JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Structural Insights into the Molecular Basis of the Ligand Promiscuity

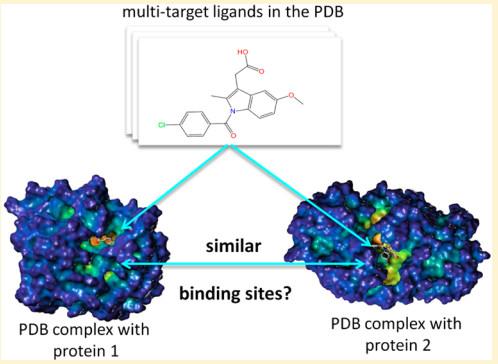
Noé Sturm,[†,‡] Jérémy Desaphy,[†] Ronald J. Quinn,[‡] Didier Rognan,[†] and Esther Kellenberger[†,*]

[†]UMR 7200 CNRS/Université de Strasbourg, MEDALIS Drug Discovery Center, 74 route du Rhin, 67401 Illkirch, France

[‡]Eskitis Institute, Griffith University, Brisbane, Qld 4111, Australia

Supporting Information

ABSTRACT: Selectivity is a key factor in drug development. In this paper, we questioned the Protein Data Bank to better understand the reasons for the promiscuity of bioactive compounds. We assembled a data set of >1000 pairs of three-dimensional structures of complexes between a "drug-like" ligand (as its physicochemical properties overlap that of approved drugs) and two distinct "druggable" protein targets (as their binding sites are likely to accommodate "drug-like" ligands). Studying the similarity between the ligand-binding sites in the different targets revealed that the lack of selectivity of a ligand can be due (i) to the fact that Nature has created the same binding pocket in different proteins, which do not necessarily have otherwise sequence or fold similarity, or (ii) to specific characteristics of the ligand itself. In particular, we demonstrated that many ligands can adapt to different protein environments by changing their conformation, by using different chemical moieties to anchor to different targets, or by adopting unusual extreme binding modes (e.g., only apolar contact between the ligand and the protein, even though polar groups are present on the ligand or at the protein surface). Lastly, we provided new elements in support to the recent studies which suggest that the promiscuity of a ligand might be inferred from its molecular complexity.



INTRODUCTION

Achieving target selectivity is often desirable in drug discovery in order to minimize side effects and possible adverse reactions due to binding to unintended targets. In recent years, much effort has been put into development of computational methods to predict all possible targets of all possible compounds,[1,2] based on the following assumptions: similar compounds share the same targets,[3] drugs with similar side-effect phenotypes share the same targets,[4] and similar protein−ligand binding sites recognize the same compounds.[5] The empirical approaches, which have benefited notably from the availability of ever growing databases collecting structure and activity data of bioactive compounds,[6,7] have proved to be successful in the identification of new targets for drugs and have also contributed to improving the understanding of the main mechanism of action of drugs as well as mechanisms of their adverse reactions.[8−13] For example, the anti-HIV drug Rescriptor, an inhibitor of the viral reverse transcriptase, was predicted and experimentally confirmed to bind to the histamine H4 receptor, thereby suggesting molecular basis for the painful rashes associated with this drug.[10] In binding and functional experiments, we recently demonstrated that some but not all protein kinase inhibitors affect the neurotransmitter release in the synapse through the binding to synapsin I, whose ATP-binding site was beforehand identified as similar to the staurosporine-binding site in Pim-1 kinase.[9]

In pharmaceutical research, the off-target activities of a compound can be characterized from in vitro testing of the compound against a panel of proteins. For example, large-scale profiling experiments are performed at the CEREP, which provides data for >2000 drugs and bioactive compounds tested in >200 assays in the BioPrint database.[14] Comprehensive analyses of BioPrint have suggested link between the chemical properties of a compound and its effects at multiple targets (i.e., its promiscuity). In particular a strong correlation was observed between lipophilicity and promiscuity.[15] The positive ionization,[16] a high number of aromatic rings,[15] and the predominance of ring systems in the compound[17] were also shown to have negative effect on compound selectivity. An independent study on data generated by GlaxoSmithKline (800 compounds tested in >490 assays) confirmed the importance of lipophilicity and aromaticity in the promiscuity of compounds.[18]

In the study presented in this paper, we investigated the reasons for which a compound can target different proteins from the structural point of view. In particular, we sought to know if the promiscuity of a compound was the consequence of the presence of similar binding sites in different proteins, or if it is due to specific characteristics of the compound itself. To this purpose, we exploited the information in the Protein Data Bank (PDB)[19] to identify ligands involved in complexes with different proteins. We then compared the different sites for the promiscuous ligands and we showed that different proteins exhibit the same binding pocket, and that some compounds can adapt to different protein cavities. We finally investigated which

Received: April 20, 2012
Published: August 25, 2012

dx.doi.org/10.1021/ci300196g | J. Chem. Inf. Model. 2012, 52, 2410−2421

molecular properties might prompt a compound to bind to dissimilar binding sites.

MATERIALS AND METHODS

Identification in the sc-PDB of Promiscuous Ligands and Their Targets. The sc-PDB[20] repository is a database built from the Protein Data Bank.[19] It exclusively contains complexes between a low molecular weight compound and its bound protein. Practically, the selection of complexes depends on physicochemical criteria for the ligand (e.g., 140 ≤ molecular weight ≤ 810, >1 carbon atoms, >1 oxygen or nitrogen atoms, <20 rotatable bonds), functional criteria for the protein (e.g., no cytochromes or immunoglobulins), and topological criteria for the binding mode (e.g., number of residues in site >7, buried surface area of the ligand >50%). Each sc-PDB entry consists of a ligand, a protein, and the corresponding binding site, which is defined as all residues with at least one atom within a 6.5 Å radius sphere centered on the ligand center of mass. The sc-PDB coordinate files include hydrogen atoms, thereby fully defining the ionization and the tautomeric state of the ligand.[21] The different proteins in the sc-PDB could be distinguished unequivocally by their name, which derived from the Uniprot[22] recommended name. The different ligands in the sc-PDB could be distinguished unequivocally by their canonical SMILES representation. The sc-PDB is a nonredundant database: for a given pair of protein and ligand, only the PDB entry with the best resolution is considered.

The data set was created from the 8166 entries of sc-PDB, release 2010. In total, 518 ligands were found in at least two complexes with different proteins. About half of them were discarded due to their high similarity with nucleic acids, peptides, monosaccharides, oligosaccharides, or fatty acids (The filtering rules are given in the Supporting Information, Table S1). The data set contains 247 promiscuous ligands.

2D-Description of Ligands. The following chemical descriptors were computed for ligands using PipelinePilot8 (Accelrys Inc., San Diego, CA, USA): molecular weight, number of hydrogen bond (H-bond) donors or acceptors, number of rotatable bonds, molecular polar surface area, ALogP (Ghose/Crippen group-contribution estimate for logP), circular FCFP_4 fingerprints, FCFP_4 density (FCFP_4 size/number of non hydrogen atoms), H-bonding propensity (number of H-bond donors and acceptors/total number of atoms), and three-dimensionality (number of sp3 carbon atoms/total number of carbon atoms).

The 247 compounds of the data set were clustered using the Jarvis-Patrick algorithm in MOE2011 (Chemical Computing Group Inc., Montreal, Canada). The MACCS keys were compared using the Tanimoto coefficient. The similarity threshold was set to 0.65 for the creation of the lists of similar compounds and for the comparison of lists ("cluster overlap" parameter).

Conformational Variability of Protein-Bound Ligands. Protein-bound ligand structures were first compared by computing the root-mean-square deviation (rmsd) of the positions of the ligand heavy atoms after the best-fit superposition of the two sets of coordinates. The rmsd was computed using the 'Match' routine of Sybyl-X1.3 (Tripos, Inc., St. Louis, MO, US), which takes into account topological symmetry within molecules. Although rmsd values are commonly used and easy to interpret, they may be biased toward low values for small molecules or toward high values if one or more of the paired atoms are at a great distance from each other.[23] In the present study, the rmsd values may be misleading for ligands which do not interact totally with their target proteins (for example a high rmsd value may be observed if the ligand moiety which interacts with the protein has a well conserved structure in the two compared complexes, whereas the ligand moiety which points outward has different structures). To overcome this limitation, we evaluated the shape similarity of the ligand part that contacts the bound protein as follows: all ligand atoms involved in nonbonded interactions with the protein were identified as previously described;[24] their coordinates were written in MOL2 format using a simplified atom typing based on the nature of protein−ligand interactions (C.3 for any atom engaged in a hydrophobic contact, N.Am for a H-bond donor, O.2 for a H-bond acceptor, N.4 for a positively charged atom, O.Co2 for a negatively charged atom) Two sets of atoms originating from the complexes of a ligand with two different proteins were 3D-aligned by optimizing the volume overlap from Gaussian functions representing the atoms.[25] The alignment routine was written using the OEChem and OEShape toolkits (OpenEye, Inc., Santa-Fe, CA, U.S.A.). The overlap of atoms was scored with a Tanimoto coefficient (shTc):

$$\mathrm{shTc}_{A,B} = \frac{\sum_i O_{A,B}}{\sum_i I_A + \sum_i I_B - \sum_i O_{A,B}}$$

Where, for each of the five above-mentioned atom types i, $O_{A,B}$ is the overlap volume between conformers A and B, and I is the self-overlap volume of each entity A and B. The shTc score is normalized and quantifies the conservation in the two complexes of the protein-interacting moiety of the ligand. For example, if all protein-interacting atoms of a ligand in complex A represent 60% of all the protein-interacting atoms of the ligand in complex B (or vice versa), the shTc value is equal to 0.6. Alternatively, if the total numbers of protein-interacting atoms of the ligand are identical in complexes A and B and if 75% of the protein-interacting atoms of the ligand are identical in the two complexes, then the shTc value is equal to 0.6 too. The Tversky coefficient (shTv) was computed in order to distinguish the different scenarios:

$$\mathrm{shTv}_{A,B} = \frac{\sum_i O_{A,B}}{\alpha \sum_i I_A + \beta \sum_i I_B - \sum_i O_{A,B}}$$

where, for each of the five above-mentioned atom types i, $O_{A,B}$ is the overlap volume between conformers A and B, I is the self-overlap volume of each entity A and B, and α and β are weights so that $\alpha \neq \beta$ and $\alpha + \beta = 1$. By contrast to a Tanimoto index ($\alpha = \beta = 1$), the Tversky index gives more importance to either the reference or the fit object by assigning different weights to the self-overlap volumes I_A and I_B. The retained Tversky coefficient was the maximal value obtained for either of the two parameter sets $\alpha = 0.05/\beta = 0.95$ or $\alpha = 0.95/\beta = 0.05$.

2D Comparison of the Targets of Promiscuous Ligands. The protein sequences in fasta format were downloaded from the RCSB PDB.[26] The comparisons of the protein sequences were performed using the default parameters of the Needle routine in the EMBOSS package.[27] Only the protein chains which form the ligand binding site were considered. If several comparisons were made for a given pair of proteins, only the highest sequence identity value was retained. A sequence identity above 30% is a good indicator of protein homology.[28] In the present analysis, we considered that

an evolutionary link exists between two proteins aligned over more than 100 residues with a sequence identity above 25%.

3D Comparison of the Targets of Promiscuous Ligands. The comparisons of the protein structures were performed using the default parameters of the CE program.[29] This program identifies the longest combination of pairs of fragments which are structurally equivalent in the two protein chains (a fragment represents the Cα atoms of 8 consecutive residues) and calculates the statistical significance of the structural alignment by evaluating the probability of finding such an alignment from a random comparison of structures (Z-score). The input files were the structure files which were downloaded from the RCSB PDB. Only the protein chains which form the ligand binding site were considered. If several comparisons were made for a given pair of proteins, only the result with the highest Z-score was retained. A Z-score value higher than 4 denotes the conservation of the overall fold of the two proteins under investigation.

3D Comparison of Binding Sites for Promiscuous Ligands. The comparisons of the binding sites were performed using three in house programs, Volsite/Shaper,[30] SiteAlign4.0,[31] and Fuzcav.[32] The sc-PDB binding site coordinates in MOL2 format were used as input files. The comparisons of sites using Shaper were repeated for hydrated binding sites. Hydrated sites were prepared using Sybyl-X1.3 and include all crystallographic water molecules whose oxygen atom is closer than 3.5 Å from any ligand polar atom and closer than 3.5 Å from at least three binding site residues. The position of water hydrogen atoms was optimized to maximize the number of H-bonds made with the protein.

In SiteAlign,[31] eight topological and physicochemical attributes are projected from the Cβ-atom of cavity-lining residues to an 80 triangle-discretized polyhedron placed at the center of the binding site, thus defining a cavity fingerprint of 640 integers. 3D alignment is performed by moving the sphere within the target binding site while keeping the query sphere fixed. After each move, the distance of the newly described cavity descriptor is compared to that of the query, the best alignment being that minimizing the distance between both cavity fingerprints. The similarity is evaluated by a "global" score which is computed by considering the pairs of aligned triangles with non null properties in the mobile sphere or the fixed sphere (D1) and a "local" score which is computed by considering only triangles with non null properties in the mobile and the fixed spheres (D2). D1 and D2 scores lower than 0.6 and 0.2, respectively, indicate that the geometry and the chemical nature of residues are similar in the two sites which are compared.[31]

From a known protein−ligand complex, Volsite[30] converts the site into a regular lattice of pseudoatoms filling the cavity. The pseudoatoms farther than 6 Å from any ligand heavy atom were discarded. To each pseudoatom is assigned a pharmacophoric type, depending on the nature of the closest protein atom (H-bond acceptor, H-bond donor, H-bond acceptor and donor, negative ionizable, positive ionizable, hydrophobic, aromatic, or none if there is no protein atoms within a 4 Å distance). Shaper then aligns two sets of cavity points using Gaussian functions (see above) and then scores the alignment according to the quality of the overlap.[30] In practice, we demonstrated that a similarity score (S) higher than 0.35 indicates that the cavity shape and pharmacophoric properties are similar in the two sites which are compared.

FuzCav[32] annotates the Cα atoms of cavity-lining amino acids with the pharmacophoric properties of its parent residue (H-bond donor, H-bond acceptor, positive ionizable, negative ionizable, aromatic, aliphatic), then enumerates all triplets of Cα (three properties, three distances ≤14.3 Å) to populate a vector of 4833 integers which encode all possible combinations of triplets. The comparison of two sites consists in the direct computing of the distance between two numerical fingerprints (it does not generate a 3D alignment of sites). The benchmarking of the program revealed that a similarity score higher than 0.16 reflects the conservation of spatial arrangement and physicochemical properties of amino acids in the two sites which are compared.

RESULTS AND DISCUSSION

Setting up a Data Set of Promiscuous Ligands and Their Bound Proteins. To better understand the molecular basis for ligand promiscuity, we searched for ligands whose crystal structure is available for complexes with two or more different proteins. We restricted our analysis to proteins which are potentially able to bind small compounds with high affinity (from here on called *druggable*)[33] and to ligands whose molecular weight ranges from 140 to 800. We did not considered monosaccharides, because they are usually weak binders and their binding sites are poorly druggable. We neither studied nucleotides nor peptides, because they are highly flexible and known to recognize conformer-specific binding pockets.[34]

Among the 4229 unique ligands in the sc-PDB, we identified 247 promiscuous ligands. The chemical diversity of the set was evaluated by a nonhierarchical clustering based on MACCS keys compared using the Tanimoto coefficient. The similarity threshold of 0.65 yielded 145 clusters which correctly grouped compounds according to biochemical scaffolds. For example, it was observed that all thiamine derivatives define a single cluster (Figure 1A). In fact, about one-third of the ligands in the data set correspond to natural lipids, amino acids, and protein cofactors, or their close analogs.

The distribution of key physicochemical properties in the data set is given in Figure 1B. The molecular size was evaluated using the molecular weight. The average molecular weight in the data set is 367 and about 90% of all the 247 promiscuous ligands have molecular weight ranging from 200 to 500. The molecular flexibility was evaluated using the number of rotatable bonds. The ligands in the data set have up to fourteen rotatable bonds. Only fifteen ligands are fully rigid whereas ten compounds have more than 10 rotatable bonds. Last, the molecular polarity was evaluated using the number of H-bond donors and acceptors, the polar surface area (PSA), and the LogP (not shown). The cumulated number of H-bond donors and acceptors ranges from 2 to 20, and approximately one-third of ligands is distributed in each of the [2; 5], [6; 10], and [11; 20] intervals. The PSA ranges from 20 to 321 Å^2, and approximately 60% of ligands is in the [50; 150] interval. A quarter of the ligands have a PSA exceeding 150 Å^2. The LogP ranges from −10.7 to +7.3 and respectively 30% and 60% of ligands are distributed in the [−5; 0] and [0; +5] intervals. About 94% of the 247 ligands comply with the Lipinski's rules of five.[35] Altogether, the area of molecular property space occupied by the molecules in the data set overlap that occupied by orally absorbed drugs (from here on this characteristics will be called *drug-like*, for a comprehensive review on drug-likeness see ref 36). There seems however to be a bias in the data set

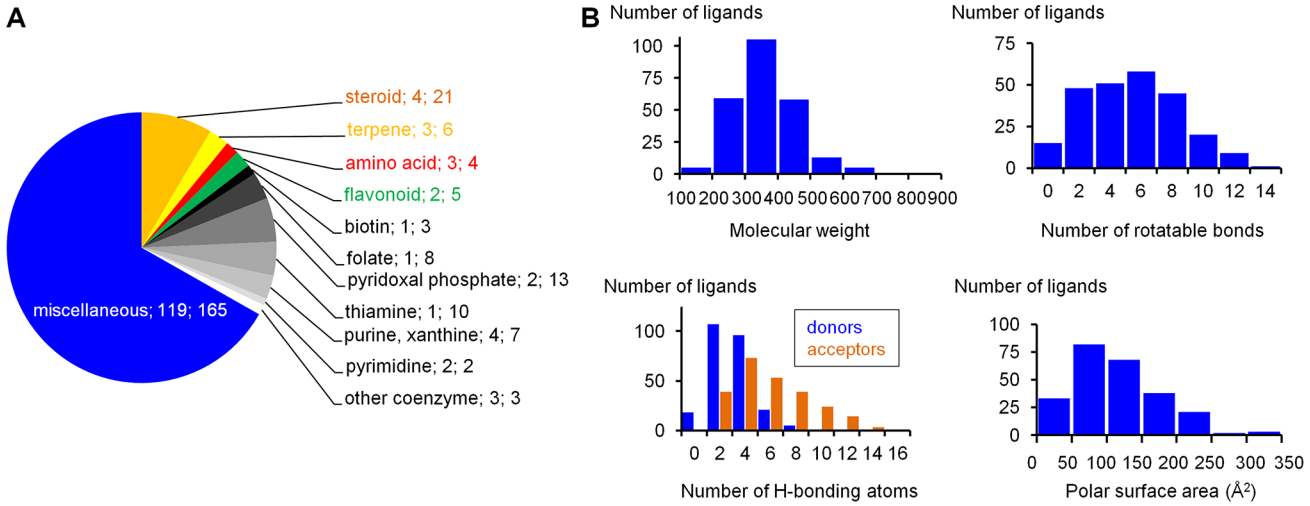


Figure 1. Description of the ligands in the data set. (A) Biochemical classification of the 247 promiscuous ligands. The name of each class is followed by the number of chemical clusters within the class, and then by the total number of members in the class. (B) Distribution of physicochemical descriptors.

toward polar compounds. To further evaluate the *drug-like* property of the 247 promiscuous ligands, we compared them to 959 drugs selected in Drugbank (FDA-approved small molecule drugs with molecular weight lower than 900, no nutraceuticals, biologics, or experimental drugs).[37] We could hence identify five drugs in our data set, namely diethylstilbestrol, progesterone, novobiocin, trimetrexate, and trimethoprim. In addition, 46 of the promiscuous ligands were found similar to 33 known drugs using circular FCFP_4 fingerprints, Tanimoto coefficient, and a similarity threshold of 0.5.

The 247 promiscuous ligands correspond to 689 PDB complexes, but only to 393 different proteins which nevertheless cover a wide range of biological functions (Figure 2A). About 65% of the ligands bind two different proteins (Figure 2B). Other ligands have up to 7 different targets, with the exception of the nonselective kinase inhibitor staurosporine which was found in complex with 23 different members of this enzyme family.

The total number of protein pairs in the data set is equal to 1070. The pairs were categorized according to their sequence identity and their global three-dimensional structure similarity: (i) 264 pairs are made of two proteins which have high sequence identity (>25% with >100 aligned residues) and a common fold (CE Z-score >4); we named them the *homologous pairs*; (ii) 478 pairs are made of two proteins which have low sequence identity but a common fold; we named them the *convergent pairs*; and (iii) 328 pairs are made of two proteins which have no sequence or fold similarities; we named them the *distant pairs*.

Binding Sites Which Accommodate the Same Ligand Are Not Necessarily Similar. We analyzed our data set in order to understand the molecular basis of the promiscuity of drug-like ligands, assuming that a ligand can associate with two different targets for one of the two following reasons: (i) the ligand-binding sites in the two proteins are similar or (ii) the ligand is able to adapt to two different binding sites. To test the first of the two hypothesis, we evaluated the similarity between the sites in each protein pairs using three different approaches: the 3D alignment of icosahedrons encoding the position and the pharmacophoric properties of the binding site-lining amino acids (SiteAlign),[31] the 3D-alignment of grid points which

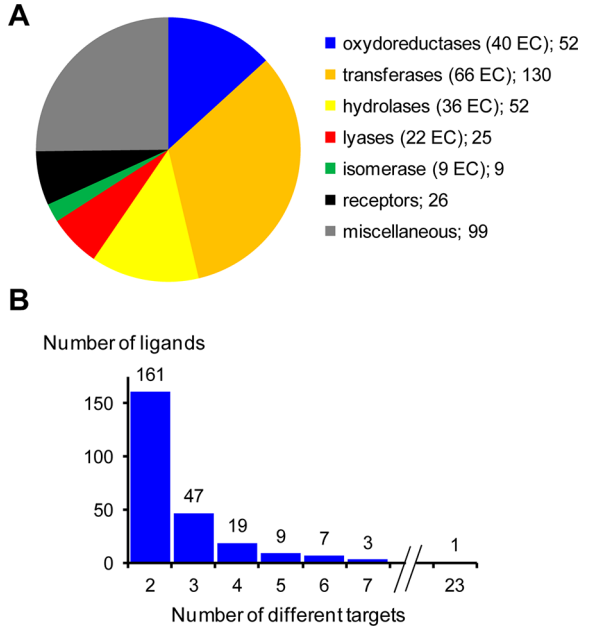


Figure 2. Description of the proteins in the data set. (A) Functional classification of the 393 proteins that are targeted by promiscuous ligands. The name of each class is followed by the number of members in the class. For the classes that group enzymes, the number of different subclasses as described by the Enzyme Commission (EC) is indicated too. (B) Level of promiscuity across the data set indicated by the number of different targets per ligand.

represent the cavity shape and the pharmacophoric properties at site surface (Shaper),[30] and the comparison of 3D-pharmacophoric fingerprints (Fuzcav).[32] Two sites were considered similar if they met the similarity criteria of at least one of the three approaches.

About three-quarters of the 1070 binding sites pairs were predicted to be similar. Similar sites were identified in all three categories of protein pairs: homologous, convergent, and distant pairs (Figure 3). Site similarity was detected in almost all homologous pairs of proteins (97%) and in about 80% of the convergent pairs. As shown in Figure 3, our data clearly revealed that the topology of the sites tend to be preserved in

dx.doi.org/10.1021/ci300196g | J. Chem. Inf. Model. 2012, 52, 2410−2421

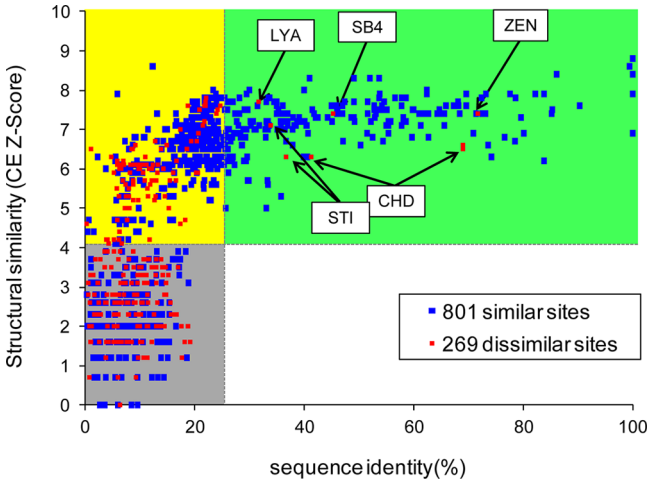


Figure 3. Sequence, fold, and site similarities in pairs of the target proteins. Three categories of protein pairs are highlighted with the different background colors: homologous proteins (green), convergent proteins (yellow), and distant proteins (gray). Boxes give the HET code of the five ligands associated to pairs of dissimilar sites of homologous proteins.

two proteins which share more than 25% of sequence identity and have a conserved fold. This observation is in line with bioinformatics studies, which demonstrated that the key functional amino acids are generally well conserved across the proteins of a functional family.[38] Interestingly, binding site similarity was also observed in about half of the distant pairs, meaning that proteins with no genetic evolutionary relationship can have a common local three-dimensional structure. This finding underlines the potency of site comparison methods to predict the ligand binding capability of a protein.

A quarter of the 1070 binding site pairs were found dissimilar with all three programs. Most of them correspond to protein pairs which have distinct sequence and fold characteristics. We cannot exclude expressly that the absence of similarity between two sites is due to methodological aspects. We nevertheless verified that crystallographic water molecules located in the cavity only have a marginal influence on site comparisons. In our data set, we found that one or several crystallographic water molecules mediate interactions between ligand and protein in 335 out of 689 PDB complexes, representing 628 pairs of sites. We repeated all Shaper calculations using as input the sites including these bridging water molecules. We observed that although the similarity score was modified in most of the comparisons (98%) which involve a "hydrated" site, the overall proportion of dissimilar and similar sites pairs in the data set was not changed upon consideration of water molecules (Table 1).

Table 1. Water at Binding Interfaces Hardly Affects Site Comparison Using Shaper

		number of similar sites pairs		
pairs of proteins	total number of site pairs	with and without water	only without water	only with water
homologous	264	242	0	0
convergent	478	331	12	29
distant	328	143	8	12
all	1070	716	20	41

In summary, the ligand information available in the PDB revealed that the promiscuity of a ligand can be explained by the presence of the same binding pocket in different proteins. However, a significant number of dissimilar sites were observed among the investigated pairs of complexes between a ligand and two different proteins, thereby supporting the assumption that the promiscuity of a ligand may solely originate from its physicochemical properties. We identified 76 ligands which have the capacity to bind to dissimilar protein sites.

Multiple Binding Modes Explain Why a Ligand Can Bind to Dissimilar Sites. In order to understand why a ligand can bind to two dissimilar sites, we compared the corresponding protein-bound ligand conformations. In particular, we scored the overlay between all heavy atoms of the ligand in the two sites (rmsd) and between the subset of heavy atoms in direct interaction with the protein (shTc). We hence could define three categories of pairs, depending on the structural and binding characteristics of the ligand: (i) in the class called *flexible ligand*, the ligands adopt different conformations in the two sites of a pair (high rmsd, low shTc), (ii) in the class called *bianchor ligand*, the ligands exhibit similar conformations in the two sites but use different sets of atoms to interact with the protein (low rmsd, low shTc), and (iii) in the class called *difficult to rationalize*, the ligands use the same moieties to interact with the two sites (high shTc). Figure 4 shows that, depending on the threshold used for rmsd and shTc, the exact number of pairs assigned to each category

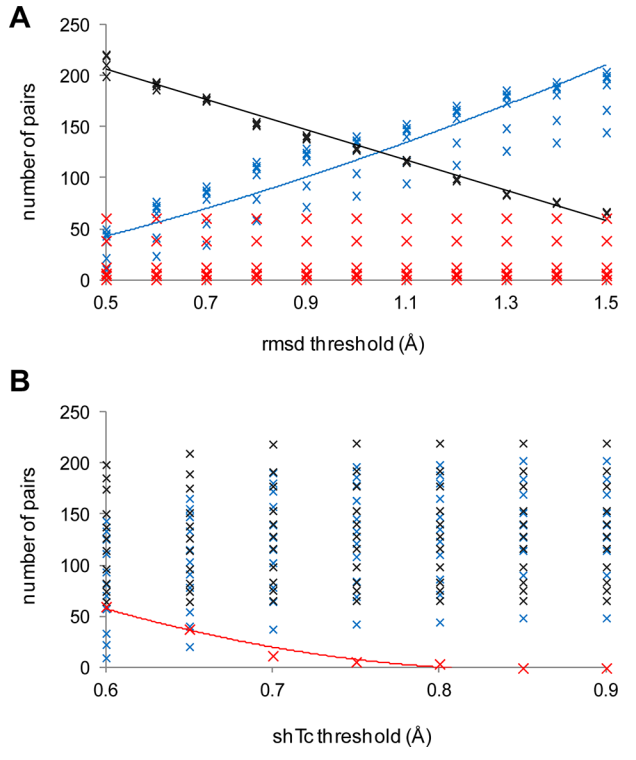


Figure 4. Structural characteristics of ligands in the pairs of dissimilar sites. The 269 pairs of dissimilar sites were classified as flexible ligand (black crosses), bianchor ligand (blue crosses), and difficult to rationalize (red crosses), for all possible combinations of thresholds for rmsd ranging from 0.5 to 1.5 Å (0.1 Å increment) and shTc ranging from 0.6 to 0.9 (0.1 increment). The linear, power, and order 2 polynomial trendlines were plotted for the flexible ligand (R^2 = 0.98), bianchor ligand (R^2 = 0.75), and difficult to rationalize (R^2 = 0.97) series, respectively.

dx.doi.org/10.1021/ci300196g | J. Chem. Inf. Model. 2012, 52, 2410−2421

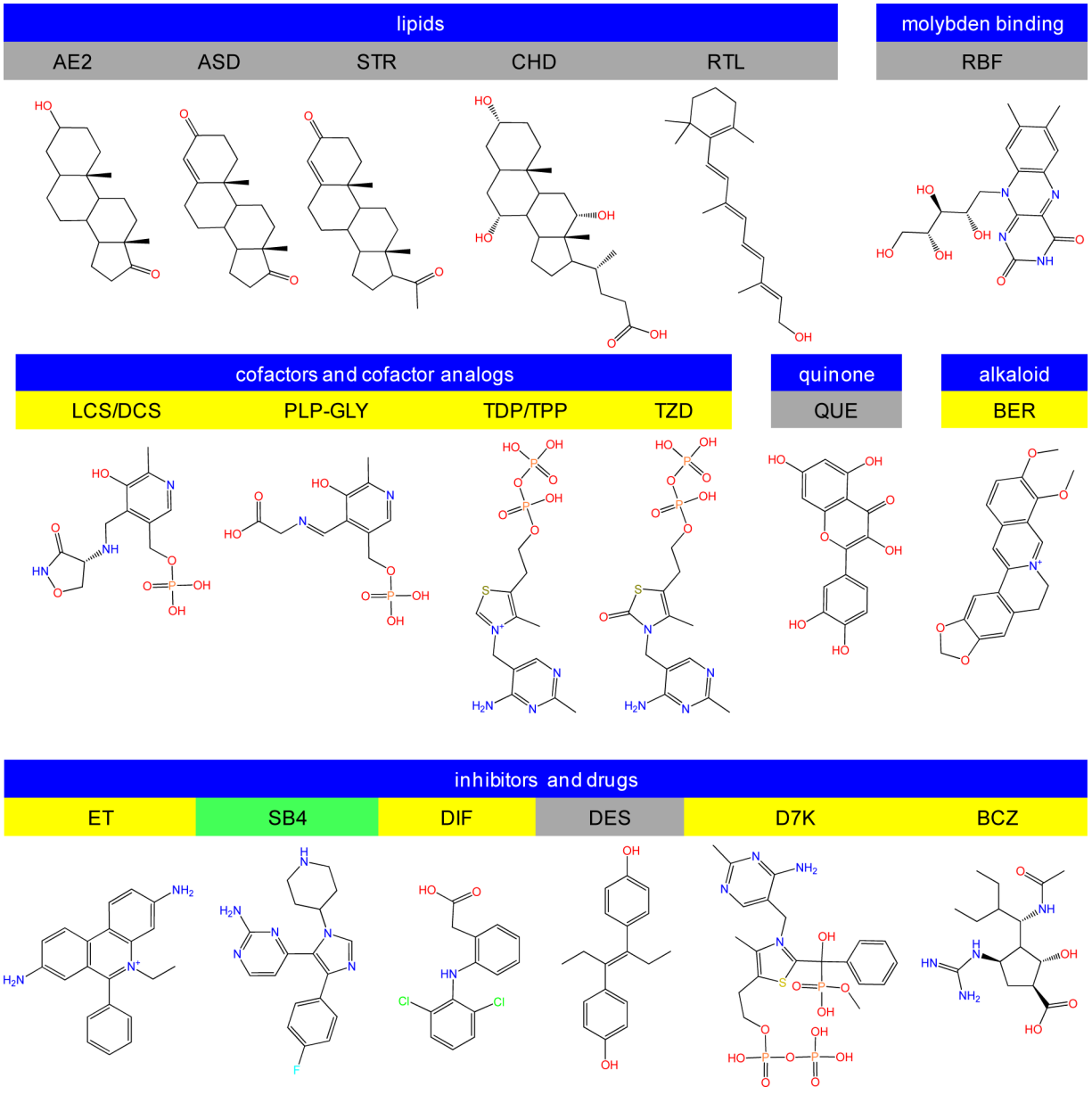


Figure 5. Chemical structure of superpromiscuous ligands. Ligands are labeled using their HET code, whose shading indicates the category of their parent pairs (green if homologous, yellow if convergent, and gray if distant), and are ordered according to their biochemical nature or their biological function.

varies, but the trends remain constant. In particular, most of the pairs of dissimilar sites correspond to ligands which adapt to the protein environment by changing their three-dimensional structure and/or their binding mode. Figure 4 indicates that the number of cases difficult to rationalize represents about 22% of the dissimilar pairs if shTc is equal to 0.6, that is, if at least 60% of the ligand atoms in interaction with one site are found among the ligand atoms in interaction with the other site of the pair. This number becomes zero if shTc is equal or higher than 0.85, meaning that at least 10% of the ligand atoms in contact with the protein are different in the two complexes.

These observations, which are based on 76 different ligands and 269 pairs of dissimilar sites, suggested that the ability of a ligand to bind to dissimilar sites principally results from its capability to modify its conformation. In addition, in about half of the pairs of dissimilar sites, the interacting atoms of the ligand in one complex constitute a subset of the interacting atoms of the ligand in the other complex (as indicated by shTv $\geq$ 1.5shTc), thus indicating that the ligand has different degrees of burial into the two proteins. The lack of similarity between sites is accordingly due to the limited size of the common ligand recognition area. In the remaining half of the pairs of dissimilar sites, the multiple possibilities of the ligand to form nonbonded interactions with a protein explain why it can bind to two topologically different sites.

Interestingly, 18 of 76 ligands were shown to use almost the same chemical moieties to bind to different sites (shTc $\geq$ 0.6, pairs which are difficult to rationalize in Figure 4), suggesting that different binding modes may be established from the same set of atoms of the ligand without significant conformational adaptation. From here on, we will call them the *superpromiscuous* ligands. Noteworthy, these ligands, with the exception of one of them, were in complex with nonhomologous proteins. They correspond to a limited number

dx.doi.org/10.1021/ci300196g | J. Chem. Inf. Model. 2012, 52, 2410−2421

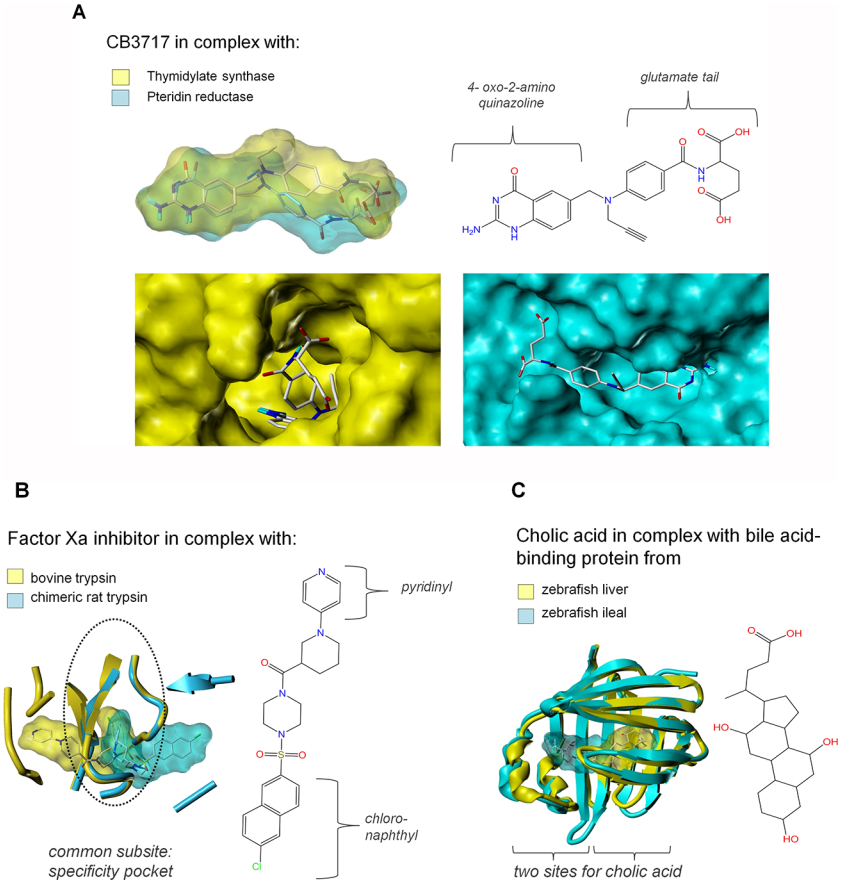


Figure 6. Examples of ligand bound to dissimilar sites. (A) HET code: CB3. PDB codes: 1an5 and 2bfa. (B) HET code: ZEN. PDB codes: 1ql8 and 1j17. (C) HET code: CHD. PDB codes: 2qo4 and 3elz. In A, B, and C, the ligand shapes are delimited by transparent solvent-excluded surfaces. In A, the protein shapes are delimited by solid solvent-excluded surfaces. In B and C, the three-dimensional structures of complexes are aligned for the best-fit of the protein backbones, as represented by ribbons (Sybyl X1.3, Tripos, Inc., St. Louis, MO, U.S.A).

of chemotypes: lipids (retinol and steroids), two cofactors, and a coenzyme which binds molybden, two natural products (a quinone and an alkaloid), and six enzyme inhibitors or drugs (Figure 5).

Examples of Multiple Binding Modes of a Ligand. An example of a ligand able to bind to dissimilar sites of a distant pair is given in Figure 6A. The ligand CB3717 binds to thymidylate synthase and pteridine reductase. The three-dimensional alignment of the two active ligand structures overlays the 4-oxo-2-amino quinazoline moiety, thus evidencing large variations in the rest of the molecule. This observation is in line with the experimental binding modes. Thymidylate synthase buries the entire CB3717 into its cofactor binding site, although the precise location of the 4-oxo-2-amino quinazoline moiety depends on the presence of a substrate.[39] In the complex between pteridine reductase and CB3717, the substrate-binding site mainly establishes nonbonded interactions with the 4-oxo-2-amino quinazoline while the glutamate tail of the ligand stretches out of the protein surface.[40]

Interestingly, we also observed ligands able to bind to dissimilar sites of homologous pairs (Figure 3). For example, the specific human factor Xa inhibitor (HET code: ZEN) is an inhibitor of bovine trypsin and of a rat trypsin mutant which was engineered to mimic factor Xa. The two enzymes have the same fold (rmsd of Cα atoms = 0.64 Å), and their amino acid sequence is highly conserved in the active site, thus defining virtually identical binding cavities. However changes in the nature of a few residues control the positioning and the affinity of ligand, so that the enzyme specificity pocket is occupied by the pyridine ring of the ligand in bovine trypsin[41] whereas it is occupied by a chloronaphthyl group in the chimeric rat trypsin (Figure 6B).[42] As a consequence, the two binding sites have only 13 residues in common, which represent only half of each site.

Alternate binding modes were also observed for the protein kinase inhibitor imatinib (HET code STI) in different tyrosine-protein kinases. Here substantial conformational changes at the secondary structure level induced either the tight binding of the inhibitor in an extended-conformation or a weaker binding to a more compact conformation (PDB codes: 1xbb, 2oiq).[43,44] In tyrosine kinases, these structural changes are involved in enzyme activation/inactivation. Changes in sequence and structure also explain the poor similarity between SB4 inhibitor-binding sites in Mitogen-activated protein (MAP) kinase 14 and MAP kinase 1 (PDB codes: 1bl7, 3erk), and between the antifolate LYA-binding site in human and protozoan thymidylate synthases (PDB codes: 1juj, 3k2h).The last example of a ligand bound to dissimilar sites in homologous pairs is cholic acid (HET code: CHD) which occupies different parts of a well conserved binding pocket in two homologue fatty acid-binding proteins (Figure 6C). Actually, in this family of enzymes, the number of cholate molecules per binding site is either one or two, and the

dx.doi.org/10.1021/ci300196g | J. Chem. Inf. Model. 2012, 52, 2410−2421

stoichiometry is finely tuned by the presence or not of a single disulfide bridge.[45]

Superpromiscuous Ligands Have Extreme Binding Modes. The first chemical class of superpromiscuous ligands (Figure 5) is made of lipids. In the studied complexes, we observed that the nonbonded interactions between lipids and their target protein involve principally hydrophobic contacts (from 9 to 18 amino acids establishing apolar contacts, none or a single amino acid establishing polar contacts). As an example, 4-androstene-3-17-dione was cocrystallized with two dehydrogenases of the steroid metabolism. The two binding sites have equivalent size and are almost exclusively lined by apolar residues, yet sequence variations cause differences in site enclosure (Figure 7A). Similar observations were made for the other steroid examples, and for retinol.

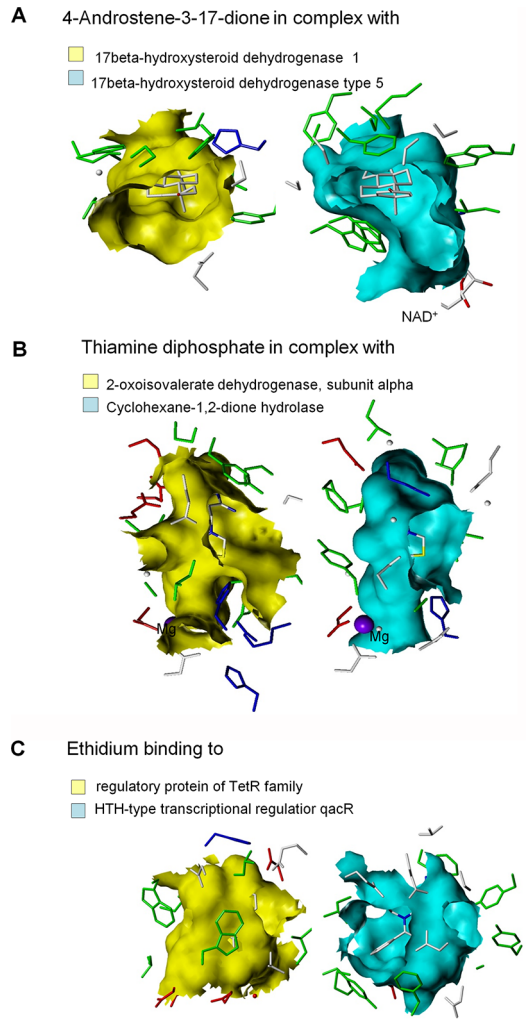


Figure 7. Examples of dissimilar sites which accommodate the same ligand while using similar binding modes. (A) HET code: ASD. PDB codes: 1qyx and 1xf0. The adenosine moiety of NAD+ cofactor is not depicted. (B) HET code: TPP. PDB codes: 1umb and 2pgn. (C) HET code: ET. PDB codes: 2zoz and 3br3. The orientation of two complexes corresponds to a fixed position of the ligand. Protein cavity shapes are delimited by solid solvent excluded surfaces. The side chains of ligand-interacting amino acids are represented by capped sticks and colored according to their property (red for acidic, blue for basic, white for neutral polar, and green for apolar). The bound ligands are represented by CPK-colored capped sticks.

Other essential natural metabolites were described in Figure 5, in particular the cofactors thiamine diphosphate (TDP, TPP, and TDZ) and pyridoxal (LCD/DCS and GLY-PLP). In the studied complexes, we observed that the molecular recognition involves numerous H-bonds and ionic interactions (more than 7 amino acids establishing polar contacts, these residues representing from 36% to 61% of the total number of residues in interaction with the ligand). The example of thiamine diphosphate in complex with a dehydrogenase and a hydrolase is given in Figure 7B. The binding pockets of the two enzymes have similar size and overall shape yet they have drastically different electrostatic properties. The binding mode is preserved because the intermolecular H-bonds involve protein backbone atoms, which anchor the aminopyrimidine moiety and a phosphate group of the coenzyme. The two enzymes also have in common a magnesium ion coordinated by the alpha and beta-phosphate groups of the cofactor.

Among the superpromiscuous ligands are also two natural products (Figure 5), the flavonoid quercetin and the alkaloid berberine, which both have a marked aromatic character. In the studied complexes, we observed that the two compounds establish none or one single H-bond to their target, even though quercetin contains 7 H-bond donors and acceptors.

These examples suggest that the interaction between natural molecules, which have numerous biological functions, and their multiple targets corresponds to an extreme binding mode, very hydrophobic or, on the contrary, very hydrophilic. Similar observations could not be made for the others superpromiscuous ligands (inhibitors and drugs). For example, ethidium which is a fluorescent DNA intercalating agent but also an antitrypanosomiasis drug, was found in two complexes with different bacterial transcriptional regulators. In the two complexes, we observed that ethidium establishes both apolar contacts and electrostatic interactions (three or eight aromatic stacking and H-bonds) with the protein, but that the two networks of nonbonded interactions are different. By considering the proteins, we could notice that the binding pockets share similar geometric features (the size and the overall shape are the same, although the opening are different) but exhibit very different electrostatic properties (Figure 7C).

At this point it is worth mentioning that crystallographic water molecules mediate up to eight intermolecular H-bonds between the superpromiscuous ligands and their target proteins. Furthermore, we noticed that the consideration of water molecules in protein sites yields a significant increase of similarity between sites for ethidium, pyridoxal, and quercetin. In detail, seven pairs of sites having as ligand ethidium, quercetin, or pyridoxal were predicted dissimilar if water is not included in proteins, whereas only three of them were predicted dissimilar if water is included in the proteins.

Do the Promiscuous Ligands Have Specific Characteristics? In this study, we considered 247 ligands capable of binding to different target proteins. Among them, 76 were demonstrated to be able to recognize different protein environments, and 18 of them were called superpromiscuous because they use almost the same anchor atoms in a preserved conformation to bind to different proteins. We have already mentioned that the superpromiscuous ligands are of limited chemical diversity. We here investigated whether the ligands in our data set possess specific chemical features that distinct them from other drug-like ligands. In particular, we compared the 247 promiscuous ligands with ligands in two other data sets: one composed of 959 approved drugs and the second one of

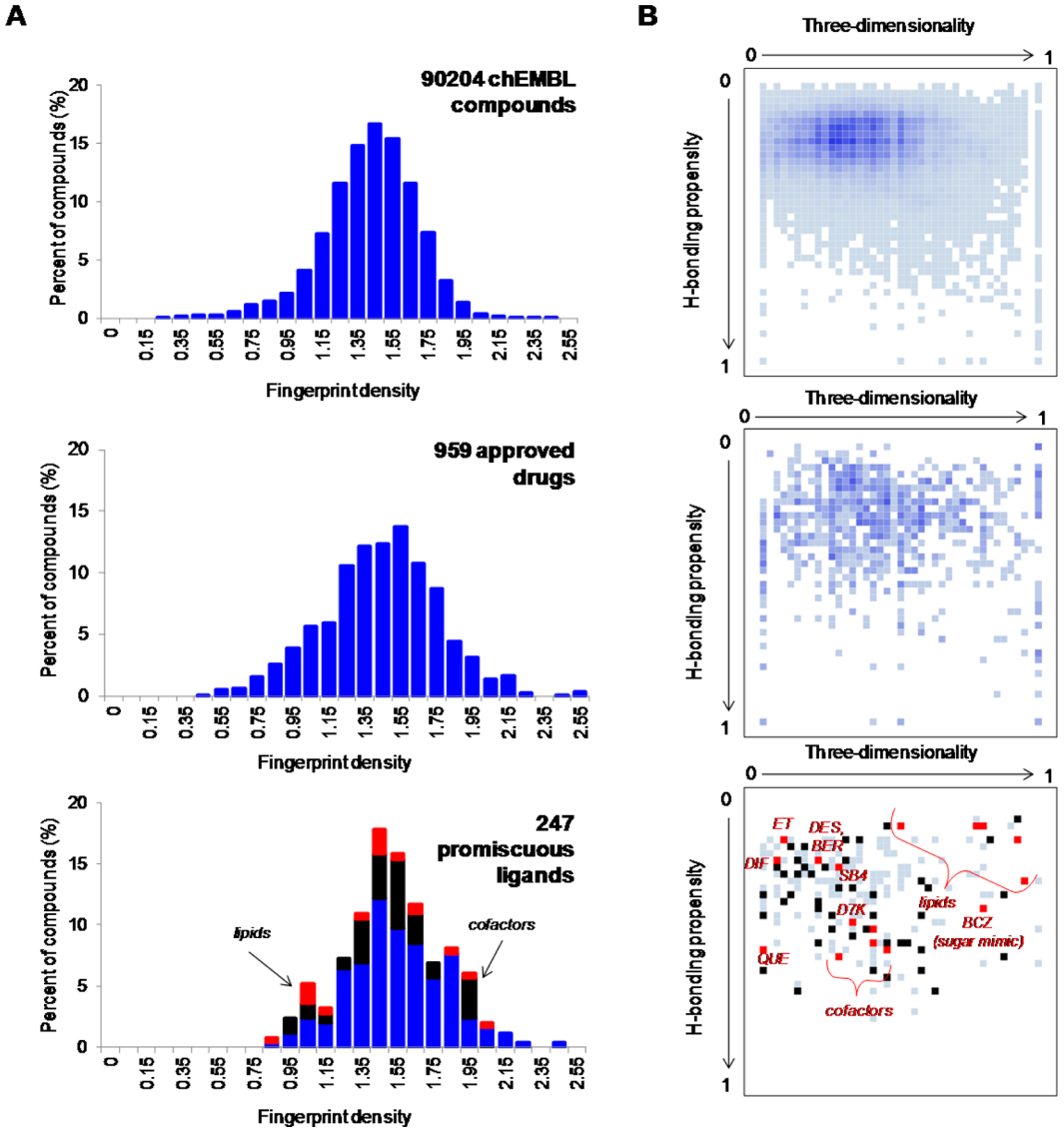


Figure 8. Comparison of the promiscuous ligands in the data set (bottom panels) with drugs (medium panels) and with bioactive compounds for which a single target is known (top panels). (A) Molecular complexity. (B) H-bonding propensity and molecular three-dimensionality. In the top and middle panels of B, the intensity of the color reflects the number of ligands in each bin. In the bottom panel of A and B, the 76 promiscuous ligands which adapt to different protein environments are highlighted in black, except the 18 superpromiscuous ligands which are colored in red. The plots were generated using the php library gg2.0.34.

90 204 bioactive compounds. The drugs were retrieved from Drugbank by querying FDA-approved "small molecule" drugs (whose molecular weight is lower than 900), but not nutraceuticals, biologics, or experimental drugs. The bioactive compounds were retrieved from ChEMBL[46] by querying compounds for which a single target has been reported and whose affinity for its target is higher than 6 (as expressed by the logarithm of a dissociation or inhibition constant).

We analyzed three molecular descriptors, molecular complexity expressed as the circular fingerprint density,[47] H-bonding propensity, and three-dimensionality (see Material and Methods). The molecular complexity in the data set of bioactive compounds follows a normal distribution (Figure 8A). In the data set of approved drugs, the molecular complexity is in the same value range as the data set of bioactive compounds, but values are more scattered around the mean value than in a normal distribution. This trend is even more pronounced in the data set of promiscuous ligands: it thus appears that this data set is rich in molecules of low complexity (including the lipids) and in molecules of high complexity (including the cofactors), which both have been classified as the superpromiscuous ligands. To further delineate molecular complexity, we partitioned the data sets according to H-bond propensity and three-dimensionality (Figure 8B). Again, the chemical space defined by the extreme values of the two properties is common to all three data sets, but the distributions of points varies significantly. The data set of promiscuous ligands especially occupies regions which are not highly populated in the two other data sets. More precisely, we could spot superpromiscuous ligands in regions of high three-dimensionality and low H-bonding propensity (including lipids), in regions of high three-dimensionality and high H-bonding propensity (including the sugar mimic BCZ), in regions of very low three-dimensionality and H-bonding propensity (including the inhibitors ET and DIF), and in

dx.doi.org/10.1021/ci300196g | J. Chem. Inf. Model. 2012, 52, 2410−2421

regions medium three-dimensionality and high H-bonding propensity (including cofactors).

Altogether, our findings suggested that drug-like compounds are able to adapt to different protein environments if they are flexible or if they possess specific chemical features, such as a high proportion of aromatic rings with few or no aliphatic hydrophobic groups, or on the opposite, a high proportion of aliphatic hydrophobic groups with few or no polar atoms.

Do the Promiscuous Ligands Have Similar Affinity for Their Different Targets? Promiscuity can be defined as the ability of a compound to exert its effects through multiple biological targets. In high-throughput screening assays, a binding affinity of 10 μM (K_i, IC_{50}) is commonly used to detect hits and therefore to consider a protein as a target. Because X-ray diffraction can observe much lower affinity complexes (mM), we investigated the binding affinity values associated with PDB files in our data set. Upon parsing bindingMOAD[48] and bindingDB[49] databases, binding affinity data could be retrieved for 265 pairs (Figure 9). All but 40 pairs

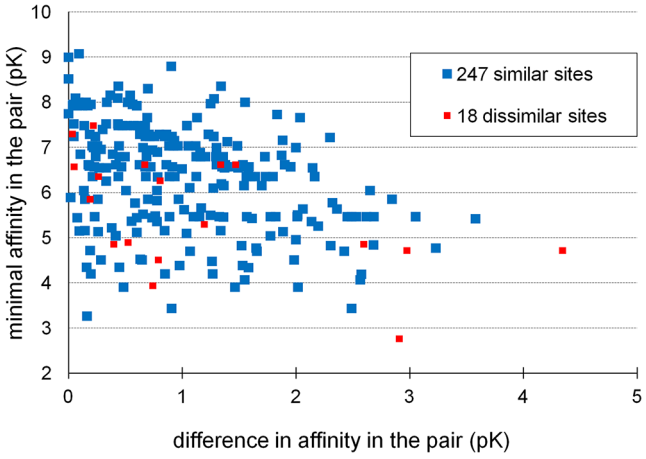


Figure 9. Ligand affinity in 265 pairs of targets for promiscuous ligands. For each ligand/protein complex, affinity represents the average value among pIC_{50}, pEC_{50}, pK_i, and pK_D data retrieved from BindingMOAD and binding databases. Differences in affinity and maximal affinities are expressed in pk units.

met the affinity threshold of 10 μM for the two targets. Although these data are not sufficient to establish robust statistics, it appeared that neither a low affinity for both sites nor a high affinity difference is somehow correlated with the degree of similarity between the corresponding binding sites.

CONCLUSIONS

In the present study, we addressed the issue of ligand promiscuity from a structural point of view. Such an approach was already carried out for a small number of primary metabolites (glucose, nucleotides, heme, estradiol) capable of binding to many different proteins.[34,50] We focused herein on *drug-like* ligands and *druggable* proteins. We proposed the critical analysis of a wide and diverse data set of ligands which are present in PDB complexes with two or more different proteins.

By comparing the different proteins targeted by a ligand at the level of their sequence, structure or binding-site conservation, we demonstrated that ligand promiscuity is either due to the presence of similar binding cavities in different proteins which do not necessarily share other evolutionary relationships (conservation of amino acid or of overall fold) or to specific characteristics of the ligand itself. The conformational flexibility of the ligand frequently explained why discrepancies in size, shape, and physicochemical properties were observed between the binding sites of different targets. Accordingly, we could also observe substantial variations in the number and nature of the ligand atoms in direct interaction with protein. Lastly, we identified a small number of ligand chemotypes that are able to remarkably adapt to different protein environments. In particular, we provided evidence that compounds of low complexity and compounds of very high complexity are prone to bind to dissimilar sites even though their conformation and their chemical moiety bound to proteins are conserved. Noteworthy, natural metabolites were the most promiscuous compounds in the studied data set, thereby suggesting that Nature has developed diverse protein architectures to bind metabolically important ligands ("hubs") like lipids, coenzymes (e.g., thiamine diphosphate participates to many enzymatic reactions like dehydrogenation, decarboxylation or transketolasation), ancient metabolites (e.g., pyridoxal[51]), and widely distributed natural products (e.g., quercetin is present in large quantity in many plants). Last, our findings are consistent with the suggested link between the lipophilicity of a compound and its prosmiscuity[15] and with the importance of three-dimensionality in advancing drug candidates to late clinical stages.[52]

ASSOCIATED CONTENT

Supporting Information

The rules for the biochemical classification of the sc-PDB ligands (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +333 68 85 42 21. Fax: +333 68 85 43 10. E-mail: ekellen@unistra.fr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The CC-IN2P3 (Villeurbanne) and GENCI (Project x2011075024) are acknowledged for providing computational resources to this study. We thank the Eskitis Institute at Griffith University for financial support.

ABBREVIATIONS

MAP, Mitogen-activated protein; PDB, Protein Data Bank; rmsd, root-mean-square deviation; shTc, shape Tanimoto coefficient; shTv, shape Tversky coefficient

REFERENCES

(1) Rognan, D. Chemogenomic approaches to rational drug design. *Br. J. Pharmacol.* **2007**, *152*, 38−52.

(2) van der Horst, E.; Peironcely, J.; van Westen, G.; van den Hoven, O.; Galloway, W.; Spring, D.; Wegner, J.; van Vlijmen, H.; Ijzerman, A.; Overington, J.; Bender, A. Chemogenomics approaches for receptor deorphanization and extensions of the chemogenomics concept to phenotypic space. *Curr. Top. Med. Chem.* **2011**, *11*, 1964−1977.

(3) Vidal, D.; Garcia-Serna, R.; Mestres, J. Ligand-based approaches to in silico pharmacology. *Methods Mol. Biol.* **2011**, *672*, 489−502.

(4) Campillos, M.; Kuhn, M.; Gavin, A.-C.; Jensen, L. J.; Bork, P. Drug Target Identification Using Side-Effect Similarity. *Science* **2008**, *321*, 263−266.
(5) Kellenberger, E.; Hofmann, A.; Quinn, R. Similar Interactions of Natural Products with Biosynthetic Enzymes and Therapeutic Targets could explain why Nature produces such a Large Proportion of Existing Drugs. *Nat. Prod. Rep.* **2011**, *28*, 1483−1492.
(6) Nicola, G.; Liu, T.; Gilson, M. K. Public Domain Databases for Medicinal Chemistry. *J. Med. Chem.* **2012**, DOI: 10.1021/jm300501t.
(7) Wassermann, A.; Bajorath, J. BindingDB and ChEMBL: online compound databases for drug discovery. *Expert. Opin. Drug. Discov.* **2011**, *6*, 683−687.
(8) Bender, A.; Scheiber, J.; Glick, M.; Davies, J. W.; Azzaoui, K.; Hamon, J.; Urban, L.; Whitebread, S.; Jenkins, J. L. Analysis of Pharmacology Data and the Prediction of Adverse Drug Reactions and Off-Target Effects from Chemical Structure. *Chem. Med. Chem.* **2007**, *2*, 861−873.
(9) Defranchi, E.; Schalon, C.; Messa, M.; Onofri, F.; Benfenati, F.; Rognan, D. Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity measurements. *PLoS One* **2010**, *5*, e12214.
(10) Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L. H.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175−181.
(11) Kinnings, S. L.; Liu, N.; Buchmeier, N.; Tonge, P. J.; Xie, L.; Bourne, P. E. Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput. Biol.* **2009**, *5*, e1000423.
(12) Lounkine, E.; Keiser, M. J.; Whitebread, S.; Mikhailov, D.; Hamon, J.; Jenkins, J. L.; Lavan, P.; Weber, E.; Doak, A. K.; Cote, S.; Shoichet, B. K.; Urban, L. Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **2012**, *486*, 361−367.
(13) Paolini, G. V.; Shapland, R. H. B.; van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L. Global mapping of pharmacological space. *Nat. Biotechnol.* **2006**, *24*, 805−815.
(14) Krejsa, C.; Horvath, D.; Rogalski, S.; Penzotti, J.; Mao, B.; Barbosa, F.; Migeon, J. Predicting ADME properties and side effects: the BioPrint approach. *Curr. Opin. Drug Discov. Devel.* **2003**, *6*, 470−480.
(15) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* **2007**, *6*, 881−890.
(16) Peters, J.-U.; Hert, J.; Bissantz, C.; Hillebrecht, A.; Gerebtzoff, G. g.; Bendels, S.; Tillier, F.; Migeon, J.; Fischer, H.; Guba, W.; Kansy, M. Can we discover pharmacological promiscuity early in the drug discovery process? *Drug Discovery Today* **2012**, *17*, 325−335.
(17) Yang, Y.; Chen, H.; Nilsson, I.; Muresan, S.; Engkvist, O. Investigation of the relationship between topology and selectivity for druglike molecules. *J. Med. Chem.* **2010**, *53*, 7709−7714.
(18) Young, R. J.; Green, D. V. S.; Luscombe, C. N.; Hill, A. P. Getting physical in drug discovery II: the impact of chromatographic hydrophobicity measurements and aromaticity. *Drug Discovery Today* **2011**, *16*, 822−830.
(19) Berman, H. M.; Kleywegt, G. J.; Nakamura, H.; Markley, J. L. The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future. *Structure* **2012**, *20*, 391−396.
(20) Meslamani, J.; Rognan, D.; Kellenberger, E. sc-PDB: a database for identifying variations and multiplicity of "druggable" binding sites in proteins. *Bioinformatics* **2011**, *27*, 1324−1326.
(21) Kellenberger, E.; Foata, N.; Rognan, D. Ranking targets in structure-based virtual screening of 3-D protein libraries: Methods and Problems. *J. Chem. Inf. Model.* **2008**, *48*, 1014−1025.
(22) The UniProt, C., Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2012**, *40*, D71−D75.
(23) Cole, J. C.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Comparing protein−ligand docking programs is difficult. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 325−332.
(24) Marcou, G.; Rognan, D. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *J. Chem. Inf. Model.* **2006**, *47*, 195−207.
(25) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653−1666.
(26) Bourne, P. E.; Beran, B.; Bi, C.; Bluhm, W. F.; Dimitropoulos, D.; Feng, Z.; Goodsell, D. S.; Prlić, A.; B. Quinn, G.; W. Rose, P.; Westbrook, J.; Yukich, B.; Young, J.; Zardecki, C.; Berman, H. M. The evolution of the RCSB Protein Data Bank website. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 782−789.
(27) Mullan, L. J.; Bleasby, A. J. Short EMBOSS User Guide. European Molecular Biology Open Software Suite. *Brief Bioinform.* **2002**, *3*, 92−94.
(28) Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **1999**, *12*, 85−94.
(29) Shindyalov, I. N.; Bourne, P. E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **1998**, *11*, 739−747.
(30) Desaphy, J.; Azdimoussa, K.; Kellenberger, E.; Rognan, D. Comparison and prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. *J. Chem. Inf. Model.* **2012**, DOI: 10.1021/ci300184x.
(31) Schalon, C.; Surgand, J.-S.; Kellenberger, E.; Rognan, D. A simple and fuzzy method to align and compare druggable ligand-binding sites. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 1755−1778.
(32) Weill, N.; Rognan, D. Alignment-free ultra-high-throughput comparison of druggable protein-ligand binding sites. *J. Chem. Inf. Model.* **2010**, *50*, 123−135.
(33) Gashaw, I.; Ellinghaus, P.; Sommer, A.; Asadullah, K. What makes a good drug target? *Drug Discovery Today* **2012**, *17*, S24−S30.
(34) Stockwell, G. R.; Thornton, J. M. Conformational diversity of ligands bound to proteins. *J. Mol. Biol.* **2006**, *356*, 928−944.
(35) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliver. Rev.* **2001**, *46*, 3−26.
(36) Ursu, O.; Rayan, A.; Goldblum, A.; Oprea, T. I. Understanding drug-likeness. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 760−781.
(37) Wishart, D. S. DrugBank and its relevance to pharmacogenomics. *Pharmacogenomics* **2008**, *9*, 1155−1162.
(38) Galperin, M. Y.; Koonin, E. V. Divergence and Convergence in Enzyme Evolution. *J. Biol. Chem.* **2012**, *287*, 21−28.
(39) Stout, T. J.; Sage, C. R.; Stroud, R. M. The additivity of substrate fragments in enzyme-ligand binding. *Structure* **1998**, *6*, 839−848.
(40) Schüttelkopf, A. W.; Hardy, L. W.; Beverley, S. M.; Hunter, W. N. Structures of Leishmania major Pteridine Reductase Complexes Reveal the Active Site Features Important for Ligand Binding and to Guide Inhibitor Design. *J. Mol. Biol.* **2005**, *352*, 105−116.
(41) Stubbs, M. T.; Reyda, S.; Dullweber, F.; Möller, M.; Klebe, G.; Dorsch, D.; Mederski, W. W. K. R.; Wurziger, H. pH-Dependent Binding Modes Observed in Trypsin Crystals: Lessons for Structure-Based Drug Design. *Chem. Bio. Chem.* **2002**, *3*, 246−249.
(42) Reyda, S.; Sohn, C.; Klebe, G.; Rall, K.; Ullmann, D.; Jakubke, H.-D.; Stubbs, M. T. Reconstructing the Binding Site of Factor Xa in Trypsin Reveals Ligand-induced Structural Plasticity. *J. Mol. Biol.* **2003**, *325*, 963−977.
(43) Jacobs, M. D.; Caron, P. R.; Hare, B. J. Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: Structure of lck/imatinib complex. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 1451−1460.
(44) Atwell, S.; Adams, J. M.; Badger, J.; Buchanan, M. D.; Feil, I. K.; Froning, K. J.; Gao, X.; Hendle, J. r.; Keegan, K.; Leon, B. C.; Müller-Dieckmann, H. J.; Nienaber, V. L.; Noland, B. W.; Post, K.;

Rajashankar, K. R.; Ramos, A.; Russell, M.; Burley, S. K.; Buchanan, S. G. A Novel Mode of Gleevec Binding Is Revealed by the Structure of Spleen Tyrosine Kinase. *J. Biol. Chem.* **2004**, *279*, 55827−55832.
(45) Capaldi, S.; Guariento, M.; Saccomani, G.; Fessas, D.; Perduca, M.; Monaco, H. L. A Single Amino Acid Mutation in Zebrafish (Danio rerio) Liver Bile Acid-binding Protein Can Change the Stoichiometry of Ligand Binding. *J. Biol. Chem.* **2007**, *282*, 31008−31018.
(46) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.
(47) Selzer, P.; Roth, H.-J. r.; Ertl, P.; Schuffenhauer, A. Complex molecules: do they add value? *Curr. Opin. Chem. Biol.* **2005**, *9*, 310−316.
(48) Benson, M. L.; Smith, R. D.; Khazanov, N. A.; Dimcheff, B.; Beaver, J.; Dresslar, P.; Nerothin, J.; Carlson, H. A. Binding MOAD, a high-quality protein-ligand database. *Nucleic Acids Res.* **2008**, *36*, D674−D678.
(49) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198−D201.
(50) Kahraman, A.; Morris, R. J.; Laskowski, R. A.; Favia, A. D.; Thornton, J. M. On the diversity of physicochemical environments experienced by identical ligands in binding pockets of unrelated proteins. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1120−1136.
(51) Kim, K. M.; Qin, T.; Jiang, Y.-Y.; Chen, L.-L.; Xiong, M.; Caetano-Anollés, D.; Zhang, H.-Y.; Caetano-Anollés, G. Protein Domain Structure Uncovers the Origin of Aerobic Metabolism and the Rise of Planetary Oxygen. *Structure* **2012**, *20*, 67−76.
(52) Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **2009**, *52*, 6752−6756.