Unbiasing Scoring Functions: A New Normalization and Rescoring Strategy

Giorgio Carta, Andrew J. S. Knox, and David G. Lloyd*

Molecular Design Group, School of Biochemistry & Immunology, Trinity College Dublin, Dublin 2, Ireland

Received October 31, 2006

Ligand bias can contribute significantly to the number of false positives observed in a virtual screening campaign. Using a receptor-based docking approach against a well-established target of therapeutic importance, estrogen receptor α (ERα), coupled with several common scoring functions (ChemGuass, ChemGauss2, ChemScore, ScreenScore, ShapeGauss, and PLP), taken both individually and as a consensus, we sought to examine the characteristics of molecules retrieved by each. It has been previously shown that scoring functions (mainly empirical) exhibit bias in prioritizing more complicated molecules arising from additive components within the function. Using Spearmen's correlation coefficient, we show that a large set of descriptors calculated for a docked set of molecules exhibit positive correlation with the ranked position in a hitlist. Moreover, most of these descriptors correlate well with MW. To this end, rather than penalizing the docked score of all heavy molecular weight (MW) molecules and rewarding those of lower MW, as is common practice, we examine the impact of penalizing the score only of those molecules which were of higher MW, leaving lower MW molecules unaffected. Here, we introduce a new power function to aid the process. Using scoring frequency analysis and SIFt fingerprints, we acheived a more meaningful analysis of virtual screening (VS) performance than with enrichment calculations, facilitating target-specific VS method development.

INTRODUCTION

Improvements in crystallography, genomics, and proteomics have permitted a deeper understanding of the interaction between proteins and their endogenous ligands. More specifically, in the realm of computational chemistry, development of algorithms and new mathematical approaches to describe the binding between receptor and ligand has allowed characterization of many important interactions such as hydrogen bonding, hydrophobic contacts, and van der Waals interactions. This knowledge is now commonly used in processes such as receptor-based virtual screening (RBVS), whereby the goal is to be able to simulate and accurately predict the docked position of a molecule inside a binding pocket allowing identification of "binders" from "nonbinders".

Subsequently, each docked compound receives a score equivalent to the free energy of binding, generated from the predicted binding pose. The general assumption is that molecules that bind tightly to the receptor should be given a higher score. Once scored, molecules can be ranked and sorted, and the top *x*% is generally retained for biological evaluation.

A large variety of methods have been described in literature to calculate and predict in silico the binding affinity between a receptor and ligand; however, it is also well-known that most fail to accurately predict it.[1] Currently, scoring functions can be classified in four main categories: force-field, semiempirical, empirical, and knowledge-based methods.[2] Empirical and knowledge-based methods are the most commonly used scoring functions in a VS campaign because of their inherent speed and ease of calculation.

In this work, we focused our attention on the empirical method, such as that described originally by Bohm,[3] who dissected the overall Gibbs free energy of binding into singular contributions

$$\Delta G_{\text{bind}} = \Delta G_0 + \Delta G_{\text{hb}} \sum_{\text{h-bonds}} f(\Delta R,\Delta\alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionic\_int}} f(\Delta R,\Delta\alpha) + \Delta G|A_{\text{lipo}}| + \Delta G_{\text{rot}} N_{\text{rot}}$$

where ΔG_{bind} is a summation of hydrogen bonds interactions, ionic interactions, and hydrophobic interactions, minus an entropy penalty depending on the number of rotatable bonds. A wide variety of empirical scoring functions now exist as a result of Bohm's pioneering work, several of which we have used for the purpose of this study (ChemGauss,[4] ChemGauss2,[4] ChemScore,[5] ScreenScore,[6] ShapeGauss,[7] and PLP[8]) to rank a set of docked molecules generated from the FRED docking algorithm. Empirical scoring functions by nature are additive and generally exhibit a dependence on molecular weight resulting in a bias toward heavier molecules in the top of a ranked list of molecules postscoring.[9] It has also been widely shown in literature how different scoring functions produce markedly different results when applied to same target.[1,10−13] The combination of two or more of these individual scoring functions into a consensus score can be more effective in the retrieval of active molecules and ameliorate this effect.

By plotting the ranked position produced by each scoring function for each molecule versus their associated calculated descriptors, we employed Spearmen's correlation coefficient to evaluate their intrinsic ligand bias. The descriptors which exhibited a trend toward the higher ranking molecules were selected, and the correlation among them was explored. A normalization procedure was applied to each of the scores,

* To whom correspondence should be addressed. Phone: +353-1-8962904. Fax: +353-1-6772400. E-mail: lloyddg@tcd.ie.

10.1021/ci600471m CCC: $37.00
Published on Web 06/07/2007

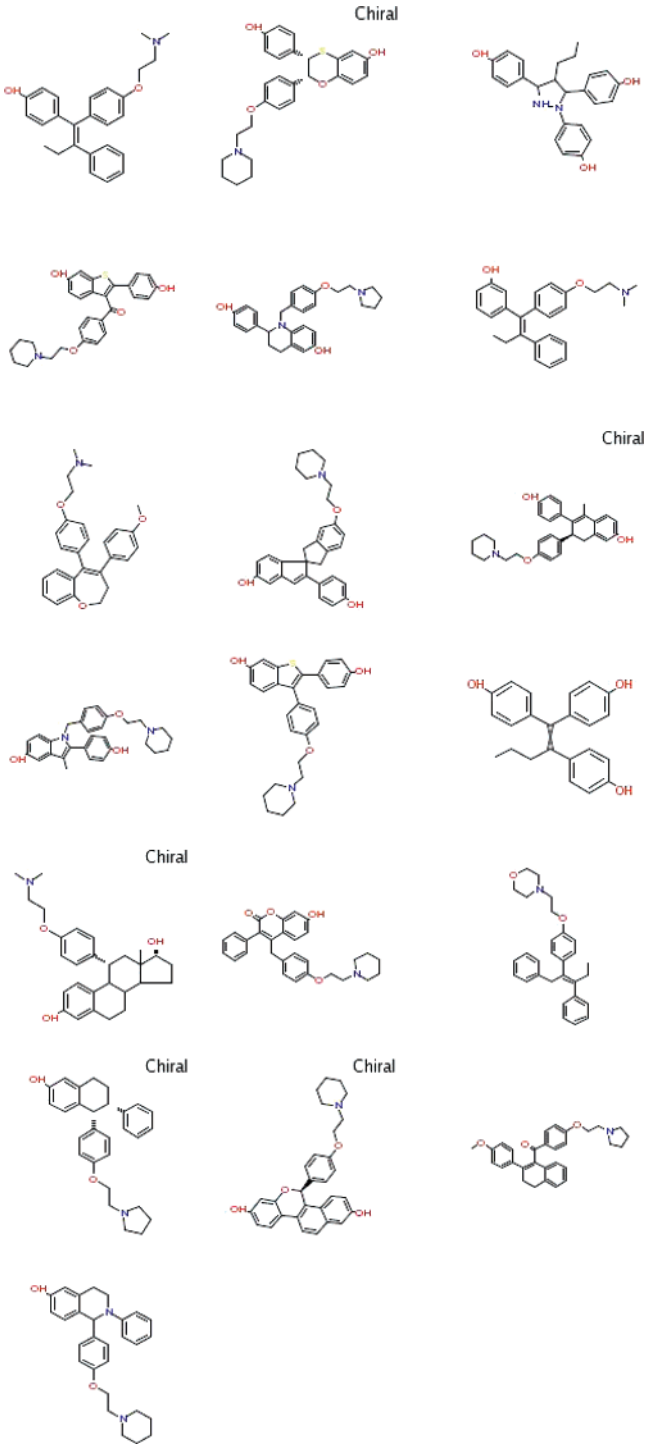

Figure 1. Structure of 19 Estrogen Receptor antagonists selected from literature.

depending on the strength of the trend exhibited by selected descriptors.

To minimize the occurrence of false positives at the top of a ranked hitlist caused by high MW, we introduce a power function whereby only high MW molecules are penalized according to their score. This serves to evenly distribute the probability of a molecule being retrieved because of disproportionate MW contribution and is shown to reduce false positives overall. The impact of this rescoring strategy within a consensus scoring scheme is also investigated.

Comparison of enrichment rates before and after the rescoring procedure was carried out to illustrate the benefits

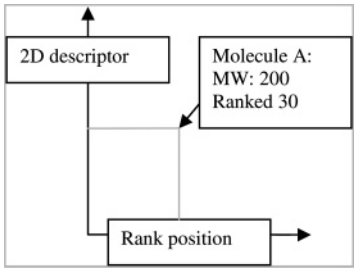

Figure 2. Descriptor (*y*-axis) plotted against ranked position (*x*-axis).

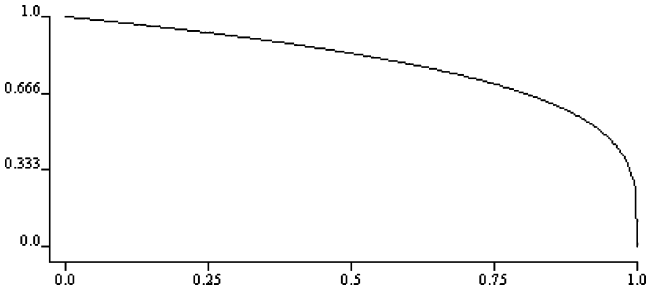

Figure 3. Plot of the function $y = (1 - x)^{0.25}$.

in reducing the number of false positives ranked highly because of increasing molecular weight. Enrichment is a commonly used measure of VS performance but it fails to depict the effect that each adjustment to a VS protocol has on the whole of the database, not just the actives. In line with this, Warren et al.[2] have observed that enrichment alone is not sufficient for optimizing a VS protocol. Thus, we propose that scoring frequency, that is, the distribution of scores observed for a set of ranked ligands, is a better method for assessing the discriminatory power of these scoring functions.

Finally, to analyze the binding modes of the ligands used in the docking studies, structural interaction fingerprints (SIFt[14]) were employed as a means of determining whether ligands with better interactions are shifted to better ranks following application of our rescoring strategy.

COMPUTATIONAL METHODS

Multiconformers Database Preparation. The following protocol was fully automated through Pipeline Pilot (PP).[15] Using PP, we trained a Bayesian to extract compounds with similar characteristics to known active ligands according to the following: number of H-Bond donors/acceptors, number of rotatable bonds, *A* log *P*, molecular Weight, and FCFP_4 fingerprint.

The Bayesian algorithm was run against Maybridge, Bionet, and IBS2004 databases to select similar molecules to known ERα actives. All molecules were standardized (e.g., keep largest fragment, check valences) and ionized at pH 7.4.

Nineteen estrogen antagonist structures were retrieved from literature (Figure 1) and standardized as above. The active and decoy set were merged and duplicates were removed to ensure that the final database size was 10 000. Omega 1.81[16] was employed for the generation of conformers. Default settings were kept with the exception of maxconfs = 10, finalopt = true, and output = mol2 to retain charges.

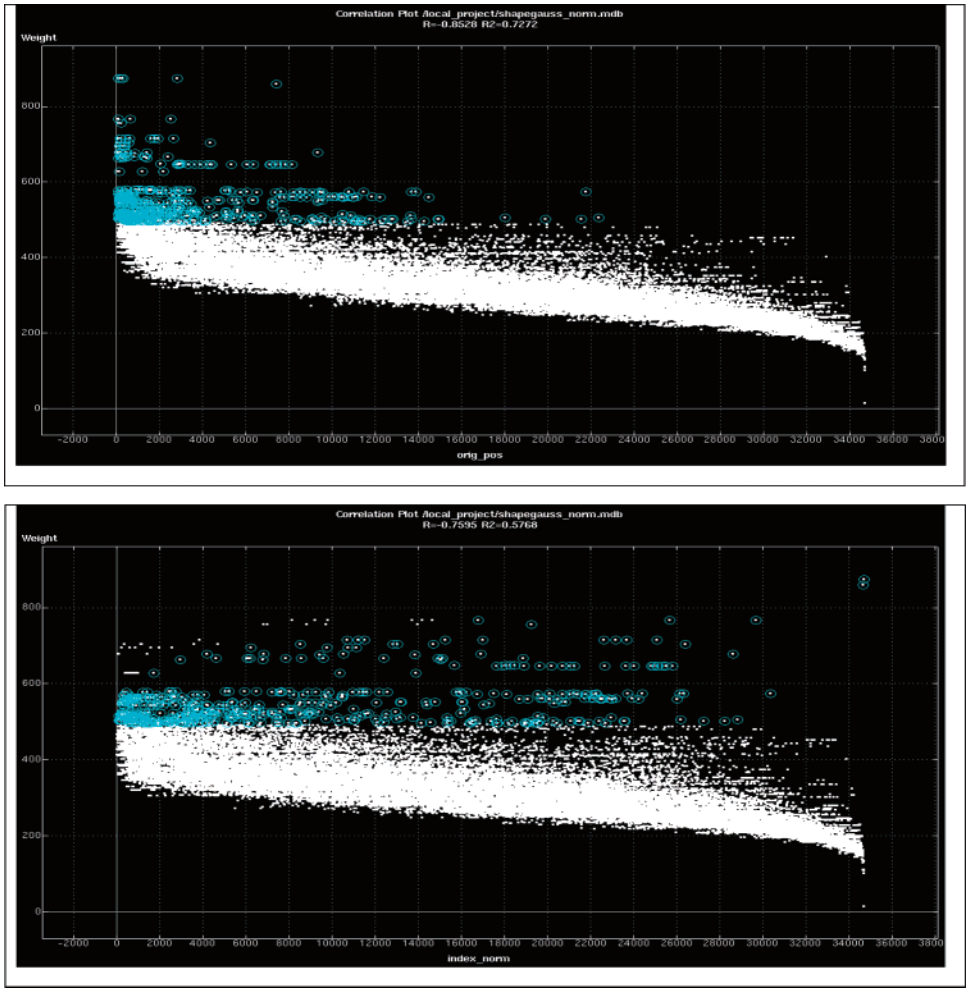

Figure 4. Effect of the rescoring procedure on the molecular-weight distribution plotted versus the rank position for the Shapegauss scored set. The *x*-axis depicts number of conformers of a 10 000 molecule set. Top: Before the normalization. Bottom: After normalization. Molecular weight above 500 is highlighted in both the plots.

Docking. The coordinates of estrogen α cocrystal structure 3ERT[17] with agonist 4-hydroxytamoxifen were downloaded from Protein Data Bank (PDB).[18] Water molecules were removed. MOE.2005.06[19] was used to add hydrogens to the protein, and a minimization protocol using MMFF94 forcefield was implemented to adjust the positions of hydrogen keeping the heavy atoms fixed at their coordinates. The ligand was extracted and used to define the 5 Å search box for docking. FRED2.11[4] was then used to generate poses inside the binding pocket using the standard default settings.

Scoring. Scoring of the docked poses for each ligand was achieved using the scoring functions implemented in FRED2.1.1 (ChemGauss2, ChemGauss, ChemScore, PLP, ScreenScore, and ShapeGauss). Consensus scoring (testconf = none) was also calculated. Knox et al.[20] have previously shown FRED2.01 in conjunction with ChemGauss and PLP to provide excellent enrichment with ERα validating our choice of docking algorithm and set of scoring functions.

Descriptor Correlation. A variety of 2D descriptors (available in MOE)[19] were calculated for all molecules and plotted against the ranked position produced by each scoring function for each molecule (Figure 2).

A trend line was calculated for each scoring function. The descriptors which exhibited a trend toward the higher ranking molecules were selected, and the correlation among them was explored. Correlation matrices were calculated using MOE.2005.06 by application of Spearmen's formula

$$R_s = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

Spearmen's rank correlation measures the correlation between two sequences' *f* values. The two sequences are ranked separately and the differences in rank are calculated at each position (*D*). Spearmen's correlation is from −1 to 1. *N* represents the total number of elements in the list.

Rescoring Procedure. To negate the trend toward higher molecular weight compounds being prioritized at the top of the database, a normalization procedure was introduced as follows. First, the scores resulting from each scoring function and the corresponding MW of each were normalized according to

$$\text{normValue} = (x - \text{min})/(\text{max} - \text{min})$$

where *x* represents either the MW or the actual score for each compound and min and max are the minimum and maximum values, respectively, of the ranked set.

The ranked list was then rescored to reduce the dependence of the scoring function on molecular weight. A new power function formula was introduced to penalize only those

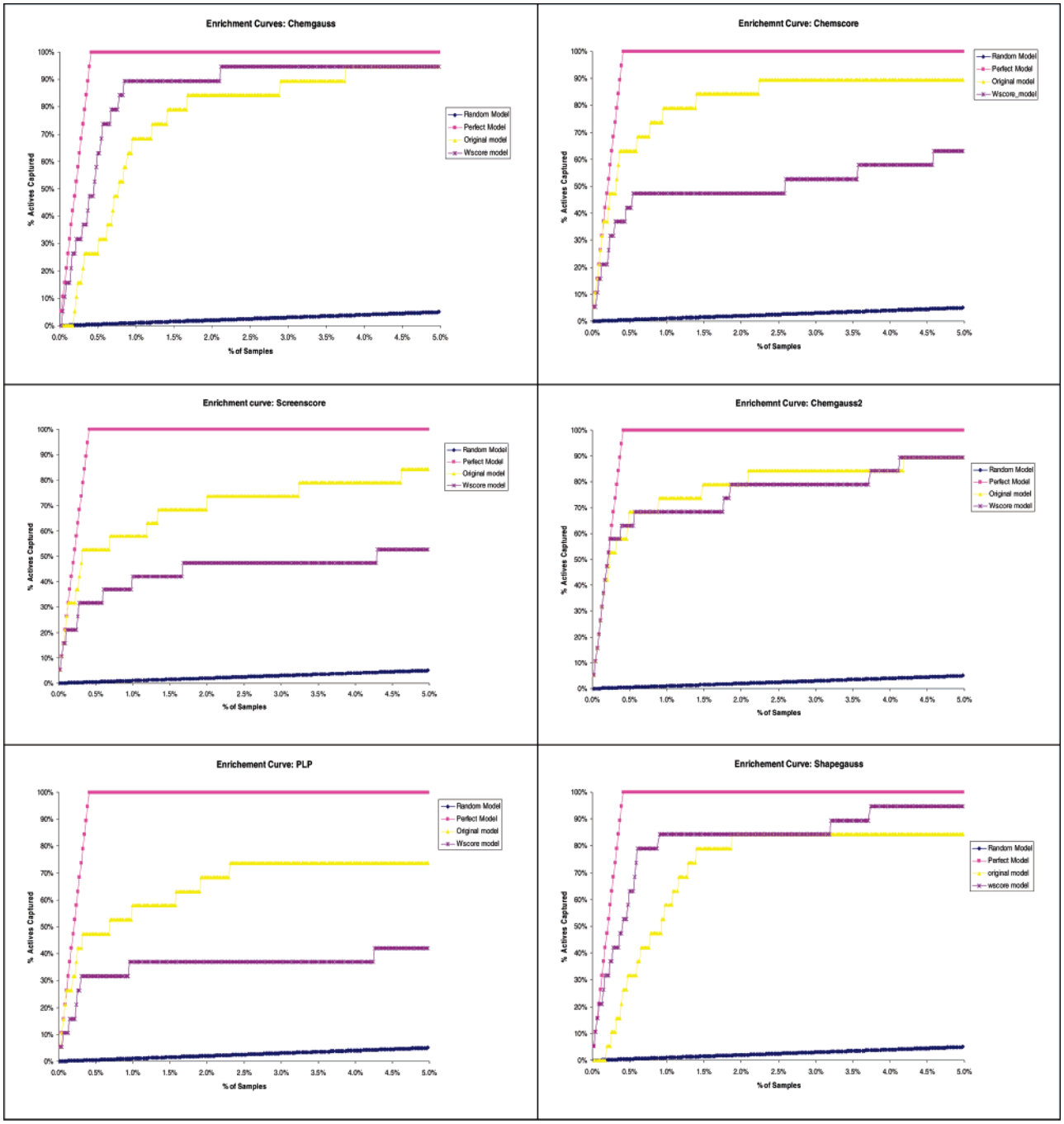

Figure 5. Enrichments curves before and after rescoring. Wscore represents the enrichment curve post rescoring. Original model represents the enrichment curve prior to rescoring.

molecules which exhibit high MW and leave the remainder relatively unchanged

$$\text{corrScore} = \text{normScore}\,(1 - \text{normMW})^{0.25}$$

where normScore is the normalized score and normMW is the normalized MW according to the previous formula.

A visual representation of the power function is depicted in the Figure 3.

ANALYSIS OF THE RESULTS

Enrichment. The potential of the scoring functions in the identification of a set of known active ligands from a set of druglike "decoys" is typically measured using the metric enrichment (E). Enrichment is a measure of the proportion of hits retrieved in a subset of compounds compared with the proportion of hits expected from a random sample of compounds

$$E = \frac{\text{Hits}_{\text{sampled}}/N_{\text{sampled}}}{\text{Hits}_{\text{total}}/N_{\text{total}}}$$

Enrichments were calculated over all the scoring functions, before and after rescoring procedure.

Scoring Frequencies. Scoring function results can be analyzed by plotting the score frequencies. In other words, how many molecules (%) receive a certain score. Scoring frequencies were normalized for ease of plotting, and results prior to and post rescoring were analyzed by this method.

SIFt Fingerprints. Structural interaction fingerprints (SIFt)[21] are used in this study to capture the interactions occurring between each docked pose of a compound and the residues of the active site of ERα. VS results for a set of protein-kinase inhibitors have previously been shown to produce superior enrichment using SIFt compared with traditional scoring functions.[14] It has also been used to design target-focused libraries because of its inherent sensitivity.

In this light, we have employed SIFt, available in MOE.2005.06, to calculate the 1D binary strings. The Tanimoto coefficient[22] was used to discern the similarity between the string generated for the PDB entry 3ERT and the ranked poses. The top scoring 200 molecules for each scoring function were assessed for similarity to the crystal structure with an overlap of 65, 70, 75, 80, and 85%.

RESULTS

A large number of descriptors correlated well with MW according to Spearmen's coefficient ($>0.60 \times 100$), and for this reason, we selected MW as the descriptor to apply our power function. It is important to note however that additional descriptors such as Oprea's number of rigid bonds (opr_brigid), number of rings (opr_nrings), and number of rotatable bonds (nrot) did not correlate well with MW, giving r^2 values of 0.59, 0.52, 0.49 $\times$ 100. Thus, limiting of the score normalization to solely MW values with a view to the removal of ligand bias may not be wholly inclusive to effect maximal impact on performance. However, given the clear correlation with MW in this study for the ER, for the purposes of this article we apply our methodology to only MW for illustrative purposes. The concept/methodology can however be expanded and applied to any descriptor with an associated score or rank in a hitlist.

The graphics depicted in Figure 4 show the effect of the rescoring procedure on the molecular-weight distribution plotted versus the rank position for the Shapegauss scored set.

Molecules with MW above 500 are highlighted, and moving from the first plot to the second, we can observe these molecules are more evenly distributed throughout the ranked set. This will serve to concentrate the scoring by each function on actual interactions rather than because of MW.

The enrichments curves retrieved for the scoring functions as implemented in FRED2.1.1 before and after rescoring, are showed in Figure 5. Shapegauss and Chemgauss rescored results (Wscore) show higher enrichment values when compared with the original scoring functions. Chemgauss2 exhibits neither improvement nor a reduction of enrichment. PLP, Chemscore, and Screenscore original score enrichments outperform the rescored enrichment results.

To further investigate the increase in enrichment observed for Shapegauss and Chemgauss, but reduction for PLP, Chemscore, Screenscore, MW versus the ranked position was plotted for all scoring functions. The upper and lower MW for the active set lies between ~340−480, and from Figure 6, it is clear that prior to the rescoring procedure, both Shapegauss and Chemgauss retrieve molecules with a MW outside this range. After implementation of our rescoring method, the molecules retrieved by Shapegauss and Chemgauss are now in the same MW range as the active set, thus producing a higher enrichment. It is also shown that

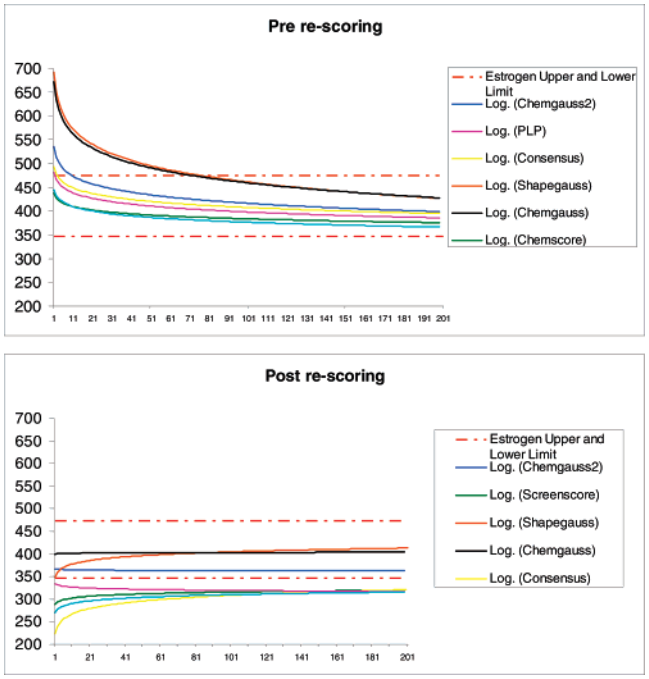

Figure 6. Molecular weight versus the ranked position of each molecule plotted for all scoring functions. Broken-line indicates the upper and lower MW for the set of ER actives.

Chemgauss2 remains unaffected, correlating well with lack of change observed in enrichment.

However, there is an appreciable shift toward the retrieval of molecules that have a lower MW than the active set. This would obviously result in an enrichment reduction for those particular scoring functions (Chemscore, PLP, and Screenscore).

To examine the full effect of our re-scoring procedure on both the active and decoy sets, it is necessary to look at the score distribution of each molecule for all the scoring functions (Figure 7).

The graphs in Figure 7 represent the score distributions for Chemgauss, Chemscore, and Chemgauss2. For conciseness, the results for just three scoring functions are depicted, being expressive of three the different scenarios obtained from the enrichment results: (1) rescoring outperforms original score (Chemgauss), (2) the original score outperforms the rescoring method (Chemscore), and (3) no appreciable change (Chemgauss2)

In the case of Chemgauss, a shift for the whole ranked set can be seen, with greater separation of the actives from decoys also. The discrimination between actives and decoys appears to decrease significantly for Chemscore, with the distribution of the estrogen remaining almost unchanged, and the decoy set shifting higher. In the case of Chemgauss2, no significant difference is observed in the scoring distribution of the dataset.

Finally, to critically analyze and classify the type of compounds being retrieved from each screen, we introduced the SIFt analysis to the procedure. The goal was to establish if compounds that exhibit the best interactions within the binding site of ERα are shifted higher in the resultant hitlist post rescoring. The *x*-axis indicates the allowed overlap and in the *y*-axis the numbers of molecules with that percentage overlap. Thus, the higher the number of molecules retrieved for a given overlap translates into closer protein−ligand

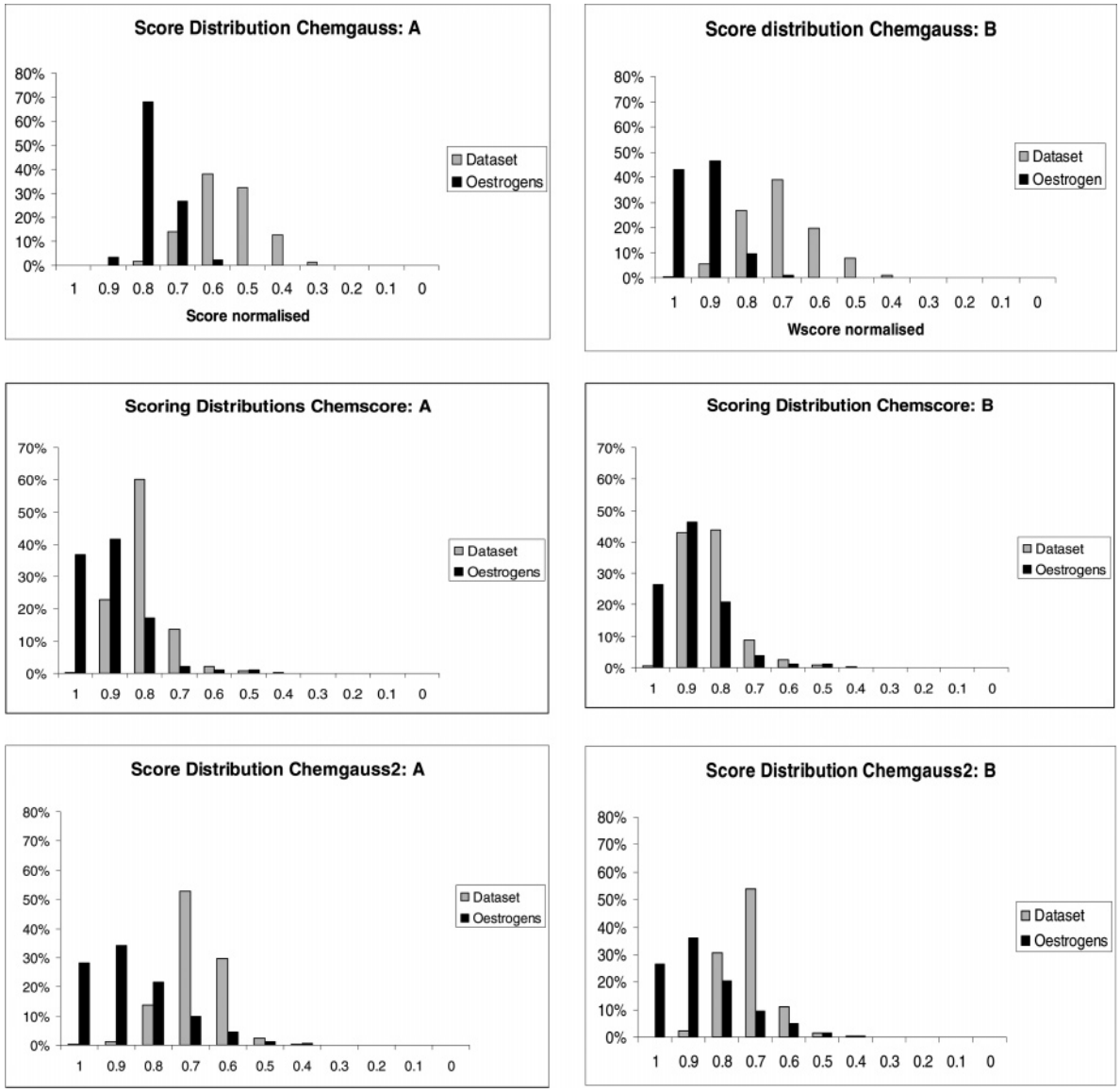

Figure 7. Score distributions for Chemgauss, Chemgauss2, and Chemscore. Set A: Before application of power function. Set B: After application of power function.

interactions to the ones observed in the crystal structure now being made for the top molecules in the ranked dataset.

The graphs (Figure 8) illustrate that after the rescoring procedure the molecules recovered with an overlap of 65−70% increase for all the scoring functions. For the scoring functions, Shapegauss, Chemgauss, and Chemgauss2, better interactions are observed for all the higher ranked molecules with better overlapping at 70−85% also.

As we have seen, after rescoring there are improvements in the abilities of Chemgauss and Shapegauss to discriminate between actives and decoys in the database used. To further illustrate the value of unbiasing the scoring functions with respect to MW, we formed a consensus score consisting of adding the rank by rank values from the original positions of PLP, Chemscore, Screenscore, and Chemgauss2 and the rescored ranks of Chemgauss and Shapegauss.

The new consensus enrichment (Table 1) obtained higher enrichment value in the first 0.20% of the database, while the values in the rest of the database remained the same. The benefit of this procedure is evident considering the formula can be tuned accordingly by altering the exponential value to penalize more heavily compounds with high or low MW. We have not attempted to tune the function to its maximum potential because the purpose of the study was not to achieve high enrichment values but to prevent high MW compounds from ranking highly in a database screen.

DISCUSSION

In this work, we propose a procedure to unbias scoring functions with regard to MW. It has been noted previously by Jacobsson et al.[23] that ligand bias is a source of false positives in structure-based virtual screening. Importantly, scoring functions are generally correlated with molecular size because of their inherently additive nature. Pan et al.[9] proposed size normalization by dividing by the square or cubic root of the heavy atom count (HAC) to negate the effect of this dependence.

Plotting the ranked position of compounds versus their associated descriptors is a fast way to inspect the characteristics of the compounds which scoring functions rank better than others. We found the Spearmen coefficient can be readily employed to explore possible relations between

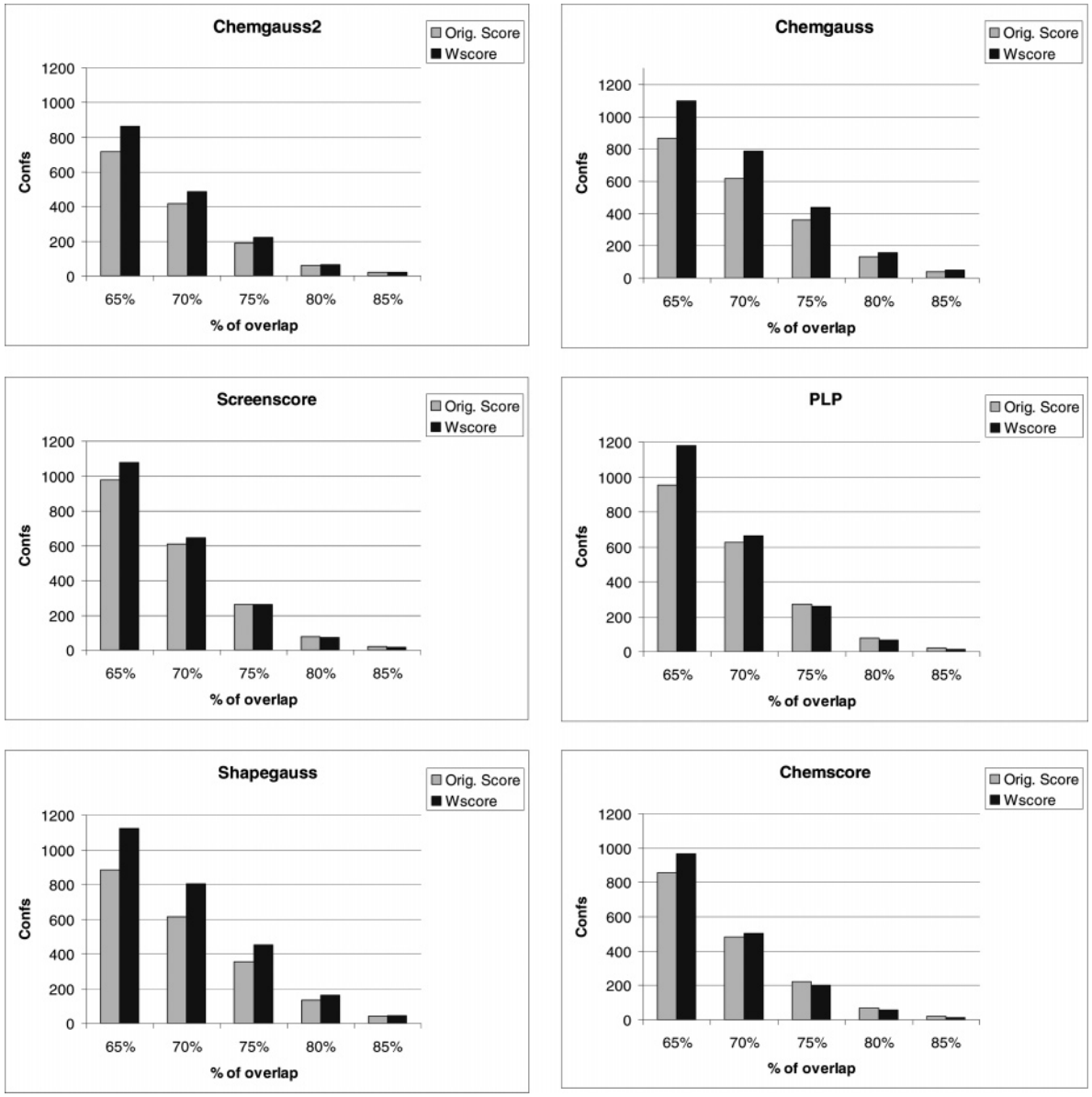

Figure 8. SIFt analysis results for all the scoring functions. The *x*-axis indicates the percent overlap of the first 200 molecules prior to and after application of our power function.

Table 1. Enrichment Calculated for 0.2, 0.5, 1, 1.5, 2, 3, and 5% of Conformer Database[a]

	0.20%	0.50%	1%	1.50%	2%	3%	5%
ligand no.	9	23	47	70	94	142	236
enrichment (pre-rescoring)	193	140	79	57	42	28	18
enrichment (post-rescoring)	220	140	79	57	42	28	18
theoretical maximum	248	204	100	67	50	33	20

[a] Consensus enrichment obtained with the original consensus scoring function versus that obtained after rescoring.

descriptors values and ranking position. Using this knowledge, we have introduced a function that can be tailored or tuned to lightly or heavily impact this MW bias. This formula penalizes exponentially higher molecular weight molecules and leaves other molecules untouched. Of great importance in this respect is that it is calculated post-scoring and thus can be reevaluated easily. The ranked set can also be easily heavily biased toward the retrieval of more leadlike MW compounds without loss in their ability to interact correctly with residues of the active site by increasing the value of the exponential in our function.

It is important to note that not all descriptors correlated well with MW, and application of our function to these descriptors would probably enhance the prioritization of actives. In fact, typical molecules containing heavy elements such as Cl or S may be treated unfavorably by this strategy, and so we suggest application of our strategy to a broader range of descriptors. We suggest its use in a consensus weighting scheme to remove ligand bias overall.

Although improvements in enrichments were observed for Chemgauss and Shapegauss after the application of our formula, a reduction was also seen for the remaining scoring functions with the exception of Chemgauss2. However, SIFt analysis showed that, although there is a fall in enrichment for some of the scoring functions, there is conversely an improvement in the quality of possible false positives in the ranked dataset.

A better understanding of the full impact of applying our formula can be derived from the score distributions because

it brings more information about how scoring functions rank all the compounds in the database and not just the positive control. Important feedback results from analysis of the scoring frequencies. For example, for Chemscore, it was clear that it outperformed most of the scoring functions in discriminating between actives and decoys. The rescoring procedure clearly decreases the enrichment value; however, analysis of the score distribution shows that on average the score of the positive controls remain more or less unchanged. Thus, the rescoring procedure brings up more molecules from the decoy set among the higher rankings. It is possible that some of these molecules that are now ranked highly with equivalent scores to those of the antiestrogens are actually actives because the database chosen was one compiled from commercially available compounds.

Finally, we demonstrate the benefit of adjusting the bias of individual scoring functions for incorporation in a consensus scheme. Using the ranks obtained for rescored Shapegauss and Chemgauss in conjunction with the ranks obtained for the original versions of Chemgauss2, PLP, Screenscore, and Chemscore, we generated higher enrichment values.

Our findings, though centered on examining the phenomenon in relation to a single receptor, are derived from an examination of the additive nature of the scoring functions selected. As such, we anticipate similar improvements can be attained in other target classes: indeed, we advocate similar target-specific examination of scoring bias during method development to maximize the performance of virtual screening.

These methods can be readily applied to any descriptors (QSAR, etc.) to tune the scoring functions toward retrieval of molecules with defined properties and unbias particular parameters within a scoring function if the user has prior knowledge of the type of ligands needed to be selected.

ACKNOWLEDGMENT

We acknowledge the funding support from Science Foundation Ireland (SFI). We thank the software vendors for their continuing support of our academic research efforts, in particular, the contributions of Openeye Scientific Software, Chemical Computing Group (CCG), and Accelrys/Scitegic (Pipeline Pilot).

REFERENCES AND NOTES

(1) Marsden, P. M.; Puvanendrampillai, D.; Mitchell, J. B.; Glen, R. C. Predicting protein−ligand binding affinities: A low scoring game? *Org. Biomol. Chem.* **2004**, *2*, 3267−73.
(2) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912−31.
(3) Bohm, 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*, 243−56.
(4) *FRED (Fast Rigid Exhaustive Docking)*, version 2.11; Openeye Scientific Software: Santa Fe, NM, 2006.
(5) 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*, 425−45.
(6) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−42.
(7) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76−90.
(8) 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*, 731−51.
(9) Pan, Y.; Huang, N.; Cho, S.; MacKerell, A. D., Jr. Consideration of molecular weight during compound selection in virtual target-based database screening. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 267−72.
(10) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287−303.
(11) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., 3rd. Comparative study of several algorithms for flexible ligand docking. *J. Comput.-Aided. Mol. Des.* **2003**, *17*, 755−63.
(12) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−67.
(13) Xing, L.; Hodgkin, E.; Liu, Q.; Sedlock, D. Evaluation and application of multiple scoring functions for a virtual screening experiment. *J. Comput.-Aided. Mol. Des.* **2004**, *18*, 333−44.
(14) Singh, J.; Deng, Z.; Narale, G.; Chuaqui, C. Structural interaction fingerprints: A new approach to organizing, mining, analyzing, and designing protein−small molecule complexes. *Chem. Biol. Drug. Des.* **2006**, *67*, 5−12.
(15) Hassan, M.; Brown, R. D.; Varma-O'Brien, S.; Rogers, D. Cheminformatics analysis and learning in a data pipelining environment. *Mol. Diversity* **2006**, *10*, 283−99.
(16) *OMEGA*, version 1.8.1; Openeye Scientific Software: Santa Fe, NM, 2006.
(17) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998**, *95*, 927−37.
(18) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−42.
(19) *MOE*, version 2006.08; Chemical Computing Group: Montreal, Quebec, Canada, 2006.
(20) Knox, A. J.; Meegan, M. J.; Carta, G.; Lloyd, D. G. Considerations in compound database preparation−"Hidden" impact on virtual screening results. *J. Chem. Inf. Model.* **2005**, *45*, 1908−19.
(21) Deng, Z.; Chuaqui, C.; Singh, J. Structural interaction fingerprint (SIFt): A novel method for analyzing three-dimensional protein−ligand binding interactions. *J. Med. Chem.* **2004**, *47*, 337−44.
(22) Tanimoto, T. T. IBM Internal Report. Nov. 17, 1957.
(23) Jacobsson, M.; Karlen, A. Ligand bias of scoring functions in structure-based virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 1334−43.

CI600471M