Rescoring Virtual Screening Results with the MM-PBSA Methods: Beware of Internal Dielectric Constants

Xiao Hu and Alessandro Contini[*]

Department of Pharmaceutical Science, Università degli Studi di Milano, 20133 Milan, Italy

E-mail: alessandro.contini@unimi.it

Abstract

With the potential of improving virtual screening outcome, MM-PB/GBSA has become a disputed method that requires extensive testing and tuning to provide the optimal results. One of the tuning factors is the internal or solute dielectric constant. We have applied three test sets with receptors of different categories and libraries from different sources to investigate the underlying issue related to this constant. We discovered that increasing internal dielectric value does not improve the virtual screening enrichment qualitatively. More interestingly, non-polar and polar calculated energies act differently in libraries with different molecular weight distributions. From this work, the performance of MM-PBSA rescoring in virtual screening is more library-than receptor-dependent.

Introduction

Virtual screening has become a widely applied method at the beginning of drug discovery projects.[1–4] Success in the prediction of hit candidates is closely related to the scoring func-

tion applied. Accuracy in predicting absolute binding free energy is not of priority; the capability of the scoring function to distinguish active from inactive compounds plays the utmost crucial part. Multiple scoring functions have been developed to satisfy such purpose. Innumerable testings, fittings, and comparisons have been carried out to prove each the more advantageous method over the others.

Due to the nature of the more intricate parameterizations, the MM-PB/GBSA free energy calculation is potentially a better tool than most scoring functions. Since its debut in the late 90s,[5,6] it has been validated on a wide variety of biomolecular systems. Both the MM-PBSA and the MM-GBSA yield moderate to good correlations to experimental data, though performance varies among different protein receptors.[7–23] A comparison to 11 scoring functions showed that the MM-GBSA method provides improved Spearman correlation coefficients to experimentally determined IC_{50}.[13] Moreover, multiple studies have also investigated the rescoring performance of the MM-PB/GBSA in virtual screening.[8,10,17,20,22] The MM-PB/GBSA model was demonstrated with significant improvements in enrichment, comparing to virtual screening carried out using Autodock program[24,25] in several systems including dihydrofolate reductase,[10] tyrosine kinases,[17,22] aldose reductase, and factor X.[17] The MM-GBSA free energy rescoring method also outperform Autodock Vina (vina hereafter)[26] in multiple DUD-E[27] protein targets.[20] Furthermore, one assessment using the mineralocorticoid receptor set showed that the MM-PBSA method did not provide better outcome than the molecular mechanics (MM) alone in rescoring virtual screening results of the mineralocorticoid receptor target.[8]

More recent studies also attempted the possibility of including explicit solvation for the MM-PB/GBSA calculations.[23,28–30] Maffucci and co-workers developed a Nwat-MMGBSA workflow by including numbers of closest water molecules centered on the ligands while complexed with the receptors.[23,30] Assessments on several systems including penicillopepsin, HIV1-protease, BCL-XL, Rac1, and β-lactamase showed significant improvements in virtual screening over standard MM-GBSA rescoring.[23,30]

The conventional MM-PB/GBSA method developed on averaging the free energy estimation of a collection of conformations. However, it was suggested that the effect of performing short MD simulation prior to the MM-PB/GBSA calculation was marginal. Studies have shown that energy minimization provided a similar outcome as averaging over short MD simulation in both docking and virtual screening assessments.[10,21,22] On the other hand, earlier works advised that some systems might require short time-scale MD for improved performance, if the docking method fails to correctly predict the binding poses.[13,19]

In addition, the MM-PB/GBSA rescoring methods also displayed some charging dependency using different partial charge models. Oehme and coworkers investigated the effect of multiple atomic partial charge methods including empirical (Gasteiger), semi-empirical (AM1-BCC), and restrained electrostatic potential (RESP) charges computed by a variety of single-point quantum mechanical methods. Six inhibitors of HIV protease were considered and the outcome showed that Gasteiger, HF/STO-3G, and B3LYP/cc-pVTZ provided the best correlations to experimental data.[16] Another work commenced by Xu et al. also suggested that the RESP charges with HF-SCF/6-31G* level of calculation gave the optimal results, while those of AM1-BCC and ESP were also competent across 5 different biomolecular systems.[19] In addition, a comparison between MM-GBSA and linear interaction energy on cucuerbit[8]uril and α-cyclodextrin systems resulted in similar predicted relative affinity between AM1-BCC, RESP, and xAvESP.[15]

The alternative choices of protein dielectric constant have also been expounded.[14,22] Yang et al. demonstrated that improved correlations to experimental data among ligands targeting 7 classes of proteins were observed with increased internal dielectric value.[14] Genheden and Ryde also confirmed the relation between increased solute dielectric to the improved MM-GBSA performance in factor Xa.[18] In a later study, Sun and co-workers also showed agreeable results in both docking and virtual screening among three tyrosine kinases, though choices of the internal dielectric values might vary between different systems.[22] Additionally, These discoveries indirectly shed lights on the importance of using relative better electrostatics in

the MM-PB/GBSA rescoring.

In this work, we investigated the roles played by different energy compositions within MM-PB/GBSA free energy estimations in virtual screening. We firstly compared PLANTS[31,32] and vina[26] for docking pose sampling. With the chosen dataset, the former provided a better agreement to experiments and was used as the method of choice prior to MM-PB/GBSA rescoring. The quality of performance was evaluated according to the capability of the rescoring method to distinguish actives from inactives or decoys. Comparisons in enrichment outcomes were performed for the polar and electrostatic contributions, the non-polar energetics, and the total binding free energy. Two popular and computationally efficient charging methods (Gasteiger and AM1-BCC) were also investigated for their influences on the electrostatics. The assessments were performed on multiple test sets derived from CCDC-Astex,[33] DUD-E,[27] and ChEMBL[34] datasets, consisting of a larger variation in protein categories. The conclusions provide further insights in applying the MM-PB/GBSA model in virtual screening rescoring.

Methods

Test set preparations

Several test sets were adopted for different validations for this work, as listed in Table 1. Test set 0 and 0' were adopted for the validation of docking performance in pose prediction with the package of choice in this study. Test set 0 was derived from the CCDC-Astex clean set[33] and test set 0' from the DUD-E target set.[27] All complexes were processed using the MOE software.[35] Any loop missing more than 4 amino acids was truncated and capped with acetyl or N-methyl amide groups. Otherwise, the loops were rebuilt during the structure preparation process. All unnatural termini were capped to avoid artifacts from charges. Moreover, all missing loops were relatively far from the ligand binding site hence the above processing should not perturb the binding site significantly. The binding site with a distance

Table 1: Test sets applied in this work.

Index	Size	Source	Protein contained (PDB ID (lowercase)/gene name (uppercase)
0	74	The CCDC-Astex clean set for pose prediction validation	1a28, 1a4q, 1a6w, 1abe, 1abf, 1aoe, 1apu, 1aqw, 1atl, 1bma, 1byb, 1c5c, 1c5x, 1c83, 1cbs, 1cil, 1cle, 1d0l, 1d3h, 1ejn, 1eta, 1f3d, 1flr, 1glp, 1glq, 1hfc, 1hpv, 1hsb, 1hsl, 1hvr, 1hyt, 1ida, 1jap, 1kel, 1lcp, 1lic, 1lst, 1mld, 1mmq, 1mrg, 1mrk, 1mts, 1nco, 1ppc, 1pph, 1qbr, 1rnt, 1rob, 1slt, 1snc, 1srj, 1tmn, 1tng, 1tnh, 1tni, 1tnl, 1tyl, 1ukz, 1xid, 1xie, 2ak3, 2cmd, 2ctc, 2fox, 2gbp, 2h4n, 2qwk, 2tmn, 3cla, 3ert, 4dfr, 5abp, 6rnt, 7tim
0´	78	The DUD-E targets	1b9v, 1bcd, 1c8k, 1d3g, 1e66, 1h00, 1j4h, 1l2s, 1mv9, 1njs, 1q4x, 1sj0, 1sqt, 1syn, 1udt, 1uyg, 1vso, 1w7x, 1xl2, 1ype, 2aa2, 2am9, 2ayw, 2azr, 2b8t, 2etr, 2gtk, 2hzi, 2i0e, 2i78, 2ica, 2nnq, 2of2, 2oi0, 2oj9, 2ojg, 2owb, 2oyu, 2p2i, 2p54, 2qd9, 2rgp, 2zdt, 2zec, 2znp, 3biz, 3bqd, 3bwm, 3bz3, 3c4f, 3ccw, 3cqw, 3d0e, 3d4q, 3e37, 3el8, 3eqh, 3f9m, 3g0e, 3g6z, 3hl5, 3hmm, 3kba, 3kgc, 3kl6, 3krj, 3l3m, 3l5d, 3lan, 3ln1, 3lpb, 3lqb, 3m2w, 3nf7, 3nxo, 4trj, 830c
1	18	Active and inactive ligands from DUD-E	ACES, CDK2, DPP4, EGFR, ESR1, FA10, FGFR1, FNTA, GRIA2, GRIK1, HIVPR, KPCB, LCK, PTN1, SRC, THRB, TRY1, VGFR2
2	19	Clustered active and inactive ligands from ChEMBL	ACES, ADORA1, AURKA, CA1, CASP1, CDK2, CTSD, EGFR, ELANE, ESR1, FGFR1, GSK3B, LCK, MAPK1, MMP1, NTRK1, PPARG, SRC, VGFR2
3	18	Same proteins as in test set 1 but using the 50 best active compounds and a drug-like 1000 decoys set[a]	Same as test set 1

[a] A decoys set from previous studies.[36,37]

range of 4.5 Å from the ligand was minimized briefly with the AMBER 10EHT force field and the Born implicit solvent included within MOE package.[35] The ligands were then stripped from the binding sites. The docking process was then repeated 12 times using the docking programs (PLANTS[31,32] or vina[26]) of choices with the original crystal ligands.

Test sets 1-3 were all adopted for virtual screening assessments. The receptors in the test set 1 were selected from DUD-E database[27] based on the number of true inactive ligands available; if this number is larger than 100 (except HIVPV which have an inactive count of 96) the receptor is included in the test set 1. The libraries included in this test set consists

of the active and inactive compounds available from DUD-E database.[27] The libraries were downloaded as they were in SMILES format and ligand 3D structures were generated and minimized using MOE.[35]

The test set 2 includes active and inactive ligands from the ChEMBL database[34] for the receptor proteins selected. Both the active and inactive sub-libraries were generated according to available experimental data. The available inhibition constant (K_i) values were considered first as they are more directly connected to the binding affinities of the compounds to the receptors. The ligands considered as the inactive in the experimental method of choice were selected into the inactive list. Moreover, one criterion for the receptor to be considered as a candidate in the test set 2 is that the size of inactive ligands should be above 200. This ensured that at least 10 and no more than 20 active compounds can be included in the final library, maintaining an active-to-inactive ratio no larger than 1:10. If the number of inactives derived from the K_i lists does not satisfy the 200-inactive criterium, as were the cases for CTSD and PPARG, the data of IC$_{50}$ were also considered. Both active and inactive ligands were then filtered to have molecular weights between 250 and 550. To remove repetitions and compounds that are too similar to each other within the libraries, each inactive library was clustered using the Tanimoto superset/subset similarity metrics on the bit-packed MACCS structural keys (MOE package[35]).

The test set 3, with the same receptors selected for the test set 1, combined the 50 most active compounds of each receptor from DUD-E database[27] with a 1000-decoy set with an average molecular weight at 400 from PDB bank applied in a previous study.[36,37] We did not use the decoys provided within the DUD-E database[27] due to the larger computational time required. Moreover, we expect this universal decoy application across different receptors would put more emphasis on the assessment of a more generalized scoring performance.

The actives in all of the test sets were selected as the n-most active compounds, from the final active information from the DUD-E database[27] or after the clustering on the ChEMBL active compounds.[34] The n is the targeted size satisfying a relative active-to-inactive ratio

according to the test set the receptor belongs. The 3D structures of the ligands were then generated from the SMILES format and minimized using MOE package.[35] The numbers of actives and inactives used for the test set 1 and 2 are listed in Table S1 and S2.

The receptor structures of all the test sets were downloaded from the Protein Data bank[38] and are listed in Tables S1 and S2. The water, salts, metals, and co-crystal reagents were removed. All post-translational modifications were far away from the binding sites and were thus all removed. Some amino acid residue names were corrected for compatibility with the *leap* program of the AMBER package.[39] The hydrogen atoms were added or corrected by using the Protonate 3D function within the MOE package.[35] In cases where crystal ligands or metals were presented within the binding site, the sites were visually inspected in case of erred hydrogen placements. The centers of the binding sites were determined using the PLANTS program.[31,32] The radii of the docking search sphere were derived by adding an additional 2 Å onto the half of the longest intramolecular atom-to-atom distance determined from the processed screening library per receptor.

Docking and virtual screening

Docking and virtual screening for all test sets (including test set 0 and 0') were performed with the same procedures as follow. Multiple reasonable tautomeric and protonation states of ligands were generated under normal physiological conditions using the UNICON program[40] for virtual screening. Only the top scored states were preserved for screening. The docking procedures were then carried out using PLANTS.[31,32] The ligand atoms were kept flexible while the ones for receptors were fixed. The cluster RMSD was adjusted to 1.0 and the sampling factor sigma was set to 7.0 in the PLANTS configuration files to improve pose sampling. The poses were initially scored with the $PLANTS_{chemplp}$ scoring functions.[31,32] The three top-ranked docking poses for each tautomeric and/or protonation state were saved for later MM-PBSA rescoring. All virtual screening and rescoring were initially performed in triplicate. However, we observed negligible variations between different runs. Hence, the

results presented in this work are shown for only one replicate.

Energy minimisation of the docked poses

Due to the empirical nature of the PLANTS scoring,[31,32] occasional structural clashes were observed in the final poses. Moreover, the fixed receptor structures would possibly introduce errors to the final pose accuracy. Hence, brief energy minimization of the binding poses was included prior to the MM-PBSA single-frame calculation for structure relaxation. The minimization was performed using the AMBER simulation package.[39] The ligands were charged with the AM1-BCC method[41] provided by antechamber[42] or the Gasteiger charges from Autodock.[25] The *ff14SB* force field[43] was applied to the proteins and *gaff*[42] to the ligands.

The minimization was performed with the GB-Neck(II) implicit solvent model ($igb = 8$) using *pmemd* of the Amber14 package.[39] The recommended mbondi3 radii[44] were applied during the preparation with *tleap*. The minimization switched from steepest descent to conjugate gradient after 500 cycles. The maximum number of minimization cycles was set to 5000 and the convergence criterion was relaxed to 0.1 kcal·mol^{-1}·Å^{-1}. A positional restraint with a weight of 100 kcal·mol^{-1}·Å^{-1} was exerted onto atoms with a distance more than 4.5 Å away from the ligand. These procedures were carried out for all the saved poses from the docking procedures.

MM-PB/GBSA rescoring

The binding free energy score was calculated using the MM-PB/GBSA methods. The scheme has been well summarised elsewhere[45,46] and the simplified cycle can be briefly described as below:

$$\Delta G_{binding} = G_{complex} - G_{receptor} - G_{ligand}$$

The free energies of complex ($G_{complex}$), receptor ($G_{receptor}$), and ligand (G_{ligand}) were calculated from a single frame, i.e. the minimised structures in this work. Free energies of complex, receptor, and ligand are calculated by a summation of multiple energy terms:

$$G_x = E_{MM} + G_{PB\text{-}polar} + G_{PB\text{-}nonpolar} - TS$$

The E_{MM} represents the gas phase molecular mechanic contributions calculated using force field definition, while G_{polar} is the polar solvation term obtained by numerically solving either the Poisson-Boltzmann or the generalized Born equations. The non-polar solvation term ($G_{non\text{-}polar}$), on the other hand, is calculated using a linear relation involving solvent accessible surface area (SASA) in the form as

$$G_{PB\text{-}nonpolar} = \gamma * SASA + b$$

with γ and b taking different values according to radii applied to atoms within the system.

The entropic contribution ($-TS$) is an important element in thermodynamic calculations. However, it has also been a debatable contribution to be included within the MM-PB/GBSA calculation. It was suggested that similar entropic contributions would likely result from similar ligand structures targeting the same receptor.[6,7,9] Oehme et al. confirmed that the entropic indiscrimination also extended to ligand with larger structural variations.[16] Moreover, the high computational expenses will very likely impair the efficiency of virtual screening. Hence, the entropic term is not considered in this work.

The Perl script *mm_pbsa.pl* provided by AMBER suite was applied for MM-PB/GBSA single-framed calculation.[39] The sander *pbsa* program was used to calculate electrostatic contribution by solving the Poisson-Boltzmann equation. The external dielectric value was set to 80.0. The rescoring processes were carried out with multiple internal/solute dielectric values as 1, 2, 4, and 6, respectively. Dielectric value for molecular mechanic (MM) part of energy evaluation was also adjusted accordingly. An ionic strength of 150 mM was used. The

Bondi radii were applied for MM-PBSA calculation and missing or wrong radii were added or corrected.[47,48] The mbondi2 set of radii were applied for the MM-GBSA calculations using the GB^{OBC} model ($igb = 5$).[49] The γ and b values were selected as 0.00542 and 0.92 for PB and 0.005 and 0.0 for GB.

All the poses generated from PLANTS virtual screening were re-evaluated with the MM-PB/GBSA method. The ligands were ranked with the value of the lowest free energy calculated from all the rescored poses and receiver operating characteristic area under curve (ROC AUC) values were calculated for assessment on enrichment performance. In this work, we also separately considered the enrichment performance of total non-polar and polar calculated energies consisting in the total MM-PB/GBSA calculated energies:

$$\Delta G_{total} = \Delta G_{polar} + \Delta G_{non\text{-}polar}$$

where

$$\Delta G_{polar} = \Delta G_{ele(gas\ phase)} + \Delta G_{polar(PB/GB)}$$

$$\Delta G_{non\text{-}polar} = \Delta G_{non\text{-}polar(gas\ phase)} + \Delta G_{non\text{-}polar(PB/GB)}$$

The energies were calculated by combining the gas phase and the corresponding PB/GB terms. This decomposition process gives the non-polar, electrostatic-polar, and total calculated energies from MM-PBSA estimation in the later analyses.

Molecular weight adjustment procedures

On a later stage of our work, we also adjusted the molecular weight (MW) composition of the test set 2 and 3 to lower the biases due to differences in molecular size between active and inactive ligands (or decoys). We did not discriminate this bias in the early studies since the assessment of MW-biased system should not be excluded. In an actual virtual screening scenario, such bias is a question that worth to be addressed. On the other hand, it was

deemed necessary to validate the unbiased circumstances. We expected that combining the two studies would provide a more complete picture of the actual performance of the MM-PBSA rescoring in virtual screening. In addition, we did not apply the same treatments to test set 1 since most of the libraries were too small to be considered for MW re-adjustment. Additional shrinking would have made the later analysis statistically meaningless. Some receptors were also removed from the analysis in the test set 2 as the number of ligands within the library has dropped to values that were too low after MW processing.

To lower the biasing effect from MW, we focused on lowering the ROC AUC values from high-to-low MW rankings of ligands within the library to about 0.5 or lower. To achieve this, the MW range of active ligands were first determined and the 30 and 100 percentile of this range was recorded as the lower and upper bound of the MW cut-offs. The inactives with MW outside the boundaries were then excluded and the size of the remaining inactive ligands was then determined; if the size was smaller than 70, the corresponding receptor was excluded from the later analysis. After determining the MW-unbiased inactive/decoy list, the size of the actives was then determined according to the size of remaining inactives to give a reasonable active-to-inactive ratio (typically around 1:10 or 1:20). The active ligands for the MW-unbiased library were then selected randomly from the original active list, resulting in an MW ROC AUC of about 0.5 or less.

Results and Discussions

Validation of PLANTS performance

The prediction of correct poses within the binding site is one of the crucial requirements to achieve a satisfactory ranking from later analyses. We are aware that multiple docking software are available,[50–54] while our extensive previous experiences were with Autodock/vina[55,56] and PLANTS.[30] Additionally, since multiple MM-PBSA rescoring studies have applied Autodock/vina as the docking software,[17,20,21,28] we therefore compared vina and PLANTS in

Table 2: The success rates of reproducing crystal poses of test sets 0 and 0' using PLANTS[31,32] and vina.[26]

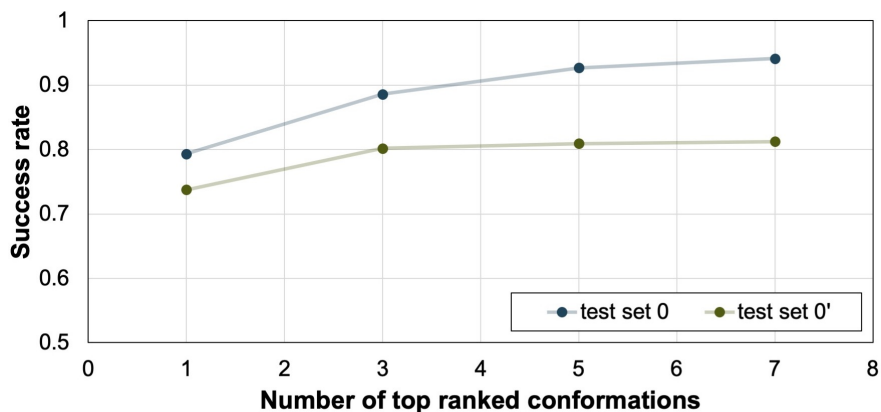| Docking method	Test set 0	Test set 0'
PLANTS	79.3 %	73.5 %
vina, exh = 10	61.7 %	62.7 %
vina, exh = 12	62.7 %	63.6 %

Figure 1: The graph showing the success rate of docking experiments when including different numbers of top ranked conformations using PLANTS for test 0 and 0'.

this study to demonstrate that PLANTS[31,32] is a compatible software for pose sampling. The performance on pose predictions was tested using 74 crystal complexes from CCDC-Astex clean set[33] against those generated by vina.[26] Moreover, since PLANTS scoring function was trained using the CCDC-Astex set, an additional test set derived from DUD-E[27] targets was also included. We also considered monomers separately, as some minor variations within the crystal structure might incur. This resulted in a total of 98 individual binding sites in test set 0 and 108 binding sites in test set 0' to dock the crystal ligands back into. For vina, testings were done for exhaustiveness at both 10 and 12.

The results showed that PLANTS performs significantly better in pose prediction. 79.3 % and 73.8 % of the 1st top ranked conformations were with an RMSD less than 2.0 Å to the original crystal poses for test set 0 and 0', respectively. On the other hand, vina gave lower success rate among the best poses using the same test sets (61.6 - 63.6 %, Table 2) for both exhaustiveness 10 and 12.

When more than one top-ranked pose were included in final results using PLANTS docking, the reproducibility of cystal poses improved accordingly (Figure 1). However, after increasing to more than 3 top ranked conformations in the docking experiments, the success rate is reaching a plateau in the graph. This is particularly true for the test set 0'. Hence, we consider it sufficient to use the top 3 PLANTS-generated poses for each protonation or tautomerization states in our rescoring assessments.

Rescoring using different energy compositions from MM-PBSA

It was suggested that increased solute dielectrics might provide improved estimations in electrostatics for some proteins using the MM-PBSA methods.[14,22] Hence, we tested the trends in different energy compositions from the MM-PBSA calculations on the assessing test sets. The ROC AUC values were calculated for total electrostatic-polar energies, total nonpolar energies, and total MM-PBSA energies. The squared Pearson correlation coefficient ($r_p{}^2$) and linear fitting of ROC AUC between higher internal dielectrics to the internal dielectric value of 1 were calculated and are shown in Figure 2. The results indicate that both electrostatics and non-polar energy compositions give almost perfect linear correlations using different internal dielectric values. This implies that the overall trending of the two energies remains consistent among different choices of dielectrics across all receptors. However, the ROC AUC correlations were drastically worsened when combining the two, i.e. the total MM-PBSA scores provide very poor enrichment correlations between different internal dielectric values.

The direct comparisons of the ROC AUC values provide some further insights (Tables 3 to 5). If only considering non-polar interactions from the MM-PBSA calculation, 61.1 - 66.7 % of the complexes in test set 1 resulted with ROC AUC over 0.70 across all the internal dielectric tested (Table 3). This rate is 44.4 % when using the $PLANTS_{chemplp}$[31,32] as the scoring function. A more significant improvement with non-polar energy calculations was observed for test set 2, from 47.4 % using the $PLANTS_{chemplp}$ scoring to 73.7 - 78.9 %

(Table 4). However, merely 1 out of 18 (5.6 %) and 3 out of 19 (15.79 %) ROC AUC values resulted in values over 0.7 when only considering electrostatic and polar interactions for the two test sets, respectively. This observation does not vary with the internal dielectric values. Furthermore, the ROC AUC results are consistent using different internal dielectric values
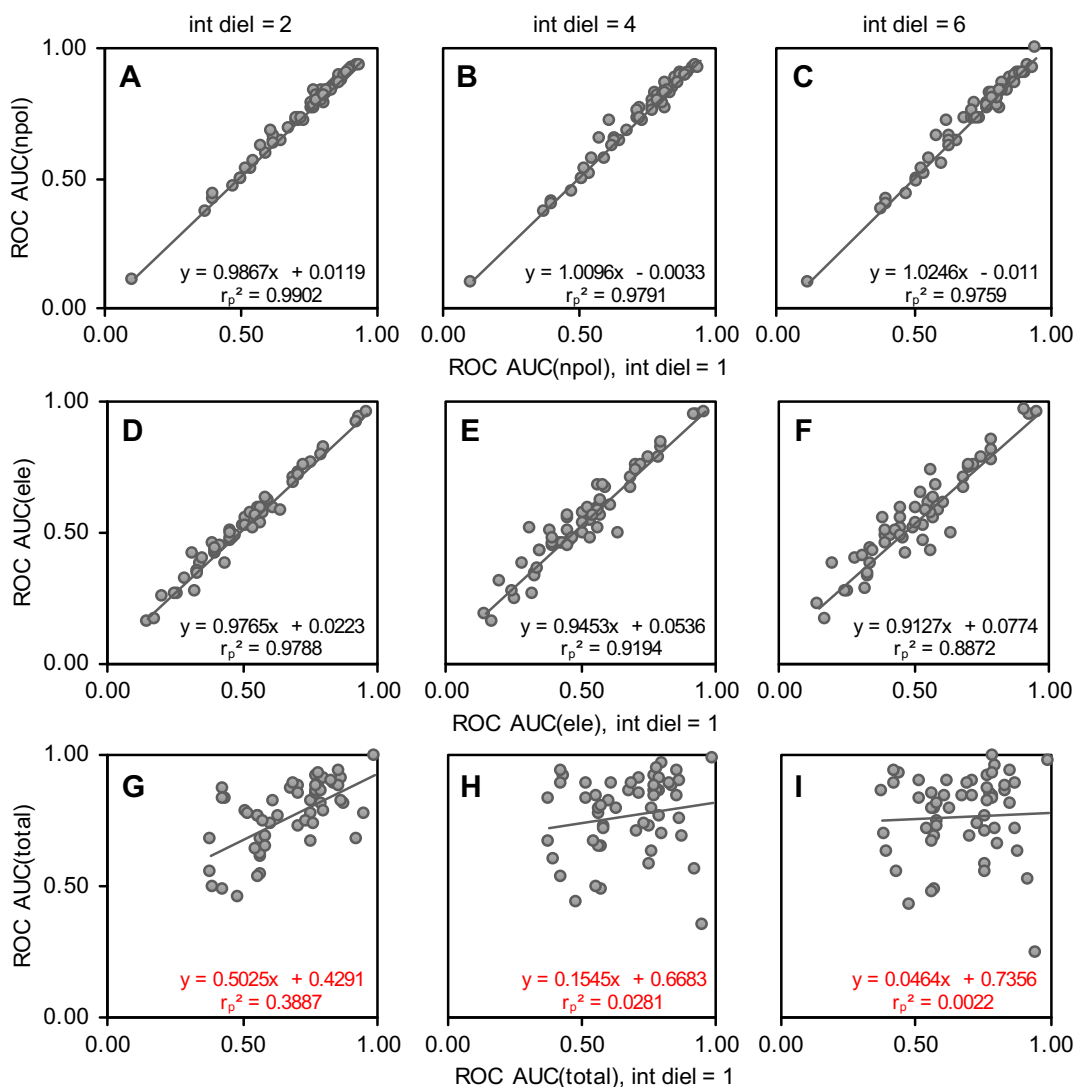
Figure 2: The correlations of ROC AUC values using internal dielectrics 2, 4, and 6 to the ones calculated with the internal dielectric value of 1 with outcomes from all test sets combined. Results from different energy compositions of MM-PBSA calculations are shown in separate rows. AM1-BCC was the charging method. A-C: the ROC AUCs calculated from electrostatic and polar energies. D-F: the ROC AUCs calculated from non-polar energies. G-I: the ROC AUCs calculated from total MM-PBSA binding free energies.

Table 3: The ROC AUC results using different internal dielectric values with MM-PBSA virtual screening on the test set 1 (DUD-E test set). Squared Pearson correlation coefficients to results from internal dielectric = 1 are also shown. AM1-BCC was the charging method.

Gene ID	ROC AUC (PLANTS)	ROC AUC, internal dielectric = 1			ROC AUC, internal dielectric = 2			ROC AUC, internal dielectric = 4			ROC AUC, internal dielectric = 6		
		non-polar	Ele + polar	total	non-polar	Ele + polar	total	non-polar	Ele + polar	total	non-polar	Ele + polar	total
ACES	0.8656	0.9209	0.2008	0.4384	0.9175	0.2464	0.8284	0.9253	0.3071	0.9172	0.9283	0.3730	0.9300
CDK2	0.3931	0.5136	0.5068	0.5707	0.5000	0.5170	0.5419	0.4949	0.4881	0.4836	0.4921	0.5130	0.4802
DPP4	0.6807	0.7756	0.4684	0.5704	0.7873	0.4476	0.6766	0.7929	0.4657	0.7688	0.7877	0.4736	0.7877
EGFR	0.6883	0.6361	0.5403	0.7052	0.6408	0.5506	0.7262	0.6500	0.5375	0.7049	0.6562	0.5177	0.6833
ESR1	0.5498	0.4050	0.9282	0.9180	0.4191	0.9338	0.6793	0.4123	0.9480	0.5600	0.4202	0.9446	0.5181
FA10	0.6149	0.8185	0.5128	0.7867	0.7904	0.5467	0.8820	0.7705	0.5683	0.8611	0.7657	0.5920	0.8367
FGFR1	0.8009	0.9364	0.1487	0.4281	0.9320	0.1531	0.8704	0.9271	0.1791	0.9344	0.9247	0.2167	0.9340
FNTA	0.6049	0.6603	0.3380	0.3945	0.6404	0.3508	0.4930	0.6428	0.3351	0.5962	0.6387	0.3304	0.6235
GRIA2	0.7128	0.8498	0.3940	0.6073	0.8417	0.4485	0.7330	0.8455	0.4972	0.8192	0.8578	0.5455	0.8440
GRIK1	0.4840	0.6026	0.4019	0.4287	0.5935	0.4137	0.4867	0.5694	0.4454	0.5327	0.5572	0.4778	0.5473
HIVPR	0.7766	0.8760	0.1771	0.4281	0.8813	0.1604	0.8281	0.8906	0.1573	0.8854	0.8802	0.1656	0.8896
KPCB	0.7394	0.6858	0.5965	0.7586	0.6880	0.6200	0.8231	0.6802	0.6658	0.7913	0.7246	0.5817	0.7651
LCK	0.9105	0.8811	0.5313	0.7705	0.8762	0.5705	0.9139	0.9026	0.5841	0.9188	0.8992	0.6448	0.9207
PTN1	0.6649	0.7197	0.5145	0.5724	0.7224	0.5184	0.6044	0.7281	0.5272	0.6496	0.7241	0.5333	0.6851
SRC	0.5334	0.7744	0.3477	0.5630	0.7648	0.3743	0.7602	0.7637	0.4219	0.7929	0.7682	0.4289	0.7894
THRB	0.8819	0.8206	0.5750	0.7741	0.8126	0.5915	0.8479	0.8231	0.5619	0.8598	0.8354	0.5455	0.8538
TRY1	0.3860	0.4059	0.5662	0.4841	0.4370	0.5303	0.4498	0.4043	0.5909	0.4338	0.3971	0.6132	0.4226
VGFR2	0.7983	0.7797	0.5629	0.7862	0.7706	0.5900	0.8853	0.7762	0.5865	0.8617	0.7827	0.6051	0.8330
% (> 0.7)	44.44	61.11	5.6	38.89	61.11	5.6	61.11	61.11	5.6	66.67	66.67	5.6	61.11
r_p^2 to data from ROC AUC, internal dielectric = 1					0.9941	0.9860	0.1962	0.9886	0.9539	0.0214	0.9811	0.8944	0.0029

Table 4: The ROC AUC results using different internal dielectric values with MM-PBSA virtual screening on the test set 2 (ChEMBL test set). Squared Pearson correlation coefficients to results from internal dielectric = 1 are also shown. AM1-BCC was the charging method.

Gene ID	ROC AUC (PLANTS)	ROC AUC, internal dielectric = 1			ROC AUC, internal dielectric = 2			ROC AUC, internal dielectric = 4			ROC AUC, internal dielectric = 6		
		non-polar	Ele + polar	total	non-polar	Ele + polar	total	non-polar	Ele + polar	total	non-polar	Ele + polar	total
ACES	0.7934	0.7762	0.2904	0.3824	0.7720	0.3194	0.5533	0.7575	0.3749	0.6651	0.7721	0.3909	0.6929
ADORA1	0.7117	0.6334	0.6180	0.7560	0.6618	0.5896	0.7733	0.6432	0.6008	0.7263	0.6436	0.6068	0.7005
AURKA	0.4995	0.7273	0.5685	0.7724	0.7320	0.5740	0.8558	0.7672	0.5065	0.8371	0.7854	0.4280	0.8242
CA1	0.4385	0.5865	0.5362	0.5680	0.6247	0.5102	0.6199	0.6481	0.4700	0.6433	0.6594	0.4594	0.6645
CASP1	0.9250	0.8580	0.9186	0.9919	0.8557	0.9148	0.9945	0.8860	0.9485	0.9837	0.8848	0.9640	0.9710
CDK2	0.6915	0.4792	0.5712	0.5607	0.4714	0.5662	0.5336	0.4491	0.5789	0.4930	0.4424	0.5707	0.4702
CTSD	0.9270	0.9067	0.7076	0.8013	0.9139	0.7190	0.9070	0.9060	0.7555	0.9652	0.9053	0.7462	0.9532
EGFR	0.6780	0.7270	0.4228	0.5832	0.7220	0.4383	0.6837	0.7301	0.4528	0.7280	0.7264	0.4843	0.7418
ELANE	0.8623	0.8230	0.4557	0.8376	0.8361	0.4597	0.8924	0.8340	0.5002	0.8742	0.8347	0.5056	0.8620
ESR1	0.5635	0.5437	0.7971	0.8781	0.5381	0.8057	0.8154	0.5204	0.8181	0.6847	0.5174	0.8064	0.6286
FGFR1	0.5270	0.8731	0.3271	0.5169	0.8942	0.2704	0.7862	0.8967	0.2566	0.8895	0.8947	0.2815	0.8986
GSK3B	0.7151	0.7402	0.3997	0.5497	0.7186	0.4204	0.6370	0.7148	0.4503	0.6698	0.7281	0.4536	0.7160
LCK	0.5812	0.8637	0.4596	0.8697	0.8665	0.5046	0.9040	0.8620	0.5496	0.8952	0.8644	0.5487	0.8847
MAPK1	0.7449	0.8244	0.2564	0.3799	0.8250	0.2590	0.6791	0.8680	0.2397	0.8285	0.8666	0.2706	0.8577
MMP1	0.8578	0.8119	0.4775	0.5836	0.8002	0.4787	0.6514	0.7851	0.4692	0.7132	0.7804	0.4130	0.7204
NTRK1	0.2840	0.7916	0.3353	0.5206	0.7918	0.3437	0.7710	0.8120	0.3245	0.8258	0.8144	0.3374	0.8269
PPARG	0.9202	0.8073	0.5774	0.7129	0.8315	0.5726	0.8823	0.8081	0.6169	0.9089	0.8089	0.6234	0.8984
SRC	0.5956	0.7819	0.4383	0.6790	0.8358	0.3750	0.8657	0.8275	0.4527	0.8561	0.8214	0.5039	0.8444
VGFR2	0.6354	0.6200	0.6411	0.7365	0.6759	0.5810	0.7436	0.7167	0.4916	0.7377	0.7167	0.4929	0.7320
% (> 0.7)	47.37	73.68	15.79	47.37	73.68	15.79	63.16	78.95	15.79	73.68	78.95	15.79	78.95
r_p^2 to data from ROC AUC, internal dielectric = 1					0.9838	0.9853	0.8025	0.9627	0.9446	0.4606	0.9599	0.9165	0.3299

for both non-polar and electrostatic-polar energies.

The ROC AUCs calculated from total MM-PBSA estimations varies between different internal dielectrics. The results, for most of the cases, are much deteriorated when internal dielectric equals to 1 and improves as internal dielectric value increases. This behavior implies

Table 5: The ROC AUC results using different internal dielectric values with MM-PBSA virtual screening on the test set 3 (decoy test set). Squared Pearson correlation coefficients to results from internal dielectric = 1 are also shown. AM1-BCC was the charging method.

Gene ID	ROC AUC (PLANTS)	ROC AUC, internal dielectric = 1			ROC AUC, internal dielectric = 2			ROC AUC, internal dielectric = 4			ROC AUC, internal dielectric = 6		
		non-polar	Ele + polar	total	non-polar	Ele + polar	total	non-polar	Ele + polar	total	non-polar	Ele + polar	total
ACES	0.7934	0.8414	0.3996	0.5611	0.8339	0.4319	0.7610	0.8278	0.4728	0.8420	0.8339	0.5035	0.8497
CDK2	0.7117	0.5148	0.6910	0.7605	0.4955	0.7058	0.7353	0.4927	0.6993	0.6241	0.4849	0.7051	0.5751
DPP4	0.4995	0.5278	0.7542	0.7939	0.5369	0.7655	0.8080	0.5305	0.7802	0.7605	0.5362	0.7825	0.7140
EGFR	0.4385	0.5576	0.6932	0.8026	0.5633	0.6865	0.7797	0.5780	0.6692	0.6935	0.5776	0.6692	0.6550
ESR1	0.9250	0.6303	0.7971	0.8687	0.6282	0.8145	0.8231	0.6243	0.8423	0.7502	0.6244	0.8506	0.7181
FA10	0.6915	0.9466	0.3157	0.7850	0.9311	0.4109	0.9269	0.9260	0.5122	0.9467	0.9995	0.4063	0.9990
FGFR1	0.9270	0.7190	0.4526	0.6305	0.7261	0.4678	0.7597	0.7604	0.4469	0.7907	0.7561	0.4822	0.7929
FNTA	0.6780	0.7940	0.3391	0.5799	0.8146	0.3389	0.7393	0.8177	0.3525	0.8001	0.8244	0.3718	0.8135
GRIA2	0.8623	0.3792	0.7900	0.7579	0.3743	0.7894	0.6703	0.3754	0.7771	0.5794	0.3818	0.7699	0.5470
GRIK1	0.5635	0.1130	0.9584	0.9460	0.1145	0.9569	0.7680	0.1017	0.9578	0.3480	0.0975	0.9577	0.2391
HIVPR	0.5270	0.8914	0.2496	0.6165	0.8875	0.2613	0.8163	0.8922	0.2719	0.8890	0.8978	0.2705	0.8978
KPCB	0.7151	0.7330	0.7240	0.8577	0.7247	0.7541	0.8743	0.7308	0.7530	0.8368	0.7266	0.7545	0.8094
LCK	0.5812	0.8281	0.5679	0.8323	0.8368	0.5912	0.8939	0.8345	0.6746	0.8915	0.8342	0.7336	0.8809
PTN1	0.7449	0.8117	0.7114	0.7898	0.8171	0.7166	0.8574	0.8103	0.7311	0.9079	0.8076	0.7537	0.9314
SRC	0.8578	0.7942	0.4562	0.7124	0.7950	0.4908	0.8458	0.7954	0.5579	0.8471	0.8007	0.5847	0.8407
THRB	0.2840	0.9001	0.5860	0.8539	0.8974	0.6280	0.9361	0.8939	0.6771	0.9394	0.9009	0.6742	0.9330
TRY1	0.9202	0.8716	0.3527	0.6916	0.8603	0.3962	0.8837	0.8628	0.4206	0.8926	0.8681	0.4268	0.8889
VGFR2	0.5956	0.8206	0.5467	0.7738	0.8158	0.5620	0.8775	0.8282	0.5617	0.8801	0.8297	0.5760	0.8694
% (> 0.7)	77.78	66.67	33.33	72.22	66.67	38.89	94.44	66.67	33.33	77.78	66.67	44.44	77.78
r_p^2 to data from ROC AUC, internal dielectric = 1					0.9980	0.9881	0.1120	0.9948	0.9281	0.1169	0.9936	0.9387	0.1528

that the variation is very likely related to electrostatics. However, the electrostatic and polar estimated energies provide the universally similar and poor capability of distinguishing actives from inactive ligands, as demonstrated by the almost perfect linear correlations of ROC AUC values using different internal dielectric value. Thus, it is highly likely that increasing the internal dielectric value does not qualitatively improve the differentiating capability of electrostatic and polar energies.

We adopted an additional decoy set (test set 3) for the purpose to confirm that the previous observations do not alter when using libraries containing decoys. In this test set, electrostatic and polar energies perform better than the previous two; the above-0.7-ROC-AUC rate increases to 33.33 - 44.44 % (Table 5). These improvements in electrostatic enrichments are very likely due to the much more distinct chemical properties of the actives to the decoys than to the experimentally tested inactive compounds. However, the non-polar energies still outperform its polar companions; 66.67 % of the ROC AUCs calculated with non-polar energies reached above 0.7 within the test set. Hence, the decoy results further confirmed the observations in the previous test sets. Moreover, the slight improvements in electrostatic predictions in the test set 3, due to the application of the less specific decoy

set, provide a much better outcome in the total score from MM-PBSA calculations. The above-0.7 success rates resulted in the 73.68 - 94.74 % range, much better than that observed in the previous cases, likely caused by the application of the universal decoys.

Rescoring using the MM-GBSA, and an alternative charging method

To confirm these distinct behaviors of different energy compositions are not singular for the MM-PBSA method, we also tested the complexes from the test set 1 using the MM-GBSA calculations. Similar results further confirmed that such differences in the decomposed energy calculations are universal within the MM-PB/GBSA models (tables 6 and 8 and S3). Virtual screening performed on the test set 1 demonstrated that close to one $r_p{}^2$ were observed only for correlations of polar and non-polar energies between different internal dielectric applied. The deteriorating effects on correlations appear once combining the two as the total binding score. The MM-GBSA model showed better robustness in reproducibility when increasing the internal dielectric value for total binding free energy calculation; using internal dielectric at 2 still provides a good correlation to the internal dielectric of 1. However, this robustness did not extend onto the other two sets of calculations. More crucially, this robustness does not provide that the MM-GBSA model essentially provides better predictability than MM-PBSA methods, only that result with increased internal dielectrics performs as bad as those with internal dielectric equals to 1. The ROC AUC outcomes also confirmed that MM-GBSA performs slightly better than MM-PBSA in terms of electrostatics estimations; 3 out of 18 polar energy calculations provided ROC AUC above 0.7 across all the dielectrics tested. A poor correlation of the decomposed energies to the ones calculated using the MM-PBSA model suggest a big variation between the two method, likely due to the intrinsic differences in the mathematics approximations.

We also investigated the possibility of improvements by using an alternative charging method. In a study on docking experiments, the Gasteiger method provided a better correlation to experimental data than AM1-BCC among six HIV protease inhibitors.[16] Due to its

Table 6: ROC AUC correlation data for MM-GBSA energy compositions with different internal dielectrics (2, 4, and 6) to results using internal dielectric = 1. The AM1-BCC charging method was applied.

MM-GBSA energy compositions	Internal dielectric = 2			Internal dielectric = 4			Internal dielectric = 6		
	r_p^{2*}	m^{**}	b^{**}	r_p^{2*}	m^{**}	b^{**}	r_p^{2*}	m^{**}	b^{**}
Ele + GB polar	0.9996	1.0019	0.0014	0.9951	1.0040	0.0011	0.9870	1.0154	0.0020
VDW + GB non-polar	0.9981	1.0168	-0.0141	0.9924	1.0111	-0.0090	0.9857	1.0140	-0.0124
Total MM-GBSA	0.9061	0.8523	0.1214	0.3562	0.4572	0.4060	0.0685	0.2127	0.5775

* Pearson correlation coefficient squared
** m and b are the fitted parameter to a linear equation of the form $y = mx + b$

Table 7: The ROC AUC correlation data for MM-PBSA energy compositions with different internal dielectrics (2, 4, and 6) to results using internal dielectric = 1. The Gasteiger charging method was applied.

MM-GBSA energy compositions	Internal dielectric = 2			Internal dielectric = 4			Internal dielectric = 6		
	r_p^{2*}	m^{**}	b^{**}	r_p^{2*}	m^{**}	b^{**}	r_p^{2*}	m^{**}	b^{**}
Ele + PB polar	0.9870	0.9954	0.0180	0.9386	0.9598	0.0501	0.8742	0.9152	0.0769
VDW + PB non-polar	0.9952	0.9982	-0.0033	0.9869	1.0098	-0.0078	0.9854	1.0194	-0.0114
Total MM-PBSA	0.3728	0.6109	0.3293	0.0628	0.2598	0.5636	0.0228	0.1595	0.6261

* Pearson correlation coefficient squared
** m and b are the fitted parameter to a linear equation of the form $y = mx + b$

efficiency and convenience in incorporating into the workflow, we also adopted the Gasteiger partial charges with the test sets included within the Autodock package.[25] Gasteiger partial charges historically are better implemented and more complete within Autodock led us to such choice over the set provided by antechamber.[39] However, the ROC AUC correlations between separate runs with different internal dielectric choices were similar to those shown for the AM1-BCC method (Table 7 and S4). Moreover, the individual ROC AUC outcomes considering different energy compositions were close to the ones calculated using the AM1-BCC partial charges, as the r_p^2 all provided values over 0.78 (Table 8). Hence, it is very likely the observed worsening effects from electrostatic and polar energies were more related to the methods applied to calculate the energy than the partial charges applied.

Table 8: The ROC AUC results using MM-PB/GBSA methods and different partial charges when internal dielectric equals to 1.

Gene ID	ROC AUC (PLANTS)	ROC AUC (MM-PBSA, AM1-BCC)			ROC AUC (MM-GBSA, AM1-BCC)			ROC AUC (MM-PBSA, Gasteiger)		
		non-polar	Ele + polar	total	non-polar	Ele + polar	total	non-polar	Ele + polar	total
ACES	0.8656	0.9209	0.2008	0.4384	0.9088	0.4445	0.6322	0.9190	0.1799	0.4434
CDK2	0.3931	0.5136	0.5068	0.5707	0.4762	0.4400	0.4429	0.5074	0.3507	0.3546
DPP4	0.6807	0.7756	0.4684	0.5704	0.7654	0.5380	0.5806	0.7590	0.5019	0.5904
EGFR	0.6883	0.6361	0.5403	0.7052	0.6402	0.5270	0.5609	0.6485	0.5883	0.7523
ESR1	0.5498	0.4050	0.9282	0.9180	0.4293	0.8399	0.8066	0.4163	0.8201	0.8360
FA10	0.6149	0.8185	0.5128	0.7867	0.7995	0.7360	0.7819	0.8046	0.4439	0.7262
FGFR1	0.8009	0.9364	0.1487	0.4281	0.9286	0.0861	0.1248	0.9320	0.1781	0.4604
FNTA	0.6049	0.6603	0.3380	0.3945	0.6445	0.6020	0.6171	0.6404	0.4073	0.4714
GRIA2	0.7128	0.8498	0.3940	0.6073	0.8645	0.6690	0.7095	0.8608	0.3443	0.5197
GRIK1	0.4840	0.6026	0.4019	0.4287	0.5896	0.5684	0.6026	0.5792	0.4300	0.4420
HIVPR	0.7766	0.8760	0.1771	0.4281	0.8396	0.7906	0.8479	0.8844	0.2021	0.5625
KPCB	0.7394	0.6858	0.5965	0.7586	0.7194	0.4854	0.5403	0.6719	0.5769	0.6850
LCK	0.9105	0.8811	0.5313	0.7705	0.8704	0.3728	0.4638	0.8792	0.4428	0.6531
PTN1	0.6649	0.7197	0.5145	0.5724	0.7399	0.6211	0.6289	0.7197	0.5465	0.5934
SRC	0.5334	0.7744	0.3477	0.5630	0.7642	0.3295	0.4004	0.7753	0.3500	0.5591
THRB	0.8819	0.8206	0.5750	0.7741	0.8201	0.5039	0.6330	0.8232	0.5110	0.7791
TRY1	0.3860	0.4059	0.5662	0.4841	0.4027	0.3931	0.3541	0.4330	0.4984	0.3533
VGFR2	0.7983	0.7797	0.5629	0.7862	0.7842	0.5900	0.6569	0.7882	0.5226	0.7470
% (> 0.7)	44.44	61.11	5.56	38.89	66.67	16.67	22.22	61.11	5.56	27.7
r_p^2 to data from ROC AUC, internal dielectric = 1					0.9932	0.3891	0.3231	0.9968	0.9451	0.8555

KDE probability density of different energy compositions

The above results suggested that the overall electrostatic and polar interactions provide poorer discrimination of active from inactive ligands and this also affected the overall ROC AUC using the total binding free energy estimated by MM-PB/GBSA methods. The strength of the effect also varies with internal dielectric values. To directly observe the effects from electrostatic and polar energies, we thus performed kernel density estimation (KDE) for the energy probability densities of different energy compositions between actives and inactives for test set 1 (Figure 3 and S3-4). Thirteen systems show that the increased internal dielectric values quantitatively decrease the deteriorating effects from electrostatic and polar energies, leading to improvements in the enrichment of total MM-PBSA estimations (as ACES, HIVPR, and SRC shown in Figure 3). However, the predictivities of each energy components remain in the same range. Same effects were observed for another 13 systems in the test set 2 (Figure S5 and S6) and 11 systems in the test set 3 (Figure S7 and S8). This quantitative change in electrostatic-polar interactions does not indicate that the increase in internal dielectric provides an improved PB estimation in electrostatic/polar interactions

qualitatively. More interestingly, the electrostatics calculated from gas phase perform better than the total electrostatic-polar calculated, in terms of ROC AUC values (Table S5). This implies that the PB model contributes to the worsening influence on the polar interaction estimation at lower internal dielectric values within the test sets tested. Hence, readers should

Figure 3: The KDE plots of different energy compositions using several example receptors picked from the test set 1. The ROC AUC values are shown above each graph. ACES, HIVPR, and SRC demonstrate the MM-PBSA rescoring trends in the majority of the receptors in test sets 1-3. CDK2 is the intermediate behavior while there is no distinct difference in enrichment performance using non-polar or electrostatic-polar energy calculations. ESR1 plots are an illustration of the cases where electro-polar estimations out-perform non-polar composition. A relatively over-smoothed bandwidth (0.5) was chosen for demonstration purpose as the plots were not intended for the accurate estimation of densities.

20

bear in mind that increasing internal dielectric values might improve the overall performance of virtual screening, but this does not change the quality of the polar energy estimated using PB model.

Considering Effects from Molecular Weight

The screening performance is closely related to the distribution of the MW within the tested libraries. Larger molecules are more likely to score higher in ranking. In this work, most libraries provided good MW overlay between actives and inactives (decoys). However, in comparison, ranking with MW would be a more direct assessment to determine biases from MW (Figure S9). Hence, we performed the ROC AUC analysis based on the ranking of the MW in each library.

The results show that within the accessed databases (both DUD-E and ChEMBL) for the receptors selected, the ROC AUCs calculated from molecular weight are very nicely correlated with non-polar interactions. The $r_p{}^2$ between the ROC AUCs generated using non-polar interactions and molecular weights are 0.939, 0.877, and 0.984 for the test set 1, 2, and 3, respectively (Table S6). However, this trend has not occurred for polar-electrostatic interactions (Table S6). Negative correlation values were even observed between the electrostatic-polar and the MW results for all the test sets assessed (-0.583, -0.264, and -0.803 for the test set 1, 2, and 3). This observation implies that the different distribution of molecular weight between active and inactive ligands could be one of the underlying causes of the out-performance of non-polar interaction in the previous analysis.

It became necessary to remove the biasing effects from molecular weight to further confirm the presumption, and to reveal the masked relation between non-polar and electrostatic-polar interactions among molecules of similar sizes. Hence, we processed the ligand library for each receptor in the test set 2 and 3 to lower the ROC AUC values of molecular weight score to less than 0.5. If non-polar interactions are more connected to the sizes of molecules, we expect this lower in MW-biases will also worsen the performance of the calculated non-
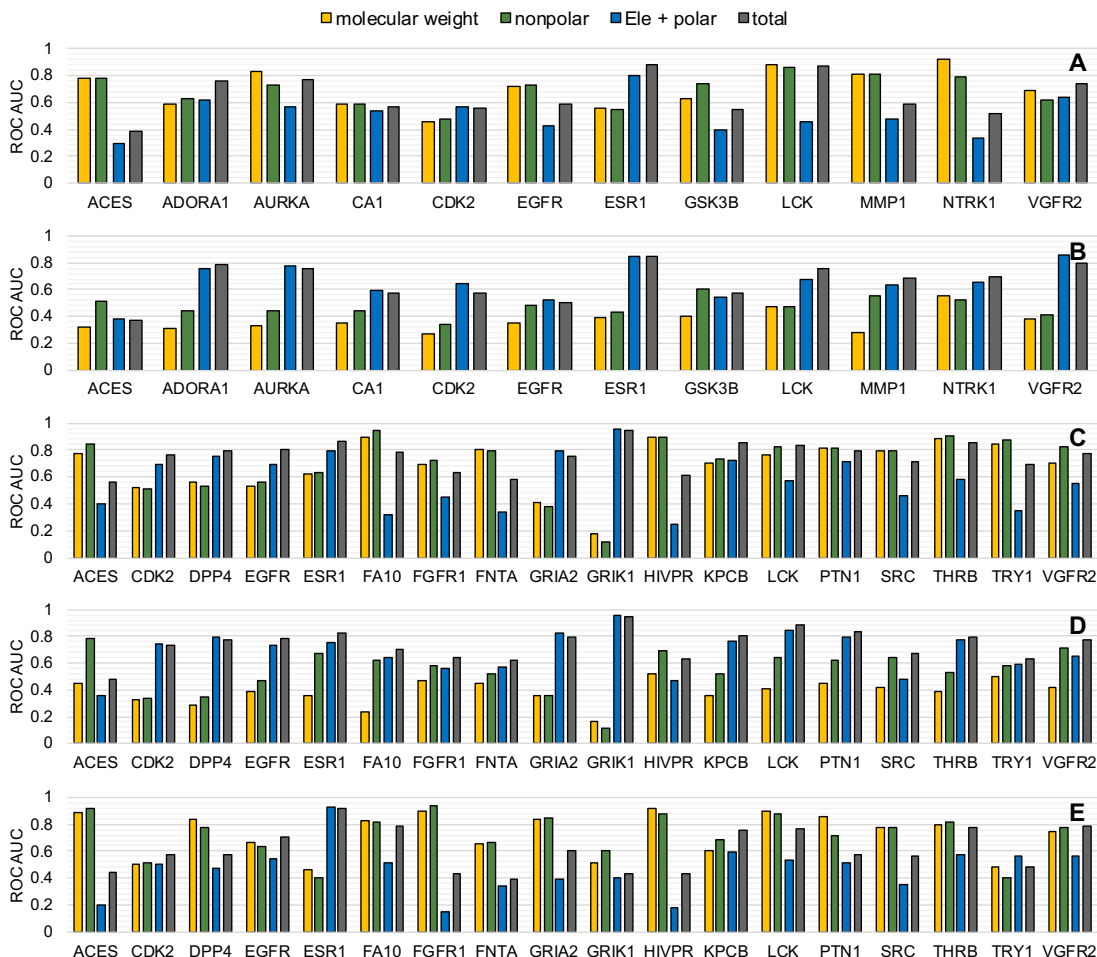
21

Figure 4: The bar charts illustrating ROC AUC results for both the virtual screening using the original libraries and the libraries with adjusted molecular weight. A and B are data generated from the test set 2 (ChEMBL test set). C and D are data generated from the test set 3 (decoy test set). For comparison purpose, the result from the test set 1 (DUD-E test set, shown as E) is also included. Considering the library type, A, C, and E are results for the original libraries, while B and D are for the molecular-weight-adjusted libraries. Internal dielectric constant = 1.

polar interactions. This might also provide insights on influences from polar-electrostatic interactions on total virtual screening performance under an MW-unbiased circumstance.

The ROC AUCs calculated from the adjusted molecular weight were plotted alongside those for non-polar, electrostatic-polar, and total MM-PBSA scores (Figure 4). The bar plots of the results from the original libraries were also included for comparison. We noticed that firstly, the decrease in ROC AUC using molecular weight does affect the enrichment

of non-polar interactions. This further confirmed that the non-polar interaction is closely related to the size of the ligands. On the other hand, the non-polar interactions still provided better enrichment for most of the receptors than using the molecular weight as the ranking method. This can be explained considering the more direct connection between non-polar to the fitness within the binding pocket than considering MW solely.

Additionally, the enrichment performance of electrostatic-polar interactions improved for the majority of the receptors, when comparing to the trends in the original libraries. These can be better observed in the plot of differential ROC AUC (Figure 5). The decrease in non-polar prediction is compensated by the improvement in the enrichment of electrostatic-polar, safeguarding the total MM-PBSA rescoring from severe deteriorating caused by the change in non-polar estimations. (Figure 5A and C). Increasing the internal dielectric values, however, impairs the safeguarding effects from the electrostatic and polar interactions (Figure 5B and D) by quantitatively decreasing the calculated electrostatic-polar energies. Hence, it seems also dangerous to increase internal dielectric values within the MM-PBSA rescoring for libraries where the active compounds are of similar or smaller size than the inactives.

Zhang et al. previously evaluated the varied performance of MM-GBSA in proteins of different categories. It was suggested that the performance of MM-GBSA was target-dependant.[20] However, by using different ligand libraries, the same proteins using the same receptors structure provided very different outcomes in this study. Comparing the ROC AUC outcomes between the test set 1 and 3, which used the same set of receptors, multiple targets (CDK2, DPP4, EGFR, GRIA2, GRIK1, and TRY1) have shown inverted behaviors of non-polar and polar-electrostatic calculated scores (Table 3 and 5). This is likely due to the fact that the libraries of each receptor have varied molecular size compositions (Figure S9). Moreover, after the MW-readjustments in the test set 3, more receptors showed the change in outcomes for rescoring assessment, such as FA10, FGFR1, FNTA, KPCB, LCK, PTN1, THRB, and TRY1. This made us suspect that the MM-PBSA method could likely behave more in a library-dependent manner.

Taken together, the assessment outcomes suggest that non-polar and electrostatic-polar interactions compensate each other to some extent. When the targeted receptor requires larger molecules within the binding site, the non-polar interactions guarantee that only molecules of decent size are forwarded within the ranking process. However, among molecules
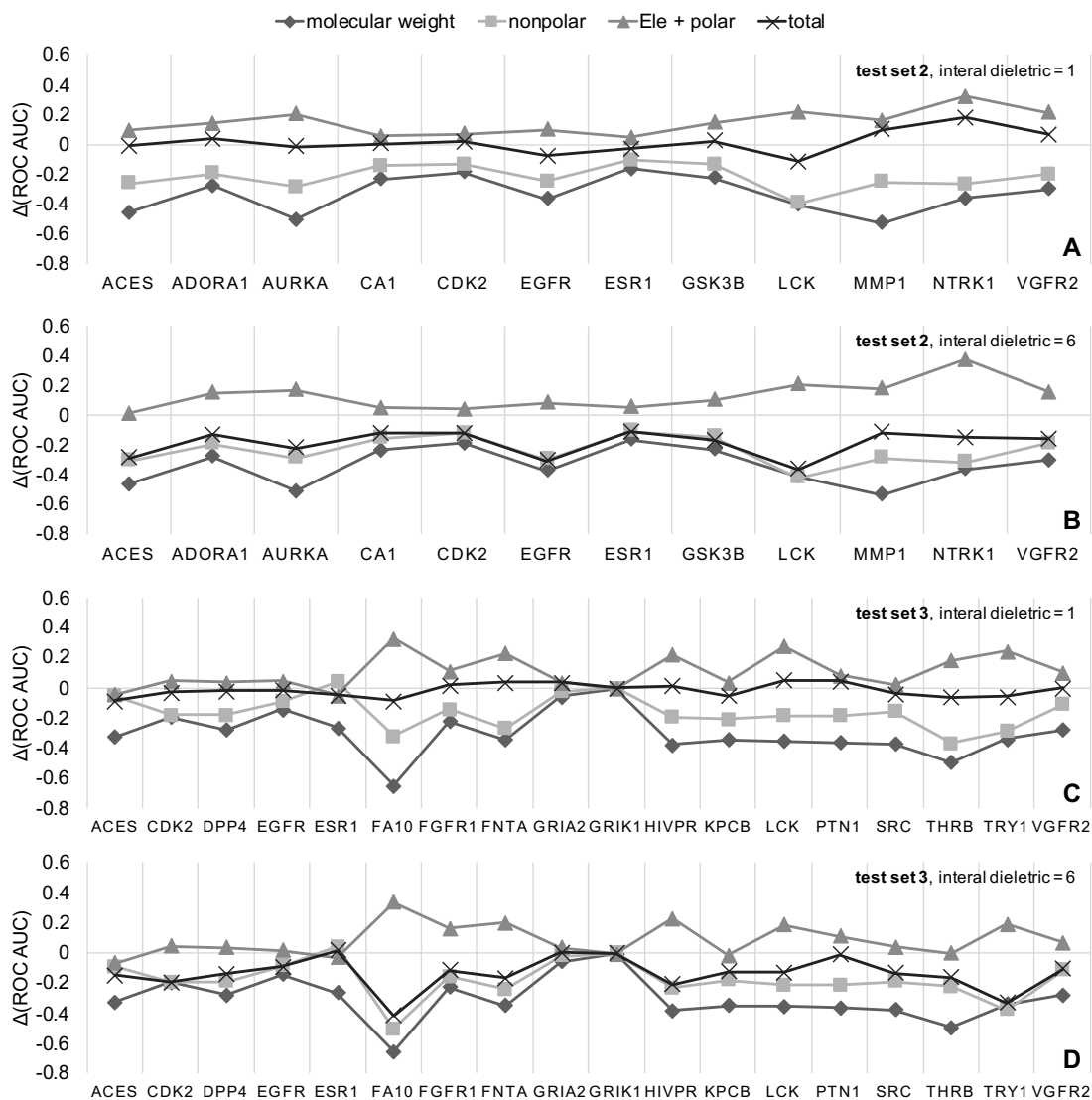

Figure 5: The differential ROC AUC of the MW-adjusted virtual screening to the original libraries. Different rescoring methods are plotted using different greyscale colours and markers as illustrated. A and B are the outcomes calculated for receptors from the test set 2 (ChEMBL test set) using internal dielectric values of 1 and 6, respectively. C and D are the same plots for receptors from the test set 3 (decoy test set). Lower internal dielectrics show less worsening in total MM-PBSA energies when impairments to the non-polar energies are caused by the change in MW distribution within the tested library.

with similar sizes, electrostatic energy become more selective for active ligands. On the other hand, this relative opposite behaviours of non-polar and polar interactions are also alerting. We observed in this study that the compensation could easily turn towards the direction of a cancelling effect when combining the two types of energies. This is particularly severe when using the default MM-PBSA on a molecular-size-unbalanced library, or a library with ligands of similar size but an unbalanced MM-PBSA evaluation (i.e. with high internal dielectric values). The beneficiary influence is no more effective in the final binding estimation. This would lead to an overall worsened outcome compared to the empirical $PLANTS_{chemplp}$ function or consider individual energy composition alone. Additionally, the distinct behaviours of the MM-PBSA rescoring on different libraries targeting the same receptors may have addressed that the issue related to the MM-PBSA rescoring have more library-originated influential factors than variation between receptors.

This study further confirmed that increasing internal dielectric value is not the universal solution for performance improvement in the MM-PBSA rescoring. From the results reported here, such improvement is likely related to the fact that increasing internal dielectric constant in MM-PBSA only quantitatively decreases the potential worsening effects from the calculated electrostatic-polar energies in a system that prefer larger molecules. However, the quality of the electro-polar estimation is not improved to provide a better outcome when considering polar-electrostatic energies alone. Increasing internal dielectric constants might also be potentially dangerous, as it also decreases the beneficiary effects provided by electrostatic-polar composition within MM-PBSA calculations. We recommend that the default internal dielectric value should not be increased without a good understanding of the chemical properties of the compounds that are capable of binding to the targeted receptor.

Conclusions

In this work, we have attempted to address the underlying causes of different responses of biomolecular systems to the change of solute dielectric constants in the MM-PBSA rescoring process. We established multiple virtual screening scenarios and larger test sets of receptors from various categories and ligand libraries of different natures. The work intended to simulate as many screening circumstances as possible to explore the possible influences from the internal dielectric changes on the virtual screening outcome.

The outcomes of this study have provided the following insights regarding MM-PBSA applications in virtual screening rescoring:

- In MW-biased systems, i.e. the receptors prefer ligands of certain sizes and the active compounds are relatively larger within the screened library, the calculated non-polar interactions play the dominant role in distinguishing active compounds from the ligand pool. Within these systems, increasing the internal (solute) dielectric constant is highly likely to improve the overall performance of the MM-PBSA method.

- In MW-unbiased systems, where the actives and inactives are of the similar sizes, the electrostatic and polar interactions safeguarding the smaller active ligands from moving backwards on the ranking list. In such circumstance, increasing internal dielectric can impair the enrichment of MM-PBSA rescoring by quantitatively decreasing the electrostatic-polar counteracting on non-polar energies.

- Using an alternative method such as MM-GBSA and the Gasteiger charges do not change the overall trend of performance within the scope of this study.

Previous studies using the MM-PBSA calculation to explore the reproducing trend within experimental data partially implied the above observations. Reproducing binding ranking generally involves a smaller library containing multiple molecules of similar sizes. Under this condition, MM-PBSA rescoring with the default internal dielectric constant works decently

well due to the dominance of electrostatic-polar estimations that secure the specificity to the receptor.

Notice that we did not perform the comparison between different PB solvers in this study, due to the availability of other solvers and also the concerns regarding the ease of use. We do not exclude the differences in results by using different methods. However, we consider the solver comparison beyond the scope and purpose of this work and shall be included in a separate study. However, some information still can be inferred from the comparison between PB and GB involved calculations. Though variations do exist, the overall trends in performance for electrostatic-polar estimations were similar between using a PB or a GB solver.

Our investigations indicate that increasing internal dielectric value might be dangerous. It is very likely that in an actual virtual screening scenario, the MWs of molecules are pre-processed to be similar. The benefits brought by non-polar interactions are diminished and the final ranking is dependent on the electrostatic-polar interactions to guarantee the selectivity of active compounds. Increasing internal dielectric might impair this selectivity, leading to a failure in the application of the virtual screening. Thus, it should not be the universal solution to improve the performance of MM-PBSA.

Author Information

Corresponding Author

E-mail: alessandro.contini@unimi.it

Notes

The authors declare no competing financial interest.

Acknowledgement

Abbreviations

MM-PBSA - Molecular mechanics Poisson-Boltzmann surface area

MM-GBSA - Molecular mechanics generalized Born surface area

MM - Molecular mechanics

vina - Autodock Vina

ROC AUC - Receiver operating characteristic area under curve

MW - Molecular weight

KDE - Kernel density estimation

$r_p{}^2$ - Squared Pearson correlation coefficient

Supporting Information Available

The following files are available free of charge at http://pubs.acs.org:

- Table S1-2: Protein receptors included in the test sets 1 and 2.

- Figure S1-2: Comparison of pose predictions between PLANTS and vina (with exhaustiveness = 12 and 10) for test set 0 and 0'.

- Table S3: The ROC AUC results using different internal dielectric values with MM-GBSA virtual screening on test set 1.

- Table S4: The ROC AUC results using different internal dielectric values with MM-PBSA virtual screening on test set 1 (Gasteiger partial charges).

- Figure S3-8: The KDE plots of different energy compositions of receptors from test set 1, 2, and 3.

- Table S5: The ROC AUC outcomes of virtual screening using different gas phase energies. Internal dielectric equals 1.

- Figure S9: Comparison between active-inactive average MW and ROC AUC using MW rankings to determine MW biases.

- Table S6: Comparing ROC AUC outcomes rescoring using MW and MM-PBSA calculated energy compositions.

- ligand_libraries.zip: the compressed .txt files of all the ligand libraries applied in this study in the SMILES format.

References

(1) Walters, W.; Stahl, M. T.; Murcko, M. A. Virtual screeningan overview. *Drug Discov. Today* **1998**, *3*, 160–178.

(2) Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discov. Today* **2002**, *7*, 1047–1055.

(3) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949.

(4) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today* **2006**, *11*, 580–594.

(5) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.

(6) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules:ǎ Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(7) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221–5230.

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

(9) Wong, S.; Amaro, R. E.; McCammon, J. A. MM-PBSA Captures Key Role of Intercalating Water Molecules at a Protein-Protein Interface. *J. Chem. Theory Comput.* **2009**, *5*, 422–429.

(10) Rastelli, G.; Rio, A. D.; Degliesposti, G.; Sgobba, M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2009**, *31*, 797–810.

(11) Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, P. Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *J. Comput. Chem.* **2009**, *30*, 1701–1711.

(12) Lindström, A.; Edvinsson, L.; Johansson, A.; Andersson, C. D.; Andersson, I. E.; Raubacher, F.; Linusson, A. Postprocessing of docked protein-ligand complexes using implicit solvation models. *J. Chem. Inf. Model.* **2011**, *51*, 267–282.

(13) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and

MM/GBSA methods: II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **2011**, *32*, 866–877.

(14) Yang, T.; Wu, J. C.; Yan, C.; Wang, Y.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Virtual screening using molecular simulations. *Proteins Struct. Funct. Bioinf.* **2011**, *79*, 1940–1951.

(15) Genheden, S. MM/GBSA and LIE estimates of host-guest affinities: dependence on charges and solvation model. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 1085–1093.

(16) Oehme, D. P.; Brownlee, R. T. C.; Wilson, D. J. D. Effect of atomic charge, solvation, entropy, and ligand protonation state on MM-PB(GB)SA binding energies of HIV protease. *J. Comput. Chem.* **2012**, *33*, 2566–2580.

(17) Sgobba, M.; Caporuscio, F.; Anighoro, A.; Portioli, C.; Rastelli, G. Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations. *Eur. J. Med. Chem.* **2012**, *58*, 431–440.

(18) Genheden, S.; Ryde, U. Comparison of end-point continuum-solvation methods for the calculation of proteinligand binding free energies. *Proteins: Structure, Function, and Bioinformatics* **2012**, *80*, 1326–1342.

(19) Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. *J. Phys. Chem. B* **2013**, *117*, 8408–8421.

(20) Zhang, X.; Wong, S. E.; Lightstone, F. C. Toward fully automated high performance computing drug discovery: a massively parallel virtual screening pipeline for docking and molecular mechanics/generalized Born surface area rescoring to improve enrichment. *J. Chem. Inf. Model.* **2014**, *54*, 324–337.

(21) Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *PCCP* **2014**, *16*, 16719–16729.

(22) Sun, H.; Li, Y.; Shen, M.; Tian, S.; Xu, L.; Pan, P.; Guan, Y.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22035–22045.

(23) Maffucci, I.; Contini, A. Improved Computation of ProteinProtein Relative Binding Energies with the Nwat-MMGBSA Method. *J. Chem. Inf. Model.* **2016**, *56*, 1692–1704.

(24) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145–1152.

(25) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.

(26) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2009**, *31*, 455–461.

(27) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594.

(28) Liu, J.; He, X.; Zhang, J. Z. H. Improving the scoring of protein-ligand binding affinity by including the effects of structural water and electronic polarization. *Journal of chemical information and modeling* **2013**, *53*, 1306–1314.

(29) Mikulskis, P.; Genheden, S.; Ryde, U. Effect of explicit water molecules on ligand-binding affinities calculated with the MM/GBSA approach. *Journal of molecular modeling* **2014**, *20*, 2273.

(30) Maffucci, I.; Hu, X.; Fumagalli, V.; Contini, A. An Efficient Implementation of the Nwat-MMGBSA Method to Rescore Docking Results in Medium-Throughput Virtual Screenings. *Front. Chem.* **2018**, *6*.

(31) Korb, O.; Stützle, T.; Exner, T. E. An ant colony optimization approach to flexible proteinligand docking. *Swarm Intell.* **2007**, *1*, 115–134.

(32) Korb, O.; Stüttzle, T.; Exner, T. E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84–96.

(33) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein-ligand interaction. *Proteins Struct. Funct. Bioinf.* **2002**, *49*, 457–471.

(34) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magariños, M. P.; Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954.

(35) Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group ULC. 2018.

(36) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem* **2004**, *47*, 1750–1759.

(37) Murphy, R. B.; Repasky, M. P.; Greenwood, J. R.; Tubert-Brohman, I.; Jerome, S.; Annabhimoju, R.; Boyles, N. A.; Schmitz, C. D.; Abel, R.; Farid, R.; Friesner, R. A.

WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in LigandReceptor Docking. *J. Med. Chem* **2016**, *59*, 4364–4384.

(38) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(39) Case, D. A.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M.; Gohlke, H.; Goetz, A.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman, P. AMBER 2018. 2018.

(40) Sommer, K.; Friedrich, N.-O.; Bietz, S.; Hilbig, M.; Inhester, T.; Rarey, M. UNICON: A Powerful and Easy-to-Use Compound Library Converter. *J. Chem. Inf. Model.* **2016**, *56*, 1105–1111.

(41) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(42) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(43) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(44) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034.

(45) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113–135.

(46) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449–461.

(47) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441–451.

(48) Batsanov, S. S. Van der Waals Radii of Elements. *Inorg. Mater.* **2001**, *37*, 871–885.

(49) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins Struct. Funct. Bioinf.* **2004**, *55*, 383–394.

(50) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *Journal of Molecular Biology* **1996**, *261*, 470–489.

(51) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **1997**, *267*, 727–748.

(52) Sanchez-Linares, I.; Perez-Sanchez, H.; Cecilia, J. M.; Garcia, J. M. High-Throughput parallel blind Virtual Screening using BINDSURF. *BMC bioinformatics* **2012**, *13 Suppl 14*, S13.

(53) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry* **2015**, *36*, 1132–1156.

(54) Imbernón, B.; Cecilia, J. M.; Pérez-Sánchez, H.; Giménez, D. METADOCK: A parallel metaheuristic schema for virtual screening methods. *The International Journal of High Performance Computing Applications* **2017**, *32*, 789–803.

(55) Ferri, N.; Corsini, A.; Bottino, P.; Clerici, F.; Contini, A. Virtual Screening Approach for the Identification of New Rac1 Inhibitors. *Journal of Medicinal Chemistry* **2009**, *52*, 4087–4090.

(56) Conforti, P.; Zuccato, C.; Gaudenzi, G.; Ieraci, A.; Camnasio, S.; Buckley, N. J.; Mutti, C.; Cotelli, F.; Contini, A.; Cattaneo, E. Binding of the repressor complex REST-mSIN3b by small molecules restores neuronal gene transcription in Huntington's disease models. *Journal of Neurochemistry* **2013**, *127*, 22–35.

Graphical TOC Entry

Virtual Screening with the MM-PBSA methods: Beware of Internal Dielectric Constants
Xiao Hu, Alessandro Contini*