JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Accounting for Intraligand Interactions in Flexible Ligand Docking with a PMF-Based Scoring Function

A. Y. Lizunov,*[,†,‡] A. L. Gonchar,[‡] N. I. Zaitseva,[¶] and V. V. Zosimov[†]

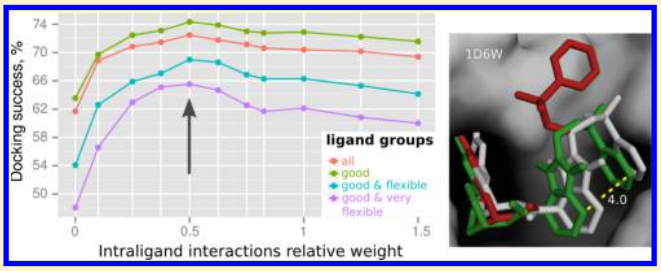
[†]Department of Mathematics, Moscow Institute of Physics and Technology, Moscow 117303, Russia

[‡]Faculty of Fundamental Medicine, Moscow State University, Moscow 119991, Russia

[¶]The Faculty of Pharmacy, First Moscow State Medical University, Moscow 119991, Russia

Supporting Information

ABSTRACT: We analyzed the frequency with which intraligand contacts occurred in a set of 1300 protein−ligand complexes [Plewczynski et al. *J. Comput. Chem.* **2011**, *32*, 742−755.]. Our analysis showed that flexible ligands often form intraligand hydrophobic contacts, while intraligand hydrogen bonds are rare. The test set was also thoroughly investigated and classified. We suggest a universal method for enhancement of a scoring function based on a potential of mean force (PMF-based score) by adding a term accounting for intraligand interactions. The method was implemented via in-house developed program, utilizing an Algo_score scoring function [Ramensky et al. *Proteins: Struct., Funct., Genet.* **2007**, *69*, 349−357.] based on the Tarasov-Muryshev PMF [Muryshev et al. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 597−605.]. The enhancement of the scoring function was shown to significantly improve the docking and scoring quality for flexible ligands in the test set of 1300 protein−ligand complexes [Plewczynski et al. *J. Comput. Chem.* **2011**, *32*, 742−755.]. We then investigated the correlation of the docking results with two parameters of intraligand interactions estimation. These parameters are the weight of intraligand interactions and the minimum number of bonds between the ligand atoms required to take their interaction into account.



1. INTRODUCTION

Computer simulation is an essential stage in modern drug development. "In silico" modeling includes various calculations, such as molecular dynamics simulations, QSAR analysis of databases, various ADMET properties prediction, and other modeling tasks. One widely used type of computer simulation is the modeling of protein−ligand interaction with the aim of binding energy and binding conformation mode prediction. In this study we consider the computer simulation of protein−ligand interaction in approximation of a rigid protein and flexible ligand. Two main problems arise from this approximation. First, docking: the problem of determining correct binding conformation of a ligand. Second, virtual screening: the problem of sorting ligands according to their binding energy with a protein. Although good results have been achieved[1,4] in this area, modeling protein−ligand interaction is still a challenging task, and there remains much room for improvement, especially for large and flexible ligands.[5−7]

Estimation of protein−ligand binding energy is made by a scoring function, which assigns a score to a given spatial conformation of a protein−ligand complex. Scoring functions are usually divided into three classes: force field-based, empirical, and knowledge-based.[8,9] This separation is not very strict. For example, empirical and knowledge-based scoring functions may include some force field terms, and construction procedure for knowledge-based scoring functions is sometimes similar to the construction of empirical scoring functions.[3,10]

In this study we consider the class of scoring functions based on a potential of mean force: so-called PMF-based scores. A potential of mean force (PMF) describes the average energy of interaction of two typed atoms, resulting from distance. PMF-based scores include all knowledge-based scores but can also include other types of scores, where PMF is derived from something other than statistical distribution of protein−ligand atom pairs. PMF-based scoring functions are discussed in detail in subsections 1.2 and 1.2.1 below.

We propose a method for enhancement of a PMF-based scoring function by taking into account intraligand interactions. Intraligand interactions are proposed to be accounted by the same potential of mean force which is used in the base scoring function. We implemented the approach for an Algo_score scoring function[2] with a Tarasov-Muryshev PMF.[3]

An essential stage of a scoring function development is validation of its quality when applied to practical problems. One usually uses a test set to validate the quality of scoring functions and to compare various scoring functions and docking algorithms. Test sets are comprised of protein−ligand pairs with known binding energy or complex spatial structure, or both.

Received: March 24, 2015
Published: September 29, 2015

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

To validate the efficiency of the proposed scoring function improvement method, we chose a test set of 1300 protein−ligand complexes[1] with both known spatial structure and binding energy. The test set was thoroughly analyzed and classified as described in the section 2.2.2 below. Frequencies of various types of intraligand contacts in the test set were also analyzed apart from the docking problem (section 3.1).

1.1. The Role of Intraligand Interactions in Protein−Ligand Complex Formation. The influence of intraligand contacts on protein−ligand complex formation was investigated mostly in connection with the estimation of ligand's internal energy increase in a protein bound conformation relative to the minimum of internal ligand's energy in the absence of the protein. The focus on the ligand's internal energy increase upon binding arose from the fact that some docking programs disregard the flexibility of the ligand, using a set of previously generated ligand conformations instead.[11] Users of such programs are faced with the question of how much ligand's internal energy increase can take place, while still not preventing the ligand conformation from binding to the protein. That means the ligand's internal energy penalty in the given conformation can be compensated by ligand−protein interaction energy.

The ligand's internal interactions affect the energy difference between various ligand conformations. In the absence of a protein, the ligand will most often take the conformation with the minimum of the ligand's internal energy with respect to surrounding medium - vacuum or solvent. Ligand conformation is significantly affected by interactions with the protein upon binding. Ligands tend to get more stretched conformations in complexes, with hydrophobic areas open for interaction with the protein.[12] According to refs 12−14, the energy of ligand conformation in a protein−ligand complex is usually higher than the minimum energy of a standalone conformation by less than 5 kcal/mol. However, in approximately 10% of the cases, the energy difference can exceed 9 kcal/mol.

Nevertheless, in a number of studies, it was shown that ligand conformation in complex with a protein is sometimes stabilized by intraligand interactions. So, for example, an inhibitor of HIV-1 reverse transcriptase PETT forms an intraligand hydrogen bond in the complex with the protein (PDB ID 1dtt[15]), and the 11-residue peptide in the complex with opsin protein adopts a helical conformation, which is held by intraligand hydrogen bonds (PDB ID 3dqb[16]). Another example with a peptide inhibitor was presented in ref 17. In this work, it was specifically noted that a 16-residue peptide's binding affinity with a BCL-XL protein depends on the peptide helix propensity, which, in turn, is determined by intraligand interactions.

1.2. PMF-Based Scoring Functions. PMF-based scoring functions are based on potentials of mean force (PMF), which estimate the interaction energy of two atoms of given types at a given distance, neglecting other characteristics of the atom's environment and the atoms' mutual orientation.

The most common examples of PMF-based scoring functions are statistical (or knowledge-based) scoring functions. In statistical scores, PMF is derived from the statistical distribution of distances between atoms of the given types in experimental data: usually the experimentally determined spatial structures from the PDB (Protein Data Bank[18]) are used. The derivation of PMF is based on the inverse Boltzmann law, which relates the energy U of a system state with the probability P of the system to occur in this state: $U = -kT \ln(P)$ (up to constant summand). The statistic of interatomic distances is used to estimate the probabilities P of corresponding states.[19,20] The Boltzmann law is stated for an ensemble of noninteracting particles in thermodynamic equilibrium. Strictly speaking, it is not applicable to the case of a large number of protein−ligand complexes from the PDB.[21−23] In this case it is accurate to call it Boltzmann hypothesis. Nevertheless, the practice shows that the scoring functions based on this hypothesis perform quite well.[24,25]

PMF-based scores are not limited to statistical scores. The underlying PMF can be derived from observations other than statistical distributions. For example, to construct a Tarasov-Muryshev (TM) PMF,[3] a procedure was used that was more like an empirical scoring function construction. A general form of PMF with several parameters was considered, and the values for the parameters were determined by fitting the predicted binding energies of 164 protein−ligand complexes to experimentally known values.

The final PMF-based score of protein−ligand interaction is calculated, in the simplest case, as the sum of interatomic interactions over all pairs of atoms, one from the protein and another from the ligand

$$S = \sum_{l \in lig, p \in prot} U_{t(l)t(p)}(r_{lp}) \quad (1)$$

where l and p denote atoms in the ligand and in the protein, $U_{t(l)t(p)}$ - PMF for atoms of types $t(l)$ and $t(p)$, and r_{lp} - the distance between the atoms.

Derivation of a PMF is based on averaging binding energy over all parameters except for the distance between ligand and protein atoms. This simple averaging approach has the advantage of an implicit account of all types of protein−ligand interactions, including Van der Waals interactions, hydrogen bond formation, and other types of interactions, which are not distinguished as separate terms in force fields. Besides that, PMF-based scoring functions can implicitly account interaction with solvent (hydrophobic interactions) and even a part of the entropy component of free energy.[24]

Weak points of PMF-based scoring functions can be separated into two groups: general disadvantages of PMF-based scores and imperfections of a particular PMF. The main general disadvantage of these scoring functions is the averaging of the "real" interaction function over all other parameters, describing the protein−ligand complex geometry, except the pairwise interatomic distances. For example, the omission of spatial orientation of atomic electron orbitals leads to a false account of hydrogen bonds in the protein−ligand complex conformations; there the hydrogen bond actually could not be formed due to wrong angles between covalent bonds, but potential hydrogen bond donor and acceptor atoms are located at the correct distance for hydrogen bond formation. This effect of false hydrogen bonds account is partially compensated by account of interactions with other atoms in the vicinity. Another general disadvantage of PMF-based scores is failure to account for intraligand energy change.

Imperfections of a particular PMF result from the PMF's calculation method, such as an insufficiently detailed atom typification or the Boltzmann approximation used for PMF derivation. The approximation of Boltzmann dependency between the frequency of protein−ligand typed atom pair occurrence at a given distance and energy of their interaction at that distance is distorted by a number of factors. An example of a factor distorting the Boltzmann distribution is the existence of covalent bonds between atoms in molecule. If a couple of atom types do not participate in a noncovalent attractive interaction like hydrogen bond formation, but atoms of these types are often

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

covalently bonded to atoms forming an hydrogen bond, then one will see a maximum in distance-dependent atom frequencies, which is not connected with interaction energy minimum of atoms of these two types but with their covalent neighbors interaction energy minimum. This is an example of how covalent bonds can give rise to an artificial maximum in distance-dependent atom frequency distribution, not associated with the two atom types interaction energy minimum.

There are a number of enhanced PMF-based scores, which includes additional terms and modifications. The enhanced PMF-based scores are thoroughly discussed below, but all of them lack a precise account of intraligand interactions. Overlooking intraligand interactions does not alter docking results much for small and rigid ligands, but it may cause significant deviations in docking and scoring of larger ligands.

1.2.1. Enhanced PMF-Based Scores. PMF-based scoring functions of the simplest form (eq 1) do not account for intraligand interactions or ligand−solvent interactions explicitly. The implicit account of these types of interactions is also questionable, because protein−ligand contacts seem to have no way to reflex the ligand−ligand interatomic interactions or ligand−solvent interactions. Nevertheless some scoring functions are limited to the simplest form and perform pretty well. So, PMF,[24] TM-Score,[3] PLASS,[26] M-Score,[27] and ITScore[28] scoring functions include only the sum of PMF estimations of interatomic interactions between protein and ligand atoms (have the form of eq 1).

To improve PMF-based scores, additional summands to eq 1 are often used. These summands usually serve to estimate ligand's internal energy change, ligand−solvent interaction energy, or entropy contributon to free energy.

Gehlhaar et al. developed a model of intermolecular recognition for flexible ligand docking into a protein active site - PLP scoring function.[29] This scoring function accounts for both inter- and intramolecular interactions. Ligand−protein interaction is evaluated by a PMF, intraligand energy is evaluated by two summands for torsion energy of sp^3-sp^3 and sp^2-sp^3 bonds and a very large penalty of 10000 units for ligand self-intersections (when the distance between two unbounded ligand atoms is less than 2.35 Å). The PLP scoring function can be written as

$$\sum_{l,p} U_{pl}(r_p l) + \sum_{rot} A(1 - \cos(n\phi - \phi_0)) + N_{clash}{*}Penalty$$

The ASP[30] scoring function utilizes summands from the ChemScore[31] empirical scoring function to estimate ligands' internal energy. Just as in the PLP scoring function, additional summands account only for torsion energy of rotatable bonds and ligand self-intersections. The formula for the ASP score is

$$\sum_{l,p} U_{pl}(r_{pl}) + \Delta E_{lig} + \Delta E_{clash}$$

.

During the calculation of the Bleep[32] scoring function, the conformation of the protein−ligand complex is filled up by explicit water molecules. These water molecules are accounted for in summation of interatomic interactions. The Bleep PMF used in the summation is also constructed in reliance on X-ray structures with artificially added water molecules. A large variety of X-ray structures with a homology account were used during the Bleep PMF construction. There are two versions of the Bleep scoring function; in the water-exclusive one, no explicit water molecules were used. The formula for the Bleep score can be written as

$$\sum_{l,p} U_{pl}(r_{pl}) + \Delta \sum_{l,w} U_{lw}(r_{lw}) + \Delta \sum_{p,w} U_{pw}(r_{pw})$$

In the DrugScore[22] scoring function, the sum of pairwise interatomic interactions between ligand and protein is added to a summand that accounts for interaction with the solvent. A monatomic potential is introduced to account for interaction with the solvent, and the potential depends on the SAS (solvent accessible surface) value for the atom. The formula for DrugScore takes the form of

$$(1 - \gamma) \sum_{l,p} U_{pl}(r_{pl}) + \gamma \sum_{l} \Delta W_l(\mathrm{SAS}, \mathrm{SAS}_0) + \gamma \sum_{p} \Delta W_p(\mathrm{SAS}, \mathrm{SAS}_0)$$

Like in DrugScore, in the improved version of ITScore[28] - ITScore/SE[33] protein and ligand interaction with the solvent is accounted for using a SAS parameter. Besides that, two additional summands are introduced in the score to account for the entropic contribution to the free energy. So the formula for ITScore/SE takes the form

$$\sum_{l,p} U_{pl}(r_{pl}) + \sum_{i} \sigma_i \Delta(\mathrm{SAS}) + \alpha \Delta S_{conf} + \beta \Delta S_{vib}$$

Zheng and Merz developed a PMF-based scoring function[10] where a novel knowledge-based and empirical combined scoring algorithm (KECSA) was used to derive the PMF. In addition to the PMF-based term for enthalpy estimation the scoring function includes an empirically derived term for entropy estimation. This entropy term is based on 9 parameters of ligand, including the number of rotatable bonds, the nonpolar buried surface area, total buried surface area, and others. So the scoring function can be written in the form of an equation

$$\sum_{l,p} U_{pl}(r_{pl}) + \Sigma_{entropy}$$

.

Summing up, PMF-based scoring functions are usually enhanced either by taking the torsion energy of ligand rotatable bonds into account (PLP,[29] ASP[30]) or by explicit estimation of solvent influence (DrugScore,[22] Bleep,[32] ITScore/SE[33,34]). A more detailed overview of the enhanced PMF-based scores is provided in the Supporting Information.

2. MATERIALS AND METHODS

2.1. Method for Accounting for Intraligand Interactions. First, we should note that in the intraligand interactions below, the ligand−solvent interaction is also implied. This is due to the fact that hydrophobic intraligand interactions are in fact the ligand−solvent interactions accounted for implicitly.

In the previous section 1.2.1 it was shown that PMF-based scoring functions either have no explicit terms for intraligand interactions account or have limited account based on rotatable bonds energy, ligand clashes, and a SAS parameter. However, as mentioned in section 1.1 above, intraligand interactions can play a role in the ligand−protein binding process.

The nature of intraligand interactions between atoms located far enough apart in the molecule structure is the same as intermolecular interactions between atoms of the protein and

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

ligand. Ligand atoms can form hydrogen bonds and hydrophobic contacts the same way as between ligand and protein. On the basis of the above, we suggest to estimate intraligand interatomic interactions by the same PMF that is used to estimate protein−ligand interaction. This estimation makes sense only for a pair of atoms, separated by a sufficient number of covalent bonds in the molecule structure to allow the two atoms to move against each other without regard to restrictions of covalent bond lengths and angles.

For a base scoring function S_0 the suggested scoring function modification can be written in the form

$$S_0 \rightarrow S_0 + \alpha \sum_{\substack{a,b\in lig \\ d(a,b)\geq n}} U_{t(a)t(b)}(r_{ab}) \tag{2}$$

where a and b denote two atoms in the ligand, $U_{t(a)t(b)}$ - PMF for atoms of types $t(a)$ and $t(b)$, r_{ab} - distance between the atoms. Later we write U_{ab} instead of $U_{t(a)t(b)}$ to simplify the notation. $d(a,b)$ denotes the number of covalent bonds separating the atoms, so the sum is taken over all pairs of atoms a and b from the ligand, where a and b are separated by no fewer than n covalent bonds with at least one not adjacent bond being rotatable. The coefficient α in front of the intraligand interactions estimation defines the "weight" of intraligand interactions in the score. This coefficient allows us to analyze the influence of intraligand interactions on docking and scoring accuracy.

The proposed modification can be used both for the simplest PMF-based scores of the form eq 1 and for any other score. The question one should consider when applying the modification eq 2 to a scoring function is whether the intraligand or ligand−solvent interactions are already accounted for in the base scoring function S_0. If these interactions are already partially accounted for in the base score, one can consider using a smaller value for the intraligand interactions weight α. For example, it relates to enhanced PMF-based scores with additional SAS based terms (section 1.2.1).

The additional term in eq 2 does not account for interactions between ligand atoms located close to each other in the molecular structure. These interactions can hardly be decomposed into pairwise interactions and should be described by quantum-mechanical models. A common way to partially account for these "close" interactions is to estimate the torsion energy of rotatable bonds. Since the additional term in eq 2 estimates "far" interactions, but the torsion energy term used in some PMF-based scores (section 1.2.1) estimates "close" interactions, these two summands can be used simultaneously.

This modification (eq 2) will not change the score value for small or rigid ligands, because their atoms cannot move against each other and form internal contacts, but for large and flexible ligands the intraligand interactions can significantly influence the estimation of binding energy and predicted binding pose.

A weak point of the proposed scoring function improvement is an inability to calculate the additional summand in eq 2 on a grid around a rigid protein, because this summand depends on relative distances between ligand atoms instead of their positions relative to the protein. This fact may complicate the implementation of the improvement with grid-based scoring functions and may slow down the docking procedure. However, we did not notice a significant (more than by a factor of 2) slowdown of the docking procedure in docking tests described below in the article.

2.1.1. Implementation of the Proposed Modification with an Example Scoring Function. To analyze the effect of accounting for intraligand interactions, we implemented the proposed improvement (eq 2) in the Algo_score scoring function[2] with Tarasov-Muryshev potential of mean force (TM PMF).[3] The Algo_score is an enhanced PMF-based score originally calculated by the following formula

$$S_0 = \sum_{l\in lig}\left(C_l(\mathbf{r}_l)\sum_{p\in prot} U_{pl}(r_{pl})\right) + S_{clash} + S_{tors} \tag{3}$$

where $U_{pl}(r_{pl})$ is Tarasov-Muryshev PMF for a pair of atoms from protein and ligand, $C_l(\mathbf{r}_l)$ - a special coefficient named "cloud score" for the given position of the ligand atom in the protein vicinity or 1 if $\sum_p U_{pl}(r_{pl}) > 0$, i.e. if the ligand atom position is not energetically favorable. Cloud score is a special Algo_score feature. For a detailed description of cloud score and Algo_score refer to ref 2. Briefly, cloud score is a multiplier before the estimation of interaction energy between a ligand atom and the protein. The multiplier takes values from 0 to 1 and shows if an atom with the same type occurs in a ligand molecule in a PDB structure surrounded with a similar protein environment. Subscripts in $U_{pl}()$ and $C_l()$ show that the potentials depend on types of atoms in TM typification.

S_{clash} in eq 3 is a penalty for molecule self-intersection that occurs if a pair of ligand atoms in the given position are closer to each other than the sum of their Van der Waals radii; S_{tors} is an estimation of torsion tensions in the ligand calculated by Amber potential.[35]

Eq 3 shows that in the original Algo_score, intraligand interactions are estimated as sum of penalties for ligand self-intersections and the torsion energy of rotatable bonds. In many cases such an estimation is quite sufficient. For example, if the ligand has a few rotatable bonds or if the structure of the protein site does not allow ligand atoms to approach each other and to form a noncovalent contact. That is why the Algo_score showed good results in tests.[2] However, as noted above, for more flexible ligands, noncovalent intraligand interactions such as hydrogen bonds and hydrophobic interactions can significantly affect the binding pose and binding energy of a complex.

With the proposed modification, the Algo_score takes the form

$$\begin{aligned} S_{new} &= S_0 + \alpha \sum_{\substack{a,b\in lig \\ d(a,b)\geq n}} U_{ab}(r_{ab}) \\ &= \sum_{l\in lig}\left(C_l(\mathbf{r}_l)\sum_{p\in prot} U_{pl}(r_{pl})\right) + S_{clash} + S_{tors} + \alpha \sum_{\substack{a,b\in lig \\ d(a,b)\geq n}} U_{ab}(r_{ab}) \end{aligned} \tag{4}$$

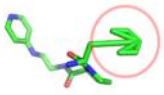
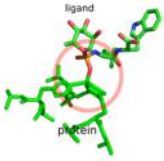
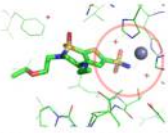
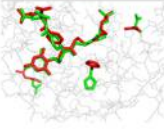
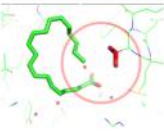
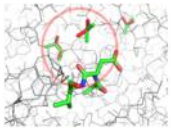
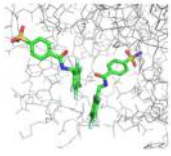
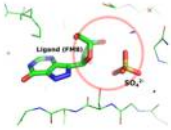
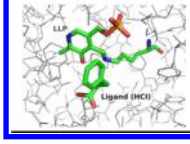
where S_0 and S_{new} are estimations of ligand−protein interaction energy at the given spatial conformation, calculated by original and by modified Algo_score scoring functions, respectively; $U_{ab}(r_{ab})$ is the TM PMF for a pair of ligand atoms a and b with TM typification. The condition $d(a,b) \geq n$ denotes that the sum is taken over pairs of ligand atoms, separated by no fewer than n covalent bonds with at least one of the central bonds being rotatable.

The weight of intraligand interactions α in the case of Algo_score have an additional role to compensate for the cloud score multiplier C_l used with ligand−protein interactions.

To determine the effect of intraligand interactions accounting, we varied the weight of intraligand interactions α and the minimum number of bonds between ligand atoms required to account for their interaction n. The results of the variations are described below in section 3.2.

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

Table 1. Description of Special Cases in the Test Set Complexes

	Special case type	Count	Processing method
	Mistakes in ligand molecule coordinates	2	Ligands are corrected manually
	Ligand is covalently bonded with protein	8	Ligand is cut from protein and docked as a separate molecule
	Ligand have a contact with metal ion (at distance less than 5Å)	208	Remove ions, neglect them while docking
	Both protein and ligand have alternative locations and there is a location with matching alternate location indicator in PDB-file for protein and ligand	97	Alternative location with matching alternate location indicator for both ligand and protein is used
	Only protein have alternative locations. Alternative locations may be in interaction site (at distance less than 5Å from ligand) or out of the site	314	One of the protein alternative locations is considered
	There are alternative locations marked with different alternate location indicator for protein and ligand	4	First alternative location for both ligand and protein is considered
	Part of ligand has alternative locations	4	One of the ligand alternative locations is considered
	More than one instance of ligand molecule are presented in interaction site (at a distance less than 5Å from each other). It is either dimer ligands or errors in PDB-file	13	One instance of the ligand is considered
	Ligand has a contact with small non-peptide molecule (less than 6 heavy atoms, water is not considered)	95	Contacts with non-peptide molecules are disregarded
	Ligand has a contact with large non-peptide molecule (more than 6 heavy atoms)	46	Contacts with non-peptide molecules are disregarded

2.2. Docking Programs Testing and Verification. *2.2.1. Overview.* Evaluating the quality of docking and scoring procedures is an essential constituent of docking programs development. We distinguish three types of docking program testing procedures: self-docking tests, virtual screening tests, and correlation tests. A test set is required to carry out a test of any of the above types, but the constituents of these sets differ for different test types.

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

One needs to know the experimentally determined spatial structure for a set of protein−ligand complexes to carry out a self-docking test. In a self-docking test, the experimentally determined protein−ligand complex conformations are compared with conformations predicted by a docking program. The success of a self-docking experiment is usually measured as the fraction of successfully docked protein−ligand complexes in the test set. Self-docking tests are used, for example, in refs 1 and 36−38.

To carry out a virtual screening test, one needs to know the spatial structure of a protein active site and experimentally determined binding constants with the protein for a set of ligands. The ligand set is sorted two ways: by experimentally determined binding energies and by calculated binding estimations. Two resulting orderings of the ligand set are compared, and their similarity is considered as the measure of success of the virtual screening experiment. A virtual screening experiment is often implemented in simplified form, where only binary information "binds - does not bind" is accounted for, instead of experimental values of binding energies. In this case, the fake ligands, which do not actually bind with the protein, are usually called "decoys". Virtual screening tests are used, for example, in refs 39−41. In this work we do not perform any virtual screening tests.

Correlation tests are less frequently used. Like in a virtual screening test, in a correlation test, the experimentally determined protein−ligand binding constants (K_d) are compared with calculated estimations of the binding energies, but unlike a virtual screening case, binding constants with different proteins are usually used in this type of test. The correlation between experimentally determined K_d and calculated scores is the measure of success in this type of testing experiment. To carry out a correlation test, one needs to have a set of protein−ligand complexes with spatial structures and binding energies both known. We presume that the lack of data about complexes with K_d and spatial structure is the main reason why correlation tests are rarely used. Another reason to avoid this type of testing experiment is the fact that constants K_d for different proteins are usually determined by different methods and work groups, so their comparison is questionable. Yet another argument against this type of experiment is the fact that the scoring procedure is often optimized for a certain type of protein, so it may give a good correlation with experimental K_d for ligands binding this particular protein (virtual screening case), but this correlation may get worse when one tries to compare scores calculated for different proteins. Nevertheless an "ideal" scoring function must score a protein−ligand complex higher, if it has higher K_d, hence the correlation is a reasonable rating criterion for scoring functions. We found only several articles where correlation criterion was used for testing and comparison of various scoring functions,[1,5,25,42] but the correlation between K_d and calculated scores is also sometimes used for optimization of score parameters.[43,44]

2.2.2. Verification Test Set. To determine the effect of modification of the scoring function, we use the test set of 1300 protein−ligand complexes with known spatial structure and binding constants.[1] This test set is based on the PDBbind database[45] and is unique in the number of complexes with both types of information known. Along with the extensive self-docking test, this test set allows for us to perform the correlation test.

To extract molecules from PDB files, a pipeline was developed. The pipeline includes parsing of PDB headers, syntax comparison of PDB group names with the given name of the ligand, analysis of PDB LINK and CONECT and CONECT sections, analysis of distances between atoms, and analysis of alternative positions of the ligand and protein. We implemented the pipeline in Python, with use of Babel[46] routines. Finally, each ligand was named as a sequence of PDB-group names, composing the ligand. The full list of ligand names is provided in the Supporting Information.

There is a number of special cases in the source PDB files of the 1300 test complexes. A short description of the special cases is presented in Table 1, while the full list of corresponding PDB IDs is supplied in the Supporting Information.

During the test set preparation, we identified a number of complexes with limited applicability for docking tests. There are a couple of cases where the ligand has an error in its geometrical structure and 8 cases of ligands covalently bonded with the protein. In 208 cases, the ligand has a contact with a metal ion (hereinafter a contact with a ligand means location at a distance less than 5 Å from a non-hydrogen atom of a ligand). For the scoring functions not designed for use with metalloproteins, these complexes may distort the test results.

In 415 cases, the ligand or protein has alternative locations in the PDB file. Among them we can distinguish 97 cases where the protein and ligand have alternative locations with a matching alternate location indicator in the PDB file, making it possible to extract a matching protein−ligand conformation pair for docking. Further, in 139 of these cases, only the protein has an alternative location, and it is located out of the protein−ligand interaction site, so the choice of an alternative location will not alter the docking results. In the remaining 179 cases, either the ligand or protein has alternative conformations in the interaction site, and it is difficult to match a pair of protein and ligand conformations for the docking test. Therefore, the 179 complexes were not included into the *good* subset, defined in Table 2 below.

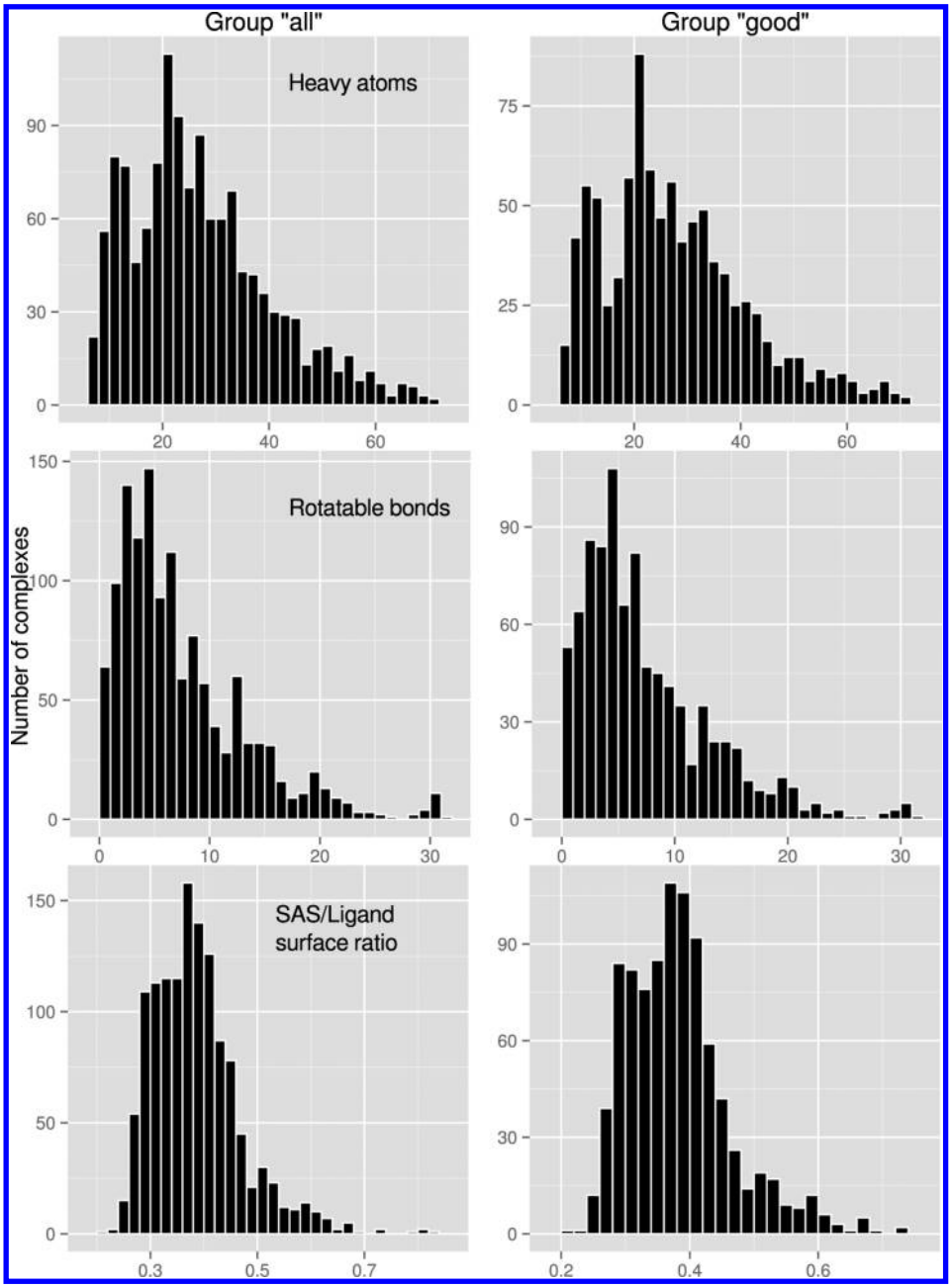
Table 2. Test Set Groups

subset short name	description	no. of complexes
all	all complexes	1300
good	• ligand and protein are not bonded covalently • ligand description has no mistake in source PDB file • there are no metal ions in active site • either there are no alternative locations of ligand and protein atoms in the area of contact or there are alternative locations with matching alternate location indicator for protein and ligand • ligand has no contacts with nonpeptide molecules with more than 5 non-hydrogen atoms	911
flex	more than 4 bonds is considered as rotatable in ligand by OpenBabel utility[46,47]	732
gflex	complexes both in *good* and *flex* group	516
fflex	more than 9 bonds is considered as rotatable in ligand by OpenBabel[46,47] utility	334
gfflex	complexes both in *good* and *fflex* group	235

There are cases where a ligand interacts not only with a protein but also with other nonpeptide molecules: 13 cases of dimer ligands (or errors in the PDB file, where a single ligand conformation is duplicated); 95 cases where a ligand has a contact with a small (less than 6 non-hydrogen atoms) nonpeptide molecule (except water molecules), the molecules

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

Chart 1. Distributions of Ligands in Groups *all* and *good*[a]



[a]The distributions over number of heavy atoms, over number of rotatable bonds, and over solvent available surface to ligand surface ratio are shown.

usually are nonmetal ions or detergent molecules; 46 cases where a ligand has a contact with a large nonpeptide molecule.

For purposes of the docking test, the source PDB files were converted into the final 1300 pairs of protein and ligand molecules, each in single conformation. Geometric mistakes were manually corrected, and covalently bonded ligands were detached from protein. In the case of alternative locations for the ligand or protein (not in the case of several protein−ligand complexes in a single PDB file, but in the case of ambiguity of X-ray structure interpretation, marked by a special alternate location indicator character in PDB format) a single conformation for the protein−ligand complex was chosen. Metal ions were discarded from the protein active site. The complexes with a broken covalent bond between protein and ligand, discarded metal ions in the active site, or complexes with ambiguous protein and ligand alternative conformations matching are not suited well for docking tests.

There was no explanation in the original work[1] as to how these problematic cases were handled. To get the docking results applicable for comparison with the results reported by Plewczynski et al., we use the whole set of 1300 complexes for docking. Additionally we consider a subset *good* of 911 complexes fully applicable for docking. All the considered subsets are described in Table 2: Detailed information is available in the Supporting Information.

The histograms in Chart 1 show the distributions of ligands in the test set groups *all* and *good* over the number of rotatable bonds, over the number of heavy (non-hydrogen) atoms, and over the relative solvent available surface (SAS) of the ligand, calculated in PyMOL.[48]

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

Table 3. Ligand Groups and Contacts[a]

group name	count of complexes	% of complexes with intraligand hbonds by Chimera[51]	% of complexes with intraligand hphob contacts	mean no. of hphob contacts	% of complexes with hphob >5	>10	>15
all	1300	6.7	65	7.96	35	23	18
good	911	6.9	67	8.75	37	25	19
gflex	516	9.3	88	14.5	59	44	34
gfflex	235	9.3	99.6	22.5	77	67	58

[a]hbonds - hydrogen bonds; hphob - hydrophobic contacts.

One can see from the histograms in Chart 1 that the distributions are qualitatively the same for subsets *all* and *good*.

The unusual maximum in distribution by rotatable bonds at the 30 rb area is due to multiple occurrences of the same ligands in the test set, so 0Q4 peptide-like ligand has 30 rb and occurs 3 times and 2NC peptide-like ligand has also 30 rb and occurs 5 times.

2.2.3. Success Criteria of the Test Computations. We use three criteria to estimate the success of a test docking computation:

dock_succ - the fraction of ligands, docked with RMSD (root-mean-square deviation) of calculated coordinates from experimentally determined coordinates of ligand atoms less than 2 Å. It is standard criterion used in multiple works.[1,2,42,49,50]

correl - the Spearman correlation between score and experimentally determined binding constant K_d. In some articles, authors use the Pearson correlation,[5] but we prefer using the Spearman correlation instead. In contrast to the Pearson correlation, the Spearman correlation compares the order of complexes, defined by a scoring function and by K_d, disregarding the values of the parameters. So the Spearman correlation will take the value of 1 if the orders match, even if the values of score and K_d do not match. The main requirement of scoring function is to order complexes by binding energy correctly, but it is not necessary to calculate binding energy value precisely, so the Spearman correlation is better suited for characterization of scoring functions. Despite all the arguments against correlation tests mentioned above in section 2.2, correlation criterion shows the positive effect of the changes in a scoring function proposed in this article as well as dock_succ criterion.

corr_succ - the Spearman correlation between score and K_d, calculated on a subset of successfully docked complexes. Successfully docked means with RMSD less than 2 Å from native position. corr_succ criterion is intended to show the quality of scoring function, unbiased by inaccuracy of docking algorithm. There is little sense in comparing K_d with the score for an incorrect ligand position.

3. RESULTS AND DISCUSSION

3.1. Analyze of Intraligand Interactions in the Test Set.

To gain an initial insight into the significance of intraligand interaction in the formation of protein−ligand complex conformation, we analyzed the frequencies of internal ligand contacts in the set of 1300 complexes we used.[1] We distinguish two types of intraligand contacts: hydrogen bonds and hydrophobic contacts.

The key values are summarized in Table 3. The main conclusion about the frequencies of intraligand contacts is that intraligand hydrogen bonds are rare, but intraligand hydrophobic contacts are frequent, especially in flexible ligands. Detailed analysis of the two types of intraligand contacts is found in the subsections below.

The analysis of ligand conformations in the test set showed that nearly 65% of all ligands and 88% of flexible ligands have intraligand contacts (Table 3). This finding clearly implies that accounting for intraligand interactions during the docking procedure can significantly alter the docking results.

3.1.1. Intraligand Hydrogen Bonds in the Test Set Complexes. Due to the complex nature of hydrogen bonds, which includes electrostatic, covalent, and Van der Waals components, the definition of hydrogen bond is nowadays ambiguous.[52] In this work we rely on the definition of hydrogen bonds implemented in the Chimera software,[51] the definition is based on the work of Mills and Dean.[53]

Some intraligand hydrogen bonds are determined by the ligand's covalent structure and would arise in any or almost any ligand conformation, we call such intraligand hydrogen bonds - "forced hydrogen bonds". There are two cases of forced hydrogen bonds: when the covalent structure of the ligand does not allow a donor and an acceptor to move relative to each other, and when the ligand is rich in hydrogen bond donors and acceptors and thereby most of ligand conformations will have intraligand hydrogen bonds (Figure 1). The forced intraligand

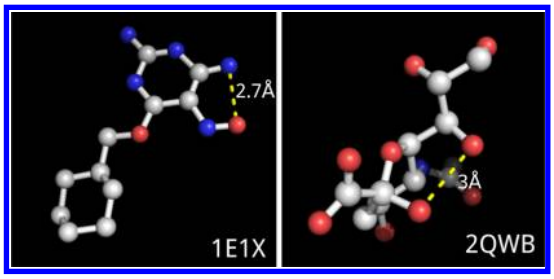


Figure 1. Two types of forced intraligand hydrogen bonds examples.

hydrogen bonds are of little interest for the docking studies, because taking them into account would change the score of any ligand conformation to the same value and therefore would not change the choice of the best scored conformation.

Initially we used Chimera software[51] to estimate the number of ligands with internal hydrogen bonds in native conformation among 1300 ligands from the test set. With default parameters Chimera identified 87 ligands as having intraligand hydrogen bonds, which is less than 7% of the cases. Visual inspection showed that in many cases the intraligand hydrogen bonds were forced. Therefore, we developed a procedure to estimate the number of nonforced intraligand hydrogen bonds taking into account the number of covalent bonds separating donor and acceptor atoms.

To estimate the number of intraligand hydrogen bonds we calculate the number of the contacts between ligand atoms of types O−N or O−O. We call them hydrogen bond-like contacts. The contacts are characterized by two parameters: geometrical

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

distance between the atoms and the number of bonds in the ligand molecule between the atoms. We abbreviate the two parameters as dist and nb. Also, the additional requirement of the presence of at least one rotatable bond between the two atoms was used. The number of complexes where the ligand has at least one hydrogen bond-like contact with dist parameter less and nb parameter strictly more than specified values are listed in Table 4.

Table 4. Number of Ligands with at Least One Hydrogen Bond-like Contact with Specified Parameters (of 1300 Total)

nb> \ dist<	3.5Å	3.2Å	3.1Å	3Å	2.9Å
3	425	297	255	189	139
4	288	178	147	107	83
5	118	(71)	58	47	41
6	(66)	**(33)**	27	22	21
7	45	22	17	13	12

Visual analysis of ligands from groups with different parameters showed that the hydrogen bond-like contacts with *dist* < 3.2 Å and *nb* > 6 (33 ligands) usually form real hydrogen bonds. Increasing the dist to 3.5 Å results in 66 ligands, but many of the newly found contacts cannot form a hydrogen bond due to the severely violated angle restrictions of a hydrogen bond. Decreasing nb to *nb* > 5 adds 38 ligands (71 complexes in total against 33). However, only half of them have hydrogen bonds with allowed angles, and these hydrogen bonds are mostly forced by a large number of donors and acceptors in the ligand (substituted carbohydrates and nucleotides like cases). So the real number of the complexes with nonforced intraligand hydrogen bonds can be estimated as 50 complexes of 1300, which is less than 4%.

Finally, we can conclude that intraligand hydrogen bonds are very rare, taking place in less than 7% of the cases, while nonforced intraligand hydrogen bonds are even more rare taking place in less than 4% of the cases. An example of this rare instance is complex 1utc (Figure 2), where a peptide ligand is stabilized by three intraligand hydrogen bonds.

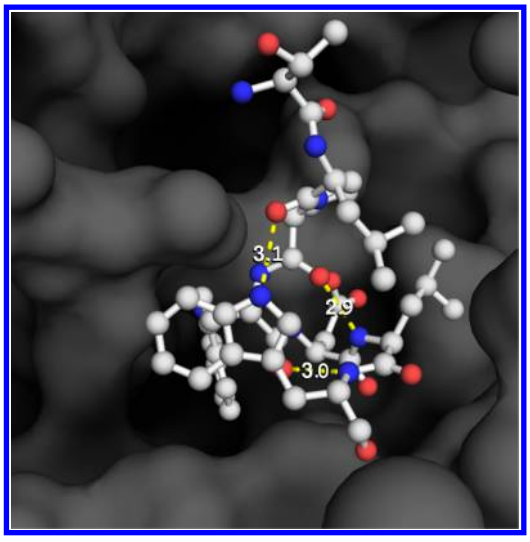


Figure 2. Complex 1utc with 3 intraligand hydrogen bonds.

3.1.2. Intraligand Hydrophobic Contacts in the Test Set Complexes. We consider two atoms to form a hydrophobic contact if they have the types C−C or C−S, the geometrical distance between the atoms is less than 5 Å, the distance between the atoms in the molecule is more than 4 covalent bonds, and at least one of the central bonds between the atoms is rotatable. The condition about the presence of a rotatable bond between the atoms is required to make the two atoms mobile relative to each other. Below, we call this type of contact the hphob contact.

As one can see in Chart 2, there are 453 of 1300 complexes in the group *all* where the ligands have no internal hphob contacts. That means 847 complexes out of 1300 have at least one internal hphob contact, so approximately 65% of complexes have intraligand hphob contacts. For the group *good* the percentage of complexes with intraligand hphob contacts is approximately 67%. The key values are summarized above in Table 3. As expected, the percentage of complexes with intraligand contacts is higher in groups *gflex* and *gfflex* of flexible ligands. So, in group *gfflex* containing 235 complexes with ligands having at least 10 rotatable bonds, almost all ligands have internal hydrophobic contacts.

The histograms in Chart 3 clearly show an increasing frequency of intraligand contacts, correlated with a greater number of rotatable bonds in the ligands. The percentage of complexes with intraligand contacts increases monotonically up to 100% with the number of rotatable bonds in the ligand. Even with 2 rotatable bonds, half of the ligands have internal contacts, while with 5 rotatable bonds, more than 80% of such ligands have internal contacts.

Unlike hydrogen bonds, hydrophobic contacts are formed by groups of atoms. To estimate the number of real hydrophobic contacts, while excluding occasional separate C−C atoms approach, we analyzed the quantities of hphob contacts.

Chart 4 presents the percentage of complexes with more than 0, 5, 10, and 15 intraligand hydrophobic contacts among complexes with different number of rotatable bonds. From the chart, one can see that more than half of ligands with more than 8 rotatable bonds have more than 5 hydrophobic contacts.

We see from the histograms in Chart 4 that ligands with more than 3 rotatable bonds have at least one internal contact in more than half of the cases, and more than 75% of ligands with more than 4 rotatable bonds have internal contacts in binding conformation. This shows that accounting for intraligand interactions is significant for flexible ligand docking, because if one does not take the interactions into account, a large energy term can be neglected.

3.2. The Effect of Intraligand Interactions Accounting on Docking and Scoring Accuracy. We use an internally developed docking program with an Algo_score scoring function to perform docking and scoring. The program was developed based on ideas from Ramensky et al.[2] The program performs flexible ligand docking with rigid protein. Only the best scored conformation is considered as a result.

To estimate the influence of intraligand interactions accounting on docking and scoring accuracy we conducted a series of docking computations on the test set of 1300 protein−ligand complexes. During the test docking computations two parameters influencing the intraligand interactions accounting were varied. These parameters are denoted as α and n in formula 4. The coefficient α has the meaning of the weight of intraligand interactions relative to ligand−protein interactions: with $\alpha = 0$ we got the initial scoring function without intraligand interactions account, while with $\alpha = 1$ the intraligand interactions

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

Chart 2. Intraligand Hydrophobic Contacts Distribution in Groups *all* and *good*

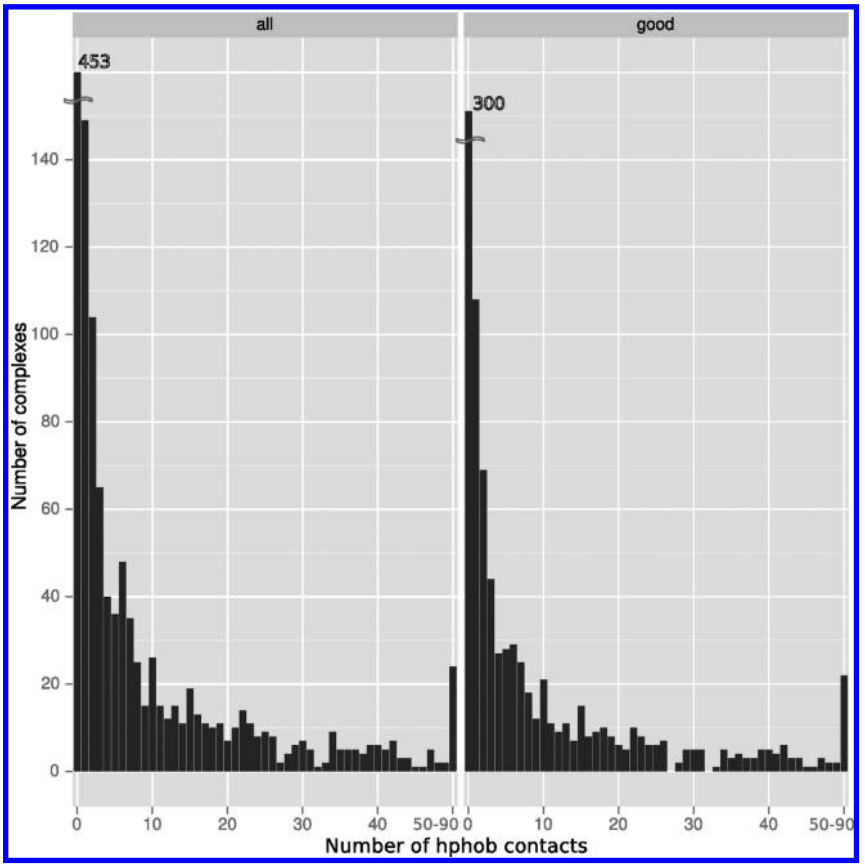
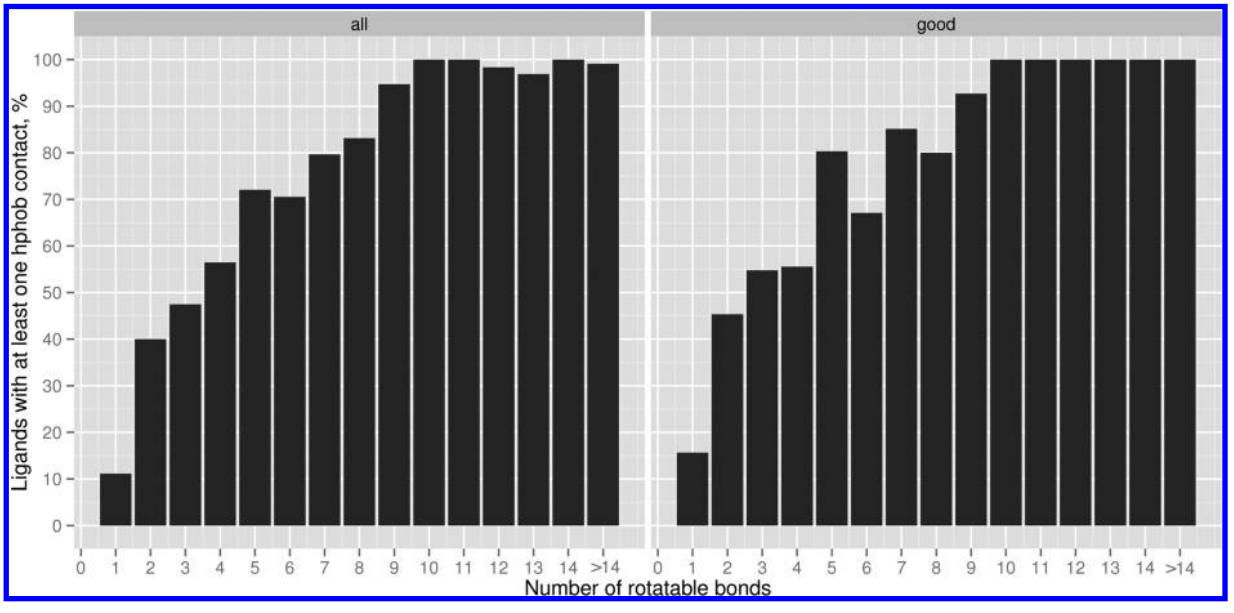


Chart 3. Percentage of Complexes with Intraligand hphob Contacts in Groups *all* and *good* among Ligands with a Given Number of Rotatable Bonds



are weighted equal to ligand−protein interactions. The parameter n defines the minimum number of bonds between ligand atoms required to account for the interaction between these two atoms in the scoring function. With a very large n we also got the initial scoring function without intraligand interactions accounted for.

For each test computation three criteria of success - `dock_succ`, `corr`, `corr_succ` - were considered, and four subsets of the entire set of 1300 complexes - *all* (1300 complexes), *good* (911 complexes), *gflex* (516 complexes), *gfflex* (235 complexes) - were considered. The test set is described in section 2.2.2 above, and the subsets and the criteria are defined there as well.

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

Chart 4. Percentage of Complexes with a Given Number of Intraligand hphob Contacts in Group *all* among Ligands with a Given Number of Rotatable Bonds

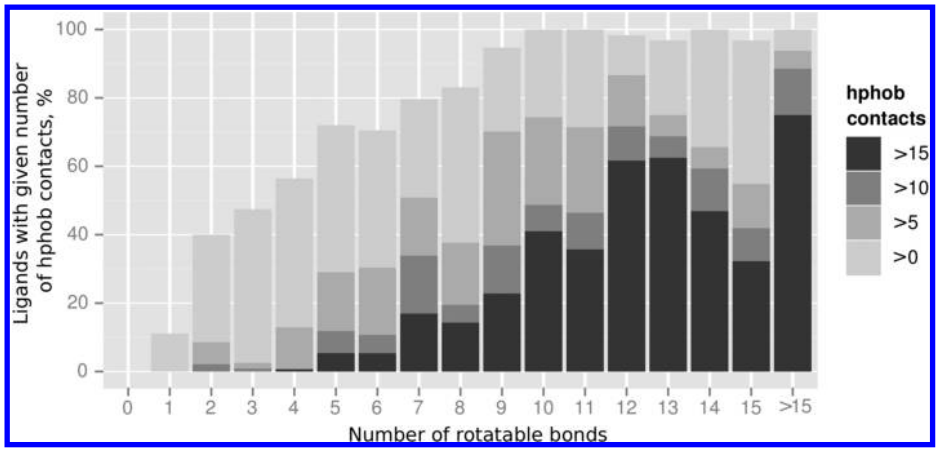
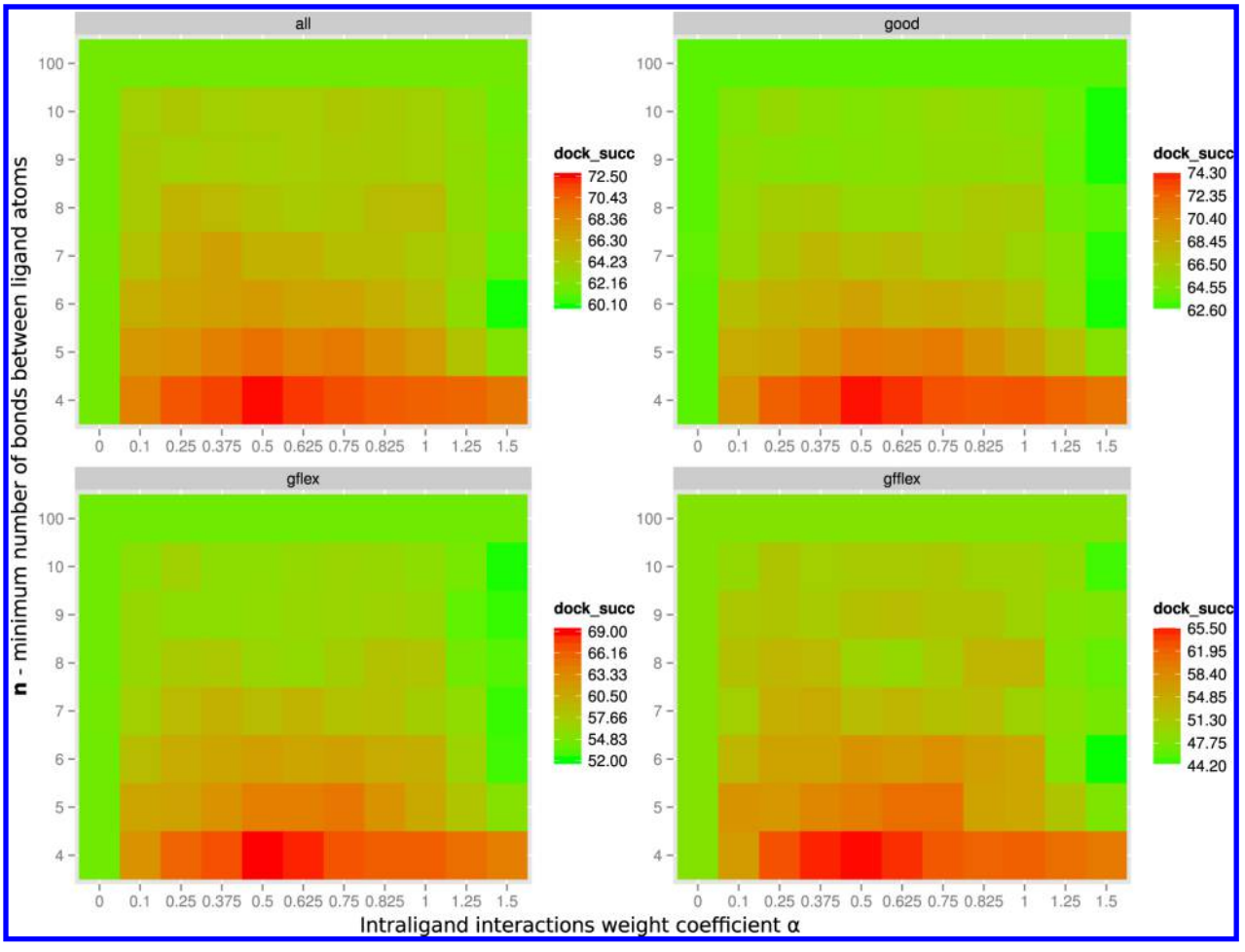


Chart 5. Docking Success by dock_succ Criterion Depending on the Weight of Intraligand Interactions α and Minimum Distance between Ligand Atoms Requiring n

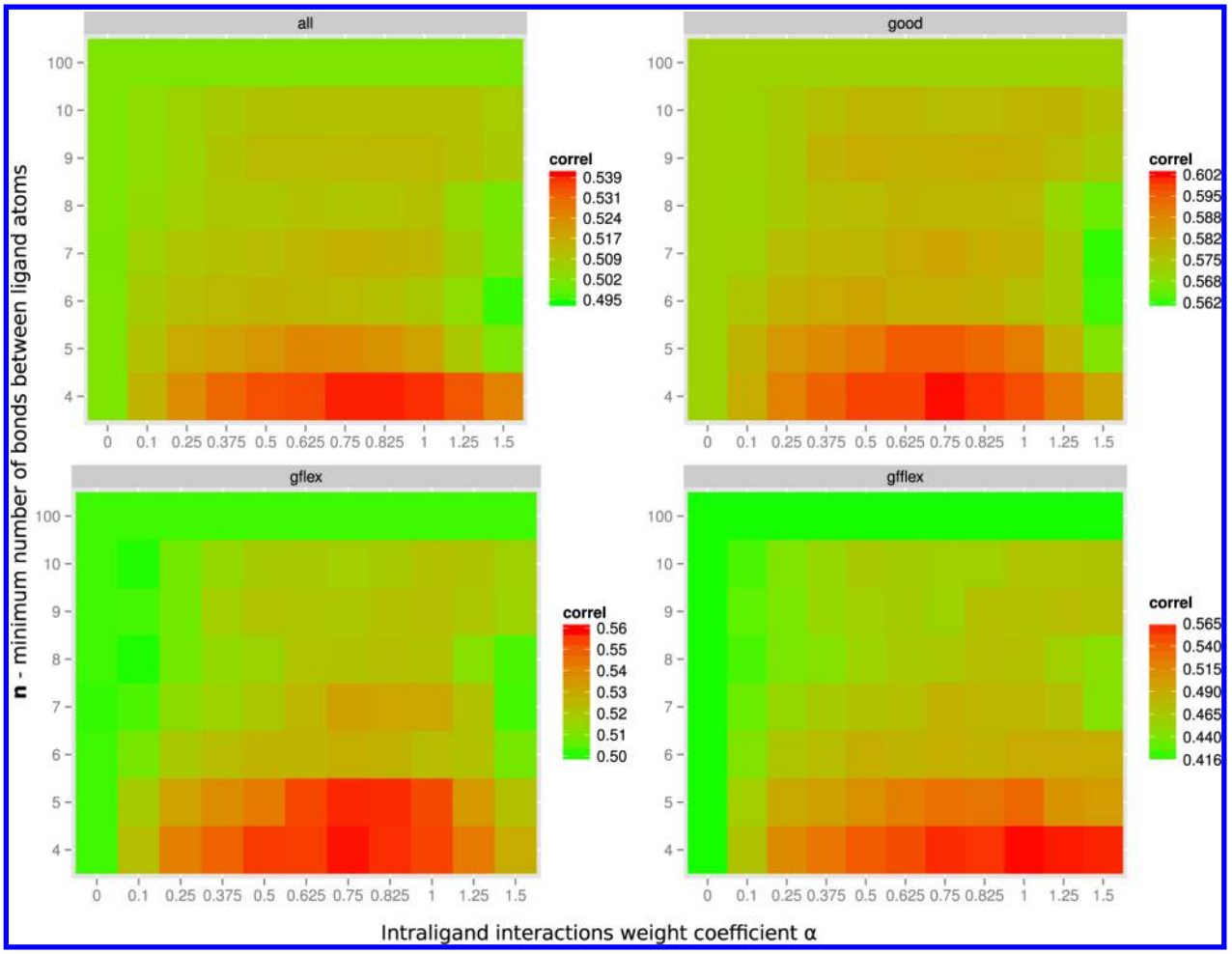


The docking computations were conducted with n = 3, 4, 5, ...,10, 100 and with a number of values for α in the range from 0 to 1.5. With a minimum value of n = 3 and with α > 0.1 the docking procedure failed for many complexes. This is due to large positive values of score, which is generated by repulsing atoms in the ligand (the positive value of score corresponds to positive energy of the conformation, i.e. to an energetically unfavorable conformation). The ligand conformations with a high score are rejected as defective by the docking program, and finally no ligand conformations are given as output of the docking procedure. For this reason we do not consider the results with n = 3.

The results of the test computations are presented in Charts 5, 6, and 7. As one can see, both the large value of n = 100 and the zero value α = 0 give the same results: this is the result without intraligand interactions accounting.

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

Chart 6. Correlation between Calculated Score and K_d Depending on the Weight of Intraligand Interactions α and the Minimum Distance between Ligand Atoms Requiring n



Precise values for the best results achieved by each criterion in each testing subset are summarized in Table 5. The values of n and α giving the result are specified within brackets next to the result value. In Table 6 the results achieved without intraligand interactions accounted for are shown (the case of $\alpha = 0$).

Accounting for intraligand interactions clearly improves the quality of docking by all criteria in all groups. As expected, the improvement is more significant for groups *gflex* and *gfflex* of flexible ligands. So, the accounting for the intraligand interactions increases the percentage of successfully docked complexes from 61.7 to 72.5 in group *all*, from 54.1 to 69.0 in group *gflex*, and from 48.1 to 65.5 in group *gfflex*. Absolute value of correlation between K_d and score of the best scored position raises from 0.501 to 0.539 in group *all*, from 0.501 to 0.560 in group *gflex*, and from 0.416 to 0.565 in group *gfflex*. Absolute value of correlation between K_d and score of the best scored position, counted on the subset of successfully docked complexes (criterion *corr_succ*) raises from 0.490 to 0.544 in group *all*, from 0.428 to 0.551 in group *gflex*, and from 0.325 to 0.585 in group *gfflex*. The specified best values are achieved with slightly different values of α and n parameters (Table 6), but as is seen from the plots in Chart 8 below, any choice of the parameters pair with n equals 4 or 5 and α lying in the range from 0.5 to 0.75 gives similar results, decisively surpassing the results without intraligand interactions accounting.

In section 3.1.1 it was shown that the number of fake hydrogen bonds strongly depends on parameter *nb* - the minimum number of bonds in the molecule between the atoms, required to consider the contact between the two atoms as not being forced by molecule structure. With $nb \leq 6$ there are a large number of fake hydrogen bonds, i.e. cases of a donor and an acceptor of the hydrogen bond approaching the distance of hydrogen bond forced by molecule structure but with wrong angles between bonds for a real hydrogen bond formation. This suggests that accounting for intraligand interactions with $n \leq 6$ in eq 4 may give wrong results and the optimal value should be $n = 7$. But the results presented in Charts 5, 6, and 7 show that the best results are achieved with $n = 4$, and they get significantly worse with $n = 7$. This may be due to the low number of intraligand hydrogen bonds (as mentioned in section 3.1.1) and the differing nature of hydrophobic interactions, giving the main contribution to intraligand score component change.

As is seen from Charts 5, 6, and 7, the best results by docking criterion `dock_succ` are achieved with $\alpha = 0.5$ and $n = 4$ in all groups. The best results by `corr` criterion are achieved with $\alpha = 0.75$ and $n = 4$ in groups *all*, *good*, and *gflex* and with $\alpha = 1.0$ and $n = 4$ in group *gfflex*. The results get worse more or less monotonically when changing α and n from the optimal values for this two criteria. The optimal values for α and n by the `corr_succ` criterion are more volatile.

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

Chart 7. Correlation between Score and K_d Calculated on the Subset of Successfully Docked Complexes Depending on the Weight of Intraligand Interactions α and the Minimum Distance between Ligand Atoms Requiring n

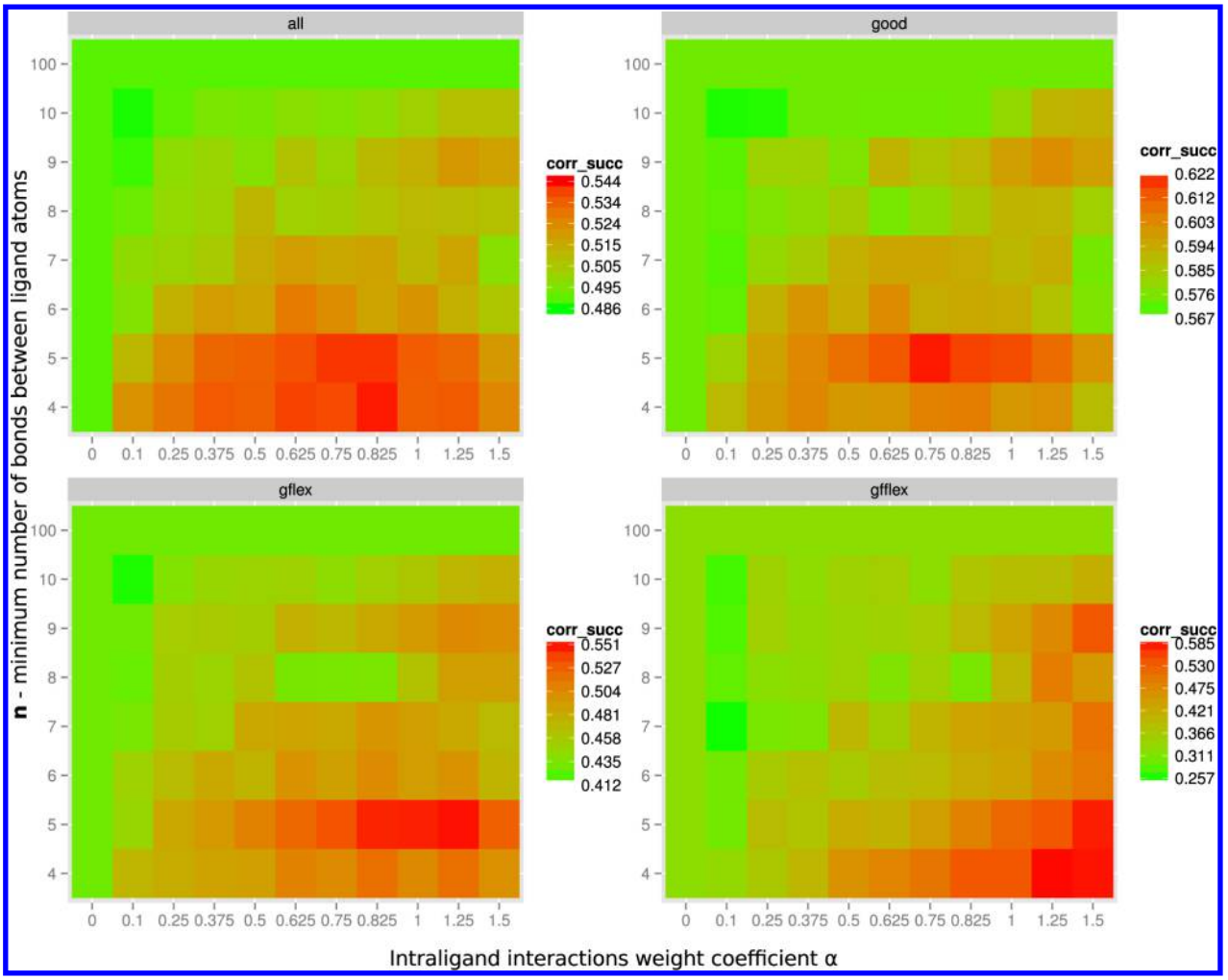


Table 5. Best Results in Test with Intraligand Interactions Accounted for[a]

	all	good	gflex	gfflex
succ	72.5 (4;0.5)	74.3 (4;0.5)	69.0 (4;0.5)	65.5 (4;0.5)
corr	0.539 (4;0.75)	0.602 (4;0.75)	0.560 (4;0.75)	0.565 (4;1.0)
corr_succ	0.544 (4;0.825)	0.622 (5;0.75)	0.551 (5;1.25)	0.585 (4;1.25)

[a]Corresponding values for n and α are in brackets.

Table 6. Testing Results without Intraligand Interactions Accounting ($\alpha = 0$)

	all	good	gflex	gfflex
succ	61.7	63.6	54.1	48.1
corr	0.501	0.572	0.501	0.416
corr_succ	0.490	0.573	0.428	0.325

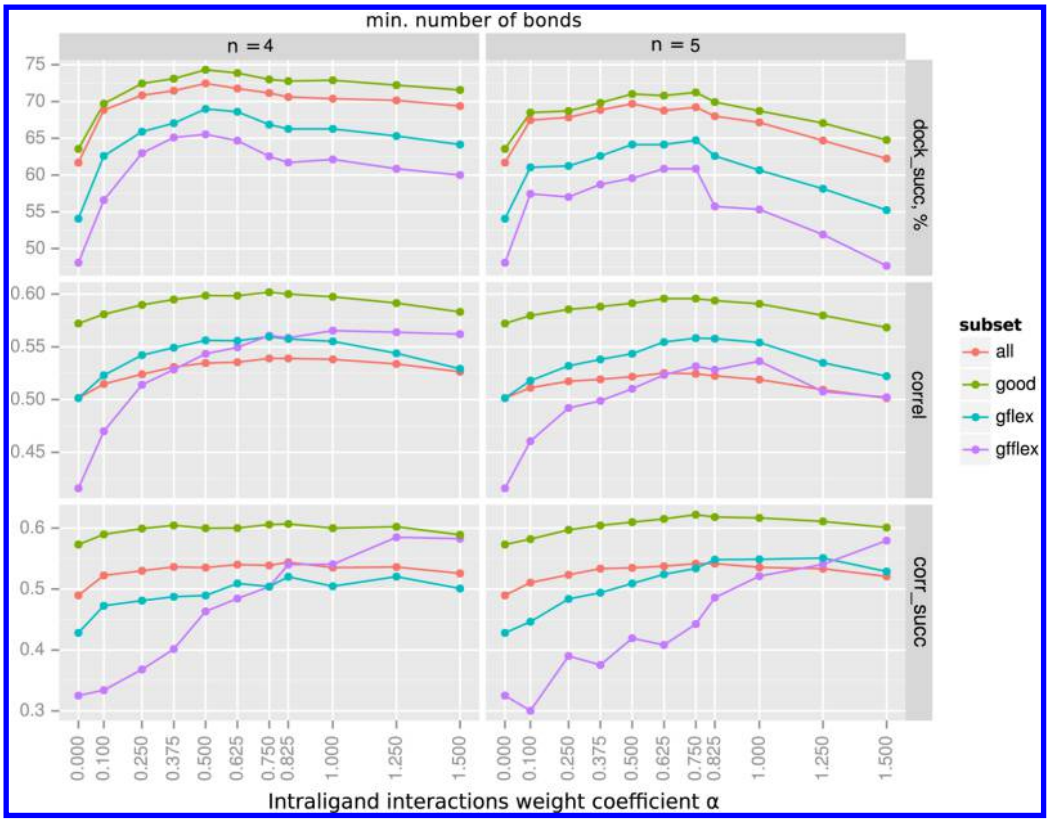
The best results are achieved with $n = 4$ and $n = 5$. The relationship of the results to the weight of intraligand interactions α in these two cases is plotted in Chart 8.

One can expect that the best docking results must be achieved with α in eq 4 equal to 1, which means that the intraligand interactions are weighted equal to ligand−protein interactions, but the plots in Chart 8 show that the best results are achieved with α less than 1. We attribute this effect mainly to the presence of the cloud score multiplier C_i in eq 4 of the Algo_score. The cloud score takes the values from 0 to 1 and effectively results in the multiplication of protein−ligand interaction by a factor in the range from 0.5 to 1, that is why the value of 1 for coefficient α in eq 4 makes the resulting score overweight the intraligand interactions. For PMF-based scoring functions of standard form eq 1, without a cloud score multiplier, the optimal value of coefficient α for intraligand interactions would probably be close to 1. Another factor that can bias the optimal value for coefficient α away from 1 is the way a potential of mean force U_{ij} in eq 1 is derived. A PMF may be optimized for protein−ligand interatomic interactions and therefore estimate the intraligand interactions with an error.

The plots in Chart 8 show that there is a general improvement of docking and scoring quality associated with accounting for intraligand interactions perturbed by random effects of the test set used. The general form of the plots with a maximum reached at $\alpha \in [0.5; 0.75]$ seems to be a nonrandom effect, independent

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

Chart 8. Docking Results Depending on the Weight of Intraligand Interactions α with a Minimum Number of Bonds between the Ligand Atoms Requiring $n = 4$ and $n = 5$



from a choice of a sufficiently large random set of test complexes. The exceptions are the results by correlation criteria for the group *gfflex* of very flexible ligands; in this situation the optimal value for α is shifted toward higher values. We suppose it to be a result of an indirect accounting of entropy loss for flexible ligands upon binding, which is not accounted for in Algo_score.

To validate the hypothesis that the optimal values for α lie in the range from 0.5 to 0.75 independently from a large random test set used, we performed an additional check. The 911 complexes of the *good* subset were split randomly into two groups, and for each group the plots similar to those in Chart 8 were built (see the Supporting Information). The best results in each group were still achieved with $\alpha \in [0.5; 0.75]$.

Some illustrations of docking result improvements when accounting for intraligand interactions are in Figure 3 below. The native position of a ligand is colored gray, the position found without accounting for intraligand interactions is colored red, and the position found with accounting for intraligand interactions is colored green. The dashed lines show the intraligand hydrophobic contacts.

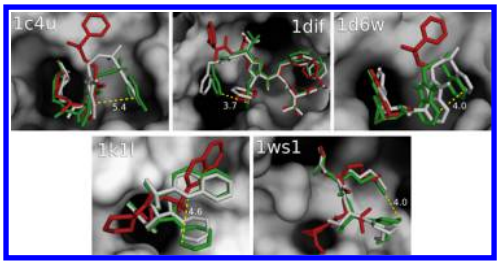


Figure 3. Illustrations of docking results improvement upon intraligand interactions account. Gray - native position of a ligand, red - position found without intraligand interactions account, green - position found with intraligand interactions accounted.

3.2.1. Comparison of the Docking Results with Other Programs. Along with evaluating the effect of accounting for intraligand interactions on the docking procedure, we compared the docking and scoring results with the results reported in the work of Plewczynski et al.[1]

Straightforward comparison of the results reported by Plewczynski and the results we got faces two problems: first, in the original work[1] there is no explanation how the problematic complexes of the test set (i.e., not belonging to the group *good* in Table 2) were handled. We believe that no matter how these *bad* complexes were handled by Plewczynski et al., our testing results have no unfair advantage over their results, since the inclusion of the *bad* complexes into the test set significantly worsens our test results (group *all* as compared to group *good* in Chart 8).

Another problem arises when we compare our results with optimized values of parameters α and n with the results reported by Plewczynski.[1] In this case the set of 1300 complexes is used formally both as a training and a testing set, but we do not strive to find the best values for the α and n at this point. The aim is to qualitatively compare our docking results with intraligand interactions accounted for with the results obtained by other scoring functions. As discussed in the above section 3.2, the belonging of the optimal α and n values to the intervals $\alpha \in [0.5; 0.75]$ and $n \in \{4;5\}$ is a nonrandom effect perturbed with a small random influence of the test set choice. To negate the possible unfair advantage from the random effects we compare the results reported by Plewczynski with the worst testing results obtained by us with $\alpha \in [0.5; 0.75]$ and $n \in \{4;5\}$. Below in the paragraph

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

"the results with intraligand interactions accounted for" means the worst results obtained with $\alpha \in [0.5; 0.75]$ and $n \in \{4;5\}$.

We note that even in the case of $\alpha = 0$ in eq 4 the resulting scoring function performs very well comparing to other scoring functions reported in ref 1. This result conforms with the original result of the Algo_score reported in ref 2, where the Algo_score based program was shown to outperform all other docking programs in Rognan's self-docking test of 100 complexes.[49]

In our study, the case of $\alpha = 0$ corresponds to the ignoring of intraligand interactions and gives the original Algo_score scoring function eq 3 as the result. The percentage of successfully docked complexes with $\alpha = 0$ is 61.7% (from 1300), which surpasses all results reported in ref 1, except the combination of GOLD + Omega_ten which gives about 62.5% of successfully docked best scored conformations in this test set. With intraligand interactions accounted for the number of successfully docked complexes comes up to 68.5%, which outperforms all the results reported in ref 1.

The Algo_score scoring function showed a very good result by correlation criteria also. The best Spearman correlation for the whole test set reported in ref 1 is 0.47 and is achieved by the eHiTS program, while the original Algo_score (case of $\alpha = 0$) gives the correlation 0.5 in modulus. With accounting for intraligand interactions, the correlation rises even higher to 0.525 in modulus.

As expected, the results are much better on the subset *good*, consisting of complexes fully applicable for docking. The docking success rate for this subset is 63.7% without intraligand interactions accounting and 71% with intraligand interactions accounted for. The best Spearman correlation for the subset is 0.575 without intraligand interactions accounting and with intraligand interactions accounted for the result rises to 0.59.

We found out that the set of 195 complexes used by Cheng et al.[42] for testing 16 scoring functions was actually a subset of 1300 complexes we use. That allowed us to compare our docking results with results reported in Cheng's article. We got the following results on Cheng's set: dock_succ is 62% without intraligand interactions accounting and 68% with intraligand interactions accounted for. This result corresponds to a middle result of the 16 scoring functions tested in the article. The Spearman correlation between K_d and Algo_score is 0.52 without intraligand interactions accounting and 0.585 with intraligand interactions accounted for, which outperforms all 16 scoring functions tested by Cheng et al., except for two scoring functions X-Score::HMScore and DrugScoreCSD, which give the correlations 0.705 and 0.627, respectively. The charts with the results obtained on Cheng's set are in the Supporting Information.

4. CONCLUSIONS

The analysis of 1300 protein−ligand complexes of Plewczynski's test set[1] shows that ligand atoms in PDB complexes rather often approach each other to a distance small enough for hydrophobic and hydrogen bond interactions. This observation provides support for the hypotheses that intraligand interactions might play a significant role in the protein−ligand binding process, and they should not be neglected in scoring functions. Detailed analysis showed that intraligand hydrogen bonds are rare, but intraligand hydrophobic contacts are frequent, especially in flexible ligands.

We propose a general method for a scoring function modification to account for intraligand interactions. The key idea is to account for interactions between different ligand atoms by the same potential of mean force, which is used in a PMF-based scoring function to estimate the interactions between protein and ligand atoms. The suggested method was implemented with the Algo_score scoring function.[2] The effectiveness of the scoring function modification was proved on a test set of 1300 protein−ligand complexes.[1]

The proposed scoring function modification would probably improve the quality of docking and scoring of flexible ligands for any PMF-based scoring function. The modification can also be effective with any scoring function (not necessary PMF-based), which does not account for ligand−ligand interactions and ligand−solvent interactions.

We analyzed in detail the relationship between docking and scoring success and the couple of parameters affecting the intraligand score component: the overall weight of intraligand interactions (coefficient α in eq 4) and the minimum number of bonds between ligand atoms required to account for their interaction (n in eq 4). The analysis showed that for the Algo_score scoring function, the optimal value for α lies in the range 0.5−0.75, and for n the optimal value is 4 or 5, depending on the considered criterion. We suppose that for other scoring functions the optimal value for n would be the same, but the optimal value for α would be closer to 1 due to specific properties of the Algo_score.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00158.

Ligand extraction algorithm, PMF-based scoring functions overview, docking results for two random groups, docking results for Cheng's subset, list of extracted ligands along with their PDB IDs and other properties, list of extracted ligands PDB IDs grouped by flexibility and quality, list of PDB IDs in two random groups (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: al_bio@mail.ru.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The study was supported by a grant from the Russian Science Foundation (14-14-00598).

■ REFERENCES

(1) Plewczynski, D.; Łaźniewski, M.; Augustyniak, R.; Ginalski, K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* **2011**, *32*, 742−755.

(2) Ramensky, V.; Sobol, A.; Zaitseva, N.; Rubinov, A.; Zosimov, V. A novel approach to local similarity of protein binding sites substantially improves computational drug design results. *Proteins: Struct., Funct., Genet.* **2007**, *69*, 349−357.

(3) Muryshev, A. E.; Tarasov, D.; Butygin, A.; Butygina, O. Y.; Aleksandrov, A.; Nikitin, S. A novel scoring function for molecular docking. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 597−605.

(4) Huang, S.-Y.; Zou, X. Advances and challenges in protein-ligand docking. *Int. J. Mol. Sci.* **2010**, *11*, 3016−3034.

(5) Li, X.; Li, Y.; Cheng, T.; Liu, Z.; Wang, R. Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. *J. Comput. Chem.* **2010**, *31*, 2109−2125.

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

(6) Li, Y.; Liu, Z.; Li, J.; Han, L.; Liu, J.; Zhao, Z.; Wang, R. Comparative assessment of scoring functions on an updated benchmark: 1. Compilation of the test set. *J. Chem. Inf. Model.* **2014**, *54*, 1700−1716.
(7) Li, Y.; Han, L.; Liu, Z.; Wang, R. Comparative assessment of scoring functions on an updated benchmark: 2. Evaluation methods and general results. *J. Chem. Inf. Model.* **2014**, *54*, 1717−1736.
(8) Huang, S.-Y.; Grinter, S. Z.; Zou, X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12899−12908.
(9) Liu, J.; Wang, R. Classification of Current Scoring Functions. *J. Chem. Inf. Model.* **2015**, *55*, 475−482.
(10) Zheng, Z.; Merz, K. M., Jr Development of the knowledge-based and empirical combined scoring algorithm (kecsa) to score protein-ligand interactions. *J. Chem. Inf. Model.* **2013**, *53*, 1073−1083.
(11) Lorber, D. M.; Shoichet, B. K. Flexible ligand docking using conformational ensembles. *Protein Sci.* **1998**, *7*, 938−950.
(12) Perola, E.; Charifson, P. S. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J. Med. Chem.* **2004**, *47*, 2499−2510.
(13) Tirado-Rives, J.; Jorgensen, W. L. Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding. *J. Med. Chem.* **2006**, *49*, 5880−5884.
(14) Boström, J.; Norrby, P.-O.; Liljefors, T. Conformational energy penalties of protein-bound ligands. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 383−383.
(15) Ren, J.; Diprose, J.; Warren, J.; Esnouf, R. M.; Bird, L. E.; Ikemizu, S.; Slater, M.; Milton, J.; Balzarini, J.; Stuart, D. I. Phenylethylthiazolylthiourea (PETT) Non-nucleoside Inhibitors of HIV-1 and HIV-2 Reverse Transcriptases STRUCTURAL AND BIOCHEMICAL ANALYSES. *J. Biol. Chem.* **2000**, *275*, 5633−5639.
(16) Scheerer, P.; Park, J. H.; Hildebrand, P. W.; Kim, Y. J.; Krauß, N.; Choe, H.-W.; Hofmann, K. P.; Ernst, O. P. Crystal structure of opsin in its G-protein-interacting conformation. *Nature (London, U. K.)* **2008**, *455*, 497−502.
(17) Petros, A. M.; Nettesheim, D. G.; Wang, Y.; Olejniczak, E. T.; Meadows, R. P.; Mack, J.; Swift, K.; Matayoshi, E. D.; Zhang, H.; Fesik, S. W. Rationale for Bcl-XL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Sci.* **2000**, *9*, 2528−2534.
(18) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank. *Eur. J. Biochem.* **1977**, *80*, 319−324.
(19) Sippl, M. J. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 473−501.
(20) Sippl, M. J. Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* **1995**, *5*, 229−235.
(21) Finkelstein, A. V.; Badretdinov, A. Y.; Gutin, A. M. Why do protein architectures have boltzmann-like statistics? *Proteins: Struct., Funct., Genet.* **1995**, *23*, 142−150.
(22) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337−356.
(23) Kirtay, C. K.; Mitchell, J. B.; Lumley, J. A. Knowledge based potentials: The reverse Boltzmann methodology, virtual screening and molecular weight dependence. *QSAR Comb. Sci.* **2005**, *24*, 527−536.
(24) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791−804.
(25) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.
(26) Ozrin, V. D.; Subbotin, M.; Nikitin, S. PLASS: Protein-ligand affinity statistical score-a knowledge-based force-field model of interaction derived from the PDB. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 261−270.
(27) 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.
(28) Huang, S.-Y.; Zou, X. An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *J. Comput. Chem.* **2006**, *27*, 1866−1875.
(29) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. R.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol. (Oxford, U. K.)* **1995**, *2*, 317−324.
(30) Mooij, W.; Verdonk, M. L. General and targeted statistical potentials for protein-ligand interactions. *Proteins: Struct., Funct., Genet.* **2005**, *61*, 272−287.
(31) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(32) Mitchell, J. B.; Laskowski, R. A.; Alex, A.; Thornton, J. M. BLEEP-potential of mean force describing protein-ligand interactions: I. Generating potential. *J. Comput. Chem.* **1999**, *20*, 1165−1176.
(33) Huang, S.-Y.; Zou, X. Inclusion of solvation and entropy in the knowledge-based scoring function for protein- ligand interactions. *J. Chem. Inf. Model.* **2010**, *50*, 262−273.
(34) Grinter, S. Z.; Yan, C.; Huang, S.-Y.; Jiang, L.; Zou, X. Automated Large-Scale File Preparation, Docking, and Scoring: Evaluation of ITScore and STScore Using the 2012 Community Structure-Activity Resource Benchmark. *J. Chem. Inf. Model.* **2013**, *53*, 1905−1914.
(35) Weiner, S. J.; Singh, U. C.; O'donnell, T.; Kollman, P. A. Quantum and molecular mechanical studies on alanyl dipeptide. *J. Am. Chem. Soc.* **1984**, *106*, 6243−6245.
(36) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(37) Pang, Y.-P.; Perola, E.; Xu, K.; Prendergast, F. G. EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J. Comput. Chem.* **2001**, *22*, 1750−1771.
(38) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(39) Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 235−249.
(40) Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. On evaluating molecular-docking methods for pose prediction and enrichment factors. *J. Chem. Inf. Model.* **2006**, *46*, 401−415.
(41) Mukherjee, S.; Balius, T. E.; Rizzo, R. C. Docking validation resources: protein family and ligand flexibility experiments. *J. Chem. Inf. Model.* **2010**, *50*, 1986−2000.
(42) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **2009**, *49*, 1079−1093.
(43) Böhm, H.-J. Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309−309.
(44) Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11−26.
(45) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.
(46) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
(47) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. OBBond Class Reference. 2012. http://openbabel.org/dev-api/classOpenBabel_1_1OBBond.shtml (accessed Oct 1, 2015).

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137

(48) The PyMOL Molecular Graphics System, Version 1.7.0.0; Schrödinger, LLC.: August 2013.
(49) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Struct., Funct., Genet.* **2004**, *57*, 225−242.
(50) Paul, N.; Rognan, D. ConsDock: A new program for the consensus analysis of protein-ligand interactions. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 521−533.
(51) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimeraa visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(52) Desiraju, G. R. Hydrogen bridges in crystal engineering: interactions without borders. *Acc. Chem. Res.* **2002**, *35*, 565−573.
(53) Mills, J.; Dean, P. M. Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 607−622.

NOTE ADDED AFTER ASAP PUBLICATION

This article was published ASAP on October 8, 2015, with errors in the Abstract. The corrected version was published ASAP on October 9, 2015.

DOI: 10.1021/acs.jcim.5b00158
J. Chem. Inf. Model. 2015, 55, 2121−2137