Title
The Recognition of Identical Ligands by Unrelated Proteins

Journal
ACS Chemical Biology, 10(12)

ISSN
1554-8929

Authors
Barelier, Sarah
Sterling, Teague
O'Meara, Matthew J
et al.

Publication Date
2015-12-18

DOI
10.1021/acschembio.5b00683

Peer reviewed

HHS Public Access

Author manuscript

ACS Chem Biol. Author manuscript; available in PMC 2016 December 18.

Published in final edited form as:

ACS Chem Biol. 2015 December 18; 10(12): 2772–2784. doi:10.1021/acschembio.5b00683.

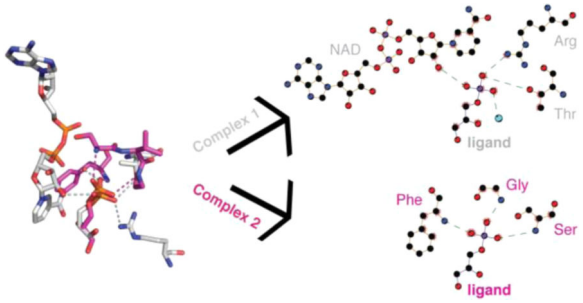
The Recognition of Identical Ligands by Unrelated Proteins

Sarah Barelier, **Teague Sterling**, **Matthew J. O'Meara**, and **Brian K. Shoichet***

Department of Pharmaceutical Chemistry, University of California San Francisco, 1700 Fourth Street, Byers Hall, San Francisco, California 94158, United States

Abstract

The binding of drugs and reagents to off-targets is well-known. Whereas many off-targets are related to the primary target by sequence and fold, many ligands bind to unrelated pairs of proteins, and these are harder to anticipate. If the binding site in the off-target can be related to that of the primary target, this challenge resolves into aligning the two pockets. However, other cases are possible: the ligand might interact with entirely different residues and environments in the off-target, or wholly different ligand atoms may be implicated in the two complexes. To investigate these scenarios at atomic resolution, the structures of 59 ligands in 116 complexes (62 pairs in total), where the protein pairs were unrelated by fold but bound an identical ligand, were examined. In almost half of the pairs, the ligand interacted with unrelated residues in the two proteins (29 pairs), and in 14 of the pairs wholly different ligand moieties were implicated in each complex. Even in those 19 pairs of complexes that presented similar environments to the ligand, ligand superposition rarely resulted in the overlap of related residues. There appears to be no single pattern-matching "code" for identifying binding sites in unrelated proteins that bind identical ligands, though modeling suggests that there might be a limited number of different patterns that suffice to recognize different ligand functional groups.



The search for ligands specific for their receptors has dominated medicinal chemistry for a century.[1] Meanwhile, a central dogma of biology has been the fidelity of information flow

*Corresponding Author Phone: +1 415 514 4126. shoichet@cgl.ucsf.edu..

The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.5b00683. Tables 1-SI to 3-SI, including the ligand codes, PDB codes, fold, Electrostatics Energy Mean Square Deviation (EEMSD), and van der Waals Energy Mean Deviation (VEMD) for all pairs of complexes; Figures 1-SI to 4-SI, including the electrostatics and van der Waals energy profiles and ligplot representation for all pairs of complexes; and the effect of minimization on the ligand electrostatics and van der Waals energies (PDF)

from gene to protein to folded structure to specific activity. Thus, when seeking “off-targets” to which drugs may bind, it has been natural to focus on proteins related in sequence and structure to the primary target. Obtaining specificity against a related human target, not involved in the disease but perhaps in an adverse reaction, or for a pathogen target while sparing the human homologue, often requires capitalizing on subtle differences in the binding sites. Such optimization can be difficult, but the nature of the challenge is well understood. More perplexing is the possibility that a ligand might modulate a protein unrelated in sequence and structure to its primary target.

The advent of large ligand-target association databases[2–4] has revealed that these off-targets can bear little relationship to their primary ones. Paolini and colleagues observed that not only do molecules targeting aminergic G Protein Coupled Receptors (GPCRs) cross-react with other GPCRs but they often are active on protein kinases, while kinase inhibitors in turn have activity on ion channels and phosphodiesterases.[5] Bork and colleagues linked related side-effects to predict that raloxifene, an estrogen nuclear hormone receptor (NHR) drug, also inhibits the $5HT_{1D}$ GPCR,[6] that the proton-pump inhibitor rabeprazole also acts on the dopamine D_3 GPCR, and that the antihistamine loratadine modulates the GABA ion channel. A large-scale study found that many approved drugs act on targets unrelated to their primary efficacy targets,[7] with over a quarter cross major target boundaries. Thus, kinase inhibitors antagonized GPCRs, GPCR ligands antagonized ion channels, ion channel modulators bound to NHRs, among others. Intriguingly, some of this polypharmacology tracks that of endogenous hormones and neurotransmitters, which also modulate multiple receptors unrelated in sequence or structure.[8] Among other examples, acetylcholine, glutamate, serotonin, and ATP all modulate both ion channels and GPCRs as primary signaling receptors, while estradiol, progesterone, and leukotrienes do the same against both NHRs and GPCRs.

If mimicry of endogenous signaling molecules may suggest an origin for drug polypharmacology,[8] it does not explain its structural basis. By analogy to convergent enzyme evolution,[9] one might imagine that two binding sites recognizing the same ligand will have similar residues making similar interactions with the ligand. If true, then the problem of predicting off-targets would resolve into detecting similar binding sites in two otherwise unrelated proteins. This is the case that would most easily fit with our current target-based approaches to medicinal chemistry and chemical biology. However, we already know that similar ligand motifs can be recognized by altogether different environments (Figure 1).[10,11] We can reasonably expect that some ligands at least will bind to unrelated binding sites. Indeed, in a study of 100 complexes involving nine cofactors, Kahraman and Thornton found that unrelated binding sites, in different folds, could recognize each cofactor.[12,13] For the prediction of enzyme function, a focus of that study, they found no simple mapping between ligands and recognition pockets.

Three possibilities can be distinguished for how an identical ligand interacts with unrelated proteins (Figure 2). The first is where the same ligand groups make similar interactions with related residues in both binding sites (class A). A second possibility is where the same ligand groups interact with dissimilar residues and environments in the two binding sites

(class B). Finally, different ligand groups may interact with the protein, such that a different part of the ligand makes the defining pharmacophore interactions in each complex (class C).

Inspection of ligand–protein complexes reveals examples of each of these cases (Figure 3). For instance, the anti-Alzheimer's drug galanthamine (GNT) binds to both the mainly-β distorted sandwich acetylcholine binding protein[14] (PDB 2ph9) and the α/β 3-layer (aba) sandwich acetylcholine esterase[15] (PDB 1dx6; Figure 3, class A). Though superposition of the two complexes finds few overlapping residues, recognition is dominated by cation–π interactions at the aminergic cation, and a similar mixture of nonpolar and hydrogen bond interactions on the other side of the ligand, in both complexes. Conversely, carboxyaminoimidazole ribonucleotide (CAIR, ligand code C2R) finds different environments in its complexes with the α/β three-layer (aba) sandwich N5-CAIR mutase[16] (PurE, PDB 2nsl) and the α/β two-layer sandwich SAICAR synthetase[17] (PurC, PDB 2gqs). CAIR binds to each protein with similar affinities (21[16,18] and 7.8 μM,[19] respectively), but different protein recognition motifs are engaged (Figure 3, class B). These differences are most prominent in the ribose-carboxyimidazole moiety, which in PurE forms a network of hydrogen bonds to main chain atoms of six different residues. In PurC, the same groups are recognized by two catalytic magnesium ions and hydrogen bonds from an arginine and an aspartate. The imidazole is partly exposed in PurC but is buried in PurE. Finally, the metabolite allantoin (2AL) binds both to the α-bundle 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU) decarboxylase (PDB 2o73)[20] and to the α/β-fold urate oxidase[21] (PDB 2fxl; Figure 3, class C). Wholly different ligand groups are implicated in each. In the OHCU decarboxylase complex, for which allantoin is the reaction product, multiple hydrogen bonds are made to the imidazole-dione core, and the ligand is entirely buried. Conversely, with urate oxidase, which is two enzymes upstream from OHCU decarboxylase in the xanthine degradation pathway, the urea tail forms the core hydrogen bonds while the imidazole-dione ring is solvent exposed.

It thus seemed interesting to interrogate the protein data bank more extensively for complexes that share the same ligand but are unrelated by sequence or fold. We compared 59 ligands binding in 116 complexes by ligand superposition and examination of the residues in each binding site, both manually and computationally. In doing so, we hoped to address the following questions: Is there a pattern or code for how an identical ligand is recognized by pockets in proteins unrelated by fold? More specifically, are there similar interaction residues and environments in the two proteins that one might hope to identify by pattern matching, if only loosely? In cases where this is not true, how can we understand the ability of unrelated binding sites to bind not only related ligands, but exactly the same ligand? Is this just a rare curiosity, or should we expect it to be common?

RESULTS

Fifty-nine pairs of proteins with unrelated folds but binding the same ligand were examined at atomic resolution, using both calculated van der Waals and electrostatic energies, and by visual inspection. Three pairs of proteins binding similar, not identical, ligands were also included (62 pairs total). Not every complex that met the criterion of identical ligand binding to two- or more-fold families is presented here; small fragments, and cofactors studied

previously like ATP, ADP, and NAD,[12,13,22] were excluded, as were promiscuous ligands. Because our priority was to ensure that all pairs had different folds, we focused on partners with 10% or less sequence identity that also had different domain descriptions; all pairs were also visually inspected (Experimental Section). Thus, the complexes presented here are not comprehensive, though they are unbiased in the sense that we did not prechoose them by category. Ultimately, 59 ligands binding in 116 complexes representing 62 pairs with different folds were fully analyzed.

Summary of the Analysis

What follows is a description of representative complexes in each of the three categories: A, B, and C. We begin with a brief overview of the main observations.

Each pair of complexes was visually inspected and placed in one of three classes: A if the same ligand groups interacted with similar protein groups (19 pairs), B if the same ligand groups interacted with different protein environments (29 pairs), C if different ligand groups interacted with the proteins (14 pairs; Figure 2). The environments of the ligand complexes were also investigated computationally, comparing the molecular potentials felt by each ligand atom in any given pair of complexes. For each ligand atom, the van der Waals and the electrostatics complementarity were compared, using the mean deviation (van der Waals) or the mean square deviation (electrostatics) of the energies in each complex (Figure 4). As expected, class A pairs, where the ligands encounter similar environments, have more similar electrostatics energies (average electrostatic mean-square deviation EEMSD of 35.3 kcal/mol) than class B and C (average EEMSD of 192.7 and 376 kcal/mol, respectively). Though the differences in the van der Waals energies were smaller in magnitude, class C pairs, where the ligands make use of wholly different groups to bind the proteins, have larger differences (average van der Waals energy mean deviation VEMD of 0.68 kcal/mol) than class A and class B pairs (average VEMD of 0.52 and 0.57 kcal/mol, respectively; Supporting Information Figures 1–3).

A key result is that for 43 out of 62 complexes investigated, the identical ligand was recognized by receptor environments that were not even approximately related, at least at the residue level, in the pairs of fold- and sequence-unrelated proteins. We examine several examples in more detail.

Class A Pairs of Complexes

Of the 62 pairs of complexes, 19 were placed into class A, where the environments experienced by the ligand were similar (characteristic examples are shown in Figure 5; the folds and a full list of complexes may be found in Table 1-SI). We had expected to find binding sites that interacted with the ligand via related amino acids with, for instance, an aspartate in one site mapping to a glutamate in the other, a serine to a threonine, and so forth. This was rarely found, and in placing pairs of complexes into class A we relied on broader environmental analogy.

An example of a class A pair was that of the anti-inflammatory drug ibuprofen (IBP) in complex with the α/β fold acyl-CoA synthetase[23] (PDB 2wd9) and with the β-barrel fatty-

acid binding protein FABP4[24] (PDB 3p6h; Figure 5A). Though superposition of the two complexes finds few residues that overlap, recognition is driven by the ligand carboxylate hydrogen-bonding with an arginine and a threonine hydroxyl in one complex, and a tyrosine hydroxyl in the other. In both cases, the ligand makes additional nonpolar interactions with hydrophobic residues. Correspondingly, the electrostatics and van der Waals energies profiles are similar (EEMSD 22.1 and VEMD 0.4, see Figure 5A and Table 1-SI).

Another pair grouped into class A was that of arachidonic acid (ACD) binding to both the β-barrel fatty acid binding protein Sm14[25] (PDB 1vyg) and to the α-orthogonal bundle prostaglandin H1 synthase[26] (PDB 1diy). The ligand carboxylate hydrogen-bonds to arginine, threonine, and tyrosine residues in one complex, and to arginine and tyrosine in the other (Figure 5B). Here, the electrostatics energies are very similar (EEMSD 7.0, Figure 4 and Table 1-SI), while the van der Waals energies differ slightly, although the van der Waals energy mean deviation score for this pair of complexes remains well within class A standards (VEMD 0.6, Figure 4 and Table 1-SI). Another lipid grouped into class A was linoleic acid (EIC), which bound to both the β-barrel lipocalin-like protein (PDB 4nyq) and to the Rossmann-fold NAD(P)-binding hydratase[27] (PDB 4ia6). Here, in contrast to arachidonic acid, the lipid's carboxylate is exposed to solvent in both complexes, while the hydrophobic tail is involved in hydrophobic interactions (Figure 5C). Both the electrostatics and van der Waals energy profiles are similar (EEMSD 15.2 and VEMD 0.4, see Table 1-SI).

Finally, the complexes between the β-blocker carazolol (CAU) with the 7-TM β2-adrenergic receptor (β2-AR)[28] (PDB 2rh1) and the antidepressant clomipramine (CXX) with the 12-TM LeuT transporter[29] (PDB 2q6h) are dominated by similar interactions (Figure 5D). Carazolol and clomipramine are not, of course, identical—they are one of the three pairs in this study that are not—but they are similar, and clomipramine, like many transporter inhibitors, also binds to GPCRs, including the β2-AR. Both molecules ion pair through their aminergic nitrogen to an aspartate and make extensive nonpolar contacts with the binding sites.

It is worth noting that in most class A complexes where the environments were analogous, they were never identical, and even residue types were not conserved.

Class B Pairs of Complexes

Twenty-nine pairs of complexes were categorized into class B, where the same ligand groups are recognized but by much different protein environments (Figure 6 and Table 2-SI). An example relevant to human therapy is the drug cycloserine (4AX), which binds both to its bacterial target alanine racemase[30] (PDB 1xql), a TIM barrel, and to the ligand binding domain of the α/β 3-layer (aba) sandwich NMDA receptor[31] (NMDAR; PDB 1pb9) at low micromolar concentrations (Figure 6A).[32] The NMDAR binding is consistent with the drug's profound and dose-limiting psychotropic side effects. Notwithstanding its small size, unrelated residues are involved in each complex. In NMDAR, the α-amino nitrogen of cycloserine hydrogen-bonds with an aspartate, a threonine, and a main chain carbonyl oxygen, while in the complex with alanine racemase the same nitrogen is uncomplemented, leaving it free to react with the pyridoxal phosphate. The ring nitrogen of the drug hydrogen-bonds with an arginine in NMDAR but with a backbone nitrogen in alanine racemase, and

the polarized ring oxygen hydrogen-bonds to a backbone nitrogen in NMDAR but accepts a hydrogen bond from a tyrosine in the racemase. The energy profiles reflect these interactions well. While the van der Waals energies are very similar in the two complexes (VEMD 0.4), as expected from a small ligand fully involved with the protein in both pairs, the electrostatics energies are distinctly different (EEMSD 300.4, Figure 6A and Table 2-SI).

The anti-inflammatory drug celecoxib (CEL) finds different environments in its complexes with the α-orthogonal bundle COX-2[33] (PDB 3ln1) and the α/β roll carbonic anhydrase[34] (PDB 1oq5; Figure 6B). The drug binds COX-2 with a K_i of 4 nM.[33] Subsequently, it was predicted and shown to bind to carbonic anhydrase with a K_i of 21 nM.[34] While the same pharmacophore is recognized in each site, the residues differ.[34,35] Most obviously, the drug's sulfonamide hydrogen-bonds with an arginine, a serine, and a backbone carbonyl in COX-2, while in carbonic anhydrase it binds as an anion to the catalytic zinc. The same ligand group is involved in both complexes, which is reflected in the van der Waals energy profile. Except for two atoms, including the nitrogen, which is in close contact with the zinc ion, explaining a much higher van der Waals energy, the van der Waals energies are similar (VEMD 0.6, Figure 6B and Table 2-SI). The electrostatics energy profiles differ (EEMSD 169.1, Figure 6B and Table 2-SI), especially in the sulfonamide group, which makes key interactions in both complexes.

The complexes of several flavonoid-like plant natural products also fall into class B (among them, emodin (EMO), flavopiridol (CPB), and naringerin (NAR), Figure 6C and Table 2-SI). In the α/β three-layer (aba) sandwich glycogen phophorylase[36] (PDB 3ebp), the flavone ring of flavopiridol (CPB) π-stacks in a sandwich with a tyrosine and a phenylalanine (Figure 6C). There are otherwise no direct polar interactions, with the cationic nitrogen and the ligand oxygens exposed to solvent. Conversely, the same ring is sandwiched by nonpolar side chains in the α/β two-layer sandwich CDK9[37] (PDB 3blr), where the ligand cation ion-pairs with an aspartate while one of the flavone carbonyls interacts with a backbone nitrogen. We do note that both the electrostatics and van der Waals energies are similar (EEMSD 25.5 and VDW 0.4, see Table 1-SI), as expected for a large class B ligand making few electrostatics interactions.

Finally, in the case of the cofactor biopterin (BIO), both the α/β two-layer sandwich tetrahydropterine synthase[38] (PDB 1b66) and the α/β three-layer sandwich sepiapterin synthase[39] (PBD 1sep) use a glutamate and an aspartate, respectively, to recognize the guanidine headgroup of the ligand, but the gemdiol side chain is recognized by a zinc ion in the former and by hydrogen bonds from a Ser/Tyr pair in the latter (Figure 6D). In this pair of complexes, most ligand atoms are engaged in differing electrostatics interactions, as indicated by dissimilar electrostatics energy profiles (EEMSD 120.0, see Figure 6D and Table 2-SI).

Class C Pairs of Complexes

For 14 pairs of complexes, not only did the protein environments differ but so did the very moieties on the ligand that were recognized; these were categorized as class C pairs (Figure 7 and Table 3-SI). Most follow the pattern set by allantoin binding to the α-bundle 2-oxo-4-

hydroxy-4-carboxy-5-ureidoimidazoline (OHCU) decarboxylase[20] (PDB 2o73) and to the α/β-fold urate oxidase[21] (PDB 2fxl; Figure 3). In the first of these complexes, allantoin's imidazole-dione ring is recognized by the protein, while in the second it is exposed to solvent.

In the complex between acetazolamide (AZM) and the TIM barrel chitiniase CTS1[40] (PDB 2uy4), the drug interacts via its acetamide moiety while the thiadiazole ring stacks to a tryptophan, and the sulfonamide is solvent-exposed (Figure 7A). When bound to its primary target, the α/β roll carbonic anhydrase[41] (PDB 3hs4), the sulfonamide is buried and interacts both with a threonine and a zinc ion; the thiadazole ring hydrogen-bonds to a threonine, and the rest of the molecule is exposed to solvent. Correspondingly, the electrostatics energy profile suggests that the ligand atoms experience wholly different physical environments (EEMSD 1877.8, Figure 7A and Table 3-SI).

Similarly, the charged amine in the antiseptic surfactant cetrimonium (16A) is buried and ion-pairs with two glutamates and an aspartate, and the hydrophobic chain is partly solvent-exposed in its complex with the β-sandwich laminarinase[42] (PDB 3b00), whereas in the complex with the immunoglobulin C1-set domain of MHC class I protein YF1[43] (PDB 3p73), the hydrophobic chain is deeply buried (Figure 7B). As might be expected, the major difference in the electrostatics energy occurs in the amine region of the ligand, the only part that is polar.

In the case of the riboflavin catabolite lumichrome (LUM), the ligand is partly exposed to solvent in both complexes, but the recognition motifs vary substantially (Figure 7C). In the complex with the α/β three-layer (aba) sandwich FAD synthetase[44] (PDB 1s4m), recognition is dominated by four hydrogen bonds down one side of the flavin. In the complex with the β-sandwich biliverdin reductase[45] (PDB 1hes), recognition is dominated by stacking to an NADP cofactor, with only one poor geometry hydrogen bond, again to the cofactor. Although the ligand adopts the same orientation in both binding sites (the same atoms of the ligands are buried or solvent-exposed in both sites) and the van der Waals energies are similar, it is put into class C because of the difference in the recognition motifs, reflected in the electrostatics energy profiles (Figure 7C).

Finally, the anti-inflammatory flufenamic acid (FLF) binds to the α/β barrel prostaglandin D2 11-ketoreductase AKR1C3[46] (PDB 1s2c) via networks of hydrogen bonds (histidine, tyrosine, and NADP cofactor), whereas in its complex with the α-orthogonal bundle androgen receptor[47] (PDB 2pix), recognition is exclusively driven by nonpolar contacts (Figure 7D). Here, the major difference in the electrostatics energy occurs in the carboxylate and at the linker amine regions, while the van der Waals energies are more similar (Figure 7D).

We note that despite the only partial complementarity of these class C complexes, affinity can be substantial, with acetazolamide binding to endochitinase in the low micromolar range[18] and to carbonic anhydrase in the low nanomolar range,[18] and flufenamic acid binding to AKR1C3 in the low micromolar range[18] and to the androgen receptor in the midmicromolar range.[47] More examples of class C complexes are available in Table 3-SI.

Effects of Energy Minimization on Complex Classification

Up until now, we have compared pairs of complexes visually, and by molecular energy potentials, essentially as they were determined experimentally. However, some crystal structures retain unfavorable ligand–protein contacts, often owing to force fields under-parametrized for ligand atoms. Accordingly, for eight pairs of complexes (16 structures), we parametrized the ligand using the GAFF procedures (http://t1.chem.umn.edu/amsol/, OEChem version 1.7.4 and reduce program[48]) and relaxed the ligand–protein complex in the AMBER force field.[49–51] The resulting complexes were reinspected, and the analysis of the molecular potential energies was repeated (Supporting Information Figure 4-SI). As expected, there was only modest visual change in the complexes after minimization. While this relaxation did lead to smoother van der Waals energies—eliminating most peaks owing to occasional close contacts—the electrostatics energies were less affected, and overall no substantial changes in patterns were seen. Thus, minimization leads to smoother energy profiles but does not change how we would classify pairs of complexes into class A, B, and C complexes (Figure 4).

A Benchmarking Set for Testing Binding Site Comparison Methods

Several methods have been recently introduced to compare binding sites.[34,52–55] One of these, or newly developed ones, may be able to recognize environmental similarities between pairs of targets that we see now as very different. To enable such comparison, we have constructed and made publicly available an open access benchmarking set of the pairs of fold-unrelated complexes described here (Directory of Unrelated Complexes, DUC, http://duc.docking.org). Each pair is organized by the single ligand they bind, with the PBB ID for each complex, the structure modified for ready energy calculation, and a 2D image of the key interactions in the binding sites. Each pair may be downloaded for comparison, as can the entire set.

Modeling How Many Different Environments May Exist to Recognize Similar Ligands

Clearly, more than one pattern of residues and of receptor environments can recognize even identical ligands; this is the case for more than two-thirds of the 62 pairs of unrelated proteins investigated here. To quantitatively model the number of possible environments per ligand, based on what we observe in this initial, and admittedly small, benchmark, we assumed that the number of possible environments follows a Poisson distribution with an unknown parameter λ that controls the mean and variance of the distribution (Figure 8A). We further assume that the ligand environments in different receptors are uncorrelated with one another. Then, if a ligand has k possible environments, there will be a 1 in k chance that two unrelated proteins will select the same environment (naturally, none of this holds for sequence-related proteins). We fit the model by counting the number of pairs in our set of 62 pairs that were similar or dissimilar, seeking the most likely value of λ (Figure 8B). Since 19 of the 62 pairs are class A, the most likely value of λ is about 4, which corresponds to each ligand, and each family of SAR-related ligands, having between two and five possible recognition environments, each essentially unrelated, in proteins unrelated by fold.

DISCUSSION

Returning to our motivating questions, a key observation is that there is no obvious conserved pattern for the recognition of identical ligands by pockets in unrelated folds. The same ligand might be recognized by different residues, with different interaction types, and even different ligand chemotypes may be engaged. Though there were pairs of complexes where the ligand encountered similar environments (class A), these were not the majority. Even among class A complexes, similarity was only approximate and environmental. Cases where similarity could be mapped residue-to-residue, even allowing for conservative substitutions, were rare. These observations suggest that one cannot reliably infer the pattern of receptor residues with which any given ligand will interact, nor from a complex with that ligand can one infer the characteristics of the same ligand in complex with an unrelated protein. At the functional group and residue level, between binding sites on proteins unrelated by fold, there is no single code for ligand recognition.[56]

A reason why there is no simple code for ligand recognition among binding sites is that proteins have found multiple, at least superficially unrelated ways to recognize most common ligand groups. Thus, cationic amines can be recognized both by anionic residues such as aspartate or glutamate, but they can also be recognized by cation–π interactions. Nucleotide phosphates can be recognized by cationic residues such as arginines, but recognition by main chain amide nitrogens in a P-loop is also common. Ligand aromatic groups can stack with tyrosines, phenylalanines and tryptophans, but they can also form cation–π interactions (Figure 1); many other variations might be mentioned. Thus, for proteins unconstrained by a common evolutionary origin, there is no strong reason to expect that when two dissimilar folds bind a common ligand; they will do so using similar interactions. This observation is not restricted to protein–ligand complexes, as it has long been known that between repressors and operators, multiple recognition patterns are observed, and unrelated repressors can bind identical operators. Here too there is “no code” for recognition.[56]

Naturally, we do not imply that the number of possible receptor environments for any family of ligands is unbounded. Working backward from the observation that only 19 of the 62 pairs belong to class A, a simple model of the distribution of environments suggests that most ligand families can be recognized by between two and five different receptor environments. Given the small size of this benchmarking set, we do not make too much of these values, other than to say that we can reject the possibility that only a single receptor environment is plausible for most ligands, and that we expect that the number of plausible dissimilar environments for a ligand is substantially smaller than what full combinatorial elaboration of environments for each chemotype would suggest.

This study does not imply that predicting off-target binding across fold families is impossible. Any approach that can calculate absolute binding affinities would recognize opportunities for off-target binding, irrespective of fold. More generally, true biophysical—as opposed to pattern matching—approaches might also do so. After all, the discovery that celecoxib not only binds to COX-2 but also to carbonic anhydrase arose from an examination of pocket structures,[34] and recently several methods have been introduced to

compare[52–55,57,58] and exploit[59,60] "off-target" binding sites, based on their structures. These and related methods may find relationships among what, by pattern matching and by mapping of molecular potentials, appear to be unrelated ligand binding sites. Thus, we have organized the 116 complexes and 62 pairs of proteins into a readily accessible set, where any given pair of fold-unrelated proteins for any of the 59 ligands, or all of them together, may be rapidly accessed and compared by investigators interested in testing new methods (http://duc.docking.org).

Several other caveats merit mentioning. Distinguishing between a class A, B, or C retains a subjective aspect, and we sometimes disagreed with the calculated van der Waals and electrostatic complementarities, with a few class A pairs having noticeably different van der Waals energies (high VEMD score) and a few class B and C pairs having similar electrostatics energies (low EEMSD score; Figure 4 and Tables S1–S3). Sometimes these discrepancies are readily explained. For example, berberine (BER) makes few electrostatics interactions with either phospholipase A2[61] (PDB 2qvd) or BmrR[62] (PDB 3d6y), explaining its low EEMSD score (Figure S3 and Table 3-SI). Still, the ligand is exposed to substantially different environments: its interaction with phospholipase A2 is dominated by its dioxolane ring while the rest of the ligand is solvent-exposed, whereas in BmrR it hydrogen-bonds to an ordered water molecule via one of the two methoxy groups, thus falling in the class C category. Also, with only 59 ligands in 116 complexes, this study does not pretend to be comprehensive, though we hope it is large and diverse enough to be representative.

These caveats should not obscure the principal observation from this study, that when identical ligands bind to unrelated targets, the binding site similarity is rarely more than approximate, and most sites differ substantially. This reflects the multiple ways a protein can recognize most ligand functional groups (Figure 1). Since ligands use multiple groups to bind to a protein, and since each group can be recognized in several unrelated ways, there is little expectation that evolutionarily unrelated proteins will use the same residue patterns to recognize similar or even identical ligands. This is important because of the profound polypharmacology of small molecule drugs and reagents, whose biological effects can rarely be understood through binding to only a single target. When off-targets are related by sequence or structure to the primary target, as is often the case, such off-target activity may be optimized against (or for), or at least accommodated.[60] However, ligands frequently bind to off-targets that are unrelated by sequence or structure, and these present deeper challenges; there is no simple pattern-matching code for protein–ligand recognition.

EXPERIMENTAL SECTION

Ligand-Protein Complex Identification

A mapping of all PDB structures organized by the ligands is available from RCSB (http://ligand-expo.rcsb.org/dictionaries/cc-to-pdb.tdd). Scripts identified those ligands that occurred in at least two or more complexes and discarded those that shared the same protein name or uniprot ID. Most ligands that had 15 or more targets were discarded, as these were typically simple salts, precipitants, or common residue modifications. The ligands were further filtered by limiting their molecular weight to 500 Da and removing most ligands with fewer than 10 non-hydrogen atoms or that were cofactors such as ATP, ADP, NAD, and

others studied by Kahraman and co-workers.[12,13] The complexes were further reduced by selecting pairs of proteins with at most 10% sequence identity, that had different PFAM, CATH, and SCOP domain descriptions,[63–65] and that were of different sizes. Finally, each pair of complexes was visually inspected. Ultimately, 59 ligands in 116 complexes were selected (62 pairs in total, some ligands bind to more than one pair of unrelated proteins). We do not pretend to have comprehensively described all pairs of unrelated proteins that bind identical ligands, nor unrelated proteins that bind similar ligands; we suspect that more such pairs may be found among determined structures.

Classification of Ligand–Protein Complexes

The 62 pairs of complexes were classified by inspection at atomic resolution. Categorization into class C, where wholly different ligand groups are engaged with each protein, was the simplest, as typically for one complex in the pair a key ligand group would be exposed to solvent, whereas in the second complex the same warhead would directly interact with the protein. Classes A and B demanded more judgment. When most ligand atoms encountered similar interaction types, irrespective of the side chains that contributed them, they were assigned to class A. Class B complexes were those where similar ligand groups were engaged in both complexes but the residues that they encountered differed substantially, at least by pattern-matching. In every case, visual classification was compared to the per atom van der Waals (AMBER potential function) and electrostatic energies of the crystallographic pose of the ligand, calculated with the scoreopt utility[66] from DOCK 3.6,[67] using the Poisson–Boltzmann method QNIFFT to calculate electrostatic potential maps.[68,69] Cofactors and ions were included in the DOCK receptor preparation when they were involved in ligand binding, as indicated by Ligplot[70] visualization of the site. In a few cases, manual manipulation was required to correct the positioning of hydrogens added during the preparation. Per atom scores for both electrostatics and van der Waals energies were plotted for each pair. The average difference for both electrostatic and van der Waals energies was calculated between the ligand atoms in each of the two structures. As only binding sites containing the same ligands are considered, atoms can be matched by name and directly compared (the similar but nonidentical ligands were excluded from this analysis). For electrostatics scores, the mean squared difference is used to highlight major differences over a large number of small fluctuations (Electrostatics Energy Mean Square Deviation, or EEMSD). Due to a smaller scale, the mean difference is sufficient to compare van der Waals scores (van der Waals Energy Mean Deviation, or VEMD).

$$EEMSD(\vec{x},\vec{y}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2$$

where *x* and *y* are the per-ligand-atom electrostatics energies for the ligand in each pair

$$VEMD(\vec{x},\vec{y}) = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i|$$

where *x* and *y* are the per-ligand-atom van der Waals energies for the ligand in each pair

Atoms with high magnitude energies (electrostatics energies over 30 kcal/mol and van der Waals energies over 5 kcal/mol) were excluded. Additionally, unusually low electrostatics energies were typically clipped at −50 kcal/mol to prevent them from dominating the average difference, unless it caused two scores to appear unnaturally similar. Since the overall energy in two binding sites may not be equivalent, yet contributions of individual atoms may be similar, scores were normalized after filtering and clipping by subtracting off the mean energy score of the ligand atoms in each structure. Per atom scores for both electrostatics and van der Waals energies were plotted for each pair and overall scores recorded. When constructing per atom plots, atoms in both structures were sorted by ascending energetic scores of the first structure to make visual inspection easier.

The 3D images of the complexes were rendered with PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4; Schrödinger, LLC) and 2D images with Ligplot.[70]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank K. Sharp for a gift of QNIFFT. We thank G. Rocklin, H. Lin, J. Irwin, and T. Balius for reading this manuscript and A. Edwards, C. Arrowsmith (Univ. of Toronto), and B. Roth (UNC-Chapel Hill) for hosting sabbatical stays, during which this work was begun. Supported by GM71896 (to J. Irwin & B. Shoichet).

ABBREVIATIONS

AUC	Area Under the Curve
β2-AR	β2-adrenergic receptor
CAIR	carboxyaminoimidazole ribonucleotide
EEMSD	Electrostatics Energy Mean Square Deviation
GPCR	G-Protein Coupled Receptors
NHR	Nuclear Hormone Receptor
NMDAR	NMDA receptor
OHCU	2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline decarboxylase
VEMD	van der Waals Energy Mean Deviation

REFERENCES

1. Ehrlich P. Address to the 17th International Medical Congress. 1913
2. Olah, M.; Mracec, M.; Ostopovici, L.; Rad, R.; Bora, A.; Hadaruga, N.; Olah, I.; Banda, M.; Simon, Z.; Mracec, M.; Oprea, TI. WOMBAT: World of Molecular Bioactivity. In: Oprea, TI., editor. Chemoinformatics in Drug Discovery. Wiley-VCH; New York: 2004. p. 223-239.
3. Warr W. ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBLEBI). Interview by Wendy A. Warr. J. Comput.-Aided Mol. Des. 2009; 23:195–198.

4. Roth BL, Sheffler DJ, Kroeze WK. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nat. Rev. Drug Discovery. 2004; 3:353–359.
5. Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL. Global mapping of pharmacological space. Nat. Biotechnol. 2006; 24:805–815. [PubMed: 16841068]
6. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side-effect similarity. Science. 2008; 321:263–266. [PubMed: 18621671]
7. Lounkine E, Keiser MJ, Whitebread S, Mikhailov D, Hamon J, Jenkins JL, Lavan P, Weber E, Doak AK, Cote S, Shoichet BK, Urban L. Large-scale prediction and testing of drug activity on side-effect targets. Nature. 2012; 486:361–367. [PubMed: 22722194]
8. Lin H, Sassano MF, Roth BL, Shoichet BK. A pharmacological organization of G protein-coupled receptors. Nat. Methods. 2013; 10:140–146. [PubMed: 23291723]
9. Wright CS, Alden RA, Kraut J. Structure of subtilisin BPN' at 2.5 angstrom resolution. Nature. 1969; 221:235–242. [PubMed: 5763076]
10. Dougherty DA. Cation-pi interactions in chemistry and biology: a new view of benzene. Phe, Tyr, and Trp. Science. 1996; 271:163–168.
11. Hirsch AK, Fischer FR, Diederich F. Phosphate recognition in structural biology. Angew. Chem., Int. Ed. 2007; 46:338–352.
12. Kahraman A, Morris RJ, Laskowski RA, Thornton JM. Shape variation in protein binding pockets and their ligands. J. Mol. Biol. 2007; 368:283–301. [PubMed: 17337005]
13. Kahraman A, Morris RJ, Laskowski RA, Favia AD, Thornton JM. On the diversity of physicochemical environments experienced by identical ligands in binding pockets of unrelated proteins. Proteins: Struct., Funct., Genet. 2010; 78:1120–1136. [PubMed: 19927322]
14. Hansen SB, Taylor P. Galanthamine and non-competitive inhibitor binding to ACh-binding protein: evidence for a binding site on non-alpha-subunit interfaces of heteromeric neuronal nicotinic receptors. J. Mol. Biol. 2007; 369:895–901. [PubMed: 17481657]
15. Greenblatt HM, Kryger G, Lewis T, Silman I, Sussman JL. Structure of acetylcholinesterase complexed with (–)-galanthamine at 2.3 A resolution. FEBS Lett. 1999; 463:321–326. [PubMed: 10606746]
16. Hoskins AA, Morar M, Kappock TJ, Mathews II, Zaugg JB, Barder TE, Peng P, Okamoto A, Ealick SE, Stubbe J. N5-CAIR mutase: role of a CO2 binding site and substrate movement in catalysis. Biochemistry. 2007; 46:2842–2855. [PubMed: 17298082]
17. Ginder ND, Binkowski DJ, Fromm HJ, Honzatko RB. Nucleotide complexes of Escherichia coli phosphoribosylaminoimidazole succinocarboxamide synthetase. J. Biol. Chem. 2006; 281:20680–20688. [PubMed: 16687397]
18. Benson ML, Smith RD, Khazanov NA, Dimcheff B, Beaver J, Dresslar P, Nerothin J, Carlson HA. Binding MOAD, a high-quality protein-ligand database. Nucleic Acids Res. 2007; 36:D674–D678. [PubMed: 18055497]
19. Nelson SW, Binkowski DJ, Honzatko RB, Fromm HJ. Mechanism of action of Escherichia coli phosphoribosyla minoimidazolesuccinocarboxamide synthetase. Biochemistry. 2005; 44:766–774. [PubMed: 15641804]
20. Cendron L, Berni R, Folli C, Ramazzina I, Percudani R, Zanotti G. The structure of 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline decarboxylase provides insights into the mechanism of uric acid degradation. J. Biol. Chem. 2007; 282:18182–18189. [PubMed: 17428786]
21. Gabison L, Chiadmi M, Colloc'h N, Castro B, El Hajji M, Prange T. Recapture of [S]-allantoin, the product of the two-step degradation of uric acid, by urate oxidase. FEBS Lett. 2006; 580:2087–2091. [PubMed: 16545381]
22. Stegemann B, Klebe G. Cofactor-binding sites in proteins of deviating sequence: comparative analysis and clustering in torsion angle, cavity, and fold space. Proteins: Struct., Funct., Genet. 2012; 80:626–648. [PubMed: 22095739]
23. Kochan G, Pilka ES, von Delft F, Oppermann U, Yue WW. Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A. J. Mol. Biol. 2009; 388:997–1008. [PubMed: 19345228]

24. Gonzalez JM, Fisher SZ. Structural analysis of ibuprofen binding to human adipocyte fatty-acid binding protein (FABP4). Acta Crystallogr., Sect. F: Struct. Biol. Commun. 2015; 71:163–170. [PubMed: 25664790]
25. Angelucci F, Johnson KA, Baiocco P, Miele AE, Brunori M, Valle C, Vigorosi F, Troiani AR, Liberti P, Cioli D, Klinkert MQ, Bellelli A. Schistosoma mansoni fatty acid binding protein: specificity and functional control as revealed by crystallo-graphic structure. Biochemistry. 2004; 43:13000–13011. [PubMed: 15476393]
26. Malkowski MG, Ginell SL, Smith WL, Garavito RM. The productive conformation of arachidonic acid bound to prostaglandin synthase. Science. 2000; 289:1933–1937. [PubMed: 10988074]
27. Volkov A, Khoshnevis S, Neumann P, Herrfurth C, Wohlwend D, Ficner R, Feussner I. Crystal structure analysis of a fatty acid double-bond hydratase from Lactobacillus acidophilus. Acta Crystallogr., Sect. D: Biol. Crystallogr. 2013; 69:648–657. [PubMed: 23519674]
28. Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Yao XJ, Weis WI, Stevens RC, Kobilka BK. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. Science. 2007; 318:1266–1273. [PubMed: 17962519]
29. Singh SK, Yamashita A, Gouaux E. Antidepressant binding site in a bacterial homologue of neuro-transmitter transporters. Nature. 2007; 448:952–956. [PubMed: 17687333]
30. Fenn TD, Holyoak T, Stamper GF, Ringe D. Effect of a Y265F mutant on the transamination-based cycloserine inactivation of alanine racemase. Biochemistry. 2005; 44:5317–5327. [PubMed: 15807525]
31. Furukawa H, Gouaux E. Mechanisms of activation, inhibition and specificity: crystal structures of the NMDA receptor NR1 ligand-binding core. EMBO J. 2003; 22:2873–2885. [PubMed: 12805203]
32. Leeson PD, Iversen LL. The glycine site on the NMDA receptor: structure-activity relationships and therapeutic potential. J. Med. Chem. 1994; 37:4053–4067. [PubMed: 7990104]
33. Portevin B, Tordjman C, Pastoureau P, Bonnet J, De Nanteuil G. 1,3-Diaryl-4,5,6,7-tetrahydro-2H-isoindole derivatives: a new series of potent and selective COX-2 inhibitors in which a sulfonyl group is not a structural requisite. J. Med. Chem. 2000; 43:4582–4593. [PubMed: 11101350]
34. Weber A, Casini A, Heine A, Kuhn D, Supuran CT, Scozzafava A, Klebe G. Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. J. Med. Chem. 2004; 47:550–557. [PubMed: 14736236]
35. Wang JL, Limburg D, Graneto MJ, Springer J, Hamper JR, Liao S, Pawlitz JL, Kurumbail RG, Maziasz T, Talley JJ, Kiefer JR, Carter J. The novel benzopyran class of selective cyclooxygenase-2 inhibitors. Part 2: the second clinical candidate having a shorter and favorable human half-life. Bioorg. Med. Chem. Lett. 2010; 20:7159–7163. [PubMed: 20709553]
36. Tsitsanou KE, Hayes JM, Keramioti M, Mamais M, Oikonomakos NG, Kato A, Leonidas DD, Zographos SE. Sourcing the affinity of flavonoids for the glycogen phosphorylase inhibitor site via crystallography, kinetics and QM/ MM-PBSA binding studies: comparison of chrysin and flavopiridol. Food Chem. Toxicol. 2013; 61:14–27. [PubMed: 23279842]
37. Baumli S, Lolli G, Lowe ED, Troiani S, Rusconi L, Bullock AN, Debreczeni JE, Knapp S, Johnson LN. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. EMBO J. 2008; 27:1907–1918. [PubMed: 18566585]
38. Ploom T, Thony B, Yim J, Lee S, Nar H, Leimbacher W, Richardson J, Huber R, Auerbach G. Crystallographic and kinetic investigations on the mechanism of 6-pyruvoyl tetrahydropterin synthase. J. Mol. Biol. 1999; 286:851–860. [PubMed: 10024455]
39. Auerbach G, Herrmann A, Gutlich M, Fischer M, Jacob U, Bacher A, Huber R. The 1.25 A crystal structure of sepiapterin reductase reveals its binding mode to pterins and brain neurotransmitters. EMBO J. 1997; 16:7219–7230. [PubMed: 9405351]
40. Hurtado-Guerrero R, van Aalten DM. Structure of Saccharomyces cerevisiae Chitinase 1 and screening-based discovery of potent inhibitors. Chem. Biol. 2007; 14:589–599. [PubMed: 17524989]

41. Sippel KH, Robbins AH, Domsic J, Genis C, Agbandje-McKenna M, McKenna R. High-resolution structure of human carbonic anhydrase II complexed with acetazolamide reveals insights into inhibitor drug design. Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun. 2009; 65:992–995.
42. Jeng WY, Wang NC, Lin CT, Shyur LF, Wang AH. Crystal structures of the laminarinase catalytic domain from Thermotoga maritima MSB8 in complex with inhibitors: essential residues for beta-1,3- and beta-1,4-glucan selection. J. Biol. Chem. 2011; 286:45030–45040. [PubMed: 22065588]
43. Hee CS, Gao S, Loll B, Miller MM, Uchanska-Ziegler B, Daumke O, Ziegler A. Structure of a classical MHC class I molecule that binds “non-classical” ligands. PLoS Biol. 2010; 8:e1000557. [PubMed: 21151886]
44. Wang W, Kim R, Yokota H, Kim SH. Crystal structure of flavin binding to FAD synthetase of Thermotoga maritima. Proteins: Struct., Funct., Genet. 2005; 58:246–248. [PubMed: 15468322]
45. Pereira PJ, Macedo-Ribeiro S, Parraga A, Perez-Luque R, Cunningham O, Darcy K, Mantle TJ, Coll M. Structure of human biliverdin IXbeta reductase, an early fetal bilirubin IXbeta producing enzyme. Nat. Struct. Biol. 2001; 8:215–220. [PubMed: 11224564]
46. Lovering AL, Ride JP, Bunce CM, Desmond JC, Cummings SM, White SA. Crystal structures of prostaglandin D(2) 11-ketoreductase (AKR1C3) in complex with the nonsteroidal anti-inflammatory drugs flufenamic acid and indomethacin. Cancer Res. 2004; 64:1802–1810. [PubMed: 14996743]
47. Estebanez-Perpina E, Arnold LA, Nguyen P, Rodrigues ED, Mar E, Bateman R, Pallai P, Shokat KM, Baxter JD, Guy RK, Webb P, Fletterick RJ. A surface on the androgen receptor that allosterically regulates coactivator binding. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:16074–16079. [PubMed: 17911242]
48. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J. Mol. Biol. 1999; 285:1735–1747. [PubMed: 9917408]
49. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009; 25:1422–1423. [PubMed: 19304878]
50. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. 2004; 25:1605–1612. [PubMed: 15264254]
51. Case, DA.; Berryman, JT.; Betz, RM.; Cerutti, DS.; Cheatham, TE.; Darden, TA.; Duke, RE.; Giese, TJ.; Gohlke, H.; Goetz, AW.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, TS.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, KM.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, HT.; Omelyan, I.; Onufriev, A.; Roe, DR.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, CL.; Smith, W.; Swails, J.; Walker, RC.; Wang, J.; Wolf, RM.; Wu, X.; York, DM.; Kollman, PA. AMBER 2015. University of California; San Francisco: 2015.
52. Schmitt S, Kuhn D, Klebe G. A new method to detect related function among proteins independent of sequence and fold homology. J. Mol. Biol. 2002; 323:387–406. [PubMed: 12381328]
53. Liu T, Altman RB. Using multiple microenvironments to find similar ligand-binding sites: application to kinase inhibitor binding. PLoS Comput. Biol. 2011; 7:e1002326. [PubMed: 22219723]
54. Konc J, Janezic D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. Bioinformatics. 2010; 26:1160–1168. [PubMed: 20305268]
55. Ito J, Tabei Y, Shimizu K, Tsuda K, Tomii K. PoSSuM: a database of similar protein-ligand binding and putative pockets. Nucleic Acids Res. 2012; 40:D541–548. [PubMed: 22135290]
56. Matthews BW. Protein-DNA interaction. No code for recognition. Nature. 1988; 335:294–295. [PubMed: 3419498]
57. Krotzky T, Rickmeyer T, Fober T, Klebe G. Extraction of protein binding pockets in close neighborhood of bound ligands makes comparisons simple due to inherent shape similarity. J. Chem. Inf. Model. 2014; 54:3229–3237. [PubMed: 25345905]

58. Morris RJ, Najmanovich RJ, Kahraman A, Thornton JM. Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. Bioinformatics. 2005; 21:2347–2355. [PubMed: 15728116]

59. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE. 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. [PubMed: 19578428]

60. Milletti F, Vulpetti A. Predicting polypharmacology by binding site similarity: from kinases to the protein universe. J. Chem. Inf. Model. 2010; 50:1418–1431. [PubMed: 20666497]

61. Chandra DN, Prasanth GK, Singh N, Kumar S, Jithesh O, Sadasivan C, Sharma S, Singh TP, Haridas M. Identification of a novel and potent inhibitor of phospholipase A(2) in a medicinal plant: crystal structure at 1.93A and Surface Plasmon Resonance analysis of phospholipase A(2) complexed with berberine. Biochim. Biophys. Acta, Proteins Proteomics. 2011; 1814:657–663.

62. Newberry KJ, Huffman JL, Miller MC, Vazquez-Laslop N, Neyfakh AA, Brennan RG. Structures of BmrR-drug complexes reveal a rigid multidrug binding pocket and transcription activation through tyrosine expulsion. J. Biol. Chem. 2008; 283:26795–26804. [PubMed: 18658145]

63. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. Pfam: the protein families database. Nucleic Acids Res. 2014; 42:D222–230. [PubMed: 24288371]

64. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees JG, Lehtinen S, Studer RA, Thornton J, Orengo CA. CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res. 2015; 43:D376–381. [PubMed: 25348408]

65. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 1995; 247:536–540. [PubMed: 7723011]

66. Lorber DM, Shoichet BK. Hierarchical docking of databases of multiple ligand conformations. Curr. Top. Med. Chem. 2005; 5:739–749. [PubMed: 16101414]

67. Mysinger MM, Shoichet BK. Rapid context-dependent ligand desolvation in molecular docking. J. Chem. Inf. Model. 2010; 50:1561–1573. [PubMed: 20735049]

68. Gallagher K, Sharp K. Electrostatic contributions to heat capacity changes of DNA-ligand binding. Biophys. J. 1998; 75:769–776. [PubMed: 9675178]

69. Sharp KA. Polyelectrolyte Electrostatics - Salt Dependence, Entropic, and Enthalpic Contributions to Free-Energy in the Nonlinear Poisson-Boltzmann Model. Biopolymers. 1995; 36:227–243.

70. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. 2011; 51:2778–2786. [PubMed: 21919503]

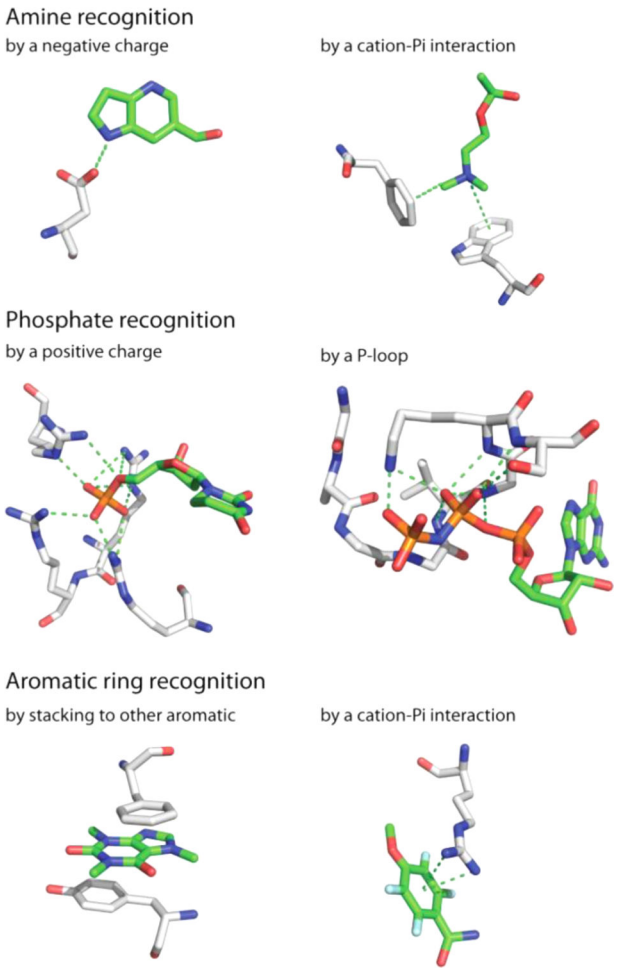


Figure 1.

Recogntion of ligand groups (carbons in green) by unrelated receptor residues (carbons in gray). A charged amine may be recognized by an aspartate (PDB 4jn0) or by aromatic rings via cation–π interactions (PDB 2ace). A phosphate may be recognized by a positive charge (PDB 2tsc) or by a P-loop (PDB 5p21). An aromatic ring may be recognized by stacking with other aromatic rings (PDB 3dds) or by a cation–π interaction (PDB 1kjr). The list is not exhaustive and only serves to illustrate the variety of ways that the same chemotype can interact with protein binding motifs.

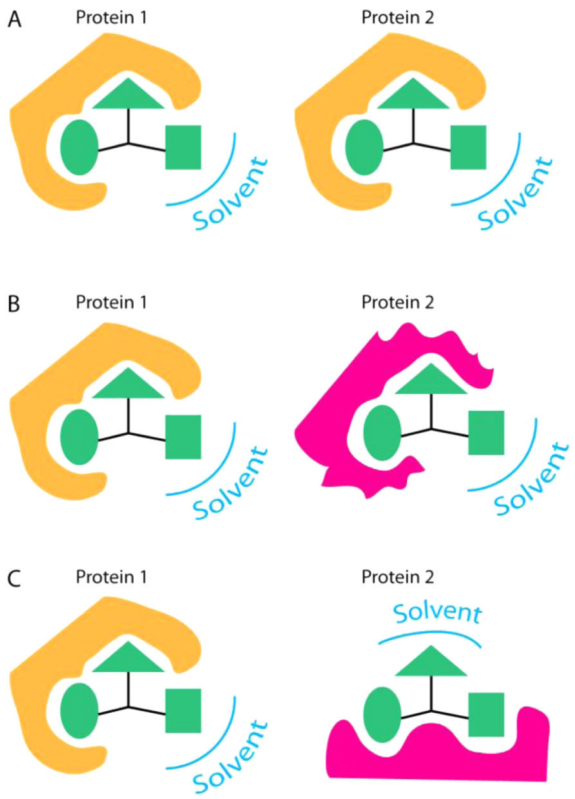


Figure 2.

Classes of recognition of an identical ligand by unrelated proteins. (A) The same ligand functional groups interact with similar protein groups in both sites. (B) The same ligand functional groups interact with different protein groups in each site. (C) different ligand functional groups interact with different (or similar) protein groups in each site.

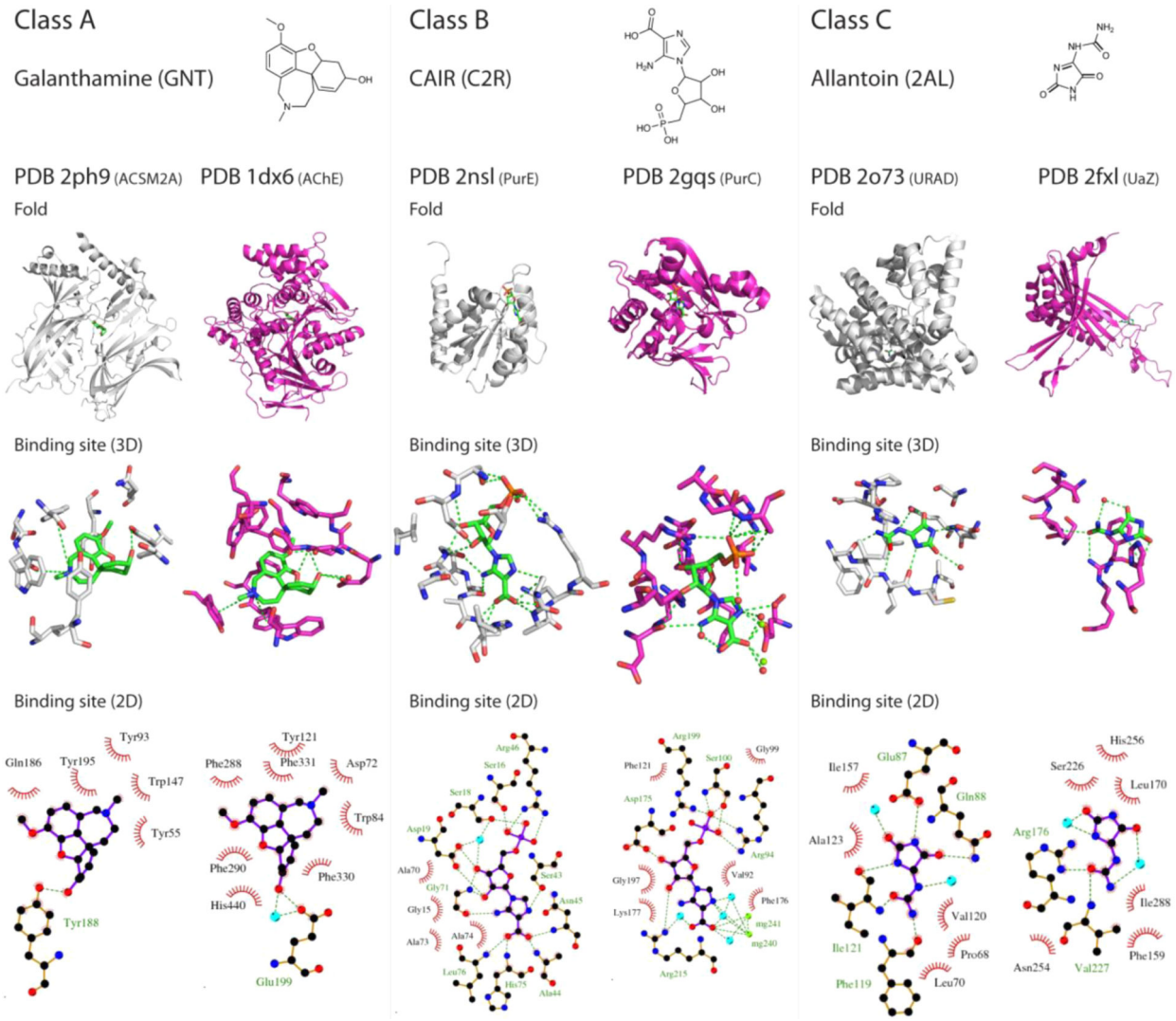


Figure 3.

Examples of class A, B, and C complexes between identical ligands and proteins unrelated by fold. The ligand, the fold, and 3D and 2D interactions are shown. Left: a class A pair, galanthamine (GNT) in complex with acetylcholine binding protein[14] (white, PDB 2ph9) and acetylcholine esterase[15] (pink, PDB 1dx6). Middle: a class B pair, carboxyaminoimidazole ribonucleotide (C2R) in complex with N5-CAIR mutase (PurE;[16] white, PDB 2nsl) and SAICAR synthetase (PurC;[17] pink, PDB 2gqs). Right: a class C pair, allantoin (2AL) in complex with OHCU decarboxylase[20] (white, PDB 2o73) and urate oxidase[21] (pink, PDB 2fxl).

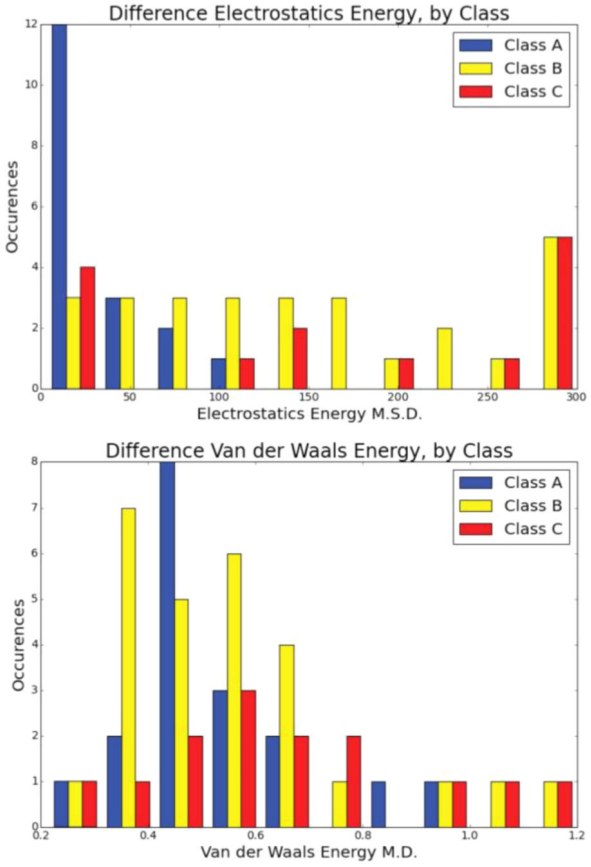


Figure 4.

Distribution of Electrostatics Energy Mean Square Deviation (EEMSD) and van der Waals Energy Mean Deviation (VEMD) for 59 pairs of complexes, colored by class.

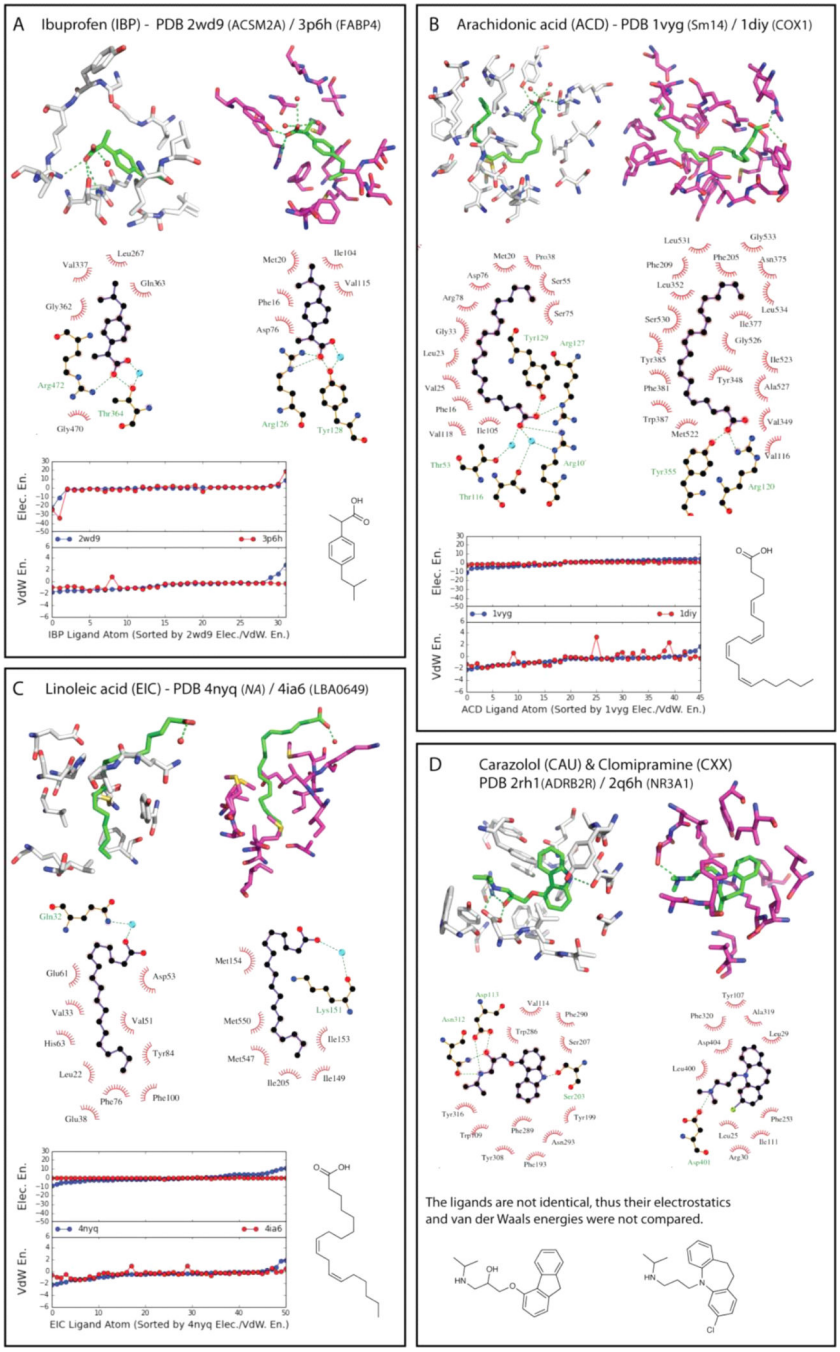


Figure 5.

Characteristic class A pairs of complexes (see also Table 1-SI). The ligand, 3D, and 2D (ligplot) interactions and the electrostatics and van der Waals energy profiles are shown. (A) Ibuprofen (IBP) in complex with acyl-CoA synthetase ACSM2A[23] (PDB 2wd9) and fatty-acid binding protein[24] (PDB 3p6h). (B) Arachidonic acid (ACD) in complex with fatty acid binding protein[25] (PDB 1vyg) and prostaglandin H synthase-1[26] (PDB 1DIY). (C) Linoleic acid (EIC) in complex with a lipid-binding lipocalin-like protein (PDB 4nyq) and

hydratase[27] (PDB 4ia6). (D) Carazolol (CAU) in complex with β2-AR[28] (PDB 2rh1) and clomipramine (CXX) in complex with the LeuT transporter[29] (PDB 2q6h).

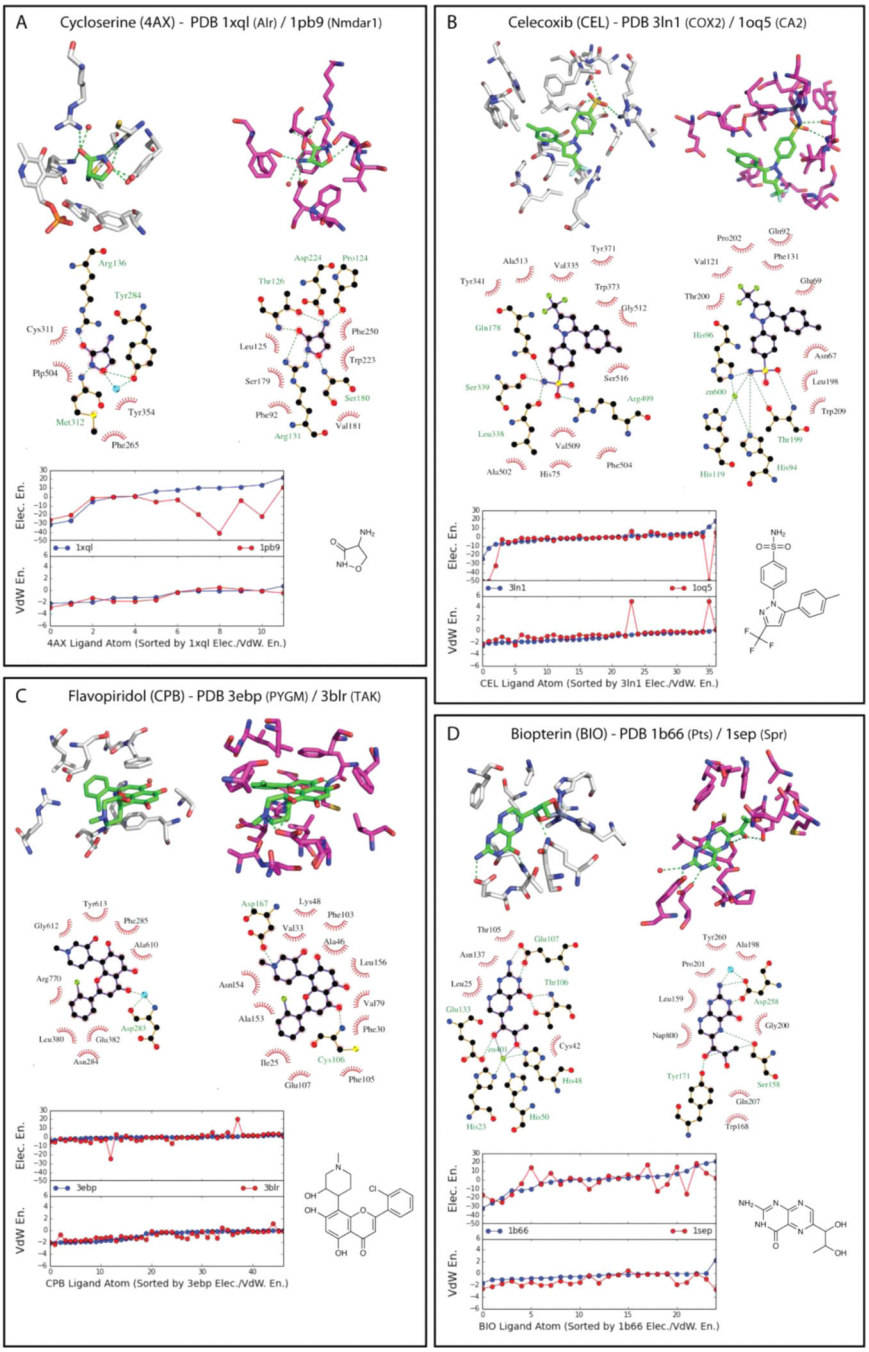


Figure 6.

Characteristic class B pairs of complexes (see also Table 2-SI). The ligand, 3D, and 2D (ligplot) interactions and the electrostatics and van der Waals energy profiles are shown. (A) Cycloserine (4AX) in complex with alanine racemase[30] (PDB 1xql) and the ligand binding domain of the NMDA receptor[31] (PDB 1pb9). (B) Celecoxib (CEL) in complex with COX-2[33] (PDB 3ln1) and carbonic anhydrase[34] (PDB 1oq5). (C) Flavopiridol (CPB) in complex with glycogen phophorylase[36] (PDB 3ebp) and CDK9[37] (PDB 3blr). (D) Biopterin

(BIO) in complex with tetrahydropterine synthase[38] (PDB 1b66) and sepiapterin synthase[39] (PBD 1sep).

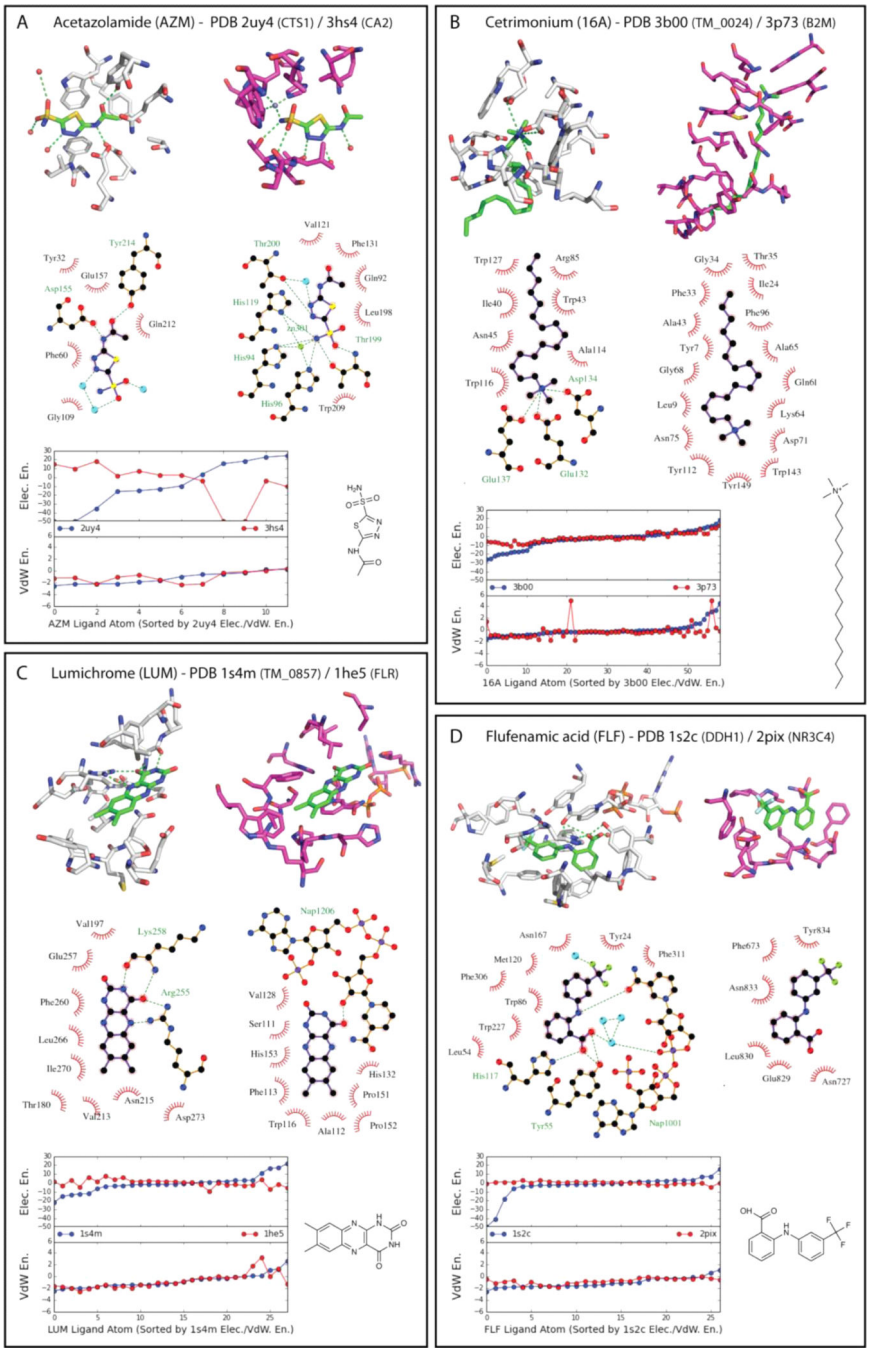


Figure 7.

Characteristic class C pairs of complexes (see also Table 3-SI). The ligand, 3D, and 2D (ligplot) interactions and the electrostatics and van der Waals energy profiles are shown. (A) Acetazolamide (AZM) in complex with chitiniase CTS1[40] (PDB 2uy4) and carbonic anhydrase[41] (PDB 3hs4). (B) Cetrimonium (16A) in complex with laminarinase[42] (PDB 3b00) and the MHC class I protein YF1 complex[43] (PDB 3p73). (C) Lumichrome (LUM) in complex with FAD synthase[44] (PDB 1s4m) and biliverdin reductase complex[45] (PDB 1hes).

(D) Flufenamic acid (FLF) in complex with prostaglandin D2 11-ketoreductase AKR1C3[46] (PDB 1s2c) and the androgen receptor[47] (PDB 2pix).

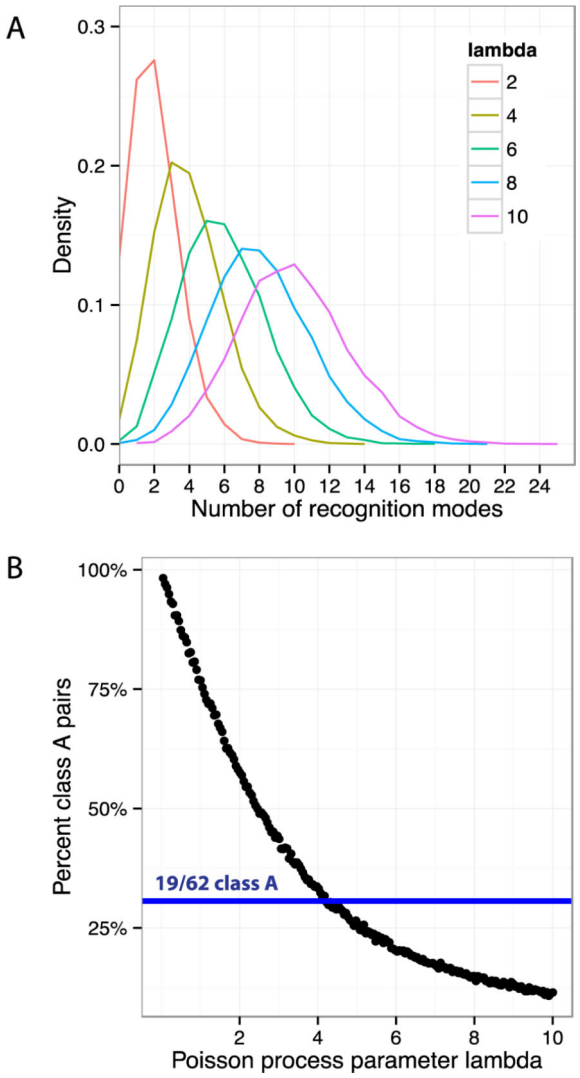


Figure 8.

A model for the number of possible ligand recognition sites among unrelated receptors. Assuming that the distribution of possible binding sites follows a Poisson distribution (A), the parameter λ can be fit by considering the percentage of observed pairs that are recognized by similar binding sites, and looking up the most likely λ (B). We observed 19 class A pairs out of 62 pairs (blue line), corresponding to a λ of about 4 and each ligand binding in sites representing two to five different environments.