Journal of Molecular Graphics and Modelling 46 (2013) 41–51

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Could MM-GBSA be accurate enough for calculation of absolute protein/ligand binding free energies?

Chandrika Mulakala*, Vellarkad N. Viswanadhan

Jubilant Biosys Limited, #96, Industrial Suburb, 2nd Stage, Yeshwanthpur, Bangalore 560 022, India

ARTICLE INFO

Article history:
Accepted 7 September 2013
Available online 18 September 2013

Keywords:
Implicit solvation
MM-GBSA
Protein/ligand binding
Free energy of binding
Absolute binding free energy

ABSTRACT

Implicit solvation methods such as MM-GBSA, when applied to evaluating protein/ligand binding free energies, are widely believed to be accurate only for the estimation of relative binding free energies for a congeneric series of ligands. In this work, we show that the MM-GBSA flavor of Prime 3.0, VSGB-2.0, with a variable dielectric model and a novel energy function, could be approaching the accuracy required for evaluating absolute binding free energies, albeit, through a linear regression fit. The data-set used for validation includes 106 protein–ligand complexes that were carefully selected to control for variability in the affinity data as well as error in the modeled complexes. Through systematic analysis, we also quantify the degradation in the R^2 of fit between experimental and calculated values with either greater variability in the affinity data or an increase in error in the modeled protein/ligand complexes. Limitations for its application in drug discovery are discussed along with the identification of areas for future development.

1. Introduction

Accurate theoretical estimation of protein–ligand binding affinities would be invaluable for drug discovery. Where X-ray structures of the protein of interest are available, docking algorithms offer a computationally efficient means of predicting the bound pose of a ligand in the active-site of the protein. Scoring functions translate this knowledge of the bound complex to an estimate of binding affinity. However, since scoring functions are typically optimized for speed, they achieve computationally efficiency through the use of empirical functions that lack a clear theoretical basis. For instance, solvent effects that strongly influence binding equilibrium are only indirectly accounted for by most empirical scoring functions [1,2].

The theoretically most realistic form of modeling the effects of solvent is through explicit solvation, which is computationally prohibitive and generally incompatible with the timelines expected of industrial drug discovery. In the past decade, implicit continuum solvation methods, such as the generalized born surface area (GBSA) methodology, have gained in popularity as they offer modeling of solvent effects at a modest computational cost [3,4]. GBSA, and its theoretical precedent, Poisson–Boltzmann surface area (PBSA) [5] assume that the total energy of a solvated molecule, $E_{tot} = E_{vac} + \Delta G_{solv}$, where, E_{vac} is the gas phase energy of the molecule, and ΔG_{solv} is the solvation free energy. As a further simplification, it is assumed that the solvation free energy can be separated into two components: $\Delta G_{Nonpolar}$ and ΔG_{El}. Where, $\Delta G_{Nonpolar}$ is the free energy of solvating the molecule after setting the partial charges of all the atoms to zero and ΔG_{El} is the energy of adding the charges in the presence of a continuum solvent. ΔG_{El} is calculated using the generalized-Born (GB) model, which is an approximation to the more general Poisson–Boltzmann equation [4]. E_{vac} can be easily computed for a given choice of a molecular mechanics (MM) force-field, and when coupled with GBSA to estimate E_{tot}, it is referred to as MM-GBSA.

In early implementations, $\Delta G_{Nonpolar}$ was approximated to be proportional to the total solvent accessible surface area. Since then, it has been found that a simple surface area term for capturing $\Delta G_{Nonpolar}$ is insufficient for realistic modeling of macromolecules and has been further parameterized [4]. Several flavors of GBSA have evolved over the years, each with their own set of empirical terms implemented to correct for observed deficiencies of the implicit assumptions made for solving the GB equation [4]. In the present work we chose to use the MM-GBSA implementations available with Prime 3.0 [6] of the Schrödinger software suite and hence focus our attention on the various empirical corrections that were applied to Prime MM-GBSA.

Abbreviations: GB, generalized-Born; MM, molecular mechanics; GBSA, generalized-Born surface area; SGB, surface-area based generalized Born; SGB-NP, SGB with hydrophobic term added to the non-polar surface area term; VSGB-1.0, SGB-NP with a variable dielectric formulation; VSGB-2.0, VSGB-1.0 with an improved energy function; BACE, β-secretase; PDPK-1, phosphoinositide-dependent kinase-1.

* Corresponding author. Tel.: +91 80 6662 8290; fax: +91 80 6662 8333.
E-mail address: chandrika_mulakala@jubilantbiosys.com (C. Mulakala).

1093-3263/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2013.09.005

1.1. Prime 3.0 MM-GBSA

Prime 3.0 supports a surface-area based generalized Born model for polar solvation (SGB), with several empirical corrections introduced to reproduce accurate solutions of the Poisson–Boltzmann equation [7]. The non-polar surface area term was subsequently augmented with a hydrophobic term to account for attractive solute-solvent interactions (SGB-NP) [8]. Later, a variable dielectric formulation was introduced to account for internal polarization effects of protein side chains (VSGB-1.0) [9]. Their most recent implementation, VSGB-2.0, includes further terms to model physics-based empirical corrections for modeling directionality of hydrogen bonding interactions, π–π stacking interactions and, a special term to account for self-hydrogen bonds in protein structures [10]. All of these empirical corrections were validated by assessing the accuracy of protein side chain and loop predictions, and VSGB-2.0 is reported to have significantly improved performance in the prediction of conformations/energetics of a large set of super long loops [10].

It is essential to note that the improved energy function of VSGB 2.0 was implemented for amino-acid residues and therefore is not applicable to ligands [10]. Despite that, we expected VSGB-2.0 to have significantly improved accuracy in calculating MM/GBSA binding free energies (methodology described in the next section) through better modeling of the protein structures, and toward that goal, we tested it against a carefully chosen set of protein/ligand complexes. For our chosen data-set, we were pleasantly surprised to find that not only was the VSGB-2.0 energy function quite accurate for estimating relative binding free energies, the derived regression equation successfully computed accurate affinities ($R^2 = 0.89$) for a data-set that covered several protein targets and a wide range of ligand chemotypes. This is consistent with the recent findings of Greenidge et al. [11], where they also find a statistically significant correlation ($R^2 = 0.63$) for estimated MM/GBSA ΔG_{bind}s with VSGB-2.0 for a diverse data-set of 855 protein–ligand complexes. In their analysis, they point out the pitfalls of automated procedures of protein preparation and the need for manual intervention to ensure that starting complexes are accurately modeled. All the studies reported prior to Greenidge et al. suggest that MM/GBSA based calculations are accurate only for the prediction of relative binding free energies of a congeneric series of ligands [12–19].

Our approach diverges from that of Greenidge et al. through a careful selection of the data-set that was used for validation in the present study. In order to extract the accuracy of MM-GBSA, without the noise cancelation available from the use of statistically large validation sets for identifying correlations, we think that is important to have a strong handle on the variability and error inherent in the data itself. Intrinsic variability in the experimental data is expected to superpose over the error-bars inherent in the GB-formulation itself, which will together reduce the R^2 of the fit in a manner that is not indicative of the accuracy of the MM-GBSA methodology. In addition, we also estimate the fall in the accuracy of MM-GBSA when docked poses are used for calculations, as opposed to X-ray structures.

1.2. Errors due to approximations in binding free energy calculations

Statistical thermodynamics postulates that accurate determination of the free energy of binding would require a complete sampling of the conformational space of the solvated ligand, solvated protein and the solvated protein/ligand complex. Rastelli G et al. [19] report that for binding free energy determination using MM-GBSA/PBSA, the accuracy achieved from using a single conformer approaches that obtained from an averaging over multiple molecular dynamics snapshots, leading to significant savings in computational time. In the present work, therefore, we have used an approximate form of the binding free energy equation that does not involve any conformational sampling. The free energy of binding is evaluated simply as:

$$\Delta G_{\text{bind}} = E_{\text{complex(minimized)}} - (E_{\text{protein(unbound,minimized)}} + E_{\text{lig(unbound,minimized)}}) \quad (1)$$

where, ΔG_{bind} is the calculated binding free energy, $E_{\text{complex(minimized)}}$ is the MM-GBSA energy of the minimized complex, $E_{\text{protein(unbound,minimized)}}$ is the MM-GBSA energy of the minimized protein after separating it from its bound ligand and $E_{\text{lig(unbound,minimized)}}$ is the MM-GBSA energy of the ligand after separating it from the crystal complex and allowing it to relax. This implementation for computation of binding free energies in the Prime 3.0 [6] module of the Schrödinger suite was designed primarily to provide a quick rescoring of docked poses generated by Glide [20]. The complete absence of any conformational sampling is likely to introduce significant errors, particularly if the global minimum conformations of the ligand or the protein differ significantly from their respective bioactive conformations in the crystal structures. In addition, it also leads to a complete neglect of the solute entropic terms in the free energy calculation.

1.3. Sources of variability and error in the data-sets

The first source of variability is in the affinity data itself. Biological assays are complex and sensitive to assay conditions and the quality of reagents and enzymes. In the industrial set-up, prior to utilization for determination of structure activity relationships, assays are rigorously statistically characterized to measure and validate reliability of potency estimates [21]. A second issue is the use of IC_{50} as a popular measure of affinity over the more objective K_i or K_d, which makes it difficult to compare data published between laboratories. In order to reduce variability due to the aforementioned causes we were careful to select data that was generated within the same laboratory for our high quality data-set. The data-set studied in this work includes four different targets – bovine trypsin, phosphoinositide-dependent kinase-1 (PDPK-1), β-secretase (BACE), and human aldose reductase. In case of the trypsin data-set, K_is were reported for all the data points by the same laboratory [22,23]. For the PDPK1 data-set, even though the reported values are IC_{50}s, we consider it to have low variability, as it was generated by the same laboratory [24,25]. In the BACE data-set, the data came from two different laboratories, however, it has a common fragment (2-aminoquinoline) with similar reported affinities ($IC_{50} \sim 22$ mM [26] and $K_i = 0.9$ mM [27]) and one crystal structure complexed with 2-aminoquinoline (2ohl) [26]. These two data-sets were therefore considered together and assumed to have low variability in reported affinities, even though they include a mix of K_is and IC_{50}s. The aldose reductase data-set [28–41] includes data reported from several laboratories over the course of a decade, with a mix of K_is and IC_{50}s. We expect this data-set to have higher variability in the affinity data, and therefore used it to quantify the degradation in the fit due to variability in the affinity data.

A second source of error is in the quality of the protein/ligand complexes used for the MM-GBSA calculation. High-resolution crystal structures of protein/ligand complexes offer the best accuracy. Even then, since X-ray structures do not have resolved hydrogen atoms, (except for structures of very high resolution i.e. <1.6 Å), deciphering the correct protonation states for titratable residues in the active site as well as the ligand is critical for accurate energy calculations. In this study, we controlled for errors arising from inaccuracy in the modeled complexes by restricting our most accurate data-set to include only those data-points that had

published X-ray complexes. In addition, for targets with titratable amino-acids and/or titratable groups in the ligand, we ensured that the reported pH for crystallization was within 1.5 pH units of the pH of the assay. This was necessary to exclude the possibility of pH-dependent conformational changes in the protein which are likely to affect computed binding energies. X-ray structures were reported for all the data-points used in the aldose reductase data-set, ensuring high accuracy of the starting complexes modeled in this data-set. Therefore, this data-set has high variability in the affinity data, while having accurate starting complexes for modeling.

Furthermore, in industrial drug discovery, docked structures are routinely used to guide synthesis. For well studied targets several crystal complexes may be available. In such a scenario, bound complexes are routinely inferred through computational docking, which results in further extrapolations and additional error in the MM-GBSA energies. In order to understand the influence of the use of docked complexes in estimated ΔG_{bind}s, we used an extended data-set of PDPK1 SAR reported by the Medina et al. [24,25], for which no X-ray structures were reported. For this data-set, docked poses were scored using VSGB-2.0, in an effort to understand the effect of inaccuracies in the modeled complexes toward estimated ΔG_{bind}s. In this data-set, the affinity data was considered to have low variability.

In addition, the data-sets have a wide-range in the size (133.2–461.4) and flexibility (0–9 torsions) of the ligands and a range of six orders of magnitude in the measured IC_{50}s or K_is (structures of modeled ligands along with protonation states of ligands and titratable amino-acids in the active site are included in supplementary data). Within each chosen target, we ensured a significant range (>3 orders of magnitude in IC_{50} or K_i) and spread in the affinity data. For BACE and PDPK-1, the data-set covers fragment hits from high-throughput screens to fully optimized lead-like compounds of low nanomolar potency.

2. Computational methods

2.1. Protein preparation

The protein preparation wizard [42,43] of Schrödinger was used for preparing proteins to assign bond orders, add hydrogens, delete waters beyond 8 Å from the ligand and generate protonation states for the ligand. The hydrogen bonding network was optimized using the 'exhaustive sampling' option at neutral pH, except for the BACE data-set, where the low-pH option was used. Hydrogen bonding was optimized in the presence of X-ray waters in order to include the immediate solvent environment in the optimization of hydrogen bonding. In case of trypsin, the protonation states of the ligands and His57 in the active-site was forced to be consistent with the X-ray structures (i.e. His57 side-chain was positively charged), which are of high resolution (1.33–1.70 Å), and therefore have resolved hydrogens for His57 as well as the ligands. The complex was then minimized with the heavy atoms restrained to remain within a 0.3 Å root-mean-square deviation of the original PDB structure.

2.2. Prime MM-GBSA

All waters were deleted from the modeled complexes prior to binding free energy computation using Prime MM-GBSA. Residues within 4.0 Å of the ligand were allowed to flex while minimizing the complex in order to relieve minor steric clashes in the modeled complex. Three different solvent models for MM-GBSA were compared: SBG-NP [8] – the surface area based generalized Born model with a hydrophobic term, VSGB-1.0 [9] – which uses a residue dependent variable dielectric, and VSGB-2.0 [10] – which incorporates further residue-dependent parameterization.

2.3. X-ray structures

The following PDB's were used for modeling starting complexes for each of the following targets:

Bovine trypsin: 1ghz, 1gi2, 1gi6, 1o2k, 1o2q, 1o2x, 1o30, 1o33, 1o36, 1o38, 1o3d, 1o3g [22,23]. Inhibitors in this data-set exhibit a parabolic pH dependence with a minimum in K_i. Only structures determined at pHs close to $K_{i(min)}$ were used to ensure that the selected crystal structures captured accurate protonation states of the inhibitors. This is the same data-set that was curated and used by one of us, elsewhere [44].

PDPK-1: 3nun, 3nus, 3nuu, 3nuy [24], 3qcq, 3qcs, 3qcx, 3qcy, 3qd0, 3qd3, 3qd4 [25].

BACE: 2ohk, 2ohl, 2ohm, 2ohn, 2ohp, 2ohq, 2ohr, 2ohs, 2ohu [26], 3rsv, 3rsx, 3rth, 3rtm, 3rtn, 3ru1, 3rvi [27]. As stated earlier, although this data-set comes from two different laboratories, it has a common fragment (2-aminoquinoline) with similar reported affinities ($IC_{50} \sim 2$ mM [26] and $K_i = 0.9$ mM [27]) and one crystal structure complexed with 2-aminoquinoline (2ohl). These two data-sets were therefore considered together and assumed to have low variability in reported affinities.

Aldose reductase: 1el3 [28], 1iei [29], 1pwl, 1pwm [30], 1t41 [31], 1us0 [32], 1x97 [33], 1z89 [34], 2fzd [35], 2i16 [36], 2ikg, 2ikh, 2iki, 2ikj [37], 2ine, 2inz, 2iq0, 2is7 [38], 2nvc, 2nvd [39], 3dn5 [40], 3g5e [41]. This data-set was considered to have significant variability in the affinity data as it is derived from the work of several laboratories and reported over a decade. 2iqd was excluded as the ligand was not fully resolved in the X-ray structure.

Figures with structures of ligands docked to each protein along with the assigned protonation state used for modeling are included in the supplementary material.

2.4. Composition of the data-sets

The bovine trypsin, BACE and PDPK-1 (with X-ray complexes) data sets together constitute our high quality data-set and are reported together as data-set 1 in Table 1. The Aldose reductase data-set with higher variability in the affinity data and with reported X-ray structures is presented as data-set 2 in Table 2. Table 3 (data-set 3) has data for docked poses of data reported by Medina et al. [24,25] for activity against PDPK-1 and includes data for reported crystal structures for completeness.

2.5. Biased docking protocol

For PDPK-1, in addition to crystal structures, docked structures of compounds with reported IC_{50}s [24,25] were used for MM-GBSA calculations. The docking protocol that was used in the study was designed to incorporate the knowledge of available crystal structures, and is similar to the protocol used by Rapp *et al.* [14] Crystal structures were available for the hinge interacting core (3nus), a core-fragment that spanned the hinge and gate-keeper (3nun) and other key compounds that extended to the glycine rich loop (3qd0, 3qd3, 3qd4). For compounds with reported IC_{50}'s but no crystal structures, we biased docking toward the observed binding modes as follows: for any given design, the crystal structure with the highest common substructure was used to create a grid for docking. Conformational sampling of the ligands was carried out by freezing atoms of the highest common substructures. The resultant conformer set was 'refined' in the active-site using Glide SP, [20] and the best binding mode was retained. The Glide refine mode allows the ligand as a whole to move rigidly by ±1 Å in Cartesian coordinates and includes a full force-field post-docking minimization that improves the geometry of the bound poses. This allows a small degree of induced-fit effects to be modeled, since binding modes are

Table 1
Data set 1: predicted ΔG_{bind}s using the SGB-NP-, VSGB 1.0- and VSGB 2.0 models for the most accurate data-set.

PDBID	IC_{50} (K_i)[a] (nM)	ΔG_{bind} [b] (exptl)	ΔG_{bind} (SGB-NP)	ΔG_{bind} (VSGB 1.0)	ΔG_{bind} (VSGB 2.0)
Trypsin					
1ghz	*16 000*	−6.54	−77.77	−61.71	−60.92
1gi2	*3600*	−7.42	−86.57	−45.46	−61.93
1gi6	*1700*	−7.86	−88.59	−47.31	−63.00
1o2k	*120*	−9.43	−99.95	−76.08	−77.04
1o2q	*21*	−10.47	−105.92	−89.73	−101.24
1o2x	*1400*	−7.98	−88.10	−54.93	−64.49
1o30	*170*	−9.23	−92.75	−76.32	−75.23
1o33	*1800*	−7.83	−91.48	−66.61	−63.28
1o36	*1100*	−8.12	−92.37	−74.19	−77.05
1o38	*150*	−9.30	−82.90	−62.43	−78.57
1o3d	*74*	−9.72	−79.84	−58.93	−77.84
1o3g	*11*	−10.85	−92.09	−83.47	−78.54
BACE					
2ohk	2 000 000	−3.68	−46.72	−24.19	−23.88
2ohl	*900 000*	−4.15	−53.28	−33.20	−23.51
2ohm	310 000	−4.78	−52.28	−49.10	−33.76
2ohn	500 000	−4.50	−66.12	−65.89	−47.43
2ohp	94 000	−5.49	−60.94	−48.37	−37.78
2ohq	25 000	−6.27	−57.42	−70.37	−52.61
2ohr	100 000	−5.45	−55.68	−52.83	−47.77
2ohs	40 000	−5.99	−55.81	−65.00	−56.16
2oht	9100	−6.87	−53.05	−59.42	−43.60
2ohu	4200	−7.33	−68.83	−69.38	−68.91
3rsv	*0.7*	−12.48	−88.92	−121.17	−93.87
3rsx	*38 800*	−6.01	−63.22	−58.51	−52.02
3rth	*6100*	−7.11	−71.14	−91.57	−61.71
3rtm	*38 400*	−6.02	−44.25	−62.53	−54.17
3rtn	*74*	−9.72	−75.30	−100.45	−78.42
3ru1	*14 200*	−6.61	−63.26	−64.89	−55.28
3rvi	*11*	−10.85	−85.77	−119.39	−94.66
PDPK-1					
3nun	370	−8.77	−39.08	−54.52	−65.38
3nus	311 000	−4.78	−23.82	−40.13	−43.78
3nuu	169 000	−5.14	−26.08	−55.33	−49.47
3nuy	345 000	−4.72	−25.37	−46.17	−42.97
3qcq	160[c]	−9.26	−41.27	−59.07	−73.51
3qcs	320[c]	−8.85	−29.84	−59.90	−75.96
3qcx	60[c]	−9.84	−44.81	−61.42	−81.58
3qcy	20[c]	−10.49	−56.19	−90.30	−94.43
3qd0	3[c]	−11.62	−71.82	−96.77	−104.09
3qd3	10[c]	−10.91	−61.29	−92.01	−108.22
3qd4	6[c]	−11.21	−65.20	−91.49	−109.63

[a] K_i's are italicized.
[b] $\Delta G_{bind} \approx RT \ln(IC_{50})$ or $RT \ln(K_i)$. All ΔG_{bind} values are in kcal/mol (= 4.184 kJ/mol).
[c] IC_{50}s have been back-calculated from reported pIC_{50}'s published in Medina et al. [25].

often under some strain. [45,46] The protein was held rigid during the docking.

3. Results

3.1. MM-GBSA scores for data set 1

Table 1 has computed binding free energy data for our high quality data-set (data-set 1) which is controlled for affinity data variability and only uses crystal structures. This is therefore our most accurate data-set and has 40 X-ray crystallographic structures. It includes data-sets for Trypsin, BACE and PDPK-1. Table 4 has the linear regression statistics for the generated models when fit to either data belonging to a specific protein (local regression model) or when fit to all of the data (global regression model). In addition, it compares the predictivity of the three different MM-GBSA models available with Prime 3.0: SGB-NP [8] – the surface area based generalized Born model with a hydrophobic term, VSGB-1.0 [9] – which uses a residue dependant variable dielectric, and VSGB-2.0 [10] – which incorporates further residue-dependent parameterization to model hydrogen bonding and π–π interactions. It shows the significant improvement in the predictivity of the successive GBSA models within Prime for data-set 1, with a linear regression $R^2 = 0.25$ for SGB-NP global regression model, which improves to an $R^2 = 0.60$ for VSGB-1.0 and further increases to a highly statistically significant $R^2 = 0.89$ for the VSGB-2.0 model. Also, a protein-wise distinction of the data (Table 4, Fig. 1) shows how a local fit (intra-protein) evolves into a global model (including all three proteins) that is target independent from SGB-NP to VSGB-2.0. Interestingly, the *y*-axis intercepts of the best-fit regression equations also show a systematic reduction from −30.41 for SGB-NP to −12.23 for VSGB-1.0 to 2.50 for VSGB-2.0. These results suggest that the empirical corrections implemented to the MM-GBSA model are progressing toward a more accurate description of amino-acid interactions in the presence of solvent, due to which the accuracy of predicted binding free energies is no longer dependent on the cancelation of systematic computational errors.

3.2. Affinity data with higher variability

Fig. 2 and Table 4 show the R^2 of the regression fit for the human aldose reductase data (data-set 2). This data-set is expected to have higher variability, since it was generated over a decade of research

Table 2
Data set 2: predicted ΔG_{bind}s using the SGB-NP-, VSGB 1.0- and VSGB 2.0 models for the human aldose reductase data set with high variability in the affinity data.

PDBID	IC_{50} (K_i)[a] (nM)	ΔG_{bind} [b] (exptl)	ΔG_{bind} (SGB-NP)	ΔG_{bind} (VSGB 1.0)	ΔG_{bind} (VSGB 2.0)
1el3	108	−9.50	−50.37	−62.12	−58.80
1iei	*44*	−10.03	−37.48	−76.49	−86.40
1pwl	73	−9.73	−52.42	−121.86	−112.02
1pwm	35	−10.16	−43.78	−75.84	−72.46
1t41	11	−10.85	−52.45	−86.39	−102.72
1us0	30	−10.25	−54.85	−100.69	−95.37
1x97	570	−8.51	−32.04	−66.39	−61.91
1z3n	5	−11.32	−70.18	−93.35	−118.82
1z89	1	−12.27	−51.24	−60.04	−91.91
2fzd	10	−10.91	−70.02	−93.98	−90.93
2i16	30	−10.25	−58.57	−95.61	−94.67
2ikg	530	−8.55	−57.37	−69.50	−78.53
2ikh	4100	−7.34	−56.92	−57.67	−89.49
2iki	*32*	−10.22	−57.18	−108.77	−109.18
2ikj	6	−11.21	−66.52	−99.85	−100.08
2ine	96 000	−5.48	−35.41	−28.80	−42.43
2inz	3 500	−7.44	−34.55	−37.40	−40.14
2iq0	68 200	−5.68	−6.23	−18.41	−28.66
2is7	4 400	−7.30	−42.78	−53.92	−63.70
2nvc	550	−8.53	−54.02	−61.60	−66.91
2nvd	140	−9.34	−69.39	−80.29	−78.90
3dn5	170	−9.23	−53.06	−82.76	−87.15
3g5e	7	−11.12	−70.09	−84.54	−115.55

[a] K_i's are italicized.
[b] $\Delta G_{bind} \approx RT \ln(K_i)$ or $RT \ln(IC_{50})$. All ΔG_{bind} values are in kcal/mol.

from several laboratories. Here too, a similar trend is observed in the performance of SGB-NP, VSGB-1.0 and VSGB-2.0. However, the best R^2 for this data with VSGB-2.0 is significantly lower at 0.63 for the regression fit, which coincidentally is similar to the regression accuracy reported by Greenidge *et al.* (also $R^2 = 0.63$) [11]. Not unexpectedly, when included along with data-set 1, the R^2 for the global fit drops down to 0.80 from 0.89 (Table 4). A concurrent increase in σ is also observed from 7.33 to 10.76, underscoring the importance of a careful consideration of the quality of the input data in validation exercises.

3.3. Rescoring of docked poses with MM-GBSA

A common scenario in drug discovery is one of structurally enabled projects, where, as the SAR progresses, key compounds are crystallized to confirm binding modes. In such situations, molecules are routinely prioritized through docking. Data-set 3 offers the opportunity to test the accuracy of VSGB-2.0 in a situation where crystal structures are available for key molecules to help ensure that hypotheses regarding binding modes are correct. In this case, we chose to modify the docking protocol to bias the sampling in favor of the crystal binding mode as described in the computational methods section. Fig. 3 shows the R^2 of fit for data-set 3. This data-set is differentiated based on three criteria – the availability of crystal complexes (◯), the availability of the crystal structure of a close analog (+) such that only conservative extrapolation of known crystal complexes is required for predicting the binding modes by docking, and highly extrapolated structures (◊). This last set consists of structures of stereoisomers of analogs that are expected to have significant influence on the geometry of the molecule and consequently their binding mode in the protein. These molecules (indicated in Table 3) were treated as outliers in the regression analysis. This data-set was designed to highlight the importance of feedback from X-ray crystal structures in order to steer the docking protocols toward more accurate predictions. For this data-set, with VSGB-2.0 (Table 4), the R^2 and σ become worse from 0.92 and 7.60 (with PDB structures) to 0.81 and 8.67 when docked poses are included, respectively.

4. DISCUSSION

The local regression models for Trypsin, BACE and PDK1 (Table 4) show the effect of the empirical corrections to the SGB-NP, VSGB-1.0 and VSGB-2.0 on the quality of the achieved regression fit. The R^2 values improve significantly between the three models for the trypsin data-set from 0.26 for SGB-NP to 0.48 for VSGB-1.0 to 0.68 for VSGB-2.0. The σ values however increase from 7.17 for SGB-NP to 10.54 for VSGB-1.0 and then fall to 6.88 for the VSGB-2.0 model. For the BACE set, the R^2 and σ values are 0.74, 0.88, 0.88 and 6.66, 9.68, 7.27, for SGB-NP, VSGB-1.0 and VSGB-2.0, respectively. In this case, the R^2 improves from SGB-NP to VSGB-1.0, but does not increase further for VSGB-2.0. The σ values show a trend similar to that seen for trypsin, with degradation in going from SGB-NP to VSGB-1.0, but improving again for VSGB-2.0. For PDPK1 (only PDBs), the R^2 and σ values are 0.82, 0.74, 0.92 and 7.77, 11.12, 7.60, for SGB-NP, VSGB-1.0 and VSGB-2.0, respectively. In this case the R^2 falls from 0.82 to 0.74 from SGB-NP to VSGB-1.0, but improves to 0.92 for VSGB-2.0. The σ values, in this case as well, increase from SGB-NP to VSGB-1.0, before decreasing again for VSGB-2.0. This trend in σ is common to all three data-sets and indicates that the variable dielectric model leads to less accurate predictions, even though R^2 values improved for the trypsin and BACE datasets. Therefore, the additional corrections introduced for aromatic stacking and hydrogen bonding with VSGB-2.0, seem to be necessary to bring overall improvement in both R^2 and σ values.

Overall, the lowest accuracy in the regression fits is seen for the trypsin data-set for all three MM/GBSA models, which could be attributed to the increased complexity of the trypsin protein/ligand complexes. The trypsin X-ray ligands are reported to have strong binding mediated by a cluster of very short hydrogen bonds (<2.3 Å, Fig. 4) [22,23]. Such short hydrogen bonds are not well modeled by the OPLS-2005 force-field and would perhaps require encoding of polarization effects in order to produce accurate binding geometries after minimization. Fig. 4 shows how the lengths of the ultra-short hydrogen bonds of the crystal structure of 1o3g increase to more standard hydrogen bond lengths after the protein preparation step. In addition, the high resolution X-ray structure (1.55 Å)

Table 3
Data set 3: predicted VSGB 2.0 ΔG_{bind}s for PDPK1 affinity data. Docked poses were generated for molecules with available affinity data but no crystal structures.

PDBID	pIC_{50} [a]	ΔG_{bind} [b] (exptl)	ΔG_{bind} (VSGB 2.0)
Fragment[c] 8 (3nus)[d]	3.5	−4.78	−43.78
Fragment 19 (3nun)	6.4	−8.77	−65.38
Fragment 20	5.6	−7.60	−60.91
Fragment 21	5.1	−7.00	−64.68
Fragment 22	4.9	−6.63	−64.12
Fragment 23	4.7	−6.37	−64.54
Fragment 24	4.6	−6.27	−59.65
Fragment25	4.9	−6.65	−59.54
Fragment11 (3nuu)	3.8	−5.14	−49.47
Fragment17 (3nuy)	3.5	−4.72	−42.97
8a[e]	6.0	−8.18	−66.58
8b	4.7	−6.41	−61.76
8c	5.1	−6.95	−65.31
8d	4.9	−6.68	−60.90
12a	6.9	−9.41	−65.47
12b	6.6	−9.00	−71.46
12c	6.7	−9.13	−70.25
12d	6.6	−9.00	−66.35
12e	7.1	−9.68	−69.18
12f (3qcq)	6.8	−9.27	−73.51
12g	6.9	−9.41	−75.02
12h	7.3	−9.95	−78.06
12i	7.7	−10.50	−73.72
12j	6.3	−8.59	−67.88
12k (3qcs)	6.5	−8.86	−75.96
12l	6.6	−9.00	−72.55
12m	6.6	−9.00	−76.28
(R)-18a (3qcx)	7.2	−9.81	−81.58
(S)-18a	5.9	−8.04	−75.44
18b	7.5	−10.22	−80.69
18c	6.7	−9.13	−79.43
18d	6.9	−9.41	−75.59
(S)-21a (3qcy)	7.7	−10.50	−94.43
(S)-21b	7.5	−10.22	−94.84
(S)-21c	7.2	−9.81	−101.46
(S)-21d	6.9	−9.41	−99.11
(3S,6R)-28a (3qd0)	8.5	−11.59	−104.09
(3S,6R)-28b	8.8	−12.00	−101.87
(3S,6R)-28c	8.1	−11.04	−114.06
(3S,6R)-28d	8.6	−11.72	−113.41
34a	8.3	−11.31	−99.67
34b	8.5	−11.59	−107.17
34c	8.5	−11.59	−107.25
34d	8.2	−11.18	−101.01
(R)-37b	7.6	−10.36	−101.74
(3R,6S)-37c (3qd3)	8.0	−10.91	−108.22
43a (3qd4)	8.2	−11.18	−109.63
43b	9.1	−12.40	−110.01
43c	9.2	−12.54	−112.22
(R)-21a[f]	6.0	−8.18	−100.82
(R)-21b[f]	6.9	−9.41	−101.89
(3R,6S)-28a[f]	5.9	−8.04	−90.99
(S)-37b[f]	6.0	−8.18	−101.13
(3S,6R)-37c[f]	6.2	−8.45	−93.54

[a] $pIC_{50} = -\log_{10}(IC_{50})$.
[b] $\Delta G_{bind} \approx RT\ln(K_i)$ or $RT\ln(IC_{50})$.
[c] Fragment numbering adapted from Medina et al. [24].
[d] PDB structures where available are mentioned in parentheses in bold font.
[e] Molecule numbering adapted from Medina et al. [25].
[f] Docked poses are unreliable due to high degree of extrapolation from available X-ray structures, these molecules were treated as outliers in the regression fit. All ΔG_{bind} values are in kcal/mol.

indicates the presence of a protonated His57. In fact, modeling a neutral His57 did not lead to a significant difference in the estimated binding free energy (not shown), implying that the error is more likely to be coming from the inability to model ultra-short hydrogen bonds, or even due to removal of the waters during binding free energy calculation. Even so, overall, the trypsin data-set seems to benefit most significantly from the improvements made to Prime MM/GBSA, as seen from the steady improvements in the R^2 values from 0.26 for SGB-NP to 0.48 for VSGB-1.0 to 0.68 for

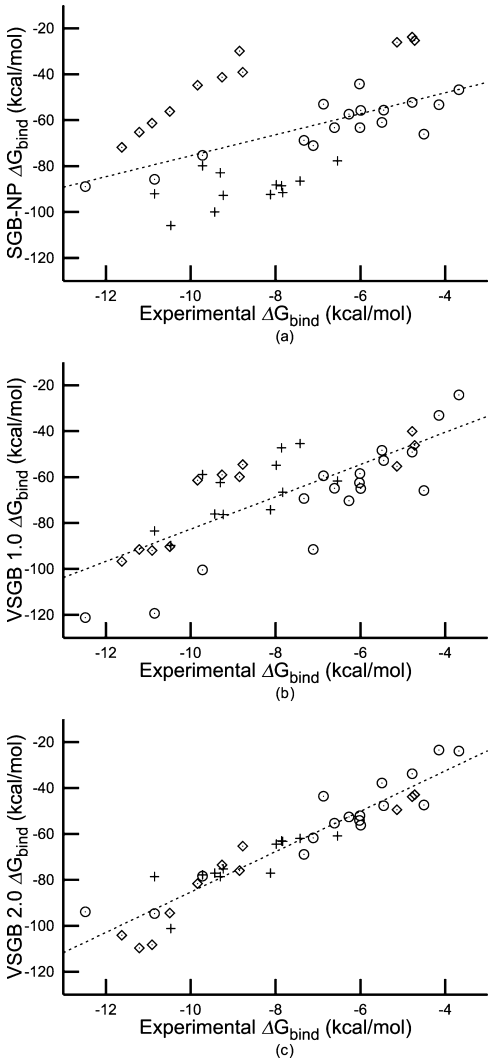

Fig. 1. Predicted ΔG_{bind}s and linear regression fits (dashed lines) of (a) SGB-NP-, (b) VSGB 1.0- and (c) VSGB 2.0 models for data set 1; (+) Bovine trpysin; (◯) BACE; (◊) PDPK-1. These plots show a global regression model emerging from VSGB 2.0 predictions that fits all the data irrespective of the protein or ligand chemotype.

VSGB-2.0. This is perhaps due to the fact that, of the four proteins studied here, only trypsin has a titratable histidine residue in its active site which benefits from the variable dielectric model as well as the improvements introduced to better account for π–π stacking interactions and directionality in hydrogen bonding.

Interestingly, as noted earlier, Fig. 1 shows the significant improvement in the predictivity of the successive GBSA models within Prime for data-set 1, with a linear regression $R^2 = 0.25$ for SGB-NP global regression model, which improves to an $R^2 = 0.60$ for VSGB-1.0 and further increases to a highly statistically significant $R^2 = 0.89$ for the VSGB-2.0 model. Since the data-set includes three different proteins, with significant diversity in ligand chemotypes, we cautiously suggest that VSGB-2.0 may be approaching the accuracy required for estimation of absolute binding free energies, albeit through a linear regression fit. As noted earlier, the observed global fit across three different targets was achieved with no conformational sampling of ligand, protein or complex. In the cases were significant reorganization of either the protein or the ligand from their respective global minima in solution is necessary for formation of the bound complex, a global model for estimation of ΔG_{bind} would not be achieved without accurately accounting for the reorganization energies and the solute entropic terms through

Table 4
Linear regression statistics for models generated from data sets 1, 2 and 3.

Data-set	R^2	σ[a]	F[b]	N[c]	Regression equation	
					Intercept	Slope
Local regression model – SGB-NP						
Trypsin	0.26	7.17	3.46	12	−62.83	3.10
BACE	0.74	6.66	41.67	17	−32.11	4.55
PDPK1 (only PDBs)	0.82	7.77	40.40	11	7.76	5.97
Aldose reductase	0.46	11.39	17.95	23	2.95	5.78
Local regression model – VSGB 1.0						
Trypsin	0.48	10.54	9.06	12	−2.17	7.36
BACE	0.88	9.68	106.65	17	2.60	10.59
PDPK1 (only PDBs)	0.74	11.12	25.26	11	−9.19	6.76
Aldose reductase	0.57	17.13	27.43	23	25.97	10.75
Local regression model – VSGB 2.0						
Trypsin	0.68	6.88	20.74	12	−9.78	7.27
BACE	0.88	7.27	114.39	17	0.43	8.23
PDPK1 (only PDBs)	0.92	7.60	99.60	11	2.47	9.17
PDPK1 (Data set 3)	0.81	8.67	204.54	49	1.07	8.89
Aldose reductase	0.63	14.98	38.13	23	21.70	11.09
Global regression model – SGB-NP						
Data set 1	0.25	19.18	12.31	40	−29.79	4.57
Data sets 1 & 2	0.14	19.42	9.56	63	−32.17	3.36
Global regression model – VSGB 1.0						
Data set 1	0.60	13.63	57.70	40	−12.39	7.03
Data sets 1 & 2	0.56	15.29	78.93	63	−6.36	7.59
Global regression model – VSGB 2.0						
Data set 1	0.89	7.33	312.46	40	2.63	8.80
Data sets 1 and 2	0.80	10.76	247.28	63	7.32	9.46
Data sets 1, 2 and 3	0.80	10.11	391.97	101	5.70	9.30

[a] σ is standard error of regression.
[b] F is the Fisher value.
[c] N is the size of the data-set.

conformational sampling. Our ability to uncover a global model in the absence of conformational sampling suggests that we may have unwittingly selected protein complexes for which the reorganization energies of both protein and ligand are low. Our method however, does account for the strain in the protein/ligand complex, which is also a component of the ligand reorganization energy.

In addition, in case of the PDPK1 data-set, we see the data fit to a global model in spite of using IC_{50}s instead of K_i's as measures of affinity. This implies that either the excellent fit we are observing with the global model is fortuitous, or the K_i for this data-set is within the 3-fold experimental variability that is common for reported IC_{50}'s. We attempted to estimate the ratio of IC_{50} and K_i for this data-set using the steady-state kinetic mechanism of PDPK1 reported by Gao et al. [47], and incorporating the reported concentrations of ATP (1uM) and the PDKtide (1 μM) used by Medina et al. [25]. Gao et al. [47] characterize the phosporylation of the PDK1-tide as a rapid equilibrium bi-bi system. Their reported values for various kinetic constants are $K_d{}^{ATP} = 0.54 \pm 0.15$ μM, $K_d{}^{Tide} = 1.0 \pm 0.3$ μM, $\alpha = 78 \pm 21$; and $\beta = 49 \pm 20$ (with ADP as inhibitor). Assuming that for the inhibitors studied here, β is same as for ADP, and using the following relationship for $IC_{50}/K_i = \beta(\alpha K_d{}^{ATP}(K_d{}^{Tide} + [\text{Tide}]) + [\text{ATP}](\alpha K_d{}^{Tide} + [\text{Tide}]))/(\alpha K_d{}^{ATP}(\beta K_d{}^{Tide} + [\text{P}])$, the calculated ratio of $IC_{50}/K_i = 3.8$. Even if we use the entire range of reported values in the literature for β of 1–49 [48], the ratio of IC_{50}/K_i varies from 1.9–3.8. It appears, therefore, that the reported IC_{50}s are within a few folds of the K_i values for the PDPK1 data-set, therefore justifying our use of IC_{50}s instead of K_is for this data-set.

For the global VSGB-2.0 model for data-set 1 the σ is 7.33 for a slope of 8.80, which translates to a standard deviation of 0.6 pIC_{50} points (~4-fold) (we used $RT\ln(K_i)$ instead of pIC_{50} in the regression fits). When data-sets 1 and 2 are clubbed together, σ is 10.76 for a slope of 9.46, which is equivalent to 0.83 pIC_{50} points (~7-fold), while for data-sets 1, 2 and 3 clubbed together, σ is 10.11 for a slope of 9.30, which is equivalent to 0.79 pIC_{50} points (~6.2-fold).

The lower accuracy achieved when docked poses are included can be partially accounted for by the fact that the docked complexes are less accurate than X-ray structures. Comparison of all the X-ray structures published for PDPK1 (51 in the PDB) shows significant flexibility in the amino acid side chains of the active site residues, such as the catalytic lysine (K111), L88, M134, V143, S160, E166, K169, E209, N210, T222 and D223, which is not accounted for while docking to a rigid protein. In addition, a second source for error could be due to the non-estimation of ligand reorganization energy, if we assume that the protein's reorganization energy is low, since they all bind the protein kinase in its DFG-in [49] conformation. In the current formulation of the binding free energy equation, the free energy of the solvated ligand is computed for the minimized ligand after extracting it from a bound complex. This conformation, is often not the global minimum conformer of the ligand in solution, [45] and for flexible ligands, the difference in energy between the minimized bioactive conformation and the global minimum conformation in solution could run into a few kcal/mol, sufficient to result in significant loss of potency. This difference is neglected by the MM/GBSA binding free energies calculated using Eq. (1).

In fact, in real life drug discovery scenarios, due to resource constraints, usually only a few representative compounds with good ligand efficiencies/potencies are prioritized for crystallography, and therefore it is conceivable that data-sets of affinity using only X-ray complexes may be biased to include ligands for which the reorganization energies (from their respective global minimum conformations in solution) are low. This bias, if it exists, will lead to good predictions and rankings for X-ray structures, but may over predict potency for design molecules with high ligand reorganization energies. Designed molecules, therefore, should be carefully analyzed for changes that might destabilize the free ligand's

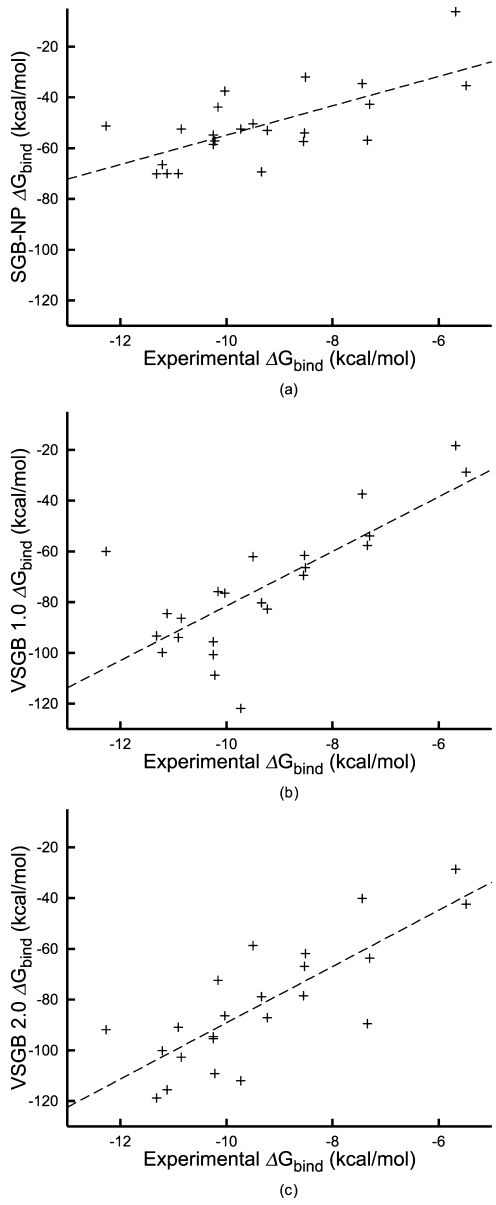

Fig. 2. Predicted ΔG_{bind}s (+) and linear regression fits (dashed lines) for the (a) SGB-NP-, (b) VSGB 1.0- and (c) VSGB 2.0 models for data set 2. VSGB 2.0 predictions are superior ($R^2 = 0.63$) when compared with either SGB-NP ($R^2 = 0.46$) or VSGB 1.0 ($R^2 = 0.57$).

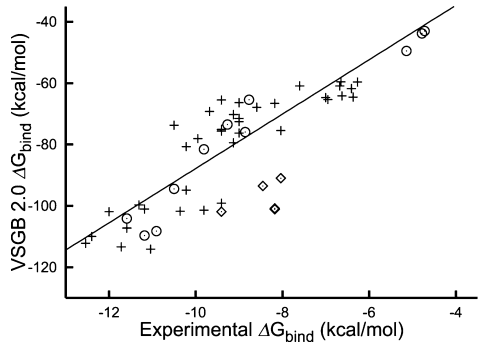

Fig. 3. Linear regression fit (solid line) between predicted and experimental ΔG_{bind}s for data-set 3 using the VSGB 2.0 model. (○) Data-points with available X-ray structures, (+) data points corresponding to docked structures with available X-ray complexes of structurally similar analogs, (◊) data-points with ligand structures that are significantly different from complexed ligands of available crystal structures, these data-points were treated as outliers in the linear regression fit.

'bioactive' conformation in solution. In such cases, even if VSGB-2.0 scores suggest high potency, false positives are possible due to not accounting for ligand reorganization costs.

For instance, a closer observation of the VSGB-2.0 global model shows that a majority (6 of 8) of the data points of higher affinity ($\Delta G > 10$ kcal/mol; 1 kcal/mol = 4.184 kJ/mol) are over-predicted (to the right of the regression line). This is perhaps due to the fact that for these larger, more flexible molecules, our protocol neglects the energetic costs of ligand reorganization and entropy. For flexible ligands, increased enthalpic components would have to offset entropic and reorganization costs to maintain ligand efficiency. Hence, it is conceivable that better enthalpic optimization of these potent compounds results in over-prediction of computed binding free energies. Also, the systematic error from under-sampling of the protein may cancel out as a difference in the binding free energy equation. However, such error correction is not likely for ligand re-organization energies.

A second consequence of the lack of conformational sampling is the high slope of the global regression equation (8.78). A high dynamic range of computed binding free energies when compared to experimental values has been observed by others [50], and has been attributed, among other factors, to the absence of conformational sampling and the associated neglect of the entropic terms in the free energy of binding. Nevertheless, these results clearly outline the dominant effect of the solvent [3,51,52] on the energetics of protein–ligand binding.

Importantly, these results provide support to the approach of injecting careful physics-based empiricism to achieve high computational efficiency while not compromising on accuracy. Interestingly, VSGB-2.0 was parameterized to optimize single side chain predictions of X-ray structures and validated on an associated but different problem, that of improving predictions of long loops in homology modeling. This implies that empiricism derived from a focus on more fundamental theory leads to solutions that can be extrapolated to auxiliary problems. Also significantly, this success is contingent upon incremental developments in associated problems in molecular modeling, such as the correct assignment of protonation states and improved accuracy of molecular mechanics force-fields [53].

In addition, since VSGB-2.0 is completely focused on improvements in the prediction of the energetics of proteins side-chains and loops, these developments will not help improve estimates of the solvation energetics of ligands. For, instance, in our experience (unpublished results) VSGB-2.0 does not sufficiently penalize hydrogen-bond donors buried in hydrophobic pockets, although they are believed to be highly detrimental to potency [54]. Since, ligands do not always bind in their lowest energy conformations [45], accurate estimations of ligand reorganization costs would be essential for detecting false positives. However, this would require estimating the relative energies of solvated conformers in the conformational ensemble of the ligand [55]. A ligand-centric effort at improving MM-GBSA estimates of solvation energies for drug-like molecules is therefore warranted.

4.1. The last mile problem of ligand binding

The PDPK-1 data-set. (data-set 3) highlights the importance of using accurate starting structures of modeled complexes in order to make reliable estimations of binding affinities. This implies that using MM-GBSA to re-score docked poses would be reasonable when the poses themselves are expected to be very accurate. Judicious use of structural biology is therefore recommended to correct predicted binding modes, as fragment hits are systematically grown to potent lead-like molecules. This is especially necessary when there is significant flexibility in the protein's structure. In PDPK-1, for instance, an analysis of all available crystal

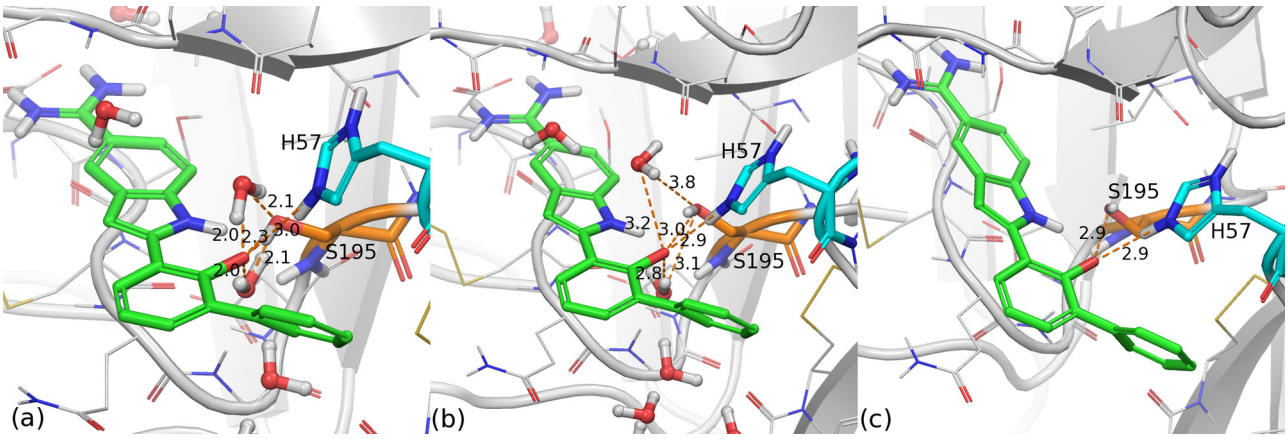

Fig. 4. The ligand of 1o3g is shown in green, H57 is in cyan and S195 is in orange; two waters HOH384 (top) and HOH383(bottom) are also shown in panels a and b (a) the network of ultra-short hydrogen bonds (<2.3 Å) in the X-ray structure of 1o3g are shown (b) after the protein preparation step the water molecules move and the hydrogen bond distances are in range of >2.8 Å, as encoded in common molecular mechanics force-fields (c) hydrogen bonding distances after the MM-GBSA step are shown, the two waters were deleted prior to minimization.

structures in the Protein Data Bank [56] shows significant flexibility in the side chains of the gatekeeper residue (V143), T222 adjacent to the gatekeeper, the catalytic lysine (K111), as well as residues of the glycine-rich loop. Even so, our results for PDPK-1 suggest that reasonably high accuracy of prediction can be achieved ($R^2 = 0.81$, $\sigma = 8.67 \approx 0.7$ pIC_{50}) when docked structures are carefully extrapolated from known X-ray structures. Structures that are expected to have higher error due to greater extrapolation from known structures should be excluded from the analysis to reduce noise.

Current docking methodologies generate reasonably accurate first guesses of binding modes [57], but are still significantly away from X-ray complexes for them to be problematic for accurate estimation of binding energies. Docking algorithms have used various metrics to measure their ability to accurately predict binding modes, with root mean square deviations (RMSDs) of <2.0 Å in ligand position from known binding modes being most popular. The present analysis demonstrates degradation in the R^2 values for estimated binding energies as we transition from crystal structures ($R^2 = 0.89$, Table 4) to docked poses ($R^2 = 0.81$). In another analysis, Singh et al. [58] show that calculated affinities deviate as much as $14.78 \pm 7.5\%$ even for highly accurate models (RMSD <2.0 Å in the protein side chains within binding-mode) when compared with those evaluated using the crystal coordinates directly. It appears, therefore, that an RMSD <2.0 Å is a very permissive metric for deviation from the X-ray coordinates for accurate estimation of binding free energies. Therefore, although very informative, docked poses lead to additional error in the estimated binding free energies.

This leads us to what we consider to be the last mile problem of protein/ligand binding – that of deriving accurate binding modes from cruder first approximations generated by docking algorithms. Generating more accurate binding complexes from docked poses would entail a complete sampling of conformational space in the vicinity of docked complexes along with high accuracy in the encoding of interaction energetics [59]. For instance, for the local regression fits, the accuracy of VSGB-2.0 is lower for the trpysin set ($R^2 = 0.68$, Table 4.) when compared to the other highly accurate data sets of BACE ($R^2 = 0.88$) and PDPK1 (only PDBs, $R^2 = 0.92$), which could, at least partially, be attributed to the network of ultra-short hydrogen bonds involving water molecules, that bridge the interactions between Trypsin (Fig. 4) and its ligands. Others have shown [60–63] that for solvent-accessible pockets, modeling of explicit solvent molecules may be unavoidable to generate correct binding modes. Current docking methods, therefore, need further refinement to generate more accurate binding modes, and induced-fit docking [64], solvated docking [65], quantum-polarized ligand docking [66,67] to account for the polarization effect of the protein on the ligand, and the use of molecular dynamics to refine docked complexes [68] are amongst significant efforts in that direction. Achieving computational efficiency in these calculations would be critical for their exploitation in virtual screening.

5. Conclusion

We show that the VSGB-2.0 MM-GBSA flavor of Prime 3.0 may be approaching the accuracy required for absolute binding free energy determination, albeit through linear regression and without any conformational sampling. Our results are consistent with recent reports by Greenidge et al. [11], which appeared during the preparation of this manuscript. We demonstrate the importance of controlling for affinity data variability and errors in the modeled complexes used for computing binding free energies, in order to recover the true accuracy of MM-GBSA. We would, however, like to assert that the data-set presented here is fairly limited and does not suggest that the regression equation obtained is generally applicable to any data-set, especially in the absence of conformational sampling. Nevertheless, the correlations extending across multiple targets make us hopeful that the MM-GBSA methodology, with its more controlled, physics-based injections of empiricism, is inching toward a more physically realistic representation of bio-molecular structures and their energetics. We conclude, optimistically, that the high accuracy achieved in binding free energy prediction at a modest computational cost implies that MM-GBSA may be poised toward generating physics-based scoring functions for docking, which should replace the rather uncontrolled approximations that are the norm in empirical scoring functions [2].

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2013.09.005.

References

[1] O. Guvench, A.D. MacKerell Jr., Computational evaluation of protein–small molecule binding, Curr. Opin. Struct. Biol. 19 (2009) 56–61.
[2] S.Y. Huang, S.Z. Grinter, X. Zou, Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions, Phys. Chem. Chem. Phys. 12 (2010) 12899–12908.
[3] M.K. Gilson, H.X. Zhou, Calculation of protein–ligand binding affinities, Annu. Rev. Biophys. Biomol. Struct. 36 (2007) 21–42.

[4] A. Onufriev, Continuum electrostatics solvent modeling with the generalized Born model, in: M. Feig (Ed.), Modeling Solvent Environments, Wiley-VCH, Weinheim, 2010, pp. 127–165.
[5] I. Massova, P.A. Kollman, Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding, Perspect. Drug Discov. Des. 18 (2000) 113–135.
[6] Prime, Version 3.0, Schrödinger LLC, New York, NY, 2011.
[7] A. Ghosh, C.S. Rapp, R.A. Friesner, Generalized Born model based on a surface integral formulation, J. Phys. Chem. B 102 (1998) 10983–10990.
[8] K. Zhu, D.L. Pincus, S. Zhao, R.A. Friesner, Long loop prediction using the protein local optimization program, Proteins: Struct. Funct. Bioinf. 65 (2006) 438–452.
[9] K. Zhu, M.R. Shirts, R.A. Friesner, Improved methods for side chain and loop predictions via the protein local optimization program: variable dielectric model for implicitly improving the treatment of polarization effects, J. Chem. Theory Comput. 3 (2007) 2108–2119.
[10] J. Li, R. Abel, K. Zhu, Y. Cao, S. Zhao, R.A. Friesner, The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling, Proteins: Struct. Funct. Bioinf. 79 (2011) 2794–2812.
[11] P.A. Greenidge, C. Kramer, J.-C. Mozziconacci, R.M. Wolf, MM/GBSA binding energy prediction on the PDBbind data set: successes failures and directions for further improvement, J. Chem. Inf. Model. 53 (2013) 201–209.
[12] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods. I. The accuracy of binding free energy calculations based on molecular dynamics simulations, J. Chem. Inf. Model. 51 (2011) 69–82.
[13] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods II. The accuracy of ranking poses generated from docking, J. Comput. Chem. 32 (2011) 866–877.
[14] C. Rapp, C. Kalyanaraman, A. Schiffmiller, E.L. Schoenbrun, M.P. Jacobson, A Molecular mechanics approach to modeling protein–ligand interactions: relative binding affinities in congeneric series, J. Chem. Inf. Model. 51 (2011) 2082–2089.
[15] C.R.W. Guimarães, M. Cardozo, MM-GB/SA rescoring of docking poses in structure-based lead optimization, J. Chem. Inf. Model. 48 (2008) 958–970.
[16] P.D. Lyne, M.L. Lamb, J.C. Saeh, Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring, J. Med. Chem. 49 (2006) 4805–4808.
[17] T.Y. Yang, J.C. Wu, C. Yan, Y. Wang, R. Luo, M.B. Gonzales, K.N. Dalby, P. Ren, Virtual screening using molecular simulations, Proteins: Struct. Funct. Bioinf. 79 (2011) 1940–1951.
[18] B. Kuhn, P. Gerber, T. Schulz-Gasch, M. Stahl, Validation and use of the MM-PBSA approach for drug discovery, J. Med. Chem. 48 (2005) 4040–4048.
[19] G. Rastelli, A.D. Rio, G. Degliesposti, M. Sgobba, Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA, J. Comput. Chem. 31 (2010) 797–810.
[20] Glide, Version 5.7, Schrödinger LLC, New York, NY, 2011.
[21] B.J. Eastwood, M.W. Farmen, P.W. Iversen, T. Craft, J.K. Smallwood, K.E. Garbison, et al., The minimum significant ratio: a statistical parameter to characterize the reproducibility of potency estimates from concentration-response assays and estimation by replicate-experiment studies, J. Biomol. Screen. 11 (2006) 253–261.
[22] B.A. Katz, K. Elrod, C. Luong, M.J. Rice, R.L. Mackman, P.A. Sprengeler, et al., A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site, J. Mol. Biol. 307 (2001) 1451–1486.
[23] B.A. Katz, K. Elrod, E. Verner, R.L. Mackman, C. Luong, W.D. Shrader, et al., Elaborate manifold of short hydrogen bond arrays mediating binding of active site-directed serine protease inhibitors, J. Mol. Biol. 329 (2003) 93–120.
[24] J.R. Medina, C.W. Blackledge, D.A. Heerding, N. Campobasso, P. Ward, J. Briand, et al., Aminoindazole PDK1 inhibitors: a case study in fragment-based drug discovery, ACS Med. Chem. Lett. 1 (2010) 439–442.
[25] J.R. Medina, C.J. Becker, C.W. Blackledge, C. Duquenne, Y. Feng, S.W. Grant, et al., Structure-based design of potent and selective 3-phosphoinositide-dependent kinase-1 (PDK1) inhibitors, J. Med. Chem. 54 (2011) 1871–1895.
[26] C.W. Murray, O. Callaghan, G. Chessari, A. Cleasby, M. Congreve, M. Frederickson, et al., Application of fragment screening by X-ray crystallography to β-secretase, J. Med. Chem. 50 (2007) 1116–1123.
[27] Y. Cheng, T.C. Judd, M.D. Bartberger, J. Brown, K. Chen, R.T. Fremeau, et al., From fragment screening to in vivo efficacy: optimization of a series of 2-aminoquinolines as potent inhibitors of beta-site amyloid precursor protein cleaving enzyme 1 (BACE1), J. Med. Chem. 54 (2011) 5836–5857.
[28] V. Calderone, B. Chevrier, M. Van Zandt, V. Lamour, E. Howard, A. Poterszman, et al., The structure of human aldose reductase bound to the inhibitor IDD384, Acta Crystallogr. D 56 (2000) 536–540.
[29] T. Kinoshita, H. Miyake, T. Fujii, S. Takakura, T. Goto, The structure of human recombinant aldose reductase complexed with the potent inhibitor zenarestat, Acta Crystallogr. D 58 (2002) 622–626.
[30] O. El-Kabbani, C. Darmanin, T.R. Schneider, I. Hazemann, F. Ruiz, M. Oka, et al., Ultrahigh resolution drug design II. Atomic resolution structures of human aldose reductase holoenzyme complexed with Fidarestat and Minalrestat: implications for the binding of cyclic imide inhibitors, Proteins: Struct. Funct. Bioinf. 55 (2004) 805–813.
[31] F. Ruiz, I. Hazemann, A. Mitschler, A. Joachimiak, T. Schneider, M. Karplus, A. Podjarny, The crystallographic structure of the aldose reductase-IDD552 complex shows direct proton donation from tyrosine 48, Acta Crystallogr. D 60 (2004) 1347–1354.
[32] E. Howard, R. Sanishvili, R. Cachau, A. Mitschler, B. Chevrier, P. Barth, et al., Ultrahigh resolution drug design I: details of interactions in human aldose reductase–inhibitor complex at 0.66 Å, Proteins: Struct. Funct. Bioinf. 55 (2004) 792–804.
[33] O. El-Kabbani, C. Darmanin, M. Oka, C. Schulze-Briese, T. Tomizaki, I. Hazemann, et al., High-resolution structures of human aldose reductase holoenzyme in complex with stereoisomers of the potent inhibitor Fidarestat: stereospecific interaction between the enzyme and a cyclic imide type inhibitor, J. Med. Chem. 47 (2004) 4530–4537.
[34] H. Steuber, M. Zentgraf, A. Podjarny, A. Heine, G. Klebe, High-resolution crystal structure of aldose reductase complexed with the novel sulfonyl-pyridazinone inhibitor exhibiting an alternative active site anchoring group, J. Mol. Biol. 356 (2006) 45–56.
[35] H. Steuber, M. Zentgraf, C. Gerlach, C.A. Sotriffer, A. Heine, G. Klebe, Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions, J. Mol. Biol. 363 (2006) 174–187.
[36] T. Petrova, S. Ginell, A. Mitschler, I. Hazemann, T. Schneider, A. Cousido, et al., Ultrahigh-resolution study of protein atomic displacement parameters at cryotemperatures obtained with a helium cryostat, Acta Crystallogr. D 62 (2006) 1535–1544.
[37] H. Steuber, A. Heine, G. Klebe, Structural and thermodynamic study on aldose reductase: nitro-substituted inhibitors with strong enthalpic binding contribution, J. Mol. Biol. 368 (2007) 618–638.
[38] J.M. Brownlee, E. Carlson, A.C. Milne, E. Pape, D.H.T. Harrison, Structural and thermodynamic studies of simple aldose reductase–inhibitor complexes, Bioorg. Chem. 34 (2006) 424–444.
[39] H. Steuber, M. Zentgraf, C. La Motta, S. Sartini, A. Heine, G. Klebe, Evidence for a novel binding site conformer of aldose reductase in ligand–bound state, J. Mol. Biol. 369 (2007) 186–197.
[40] M. Eisenmann, H. Steuber, M. Zentgraf, M. Altenkämper, R. Ortmann, J. Perruchon, et al., Structure-based optimization of aldose reductase inhibitors originating from virtual screening, ChemMedChem 4 (2009) 809–819.
[41] M.C. Van Zandt, B. Doan, D.R. Sawicki, J. Sredy, A.D. Podjarny, Discovery of [3-(4,5,7-trifluoro-benzothiazol-2-ylmethyl)-pyrrolo[2,3-b]pyridin-1-yl] acetic acids as highly potent and selective inhibitors of aldose reductase for treatment of chronic diabetic complications, Bioorg. Med. Chem. Lett. 19 (2009) 2006–2008.
[42] Protein Preparation Wizard-Epik, Version 2.2, Impact, Version 5.7, Prime, Version 3.0, Schrödinger LLC, New York, NY, 2011.
[43] G.M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, J. Comput. Aided Mol. Des. (2013), http://dx.doi.org/10.1007/s10822-013-9644-8.
[44] C. Mulakala, Y.N. Kaznessis, Path-integral method for predicting relative binding affinities of protein–ligand complexes, J. Am. Chem. Soc. 131 (2009) 4521–4528.
[45] E. Perola, P.S. Charifson, Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding, J. Med. Chem. 47 (2004) 2499–2510.
[46] K.T. Butler, F. Luque, X. Barril, Toward accurate relative energy predictions of the bioactive conformation of drugs, J. Comput. Chem. 30 (2009) 601–610.
[47] X. Gao, T.K. Harris, Steady-state kinetic mechanism of PDK1, J. Biol. Chem. 281 (2006) 12681–21670.
[48] P. Chène, Can biochemistry drive drug discovery beyond simple potency measurements, Drug Discov. Today 17 (2012) 388–395.
[49] F. Zuccotto, E. Ardini, E. Casale, M. Angiolini, Through the gatekeeper door: exploiting the active kinase conformation, J. Med. Chem. 53 (2010) 2681–2694.
[50] C.R.W. Guimarães, A.M. Mathiowetz, Addressing limitations with the MM-GB/SA scoring procedure using the WaterMap method and free energy perturbation calculations, J. Chem. Inf. Model. 50 (2010) 547–559.
[51] L. Wang, B.J. Berne, R.A. Friesner, Ligand binding to protein-binding pockets with wet and dry regions, Proc. Natl. Acad Sci. USA 108 (2011) 1326–1330.
[52] P.W. Snyder, J. Mecinović, D.T. Moustakas, S.W. Thomas III, M. Harder, E.T. Mack, et al., Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase, Proc. Natl. Acad Sci. USA 108 (2011) 17889–17894.
[53] D. Shivakumar, E. Harder, W. Damm, R.A. Friesner, W. Sherman, Improving the prediction of absolute solvation free energies using the next generation OPLS force field, J. Chem. Theory Comput. 8 (2012) 2553–2558.
[54] C. Bissantz, B. Kuhn, M. Stahl, A medicinal chemist's guide to molecular interactions, J. Med. Chem. 53 (2010) 5061–5084.
[55] J. Kongsted, P. Söderhjelm, U. Ryde, How accurate are continuum solvation models for drug-like molecules? J. Comput. Aided Mol. Des. 23 (2009) 395–409.
[56] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, et al., The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[57] D. Plewczynski, M. Łaźniewski, R. Augustyniak, K. Ginalski, Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database, J. Comput. Chem. 32 (2011) 742–755.
[58] M.K. Singh, B.N. Dominy, Thermodynamic resolution: How do errors in modeled protein structures affect binding affinity predictions, Proteins: Struct. Funct. Bioinf. 78 (2010) 1613–1617.
[59] D.L. Mobley, K.A. Dill, Binding of small-molecule ligands to proteins: "What You See" Is Not Always "What You Get", Structure 17 (2009) 489–498.
[60] C.R. Corbeil, C.I. Williams, P. Labute, Variability in docking success rates due to dataset preparation, J. Comput. Aided Mol. Des. 26 (2012) 775–786.

[61] R. Abel, N.K. Salam, J. Shelley, R. Farid, R.A. Friesner, W. Sherman, Contribution of explicit solvent effects to the binding affinity of small-molecule inhibitors in blood coagulation factor serine proteases, ChemMedChem 6 (2011) 1049–1066.
[62] J. Michel, J. Tirado-Rives, W.L. Jorgensen, Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization, J. Am. Chem. Soc. 131 (2009) 15403–15411.
[63] T. Beuming, R. Farid, W. Sherman, High-energy water sites determine peptide binding affinity and specificity of PDZ domains, Prot. Sci. 18 (2009) 1609–1619.
[64] W. Sherman, T. Day, M.P. Jacobson, R.A. Friesner, R. Farid, Novel procedure for modeling ligand/receptor induced fit effects, J. Med. Chem. 49 (2006) 534–553.
[65] A.D.J. Van Dijk, A.M.J.J. Bonvin, Solvated docking: introducing water into the modelling of biomolecular complexes, Bioinformatics 22 (2006) 2340–2347.
[66] A.E. Cho, V. Guallar, B.J. Berne, R. Friesner, Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach, J. Comput. Chem. 26 (2005) 915–931.
[67] J. Du, H. Sun, L. Xi, J. Li, Y. Yang, H. Liu, X. Yao, Molecular modeling study of checkpoint kinase 1 inhibitors by multiple docking strategies and prime/MM-GBSA calculation, J. Comput. Chem. 32 (2011) 2800–2809.
[68] G. Rastelli, G. Degliesposti, A. Del Rio, M. Sgobba, Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening, Chem. Biol. Drug Des. 73 (2009) 283–286.