

Consideration of Molecular Weight during Compound Selection in Virtual Target-Based Database Screening

Yongping Pan, Niu Huang, Sam Cho, and Alexander D. MacKerell, Jr.*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201

Received September 10, 2002

Virtual database screening allows for millions of chemical compounds to be computationally selected based on structural complementarity to known inhibitors or to a target binding site on a biological macromolecule. Compound selection in virtual database screening when targeting a biological macromolecule is typically based on the interaction energy between the chemical compound and the target macromolecule. In the present study it is shown that this approach is biased toward the selection of high molecular weight compounds due to the contribution of the compound size to the energy score. To account for molecular weight during energy based screening, we propose normalization strategies based on the total number of heavy atoms in the chemical compounds being screened. This approach is computationally efficient and produces molecular weight distributions of selected compounds that can be selected to be (1) lower than that of the original database used in the virtual screening, which may be desirable for selection of leadlike compounds or (2) similar to that of the original database, which may be desirable for the selection of drug-like compounds. By eliminating the bias in target-based database screening toward higher molecular weight compounds it is anticipated that the proposed procedure will enhance the success rate of computer-aided drug design.

INTRODUCTION

Computer-based virtual database screening for lead compounds has become one of the integral approaches in the structure-based drug discovery process.¹ Database searching or docking approaches that select small molecular weight compounds that are structurally or energetically complementary to a putative binding site on a biological macromolecule have been reasonably successful in the past few years.^{2–6} However, the hit rate, or percentage of biologically active compounds out of those selected via the screening procedure, is typically 20% or less indicating the need for further improvements in the selection procedure.^{6–8}

A key element in determining the hit rate of virtual screening is the energy function used for prediction of the binding orientation and energy.^{7,9,10} Currently, compounds are selected based on a variety of energetic scoring methods. For example, Chen et al.⁸ and Enyedy et al.¹¹ identified lead inhibitors targeting HIV-1 integrase (IN) and Bcl-2, respectively, based on the empirical energy function implemented in DOCK.¹² A similar energy function with preferences for hydrogen bonds is implemented in FlexX.¹³ Other energy based scoring methods include potentials of mean force (PMF),^{14–16} free energy grids¹⁷ and many others.^{2,9} An extension of these approaches is the consensus-scoring method,^{18,19} that combines two or more scoring functions. This method was suggested to be more robust than using the individual scoring functions, but the hit rate was comparable due to the inherent limitation of the energy functions.²⁰

One important issue that has not been adequately dealt with to date in most docking approaches, and that may adversely influence the hit rate, is the molecular weight (MW) of the selected molecules, though several studies have mentioned its impact on compound selection.^{14,21} It is accounted for partially in the energy function implemented in FlexX¹³ by including the number of rotatable bonds in the scoring criteria, which was intended to account for the entropy penalty for constraining the rotatable bonds upon binding. While such a correction is appropriate it does not account for the fact that the overall van der Waals (VDW) interaction energy, a term suggested to be important for ligand affinity,²² is the sum over all pairs of ligand and target protein atoms within a specified cutoff distance. Therefore, energetic selection of compounds that includes a VDW attractive contribution, or similar term, will favor larger MW compounds since they have a larger number of atoms interacting with the target molecule. Thus, there will be a tendency for the selection of larger molecules even though they may not necessarily be as structurally complementary to the target binding site as smaller compounds. Adjustment of the energy score to account for the size of the molecule may correct for this problem. In the present manuscript, a normalization method to adjust the energy score based on the total number of heavy atoms in a molecule is presented. Based on this approach the tendency of energy based scoring methods to bias toward the selection of higher molecular weight compounds is controlled.

Previous work⁸ in our laboratory has targeted the Y3-binding site on HIV-1 integrase²³ via database screening of the National Cancer Institute (NCI) nonproprietary database.²⁴ Here we chose the same target and two other databases, the Derwent World Drug Index (WDI)²⁵ and the

* Corresponding author phone: (410)706-7442; fax: (410)706-0346; e-mail: amackere@rx.umaryland.edu. Corresponding author address: 20 N. Pine Street, Baltimore, MD 21201.

CHEMDIV²⁶ database, to test the influence of MW on the compounds selected using the DOCK energy scoring function¹² directly and following normalization. It is shown that based on the normalization strategy applied, compound selection could be biased to (1) a distribution that had MWs reminiscent of those associated with leadlike compounds²⁷ or (2) a distribution of MWs that significantly overlapped with the entire database MW distribution.

METHODS

Two chemical databases, WDI and CHEMDIV with approximately 50 000 and 200 000 compounds, respectively, were used for docking to the Y3 binding site of HIV-1 integrase. Three-dimensional (3D) structures of the compounds in the databases were generated as follows: the 2-dimensional (2D) structural data files (SDF) from the vendors were first converted to 2D MOL2 format using the program SDF2MOL2 implemented in DOCK 4.0.1.¹² Hydrogens and Gasteiger charges were then calculated and added to the 2D MOL2 structures followed by 500 steps of minimization with the POWELL conjugate gradient algorithm using the SYBYL6.7²⁸ molecular modeling package. The protonation state of the ligands was taken directly from the suppliers.

DOCK 4.0.1¹² was used to carry out the database screening. Modeling of the binding site and selection of the sphere set were described previously.⁸ The anchor-search-first algorithm was used to initially place the ligands and their conformational space was searched via the standard torsion drive. While the added segment was minimized at every intermediate stage, re-minimize-layer-numbers of 3 and 5 were used for the large database screening (method one) and for the second, more rigorous search (method two), respectively. Anchor re-minimization during the conformational search was also used in method two. A dielectric function of 4r and a cutoff of 10.0 Å were used for generating the steric and electrostatic environment at the binding site with the GRID²⁹ module of DOCK. In all cases the docking and resulting binding orientations were based on the total energy, such that selection of compounds via the VDW attractive energy used the docked orientations based on the total energy. Heavy atom cutoff numbers of 40 and 60 were used to compare its influence on compound selection. Solvent accessible surface areas³⁰ of the small molecules were calculated with CHARMM³¹ program and the MMFF³² force field implemented in CHARMM using a 1.4 Å probe radius.

RESULTS AND DISCUSSIONS

In the following, results are presented showing how database screening using the program DOCK leads to the selection of compounds with a higher MW distribution than those in the original database. A normalization procedure is then proposed and tested that corrects for this bias. Two databases have been selected for the study. The WDI was selected as it contains a collection of pharmacologically active, or druglike, compounds, including therapeutic agents currently on the market. Since it may be considered undesirable to shift the MW distribution of selected compounds above that of known pharmacologically active compounds, inclusion of a database of biologically active compounds is important. Alternatively, it is necessary to check if the

Table 1. Mean Molecular Weights and Standard Errors for the WDI and CHEMDIV Databases and for the 5000 Compounds Selected Using Different Scoring and Normalization Procedures^a

| normalization | vdw attractive | | total energy | |
|------------------------------------|----------------|----------|--------------|----------|
| | 40 | 60 | 40 | 60 |
| WDI Full Database Mean 359 ± 8 | | | | |
| none | 369 ± 12 | 382 ± 12 | 389 ± 15 | 403 ± 16 |
| N | 258 ± 24 | 258 ± 24 | 284 ± 7 | 284 ± 7 |
| N ^{2/3} | 289 ± 6 | 290 ± 6 | 320 ± 7 | 322 ± 7 |
| N ^{1/2} | 310 ± 5 | 313 ± 5 | 339 ± 10 | 343 ± 9 |
| N ^{1/3} | 332 ± 10 | 338 ± 9 | 359 ± 11 | 365 ± 12 |
| normalization | vdw attractive | | total energy | |
| | 40 | 60 | 40 | 60 |
| CHEMDIV Full Database Mean 385 ± 7 | | | | |
| none | 431 ± 11 | 440 ± 13 | 442 ± 14 | 453 ± 19 |
| N | 305 ± 13 | 305 ± 12 | 309 ± 13 | 309 ± 13 |
| N ^{2/3} | 346 ± 8 | 346 ± 7 | 350 ± 10 | 350 ± 10 |
| N ^{1/2} | 369 ± 2 | 370 ± 2 | 375 ± 5 | 377 ± 4 |
| N ^{1/3} | 393 ± 3 | 395 ± 3 | 400 ± 5 | 404 ± 6 |

^a Mean and standard error values were calculated by separating each data set into 5 independent data sets (e.g. a set of 5000 compounds MWs were separated into 5 1000 compound MW sets), the mean obtained for each individual set and then obtaining the mean and the standard error over the 5 individual means.³⁶

problem and resulting solution are also applicable to a chemical database typically used for database searching. This motivated the selection of CHEMDIV, a diverse collection of compounds designed for database searching.

Molecular Weight Distributions from Database Screening. The MW distributions of the 5000 selected compounds based on both the VDW attractive and total energy scores are presented in Figure 1A,B, for the WDI and CHEMDIV databases, respectively. A significant observation is that all the distributions of the docked compounds are shifted toward the high MW region relative to the full database distributions in both cases. These results are quantitated in Table 1, where it can be seen that the mean MW of the 5000 selected compounds (normalization: none) are larger than that of the original database in all cases. For example, with the WDI the mean MW of the entire database is 359, while that from the scoring based on the VDW attractive term with a 40 non-hydrogen atom cutoff is 369 and increases to 382 with a 60 atom cutoff. Similar increases in the mean MW are seen with the total energy scoring as well as with the CHEMDIV database.

The results based on the VDW attractive or total energy scores are similar since total scores for most molecules were dominated by the VDW attractive energy (see Supporting Information Figures S1 and S2). These results show that there is a bias toward the selection of high MW compounds when doing target-based database screening, leading to MW distributions skewed to values larger than those typically seen for pharmacologically active compounds as judged by the WDI. As mentioned above, similar upshifting of the MW of selected compounds has also been reported in database screening work based on PMF scoring.¹⁴ Clearly, a normalization procedure to correct for the bias toward higher MW compounds would be desirable, as lower MW compounds typically have improved absorption and disposition properties.³³ In addition, when considering the use of database screening in drug design, it may be considered

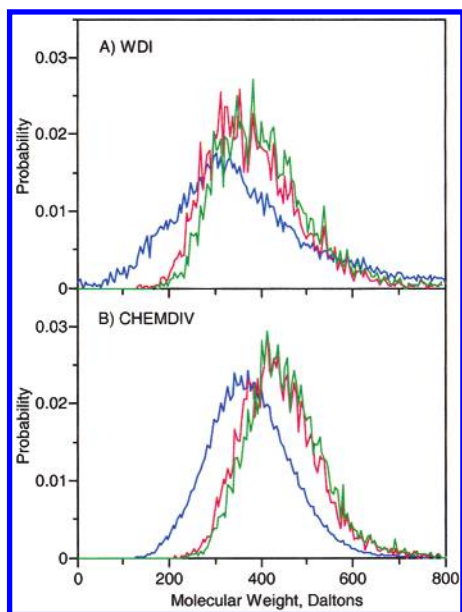


Figure 1. Molecular weight distribution for the top 5000 molecules selected from the (A) WDI and (B) CHEMDIV databases based on different scoring: VDW attractive energy and cutoff 60 (red) and total energy and cutoff 60 (green). For comparison, the original database distributions are included (blue).

desirable to select compounds with lower MW than those seen in the WDI as they may be more appropriate for lead compound structural optimization.³⁴

Score Normalization Based on the Number of Non-Hydrogen Atoms. Empirically, the bias toward the selection of higher MW compounds in database screening is not surprising given that additional favorable interactions between a ligand and a receptor are available to larger compounds. This is especially true for the VDW attractive term, which is favorable for all atom pairs and leads to consideration of a normalization procedure based on the number of non-hydrogen atoms, N , in each ligand. This is similar to the use of scoring on a per heavy atom basis as used in the SmoG scoring function.³⁵ Accordingly, the VDW attractive or total energy for each compound was divided by the number of non-hydrogen atoms in each respective compound and the top 5000 compounds selected based on those normalized scores. Presented in Figure 2 are the resulting MW distributions for the two databases and for the different scoring criteria. As is evident, score normalization based on N leads to a significant downshift in the MW of the selected compounds as compared to the total databases. Consistent with this shift are the mean MWs reported in Table 1 for N normalization. Thus, use of the number of non-hydrogen atoms for score normalization leads to a significant overcorrection, leading to a strong bias toward low MW compounds. However, such an overshift may be considered desirable in database screening efforts where the goal is the identification of leadlike versus druglike compounds.³⁴

Alternatively, a normalization procedure may be considered that is based on the surface area of a molecule. In this model, the increase in the surface area of a molecule as the number of atoms increases would allow for more favorable interactions, thereby biasing toward high MW compounds during screening. However, determination of the solvent accessible surface area of a molecule is computationally

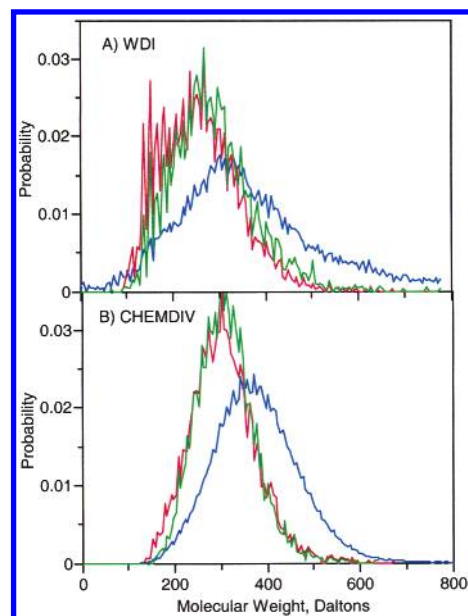


Figure 2. Molecular weight distribution for the top 5000 molecules selected using the N normalization scoring method from the (A) WDI and (B) CHEMDIV databases based on different scoring: normalized VDW attractive energy and cutoff 60 (red) and normalized total energy and cutoff 60 (green). For comparison, the original database distributions are included (blue).

expensive, especially in situations where it must be determined for a million or more molecules, a situation common in virtual database screening. Instead, one may consider that N is directly proportional to the volume of a molecule and, assuming a simple geometry, is also proportional to the radius cubed, r^3 . Based on r being proportional to $N^{1/3}$ and the surface area being proportional to r^2 , the surface area is proportional to $N^{2/3}$. To test this model, the solvent accessible surface area of 4523 randomly selected molecules from the CHEMDIV database was calculated and is plotted against $N^{2/3}$ in Figure 3. As expected a linear relationship is present ($R = 0.91$). Accordingly, the VDW attractive and total energy scores were renormalized using $N^{2/3}$ and the resulting MW distributions presented in Figure 4. For both WDI and CHEMDIV, there is still a downshifting of the MW distribution below those in the entire databases, though the effect is smaller than that observed with N normalization (Figure 2). The mean MWs in Table 1 show the values for the $N^{2/3}$ selected compounds to be smaller than the mean values from the entire databases.

Motivated by the shift in MW distributions upon going from N to $N^{2/3}$ normalization, tests using $N^{1/2}$ normalization were undertaken. Presented in Figure 5 and Table 1 are the MW distributions and mean MWs, respectively, following rescoring with normalization by $N^{1/2}$. Both Figure 5 and Table 1 show that this normalization brings the MW distribution of selected compounds into better agreement with that of the entire databases, although a shift to small MWs is evident. Therefore, normalization via $N^{1/3}$ was tested, with the resulting MW distributions shown in Figure 6 and the mean values in Table 1. With the WDI, the selected compounds are still shifted to lower MWs, while with CHEMDIV the MW distribution of the selected compounds is in good agreement with that of the full database. The larger discrepancy with the WDI appears due to the number of high MW compounds in the entire database, which increases the mean.

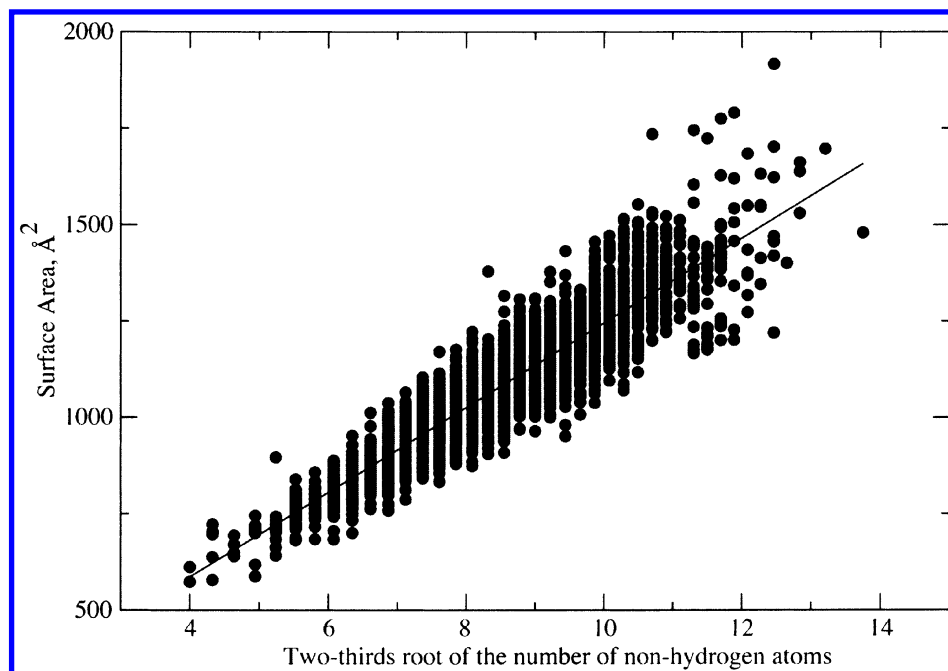


Figure 3. Two-thirds root of the number of heavy atoms, $N^{2/3}$, versus the solvent accessibility for 4523 compounds randomly selected from the ChEMDIV database. The line represents the least-squares fit to the data.

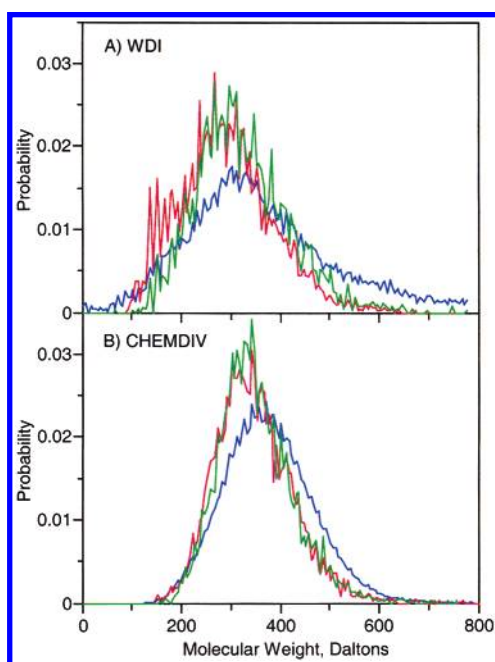


Figure 4. Molecular weight distribution for the top 5000 molecules selected using the $N^{2/3}$ normalization scoring method from the (A) WDI and (B) ChEMDIV databases based on different scoring: normalized VDW attractive energy and cutoff 60 (red) and normalized total energy and cutoff 60 (green). For comparison, the original database distributions are included (blue).

Thus, the empirically motivated use of $N^{1/3}$ for score normalization yields a MW distribution of selected compounds that is most similar to that of the original database being screened. However, the use of $N^{1/3}$ is difficult to justify based on physical considerations. It is therefore suggested that the $N^{1/3}$ normalization factor represents a collection of physical phenomena, including the surface area term suggested above. It should be noted that the relationship of surface area to $N^{1/3}$ is almost identical to that of $N^{2/3}$, with $R = 0.91$.

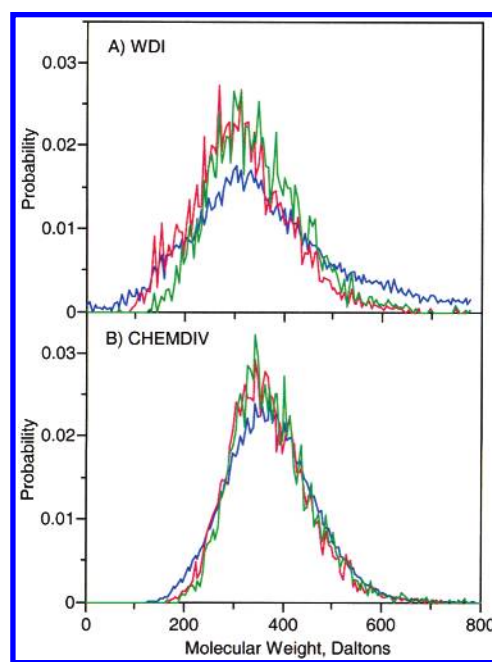


Figure 5. Molecular weight distribution for the top 5000 molecules selected using the $N^{1/2}$ normalization scoring method from the (A) WDI and (B) ChEMDIV databases based on different scoring: normalized VDW attractive energy and cutoff 60 (red) and normalized total energy and cutoff 60 (green). For comparison, the original database distributions are included (blue).

The “rule of five” used for lead identification and optimization³³ lends additional support to the advantage of the proposed normalization procedure. The decreased MW distributions may be considered desirable as compounds with MWs greater than 500 daltons will typically have lower biological membrane permeability and therefore lower bio-availability.³³ This is supported by a previous database screening study where selected compounds that were biologically active had a lower MW distribution than drugs on the market.³⁴

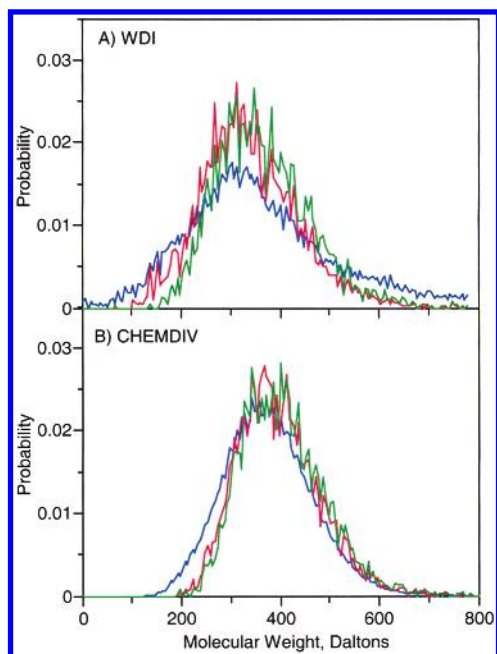


Figure 6. Molecular weight distribution for the top 5000 molecules selected using the $N^{1/3}$ normalization scoring method from the (A) WDI and (B) CHEMDIV databases based on different scoring: normalized VDW attractive energy and cutoff 60 (red) and normalized total energy and cutoff 60 (green). For comparison, the original database distributions are included (blue).

To further illustrate the influence of $N^{1/3}$ normalization on the MW distribution of the selected compounds, the 5000 compounds selected from the WDI database using the method one screening protocol based on VDW attractive energy were further screened using method two with 500 compounds selected based on the total energy score. Without normalization after method one and method two screening, 121 out of 500 compounds had MWs greater than 500 daltons, compared with 38 out of 500 compounds with $N^{1/3}$ normalization used in both steps. The mean molecular weights of the distributions were 447 and 367 daltons for un-normalized and normalized cases, respectively, versus 359 (Table 1) for the entire WDI database. Consistent with the results presented above, the more rigorous method two docking protocol was again biased toward the higher MW region without normalization.

Influence of Normalization on Ligand-Target Complementarity. A final question was whether the proposed normalization procedure enhances the complementarity between the ligand and the target binding site. This was addressed by analyzing the ratio of the buried surface areas of the ligands in the bound state relative to that of the unbound ligand for the unnormalized and $N^{1/2}$ normalized scores for the WDI method one screen using VDW attractive energy and 60 atom cutoff. As is shown in Figure 7, the normalization procedure shifts the distribution toward compounds with a higher ratio of buried surface area. This indicates that compounds selected via the normalized score show a tendency to have a higher percentage of their surface area buried in the bound state versus the unbound state than those selected via the un-normalized scores. While certainly not an unequivocal measure of complementarity, these results suggest that the normalization procedure may favor the selection of compounds with more contacts per unit surface area with the target binding site when MW is taken into

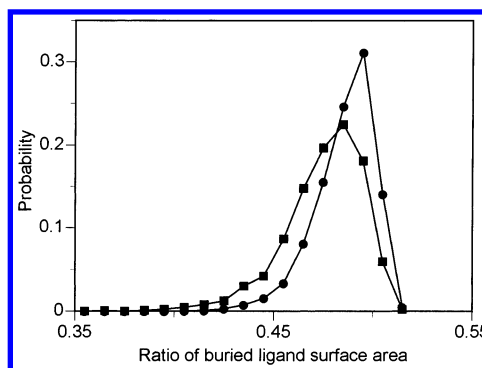


Figure 7. Probability distributions of the ratio of the buried surface area upon going from the unbound to bound states for the original (squares) and normalized (circles) energy scores. Results are based on the change in the surfaces areas for the bound conformations of the ligands to IN from the WDI VDW energy database screen.

account. The higher ratio of contacts to surface area may facilitate the selection of compounds with a higher probability of being specific for the target molecule.

SUMMARY

The most important issue in virtual database screening for lead compounds is the ability of the protocol used to generate a satisfying hit rate. In this work it is shown that use of interaction energy scoring alone for choosing molecules favors the selection of high MW compounds. However, by normalizing the energy score based on the number of heavy atoms, N , in the ligands the bias toward higher MW compounds can be eliminated. Moreover, by selection of the normalization procedure based on the power of N , a target range for the MW distribution can be selected. For example, if the goal of a study is the identification of druglike compounds,³⁴ normalization via $N^{1/3}$ would be appropriate as it yields a distribution similar to that of the entire database, whereas if leadlike molecules are the goal, normalization via N is appropriate as it yields a MW distribution for the selected compounds with mean values of 300 daltons or less (Table 1). It is anticipated that this approach may enhance the hit-rate of database screening efforts.

In the present study we have applied our normalization to only a single binding site on HIV IN using the program DOCK and its energy scoring function.²⁹ However, in work by Muegge et al.,¹⁴ a similar upshift of the MW distribution was observed using DOCK with a PMF based scoring function.¹⁵ Thus, the proposed approach is anticipated to be applicable to a variety of target-based database screening studies, although it is suggested that N to various powers be tested to identify the value that yields the desired MV distribution appropriate for each particular project.

ACKNOWLEDGMENT

The authors thank Dr. Holger Gohlke for helpful discussions and financial support from the NIH (GM CA95200), and the Greenebaum Cancer Center, the Department of Pharmaceutical Sciences, School of Pharmacy, Baltimore, and the Computer Aided Drug Design Center, School of Pharmacy, University of Maryland, Baltimore.

Supporting Information Available: Figures of energy component distribution for the 5000 compounds selected from the CHEMDIV and WDI databases. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- (1) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening-an overview. *Drug Discovery Today* **1998**, 3, 160–178.
- (2) Gane, P. J.; Dean, P. M. Recent advances in structure-based rational drug design. *Curr. Opin. Struct. Biol.* **2000**, 10, 401–404.
- (3) Burkhard, P.; Hommel, U.; Sanner, M.; Walkinshaw, M. D. The discovery of steroids and other novel FKBP inhibitors using a molecular docking program. *J. Mol. Biol.* **1999**, 287, 853–858.
- (4) Godden, J. W.; Stahura, F.; Bajorath, J. Evaluation of docking strategies for virtual screening of compound databases: camp-dependent serine/threonine kinase as an example. *J. Mol. Graphics Mod.* **1999**, 16, 139–143.
- (5) Filikov, A. V.; Monan, V.; Vickers, T. A.; Griffey, R. H.; Cook, P. D.; Abagyan, R. A.; James, T. L. Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J. Comput. Aided Mol. Des.* **2000**, 14, 593–610.
- (6) Perola, E.; Xu, K.; Kollmeyer, T. M.; Kaufmann, S. H.; Prendergast, F. G.; Pang, Y.-P. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J. Med. Chem.* **2000**, 43, 401–408.
- (7) 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, 4759–4767.
- (8) Chen, I.-J.; Neamati, N.; Nicklaus, M. C.; Or, A.; Anderson, L.; Barchi, J. J., Jr.; Kelley, J. A.; Pommier, Y.; MacKerell, A. D., Jr. Identification of HIV-1 Integrase Inhibitors via Three-Dimensional Database Searching Using ASV and HIV-1 Integrases as Targets. *Bioorg., Med. Chem.* **2000**, 8, 2385–2398.
- (9) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, 44, 1035–1042.
- (10) Pearlman, D. A.; Charifson, P. S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* **2001**, 44, 3417–3423.
- (11) Enyedy, I. J.; Ling, Y.; Nacro, K.; Tomita, Y.; Wu, X.; Cao, Y.; Guo, R.; Li, B.; Zhu, X.; Huang, Y.; Long, Y.-Q.; Roller, P. P.; Yang, D.; Wang, S. Discovery of Small-Molecule Inhibitors of Bcl-2 through Structure-Based Computer Screening. *J. Med. Chem.* **2001**, 44, 4313–4324.
- (12) Makino, S.; Kuntz, I. D. Automated Flexible Ligand Docking Method and Its Application for Database Search. *J. Comput. Chem.* **1997**, 18, 1812–1825.
- (13) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, 261, 470–489.
- (14) Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. Evaluation of PMF Scoring in Docking Weak Ligands to the FK506 Binding Protein. *J. Med. Chem.* **1999**, 42, 2498–2503.
- (15) 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.
- (16) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based Scoring Function to Predict Protein–Ligand Interactions. *J. Mol. Biol.* **2000**, 295, 337–356.
- (17) Pearlman, D. A.; Charifson, P. S. Improved Scoring of Ligand-Protein Interactions Using OWFEG Free Energy Grids. *J. Med. Chem.* **2001**, 44, 502–511.
- (18) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, 42, 5100–5109.
- (19) Gohlke, H. and Klebe, G. Statistical potentials and scoring functions applied to protein–ligand binding. *Curr. Opin. Struct. Biol.* **2001**, 11, 231–235.
- (20) Wang, R.; Wang, S. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1422–1426.
- (21) Pegg, S. C.-H.; Haresco, J. J.; Kuntz, I. D. A genetic algorithm for structure-based de novo design. *J. Comput.-Aided Mol. Des.* **2001**, 15, 911–933.
- (22) Kuntz, I. D.; Chen, K.; Sharp, K.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 9997–10002.
- (23) Craigie, R. HIV Integrase, a brief overview from chemistry to therapeutics. *J. Biol. Chem.* **2001**, 276, 23213–23216.
- (24) Milne, G. W. A.; Nicklaus, M. C.; Driscoll, J. S.; Wang, S.; Zaharevitz, D. National Cancer Institute Drug Information System 3D Database. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1219–1224.
- (25) WDI: World Drug Index; Version 4/99, Derwent Information, 1999 (www.derwent.com).
- (26) See <http://www.chemdiv.com>.
- (27) Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. The Design of Leadlike Combinatorial Libraries. *Angew. Chem., Int. Ed. Engl.* **1999**, 38, 3743–3748.
- (28) SYBYL, version 6.7; Tripos Associates: St. Louis MO, 2000.
- (29) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automatic docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, 13, 505–524.
- (30) Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **1971**, 55, 379–400.
- (31) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, 2, 187–217.
- (32) Halgren, T. A. Merck molecular force field: I. Basis, form, scope, parametrization and performance of MMFF. *J. Comput. Chem.* **1996**, 17, 490–519.
- (33) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, 46, 3–26.
- (34) Oprea, T. I.; Davis, A.; Teague, S. J.; Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1380–1315.
- (35) DeWitte, R. S.; Shakhovitch, 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, 11733–11744.
- (36) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanine-N'-Methylamide. *Biopolymers* **1992**, 32, 523–535.

CI020055F