

What Can We Learn from the Evolution of Protein-Ligand Interactions to Aid the Design of New Therapeutics?

Alicia P. Higueroel^{1*}, Adrian Schreyer¹, G. Richard J. Bickerton^{1‡}, Tom L. Blundell¹, Will R. Pitt^{1,2}

¹ Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom, ² UCB Pharma, Slough, United Kingdom

Abstract

Efforts to increase affinity in the design of new therapeutic molecules have tended to lead to greater lipophilicity, a factor that is generally agreed to be contributing to the low success rate of new drug candidates. Our aim is to provide a structural perspective to the study of lipophilic efficiency and to compare molecular interactions created over evolutionary time with those designed by humans. We show that natural complexes typically engage in more polar contacts than synthetic molecules bound to proteins. The synthetic molecules also have a higher proportion of unmatched heteroatoms at the interface than the natural sets. These observations suggest that there are lessons to be learnt from Nature, which could help us to improve the characteristics of man-made molecules. In particular, it is possible to increase the density of polar contacts without increasing lipophilicity and this is best achieved early in discovery while molecules remain relatively small.

Citation: Higueroel AP, Schreyer A, Bickerton GRJ, Blundell TL, Pitt WR (2012) What Can We Learn from the Evolution of Protein-Ligand Interactions to Aid the Design of New Therapeutics? PLoS ONE 7(12): e51742. doi:10.1371/journal.pone.0051742

Editor: Narayanaswamy Srinivasan, Indian Institute of Science, India

Received: August 29, 2012; **Accepted:** November 8, 2012; **Published:** December 11, 2012

Copyright: © 2012 Higueroel et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding was provided by the BBSRC, Wellcome Trust, UCB Pharma, and the University of Cambridge. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: WRP is an employee of UCB Pharma. However, the work was carried out while he was seconded to the University of Cambridge and he has no competing interests in the material in this manuscript. Therefore this does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials. The authors have declared that no other competing interests exist.

* E-mail: alicia@cryst.bioc.cam.ac.uk

‡ Current address: Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee, United Kingdom

Introduction

The decrease in productivity in drug discovery and development, as measured by the cost of bringing a new medical entity (NME) to the market, is a many fold problem [1]. One of the main reasons for this well-documented decline is the poor physical properties of drug candidates entering into clinical trials ([2–5] and the references therein). Current leads are far too lipophilic to have good chances of success as safe drugs, as logP (calculated octanol-water partition coefficient) correlates positively with *in vitro* compound promiscuity [2]. For oral drugs, lipophilicity is generally considered a requirement for their absorption by passive diffusion in membranes, although there is some debate regarding the contribution of carrier-mediated transport to cellular uptake [6,7]. Standard medicinal chemistry practices, driven by a “perceived need for potency” [8] and a sense of urgency to deliver leads into development, also contribute. In practical terms this has led to more lipophilic candidates, as the exploration of more risky molecules is discouraged [3]. This realization has led to the concept of lipophilic efficiency [2].

Nevertheless, trends in target selection are now providing further challenges. Although the focus on enzyme superfamilies such as protein kinases [9] and aspartic proteinases [10] has allowed knowledge and expertise to be transferred from one target to the next superfamily member, selectivity has proved more elusive than hoped. This has led to increasing interest in targeting protein-protein interactions [11], where the varied regulatory mechanisms within superfamilies promise a new avenue to selective drugs [12]. However, successful small molecule inhibitors of multi-protein complexes tend to be larger lipophilic molecules

with few polar features [13], representing challenging starting points for the development of safe drugs. Here we ask whether this size and lipophilicity is a requirement that small molecules need to fulfill in order to bind to protein interfaces or merely a reflection on the nature of chemical matter that has been explored to date.

The aim of this analysis is to understand how Nature effects interactions and to migrate this knowledge to the design of small molecule modulators of biological targets. Although molecular recognition laws are far from simple, one can elucidate general trends in terms of atomic interactions from experimentally determined structures of natural protein complexes (not only multi-protein complexes but also endogenous small-molecule protein complexes) and compare them with trends from synthetic small molecule protein complexes. We therefore define interaction profiles in terms of polar and apolar contacts, with the aim of learning from natural patterns and incorporating them into the design of new therapeutics.

Methods

Compound set definition

Description and examples for each molecular set studied here can be found in File S1. PDB codes and redundancy flags can be downloaded from <http://www-cryst.bioc.cam.ac.uk/members/alicia>.

Contact definition

Software to calculate hydrogen bonds for all types of molecules (proteins, nucleic acids and small molecules), with the same level of

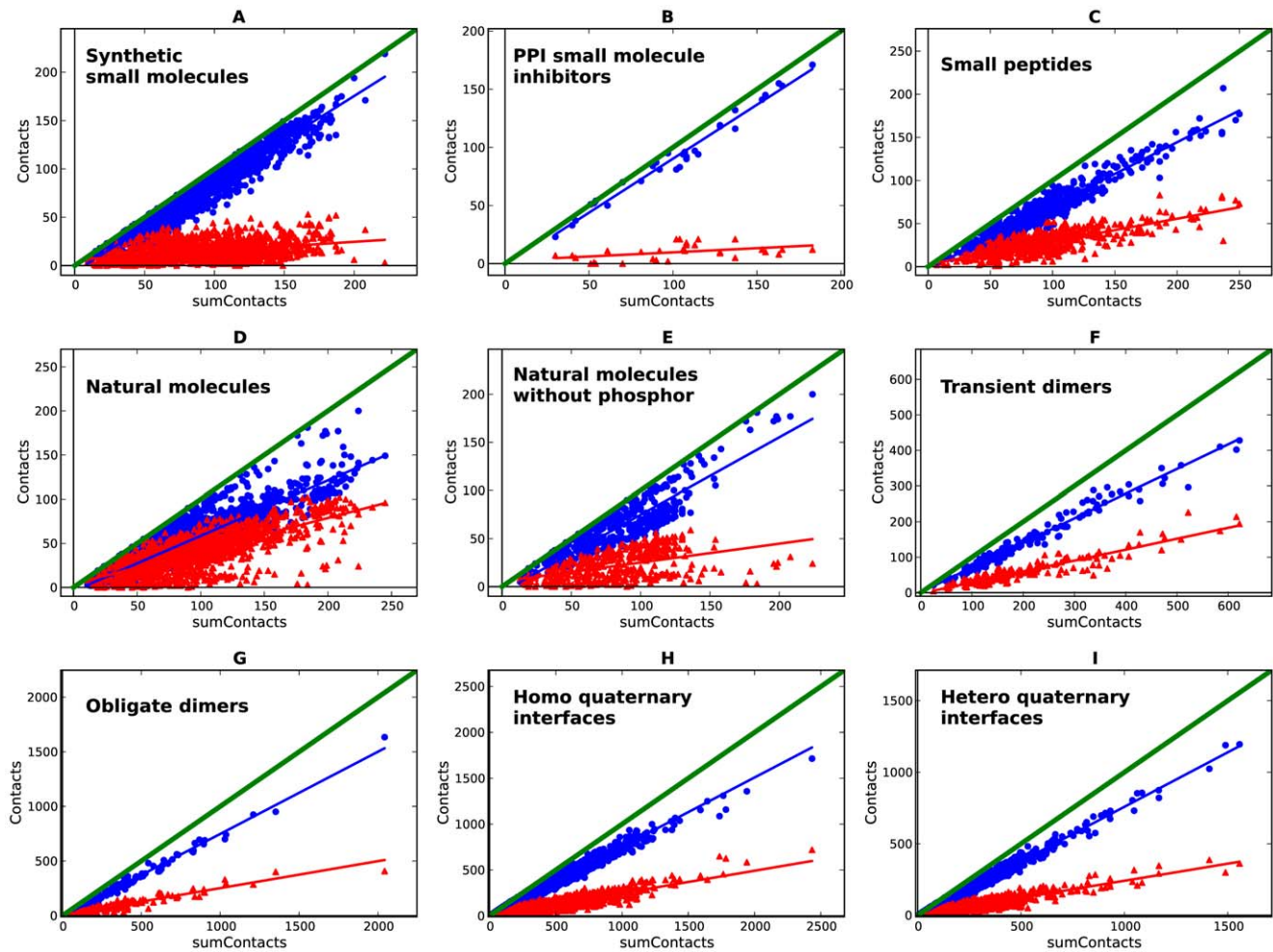


Figure 1. Scissors plots for the non-redundant-by-complex (Table S2) sets of protein complexes. a: synthetic small molecules bound to proteins; b: protein-protein interactions small molecule inhibitors bound to proteins; c: small peptides bound to proteins; d: natural small molecules bound to proteins; e: natural small molecules not containing phosphorus bound to proteins; f: transient protein-protein dimers; g: obligate protein-protein dimers; h: homo protein-protein interfaces from quaternary structures; i: hetero protein-protein interfaces from quaternary structures; polar (red) and apolar (blue) contacts are scattered against sum of contacts. doi:10.1371/journal.pone.0051742.g001

specificity, is not available at the moment. For this reason, simple polar and apolar contacts were defined using a simple distance cut-off of 4.5 Å. Contacts are labeled depending on the atom types of the proximal pair as follows:

Protein-protein complexes. Apolar contacts: C...C, C...S, S...S (not in Cys-Cys bridges).

Polar contacts: N...O, O...O, N...N, O...S, N...S (S from Cys).

Protein-small molecules complexes. Apolar contacts: C...C, C...S, C...X, S...X (X = Cl, Br, I).

Polar contacts: N...O, O...O, N...N, O...S, N...S, N...F, O...F, S...F (S from Cys).

This discrete count of atomic interactions correlates with buried surface area, which is used in other related studies. See File S2.

Calculation of molecular properties

Molecular properties have been calculated using Pipeline Pilot (<http://accelrys.com/products/pipeline-pilot/>). The partition coefficient used in this study is AlogP. Rotatable bonds are defined as single bonds between heavy atoms. This does not include ring bonds (regardless of ring size) or those bonds that connect two

heavy atoms, one of which is attached to only hydrogens. Furthermore, amide bonds are not considered to be rotatable.

Statistical analysis

Distribution of polar/apolar contact ratio (normalized as polar/[polar+apolar]) between sets has been analyzed with non-parametrical tests (Kruskal-Wallis test for comparison of medians), as not all the sets have a normal distribution of the contact ratio.

Results

Data sets

A non-redundant set of protein-ligand complexes was extracted from the CREDO database [14] and classified into six sets by the type of small molecule involved: approved drugs, oral drugs, synthetic small molecules (for want of a better term, see File S1), protein-protein interactions inhibitors, natural small molecules and small peptides. These terms are used to describe the categories in the remaining sections of this paper. Five non-redundant sets of protein-protein complexes were extracted from the PICCOLO database [15] and classified as obligate dimers, transient dimers,

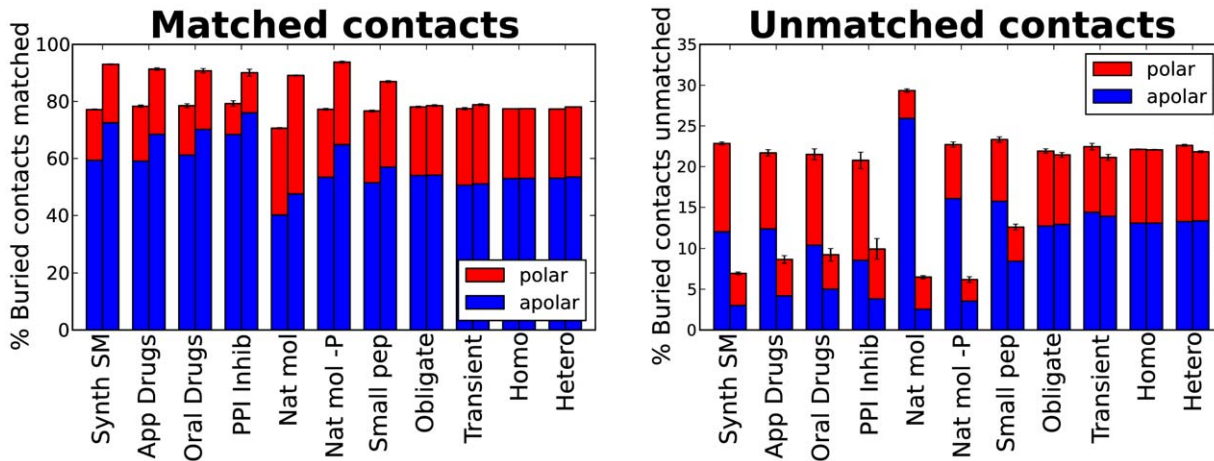


Figure 2. Matched and unmatched contacts. Mean of the percentage of buried atoms per molecule engaged in successful interactions (Matched contacts, left chart) and mean of the percentage of buried atoms without an appropriate partner in the other side of the interface (Unmatched contacts, right chart). The percentage is divided into polar (red) and apolar (blue) contribution. Each set has two bars, one on the left for the atoms in the protein and one on the right for the atoms in the ligand or smaller protein in the case of protein complexes. Error bars denote the standard error of the mean. Sets are ordered from left to right: Synthetic small molecules, approved drugs, oral drugs, protein-protein interaction small molecule inhibitors, natural molecules, natural molecules without phosphorus, small peptides, obligate protein-protein dimers, transient protein-protein dimers, homo quaternary protein-protein interfaces and hetero quaternary protein-protein interfaces. doi:10.1371/journal.pone.0051742.g002

homo and hetero pairwise interfaces from quaternary assemblies, and protein complexes successfully inhibited by small molecules. See File S1. PDB [16] codes for each subset are available to download at: <http://www-cryst.bioc.cam.ac.uk/members/alicia>.

Polar and apolar contacts with protein molecules

Comparisons between different sets of molecules are based upon the number of polar and apolar atomic contacts. This discrete count of atomic interactions is analogous to the measurement of buried surface area used in other studies, providing a coarse description of the binding interfaces (see File S2).

For each set of interfaces, plotting the total number of contacts (as sum of contacts) versus the number of both polar and apolar contacts generates the ‘scissors plot’. In such scissors plots, the

relationship between the number of contacts and the polar and apolar contributions can be shown to be linear, with the gradient varying depending of the interaction type. The angle between the trend lines reflects the ratio of polar versus apolar contributions. Figure 1 shows the scissors plots for synthetic small molecules and PPI inhibitors (scissors open), natural small molecules (scissors closed) and small peptides and protein complexes (scissors half way). For the synthetic small molecules, the larger the interface, the greater the degree to which molecular interactions are dominated by apolar contacts, whereas the polar contacts remain relatively constant. The same conclusion was reached by Olsson and co-workers in their analysis of the SCORPIO database [17]. This observation is more pronounced in small molecules inhibiting protein-protein interactions as shown previously [13]. Conversely, natural small molecules, small peptides and protein complexes follow a different trend, where the polar interactions proportionally make a greater contribution. Interestingly, the lower left quadrant of the graphs, where the smaller ligands (fragments) are mostly located, shows a more balanced ratio between polar and apolar contacts. See File S3 for a different representation of these data.

This result agrees with the Hann complexity model [18], which maintains that it is easier for a smaller molecule to match target features, and also it supports the strategy of fragment-based drug design where the initial fragments anchor in a site with specific interactions [19] and where the step-by-step “growing” of the molecules, often guided by structure, tends to deliver less lipophilic hits [3],[5]. Natural molecules have a bimodal distribution as shown in Figure 1 (d and e) and Figure S1 (a); this is due to the presence of a subpopulation of steroid-like molecules, which have an apolar profile, amongst the population as a whole which tend to have more polar interactions. To evaluate whether this relatively high proportion of polar contacts is because many of the natural molecules have phosphate groups, Figure 1e shows the scissors plot for the set of natural molecules, which do not contain phosphorus. Despite the smaller sample size the same trend is observed, and the bimodal distribution is maintained. See File S4 for statistical significance of these plots.

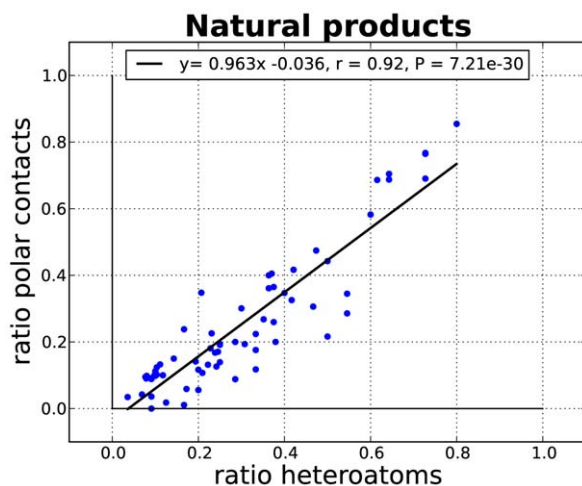


Figure 3. Linear regression of the proportion of heteroatoms to the polar ratio (as number of polar contacts divided by the total [polar + apolar] number of contacts) for the natural-product-like subset of molecules. doi:10.1371/journal.pone.0051742.g003

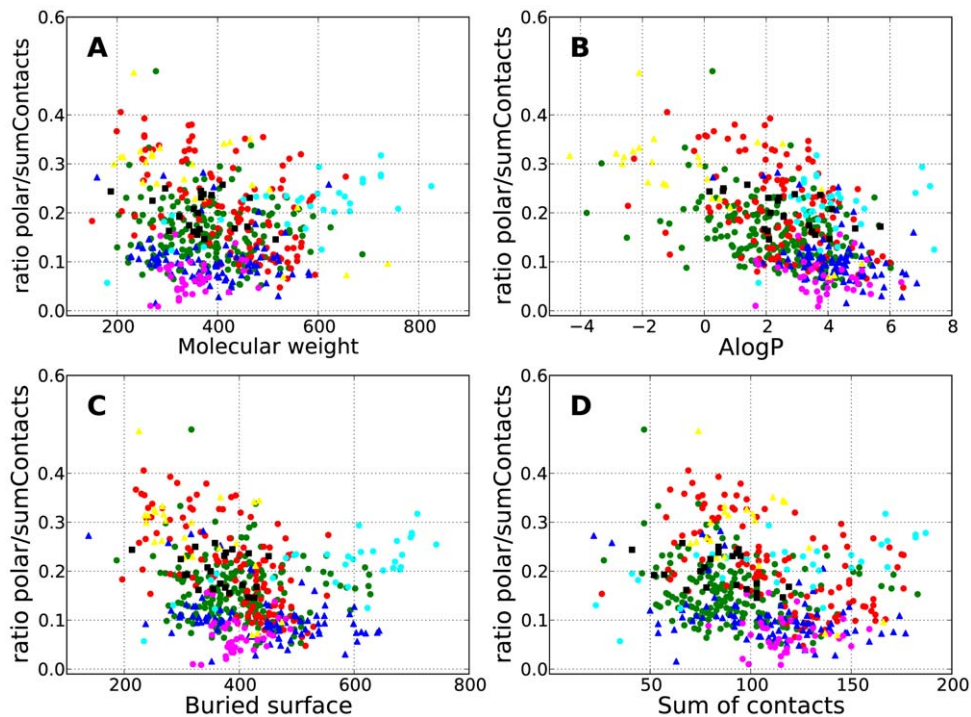


Figure 4. Calculated properties versus interaction profile. The polar ratio (as number of polar contacts divided by the total [polar + apolar] number of contacts) versus molecular weight (A), AlogP (B), buried area upon binding (C) and sum of contacts (D) for protein complexes with synthetic small molecules. Different colors denote SCOP families: Protein kinase catalytic subunit (green), nuclear receptor ligand-binding domain (blue), eukaryotic proteases (red), retroviral proteases – retropepsin (cyan), reverse transcriptase (magenta), Higher-molecular weight phosphotyrosine protein phosphatases (yellow), HSP90 N-terminal domain (black). For clarity, only SCOP families binding to more than 20 different ligands are shown.

doi:10.1371/journal.pone.0051742.g004

Comparison of Figure 1a with Figure 1d and Figure 1e illustrates that natural molecules have a higher proportion of polar contacts than their synthetic counterparts. Likewise on average small peptides also engage in more polar contacts with their targets than synthetic molecules (Figure 1c versus 1a). In order to understand the determinants of these observed differences, the chemical structures of the different ligand sets were analyzed in terms of heteroatom content, rotatable bonds and the proportion of matched and unmatched atoms at the interfaces. The results of these analyses are shown below.

Atomic composition and molecular flexibility

The heteroatom content of each molecule was measured by calculating the ratio of the number of non-carbon atoms to the total number of heavy (i.e. non hydrogen) atoms. The flexibility of each molecule was estimated by dividing the number of rotatable bonds by the number of heavy atoms. The average values for these metrics were then calculated for each of our sets, see File S5. The more polar interaction profile presented by natural molecules compared to synthetic small molecules coincides with a higher content of heteroatoms (26% on average in synthetic small molecules raises to 45% on average in natural molecules). Whereas peptides have a lower content of heteroatoms compared to natural molecules (36% on average). Peptides are also more flexible, the proportion of rotatable bonds by heavy atom is on average 18%, 22% and 42% for synthetic, natural molecules and small peptides respectively. This flexibility should make them more able to match the more directionally constrained polar interactions. In contrast to the natural molecules and peptide sets, the predominately apolar nature of the interactions for synthetic small molecules

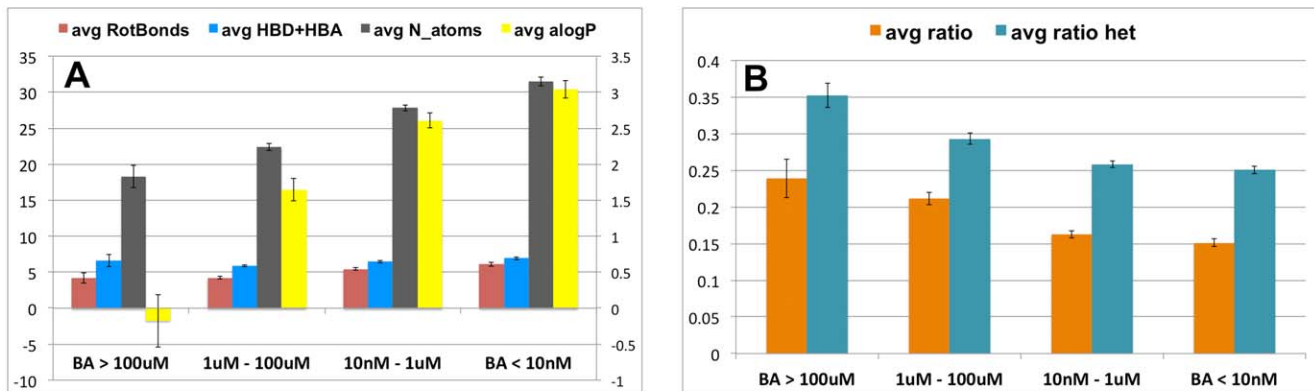
coincides with a lower heteroatom content and greater rigidity. However, examples exist of natural molecules that are rigid and lipophilic, including steroids such as testosterone. In fact, natural molecules cover a broad range of polarity, but the overall character is predominantly polar, especially when compared with synthetic molecules.

Matched and unmatched atoms at the binding interfaces

As we have shown, the higher content of heteroatoms in non-peptidic natural molecules often leads to more polar interactions with the protein. It is also clear from our analyses that this is not the case for small peptides but the lower heteroatom content may be compensated for by their greater flexibility, allowing them to form more polar interactions. However, it is important to understand what proportion of heteroatoms in each class of molecule is making polar contacts or is unmatched (i.e. heteroatoms within the 4.5 Å threshold that do not satisfy any polar criteria). In other words, what are the polar contact efficiencies?

Figure 2 shows the means of the proportions of matched and unmatched buried atoms. Small molecule ligands, which are represented on the left hand side of the figure, are more contact efficient than the protein to which they are bound. On average around 90% of the ligand atoms are matched in all sets, whereas only 70–80% of the protein atoms are matched. Natural molecules without phosphorus are the most contact efficient, 70–80% of the protein atoms are matched. The small molecule atoms are more exposed and able to contact the protein, whereas the atoms in the protein can be less accessible probably due to the concave nature of ligand binding pockets. Furthermore, studies of ligand-binding

Synthetic small molecules



Small peptides

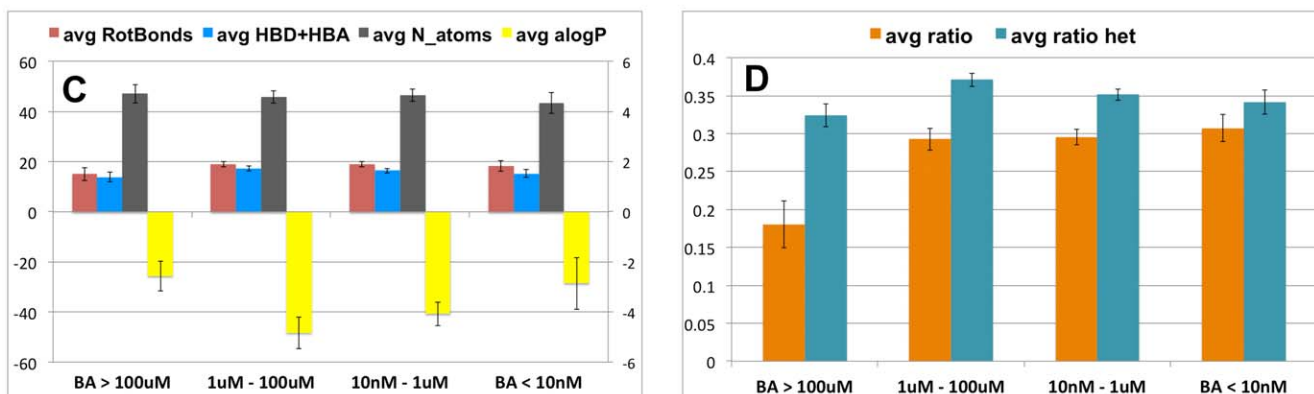


Figure 5. Binding affinity versus interaction profile. Binned binding affinity (BA) data for synthetic small molecules (A and B) and for small peptides (C and D). Bars in (A) and (C) denote the average of molecular properties for each affinity bin: rotatable bonds (red), sum of hydrogen bond donors and acceptors (blue), number of atoms (black) and AlogP (yellow, in the secondary right axis scale for clarity). Bars in (B) and (D) denote the average of polar ratio (orange) and the average heteroatom content (cyan). Error bars are the standard error of the mean. doi:10.1371/journal.pone.0051742.g005

pockets have shown that they are on average one and a half to three times bigger than the ligand they encapsulate [20,21] (the exact ratio varies depending on the dataset studied and the algorithm used). Therefore, in proportion more atoms in the

protein will be at the periphery of the ligand (our cut-off here was 4.5 Å) without making useful interactions. Another interesting result from this analysis is that synthetic molecules have a higher proportion of unmatched polar atoms (on both ligand and protein

Table 1. Examples of polar ratio in protein complexes.

Target	PDB p-p	Ratio p-p	Affinity	PDB p-sm	Ratio p-sm	Affinity	refs
IL-2	1Z92 (A:B)	0.35	10nM (Kd)	1PY2 (A)	0.21	60nM (IC50)	[27],[28]
Bcl-XL	2BZW (A:B)	0.19	6nM (Kd)	2YXJ (B)	0.08	0.6nM	[29]
MDM2	1YCR (A:B)	0.14	600nM (Kd)	1T4E (A)	0.03	67nM (Kd)	[30],[31]
XIAP	1G3F (A:B)	0.22	-	1TFT (A)	0.12	-	[32],[33]
ZipA	1F47 (B:A)	0.10	21.6uM (Kd)	1Y2F (A)	0.00	12uM (Kd)	[34],[35]
TNF	1TNF (AB:C)	0.30	-	2AZ5 (C&D)	0.12	13uM	[36],[37]
S100B	1DT7 (A:X)	0.34	-	3GK1 (A)	0.12	-	[38],[39]

Polar ratio is the number of polar contacts divided by the total [polar + apolar] number of contacts. These protein complexes are for proteins that bind to both protein partners (ratio p-p, left) and synthetic small molecules (ratio p-sm, right). The PDB code includes the interacting chains, for example 1TNF(AB:C) denotes chain A and B interacting with chain C of the TNF trimer, whereas 2AZ5 (C&D) denotes chains C and D interacting with the small molecule. When available, affinity measure and units is specified in table.

doi:10.1371/journal.pone.0051742.t001

side) than the natural ones. Therefore, if one wants to increase the polar contacts synthetic molecules make, then there is opportunity for improvement. Oral drugs need to have a limited polar surface area in order to be absorbed by the body [22]. However, Figure 2 shows that approved and oral drugs are not making the most of their polar composition. Nevertheless, improving enthalpic contacts is not a trivial task, not only because of the difficulty of designing chemical geometries that match polar constraints, but also because of the increased desolvation penalty and the reduction of conformational entropy [23] that can result.

The question remains, does Nature make the most of the polar composition of its ligands? Figure 3 shows a linear correlation between heteroatom content (the number heteroatoms divided by the total number of atoms) and the ratio of polar interactions (the number polar contacts divided by the sum of contacts, henceforth referred to as the polar ratio), for the natural-product-like subset in natural molecules. For these small molecules, the increase in polar features translates into more polar interactions with the protein.

Synthetic small molecule complexes. Calculated properties versus interaction profile

An analysis of the distribution of the polar ratio plotted against molecular weight, AlogP, surface area buried upon binding and sum of contacts has been carried out for the synthetic small molecules. Figure 4 shows these distributions color-coded by SCOP [24] family.

Synthetic molecules bound to protein kinases (green dots in Figure 4) tend to have a high AlogP and a low polar ratio (usually below 30%). Even those molecules with a negative AlogP do not appear to make many polar contacts. However, there are a few molecules with 50% of polar contacts, for example HET-ID 3C3 in PDB entry 2CGW. In the case of nuclear receptor ligand-binding domain (NR-LBD, blue triangles in Figure 4), all the molecules have $AlogP > 1$ and most have less than 15% of polar contacts, but as with the kinase ligands there are examples that buck the trend, for example HET-ID 444 in PDB entry 1UHL with a polar ratio of 34%. Eukaryotic proteases (red dots in Figure 4) bind to a wide range of molecules with molecular weights ranging from 200 to 700 Da and with AlogPs between -2 and 6 . They also have a wide range of polar ratios. Bigger and more lipophilic molecules are found to bind to retroviral proteases (cyan dots in Figure 4) with similar polar binding patterns as found for eukaryotic enzymes, although polar fragments are also found. The proteins belonging to the reverse transcriptase SCOP family have similar characteristics to nuclear receptor ligand-binding domain, and bind to molecules with $AlogP > 1$, all of which have less than 15% polar contacts. Most of the molecules bound to phosphotyrosine protein phosphatases (PTPP, yellow triangles in Figure 4) have around 30% of polar contacts with a low AlogP range (-2 to 2), although there are also three apolar binders (HET-ID 892 in PDB entry 1T49, HET-ID BB3 in PDB entry 1T48 and HET-ID FRJ in PDB entry 1T4J). However, these apolar molecules are inhibitors binding to an allosteric site. Finally, HSP90 domains (black squares in Figure 4) bind to molecules with a wide range of AlogP ($0-6$) with between 15–25% of polar contacts.

The characteristics of synthetic small molecules depend heavily upon their targets. However, there is no correlation between calculated lipophilicity or molecular weight and the polar ratio. Thus, it is possible to achieve more polar contacts without compromising the physical properties of the drug leads. See Table S1 for a list of approved and oral drugs with more than 40% of polar contacts. The low lipophilicity of these molecules is noteworthy. In a recent analysis of patented small molecules from 18 pharmaceutical companies, Leeson and St-Galley [5] show that

here is indeed a target dependency of the lead lipophilicity. However, the outcomes of different companies working on the same target also highlight that the major determinant of the lipophilicity of the compounds is the discovery practices of each organization and not target requirements.

Synthetic small molecule complexes. Affinity versus interaction profile

Affinity data from the implementation of PDBBind [25] in CREDO [14] are available for almost 700 of the synthetic small molecule set and more than one hundred of the small peptides. For these molecules there is no correlation between the binding energy and the proportion of polar contacts they make with their protein partners (File S6). Small, weakly binding synthetic molecules have high polar ratios, whereas small peptides can engage more polar contacts across a wide range of affinities.

Indeed, Figure 5a shows that the most potent synthetic molecules have the highest average number of atoms and AlogP while the average count of hydrogen bond acceptors and donors remains constant across the whole potency range. This result is in agreement with the much discussed general tendency in drug discovery of gaining potency by adding lipophilic groups to the small molecules; see for example [2]. In the set studied here, this translates (Fig. 5b) into lower heteroatom content and a lower polar ratio.

Figure 5 (c and d) shows that small peptides have on average the same property profile regardless of their potency. This result suggests that small peptides do not achieve tight binding through increase of lipophilicity. Furthermore the proportion of polar contacts and the heteroatom content is maintained across the whole range of affinities with the exception of weak binders where the polar ratio is lower (although there are only seven complexes in this category). The important point to highlight here is that peptides, unlike synthetic small molecules, can bind specifically to their receptors and have a high number of polar contacts. We argue that this is in part due to their inherent flexibility and low lipophilicity. Note that the average number of rotatable bonds in small peptides is four-fold higher than in synthetic molecules. In fact, the small peptides studied here are often bigger than the synthetic small molecules, as no size limit was applied to select the small peptide set, whereas synthetic molecules larger than 900 Da were omitted. However, there is no correlation between the number of atoms and the free energy of binding for small peptides, see File S7, in agreement with the conclusions previously reached by Kuntz [26]. Nevertheless, they are less efficient than small synthetic molecules as they use more atoms to achieve the same affinity.

Small molecule inhibitors of protein-protein interactions

We now return to the question of whether small molecules inhibiting protein-protein interactions need to be large and lipophilic. Although these molecules tend to be lipophilic with few polar features, Figure 2 shows that they have polar atoms unmatched in the binding site. Using the TIMBAL database [13], we have extracted the seven cases where there is structural information for both the small molecule-protein and the protein-protein complexes. In all cases studied, there are unutilized potential polar contacts on the protein interface (see Table 1 and File S8 for a graphical representation of these data). These cases illustrate a common pattern for synthetic molecules: they typically comprise few anchor points (which can be considered as more constrained polar contacts) and more potency-boosting hydrophobic interactions. Small molecule inhibitors of protein-protein interactions do not take advantage of all the available opportu-

nities for polar contacts in the interfaces as only a few of them are engaged. For these seven cases, comparison of the interacting residues in the target protein highlights that small molecules tend to use more aromatic and fewer charged residues than the protein partner, see Figure S2.

Discussion

In this study we have analyzed the atomic contacts between different sets of molecules, some natural and some synthetic. The results presented here show that natural complexes typically engage in more polar contacts than synthetic molecules bound to proteins. This latter set, also have a higher proportion of unmatched heteroatoms than the natural sets and, probably for this reason, show no correlation between lipophilicity and proportion of polar contacts. These differences may be due to the fact that the dynamics of evolutionary processes are more likely to provide a more highly optimized configuration than is possible by medicinal chemistry efforts. Importantly it must be borne in mind that endogenous molecules have not been constrained by the requirements for oral absorption and distribution across membranes into cellular compartments.

For synthetic small molecules in general, but in particular for the inhibitors of protein-protein interactions, we conclude that efforts should be invested to maximize polar contacts to better resemble the interaction patterns that natural molecules present. As the ratio of polar versus apolar contacts is greater when the size of the synthetic molecules is smaller, this will be easiest to achieve when ligands remain small. However, it will almost always be harder to match polar contacts than to gain affinity through the addition of hydrophobic substituents. Nevertheless, the recent analysis of Leeson and St-Gallay [5] shows that for the same target, a range of compound lipophilicity can be achieved using different strategies, thus encouraging efforts that may lead to better molecules. If we are to minimize ligand promiscuity and ultimately lower the costs of delivering safe and effective drugs to the market, it must be worthwhile to “go back to Nature” and see what can be learnt from the exquisite specificity – albeit low ligand efficiency – of protein-ligand interactions that has evolved to be selectively advantageous in living organisms.

Supporting Information

Figure S1 Normalized distributions of the ratio of polar contacts (represented by $\text{polar}/[\text{polar}+\text{apolar}]$), each chart compares synthetic small molecules against the other sets. (PDF)

References

- Garnier JP (2008) Rebuilding the R&D engine in big pharma. *Harv Bus Rev* 86: 68–70.
- Leeson P, Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* 6: 881–890.
- Keserü GM, Makara GM (2009) The influence of lead discovery strategies on the properties of drug candidates. *Nat Rev Drug Discov* 8: 203–212.
- Gleeson MP, Hersey A, Montanari D, Overington J (2011) Probing the links between *in vitro* potency, ADMET and physicochemical parameters. *Nature Reviews Drug Discovery* 10: 197–208.
- Leeson PD, St-Gallay SA (2011) The influence of the ‘organizational factor’ on compound quality in drug discovery. *Nat Rev Drug Discov* 10: 749–765.
- Dobson PD, Kell DB (2008) Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? *Nat Rev Drug Discov* 7: 205–220.
- Sugano K, Kansy M, Artursson P, Avdeef A, Bendels S, et al. (2010) Coexistence of passive and carrier-mediated processes in drug transport. *Nat Rev Drug Discov* 9: 597–614.
- Hamm MM (2011) Molecular obesity, potency and other addictions in drug discovery. *MedChemComm* 2: 349–355.
- Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, et al. (2011) Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* 29: 1046–1051.
- Deu E, Verdoes M, Bogoy M (2012) New approaches for dissecting protease functions to improve probe development and drug discovery. *Nat Struct Mol Biol* 19: 9–16.
- Wells J, McClendon C (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 450: 1001–1009.
- Cochran AG (2000) Antagonists of protein-protein interactions. *Chem Biol* 7: R85–R94.
- Higueruelo AP, Schreyer A, Bickerton GRJ, Pitt WR, Groom CR, et al. (2009) Atomic Interactions and Profile of Small Molecules Disrupting Protein-Protein Interfaces: the TIMBAL Database. *Chem Biol Drug Des* 74: 457–467.
- Schreyer A, Blundell T (2009) CREDO: A Protein-Ligand Interaction Database for Drug Discovery. *Chem Biol Drug Des* 73: 157–167.
- Bickerton GR, Higueruelo AP, Blundell TL (2011) Comprehensive, atomic-level characterization of structurally characterized protein-protein interactions: the PICCOLO database. *BMC Bioinformatics* 12: 313.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242.

Figure S2 Comparisons of residue propensities at the binding sites for small molecule protein-protein inhibitors versus protein-protein complexes inhibited by them. (PDF)

File S1 Detailed description of each data set. (PDF)

File S2 Buried surface area versus atomic contacts. (PDF)

File S3 More polar contacts for synthetic fragments. (PDF)

File S4 Comparison of interaction profiles. (PDF)

File S5 Heteroatom content and rotatable bond count. (PDF)

File S6 Free energy of ligand binding versus the polar ratio of contacts [polar/(polar+apolar)] for the synthetic small molecules set and the small peptide set. (PDF)

File S7 Free energy of ligand binding versus the number of atoms of the ligand for the synthetic small molecules set and the small peptide set. (PDF)

File S8 Protein-protein complexes inhibited by small molecules. (PDF)

Table S1 Approved and oral small molecule drugs that engage more than 40% polar contacts with their bound protein. (PDF)

Table S2 Number of entries in each set of molecules. (PDF)

Acknowledgments

APH would like to thank Nikolay Todorov and Colin Groom for the many helpful discussions.

Author Contributions

Conceived and designed the experiments: APH TLB WRP. Performed the experiments: APH. Analyzed the data: APH. Contributed reagents/materials/analysis tools: AS GRJB. Wrote the paper: APH TLB WRP. Reviewed and edited the final version of the manuscript: APH AS GRJB TLB WRP.

17. Olsson TSG, Williams MA, Pitt WR, Ladbury JE (2008) The Thermodynamics of Protein-Ligand Interaction and Solvation: Insights for Ligand Design. *J Mol Biol* 384: 1002–1017.
18. Ham MM, Leach AR, Harper G (2001) Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J Chem Inf Comput Sci* 41: 856–864.
19. Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent Developments in Fragment-Based Drug Discovery. *J Med Chem* 51: 3661–3680.
20. An J, Totrov M, Abagyan R (2005) Pocketome via Comprehensive Identification and Classification of Ligand Binding Envelopes. *Molecular & Cellular Proteomics* 4: 752–761.
21. Kahraman A, Morris RJ, Laskowski RA, Thornton JM (2007) Shape variation in protein binding pockets and their ligands. *J Mol Biol* 368: 283–301.
22. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, et al. (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45: 2615–2623.
23. Ferenczy GG, Keserü GM (2010) Thermodynamics guided lead discovery and optimization. *Drug Discov Today* 15: 919–932.
24. Murzin A, Brenner S, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247: 536–540.
25. Wang R, Fang X, Lu Y, Wang S (2004) The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J Med Chem* 47: 2977–2980.
26. Kuntz ID, Chen K, Sharp KA, Kollman PA (1999) The maximal affinity of ligands. *Proc Natl Acad Sci U S A* 96: 9997–10002.
27. Rickert M, Wang X, Boulanger MJ, Goriatheva N, Garcia KC (2005) The structure of interleukin-2 complexed with its alpha receptor. *Science* 308: 1477–1480.
28. Thanos CD, Randal M, Wells JA (2003) Potent Small-Molecule Binding to a Dynamic Hot Spot on IL-2. *J Am Chem Soc* 125: 15280–15281.
29. Lee EF, Czabotar PE, Smith BJ, Deshayes K, Zobel K, et al. (2007) Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of antagonists of the Bcl-2 family. *Cell Death Differ* 14: 1711–1713.
30. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, et al. (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274: 948–953.
31. Grasberger BL, Lu T, Schubert C, Parks DJ, Carver TE, et al. (2005) Discovery and cocystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. *J Med Chem* 48: 909–912.
32. Liu Z, Sun C, Olejniczak ET, Meadows RP, Betz SF, et al. (2000) Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain. *Nature* 408: 1004–1008.
33. Oost TK, Sun C, Armstrong RC, Al-Assaad AS, Betz SF, et al. (2004) Discovery of potent antagonists of the antiapoptotic protein XIAP for the treatment of cancer. *J Med Chem* 47: 4417–4426.
34. Mosyak L, Zhang Y, Glasfeld E, Hancy S, Stahl M, et al. (2000) The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by X-ray crystallography. *EMBO J* 19: 3179–3191.
35. Rush TS, Grant JA, Mosyak L, Nicholls A (2005) A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J Med Chem* 48: 1489–1495.
36. Eck MJ, Sprang SR (1989) The structure of tumor necrosis factor-alpha at 2.6 Å resolution. Implications for receptor binding. *J Biol Chem* 264: 17595–17605.
37. He M, Smith A, Oslob J, Flanagan W, Braisted A, et al. (2005) Small-Molecule Inhibition of TNF-alpha. *Science* 310: 1022–1025.
38. Rustandi RR, Baldisseri DM, Weber DJ (2000) Structure of the negative regulatory domain of p53 bound to S100B(beta-beta). *Nat Struct Biol* 7: 570–574.
39. Charpentier T, Wilder P, Liriano M, Varney K, Zhong S, et al. (2009) Small Molecules Bound to Unique Sites in the Target Protein Binding Cleft of Calcium-Bound S100B As Characterized by Nuclear Magnetic Resonance and X-ray Crystallography. *Biochemistry* 48: 6202–6212.