

## Differences between High- and Low-Affinity Complexes of Enzymes and Nonenzymes

Heather A. Carlson,<sup>\*,†</sup> Richard D. Smith,<sup>‡</sup> Nickolay A. Khazanov,<sup>§</sup> Paul D. Kirchhoff,<sup>†</sup> James B. Dunbar, Jr.,<sup>†</sup> and Mark L. Benson<sup>§</sup>

Department of Medicinal Chemistry, Biophysics Research Division, and Bioinformatics Graduate Program, University of Michigan, Ann Arbor, Michigan 48109

Received May 29, 2008

Physical differences in small molecule binding between enzymes and nonenzymes were found through mining the protein–ligand database, Binding MOAD (Mother of All Databases). The data suggest that divergent approaches may be more productive for improving the affinity of ligands for the two classes of proteins. High-affinity ligands of enzymes are much larger than those with low affinity, indicating that the addition of complementary functional groups is likely to improve the affinity of an enzyme inhibitor. However, this process may not be as fruitful for ligands of nonenzymes. High- and low-affinity ligands of nonenzymes are nearly the same size, so modest modifications and isosteric replacement might be most productive. The inherent differences between enzymes and nonenzymes have significant ramifications for scoring functions and structure-based drug design. In particular, nonenzymes were found to have greater ligand efficiencies than enzymes. Ligand efficiencies are often used to indicate druggability of a target, and this finding supports the feasibility of nonenzymes as drug targets. The differences in ligand efficiencies do not appear to come from the ligands; instead, the pockets yield different amino acid compositions despite very similar distributions of amino acids in the overall protein sequences.

### Introduction

Both enzymatic and nonenzymatic proteins can bind small molecules, but enzymes catalyze reactions and have a fundamentally different role from nonenzymes, which may have an impact on their recognition of ligands. Do these two types of binding events have the same physical characteristics? Furthermore, are there any differences between high-affinity complexes and weaker binding events that can be linked to their physical contacts? To answer these questions, physicochemical patterns were mined from our protein–ligand database Binding MOAD<sup>a</sup> (Mother of All Databases), where MOAD is pronounced “mode” as a pun on a ligand’s mode of binding.<sup>1,2</sup>

Binding MOAD is the largest curated database of high-resolution protein–ligand complexes from the Protein Data Bank (PDB).<sup>3</sup> Although it only reflects proteins that can be crystallized, these are the exact systems where structure-based insights will be used. The PDB is the source of all structures used for docking and scoring development by academics. However, the data used here are significantly larger than most sets used to develop existing scoring functions, which are typically sets of <300 complexes of <50 unique proteins. We use 2214 structures: 1790 enzymes and 424 nonenzymes (512 unique enzymes and 176 unique nonenzymes). This study provides an important benchmark of the current landscape available from structural biology (incomplete and/or biased as it may be).

For this study, we have compared distributions of various properties between four classes of protein complexes. Distribu-

tion analysis is used widely in many fields, and it is important to stress that it does not define “absolute rules” nor are the data presented as such. These are general guidelines and, of course, there will be exceptions to those trends. Distribution analysis can show that “men are taller than women” and “women live longer than men.” Those trends are true even though some women are 6 ft tall and some men live to 100.

Empirically derived rules can be very useful in discovering and applying new principles in chemistry. One of the most well-known examples is Lipinski’s Rule of Five, which describes the physical properties of orally available drugs.<sup>4,5</sup> These rules provide general guidelines for size, lipophilicity, and hydrogen-bonding characteristics that correlate with the likelihood that a molecule can be orally absorbed into the body. The findings are based on distribution data of the chemical characteristics of orally absorbed molecules going into phase-II testing. The data set is biased by issues outside of pharmacokinetics such as the need for good synthesis (not just accessible chemistry, but few steps in high yield) and market considerations (completely economic, no basis in the thermodynamics of protein–ligand binding). The rules do not hold for natural products, actively transported molecules, molecules that require metabolism for activation, or most antibiotics, antifungals, vitamins, and cardiac glycosides. There are plenty of molecules in Lipinski space that are not drugs, and many molecules outside that space that are. Despite these limitations and biases, the Rule of Five is used widely in the pharmaceutical industry.

We hope that the present work will also aid drug discovery. In this study, we provide new patterns that describe high-affinity, protein–ligand binding and outline differences between enzymes and nonenzymes. Of course, there will be examples that fall outside the typical pattern, but these relationships provide a good description of the general landscape that structural biology can provide at this time. We expect that our understanding will grow as more structures become available through the various protein structure initiatives.<sup>6</sup> These guiding principles may be useful in designing targeted libraries for drug discovery and improving

\* To whom correspondence should be addressed. Phone: 1-734-615-6841. Fax: 734-763-2022. E-mail: carlsonh@umich.edu.

<sup>†</sup> Department of Medicinal Chemistry, University of Michigan.

<sup>‡</sup> Biophysics Research Division, University of Michigan.

<sup>§</sup> Bioinformatics Graduate Program, University of Michigan.

<sup>a</sup> Abbreviations: MOAD, Mother of All Databases; PDB, Protein Data Bank; BSA, buried surface area; ESA, exposed surface area; MSA, molecular surface area; SASA, solvent-accessible surface area;  $\Delta G_{\text{bind}}$ , free energy of binding.

scoring functions. They are also important to advancing our fundamental understanding of chemical biology, protein–ligand binding, and the biophysics that dictate molecular recognition.

Noncovalent, small molecule binding is a tradeoff between the enthalpy gained by making specific contacts between functional groups of the ligand and the protein and entropy lost by forcing the ligand and protein into a specific conformation.<sup>7,8</sup> Because this study uses crystal structures, it is difficult to fully account for the effect caused by entropy. However, it is possible to determine the physical characteristics of the small molecule and the protein that leads to the binding affinity.

Other studies<sup>9,10</sup> have noted an inherent limitation in mining protein structures for physical characteristics of binding. When a pocket is discovered on a protein surface, it is difficult to identify whether it is a true binding site or if it is capable of high-affinity binding appropriate to represent druglike binding. This study does not suffer from these limitations; all sites have been curated to ensure that they are true binding pockets, and the high-affinity complexes are separated from those with low affinity.

Only complexes with binding data ( $K_d$ ,  $K_i$ , or  $IC_{50}$ ) were used for this study. No complexes in MOAD are annotated with  $K_m$  data, so almost all ligands are inhibitors, agonists, or antagonists (a small number are cofactors, 5%, included only for systems where affinity data is appropriate). We specifically focused on the contacts between the ligand and the protein, excluding any structure with poorly defined contacts such as missing atoms from under-resolved density or ligands and side chains resolved in multiple orientations. Distributions of ligand size, buried surface area (BSA), exposed surface area (ESA), and other physical characteristics were examined for statistically significant differences between four subsets of the complexes: high-affinity binding to enzymes, high-affinity nonenzymes, low-affinity enzymes, and low-affinity nonenzymes. A common metric to evaluate lead compounds is ligand efficiency.<sup>11–14</sup> In this study, ligand efficiencies for the different classes of proteins are reported as affinity per size ( $-\Delta G_{\text{bind}}$  divided by the number of non-hydrogen atoms) and per the degree of contact between the ligand and the pocket ( $-\Delta G_{\text{bind}}/BSA$ ).

Here, we focus on the most significant differences between molecular recognition of tight and weak binding to enzymes and nonenzymes.

## Methods

Data for this study come from the largest comprehensive database of protein–ligand crystal structures with binding data, Binding MOAD. The latest version of Binding MOAD was created from structures released on 12/31/2006 or earlier; it contains 9836 complexes, comprising 3151 unique protein families binding to 4659 unique ligands. The great care taken in curating this data set has been outlined elsewhere,<sup>1</sup> but it should be noted for these purposes that ~9000 crystallography papers have been examined to determine the appropriateness of every ligand (crystallographic additives, post-translational modifications, and covalently bound ligands are excluded from consideration). From these efforts, binding affinity data is available for 30% of the entries, with a preference for  $K_d$  data over  $K_i$  data over  $IC_{50}$  values. The affinities were converted to free energies of binding by  $\Delta G_{\text{bind}} = RT \times \ln(K_d)$  or simply approximated by  $\Delta G_{\text{bind}} = RT \times \ln(K_i \text{ or } IC_{50})$  with a temperature of 298 K.

High-affinity binding was defined as  $K_d$ ,  $K_i$ , or  $IC_{50} \leq 250$  nM ( $\Delta G_{\text{bind}} \leq -9$  kcal/mol), which is approximately the average of all the complexes with binding data in Binding MOAD. Enzyme complexes were defined from the enzyme classification number in the PDB file. The nonenzymes were annotated by hand using keywords reported in the remarks section of the PDB entry. Binding MOAD's

high-affinity nonenzymes and enzymes are listed in the Supporting Information. All complexes and binding data are available at the Binding MOAD Web site, [www.BindingMOAD.org](http://www.BindingMOAD.org).

To calculate surface areas, BSA and ESA were calculated with GoCAV using radii based on united-atom OPLS parameters.<sup>2</sup> This code reports buried molecular surface area (MSA) of the pocket and also defines ESA of the binding site, bounded by the 3D coordinates of the ligand.

The SlogP for the ligands was calculated using MOE<sup>15</sup> based on the method developed by Wildman and Crippen.<sup>16</sup> For the 2D and 3D descriptors calculated with MOE, the idealized SDF files from the PDB were used if available; otherwise, the coordinates of the ligand from the protein's structure were taken. Hydrogens were added with MOE. In an effort to identify any differences, all 2D and 3D ligand characteristics available within MOE were compared for the four groups of complexes: high-affinity enzyme, low-affinity enzyme, high-affinity nonenzyme, and low-affinity nonenzyme.

**Statistical Analysis.** Statistical significance was assessed with the programs SAS<sup>17</sup> and JMP.<sup>18</sup> Initial assessments used JMP to calculate all pairwise correlations for the over 200 descriptors calculated. For the descriptors showing interesting trends, the significance of the differences between the distributions of physical properties were determined by the Wilcoxon rank-sum test, which is most appropriate given the non-Gaussian distributions of the data. We also performed one-way ANOVA, two-way ANOVA, and Tukey–Kramer HSD tests between the four classifications. Because these second series of tests require near-normal distributions, the square-root transform was applied to reduce the skew and bring the distributions closer to normal. For the important descriptors, distribution analyses from JMP are included in the Supporting Information (Supporting Information, Figures S1–S7), and each includes the mean, median, quantiles, distribution histogram, and outlier box plot. The results of the Tukey–Kramer HSD test are presented in Supporting Information (Supporting Information, Tables S1–S5).

Histograms of the distributions of ligand size were binned in increments of 5 heavy atoms. Distributions of BSA and ESA were binned by 50 Å<sup>2</sup>. Those plotting ligand efficiency were binned by 0.1 kcal/mol-atom for affinity per size or 10 cal/mol-Å<sup>2</sup> for affinity per degree of contact. Distributions of SlogP were binned by 2 log units. These bin sizes were in proportion to the size of the data sets and were consistent with those automatically generated by JMP.

## Results and Discussion

Considerable effort was made to determine direct mathematical relationships between affinity and surface area, ligand size, or other characteristics of protein–ligand interactions, but there was no global correlation across all complexes. Recent work by Coleman and Sharp<sup>19</sup> based on the PDBbind data set<sup>20</sup> also found no correlation between affinity and surface area or depth of the binding pocket. Inspired by analyses of distributions of ligand efficiencies from screening data,<sup>11</sup> we changed our approach and focused on distributions of the properties between subsets of protein–ligand complexes.

Table 1 outlines the characteristics that differ between high-affinity and low-affinity binding for enzymes and nonenzymes; all emphasized differences in the data sets have a statistical significance >99.99% ( $p < 0.0001$ ) based on a two-tailed, Wilcoxon rank-sum test. Figure 1 shows a comparison between each of the subsets of complexes, examining the distribution of ligand sizes, BSA, SlogP, and ESA. Many of the low-affinity complexes have ~300 Å<sup>2</sup> of BSA, but the high-affinity complexes display more contact. It has been estimated that druglike binding sites have ~300 Å<sup>2</sup> of solvent-accessible surface area (SASA).<sup>9</sup> Our measurement for BSA is based on MSA, and so the slightly higher values of the high-affinity complexes are appropriately comparable.<sup>9</sup>

**Table 1.** Characteristics of Protein–Ligand Binding for Enzymes and Nonenzymes in the Full Dataset<sup>a</sup>

median physical properties	low affinity > 250 nM $\Delta G_{\text{bind}}$ > -9 kcal/mol	high affinity $\leq$ 250 nM $\Delta G_{\text{bind}}$ $\leq$ -9 kcal/mol	comparison <sup>b</sup>
<b>enzymes</b>	<b>1048 complexes</b>	<b>742 complexes</b>	
$\Delta G_{\text{bind}}$	-6.6 kcal/mol	-10.9 kcal/mol	
size <sup>c</sup>	21 atoms	32 atoms	
BSA	305 Å <sup>2</sup>	419 Å <sup>2</sup>	
ESA (%ESA) <sup>d</sup>	87 Å <sup>2</sup> (22%)	144 Å <sup>2</sup> (24%)	high-affinity ligands are 52% larger and more hydrophobic
SlogP	0.3	2.4	
$-\Delta G_{\text{bind}}/\text{atom}$	0.31 kcal/mol-atom	0.36 kcal/mol-atom	
$-\Delta G_{\text{bind}}/\text{BSA}$	21 cal/mol-Å <sup>2</sup>	26 cal/mol-Å <sup>2</sup>	
<b>nonenzymes</b>	<b>234 complexes</b>	<b>190 complexes</b>	
$\Delta G_{\text{bind}}$	-7.2 kcal/mol	-10.4 kcal/mol	
size <sup>c</sup>	22 atoms	25 atoms	
BSA	265 Å <sup>2</sup>	361 Å <sup>2</sup>	
ESA (%ESA) <sup>d</sup>	118 Å <sup>2</sup> (33%)	45 Å <sup>2</sup> (11%)	Low-affinity ligands are three times more exposed and more hydrophilic
SlogP	-2.2	1.5	
$-\Delta G_{\text{bind}}/\text{atom}$	0.28 kcal/mol-atom	0.41 kcal/mol-atom	
$-\Delta G_{\text{bind}}/\text{BSA}$	22 cal/mol-Å <sup>2</sup>	31 cal/mol-Å <sup>2</sup>	
<b>comparison<sup>b</sup></b>		Nonenzymes have 17% greater ligand efficiencies	in enzymes, high-affinity in nonenzymes, low-affinity

<sup>a</sup> Values presented are medians for each population. <sup>b</sup> All differences noted in the comparisons sections have a statistical significance of >99.99% ( $p < 0.0001$ ). <sup>c</sup> Ligand size is given in the number of non-hydrogen atoms. <sup>d</sup> Percent exposure is ESA/(ESA + BSA).

**Different Approaches for Improving Inhibitors of Enzymes versus Nonenzymes.** For enzymes, there is a significant difference in the size of the ligands in high- and low-affinity complexes (Figure 1a). High-affinity ligands are much larger (11 more non-hydrogen atoms). However, nonenzymes display very little difference in the size of the ligands between high-affinity and low-affinity complexes (Table 1, Figure 1b). These differences do not come from any influence of the inclusion of cofactors in the set. The medians are nearly unchanged if they are removed from the data set (see Supporting Information, Table S6).

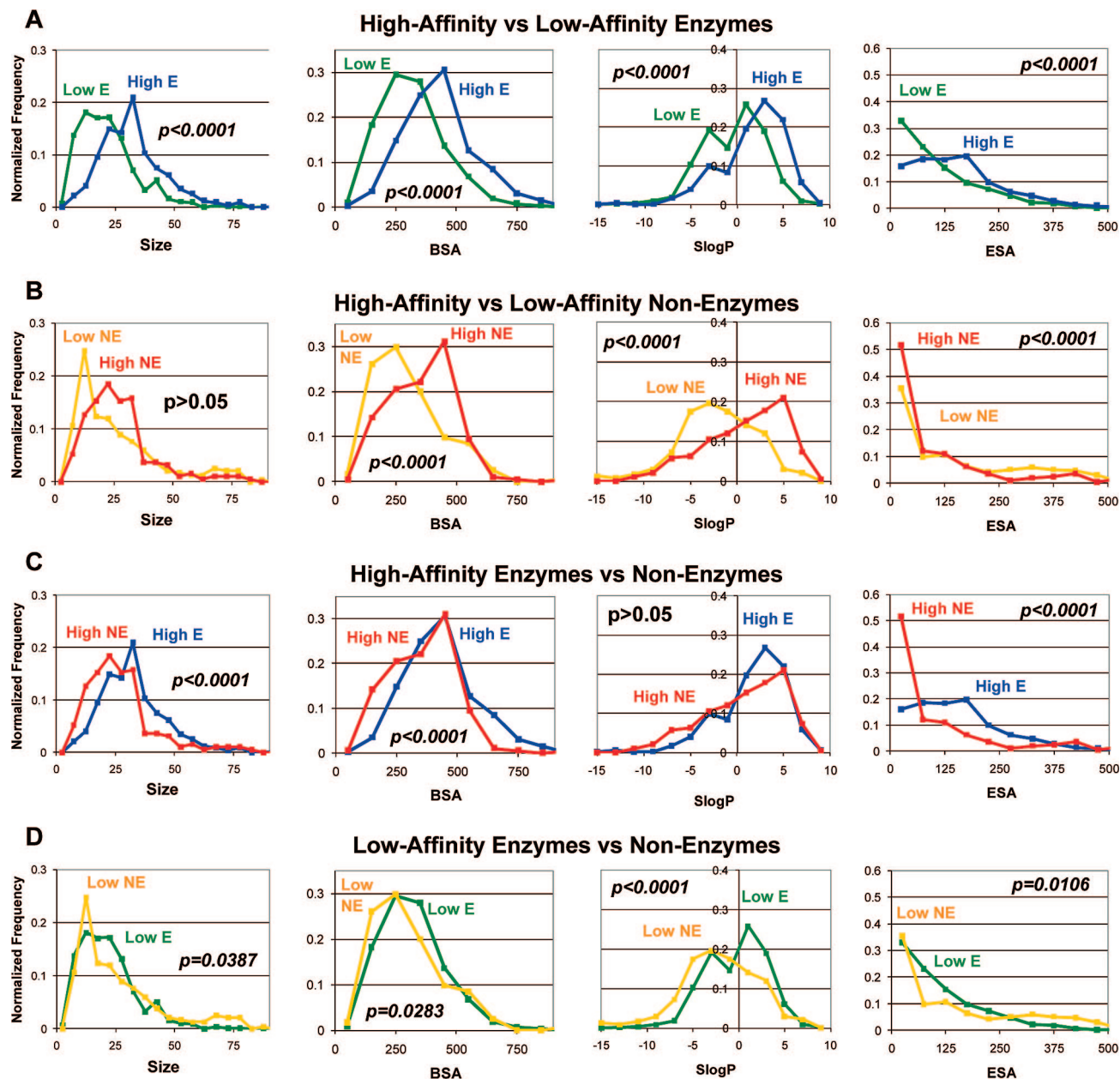
Sizes of the ligands point to a strong difference in the complexes, particularly in how to improve an inhibitor for enzymes versus nonenzymes. To improve the affinity of an enzyme inhibitor, it appears fruitful to add functional groups to increase the complementary contact between the inhibitor and the protein. In contrast, improving ligands for nonenzymes may best involve conservative changes that maintain the ligand's size. Tight binders for nonenzymes are less exposed than the low-affinity ligands, making them more sequestered from the surrounding solvent (Table 1). Distributions of the calculated octanol/water partition ratios (Figure 1a,b) show that high-affinity ligands are more hydrophobic than those with low affinity, but there is no significant difference between enzymes and nonenzymes in this regard. It appears that "adding grease" equally improves binding to both enzymes and nonenzymes, consistent with a general desolvation effect.<sup>7</sup>

The above trends for improving inhibitors for enzymes versus nonenzymes come from observing patterns across different proteins (interprotein relationships), but information to improve inhibitors for a specific target must come from observing trends of one protein binding a variety of ligands (intraprotein binding trends). This is a more difficult comparison to make because few proteins are crystallized with a significant range of bound ligands. For the few that exist, we must divide them into enzymes and nonenzymes, further reducing the sizes of the available data sets. The findings below are qualitative in nature.

Overall, our data show that enzymes appear to have better correlations between size and affinity than nonenzymes.

To determine a relationship between ligand size and affinity within a protein family (Figures 2 and 3), the complexes were grouped by 100% sequence identity. This organization ensures that changes in affinity are the result of changes in the ligand and not a mutation within the binding site. (For a few proteins, we were able to combine two sets when the mutations were far from the active site and inconsequential.) Groups that contained  $\geq 5$  complexes were examined. For nonenzymes, there were only a few proteins available: oligopeptide-binding protein, glutamate receptor 2, estrogen receptor  $\alpha$ , estrogen receptor  $\beta$ , arabinose-binding protein, mannose-binding protein, maltose-binding protein, and src SH2-binding domain. For most of the nonenzymes, the ligands are very similar in size and affinity. Six of the eight proteins have a small range of ligand sizes, which shows little correlation to affinity (Figure 2a,b). The small range of observed ligand sizes supports the idea that conservative changes are most appropriate for trying to improve ligands for nonenzymes. However, the lack of a distinct trend between ligand size and affinity does not necessarily prove that a trend could not be observed. It is unclear if the small range of ligands is the result of the specificity of the protein systems or whether more diverse complexes are simply not available from the PDB.

Only maltose-binding protein (Figure 2c) and the nonenzymatic site on the SH2 domain of <sup>PP60</sup>src tyrosine kinase (Figure 2d) have a significant range of ligand sizes. The maltose-binding protein complexes contain sugar chains of varying length. Almost all bind with roughly the same affinity, and this may be explained by the fact that the larger ligands show little difference in the BSA contact despite the very large range of sizes. The nonenzymatic site on the SH2 domain of <sup>PP60</sup>src tyrosine kinase is the only nonenzyme complex showing some correlation between ligand size and binding affinity. It is interesting that the only exception in nonenzymes is a regulatory site on an enzyme. These linear correlations reflect a trend across several ligands,  $\Delta(\Delta G_{\text{bind}}/\text{size})$ , which is slightly different than



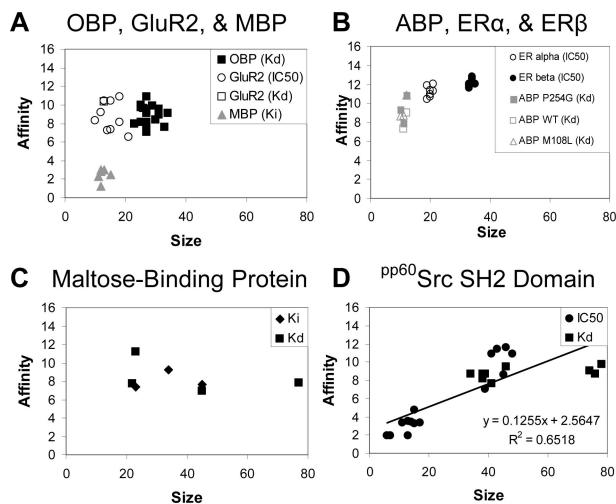
**Figure 1.** Comparisons of (A) enzyme complexes, (B) nonenzyme complexes, (C) high-affinity complexes, and (D) low-affinity complexes are presented. High-affinity enzymes are shown in dark blue, and low-affinity enzymes are in green. High-affinity nonenzymes are in red, and low-affinity nonenzymes are in gold. Distribution of ligand sizes (number of non-hydrogen atoms), buried surface area of the pocket ( $\text{\AA}^2$ ), SlogP, and exposed surface area ( $\text{\AA}^2$ ) are given in normalized percent frequencies. *P*-values show the significance of the difference in the medians of the distributions, as determined by a two-tailed Wilcoxon rank-sum evaluation (insignificant differences have  $p > 0.05$ ).

the ligand efficiency of an individual ligand,  $\Delta G_{\text{bind}}/\text{size}$ . In the discussions below, we will use the terms “trend” or “correlation” when comparing across several ligands bound to the same protein,  $\Delta(\Delta G_{\text{bind}}/\text{size})$ .

In the case of enzymes in MOAD, 37 proteins were available with five complexes or more. Unlike nonenzymes, over half of the families showed correlations between size and affinity. For brevity, only seven examples of MOAD’s enzymes are given in Figure 3. One of the most interesting features of the data in Figure 3 is that the slopes, the overall trend for each set, significantly vary! Although a linear correlation can be found for a good number of enzymes, the additive contributions of more functional groups appear to be system dependent, with some contributions being rather small. The trends range from 0.44 kcal/mol-atom for carboxypeptidase A (Figure 3b) to 0.09 kcal/mol-atom for FK506-binding protein (Figure 3f). Most

scoring functions use additive terms, and these findings underscore the difficulty in developing a universal scoring function appropriate for all protein systems. Yang et al. have also noted these difficulties in development of their M-Score scoring function.<sup>21</sup>

However, for 11 enzymes, there was no correlation; the ligands had roughly comparable affinity and sizes, much like the nonenzyme examples. Three enzymes showed a very small range of ligand sizes and a large range in binding affinity (Supporting Information). It is debatable whether these trends are exceptional examples of the correlation expected for enzymes or whether they indicate cases where only conservative changes in sizes are allowed, as would be expected for nonenzymes. It is also possible that they result from an unusual set of ligands from one chemical class.

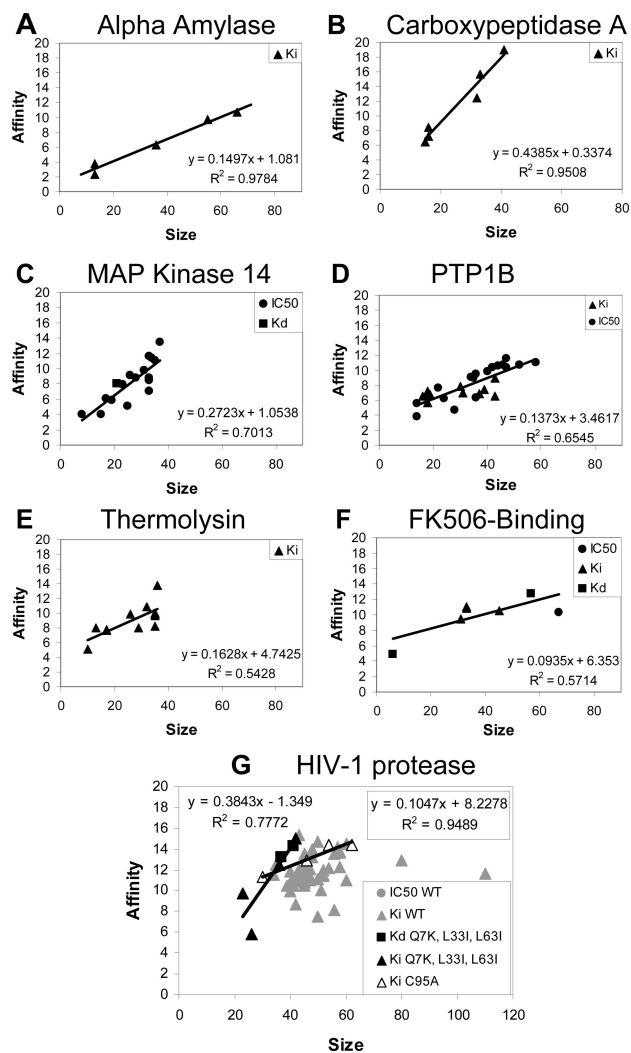


**Figure 2.** Limited correlation is seen between size and affinity in nonenzymes (A and B). The proteins with “clusters” of points have smaller binding sites and no ligands over 40 non-hydrogen atoms. The ligands have similar sizes and affinities for oligopeptide-binding protein (OBP), glutamate receptor 2 (GluR2), and mannose-binding protein (MBP), arabinose-binding protein (ABP), and estrogen receptors (ER)  $\alpha$  and  $\beta$ . The only nonenzymes with a range of ligand sizes are maltose-binding protein and the nonenzymatic site on the SH2 domain of pp60src tyrosine kinase (C and D, respectively).

Although Babaoglu and Shoichet have used fragments of inhibitors of  $\beta$ -lactamase to show that ligand efficiency is not necessarily additive within a binding site,<sup>22</sup> fragment-based design often couples these small building blocks in the pursuit of high-affinity ligands.<sup>23</sup> From our data above, one might expect greater success for this strategy when targeting enzymes where increasing size generally leads to increasing affinity. A recent study by Hajduk compared fragment-based design for 14 enzymes and four nonenzymes to show that ligand efficiency remained rather constant as the optimal leads were increased in size.<sup>24</sup> The contributions were roughly additive for the best functional groups. The average trend across these systems was 0.3 kcal/mol-atom, with individual systems showing trends from approximately 0.23 to 0.51 kcal/mol-atom (reported as binding efficiency indices of 11–28 pK<sub>d</sub> units per MW in kDa). It is encouraging that the values are comparable to the ligand efficiencies reported in Table 1.

Hajduk's trends were presented for the most efficient ligands for each protein, emphasizing the most ideal cases of improving a ligand.<sup>24</sup> However, his data for Bcl-xL, a nonenzyme with a large binding cleft, showed that many changes will not be optimal. A detailed analysis for >2300 additional molecules showed that many had significantly lower efficiencies. In fact, he suggests that chemical modifications that reduce the ligand efficiency by >10% deviate too much from the ideal and indicate that either the location or chemical nature of the modification is less desirable.

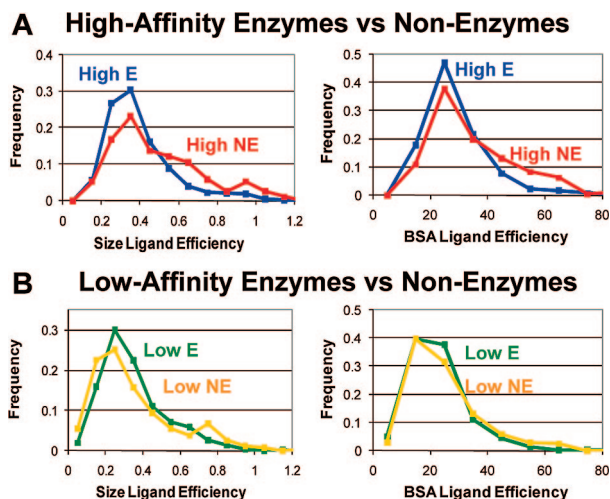
The HIV-1 protease data (Figure 3g) shows that there is a large scatter of inhibitor sizes and affinities, but two subsets of data (from mutants of HIV-1 protease) show strong linearity. This could demonstrate the same issue seen in Hajduk's detailed analysis of Bcl-xL.<sup>24</sup> The full set of data shows wide scatter and little trend, but a carefully chosen subset could reveal idealized trends for a particular protein system or class of ligands from a specific synthetic series. For HIV-1 protease, the compensation between enthalpy and entropy can be hard to control. Lafont et al. have demonstrated that an increase in size from the inhibitor (4*R*)-*N*-[(1*S*,2*R*)-2-hydroxy-2,3-dihydro-1*H*-



**Figure 3.** Many examples are available of enzyme complexes that show a strong correlation between size and affinity of the ligands; seven are given here (A–G). HIV-1 protease (G) demonstrates that a large collection of ligands may show no correlation, but subsets of data may reveal strong trends (data for the C95A and Q7K/L331/L63I mutants). It is interesting that even small binding sites with ligands of 40 non-hydrogen atoms or less (B,C,D) show a linear trend with affinity; this was not seen for nonenzymes with small binding sites.

inden-1-yl]-3-[(2*S*,3*S*)-2-hydroxy-3-[(2*R*)-2-(2-isoquinolin-5-yloxyethanoylamino)-3-methylsulfanyl-propanoyl]amino]-4-phenyl-butanoyl]-5,5-dimethyl-1,3-thiazolidine-4-carboxamide (KNI-10033, **1**) to the inhibitor (4*R*)-*N*-[(1*S*,2*R*)-2-hydroxy-2,3-dihydro-1*H*-inden-1-yl]-3-[(2*S*,3*S*)-2-hydroxy-3-[(2*R*)-2-(2-isoquinolin-5-yloxyethanoylamino)-3-methylsulfonyl-propanoyl]amino]-4-phenyl-butanoyl]-5,5-dimethyl-1,3-thiazolidine-4-carboxamide (KNI-10075, **2**) did not increase binding affinity despite a more favorable enthalpy from a strong hydrogen bond.<sup>25</sup> The entropic penalty of changing a thio ether (two heavy atoms) in **1** to a sulfonyl group **2** (four heavy atoms) is responsible for the lack of change in binding affinity. That study noted that, although others have been able to optimize certain HIV-1 protease inhibitors with respect to enthalpy, the enthalpy–entropy compensation could make optimization of affinity impossible for some chemical series.

An important caveat should be considered in the preceding discussion. It is possible that strong correlations between size and affinity can only be easily determined for large binding sites. Large ligands can be truncated to provide smaller, weaker ligands that bind to subsites. This would give a wide range of



**Figure 4.** Distribution of ligand efficiencies per size ( $-\text{kcal/mol-atom}$ ) and per contact ( $-\text{cal/mol-}\text{\AA}^2$ ), given in normalized percent frequencies. Distributions present comparisons of (A) high-affinity complexes ( $p < 0.0001$  in both cases) and (B) low-affinity complexes. High-affinity enzymes are shown in dark blue, and low-affinity enzymes are in green. High-affinity nonenzymes are in red, and low-affinity nonenzymes are in gold.

ligand sizes and affinities, allowing a definite size–affinity relationship to emerge from the data. It may be more difficult to determine a trend for a small binding site. This would still imply that enzyme inhibitors are more likely to be improved through the addition of functional groups simply because the binding sites in enzymes are generally larger than those of nonenzymes. However, if this were the case, the trend would be due to the size of the binding site and not necessarily the protein’s basic function.

Although the size argument above is important to note, it is most likely not the cause of the difference between enzymes and nonenzymes. Several examples of smaller binding sites, characterized by ligands of 40 non-hydrogen atoms or fewer, are presented in Figures 2 and 3. For small nonenzymes, there are no proteins that show a correlation between size and affinity. Conversely, there are several enzymes with small binding sites that do show a good correlation of increased affinity with increased size.

**Ligand Efficiencies.** Distributions of ligand efficiencies are given in Figure 4. Ligand efficiency based on contact ( $-\Delta G_{\text{bind}}/\text{BSA}$ ) can be compared to established values for the desolvation effect. The free energy of transferring a hydrophobic molecule from a hydrophobic solvent into water has been estimated as 24–47  $\text{cal/mol-}\text{\AA}^2$ , with the higher value being the most widely accepted.<sup>26–28</sup> Honig and co-workers have noted that this is lower than the value of 72  $\text{cal/mol-}\text{\AA}^2$ , derived from the surface tension of a hydrocarbon–water interface.<sup>28</sup> Only 0.8% of the complexes in this study have ligand efficiencies that exceed 72  $\text{cal/mol-}\text{\AA}^2$  (i.e., greater than Honig’s value), and many have efficiencies ranging between 20 and 40  $\text{cal/mol-}\text{\AA}^2$ . The low-affinity complexes are roughly bounded by the 47  $\text{cal/mol-}\text{\AA}^2$  value (only 4.1% have greater efficiencies), but the high-affinity complexes have large populations greater than that value. Although the complexes in Binding MOAD are not exclusively driven by hydrophobic association, these values provide a yardstick for comparisons. However, it should be noted that the range of values from the literature are based on SASA of small molecule ligands in differing environments, and our values are based on MSA of the contacts within the pockets. While the comparison is not ideal, MSA-based values

for ligands are not prevalent in the literature, and SASA of a pocket is not equivalent to SASA of a ligand.

For low-affinity complexes, the ligand efficiencies are basically the same for enzymes and nonenzymes (Table 1, Figure 4b). However, the differences are significant in high-affinity complexes ( $p < 0.0001$  for both efficiencies). The ligand efficiencies for high-affinity, nonenzyme complexes are  $\sim 17\%$  greater than those of high-affinity enzyme complexes (Table 1). Nonenzymes in Figure 4a show a broader distribution of efficiencies and much higher populations above 0.4  $\text{kcal/mol-atom}$  (55% of high-affinity nonenzyme complexes vs 37% of high-affinity enzyme complexes) and 30  $\text{cal/mol-}\text{\AA}^2$  (51% of nonenzymes vs 35% of enzymes). *On average over the high-affinity complexes, every atom and square Angstrom of buried cavity surface is worth more free energy in nonenzymes!*

The differences in efficiencies between high-affinity enzymes and nonenzymes are not dependent on the choice of cutoff between high- and low-affinity complexes. Even if the full set of enzymes is compared to the full set of nonenzymes, the ligand efficiencies are better for nonenzyme complexes. For the 1790 enzyme complexes, the median ligand efficiencies are 0.33  $\text{kcal/mol-atom}$  and 23  $\text{cal/mol-}\text{\AA}^2$ ; the median ligand efficiencies for the 424 nonenzymes are 0.36  $\text{kcal/mol-atom}$  and 26  $\text{cal/mol-}\text{\AA}^2$ .

The same patterns for enzymes and nonenzymes are observed when redundancy is removed (Supporting Information, Table S7, Figures S8 and S9). This is important because it corrects for some biases in the data set by using only one complex of a protein (some proteins have hundreds of entries and are heavily represented in the PDB). The nonredundant data set in Binding MOAD is obtained by grouping the proteins into families of 90% sequence identity and representing that family by the single complex with the highest-affinity ligand—in essence, the optimal binding event available for that individual protein. There are 688 unique complexes in this data set, 512 enzymes and 176 nonenzymes. Again, the high-affinity enzymes (235 complexes) have poorer ligand efficiency than the high-affinity nonenzymes (85 complexes). For the nonredundant data sets, the median ligand efficiencies for high-affinity enzyme complexes are 0.39  $\text{kcal/mol-atom}$  and 28  $\text{cal/mol-}\text{\AA}^2$ . The median ligand efficiencies for the nonredundant, high-affinity, nonenzyme complexes are still larger at 0.44  $\text{kcal/mol-atom}$  and 34  $\text{cal/mol-}\text{\AA}^2$ . The smaller number of complexes produces nearly identical distributions, and although the  $p$ -value of the comparison is slightly poorer ( $p = 0.04$ ), it is still significant (96%).

**Efficiencies, Evolution, and Druggability.** The significant differences in ligand efficiencies suggest a differentiation in the binding sites of these two classes of proteins based on their function. This may reflect the different evolutionary pressures upon enzymes and nonenzymes. The higher ligand efficiencies of nonenzymes make them, in essence, more responsive to low concentrations of ligand molecules. This is fitting, given their roles in signaling and regulatory control of cellular function in response to stimuli. Conversely, enzymes are optimized to bind molecules, change them, and release them again.

Ligand efficiencies are one key factor in describing the druggability of a target. Does this imply that nonenzymes may be more druggable? In general, higher ligand efficiencies mean that druglike affinities can be obtained with smaller molecules. Smaller molecules would tend to provide better oral absorption and fewer functional groups for toxicity concerns.<sup>10,29–31</sup> Of course, ligand efficiencies reflect “bindability”, and it is important to recognize that there are additional properties that make a protein a suitable drug target. It must be essential to

the disease state. Leads must show selectivity to avoid any negative consequences of off-target binding events. There are a myriad of ADME and pharmacokinetic properties to be considered. However, the differences in ligand efficiencies do indicate a greater likelihood to have better druglike properties for inhibitors, agonist, and antagonists of nonenzyme targets.

Many nonenzymes are the subject of intense drug discovery efforts in both the private and public sectors; for instance, hormone receptors, signaling proteins, and transcription regulators are targets for anticancer treatment.<sup>32,33</sup> Recent discussions on the druggability of protein–protein interfaces note that these difficult targets may be more amenable than originally thought.<sup>34,35</sup> Small molecules have been developed that bind to key hot-spot regions with greater efficiencies and deeper burial than the natural partner. Furthermore, many of the nonenzymes not represented in the PDB are membrane-bound receptors. Even though they are not included here, it is likely that the additional information would support the hypothesis that nonenzymes are more druggable because they are the target of many drugs. G-protein coupled receptors alone constitute 30% of the drugs on the market,<sup>30</sup> and genomic analysis has indicated many more receptors are druggable.<sup>36</sup>

Our results are also in good agreement with a recent study that estimated the druggability of 1096 nonredundant human proteins.<sup>10</sup> The predictions used a statistical model trained on NMR-screening data using a small fragment library.<sup>37</sup> Four of the top six classes were nonenzymes: vitamin-binding, steroid-binding, lipid-binding, and nucleotide-binding proteins.<sup>10</sup> The nonenzymes that were predicted to be the least druggable were large macromolecular complexes and are not reflected in Binding MOAD and this study.

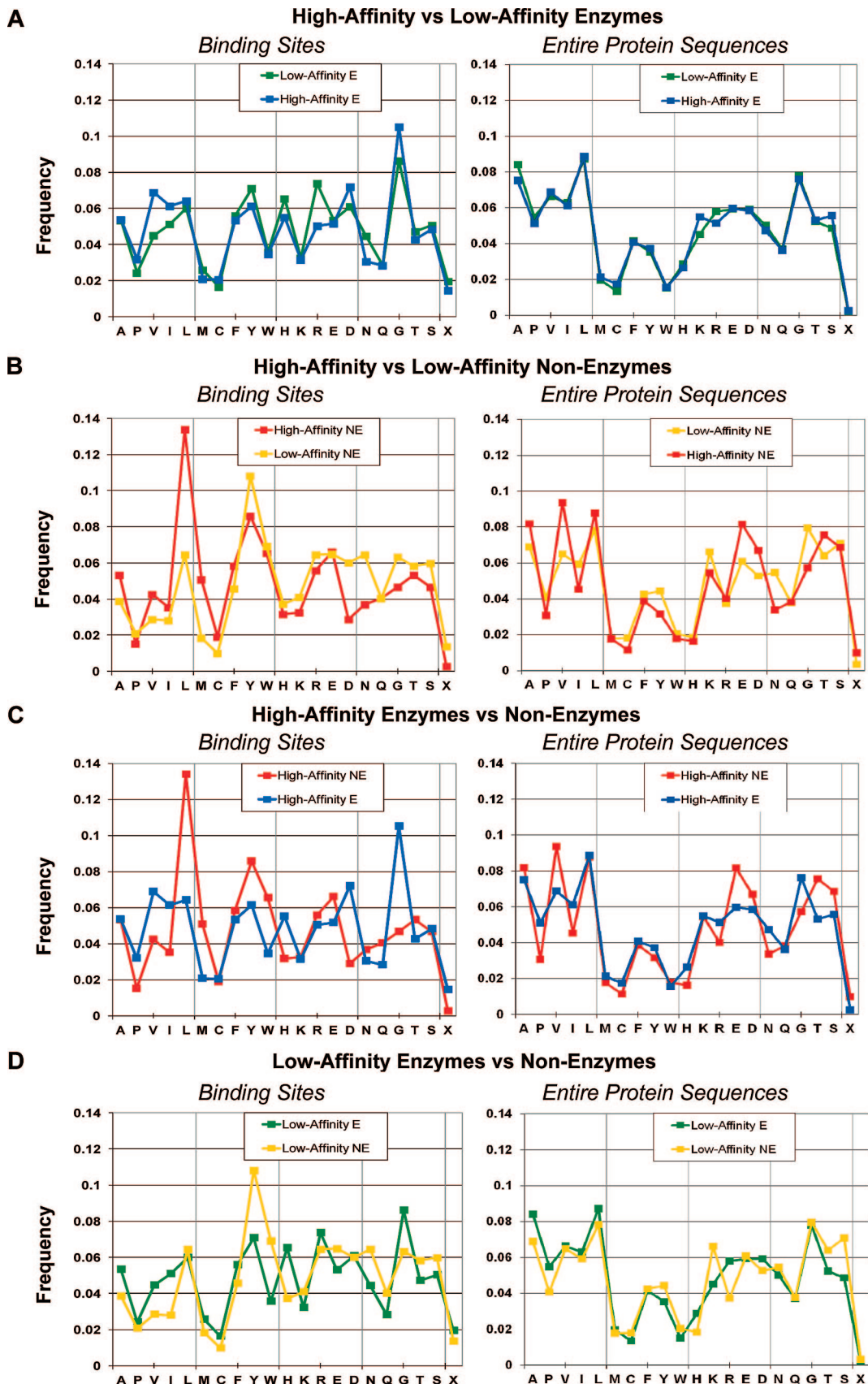
**What Produces the Higher Ligand Efficiencies in Nonenzymes?** Obviously, the root cause of the disparity in ligand efficiencies between enzymes and nonenzymes is of paramount interest. Although the ligands for nonenzymes are smaller, the SlogP characteristics are roughly the same for high-affinity ligands of enzymes and nonenzymes (Figure 1c). If the ligands are chemically similar, then the difference in efficiencies must come from the protein pocket. The most significant difference is the degree of exposure for ligands of nonenzymes versus enzymes. High-affinity ligands have a median exposure of only 11% in nonenzymes but 25% in enzymes (note that %ESA are used instead of ESA to correct for the difference in sizes of the ligands). Low-affinity ligands for nonenzymes are significantly more exposed (median of 33%), even more than the low-affinity ligands for enzymes (22%). Tight and weak inhibitors have the same degree of exposure in enzymes, but tight ligands for nonenzymes are much more encapsulated than the weak ligands ( $p < 0.0001$ ). Other 2D and 3D ligand descriptors displayed no significant patterns. This comparison was cognizant of correlations between characteristics; for instance, differences in surface area are correlated to size and were not “double counted” as additional differences between high-affinity ligands of enzymes vs nonenzymes.

Amino acid composition of the binding sites was examined (Figure 5, left column). There is little difference between the binding sites of high- and low-affinity enzyme complexes. The largest differences are an increase in Val content in high-affinity enzymes and an increase in Arg in the low-affinity complexes. For enzymes, the hydrophobic residues (Ala through Trp) in Figure 5 are 47.0% of the binding sites for high-affinity complexes but 43.9% for low-affinity ones. This is fitting with the aforementioned finding that the high-affinity ligands are slightly more hydrophobic. The comparison between binding

sites of high- and low-affinity nonenzyme complexes shows more pronounced variation but also holds the general pattern of high-affinity complexes having more hydrophobic content. The Ala-Trp residues are 55.9% of the binding sites for high-affinity complexes but 43.2% for low-affinity ones. What is most interesting is the comparison between enzymes and nonenzymes, particularly for the high-affinity complexes. The hydrophobic content is higher for nonenzymes (55.9% vs 47.0%), but the reader should recall that there is no significant difference in the SlogP of the ligands (in fact, the median value for nonenzymes is more hydrophilic). Why are more hydrophobic sites recognizing slightly more hydrophilic molecules with better affinity? The answer may lie in the fact that the amino acids making the contacts are significantly different. In high-affinity nonenzymes, Leu and Met provide a large portion of the hydrophobic contacts at the expense of Val and Ile. The nonenzyme's preference for Glu over Asp is reversed in high-affinity enzyme complexes, yet the use of Lys and Arg is the same. Leu, Met, and Glu are larger than their counterparts Val, Ile, and Asp. It is possible that those residues are slightly more polarizable. (Confirmation will have to come from in-depth examinations of fully modeled complexes, inclusive of added hydrogens, detailed atom typing, and possibly polarizable force fields. To do this for thousands of complexes is a sizable effort and outside the scope of the present study.) It should be noted that differences in the binding sites are not correlated with differences in the overall amino acid content; the reader should compare the left and right columns in Figure 5. Leu, Met, Phe, Tyr, and Trp make up nearly the same percentage of residues in the protein sequences but not the binding sites. This selective placement of differing residues within binding pockets may have direct relevance to analyses of hot-spot regions and potential binding sites on proteins.<sup>38–40</sup>

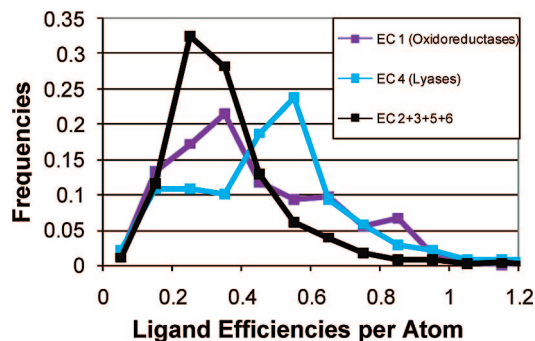
**Most Druggable Enzymes.** Of course, many pharmaceutically relevant targets are enzymes. By no means is it suggested that they are not appropriate drug targets, especially when they constitute 47% of the drugs on the market<sup>30</sup> and a large percentage of new targets identified through genomic analysis.<sup>36</sup> The distribution of ligand efficiencies for the enzyme classes suggests that lyases and oxidoreductases are the most druggable enzymes, Figure 6. The distribution of lyases is significantly shifted to higher efficiencies, standing out from the other data. The better efficiencies for oxidoreductases come from an increased population in the tail of the distribution. The median ligand efficiencies for the 139 lyases are 0.50 kcal/mol-atom and 33 cal/mol-Å<sup>2</sup>; and the median ligand efficiencies for the 256 oxidoreductases are 0.39 kcal/mol-atom and 26 cal/mol-Å<sup>2</sup>. The 1395 enzymes from the other four classes have median efficiencies of 0.31 kcal/mol-atom and 23 cal/mol-Å<sup>2</sup>, which are significantly lower (significance of  $\geq 99.99\%$  using the Wilcoxon test). It should be noted that the two enzymes that were predicted to be most druggable in the aforementioned study were also lyases and oxidoreductases, in that order.<sup>10</sup>

Recently, a new method was introduced to predict druggability of a binding site by estimating the site's maximum  $K_d$  based on the percent hydrophobic SASA and a scaling factor for efficiency that is dependent on the curvature of the site.<sup>9</sup> The model was trained on eight enzymes and applied to 63 structures, comprising complexes of 26 enzymes and a single structure of the nonenzyme mdm2.<sup>41</sup> An important goal of the study was to fit a predictive equation to assess druggability of a site based on protein–ligand structures of orally available compounds. This feature of the study is important to note because the contributions of various physical characteristics



**Figure 5.** The binding sites (left) and the entire protein sequences (right) are analyzed for amino acid content. Distributions are given in normalized percent frequencies. Amino acids within 4 Å of the ligands are considered to comprise the binding site. Distributions of (A and B) low- and high-affinity complexes of the same class show smaller differences than comparisons between enzymes and nonenzymes (C and D). Amino acids are listed by hydrophobic, aromatic, cationic, anionic, and hydrophilic nature. “X” denotes contacts with cofactors, unnatural amino acids, and covalent modifications on the protein.





**Figure 6.** Distribution of ligand efficiencies (–kcal/mol-atom) for enzymes, given in percent frequencies normalized for the different number of complexes in each enzyme class. The distribution of transferases (EC 2, 468 complexes), hydrolases (EC 3, 843 complexes), isomerases (EC 5, 60 complexes), and ligase (EC 6, 17 complexes) are the same and have been added together for this example (black line). Oxidoreductases (EC 1, purple line, 256 complexes) have larger populations in the higher efficiencies ( $p < 0.0001$ ). The distribution of lyases (EC 4, blue line, 139 complexes) is notably shifted ( $p < 0.0001$ ).

within the model should reflect both high-affinity binding and oral bioavailability of the ligand. The model was fit under the assumption that hydrophobic desolvation is the major driving force of binding, so terms based on electrostatics were not included. The model was able to properly rank the training set, noting that outliers were compounds with strong electrostatic components, prodrugs, or ligands that are actively transported. The model was then used to identify new, druggable structures from the PDB. It was interesting that the two newly identified targets were both enzymes. With only two new targets presented, it is not clear whether the model preferentially identifies enzymes over nonenzymes, but a preference toward enzymes may be expected from their model given the training and test sets used. Our data indicate that enzymes and nonenzymes may require different models in such analyses. Furthermore, many of the ligand efficiencies in our set exceed the established values for hydrophobic association, indicating that the most efficient complexes have additional factors that contribute to their affinity. The affinity of these complexes may not be well described by models based solely on hydrophobic SASA.

## Conclusion

We have presented a substantial mining study of Binding MOAD, the largest public database of curated protein–ligand structures with binding data. Physical characteristics of bound ligands were compared between enzymes and nonenzymes as well as high-affinity and low-affinity complexes. The comparison between ligand sizes for low-affinity versus high-affinity binding shows that divergent approaches are likely needed to improve the affinity of enzyme inhibitors versus those for nonenzymes. The traditional approach of adding functional groups to fill more of the pocket may work for enzymes, but it may not be as appropriate for nonenzyme systems. However, making ligands more hydrophobic appears to aid binding in both enzymes and nonenzymes.

Nonenzymes have higher ligand efficiencies than enzymes, which may be a reflection of their biological roles. This is also encouraging when considering the druggability of nonenzymes. In the pharmaceutical industry, ligand efficiencies have become a metric for evaluating hits from screening campaigns and even candidate compounds.<sup>12</sup> Our results would caution against applying a rigid standard across all protein targets. At the very

least, a cutoff based on ligand efficiency should differ between enzymes and nonenzymes. Ideally, cutoffs would differ between protein families and only be considered as one of several guidelines in a selection process.

Binding MOAD provides strong support of several mathematical models cited above,<sup>10,24,41</sup> particularly those of Hajduk and co-workers. Our results have implications for the development of scoring functions for docking and predicting druggability of a binding site.<sup>42–45</sup> The differences between nonenzymes and enzymes, as well as the differences across enzymatic systems, underscore the challenges of developing universal functions that perform well across all systems. Modest improvement might be achieved by developing separate functions for enzymes and nonenzymes, with even greater improvement expected for functions trained on specific protein families.

**Acknowledgment.** This work was funded by a Beckman Young Investigator Award, an NSF CAREER Award (MCB-0546073), and the National Institutes of Health (HG003890). R.D.S. thanks the Molecular Biophysics Training Program for support (GM008270). N.A.K. thanks the Training Program in Bioinformatics for support (GM070449). We thank Chemical Computing Group, Inc., for their generous donation of MOE for calculating the physical properties of the ligands. The authors greatly appreciate Prof. Bruce Mueller and Dr. Jignesh Patel for their invaluable help with statistical analysis and the SAS program. Thanks are also due to many other colleagues at the University of Michigan who have provided valuable feedback.

**Supporting Information Available:** Box plots and distribution analysis from JMP, Tukey–Kramer HSD analysis, patterns obtained from the nonredundant data set and the full data set without protein–cofactor complexes, data for the three enzymes with a large range in affinities for a small range of ligand sizes, listing of the high-affinity complexes. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. Binding MOAD (Mother Of All Databases). *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 333–340.
- Smith, R. D.; Hu, L.; Falkner, J. A.; Benson, M. L.; Nerothin, J. P.; Carlson, H. A. Exploring protein–ligand recognition with Binding MOAD. *J. Mol. Graphics Modell.* **2006**, *24*, 414–425.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26.
- Sugiyama, Y. Druggability: Selecting optimized drug candidates. *Drug Discovery Today* **2005**, *10*, 1577–1579.
- Norvell, J. C.; Machalek, A. Z. Structural genomics programs at the US National Institute of General Medical Sciences. *Nat. Struct. Biol.* **2000**, *7*, 931.
- Luque, I.; Freire, E. Structure-based prediction of binding affinities and molecular design of peptide ligands. *Methods Enzymol.* **1998**, *295*, 100–127.
- Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. Understanding noncovalent interactions: Ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors. *Angew. Chem., Int. Ed.* **2004**, *43*, 6596–6616.
- Coleman, R. G.; Salzberg, A. C.; Cheng, A. C. Structure-based identification of small molecule binding sites using a free energy model. *J. Chem. Inf. Model.* **2006**, *46*, 2631–2637.
- Hajduk, P. J.; Huth, J. R.; Tse, C. Predicting protein druggability. *Drug Discovery Today* **2005**, *10*, 1675–1682.
- Abad-Zapatero, C.; Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today* **2005**, *10*, 464–469.
- Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004**, *9*, 430–431.

- (13) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9997–10002.
- (14) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discovery* **2004**, *3*, 660–672.
- (15) *Molecular Operating Environment (MOE)*, 2007.08; Chemical Computing Group, Inc.: Montreal, 2007.
- (16) Wildman, S. A.; Crippen, G. M. Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873.
- (17) SAS, release 9.1; SAS Institute Inc.: Cary, NC., 2002–2003.
- (18) JMP, release 7.01; SAS Institute Inc.: Cary, NC, 2007.
- (19) Coleman, R. G.; Sharp, K. A. Travel depth, a new shape descriptor for macromolecules: application to ligand binding. *J. Mol. Biol.* **2006**, *362*, 441–458.
- (20) Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111–4119.
- (21) Yang, C.-Y.; Wang, R.; Wang, S. M-Score: A knowledge-based potential scoring function accounting for protein atom mobility. *J. Med. Chem.* **2006**, *49*, 5903–5911.
- (22) Babaoglu, K.; Shoichet, B. K. Deconstructing fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2006**, *2*, 720–723.
- (23) Carr, R. A.; Congreve, M.; Murray, C. W.; Rees, D. C. Fragment-based lead discovery: leads by design. *Drug Discovery Today* **2005**, *10*, 987–992.
- (24) Hajduk, P. J. Fragment-based drug design: how big is too big? *J. Med. Chem.* **2006**, *49*, 6972–6976.
- (25) Lafont, V.; Armstrong, A. A.; Ohtaka, H.; Kiso, Y.; Amzel, L. M.; Freire, E. Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chem. Biol. Drug Des.* **2007**, *69*, 413–422.
- (26) Chothia, C. Hydrophobic bonding and accessible surface area in proteins. *Nature* **1974**, *248*, 338–339.
- (27) De Young, L. R.; Dill, K. A. Partitioning of nonpolar solutes into bilayers and amorphous *n*-alkanes. *J. Phys. Chem.* **1990**, *94*, 801–809.
- (28) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* **1991**, *252*, 106–109.
- (29) An, J.; Totrov, M.; Abagyan, R. Comprehensive identification of “druggable” protein ligand binding sites. *Genome Inf.* **2004**, *15*, 31–41.
- (30) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discovery* **2002**, *1*, 727–730.
- (31) Kubinyi, H. Drug research: myths, hype and reality. *Nat. Rev. Drug Discovery* **2003**, *2*, 665–668.
- (32) Strachan, R. T.; Ferrara, G.; Roth, B. L. Screening the receptorome: an efficient approach for drug discovery and target validation. *Drug Discovery Today* **2006**, *11*, 708–716.
- (33) Whitty, A.; Kumaravel, G. Between a rock and a hard place. *Nat. Chem. Biol.* **2006**, *2*, 112–118.
- (34) Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* **2007**, *450*, 1001–1009.
- (35) Thanos, C. D.; DeLano, W. L.; Wells, J. A. Hot-spot mimicry of a cytokine receptor by a small molecule. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15422–15427.
- (36) Russ, A. P.; Lampel, S. The druggable genome: an update. *Drug Discovery Today* **2005**, *10*, 1607–1610.
- (37) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **2005**, *48*, 2518–2525.
- (38) Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280*, 1–9.
- (39) DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14–20.
- (40) Soga, S.; Shirai, H.; Kobori, M.; Hirayama, N. Use of amino acid composition to predict ligand-binding sites. *J. Chem. Inf. Model.* **2007**, *47*, 400–406.
- (41) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Souillard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **2007**, *25*, 71–75.
- (42) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
- (43) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Struct., Funct., Genetics* **2002**, *47*, 409–443.
- (44) Krovat, E. M.; Steindl, T.; Langer, T. Recent advances in docking and scoring. *Curr. Comput.-Aided Drug Des.* **2005**, *1*, 93–102.
- (45) Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. Docking: successes and challenges. *Curr. Pharm. Des.* **2005**, *11*, 323–333.

JM8006504