Check for updates

A systematic analysis of atomic protein–ligand interactions in the PDB†

Renato Ferreira de Freitas [ID][a] and Matthieu Schapira [ID]*[ab]

As the protein databank (PDB) recently passed the cap of 123 456 structures, it stands more than ever as an important resource not only to analyze structural features of specific biological systems, but also to study the prevalence of structural patterns observed in a large body of unrelated structures, that may reflect rules governing protein folding or molecular recognition. Here, we compiled a list of 11 016 unique structures of small-molecule ligands bound to proteins – 6444 of which have experimental binding affinity – representing 750 873 protein–ligand atomic interactions, and analyzed the frequency, geometry and impact of each interaction type. We find that hydrophobic interactions are generally enriched in high-efficiency ligands, but polar interactions are over-represented in fragment inhibitors. While most observations extracted from the PDB will be familiar to seasoned medicinal chemists, less expected findings, such as the high number of C–H⋯O hydrogen bonds or the relatively frequent amide–π stacking between the backbone amide of proteins and aromatic rings of ligands, uncover underused ligand design strategies.

Introduction

Significant progress in high-throughput X-ray crystallography[1,2] combined with advances in structural genomics[3–5] have led to an explosion in the number of structures publicly available in the protein data bank (PDB).[6] At the time this manuscript was written, more than 123 456 structures had been deposited in the PDB,[6] including 76 056 protein–small molecule complexes, of which 13 000 have a reported binding potency.[7,8] This large body of data contains important information on the nature, geometry, and frequency of atomic interactions that drive potent binding between small molecule ligands and their receptors. Systematic analysis of this data will lead to a better appreciation of intermolecular interactions between proteins and their ligands and can inform structure-based design and optimization of drugs.[9]

Several approaches have been developed for large-scale analysis of protein–small molecule interactions, such as SuperStar, or the method implemented to build the Relibase database.[10,11] PDBeMotif[12] and the recently published PELIKAN[13] are two examples of free tools that can search for patterns in large collections of protein–ligand interfaces. Structural interaction fingerprints (SIF)[9] are another method of representing and analyzing 3D protein–ligand interactions where the pres-

ence or absence of interactions between distinct residues and ligand atoms are represented as bit strings that can be compared rapidly.[14] In addition, there has been an increase in the number of free tools to fully automate the detection and visualization of relevant non-covalent protein–ligand contacts in 3D structures.[15–17]

A statistical analysis of the nature, geometry and frequency of atomic interactions between small molecule ligands and their receptors in the PDB could inform the rational optimization of chemical series, help in the interpretation of difficult SAR, aid the development of protein–ligand interaction fingerprints, and serve as a knowledge-base for the improvement of scoring functions used in virtual screening. To the best of our knowledge, such public resource is currently missing.

Here, we analyze the frequency of common atomic interactions between protein and small molecules observed in the PDB. We find that some interactions occur more frequently in fragments than drug-like compounds, or in high-efficiency ligands than low-efficiency ligands. We next review in detail each of the most frequent interactions and use matched molecular pairs to illustrate the impact of these atomic interactions on binding affinity.

Most frequent protein–ligand atomic interactions

We extracted from the PDB all X-ray structures of small-molecules in complex with proteins, with a resolution ≤2.5 Å, resulting in a collection of 11 016 complexes. To be

[a] *Structural Genomics Consortium, University of Toronto, Toronto, ON M5G 1L7, Canada. E-mail: matthieu.schapira@utoronto.ca*
[b] *Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7md00381a

considered as a ligand, the compound had to meet several criteria such as being a small molecule and be of interest for medicinal chemistry applications (buffers or part of crystallization cocktails were excluded. See ESI† for more details). This collection contained 750 873 ligand–protein atom pairs, where a pair of atoms is defined as two atoms separated by 4 Å or less. The top-100 most frequent ligand–protein atom pairs (Table S1†) can be clustered into seven interaction types (Fig. 1). Among the most frequently observed are interactions that are well known and widely used in ligand design such as hydrophobic contacts, hydrogen bonds and π-stacking.[18,19] These are followed by weak hydrogen bonds, salt bridges, amide stacking, and cation–π interactions.

More than 500 protein families were present in our dataset. The distribution of the ten most frequent protein families (Fig. 2a) shows that kinases are overrepresented with 1588 structures, followed by trypsin-like serine and aspartyl proteases with 637 and 631 structures, respectively. The top-10 protein families were all enzymes with the exception of the nuclear hormone receptor and the bromodomain families.

We selected three unrelated protein families to evaluate the differences in interaction frequencies. First, we observed that the relative frequency of salt bridge and cation–π interactions was very low in all families (Fig. 2b). The relative frequency of π-stacking interactions was similar among the three families ranging from 12% to 16%. On the other hand, weak hydrogen bonds were two times more frequent in kinases and trypsin-like proteins than in nuclear hormone receptor (weak hydrogen bonds are frequently observed between kinase inhibitors and the canonical hinge region of kinases, as discussed below). Finally, the most striking finding was that the relative frequency of hydrogen bonds and amide stacking interactions were much higher in trypsin-like

proteins (42% and 32%) than in kinases (16% and 6%) and nuclear hormone receptors (10% and 4%). In fact, the trypsin-like family alone contributed to 25.5% of all the amide stacking interactions. The lower frequency of hydrogen bonds and amide stacking interactions in kinases and nuclear hormone receptors reflects the fact that the binding pocket of these protein families are more hydrophobic and that the polar π-surface of protein amide groups is less exposed than in trypsin-like proteins.

We next asked whether some interactions types were more frequently observed in high-efficiency ligands. Experimental binding affinity for 6444 protein–ligands in the PDB were retrieved from the PDBbind database,[7,8] and a fit quality (FQ) score – a size-adjusted calculation of ligand efficiency – was used to evaluate how optimally a ligand binds relative to other ligands of any size.[20] The frequency of each interaction type was calculated for the 1500 protein–ligand complexes with the best FQ score (FQ > 0.81) and the 1500 complexes with the worst FQ scores (FQ < 0.54) (Fig. 3a).

We find that hydrophobic interactions are more frequent in high-efficiency ligands. In particular, the frequency of hydrogen bonds is reduced from 59% to 34% of that of hydrophobic contacts in efficient binders, and the frequency of salt bridges is more than halved, from 13% to 7% (Fig. 3a). This observation probably reflects the fact that most ligands in the PDB are the product of lead optimization strategies that aim at increasing the number of favorable hydrophobic interactions, which is less challenging than optimizing directionality-constrained hydrogen-bonds (discussed in more details in a later section).[21] We also find that efficient ligands are more hydrophobic, as the median number of heavy atoms and log D (ChemAxon) for compounds with high FQ are 27 and 1.7, respectively, and 21 and 0.2, respectively, for compounds with low FQ. Both groups showed similar profiles for other properties like polar surface area (median PSA: 95.3 vs. 89.8 Å^2), hydrogen bond acceptors (median HBA: 5 for both), and hydrogen bond donors (median HBD: 2 for both). Taken together these results show that small-molecule ligands that bind their target with high efficiency are more hydrophobic, and that hydrophobic interactions are a driving factor for the increased ligand efficiency.

Since fragments are typically binding their targets with higher ligand efficiency than larger ligands, we asked whether hydrophobic interactions were also more frequent in protein-fragment complexes. The frequency of each interaction type was calculated for two random groups of 1500 protein–ligand complexes, one with fragment molecules (HA ≤ 20), the other with drug-like molecules (30 ≤ HA ≤ 50) (Fig. 3b). Unlike high-efficiency ligands, we find that protein-fragment complexes are enriched in polar interactions: the frequency of hydrogen bond is doubled, from 31% to 62%, compared to that of hydrophobic contacts, and the frequency of electrostatic interactions is multiplied by three, from 5% to 17% (Fig. 3b). To compensate for their low number, interactions made by fragments need to be highly efficient.[22] We note that electrostatic interactions define maximum efficiency of ligand binding.[23]
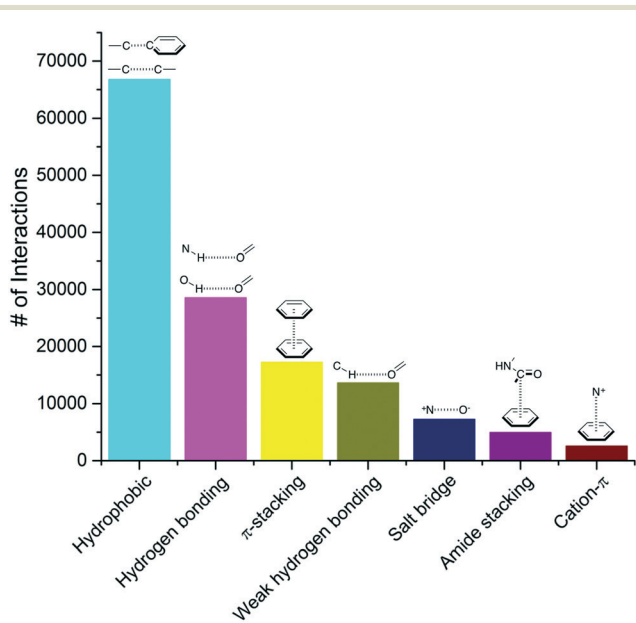

Fig. 1 Frequency distribution of the most common non-covalent interactions observed in protein–ligands extracted from the PDB.

This journal is © The Royal Society of Chemistry 2017
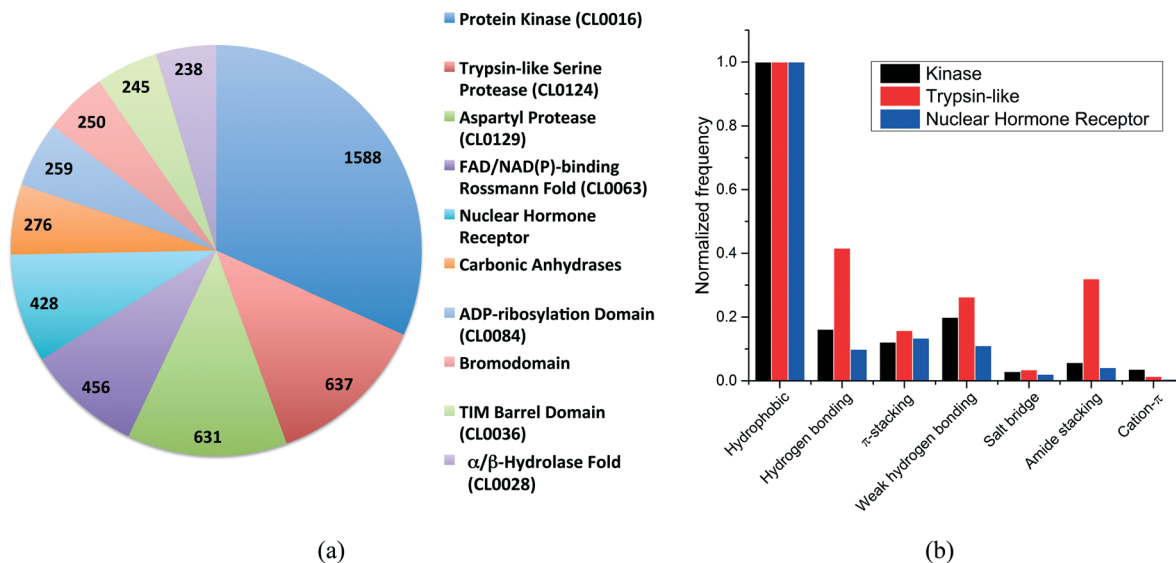
Med. Chem. Commun., 2017, 8, 1970–1981 | 1971

Fig. 2 (a) Distribution of the ten most frequent protein families in the dataset. The CLAN number is provided (when available) in parenthesis; (b) frequency distribution of the most common non-covalent interactions observed for the three unrelated protein families.
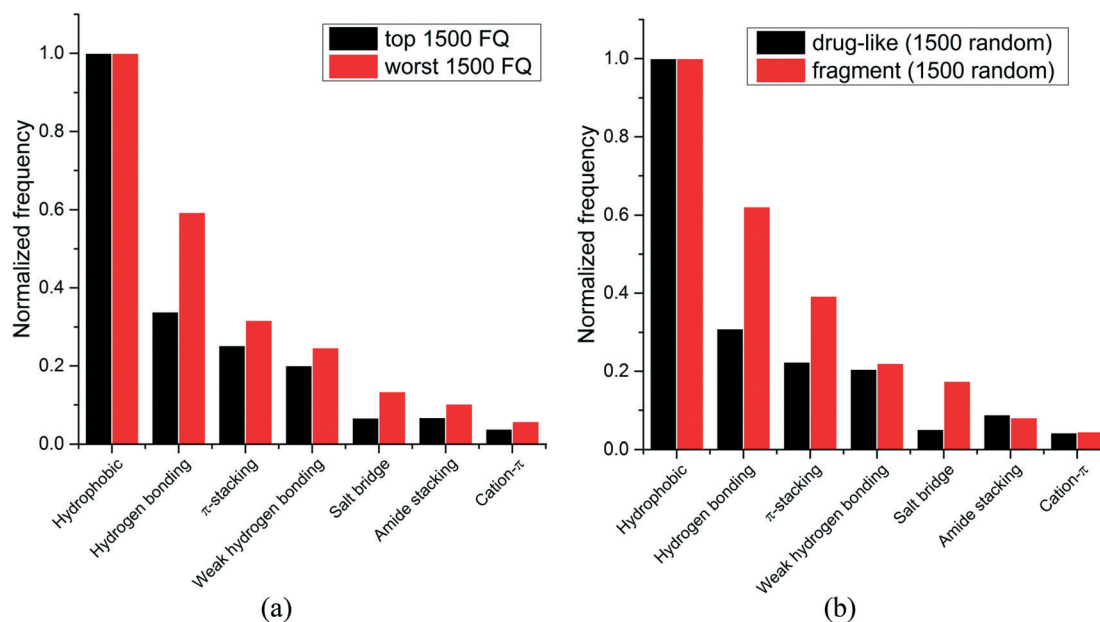

Fig. 3 Relative frequency distribution of the most common non-covalent interactions observed in: (a) ligands with high *vs.* low fit quality (FQ); (b) fragments *vs.* drug-like compounds. (1500 random molecules were selected for each group).

The higher prevalence of polar interactions in fragments compared to drug-like compounds could be seen as a requirement for high solubility, as fragments are tested at high concentrations. It also reflects the fact that fragments are freer than larger compounds to adopt binding poses that will optimally satisfy the geometric constraints of high-efficiency interactions, such as electrostatic or hydrogen-bonds.[24]

Together, these results show that fragments are using polar interactions to gain maximum binding efficiency from a limited number of interactions, but as small-molecule ligands are optimized, geometric constraints associated with

polar bonds are more challenging to satisfy, and the contribution of hydrophobic interactions increases.

To gain further insight, we next analyzed in detail the composition, geometry, frequency, protein side-chain preference, and impact towards binding affinity of each protein–ligand interaction type in the PDB.

Specific intermolecular interactions

Hydrophobic interactions

From our analysis, hydrophobic contacts are by far the most common interactions in protein–ligand complexes, totalizing

1972 | *Med. Chem. Commun.*, 2017, **8**, 1970–1981

This journal is © The Royal Society of Chemistry 2017

66 772 contacts between a carbon and a carbon, halogen or sulfur atom (the distance cut-off of 4.0 Å allows the implicit inclusion of hydrogen atoms) (Fig. 1). Hydrophobic interactions were separated into five groups (Table S2†). The most populated group is the one formed by an aliphatic carbon in the receptor and an aromatic carbon in the ligand, which alone accounts for more than 42 000 interactions (Table S2†). This is an indication that aromatic rings are prevalent in small molecule inhibitors. In fact, 76% of the marketed drugs contain one or more aromatic ring, with benzene being by far the most frequently encountered ring system.[25,26] Not surprisingly, leucine, followed by valine, isoleucine and alanine side-chains are the most frequently engaged in hydrophobic interactions (Fig. S4†).

Contacts involving an aromatic or aliphatic carbon in the receptor and an aliphatic carbon in the ligand were observed in 8899 and 8974 instances, respectively (Table S2†). We observed that aliphatic carbons were distributed mostly above or below the plane of the aromatic ring, rather than at the edge (Fig. S5†). Interactions involving an aliphatic or aromatic carbon in the protein and a chlorine or fluorine in the ligand were the second most common hydrophobic contacts (observed in 5147 complexes) followed by interactions involving a sulfur atom from the side chain of methionine and an aromatic carbon from the ligand (observed in 1309 complexes) (Table S2†). Although methionine is classified as a hydrophobic residue, a recent study shows that the Met S···C(aro) interaction yields an additional stabilization energy of 1–1.5 kcal mol^{-1} compared with a purely hydrophobic interaction.[27]

Hydrophobic interactions are the main driving force in drug–receptor interactions. The benefit of burying a solvent-exposed methyl group on a ligand into a hydrophobic pocket of a protein is about 0.7 kcal mol^{-1} or a 3.2-fold increase in binding constant per methyl group.[28] However the effect of replacing a hydrogen atom with a methyl group is highly context dependent, and potency losses are as common as gains. Ten-fold and 100-fold gains in potency are observed in 8% and 0.5% of cases, respectively.[29,30] For instance, addition of a single methyl group improves by 50 fold the potency of a tankyrase-2 (TNKS2) inhibitor.[31] The added methyl group occupies a small hydrophobic cavity and potentially releases unfavorably bound water molecules (Fig. 4b). In rare cases, the increase in potency due to the introduction of a "magic methyl" exceeds two orders of magnitude.[32] This is generally due to the combined entropic effect of lowering the conformational penalty paid by the ligand upon binding, and the desolvation effect of burying the methyl group in a hydrophobic pocket.[29]

Hydrogen bonds

We find that hydrogen bonds were the second most frequent type of interactions observed in our collection of protein–ligand complexes, with a total of 28 577 (Fig. 1). N–H···O interactions were more frequent (15 105 interactions) than O–
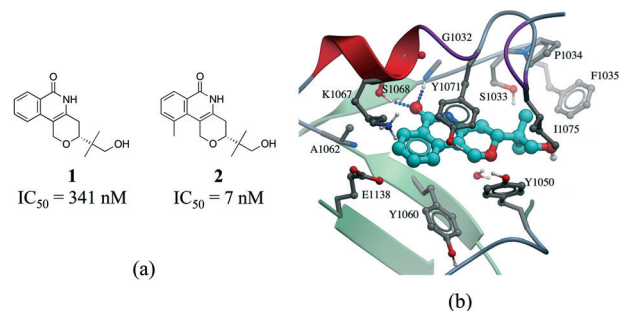

Fig. 4 Magic methyl effect: (a) chemical structure of two TNKS2 inhibitors; (b) crystal structure of **1** (carbon atoms in cyan) bound to TNKS2 (PDB: 5C5P).

H···O (8251 interactions) and N–H···N (333 interactions) (Table S2†). Among the N–H···O interactions, the number of neutral and charged hydrogen bonds were almost equal (7554 vs. 7551, respectively). Proteins were more often hydrogen-bond donors than acceptors (9217 vs. 5888, respectively). Surprisingly, glycine was the most frequent hydrogen-bond acceptor, and the second most frequent donor, probably due to the absence of side-chain to mask backbone atoms, and increased backbone flexibility to better satisfy the spatial constraints of hydrogen-bonds (Fig. S6†). Arginines were engaged in more hydrogen-bonds than lysines, probably reflecting the presence of 3 nitrogen atoms in the guanidinium group of arginine side-chains (Fig. S6†). Among O–H···O interactions, charged hydrogen bonds (typically between an alcohol and a carboxylic acid) were 3 times more frequent than neutral ones, and ligands more often behaved as donors than acceptors (Table S2†). The most common acceptors were aspartic acids in charged hydrogen bonds, and asparagine, glycine and glutamine in neutral interactions (Fig. S7†). Serine was the most usual donor (Fig. S7†). Finally, a total of 4888 protein–ligand hydrogen bonds mediated by water were observed in our analysis. Of these, water-mediated hydrogen bonds involving an oxygen in the ligand were roughly two times more frequent than those involving a nitrogen (3131 vs. 1757).

We found that heavy atoms in N–H···O, N–H···N, and O–H···O hydrogen bonds were all separated by similar median distances of approximately 3.0 Å (Fig. 5). This value is slightly higher (~0.1–0.2 Å) than previously reported for hydrogen bonds between amide C=O and OH/NH.[33] In addition, the median distances of neutral and charged hydrogen bonds were almost identical (0.1 Å difference, data not shown). The D–H···A angle usually peaked at 130–180°, and the preferred angle for N–H···O hydrogen bonds was around 180° (data not shown).

Hydrogen bonds are the prevailing directional intermolecular interactions in biological complexes,[34,35] and the predominant contribution to the specificity of molecular recognition.[36] The free energy for hydrogen bonding can vary between −1.5 kcal mol^{-1} to −4.7 kcal mol^{-1}.[28] However, the contribution of a hydrogen bond to binding can be very modest (or penalizing) if the new interaction formed does not
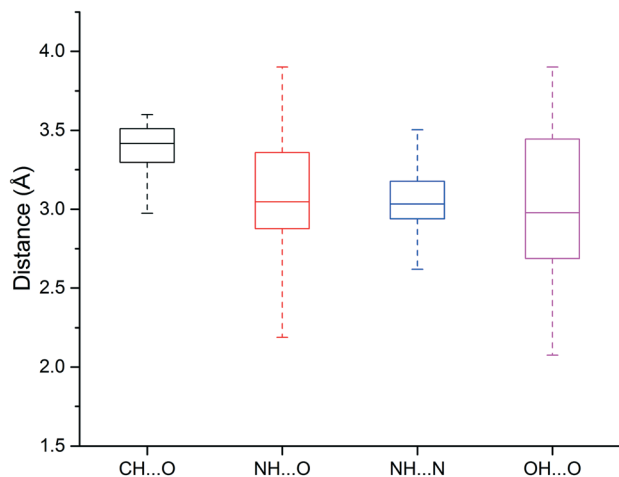
This journal is © The Royal Society of Chemistry 2017

Med. Chem. Commun., 2017, **8**, 1970–1981 | **1973**

Fig. 5 Box plot of hydrogen bond length distributions for the weak (C–H⋯O) and strong hydrogen bonds (N–H⋯O, N–H⋯N, O–H⋯O).

Fig. 6 Effect of adding a hydrogen bond in a thrombin inhibitor: a) chemical structure of a pair of thrombin inhibitors; b) crystal structure of **4** (cyan carbons) in complex with thrombin (PDB: 2ZC9). Hydrogen bonds are displayed in dotted green lines.

outweigh the desolvation penalty upon ligand binding.[37] Also, the contribution of a hydrogen bond is dependent on the local environment: a solvent-exposed hydrogen-bond contributes significantly less to net interaction energy than the same hydrogen-bond in a buried hydrophobic pocket.[38] Consequently, optimizing hydrophobic interactions is generally considered easier than hydrogen bonds.[28] In drug design, hydrogen bonds are exploited to gain specificity owing to their strict distance and geometric constraints.[39]

Among numerous examples, a series of potent thrombin inhibitors shows a remarkable increase in binding affinity (>500-fold) through simple addition of hydrogen-donating ammonium group (Fig. 6a).[40] In the crystal structure, the ammonium group forms a charge-assisted hydrogen bond with the carbonyl oxygen of Gly216 and surrounding waters (Fig. 6b).[41]

π-Stacking interactions

The third most frequent protein–ligand contacts in the PDB were aromatic interactions (Fig. 1). Interactions involving aromatic rings are ubiquitous in chemical and biological systems and can be considered a special case of hydrophobic interactions.[42] We found that edge-to-face and face-to-face interactions were equiprobable (8704 and 8537 contacts respectively) (Table S2†). This is in agreement with quantum mechanical calculations of the interaction energy of benzene dimers that predict the edge-to-face and parallel displaced face-to-face as being isoenergetic, and more stable than the eclipsed face-to-face π-stacking.[43] Almost 50% of all π-stacking interactions are observed between the aromatic ring of phenylalanine and an aromatic ring in the ligand, followed by tyrosine (36.8%), tryptophan (8.7%) and histidine (5.1%) (Fig. S8†).

Interactions involving aromatic rings are major contributors to protein–ligand recognition and concomitantly to drug design.[42,44] An example of the strong gain in binding affinity
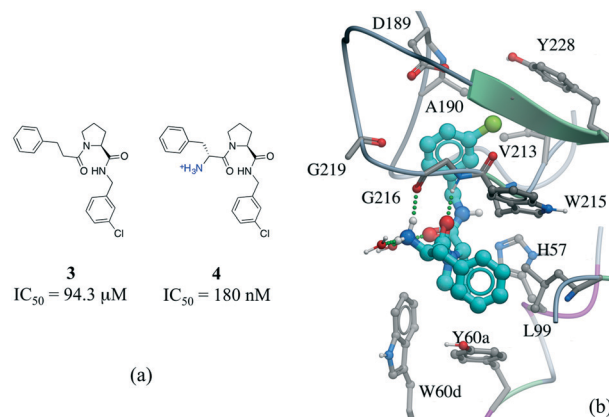
that can be obtained by forming a π-stacking interaction is illustrated in a series of soluble epoxide hydrolase (sEH) inhibitors.[45] In the X-ray cocrystal structure of human sEH and **6** (IC$_{50}$ = 7 nM), the phenyl ring is positioned to allow π-stacking interaction with H524 (Fig. 7b), while an analog (**5**, IC$_{50}$ = 700 nM) without the phenyl ring is l00-fold less potent. While π-stacking interaction can increase the binding affinity of the inhibitor for its target, it has been pointed out that reducing the number of aromatic rings of a molecule might improve its physicochemical properties, such as solubility.[46,47]

Weak hydrogen bonds

The fourth most frequent interactions (13 600 contacts) were C–H⋯O hydrogen bonds, the existence of which is well documented (Fig. 1).[48,49] When the interacting carbon was aromatic, protein oxygens were found to be acceptors much more often than ligand oxygen atoms (4927 vs. 708 interactions, Table S2†). This simply reflects the fact that most
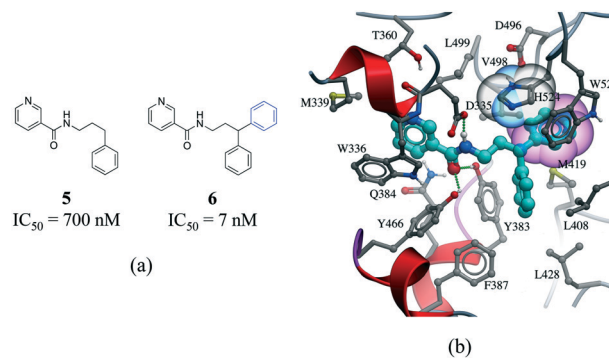

Fig. 7 a) Chemical structure of two inhibitors of human sEH; b) X-ray cocrystal structure of human sEH and **6** (cyan carbons, PDB: 3I1Y). The phenyl ring (transparent CPK magenta) is positioned to allow a π-stacking interaction with H524 (shown as transparent CPK). Hydrogen bonds are displayed in dotted green lines.

1974 | *Med. Chem. Commun.*, 2017, **8**, 1970–1981

This journal is © The Royal Society of Chemistry 2017

ligands have aromatic rings, while most side-chains don't. Glycine, aspartic acid and glutamic acid were always the most frequent acceptors in C–H⋯O interactions, while leucine was the most frequent donor (Fig. S9†).

The median distance of the C–H⋯O hydrogen bonding was 3.4 Å, which is 0.4 Å longer than traditional hydrogen bonds (N–H⋯O, N–H⋯N, O–H⋯O), and distances separating the two heavy atoms were rarely lower than 3.2 Å (Fig. 3). The angle distribution of C–H⋯O interactions peaked around 130° (data not shown), which is in agreement with previous work.[50]

The existence of weak hydrogen bonds has been extensively analyzed and reviewed.[51–54] Calculations indicate that the magnitude of the C_α–H⋯O=C interactions are about one-half the strength of an NH⋯O=C hydrogen bond.[55] In addition, an analysis of protein–ligand complexes revealed that C_α–H⋯O hydrogen should be better interpreted as secondary interactions, as they are frequently accompanied by bifurcated N–H⋯O hydrogen bonds.[56] However, it is increasingly recognized that C–H⋯O hydrogen bonds play an important role in molecular recognition processes,[57] protein folding stabilization,[58] in the interaction of nucleic acids with proteins,[59] in enzyme catalysis,[60] and in the stabilization of protein–ligand binding complexes.[61,62] A matched pair of CDK2 inhibitors illustrates the contribution of C–H⋯O hydrogen bonds to protein–ligand complexes (Fig. 8).[63] The only difference between the two inhibitors is the substitution of a NH_2 by a methyl group on the thiazole ring of compound 8 (Fig. 8a). Although the N–H⋯O hydrogen bond of 7 is stronger than the C–H⋯O hydrogen bond of 8, the latter compound is more potent probably due to the penalty associated with desolvating the NH_2 of 7 upon binding (Fig. 8b).

Salt bridges

The contact between a positively charged nitrogen and a negatively charged oxygen (*i.e. salt bridge*) was the fifth most frequent interaction type in our analysis (7276 interactions) (Fig. 1). The number of salt bridge interactions with a positive nitrogen coming from the protein and the negative oxygen coming from the ligand was two times higher than the

opposite (4882 *vs.* 2394 interactions, Table S2†). This probably reflects the higher number of ligands containing carboxylic acids (1849) than ammonium groups (1103) in the PDB, as the frequency of arginine (5.6%) and lysine (5.0%) in proteins is similar to that observed for aspartic acid (5.4%) and glutamic acid (3.8%) (UniProtKB/TrEMBL UniProt release 2017_03).[64] Arginine was the cation in 83.6% of all interactions (Fig. S10†). This seems to be agreement with quantum mechanical calculations, which predict that arginine are more inclined than lysine side-chains to form salt bridges.[65] Finally, the distribution of negatively charged oxygens around the guanidinium group of arginine shows a higher density around the terminal (ω) nitrogens than at the secondary amine (ε) nitrogen (Fig. S10†).

Salt bridges contribute little to protein stability as the favorable binding energy obtained from forming a salt bridge is not sufficient to offset the energetic penalty of desolvating charged groups.[66,67] However, the strength of salt bridge interactions is strongly dependent on the environment. In particular, buried salt-bridges can make crucial contributions to ligand binding.[68–70] For example, the terminal *N,N*-dimethylamino tail of 10 forms a salt bridge with D831 in the kinase domain of epidermal growth factor receptor (EGFR) (Fig. 9). When the nitrogen atom of the terminal *N,N*-dimethylamino group was replaced with a carbon (9) potency was reduced by more than 800-fold.[71]

Amide⋯π stacking

Interactions between an amide group and an aromatic ring were the sixth most frequently observed (Fig. 1). In these interactions, which are related to canonical aromatic π-stacking, the π-surface of the amide bond stacks against the π-surface of the aromatic ring.[72,73] As previously observed for π–π stacking interactions, we did not find significant preference for face-to-face over edge-to-face arrangement (2907 and 2060 interactions respectively) (Table S2†). The most frequent amino acids participating in face-to-face amide⋯π stacking were glycine (19.4%) and tryptophan (17.9%), while glycine (20.1%) and leucine (13.0%) were the most often observed in edge-to-face geometry (Fig. S11†). The fact that
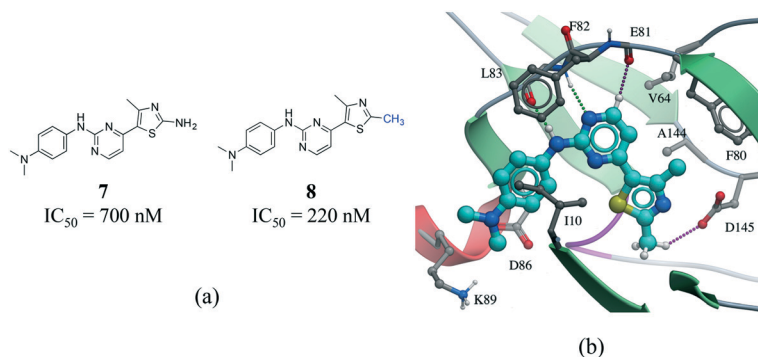

Fig. 8 a) Chemical structure of two CDK2 inhibitors; b) X-ray cocrystal structure of the human CDK2 and 8 (PDB: 1PXP, cyan carbons). The N–H⋯O and CH⋯O hydrogen bonds are displayed as green and magenta dotted lines, respectively.
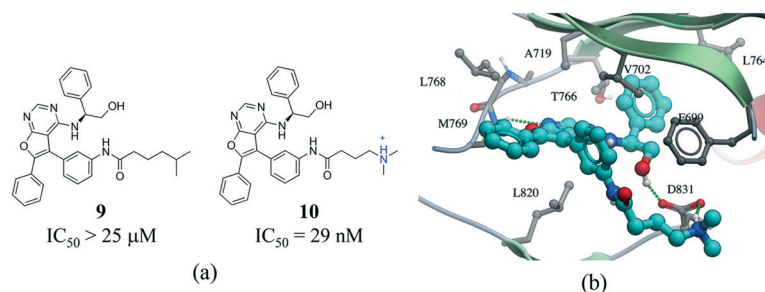
Fig. 9 a) Chemical structure of two inhibitors of human EGFR; b) X-ray cocrystal structure of the kinase domain of EGFR and **10** (PDB: 4JRV, cyan carbons), the terminal *N,N*-dimethylamino tail of **10** forms a salt bridge with D831. Hydrogen bonds are displayed in dotted green lines.

88.5% of all amide⋯π stacking interactions occurred between the backbone amide group of a protein (generally a glycine) and the aromatic ring of a ligand points at a strategy to exploit peptide bonds in binding sites that is probably underused in structure-based drug design.

Amide⋯π stacking interactions are common and significant in protein structures.[74] These interactions were also shown to sometimes play an important role in ligand binding.[75–77] For example, the 11-fold difference in K_i between a matched pair of oxazole-containing factor Xa inhibitors was attributed to the influence of the dipole of the oxazole ring on the amide⋯π interaction (Fig. 10).[78]

Cation–π

We found 2577 interactions between a positively charged nitrogen and an aromatic ring (Fig. 1). These cation–π interactions are essentially electrostatic due to the negatively charged electron cloud of π systems.[79] In more than 90% of these interactions, the nitrogen came from the receptor and

the aromatic ring from the ligand, reflecting, as previously noted, that drug-like compounds have often aromatic rings while ammonium groups are more rare (Table S2†). Arginines were 3 times more frequently engaged in cation–π interactions than lysine side-chains. (Fig. S12†). A similar trend was previously observed for peptidic interactions.[80] This preference has been attributed to the fact that the guanidinium group of arginines can donate several hydrogen bonds while simultaneously binding to an aromatic ring.[73] When the positive nitrogen came from the ligand, tyrosine side-chains were the most common partner with 156 interactions, followed by phenylalanine and tryptophan (59 and 24 interactions respectively) (Fig. S12†). Potentiation of the cation–π binding ability of the tyrosine upon hydrogen bonding of its hydroxyl group was proposed to be at the origin of a similar bias in peptidic interactions.[80]

Cation–π interactions are widespread in proteins and are important determinants of the structure, stability, and function of proteins.[81] An example that is especially compelling is the Royal family of epigenetic reader proteins, that feature an
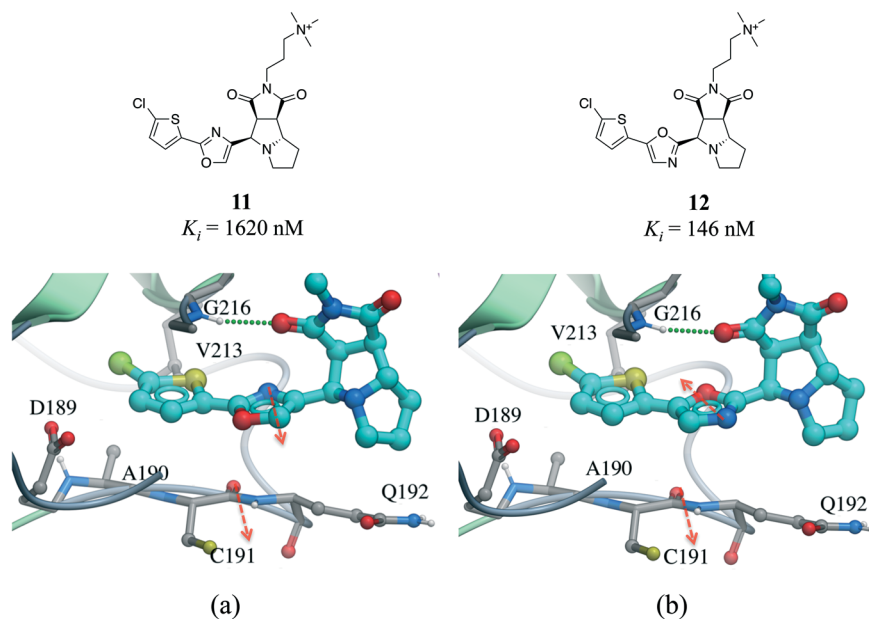

Fig. 10 X-ray cocrystal structure of (a) **11** (PDB: 2Y5H) and (b) **12** (PDB: 2Y5G) bound at the active site of factor Xa. The amide⋯π stacking interaction is shown as dotted green lines. The dipoles of the oxazole ring and peptide amide (red arrows) are parallel in **11** and anti-parallel in **12**.

aromatic cage composed of two to four aromatic residues that make cation–π and hydrophobic interactions with postranslationally methylated lysines or arginines side-chains.[82]

Many drug–receptor interactions involve cation–π interactions. One of the earliest examples is the recognition of acetylcholine (ACh) by the nicotinic acetylcholine receptor (nAChR). Similarly, GABA,[83] glycine,[84] and 5-HT$_3$ (ref. 85) receptors have all been shown to participate in cation–π interactions with neurotransmitters. In a series of insightful experiments, sequential methylation of an ammonium group in a series of potent factor Xa inhibitors gradually increased the binding affinity by 3 orders of magnitude.[86] Comparing the affinity of a *tert*-butyl analog (compound **15**) with the trimethylated ammonium group (compound **17**), indicated that the cation–π interaction contributed to a 60 fold increase in potency (Fig. 11).

Halogen bonding

Although specific interactions involving halogen atoms were much less frequent than the other interactions discussed above we included them in our analysis as the impact of these interactions is regularly debated among medicinal chemists.[87–89]

We found 351 interactions of the type C–X⋯Y (X = Cl, Br, I; Y = O, N, S) where Y was either from protein side chain or backbone. These halogen bonding (XB) interactions[90–92] occur between the σ-hole (positive electrostatic potential) of a halogen atom (XB donor) and a nucleophile (XB acceptor).[93–95] Fluorine is not able to form halogen bonding interactions due to its higher electronegativity and lack of polarizability, and only heavier halogens (Cl, Br, and I) are considered in the analysis.[96]

From the 351 interactions, those involving a chlorine atom were the most frequent (222 interactions), followed by bromine (91 interactions) and iodine (38 interactions) (Table S2†). This is in agreement with other surveys and reflects the relative prevalence of these three halogen atoms in small

molecule ligands.[97,98] The C–X⋯Y angle had a median value of 156°, indicating a preferred near linear arrangement. Oxygen atoms were by far the most common XB acceptors (~90% of all interactions), followed by sulfur (~9%) and nitrogen (~1%) (Table S2 and Fig. S13†). Overall, approximately 71% (251 interactions) of all halogen bonds were engaged with backbone carbonyl oxygen atoms, while asparagine, proline, arginine, and tryptophan residues were underrepresented (Fig. S13†).

Halogen bonds are well-characterized intermolecular interactions in small molecules, and have many applications in fields as diverse as crystal engineering and supramolecular chemistry.[99,100] The introduction of halogens in small molecules is largely used in medicinal chemistry programs to increase not only the affinity but also the membrane permeability and metabolic stability of compounds. Usually, the insertion of halogen atoms on lead compounds is used to explore their steric and electronic effects.[101] Only recently was it recognized that halogens can form distinct molecular interactions that contribute to the recognition of ligands by proteins.[102]

Several examples of the impact of halogen bonds in protein–ligand complexes have been reported.[103–105] A revealing example is provided in a series of potent and selective [1,2,4]triazolo[1,5-*a*]pyrimidine PDE2a inhibitors.[106] In this work, a systematic analysis was conducted to investigate the effect of halogens on the *meta* position of inhibitor **19** (Fig. 12a). Additional analogues were synthesized where the hydrogen was replaced with F, Cl, Br and I. All compounds bound with a similar pose with no noticeable conformational changes in the binding site residues. An increase in the activity of the compound was observed in the following order H–F ≪ Cl < Br < I, corroborating the presence of a halogen bonding with the side chain oxygen of Y827 (Fig. 12b), although electronic effects at the aromatic ring are probably also contributing to the change in potency.

Halogen multipolar interactions

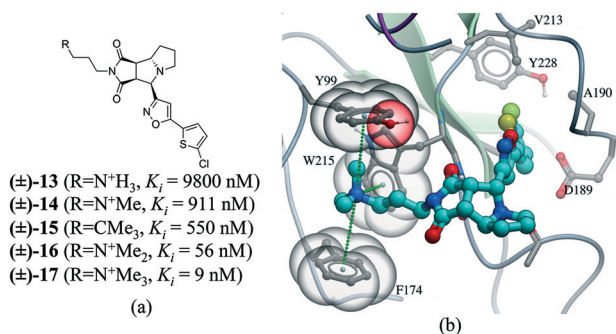Related to, but distinct from halogen bonds are multipolar interactions between halogen atoms and carbonyl carbon or

(±)-**13** (R=N$^+$H$_3$, K_i = 9800 nM)
(±)-**14** (R=N$^+$Me, K_i = 911 nM)
(±)-**15** (R=CMe$_3$, K_i = 550 nM)
(±)-**16** (R=N$^+$Me$_2$, K_i = 56 nM)
(±)-**17** (R=N$^+$Me$_3$, K_i = 9 nM)

(a)

(b)

Fig. 11 a) Chemical structure of a series inhibitors of human factor Xa; b) in the X-ray cocrystal structure of human factor Xa and **17** (PDB: 2JKH, cyan carbon), the quaternary ammonium ion fill the aromatic box (Y99, F174, and W215 are shown as transparent CPK). The cation–π interaction is displayed as dotted green lines.
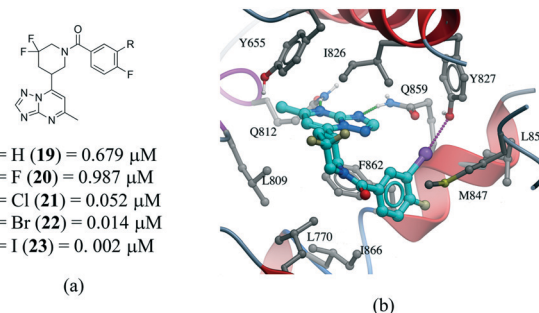

R = H (**19**) = 0.679 μM
R = F (**20**) = 0.987 μM
R = Cl (**21**) = 0.052 μM
R = Br (**22**) = 0.014 μM
R = I (**23**) = 0. 002 μM

(a)

(b)

Fig. 12 a) Chemical structure of a series of PDE2a inhibitors; b) X-ray cocrystal structure of PDE2a and **23** (PDB: 5U00, cyan carbons). The halogen bond interaction is shown as a dotted magenta line.
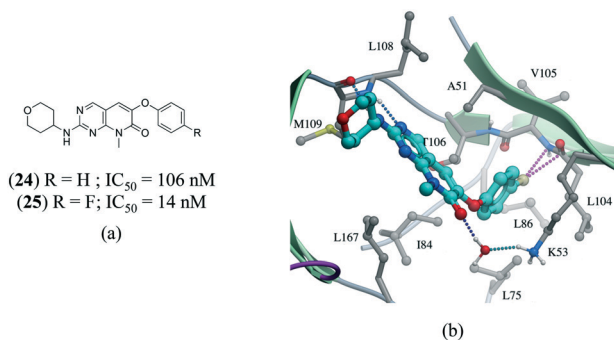
(24) R = H ; IC$_{50}$ = 106 nM
(25) R = F; IC$_{50}$ = 14 nM

(a)

(b)

Fig. 13 a) Chemical structure of p38α inhibitors; b) X-ray cocrystal structure of p38a and **24** (PDB: 3FLZ). The hydrogen bonding is shown as a dotted green line. The ligand **24** was modified to **25** to show the fluorine multipolar interaction as a dotted magenta line.

amide nitrogen.[107,108] These are favorable dipolar interactions between a C–X group (mainly with fluorine) and an electrophilic center such as the amide group in the backbone or side chain of proteins.[109] Instead of approaching the negatively polarized center in a head-to-head manner, the C–X interacts orthogonally with the carbonyl group.[110]

We found 109 multipolar interactions involving fluorine atoms, 65 chlorine atoms, and hardly any with bromine or iodine. The C–X···Y (C, N) and X···C=O (or N–C) (Θ_1 and Θ_2 in Fig. S2†) angles had median values of 148° and 88° respectively, suggesting the preference for an orthogonal geometry. More than 93% were formed with protein main-chain carbon and nitrogen, with a strong preference for glycine (Fig. S14†).

Compared with other interactions, little attention has been given to the role of multipolar interactions in molecular recognition events of chemical and biological systems.[107,111] Previous reports indicate that this interaction may substantially contribute to the affinity of small molecule inhibitors.[112,113] However, a systematic analysis of a large data set revealed only a modest improvement in potency (0.3–0.6 kcal mol^{-1}) associated with fluorine multipolar interaction.[108]

A series of p38α inhibitors recently illustrated the potential impact of a fluorine multipolar interaction.[114] Replacement of a hydrogen in **24** (IC$_{50}$ = 106 nM) by fluorine in **25** (IC$_{50}$ = 14 nM) improved the potency by approximately 8-fold (Fig. 13a). Introducing a fluorine atom at the *para*-position of the ring in the crystal structure of **24** confirms a short distance from the peptide carbonyl carbon and amide nitrogen of L104 and V105, respectively, indicative of a multipolar interaction (Fig. 13).

Conclusion

We presented here a statistical analysis of the nature, geometry and frequency of atomic interactions between small molecule ligands and their receptors available in the PDB. The enrichment of polar interactions in bound fragments, but hydrophobic contacts in optimized compounds reflects the challenge of overcoming desolvation penalty during lead optimization. This unbiased census recapitulates well-known

rules driving ligand design, but also uncovers some interaction types that are often overlooked in medicinal chemistry. This analysis will help in the interpretation of difficult SAR, and may serve as a knowledgebase for the improvement of scoring functions used in virtual screening.

Conflicts of interest

The author declare no competing interests.

Acknowledgements

References

1 A. Sharff and H. Jhoti, *Curr. Opin. Chem. Biol.*, 2003, 7, 340–345.

2 J.-P. Renaud, C. Chung, U. H. Danielson, U. Egner, M. Hennig, R. E. Hubbard and H. Nar, *Nat. Rev. Drug Discovery*, 2016, 15, 679–698.

3 A. E. Todd, R. L. Marsden, J. M. Thornton and C. A. Orengo, *J. Mol. Biol.*, 2005, 348, 1235–1260.

4 A. Yee, K. Pardee, D. Christendat, A. Savchenko, A. M. Edwards and C. H. Arrowsmith, *Acc. Chem. Res.*, 2003, 36, 183–189.

5 M. Schapira and D. J. Abraham, in *Burger's Medicinal Chemistry and Drug Discovery*, John Wiley & Sons, Inc., 2003, pp. 569–600.

6 H. M. Berman, *Nucleic Acids Res.*, 2000, 28, 235–242.

7 R. Wang, X. Fang, Y. Lu and S. Wang, *J. Med. Chem.*, 2004, 47, 2977–2980.

8 R. Wang, X. Fang, Y. Lu, C. Y. Yang and S. Wang, *J. Med. Chem.*, 2005, 48, 4111–4119.

9 Z. Deng, C. Chuaqui and J. Singh, *J. Med. Chem.*, 2004, 47, 337–344.

10 M. L. Verdonk, J. C. Cole and R. Taylor, *J. Mol. Biol.*, 1999, 289, 1093–1108.

11 M. Hendlich, A. Bergner, J. Günther and G. Klebe, *J. Mol. Biol.*, 2003, 326, 607–620.

12 A. Golovin and K. Henrick, *BMC Bioinf.*, 2008, 9, 1–11.

13 T. Inhester, S. Bietz, M. Hilbig, R. Schmidt and M. Rarey, *J. Chem. Inf. Model.*, 2017, 57, 148–158.

14 G. Marcou and D. Rognan, *J. Chem. Inf. Model.*, 2007, 47, 195–207.

15 S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme and M. Schroeder, *Nucleic Acids Res.*, 2015, 43, W443–W447.

1978 | *Med. Chem. Commun.*, 2017, **8**, 1970–1981

This journal is © The Royal Society of Chemistry 2017

16 H. C. Jubb, A. P. Higueruelo, B. Ochoa-Montaño, W. R. Pitt, D. B. Ascher and T. L. Blundell, *J. Mol. Biol.*, 2016, **429**, 365–371.

17 A. M. Schreyer and T. L. Blundell, *Database*, 2013, **2013**, 1–9.

18 R. Mohamed, J. Degac and V. Helms, *PLoS One*, 2015, **10**, 1–18.

19 K. Chen and L. Kurgan, *PLoS One*, 2009, **4**, 1–14.

20 C. H. Reynolds, B. A. Tounge and S. D. Bembenek, *J. Med. Chem.*, 2008, **51**, 2432–2438.

21 M. M. Hann, *Multifaceted Roles Crystallogr. Mod. Drug Discov.*, 2015, vol. 2, pp. 183–196.

22 D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discovery*, 2016, **15**, 605–619.

23 R. D. Smith, A. L. Engdahl, J. B. Dunbar and H. A. Carlson, *J. Chem. Inf. Model.*, 2012, **52**, 2098–2106.

24 A. P. Higueruelo, A. Schreyer, G. R. J. Bickerton, T. L. Blundell and W. R. Pitt, *PLoS One*, 2012, **7**, 1–8.

25 T. J. Ritchie and S. J. F. Macdonald, *J. Med. Chem.*, 2014, **57**, 7206–7215.

26 R. D. Taylor, M. MacCoss and A. D. G. Lawson, *J. Med. Chem.*, 2014, **57**, 5845–5859.

27 C. C. Valley, A. Cembran, J. D. Perlmutter, A. K. Lewis, N. P. Labello, J. Gao and J. N. Sachs, *J. Biol. Chem.*, 2012, **287**, 34979–34991.

28 A. M. Davis and S. J. Teague, *Angew. Chem., Int. Ed.*, 1999, **38**, 736–749.

29 C. S. Leung, S. S. F. Leung, J. Tirado-Rives and W. L. Jorgensen, *J. Med. Chem.*, 2012, **55**, 4489–4500.

30 P. J. Hajduk and D. R. Sauer, *J. Med. Chem.*, 2008, **51**, 553–564.

31 J. de Vicente, P. Tivitmahaisoon, P. Berry, D. R. Bolin, D. Carvajal, W. He, K.-S. Huang, C. Janson, L. Liang, C. Lukacs, A. Petersen, H. Qian, L. Yi, Y. Zhuang and J. C. Hermann, *ACS Med. Chem. Lett.*, 2015, **6**, 1019–1024.

32 H. Schönherr and T. Cernak, *Angew. Chem., Int. Ed.*, 2013, **52**, 12256–12267.

33 C. Bissantz, B. Kuhn and M. Stahl, *J. Med. Chem.*, 2010, **53**, 5061–5084.

34 T. Steiner, *Angew. Chem., Int. Ed.*, 2002, **41**, 49–76.

35 E. Nittinger, T. Inhester, S. Bietz, A. Meyder, K. T. Schomburg, G. Lange, R. Klein and M. Rarey, *J. Med. Chem.*, 2017, **60**, 4245–4257.

36 A. R. Fersht, *Trends Biochem. Sci.*, 1987, **12**, 301–304.

37 D. H. Williams, M. S. Westwell, K. Pawlak, R. L. Bruening, B. J. Tarbet, G. R. Marshall, P. Brick, P. Carter, M. M. Y. Waye and G. Winter, *Chem. Soc. Rev.*, 1998, **27**, 57–64.

38 B. K. Shoichet, *Nat. Biotechnol.*, 2007, **25**, 1109–1110.

39 E. Freire, *Drug Discovery Today*, 2008, **13**, 869–874.

40 L. Muley, B. Baum, M. Smolinski, M. Freindorf, A. Heine, G. Klebe and D. G. Hangauer, *J. Med. Chem.*, 2010, **53**, 2126–2135.

41 B. Baum, L. Muley, A. Heine, M. Smolinski, D. Hangauer and G. Klebe, *J. Mol. Biol.*, 2009, **391**, 552–564.

42 E. A. Meyer, R. K. Castellano and F. Diederich, *Angew. Chem., Int. Ed.*, 2003, **42**, 1210–1250.

43 S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami and K. Tanabe, *J. Am. Chem. Soc.*, 2002, **124**, 104–112.

44 L. M. Salonen, M. Ellermann and F. Diederich, *Angew. Chem., Int. Ed.*, 2011, **50**, 4808–4842.

45 A. B. Eldrup, F. Soleymanzadeh, S. J. Taylor, I. Muegge, N. A. Farrow, D. Joseph, K. McKellop, C. C. Man, A. Kukulka and S. De Lombaert, *J. Med. Chem.*, 2009, **52**, 5880–5895.

46 T. J. Ritchie and S. J. F. Macdonald, *Drug Discovery Today*, 2009, **14**, 1011–1020.

47 F. Lovering, J. Bikker and C. Humblet, *J. Med. Chem.*, 2009, **52**, 6752–6756.

48 T. Steiner, *New J. Chem.*, 1998, **22**, 1099–1103.

49 S. Scheiner, *Phys. Chem. Chem. Phys.*, 2011, **13**, 13860–13872.

50 S. K. Panigrahi and G. R. Desiraju, *Proteins: Struct., Funct., Genet.*, 2007, **67**, 128–141.

51 S. Sarkhel and G. R. Desiraju, *Proteins: Struct., Funct., Genet.*, 2003, **54**, 247–259.

52 Z. S. Derewenda, L. Lee and U. Derewenda, *J. Mol. Biol.*, 1995, **252**, 248–262.

53 M. Wahl, *Trends Biochem. Sci.*, 1997, **22**, 97–102.

54 L. Jiang and L. Lai, *J. Biol. Chem.*, 2002, **277**, 37732–37740.

55 R. Vargas, J. Garza, D. A. Dixon and B. P. Hay, *J. Am. Chem. Soc.*, 2000, **122**, 4750–4755.

56 Z. Liu, G. Wang, Z. Li and R. Wang, *J. Chem. Theory Comput.*, 2008, **4**, 1959–1973.

57 E. M. D. Keegstra, A. L. Spek, J. W. Zwikker and L. W. Jenneskens, *J. Chem. Soc., Chem. Commun.*, 1994, 1633–1634.

58 S. Aravinda, N. Shamala, A. Bandyopadhyay and P. Balaram, *J. Am. Chem. Soc.*, 2003, **125**, 15065–15075.

59 Y. Mandel-Gutfreund, H. Margalit, R. L. Jernigan and V. B. Zhurkin, *J. Mol. Biol.*, 1998, **277**, 1129–1140.

60 S. Horowitz and R. C. Trievel, *J. Biol. Chem.*, 2012, **287**, 41576–41582.

61 R. A. Musah, G. M. Jensen, R. J. Rosenfeld, D. E. McRee and D. B. Goodin, *J. Am. Chem. Soc.*, 1997, **119**, 9083–9084.

62 A. C. Pierce, K. L. Sandretto and G. W. Bemis, *Proteins: Struct., Funct., Genet.*, 2002, **49**, 567–576.

63 S. Wang, C. Meades, G. Wood, A. Osnowski, S. Anderson, R. Yuill, M. Thomas, M. Mezna, W. Jackson, C. Midgley, G. Griffiths, I. Fleming, S. Green, I. McNae, S. Y. Wu, C. McInnes, D. Zheleva, M. D. Walkinshaw and P. M. Fischer, *J. Med. Chem.*, 2004, **47**, 1662–1675.

64 UniProtKB/TrEMBL 2017_03, http://www.uniprot.org/statistics/TrEMBL, (accessed 28 March 2017).

65 P. I. Nagy and P. W. Erhardt, *J. Phys. Chem. B*, 2010, **114**, 16436–16442.

66 C. D. Waldburger, J. F. Schildbach and R. T. Sauer, *Nat. Struct. Biol.*, 1995, **2**, 122–128.

67 Z. S. Hendsch and B. Tidor, *Protein Sci.*, 1994, **3**, 211–226.

68 E. Segala, D. Guo, R. K. Y. Cheng, A. Bortolato, F. Deflorian, A. S. Doré, J. C. Errey, L. H. Heitman, A. P. Ijzerman, F. H. Marshall and R. M. Cooke, *J. Med. Chem.*, 2016, **59**, 6470–6479.

This journal is © The Royal Society of Chemistry 2017

Med. Chem. Commun., 2017, **8**, 1970–1981 | 1979

69 Y. Miyamoto, Y. Banno, T. Yamashita, T. Fujimoto, S. Oi, Y. Moritoh, T. Asakawa, O. Kataoka, H. Yashiro, K. Takeuchi, N. Suzuki, K. Ikedo, T. Kosaka, S. Tsubotani, A. Tani, M. Sasaki, M. Funami, M. Amano, Y. Yamamoto, K. Aertgeerts, J. Yano and H. Maezaki, *J. Med. Chem.*, 2011, **54**, 831–850.

70 G. Weltrowska, N. N. Chung, C. Lemieux, J. Guo, Y. Lu, B. C. Wilkes and P. W. Schiller, *J. Med. Chem.*, 2010, **53**, 2875–2881.

71 Y. H. Peng, H. Y. Shiao, C. H. Tu, P. M. Liu, J. T. A. Hsu, P. K. Amancha, J. S. Wu, M. S. Coumar, C. H. Chen, S. Y. Wang, W. H. Lin, H. Y. Sun, Y. S. Chao, P. C. Lyu, H. P. Hsieh and S. Y. Wu, *J. Med. Chem.*, 2013, **56**, 3889–3903.

72 M. Harder, B. Kuhn and F. Diederich, *ChemMedChem*, 2013, **8**, 397–404.

73 J. B. Mitchell, C. L. Nandi, I. K. McDonald, J. M. Thornton and S. L. Price, *J. Mol. Biol.*, 1994, **239**, 315–331.

74 G. Duan, V. H. Smith and D. F. Weaver, *Chem. Phys. Lett.*, 1999, **310**, 323–332.

75 M. Giroud, J. Ivkovic, M. Martignoni, M. Fleuti, N. Trapp, W. Haap, A. Kuglstatter, J. Benz, B. Kuhn, T. Schirmeister and F. Diederich, *ChemMedChem*, 2016, **12**, 257–270.

76 M. Giroud, M. Harder, B. Kuhn, W. Haap, N. Trapp, W. B. Schweizer, T. Schirmeister and F. Diederich, *ChemMedChem*, 2016, **11**, 1042–1047.

77 Z. Qiu, B. Kuhn, J. Aebi, X. Lin, H. Ding, Z. Zhou, Z. Xu, D. Xu, L. Han, C. Liu, H. Qiu, Y. Zhang, W. Haap, C. Riemer, M. Stahl, N. Qin, H. C. Shen and G. Tang, *ACS Med. Chem. Lett.*, 2016, **7**, 802–806.

78 L. M. Salonen, M. C. Holland, P. S. J. Kaib, W. Haap, J. Benz, J.-L. Mary, O. Kuster, W. B. Schweizer, D. W. Banner and F. Diederich, *Chem. – Eur. J.*, 2012, **18**, 213–222.

79 A. S. Mahadevi and G. N. Sastry, *Chem. Rev.*, 2013, **113**, 2100–2138.

80 J. P. Gallivan and D. A. Dougherty, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9459–9464.

81 D. A. Dougherty, *Acc. Chem. Res.*, 2013, **46**, 885–893.

82 J. J. A. G. Kamps, J. Huang, J. Poater, C. Xu, B. J. G. E. Pieters, A. Dong, J. Min, W. Sherman, T. Beuming, F. Matthias Bickelhaupt, H. Li and J. Mecinović, *Nat. Commun.*, 2015, **6**, 1–12.

83 S. C. R. Lummis, N. J. Harrison, J. Wang, J. A. Ashby, K. S. Millen, D. L. Beene and D. A. Dougherty, *ACS Chem. Neurosci.*, 2012, **3**, 186–192.

84 S. A. Pless, A. P. Hanek, K. L. Price, J. W. Lynch, H. A. Lester, D. A. Dougherty and S. C. R. Lummis, *Mol. Pharmacol.*, 2011, **79**, 742–748.

85 N. H. Duffy, H. A. Lester and D. A. Dougherty, *ACS Chem. Biol.*, 2012, **7**, 1738–1745.

86 L. M. Salonen, C. Bucher, D. W. Banner, W. Haap, J. L. Mary, J. Benz, O. Kuster, P. Seiler, W. B. Schweizer and F. Diederich, *Angew. Chem., Int. Ed.*, 2009, **48**, 811–814.

87 L. A. Hardegger, B. Kuhn, B. Spinnler, L. Anselm, R. Ecabert, M. Stihle, B. Gsell, R. Thoma, J. Diez, J. Benz, J.-M. Plancher, G. Hartmann, Y. Isshiki, K. Morikami, N. Shimma, W. Haap, D. W. Banner and F. Diederich, *ChemMedChem*, 2011, **6**, 2048–2054.

88 C. Dalvit, C. Invernizzi and A. Vulpetti, *Chem. – Eur. J.*, 2014, **20**, 11058–11068.

89 M. O. Zimmermann, A. Lange, R. Wilcken, M. B. Cieslik, T. E. Exner, A. C. Joerger, P. Koch and F. M. Boeckler, *Future Med. Chem.*, 2014, **6**, 617–639.

90 Y. Lu, T. Shi, Y. Wang, H. Yang, X. Yan, X. Luo, H. Jiang and W. Zhu, *J. Med. Chem.*, 2009, **52**, 2854–2862.

91 S. Sirimulla, J. B. Bailey, R. Vegesna and M. Narayan, *J. Chem. Inf. Model.*, 2013, **53**, 2781–2791.

92 P. Auffinger, F. A. Hays, E. Westhof and P. S. Ho, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 16789–16794.

93 T. Clark, M. Hennemann, J. S. Murray and P. Politzer, *J. Mol. Model.*, 2007, **13**, 291–296.

94 P. Politzer, J. S. Murray and T. Clark, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11178–11189.

95 Q. Zhang, Z. Xu, J. Shi and W. Zhu, *J. Chem. Inf. Model.*, 2017, **57**, 1529–1534.

96 L. A. Hardegger, B. Kuhn, B. Spinnler, L. Anselm, R. Ecabert, M. Stihle, B. Gsell, R. Thoma, J. Diez, J. Benz, J.-M. Plancher, G. Hartmann, D. W. Banner, W. Haap and F. Diederich, *Angew. Chem., Int. Ed.*, 2011, **50**, 314–318.

97 Z. Xu, Z. Yang, Y. Liu, Y. Lu, K. Chen and W. Zhu, *J. Chem. Inf. Model.*, 2014, **54**, 69–78.

98 Y. Lu, Y. Wang and W. Zhu, *Phys. Chem. Chem. Phys.*, 2010, **12**, 4543–4551.

99 C. B. Aakeröy, N. R. Champness, C. Janiak, R. W. Gable, B. F. Hoskins, J. P. Liu, D. B. Varshney, I. G. Georgiev, D.-Z. Liao, S. Parsons, E. K. Brechin, N. Crivillers, J. Veciana, C. Rovira, M. Leufgen, G. Schmidt and L. W. Molenkamp, *CrystEngComm*, 2010, **12**, 22–43.

100 F. Zordan, L. Brammer and P. Sherwood, *J. Am. Chem. Soc.*, 2005, **127**, 5979–5989.

101 M. Z. Hernandes, S. M. T. Cavalcanti, D. R. M. Moreira, W. F. de Azevedo Junior and A. C. L. Leite, *Curr. Drug Targets*, 2010, **11**, 303–314.

102 M. R. Scholfield, C. M. Vander Zanden, M. Carter and P. S. Ho, *Protein Sci.*, 2013, **22**, 139–152.

103 Z. Xu, Z. Liu, T. Chen, T. Chen, Z. Wang, G. Tian, J. Shi, X. Wang, Y. Lu, X. Yan, G. Wang, H. Jiang, K. Chen, S. Wang, Y. Xu, J. Shen and W. Zhu, *J. Med. Chem.*, 2011, **54**, 5607–5611.

104 M. Bollini, R. A. Domaoal, V. V. Thakur, R. Gallardo-Macias, K. A. Spasov, K. S. Anderson and W. L. Jorgensen, *J. Med. Chem.*, 2011, **54**, 8582–8591.

105 E. Y. Cotrina, M. Pinto, L. Bosch, M. Vilà, D. Blasi, J. Quintana, N. B. Centeno, G. Arsequell, A. Planas and G. Valencia, *J. Med. Chem.*, 2013, **56**, 9110–9121.

106 L. Gomez, M. E. Massari, T. Vickers, G. Freestone, W. Vernier, K. Ly, R. Xu, M. McCarrick, T. Marrone, M. Metz, Y. G. Yan, Z. W. Yoder, R. Lemus, N. J. Broadbent, R. Barido, N. Warren, K. Schmelzer, D. Neul, D. Lee, C. B. Andersen, K. Sebring, K. Aertgeerts, X. Zhou, A. Tabatabaei, M. Peters and J. G. Breitenbucher, *J. Med. Chem.*, 2017, **60**, 2037–2051.

107 R. Paulini, K. Müller and F. Diederich, *Angew. Chem., Int. Ed.*, 2005, **44**, 1788–1805.

1980 | *Med. Chem. Commun.*, 2017, **8**, 1970–1981

This journal is © The Royal Society of Chemistry 2017

108 L. Xing, C. Keefer and M. F. Brown, *J. Fluorine Chem.*, 2017, **198**, 47–53.

109 J. A. Olsen, D. W. Banner, P. Seiler, B. Wagner, T. Tschopp, U. Obst-Sander, M. Kansy, K. Müller and F. Diederich, *ChemBioChem*, 2004, **5**, 666–675.

110 C. Dalvit and A. Vulpetti, *ChemMedChem*, 2011, **6**, 104–114.

111 E. P. Gillis, K. J. Eastman, M. D. Hill, D. J. Donnelly and N. A. Meanwell, *J. Med. Chem.*, 2015, **58**, 8315–8359.

112 J. Pollock, D. Borkin, G. Lund, T. Purohit, E. Dyguda-Kazimierowicz, J. Grembecka and T. Cierpicki, *J. Med. Chem.*, 2015, **58**, 7465–7474.

113 J. A. Olsen, D. W. Banner, P. Seiler, U. Obst Sander, A. D'Arcy, M. Stihle, K. Müller and F. Diederich, *Angew. Chem., Int. Ed.*, 2003, **42**, 2507–2511.

114 D. M. Goldstein, M. Soth, T. Gabriel, N. Dewdney, A. Kuglstatter, H. Arzeno, J. Chen, W. Bingenheimer, S. A. Dalrymple, J. Dunn, R. Farrell, S. Frauchiger, J. La Fargue, M. Ghate, B. Graves, R. J. Hill, F. Li, R. Litman, B. Loe, J. McIntosh, D. McWeeney, E. Papp, J. Park, H. F. Reese, R. T. Roberts, D. Rotstein, B. San Pablo, K. Sarma, M. Stahl, M.-L. Sung, R. T. Suttman, E. B. Sjogren, Y. Tan, A. Trejo, M. Welch, P. Weller, B. R. Wong and H. Zecic, *J. Med. Chem.*, 2011, **54**, 2255–2265.