JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

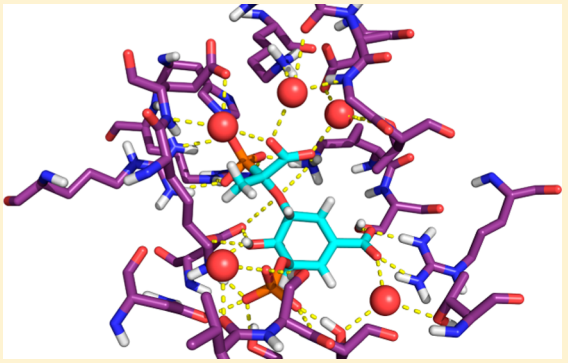
Hydration Properties of Ligands and Drugs in Protein Binding Sites: Tightly-Bound, Bridging Water Molecules and Their Effects and Consequences on Molecular Design Strategies

Alfonso T. García-Sosa*

Institute of Chemistry, University of Tartu, Ravila 14a, Tartu 50411, Estonia

Supporting Information

ABSTRACT: Some water molecules in binding sites are important for intermolecular interactions and stability. The way binding site explicit water molecules are dealt with affects the diversity and nature of designed ligand chemical structures and properties. The strategies commonly employed frequently assume that a gain in binding affinity will be achieved by their targeting or neglect. However, in the present work, 2332 high-resolution X-ray crystal structures of hydrated and nonhydrated, drug and nondrug compounds in biomolecular complexes with reported K_i or K_d show that compounds that use tightly bound, bridging water molecules are as potent as those that do not. The distribution of their energies, physicochemical properties, and ligand efficiency indices were compared for statistical significance, and the results were confirmed using 2000 permutation runs. Ligand cases were also split into agonists and antagonists, and crystal structure pairs with differing tightly bound water molecules were also compared. In addition, agonists and antagonists that use tightly bound water bridges are smaller, less lipophilic, and less planar; have deeper ligand efficiency indices; and in general, possess better physicochemical properties for further development. Therefore, tightly bound, bridging water molecules may in some cases be replaced and targeted as a strategy, though sometimes keeping them as bridges may be better from a pharmacodynamic perspective. The results suggest general indications on tightly hydrated and nontightly hydrated compounds in binding sites and practical considerations to adopt a strategy in drug and molecular design when faced with this special type of water molecules. There are also benefits of lower log P and better developability for tightly hydrated compounds, while stronger potency is not always required or beneficial. The hydrated binding site may be one of the many structure conformations available to the receptor, and different ligands will have a different ability to select either hydrated or nonhydrated receptor binding site conformations. Compounds may thus be designed, and if a tightly bound, bridging water molecule is observed in the binding site, attempts to replace it should only be made if the subsequent ligand modification would improve also its ligand efficiency, enthalpy, specificity, and pharmacokinetic properties. If the modification does succeed in replacing the tightly bound, bridging water molecule, it will have at least achieved benefits for ligand optimization and development independently of either positive or negative change in binding affinity outcome.



INTRODUCTION

Water plays a fundamental and decisive role in biomolecular association.[1−22] From the several types of hydrophobic effect[4,23] to discrete hydrogen bonded water networks,[1] water also plays an important and major part in the structures available to ligands and biomolecules.[24,25] Water molecules are found structurally in hidden (buried) pockets of biomolecules, as well as tightly associated with the biomolecular surface, bridging biomolecule-ligand interactions.[26] Water molecules can act as a "glue," or third partner, between the ligand and biomolecule, and they have been observed to dictate specificity,[27,28] as well as facilitate the binding of very different peptides to the same partner, i.e., nonselective binding to the OppA peptide binding protein.[29]

Explicit, tightly bound water molecules are observed in high resolution X-ray crystal structures.[2,26] Drug design projects routinely remove them from a biomolecule binding site, given the added complexity they represent. These water molecules define different volumes, shapes, and physicochemical properties of the binding sites. However, they can be involved in ligand design, either by conserving their hydration sites, or targeting them with ligand functional groups.[30,31] Water is also intrinsically involved in the energetics of biomolecular association since all binding partners require being solvated in a free state and are then either fully or partially desolvated when bound. The hydrophobic effect describes the contribution to binding energy from the favorable desolvation of hydrophobic surfaces, and their subsequent attachment.[32] The contributions to binding energy have an enthalpic and entropic component. The estimated entropy gain from removing one water molecule from a binding site to the bulk liquid water has been estimated from crystal structures of salts to be from 0 to not more than 2 kcal/mol.[33] As

Received: December 5, 2012
Published: May 13, 2013

   dx.doi.org/10.1021/ci3005786 | J. Chem. Inf. Model. 2013, 53, 1388−1405

a comparison, one hydrogen bond may provide an energy gain of up to 6 kcal/mol,[34] though desolvation (including cavity formation) and conformational effects also affect the magnitude of energetic contributions. A buried hydrogen bond has been suggested to positively (favorably) contribute around 1 kcal/mol to the overall binding energy of a biomolecule,[35] or near 2 kcal/mol in model cavities composed of small molecules.[36] Therefore, even if some water molecules would experience a decrease in entropy transferred from bulk into a ternary association with a ligand and biomolecule, this is not always the case, and some water molecules have a much more reduced entropy, balanced by an even deeper enthalpy, close to that of ice.[37] Some water molecules are thus tightly bound to the biomolecule and/or ligand and will have different properties than loosely bound waters, or buried water molecules.[26] Several ways of describing tight as opposed to loosely bound water molecules exist: Consolv using k-nearest neighbors,[38] WaterScore using multivariate logistic regression of physicochemical properties of discrete and explicit water molecules,[26] Bayesian classifiers on free energy calculations,[39] as well as calculation of local enthalpy and entropy of waters with MD simulations.[19,40−42] In WaterMap, those waters that have less local enthalpy and energy compared to bulk waters are used to position ligand atoms.[19,41−43] In fact, the transition of water molecules in a binding site from loose to tightly ordered can be the driving force in enthalpy and energy changes in a biomolecular association.[44] A delicate interplay of effects occurs between biomolecule−water−ligand systems that include differences in bulk solvent and in the binding site, in addition to conformational states.[45] Taking account of these effects and their change due to modification of a ligand with chemically relevant series of functional groups may be used to design compounds.[46] A classical example of tightly bound molecules having been targeted with ligand functional groups is the case of cyclical urea inhibitors of HIV-1 protease, where a larger hydrophobic effect and lower ligand entropy aided in increasing affinity.[47] However, they may be included in pharmacophores,[48] docking,[49] and *de novo* design,[30,31] among others.[50−84]

The solvation properties and preferences for amino acid groups of water molecules can also change with temperature.[13] Explicit water has also been found to be required to describe and predict accurately the nonpolar solvation free energy,[18] which is often calculated solely with continuum methods including surface areas or volumes. Water molecules in the binding site, even in a solvent-exposed binding site, were thus found to have distinct properties from bulk water. The strategies commonly employed for using explicit water molecules frequently assume a gain in binding affinity will be achieved by their explicit targeting or implicit neglect. This is not always the case,[85−89] and many other effects may be in play.[4,46] In addition, stronger potency is not always required nor beneficial (not even from a putative lower dosing perspective), especially when stronger binding affinity comes from increases in entropy that contribute to a lack of specificity and molecular obesity.[90]

There are two ways of studying protein binding data. One includes experimentally or computationally addressing the enthalpic and entropic changes upon binding, including water. The other way of studying binding phenomena is through statistical analysis of large data sets of protein−ligand data.[26,91] These strategies dealing with water molecules have an influence on the chemical compounds designed and their features. Conserving or targeting tightly bound, bridging water molecules has been observed to modify the shape, volume, and locations of hydrogen bond partners in the binding site, as well as the chemical diversity of the de novo designed ligands.[30,31] In the present work, a statistical approach has been employed to study the properties of biomolecule−ligand associations in high resolution X-ray complexes, including their binding affinity, physicochemical properties of ligands, and their associated ligand efficiencies,[92−98] as well as tightly bound, bridging water molecules. The results suggest general indications on hydrated and nonhydrated compounds in binding sites and practical considerations when adopting a strategy when faced with this special type of water molecule.

METHODS

Structures of biomolecular-ligand complexes were obtained from the largest set of complexes (called "general index") in the PDBBind database.[99] The binding affinity, ΔG_{bind}, was calculated using inhibition, K_i, or dissociation, K_d, values,[100−102] using T = 300 K, through $\Delta G_{bind} = -RT \ln K_i = RT \ln K_d$, where R is the ideal gas constant. High resolution X-ray crystal structures (resolution ≤2.5 Å) were selected, since these have more confidence in their structure determination and may provide less uncertainty about the position or assignment of electron cloud density to water molecules.[103]

Programs were written in python to count the number of tightly bound, bridging water molecules in the binding site, using WaterScore,[26] that have parameters of crystallographic motion or temperature (B-factors, also called Debye−Waller factors, in Å^2), number of protein contacts within 3.5 Å, and solvent contact surface area (in Å^2).

Box plots, densities, and statistical tests were carried out using the statistical computing package R.[104] t tests were computed using Welch, two-sided, unequal variance, two-sample tests using the null hypothesis that both distributions could be the same (i.e., their differences due to random variation), using a 95% confidence level ($p < 0.05$) against the alternative hypothesis that the true difference in means of both groups is not zero. These tests were also confirmed or challenged by permutation, two-sided, independent, unequal variance t tests using 2000 simulations, and testing for statistical significance at the same confidence level ($p < 0.05$). By taking all possible samples of values of one group from the total of values of both groups, the permutation distribution is obtained.

That status of drug compounds was checked in the DrugBank database[105] and the Protein DataBank (PDB).[106] Physicochemical properties of ligands were calculated using XLOGP[107] for the logarithm of the octanol/water partition coefficient (log $P_{octanol/water}$), and Marvin Beans[108] for molecular mass (MW), number of heavy atoms (NHA), number of carbons (NoC), number of atoms, number of hydrogens, number of bonds, number of rings, aliphatic rings, aromatic rings, aromatic atoms, hydrogen bond donors, hydrogen bond acceptors, rotatable bonds, molecular surface area (MSA), polar surface area (PSA), molecular polarizability, Wiener index, Balaban index, Harary index, hyper-Wiener index, Platt index, Randic index, Szeged index, and Wiener polarity.

Ligand efficiencies (LEI) were calculated by dividing binding energy by a normalization factor (NF) as ΔG_{bind}/NF, where NF is one of MW, NHA, NoC, MSA, PSA, or Wiener, as well as a lipophilic efficiency index:[97] $\log(-\Delta G_{bind}/P)$.

dx.doi.org/10.1021/ci3005786 | J. Chem. Inf. Model. 2013, 53, 1388−1405

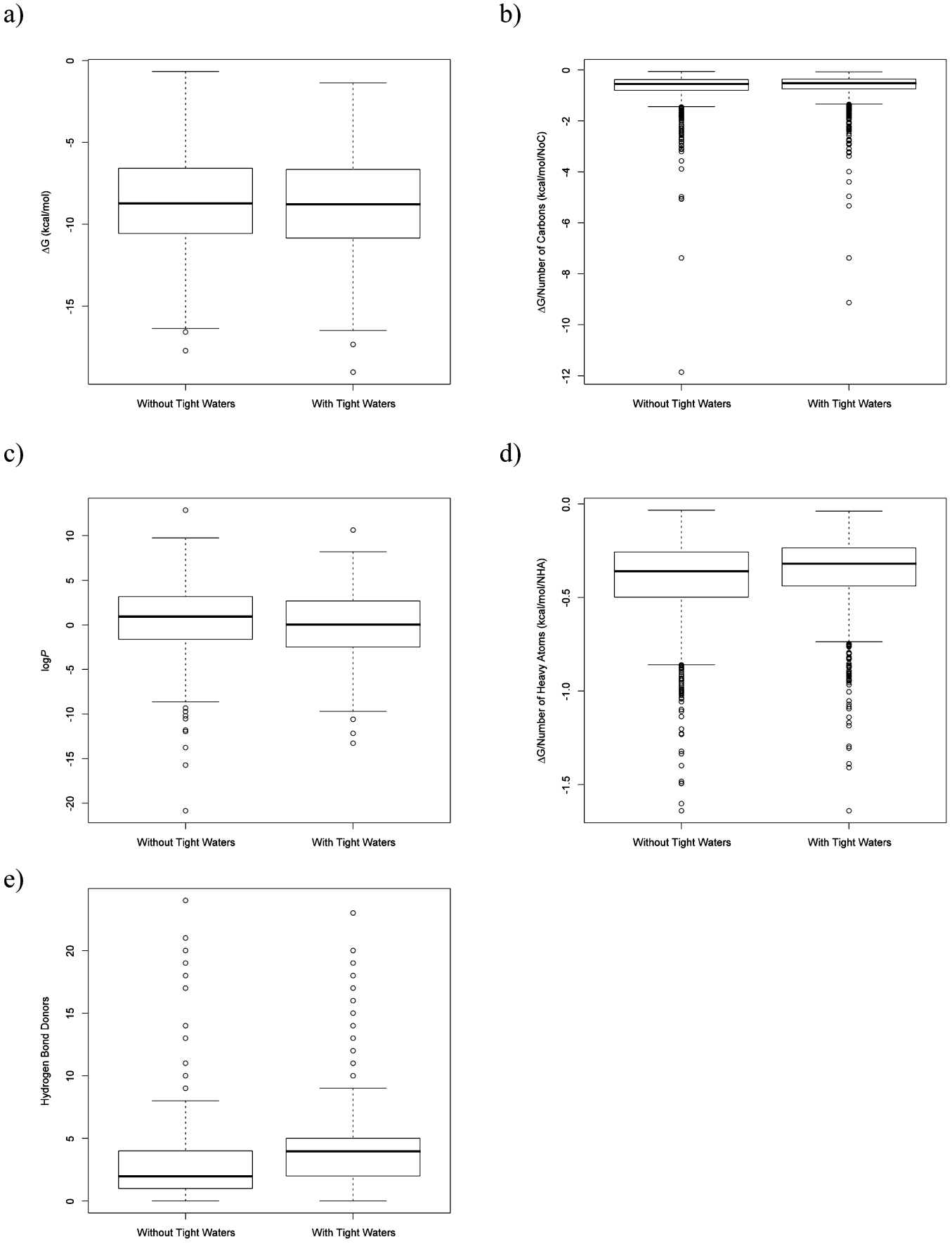


Figure 1. Box plots for comparisons between tightly hydrated and nontightly hydrated compounds: (a) ΔG_{bind}, (b) ΔG_{bind}/NoC, (c) log P, (d) ΔG_{bind}/NHA, (e) hydrogen bond donors.

dx.doi.org/10.1021/ci3005786 | J. Chem. Inf. Model. 2013, 53, 1388−1405

RESULTS AND DISCUSSION

From the total amount of 5678 entries in the largest (general) collection of the PDBbind,[99] only those that had a reported K_i or K_d value as well as biomolecule and ligand structure from high-resolution (lower than 2.5 Å) X-ray crystallography determination were retained. This left 2332 high quality complexes for study.

High Resolution X-Ray Structures. From the 2332 complexes, a program written to identify tightly bound and bridging water molecules in complexes' binding sites using previously determined criteria of a high number of contacts, low B-factor, and low solvent contact surface area[26] obtained 1119 complexes including them, and 1213 complexes without.

The distributions for the properties for each class of compound were studied, and a selection of comparisons is presented in Figure 1.

For all of these properties, two-sided, unequal variance, Welch *t* tests were carried out to determine those properties whose difference between their distributions in the tightly hydrated versus the nontightly hydrated complexes could have arisen through random chance (null hypothesis), or that there was a statistical significant difference at the 95% confidence level between the two. These results are summarized in Table 1.

Table 1. Statistically Significant Differences between Compounds with Tightly Bound, Bridging Water Molecules and Compounds without Water Bridges

property	without tightly bound waters	with tightly bound waters
• ΔG_{bind}, ΔG_{bind}/NoC, aliphatic rings, aromatic rings, aromatic atoms	statistically indistinguishable	
• ΔG_{bind}/NHA, ΔG_{bind}/MW, ΔG_{bind}/MSA, ΔG_{bind}/PSA, ΔG_{bind}/Wiener, hydrogen bond donors, hydrogen bond acceptors, Wiener index	lower	
• log *P*		lower

A superposition of the density for ΔG_{bind}, ΔG_{bind}/NoC, log *P*, ΔG_{bind}/NHA, and hydrogen bond donors can be seen in Figure 2.

Of special importance from these measurements is the indication that ΔG_{bind} is essentially the same for tightly hydrated and nontightly hydrated complexes. This would imply that there would be no need to specifically replace all water molecules in the binding site of biomolecular complexes in a given molecule design project, which is common procedure, since there may be no significant gain in ΔG_{bind}. This lack of statistical significance was also confirmed by permutation tests of 2000 permutations. The finding of no statistical significance between the binding free energies of binding sites containing tightly bound water molecules, and those without, is indeed of value, given that the widespread paradigm is to exclude water molecules or target them, irrespective of their binding nature and the difference between classes of water molecules, and the assumption that an increase in binding energy will result because of this neglect or targeting. These assumptions are not borne out by data available. Indeed, there are some water molecules that are more favorably placed in the bulk than in the binding site, and others that are more tightly associated and where this balance is more critical, since few functional groups will replace it successfully. A more careful and discerning approach than commonly used should be conducted when confronting water molecules in binding sites, where they are a frequent occurrence.

Another conclusion is that log *P* is lower in the ligands with tightly bound, bridging water molecules, and this may be helpful in designing ligands that are less lipophilic (hydrophobic) yet retain strong affinity (ΔG_{bind}). Another highlight is the observation that ligand efficiency indices (LEI) are deeper for ligands that do not have tightly bound, bridging water molecules, and this is achieved through their normalization factor (NF), not through ΔG_{bind}. It could indicate that proteins may be capable of enveloping smaller ligands and that flexibility of biomolecule receptors plays an important role. This in turn suggests that displacing or targeting all water molecules in a binding site may not be strictly required as a strategy, since a protein may wrap itself around a smaller ligand (alternatively, a smaller ligand may select a conformation of the protein that has a smaller, nontightly hydrated, binding site).

Tightly Hydrated and Nontightly Hydrated Drug High Resolution X-Ray Structures. From the high resolution complexes, the drug compounds in the binding site were identified using the DrugBank database.[105] This resulted in identifying 72 drugs. The same program written and described in the previous section was run to identify the number of tightly bound, bridging water molecules in the binding sites and found 36 tightly hydrated and 39 nontightly hydrated.

The importance of tightly bound, bridging water molecules is seen for half of the drug compounds in the data set. They are also important from the specificity they impart on the protein−ligand interaction, since their hydrogen bonds have particular spatial orientations. A selection of these bridges can be seen in Figure 3.

Comparisons of box plots for some properties are shown in Figure 4.

Superpositions of densities for several properties were also calculated and are shown in Figure 5.

Statistical comparisons of two-sided, unequal variance, *t* tests were again conducted, with the main results presented in Table 2.

Important to notice is that again, as was the case with all the complexes, tightly bound hydrated drug complexes had the same, statistically indistinguishable distributions of binding energy as nontightly hydrated drugs (with even a slightly smaller amount of weak binders than nontightly hydrated drugs). This supports the indication that replacing or targeting all of the tightly bound, bridging water molecules in a protein−ligand binding site is not a strictly required strategy, inasmuch as binding affinity is concerned. The vice versa situation may also be true. That is, on occasion, a ligand modification or a different ligand may introduce a tightly bound water into the binding site that had not been observed in the binding site structure before. These situations confirm the trend that ligand−biomolecule associations with and without tightly bound waters may be similar with respect to their binding energy, while still considering that the hydrophobic effect will be an important driver of protein−ligand association. Further confirmation comes from the fact that hydrogen bond acceptors and donors were statistically indistinguishable between tightly hydrated and nontightly hydrated drugs. This would imply that their desolvation energies would be similar. The available conformations for ligands and proteins in bulk solution, as well as in complex, will also be critical, where bulk water plays an important role in the energetic and structural properties of the molecules and their association, and where tightly bound water appears to be another variable to consider, with different properties to bulk and loosely associated water.

dx.doi.org/10.1021/ci3005786 | J. Chem. Inf. Model. 2013, 53, 1388−1405

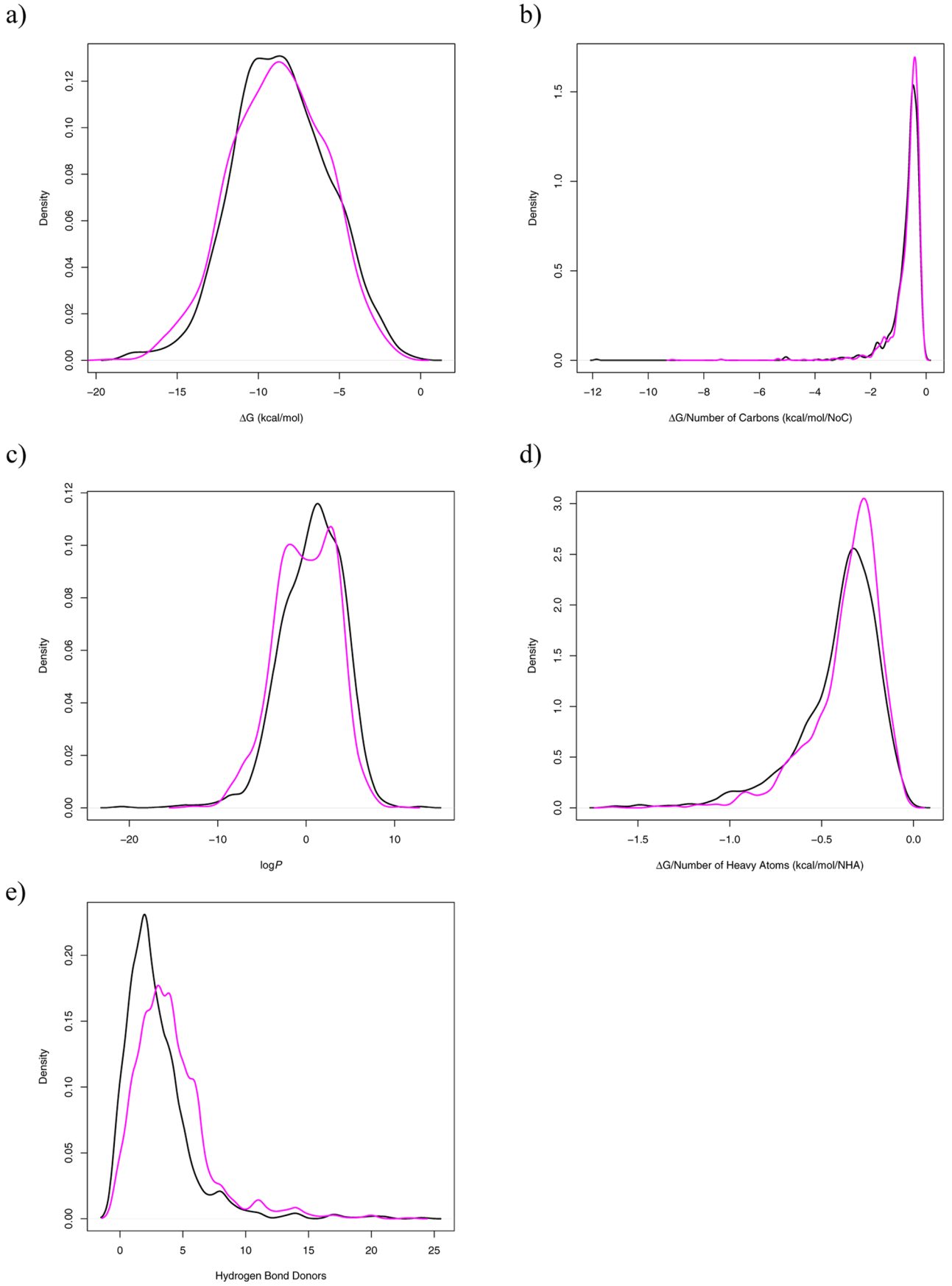


Figure 2. Densities for distributions of ligand properties of tightly hydrated (magenta) and nontightly hydrated (black) biomolecular binding sites: (a) ΔG_{bind}, (b) ΔG_{bind}/NoC, (c) log *P*, (d) ΔG_{bind}/NHA, (e) hydrogen bond donors.

dx.doi.org/10.1021/ci3005786 | *J. Chem. Inf. Model.* 2013, 53, 1388−1405

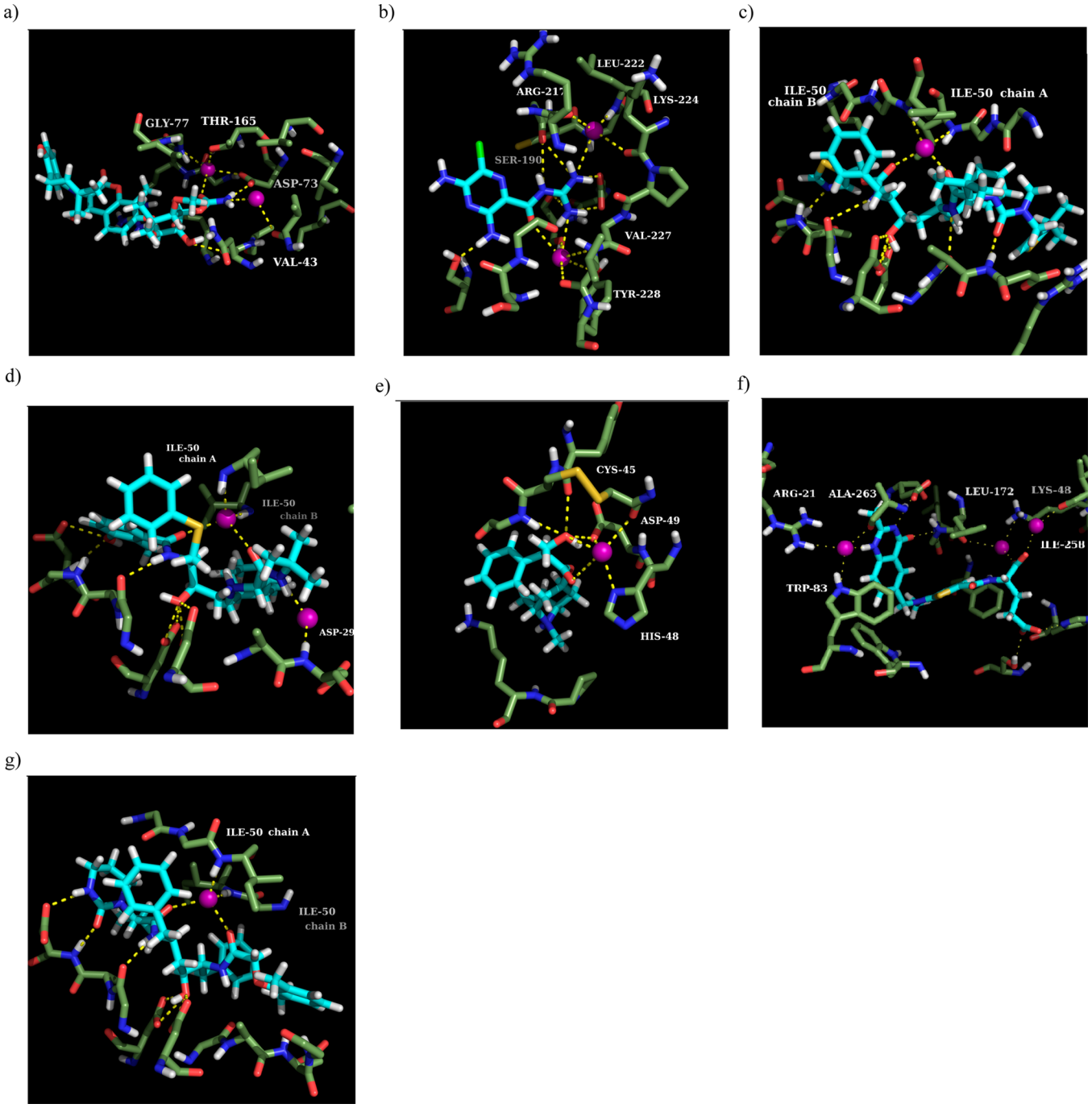


Figure 3. Intermolecular interactions for drug compounds in high resolution X-ray complex structures. (a) Novobiocin, (b) Amiloride, (c) Ritonavir, (d) Viracept, (e) Atropine, (f) Raltitrexed, (g) Lopinavir. Tightly bound, bridging water molecules are shown as purple spheres. The residues engaging in tightly bound water bridges are labeled. Hydrogen bonds are shown as yellow dashed lines, ligand in cyan, and protein in olive. Hydrogens are explicit for the ligand and for the protein polar groups.

Contrary to common thought, drugs that do not make use of tightly bound water molecules were slightly smaller than those that do. However, the ligand efficiencies were statistically indistinguishable between tightly hydrated and nontightly hydrated drug compounds in their binding sites, showing that size-dependent NFs do not overrule the binding affinity in the LEI term. Therefore, replacing a tightly bound water molecule may not increase binding affinity or ligand efficiency.

Another property that was statistically significant was lower number of heavy atoms for nontightly hydrated binding sites. This may be due to the fact that tightly hydrated binding sites require polar groups and hydrogen partners to accommodate water molecules. Although this may increase the number of atoms in a compound, it may also impart specificity on a ligand, since lipophilicity and small molecular size have been implied in ligand promiscuity.[109,110]

Drug and Nondrug High Resolution X-Ray Structures. From the high resolution complexes, the number of drugs and nondrugs that had or did not have tightly bound, bridging water molecules in their binding site are presented in Table 3. The drug names and their HET-ID are shown in Table S1 in the Supporting Information.

dx.doi.org/10.1021/ci3005786 | J. Chem. Inf. Model. 2013, 53, 1388−1405

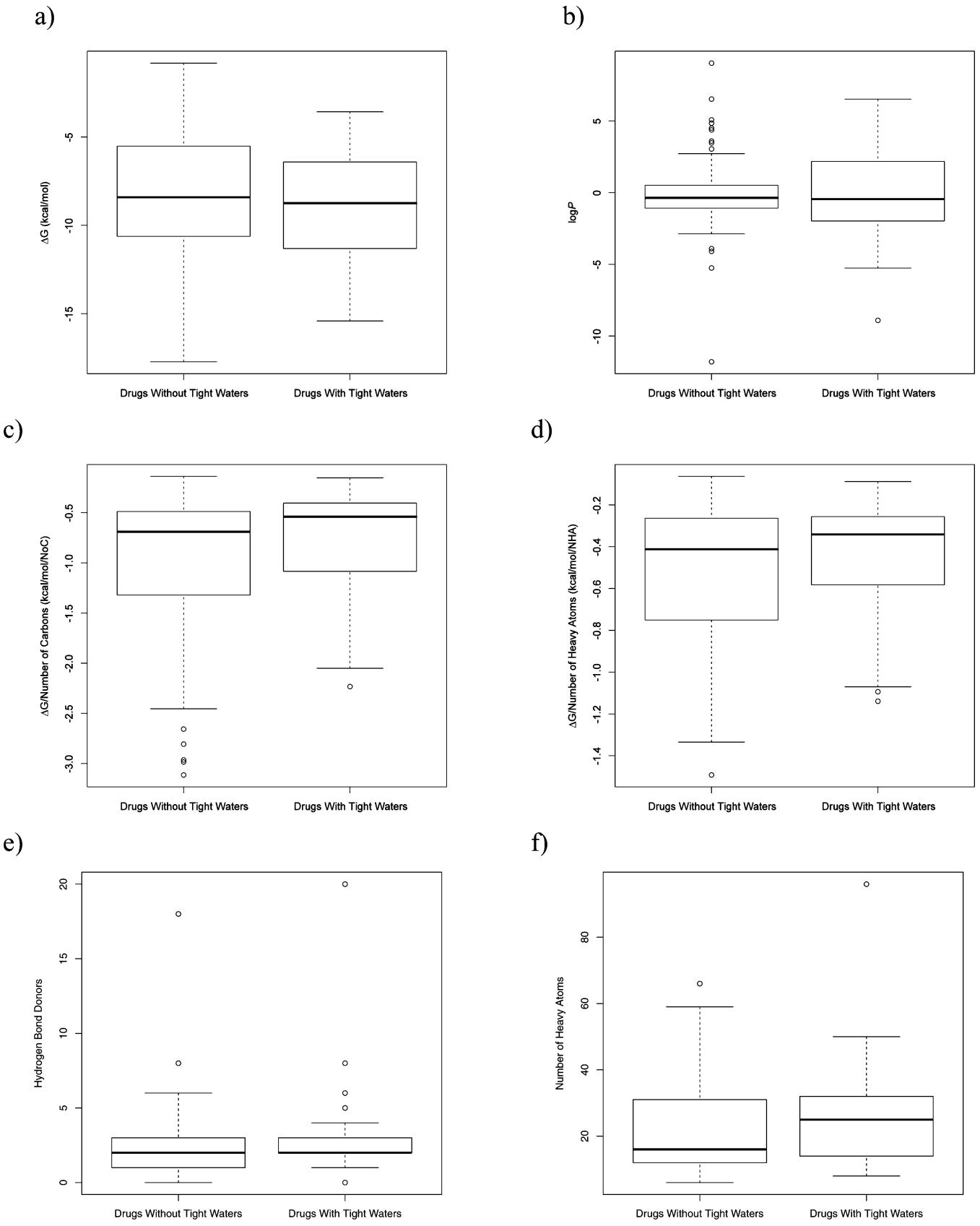


Figure 4. Box plots for comparisons between tightly hydrated and nontightly hydrated drug compounds: (a) ΔG_{bind}, (b) log *P*, (c) ΔG_{bind}/NoC, (d) ΔG_{bind}/NHA, (e) hydrogen bond donors, (f) number of heavy atoms.

Tightly bound, bridging waters were observed in nearly half of the drug structures. An example of both direct and water mediated contacts for a nondrug compound in its protein binding site can be seen in Figure 6.

dx.doi.org/10.1021/ci3005786 | *J. Chem. Inf. Model.* 2013, 53, 1388−1405

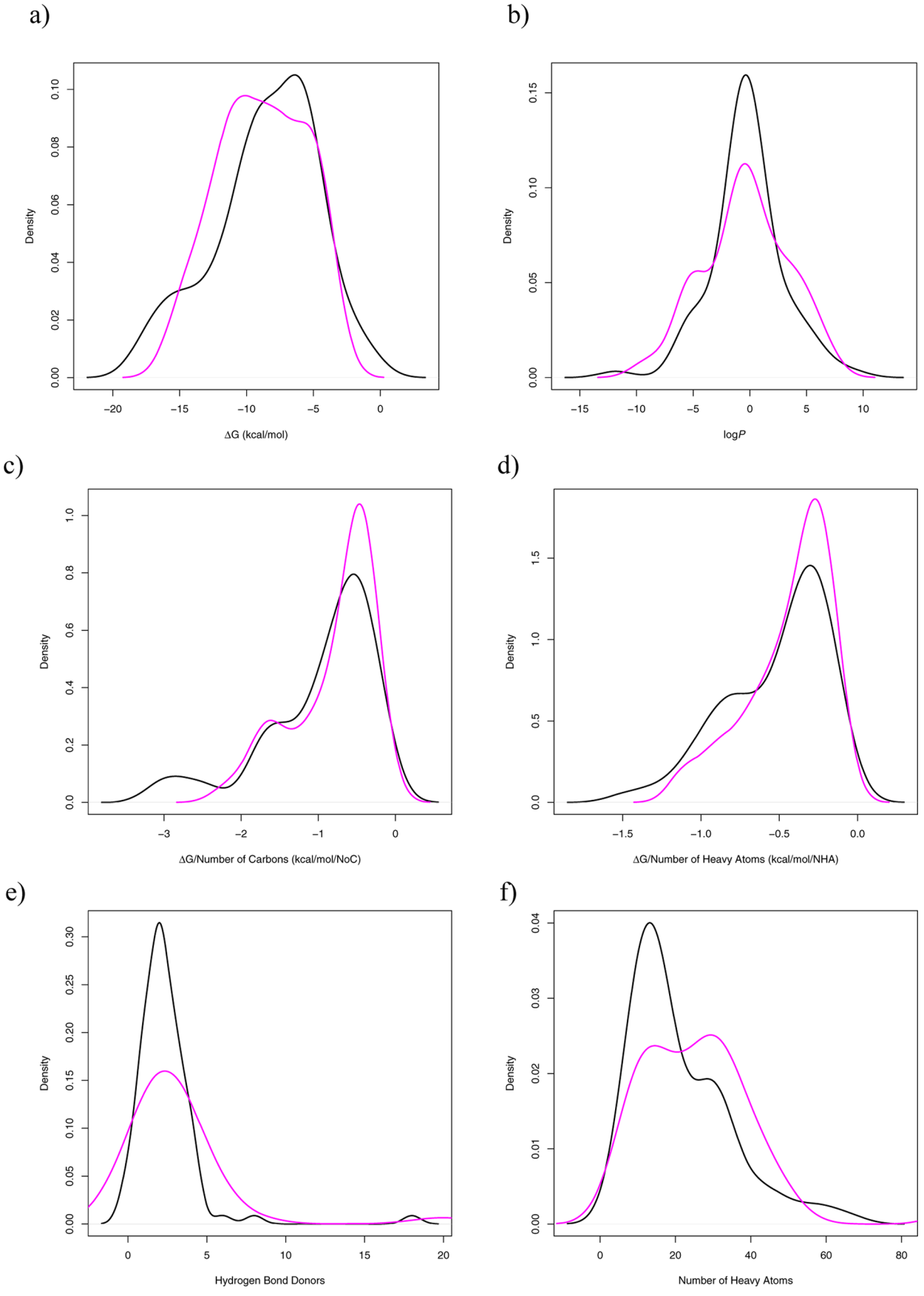


Figure 5. Densities for distributions of drug properties of tightly hydrated (magenta) and nontightly hydrated (black) biomolecular binding sites: (a) ΔG_{bind}, (b) log *P*, (c) ΔG_{bind}/NoC, (d) ΔG_{bind}/NHA, (e) hydrogen bond donors, (f) number of heavy atoms.

dx.doi.org/10.1021/ci3005786 | *J. Chem. Inf. Model.* 2013, 53, 1388−1405

Table 2. Statistically Significant Differences between Drugs with Tightly Bound, Bridging Water Molecules and Drugs without Water Bridges

property	drugs without tightly bound waters	drugs with tightly bound waters
• ΔG_{bind}, ΔG_{bind}/MW, ΔG_{bind}/NHA, ΔG_{bind}/PSA, log(−ΔG_{bind}/P), log P, PSA, hydrogen bond donors, hydrogen bond acceptors, number of hydrogens, aliphatic rings, Balaban index, hyper-Wiener index, Platt index, number of rings, Wiener index, Szeged index	statistically indistinguishable	
• MW, NHA, MSA, rotatable bonds, molecular polarizability, aromatic rings, aromatic atoms	lower	

Table 3. Number of Tightly Hydrated and Nontightly Hydrated Drug and Nondrug Compounds in 2332 Complexes of High Resolution X-Ray Structure Determination with Reported K_i or K_d

	with tightly bound, bridging water molecules	without tightly bound, bridging water molecules
drugs	61	82
nondrugs	939	1250

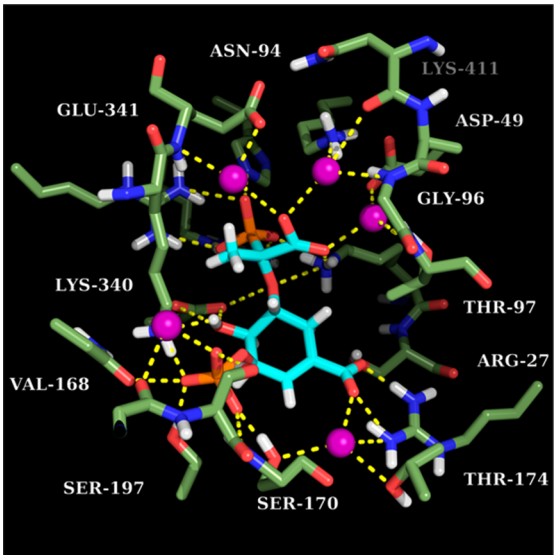


Figure 6. Intermolecular interactions for the nondrug compound in complex structure 1x8t. Tightly bound, bridging water molecules are shown in purple spheres. The residues engaging in tightly bound water bridges are labeled. Hydrogen bonds are shown as yellow dashed lines, ligand in cyan, and protein in olive. Hydrogens are explicit for the ligand and for the protein polar groups.

Statistical comparisons between properties of both compound groups were conducted, and the main results are presented in Table 4.

From Table 4, it can be seen that for the cases of tightly hydrated drugs and tightly hydrated nondrugs, almost all of the property distributions were statistically indistinguishable between both groups of compounds. This underlines the remarkable similarity in properties between compounds, be they drugs or nondrugs, when their binding sites are tightly hydrated, and suggests that compound optimization strategies need not remove or add tightly bound, bridging water molecules in the binding site. They also suggest that a change in these properties is not necessarily beneficial in order to optimize a compound to become a drug. Alternatively, compounds that make good use of tightly bound, bridging water molecules may be well-optimized, and therefore their properties are similar between drugs and nondrugs. Another alternative explanation is that tightly bound, bridging water molecules serve as a good "glue" or structure and energy modifier inside binding sites, allowing the molding of the protein−water−ligand interaction in the best way to accommodate and favor the binding interaction, "stepping aside" as it were, where it is needed and beneficial to the interaction.

Box plots of comparisons of several properties are shown in Figure 7, while superpositions of densities of these properties are shown in Figure 8.

An interesting result for these comparisons is the fact that drugs without tightly bound, bridging water molecules have more total water molecules in their binding site than nondrugs without tightly bound, bridging water molecules. This may be a reflection of a better use of their contacts inside the binding site, having been optimized for biomolecular−ligand−water interactions. Also important is the fact that nontightly hydrated drugs have lower MW, log P, aromatic rings, aromatic atoms, and number of bonds than nontightly hydrated nondrugs. This may be a reflection that many nontightly hydrated nondrugs are peptide ligands, which are generally larger than drug compounds. As in previous cases, binding energy is not statistically significant for both tightly hydrated drugs and tightly hydrated nondrugs. Contrary to the case of tightly hydrated and nontightly hydrated ligands, there is no discernible difference for the number of hydrogen bond acceptors between the tightly hydrated drugs and tightly hydrated nondrugs. This may be due to both tightly hydrated classes of compounds using optimally their polar atoms.

Thus, considering the results of all three comparison cases, general implications and suggestions appear in the use of tightly hydrated and nontightly hydrated, drugs and nondrug compounds. Adopting features present in tightly hydrated and nontightly hydrated drug compounds may help direct ligand design and optimization projects.

Pair-Wise Comparisons. Protein binding sites prior to binding may be partially or fully hydrated, or even fully nonhydrated.[18] Even if molecular dynamics simulations show that water molecules can enter and leave buried sites, ligand atoms compete at different levels of difficulty to replace either bulk, loosely associated, or tightly bound water molecules. Pairwise comparisons between pairs of structures of protein−ligand complexes may show the conservation of some water molecules, such as was done for WaterScore.[26] Using the knowledge of particular tightly bound water sites including dynamic conformational, enthalpic and entropic effects together with standard state corrections and different chemical functional groups in the ligand or protein was useful for designing ligand modification.[46] A collection of structures for pairwise comparison was compiled in the present work in order to study the change in the binding energy due to single changes in either ligand groups or protein groups and associated tightly bound waters. The comparisons are shown in Table 5.

The results from Table 5 show that at different numbers of tightly bound waters, the binding energy can either increase with fewer tight waters, increase with more tight waters, or remain

Table 4. Statistically Significant Differences between Drugs with Tightly Bound, Bridging Water Molecules and Nondrugs with Water Bridges As Well As Those between Drugs without Tightly Bound, Bridging Water Molecules and Nondrugs without Water Bridges as well as General Comparisons between Drugs and Nondrugs

property	drugs	nondrugs
with tightly bound, bridging waters:		
• ΔG_{bind}/NHA, ΔG_{bind}/MW, ΔG_{bind}/Wiener, ΔG_{bind}/NoC, hydrogen bond donors, number of hydrogens	lower	
• ΔG_{bind}, number of total waters in pocket, number of tight waters in pocket, MW, log P, Wiener index, MSA, PSA, hydrogen bond acceptors, rotatable bonds, NoC, NHA, number of carbons, molecular polarizability, aliphatic rings, aromatic rings, aromatic atoms, Balaban index, Harary index, number of bonds, hyper-Wiener index, Platt index, Randic index, number of rings, Szeged index, Wiener polarity, ΔG_{bind}/PSA, $\log(-\Delta G_{bind}/P)$.	statistically indistinguishable	
without tightly bound, bridging waters:		
• ΔG_{bind}, ΔG_{bind}/PSA, PSA, hydrogen bond acceptors, aliphatic rings, Balaban index, Szeged index	statistically indistinguishable	
• ΔG_{bind}/NHA, ΔG_{bind}/Wiener, ΔG_{bind}/MW, ΔG_{bind}/NoC, MW, Wiener index, MSA, hydrogen bond donors, rotatable bonds, number of atoms, number of hydrogens, NHA, NoC, number of total waters in pocket, molecular polarizability, aromatic rings, aromatic atoms, Harary index, number of bonds, Platt index, Randic index, number of rings, Wiener polarity, log P, $\log(-\Delta G_{bind}/P)$	lower	
all together		
• ΔG_{bind}, ΔG_{bind}/PSA, PSA, hydrogen bond acceptors, aliphatic rings, Balaban index, hyper-Wiener index, number of rings, Szeged index	statistically indistinguishable	
• number of total waters in pocket, ΔG_{bind}/NHA, ΔG_{bind}/Wiener, ΔG_{bind}/MW, ΔG_{bind}/NoC, MW, MSA, hydrogen bond donors, rotatable bonds, number of atoms, number of hydrogens, NHA, NoC, molecular polarizability, aromatic rings, aromatic atoms, Harary index, number of bonds, Platt index, Randic index, Wiener index, Wiener polarity, log P, $\log(-\Delta G_{bind}/P)$	lower	

similar. Note must be taken that different tools may provide a slightly different definition of tightly bound water molecule clusters.

Since prior to ligand binding protein binding sites may be partially hydrated, fully hydrated, or fully nonhydrated, ligand or protein atoms may sometimes be displacing an "empty site." However, conformational and cavity effects are still present. In addition, it may be extremely difficult to disentangle the contribution to energy components from changes in a hydration site, given that the resulting enthalpy, entropy, cavitation, and conformational effects are closely intertwined.[45] Further still, even if hydration site changes can dictate the specificity of an interaction,[27,28] at other times, changes in ligand structure will not be dependent on hydration structure changes, since these can accommodate the interactions of different partner ligands, acting as a glue, "brick wall," or screen.[29] Ligand binding energy can come from displacing loosely associated waters.[41,42] Here, we compare displacing tightly bound waters, which are different in nature, and where the results show less of an advantage of displacement. The study of Timson et al.[44] shows that reordering of water molecules from loose-bound to tight-bound can drive a biomolecular association. Thus, tightly bound water replacement must be especially done in a way that best reproduces or improves on the tightly bound water molecules interactions, or else the change in energy will not be favorable.

Ligand Types. A breakdown of the ligands by type showed 35% of them are inhibitors and 33% ligands. A total of 32 compounds were agonists and six antagonists, as well as nine products, nine substrates, and three antigens. Product ligands had lower binding affinity than the other types of ligand, while inhibitor ligands had the strongest binding affinity. Among the agonists and antagonists, 20 agonists and four antagonists had tightly bound waters. Since agonists and antagonists are frequently bound to membrane receptors that are hard to crystallize, there were fewer high resolution crystal structures of them as compared to enzyme ligands. The means for the binding energy per class are shown in Table 6.

Form Table 6, it can be seen that agonists with tightly bound waters had the same binding energies as those without. For antagonists, those without tightly bound water molecules were slightly more favorable than those with. However, the difference in binding energy between antagonists and agonists was not statistically significant, as was also the case between tightly and nontightly hydrated agonists, and the same comparison for antagonists. Table 7 shows the means for these groups of compounds and binding sites.

From Table 7, it can be seen that agonists and antagonists without tightly bound waters are larger and more hydrophobic, have larger steric effects (as measured in their Wiener polarity), higher complexity (higher number of fragments, as reflected in their Platt indices), and less branching (as measured in their Wiener and Randic indices) than those that use tight waters, or agonists or antagonists in general. This means that they may be less easy to develop further. Also, they contain more aromatic rings and atoms, which mean that they are flatter molecules, associated with lower solubility and more problems in transitioning through development phases.[111] On the other hand, those agonists and antagonists that use tightly bound waters are smaller, less hydrophobic, less flat, and have deeper LEI than the compounds that do not use tight waters and than the overall sets. The largest and most obese compounds are the antagonists that do not use tightly bound water molecules. Hence, compounds that use tightly bound water molecules may have advantages of developability over those compounds that do not. Agonists with tightly bound water bridges were similar in properties to antagonists with tightly bound water bridges, and agonists without tightly bound water bridges were also similar to antagonists without tightly bound water bridges.

The overall results show that compounds and binding sites that possess tightly bound water molecule bridges tend to have similar energy and physicochemical properties as those that do not. This stresses the point that water molecules in binding sites may be used to guide ligand optimization, but that their use must be well reasoned. If the water molecule is loosely bound, no major problem may be expected to arise by their substitution for other groups. However, if they form a tightly bound water bridge, then the substitution requires a much better analysis, since it can easily occur that it may not be more favorable than the original water bridge, in terms of energy, or of physicochemical properties, or of ligand efficiency, or of developability (further modification). Hence, sometimes tightly bound water molecules may be left as bridges, or else substituted properly, taking into account desolvation, other enthalpic and entropic, including conformational and cavitation effects of the ligand, protein, as

dx.doi.org/10.1021/ci3005786 | J. Chem. Inf. Model. 2013, 53, 1388−1405

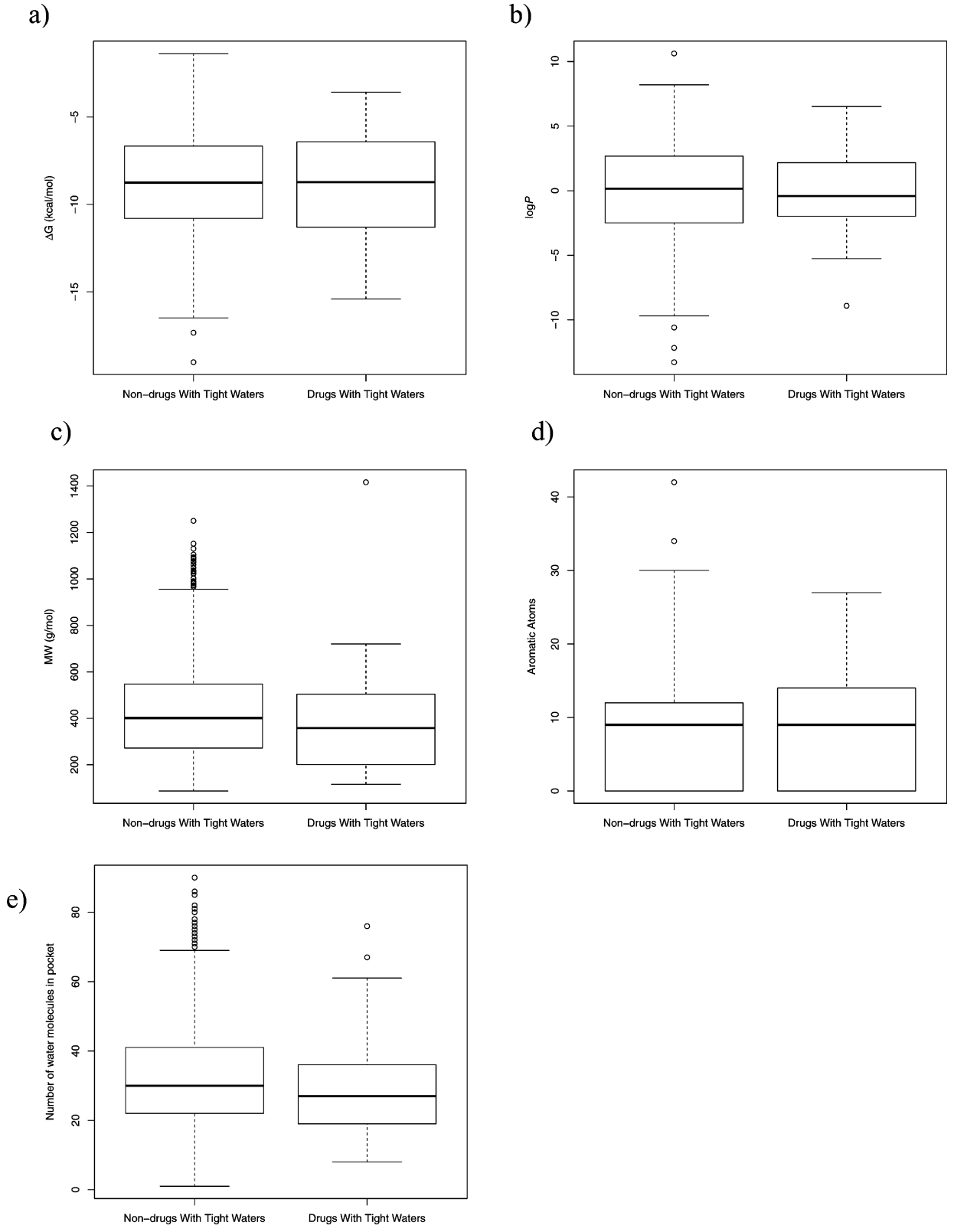


Figure 7. Box plots for comparisons between tightly hydrated drug and tightly hydrated nondrug compounds: (a) ΔG_{bind}, (b) log *P*, (c) MW, (d) number of aromatic atoms, (e) number of total waters.

well as tightly bound water molecule that may behave differently than a bulk or a loosely bound water molecule. The results also show that important information and optimization techniques can be gleaned from using protein and ligand and water structural information together.

CONCLUSIONS

The general strategy of replacing a tightly bound, bridging water molecule in a receptor−water−ligand binding site may lead in some occasions to increases in binding energy, such as the

dx.doi.org/10.1021/ci3005786 | *J. Chem. Inf. Model.* 2013, 53, 1388−1405

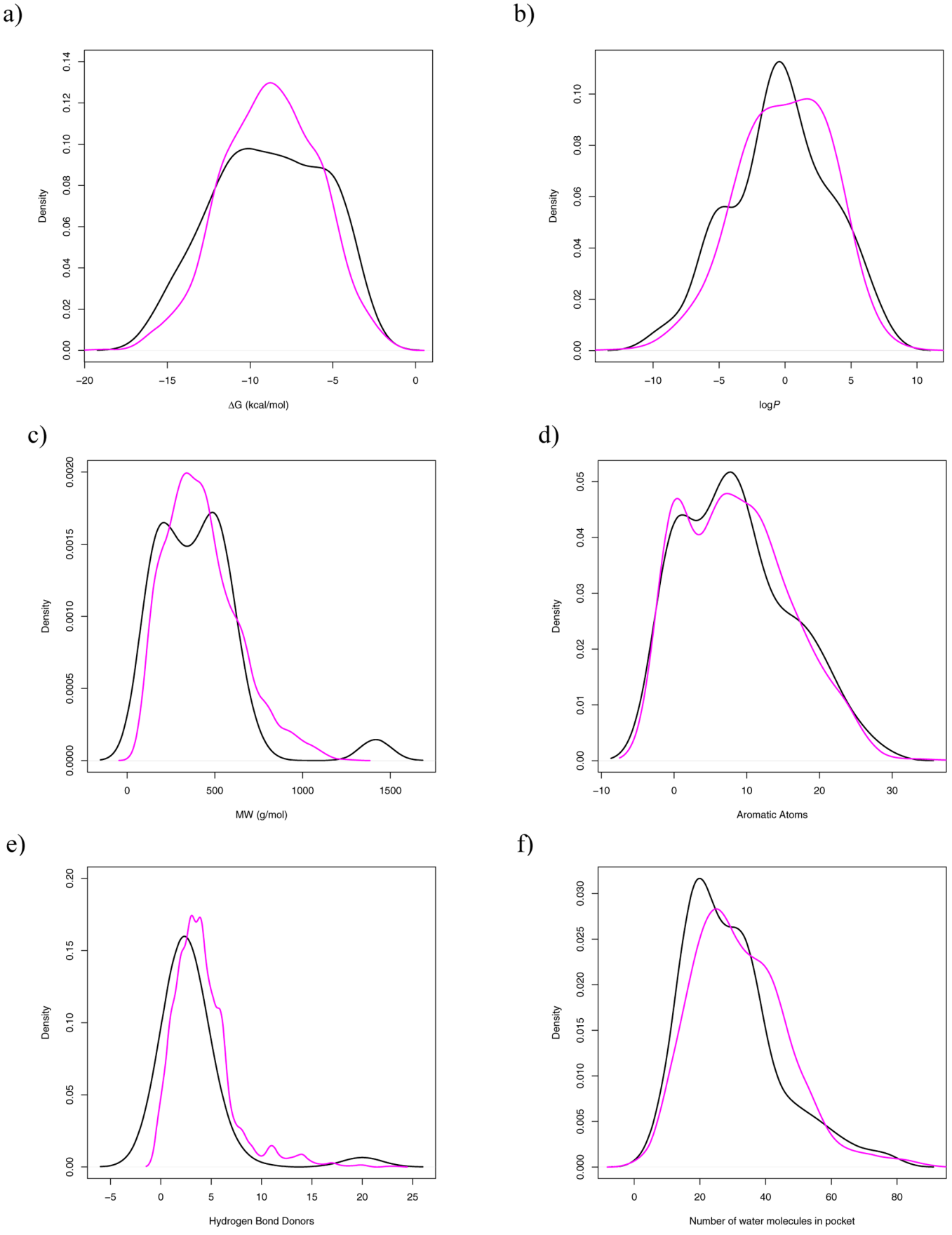


Figure 8. Densities for distributions of drug properties of tightly hydrated drug (black) and tightly hydrated nondrug (magenta) biomolecular binding sites: (a) ΔG_{bind}, (b) log *P*, (c) MW, (d) number of aromatic atoms, (e) number of hydrogen bond donors, (f) number of total waters.

dx.doi.org/10.1021/ci3005786 | *J. Chem. Inf. Model.* 2013, 53, 1388−1405

Table 5. Comparisons of Protein/Ligand Pairs with Different Numbers of Tight Water Clusters (ΔG_{bind} in kcal/mol)

PDB code	ligand	protein	tight waters	ΔG_{bind}	PDB code	ligand	protein	tight waters	ΔG_{bind}
1nxy	oxacillin	β-lactamase	0	−9.81	2o4k	atazanavir	HIV protease	2	−14.32
1ny0	(2-ethoxy-1-naphthoyl)amino]methylboronic acid	β-lactamase	3	−7.98	2o4p	tipranavir	HIV protease	0	−14.62
1nym	[(2-amino-alpha-methoxyimino-4-thiazolylacetyl)amino]methylboronic acid	β-lactamase	2	−7.36	2o4s	lopinavir	HIV protease	2	−14.33
1nyy	N-[5-methyl-3-o-tolyl-isoxazole-4-carboxylic acid amide] boronic acid	β-lactamase	0	−6.38	1cea	aminocaproic acid	plasminogen	0	−6.76
2rkf	lopinavir	HIV protease	1	−12.38	1ceb	trans-4-aminomethylcyclohexane-1-carboxylic acid	plasminogen	19	−8.18
2rkg	lopinavir	HIV protease	0	−11.02	1laf	arginine	LAO binding protein	2	−10.71
1f5k	benzamidine	urokinase	0	−5.11	1lag	histidine	LAO binding protein	0	−8.59
1f5l	amiloride	urokinase	2	−7.19	1lah	ornithine	LAO binding protein	1	−10.26
1yda	acetazolamide	carbonic anhydrase II variant	1	−8.93	1n9m	biotin	streptavidin	1	−14.94
1ydb	acetazolamide	carbonic anhydrase II variant	0	−11.23	1swg, 1swk, 1swn, 1swp, 1swr, 1df8, 1ndj, 1f27, 1n43, 2f01, 2gh7	biotin	streptavidin	0	−14.68 (average)
1ydd	acetazolamide	carbonic anhydrase II variant	0	−9.63	2aou	amodiaquine	histamine methyltransferase	1	−10.54
1l5q	caffeine	glycogen phosphorylase	3	−5.41	2aov	metoprine	histamine methyltransferase	1	−9.60
1l7x	caffeine	glycogen phosphorylase	0	−5.5	2aow	tacrine	histamine methyltransferase	0	−10.11
1gfz	caffeine	glycogen phosphorylase	1	−5.45	2aox	tacrine	histamine methyltransferase	0	−10.11
1i00	raltitrexed	thymidylate synthase	0	−8.64	2gss	ethacrynic acid	glutathione-S-transferase	2	−6.73
2kce	raltitrexed	thymidylate synthase	2	−8.42	3gss	ethacrynic acid-glutathione	glutathione-S-transferase	5	−7.94
1wdn	glutamine	Glu binding protein	0	−8.59	1n0s	fluorescein	lipocalin	0	−10.16
1ii5	glutamine	GluR0 ion channel	1	−9.02	1lke	digoxigenin	lipocalin	1	−10.25
1xff	glutamine	glucosamine 6-phosphate synthase	0	−6.58	1lnm	digitoxigenin	lipocalin	0	−11.86
1o86	lisinopril	testicular angiotensin-converting enzyme	4	−13.05	2pou	dichlorphenamide	carbonic anhydrase	1	−10.12
1j36	lisinopril	Drosophila angiotensin-converting enzyme	2	−10.56	2pov	4-amino-6-chloro-benzene-1,3-disulfonamide	carbonic anhydrase	1	−9.71
2arm	atropine	phospholipase A2	2	−11.08	2pow	4-amino-6-trifluoromethyl-benzene-1,3-disulfonamide	carbonic anhydrase	2	−9.82
1sv3	4-methoxybenzoic acid	phospholipase A2	0	−6.47	1h60	progesterone	pentaerythritol tetranitrate reductase	0	−6.41
1fv0	9-hydroxy aristolochic acid	phospholipase A2	0	−8.08	1h61	prednisone	pentaerythritol tetranitrate reductase	1	−6.66
1q7a	oxyphenbutazone	phospholipase A2	0	−9.81	1h62	1,4-androstadien- 3,17-dione	pentaerythritol tetranitrate reductase	0	−6.50
1kpm	vitamin E	phospholipase A2	0	−7.91	1b38	ATP	CDK2	1	−8.99

dx.doi.org/10.1021/ci3005786 | J. Chem. Inf. Model. 2013, 53, 1388−1405

Table 5. continued

PDB code	ligand	protein	tight waters	ΔG_{bind}	PDB code
		c-SRC kinase	0	−6.82	1b39
		tyrosine kinase	1	−7.23	1xbb
		neuraminidase	4	−11.89	1xbc
		neuraminidase	0	−9.64	1hxw
					1sh9

PDB code	ligand	protein	tight waters	ΔG_{bind}
2oiq	imatinib	CDK2	0	−9.43
1xbb	imatinib	tyrosine kinase	1	−7.23
3ckz	zanamivir	tyrosine kinase	0	−10.80
3cl0	oseltamivir	HIV protease	3	−14.76
		HIV protease	0	−8.22

	ligand
	ATP
	imatinib
	staurosporin
	ritonavir
	ritonavir

Table 6. Means of ΔG_{bind} in kcal/mol for Tightly Hydrated and Nontightly Hydrated Agonist and Antagonist Compounds

	all	with tightly bound, bridging water molecules	without tightly bound, bridging water molecules
agonists	−9.51	−9.50	−9.52
antagonists	−7.56	−7.27	−8.16

classical case of HIV-1 protease cyclic ureas,[47] where there was an increase in hydrophobic interactions as well as the added benefit of increased rigidity in the ligand as compared to the noncyclic, hydrated ligand complexes. However, the present results show that this strategy is not strictly required, since the data show that tightly hydrated ligands are as strong in binding affinity as nontightly hydrated ligands. This is also supported by observations where targeting and replacing a tightly bound, bridging water molecule did not lead to major increases in energy,[85−88] or indeed, may even result in a loss of binding affinity.[89] These inhibition effects may be due to changing effects of enthalpy and entropy, both in the binding site,[46] as well as in the bulk solvent.[46] The binding site may have difficulty accepting different ligand modifications.[46,89] In addition, the hydrated binding site may be one of the many structure conformations available to the receptor, and different ligands will have a different ability to select either tightly hydrated or nontightly hydrated receptor binding site conformations. Therefore, tightly bound, bridging water molecules may sometimes be replaced and targeted as a strategy, though sometimes keeping them as bridges may be better from a pharmacodynamic perspective.

From a pharmacokinetic point of view, compounds that make use of tightly bound, bridging water molecules have the advantage of lower log *P* (and some classes of compounds are smaller and less planar), than nontightly hydrated ones. This may be very useful when deciding to advance compounds for development, since they may retain low lipophilicity which may more easily allow further modification and specificity than more lipophilic (nontightly hydrated) compounds that may have problems of lack of specificity and promiscuity.[109,110] In addition, more polar ligands may be beneficial to avoid "molecular obesity"[90] and increase specificity since polar interactions are more directed in space than hydrophobic interactions, and place the emphasis on improving enthalpic interactions rather than presupposed entropic gains through larger ligand molecular size and lipophilicity. Conserving tightly bound, bridging water molecules in the receptor−ligand binding site may also favor these properties in the ligands.

Tightly hydrated drugs and tightly hydrated nondrugs show that the former have lower MW, NHA, and NoC and thus increase their ligand efficiency indices. This improvement of ligand efficiency indices is not required to come from water molecule substitution, since binding affinity is not significantly different for both groups of compounds.

From all the above, the results suggest that conserving or displacing tightly bound water molecules in the binding site may have unpredictable or even negative consequences on the binding affinity of a compound, as well as on its physicochemical properties, which in turn affect its pharmacokinetics, as well as its specificity and safety properties. Indeed, each ligand and binding site may behave uniquely, with tightly bound, bridging water molecules being an extra factor to be considered in the structural, conformational, energetic (enthalpic and entropic), and physicochemical variables of receptor and ligand. Pragmatically,

dx.doi.org/10.1021/ci3005786 | J. Chem. Inf. Model. 2013, 53, 1388−1405

Table 7. Means of Physicochemical Properties of Hydrated and Nonhydrated Agonist and Antagonist Compounds

property	agonists	antagonists	agonists with tightly bound waters	agonists without tightly bound waters	antagonists with tightly bound waters	antagonists without tightly bound waters
MW	377.8	356.1	359.0	409.3	321.8	424.7
NHA	27.1	24.5	25.6	29.5	21.5	30.5
NoC	19.8	17.2	18.3	22.3	14	23.5
tight waters	1.12	1.33	1.8	0	2	0
total waters	26.22	28	33.35	14.33	35.25	13.5
Wiener index	4380.2	2561.2	4136	4787.2	2014.5	3654
log *P*	2.44	2.09	1.70	3.685	1.39	3.485
MSA	494.6	450.2	436.6	546.3	383.8	583
PSA	81.35	76.68	87.88	70.48	60.74	108.56
donors	2.38	2.0	2.65	1.92	1.25	3.5
acceptors	5.75	6.5	5.7	5.83	6.5	6.5
rotatable bonds	7.44	7.33	6.4	9.2	5	12
atoms	52.4	45.7	49.4	57.25	36.5	64
hydrogens	25.3	21.2	23.8	27.75	15	33.5
molecular polarizability	39.59	36.52	37.09	43.76	31.67	46.23
aliphatic rings	0.69	0.33	0.75	0.58	0.25	0.50
aromatic rings	1.78	2.0	1.5	2.25	2.0	2.0
aromatic atoms	9.5	11	8	12	10.5	12
Balaban index	1.872	1.838	1.952	1.738	1.927	1.660
Harary index	111.87	89.82	104.88	123.52	77.05	115.37
Bonds	53.8	47.0	50.7	59.1	37.75	66.5
hyper-Wiener	29602	15171	28718	31076	11797.8	21918
Platt index	193.1	165.0	182.6	210.5	130.5	234
Randic index	22.94	20.05	21.63	25.13	16.03	28.09
rings	2.47	2.33	2.25	2.83	2.25	2.50
Szeged index	5182	3549.7	4818	5789	2960.2	4728
Weiner polarity	41.84	35.5	40.35	44.33	30.75	45.0
ΔG_{bind}/NHA	−0.464	−0.322	−0.482	−0.434	−0.352	−0.264
ΔG_{bind}/MW	−0.033	−0.022	−0.034	−0.031	−0.023	−0.019
ΔG_{bind}/NoC	−0.740	−0.501	−0.784	−0.667	−0.563	−0.377
ΔG_{bind}/PSA	−0.231	−0.131	−0.262	−0.178	−0.137	−0.120
ΔG_{bind}/Wiener	−0.021	−0.011	−0.017	−0.027	−0.015	−0.003
$\log(-\Delta G_{bind}/P)$	−1.478	−1.259	−0.730	−2.725	−0.593	−2.593

replacing a tightly bound, bridging water molecule may only be adequate when the new ligand series actually derives from this both stronger binding affinity, as well as favorable physicochemical and specificity properties, and predicting these effects is not straightforward. That is, each tightly hydrated or nontightly hydrated ligand−receptor interaction will have its own defining causes and consequences, which is in agreement with the central role of water as a solvent and its close role in biomolecular association. Picking apart this role allows better understanding and use of water molecules in biomolecular association.

A fair strategy may thus be comprised of designing compounds and obtaining structural information of their complexes with receptor, and if a tightly bound, bridging water molecule is observed in the binding site, only attempt to replace it if the subsequent ligand modification would also improve its ligand efficiency, enthalpy, entropy, specificity, and pharmacokinetic properties. If the modification does succeed in replacing the tightly bound, bridging water molecule, it will have at least achieved benefits for the ligand optimization and development independently of either positive or negative change in binding affinity outcome.

■ ASSOCIATED CONTENT

Supporting Information

Drug names and HET-IDs found in high resolution X-ray crystal structures with reported K_i or K_d. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +372 737 5270. Fax: +372 737 5264. E-mail: alfonsog@ut.ee.

Notes

The author declares no competing financial interest.

■ ACKNOWLEDGMENTS

Estonian Ministry for Education and Research (Grant SF0140031Bs09) and the University of Tartu High Performance Computing Centre for computing resources.

■ REFERENCES

(1) Lemieux, R. U. How water provides the impetus for molecular recognition in aqueous solution. *Acc. Chem. Res.* **1996**, *29*, 373−380.
(2) Tame, J. R. H.; Sleigh, S. H.; Wilkinson, A. J.; Ladbury, J. E. The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat. Struct. Biol.* **1996**, *3*, 998−1001.

(3) Li, Z.; Lazaridis, T. The effect of water displacement on binding thermodynamics: concanavalin A. *J. Phys. Chem. B* **2005**, *109*, 662−670.
(4) Snyder, P. W.; Mecinovic, J.; Moustakas, D. T.; Thomas, S. W.; Harder, M.; Mack, E. T.; Lockett, M. R.; Heroux, A.; Sherman, W.; Whitesides, G. M. Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17889−17894.
(5) Beuming, T.; Farid, R.; Sherman, W. High-energy water sites determine peptide binding affinity and specificity of PDZ domains. *Protein Sci.* **2009**, *18*, 1609−1619.
(6) Wang, L.; Berne, B. J.; Friesner, R. A. Ligand binding to protein-binding pockets with wet and dry regions. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 1326−1330.
(7) Setny, P.; Baron, R.; McCammon, J. A. How can hydrophobic association be enthalpy driven? *J. Chem. Theory Comput.* **2010**, *6*, 2866−2871.
(8) Mobley, D. L.; Dill, K. A. Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure* **2009**, *17*, 489−498.
(9) Gilson, M. K.; Zhou, H.-X. Calculation of water molecules in protein binding sites. *J. Am. Chem. Soc.* **2007**, *129*, 2577−2587.
(10) Olano, L. R.; Rick, S. W. Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.* **2004**, *25*, 7991−8000.
(11) Meagher, K. L.; Carlson, H. A. Solvation influences flap collapse in HIV-1 protease. *Proteins: Struct., Funct., Bioinf.* **2005**, *58*, 118−125.
(12) Bren, U.; Kržan, A.; Mavri, J. Microwave catalysis through rotationally hot reactive species. *J. Phys. Chem. A.* **2008**, *112*, 166−171.
(13) Bren, U.; Janežič, D. Individual degrees of freedom and the solvation properties of water. *J. Chem. Phys.* **2012**, *137*, 024108.
(14) Palencia, A.; Cobos, E. S.; Mateo, P. L.; Martinez, J. C.; Luque, I. Thermodynamic dissection of the binding energetics of proline-rich peptides to the Abl-SH3 domain: implications for rational ligand design. *J. Mol. Biol.* **2004**, *336*, 527−537.
(15) Ahmed, M. H.; Spyrakis, F.; Cozzini, P.; Tripathi, P. K.; Mozzarelli, A.; Scarsdale, J. N.; Safo, M. A.; Kellogg, G. E. Bound water at protein-protein interfaces: partners, roles and hydrophobic bubbles as a conserved motif. *PLoS One* **2011**, *6*, e24712.
(16) Vaiana, A. C.; Westhof, E.; Auffinger, P. A molecular dynamics simulation study of an aminoglycoside/A-site RNA complex: conformational and hydration patterns. *Biochimie* **2006**, *88*, 1061−1073.
(17) Bissantz, C.; Kuhn, B.; Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **2010**, *53*, 5061−5084.
(18) Genheden, S.; Mikulskis, P.; Hu, L. H.; Kongsted, J.; Soderhjelm, P.; Ryde, U. Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *J. Am. Chem. Soc.* **2011**, *133*, 13081−13092.
(19) Abel, R.; Salam, N. K.; Shelley, J.; Farid, R.; Friesner, R. A.; Sherman, W. Contribution of explicit solvent effects to the binding affinity of small-molecule inhibitors in blood coagulation factor serine proteases. *ChemMedChem* **2011**, *6*, 1049−1066.
(20) Conte, L. L.; Chothia, C.; Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **1999**, *285*, 2177−2198.
(21) Stegmann, C. M.; Seeliger, D.; Sheldrick, G. M.; de Groot, B. L.; Wahl, M. C. The thermodynamic influence of trapped water molecules on a protein-ligand interaction. *Angew. Chem., Int. Ed.* **2009**, *48*, 5207−5210.
(22) Biela, A.; Sielaff, F.; Terwesten, F.; Heine, A.; Steinmetzer, T.; Klebe, G. Ligand binding stepwise disrupts water network in thrombin: enthalpic and entropic changes reveal classical hydrophobic effect. *J. Med. Chem.* **2012**, *55*, 6094−6110.
(23) Biela, A.; Nasief, N. N.; Betz, M.; Heine, A.; Hangauer, D.; Klebe, G. Dissecting the hydrophobic effect on the molecular level: The role of water, enthalpy, and entropy in ligand binding to thermolysin. *Angew. Chem., Int. Ed.* **2013**, *52*, 1822−1828.
(24) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580−594.
(25) Lu, Y.; Wang, R.; Yang, C. -Y.; Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.* **2007**, *47*, 668−675.
(26) García-Sosa, A. T.; Mancera, R. L.; Dean, P. M. WaterScore: A novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein−ligand complexes. *J. Mol. Model* **2003**, *9*, 172−182.
(27) Swaminathan, C. P.; Nandi, A.; Visweswariah, S. S.; Surolia, A. Thermodynamic analyses reveal role of water release in epitope recognition by a monoclonal antibody against the human guanyl cyclase C receptor. *J. Biol. Chem.* **1999**, *274*, 31272−31278.
(28) Katz, B. A.; Luong, C.; Ho, J. D.; Somoza, J. R.; Gjerstad, E.; Tang, J.; Williams, S. R.; Verner, E.; Mackman, R. L.; Young, W. B.; Sprengeler, P. A.; Chan, H.; Mortara, K.; Janc, J. W.; McGrath, M. E. Dissecting and designing inhibitor selectivity determinants at the S1 site using an artificial ala190 protease (ala190 uPA). *J. Biol. Chem.* **2004**, *344*, 527−547.
(29) Sleigh, S. H.; Seavers, P. R.; Wilkinson, A. J.; Ladbury, J. E.; Tame, J. R. H. Crystallographic and calorimetric analysis of peptide binding to OppA protein. *J. Mol. Biol.* **1999**, *291*, 393−415.
(30) García-Sosa, A. T.; Mancera, R. L. The effect of a tightly bound water molecule on scaffold diversity in the computer-aided de novo design of CDK2 inhibitors. *J. Mol. Model.* **2006**, *12*, 422−431.
(31) García-Sosa, A. T.; Firth-Clark, S.; Mancera, R. L. Including tightly-bound water molecules in de novo drug design. Exemplification through the in silico generation of poly(ADP-Ribose) polymerase ligands. *J. Chem. Inf. Model.* **2005**, *45*, 624−633.
(32) Baron, R.; Setny, P.; McCammon, J. A. Hydrophobic association and volume-confined water molecules. In *Protein-Ligand Interactions*, 1st ed.; Gohlke, H., Ed.; Wiley-VCH Verlag: Weinheim, Germany, 2012; Series: Methods and Principles in Medicinal Chemistry.
(33) Dunitz, J. D. The entropic cost of bound water in crystals and biomolecules. *Science* **1994**, *264*, 670.
(34) Hooft, R. W. W.; Sander, C.; Vriend, G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins: Struct., Funct., Genet.* **1996**, *26*, 363−376.
(35) Pace, N. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K. Forces contributing to the conformational stability of proteins. *FASEB J.* **1996**, *10*, 75−83.
(36) Yu, H.; Rick, S. W. Free energy, entropy, and enthalpy of a water molecule in various protein environments. *J. Phys. Chem. B* **2010**, *114*, 11552−11560.
(37) Denisov, V. P.; Venu, K.; Peters, J.; Horlein, H. D.; Halle, B. Orientational disorder and entropy of water in protein cavities. *J. Phys. Chem. B* **1997**, *101*, 9380−9389.
(38) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting conserved water-mediated and polar ligand interactions and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997**, *265*, 445−464.
(39) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of water molecules in protein binding suites. *J. Am. Chem. Soc.* **2007**, *129*, 2577−2587.
(40) Lazaridis, T. Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102*, 3531−3541.
(41) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808−813.
(42) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friessner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817−2831.
(43) Guimaraes, C. R.; Mathiowetz, A. M. Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free energy perturbation calculations. *J. Chem. Inf. Model* **2010**, *50*, 547−559.
(44) Timson, M. J.; Duff, M. R.; Dickey, G.; Saxton, A. M.; Reyes-De-Corcuera, J. I.; Howell, E. E. Further studies on the role of water in R67 dihydrofolate reductase. *Biochemistry* **2013**, *52*, 2118−2127.
(45) Baum, B.; Muley, L.; Smolinski, M.; Heine, A.; Hangauer, D.; Klebe, G. Non-additivity of functional group contributions in protein-ligand binding: A comprehensive study by crystallography and isothermal titration calorimetry. *J. Mol. Biol.* **2010**, *397*, 1042−1054.

(46) García-Sosa, A. T.; Mancera, R. L. Free energy calculations of mutations involving a tightly bound water molecule and ligand substitutions in a ligand-protein complex. *Mol. Inf.* **2010**, *29*, 589−600.
(47) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Ericksonviitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380−384.
(48) Lloyd, D. G.; García-Sosa, A. T.; Alberts, I. L.; Todorov, N. P.; Mancera, R. L. The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 89−100.
(49) García-Sosa, A. T.; Sild, S.; Takkis, K.; Maran, U. Combined approach using ligand efficiency, cross-docking, and antitarget hits for wild-type and drug-resistant Y181C HIV-1 reverse transcriptase. *J. Chem. Inf. Model.* **2011**, *51*, 2595−2611.
(50) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins: Struct., Funct., Bioinf.* **1999**, *34*, 17−28.
(51) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504−6515.
(52) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the water content in protein binding sites. *J. Phys. Chem. B* **2010**, *113*, 13337−13346.
(53) Ross, G. A.; Morris, G. M.; Biggin, P. C. Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS One* **2012**, *7*, e32036.
(54) Huggins, D. J.; Tidor, B. Systematic placement of structural water molecules for improved scoring of protein-ligand interactions. *Protein Eng., Des. Sel.* **2011**, *24*, 777−789.
(55) Geroult, S.; Hooda, M.; Virdee, S.; Waksman, G. Prediction of solvation sites at the interface of src SH2 domain complexes using molecular dynamics simulations. *Chem. Biol. Drug Des.* **2007**, *70*, 87−99.
(56) Hamelberg, D.; McCammon, J. A. Standard free energy of releasing a localized water molecule from the binding pockets of protein: double-decoupling method. *J. Am. Chem. Soc.* **2004**, *24*, 7683−7689.
(57) Helms, V.; Wade, R. C. Computational alchemy to calculate absolute protein-ligand binding free energy. *J. Am. Chem. Soc.* **1998**, *120*, 2710−2713.
(58) de Beer, S. B. A.; Vermeulen, N. P. E.; Oostenbrink, C. The role of water molecules in computational drug design. *Curr. Top. Med. Chem.* **2010**, *10*, 55−66.
(59) Huang, N.; Shoichet, B. K. Exploiting ordered waters in molecular docking. *J. Med. Chem.* **2008**, *51*, 4862−4865.
(60) Villacanas, O.; Madurga, S.; Giralt, E.; Belda, I. Explicit treatment of water molecules in protein-ligand docking. *Curr. Comput.-Aided Drug Des.* **2009**, *5*, 145−154.
(61) Ellermann, M.; Jakob-Roetne, R.; Lerner, C.; Borroni, E.; Schlatter, D.; Roth, D.; Ehler, A.; Rudolph, M. G.; Diederich, F. Molecular recognition at the active site of the catechol-o-methyl-transferase: energetically favorable replacement of a water molecule imported by a bisubstrate inhibitor. *Angew. Chem., Int. Ed.* **2009**, *48*, 9092−9096.
(62) Forli, S.; Olson, A. J. A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. *J. Med. Chem.* **2012**, *55*, 623−638.
(63) Rossato, G.; Ernst, B.; Vedani, A.; Smiesko, M. AcquaAlta: a directional approach to the solvation of ligand-protein complexes. *J. Chem. Inf. Model.* **2011**, *51*, 1867−1881.
(64) Rechfeld, F.; Gruber, P.; Hofmann, J.; Kirchmair, J. Modulators of protein-protein interactions − novel approaches in targeting protein kinases and other pharmaceutically relevant biomolecules. *Curr. Top. Med. Chem.* **2011**, *11*, 1305−1319.
(65) Waszkowycz, B.; Clark, D. E.; Gancia, E. Outstanding challenges in protein-ligand docking and structure-based virtual screening. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 229−259.
(66) Wong, S. E.; Lightstone, F. C. Accounting for water molecules in drug design. *Expert Opin. Drug Discovery* **2011**, *6*, 65−74.
(67) Clark, D. E. What has virtual screening ever done for drug discovery? *Expert Opin. Drug Discovery* **2008**, *3*, 841−851.
(68) Bottcher, J.; Blum, A.; Heine, A.; Diederich, W. E.; Klebe, G. Structural and kinetic analysis of pyrrolidine-based inhibitors of the drug-resistant Ile84Val mutant of HIV-1 protease. *J. Mol. Biol.* **2008**, *383*, 347−357.
(69) Raub, S.; Steffen, A.; Kamper, A.; Marian, C. M. AIScore − chemically diverse empirical scoring function employing quantum chemical binding energies of hydrogen-bonded complexes. *J. Chem. Inf. Model.* **2008**, *48*, 1492−1510.
(70) Amadasi, A.; Surface, J. A.; Spyrakis, F.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. Robust classification of "relevant" water molecules in putative protein binding sites. *J. Med. Chem.* **2008**, *51*, 1063−1067.
(71) Graneto, M. J.; Kurumbail, R. G.; Vazquez, M. L.; Shieh, H. S.; Pawlitz, J. L.; Williams, J. M.; Stallings, W. C.; Geng, L. F.; Naraian, A. S.; Koszyk, F. J.; Stealey, M. A.; Xu, X. D. D.; Weier, R. M.; Hanson, G. J.; Mourey, R. J.; Compton, R. P.; Mnich, S. J.; Anderson, G. D.; Monahan, J. B.; Devraj, R. Synthesis, crystal structure, and activity of pyrazole-based inhibitors of p38 kinase. *J. Med. Chem.* **2007**, *50*, 5712−5719.
(72) Kroemer, R. T. Structure-based drug design: docking and scoring. *Curr. Protein. Pept. Sci.* **2007**, *50*, 5712−5719.
(73) Monecke, P.; Borosch, T.; Brickmann, R.; Kast, S. M. Determination of the interfacial water content in protein-protein complexes from free energy simulations. *Biophys. J.* **2006**, *3*, 841−850.
(74) Hattotuwagama, C. K.; Davies, M. N.; Flower, D. R. Receptor-ligand binding sites and virtual screening. *Curr. Med. Chem.* **2006**, *13*, 1283−1304.
(75) Branson, K. M.; Smith, B. J. The role of virtual screening in computer-aided structure-based drug design. *Aust. J. Chem.* **2004**, *57*, 1029−1037.
(76) Knegtel, R. M. A.; Robinson, D. D. A role for hydration in interleukin-2 inducible T cell kinase (lth) selectivity. *Mol. Inf.* **2011**, *30*, 950−959.
(77) Loving, K.; Alberts, I.; Sherman, W. Computational approaches for fragment-based and de novo design. *Curr. Top. Med. Chem.* **2010**, *10*, 14−32.
(78) Meanwell, N. A. Synopsis of some recent tactical application of bioisosteres in drug design. *J. Med. Chem.* **2011**, *54*, 2529−2591.
(79) Kutchukian, P. S.; Shakhnovich, E. I. De novo design: balancing novelty and continued chemical space. *Expert Opin. Drug Discovery* **2010**, *5*, 789−812.
(80) Huang, H. C.; Jupiter, D.; Qiu, M.; Briggs, J. M.; VanBuren, V. Cluster analysis of hydration waters around the active sites of bacterial alanine racemase using a 2-ns MD simulation. *Biopolymers* **2008**, *89*, 210−219.
(81) Zheng, M.; Li, Y.; Xiong, B.; Jiang, H.; Shen, J. Water PMF for predicting the properties of water molecules in protein binding site. *J. Comput. Chem.* **2013**, *34*, 583−592.
(82) de Courcy, B.; Piquemal, J.-P.; Garbay, C.; Gresh, N. Polarizable water molecules in ligand-macromolecule recognition. Impact on the relative affinities of competing pyrrolopyrimidine inhibitors for FAK kinase. *J. Am. Chem. Soc.* **2010**, *132*, 3312−3320.
(83) Gresh, N.; de Courcy, B.; Piquemal, J.-P.; Foret, J.; Courtiol-Legourd, S.; Salmon, L. Polarizable water networks in ligand-metalloprotein recognition. Impact on the relative complexation energies of Zn-dependent phosphomannose isomerase with D-mannose 6-phosphate surrogates. *J. Phys. Chem. B* **2011**, *115*, 8304−8316.
(84) Calderon-Kawasaki, K.; Kularatne, S.; Li, Y. H.; Noll, B. C.; Scheidt, W. R.; Burns, D. H. Synthesis of urea picket porphyrins and their use in the education of the role buried solvent plays in the selectivity and stoichiometry of anion binding receptors. *J. Org. Chem.* **2007**, *72*, 9081−9087.
(85) Mikol, V.; Papageorgiou, C.; Borer, X. The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporin: synthesis, biological activity, and crystallographic analysis with cyclophilin A. *J. Med. Chem.* **1995**, *38*, 3361−3367.

(86) Seo, J.; Igarashi, J.; Li, H.; Martásek, P.; Roman, L. J.; Poulos, T. L.; Silverman, R. B. Structure-based design and synthesis of Nw-nitro-L-arginine-containing peptidomimetics as selective inhibitors of neuronal nitric oxide synthase. Displacement of the heme structural water. *J. Med. Chem.* **2007**, *50*, 2089−2099.
(87) Andaloussi, M.; Henriksson, L. M.; Wieckowska, A.; Lindh, M.; Bjorkelid, C.; Larsson, A. M.; Surech, S.; Iyer, H.; Srinivasa, B. R.; Bergfors, T.; Unge, T.; Mowbray, S. L.; Larhed, M.; Jones, T. A.; Karlen, A. Design, synthesis, and X-ray crystallographic studies of alpha-aryl substituted fosmidomycin analogues as inhibitors of mycobacterium tuberculosis 1-deoxy-D-xylulose 5-phosphate reductoisomerase. *J. Med. Chem.* **2011**, *54*, 4964−4976.
(88) Ekstrom, J. L.; Pauly, T. A.; Carty, M. D.; Soeller, W. C.; Culp, J.; Danley, D. E.; Hoover, D. J.; Treadway, J. L.; Gibbs, E. M.; Fletterick, R. J.; Day, Y. S. N.; Day; Myszka, D. G.; Rath, V. L. Structure-activity analysis of the purine binding site of human liver glycogen phosphorylase. *Chem. Biol.* **2002**, *9*, 915−924.
(89) Trujillo, J. I.; Kiefer, J. R.; Huang, W.; Day, J. E.; Moon, J.; Jerome, G. M.; Bono, C. P.; Kornmeier, C. M.; Williams, M. L.; Kuhn, C.; Rennie, G. R.; Wynn, T. A.; Carron, C. P.; Thorarensen, A. Investigation of the binding pocket of human hematopoietic prostaglandin (PG) D2 synthase (hH-PGDS): A tale of two waters. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 3795−3799.
(90) Hann, M. M.; Keserü, G. M. Finding the sweet spot: the role of nature and nurture in medicinal chemistry. *Nat. Rev. Drug Discovery* **2012**, *11*, 355−365.
(91) Konc, J.; Janežič, D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* **2010**, *26*, 1160−1168.
(92) 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.
(93) Hopkins, A. L.; Groom, C. R. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* **2004**, *9*, 430−431.
(94) Hetényi, C.; Maran, U.; García-Sosa, A. T.; Karelson, M. Structure-based calculation of drug efficiency indices. *Bioinformatics* **2007**, *23*, 2678−2685.
(95) García-Sosa, A. T.; Sild, S.; Maran, U. Design of multi-binding-site inhibitors, ligand efficiency, and consensus screening of avian influenza H5N1 wild-type neuraminidase and of the oseltamivir-resistant H274Y variant. *J. Chem. Inf. Model.* **2008**, *48*, 2074−2080.
(96) García-Sosa, A. T.; Sild, S.; Maran, U. Docking and virtual screening using distributed grid technology. *QSAR Comb. Sci.* **2009**, *28*, 815−821.
(97) García-Sosa, A. T.; Hetényi, C.; Maran, U. Drug efficiency indices for improvement of molecular docking scoring functions. *J. Comput. Chem.* **2010**, *31*, 174−184.
(98) García-Sosa, A. T.; Oja, M.; Hetényi, C.; Maran, U. DrugLogit: Logistic discrimination between drugs and nondrugs including disease-specificity by assigning probabilities based on molecular properties. *J. Chem. Inf. Model.* **2012**, *52*, 2165−2180.
(99) Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBBind database: methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111−4119.
(100) García-Sosa, A. T.; Maran, U.; Hetényi, C. Molecular property filters describing pharmacokinetics and drug binding. *Curr. Med. Chem.* **2012**, *19*, 1646−1662.
(101) García-Sosa, A. T.; Oja, M.; Hetényi, C.; Maran, U. Disease-specific differentiation between drugs and non-drugs using principal component analysis of their molecular descriptor space. *Mol. Inf.* **2012**, *31*, 369−383.
(102) García-Sosa, A. T.; Maran, U. Drugs, non-drugs, and disease category specificity: organ effects by ligand pharmacology. *SAR QSAR Environ. Res.* **2013** [online] DOI: http://dx.doi.org/10.1080/1062936X.2013.773373.
(103) Davis, A. M.; St-Gallay, S. A.; Kleywegt, G. J. Limitations and lessons in the use of X-ray structural information in drug design. *Drug Discovery Today* **2008**, *13*, 831−841.
(104) *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria. http://www.R-project.org (accessed January 1, 2011).
(105) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668−D672.
(106) Berman, H. M.; Westbrook, J. D.; 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.
(107) Wang, R.; Gao, Y.; Lai, L. Calculating partition coefficient by atom-additive method. *Perspect. Drug Discovery Des.* **2000**, *19*, 47−66.
(108) *Marvin Beans*, version 5.6.0.1; ChemAxon: Budapest, Hungary, 2011.
(109) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery* **2007**, *6*, 881−890.
(110) Chen, Y.; Shoichet, B. K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2009**, *5*, 358−364.
(111) Lovering, F.; Bikker, J.; Humblet, C. Escape from flatland: Increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **2009**, *52*, 6752−6756.