JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

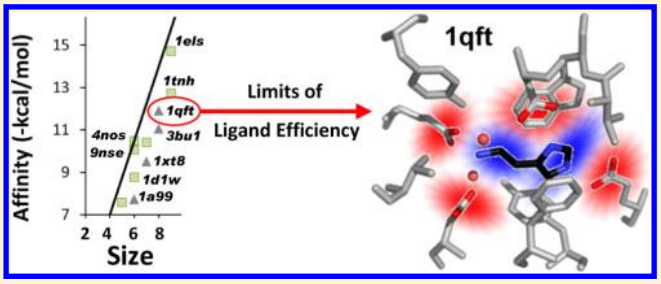
# Biophysical Limits of Protein−Ligand Binding

Richard D. Smith,[†] Alaina L. Engdahl,[‡] James B. Dunbar, Jr.,[†] and Heather A. Carlson*[,†]

[†]Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1065, United States

[‡]Department of Chemistry, Wittenberg University, Springfield, Ohio 45504, United States

**ABSTRACT:** In classic work, Kuntz et al. (*Proc. Nat. Acad. Sci. USA* **1999**, *96*, 9997−10002) introduced the concept of ligand efficiency. Though that study focused primarily on drug-like molecules, it also showed that metal binding led to the greatest ligand efficiencies. Here, the physical limits of binding are examined across the wide variety of small molecules in the Binding MOAD database. The complexes with the greatest ligand efficiencies share the trait of being small, charged ligands bound in highly charged, well buried binding sites. The limit of ligand efficiency is −1.75 kcal/mol·atom for the protein−ligand complexes within Binding MOAD, and 95% of the set have efficiencies below a "soft limit" of −0.83 kcal/mol·atom. On the basis of buried molecular surface area, the hard limit of ligand efficiency is −117 cal/mol·Å$^2$, which is in surprising agreement with the limit of macromolecule−protein binding. Close examination of the most efficient systems reveals their incredibly high efficiency is dictated by tight contacts between the charged groups of the ligand and the pocket. In fact, a misfit of 0.24 Å in the average contacts inherently decreases the maximum possible efficiency by at least 0.1 kcal/mol·atom.




## INTRODUCTION

Protein−ligand binding is a delicate balance between the loss of entropy resulting from complexation and the enthalpy gained by forming favorable contacts with the protein.[1,2] The precise contribution of these contacts is a source of debate and has provided a significant obstacle in the ability to predict how small molecules will bind.[3−5] The interplay between entropy and enthalpy is difficult to determine since they are influenced by several factors. For entropy, binding two entities results in a loss of six degrees of freedom, and a change in the internal flexibility of the protein and ligand must be taken into account. Furthermore, the reorganization of water around the ligand and within the binding site has significant implications. In the case of enthalpy, several types of contacts can be made to varying degrees in the binding site.[1] Current thinking is that van der Waals forces are the most significant factor for binding due to tight packing between the small molecule and protein.[1,6] Hydrogen-bonding and electrostatic interactions are thought to contribute more to the specificity of binding.[1] Since these interactions are also present with water and counterions in the unbound state, they are thought to have a smaller impact on affinity.[1]

Highlighting the different interpretations regarding the drive for efficient binding, there has been contradictory evidence as to which types of interactions play the most significant roles in the binding of biotin to streptavidin, the tightest known natural complex. In 1993, Miyamoto and Kollman used free energy perturbation on biotin·streptavidin to show that the increased binding affinity for the biotin·streptavidin system can be accounted for by van der Waals contacts made in the biotin·streptavidin complex where the pocket in streptavidin is preformed as in the traditional lock-and-key theory.[7] However, newer work has shown that networks of hydrogen bonds are responsible for the strong binding in the biotin·streptavidin complex.[8]

A common metric to evaluate a small molecule's ability to bind is "ligand efficiency". This metric is defined as binding affinity per number of non-hydrogen atoms.[9−11] It was first introduced by Kuntz et al. in 1999,[12] where they analyzed 159 tightest-binding complexes and the relationship between affinities and the number of heavy (non-hydrogen) atoms present in a ligand. They showed that each heavy atom can provide at most −1.5 kcal/mol of binding affinity.[12] This maximum was consistent with their predictions of the maximum affinity obtainable by van der Waals and hydrophobic interactions.[12] Though many of the most efficient ligands were metals and small ions, electrostatics was given little attention. Even in recent investigations, this class has been ignored because they are not "drug-like".[13−15] More recently efficiency metrics have been expanded to include affinity per squared angstroms of polar surface area (PSA).[16,17] AtlasCBS was developed to represent complexes based on pairs of efficiency indexes with the affinity per heavy atom or molecular weight being plotted versus the affinity per polar atom or affinity per PSA in order to map the "chemico-biological space (CBS)", of known complexes.[16,17] Efficiencies have also been calculated based on the entropy per heavy atom and enthalpy per heavy atom.[18]

In this study, we investigated which properties lead to an optimal efficiency and what defines the physical limits of binding. To study general patterns with regard to binding affinity and efficiency, it is necessary to use a large set of protein−ligand

**Received:** December 19, 2011

   dx.doi.org/10.1021/ci200612f | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

complexes for which a structure has been solved and an experimentally derived binding constant ($K_d$, $K_i$, or $IC_{50}$) has been determined. We used the largest data set available, Binding MOAD,[19,20] to explore the relationship between structure and binding affinity, extending Kuntz's examination to include all available binding events in the Protein Data Bank.[21] By looking at the most efficient ligands and the characteristics of their binding pockets, we reveal which interactions are most important to provide the highest binding affinity and efficiency. This study explores all binding events with the goal of examining fundamental biophysical properties, rather than focusing solely on properties of drug-like chemical space.

## METHODS

Structural properties were derived from the complexes in our 2010 release of Binding MOAD (Mother of All Databases), which is based on all PDB entries in 2009 and earlier.[19,20] Binding MOAD is the largest database of high-resolution protein−ligand complexes annotated with binding data from the PDB[21] (14 720 complexes comprised of 4624 unique protein families binding 7064 unique ligands). We have compiled binding affinity data for 32% of the entries (4782 complexes), with a preference for $K_d$ data over $K_i$ data over $IC_{50}$ data. For this study, no $IC_{50}$ data was used, so only the 2298 complexes with $K_d$ and $K_i$ data were considered. (Including the complexes with $IC_{50}$ data resulted in nearly identical values and did not alter the findings at all. However, their inclusion introduces too much uncertainty as the conversion to $\Delta G_{bind}$ is difficult to do accurately in many cases.) The free energy of binding was determined directly from $K_d$ values by $\Delta G_{bind} = RT \ln(K_d)$, and in the case $K_d$ was not available, we approximated the free energy of binding using $\Delta G_{bind} = RT \ln(K_i)$. All structures and affinity data are freely available at http://www.BindingMOAD.org.

Coordinates of the complexes were taken from the biological unit files provided by the PDB, which display the functional form of the protein. These files were processed to remove artifacts. We specifically focused on the size of the ligand and its contact surface with the protein, so any structure with poorly defined contacts were not considered. Therefore, we excluded structures with partially occupied or missing atoms from under-resolved ligands or side chains, as well as structures with extra atoms from ligands or side chains resolved in multiple orientations. A ligand was determined to have too many or too few atoms if the number of atoms in the formula did not match the number of atoms in the coordinate section of the pdb file.

Ligand efficiency is the free energy of binding divided by the number of non-hydrogen atoms in the ligand.[9−12] Hence, a ligand with 10 atoms is twice as efficient as a ligand with 20 atoms if they bind with the same affinity. In this study, ligand efficiencies are reported as affinity per size ($-\Delta G_{bind}$/atoms) and per degree of contact between the ligand and the pocket ($-\Delta G_{bind}$/BSA). PyMol[22] was used to make figures of the binding sites and calculate the electrostatics of the pockets using the APBS wizard.[23]

Surface areas were calculated using OPLS-based radii[24] with our code GoCAV which reports buried molecular surface area (BSA) of the pocket.[19] Variation in BSA occurs when several examples of ligand binding occur in the biological unit (i.e., slightly different interactions for three ligands in the three binding sites of a homotrimer). This variation is represented by error bars on the graph of BSA. The exposed surface area (ESA) is also computed from the total surface area minus the BSA.

## RESULTS AND DISCUSSION

**Maximum and Average Ligand Efficiencies.** If van der Waals terms are the definitive contribution, then we may expect to see a correlation between affinity and contact surface area between the protein and ligand. However, no correlation is seen between affinity and size or contact area (Figure 1A and B).

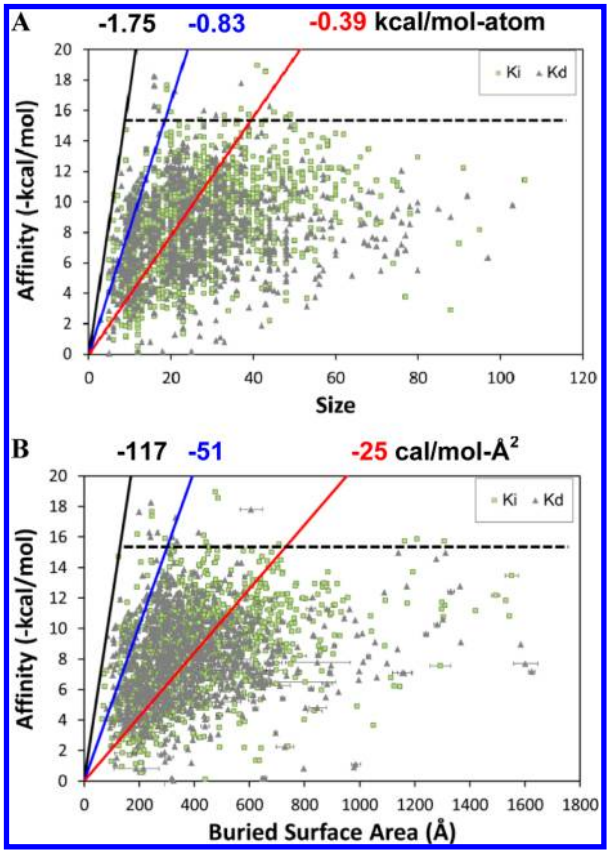



**Figure 1.** Plot of the affinity of the complexes versus their physical characteristics revealing the limiting cases as well as the general trends. Measurements used for affinity data are noted as $K_i$ (light green squares) or $K_d$ (gray diamonds). (A) Affinity versus size of the ligand, where size is given as the number of non-hydrogen atoms. (B) Affinity versus the buried surface area of the binding site. The "hard limits" of ligand efficiency are denoted with black lines and values; the "soft limits" which bound 95% of the data are denoted with blue lines and values; the average ligand efficiencies are given with red lines and values. The horizontal dashed black line denotes how few of the complexes have affinities greater than 15 kcal/mol.

Our data set is significantly larger than that of Kuntz et al.,[12] and we find a slightly higher maximal efficiency for ligands of −1.75 kcal/mol·atom. This "hard" limit is set by several systems, but we can also designate an alternative "soft" limit of −0.83 kcal/mol·atom, which is the upper bound of 95% of the data in Figure 1. The soft limit is roughly half the value of the hard limit, and it is rather surprising to see such a dramatic change over the 115 complexes that compose the top-5% most efficient complexes. The large drop in efficiency between the two limits highlights how rare it is to find exceptional ligands and suggests that −0.83 kcal/mol·atom is a sufficient limit for most uses. In particular it would be rare to create a drug molecule with ligand efficiency in excess of the soft limit.

The average and median efficiencies of our data set are −0.39 and −0.34 kcal/mol·atom, respectively. These averages are in


dx.doi.org/10.1021/ci200612f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

agreement with average values for ligand efficiency of −0.37 kcal/mol·atom for enzymes (median = −0.33 kcal/mol·atom) and −0.42 kcal/mol·atom for nonenzymes (median = −0.36 kcal/mol·atom), as reported in our previous work.[25] Accurate benchmarks for ligand efficiencies are very important because these values define physical limits of ligand binding. Furthermore, ligand efficiencies are often used to evaluate high throughput screening (HTS) data or to eliminate lead compounds during a drug development cycle.[9−11,26] Anecdotally, the best ligand efficiencies from HTS data approach −0.6 kcal/mol·atom.[9,26] Pushing for leads with ligand efficiencies near −0.3 or −0.4 kcal/mol·atom from a simple combinatorial library may be too restrictive for some systems as this is near the average for all good structures, as noted above.[10] However, ligand efficiencies of candidate compounds must often be higher to allow for changes during further drug development.[26,27]

We can also define ligand efficiency in terms of BSA of the binding site. Others have proposed metrics for ligand efficiency based on free energy of binding per surface area of the ligand, but these have been based on pharmacokinetic considerations and are not equivalent to contact surface area between the ligand and its protein target.[9−11] Recently, Nissink has proposed that the maximal ligand efficiency should be proportional to protein−ligand contact area and volume.[15] That work further suggests using a modified measure of ligand efficiency based on affinity/$N^{1/3}$. On the basis of a spherical approximation, the cubed root of $N$ estimates the ratio of area to volume of a ligand.[15] This metric is also useful for reducing the over emphasis of small ligand size that results from the traditional definition of ligand efficiency = affinity/$N$, where $N$ is the number of non-hydrogen atoms.[15]

Estimates based only on the ligand ignore a large portion of the interaction with the protein. Instead, we have chosen to measure the contacts directly. In our description based on the BSA of the binding site, the average efficiency is −25 cal/mol·Å$^2$ (median = −23 cal/mol·Å$^2$). Houk and co-workers coupled structure and affinity data for a moderate set of over 1000 host−guest, 175 antibody−antigen, and 176 enzyme−inhibitor complexes to propose that affinity is proportional to BSA of the ligand.[28,29] Their data implies a relationship, equivalent to 7 cal/mol·Å$^2$ (reported as approximately 1 log $K_a$ for every 90 Å$^2$ of buried surface). This average is approximately one-fourth of our average, but Houk's trend is for solvent accessible surface area of the ligand and ours is for molecular BSA of the binding site. Other reported values of the relationship of surface area versus free energy for transferring a hydrophobic solvent into water range from 24 to 47 cal/mol·Å$^2$,[30,31] which is in excellent agreement with the range between our average and soft-limit efficiencies.

In Figure 1B, the hard limit for efficiency is −117 cal/mol·Å$^2$ and the soft limit that bounds 95% of the data is −51 cal/mol·Å$^2$. We were surprised to find that the maximum efficiency with respect to BSA was in agreement with limits proposed for macromolecular binding.[32] In a follow-up work examining protein−protein, protein−RNA, and protein−DNA complexes, Brooijmans et al. established the limit of 120 cal/mol for every square angstrom of BSA.[32] Macromolecular recognition generally involves large, flat regions of a protein surface,[33] but despite that large contact surface, macromolecules do not inherently bind with higher affinities than small molecule ligands.[32] Keil et al. have shown that binding sites for ligands are deeper and more concave than binding sites for protein−DNA or protein−protein associations, implying a good degree of burial for small molecules despite their smaller size.[34] It is rather remarkable that the 120 cal/mol·Å$^2$ limit of binding efficiency appears to be universal across all varieties of binding interfaces on proteins.

**Electrostatic Interactions Define Maximal Efficiency.** Structures which define the limit of ligand efficiency share a distinct characteristic: 90% of the systems involve a charged ligand in contact with a charged protein residue or a metal ion cofactor. In fact, many of the ligands with the best efficiencies have two or three charge centers, and they are complemented in their binding sites by several charged side chains and/or dicationic ions. Figure 2 shows the systems with the maximum

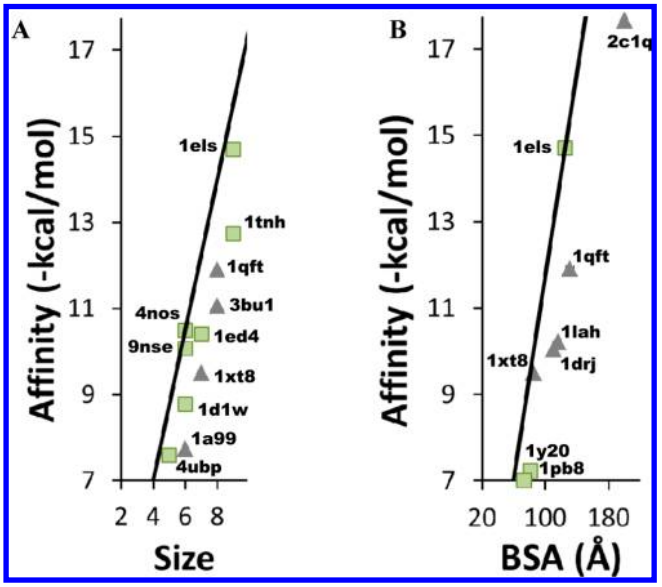



**Figure 2.** Close up view of the complexes with the highest ligand efficiencies. (A) Affinity compared to size as in Figure 1A. (B) Affinity compared to BSA as in Figure 1B. Complexes are labeled with their PDB codes.

efficiencies, annotated with their PDB codes. The highest efficiency is seen for a phosphonoacetohydroxamate compound with a −3 charge that is sandwiched between two dications in yeast enolase (PDB code 1els Figure 3).[35] The crystal structure shows several unusually tight contacts in the chelation (2.1 Å) which create very small contact surfaces. Not only is the small molecule bound by two manganese ions, there are two charged aspartates, two glutamates, two lysines, an arginine, and a histidine (that potentially could be charged) in the vicinity.

Other high efficiency complexes include a charged benzylamine coordinated to an acidic side chain in trypsin (1tnh),[36] a dicationic histamine complexed by four acidic side chains in tick histamine-binding protein (1qft),[37] a dicationic histamine which has charge−charge interactions with an acidic residue, and a cation-$\pi$ interaction with the aromatic ring of a Tyr residue (3bu1),[38] nitric oxide synthase binding +1-charged isothioureas (4nos, 1ed4, 1d1w, and 9nse—the natural substrate for this enzyme is arginine, which has a positively charged side chain and a zwitterionic core),[39,40] a zwitterionic cystine complexed by four charged side chains in the cystine transporter (1xt8),[41] a +2-charged 1,4-diaminobutane in the putrescine receptor (1a99),[42] and an anionic acetohydroxamic acid inhibitor sandwiched between two $Ni^{2+}$ in urease (4ubp).[43] Each of these binding sites can be viewed in Figure 3. Even though some of these structures contain metal ions and may be considered partially covalent by some, each structure in Binding MOAD has been verified to be


dx.doi.org/10.1021/ci200612f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

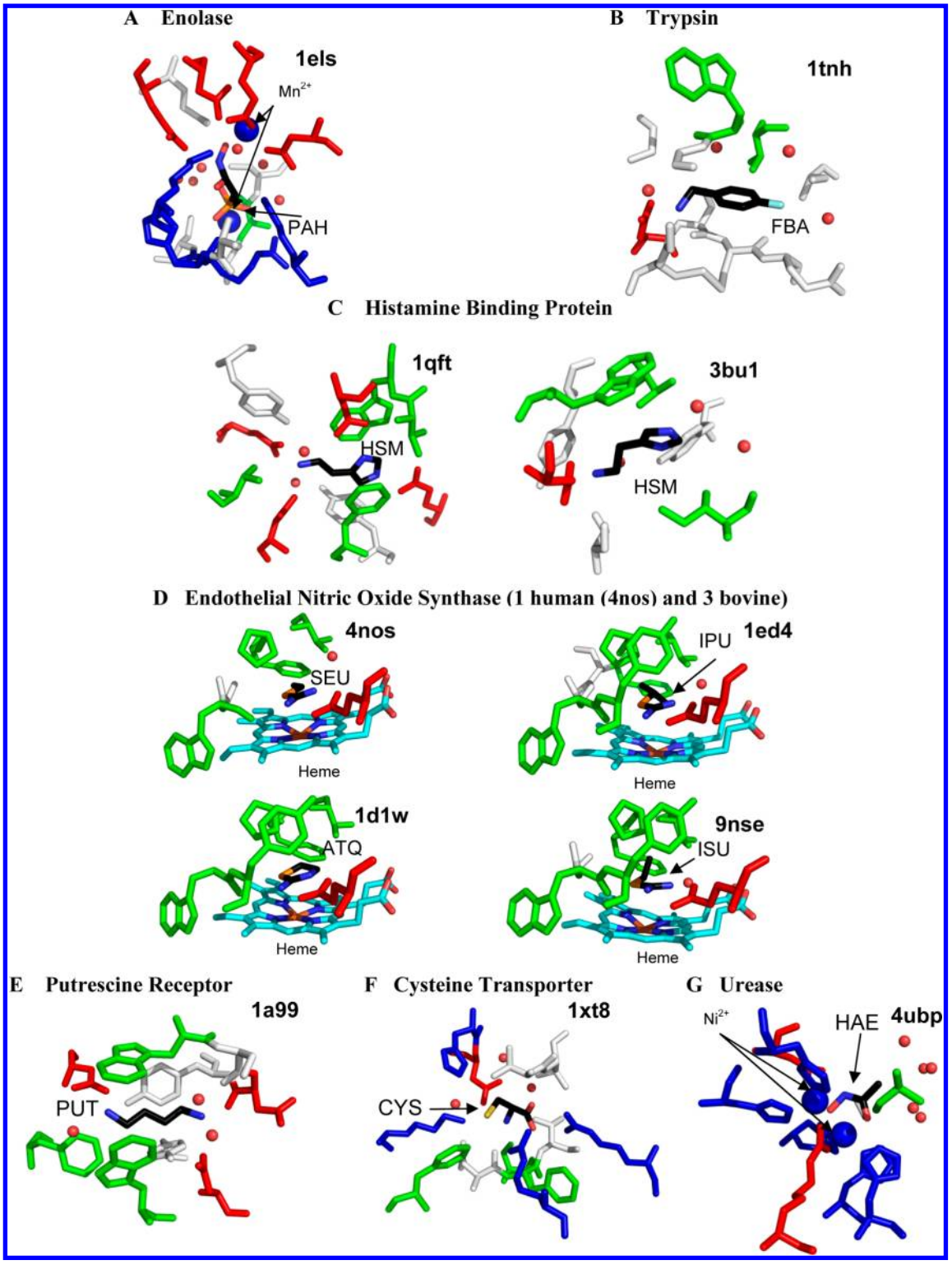



**Figure 3.** Binding sites of the 11 most efficient complexes. Figures show all residues within 4 Å of the small molecule ligand. The ligand is colored by atom type. The water is colored red and shown in small spheres. Metal ions are shown in larger blue spheres. Acidic residues (Asp and Glu) are colored red; basic residues (His, Lys, and Arg) are colored blue; hydrophobic residues (Ala, Ile, Leu, Met, Phe, Pro, and Val) are colored green; hydrophilic residues (Cys, Gly, Asn, Gln, Ser, and Thr) are colored white; and Tyr and Trp are colored either green or white depending on the interaction made with the ligand. The heme is colored with C = light blue and the iron = brown.

noncovalently bound, according to the primary citation listed in the PDB for the structure.[20]

Three of the structures that are at the maximum limit of efficiency per BSA (1els, 1qft, and 1xt8 in Figure 2B) are also seen at the limits of efficiency per atom in Figure 2A. Charged ligands are seen in most of the other systems in Figure 2B; see Figure 4. These additional complexes include 1lah (an ornithine with three charge sites bound to the lysine-arginine-ornithine-binding protein)[44] and 1y20 and 1pb8 (the glutamate NMDA receptor binding a zwitterion and D-serine, respectively).[45,46]


dx.doi.org/10.1021/ci200612f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

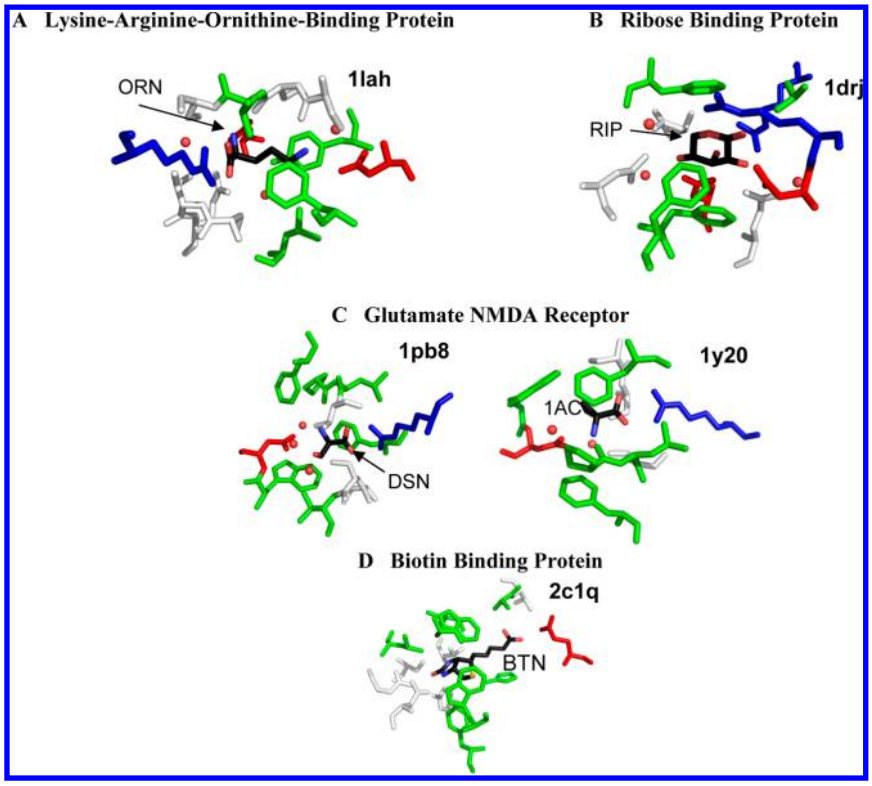



**Figure 4.** Binding sites of highly efficient complexes (affinity per buried cavity surface area). Figures show all residues within 4 Å of the small molecule ligand. The ligand is colored by atom type. Acidic residues are colored red; basic residues are colored blue; hydrophobic residues are colored green; and hydrophilic residues are colored white as in Figure 3. Water is colored red and shown in small spheres.

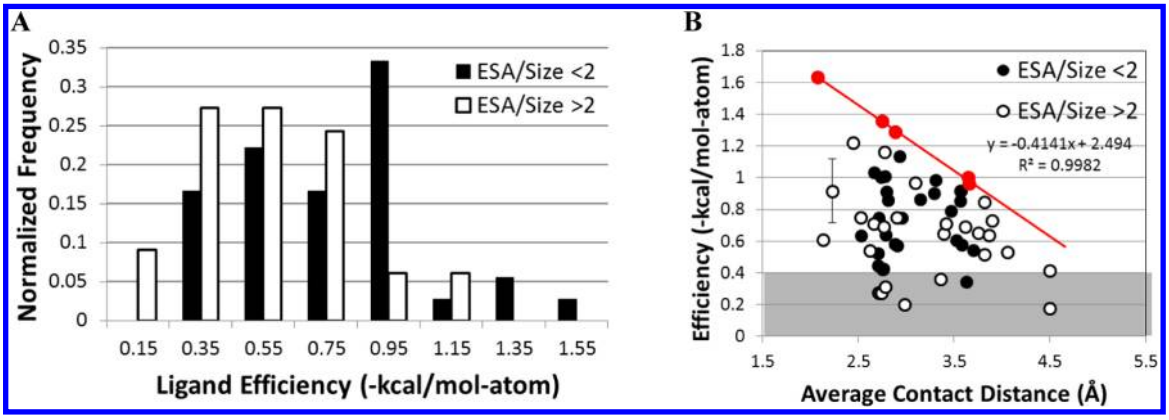



**Figure 5.** Relationship between efficiency, exposure, and protein contacts for ligands with 5−10 atoms and more than one charge site. (A) The distribution of efficiencies is compared for systems with well-buried (black) versus more exposed sites (white); a cutoff of 2 Å$^2$/atom is used to define the two sets. (B) Efficiencies are compared to the average contact distance between charged groups (black circles denote systems with ESA/size < 2 Å$^2$/atom, and white circles are ESA/size > 2 Å$^2$/atom). The line highlights the drop in maximal efficiency as the contacts become less favorable: roughly 0.41 kcal/mol·atom for every 1 Å increase in the average contact distance. The points used to set the line are colored red. The gray background notes systems with more modest efficiencies. The error bar indicates the standard deviation of the average of two affinity values reported in the literature.[70,71]

There are two structures involved in defining the maximal limits of ligand efficiency which do not have charge−charge interactions between the protein and the ligand. These structures are ribose bound to D-ribose-binding protein (1drj)[47] and a biotin-binding complex (2c1q).[48] Although the ligand is not charged in 1drj, the binding site in this structure contains four charged residues (two arginines and two aspartates) each making hydrogen bonds with the ribose (Figure 4B). Hydrogen bonds are very short and favorable for a hydroxyl group that glues a salt bridge together, and there are several pathways across the ribose that fill this role. This makes the packing in the site very tight and the BSA very low, which results in the exceptional ligand efficiency despite the modest affinity. As for 2c1q, biotin has a remarkably high binding affinity, but it is still not fully understood by the scientific community. We will not speculate reasons here.

On the basis of the known size dependence of ligand efficiency, it is not surprising that the best ligands are small, but it is important to note that not all small, charged molecules in charged binding sites are highly efficient. To determine why all charged ligands are not highly efficient, we examined all ligands with 5−10 non-hydrogen atoms and more than one charged site


dx.doi.org/10.1021/ci200612f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(61 complexes, of which only 9 have efficiencies of −0.4 kcal/mol·atom or weaker). Note that we are using a rather high cutoff here to define "less efficient binding"; drugs often have efficiencies near this value (a 1-nM ligand with ≥31 non-hydrogen atoms has ≥400 MW and an efficiency of −0.4 kcal/mol·atom or less). We are particularly interested in the electrostatic interactions of the ligand and pocket with respect to efficiency. The Coulombic potential is dependent on distance, so we examined the minimum distance of each charge of the ligand to the charged residues of the pocket, including any metal atoms that may be present in the binding site. The contacts were determined by investigating the structure of the complex visually using MOE.[49] We also calculated the exposed surface area (ESA) using our code GoCAV,[19] which bases ESA on the surface of the ligand that is not complemented by the protein. In this comparison, we chose to use ESA because it is a direct measure of the charged ligand's interaction with water, rather than use BSA which is a measure of interaction with the protein. Again, we have normalized this measure by the number of non-hydrogen atoms (ESA/size). Figure 5 presents the relationship of efficiency to these metrics. It should be noted that there are five systems that are not included in Figure 5 because they do not fit our definition of multiply charged. Though each has two titratable sites, all include an amine that is tightly coordinated to a metal cofactor, making it neutral. These systems have very short average contact distances, but very poor ligand efficiencies. The binding event must include a change in ionization, which is unfavorable and leads to reduced binding. To avoid confusion, these have been excluded.

There is a very significant difference (two-sided Wilcoxon $p$-value = 0.005) in the efficiencies of complexes that are less exposed (ESA/size < 2 Å$^2$/atom) versus those that are more exposed, Figure 5A. The 2 Å$^2$/atom cut is chosen since it is approximately 10% of the surface area of a water molecule. The median efficiency of well-buried ligands is −0.75 kcal/mol·atom versus a median efficiency of −0.57 kcal/mol·atom for those with ESA/size > 2 Å$^2$/atom (mean efficiencies are −0.79 versus −0.58 kcal/mol·atom, respectively). Furthermore, if those efficiencies are compared to the average distance between charged groups (Figure 5B), it appears that longer distances severely limit the maximum efficiency possible for the system. This is in keeping with Nissink's proposal that maximal efficiency should be proportional to contacts normalized for ligand size.[15]

For every increase of 1 Å in the average contact distance, the maximum efficiency drops by 0.41 kcal/mol·atom. Perhaps a more appropriate view is that a ligand's maximum efficiency is reduced by at least 0.1 kcal/mol·atom for a misfit as small as 0.24 Å in the average contacts between its charged groups and the protein's. For a 10-atom ligand, this small misfit reduces the affinity by 1 kcal/mol or more. Such significant gains/losses for such small displacements in the charges may explain why synthetic modifications to ligands that alter polarization and charge distribution can be so effective. The importance of charge interactions may support the ideas of optimizing charge complementarity that has been developed by Tidor and co-workers.[50−52] They developed an analytical solution of the Poisson−Boltzmann equation to model the electrostatics of the binding site and an analytical method of optimizing the charge profile of the ligand to match the calculated electrostatics of the binding site while also accounting for the desolvation penalty.[50−52]

The importance of charge complementarity in ligand binding can be supported by other biological binding events. The ability of salt bridges to improve the stability of protein−protein interactions in protein folding or protein−protein binding may be supportive.[53] Networks of salt bridges have been shown to stabilize proteins, although the majority of individual salt bridges have been shown to be destabilizing in proteins.[53,54] In a statistical study of 94 proteins from the PDB, Musafia et al. found that one-third of all residues participating in salt-bridges were involved in "complex" salt bridges, which they defined as ones involving three or more amino acids.[55] Olson et al. were able to stabilize $\alpha$-helical peptides by engineering multiple salt bridges and found that the amount of stability obtained was cooperative.[56] Networks of salt bridges were also found to be stabilizing by Kumar and Nussinov using continuum electrostatics to computationally determine the difference in energy of salt bridges compared to their hydrophobic isosteres, where the partial charges on the residue were set to zero.[57] They found that the stability for most salt bridges was determined largely by the desolvation penalty; however, the networked salt bridges were an exception to this phenomenon. In all cases, the networked salt bridge was found to be stabilizing, despite a large desolvation penalty.[57] These networked salt bridges are homologous to our charged ligands complemented by multiple charged residues in their binding sites. Furthermore, having a higher charge has been noted to be beneficial in metal ion binding to DNA/RNA. In these cases, the dicationic $Mg^{2+}$ is the preferred counterion, compared to $Na^+$, for binding to and stabilizing the phosphate backbone of the nucleic acid.[58]

Coulombic forces are the strongest nonbonded interactions that can be made, and it may not be surprising that highly efficient molecules utilize the strongest forces per atom. However, it is surprising that the free energy of binding is high in these complexes because the desolvation penalty for charged molecules is significant.[59] Their relatively high affinity indicates that the penalty must be compensated in some fashion. Almost all of the structures contain water in the binding sites (Figures 3 and 4), so not all of the water molecules are displaced. This reduces the desolvation penalty.

Another reason the desolvation penalty may be lower than initially thought is that water cannot completely solvate the charges in these systems. Many have ligands and pockets with charges that are closely spaced—too close for water to pack around each charge independently. It has been shown that the multiply charged phosphate backbone of DNA, which puts charges close together, leads to "frustrated water" around the DNA. The restructuring of water was determined to dominate the interaction of polyols with DNA.[60]

It has also been shown that contact between two oppositely charged particles are stabilized by concave surfaces.[61] Chorny et al. utilized molecular dynamics simulations with both explicit and implicit solvation models to create potentials of mean force for association of either a positively or negatively charged ion to an ion of opposite charge buried in a variety of hydrophobic surfaces representing multiple curvatures.[61] They found that the surfaces that had positive curvature, or were "receptor" like, stabilized the state in which the ions were in contact with each other. The authors suggested this was due to an image charge formed by the surface which effectively increases the electrostatic effect. Even planar surfaces and those with slight negative curvature stabilized the ion pair but required a water molecule to mediate the interaction.[61]

The limits of efficiency may be set by closely packed charged molecules because they are approaching covalent bonding. Zhang and Houk investigated 1017 enzymes-transition state

complexes as well as 160 enzyme/inhibitor complexes. They found that transition states, which tend to have covalent or partially covalent bonds to the protein, had affinities of $K_a = 10^{16}$ $M^{-1}$, while the inhibitors only bound with $K_a = 10^9$ $M^{-1}$. Additionally, they proposed that any enzyme proficiencies and affinities of greater than $10^{11}$ $M^{-1}$ (~15 kcal/mol) would exhibit covalent or partial covalent bonding.[28,62] At heavy atom distances less than 2.5 Å, low-barrier hydrogen bonds exhibit at least a partial covalent nature and provide stability of 10−20 kcal/mol.[63,64] Additionally, metals have the ability to exhibit coordinate-covalent bonding to ligands.[65] In a few highly efficient complexes, we observe distances less than 2.5 Å between atoms capable of hydrogen bonding, and some cases have metals involved in coordinating the ligand. We should note that we do not believe these systems to be overly influenced by partial bonding characteristics because all are reversibly bound, many with affinities in the micromolar and nanomolar range. Furthermore, in the NOS system (4nos, 1ed4, 1d1w, and 9nse) where the small molecule is near a heme, the distances to the iron are greater than 4 Å. Also, investigation of the available electron densities does not indicate partial bonding between the heme and ligand.

**Maximum Affinity of Ligands.** What defines the maximum binding affinity of ligands? Kuntz et al. found that binding affinity plateaus after ~15 atoms and little improvement is seen for larger ligands.[12] No ligand has a binding affinity of −20 kcal/mol or better. In fact, it is rare to exceed −15 kcal/mol (1.0% of the complexes in Figure 1). Kuntz and co-workers suggested that other biological factors may be the cause of the limit; for instance, molecules with too high of a binding affinity can exhibit clearance problems in the body.[12] Nature would tend to disfavor such molecules. Zhang and Houk might argue that affinities beyond −15 kcal/mol would require potential covalent binding.[28]

Affinities better than −15 kcal/mol are so tight that a ligand will most likely never dissociate before the protein is degraded. In an investigation of reported protein half-lives, stable protein half-lives range from 16 to 210 h with an average of ~105 h or just less than 4 and $^1/_2$ days, although there are some proteins that are degraded more rapidly.[66] In fact, the mean half-life of 3751 proteins in budding yeast is only 43 min.[67] The average half-life of 100 proteins in living human cancer cells was 9 h.[68] Therefore, it appears rare to find proteins with half-lives greater than 1 day. If we assume that $k_{on}$ is the rate of diffusion of ~$10^6$ 1/M·s, then an affinity of −15 kcal/mol ($K_d$ ~ 10 pM) would correspond to an average bound lifetime of ~1 day,[66,69] but at −16 or −18 kcal/mol, the lifetimes would be approximately 6 and 187 days, respectively. However, we do not agree that clearance issues limit binding because protein binding predates complex organisms. Instead, we hypothesize that once a ligand is bound for the lifetime of a protein (~ 1 day), there is no evolutionary pressure to coax ligands and proteins to associate more tightly. Therefore, affinities beyond −15 kcal/mol are serendipitously random or manmade. They require the artificial "selective pressure" of a scientist's goals.

Reynolds et al. have also discussed the plateau at −15 kcal/mol.[13] They noted that as size increased, the maximal efficiency would decrease. They suggested the reason for the drop in efficiency was that larger ligands would need to optimize a larger number of contacts with the protein that would lead to structural compromises and thus a reduced affinity.[13] We acknowledge that our data could also support this proposal because significant drops in efficiency can come from rather minor misfits in charge complementarity. Also, larger sites and ligands typically have more conformational freedom through more rotatable bonds; therefore, competing entropic penalties may further limit binding efficiency.

Several other factors may also contribute to the −15 kcal/mol limit to binding. First, assays which are used to determine binding constants have inherent limitations when measuring high affinity. We do not believe that this is the cause of the limit. If it was the cause, the distribution of binding affinities would drop off sharply as one approaches the limit, providing a Poisson distribution skewed toward tight binding. However, this is not the case. The distributions in MOAD follow a near-normal distribution centered at ~9 kcal/mol. Second, our study has the limitation of examining only proteins and ligands that can be crystallized, which may bias the analysis in unknown ways. Third, most of the high-affinity complexes are manmade compounds, and ADME/Tox issues for some targets may discourage pursuing molecules that bind with affinities better than low pM. Lastly, some affinities of greater than −15 kcal/mol may be incorrectly considered covalent.[28,62] We may see a limit because Binding MOAD does not contain covalently bound ligands.

## CONCLUSIONS

It is important to understand the limits of binding affinity and efficiency in order to properly describe the biophysics of protein−ligand molecular recognition. The difficulty in determining which interactions dominate free energy of binding has limited the ability of researchers to predict a priori which small molecules will bind to a target and how tightly. Our study and other recent studies have pointed to the importance of electrostatics in driving tight interactions.[50−52,55−57] We have looked at the most efficient protein−ligand complexes and have noted that the overwhelming majority of these complexes are small molecules that have at least one charge−charge interaction, and several of them have multiple charge interactions. Although desolvation of charged molecules is a barrier to binding, it appears that the small size of the ligands and the close proximity of their charges lead to water's inability to fully solvate the ligand and its binding site. Desolvation of the cramped, charged pockets should not be as difficult to overcome as it is with widely spaced charges, and many of the examined systems even retain some water in their sites, further reducing the desolvation penalty. We highlight the importance of tightly pairing the charges of the ligand and the pocket in order to reach the biophysical limits of binding. Ligand efficiency drops by at least 0.1 kcal/mol·atom for misfits as small as 0.24 Å in average contact between charged groups of the ligand and protein (i.e., a 10-atom ligand would drop affinity by 1 kcal/mol or more).

Several points should be considered for drug development. Of course, charged moieties are disfavored in drug development, and drug-like molecules are very different than the most efficient ligands examined here. We proposed that our soft limits of −0.83 kcal/mol·atom and −51 cal/mol·Å$^2$ are reasonable upper bounds for drug development, given the less drug-like character of the top 5%. On the basis of the average efficiencies, we also caution that pushing for ligand efficiencies in excess of −0.4 kcal/mol·atom from a simple combinatorial library will be too restrictive for many systems. Furthermore, we stress that optimizing charge complementarity may be more important for drug design than is currently stressed.

Lastly, we suggest that the ~15 kcal/mol limit of binding may be due to the fact that there is no evolutionary pressure to create tighter binding small molecules once the bound lifetime exceeds the lifetime of the protein. In fact, most ligands with affinities in

excess of −15 kcal/mol are manmade, pushed to extremes that are inaccessible by natural processes.

## AUTHOR INFORMATION


### Corresponding Author

*Phone: 1-734-615-6841. Fax: 734-763-2022. E-mail: carlsonh@umich.edu.


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work has been supported by a National Science Foundation CAREER Award to HAC (MCB 0546073). R.D.S. would like to thank the University of Michigan's Molecular Biophysics Training Program for support (NIGMS grant GM008270). A.L.E. was supported through the National Science Foundation Interdisciplinary Research Experiences for Undergraduates (REU) Program in the Structure and Function of Proteins at the University of Michigan's College of Pharmacy (0851723). The authors would also like to thank Nickolay A. Khazanov and Mark L. Benson for helpful discussions and codevelopment of Binding MOAD and its applications.

## REFERENCES


(1) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644−2676.

(2) 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.

(3) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **2007**, *371*, 1118−1134.

(4) Marsden, P. M.; Puvanendrampillai, D.; Mitchell, J. B. O.; Glen, R. C. Predicting protein-ligand binding affinities: a low scoring game? *Org. Biomol. Chem.* **2004**, *2*, 3267−3273.

(5) 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.

(6) Xu, J.; Deng, Q.; Chen, J.; Houk, K. N.; Bartek, J.; Hilvert, D.; Wilson, I. A. Evolution of shape complementarity and catalytic efficiency from a primordial antibody template. *Science* **1999**, *286*, 2345−2348.

(7) Miyamoto, S.; Kollman, P. A. What determines the strength of noncovalent association of ligands to proteins in aqueous solution? *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8402−8406.

(8) DeChancie, J.; Houk, K. N. The origins of femtomolar protein-ligand binding: hydrogen-bond cooperativity and desolvation energetics in the biotin-(strept)avidin binding site. *J. Am. Chem. Soc.* **2007**, *129*, 5419−5429.

(9) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discovery* **2004**, *3*, 660−672.

(10) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004**, *9*, 430−431.

(11) Abad-Zapatero, C.; Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today* **2005**, *10*, 464−469.

(12) 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.

(13) Reynolds, C. H.; Tounge, B. A.; Bembenek, S. D. Ligand binding efficiency: trends, physical basis, and implications. *J. Med. Chem.* **2008**, *51*, 2432−2438.

(14) Reynolds, C. H.; Bembenek, S. D.; Tounge, B. A. The role of molecular size in ligand efficiency. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4258−4261.

(15) Nissink, J. W. M. Simple size-independent measure of ligand efficiency. *J. Chem. Inf. Model.* **2009**, *49*, 1617−1622.

(16) Abad-Zapatero, C.; Blasi, D. Ligand Efficiency Indices (LEIs): More than a Simple Efficiency Yardstick. *Mol. Inf.* **2011**, *30*, 122−132.

(17) Abad-Zapatero, C.; Perisic, O.; Wass, J.; Bento, A. P.; Overington, J.; Al-Lazikani, B.; Johnson, M. E. Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation. *Drug Discovery Today* **2010**, *15*, 804−811.

(18) Reynolds, C. H.; Holloway, M. K. Thermodynamics of Ligand Binding and Efficiency. *ACS Med. Chem. Lett.* **2011**, *2*, 433−437.

(19) 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.

(20) Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. Binding MOAD (Mother Of All Databases). *Proteins* **2005**, *60*, 333−340.

(21) 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.

(22) Delano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific LLC:: San Carlos, CA, USA, 2002.

(23) Lerner, M. G.; Carlson, H. A. *APBS Plugin for PyMol*; Ann Arbor, MI, 2008.

(24) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(25) Carlson, H. A.; Smith, R. D.; Khazanov, N. A.; Kirchhoff, P. D.; Dunbar, J. B.; Benson, M. L. Differences between high- and low-affinity complexes of enzymes and nonenzymes. *J. Med. Chem.* **2008**, *51*, 6432−6441.

(26) Verdonk, M. L.; Rees, D. C. Group efficiency: a guideline for hits-to-leads chemistry. *ChemMedChem* **2008**, *3*, 1179−1180.

(27) Keserü, G. M.; Makara, G. M. The influence of lead discovery strategies on the properties of drug candidates. *Nat. Rev. Drug Discovery* **2009**, *8*, 203−212.

(28) Zhang, X.; Houk, K. N. Why enzymes are proficient catalysts: beyond the Pauling paradigm. *Acc. Chem. Res.* **2005**, *38*, 379−385.

(29) Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 4872−4897.

(30) DeYoung, L. R.; Dill, K. A. Partitioning of nonpolar solutes into bilayers and amorphous n-alkanes. *J. Phys. Chem.* **1990**, *94*, 801−809.

(31) Chothia, C. Hydrophobic bonding and accessible surface area in proteins. *Nature* **1974**, *248*, 338−339.

(32) Brooijmans, N.; Sharp, K. A.; Kuntz, I. D. Stability of macromolecular complexes. *Proteins* **2002**, *48*, 645−653.

(33) Deremble, C.; Lavery, R. Macromolecular recognition. *Curr. Opin. Struct. Biol.* **2005**, *15*, 171−175.

(34) Keil, M.; Exner, T. E.; Brickmann, J. Pattern recognition strategies for molecular surfaces: III. Binding site prediction with a neural network. *J. Comput. Chem.* **2004**, *25*, 779−789.

(35) Zhang, E.; Hatada, M.; Brewer, J. M.; Lebioda, L. Catalytic metal ion binding in enolase: the crystal structure of an enolase-Mn2+-phosphonoacetohydroxamate complex at 2.4-A resolution. *Biochemistry* **1994**, *33*, 6295−6300.

(36) Kurinov, I. V.; Harrison, R. W. Prediction of new serine proteinase inhibitors. *Nat. Struct. Biol.* **1994**, *1*, 735−743.

(37) Paesen, G. C.; Adams, P. L.; Harlos, K.; Nuttall, P. A.; Stuart, D. I. Tick histamine-binding proteins: isolation, cloning, and three-dimensional structure. *Mol. Cell* **1999**, *3*, 661−671.

(38) Mans, B. J.; Ribeiro, J. M.; Andersen, J. F. Structure, function, and evolution of biogenic amine-binding proteins in soft ticks. *J. Biol. Chem.* **2008**, *283*, 18721−18733.

(39) Fischmann, T. O.; Hruza, A.; Niu, X. D.; Fossetta, J. D.; Lunn, C. A.; Dolphin, E.; Prongay, A. J.; Reichert, P.; Lundell, D. J.; Narula, S. K.; Weber, P. C. Structural characterization of nitric oxide synthase isoforms reveals striking active-site conservation. *Nat. Struct. Biol.* **1999**, *6*, 233−242.

(40) Li, H.; Raman, C. S.; Martásek, P.; Král, V.; Masters, B. S.; Poulos, T. L. Mapping the active site polarity in structures of endothelial nitric

oxide synthase heme domain complexed with isothioureas. *J. Inorg. Biochem.* **2000**, *81*, 133−139.
(41) Müller, A.; Thomas, G. H.; Horler, R.; Brannigan, J. A.; Blagova, E.; Levdikov, V. M.; Fogg, M. J.; Wilson, K. S.; Wilkinson, A. J. An ATP-binding cassette-type cysteine transporter in Campylobacter jejuni inferred from the structure of an extracytoplasmic solute receptor protein. *Mol. Microbiol.* **2005**, *57*, 143−155.
(42) Vassylyev, D. G.; Tomitori, H.; Kashiwagi, K.; Morikawa, K.; Igarashi, K. Crystal structure and mutational analysis of the Escherichia coli putrescine receptor. Structural basis for substrate specificity. *J. Biol. Chem.* **1998**, *273*, 17604−17609.
(43) Benini, S.; Rypniewski, W. R.; Wilson, K. S.; Miletti, S.; Ciurli, S.; Mangani, S. The complex of Bacillus pasteurii urease with acetohydroxamate anion from X-ray data at 1.55 A resolution. *J. Biol. Inorg. Chem.* **2000**, *5*, 110−118.
(44) Oh, B. H.; Ames, G. F.; Kim, S. H. Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein. *J. Biol. Chem.* **1994**, *269*, 26323−26330.
(45) Furukawa, H.; Gouaux, E. Mechanisms of activation, inhibition and specificity: crystal structures of the NMDA receptor NR1 ligand-binding core. *EMBO J.* **2003**, *22*, 2873−2885.
(46) Inanobe, A.; Furukawa, H.; Gouaux, E. Mechanism of partial agonist action at the NR1 subunit of NMDA receptors. *Neuron* **2005**, *47*, 71−84.
(47) Björkman, A. J.; Binnie, R. A.; Zhang, H.; Cole, L. B.; Hermodson, M. A.; Mowbray, S. L. Probing protein-protein interactions. The ribose-binding protein in bacterial transport and chemotaxis. *J. Biol. Chem.* **1994**, *269*, 30206−30211.
(48) Hytonen, V. P.; Maatta, J. A.; Niskanen, E. A.; Huuskonen, J.; Helttunen, K. J.; Halling, K. K.; Nordlund, H. R.; Rissanen, K.; Johnson, M. S.; Salminen, T. A.; Kulomaa, M. S.; Laitinen, O. H.; Airenne, T. T. Structure and characterization of a novel chicken biotin-binding protein A (BBP-A). *B.M.C. Struct. Biol.* **2007**, *7*, 8.
(49) *Molecular Operating Environment (MOE)*, version 2010.10; Chemical Computing Group: Montreal, C.N., 2010.
(50) Chong, L. T.; Dempster, S. E.; Hendsch, Z. S.; Lee, L. P.; Tidor, B. Computation of electrostatic complements to proteins: a case of charge stabilized binding. *Protein Sci.* **1998**, *7*, 206−210.
(51) Lee, L. P.; Tidor, B. Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. *Protein Sci.* **2001**, *10*, 362−377.
(52) Kangas, E.; Tidor, B. Optimizing electrostatic affinity in ligand-receptor: Theory computation and ligand properties. *J. Chem. Phys.* **1998**, *109*, 7522−7545.
(53) Kumar, S.; Nussinov, R. Close-range electrostatic interactions in proteins. *ChemBioChem* **2002**, *3*, 604−617.
(54) Hendsch, Z. S.; Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **1994**, *3*, 211−226.
(55) Musafia, B.; Buchner, V.; Arad, D. Complex salt bridges in proteins: statistical analysis of structure and function. *J. Mol. Biol.* **1995**, *254*, 761−770.
(56) Olson, C. A.; Spek, E. J.; Shi, Z.; Vologodskii, A.; Kallenbach, N. R. Cooperative helix stabilization by complex Arg-Glu salt bridges. *Proteins* **2001**, *44*, 123−132.
(57) Kumar, S.; Nussinov, R. Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **1999**, *293*, 1241−1255.
(58) Tan, Z. J.; Chen, S. J. Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *J. Chem. Phys.* **2005**, *122*, 44903.
(59) Fonseca, T.; Ladanyi, B. M.; Hynes, J. T. Solvation free energies and solvent force constants. *J. Phys. Chem.* **1992**, *96*, 4085−4093.
(60) Stanley, C.; Rau, D. C. Preferential hydration of DNA: the magnitude and distance dependence of alcohol and polyol interactions. *Biophys. J.* **2006**, *91*, 912−920.
(61) Chorny, I.; Dill, K. A.; Jacobson, M. P. Surfaces affect ion pairing. *J. Phys. Chem. B* **2005**, *109*, 24056−24060.
(62) Smith, A. J.; Zhang, X.; Leach, A. G.; Houk, K. N. Beyond picomolar affinities: quantitative aspects of noncovalent and covalent binding of drugs to proteins. *J. Med. Chem.* **2009**, *52*, 225−233.
(63) Cleland, W. W.; Kreevoy, M. M. Low-barrier hydrogen bonds and enzymic catalysis. *Science* **1994**, *264*, 1887−1890.
(64) Schiøtt, B.; Iversen, B. B.; Madsen, G. K.; Larsen, F. K.; Bruice, T. C. On the electronic nature of low-barrier hydrogen bonds in enzymatic reactions. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12799−12802.
(65) Haaland, A. Covalent Versus Dative Bonds to Main Group Metals. *Angew. Chem., Int. Ed. Engl.* **1989**, *1989*, 992−1007.
(66) Rogers, S.; Wells, R.; Rechsteiner, M. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **1986**, *234*, 364−368.
(67) Belle, A.; Tanay, A.; Bitincka, L.; Shamir, R.; O'Shea, E. K. Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13004−13009.
(68) Eden, E.; Geva-Zatorsky, N.; Issaeva, I.; Cohen, A.; Dekel, E.; Danon, T.; Cohen, L.; Mayo, A.; Alon, U. Proteome half-life dynamics in living human cells. *Science* **2011**, *331*, 764−768.
(69) Corzo, J. Time the Forgotten Dimension of Ligand Binding Teaching. *Biochem. Mol. Biol. Edu.* **2006**, *34*, 413−416.
(70) Ash, D. E.; Emig, F. A.; Chowdhury, S. A.; Satoh, Y.; Schramm, V. L. Mammalian and avian liver phosphoenolpyruvate carboxykinase. Alternate substrates and inhibition by analogues of oxaloacetate. *J. Biol. Chem.* **1990**, *265*, 7377−7384.
(71) Stiffin, R. M.; Sullivan, S. M.; Carlson, G. M.; Holyoak, T. Differential inhibition of cytosolic PEPCK by substrate analogues. Kinetic and structural characterization of inhibitor recognition. *Biochemistry* **2008**, *47*, 2099−2109.