J Comput Aided Mol Des (2014) 28:201–209
DOI 10.1007/s10822-014-9722-6

SAMPL4 & DOCK3.7: lessons for automated docking procedures

Ryan G. Coleman · Teague Sterling · Dahlia R. Weiss

Received: 13 November 2013 / Accepted: 28 January 2014 / Published online: 11 February 2014

Abstract The SAMPL4 challenges were used to test current automated methods for solvation energy, virtual screening, pose and affinity prediction of the molecular docking pipeline DOCK 3.7. Additionally, first-order models of binding affinity were proposed as milestones for any method predicting binding affinity. Several important discoveries about the molecular docking software were made during the challenge: (1) Solvation energies of ligands were five-fold worse than any other method used in SAMPL4, including methods that were similarly fast, (2) HIV Integrase is a challenging target, but automated docking on the correct allosteric site performed well in terms of virtual screening and pose prediction (compared to other methods) but affinity prediction, as expected, was very poor, (3) Molecular docking grid sizes can be very important, serious errors were discovered with default settings that have been adjusted for all future work. Overall, lessons from SAMPL4 suggest many changes to molecular docking tools, not just DOCK 3.7, that could improve the state of the art. Future difficulties and projects will be discussed.

Keywords Molecular docking · Solvation · SAMPL · First-order models

Electronic supplementary material The online version of this article (doi:10.1007/s10822-014-9722-6) contains supplementary material, which is available to authorized users.

R. G. Coleman (✉) · T. Sterling · D. R. Weiss
Department of Pharmaceutical Chemistry, University of California, San Francisco, 1700 4th St, Box 2550, San Francisco, CA 94158, USA
e-mail: rgc@blur.compbio.ucsf.edu; ryan.g.coleman@gmail.com

Introduction

The SAMPL challenges have provided an excellent opportunity for prospective testing of computational methods against not just new experimental data, but also against many other computational methods and techniques [1–4], not unlike the Critical Assessment of protein Structure Prediction (CASP) [5] experiment. The SAMPL4 challenge [6] presented a unique opportunity to test methods on several new small molecule solvation energies [6, 7], two artificial, non-protein host–guest systems [8–10] and a protein system with several series of ligands designed and synthesized for an allosteric site [11, 12], with crystallographic pose data and binding affinity data.

The ambition of this paper was to compare the current molecular docking tool, DOCK3.7 [13] against both these experimental measures and against the other computational methods in the field. A slightly modified version of the ZINC Processing Pipeline [14] was used to build ligands and calculate solvation energies with AMSOL [15]. Virtual screening, pose prediction and affinity prediction were all attempted with this same automated protocol on HIV integrase, completely blindly (poses and affinities were predicted for all tested ligands, without knowledge of which ligands bound and in what positions). In this pursuit, we hope to assess DOCK3.7 against both experimental data and against other theories and methods. We discovered many ways in which our methods were insufficient or could use improvement, and found some hope in continued efforts on these methods as they were sometimes on par with the best methods of the field.

Methods

Methods used in this research follow recently published ligand preparation [14] and docking protocols [13], with

one important advancement, using ChemAxon's Marvin and cxcalc programs to compute ligand tautomers and protomers [16–18]. Anecdotally, cxcalc matched our chemical intuition for many molecules that previously had been wrongly tautomerized and protonated. The number of egregiously tautomerized and protonated molecules was low, as this tends to become all too obvious in docking where incorrect molecules often rise to the top of the list.

The specific combination of commands used was:

$$cxcalc\ in.smi\ dominanttautomerdistribution\ -H\ 7.4\ -C\ false\ -t\ 20\ \%$$

which was then parsed to keep only the molecules present at least 20 % of the time. We used pH = 7.4 for all the calculations in this work, which may have been a mistake for some systems and will be discussed later.

The rest of the ligand preparation using the ZINC Processing Pipeline [14] remains the same. Input in SMILES are converted to 3D using CORINA [19, 20] and OMEGA [21, 22], now using ChemAxon tools for protonation and tautomerization [16–18], using AMSOL [15] for partial charge and solvation calculations, and finally mol2db2 [13] to save multiple conformations, partial charge and solvation information into a single DB2 file for docking.

Docking proceeded automatically based on DOCK3.7 [13]. Receptor and ligand information from the crystal structures provided was used to start the calculations. For the HIV Integrase system, the PDB files 3nf7 and 3nf8 [23] were used as input. 3nf7 was used as it was a part of the Auto-DUD-E test set that was used in retrospective studies of DOCK3.7, though only the orthosteric site was used in that study. 3nf8 was suggested by the SAMPL4 organizers, and all three binding sites were used for automated docking.

Result and discussion

First-order models for binding affinity

Here we propose a first-order model for binding affinity prediction that all methods should be compared to. Rather than a basis in statistics like null models, this is a first-order model based on the maximal affinity of ligands concept [24]. The observation was that each heavy atom in a ligand could contribute up to 1.5 kcal/mol, up to a maximum of 15 heavy atoms where the affinities leveled off and more heavy atoms did not contribute to the total protein–ligand affinity. Extending this analysis into a first-order model of binding results in an equation where heavy atoms in each ligand are simply counted and converted to the maximal affinity that such a ligand could have in complex with a protein. This model was applied to all the affinity prediction challenges in SAMPL4, HIV integrase [11], octa-acid [9], and cb7 [8], and is shown in Fig. 1. Also of note is that for all these systems, the maximal affinity predicts a stronger binding affinity than was achieved, suggesting that none of these ligands have been completely optimized for the binding site. Of course, two of these are non-protein host–guest systems, not included in the original analysis, which may have significant differences in the achievable affinity. Pearson R values for the cb7 [8] system are 0.72, 0.5 for octa-acid [9], and 0.24 for the HIV integrase [11] system (though, really it is undefined as all the ligands have over 15 heavy atoms). This first-order model of binding is easy to calculate and represents an important benchmark for any method of calculating binding affinity.

Small molecule solvation

Small molecule solvation aims to measure the energies required to remove a small molecule from water (going

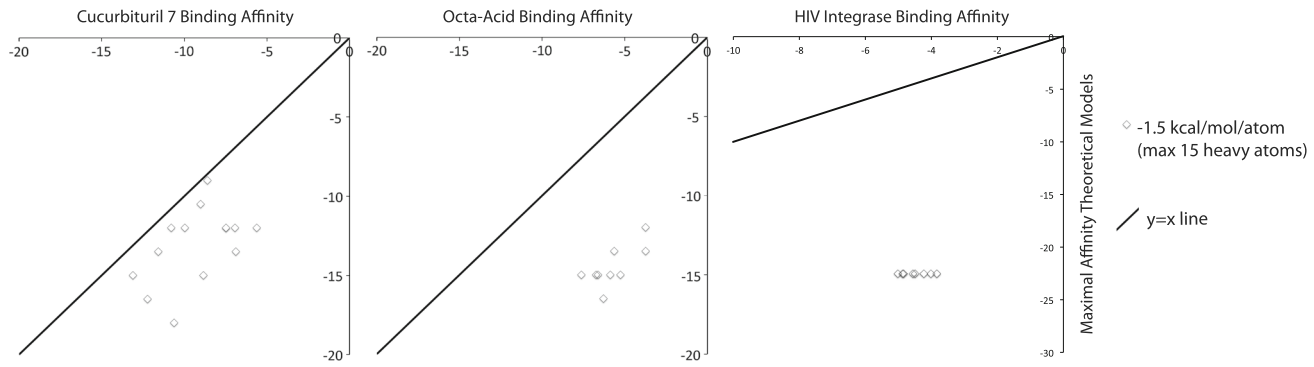

Fig. 1 The first-order affinity model [24], applied to the SAMPL4 systems

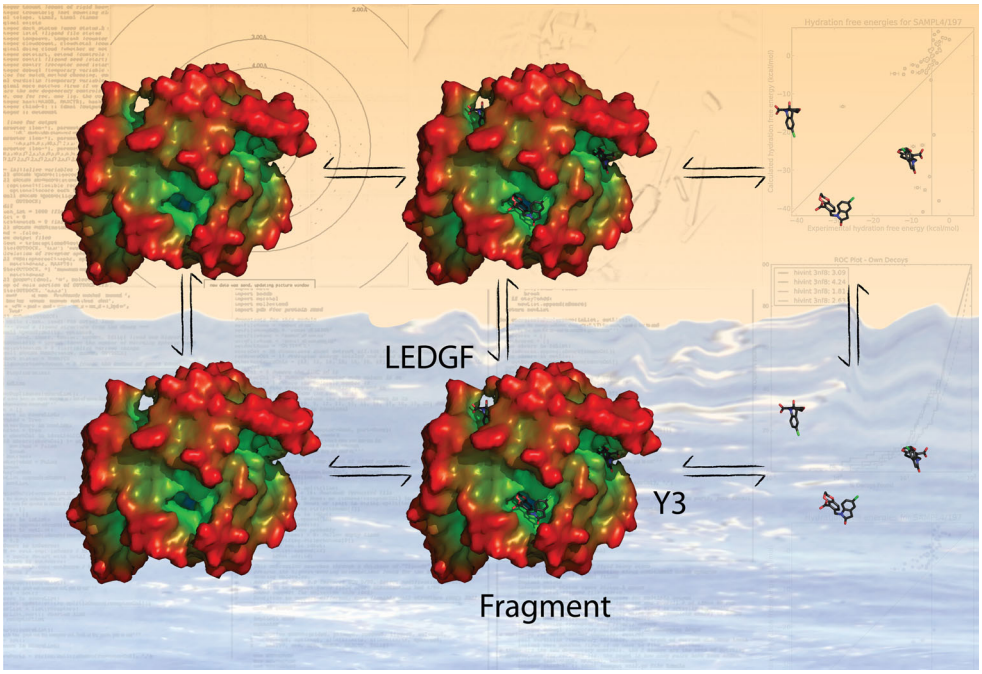

Fig. 2 Schematic view of binding, transitioning from vacuum (at *top*) to water (at *bottom*), and transitioning from the protein–ligand complex (*middle column*) to protein (*left column*) and ligand (*right column*). HIV integrase along with the crystallographic ligands in each of the three labeled sites are shown [23], colored according to Travel Depth [49]

from top to bottom in the right panel of Fig. 2). Forty-seven newly found measurements from the literature were the subject of this experiment [7]. As described in the method section, submissions using the ZINC Processing Pipeline [14], with a final solvation calculation done with AMSOL [15] were submitted. We used three different methods to provide AMSOL with starting conformations, as a means to test whether the initial input conformation was a factor in overall performance. Therefore, three submissions were sent into the SAMPL4 challenge: 158-where an arbitrary conformation from CORINA [19, 20] was used as an input for AMSOL (this is the ZINC default behavior), 196-where the lowest energy conformation from OMEGA was used as the AMSOL input, 197-where all conformations from OMEGA were used separately and the mean of the AMSOL solvation energies was reported. The advantage of this pipeline is that these calculations are very fast, fast enough that they can be accomplished for a very large library of purchasable molecules.

Briefly, AMSOL [15] uses semiempirical quantum chemistry to optimize small molecules and their partial charges, then uses a Generalized Born model to compute polar desolvation terms and a Solvent Accessible Surface Area model to compute the apolar desolvation terms. In molecular docking with the DOCK 3.5.54 [25] (later 3.6 [26, 27] and 3.7 [13]) line of programs, the partial charges are used for each small molecule to compute the electrostatic interaction energy with DelPhi [13]. The desolvation energies are used as-is (assuming full desolvation, true only for extremely buried binding sites) or with a Generalized Born partial desolvation model [26, 27]. While use of desolvation energies improves molecular docking in terms of retrospective enrichment or prospectively finding new ligands, testing the desolvation terms explicitly was not attempted.

As it turned out, AMSOL gave the worst predictions of solvation energy, by any method, submitted to the SAMPL4 challenge [6]. The three methods for providing starting conformations resulted in very similar predictions, indicating that the choice of input conformation was likely not a major factor in the inaccuracy of the calculation. The Pearson R of the experimental data to the three submissions ranged from 0.38 to 0.48. The mean unsigned error was between 7 and 8 kcal/mol. In comparison, other methods had a Pearson R of above 0.5, with well over half the submissions being above 0.8. The mean unsigned error of the next worst method was 4 kcal/mol, again the majority of submissions were below 2 kcal/mol.

On an independent test of AMSOL [28], it was quite competitive with other methods, however this was for a retrospective test set that likely contained some of the training molecules used to parameterize AMSOL itself. The authors remark that the "mean unsigned error of only 0.65 kcal/mol and a $R^2 = 0.88$", much better performance on the test set than the SAMPL4 set. This certainly reflects the difficulty of the SAMPL4 prospective set, and emphasizes the importance of using a prospective set in evaluating performance. Older test sets, limited by availability of experimental data, used smaller, simpler molecules. AMSOL, unlike some other techniques used in the SAMPL4 challenge, has not had the benefit of re-training to newer data, and had not been used in prior SAMPL

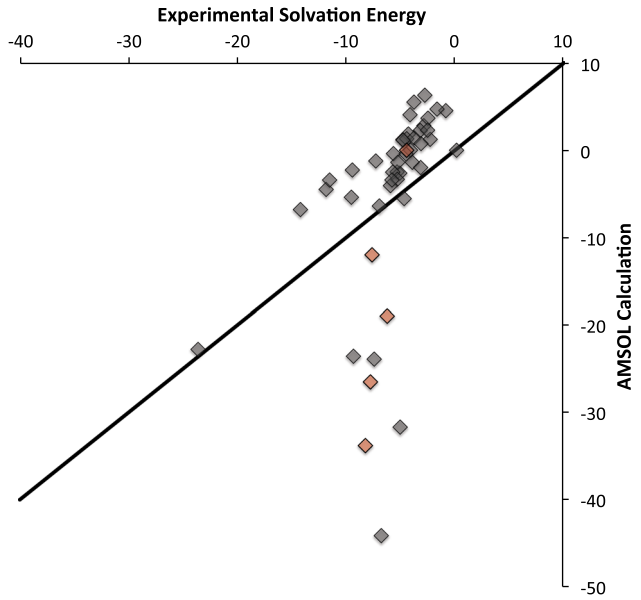

Fig. 3 ZINC [14] with AMSOL [15] computed solvation energies, compared to experimental values. Data points in red are errors definitely due to the ZINC Processing Pipeline

challenges. Though AMSOL has been in use as the desolvation term of the DOCK 3.x scoring function for over a decade [25], none of the follow-up experiments explicitly tested the desolvation term. It should be noted, however, that DOCK 3.7 does take into account a desolvation term, and this may overall contribute to the success of the docking method. It is not unreasonable to suppose that use of a more modern solvation energy calculation may improve the scoring function further.

Upon careful inspection of the data, it was realized that a small mistake was made, all molecules should have been neutral instead of 4 of them that were charged (at pH = 7.4). Even removing those 4 from consideration, and an additional molecule that was not calculated as it failed to process properly, the mean unsigned error fell only from 7.1 to 6.4 kcal/mol, still far above any other method. In further examination of the data, AMSOL appears to underestimate the solvation energy in most cases, except for molecules with many polar groups where it severely overestimates the solvation energy, see Fig. 3. In Fig. 3, the 5 molecules where the errors are certainly due to problems with the ZINC Processing Pipeline or with not being neutral are shown in red.

Though it may not be surprising that more advanced methods did well, it is somewhat interesting that many other fast methods performed among the top. These methods are: Semi-Explicit Assembly [29], various implementations of ZAP [30–32] and a parameterized solvation model [33, 34]. These methods all run in a few seconds per molecule, as fast as AMSOL and fast enough for running on a large test set. Additionally, all of these other fast methods had a Pearson R between each other of above 0.9, indicating that they agree quite often. In contrast, the Pearson R between the submissions using AMSOL and any of the other fast methods is quite low, on the order of 0.3–0.6.

It would seem advantageous to use other solvation methods in a physics-based docking method, however this was outside the scope of the SAMPL4 experiment, but will be the subject of further inquiry.

HIV integrase, virtual screening, pose prediction and affinity prediction

HIV Integrase presented a difficult challenge [11] for any computational method. First, the HIV Integrase has 1 orthosteric site and 2 allosteric sites, along with a symmetric dimer presenting 3 additional sites, see the protein depiction in Fig. 2. This makes predicting to which, if any, sites a small molecule will bind challenging. Additionally, the ligands and the non-binders from this set were very similar, as they were essentially a medicinal chemistry campaign of a few series. Many benchmark sets, including DUD-E [35], based on ChEMBL [36], do not discriminate orthosteric and allosteric binders, and therefore may actually not help discriminate and test computational methods.

Despite this challenge, automated molecular docking was used on the input PDB files to evaluate the performance of the method without human intervention. Ligands were prepared as described in methods, according to the ZINC Processing Pipeline [14] but with ChemAxon tools for protonation and tautomerization [16–18]. One docking run was done directly against the Auto-DUD-E preparation of 3NF7 [23] to the orthosteric site, exactly on the precomputed grids used in the DOCK3.7 paper [13]. To make pose analysis possible, the 3NF7 structure was aligned to the 3NF8 structure provided by the SAMPL4 organizers. The analysis of these two runs will be the subject of the next paragraph. Additionally, one ‘manual’ preparation was attempted, where an ordered water was added. Finally, we ran automated docking to each of the 3 binding sites in 3NF8. Additionally, a merged docking run was submitted, where each ligand’s best score & pose to any of the 3 sites was used. For each run except the first where pose analysis would have proven difficult, virtual screening results were submitted, along with pose predictions and affinity predictions, all done blindly without knowledge of which ligands were binders and what their poses were. Though this was an additional challenge, the automated methods have no way to take advantage of this additional information of which ligands were binders (for the pose prediction challenge) and the poses of the known binders (for the affinity challenge).

Table 1 Auto-DUD-E comparison of grid sizes over 5 sampling parameters and 102 proteins

	0.33 Å	0.2 Å	Difference in means	*p* value	Cohen effect size
Mean adj. logAUC	15.6	16.4	0.8	0.16 (not significant)	0.088 (trivial)
Mean AUC	66.8	67.4	0.6	0.22 (not significant)	0.069 (trivial)
Mean EF1	13.2	14.0	0.8	0.19 (not significant)	0.078 (trivial)

Table 2 Summary of SAMPL4 solvation submissions

SAMPL4 ID	Description
158	ZPP [14], Marvin [16–18], Arbitrary Conformation (default in ZINC)
196	ZPP [14], Marvin [16–18], Mean over 3D conformations
197	ZPP [14], Marvin [16–18], Best conformation

Upon analysis of the 3NF7/3NF7-rotated submission (SAMPL IDs 198 & 200), a discrepancy was easily noticed. A reasonable assumption would be that, barring numerical errors, the docking virtual screening results for these two would be the same. After extensive analysis, we discovered the source of the error, the Chemgrid program used to generate van der Waals grids [37], used a default setting of 3 grids per Å, or a grid spacing of 0.33 Å. Given the exponentials of r^6 and r^{12} in the Lennard–Jones equations, this grid spacing is almost surely unsuitable for accurately approximating the van der Waals forces. Even small rotations of 10 degrees caused extreme differences in the van der Waals scores. For this reason, a lower grid spacing of 0.2 Å was generated (after SAMPL4 submission) across the entire Auto-DUD-E test set, and a complete run across five sampling parameters was conducted. This grid spacing did increase the memory usage, but not substantially, indicating that smaller grid spacings could be examined eventually. Complete results are shown in Supplementary Tables 1–3. In Table 1, the results in terms of means over AUC, adjusted logAUC [27] and EF1 are shown. Though the vast majority of the docking runs in Auto-DUD-E were better with grid spacing at 0.2 Å, and the mean difference in all the metrics of virtual screening success showed improvement, the result was statistically insignificant. Given the insignificance and the trivial Cohen effect size [38, 39], we cannot conclude that a finer grid spacing is better, but it is demonstrably not worse than a coarser grid spacing. Since the memory and speed are not seriously affected by a finer grid spacing, further work in refining the optimum grid spacing would seem to be worth pursuing. Perhaps most importantly, a reduction in the grid spacing removes a source of inconsistency in docking poses and scores between what should be identical docking runs (Table 2).

Table 3 Summary of SAMPL4 HIV integrase submissions

	Virtual screening	Pose prediction[a]	Affinity prediction[b]
3NF7 [23], Auto-DUD-E preparation [13]	198	n.a.[c]	199
3NF7 [23], Aligned to 3NF8, DOCK3.7 automatic preparation [13]	200[d]	539	201[d]
3NF7 [23], Aligned to 3NF8, ordered water added by hand, otherwise automatic DOCK3.7preparation [13]	238	540	233
3NF8 [23], 257/Y3 site, DOCK3.7 automatic preparation [13]	239	537	234
3NF8 [23], 267/LEDGF site, DOCK3.7 automatic preparation [13]	240	536	235
3NF8 [23], 277/Fragment site, DOCK3.7 automatic preparation [13]	241	535	236
3NF8 [23], Combination of separate runs to all 3 binding sites (257, 267, 277) [13]	242	538	237

[a] Note that these predicted poses were made for all compounds, without knowledge of which ones bound, as opposed to most other submissions

[b] Note that these predicted affinities were made for all compounds, without knowledge of the poses of the binders, as opposed to most other submissions

[c] Poses were not submitted, as they would have been in the wrong frame of reference

[d] Submissions 198 & 200 should have been identical, as well as 199 & 201, but were not as discussed. Pearson Correlation Coefficient = 0.93304677

As it turns out, only virtual screening SAMPL ID 240 (see Table 3) was correct, as the majority of ligands bound to the LEDGF site used in docking (in fact most were designed to bind to this site). However, even entry 240 did not do much better than random as far as virtual screening goes, again likely owing to the complicated and difficult challenge. However, given this, it is interesting that docking to the other sites was not much worse than the automatic docking to the LEDGF, entry 240, see Fig. 4; Table 3, in fact some were slightly better. Entry 240 (LEDGF) had a ROC AUC of 0.53, while other entries to

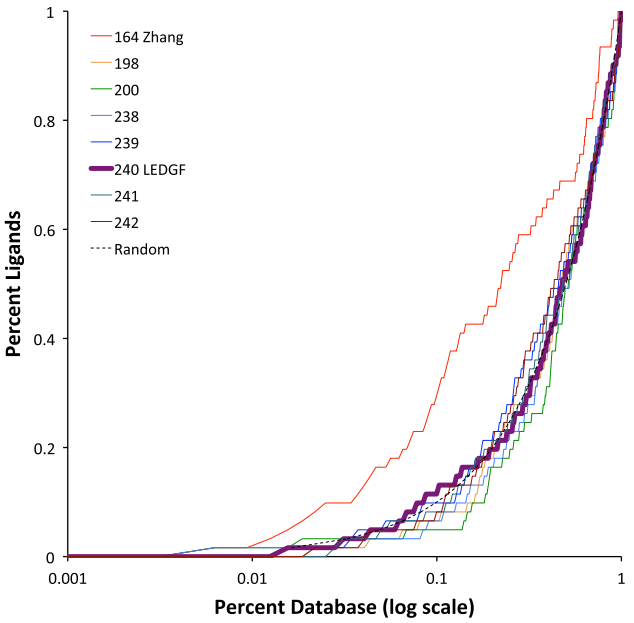

Fig. 4 Log ROC curves for all our virtual screening results submitted to SAMPL4, along with the winning run [40]. The docking run to the correct LEDGF allosteric site with DOCK3.7 is highlighted with a *thicker line*. See Table 3 for a full description of the SAMPL ID submissions

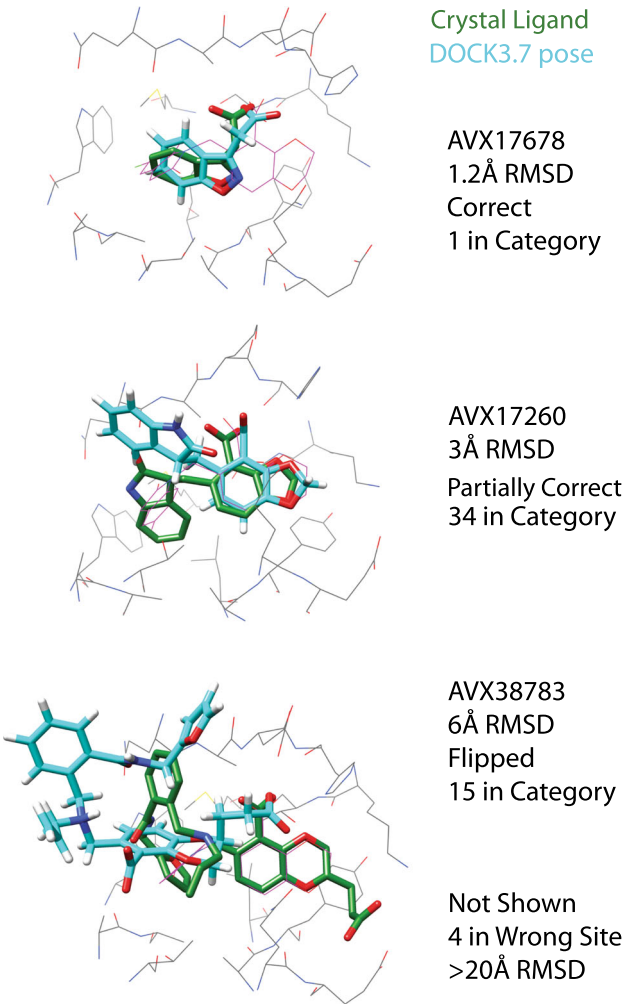

Fig. 5 Pose analysis of SAMPL4 entry 536, the best DOCK3.7 run. DOCK3.7 generated poses shown in *cyan*, crystallograhic ligands shown in *forest green*. The 3NF8 [23] ligand is shown in *purple lines*, the protein is shown in *gray lines*. 4 categories are shown: (1) Correct ligands with low RMSD, (2) Partially Correct ligands where one end of the ligand has been placed correctly (3) Incorrect ligands placed in a flipped manner in the binding site, with no overlap, (4) Ligands not placed in the correct site, of which there are 4, not shown

other sites were 0.48–0.56 [12]. The best submission for virtual screening used a combination of many methods, including a lot of analysis done carefully by hand [40], suggesting that computational methods are still not as good as careful analysis by experts. DOCK3.7 did not do well even compared to other methods at virtual screening from this challenging set, likely due to the simplicity of the energy function or other factors as discussed later.

For pose prediction analysis, our ligand placement was very good, only the best submission was better than the automated procedure we used [40]. The automatic docking to the LEDGF site was the best submission among ours, entry number 536 (see Table 3), as the ligands were limited to being placed in the allosteric site. Our RMSD calculations were done using our own code [13], in a way that differs slightly from the official RMSD calculations done by the SAMPL4 organizers [12]. Though our numbers may differ slightly they seem to agree, DOCK3.7 predicted many ligands correctly in the main interacting group, but many of the extended tails were incorrectly predicted, as shown in Fig. 5. Only one ligand could truly be considered correctly predicted, one of the smallest fragments, shown at the top of Fig. 5. When there were two carboxylate groups on the ligand, it seems that DOCK3.7 often had the incorrect one interacting with the protein, as shown at the bottom of Fig. 5. Given the difficulties with multiple carboxylate groups, and the earlier discovery that AMSOL is overestimating the solvation energy of these groups, it makes some sense that DOCK3.7 would find posing and scoring these ligands challenging. As mentioned, a few ligands experimentally bound in the other two sites, but this docking run precluded binding there as it targeted only the LEDGF site. Properly choosing which binding site to place each ligand is likely a difficult problem, especially with the current generation of DOCK energy functions, which are at best approximating relative enthalpies and not absolute energies of binding.

The first-order model for HIV integrase binding affinity had a Pearson R = 0.24 and MUE = 18.8 kcal/mol. Even the docking affinity had a Pearson R = 0.64 and MUE = 16.6 kcal/mol, despite the poor performance as shown in Fig. 6. Though the binding affinity correlated

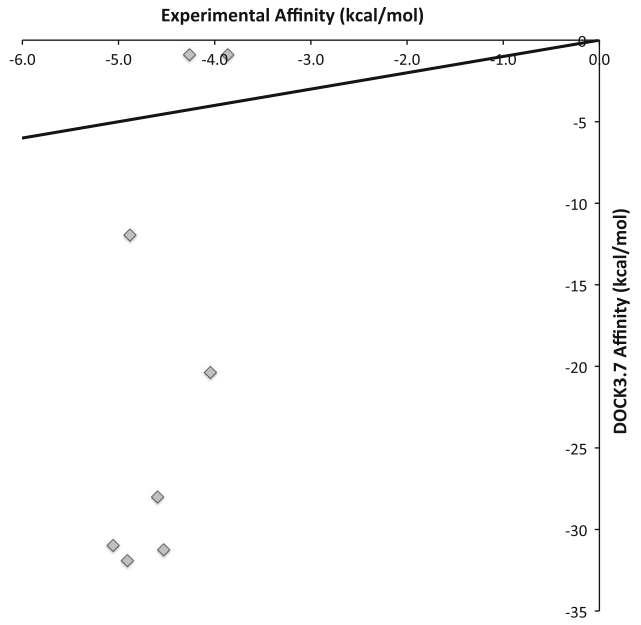

Fig. 6 DOCK3.7 calculated binding affinities versus experimental affinities for HIV Integrase. Y = X line shown

better with docking score than the first-order model, there were significant absolute errors in the binding affinities. Again, these predictions were made without knowledge of the poses of the ligands, which only one other group attempted. However, their errors were much lower than ours, and the errors of the other groups where poses were known, were even lower. Given the complexity of the methods and the uses of prior information involved in the calculations, the ordering of the performance of the methods is in line with expectations.

It is quite understandable that the affinities calculated by DOCK3.7 are wrong, as at best, it calculates a relative enthalpy in the current version of the scoring function. Many terms are left out, including protein & ligand internal energy, hydrophobic effect of the protein, hypothesized to be a large contribution [41], and receptor desolvation. Receptor desolvation is depicted in the left column of Fig. 2, and even Fig. 2 neglects entropic terms in the schematic view of protein–ligand association. As discovered earlier, the ligand desolvation terms suffer from large errors as well, which could easily be contributing to the errors in protein–ligand binding affinity.

Conclusions

For physics-based docking programs, the SAMPL challenges present an excellent opportunity for prospective testing of methods against experiment, and against other computational and theoretical techniques. Lessons learned from the solvation transfer energy part of the challenge mostly illuminate the failures of AMSOL, a comparatively older technique than others used in the challenge, to meet the challenges of more difficult molecules. The obvious question of whether other fast computational methods that predict the solvation energy more accurately will improve docking is left open for further investigation. First-order models of binding affinity, taken from previously published work, represent an important, biophysically-based model for all methods to compare against [24].

The binding of a few series of similar small molecules to a protein with multiple binding sites remains a challenging task for any computational method to predict. The HIV integrase portion of the SAMPL4 challenge presented just such a challenge. Despite this challenge, automated docking to the allosteric site performed adequately in terms of pose reproduction but barely better than random in terms of virtual screening. In both these tasks, it was outperformed by one group who used a more manual method [40], reiterating the importance that expert review can have on these challenges. No other docking method attempted to compare the predicted binding affinities with those gathered from experiments, docking unsurprisingly did quite poorly, though it was a small improvement over the first-order binding models.

A final lesson was learned that the approximations in grid-based energy calculations, often used in docking, should be carefully checked as we discovered large errors between runs caused by grid sizes that were insufficiently small to reproduce the underlying steep van der Waals function. Though smaller grid sizes showed improvement in automated testing, it was not statistically significant. This is an important lesson for any docking or scoring program, one which will be further explored.

The failures of AMSOL to correctly predict the solvation energy are not unexpected here. The molecules tested in the SAMPL4 challenge were chosen to be larger, have more polarity and generally challenge existing techniques [7]. Though AMSOL has been used in the DOCK 3.x scoring function for over a decade [25], the desolvation values have never been directly experimentally tested. Indeed, as the prediction made and tested is whether a small molecule will bind to a given protein, errors in desolvation energy alone are impossible to determine from a binding experiment.

Despite errors—including the desolvation energy, grid errors in the van der Waals term, etc.—the DOCK 3.x scoring function has shown some promise on ligand discovery: 14 % [42] to 48 % hit rate [43] for fragments against beta-lactamases, 20–40 % hit rates against G-Protein Coupled Receptors [35, 44–48]. One can only imagine how improvements in the energy function, as well as other aspects of virtual screening, will improve the results in terms of finding new ligands and perhaps even harder tasks

like pose prediction and affinity prediction. It is not just our hope, but the hope of the field, that further improvements to accurately model the energies involved in protein–ligand (and ligand–water) interactions will improve our ability to predict new ligands, which the future SAMPL challenges will surely play a part.

Acknowledgments We acknowledge financial support from National Research Service Award-Kirschstein fellowships F32GM096544 (to RGC) and F32GM093580 (to DRW), and US National Institutes of Health grants GM59957 and GM71630 (to Brian K. Shoichet). We would also like to thank the SAMPL4 organizers, as well as ChemAxon for Marvin & cxcalc, OpenEye for OMEGA and OEtools. Final acknowledgements to Paul Hawkins for motivation and guidance on statistical methods.

References

1. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS (2008) Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. J Med Chem 51(4):769–779. doi:10.1021/jm070549+
2. Skillman AG, Geballe M, Nicholls A (2010) SAMPL2 challenge: prediction of solvation energies and tautomer ratios. J Comput Aided Mol Des 24(4):257–258. doi:10.1007/s10822-010-9358-0
3. Skillman AG (2012) SAMPL3: blinded prediction of host-guest binding affinities, hydration free energies, and trypsin inhibitors. J Comput Aided Mol Des 26(5):473–474. doi:10.1007/s10822-012-9580-z
4. Guthrie JP (2009) A blind challenge for computational solvation free energies: introduction and overview. J Phys Chem B 113(14):4501–4507. doi:10.1021/jp806724u
5. Kryshtafovych A, Fidelis K, Moult J (2013) CASP10 results compared to those of previous CASP experiments. Proteins Struct Func Bioinf n/a–n/a. doi:10.1002/prot.24448
6. Mobley DL, Wymer KL, Lim NM, Guthrie JP (2014) Blind prediction of solvation free energies from the SAMPL4 challenge. J Comput Aided Mol Des. doi:10.1007/s10822-014-9718-2
7. Guthrie JP (2014) SAMPL4, a blind challenge for computational solvation free energies: the compounds considered. J Comput Aided Mol Des 28 (in press)
8. Cao L, Isaacs L (2013) Absolute and relative binding affinity of cucurbit[7]uril toward a series of cationic guests. Supramol Chem 1:1–9
9. Gibb CLD, Gibb BC (2014) Binding of cyclic carboxylates to octa-acid deep-cavity cavitand. J Comput Aided Mol Des. doi:10.1007/s10822-013-9690-2
10. Muddana HS, Fenley AT, Mobley DL, Gilson MK (2014) The SAMPL4 host-guest blind prediction challenge: an overview. J Comput Aided Mol Des 28 (in press)
11. Peat TS, Dolezal O, Newman J, Mobley D, Deadman JJ (2014) Interrogating HIV integrase for compounds that bind- a SAMPL challenge. J Comput Aided Mol Des. doi:10.1007/s10822-014-9721-7
12. Mobley DL, Liu S, Lim NM, Deng N, Branson K, Perryman SF, Levy RM, Gallicchio E, Olson AS (2014) Blind prediction of HIV integrase binding from the SAMPL4 challenge. J Comput Aided Mol Des. doi:10.1007/s10822-014-9723-5
13. Coleman RG, Carchia M, Sterling T, Irwin JJ, Shoichet BK (2013) Ligand pose and orientational sampling in molecular docking. PLoS ONE 8(10):e75992. doi:10.1371/journal.pone.0075992
14. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) ZINC: a free tool to discover chemistry for biology. J Chem Inf Model. doi:10.1021/ci3001277
15. Hawkins GD, Giesen DJ, Lynch GC, Chambers CC, Rossi I, Storer JW, Li J, Zhu T, Thompson JD, Winget P, Lynch BJ, Rinaldi D, Liotard DA, Cramer CJ, Truhlar DG (2004) AMSOL. 7.1 edn. doi:http://comp.chem.umn.edu/amsol/
16. Csizmadia F, Tsantili-Kakoulidou A, Panderi I, Darvas F (1997) Prediction of distribution coefficient from structure. 1. Estimation method. J Pharm Sci 86(7):865–871. doi:10.1021/js960177k
17. Szegezdi S, Csizmadia F (2004) Prediction of dissociation constant using microconstants. In: Paper presented at the 227th American Chemical Society National Meeting, Anaheim, CA
18. Szegezdi S, Csizmadia F (2007) Calculating pKa values of small and large molecules. In: Paper presented at the 233rd American Chemical Society National Meeting, Chicago, IL
19. Sadowski J, Gasteiger J, Klebe G (1994) Comparison of automatic three-dimensional model builders using 639 X-ray structures. J Chem Inf Comput Sci 34(4):1000–1008. doi:10.1021/ci00020a039
20. Tetko I, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin V, Radchenko E, Zefirov N, Makarenko A, Tanchuk V, Prokopenko V (2005) Virtual computational chemistry laboratory, design and description. J Comput Aided Mol Des 19(6):453–463. doi:10.1007/s10822-005-8694-y
21. OMEGA (2013) OpenEye software. Santa Fe, NM
22. Toolkits OpenEye (2013) OpenEye scientific software. Santa Fe, NM
23. Rhodes D, Peat T, Vandegraaff N, Jeevarajah D, Le G, Jones E, Smith J, Coates J, Winfield L, Thienthong N, Newman J, Lucent D, Ryan J, Savage G, Francis C, Deadman J (2011) Structural basis for a new mechanism of inhibition of HIV-1 integrase identified by fragment screening and structure-based design. Antivir Chem Chemother 21(4):155–168
24. Kuntz ID, Chen K, Sharp KA, Kollman PA (1999) The maximal affinity of ligands. Proc Natl Acad Sci U S A 96(18):9997–10002
25. Wei BQ, Baase WA, Weaver LH, Matthews BW, Shoichet BK (2002) A model binding site for testing scoring functions in molecular docking. J Mol Biol 322(2):339–355
26. DOCK3.6 (2011) http://dock.compbio.ucsf.edu/DOCK3.6/
27. Mysinger MM, Shoichet BK (2010) Rapid context-dependent ligand desolvation in molecular docking. J Chem Inf Model 50(9):1561–1573
28. Yang Q, Sharp KA (2006) Atomic charge parameters for the finite difference poisson-boltzmann method using electronegativity neutralization. J Chem Theory Comput 2(4):1152–1167. doi:10.1021/ct060009c
29. Fennell CJ, Kehoe CW, Dill KA (2011) Modeling aqueous solvation with semi-explicit assembly. Proc Natl Acad Sci 108(8):3234–3239
30. Grant JA, Pickup BT, Nicholls A (2001) A smooth permittivity function for Poisson-Boltzmann solvation methods. J Comput Chem 22(6):608–640. doi:10.1002/jcc.1032
31. Ellingson B, Skillman AG, Nicholls A (2010) Analysis of SM8 and Zap TK calculations and their geometric sensitivity. J Comput Aided Mol Des 24(4):335–342. doi:10.1007/s10822-010-9355-3
32. Ellingson BA, Bayly CI, Wlodek S, Geballe MT, Skillman AG, Nicholls A (2014) Placeholder for Ellingson/ZAP solvation paper
33. Park H (2014) Extended solvent-contact model approach to SAMPL4 blind prediction challenge for hydration free energies. J Comput Aided Mol Des 28 (in press)
34. Choi H, Kang H, Park H (2013) New solvation free energy function comprising intermolecular solvation and intramolecular self-solvation terms. J Cheminform 5(1):1–13

35. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem 55(14):6582–6594. http://pubs.acs.org/doi/abs/10.1021/jm300687e
36. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2011) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40(D1):D1100–D1107
37. Meng EC, Shoichet B, Kuntz ID (1992) Automated docking with grid-based energy evaluation. J Comp Chem 13:505–524
38. Cohen J (1988) Statistical power analysis for the behavioral sciences. Routledge, London
39. Hawkins PC (2014) Searching with statistics: is pose prediction perfectible? Abstracts of Papers of the American Chemical Society 247. (ACS Meeting Indianapolis, Indiana 2013)
40. Voet ARD, Kumar A, Berenger F, Zhang KYJ (2014) Combining in silico and in cerebro approaches for virtual screening and pose prediction in SAMPL4. J Comput Aided Mol Des. doi:10.1007/s10822-013-9702-2
41. Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, Salzberg AC, Huang ES (2007) Structure-based maximal affinity model predicts small-molecule druggability. Nat Biotech 25(1):71–75
42. Chen Y, Shoichet BK (2009) Molecular docking and ligand specificity in fragment-based inhibitor discovery. Nat Chem Biol 5(5):358–364. http://www.nature.com/nchembio/journal/v5/n5/suppinfo/nchembio.155_S1.html
43. Teotico DG, Babaoglu K, Rocklin GJ, Ferreira R, Giannetti AM, Shoichet BK (2009) Docking for fragment inhibitors of AmpC β-lactamase. Proc Natl Acad Sci USA 106(18):7455–7460
44. Carlsson J, Coleman RG, Setola V, Irwin JJ, Fan H, Schlessinger A, Sali A, Roth BL, Shoichet BK (2011) Ligand discovery from a dopamine D3 receptor homology model and crystal structure. Nat Chem Biol 7(11):769–778. http://www.nature.com/nchembio/journal/v7/n11/abs/nchembio.662.html-supplementary-information
45. Carlsson J, Yoo L, Gao ZG, Irwin JJ, Shoichet BK, Jacobson KA (2010) Structure-based discovery of A2A adenosine receptor ligands. J Med Chem 53(9):3748–3755. doi:10.1021/jm100240h
46. Kolb P, Rosenbaum DM, Irwin JJ, Fung JJ, Kobilka BK, Shoichet BK (2009) Structure-based discovery of β2-adrenergic receptor ligands. Proc Natl Acad Sci USA 106(16):6843–6848. http://www.pnas.org/content/106/16/6843.short
47. Kruse AC, Weiss DR, Rossi M, Hu J, Hu K, Eitel K, Gmeiner P, Wess Jr, Kobilka BK, Shoichet BK (2013) Muscarinic receptors as model targets and antitargets for structure-based ligand discovery. Mol Pharmacol 84(4):528–540
48. Weiss DR, Ahn S, Sassano MF, Kleist A, Zhu X, Strachan R, Roth BL, Lefkowitz RJ, Shoichet BK (2013) Conformation guides molecular efficacy in docking screens of activated Œ ≤ -2 adrenergic G protein coupled receptor. ACS Chem Biol 8(5):1018–1026. doi:10.1021/cb400103f
49. Coleman RG, Sharp KA (2006) Travel depth, a new shape descriptor for macromolecules: application to ligand binding. J Mol Biol 362(3):441–458