A Hybrid Knowledge-Based and Empirical Scoring Function for Protein–Ligand Interaction: SMoG2016

Théau Debroise, Eugene I Shakhnovich, Nicolas Chéron

A Hybrid Knowledge-Based and Empirical Scoring Function for Protein-Ligand Interaction: SMoG2016

*[1]Théau Debroise, [1]Eugene I. Shakhnovich and [1,2]Nicolas Chéron**

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge MA 02138, USA; [2]Ecole Normale Supérieure, PSL Research University, UPMC Univ. Paris 06, CNRS, Département de Chimie, UMR 8640 PASTEUR, 24 rue Lhomond, 75005 Paris, France.

**Correspondence: nicolas.cheron@ens.fr*

Abstract: We present the third generation of our scoring function for the prediction of protein-ligand binding free energy. This function is now a hybrid between a knowledge-based potential and an empirical function. We constructed a diversified set of ~1000 complexes from the PDBBinding-CN database for the training of the function and we show that this number of complexes generate enough data to build the potential. The occurrence of 420 different types of atomic pair wise interactions is computed in up to five different ranges of distances to derive the knowledge-based part. All parameters were optimized and we were able to considerably improve the accuracy of the scoring function with a Pearson correlation coefficient against experimental binding free energies of up to 0.57, which ranks our new scoring function as one of the best currently available and the second-best in term of standard deviation (SD=1.68). The function is then further improved by inclusion of different terms taking into account repulsion and loss of entropy upon binging, and we show it is capable of recovering native binding pose up to 80% of times. All programs, tools and protein sets are released in Supporting Information or as open-source programs.

Introduction

Predicting the free energy of binding between a protein and a ligand is a crucial part of computational structure-based drug design. In the context of the development of a new *de novo* drug design tool called OpenGrowth[1] which uses a combinatorial approach to fragment-based design, there was a need to construct a new accurate function that would be also fast and simple. Indeed, during the development of a scoring function, one usually wants to increase the balance between accuracy and speed, and compromises have to be found. For example, for a use in high-throughput virtual screening, speed is the prime objective and accuracy is often lost during the development. As a consequence, results obtained after docking large libraries may be inaccurate and always need to be refined. Moreover, the actual mathematical form of a scoring function is also of a crucial importance: if a given function has a complicated form and is hard to implement in various programs, it is likely that it will not be used. This point is often neglected in the literature, but we consider that a new function must also use a form allowing an easy implementation or be released as an open-source.

Scoring functions in the literature can be roughly divided into three categories: physical-based, empirical, and knowledge-based[2] (even though this classification is not rigid and one may add a descriptor-based category[3]). Each of these scoring methods try in some way to approximate the binding free energy whose explicit form is unknown. However, they usually contain parameters accounting for the complex association in vacuum and for the solvation/desolvation process. Physical-based scoring functions account for van der Waals and electrostatic interactions, and other terms can be added to take into account hydrogen bonds or solvation effects (GoldScore is such a function[4]). They often attempt to integrate high level of theory in the predictions (such as quantum mechanics) and they can also take advantage of the progress of modern force fields (such

as CHARMM, Amber, OPLS) or solvation models. However, a limited success has been achieved so far by using this kind of approach[5]. The main reason of this failure is the lack of description of the entropic part of the binding process. Another reason is that it may be difficult to obtain reliable parameters when there are not enough experimental data available.

The empirical or regression-based scoring functions compute the free energy of binding by summing terms accounting for different energy factors such as hydrogen bond, metal contact or buried surface area for example. Each energetical contribution is fixed and is coming from a multiple linear regression analysis done over a set of complexes with known binding affinity, and is then multiplied by a term taking into account geometrical information in the protein-ligand complex. Therefore, both structural and energetical information about the complexes are needed to develop the function (ChemScore is an example of this kind of function[6,7]).

Finally, knowledge-based scoring functions take advantage of structural information in a database of protein-ligand crystallographic structures. Using the pairwise approximation, the total score *S* can be expressed as a sum over all contacts whose energies $F(\sigma_p, \sigma_l)$ are usually derived from the inverse Boltzmann-like analysis:

$$S = \sum_p \sum_l \Delta F(\sigma_p, \sigma_l) = \sum_p \sum_l -RTln \frac{p(\sigma_p,\sigma_l)}{p_{ref}} \qquad (1)$$

where *p* are the protein atoms of type σ_p, *l* are the ligand atoms of type σ_l, $p(\sigma_p, \sigma_l)$ is the frequency of the contact between the atoms of type σ_p and σ_l in the training database, and p_{ref} is the frequency of this same contact in a reference state. One of the first knowledge-based protein-ligand potential (SMoG96) was developed in 1996 by De Witte and Shakhnovich[8] and was then improved in 2001 (SMoG2001[9]) by taking into account different shells of contact and a better description of the reference state. The detailed theory behind these functions is well described in the SMoG96 and SMoG2001 papers[8,9] and we point the interested reader to these articles and to the Supporting

Information where it is summed up. In the current work, we want to present the development of a new function that uses the same analytical form for the knowledge-based potential than SMoG2001. Indeed, when compared with more modern scoring functions[10], it appeared that SMoG2001 is not accurate enough, mainly due to the fact that it misses long-range interactions. To derive this new function (SMoG2016), all the different parameters have been updated and new contributions have been added.

Methods and results

Construction of the Databases

Li *et al.* have analyzed the Brookhaven Protein Data Bank and have created subsets of it. One of them, called the "core set", is made of 195 complexes with accurate structural and energetic information[11]. This set contains complexes which are diverse enough (they span 10 orders of magnitude of binding and form 65 clusters of protein sequences). Since the accuracy of 21 currently available scoring functions has been tested against this core set[10], it is natural to use this set as a testing set to be able to provide an unbiased comparison of our new scoring function with existing ones. Thus, this testing set was solely used at the end of the development to compare the final form of the new scoring function with other functions. The full set can be downloaded online (http://www.pdbbind.org.cn/) and the list of complexes contained in this set is provided in Supporting Information. Another protein-ligand set prepared by the same authors is called the "general dataset" and consists of 10,662 complexes (version 2014)[11]. To refine the training set, we sorted it in the following way: (1) all the complexes with a resolution higher than 3 Å were removed; (2) all the complexes with ligands with a molar mass greater than 1000 g/mol were removed; (3) only the structures with available experimental binding free energy (ΔG) were kept; (4) all complexes from the testing set were removed. The first two criteria pruned the set to 9,553

complexes; after applying the last two criteria, we ended up with 5,948 complexes. Then, two kinds of sorting were applied to ensure that the remaining complexes are diversified enough: (1) the Tanimoto scores between all pairs of ligands ($1.8*10^7$ scores) were calculated using OpenBabel[12] with the default FP2 fingerprint (the Tanimoto score measures the similarity between two molecules, a score of 1 meaning very similar molecules and a score of 0 meaning very different molecules[13]); (2) sequence similarity analysis between all pairs of proteins were performed with the "Pairwise Sequence Alignment" tool of ClustalW2[14]. We then removed complexes for which either the protein or the ligand were too similar to other complexes in the current set. Finally, we removed all complexes with covalent ligands (defined as complexes for which the shortest distance between all heavy atoms of the protein and all heavy atoms of the ligand is smaller than 1.75 Å) as well as complexes for which intra-ligand distances were too small and complexes containing boron, nickel or manganese because we couldn't find reliable Lennard-Jones parameters for these elements. We ended up with a set of 1,038 diverse complexes which form our training set, in which all pairs of ligands have a Tanimoto score lower than 0.95 (chosen arbitrary) and all pairs of proteins has a sequence similarity inferior to 90% (same value as the one used by Li *et al.* to cluster the proteins while preparing PDBbind-CN[11]). The list of the complexes forming this set is provided in SI. For the development of the knowledge-based potential, we randomly split the set of 1,038 complexes in one set of 938 complexes and one of 100: the first set served to train the function and the second one to test it in order to select the optimal parameters. The Pearson correlation coefficient between scores and experimental binding free energies ($\Delta G=RT*\ln(K_d)$, with R the gas constant and T=293.15K) was used to evaluate the accuracy of the function. To ensure that the choice of parameters does not depend on the training set, we performed the random splitting 20 times and we report the average correlation coefficient over the 20 testing sets. Errors are estimated

with the standard deviation (SD), calculated as (where a and b are the intercept and slope of the linear regression, x_i are the experimental data and y_i the computed scores)[10]:

$$SD = \sqrt{\frac{[y_i - (a + b * x_i)]^2}{N - 1}}$$

Atom types

In SMoG2001, a set of 13 atom types was used for the ligands; one major drawback was the absence of any type for the halides. Moreover, sulphur and phosphorus had been gathered in a unique type. We thus decided to use a more precise set of atom types by upgrading the types from SMoG2001[9]. Different variations of the atom types were tested by first preparing the potential on the training set then calculating the correlation coefficient in the testing set. The shell configurations of SMoG2001 (see below) were used during this stage. An increase in the correlation coefficient in the testing set was considered a sign of a favorable modification. A compromise had to be found between using excessive number of atom types for a better description but with low statistics or fewer atom types with better statistics. For example, we tried to use four different atom types for each halide, but it appeared that there are not enough halides in the database to obtain good statistics. Consequently, only one atom type accounting for all the halides is implemented in SMoG2016. Since the tools to derive the scoring function are released as open-source programs, a user could easily decide to use different atom types for all the halides. We ended up with a set of 14 atom types for the ligand, listed in Table 1. Note that we didn't include types for the following atoms (among others): Si, Al, Pt, As, Ru, V, Se, Cu, Fe, Hg (since they are rarely found in ligands). Depending on the code used to assign the atom types, different results may be obtained (for example, Li *et al.* have obtained different scores for the same complexes with the ChemScore function depending on the software used[10]). To avoid discrepancies, we release the

code used to assign the types together with this article (it uses the OpenBabel library code and a combination of internal rules[12]).

Table 1. Ligand atom types.

Type	Description
1	Non-polar sp^3 carbon
2	Non-polar sp^2 or sp carbon
3	Carbonyl carbon, thioketone carbon, guanidine sp^2 carbon
4	Other polar carbon (connected to atoms different than C or H)
5	Hydrogen bond donor nitrogen (e.g. secondary amine or pyrrol)
6	Hydrogen bond acceptor nitrogen (e.g. pyridine)
7	Amide nitrogen
8	Carbonyl oxygen (aldehyde, ketone, amide, ester)
9	Hydrogen bond donor oxygen (hydroxyl)
10	Hydrogen bond acceptor oxygen (ether, sp^3 in ester)
11	Charged oxygen (carboxylate, phosphate, nitro)
12	Phosphorus
13	Sulfur
14	Halides
0	H, Si, Al, Pt, As, Ru, V, Se, Cu, Fe, Hg

13 atom types were used for the protein in SMoG2001. To be more accurate, we have decided to improve the set designed by Chen *et al.* during the construction of a potential for protein folding[15]. This set originally contained 23 types "chosen to reflect either physico-chemical similarity or positional equivalence"[15], which we increased to 30 atom types for the protein (only heavy atoms are considered, see Figure 1).

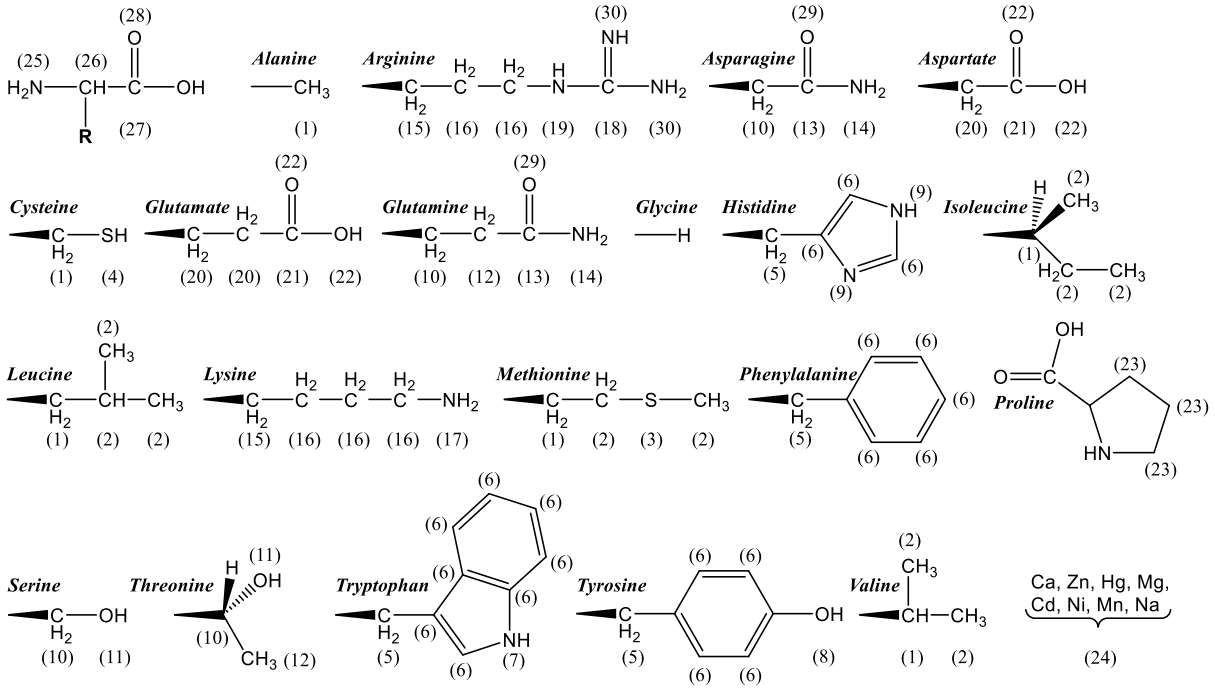

Figure 1. Protein atom types (4 for the main chain + 25 for the side chains + 1 for metals).

Shells configuration

As stated previously, SMoG2001 used two shells of distance for the contact between atoms: 0-3.5Å and 3.5-4.5Å. Using the testing set of 195 complexes, it leads to a correlation of R=0.418, whereas more recent scoring functions lead to correlation around R≈0.60 in the same testing set. To improve the function, we tested 272 new sets of intervals, using from 2 to 5 shells. All results are reported in SI: the best average correlation coefficient on the 20 testing sets was found to be 0.401 and is achieved with seven shell configurations (highlighted in orange in Tables S2/S3/S4). They all share a first shell that ends at 3.0Å (probably taking into account hydrogen bonds) and a last shell that ends at 8.5 or 9.0Å (taking into account long range effects). Four out of the seven configurations are made with three shells (the second shell ending between 5.0 and 6.0Å, probably taking into account desolvation) and the remaining three have four or five shells. We note that other configurations provide very similar results, for example 23 configurations lead to an average correlation of 0.400. In order to weight the contributions to the score from the

different shells, it appeared that dividing the score of each shell by the middle distance of the shell improved slightly the correlation (for example for a shell 5.0-9.0Å, we divided the score by 7.0).

Contact between atoms

In SMoG2001, two shells of contact were used using a step function $\Delta(r)$ (see left panel of Figure 2). For example, if $r < 3.5$Å then $\Delta(p) = 1$ for the first shell and $\Delta(p) = 0$ for the second shell. We started the development of a new scoring function to use it in a drug design software where the position of the ligand is optimized at every step. As such, it seemed more natural to use a continuous function that will result in a continuous binding free energy changes upon movements of the ligand atoms. We have thus used a function such as the one presented on the right panel of Figure 2 for each shell (we used a value of $\Delta r = 0.2$ Å for the previous comparison between all the shell configurations). We note that other approaches have been used to obtain continuous functions for contact, such as smoothing intervals by gaussians as used in the DSX function[16]. The influence of Δr (which defines the slope at each border) on the correlation for the seven shells that lead to the highest average correlation was then assessed and was found to be very low (see SI) and the average correlation coefficient never exceeded 0.401. Finally, we retained the shell configuration 0-3.0-5.0-8.5 Å and the value $\Delta r = 0.2$ Å, and call it KBP2016. Having fewer shells leads to better statistics, we thus preferred to keep a configuration with three shells over those with four or five shells. As stated above, very similar results would be obtained with the four configurations made with three shells. We retained the one where the second shell ends at 5.0 Å (and not 5.5 or 6.0) because if we assume that this shell serves to describe the desolvation effect, a distance of 5.0 Å for the first solvation shell is more plausible than 5.5 or 6.0 Å.

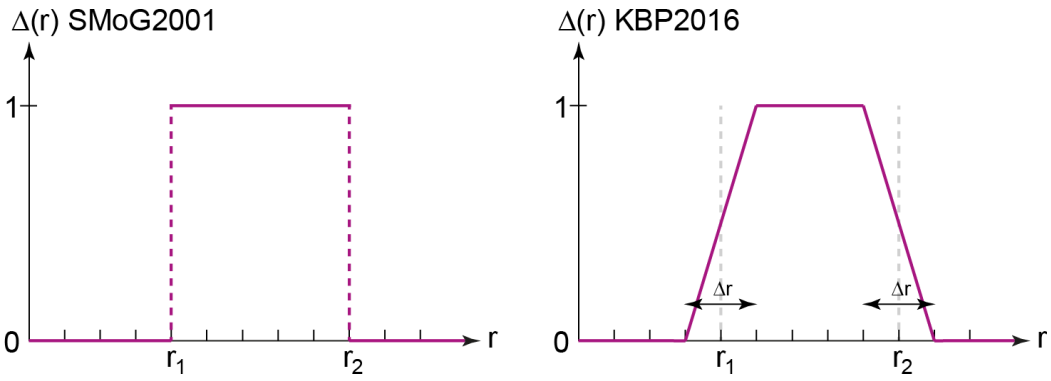

Figure 2. Contact function for the scoring functions.

Validation of the Training Set Size

During the construction of the training set, we went from 10,662 to 1,038 complexes. One may wonder if this process resulted in a loss of available data that could have led to a more accurate potential. To validate the procedure, we randomly picked different number of complexes from the 20 training sets of 938 complexes, trained the scoring function on these subsets and then computed the correlation coefficient against the 20 testing sets. Results are reported in Figure 3: when the training can be performed, the correlation coefficient converges very quickly and 300 complexes are enough to reach $\langle R \rangle \approx 0.4$. However, when the training set is too small, it may not be possible to train the function because some contacts never occur. For each value of the training set size, we report in Figure 3 the number of successful trainings (out of the 20 performed), and we can see that at least 800 complexes are needed to train the 20 potentials. When a training cannot be performed on a given set, adapting the atomtypes by reducing their number will solve the issue. Thus, our training set is large enough to derive the statistics needed to build the potential. One of the main advantages of knowledge-based potential is that only structural information is needed. As such, at the beginning of the development of knowledge-based potentials it was predicted that the accuracy of these functions would increase with the number of available crystallographic structures. However, we show here that we have reached a state where improvements with more

complexes in the training set will be very low. This is probably also due to the fact that our training set has explicitly been constructed to contain very diverse complexes. Thus, we expect that our scoring function will be useful for very diverse proteins (except when covalent ligands are involved), even if exceptions will always occur[17].

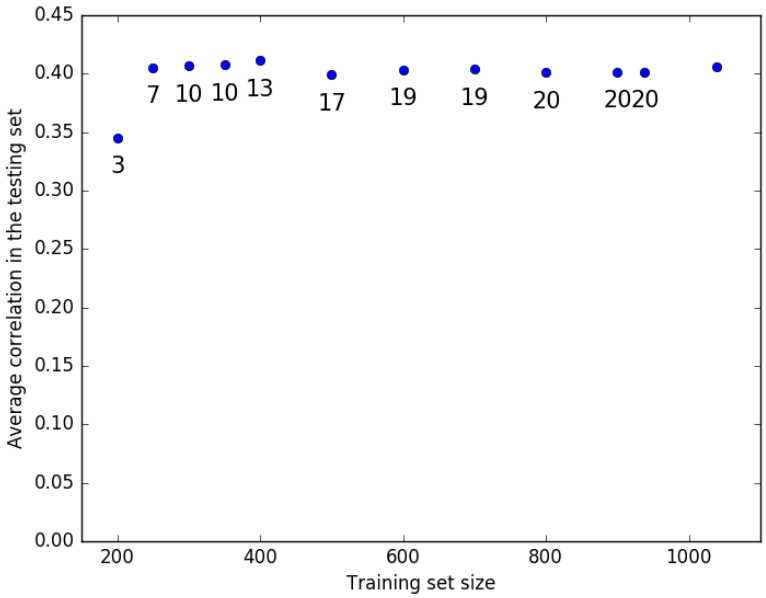

Figure 3. Correlation in the testing set with different size of the training set. Under each point is reported the number of successful training (out of the 20 performed). The last point corresponds to the correlation obtained when the potential is trained and tested on the full training set of 1,038 complexes.

Repulsion term

A main drawback of the SMoG2001 scoring function is that there is no term accounting for repulsion: the knowledge-based potential does not make a difference between a contact at 0.5 Å and at 3 Å for example. This is not a problem when the goal is only to score the interaction from a crystallographic structure because the repulsion has usually already been minimized in such a structure. However, when new ligands are generated inside the active site of a protein (which is the goal of our *de novo* drug design program OpenGrowth[1]), there is no way to avoid clashes. To circumvent this problem, a hard-wall potential was used in SMoG2001: when the distance between two atoms was smaller than the sum of the van der Waals radii multiplied by a scaling factor, then

a steric clash was found and the structure was discarded. However, this approximation is too crude. We have thus decided to add a repulsion term to the knowledge-based potential KBP2016 using the repulsive part of a Lennard-Jones potential (the A_{ij} term is computed with the Amber van der Waals parameters[18]):

$$E_{Repulsion} = \sum_{i,j} \frac{A_{ij}}{r_{ij}^{12}} \qquad (2)$$

The pair-wise energy is added to the knowledge-based score with a multiplicative coefficient α determined empirically. To determine this coefficient, we have proceeded in two ways. First, we used the L-BFGS optimization algorithm implemented in the dlib C++ library[19] to find the value of α that maximizes the correlation of the function $KBP2016+\alpha*E_{repulsion}$ with experimental binding free energies. Here, we used the full training set of 1,038 complexes to perform the optimization. With α=0 the correlation is R=0.406 and increases to R=0.426 with α=0.535. We also computed the correlation over the 20 randomly generated testing sets for the function $KBP2016+\alpha*E_{repulsion}$ with different values of α. We found that α=0.485 gives the highest average correlation (<R>=0.416), i.e. a very similar value than the optimized one.

Secondly, we have mimicked a docking procedure. For each ligand in the full training set, we have looked for all the rotor bonds (as defined by OpenBabel). When a rotatable bond was found, we prepared rotamers by rotating the fragments on each side of the rotatable bond by steps of 10°. This was only done if the fragment to rotate contained between 4 and 20 atoms in total. The hybrid score $KBP2016+\alpha*E_{repulsion}$ was then computed for each rotamer with different values of α. To relax the hard-wall potential approximation of SMoG2001, here both intermolecular and intramolecular interactions are calculated with the Lennard-Jones equation (equation (2)). We note θ_{opt} the value of the rotation angle for which the rotamer score is the lowest, and we looked at

how often $\theta_{opt} = 0°$ (meaning that the rotamer with the lowest energy is the original one, which is the expected behavior). We also allowed small deviations, and looked at how often $\theta_{opt} \in \{-10°, 10°\}$ (meaning that the rotation that leads to the lowest score is -10°, 0° or +10°), or $\theta_{opt} \in \{-20°, 20°\}$. The degeneracy of the lowest energy was also calculated for each case: ideally, there should be only one rotamer with the lowest energy. Results are presented in Table 2. For each target angle (0°, $\{-10°, 10°\}$, or $\{-20°, 20°\}$) and for various α values, we report the percentage of success i.e. the percentage of times the rotamer with the lowest score falls into the target angle range. Amongst the rotamers that fall into the target angle range, we calculated the fraction that has a degeneracy of 1 (meaning that only 1 out of the 36 rotamers has the lowest score). We observed that higher α values increase the success rate to find the optimal rotamer and are more likely to have a degeneracy of 1 (with up to ~80% of success when variations of +/-20° are allowed, and with a degeneracy of 1 ~80% of the times), however convergence seems to be reached for α=0.35. Even if direct comparisons between protein rotamers and ligand rotamers cannot be done, we want to point out that these results are in the same range as the Dunbrack's rotamers library where the overall prediction rate is 73%[20]. For the remaining of the study, we have used a value of α=0.535 since it allows to obtain both the optimal prediction of binding free energies and the best structure predictions. It is interesting to note that similar optimal values are found by two different means (optimal correlation in the sets and docking-like procedure).

Table 2. Influence of the α coefficient on the probability to find the correct rotamer. For each target angle, we report the percentage of times the rotamer with the lowest score falls into the target angle range (%(success)). Amongst the rotamers which falls into the target angle range, we also report the fraction for which only 1 out of the 36 rotamers has the lowest score (%(g=1)).

α	$\theta_{opt} = 0°$		$\theta_{opt} \in \{-10°, 10°\}$		$\theta_{opt} \in \{-20°, 20°\}$	
	%(success)	%(g=1)	%(success)	%(g=1)	%(success)	%(g=1)
0.00	31	45	44	62	49	65
0.05	45	63	62	73	70	76

0.10	49	65	66	74	74	77
0.15	51	67	67	75	75	78
0.20	52	68	69	76	77	78
0.25	53	68	70	76	78	78
0.30	54	69	71	76	79	79
0.35	55	69	72	77	79	79
0.40	54	69	71	77	79	79
0.45	54	69	72	77	79	79
0.50	54	69	71	77	79	79
0.55	53	69	72	77	79	79
0.60	54	69	72	77	78	79
0.65	54	69	73	77	78	79
0.70	54	69	73	77	78	79
0.75	53	68	72	77	78	78
0.80	53	68	72	77	78	78
0.85	53	68	72	77	78	78
0.90	52	68	72	77	78	78
0.95	52	68	73	77	78	79
1.00	52	68	72	77	78	79

When Li *et al.* developed their testing set and compared several scoring functions against it, they proposed the use of a simple descriptor (variations of the solvent-accessible surface area upon binding) and showed that its correlation against experimental binding free energies is R=0.606 i.e. the second-best currently available at that time[10]. One can thus wonder why we developed a new scoring function. We performed the same structural analysis as with KBP2016 (full results are presented in SI): when using no repulsion term, ΔSASA[21] is capable of recovering the native binding pose only 24% of times (48% if we allow deviations up to +/-20°) and thus should not be used directly for growing or docking ligands (as already pointed out by Li *et al.*[10]). If a repulsion term is included (similarly to what was done with KBP2016), the optimal value is obtained with α=0.80. On average, all percentages of success (whatever the target angle) are lower by 5% than with the KBP2016 hybrid function. Thus, the hybrid KBP2016 function provides better

structural prediction than the hybrid ΔSASA. Moreover, since the mathematical form of KBP2016 is simple, it is faster to compute a KBP2016 score than computing ΔSASA (test performed with the freeSASA library[21]).

Additional terms

We also tried to add to the new scoring function (in addition to the Lennard-Jones repulsion term) empirical terms to account explicitly for hydrogen bonds, desolvation, ligand constraint (i.e. enthalpic deformation) or fixation penalty (i.e. loss of ligand entropy upon binding). These terms aim to account for energy terms that the knowledge-based potential fails to incorporate. We added these terms one by one by optimizing their weights on the full training set with the L-BFGS algorithm implemented in the dlib C++ library[19]. To take into account hydrogen bonds, we implemented the function proposed by X-CScore[22]. For desolvation, the difference of Solvent Surface Accessible Area (ΔSASA) upon binding was used[21,23]:

$$\Delta SASA = SASA(complex) - SASA(protein) - SASA(ligand) \qquad (3)$$

Ligand constraints were evaluated by optimizing the geometry of the ligands and taking the difference of energy between the optimized geometry and the geometry of the ligand in the complex. Different levels of geometry were considered, from force fields (UFF, MMFF94, GAFF or Ghemical), to semi-empiral (PM6) or DFT (M06-2X/6-31+G** in PCM).

For the hydrogen bond, desolvation and ligand constraint, no conclusive results were obtained since the optimal weight of each term was very close to 0 and the correlation coefficient with experimental data barely increased. For example, for the ligand constraint the correlation coefficient increased by less than 0.001 unit. For the desolvation, it increased by 0.006 unit, and due to the additional computational cost we decided not to include such a term in the new scoring function. Consequently, these terms were not added in the new scoring function. Regarding the

entropic fixation penalty, it was included *via* the number of rotatable bonds in the ligands[24]. We considered two approaches: either directly adding the number of rotatable bonds (as used in AutoDock[25]) or the function developed in X-CScore[22] where $\Delta G_{Fixation} = \sum_{Atoms} \alpha_i$. If the atom i is not involved in any rotor, $\alpha_i = 0$; if the atom i is involved in one or more than two rotors $\alpha_i = 0.5$; if the atom i is involved in two rotors, $\alpha_i = 1$. We added either the number of rotor or the X-CScore rotor function to the new scoring function by optimizing the weight of each term over the training set. Thus, the new hybrid function has the form: $KBP2016+\alpha*E_{repulsion}+\beta*Rotor$. We observed that these terms add a small improvement to the scoring function since the correlation increases, from 0.426 to 0.439 (the same correlation is obtained for both the number of rotatable bonds or the X-CScore function). Considering the more arbitrary form of the X-CScore function, we decided to retain the number of rotors as a measure of the ligand fixation penalty (i.e. loss of conformational ligand entropy) with β=1.913.

When a ligand binds a protein, not only there is a loss of conformational entropy, but there is also a loss of translational entropy (by loss, we do not mean "complete loss" but more "significant decrease"). Following standard statistical mechanics[26], the translation entropy S_t of a molecule with mass *m* can be written as:

$$S_t = R * \left[\frac{5}{2} + ln\left(\frac{k_B T}{P}\right) + \frac{3}{2} * ln\left(\frac{2\pi m k_B T}{h^2}\right)\right] \qquad (4)$$

Thus, the variation of translational entropy can be written as: $\Delta S_t \approx u + \gamma * \ln\left(\frac{m_L * m_P}{m_C}\right)$, where m_L, m_P and m_C are respectively the mass of the ligand, the protein and the complex and we investigated if an additional term $\ln\left(\frac{m_L * m_P}{m_C}\right)$ may further improve the scoring function[27]. When using the same procedure as before (with the L-BFGS optimization algorithm on the full training set), we found that the correlation slightly increases from 0.439 to 0.445 and a similar result is

obtained when using only $\ln(m_L)$ (as expected, since $m_C \approx m_P$). With a term $\gamma * m_L$ the correlation increases by less than 0.001 meaning that the new term is not due to another way of considering enthalpic interactions (whereby the bigger the ligand, the higher the number of contacts and the more favorable the binding), but more likely has an origin in translational entropy. Since similar results are obtained with $\ln\left(\frac{m_L * m_P}{m_C}\right)$ and with $\ln(m_L)$, we decided to keep an additional term $\gamma * \ln(m_L)$, with γ=-21.974. The dependence of the binding entropy on molecular mass has been controversial[28], and we will discuss the validity of this assumption in the following as well as the sign of γ.

Final form of the SMoG2016 function

At the end, our new scoring function SMoG2016 has the following form, where KBP2016 is the knowledge-based potential, Rotor is the number of rotatable bonds in the ligand and m_l is the ligand mass:

$$SMoG2016 = KBP2016 + 0.535 * \sum_{i,j} \frac{A_{ij}}{r_{ij}^{12}} + 1.913 * Rotor - 21.974 * \ln(m_L) \qquad (5)$$

Raw scores coming from knowledge-based potentials have no units and can be in a large range. Thus, the scores can be scaled by the following function to place them in a realistic range of kcal/mol (where the ⟨ ⟩ means an average on the training set of 1038 complexes):

$$Final\ score = \frac{\langle Experimental\ energies \rangle}{\langle Raw\ scores \rangle} * Raw\ score = 0.032 * Raw\ score \qquad (6)$$

To illustrate the improvement made from SMoG2001 to SMoG2016, we present in Figure 4 and Figure 5 the correlations between experimental data and these two scoring functions in the testing set of 195 complexes from Li *et al.*. The correlation coefficient increased from R=0.418 to R=0.570 (with KBP2016, R=0.587). Moreover, the SMoG2016 plot clearly appears less dispersed, as

illustrated by a decrease in standard deviation 3.39 to 1.68 when going from SMoG2001 to SMoG2016. We also present in Figure 6 the correlation of SMoG2016 in the training set.

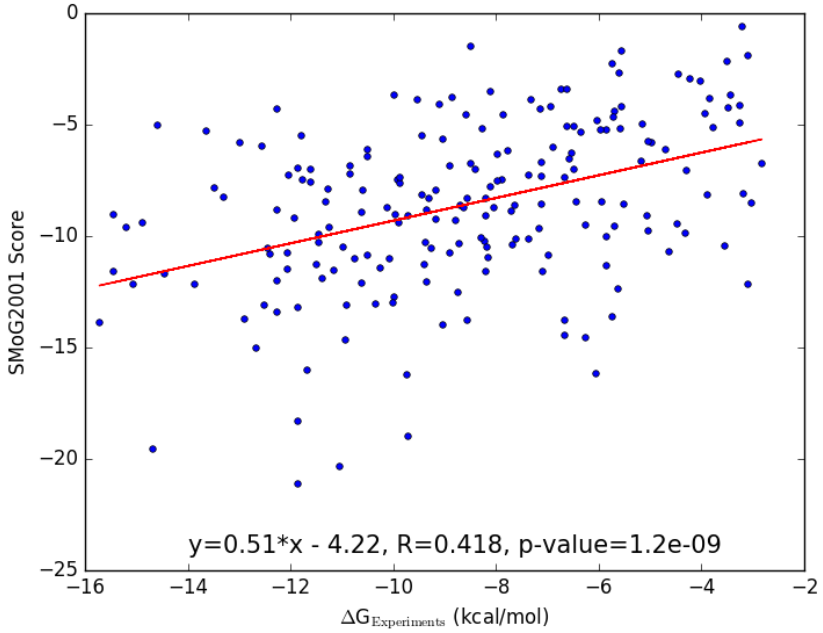

Figure 4. Correlation between SMoG2001 score and experimental binding free energies in the testing set (195 complexes).

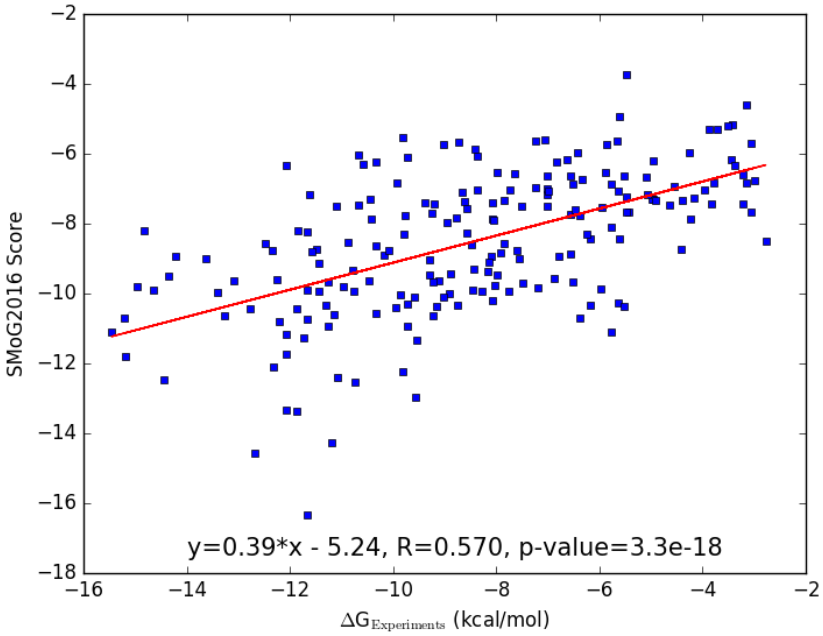

Figure 5. Correlation between SMoG2016 score and experimental binding free energies in the testing set (195 complexes).

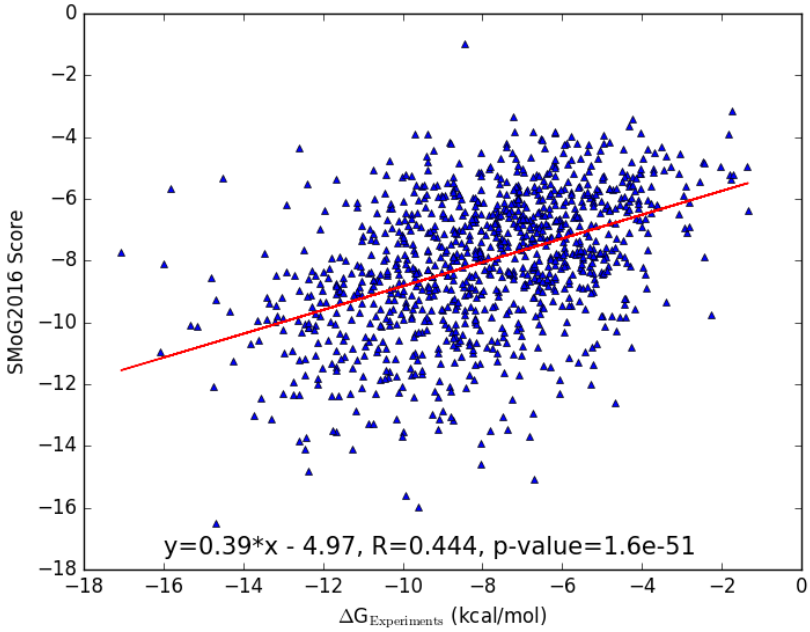

Figure 6. Correlation between SMoG2016 score and experimental binding free energies in the training set (1,038 complexes).

Program description

The function SMoG2016 is available in a program made in C++ which uses the OpenBabel library (v2.4.1)[12]. The user can directly use the scoring function with a mol2 or sdf ligand file and with a pdb protein file as inputs. Over the full training set, it takes from 1.4 to 114 ms to compute a score, with an average of 15.6ms (on a single core of an AMD Opteron 6376 CPU @ 2.3GHz). It is also possible to create a new potential from a set of complexes that was previously constructed (it takes ~20 minutes to train the potential over 1,038 complexes). All parameters files (such as those for the knowledge-based potential) are available together with the source code under the name SMoG2016.tar.gz at the address https://sourceforge.net/projects/opengrowth/files/.

Discussion and conclusion

During the development of the knowledge-based potential, we have shown that it is highly important to take into account long-range contacts. Using three shells (with the configuration 0-3.0-5.0-8.5Å) proved to be the most efficient way to reproduce experimental data since using more

shells would result in less accurate statistics. The three shells correspond to physically understandable terms: short range interactions (hydrogen bonds), desolvation and medium range interactions, long range interactions (such as electrostatic). The approach we have used (average correlation coefficient over 20 testing sets) ensures that our decision for the shell configurations are not biased and are independent of the training set. The correlation coefficient in the testing set of 195 complexes increased from R=0.418 with SMoG2001 to R=0.570 in SMoG2016 after inclusion of additional parameters to the KBP2016 function. The use of a standard test set allows us to compare our new scoring function to existing ones: according to the Pearson correlation coefficient, it is ranked the 7th one over 25 scoring functions (21 compared by Li *et al.*[10], ID-Score[29], the new hybrid AutoDock/AutoDockVina function[17], SMoG2001[9] and SMoG2016). However, we note that the function with the highest correlation coefficient (ID-Score[29]) is based on a huge number of descriptors (50) which can make its transferability and portage to new programs complicated. For the secondly ranked function, the testing set used is included in the training set which may explain their strong results (as pointed out by the authors[17]). The functions ranked 3rd and 4th are respectively X-CScore[30] and the variation of solvent-accessible surface area ΔSAS (a simple descriptor used by Li *et al.*), and we have shown that the structure prediction of ΔSAS is outperformed by SMoG2016, even when an additional term taking into account repulsion is added to ΔSAS. Based on the standard deviation criteria, SMoG2016 (SD=1.68) is ranked the 2nd best after ID-Score (SD=1.63), whereas all functions compared by Li *et al.* have standard deviations higher than 1.78 (see Table 3). Interestingly, SMoG2016 is better ranked than other famous scoring functions such as GlideScore (both SP and XP), both in term of Pearson correlation coefficient and standard deviations. We also note that it is the highest ranked knowledge-based

function. Finally, the fact that our function is very simple, very fast and available as open-source will make it a valuable tool in drug design.

One of the reasons why knowledge-based scoring functions have attracted a lot of attention at their early ages is that only structural information are needed, and no experimental binding free energies are needed to derive the different energy contributions used. Thus, it was expected that they would become more and more accurate when enough structural data (X-Ray or NMR) would become available. However, it seems that we have reached a state where improvements for generic potential can now only be very small. Indeed, we report in Table 3 the correlation coefficient of various scoring functions in the same testing set, as well as standard deviations and the year of publication. Besides a few exceptions, it seems that the community has converged towards a plateau at 0.6 (in terms of correlation coefficient). Several hypotheses can be suggested to explain why we have reached this limit. The simplicity of the models (either knowledge-based potentials, empirical scores or physics-based functions) is the first explanation that comes to mind. Widespread conformational changes in proteins upon binding or some crucial terms may still be missing in the binding free energy estimations, very likely entropic terms. By using another approach based on descriptors, Li *et al.* were able to overcome the R=0.6 limitation with ID-Score. Such a function is very useful for scoring but may be more complicated to use when docking or growing new compounds in the active site. The reason why we have reached a limit may also be that it may be crucial for a given family of protein to include one type of energetic term, whereas for another one the incorporation of another term will be more important. Thus, we believe that it may now be more suitable in some cases to derive potentials for specific family of proteins whenever possible (albeit at the expense of transferability). This will be possible because a knowledge-based potential can be calculated from any database of crystallographic or NMR

structures and there is no need of consistent experimental data on binding free energies. Finally, another possible explanation for the limit reached with the correlation coefficient is the use of a single point approach to estimate a value that is a thermodynamic average over many conformations. To overcome this approximation, it may be needed to use a new approach and perform averages on more protein structures, which will only be of interest if the scoring function is fast enough.

Table 3. Comparison of performances of 25 scoring functions. Data from 20 of the functions are coming from Li *et al.*[10]. R=Pearson correlation coefficient against the testing set of 195 complexes. SD=Standard Deviation. n.a.=not available.

Function	R	SD	Year	Ref.
ID-Score	0.753	1.63	2013	29
AutoDockHybrid	0.640	n.a.	2016	17
X-Score	0.614	1.78	2002	30
ΔSAS	0.606	1.79	2014	10
ChemScore@SYBYL	0.592	1.82	1998	7
ChemPLP@GOLD	0.579	1.84	2009	31
SMoG2016	0.570	1.68	2016	This work
PLP1@DS	0.568	1.86	2000	32,33
G-Score@SYBYL	0.558	1.87	1997	4
ASP@GOLD	0.556	1.88	2005	34
ASE@MOE	0.544	1.89	n.a.	n.a.
ChemScore@GOLD	0.536	1.90	2003	7
D-Score@SYBYL	0.526	1.92	2001	35
Alpha-HB@MOE	0.511	1.94	n.a.	n.a.
LUDI3@DS	0.487	1.97	1998	36,37
GoldScore@GOLD	0.483	1.97	1997	4
Affinity-dG@MOE	0.482	1.98	n.a.	n.a.
LigScore2@DS	0.456	2.02	2005	38
GlideScore-SP	0.452	2.03	2006	39,40
SMoG2001	0.418	3.39	2001	9
Jain@DS	0.408	2.05	2006	27
PMF@DS	0.364	2.11	2006	41–44
GlideScore-XP	0.277	2.18	2004	45
London-dG@MOE	0.242	2.19	n.a.	n.a.
PMF@SYBYL	0.221	2.20	1999	41–44

Even though experimental binding free energies are not needed to derive a knowledge-based potential, we took advantage to the fact that they are now more easily available thanks to the work of the Wang's group with the database of complexes PDBbind-CN[11]. These data lead us to try to improve even further the new function by including additional terms empirically. Surprisingly, it appeared that addition of some terms seemingly important such as hydrogen bonds, desolvation or ligand deformation energy failed to improve the correlation with experimental data. We were surprised by the lack of significant improvement upon adding ligand constraint to take into account the enthalpic deformation of the ligand upon binding (even when this constraint is calculated at a high level of DFT such as M06-2X/6-31+G** in PCM), and we currently have no explanations for this observation. As already pointed out, the use of a simple descriptor (ΔSAS) can provide a correlation of up to 0.606, we were thus also surprised to observe that the inclusion of a term based on ΔSAS on the hybrid function improved the correlation coefficient by only 0.006 units. This can be seen as a confirmation that the desolvation is already taken into account in the knowledge-based potential, as pointed out by Ishchenko and Shakhnovich[9]. Similarly, the fact that a term describing explicitly hydrogen bonds didn't improve the scoring function may be due to the fact that the first shell ends at 3.0 Å and thus describes mainly this kind of interactions. On the other hand, taking into account the loss of entropy (both conformational and translational) upon binding proved to be useful: correlation coefficient increased from 0.426 to 0.439 in the training set after including a rotor contribution, and from 0.439 to 0.445 after including a term for translational entropy. We didn't consider the inclusion of rotational and vibrational entropy yet since this would require higher computational cost and would slow our function. The inclusion of entropy can be related to the work of Jenck on enzymatic reaction and the so-called Circe effect[46,47]: one of the reason why enzymes can accelerate reactions is through the destabilization

of the substrate in the active site, decreasing the actual energy barrier needed for the reaction to occur. Additionally, it was proposed that the destabilization occurs mainly *via* the entropy loss of the ligand upon binding. Thus, entropy plays a major role in K_M. It is safe to assume that the same behavior will occur for all proteins ligands and that entropy loss is the most important term in K_d or ΔG, which explains why its inclusion in the scoring function was successful.

Gilson *et al.* discussed the dependence of energetic terms on ligand mass and concluded that "molecular mass has a negligible effect upon the standard free energy of binding for biomolecular systems"[28]. Even though some energetic decomposition may lead to terms that depend on the mass, they should in the end cancel out. However, we observed an improvement of the scoring function when a term $\gamma * ln(m_L)$ was included with an optimal value of γ=-21.974 i.e. $\gamma<0$. It is important to recall here that knowledge-based potentials directly give rise to binding free energies (and not binding energies)[48]. Thus, the variation of entropy upon binding is already partially taken into account *via* KBP2016, with probably a too high weight for the translational entropy. The negative value of γ means that a part of the translational entropy has to be removed from KBP2016 to avoid over counting.

Finally, we note that the quality of a scoring function is usually judged according to two parameters: (1) its ability to predict the most accurate free energy of interaction, (2) its docking accuracy, i.e. its ability to predict the preferred orientation of the ligand inside the active site[5]. Our main goal in developing SMoG2016 was the former since the initial purpose of this function is to include it in the OpenGrowth software[1]. However, we have shown that after inclusion of a repulsive Lennard-Jones, we are capable of finding the most favorable rotamer ≈80% of the times (if we allow small variations on the optimal angle), and ≈80% of the times the most favorable rotamer is the only one to have the lowest energy (degeneracy of 1) (see Table 2).

Supporting Information

Description of the SMoG2001 scoring function. Evaluation of different shells for the knowledge-based potential. Influence of the slope of the contact function. Influence of the α parameter on the correlation for hybrid scores with ΔSASA. List of complexes contained in the testing set (195 complexes). List of complexes contained in the training set (1,038 complexes).

Acknowledgments

This work was supported by Defense Advanced Research Projects Agency (DARPA) Contract HR0011-11-C-0093. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University.

References

(1) Chéron, N.; Jasty, N.; Shakhnovich, E. I. OpenGrowth: An Automated and Rational Algorithm for Finding New Protein Inhibitors. *J. Med. Chem.* **2016**, *59* (9), 4171–4188.

(2) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discov.* **2004**, *3* (11), 935–949.

(3) Liu, J.; Wang, R. Classification of Current Scoring Functions. *J. Chem. Inf. Model.* **2015**, *55* (3), 475–482.

(4) 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* (3), 727–748.

(5) Liu, J.; Wang, R. Classification of Current Scoring Functions. *J. Chem. Inf. Model.* **2015**, *55* (3), 475–482.

(6) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein–ligand Docking Using GOLD. *Proteins Struct. Funct. Bioinforma.* **2003**, *52* (4), 609–623.

(7) 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* (5), 425–445.

(8) DeWitte, R. S.; Shakhnovich, E. I. SMoG: De Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *J. Am. Chem. Soc.* **1996**, *118* (47), 11733–11744.

(9) Ishchenko, A. V.; Shakhnovich, E. I. SMall Molecule Growth 2001 (SMoG2001): An Improved Knowledge-Based Scoring Function for Protein−Ligand Interactions. *J. Med. Chem.* **2002**, *45* (13), 2770–2780.
(10) 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* (6), 1717–1736.
(11) 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* (6), 1700–1716.
(12) O'Boyle, N.; Banck, M.; James, C.; Morley, C.; Vandermeersch, T.; Hutchison, G. Open Babel: An Open Chemical Toolbox. *J. Cheminformatics* **2011**, *3* (1), 33.
(13) Tanimoto, T. An Elementary Mathematical Theory of Classification and Prediction. *Internal IBM Technical Report*. 1958.
(14) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23* (21), 2947–2948.
(15) Chen, W. W.; Shakhnovich, E. I. Lessons from the Design of a Novel Atomic Potential for Protein Folding. *Protein Sci.* **2005**, *14* (7), 1741–1752.
(16) Neudert, G.; Klebe, G. DSX: A Knowledge-Based Scoring Function for the Assessment of Protein–Ligand Complexes. *J. Chem. Inf. Model.* **2011**, *51* (10), 2731–2745.
(17) Tanchuk, V. Y.; Tanin, V. O.; Vovk, A. I.; Poda, G. A New, Improved Hybrid Scoring Function for Molecular Docking and Scoring Based on AutoDock and AutoDock Vina. *Chem. Biol. Drug Des.* **2016**, *87* (4), 618–625.
(18) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999–2012.
(19) King, D. E. Dlib-Ml: A Machine Learning Toolkit. *J Mach Learn Res* **2009**, *10*, 1755–1758.
(20) Dunbrack, R. L.; Cohen, F. E. Bayesian Statistical Analysis of Protein Side-Chain Rotamer Preferences. *Protein Sci.* **1997**, *6* (8), 1661–1681.
(21) Mitternacht, S. FreeSASA: An Open Source C Library for Solvent Accessible Surface Area Calculations. *F1000Research* **2016**, *5*, 189.
(22) 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* (1), 11–26.
(23) Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55* (3), 379–IN4.
(24) Andrews, P. R.; Craik, D. J.; Martin, J. L. Functional Group Contributions to Drug-Receptor Interactions. *J. Med. Chem.* **1984**, *27* (12), 1648–1657.
(25) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A Semiempirical Free Energy Force Field with Charge-Based Desolvation. *J. Comput. Chem.* **2007**, *28* (6), 1145–1152.
(26) McQuarrie, Donald A.; Simon, John D. *Molecular Thermodynamics*; University Science Books: Herndon, Virginia, USA, 1999.

(27) Jain, A. N. Scoring Noncovalent Protein-Ligand Interactions: A Continuous Differentiable Function Tuned to Compute Binding Affinities. *J. Comput. Aided Mol. Des.* **1996**, *10* (5), 427–440.
(28) Zhou, H.-X.; Gilson, M. K. Theory of Free Energy and Entropy in Noncovalent Binding. *Chem. Rev.* **2009**, *109* (9), 4092–4107.
(29) Li, G.-B.; Yang, L.-L.; Wang, W.-J.; Li, L.-L.; Yang, S.-Y. ID-Score: A New Empirical Scoring Function Based on a Comprehensive Set of Descriptors Related to Protein–Ligand Interactions. *J. Chem. Inf. Model.* **2013**, *53* (3), 592–600.
(30) 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* (1), 11–26.
(31) Korb, O.; Stützle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49* (1), 84–96.
(32) Verkhivker, G.; Appelt, K.; Freer, S. T.; Villafranca, J. E. Empirical Free Energy Calculations of Ligand-Protein Crystallographic Complexes. I. Knowledge-Based Ligand-Protein Interaction Potentials Applied to the Prediction of Human Immunodeficiency Virus 1 Protease Binding Affinity. *Protein Eng. Des. Sel.* **1995**, *8* (7), 677–691.
(33) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. Deciphering Common Failures in Molecular Docking of Ligand-Protein Complexes. *J. Comput. Aided Mol. Des.* **2000**, *14* (8), 731–751.
(34) Mooij, W. T. M.; Verdonk, M. L. General and Targeted Statistical Potentials for Protein-Ligand Interactions. *Proteins Struct. Funct. Bioinforma.* **2005**, *61* (2), 272–287.
(35) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search Strategies for Automated Molecular Docking of Flexible Molecule Databases. *J. Comput. Aided Mol. Des.* **2001**, *15* (5), 411–428.
(36) Böhm, H.-J. The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein-Ligand Complex of Known Three-Dimensional Structure. *J. Comput. Aided Mol. Des.* **1994**, *8* (3), 243–256.
(37) 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* (4), 309–309.
(38) Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. LigScore: A Novel Scoring Function for Predicting Binding Affinities. *J. Mol. Graph. Model.* **2005**, *23* (5), 395–407.
(39) 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.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749.
(40) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759.
(41) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42* (5), 791–804.
(42) Muegge, I. Effect of Ligand Volume Correction on PMF Scoring. *J. Comput. Chem.* **2001**, *22* (4), 418–425.

(43) Muegge, I. A Knowledge-Based Scoring Function for Protein-Ligand Interactions: Probing the Reference State. In *Virtual Screening: An Alternative or Complement to High Throughput Screening?*; Klebe, G., Ed.; Kluwer Academic Publishers: Dordrecht, 2002; Vol. 20, pp 99–114.
(44) Muegge, I. PMF Scoring Revisited. *J. Med. Chem.* **2006**, *49* (20), 5895–5902.
(45) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196.
(46) Jencks, W. P. Binding Energy, Specificity, and Enzymic Catalysis: The Circe Effect. In *Advances in Enzymology and Related Areas of Molecular Biology*; John Wiley & Sons, Inc.: Hoboken, New Jersey, États-Unis, 1975; pp 219–410.
(47) Jencks, W. P. Reaction Mechanisms, Catalysis, and Movement. *Protein Sci.* **1994**, *3* (12), 2459–2464.
(48) DeWitte, R. S.; Shakhnovich, E. I. SMoG: De Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *J. Am. Chem. Soc.* **1996**, *118* (47), 11733–11744.

For Table of Contents Use Only

A Hybrid Knowledge-Based and Empirical Scoring Function for Protein-Ligand Interaction: SMoG2016, by Théau Debroise, Eugene I. Shakhnovich, and Nicolas Chéron