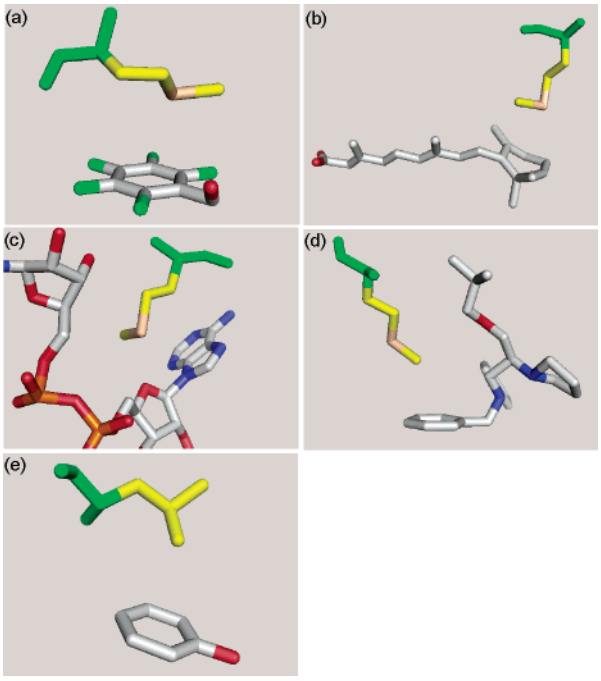


Figure 3. Met side chain-related interactions. S−π interaction between Met sulfur atom and perfluoro phenyl group (PDB 1MGO, (a)), CH−S interaction (PDB 1FEM, (b)), S−ether oxygen interaction (PDB 1GUF, (c)), CH−π interaction involving a terminal methyl carbon atom in Met (PDB 1DTL, (d)), and in Leu (PDB 1MPJ, (e)).

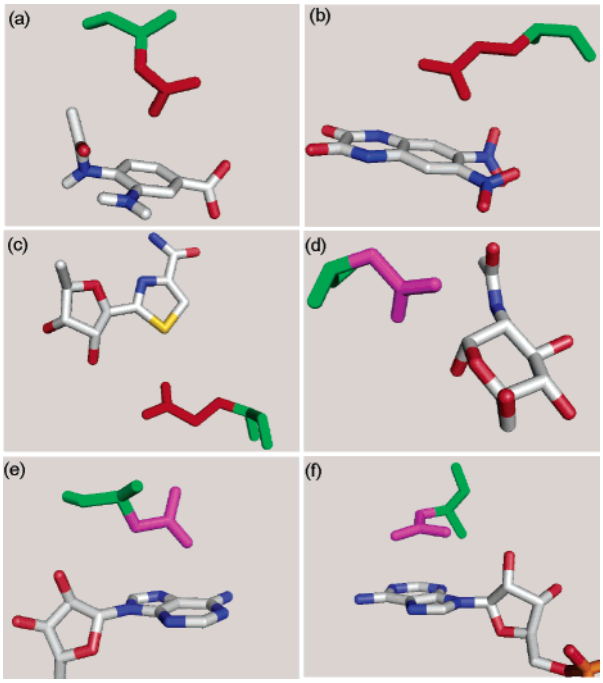


Figure 4. Asp/Glu and Asn/Gln side chain-related interactions. Interaction between the carboxyl oxygen of Asp side chain and the aromatic ring of the ligand (PDB 1IVE, (a)), interaction between carboxyl oxygen of Glu side chain and aromatic ring (PDB 1FTL, (b)), thioether sulfur (PDB 1AER, (c)). Interaction between the carbamoyl group of Asn side chain and a CH group of the ligand (PDB 1USR, (d)). Two predominant geometries in the interactions between the carbamoyl groups and the aromatic rings found in the contacts of Asn/Gln side chains, T-shaped geometry in PDB 2MPJ (e), and parallel geometry in PDB 1KVT (f).

engaged in a completely different kind of electrostatic intermolecular interactions, that is $>S\delta+...X\delta-$,[41] where the π donor may be involved as a carrier of a negative partial charge.

Surprisingly, CH−S interactions (ca. 40%) accounted for the majority of all Met sulfur atom-related interactions. Because divalent sulfur may carry a small positive partial charge and is thus unsuitable as a hydrogen bond acceptor as mentioned above, the hydrogen bonds from the weak donor CH to the weak acceptor S must be considerably weaker than CH−O interactions. However, as the counterpart of the contact was the CH moiety derived from a methylene group in almost every case, the divalent sulfur atom of the Met side chain may assume a vital role in the recognition of methylene groups (e.g., PDB 1FEM, Figure 3(b)). Since these contacts have not yet been the subject of systematic investigations and little convincing literature has appeared, this work may be a first observation regarding the preference for CH−S interactions in Met.

S−O interactions, especially the contact between the sulfur atom in Met and an ether oxygen atom in a ligand, were also observed in many cases (e.g., PDB 1GUF, Figure 3(c)), and these contacts may be due to a type of electrostatic interaction. Involvement of the methyl or methylene groups of the Met side chain in CH−π or CH−O interactions was preferentially observed. For the CH−π interactions, contacts in which the terminal CE atom of the Met side chain comes close to an aromatic ring of the ligand were found frequently (e.g., PDB 1DTL, Figure 3(d)). For comparison, the contacts of terminal carbon atoms (CD1 or CD2) in Leu side chains were examined. The propensity for CH−π interactions was also observed frequently (e.g., PDB 1MPJ, Figure 3(e)), and these contacts in the terminal carbon atoms of the side chains were suggested to be not specific for the Met side chain.

Cys. For the contacts of the Cys side chain, S(H)−π interactions (ca. 30%) were frequently observed as well as CH−S interactions (ca. 20%) and S(H)−O interactions (ca. 17%). Although the frequencies of SH−O interactions was much less than those of OH−O interactions, the SH group is one of the conventional hydrogen-bonding functional groups, and its donor and acceptor potentials have been studied in the 1960s by vibrational spectroscopy.[42] SH−π interactions were frequently observed and considered in analogy to OH−π interactions, where the SH group may be engaged as a weak hydrogen donor. The S(H)−π interaction has been proposed to be a favorable interaction because Met and Cys residues are commonly in close proximity to aromatic residues in protein crystal structures or small molecule crystal structures.[43−46] These amino acids may play an important role in protein−ligand interactions.

Asp. Regarding contacts of the Asp side chain, OH−O interactions (ca. 60%) occupied the majority of the interactions, followed by N−O (e.g., NH−O or OH−N) interactions (ca. 30%). The CH−O interactions were observed relatively frequently within the noncanonical interactions, which represent a small proportion of the total Asp-related interactions (ca. 1%). Surprisingly, a few cases (ca. 3%) were found where an aromatic ring was located within the van der Waals distance of carboxylic oxygens of an Asp side chain. In every case, the carboxylic oxygen atoms are perpendicular to the aromatic ring plane (e.g., PDB 1IVE, Figure 4(a)). If the Asp side chain is protonated, these contacts are categorized as OH−π interactions, but the hydrogen atoms were not included in any data. We are therefore unable to make any conclusion regarding this type of contact.

Glu. As well as the strong preference for OH−O interactions (ca. 35%) and N−O (e.g., NH−O or OH−N) interactions (ca. 30%) in the conventional interactions, Glu also exhibited CH−O interactions (ca. 15%) in noncanonical interactions. The preference of the CH−O interactions appears to be characteristic to Glu. The nature of the contact between the carboxylic oxygens of the Glu side chain and the aromatic ring, observed in

Table 2. Examples of Noncanonical Interactions Found in This Report

interaction	PDB code	subunit	protein[a]	SCOP class[b]	ligand[c]	resolution (Å)	Figure
NH−π	1EC3	-	HIV-1 protease	all β	MSA367	1.8	1(c)
NH−π	1ANK	B	adenylate kinase	α/β	ANP	2.0	1(d)
CH−N	1O6I	A	chitinase	α/β	glycerol	1.7	1(e)
CH−N	1KTB	-	acetyl galactosaminidase	α/β	propane-1,2,3-tri ol	1.9	1(f)
CH−π	1AGP	-	p21-H-ras	α/β	GNP	2.3	2(a)
CH−π	1NO0	-	thermolysin	$\alpha+\beta$	DMS	2.1	2(b)
OH−π	1KJQ	-	PRGA formyltransferase	all β	ethylene glycol	1.05	2(c)
OH−π	1KFW	-	chitinase B	α/β	glycerol	1.74	2(d)
Br−π	1D0C	-	NO synthase	$\alpha+\beta$	INE	1.65	2(e)
F−π	1KAK	-	PTP 1B	α/β	FNP	2.5	2(f)
S−π	1MGO	-	alcohol dehydrogenase	all β	PFB	1.2	3(a)
CH−S	1FEM	-	PRBP	all β	Retinoic acid	1.9	3(b)
S−O	1GUF	B	enoyl thioester reductase	α/β	NDP	2.25	3(c)
CH−π	1DTL	-	troponin C	all α	Bepridil	2.15	3(d)
CH−π	1MPJ	-	phenol insulin	small protein	phenol	2.3	3(e)
COO^-−π	1IVE	-	neuraminidase	all β	ST3	2.4	4(a)
COO^-−π	1FTL	-	glutamate receptor	α/β	DNQ	1.8	4(b)
S−O	1AER	B	*Pseudomonas* exotoxin	$\alpha+\beta$	TIA	2.3	4(c)
CH−π	1USR	A	HNG	all β	NDG	2.0	4(d)
$\pi-\pi$	2MJP	-	pyrophosphatase	α/β	ANP	2.2	4(e)
$\pi-\pi$	1KVT	-	UDP-Gal-epimerase	α/β	NAD	2.15	4(f)

[a] Abbreviations in protein names: PRGA formyltransferase: phosphoribosylglycinamide formyltransferase, PTP 1B, protein tyrosine phosphatase 1B; NO synthase, Nitric oxide synthase; HNG, hemagglutinin neuraminidase glycoprotein; PRBP, plasma retinol binding protein. [b] Structural classification of proteins (**46**). Available on the Internet under http://scop.mrc-lmb.cam.ac.uk/scop/. The SCOP class definition is all α, all β, α/β for mainly parallel β-sheets, e.g., in $\beta\alpha\beta$-units, $\alpha+\beta$ for mainly antiparallel β-sheets and segregated α and β-regions, peptides. [c] Abbreviations in ligand names: ANP, phosphoaminophosphonic acid adenylate ester; GNP, phosphoaminophosphonic acid guanylate ester; GMS, dimethyl sulfoxide; INE, 3-bromo-7-nitroimdazole; FNP, {[7-(difluorophosphonomethyl)naphthalene-2-yl]difluoromethyl}phosphonic acid; PFB, 2,3,4,5,6-pentafluorobenzyl alcohol; NDP, NADPH dihydronicotinamide adenine dinucleotide phosphate; ST3, 4-(acetylamino)-3-aminobenzoic acid; DNQ, 6,7-dinitroquinoxaline-2,3-dione; TIA, 2-(1,5-dideoxyribose)-4-amidothiazole; NDG, 2-(acetylamino)-2-deoxy-α-D-glucopyranose; NAD, nicotinamide adenine dinucleotide.

approximately 4% of cases like in Asp, could not be understood (e.g., PDB 1FTL, Figure 4(b)). Strong S−O interactions between the carboxylic oxygens of the Glu side chain and a divalent sulfur atom of the ligand were found in several cases, as in the case of Met (e.g., PDB 1AER, Figure 4(c)).

Asn and Gln. Because X-ray crystal diffraction usually cannot reliably distinguish nitrogen from oxygen, the chemical identity of the terminal side chain atoms is uncertain for Asn and Gln. This means that the analysis for these two amino acids based on the atom species derived from the PDB data could demonstrate no clear preference. As a carbamoyl group, these two side chains have preferences for the interaction with oxygen atoms of ligands, with either N−O (e.g., NH−O, or OH−N) or OH−O interactions. For noncanonical interactions, the contacts with a CH group of the ligand were frequently observed (e.g., PDB 1USR, Figure 4(d)). Unusual interactions between the carbamoyl group of the Asn/Gln side chain and an aromatic ring of the ligand were observed, where the carbamoyl group may bind to the aromatic ring as a π donor, and both T-shaped and parallel geometries were predominantly found (e.g., PDB 2MJP, 1KVT, Figure 4(e)(f)). These contacts have not been reported in the literature, although a theoretical study of the interaction between the carbamoyl group and an acidic π-ring exists.[47]

Reliabilities of the Noncanonical Interactions Found in This Work. Although we set limits on the resolution of the crystal structures and temperature factors of contact atoms (see Experimental Section) in the search criteria, we reexamined the resolution of each crystal structure which contained a noncanonical interaction exemplified in this report to confirm its reliability (Table 2). For the rarely observed contacts such as halogen−π and carboxyl−π interactions, some data were over 2.3 Å resolution. However, the existence of high-resolution data under 2.0 Å for these interactions demonstrate that even the noncanonical interactions found in this work are sufficiently reliable and are not derived from errors or artifacts in crystallography. The distribution of SCOP classes[48] for the proteins has also demonstrated that these interactions are not confined to particular protein−ligand systems but are universal.

Conclusions

The nonbonded contacts analysis of 14 polar and aromatic amino acid side chains was carried out for protein−ligand complexes derived from the crystal structures in the PDB. Through the exhaustive analysis of the Relibase+ database, several novel interactions were observed and recognized as well as the well-known interactions. It was revealed that NH−O, OH−N, and OH−O interactions are the major interactions involved in molecular recognition, regardless of the type of amino acids studied in this work. However, the propensities in noncanonical interactions are not always consistent among the side chains with similar characteristics. Novel interaction modes were observed for $\pi-\pi$ interactions, guanidinium−π interactions, etc. The preferences for CH−N interactions of Lys and Arg and CH−S interactions of Met, where the nitrogen/sulfur atom of their side chains and a methylene group of the ligand are located within the sum of the van der Waals radius of each atom are to our knowledge novel and reported here for the first time. As for the analysis based on the data derived from the PDB, some individual surveys such as for halogen bonds,[49] NH−O, OH−O, and CH−O hydrogen bonds in the protein−ligand complexes,[50] interactions involving amide NH and C=O groups in protein−ligand complexes,[51] CH−O interactions,[52−55] and CH−π interactions[56] in biological molecules have been reported. However, this is the first exhaustive analysis of the propensities in the nonbonded contacts of the protein−ligand interactions based on the type of the side chains. Recently, a series of web-based search tools for protein−ligand interactions have been reported.[57,58] These tools provide the analysis methodology and ligand-based statistics of the interactions. Although problems regarding accuracy in the frequency and verification of the data remain, knowledge of the propensities based on the type of interactions for each side chain makes a significant contribution toward improving structure-based drug

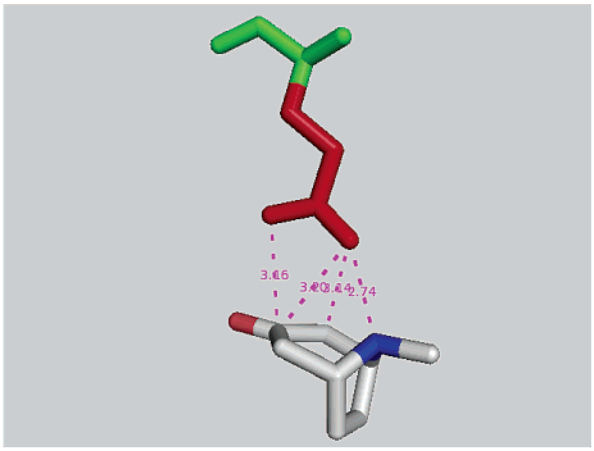
Table 3. Sybyl Atom Types Used in This Work

code	definition
nitrogens	
N.pl3	nitrogen trigonal planar, e.g., in nitro, pyrrole
N.am	nitrogen amide
N.4	nitrogen sp^3 positively charged, e.g., in lysine
N.3	nitrogen sp3
N.2	nitrogen sp^2
N.1	nitrogen sp
N.ar	nitrogen aromatic, e.g., in pyridine
carbons	
C.3	carbon sp^3
C.2	carbon sp^2
C.1	carbon sp
C.ar	carbon aromatic
C.cat	carbocation (guanadinium)
oxygens	
O.3	oxygen sp^3
O.2	oxygen sp^2
O.co2	oxygen in carboxylates and phosphates
sulfurs	
S.3	sulfur sp^3
S.2	sulfur sp^2
S.0	sulfoxide sulfur
S.o2	sulfone sulfur
halogens	
F	fluorine
Cl	chlorine
Br	bromine
I	iodine

design. Furthermore, unpredicted propensities in noncanonical interactions observed in this work demonstrate the importance of such weak interactions in molecular recognition. Exploitation of these comparatively little known interactions may increase novelty in drug design.

Experimental Section

We have searched and selected combinations of nonbonded contacts of 14 amino acids including polar and aromatic side chains (Asp, Glu, Arg, Lys, His, Asn, Gln, Ser, Cys, Thr, Tyr, Phe, Trp, and Met) in protein−ligand complex structures for the subset of the PDB structures in Relibase+ (The Cambridge Crystallographic Data Centre, Cambridge, UK., 26 538 PDB structures, 154 146 ligand data, up to August 2004) satisfying the following criteria: (i) crystallographic data only (no NMR structures or theoretical models were used); (ii) resolution <2.7 Å; (iii) temperature factors of the contact atoms <60; (iv) a ligand was defined as a nonpeptide molecule of at least two atoms with atomic species confined to C, N, O, S, F, Cl, Br, and I. Organometallic compounds such as heme were precluded; (v) ligands defined as sugar, DNA, and RNA in Relibase+ were also excluded, and only noncovalently bonded complexes were selected. In the Relibase+ definition, a ligand including a sugar or a nucleobase moiety as a substructure, such as phosphoaminophosphonic acid adenylate ester, is not categorized as sugar, DNA, or RNA. Therefore, there are still cases where sugar, carbohydrate, or nucleobase may be included in the ligands analyzed here. Throughout this work, a nonbonded contact was defined as any contact where the interatomic distance between the side chain atom and ligand-constituent atom is shorter than the sum of each van der Waals radius, and only non-hydrogen atoms were included in the calculations. At first we applied protein−ligand interaction criteria based on Miyazawa and Jernigan's definition of interresidue contacts in proteins:[59] residues whose centers are closer than 6.5 Å are defined to be in contact. However, enormous quantities of data were hit and a lot of noisy data were found in the results. As highly reliable information is required in medicinal chemistry, we employed more restricted criteria as mentioned above. This resulted in less noisy and more meaningful data. For the amino acids



Atom Glu_B_156	Atom TNE_1262 (atom type)	Distance (Å)
OE1	C3 (C.2)	3.16
OE2	C3 (C.2)	3.20
OE2	C4 (C.3)	3.14
OE2	N8 (N.3)	2.74

PDB code: 1IPF

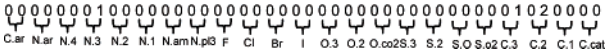


Figure 5. An example of data and the 46-digit fingerprint used in this work.

possessing an aromatic ring or a methylene chain, the contacts of carbon atoms in the side chain were also included. When each amino acid was searched, both the primary query search via the graphical user interface and the secondary search with in-house python scripts implemented by reliscript, a Relibase library for python script, were employed. In the primary search, a sub-database was prepared, which includes only ligands possessing any atoms within 6.5 Å of side chain atoms of the relevant amino acid. The secondary search was performed with more complex queries for checking interactions with the criteria mentioned above (sum of van der Waals radii, resolution, etc.). The search results contained contact atom names, residue names, Sybyl atom types, and interatomic distances as well as the PDB code. The types of the interactions were classified using the fingerprinting of the Sybyl atom types (Table 3). The atom types of the contact atoms were counted, and a 46-digit fingerprint was generated per each side chain as shown in Figure 5. These fingerprints were sorted and classified, and duplicates were excluded. The "duplicate" was defined as follows: (i) equivalent contacts in a different subunit within the identical PDB code or the redundant proteins; (ii) contacts between the identical ligand name and residue number within the redundant proteins. The redundant protein was defined based on the 90% similarity cluster list[60] distributed from the PDB. These operations were performed with in-house Perl scripts. Finally, each complex was checked by manual inspection to correct errors that were still not excluded automatically due to inconsistencies in the PDB data. The structure search using Relibase+ version 1.2 was performed on the IRIX64 Origin 200 SGI server (Silicon Graphics, Inc. Mountain View, CA). Perl scripts and python scripts were executed on the same machine. All figures were prepared by using the program PyMol.[61]

Acknowledgment. The authors gratefully acknowledge Drs. Kaneyoshi Kato and Tetsuo Miwa for their suggestions and encouragement, and Drs. Satoshi Sogabe and Akira Fujishima for helpful discussions. The authors thank Dr. Nicholas Hird for critical reading of the manuscript.

References

(1) Pierce, A. C.; Sandretto, K. L.; Bemis, G. W. Kinase Inhibitors and the Case for CH···O Hydrogen Bonds in Protein-Ligand Binding. *Proteins* **2002**, *49*, 567−576.

(2) Umezawa, Y.; Nishio, M. CH/π interactions in the crystal structure of class I MHC antigens and their complexes with peptides. *Bioorg. Med. Chem.* **1998**, *6*, 2507−2515.

(3) Nishio, M.; Hirota, M.; Umezawa, Y. *The CH/π Interaction*; Wiley-VCH: New York, 1998.
(4) Umezawa, Y.; Nishio, M. CH/π interactions as demonstrated in the crystal structure of guanine-nucleotide binding proteins, Src homology-2 domains and human growth hormone in complex with their specific ligands. *Bioorg. Med. Chem.* **1998**, *6*, 493−504.
(5) Boehr, D. D.; Farley, A. R.; Wright, G. D.; Cox, J. R. Analysis of the π-π stacking interactions between the aminoglycoside antibiotic kinase APH(3′)-IIIa and its nucleotide ligands. *Chem. Biol.* **2002**, *9*, 1209−1217.
(6) Kryger, G.; Silman, I.; Sussman, J. L. Structure of acetylcholinesterase complexed with E2020 (Aricept): implications for the design of new anti-Alzheimer drugs. *Struct. Fold. Des.* **1999**, *7*, 297−307.
(7) Desiraju, G. R.; Steinter, T. *The weak hydrogen bond in structural chemistry and biology*; Oxford University Press: Oxford, UK, 1999.
(8) The Protein Data Bank: http://www.rcsb.org/pdb/. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. N. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(9) Hendlich, M.; Bergner, A.; Gunther, J.; Klebe, G. Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.* **2003**, *326*, 607−620.
(10) Gallivan, J. P.; Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459−9464.
(11) Sun, S.; Bernstein, E. R. Aromatic van der Waals Clusters: Structure and Nonrigidity. *J. Phys. Chem.* **1996**, *100*, 13348−13366.
(12) Duffy, E. M.; Kowalczyk, P. J.; Jorgensen, W. L. Do denaturants interact with aromatic hydrocarbons in water? *J. Am. Chem. Soc.* **1993**, *115*, 9271−9275.
(13) Singh, J.; Thornton, J. M. SIRIUS. An automated method for the analysis of the preferred packing arrangements between protein groups. *J. Mol. Biol.* **1990**, *211*, 595−615.
(14) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. Amino/Aromatic Interactions in Proteins: Is the Evidence Stacked Against Hydrogen Bonding? *J. Mol. Biol.* **1994**, *239*, 315−331.
(15) Flocco, M. M.; Mowbray, S. L. Planar stacking interactions of arginine and aromatic side-chains in proteins. *J. Mol. Biol.* **1994**, *235*, 709−717.
(16) Taylor, R.; Kennard, O. Crystallographic evidence for the existence of C−H···O, C−H···N, and C−H···Cl hydrogen bonds. *J. Am. Chem. Soc.* **1982**, *104*, 5063−5070.
(17) Steiner, T. Donor and acceptor strengths in C−H···O hydrogen bonds quantified from crystallographic data of small solvent molecules. *New J. Chem.* **1998**, *22*, 1099−1103.
(18) Waksman, G.; Kominos, D.; Robertson, S. C.; Pant, N.; Baltimore, D.; Birge, R. B.; Cowburn, D.; Hanafusa, H.; Mayer, B. J.; Overduin, M.; Resh, M. D.; Rios, C. B.; Silverman, L.; Kuriyan, J. Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **1992**, *358*, 646−653.
(19) Henderson, J. N.; Remington, S. J. Crystal structures and mutational analysis of amFP486, a cyan fluorescent protein from Anemonia majano. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12712−12717.
(20) Hunter, C. A.; Singh, J.; Thornton, J. M. π-π interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J. Mol. Biol.* **1991**, *218*, 837−846.
(21) Bruno, I. J.; Cole, J. C.; Lommerse, J. P.; Rowland, R. S.; Taylor, R.; Verdonk, M. L. IsoStar: A library of information about nonbonded interactions. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 525−537.
(22) Burley, S. K.; Petsko, G. A. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **1985**, *229*, 23−28.
(23) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. Estimates of the Ab Initio Limit for π-π Interactions: The Benzene Dimer. *J. Am. Chem. Soc.* **2002**, *124*, 10887−10893.
(24) Sinnokrot, M. O.; Sherrill, C. D. Substituent Effects in π−π Interactions: Sandwich and T-Shaped Configurations. *J. Am. Chem. Soc.* **2004**, *126*, 7690−7697.
(25) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. The Magnitude of the CH/π Interaction between Benzene and Some Model Hydrocarbons. *J. Am. Chem. Soc.* **2000**, *122*, 3746−3753.
(26) Muraki, M. The Importance of CH Interactions to the Function of Carbohydrate Binding Proteins. *Protein Pept. Lett.* **2002**, *9*, 195−209.
(27) Schmidt, G. M. J. Topochemistry. III. The crystal chemistry of some trans-cinnamic acids. *J. Chem. Soc.* **1964**, 2014−2021.
(28) Codding, P. W. Halogen-substituted Drugs and their Intermolecular Interactions. Presented at the XX Congress of the International Union of Crystallography, Florence, Italy, 23−31 August 2005. http://xxiucr.iccom.cnr.it/pdf/762.pdf.
(29) Brammer, L.; Bruton, E. A.; Sherwood, P. Understanding the behavior of halogens as hydrogen bond acceptors. *Cryst. Growth Des.* **2001**, *1*, 277−290.
(30) Dunitz, J. D. Organic fluorine: odd man out. *Chembiochem* **2004**, *5*, 614−621.
(31) Lommerse, J. P. M.; Stone, A. J.; Taylor, R.; Allen, F. H. The Nature and Geometry of Intermolecular Interactions between Halogens and Oxygen or Nitrogen. *J. Am. Chem. Soc.* **1996**, *118*, 3108−3116.
(32) van den Berg, J. A.; Seddon, K. R. Critical evaluation of C-H···X hydrogen bonding in the crystalline state. *Cryst. Growth Des.* **2003**, *3*, 643−661.
(33) Price, S. L.; Stone, A. J.; Lucas, J.; Rowland, R. S.; Thornley, A. E. On the nature of −Cl···Cl− intermolecular interactions. *J. Am. Chem. Soc.* **1994**, *116*, 4910−4918.
(34) Morgan, R. S.; Tatsch, C. E.; Gushard, R. H.; McAdon, J. M.; Warme, P. K. Chains of alternating sulfur and π-bonded atoms in eight small proteins. *Int. J. Pept. Protein Res.* **1978**, *11*, 209−217.
(35) Morgan, R. S.; McAdon, J. M. Predictor for sulfur-aromatic interactions in globular proteins. *J. Pept. Protein Res.* **1980**, *15*, 177−180.
(36) Reid, K. S. C.; Lindley, P. F.; Thornton, J. M. Sulfur aromatic interactions in proteins. *FEBS Lett.* **1985**, *190*, 209−213.
(37) Zauhar, R. J.; Colbert, C. L.; Morgan, R. S.; Welsh, W. J. Evidence for a strong sulfur-aromatic interaction derived from crystallographic data. *Biopolymers* **2000**, *53*, 233−248.
(38) Nemethy, G.; Scheraga, H. A. Strong interaction between disulfide derivatives and aromatic groups in peptides and proteins. *Biochem. Biophys. Res. Commun.* **1981**, *98*, 482−487.
(39) Cheney, B. V.; Schulz, M. W.; Cheney, J. Complexes of benzene with formamide and methanethiol as models for interactions of protein substructures. *Biochim. Biophys. Acta* **1989**, *996*, 116−124.
(40) Pranata, J. Sulfur aromatic interactions: A computational study of the dimethyl sulfide benzene complex. *Bioorg. Chem.* **1997**, *25*, 213−219.
(41) Allen, F. H.; Bird, C. M.; Rowland, R. S.; Raithby, P. R. Hydrogen-bond acceptor and donor properties of divalent sulfur (Y-S-Z and R-S-H). *Acta Crystallogr.* **1997**, *B53*, 696−701.
(42) Oimentel, G. C.; McClellan, A. L. *The hydrogen bond*; W.H.Freeman: San Francisco, 1960.
(43) Petersen, M. T.; Jonson, P. H.; Petersen, S. B. Amino acid neighbours and detailed conformational analysis of cysteines in proteins. *Protein Eng.* **1999**, *12*, 535−548.
(44) Pal, D.; Chakrabarti, P. Different types of interactions involving cysteine sulfhydryl group in proteins. *J. Biomol. Struct. Dyn.* **1998**, *15*, 1059−1072.
(45) Zauhar, R. J.; Colbert, C. L.; Morgan, R. S.; Welsh, W. J. Evidence for a strong sulfur-aromatic interaction derived from crystallographic data. *Biopolymers* **2000**, *53*, 233−248.
(46) Pal, D.; Chakrabarti, P. Non-hydrogen bond interactions involving the methionine sulfur atom. *J. Biomol. Struct. Dyn.* **2001**, *19*, 115−128.
(47) Li, Y.; Snyder, L. B.; Langley, D. R. Electrostatic interaction of π-acidic amides with hydrogen-bond acceptors. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3261−3266.
(48) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **1995**, *247*, 536−540.
(49) 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.
(50) Sarkhel, S.; Desiraju, G. R. N-H···O, O-H···O, and C-H···O hydrogen bonds in protein-ligand complexes: strong and weak interactions in molecular recognition. *Proteins* **2004**, *54*, 247−259.
(51) Cotesta, S.; Stahl, M. The environment of amide groups in protein-ligand complexes: H-bonds and beyond. *J. Mol. Model.* (online) **2005**, 1−9.
(52) Jiang, L.; Lai, L. CH···O hydrogen bonds at protein-protein interfaces. *J. Biol. Chem.* **2002**, *277*, 37732−37740.
(53) Manikandan, K.; Ramakumar, S. The occurrence of C−H···O hydrogen bonds in α-helices and helix termini in globular proteins. *Proteins* **2004**, *56*, 768−781.
(54) Singh, S. K.; Babu, M. M.; Balaram, P. Registering α-helices and β-strands using backbone C-H···O interactions. *Proteins* **2003**, *51*, 167−171.
(55) Scheiner, S.; Kar, T.; Gu, Y. Strength of the CαH···O hydrogen bond of amino acid residues. *J. Biol. Chem.* **2001**, *276*, 9832−9837.
(56) Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; Hilgenfeld, R. C-H···π-interactions in proteins. *J. Mol. Biol.* **2001**, *307*, 357−377.
(57) Relibase: http://relibase.ccdc.cam.ac.uk/.
(58) Protein-Ligand, Database (PLD): http://www-mitchell.ch.cam.ac.uk/pld/. Puvanendrampillai, D.; Mitchell, J. B. L/D Protein Ligand Database (PLD): additional understanding of the nature and specificity of protein-ligand complexes. *Bioinformatics* **2003**, *19*, 1856−1857.

(59) Miyazawa, S.; Jernigan, R. L. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **1985**, *18*, 534−552.
(60) Cluster list: ftp://ftp.rcsb.org/pub/pdb/derived_data/NR/clusters90.txt.
(61) PyMOL: http://pymol.sourceforge.net/. DeLano Scientific LLC, South San Francisco, CA.

JM061038A