

Published in final edited form as:

J Chem Theory Comput. 2011 March 8; 7(3): 790–797. doi:10.1021/ct100563b.

Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein-ligand Complexes

John C. Faver^a, Mark L. Benson^a, Xiao He^a, Benjamin P. Roberts^a, Bing Wang^a, Michael S. Marshall^b, Matthew R. Kennedy^b, C. David Sherrill^b, and Kenneth M. Merz Jr.^{a,*}

^a Quantum Theory Project, The University of Florida. 2328 New Physics Building P.O. Box 118435. Gainesville, FL 32611-8435

^b Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry, and School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0400

Abstract

A largely unsolved problem in computational biochemistry is the accurate prediction of binding affinities of small ligands to protein receptors. We present a detailed analysis of the systematic and random errors present in computational methods through the use of error probability density functions, specifically for computed interaction energies between chemical fragments comprising a protein-ligand complex. An HIV-II protease crystal structure with a bound ligand (indinavir) was chosen as a model protein-ligand complex. The complex was decomposed into twenty-one (21) interacting fragment pairs, which were studied using a number of computational methods. The chemically accurate complete basis set coupled cluster theory (CCSD(T)/CBS) interaction energies were used as reference values to generate our error estimates. In our analysis we observed significant systematic and random errors in most methods, which was surprising especially for parameterized classical and semiempirical quantum mechanical calculations. After propagating these fragment-based error estimates over the entire protein-ligand complex, our total error estimates for many methods are large compared to the experimentally determined free energy of binding. Thus, we conclude that statistical error analysis is a necessary addition to any scoring function attempting to produce reliable binding affinity predictions.

Keywords

Scoring; Protein-Drug Interaction; Error Analysis; Force Fields; Electronic Structure

Introduction

One of the most challenging problems in computational biochemistry is the accurate prediction of the binding affinity of a small ligand to a protein receptor.^{1,2} This problem is challenging because of the role active site water plays in the binding process³, the large

*Corresponding Author: Kenneth M. Merz Jr. 2328 New Physics Building P.O. Box 118435 Gainesville, FL 32611-8435. Phone: 352-392-6973 Fax: (352) 392-8722 merz@qtp.ufl.edu.

Supporting information

The supporting information packet contains the information available on our dedicated website for the project at <http://crunch.qtp.ufl.edu/~faver/error/1hsg>. The packet contains an index of the fragment-based interactions, followed by individual pages for each interaction containing (1) a visual representation of the system, and (2) a table of calculated interaction energies from various computational methods. It also contains a table of error statistics over all interactions, and a table of histograms and error probability density functions plotted for each computational method studied. This information is available free of charge via the Internet at <http://pubs.acs.org>

number of conformational states of the protein and ligand⁴, tautomeric states of the ligand⁵, and the computation of accurate interaction energies for each of the hydrogen bonding and non-polar contacts at the protein-ligand interface.⁶ In order to compute an accurate binding free energy each of these factors needs to be reliably modeled to chemical accuracy (± 1 kcal/mol relative to experiment) or beyond.⁶ Reaching acceptable levels of accuracy in the computational modeling of large bio-molecular systems is indeed a challenging effort. Dill has even proposed that a protein with 100 amino acid residues should have each one modeled to within 0.1 kcal/mol accuracy to yield an acceptable level of error for the entire protein.⁷

Over the years, many different types of scoring functions have been developed and used to model protein-ligand interactions, ranging from empirical scores to physics-based functions with some even incorporating quantum mechanical calculations.^{8–11} The most popular docking and scoring algorithms have been extensively reviewed, and their success rates are variable across different score functions and different protein-ligand systems.^{2,12–16} Because of this inconsistency, the notion of consensus scoring, in which several different docking/scoring programs are used to collectively rank ligands by binding affinity, emerged.^{17,18} Consensus scoring generally yields better ordering of ligands by binding affinity, largely due to the cancellation of errors from each individual scoring function.¹⁹ Free-energy perturbation methods (FEP), which are often used to predict differences in binding affinity between two similar ligands, also have been successful largely due to the cancellation of errors.^{20–23}

The evaluation of error in docking and scoring studies, and indeed, in computational biology in general, has been rather limited. Nonetheless, error analysis can be used to clarify the ability of a method to reach a defined outcome (e.g. the computation of relative free energies of binding). For example, simple statistical error analysis can provide insight into how computed binding affinity errors are diminished in consensus scoring and how relative free energies obtained from free-energy perturbation calculations have reduced uncertainties. More importantly, it can form a hypothesis by which improvements in the desired outcome can be realized.

Errors involved in any kind of calculation or measurement can be categorized as being systematic or random. Systematic errors are predictable in both sign and magnitude, while random errors are unpredictable in both sign and magnitude. It is important to note that the propagation of these types of errors differs in an important way - systematic errors accumulate as a simple sum (Equation 1) while random errors propagate as the square root of the sum of squares²⁴ (Equation 2). The systematic error sum over individual sources of error is simple to correct for, since it represents a predictable shift in the measured (or calculated) value from the true value. Furthermore, systematic errors largely vanish (due to error cancellation) when differences in measurements are calculated. On the other hand, the random error of equation 2 represents a maximum accumulation of individual random error sources. Large random error is a consequence of using a very imprecise method of measurement.

$$Error_{Systematic} = Err_1 + Err_2 + Err_3 + \dots \quad (1)$$

$$Error_{Random} = \left[(Err_1)^2 + (Err_2)^2 + (Err_3)^2 + \dots \right]^{1/2} \quad (2)$$

Another important tool in statistical error analysis is the construction of probability density functions (pdf's). A series of repeated measurements can be used to construct a pdf for a given type of measurement, with which one is able to infer characteristics such as the mean, standard deviation, variance, skewness, and other properties. For example, the widely applicable Gaussian distribution (Equation 3) can be used to make inferences about both systematic and random errors. The mean error (μ) in a set of measurements represents a systematic shift in the pdf while the variance about the mean error (σ^2) determines the "width" of the distribution due to the accumulation of random errors.

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

With these ideas in mind, the success of consensus scoring can readily be rationalized in the terminology of error analysis. As more scoring functions are included in the consensus score, the predicted value of the binding affinity for a given ligand approaches the mean value across all scoring function predictions. As we approach the mean binding affinity, random error from individual scoring functions is minimized, systematic errors largely cancel, and ranking ligands by binding affinity is much more successful. When free energy perturbation methods are used to calculate relative free energies of binding, usually only one functional group interaction is modified between the ligands being compared. This essentially reduces the dimensionality of the error sources to one, which decreases random error. Since these methods only calculate differences in energies, the random error in the final answer is usually very small, and systematic errors mostly cancel.

An application of simple error analysis to the calculation of interaction energies of protein-ligand systems has been outlined in a Gedanken experiment by Merz.⁶ It involves decomposing a protein-ligand complex into independent, chemically meaningful fragment pairs. For example, a ligand may have a hydroxyl moiety hydrogen bound to the backbone carbonyl of a residue in the binding pocket of the receptor, while also containing a hydrophobic carbon chain several angstroms away exhibiting favorable van der Waals interactions with a phenylalanine side chain.²⁵ These are clearly distinguishable interacting fragments both contributing to the total binding affinity of the whole ligand. However, it is well known that functional groups are not completely independent entities with respect to total free energy of binding. The concept of cooperativity (also termed additivity/non-additivity) has been described and examined in several protein-ligand systems.^{7,26} The free energy changes of individual interacting fragments cannot simply be summed to produce a global free energy of binding for a protein-ligand complex. However, evidence from statistical mechanics²⁶ and isothermal titration calorimetry experiments²⁷ suggests that enthalpy changes *are* additive across several interactions in a protein-ligand complex. Even so, in some cases strong secondary electrostatic interactions can arise in systems with multiple proximally close polar groups (DNA base pairs²⁸, for example) and can introduce non-additivity in the computation of enthalpy of interaction.

As discussed in Merz's Gedanken experiment and elsewhere, through the use of the thermodynamic cycle outlined in Figure 1 the free energy change of a ligand upon binding to a protein receptor can be obtained. The final expression (equation 4) contains an electronic interaction energy term, an enthalpy correction term, an entropy term, and a change in the solvation free energy term.

$$\Delta G_b^S = \Delta E + \Delta H_{corr} - T\Delta S + \Delta \Delta G_{Solv} \quad (4)$$

$$\Delta E = E_{PS} - E_P - E_S \quad (5)$$

$$\Delta H = H_{PS} - H_P - H_S \quad (6)$$

$$T\Delta S = T\Delta S_{PS} - T\Delta S_P - T\Delta S_S \quad (7)$$

$$\Delta \Delta G_{Solv} = \Delta G_{Solv}^{PS} - \Delta G_{Solv}^P - \Delta G_{Solv}^S \quad (8)$$

The present work concentrates only on errors from the electronic interaction energy, ΔE . In doing so we assume that the errors from ΔH_{corr} , ΔS , and $\Delta \Delta G_{Solv}$ are zero (which is certainly not the case, but serves as a computational expediency). We refer to such an error estimate as the best-case scenario error, BCS_{error} , because it ignores the error from these three remaining terms of equation 4. For the sake of simplicity, we assume here as a first approximation that errors in the calculated interaction energies are purely random (even for variational quantum-based methods since the variational principle does not apply to differences between ground state energies). The random errors in the calculation of interaction energies of chemical fragments 1, 2, and 3, *etc.* now propagate as

$$BCS_{error} = \left[(\Delta E_{calc}^1 - \Delta E_{ref}^1)^2 + (\Delta E_{calc}^2 - \Delta E_{ref}^2)^2 + (\Delta E_{calc}^3 - \Delta E_{ref}^3)^2 + \dots \right]^{1/2} \quad (9)$$

By neglecting the random error from the enthalpy, entropy, and solvation energies, the overall propagated error (Equation 9) is a lower bound to the free energy error estimate and therefore can be thought of as the best-case scenario estimate. In addition to calculating the BCS_{error} , we also decompose errors from computational methods into systematic and random portions by using Gaussian error probability density functions. By using error probability density functions, we show that it is possible to remove the estimated systematic error for each interaction and thereby decrease the estimation of remaining random error.

It is well known that enthalpy and entropy tend to act against one another in free energy calculations in a phenomenon called enthalpy-entropy compensation. These two terms often have opposite signs and similar magnitudes and thus largely cancel each other in Equation 4. In terms of error estimation, however, the magnitudes of random errors in each term are not known. We can only estimate their probable range based on studies such as the current one. When calculating overall random error bars in the free energy, estimated errors from the individual terms must be propagated according to equation 2, which only increases with the addition of terms. In addition, while it is true that the errors in any of the three remaining terms of Equation 4 could cancel favorably with a portion of the predicted error in ΔE , by exploiting this we would be calculating the (sometimes) correct ΔG_{bind} for clearly the wrong reasons. The goal of molecular docking is to routinely predict binding free energies both

accurately and precisely, so the error should be diminished in each of the terms of equation 4. As a side note, any calculation of the entropy term depends on the potential energy surface of the system and thus depends on enthalpy measurements. Because of this, errors in enthalpy estimations can lead to distorted potential energy surfaces and have unpredictable effects on the entropy term.

Clearly, experimental interaction energies/enthalpies would be desirable to use as a reference for these error functions, but in most cases this will not be possible. As a substitute, we use converged quantum mechanical methods such as CCSD(T) with complete basis set extrapolations (CBS) to provide the reference energy we need to make meaningful error estimates.^{29,30} Such methods have been known to routinely achieve what is termed “chemical accuracy” or the ability to compute energies to within ± 1 kcal/mol of experimental observations.³¹

Prior to presenting the results it is worth hypothesizing what the expected error types would be for computed interaction energies of hydrogen bonded and non-polar complexes at different levels of theory. Highly parametric methods like semiempirical or force field methods would be expected, on the face of it, to have very significant random errors rather than systematic errors. On the other hand, quantum chemical methods like Hartree-Fock (HF) or density functional theory (DFT), which do not correct for dispersion, would be expected to have very large systematic errors and relatively smaller random errors. One might suppose that more sophisticated correlated methods such as second-order perturbation theory (MP2) would have smaller systematic errors than HF or DFT; however, MP2 is known to exhibit significant overbinding of dispersion-bound complexes (especially with larger basis sets), and hence MP2 may also exhibit significant systematic errors. Because MP2 performs better for hydrogen-bonding interactions than for dispersion-bound interactions, these errors may also appear to have a significant random component. Interestingly, our analysis shows that systematic errors are a significant component in all cases, which arises because all methods examined tend to give too weak interaction energies with respect to our reference level of theory. If the computed interaction energy errors were better centered at the zero error point then error cancellation would possibly be beneficial in the computation of the total interaction energy, but this was found to not be the case for the model systems examined. The good news is that the systematic error can be estimated and corrected for while reducing the remaining random error and this then becomes the main challenge in improving the prediction of interaction energies.

In order to demonstrate our error hypothesis on a real system, we have chosen to examine the HIV-II/indinavir crystal structure³² in detail (PDBID: 1HSG, see Figure 2). This system was chosen because of its large number and diversity of chemical contacts. In addition, the experimentally observed binding free energy has been reported^{33,34} to be -12.8 kcal/mol indicating an overall strong level of interaction between the protein and indinavir. We have decomposed the protein-ligand complex into 21 distinct, chemically important fragment-based interactions (Figure 3). Each pair of interacting fragments was evaluated in terms of gas-phase electronic interaction energies with a number of different computational methods including force field methods, semiempirical quantum mechanical methods, density functional methods, and *ab initio* wave function methods. By comparing the calculated interaction energy of each fragment pair for a given method to a reference, one can calculate the BCS_{error} and generate pdf's to describe the error for the full protein-ligand interaction.

Methods

The HIV II-protease/indinavir complex structure was obtained from the protein databank (PDBID: 1HSG).³² Hydrogen atoms were added to the structure with the program Reduce³⁵,

followed by optimization of their positions with the AMBER FF99SB force field.³⁶ Close contacts (less than 3 Å) in the resulting structure were highlighted in the visualization program Chimera³⁷, which were then manually examined. In many cases, one member of an interacting fragment pair would be adjacent to a polarizing group such as the peptide bond along the protein backbone. These polarizing groups, although not directly interacting with the other fragment partner, were included in the fragment structures to model the polarization effect on the fragments. Any time an aromatic fragment was involved the entire aromatic ring was included.

After manually determining all the chemically important fragment pairs, the fragments were generated by cutting the covalent bonds joining them to the remainder of the molecule, and replacing the bonds with linker hydrogen atoms. The hydrogen bond distances were set to 1.1 Å for carbon-hydrogen bonds and 1.0 Å for nitrogen-hydrogen bonds. The resulting 21 interacting fragment pair structures are displayed in Figure 3 and their interaction energies can be found in the supplementary material, available free of charge via the Internet at <http://pubs.acs.org>, or for an interactive version visit <http://crunch.qtp.ufl.edu/~faver/error/1hsg>.

The gas-phase electronic interaction energies of the 21 independent chemically meaningful fragment complexes (Figure 3) were calculated using the following computational methods: the generalized AMBER force field (GAFF)³⁸, AMOEBA³⁹, MMFF⁴⁰, AM1⁴¹, AM1-FS1⁴², PM3⁴³, PM6⁴⁴, PM6-DH2⁴⁵, PDDG⁴⁶, Hartree-Fock (HF), second-order Møller-Plesset perturbation theory (MP2), the M06-L density functional⁴⁷, B97-D⁴⁸, and coupled-cluster²⁹ with single, double, and perturbative triple excitations (CCSD(T)). Basis sets for the quantum mechanical methods ranged from the 6-31G* Pople type basis set⁴⁹ to the aug-cc-pVXZ (X=D,T,Q - hereafter referred to as aXZ) Dunning type basis sets.⁵⁰ In addition, extrapolation to the complete basis set (CBS) limit was also used.³⁰ The CCSD(T) calculations were evaluated using the “heavy augmented” haDZ and haTZ basis sets, which place diffuse functions only on heavy (non-hydrogen) atoms.^{51,52} In our error analysis, the CCSD(T)/CBS energies were used as our reference, which were evaluated as

$$\Delta E_{\text{CCSD(T)/CBS}} = \Delta \text{CCSD(T)} + \Delta E_{\text{MP2/CBS[aTZ-aQZ]}} \quad (10)$$

$$\Delta \text{CCSD(T)} = \Delta E_{\text{CCSD(T)/CBS[haDZ-haTZ]}} - \Delta E_{\text{MP2/CBS[haDZ-haTZ]}} \quad (11)$$

where the notation CBS[aTZ-aQZ] represents the complete basis set extrapolation from energies calculated with the aTZ and aQZ basis sets. In each quantum-based method, basis set superposition error (BSSE) was accounted for using the counterpoise correction. The Hartree-Fock and MP2 interaction energies were calculated with the Gaussian 03 program.⁵³ The M06-L and PM6 interaction energies were calculated with Gaussian 09.⁵⁴ AM1, PM3, and PDDG interaction energies were evaluated with the DivCon program.⁵⁵ The GAFF energies were calculated using the AMBER 10 suite.⁵⁶ Most of the CCSD(T) computations were performed using MOLPRO.⁵⁷ The larger CCSD(T) computations were performed using NWChem⁵⁸ on the Jaguar supercomputer at Oak Ridge National Laboratory. MMFF calculations were performed with the Schrödinger suite.⁵⁹

Results and Discussion

Our analysis of various error scoring functions for the set of computational methods across all fragment pairs is shown in Table 1. The second column of Table 1, the BCS random error

bar, is the evaluation of equation 9 and gives a measure of error assuming that all errors in the electronic interaction energies are purely random. If this were true it would suggest that, for example, PM3 total interaction energies for ligand binding can be off by as much as 14.7 kcal/mol and we would be left with no reliable statistical way to narrow this error window. This would of course be unacceptable because this error bar is actually larger in magnitude than the experimental binding affinity itself (−12.8 kcal/mol). The random error bar for GAFF is about half as large in magnitude, but still represents a large portion of the experimental binding affinity. When quantum effects are accounted for, e.g. MP2/CBS, we are able to narrow the range of accumulated random errors down to ~1 kcal/mol if we use the complete basis set extrapolations.

Upon examination of the mean errors over the 21 fragment systems (see column 3 of Table 1), it is apparent that there is a significant systematic error component in the majority of the methods studied. Further evidence is shown in Figure 4, which plots the interaction energy deviations from the CCSD(T)/CBS reference for each fragment pair for a subset of the methods examined. Interestingly, most of the calculated interaction energies are less strongly binding than those calculated with the CCSD(T)/CBS reference indicating far less random error character in the computed interaction energy errors. Systematic errors can be distinguished from the random sources of error by choosing an appropriate probability distribution function to describe the measurement data. If we choose the Gaussian distribution as defined by equation 3, we need to calculate two parameters, the mean error μ , and the variance σ^2 . The values for each computational method are shown in columns three and four of Table 1.

If we choose to model our calculations (measurements) with the Gaussian pdf, the calculated value of μ represents a shift of the distribution (systematic error), and σ^2 represents a measure of the width of the distribution (random error). For example, GAFF interaction energies have an average deviation from our reference of 0.94 kcal/mol, which shifts the Gaussian distribution by this amount. The variation about this mean error is 1.66 (kcal/mol)², which gives a measure of the width of the distribution and the random error of measurement (see Figure 5). By using these values, we can also obtain an estimate of the total systematic and random error from all fragment interactions comprising the protein/ligand complex by evaluating functions similar to equations 1 and 2. However, now we are using the parameters of the probability distribution function, μ and σ^2 , assuming for now that they are representative of all the chemical interactions.

$$\text{Error}_{\text{Systematic}} = \sum_i \mu_i = N\mu \quad (12)$$

$$\text{Error}_{\text{Random}} = \left[\sum_i \sigma_i^2 \right]^{1/2} = \sqrt{N\sigma^2} \quad (13)$$

Note that both equations 12 and 13 depend on the number of chemical interactions. This shows the size-dependence of our error estimates as discussed earlier by Merz.⁶ Errors in energy calculations may be inconsequential when dealing with small systems, but as more molecular contacts become involved, propagated error may have significant effects on energy calculations.

Using equations 12 and 13, we find that the predicted systematic error for the total interaction energy of the protein-ligand complex with GAFF is 19.74 kcal/mol (column 5 of Table 1). The predicted overall random error bar is 5.90 kcal/mol (column 6). These values of random error are lower than the BCS_{error} values because we have removed the systematic components of the overall errors in each method. We can narrow this random error bar even further by constructing pdf's for different types of interactions – e.g. nonpolar and polar interactions. In this way, the systematic error for each type of chemical interaction could be removed independently, while leaving less random error in the overall interaction energies.

Keeping with the GAFF energies as an example, we calculated the mean error and variance of the 15 nonpolar interactions to be 1.17 kcal/mol and $1.83 \text{ (kcal/mol)}^2$ and the mean error and variance of the 6 polar interactions to be 0.35 kcal/mol and $1.03 \text{ (kcal/mol)}^2$. To compute the total systematic and random errors due to all interactions, we use equations 12 and 13, but with different values of μ and σ^2 depending the interaction type. The overall systematic error by differentiating between nonpolar and polar interactions is predicted to be 19.65 kcal/mol, slightly less than the mean error by averaging over all interactions. The overall random error bar is now calculated to be 5.80 kcal/mol, less than the value calculated previously using one value of variance for all interactions. As more pdf's for different classes of interactions are used, and more data points are added to them, we expect that the overall systematic error in interaction energies can be more reliably removed and the random error bars can be minimized.

Conclusion

Using benchmark-quality quantum mechanical CCSD(T)/CBS results we have been able to establish reference values for 21 fragment interaction energies present in the complex of indinavir and HIV-II protease. Using these reference values we established the errors associated with the use of less sophisticated computational techniques including force fields, semiempirical methods, and DFT calculations. Using error analysis tools we estimated random and systematic errors present in each computational method by constructing probability density functions. We found that all computational methods predominantly gave interaction energies that were less binding than the reference energies yielding a significant systematic error component in each case. Semiempirical methods performed the poorest overall, yielding large random and systematic errors in all cases. The GAFF force field significantly outperformed the semiempirical methods. The HF calculations had large systematic errors due to the lack of dispersion, but had relatively small random errors. The MO6-L functional performed impressively, even with the small 6-31G* basis set, giving remarkably small systematic and random errors.

Our error analysis suggests one way in which any scoring algorithm based on energy could be improved. Given a scoring function, comparisons of interaction energies for many types of interactions in different geometries with chemically accurate quantum mechanical calculations could be used to derive error probability density functions for several classes of interactions. By constructing these pdf's, the predicted systematic error involved in each interaction present could be removed. The remaining random component of error could then be reported with the energy scores to provide insight into their reliability. The results presented herein offer only a start along this direction in the sense that the number of interactions, their type and geometry distribution is largely limited to the indinavir/HIV-II system studied. In order develop accurate pdf's to correct a broad range of interactions one would require the collection of a large number of unique interacting fragments to serve as reference systems to assemble the pdf's for any given scoring function. Such a set of interacting fragments could then provide a basis by which any systematic errors in a physics-based score function could be corrected for assuming that additivity is largely

dominant in the computation of interaction energies. The additivity or non-additivity of the interactions for a given protein-ligand complex is poorly understood and understanding this aspect of the problem would be essential to effectively apply any systematic error correction scheme.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Professors Adrian Roitberg and Erik Deumens for helpful discussions. KMM thanks the NIH (GM044974 and GM066689) for funding the present research. CDS gratefully acknowledges support from NSF (CHE-0715268 and CHE-04-43564). This work is also funded in part by the U.S. Department of Energy, the division of Advanced Scientific Computing Research as a component of the Computational Chemistry Endstation project under subcontract from the University of Tennessee. This research was performed in part using resources of the National Center for Computational Sciences at Oak Ridge National Laboratory under contract DE-AC05-00OR22725.

References

1. Gilson MK, Zhou HX. *Annu Rev Biophys Biomol Struct.* 2007; 36:21–42. [PubMed: 17201676]
2. Leach AR, Shoichet BK, Peishoff CE. *J Med Chem.* 2006; 49:5851–5855. [PubMed: 17004700]
3. Abel R, Young T, Farid R, Berne BJ, Friesner RA. *J Am Chem Soc.* 2008; 130:2817–2831. [PubMed: 18266362]
4. Deng YQ, Roux B. *J Phys Chem B.* 2009; 113:2234–2246. [PubMed: 19146384]
5. Martin YC. *J Comput-Aided Mol Des.* 2009; 23:693–704. [PubMed: 19842045]
6. Merz KM. *J Chem Theory Comput.* 2010; 6:1769–1776.
7. Dill KA. *J Biol Chem.* 1997; 272:701–704. [PubMed: 8995351]
8. Raha K, Peters MB, Wang B, Yu N, Wollacott AM, Westerhoff LM, Merz KM. *Drug Discovery Today.* 2007; 12:725–731. [PubMed: 17826685]
9. Peters MB, Raha K, Merz KM. *Curr Opin Drug Discovery Dev.* 2006; 9:370–379.
10. Raha K, Merz KM. *J Med Chem.* 2005; 48:4558–4575. [PubMed: 15999994]
11. Fukuzawa K, Kitaura K, Uebayasi M, Nakata K, Kaminuma T, Nakano T. *J Comput Chem.* 2005; 26:1–10. [PubMed: 15521089]
12. Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP. *J Med Chem.* 2005; 48:962–976. [PubMed: 15715466]
13. Guvench O, MacKerell AD. *Curr Opin Struct Biol.* 2009; 19:56–61. [PubMed: 19162472]
14. Taylor RD, Jewsbury PJ, Essex JW. *J Comput-Aided Mol Des.* 2002; 16:151–166. [PubMed: 12363215]
15. Kontoyianni M, McClellan LM, Sokol GS. *J Med Chem.* 2003; 47:558–565. [PubMed: 14736237]
16. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. *J Med Chem.* 2006; 49:5912–5931. [PubMed: 17004707]
17. Charifson PS, Corkery JJ, Murcko MA, Walters WP. *J Med Chem.* 1999; 42:5100–5109. [PubMed: 10602695]
18. Oda A, Tsuchida K, Takakura T, Yamaotsu N, Hirono S. *J Chem Inf Model.* 2006; 46:380–391. [PubMed: 16426072]
19. Wang RX, Wang SM. *J Chem Inf Comput Sci.* 2001; 41:1422–1426. [PubMed: 11604043]
20. Kollman PA. *Chem Rev.* 1993; 93:2395–2417.
21. Huang N, Kalyanaraman C, Bernacki K, Jacobson MP. *Phys Chem Chem Phys.* 2006; 8:5166–5177. [PubMed: 17203140]
22. Foloppe N, Hubbard R. *Curr Med Chem.* 2006; 13:3583–3608. [PubMed: 17168725]

23. Lybrand TP, McCammon JA, Wipff G. *Proc Natl Acad Sci U S A*. 1986; 83:833–835. [PubMed: 3456569]
24. Taylor, JR. Propagation of Uncertainties. In: McGuire, A., editor. *An introduction to error analysis: the study of uncertainties in physical measurements*. 2. University Science Books; Sausalito, California: 1996. p. 45–93.
25. Bissantz C, Kuhn B, Stahl M. *J Med Chem*. 2010; 53:5061–5084. [PubMed: 20345171]
26. Mark AE, Vangunsteren WF. *J Mol Biol*. 1994; 240:167–176. [PubMed: 8028000]
27. Baum B, Muley L, Smolinski M, Heine A, Hangauer D, Klebe G. *J Mol Biol*. 2010; 397:1042–1054. [PubMed: 20156458]
28. Jorgensen WL, Pranata J. *J Am Chem Soc*. 1990; 112:2008–2010.
29. Bartlett RJ, Musial M. *Rev Mod Phys*. 2007; 79:291–352.
30. Halkier A, Helgaker T, Jorgensen P, Klopper W, Koch H, Olsen J, Wilson AK. *Chem Phys Lett*. 1998; 286:243–252.
31. Helgaker T, Klopper W, Tew DP. *Mol Phys*. 2008; 106:2107–2143.
32. Chen ZG, Li Y, Chen E, Hall DL, Darke PL, Culberson C, Shafer JA, Kuo LC. *J Biol Chem*. 1994; 269:26344–26348. [PubMed: 7929352]
33. Dorsey BD, Levin RB, Mcdaniel SL, Vacca JP, Guare JP, Darke PL, Zugay JA, Emini EA, Schleif WA, Quintero JC, Lin JH, Chen IW, Holloway MK, Fitzgerald PMD, Axel MG, Ostovic D, Anderson PS, Huff JR. *J Med Chem*. 1994; 37:3443–3451. [PubMed: 7932573]
34. Vacca JP, Dorsey BD, Schleif WA, Levin RB, Mcdaniel SL, Darke PL, Zugay J, Quintero JC, Blahy OM, Roth E, Sardana VV, Schlabach AJ, Graham PI, Condra JH, Gotlib L, Holloway MK, Lin J, Chen IW, Vastag K, Ostovic D, Anderson PS, Emini EA, Huff JR. *Proc Natl Acad Sci U S A*. 1994; 91:4096–4100. [PubMed: 8171040]
35. Word JM, Lovell SC, Richardson JS, Richardson DCJ. *Mol Biol*. 1999; 285:1735–1747.
36. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. *Proteins: Struct Funct Bioinf*. 2006; 65:712–725.
37. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem*. 2004; 25:1605–1612. [PubMed: 15264254]
38. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA. *J Comput Chem*. 2004; 25:1157–1174. [PubMed: 15116359]
39. Ponder JW, Wu CJ, Ren PY, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T. *J Phys Chem B*. 2010; 114:2549–2564. [PubMed: 20136072]
40. Halgren TA. *J Comput Chem*. 1999; 20:720–729.
41. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. *J Am Chem Soc*. 1985; 107:3902–3909.
42. Foster M, Sohlberg K. *J Chem Theory Comput*. 2010; 6:2153–2166.
43. Stewart JJP. *J Comput Chem*. 1989; 10:209–220.
44. Stewart JJP. *J Mol Model*. 2007; 13:1173–1213. [PubMed: 17828561]
45. Korth M, Pitonák M, Rezák J, Hobza P. *J Chem Theory Comput*. 2010; 6:344–352.
46. Repasky MP, Chandrasekhar J, Jorgensen WL. *J Comput Chem*. 2002; 23:1601–1622. [PubMed: 12395428]
47. Zhao Y, Truhlar DG. *Theor Chem Acc*. 2008; 120:215–241.
48. Grimme S. *J Comput Chem*. 2006; 27:1787–1799. [PubMed: 16955487]
49. Ditchfie R, Hehre WJ, Pople JA. *J Chem Phys*. 1971; 54:724–728.
50. Kendall RA, Dunning TH, Harrison RJ. *J Chem Phys*. 1992; 96:6796–6806.
51. Sherrill CD, Takatani T, Hohenstein EG. *J Phys Chem A*. 2009; 113:10146–10159. [PubMed: 19689152]
52. Elsohly AM, Tschumper GS. *Int J Quantum Chem*. 2009; 109:91–96.
53. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JJA.; Vreven, T.; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.;

- Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J.E.; Hratchian, H.P.; Cross, J.B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.E.; Yazyev, O.; Austin, A.J.; Cammi, R.; Pomelli, C.; Ochterski, J.W.; Ayala, P.Y.; Morokuma, K.; Voth, G.A.; Salvador, P.; Dannenberg, J.J.; Zakrzewski, V.G.; Dapprich, S.; Daniels, A.D.; Strain, M.C.; Farkas, O.; Malick, D.K.; Rabuck, A.D.; Raghavachari, K.; Foresman, J.B.; Ortiz, J.V.; Cui, Q.; Baboul, A.G.; Clifford, S.; Cioslowski, J.; Stefanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R.L.; Fox, D.J.; Keith, T.; Al-Laham, M.A.; Peng, C.Y.; Nanayakkara, A.; Challacombe, M.; Gill, P.M.W.; Johnson, B.; Chen, W.; Wong, M.W.; Gonzalez, C.; Pople, J.A. Gaussian 03, Revision E.01. Gaussian, Inc; Wallingford, CT: 2004.
54. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.P.; Izmaylov, A.F.; Bloino, J.; Zheng, G.; Sonnenberg, J.L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J.A., Jr; Peralta, J.E.; Ogliaro, F.; Bearpark, M.; Heyd, J.J.; Brothers, E.; Kudin, K.N.; Staroverov, V.N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.C.; Iyengar, S.S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N.J.; Klene, M.; Knox, J.E.; Cross, J.B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.E.; Yazyev, O.; Austin, A.J.; Cammi, R.; Pomelli, C.; Ochterski, J.W.; Martin, R.L.; Morokuma, K.; Zakrzewski, V.G.; Voth, G.A.; Salvador, P.; Dannenberg, J.J.; Dapprich, S.; Daniels, A.D.; Farkas, Ö.; Foresman, J.B.; Ortiz, J.V.; Cioslowski, J.; Fox, D.J. Gaussian 09. Gaussian, Inc; Wallingford, CT: 2009.
55. Dixon, S.L.; van der Vaart, A.; Gogonea, V.; Vincent, J.J.; Brothers, E.N.; Suárez, D.; Westerhoff, L.M.; Merz, K.M., Jr. DIVCON99. The Pennsylvania State University; 1999.
56. Case, D.A.; Darden, T.A.; Cheatham, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Ross, C.; Walker, W.Z.; Merz, K.M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K.F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D.H.; Seetin, M.G.; Sagui, C.; Babin, V.; Kollman, P.A. AMBER 10. University of California; San Francisco: San Francisco, CA: 2008.
57. Werner, H.-J.; Knowles, P.J.; Lindh, R.; Manby, F.R.; Schütz, M.; Celani, P.; Knizia, G.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Adler, T.B.; Amos, R.D.; Bernhardsson, A.; Berning, A.; Cooper, D.L.; Deegan, M.J.O.; Dobbyn, A.J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A.W.; Mata, R.A.; May, A.J.; McNicholas, S.J.; Meyer, W.; Mura, M.E.; Nicklass, A.; Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A.J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. MOLPRO. 2009. see <http://www.molpro.net>
58. Kendall RA, Apra E, Bernholdt DE, Bylaska EJ, Dupuis M, Fann GI, Harrison RJ, Ju JL, Nichols JA, Nieplocha J, Straatsma TP, Windus TL, Wong AT. Comput Phys Commun. 2000; 128:260–283.
59. MacroModel, version 9.8. Schrodinger, LLC; New York, NY: 2010.

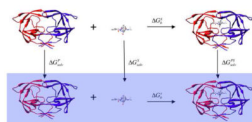


Figure 1.
Thermodynamic cycle used to estimate the free energy of binding of a drug molecule to a protein receptor. PS=protein-substrate complex, P=protein and S=substrate

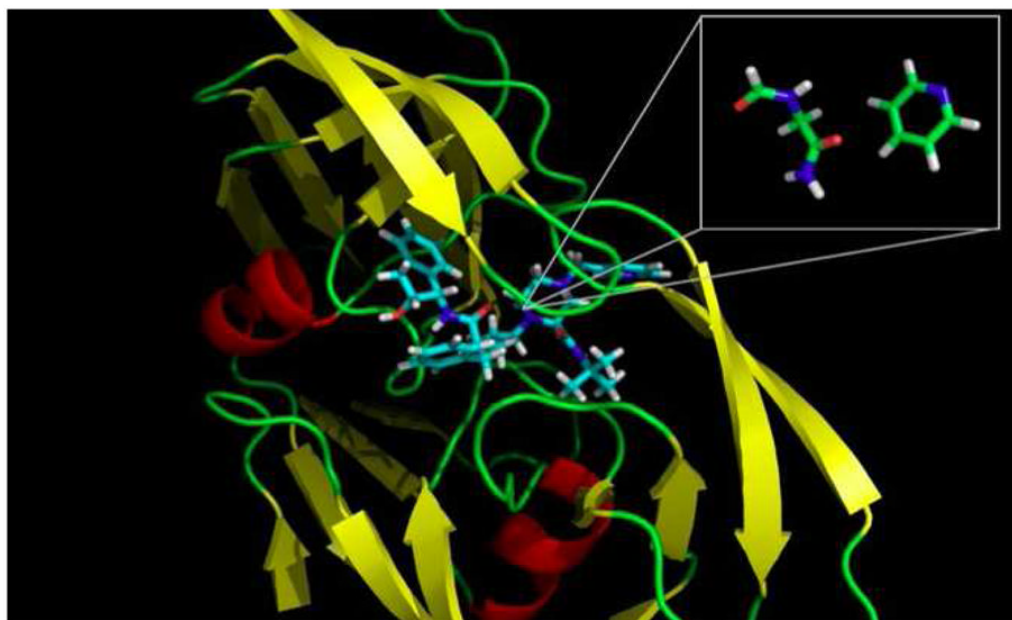


Figure 2. Crystal structure of HIV-II protease bound to indinavir (PDBID: 1HSG). The inner window displays an example of the fragment systems studied, which includes a glycine residue interacting with a pyridine ring in the ligand.

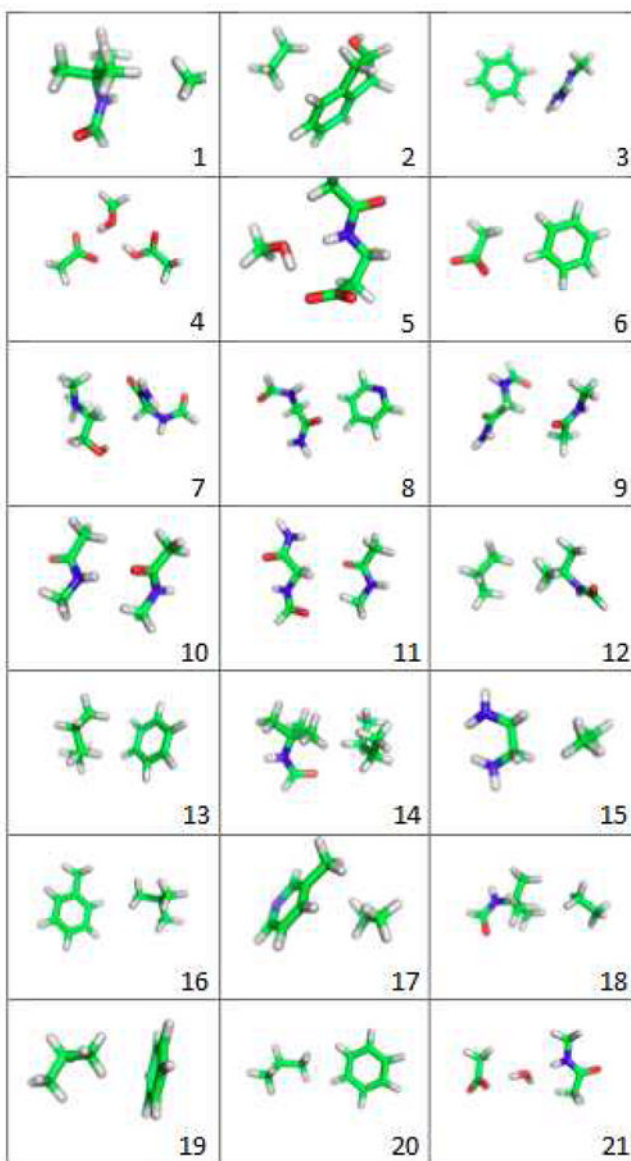


Figure 3.

The 21 chemical fragment pairs comprising the 1HSG protein/ligand complex used in this study.

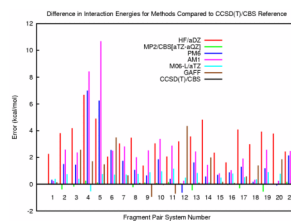


Figure 4.

The difference between example methods and the CCSD(T)/CBS reference method in interaction energy of each of the 21 fragment systems. Most methods generally underestimate the interaction energy, which implies a portion of the errors is systematic in nature

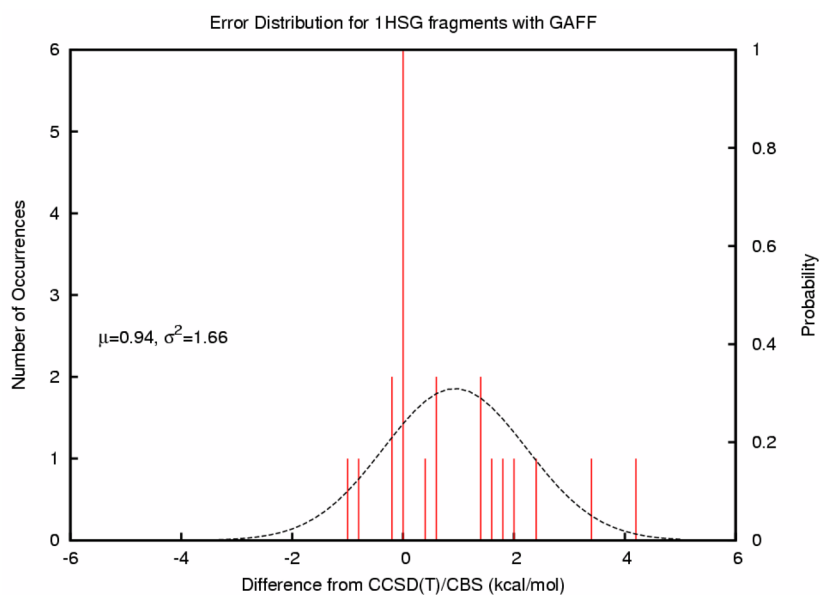


Figure 5. Error distribution of GAFF interaction energies of the 21 fragment systems fit to a Gaussian probability density function. The systematic error in the GAFF interaction energies (using CCSD(T)/CBS as a reference) is predicted to be 0.94 kcal/mol per interaction and the variance (random error) about this shift is 1.66 (kcal/mol)².

Table 1

Error Analysis of the Absolute Interaction Energies from Various Methods. Various scores (based on kcal/mol units) measuring the error of gas-phase electronic interaction energies of 21 independent chemical fragments involved in the binding of indinavir to IHSG using CCSD(T)/CBS calculations as reference. The error scores are (from left to right): BCS_{error} (equation 9), mean error, variance about the mean error, overall expected systematic error, overall expected random error, root mean square error, and maximum error.

Method	BCS_{error}	ME (μ)	Var (σ^2)	$N\mu$	$\sqrt{N\sigma^2}$	RMSE	Max Err
GAFF	7.41	0.94	1.66	19.74	5.90	1.62	4.35
AMOEBA ^a	2.58	0.36	0.24	7.56	2.24	0.75	2.23
MMFF	20.81	-0.61	20.25	-12.81	20.62	4.54	11.43
AM1	16.39	2.57	5.89	53.97	11.12	3.58	10.69
AM1-FS1	11.44	0.03	6.23	0.63	11.44	2.50	7.97
PM3	14.70	1.82	6.67	38.22	11.84	3.21	8.24
PDDG	20.46	1.69	16.29	35.49	18.50	4.47	14.8
PM6	10.94	1.56	3.13	32.76	8.11	2.39	6.99
PM6-DH2	4.89	-0.25	1.11	-5.25	4.83	1.07	2.94
B97-D/							
TZVP	2.53	0.05	0.30	1.05	2.51	0.55	2.12
M06-L/							
6-31G*	10.94	0.44	0.48	9.24	3.17	0.83	1.92
aTZ	3.25	0.62	0.11	13.02	1.52	0.71	1.16
HF/							
6-31G*	15.22	3.18	0.87	66.78	4.27	3.32	4.91
aDZ	16.25	3.33	1.42	69.93	5.46	3.55	6.68
aTZ	16.22	3.32	1.42	69.72	5.46	3.54	6.77
aQz	16.16	3.31	1.39	69.51	5.40	3.53	6.70
MP2/							
6-31G*	7.45	1.58	0.13	33.18	1.65	1.63	2.35
aDZ	2.89	0.41	0.22	8.61	2.15	0.63	1.83
aTZ	1.35	0.09	0.08	1.89	1.30	0.30	0.77
aQZ	1.02	-0.02	0.05	-0.42	1.02	0.22	0.53

Method	BCS_{error}	ME (μ)	Var (σ^2)	$N\mu$	$\sqrt{N\sigma^2}$	RMSE	Max Err
CBS[aDZ-aTZ]	1.07	-0.04	0.05	-0.84	1.02	0.23	0.57
CBS[aTZ-aQZ]	1.02	-0.05	0.04	-1.05	0.92	0.22	0.56
CCSD(T)/							
haDZ	3.80	0.69	0.20	14.49	2.05	0.83	2.08
haTZ	1.28	0.23	0.03	4.83	0.74	0.28	0.74
CBS[haDZ-haTZ] ^b	0.28	0.03	0.00	0.63	0.00	0.06	0.14

^a AMOEBA results were calculated for 12 of the 21 fragments, due to lack of parameters for atoms in the remaining 9 systems.

^b haDZ-haTZ extrapolation to the CCSD(T) complete basis set limit; this approximates our benchmark CCSD(T)/CBS values but lacks the MP2 estimate of basis set effects beyond haTZ.