JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

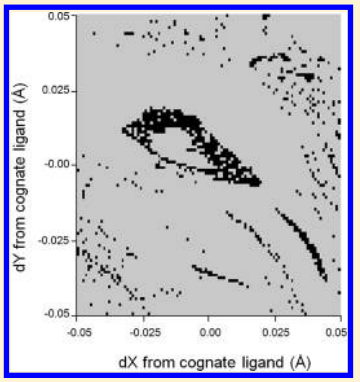
Numerical Errors in Minimization Based Binding Energy Calculations

Miklos Feher*,[†] and Christopher I. Williams[‡]

[†]Campbell Family Institute for Breast Cancer Research, University Health Network, Toronto Medical Discovery Tower, 101 College Street, Suite 5-361, Toronto, ON M5G 1L7, Canada

[‡]Chemical Computing Group, Suite 910, 1010 Sherbrooke Street West, Montreal, QC H3A 2R7, Canada

ABSTRACT: This work examines the effect of small input perturbations on binding energies computed from differences between energy minimized structures, such as the Prime MM-GBSA and MOE MM-GB/VI methods. The applied perturbations include translations of the cognate ligand in the binding site by a maximum of 0.1 Å along each coordinate or the permutation of the order of atoms of the cognate ligand without any changes to the atom coordinates. These seemingly inconsequential input changes can lead to as much as 17 kcal/mol differences in the computed binding energy. The calculated binding energies cluster around discrete values, which correspond to specific ligand poses. It appears that the largest variations are observed for target-ligand systems in which there is a possibility for multiple poses with strong hydrogen bonds. The barriers between different poses can appear fractal-like, making it difficult to predict which solution will be produced from a given input. Including protein flexibility in MM-GBSA calculations further increases numerical instability, and the protein strain terms seem to be the major factor contributing to this sensitivity. In such calculations it appears unwise to extend the flexible region beyond 6 Å.



INTRODUCTION

Accurate prediction of protein−ligand binding affinity ($\Delta G_{binding}$) is considered by many to be the holy grail of computational drug discovery, as it would help tremendously with ranking candidate compounds and selecting the most promising ones for synthesis. This has led to the development of a plethora of computational methods for predicting $\Delta G_{binding}$ (see e.g. refs 1−3 and references therein), ranging from simple empirical scoring functions to more sophisticated free energy techniques. Simple empirical scoring functions (typically applied in docking programs) are somewhat inaccurate but are popular due to their ease of computation and their applicability across a wide range of chemical classes. More sophisticated methods such as free energy perturbation and thermodynamic integration[4−7] can be more accurate, but they are quite computationally intensive and are usually only applied to a series a highly similar molecules, which limits their scope and applicability.

In between the simple and more sophisticated approaches to $\Delta G_{binding}$ calculations lie methods that are more complex than empirical scoring functions but are still computationally tractable and somewhat realistic for a wide range of ligand classes.[8−11] The MM-GBSA[12] method is probably the most popular of these approaches (together with MM-PBSA), mainly because of its conceptual simplicity as a combination of an all-atom molecular mechanics force field (e.g., Amber, CHARMM, OPLS-AA) and the Generalized Born solvation method GB-SA.[13] In MM-GBSA calculations the target structure can be treated as rigid, partially flexible, or fully flexible. Combining MM with GB-SA solvation is believed to capture ligand strain, protein−ligand interactions, solvation/desolvation effects, and possibly protein strain, yielding a prediction of $\Delta G_{binding}$ that includes all these energetic contributions. In a popular variety of the MM-GBSA, the molecular dynamics simulation is replaced by single energy minimizations, as is the case in the two methods studied in this work, Prime MM-GBSA and MOE MM-GB/VI.[14] It has been previously pointed out that using energy minimized energies instead of snapshots from MD simulations often leads to correct energies but occasionally produces entirely wrong energies,[15] but the effect was not studied systematically. The most popular use of such binding energy calculations is as a fast rescoring protocol after docking.[16−19]

We have recently investigated the effect of small input perturbations on the output of deterministic simulations such as docking and MM protein minimizations.[20−22] Given the exact same input, repeated runs of a deterministic simulation on a single CPU will produce the exact same output, because no stochastic elements are used in their calculation. However, deterministic simulations can be quite sensitive to initial conditions, to the extent that small input perturbations can give rise to wildly different and unexpected simulation outcomes. With docking calculations, tiny perturbations to the ligand input structure can produce a number of significantly different top-scoring poses, whereas for MM protein minimizations, small perturbations to the initial atomic coordinates can have a dramatic effect on the energy and geometry of the final optimized structure.[23] The magnitude of output variations in protein minimizations also showed a strong size dependency, with larger structures producing greater output variations than smaller structures.

Received: June 26, 2012

   dx.doi.org/10.1021/ci300298d | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

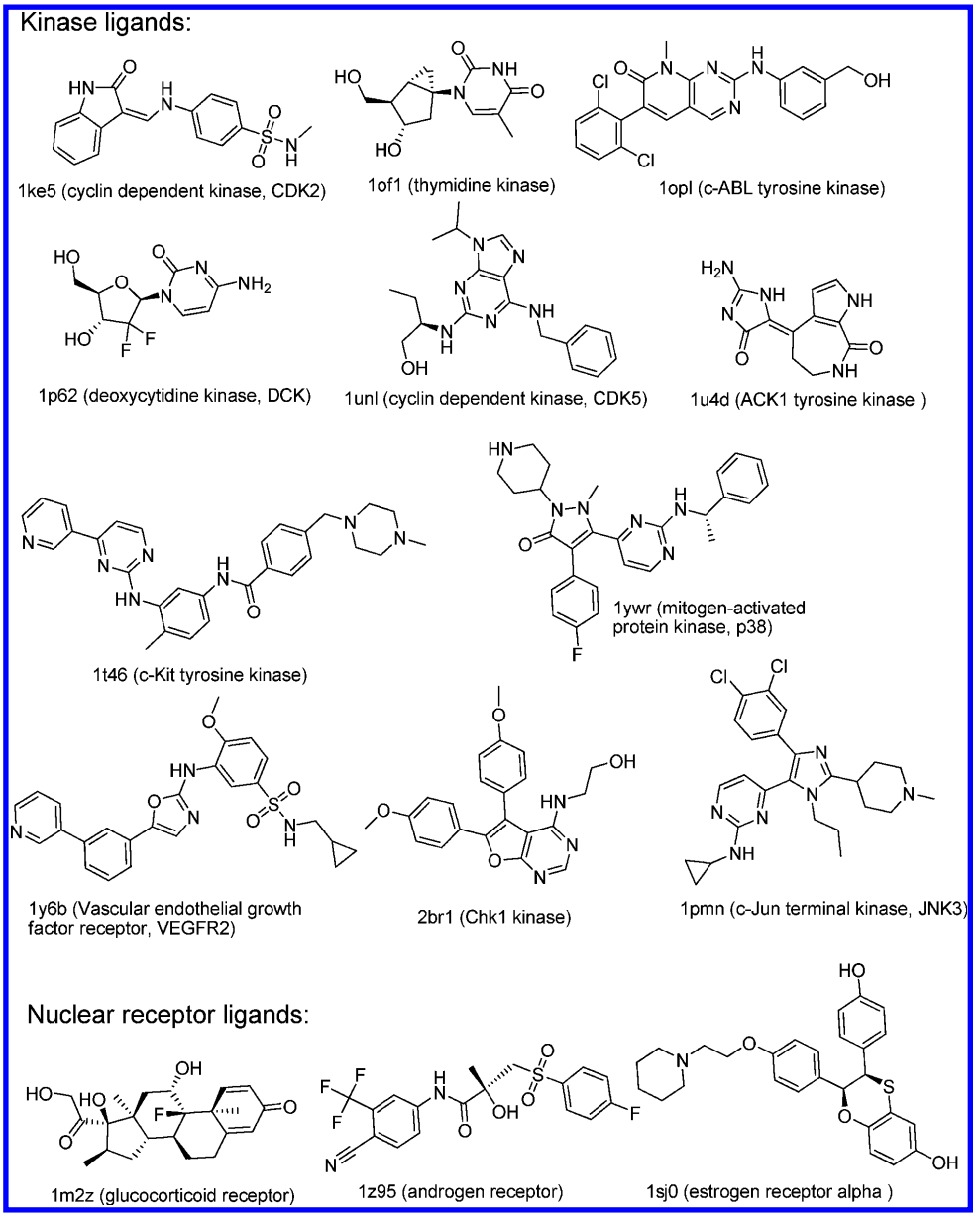


Figure 1. Ligand structures (with target names and PDB codes) used in this study.

Even though MM-GBSA calculations do not contain stochastic elements, we speculated that MM-GBSA calculations using a single minimized structure could exhibit input sensitivities similar to those observed in docking and protein minimizations. To our knowledge, no systematic studies of MM-GBSA input sensitivity have been published, but one relevant study performed by Genheden and Ryde[24] used four equivalent avidin binding pockets to estimate binding affinity using MM-GBSA. The four equivalent pockets (obtained from a tetrameric crystal structure) were expected to give the same estimated binding energy, but they did not. After testing different simulation protocols to derive a more reliable binding energy, the authors finally concluded that in this context, it is better to average the results of several independent simulations rather than rely on the results of only one run. This conclusion is somewhat similar to our proposed multiple input docking approach,[20] aimed at obtaining the best correlation between docking scores and binding affinities. The variation in MM-GBSA binding energy between four equivalent avidin pockets suggests that MM-GBSA binding energies are indeed highly sensitive to input variations, and our objective here is a systematic examination of this effect.

METHODS

Protein Preparation. For comparative purposes, we used the same set of 10 kinase and 3 nuclear receptor targets as in our previous work,[20−22] the X-ray structures of which were carefully checked for errors.[25] The pdb codes and the names of the considered targets are as follows: 1ke5 (cyclin dependent kinase, cdk2), 1of1 (thymidine kinase), 1opl (c-abl tyrosine kinase), 1p62 (deoxycytidine kinase, dck), 1unl (cyclin dependent kinase, ckd5), 1t46 (c-kit tyrosine kinase), 1ywr (mitogen-activated protein kinase, p38), 1y6b (vascular endothelial growth factor receptor, vegfr2), 2br1 (checkpoint kinase, chk1), 1pmn (c-jun terminal kinase, jnk3), 1m2z (glucocorticoid receptor), 1z95 (androgen receptor), and 1sj0 (estrogen receptor-α). However, the preparation of these structures was somewhat different from our previous work, due

dx.doi.org/10.1021/ci300298d | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Percentage Range in Binding Free Energies[a] from the Prime MM-GBSA and MOE MM-GB/VI Optimization of Four Ensembles[b] of near-Identical Input Structures

	$\Delta G_{binding}$ (percentage variation with ligand strain) Prime MM-GBSA				$\Delta G_{binding}$ (percentage variation with no ligand strain) Prime MM-GBSA				$\Delta G_{binding}$ (percentage variation with ligand strain) MOE MM-GBVI	
targets	0.1 Å	0.05 Å	0.01 Å	A.S.[b]	0.1 Å	0.05 Å	0.01 Å	A.S.[b]	0.01 Å	A.S.[b]
1ke5_cdk2	8.23	0.49	0.40	5.76	10.0	9.7	9.6	1.05	0.11	0.09
2br1_chk1	0.73	0.77	0.65	0.54	0.5	0.8	0.7	0.67	0.08	0.08
1opk_abl	0.78	0.57	0.32	0.24	0.7	0.5	0.3	0.13	0.17	0.02
1ywr_p38	0.59	0.37	0.26	0.74	0.5	0.3	0.2	0.67	0.60	0.53
1of1_thym	0.80	0.69	0.33	0.37	0.8	0.7	0.3	0.39	0.08	0.07
1pmn_jun3	5.42	5.41	0.46	9.88	0.3	0.2	0.1	4.34	4.93	4.96
1p62_dck	15.96	13.83	14.01	14.44	10.2	9.7	8.8	9.00	0.12	0.11
1t46_ckit	0.59	0.42	0.31	0.35	0.6	0.4	0.2	0.25	0.15	0.12
1unl_cdk5	36.49	32.17	30.65	32.37	17.6	17.8	17.5	17.71	18.70	0.14
1y6b_vegf2	6.80	6.71	6.45	4.93	1.2	1.1	0.7	0.68	0.08	0.08
1m2z_gr	0.82	0.56	0.25	0.26	0.8	0.5	0.3	0.23	11.42	11.46
1sj0_er	0.39	7.75	0.26	0.37	0.3	4.8	0.3	0.31	0.09	0.09
1z95_ar	10.25	0.39	0.29	0.26	7.9	0.3	0.2	0.13	0.38	0.36

[a]100 times the range of ΔG's over mean of ΔG's. [b]Three of the input ensembles consisted of 729 copies of the cognate ligand, translated within a cube with a side of 0.1 Å, 0.05 Å, and 0.01 Å. The fourth input ensemble (A.S.) consisted of 729 identical structures which differed only by the order in which the atoms were listed in the input file. See text for further details.

to the specific requirements of Prime[26] related to the separation and naming of chains and residues, as well as the naming of hydrogen atoms. As before, we used the protein preparation wizard in Maestro (hydrogen bond optimization and restrained structure minimization with the OPLS-AA force field to a maximum RMSD of 0.3 Å). To standardize repeated calculations on the same sets of ligands, the prepared target and sets of ligands were stored together in a Maestro (.mae) file and used in subsequent calculations. Structures to be used in energy calculations in MOE were prepared with the Protonate3D algorithm.

Ligand Ensemble Preparation. The kinase and nuclear receptor ligands were the cognate ligands from the X-ray structures described above (see Figure 1). Since the purpose of these MM-GBSA calculations was to calculate the binding energy of the cognate ligand, it was kept in its original position. Hydrogen atoms were added, and their proper protonation state was determined inside the target using the Protonate3D process in MOE. From this starting point, three ensembles of alternative ligand input positions were generated by translating the cognate ligand within a cube with a side of 0.01 Å, 0.05 Å, and 0.1 Å with 9 steps in each direction (corresponding to 729 structures per ensemble). These translations are very small, and the resulting structures are indistinguishable when overlaid on a normal computer screen. A fourth ensemble of 729 input ligands was generated by keeping the cognate ligand in its original position (i.e., all atomic coordinates remained identical) and only shuffling the order of atoms within the input file. This process mimics the real life situation of changing the order in which the atoms are drawn and connected in a molecular drawing program. This set is henceforth referred to as the *atom shuffled* (A.S.) ensemble.

Binding Energy Calculations. MM-GBSA calculations were performed with the Prime program[26] within the Schrödinger suite. This choice was made because of the popularity of this package within the industry and its ability to treat part of the protein structure as flexible. The OPLS-AA[27] force field was applied in conjunction with the SGB continuous solvation model.[28,29] In these calculations the protein structure was either kept fixed, or regions within a certain distance from the ligand were treated as flexible. In these latter cases, no constraints were applied on protein atoms, and all protein atoms (i.e., both backbone and side chain atoms) were included in the minimizations. The gradient cutoff was 0.01 kcal/mol Å^2 in these calculations.

Alternatively, MM-GB/VI (Molecular Mechanics, Generalized Born Volume Integral) binding energies were also calculated, available in MOE[14] as part of the Pose Rescoring protocol and utilizing the Generalized Born Volume Integral implicit solvent model.[30] The MM-GB/VI binding energies are also computed from the difference between the single point energies of the complex, the unbound ligand, and the unbound target. The MM-GB/VI energy considers only absolute ligand strain, not strain relative to the global minimum ligand conformation and ligand and protein entropies are ignored.

It may seem reasonable that the convergence criterion of the minimization might affect the output variabilities described in this work. Hence the effect of varying the convergence criterion on the calculated energies was studied in one example (1pmn) between 10^{-6}−10 kcal/mol Å^2. It was found that below a gradient cutoff of 1 kcal/mol Å^2, the variations in final energies did not change significantly, indicating that minimizations to smaller gradients do not reduce the variation in output energies. Thus the applied gradient was left at its default value of 0.01 kcal/mol Å^2.

Note that alternative binding energy calculation methods were also tested, including the Structure Based Design suite in Discovery Studio[31] and the MINTA package within MacroModel.[32] Although the results are not presented here, it was found that the magnitude of variability produced by these packages (using the 1ke5 system) is similar to Prime MM-GBSAand MOE GB/VI results, i.e. this variability does not arise as a result of artifacts within these particular implementations of the method.

RESULTS AND DISCUSSION

The MM-GBSA calculation on each structure proceeds in the following stages: minimization of the target alone, the ligand

dx.doi.org/10.1021/ci300298d | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

alone, and the target-ligand complex, followed by separate energy calculations of the ligand and the target extracted from the optimized target-ligand complex. The results of all these energy values and their components are available and can be compared individually with regards to input variations. Since the most important end points of these calculations are the final binding energy ($\Delta G_{binding}$) and the optimized ligand position, the variations in $\Delta G_{binding}$ and range of ligand RMSD values will be used to assess MM-GBSA output variations when the protein is held rigid.

MM-GBSA Variations with a Rigid Protein Target. This section considers binding energy calculations with a *rigid* protein target − no protein atoms are allowed to move during the MM-GBSA and MM-GB/VI optimization (the effect of receptor flexibility will be discussed in a following section). The variations in calculated binding free energies ($\Delta G_{binding}$) produced using the four input ensembles of 729 near-identical ligand structures are shown in Table 1. Three of the near-identical structure input ensembles were generated by translating the cognate ligand in 9 steps along the sides of a cube with dimensions 0.1 Å, 0.05 Å, and 0.01 Å, respectively. The fourth ensemble of 729 input structures is the *atom shuffled* (A.S.) set, where the structures all have atomic coordinates identical to the cognate ligand, but the order of atoms in the input file differs. The variations in $\Delta G_{binding}$ energies are reported as a percentage −100 times the range of $\Delta G_{binding}$ divided by the mean value of $\Delta G_{binding}$. The corresponding pose variations are reported in Table 2 as a range of RMSD values, as measured from the structures obtained after the optimization of the unperturbed cognate ligand.

Table 2. Range in RMSD Values (in Å) from the Prime MM-GBSA and MOE MM-GB/VI Optimization of Four Ensembles[a] of near-Identical Input Structures

	Prime MM-GBSA				MOE MM-GBVI	
targets	0.1 Å	0.05 Å	0.01 Å	A.S.[a]	0.01 Å	A.S.[a]
1ke5_cdk2	0.320	0.330	0.320	0.169	0.012	0.013
2br1_chk1	0.040	0.047	0.048	0.058	0.010	0.008
1opk_abl	0.014	0.009	0.002	0.004	0.006	0.003
1ywr_p38	0.004	0.004	0.001	0.002	0.019	0.015
1of1_thym	0.003	0.002	0.003	0.004	0.002	0.002
1pmn_jun3	0.006	0.005	0.002	0.290	0.450	0.450
1p62_dck	0.219	0.166	0.150	0.151	0.002	0.002
1t46_ckit	0.026	0.030	0.011	0.011	0.004	0.004
1unl_cdk5	0.841	0.785	0.842	0.785	0.763	0.012
1y6b_vegf2	0.106	0.095	0.087	0.088	0.005	0.004
1m2z_gr	0.025	0.027	0.005	0.004	0.053	0.053
1sj0_er	0.016	0.256	0.019	0.003	0.002	0.002
1z95_ar	0.362	0.003	0.003	0.003	0.009	0.006

[a]Three of the input ensembles consisted of 729 copies of the cognate ligand, translated within a cube with a side of 0.1 Å, 0.05 Å, and 0.01 Å. The fourth input ensemble (A.S.) consisted of 729 identical structures which differed only by the order in which the atoms were listed in the input file. See text for further details.

The results in Tables 1 and 2 show two types of behavior with respect to input perturbations and atom shuffling. In some systems, output variations are small, with ΔG variations less than 1% (with the corresponding energy ranges <1 kcal/mol) and RMSD output variations similar to and often smaller in magnitude than the input perturbations (<0.01 Å in many cases). This is an expected result for any well-behaved simulation; the perturbed input structures all converge to a common solution, and the simulation seems *insensitive* to initial conditions. In contrast, some systems exhibit large ΔG variations (>5%, with the corresponding ranges exceeding 10 kcal/mol) and output RMSD variations greater than the input perturbations. This is an unexpected result; one might expect energy and RMSD variations to correlate with the size of the input perturbation (i.e., 0.1 Å vs 0.05 Å vs 0.01 Å) and thus be small, but this is not the case for a good number of systems. Perhaps the most striking result in Tables 1 and 2 is that atom shuffling (the A.S. column) also gives rise to significant ΔG and pose variability, despite the fact one would expect *no* output variation from atom shuffling because the atomic coordinates are identical for all input structures (only their order in the input file differs). Results obtained with MM-GB/VI on the same systems are also listed in Tables 1 and 2. It is interesting to note that although the behavior and even the order of magnitude of the variability in Prime and MOE are mostly similar on the studied systems, there are also differences (see e.g. the 1ke5 and 1m2z systems).

Comparing ΔG variations with and without ligand strain (Table 1) shows the ligand strain energy term can contribute significantly to output variability in some cases. The most extreme cases of ligand strain variability are the 1pmn and 1y6b structures. In contrast, ligand strain energy has little effect on ΔG variations in the 1of1, 1t46, and 1m2z structures.

Pose variations in MM-GBSA and MM-GB/VI optimized structures are reported in Table 2 as the range of RMSDs from the optimized cognate ligand. We might expect flexible ligands to show the largest dispersion in the final minimized pose, but in fact the greatest values are seen for the 1ke5, 1p62, 1unl, and 1z95 systems. There also appears to be no significant correlation between the RMSD range and the size of the initial perturbation. The atom shuffled (A.S.) ensemble also shows RMSD variations, despite the fact these input structures all have the same atomic coordinates.

We speculate that variations in final energies that arise from tiny input coordinate perturbations are caused by accumulation of errors over the course of the optimization. Coordinate perturbations in the starting structures produce small differences in the initial energy and gradients, which lead to different optimization trajectories that can diverge significantly. In some sense this is widely recognized phenomenon in computational chemistry - different starting structures can optimize to different local minima − but the degree of sensitivity to small input differences observed here is somewhat unexpected.

Output variations that result from changing the atom order in an input file are more subtle in cause, arising from the properties of computer arithmetic. It is well-known that due to finite precision in digital computers, different ordering of computations with nonintegers will lead to slightly different results because of rounding errors.[33] For example, due to rounding effects, the results of summing a long list of double-precision numbers can differ in the last few decimal places of precision, depending on the order in which the numbers were added. Thus, changing the atom order in an input file can change the order in which the energy and gradient terms are evaluated in the program, producing small differences in the starting energies and gradients, which can lead to divergent optimization trajectories. This is well demonstrated in Figure 2, which shows how different permuted input files start at near-identical points but eventually follow diverging trajectories. As an aside, the effect of atom order on output variations can be

dx.doi.org/10.1021/ci300298d | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

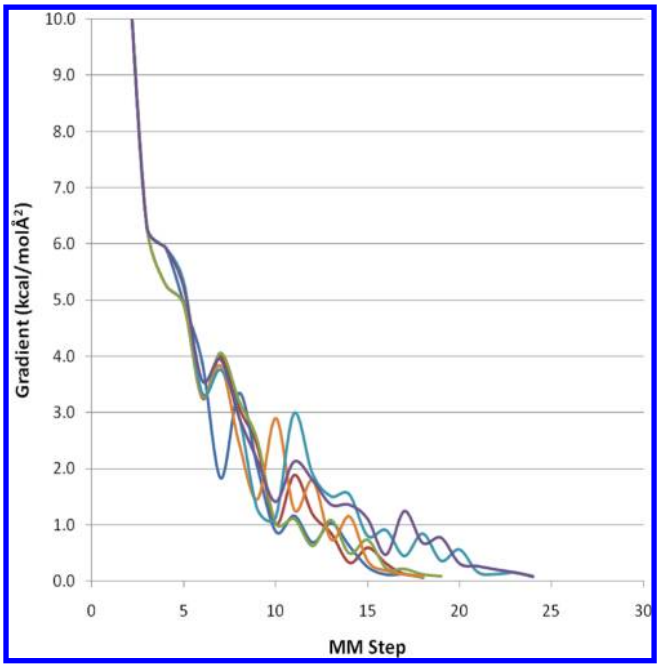


Figure 2. Gradient trajectories of selected minimizations for the 1m2z structure (glucocorticoid receptor) from MM-GB/VI calculations. In all cases the input geometries were identical; the only difference between the examples was the order of atoms in the input file. Similar gradient trajectories were observed in the other examples.

"hidden" by a computer program; for example, processing the input into some canonical form, such as sorting the coordinates prior to energy and gradient evaluations, will eliminate the effect. However, although such a procedure would *appear* to eliminate output variations due to atom permutations, it would really just mask the underlying instability of the calculation from the user, effectively hiding the problem instead of facing it.

At this point it might be worthwhile to make a note on the reproducibility of these calculations with respect to using a single versus multiple cpus, studied in some detail using MOE GB/VI in the 1pmn example. We found that when the exact same input was used and the calculations were restricted to a single processor, the results were completely reproducible on the same computer. If however the calculation was allowed to use multiple processors, the exact reproducibility was lost. It was found that in the majority of cases (>75%) the difference between the first and second calculated energy was less than 0.05 kcal/mol. However, in the remainder of cases the difference was greater, and in some cases (<10%) it was greater than 1 kcal/mol. It was also found that there was no significant difference between repeating the calculations on the same or different PCs and under linux or windows.

Again, this is a property of digital computer arithmetic: running a deterministic program with the exact same input on a single CPU will produce the exact same output, with no output variations. However, once multiple processors are brought online, each processor is assigned a portion of the calculation, and the results are then combined in the order they are received from each processor. The order in which the processors finish their tasks can vary between runs, leading to slight differences between runs.[34] Thus, with multiple processors, the same input can produce slightly different starting energies and gradients with each run, which can follow divergent trajectories in a manner similar to the coordinate perturbed and the atom-permutated input files.

In Tables 1 and 2 only the overall output variations are shown, not the distribution of solutions. The variations could represent a single outlier, a loose distribution of solutions around a single central minimum, or the existence of multiple competing distinct minima. For example, the 1z95 system with 0.1 Å input perturbations shows much larger ΔG variations than the 0.05 Å and 0.01 Å sets. Examining the individual results reveals a single outlier data point with an energy ~10 kcal/mol higher than the rest of the 728 points. If that single point is removed, the calculated relative energy range is modified from 10.25 to 0.69, which is in line with the rest of the other input variations for 1z95. Note however that the 1z95 outlier point is a genuine data point; if this single point had been used as the sole input in an MM-GBSA study of this system, the results would be grossly different from those obtained if the other input structures were used.

To gain further insight into the distributions of MM-GBSA solutions, histograms of binding energies and their corresponding RMSDs were computed for the 1p62 (dck) and 1unl (cdk5) structures (see Figures 3 and 4 respectively). These two cases were selected as examples because the solutions cover a relatively large range of energies and RMSDs, and they exhibit features seen in other systems. When comparing Figure 3 and 4 with Table 1 note that the figures display the actual calculated binding energies on the *y*-axis, while Table 1 contains the percentage energy range (i.e., the energy differences divided by the mean energy and multiplied by 100.)

The histograms in Figures 3 and 4 have well-defined peaks which correspond to different distinct minima identified by the searches. Note that if there were *no* output variations, the ΔG and RMSD histograms would both consist of a single vertical line. The peak positions in these histograms are similar for all four input ensembles (0.1 Å, 0.05 Å, and 0.01 Å pulses and A.S. - atom shuffling) indicating the same distinct minima are found by all input ensembles. However, the peak sizes differ, reflecting the different number of structures produced in each minimum by each ensemble. The energy histograms also show that by including or excluding ligand strain energy, the number and relative positions of the peaks can change; in the 1p62 case the number of peaks changes, while in the 1unl case the order of the peaks changes.

The $\Delta G_{binding}$ histogram for the 1p62 system in Figure 3 is an example of having a dominant peak corresponding to the highest (most negative) binding energy and three smaller peaks corresponding to binding energies around 6−10 kcal/mol lower than the largest binding energy. These three peaks merge into one when the ligand strain is excluded from the calculations. There is also a small scattering of data points between these larger peaks. The histogram of RMSD values in Figure 3c also shows two peaks, corresponding to two distinct minima found by the searches. As indicated on the molecule depicted above the histogram, these minima differ only in the position of two hydrogen atoms marked on the molecule as *1* and *2*. When bound to the dck enzyme these hydrogen atoms participate in strong hydrogen bonds with Glu53, Tyr86, and Glu197, which explains why they have such a major influence on the calculated binding energies. Molecules of cluster *1* only hydrogen bond to Glu53, whereas those in cluster *2* interact with all three residues, similarly to that seen for the cognate ligand after hydrogen addition and protonation with MOE's Protonate3D process. Note that the bulk of the solutions are at

dx.doi.org/10.1021/ci300298d | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

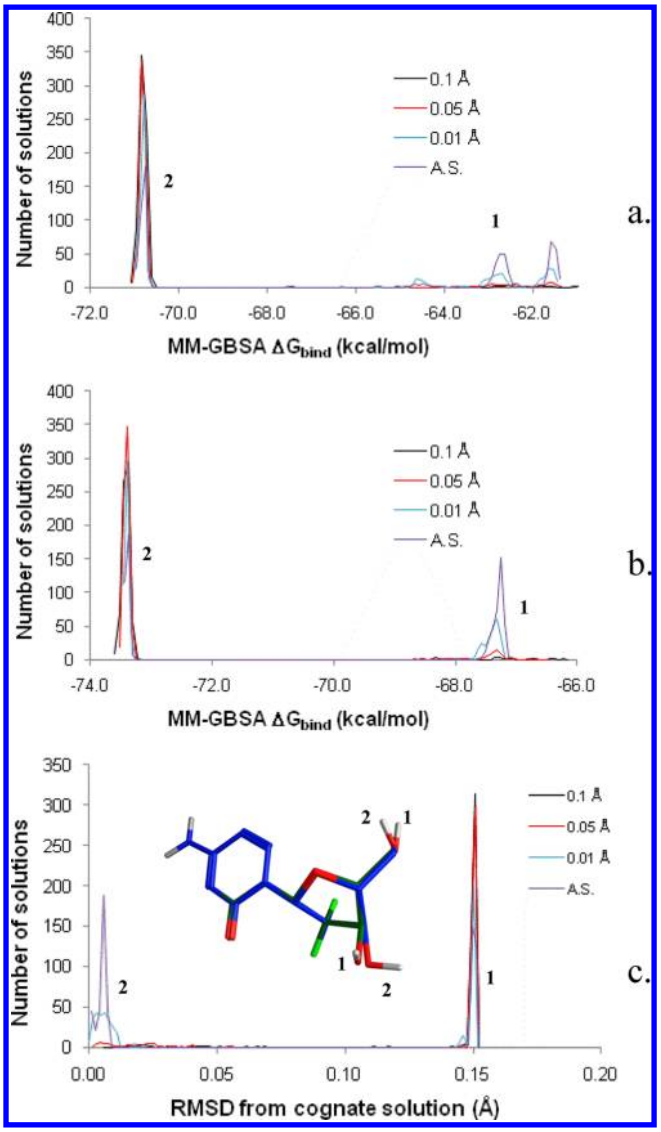


Figure 3. Histogram of the distribution of results from MM-GBSA optimizations of 729 highly similar 1p62 (dck) enzyme input structures. To obtain these histograms the cognate ligand structure was either perturbed in three perpendicular directions in 9 equal steps (9^3 = 729) by a maximum of ±Δ/2 (with Δ marked in the plot), or the order of the atoms in the input file was permuted leaving the coordinates unchanged (marked as A.S. in the plot). Each histogram contains 100 bins. Three curves correspond to Δ's of 0.1 Å, 0.05 Å, and 0.01 Å. The fourth curve is for the atom shuffled (A.S.) set. a) Distribution of MM-GBSA energies with ligand strain included. b) Distribution of MM-GBSA energies with ligand strain excluded. c) Distribution of optimized ligand RMSDs relative to the MM-GBSA optimized cognate ligand. The separate peaks in the distribution correspond to ligand poses that differ in the position of two hydrogen atoms marked on the molecule. Since these hydrogens participate in strong hydrogen bonding (to Glu53 and Glu197), they have a major influence on the calculated binding energies.

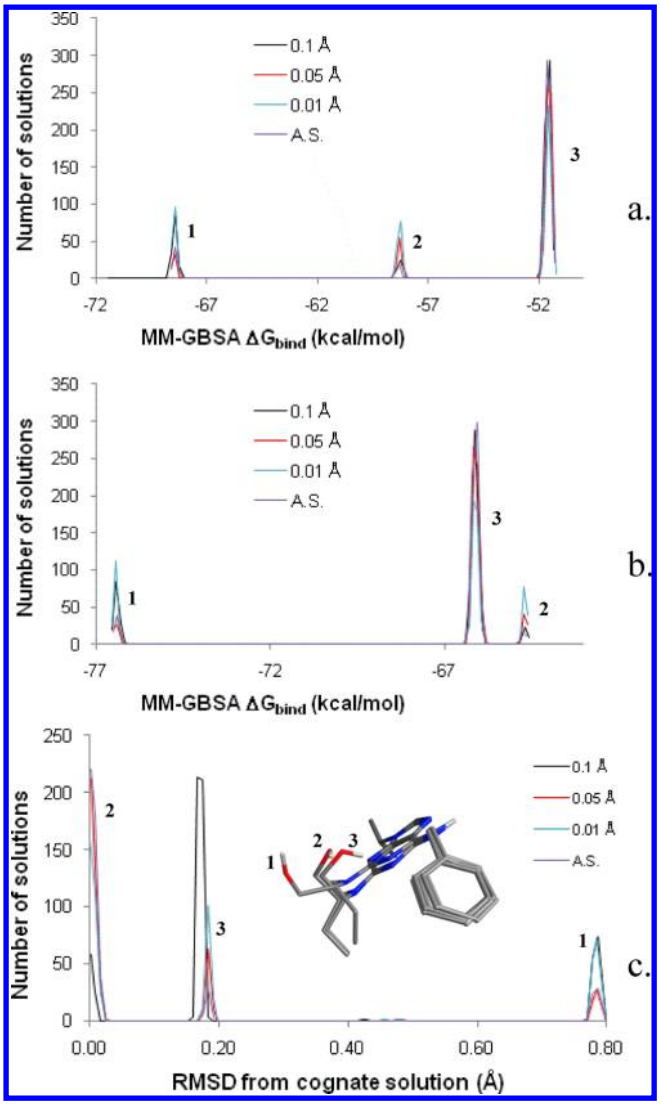


Figure 4. Histogram of the distribution of results from MM-GBSA optimizations of 729 highly similar 1unl (cdk5) enzyme input structures. To obtain these histograms, the cognate ligand structure was either perturbed in three perpendicular directions in 9 equal steps (9^3 = 729) by a maximum of ±Δ/2 (with Δ marked in the plot), or the order of the atoms in the input file was permuted leaving the coordinates unchanged (marked as A.S. in the plot). Each histogram contains 100 bins. Three curves correspond to Δ's of 0.1 Å, 0.05 Å, and 0.01 Å. The fourth curve is for the atom shuffled (A.S) set. a) Distribution of MM-GBSA energies with ligand strain included. b) Distribution of MM-GBSA energies with ligand strain excluded. c) Distribution of optimized ligand RMSDs relative to the MM-GBSA optimized cognate ligand. The separate peaks in the distribution correspond to ligand poses that differ mainly in the orientation of a hydrogen atom. The ligand poses in pose cluster 3 also differ in a few heavy atom positions. The solutions are marked on the depicted molecule and on the corresponding histogram peaks. The strong influence of hydrogen position on these MM-GBSA binding energies arises because different orientations of this hydrogen lead to different hydrogen bonding patterns (cluster *1* H-bonds to Asp86, cluster *2* has no corresponding H-bond, cluster *3* H-bonds to Gln130).

0.15 Å RMSD relative to the MM-GBSA optimized pose of the cognate ligand and are lower in energy; thus, if only the cognate ligand had been used as input in a study, the MM-GBSA binding energy would be estimated as −67.3 kcal/mol, instead of the −71 kcal/mol value produced by the majority of the calculations (an error of ∼3.7 kcal/mol).

The 1unl energy histograms in Figure 4 display 3 major peaks, covering an energy range greater than 17 kcal/mol. Each peak corresponds to a unique minimum, or pose, with each pose representing at least 20% of the solutions. As with the 1p62 system, the results with the four different input ensembles (maximum perturbations of 0.1 Å, 0.05 Å and 0.01 Å and shuffling the atom order) are qualitatively similar; the same 3

dx.doi.org/10.1021/ci300298d | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

unique poses are found in all cases, albeit with different frequencies, as reflected in the relative heights of the histogram peaks. When internal strain is excluded from the calculation, the relative peak positions change somewhat, inverting the rank order of poses 2 and 3. One notable outcome seen in Figure 4 is that the majority of solutions are at the *lower* binding energy range. Assuming that higher (more negative) binding energies arise because of a better fit of the ligand into the pocket, this result means that the minimization fails to find the lower energy minimum more often than it succeeds in doing so.

The three RMSD histogram maxima in Figure 4 also reflect the three distinct poses found for 1unl. As marked on the depicted molecule in Figure 4, the maxima correspond to three MM-GBSA solutions, two of which differ only in the orientation of hydrogen atoms (*2* and *3*), and a third which also differs in the positions of a few heavy atoms (peak *1*). The largest portions of the solutions are located in a minimum that is ∼0.2 Å from the optimized cognate ligand. The strong influence of RMSD on the MM-GBSA binding energy arises because different orientations of the hydrogen atom lead to different hydrogen bonding patterns; cluster *3* has an H-bond to Asp86 (similarly to that seen for the cognate ligand after hydrogen addition and protonation with MOE's Protonate3D process), cluster *2* has no corresponding H-bond, and cluster *1* forms an H-bond with Gln130.

Experimental protein−ligand binding energies typically range from −2 to −15 kcal/mol,[35] and a binding energy difference of about 1.4 kcal/mol represents a 10-fold change in protein−ligand affinity. It is understood that the absolute values of MM-GBSA binding energies are not necessarily in agreement with experimental binding energies, but the ranking of ligands based on the calculated binding energies are often expected to agree reasonably well with ranking based on experimental binding affinity.[36] However, the energy errors that arise from small input perturbations to MM-GBSA input are large enough to significantly affect compound ranking exercises. The range of ΔG values observed in the above examples (∼9 and ∼17 kcal/mol) represent 14% and 34% of the lowest energy outcome respectively. The same energy variations also arise when using the atom shuffled (A.S.) ensemble, demonstrating that even identical input coordinates can give rise to these variations in ΔG and RMSD. For comparison, a 20% error on an *experimental* binding energy of e.g. −10 kcal/mol represents 2 kcal/mol or almost 2 log units in activity. Even in a best-case scenario, where MM-GBSA energies perfectly correlate with experimental values and differ only by a scaling factor, errors of 14−34% in MM-GBSA energies (due to input variations) would be sufficiently large to affect compound ranking.

MM-GBSA Variations with a Flexible Protein Target. To understand the effects of protein flexibility on variations in MM-GBSA energies and placements, the same systems used in the previous section were subjected to MM-GBSA optimizations that allow parts of the protein within a given radius of the ligand to be flexible in the minimization. Figure 5 displays the calculated maximum and minimum MM-GBSA energies as a function of this radius for the 1ke5 and 1z95 structures (cdk2 enzyme and androgen receptor, respectively), obtained using 729 structures with a maximum perturbation of 0.01 Å. Without protein flexibility, these systems showed little variation in output when provided with perturbed ligand input files (Table 2). However, it is clear from these calculations that the greater the part of the protein that is treated as flexible, the larger the variation in the final energy. The separation of the highest and

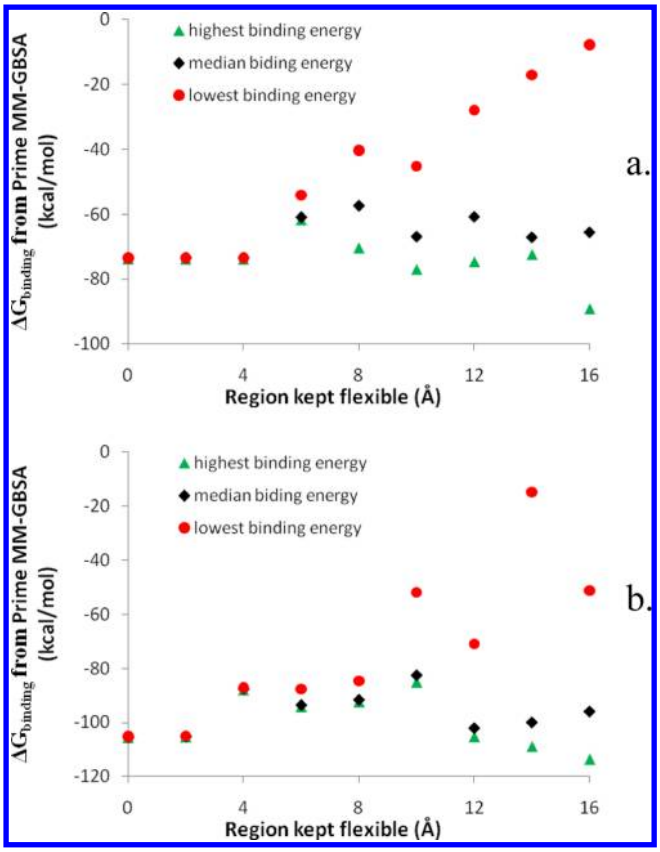


Figure 5. Minimum, median, and maximum MM-GBSA binding energies from 729 (=9^3) highly similar input structures, as a function of the radius around the ligand where the target protein is considered as flexible. These 729 inputs were obtained by perturbing the cognate ligand by a maximum of ±Δ/2 (where Δ was 0.01 Å) in three perpendicular directions in 9 equal steps. (Note that these systems exhibit very little output variation when the target protein is held rigid). a) 1ke5 structure (cdk2 enzyme). b) 1z95 structure (androgen receptor).

lowest calculated binding energies is relatively small up to a radius of 4 Å, above which it generally increases, but with considerable fluctuations in the highest, lowest, and median energies (indicating the significance of 'random' factors contributing to the final energies). Alarmingly, using a 16 Å flexibility radius produces an energy difference of 81 kcal/mol between the highest and lowest MM-GBSA energies obtained from 729 near-identical sample inputs. This is greater than the mean MM-GBSA binding energy of −62.5 kcal/mol. Even at a radius of 8 Å, the binding energy range is 30 kcal/mol, which is significant when compared with the mean binding energy of −57 kcal/mol.

Prime allows the calculation of the binding energy with and without the inclusion of ligand and target strain, and this is helpful for a better understanding of the source of variability. The flexible region of the target is determined using a radius around the ligand. Figure 6 displays binding energy ranges (i.e., the difference between the highest and lowest binding energies from the 729 MM-GBSA calculations) for the 1ke5 and 1z95 targets as a function of the radius around the ligand where the target is considered to be flexible. Figure 7 displays the corresponding RMSD ranges from the optimized ligands as a function of the same radius. If there were *no* output variability, the binding energies and RMSD plots in Figures 6 and 7 would

dx.doi.org/10.1021/ci300298d | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

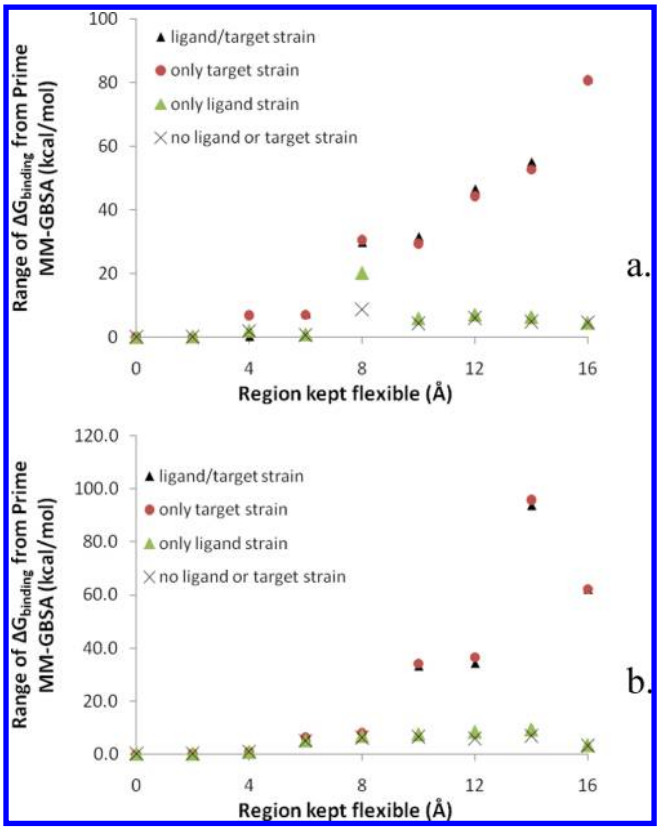


Figure 6. Range of MM-GBSA binding energies from 729 (=9^3) highly similar input structures, obtained as a function of the radius around the ligand where the target protein is considered as flexible. These 729 inputs were obtained by perturbing the cognate ligand by a maximum of $\pm\Delta/2$ (where Δ was 0.01 Å) in three perpendicular directions in 9 equal steps. The plots show the effect of including ligand and/or target strain into the calculations. The results indicate that the lack of robustness in protein strain is the main factor behind the poor reproducibility of MM-GBSA calculations with a flexible protein. a) 1ke5 structure (cdk2 enzyme). b) 1z95 structure (androgen receptor).

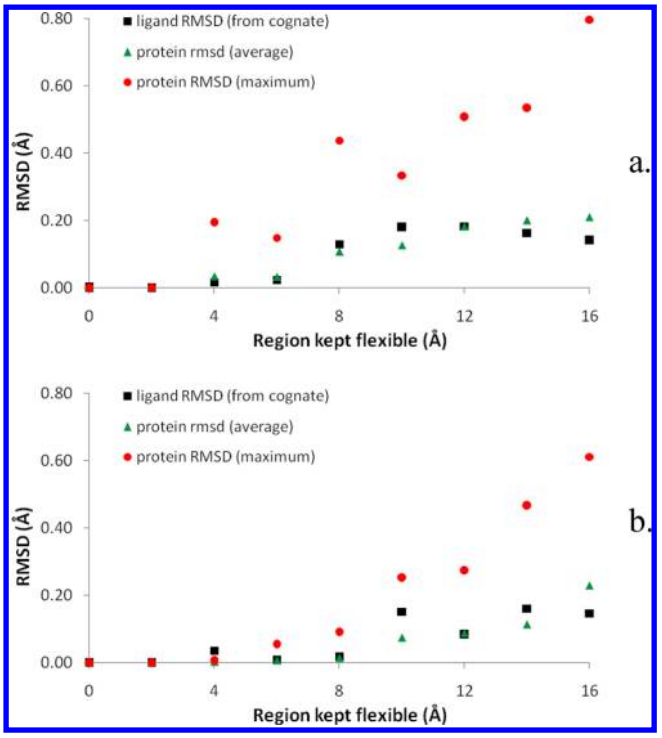


Figure 7. Root-mean square distance (RMSD) between the final minimized structures resulting from an MM- GBSA optimization of 729 (=9^3) highly similar input structures. Results are plotted as a function of the radius around the ligand where the target is considered as flexible. These 729 inputs were obtained by perturbing the cognate ligand by a maximum of $\pm\Delta/2$ (where Δ was 0.01 Å) in three perpendicular directions in 9 equal steps. The three curves show the average RMSD calculated from the cognate ligand, and the average and maximum RMSD of the target, calculated in a 20 Å box surrounding the ligand. The plot indicates that RMSD values calculated for the protein target are 'comparable' to those calculated from the ligand. The poor reproducibility of the energy values (in Figure 6) arises due to the movement of individual flexible side chains, rather than the entire protein. a) 1ke5 structure (cdk2 enzyme). b) 1z95 structure (androgen receptor).

be horizontal lines at zero. The total score variability in Figure 6 seems relatively small (<1 kcal/mol) up to a 4 Å radius around the ligand but then starts increasing rapidly. A similar effect is seen for the RMSD variations of the minimized ligands (denoted with squares); below the 4 Å flexibility radius, the range of RMSD values is below 0.005 Å, but it increases rapidly up to a range of about 0.5 Å as the flexible radius increases. The change is not monotonous, and there are large fluctuations in the actual ΔG and RMSD ranges. However it is clear that the ΔG variability is consistently of a sufficient magnitude to render the calculated MM-GBSA binding energies rather unreliable if more than a handful of protein residues are considered flexible.

It is also interesting to look at the components of this flexibility in Figure 6a. The variability of the total binding energy (both ligand and target strain included) is described by almost the same curve as when only the protein strain is considered (i.e., ligand strain is neglected). The other two curves which consider only ligand strain (i.e., protein strain neglected) or no strain (i.e., both protein and ligand strain neglected) also run close to each other, showing that ligand strain has little effect on the binding energy variability in this example. The fact that much of the variability comes from the protein strain is not surprising, given the complexity of the protein potential energy surface compared to that of the ligand.

We have previously shown that calculated total energies and resulting protein structures vary greatly after a geometry optimization,[23] and the effect is dependent on the size of the protein; this is in agreement with current observations. In summary we can conclude that much of the unacceptably large binding energy variability in flexible protein MM-GBSA calculations could be eliminated if protein strain is excluded from the calculations, although the value of this approach would obviously be questionable.

To gain better insight into the variations in minimized protein structures, per residue RMSDs were also evaluated. The average and maximum variations in residue placement for 1z95 and 1ke9 after MM-GBSA optimization are plotted vs residue number for the 4, 10, and 16 Å flexible regions (Figures 8 and 9). The pocket residues (all within 4.5 Å of the ligand) are marked with dots. The plots show the general increase in residue variation as the size of the flexible region increases. The degree of variation is also not uniform, with some residues exhibiting more variation than others. Inspecting individual residues reveals that hotspots for variations are almost always on the protein surface. In many cases it is easy to understand the movement, e.g. residues near termini can move more freely, and there can be significant movement of charged groups near

dx.doi.org/10.1021/ci300298d | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

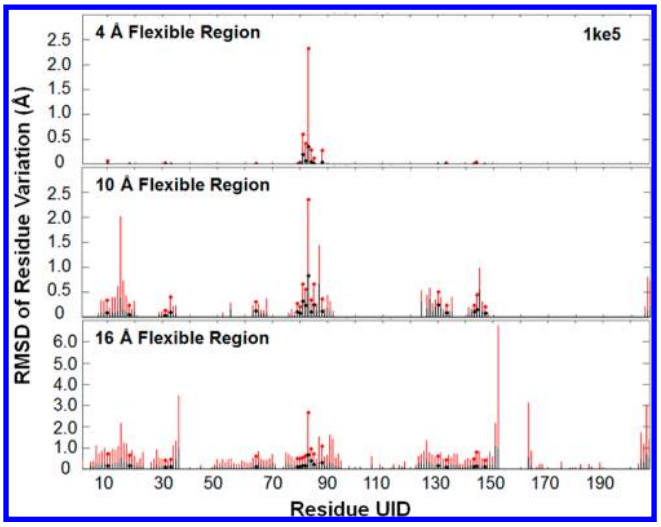


Figure 8. Variation in protein residue positions of MM-GBSA optimized 1ke5 structures. The average (in black) and maximum (in red) variations in residue position (in Å) of MM-GBSA optimized structures are given on a per-residue basis for three different sized regions of protein flexibility −4 Å, 10 Å, and 16 Å. The pocket residues are denoted with small diamonds in the plot. If there were no output variations in residue positions, these plots would all be at zero.

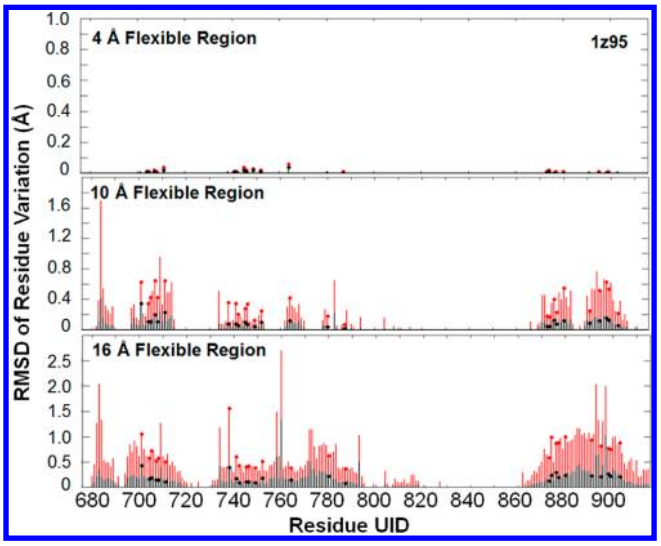


Figure 9. Variation in protein residue positions of MM-GBSA optimized 1z95 structures. The average (in black) and maximum (in red) variations in residue position (in Å) of MM-GBSA optimized structures are given on a per-residue basis for three different sized regions of protein flexibility −4 Å, 10 Å, and 16 Å. The pocket residues are denoted with small diamonds in the plot. If there were no output variations in residue positions, these plots would all be at zero.

the surface (e.g., Arg, Glu, and Asp residues). Significant movements can also occur when residues are in exposed loops, such as Gly683 and Val684 in 1z95, whose positions can vary by up to 2 Å. In most cases, a range of residues flanking these hotspots also moves by a significant amount.

One interesting case is HIS84 in 1ke5, which exhibits significant variation, even with the 4 Å flexibility radius. This is because although the backbone of the residue points into the pocket, the side-chain is exposed to solution and adopts different distinct rotamers as a result of the ligand input perturbation. Interestingly, the placement variation of this residue does not change much when the flexibility radius is increased to 16 Å, presumably because its full range of variation has already been explored at the 4 Å flexibility radius.

Relationship between Input Perturbation and MM-GBSA Optimized Structures. The histograms in Figures 3 and 4 demonstrate that ΔG and RMSD variations in the 1p62 and 1unl systems arise from the production of two or more distinct minima, rather than from a normal distribution of solutions around a single minimum. This suggests the cognate ligand geometry is located in a cusplike region on the MM-GBSA optimization surface, where simulation outcomes are very sensitive to initial conditions. To better visualize this sensitivity, plots of the final MM-GBSA solution as a function of x and y coordinate perturbations were made for the 1p62 (Figure 10) and 1unl (Figure 11) systems. The cognate ligand input structure is at the center of these plots (dX, dY = 0), while other points represent input structures with varying dX and dY translations. The color of a point in the plot indicates which distinct minimum was produced by that input structure. If all input structures optimized to the same minimum, the plot would be one solid color. In well-behaved systems with multiple distinct minima, the plot would consist of contiguous regions of points that optimize to each of the distinct minima, with smooth boundaries between these regions. In chaotic systems, regions that optimize to different minima can be noncontiguous, and boundaries between the regions can have fractal qualities.[37] The plots in Figures 10 and 11 clearly show that input regions that optimize to a given cluster of MM-GBSA solutions are not contiguous, and boundaries between regions appear to have fractal properties in places. The plots resemble both plots of fractal basin boundaries reported for simple MM minimizations[38] as well as plots of docking output as a function of input torsions angles we reported in a recent publication.[22]

The plot in Figure 10 shows that while most input structures optimize to the *Cluster1* minimum (*region 1*), a significant portion optimize to the *Cluster2* minimum (*region 2*). Furthermore, the cluster 1 input structures are not localized in one region of the plot but instead form a complex pattern in input space, with discontinuous, fractal-like boundaries between points that optimize to either cluster 1 or cluster 2. The plot in Figure 11 shows three regions that optimize to three distinct clusters in MM-GBSA solution space. The Cluster 1 region is mostly contiguous and in many places has smooth nonfractal boundaries with other regions. This suggests input structures which optimize to Cluster 1 solutions could potentially be identified by ranges of dX and dY values. In contrast, regions that optimize to Cluster 2 and Cluster 3 structures are intertwined and discontinuous, with fractal-like barriers between them. Unlike the Cluster 1 region, it would be difficult to define simple ranges of (dX,dY) values that separate input structures which optimize to either cluster 2 or cluster 3 solutions.

As noted before, although most of the studies in this work were performed using Prime MM-GBSA, we found input sensitivity with all other programs we tested for calculating binding energies. In order to establish whether similar fractal-like behavior may also be present in other programs, the above 1unl example was reproduced using the MOE implementation of the same method (MM-GB/VI). The resulting plot is presented in Figure 12. Although the effect in Figure 12 is not as striking as some of the previous examples, the plot clearly shows that MM-GB/VI optimizations in MOE can also

dx.doi.org/10.1021/ci300298d | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

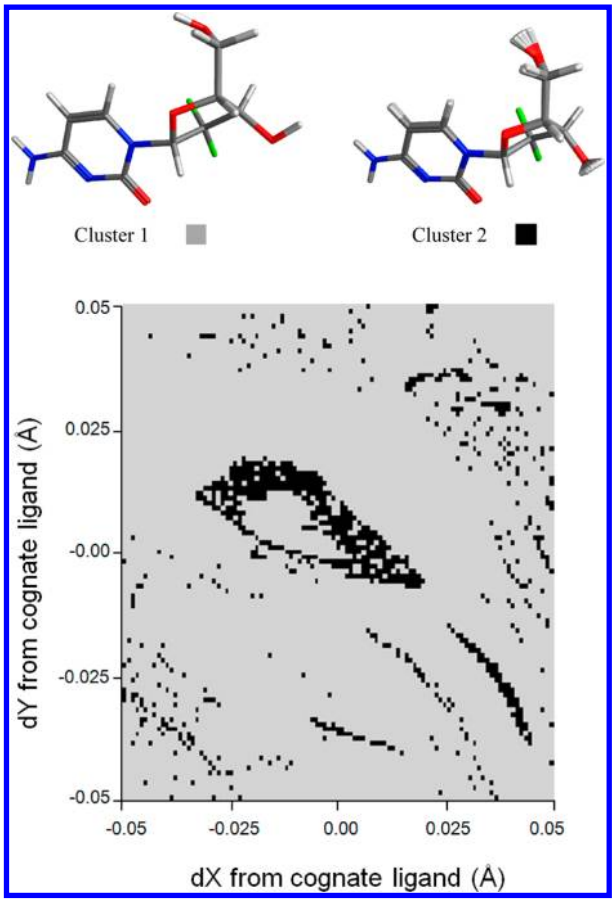


Figure 10. Cluster membership of MM-GBSA optimized 1p62 ligand structures from an input grid. The cognate ligand of the 1p62 (dck) structure was translated by ±0.05 Å in 100 steps along both the x- and y-directions, leading to 10201 starting geometries (i.e., 101 × 101 matrix). MM-GBSA calculations were initiated from each of these points, and the cluster membership of the minimized structure was established from the conformation. The cluster membership of the optimized ligand is plotted as a function of the (dX,dY) input perturbation. Cluster 1 (shown as gray squares) has 9334 members and covers the binding energy range −71.06 to −70.68 kcal/mol (with ligand strain, −73.55 to −73.16 kcal/mol without). Cluster 2 (shown as black squares) has 867 members and covers the binding energy range −67.8 to −59.8 kcal/mol (with ligand strain, −69.7 to −65.0 kcal/mol without). As can be seen from the overlaid structures of both clusters, the geometry of cluster 1 is well-defined, whereas three hydrogen positions (and corresponding energies) are somewhat dispersed in cluster 2. Furthermore, the regions of the plot which optimize to cluster 1 and cluster 2 form discontinuous regions with fractal-like boundaries between regions.

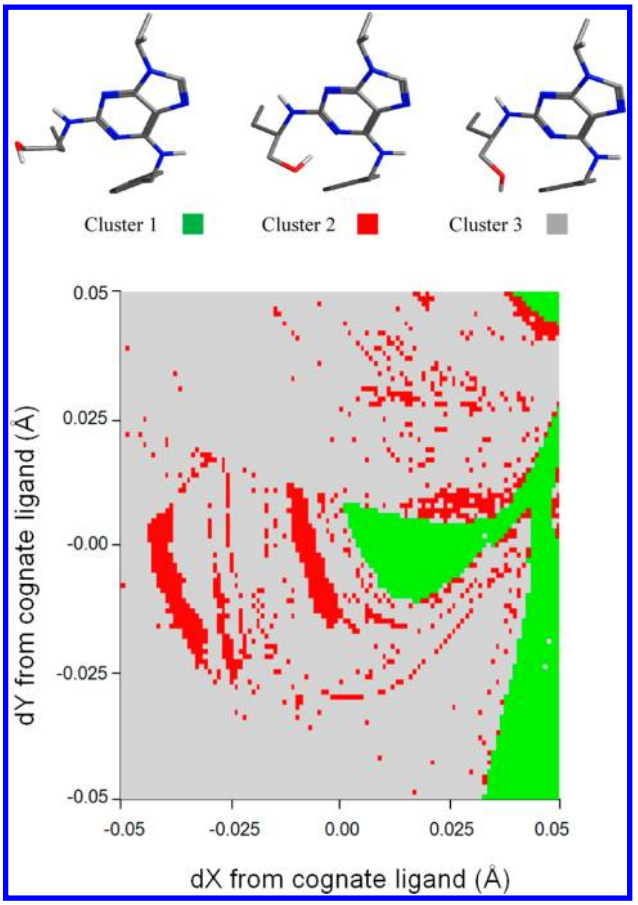


Figure 11. Cluster membership of MM-GBSA optimized 1unl ligand input structures from an input grid. The cognate ligand of the 1unl (cdk5) structure was translated by ±0.05 Å in 100 steps along both the x- and y-directions, leading to 10201 starting geometries (i.e., 101 × 101 matrix). MM-GBSA calculations were initiated from each of these points, and the cluster membership of the minimized structure was established from the conformation. The cluster membership of the optimized ligand is plotted as a function of the (dX,dY) input perturbation. Cluster 1 (shown as green squares) has 1209 members and covers the binding energy range −71.64 to −68.15 kcal/mol (with ligand strain, 76.59−76.10 kcal/mol without). Cluster 2 (shown as red squares) has 1043 members and covers the binding energy range −58.45 to −58.08 kcal/mol (with ligand strain, −64.79 to −64.54 kcal/mol without). Cluster 3 (shown as gray squares) has the largest population with 7949 members, covering a binding energy range of −54.32 to −52.22 kcal/mol (with ligand strain, −66.70 to −65.85 kcal/mol without). The cognate ligand also minimizes to this cluster. It can be seen from overlaid clusters that all three clusters are well-defined. Regions of the plot which optimize to cluster 1, 2, or 3 form many discontinuous regions with fractal-like boundaries between them, reflecting the extreme sensitivity of these calculations to initial conditions. There are also large contiguous regions in each cluster, representing regions of input space where the optimization is 'well-behaved', and similar input structures all converge to the same cluster.

produce different minima from very small input perturbations. There are three poses present among the solutions, with Cluster 1 representing the majority (>99%) of the solutions. The heavy atom positions in this cluster are nearly identical to the cognate ligand (RMSD 0.3 Å), but one of the hydrogen atoms is involved in a different hydrogen bond (Gln130 instead of Asp86 in the cognate ligand as seen after using the Protonate3D process on the latter). The calculated binding energies within these clusters differ by up to 4.5 kcal/mol, somewhat less than with MM-GBSA variation of up to 17 kcal/mol but still of significant size to affect binding energy predictions. Note that the absolute MM-GBVI binding energies (∼−25 kcal/mol) are closer in magnitude to experimental binding energies that the MM-GBSA energies but are still too large when compared with experimental values. On a percentage basis the errors in MM-GBSA and MM-GBVI energies are similar; the 4.5 kcal/mol binding energy variations in MM-GBVI represent ∼18% of the binding energy, while the 17 kcal/mol binding energy variations in MM-PBSA represent ∼22% of the total binding energy. Furthermore, regardless of their different magnitudes, these output variations are, in general, to be expected; although some portion of observed input sensitivity may be due to peculiarities in a given software

dx.doi.org/10.1021/ci300298d | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

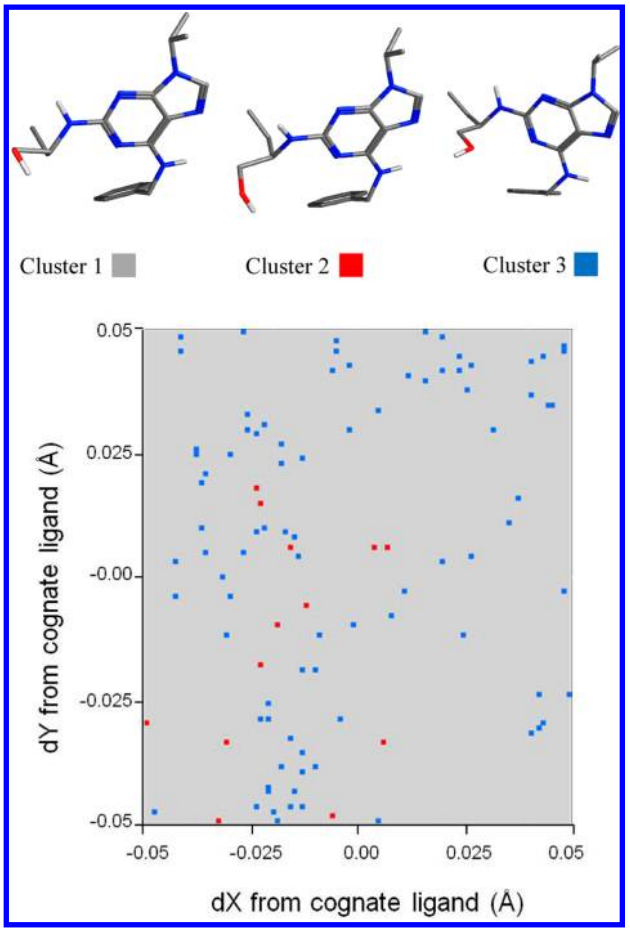


Figure 12. Cluster membership of MOE MM-GB/VI optimized 1unl ligand input structures from an input grid. The input was identical to that used to generate Figure 11. The cognate ligand of the 1unl (cdk5) structure was translated by ±0.05 Å in 100 steps along both the x- and y-directions, leading to 10201 starting geometries (i.e., 101 × 101 matrix). MM-GB/VI calculations were initiated from each of these points, and the cluster membership of the minimized structure was established from the conformation. The cluster membership of the optimized ligand is plotted as a function of the input perturbation. Cluster 1 (shown as gray squares) has 10101 members and is around a binding energy of −25 kcal/mol (the cognate ligand also minimizes to this cluster). Cluster 2 (shown as red squares) has 13 members and has a binding energy of around −22.4 kcal/mol. Cluster 3 (shown as green squares) has 87 members with a binding energy of around −20.7 kcal/mol.

package, a large contributor to input sensitivity is the nature of the mathematical problem and how it is handled in digital computers.

CONCLUSIONS

This work clearly shows that the results of minimization-based binding energy calculations that depend on optimizing a single complex, such as the MM-GBSA and MM-GB/VI calculations studied in this work, can be highly dependent on tiny input perturbations. Perturbations to input coordinates as small as 0.01 Å can give rise to optimized solutions that differ by as much as 0.8 Å in atom placement and up to 17 kcal/mol in binding energy. These large output variations are unexpected based on the size of the input perturbations and reflect the binding energy variations observed by Weis et al.[15] and Genhenden et al.[24] In many cases the output variations lead to different ligand poses with different binding energies and hydrogen-bonding patterns. Optimization of highly similar structures shows that the barriers between different poses can appear fractal-like, making it difficult to predict which solution will be produced from a given input. Observing output variations from atom shuffling highlights the fact that these variations arise from small initial numerical differences which, over the course of a simulation, cause the minimization trajectories to diverge. It is worth noting that these large errors were obtained by assuming a 'perfect' protein structure, whereas in reality experimental structures have limited resolution and a large percentage of them include major errors.[39,40] If the binding model applied is inaccurate, significant further errors in the calculated binding energy are expected.[41]

The large variations in computed binding energies that result from allowing part of the protein to be flexible suggest that protein flexibility should be included with extreme caution, if at all, in MM-GBSA calculations based on the minimization of a single complex structure. If protein flexibility is included, the size of the flexible region will need to be kept small, definitely less than 8 Å. The breakdown of binding energy variations shown in Figure 6 suggests that protein strain energy terms should be *excluded* from binding energy calculation if protein flexibility is allowed, as they are by far the largest source of variation in the output binding energy.

Since running MM-GBSA calculations based on a single minimized complex can result in large and unpredictable variations in pose and binding energy, it seems prudent to run multiple copies of near-identical input to produce a range of values for a binding energy prediction. This approach will better ensure more relevant minima are found (if there is more than one) and less low-energy minima are missed. Examining the degree of variation in the results will also give an indication of the stability of calculation and possibly an estimate of the error in the calculation.

Although it may seem sensible to assume that MM-GBSA optimizations initiated from different starting points would all converge to more-or-less the same optimal binding pose and score, this study shows that this is clearly not so. With the 1unl example, it was shown that the solution corresponding to optimizing the cognate ligand was >10 kcal/mol above the lowest energy solution found by other optimizations initiated from near-identical starting points. The reason for the large energy differences between similar solutions probably lies in the misprediction of strong hydrogen bond energies; it appears that the largest energy variations are observed when there is a possibility for multiple poses with near-identical patterns of strong hydrogen bonds. One possible solution to reduce variations from this effect would be to improve the estimates of these hydrogen bond energies. Furthermore, since we do not know *a priori* the correct orientation of the hydrogens, another approach using current hydrogen bond estimates would be to calculate a weighted average of energies from different simulations to produce a more statistically converged energy estimate. Further studies will be needed to establish whether this will contribute to better correlations between MM-GBSA binding energies and experimental binding affinities. As a final note, although MM-GBSA was the focus of this study, we must emphasize that we believe *any* binding energy calculation that includes ligand and/or target minimizations could potentially display significant output variations when faced with minor input perturbations. Regardless of the software being used, researchers performing binding energy calculations should be

dx.doi.org/10.1021/ci300298d | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

aware of output variations and design their experiments and make their conclusions accordingly.

AUTHOR INFORMATION

Corresponding Author

*Phone: +1 416 581 7611. E-mail: mfeher@uhnres.utoronto.ca.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Chris Williams would like to thank Martin Santavy of Chemical Computing Group for many detailed discussions about numerical processes and their effect on the reproducibility of digital calculations.

REFERENCES

(1) Singh, N.; Warshel, A. Absolute binding free energy calculations: on the accuracy of computational scoring of protein−ligand interactions. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1705−1723.

(2) Ruiter, A.; Oostenbrink, C. Free energy calculations of protein−ligand interactions. *Curr. Opin. Chem. Biol.* **2011**, *16*, 547−552.

(3) Chodera, J. D.; Mobley, D. L.; Shirts, M. L.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.

(4) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Basic ingredients of free energy calculations: a review. *J. Comput. Chem.* **2010**, *31*, 1569−1582.

(5) de Ruiter, A.; Oostenbrink, C. Free energy calculations of protein−ligand interactions. *Curr. Opin. Chem. Biol.* **2011**, *15*, 547−552.

(6) Michel, J.; Essex, J. W. Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 639−658.

(7) Shirts, M. R.; Mobley, D. L.; Brown, S. P. Free energy calculations in structure-based drug design. Merz, K. M., Ringe, D., Reynolds, C. H., Eds.; *Drug Design: Structure- and Ligand-Based Approaches*; Cambridge University Press: 2010; pp 61−86.

(8) Deng, Y.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.

(9) Knight, J. L.; Brooks, C. L. λ-Dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30*, 1692−1700.

(10) Michel, J.; Essex, J. W. Hit identification and binding mode predictions by rigorous free energy simulations. *J. Med. Chem.* **2008**, *51*, 6654−6664.

(11) Thompson, D. C.; Humblet, C.; Joseph-McCarthy, D. Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.* **2008**, *48*, 1081−1091.

(12) Nu, H.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-based scoring of protein-ligand complexes: enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 243−253.

(13) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(14) Molecular Operating Environment (MOE), 2011.10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2011.

(15) Weis, A.; Katebzadeh, K.; Soderhjelm, P.; Nilsson, I.; Ryde, U. Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* **2006**, *49*, 6596−6606.

(16) Guimarães, C. R. W.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* **2008**, *48*, 958−970.

(17) Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. Rescoring docking hit lists for model cavity sites: predictions and experimental testing. *J. Mol. Biol.* **2008**, *377*, 914−934.

(18) Kawatkar, S.; Wang, H.; Czerminski, R.; Joseph-McCarthy, D. Virtual fragment screening: an exploration of various docking and scoring protocols for fragments using Glide. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 527−539.

(19) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* **2006**, *49*, 4805−4808.

(20) Feher, M.; Williams, C. I. Effect of input differences on the results of docking calculations. *J. Chem. Inf. Model.* **2009**, *49*, 1704−1714.

(21) Feher, M.; Williams, C. I. Reducing docking score variations arising from input differences. *J. Chem. Inf. Model.* **2010**, *50*, 1549−1560.

(22) Feher, M.; Williams, C. I. Numerical errors and chaotic behavior in docking simulations. *J. Chem. Inf. Model.* **2012**, *52*, 724−738.

(23) Williams, C. I.; Feher, M. The effect of numerical error on the reproducibility of molecular geometry optimizations. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 39−51.

(24) Genheden, S.; Ryde, U. How to obtain statistically converged MM/GBSA results. *J. Comput. Chem.* **2009**, *31*, 837−846.

(25) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.* **2007**, *50*, 726−741.

(26) *Prime*, version 3.0; Schrödinger, LLC: New York, NY, 2011.

(27) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(28) Ghosh, A.; Rapp, C. S.; Friesner, R. A. Generalized Born model based on a surface integral formulation. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.

(29) Li, J.; Abel, R.; Zhu, K.; Cao, Y.; Zhao, S.; Friesner, R. A. The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 2794−2812.

(30) Labute, P. The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J. Comput. Chem.* **2008**, *29*, 1693−1698.

(31) *Discovery Studio*, version 3.1; Accelrys Inc., San Diego, CA.

(32) *MacroModel*, version 9.9; Schrödinger, LLC, New York, NY, 2011.

(33) He, Y.; Ding, C. H. Q. Using accurate arithmetics to improve numerical reproducibility and stability in parallel applications. *J. Supercomput.* **2001**, *18*, 259−277.

(34) Blackford, L. S.; Cleary, A.; Petitet, A.; Whaley, R. C.; Demmel, J.; Dhillon, I.; Ren, H.; Stanley, K.; Dongarra, J.; Hammarling, S. Practical experience in the numerical dangers of heterogeneous computing. *ACM Trans. Math. Software* **1997**, *23*, 133−147.

(35) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9997−10002.

(36) Schrodinger knowledge base, article ID 1647. http://www.schrodinger.com/kb/1647 (accessed June 18, 2012).

(37) McDonald, S. W.; Gregbogi, C.; Ott, E.; Yorke, J. A. Fractal basin boundaries. *Physica D* **1985**, *17*, 125−153.

(38) Xu, Y. Z.; Ouyang, Q.; Wu, J. G.; Yorke, J. A.; Xu, G. X.; Xu, D. F.; Soloway, R. D.; Ren, J. Q. Using fractal to solve the multiple minima problem in molecular mechanics calculation. *J. Comput. Chem.* **2000**, *21*, 1101−1108.

(39) Davis, A. M.; St-Gallay, S. A.; Kleywegt, G. J. Limitations and lessons in the use of X-ray structural information in drug design. *Drug Discovery Today* **2008**, *13*, 831−841.

(40) Søndergaard, C. R.; Garrett, A. E.; Carstensen, T.; Pollastri, G.; Nielsen, J. E. Structural artifacts in protein-ligand X-ray structures: implications for the development of docking scoring functions. *J. Med. Chem.* **2009**, *52*, 5673−5684.
(41) Singh, M. K.; Dominy, B. N. Thermodynamic resolution: how do errors in modeled protein structures affect binding affinity predictions? *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1613−1617.