Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System†

David A. Pearlman* and Paul S. Charifson

Vertex Pharmaceuticals Incorporated, 130 Waverly Street, Cambridge, Massachusetts 02139-4242

Received January 22, 2001

Precise thermodynamic integration free energy simulations have been applied to a congeneric series of 16 inhibitors to the p38 MAP kinase protein for which the experimental binding data (IC_{50}) is known. The relative free energy of binding for each compound has been determined. For comparison, the same series of compounds have also scored using the best rapid scoring functions used in database screening. From the results of these calculations, we find (1) that precise free energy simulations allow predictions that are reliable and in good agreement with experiment; (2) that predictions of lower reliability, but still in good qualitative agreement with experiment, can be obtained using the OWFEG free energy grid method, at a much lower computational cost; (3) and that other methods, not based on free energy simulations yield results in much poorer agreement with experiment. A new predictive index, which measures the reliability of a prediction method in the context of normal use, is defined and calculated for each scoring method. Predictive indices of 0.84, 0.56, 0.04, −0.05, and 0.25 are calculated for thermodynamic integration, OWFEG, ChemScore, PLPScore, and Dock Energy Score, respectively, where +1.0 is perfect correct prediction, −1.0 is perfect incorrect prediction, and 0.0 is random.

Introduction

Database screening has become a routine component of drug discovery. To hasten the identification of a lead candidate, very large numbers of compounds are now passed through various types of rapid theoretical screens. Each screen is based on some sort of scoring function and/or acceptable property range filter. The much reduced set of compounds that survives these filters is subjected to more detailed, slower, and considerably more expensive experimental analysis.[1]

Of overriding importance in choosing a scoring function is that the function be fast. It is not unusual to pass hundreds of thousands or *millions* of compounds through the theoretical screen.[1,2] Only very rapid functions can possibly be used with these types of numbers of compounds. A variety of appropriate functions have been developed.[3] But while the best of these are able to predict the better compounds in the database with accuracy of around 70%—which makes them valuable for wholesale scoring—these functions are not accurate enough to be particularly useful in answering questions where a precise answer is required.[4] Yet, the latter types of questions nearly always arise in drug discovery after selection of candidates for further optimization has been performed using the rapid screens. Screening can bring us reasonably close to an appropriate drug candidate by suggesting appropriate scaffolds and pendent functionalities. But additional chemistry is ultimately required, which can take the form of either traditional bench synthesis or else combinatorial synthesis when an amenable scaffold is present. With rare exception, the final compounds sent to development reflect contributions from nonbatch bench synthesis.[2]

Bench synthesis is much slower and much more expensive on a compound basis than the initial screening. By the time a project has migrated to the bench, the goal is usually to reduce a micromolar range inhibitor to one that is nanomolar or better. The rapid scoring functions amenable to screening are simply not reliable enough to use in this regime. But the value of a theoretical approach at this stage is potentially tremendous. Bench synthesis often follows a simple binary search path, with each step performed to answer questions of the following form: Should we change substituent "A" to "B"? Even conceptually simple changes can be extremely difficult to make, depending on the specifics of the system under consideration. The time (and financial) savings possible by being able to *reliably* answer some of these questions using theoretical methods can be enormous.

Fortunately, a theoretical method exists that can—in principle—address these types of questions at the required level of precision. The method is free energy calculations.[5] Free energy calculations use statistical mechanical information generated during a molecular dynamics (MD) simulation to calculate the free energy difference between two molecules. Though free energy calculations are much too slow to be used directly in screening, they can provide answers to specific questions more quickly than experimental synthesis and assay and at a trivial fraction of the cost. These calculations

* Corresponding author. Phone: 617-577-6522. E-mail: dap@vpharm.com. Fax: 617-444-6680.

† The authors dedicate this work to the memory of Professor Peter A. Kollman, who passed away last spring at the much-too-young age of 56. Peter's keen intellect and relentless enthusiasm had an immense impact on both of the authors—who had the privilege to work with him—as well as on the whole field of computational chemistry. It is no exaggeration to say that without Peter the face of modern computational chemistry would be substantially different and that work such as this might not exist. Peter's contributions and joyful demeanor are sorely missed.

10.1021/jm0100279 CCC: $20.00
Published on Web 09/13/2001

also have the advantage of being applicable to a vast array of systems. This distinguishes them from even more precise pure quantum-mechanically derived approaches which can rarely be applied to such questions without wholesale approximation in terms of the system itself (i.e.. no explicit solvent, no atomic flexibility, only a cursory representation of the protein, etc.).[6−8]

While the master equations of free energy simulations are exact, reliability, precision, and accuracy issues arise from a number of sources.[5,9−11] Primary among these are sampling and force field. The free energy is related to the ensemble average of a quantity that is typically accumulated during molecular dynamics. If the sampling performed by MD is insufficient, the required ensemble average (and derived free energy) will be imprecise. Similarly, even if the sampling is sufficient, the free energy can be incorrect if the force field harbors errors or if errors in the model for the system are made. Therefore, it is not a given that *in practice* the predictions of free energy calculations will be accurate.

The literature is scattered with isolated examples where free energy simulations appear to have reproduced experimental behavior well. However, there has not been an attempt to apply free energy calculations systematically to a series of related protein−inhibitor complexes for which experimental binding data is known and has been derived in a consistent manner. Our goal here is to perform such a test. We have chosen the p38 MAP kinase protein system[12] for our test. This system was chosen for several reasons. These include the following: availability of appropriate experimental IC_{50} data for a series of compounds measured in a consistent fashion; a congeneric series of inhibitors that span roughly 2 orders of magnitude in binding efficacy; and binding behavior that is not intuitively predicatable, nor well explained by many qualitative scoring schemes.

Our goal in performing this test is to determine just how well exact free energy simulations can perform. Sixteen full thermodynamic integration free energy simulations studies have been performed for 16 inhibitors to p38. These 16 compounds represent a congeneric series chosen both to be representative of the incremental changes one considers during the ligand optimization process once a scaffold is identified and to span a significant range of IC_{50} ranging from 36 nM to 1.9 μM. For comparison and contrast, the same compounds have been scored with widely used empirical scoring functions[13−17] and with the recently described OWFEG free energy grid scoring method.[18] From these results, we can obtain guidance as to when it is appropriate to rely on the fast-but-qualitative simulations, and how well we can hope to do with the better exact calculations.

Methods

All high precision free energy calculations were carried out using the Amber/Gibbs 5.0 program.[19] The thermodynamic integration (TI) method[5] was used, wherein the free energy difference between two molecules is evaluated using the statistical mechanics equation.

$$G_B - G_A = \Delta G = \int_0^1 \left\langle \frac{dV(\lambda,\mathbf{x})}{d\lambda} \right\rangle_\lambda d\lambda \qquad (1)$$

where "A" and "B" are two molecules in a selected environment (e.g., within the protein binding site or free in solution) and G_B and G_A are their respective free energies. A parameter λ is introduced to the potential energy function such that $V(\mathbf{r}, \lambda = 0)$ is the potential energy function appropriate for molecule A, $V(\mathbf{r}, \lambda = 1)$is the potential energy function appropriate for molecule B, and $V(\mathbf{r}, \lambda)$ is continuous between A and B. The specific form has been described elsewhere.[20] "$\langle\ \rangle_\lambda$" indicates that the average of the quantity with the brackets must be determined from an ensemble that reflects the intermediate λ state potential function $V(\mathbf{r}, \lambda)$. The value of the integrand is determined at a series of discrete λ points, and the total integral is approximated from these data using trapezoidal numerical integration.

Table 1. Ligands for Which Free Energies of Binding Were Calculated

sequence number	R_1	R_2	R_3	pIC_{50}
1	H	H	H	6.602
2	H	H	F	7.000
3	H	H	CH_3	5.854
4	H	Cl	Cl	6.097
5	H	CH_3	H	5.854
6	H	CH_3	CH_3	5.721
7	H	F	H	6.347
8	CH_3	H	H	6.699
9	H	Cl	F	6.301
10	H	Cl	H	6.553
11	CH_3	H	Cl	6.745
12	Br	H	H	6.602
13	CH_3	H	CH_3	6.577
14	OH	H	H	6.444
15	NH_2	H	F	6.658
16	Cl	H	F	7.444

[a] R_1, R_2, and R_3 refer to substitution positions on an aromatic ring in the basic scaffold. Refer to Figure 1b.

TI calculations were performed for 16 bound inhibitors of p38, with IC_{50} values ranging from 36 nM to 1900 nM (1.9 μM). The reference state "A" for all free energy simulations was the molecule shown in Figure 1A. All binders "B" are all variants on the basic scaffold shown in Figure 1b, differing from the reference state only in relatively small substituents on the ring as shown and listed in Table 1. To construct the initial "B" states, each of the binders was flexibly docked[21] in the active site of the crystal conformation of the p38 protein. A spherical 16Å cap of TIP3P solvent water,[22] centered on the binder, was then added to the complex. For all simulations, a moving "belly" was defined to include all protein atoms within 12Å of any atom of the inhibitor. The water cap was restrained to the 16Å barrier using a single-sided harmonic restraint with a 0.5 kcal/mol force constant. All solvent water and all atoms of the inhibtor were free to move. All atoms of the protein within the belly could also move. Unless otherwise noted, position restraints with a positional force constant of $K = 0.5$ kcal/mol were imposed on the protein atoms to keep the protein conformation near the initial structure.

Protein and water nonbonded parameters are from the Cornell et al. force field[23] as are van der Waals parameters for the inhibitors. Charges for the inhibitors were derived from ESP-fitted ab initio charges (3-21g*/6−31g* MK charges[24]). Internal (bond, angle, torsion) parameters for the inhibitors were assigned according to the Charmm Parm22 force field.[25]

For each simulation, the protein/inhibitor/water complex was minimized for 1000 steps, and then 500 ps of MD equilibration was performed. This was followed by the "forward" TI free energy simulation ($\lambda = 0 \rightarrow \lambda = 1$; state B to reference state A). The ensemble in eq 1 was evaluated at 21 equally spaced λ points between 0 and 1 to approximate the integral. For most simulations, 10 ps of equilibration and 10 ps of data collection were performed at each integration point. For one set of simulations (noted in the text), equilibration and data collection were doubled to 20 ps each, to affirm that the results obtained were converged. At the end of the "forward" simulation, an additional 100 ps of equilibration was performed, and then the TI simulation was run in the reverse direction ($\lambda = 1 \rightarrow \lambda = 0$; state A to state B). The forward and reverse simulations are averaged to reduce the net error. All

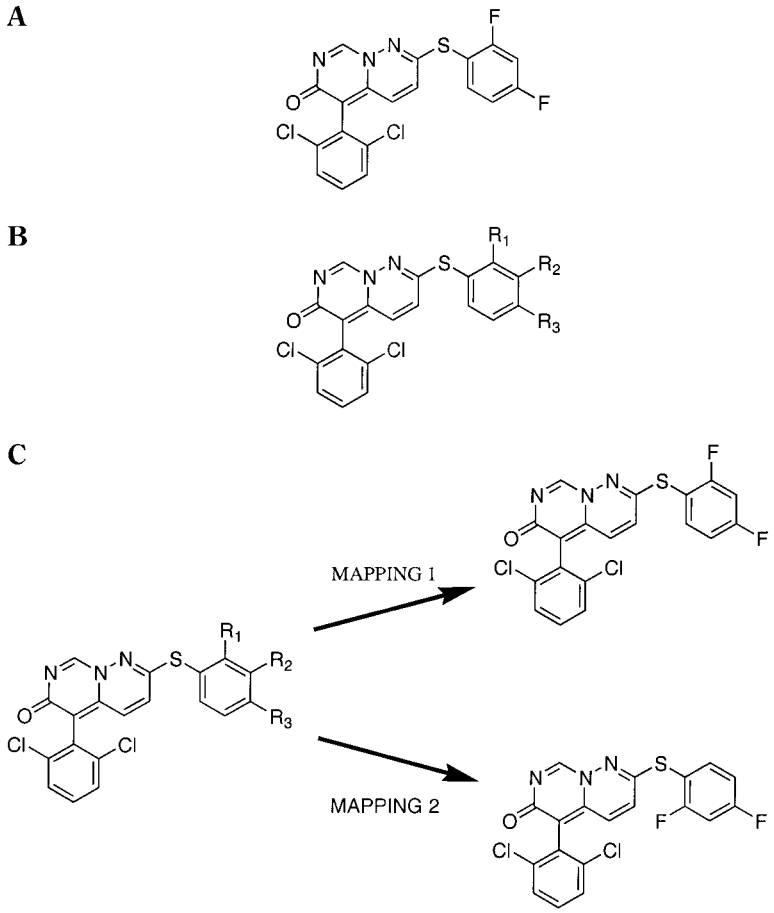
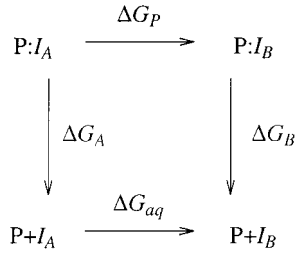


Figure 1. (A) Reference state for all the free energy simulations performed. This molecule binds to p38 with an experimental IC_{50} of 9 nM. (B) Scaffold class for all variants examined. All substitutions relative to the molecule in A are at the positions noted. (C) The "right-hand" aromatic group of all molecules of the series can be mapped onto the reference state in two ways, as shown.

MD simulations were run using a 2 fs time step and SHAKE[26] bond constraints. Simulations were run at a constant temperature of 300 K, using the Berendsen temperature coupling method. A nonbonded cutoff of 8 Å was used in all cases.

Note that it is possible to map each binder "B" onto the reference state molecule "A" in two ways, reflecting a 180° flip along the torsion connecting the right-hand aromatic ring in Figure 1b to the remainder of the molecule. This is shown in Figure 1c. In theory, it makes no difference how this mapping is performed. In practice, it *might* make a difference if the pendant ring is not able to appropriately reorient itself within the binding site as the simulation progresses. For this reason, four simulations were run for each free energy change: A↔B with mapping "1" (two simulations for forward and reverse) and A ↔ B with mapping "2" (two more simulations for forward and reverse).

The differential free energy of binding is determined according to the following thermodynamic cycle where "P" is the



protein and "I_A" is inhibtor "A" and "I_B" is inhibitor "B". P:I is a protein−inhibitor complex. Experimentally, we measure ΔG_A and ΔG_B, the free energies of association for substrates A and B, respectively, with the protein. Theoretically, we calculate ΔG_P and ΔG_{aq}, which are related to the experimental values through the thermodynamic cycle:

$$\Delta\Delta G = \Delta G_P - \Delta G_{aq} = \Delta G_B - \Delta G_A \qquad (2)$$

To complete this cycle, we therefore need to perform a second set of simulations to calculate the value of ΔG_{aq} for each A-to-B conversion. These simulations were carried out by placing each binder in a periodic box box of water of approximately 28 Å on a side. The box was minimized for 1000 steps, followed by 500 ns of equilibration, at a constant temperature of 300 K and a constant pressure of 1 atm. The free energy simulations were carried out using 21 integration points, with 20 ps of equilibration, and 20 ps of data collection at each point. In each case, both "forward" and "backward" TI simulations were performed, separated by 500 ps of reequilibration.

For comparison, the same set of 16 inhibitors has been scored using the OWFEG free energy grid approach[27] and also using three of the most commonly applied scoring methods: ChemScore,[13,14] Piecewise Linear Potential (PLP),[15] and Dock scoring grid.[16,17] Of particular note among these is the OWFEG grid approach. We have recently documented this method.[18,27] The basic idea is that a qualitative grid is generated within the active site of interest, using a one window free energy perturbation calculation at each grid point to approximate the free energy change for adding a probe group at that point. We have found that this method produces results in better agreement with experiment than other rapid qualtivative scoring methods in common use. In addition, the grid used in this method is generated from statistics obtained during an MD simulation, using the same atomic parameters as those used for the precise TI free energy calculations. Therefore, the

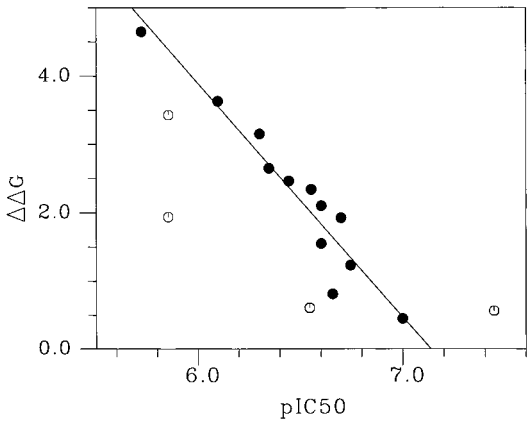


Figure 2. Relative free energy of binding, $\Delta\Delta G$, for each p38 inhibitor as calculated using TI versus the experimentally measured pIC_{50}. Free energies are relative to the reference state molecule in Figure 1a. For visual emphasis, the 12 points that suggest a very strong linear correlation are shown as solid circles, with the remaining four points shown by open circles. The line shown is a least-squares fit to the 12 filled circles. Results correspond to a model with 0.5 kcal/mol positional restraints on the moving atoms of the protein. Each data point represents the average of two sets of two simulations (four simulations total): forward and reverse with both possible mappings (Figure 1c).

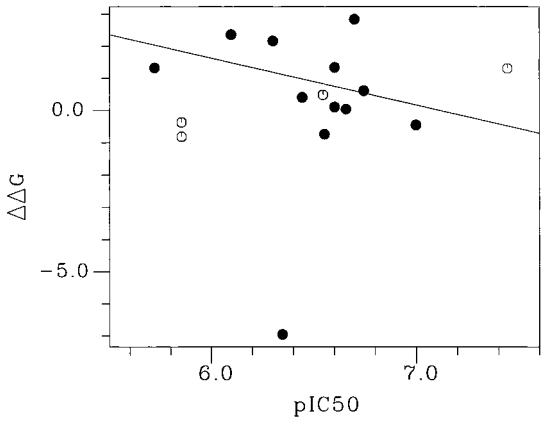


Figure 3. Relative free energy of binding, $\Delta\Delta G$, for each p38 inhibitor as calculated using TI versus the experimentally measured pIC_{50}. Details are the same as in Figure 2, but correspond to a model with no positional restraints on the moving atoms of the protein. To aid in comparison with Figure 2, the points represented by open circles here are the same as in that figure.

results using OWFEG and using the precise TI simulations should be directly comparable, since both are based on the same parameter set. Note that a free energy grid is converted into an OWFEG scoring grid using a linear translation function which requires three fitted parameters. For the OWFEG results reported here, these parameters were derived from a fit to experimental data for IMPDH and HIV-1 aspartyl protease protein systems.[18]

Results

The free energy differences $\Delta\Delta G$ were calculated according to eqs 1−2 for all 16 inhibitors of p38. These have been plotted against the experimentally measured values of $p(IC_{50})$ in Figure 2. For visual emphasis, points have been plotted with two types of symbols: Those that appear to be strongly correlated with the experimental data are shown as filled circles; those that appear to fall off the correlation line are plotted as open circles. Crude error estimates in the calculated $\Delta\Delta G$ values, based on comparison of the energies calculated in the forward and reverse directions, are on the order of 0.2 kcal/mol or smaller. The correlation between the well fit points and experiment is strikingly good.

The first reaction to Figure 2 is to question whether the small number of compounds that fall away from the correlation line are somehow obviously different than those that are well fit. The answer is no: these compounds are not different in any obvious manner that would lead to expectations that they might present difficulties in calculation. The compounds that fall off the line (compounds **3**, **5**, **13**, and **16** in Table 1) all contain small organic and halogen substitutions that do not differ from other compounds in the series in any obvious way. Nor do these points apparently reflect a simple lack of convergence. We re-ran the simulations for the off-line points, doubling the amount of equilibration and data collection at each integration point to 20 ps each. The results were essentially unchanged from those in Figure 2.

To test if perhaps these points reflect problems with the model of the system, additional simulations were run. For these, the positional restraints on the protein were removed, and the sampling and equilibration at each integration point were doubled to 20 ps each. All other parameters and simulation details were kept the same. The results in this case are shown in Figure 3. As can be seen, not only does this model not improve the fit of the "off-line" points in Figure 2, correlation among the reamining points is siginficantly poorer. This most likely reflects the much greater number of conformational substates available in a no-restraint model, which means our ability to fully sample important configurations in a reasonable amount of simulation time is curtailed. The result would be greater imprecision in the calculated values.

Since they are much more computationally expensive, there would be no point to running precise free energy calculations if they did not provide more reliable predictions than rapid scoring schemes. To determine if this advantage exists, the same 16 compounds have been scored using four scoring functions: OWFEG, ChemScore, PLP, and Dock Energy Score. OWFEG is a new scoring method that is based on an approximate free energy grid derived from a single MD simulation. ChemScore, PLP, and Dock Energy Score are the rapid scoring methods currently in wide use that afforded the best overall results in a recent survey.[4] In another recent paper, we demonstrated that OWFEG appears to be more predictive over multiple protein systems than any of these other methods.[18]

The results using OWFEG, ChemScore, PLP, and Dock Energy Score are shown in Figure 4a−d. To aid in comparison to Figure 2, the same compounds that were represented by closed and open circles in that figure are so-represented in here. In each plot, the subset of points that suggest a linear correlation have been fitted to a line. This line has no purpose except to provide visual guidance in interpreting the plot. It is immediately apparent that in no case does a substantial subset of the scored points suggest a correlation as tight as was obtained using the exact TI simulations. This is

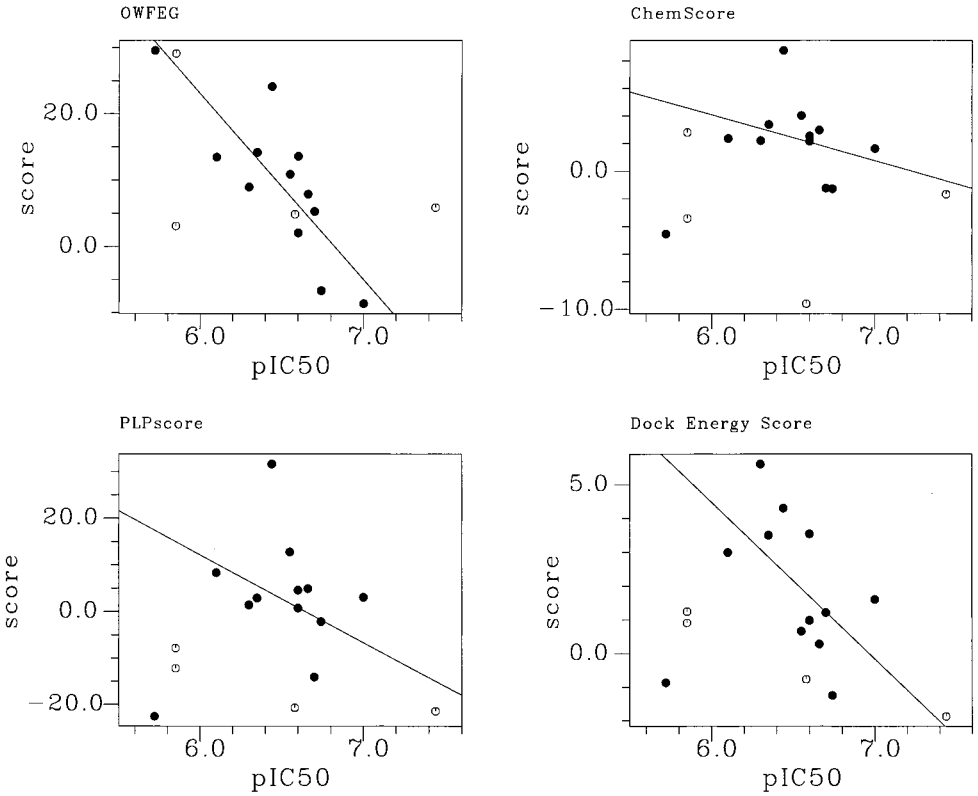


Figure 4. Predicted binding scores for the p38 ligands as calculated using four scoring schemes: OWFEG, ChemScore, PLPScore, and Dock Energy Score. The line in each plot is a fit to those points in the plot that most strongly suggest, on an empirical qualitative basis, a linear correlation. The line is shown for visual emphasis. The same inhibitors represented by open circles in Figure 2 are represented by open circles in these plots to aid in comparisons to the Figure 2 data. (A) OWFEG scoring; (B) ChemScore scoring; (C) PLPscore scoring; (D) Dock Energy Score scoring.

not surprising, given the many approximations made in the empirical schemes. What *is* surprising is how generally good the OWFEG scored data are, since the OWFEG grid is essentially a qualitative version of the TI simulation. The other scoring methods, ChemScore, PLP, and Dock Energy Score, all appear to perform much worse than OWFEG for this data set.

It is important to consider the scoring abilities of TI and the other various methods in the context of how they will actually be used. Precise free energy calculations are generally much too costly to apply to broad questions such as "what should we do"? Instead, they are most commonly applied to answer specific multiple choice questions, e.g., 'Should group A be changed to group B, to group C, or not changed at all?' Within this context, the feature of greatest importance is not the exact free energy calculated by the method, but rather the reliability of the method in correctly choosing A, B, or C (i.e., not the absolute free energy values, but their relative order).

To quantify this aspect of the models, we have developed the predictive index, PI. For a set of experimental pK_i (or pIC_{50}) values, $E(i)$, and corresponding predicted scores $P(i)$, the index is defined as follows:

$$\mathrm{PI} = \sum_{j>i}\sum_{i} w_{ij}\, C_{ij} \Big/ \sum_{j>i}\sum_{i} w_{ij} \tag{3}$$

with

$$w_{ij} = |E(j) - E(i)| \tag{4}$$

and

$$C_{ij} = \begin{cases} 1 & \text{if } [E(j) - E(i)]/[P(j) - P(i)] < 0 \\ -1 & \text{if } [E(j) - E(i)]/[P(j) - P(i)] > 0 \\ 0 & \text{if } [P(j) - P(i)] = 0 \end{cases} \tag{5}$$

$P(i)$ is the model score assigned to molecule i and $E(i)$ is the experimental pK_i (or pIC_{50}). This index ranges from −1 to +1, depending on how well the predictions track the order of experiment: +1 arises from perfect prediction, −1 arises from predictions that are always wrong, and 0 arises from a predictions that are completely random. Note that this function includes a weighting term that depends on the difference between the experimental values. This reflects the fact that a good function should be able to differentiate between changes that result in large differences in binding. On the other hand, for changes that result in little or no change in the experimental binding, selection of A or B by the prediction algorithm is basically irrelevant, since either choice will ultimately result in a compound with the same binding efficacy. Note also that the sign convention in eq 5 reflects the fact that here $E(i)$ represents pK_i or pIC_{50}, i.e., the log of the binding efficacy. The sign convention would be reversed if we were to do the comparisons with K_i or IC_{50} directly.

The calculated values of the predictive index for precise TI, as well as for OWFEG, ChemScore, PLP, and Dock Energy Score, are presented in Table 2. These data make it clear that TI preforms *substantially* better as a predictive tool than do the qualitative schemes. The weighted predictive index for TI is quite high, 0.843.

Table 2. Predictive Indices (PI) for Free Energy Calculations and Empirical Scoring Methods

scoring method	PI
thermodynamic integration	0.843
OWFEG	0.563
ChemScore	0.043
PLPScore	−0.052
Dock Energy Score	0.248

[a] The predictive index is defined in eqs 3−5. The predictive index measures how reliably a function can correctly choose the better binder for any pair of molecules. A predictive index of +1.0 would mean the function was correct 100%, an index of −1.0 would mean the index was wrong 100%, and an index of 0.0 would result from random selection.

By comparison, OWFEG is 0.563, the Dock Energy Score is 0.248, ChemScore gives 0.043, and PLP score yields −0.052. Since by definition pure random choice would allow a PI of 0.00, ChemScore is little better than random and PLP score is actually slightly worse than random for this data set. Given the small total number of compounds and associated errors in the calculated values of the PI, we can broadly conclude that ChemScore and PLP have no ability to make the subtle differentiations required to rank this series, while the abilities of the Dock Energy Score are modest. This is perhaps not surprising, since these methods have been developed for rapid screening, but it does emphasize caveats regarding overinterpretation of compound rankings when screening using these methods.

Discussion

It is clear that we can obtain predictions from exact free energy simulations that are of sustained very high reliability, at least within a congeneric series. The quantitative results using TI to predict IC_{50} values for the set of p38 binds are quite impressive. This is seen both visually (Figure 2) and numerically via the predictive index (PI).

The results using TI are even more impressive when one considers the various limitations and approximations that have been made in performing these calculations. Among these are the use of a moving protein "belly" surrounded by a sphere of moving water in lieu of a full free protein and water periodic boundary simulation; the use of modest positional restraints on the moving protein belly; and the use of somewhat generic force field parameters for the various inhibitors (which have not been optimized for this particular congeneric series). In addition, it is important to note that the calculated free energies are being compared to experimentally measured IC_{50} values. While, empirically, IC_{50} and K_i are roughly related by a proportionality factor, the relationship is not typically precise. Thus, while pK_i and ΔG form a strict linear correlation ($\Delta G = -RT \ln(K)$), the correlation between pIC_{50} and ΔG is not expected to be exactly linear.

The PI obtained using the OWFEG method is very good, given the qualitative and rapid nature of the approach. This begs the question of whether we should contemplate using a method such as OWFEG in lieu of the considerably more expensive precise TI approach. We emphatically believe the answer is no. Recall that the reason these types of simulations are being applied is to help direct synthesis in the end stages of compound optimization. If these simulations are to have any real impact, then one must be willing to base synthetic strategy on the suggestions they provide. It is a nice academic exercise to make predictions, then test out all the choices, but this saves nothing and ultimately renders the theoretical prediction valueless in the real world. If we are going to make choices based on these calculations, then we must strive to base those choices on the best theoretical values available; the cost of a bad choice in direction can be enormous in both resources wasted and lost time. While a PI of 0.563 is good by the standards of qualitative schemes, it is *very* much inferior to the PI of 0.843 available using precise TI simulations. Within the context of how and why these calculations might be applied, the extra costs of carrying out the TI simulations are slight compared to the downstream costs of a inaccurate decision. The superior predictive index for OWFEG versus the other rapid approaches does, however, further strengthen the arguments made in our previous paper:[18] that OWFEG can be a valuable component of any theoretical screen program, and that the OWFEG approach appears to have strengths that the methods now in common use for such screening do not share.

In the end, the results presented here lend credence to the assertion that precise free energy calculations *should* routinely be added to the toolbox of simulation tools that a modeler utilizes in the process of drug design. The key is knowing how *and when* to apply them. They are not appropriate for the initial stages of drug discovery. But nearing the finish line, when the path of synthesis is frequently characterized by a series of multiple choice questions within a congeneric series, free energy calculations can provide valuable guidance that is not available using the more empirical (and faster) tools that are so useful in the early stages of a project.

Acknowledgment. We are grateful to Dr. Ajay for numerous enlightening discussions and to Dr. Vicki Sato for critical reading of the manuscript.

References

(1) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening − An Overview. *Drug Discovery Today* **1998**, *3*, 160−178.
(2) Leach, A. R.; Hann, M. M. The in silico world of virtual libraries. *Drug Discovery Today* **2000**, *5*, 323−336.
(3) Ajay,; Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand−Receptor Complexes. *J. Med. Chem.* **1995**, *38*, 4953−4967.
(4) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A method of obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.
(5) Pearlman, D. A.; Rao, B. G. Free energy calculations: methods and applications. In *Encyclopedia of Computational Chemistry*; Schleyer, P. V. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, R. P., Eds.; John Wiley & Sons: New York, 1998; pp 1036−1061.
(6) *Methods of Electronic Structure Theory*; Schaefer, H. F., Ed.; Modern Theoretical Chemistry, Vol. 3; Plenum: New York, 1977.
(7) *Methods of Electronic Structure Theory*; Schaefer, H. F., Ed.; Modern Theoretical Chemistry, Vol. 4; Plenum: New York, 1977.
(8) *Semiempirical Methods of Electronic Structure Calculation*; Segal, G., Ed.; Modern Theoretical Chemistry, Vol. 7−8; Plenum: New York, 1977.
(9) Shi, Y.-Y.; Mark, A. E.; Cun-xin, W.; Fuhua, H.; Berendsen, H. J. C.; van Gunsteren, W. F. Can the stability of protein mutants be predicted by free energy calculations? *Protein Eng.* **1993**, *6*, 289−295.
(10) Mark, A. E.; van Helden, S. P.; Smith, P. E.; Janssen, L. H. M.; van Gunsteren, W. F. Convergence properties of free energy calculations: Alpha-cyclodextrin complexes as a case study. *J. Am. Chem. Soc.* **1994**, *116*, 6293−6302.

(11) Wade, R. C.; McCammon, J. A. Binding of an antiviral agent to a sensitive and a resistant human rhinovirus. Computer simulation studies with sampling of amino acid side-chain conformations. I. Mapping the rotomers of residue 188 of viral protein 1. *J. Mol. Biol.* **1992**, *225*, 679−696.
(12) Wilson, K. P.; McCaffrey, P. G.; Hsiao, K.; Pazhanisamy, S.; Galullo, V.; Bemis, G. W.; Fitzgibbon, M. J.; Caron, P. R.; Murcko, M. A.; Su, M. S. The structural basis for the specificity of pyridinylimidazole inhibitors of p38 MAP kinase. *Chem. Biol.* **1997**, *4*, 423−431.
(13) Murray, C. W.; Auton, T. R.; Eldridge, M. D. Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand−receptor binding affinities and the use of Bayesian regression to improve the quality of the model. *J. Comput.-Mol. Des.* **1998**, *12*, 503−519.
(14) 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 Des.* **1997**, *11*, 425−445.
(15) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Freer, S. T. Molecular Recognition of the Inhibitor AG-1343 by HIV-1 Protease − Conformationally Flexible Docking by Evolutionary Programming. *Chem. Biol.* **1995**, *2*, 317−324.
(16) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking With Grid-Based Energy Evaluation. *J. Comput. Chem.* **1992**, *13*, 505−524.
(17) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18*, 1175−1189.
(18) Pearlman, D. A.; Charifson, P. S. Improved scoring of ligand-protein interactions using OWFEG free energy grids. *J. Med. Chem.* **2001**, *44*, 502−511.
(19) Case, D. A.; Pearlman, D. A.; Caldwell, J. C.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. AMBER 5.0. University of California, San Francisco, 1997.
(20) Pearlman, D. A.; Kollman, P. A. A new method for carrying out free energy perturbation calculations: Dynamically modified windows. *J. Chem. Phys.* **1989**, *90*, 2460−2470.
(21) Walters, W. P.; Corkery, J. J.; Charifson, P. S.; Murcko, M. A. Skizmo-A new program for rapid conformational analysis and structure-based design of combinatorial libraries. *Abstracts of Papers*, 217th National Meeting of the American Chemical Society, Anaheim, CA, March 21−25, 1999; American Chemical Society: Washington, DC, 1999; Abstract COMP-012.
(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(23) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(24) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98 (Revision A.3). Gaussian Inc.: Pittsburgh, PA, 1998.
(25) Molecular Simulations Inc., Charmm Parm22, Burlington, MA.
(26) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(27) Pearlman, D. A. Free energy grids: A practical qualitative application of free energy perturbation to ligand design using the OWFEG method. *J. Med. Chem.* **1999**, *42*, 4313−4324.

JM0100279