Standardization of virtual-screening and post-processing protocols relevant to in-silico drug discovery

Sunita Gupta¹ · Andrew M. Lynn² · Vibha Gupta¹

Received: 4 May 2018 / Accepted: 17 November 2018 / Published online: 30 November 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Structure-based drug discovery has emerged as a powerful tool in computational drug discovery and has gained rapid acceleration due to the development of better algorithms for high-end computation in an affordable time. Molecular docking and virtual screening methods are routinely used for the purpose but computing the ligand binding energies with inbuilt scoring functions accurately is still a limitation. Although, MMPBSA and MMGBSA are routinely employed tools for achieving accurate binding free energies, they are applied on well-equilibrated explicitly solvated systems and are computationally expensive and time-consuming. This study compares different post-processing protocols performed on an in-silico screened benchmarked *P. falciparum* Dihydrofolate reductase (PfDHFR) dataset with AutoDockVina. The docked and implicitly solvated complexes were subjected to (1) rescoring, (2) energy minimization and (3) Binding Estimation After Refinement (BEAR) algorithm. Subsequently, binding free energies were computed using three different tools—MMPBSA, MMGBSA and “g_mmpbsa”. Surprisingly, rescoring alone displays lower accuracy than the inherent scoring function of the AutoDockVina. However, encouraging results were seen after post-processing with the other two protocols. The results suggest that MMPBSA applied on energy minimized conformations is able to achieve 42-fold reductions in computational time as opposed to the BEAR algorithm with comparable accuracy.

Keywords Standardization · AutoDockVina · BEAR algorithm · MM(PB/GB)SA · g_mmpbsa

Introduction

Since past few decades virtual screening (VS) has gained importance in the field of computational drug discovery. This technique has accelerated the process of screening millions of compounds due to the availability of faster

computers and new algorithms. VS involves estimating the binding poses using approximated scoring functions and binding energies corresponding to the generated conformations. Major limitations lie in reliable simulations of the flexibility of receptor and small-molecules and in calculating the binding energies close to experimental values. As a result, false-negative compounds are also scored as top hits and community is constantly working on enhancing the accuracy of predictions. Post-processing of the docked complexes is one technique that is employed for accurate docking orientations and filtering out false-negative hits from the docking ensembles as well as rank the top hits on the basis of calculated binding free energies. Various methods are available depending on the levels of accuracy and computational expenses like free energy perturbation (FEP) (Kollman 1993), linear response (LR) (Lybrand et al. 1986), thermodynamics integration (TI) (Åqvist et al. 1994), molecular mechanics Poisson–Boltzmann surface area (MMPBSA), and molecular mechanics generalized born surface area (MMGBSA) (Kollman et al. 2000). Among the available methods, MMPBSA and

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-018-1523-5>) contains supplementary material, which is available to authorized users.

✉ Andrew M. Lynn
andrew@jnu.ac.in

✉ Vibha Gupta
vibha.gupta@jiit.ac.in

Sunita Gupta
sunita.bio@gmail.com

¹ Department of Biotechnology, Jaypee Institute of Information Technology, A-10, Sector-62, Noida, Uttar Pradesh 201309, India

² School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India

MMGBSA are widely used for predicting binding affinities of receptor-small-molecule complexes as they are faster than others (Rastelli et al. 2009). However, both these protocols work on well-converged systems of receptor-small-molecule complexes generated during explicit molecular dynamics (MD) simulations. The explicit equilibration is highly time-consuming and requires huge computation, especially when dealing with large number of compounds. This study addresses these limitations and proposes a post-docking protocol that is accurate as well as requires reduced computational resources. The protocol consists of rescoring of implicitly solvated docked and refined complexes and plotting the energies in the form of a metric “Enrichment curve”. This curve evaluates the performance of the screening strategy by its capacity to enrich a small number of true binders in the top-rank from the larger number of non-binders (Huang et al. 2006). *P. falciparum*, the main causative organism for malaria, is responsible for 1–3 million deaths every year. In addition due to emerging drug resistance strains, novel drug targets and the development of effective anti-malarials is urgently required (Sharma 2017; de Ruyck and Wouters 2008). PfDHFR is one such well-documented drug target where structure-based drug discovery efforts are ongoing. Directory of Useful Decoys (DUD) is a benchmarking dataset, available at <http://blaster.docking.org/dud/> (Huang et al. 2006). In the present study, PfDHFR dataset developed by Huang et al. 2006 is employed for the testing of virtual screening and post-processing protocols which include 201 positives and 7145 negative controls. Along with 201 positive controls, WR99210, a known potent inhibitor of PfDHFR, was also included in the data under study (Yuvaniyama et al. 2003). Virtual screening was performed using AutoDockVina (Trott and Olson 2010) and its docked poses were considered as one of the conformations for rescoring. Further, two other conformations were generated using—(1) implicit energy minimization for 2000 steps and (2) refinement using BEAR algorithm (Rastelli et al. 2009). On the above generated post-refined conformations, energy calculations were performed using MMPBSA/MMGBSA methods implemented in Amber16 (Case et al. 2017) and another Gromacs (Pronk et al. 2013) compatible tool g_mmpbsa (Kumari et al. 2014). The benchmarking on PfDHFR dataset using AutoDock (Huey et al. 2007) and BEAR algorithm has already been reported and show excellent performance of BEAR as compared to virtual screening (Rastelli et al. 2009). The enrichment curve obtained in the present study displays further improvement in ranking of known ligands after docking with AutoDockVina and post-processing employing energy minimization alone. The work highlights that even the energy minimization of docked complexes is able to achieve accuracy comparable to BEAR protocol with

outstanding 42-folds reduction in computational time as bonus.

Materials and methods

Database selection and virtual screening

In this study, PfDHFR dataset with known positive and negative controls available at DUD was taken for standardization of virtual screening and post-processing techniques. The B-chain of the crystal structure of PfDHFR (PDB-id 1J3I) complexed with NADPH was selected for virtual screening and had four missing residues (87–90) which were modeled using Modeller9v5 (Šali and Blundell 1993; Martí-Renom et al. 2000). The model having least DOPE score was considered best and further validated using RAMPAGE (Lovell et al. 2003) and ProSA (Wiederstein et al. 2007). All the positive and negative control compounds were docked into modeled structure using AutoDockVina. A cubic box of 25*25*25 Å was centered on the binding site and AutoDockTools4 (Morris et al. 2009) were used to assign Gasteiger charges to all atoms with correct rotatable bonds of receptor as well as small-molecules. AutoDockVina generates multiple conformations and least energy conformation was selected for post-processing.

System preparation

The post-processing was performed by estimating the binding free energies using MM(PB/GB)SA methods available in AMBER package and g_mmpbsa tool which is compatible with Gromacs software. Before proceeding for binding free energy estimation, the molecular mechanics parameters for small-molecules were assigned with the antechamber module of Amber16 using generalized amber force field (gaff) (Wang et al. 2004) and AM1-BCC (Jakalian et al. 2002) charge methods. Parmchk was used to assign missing force field parameters and the receptor was parameterized using ff99SB forcefield of AMBER. The parameters for NADPH were taken from AMBER parameter database (<http://research.bmh.manchester.ac.uk/bryce/amber/>) using gaff force field and AM1-BCC charge method. Acpye.py (da Silva et al. 2012), a python script was used for automatic generation of topology and parameters from Amber to Gromacs compatible format.

Structure refinement

Apart from receptor-small-molecule docked conformations, two other sets of conformations were obtained with different strategies, namely (1) Implicit energy minimization for 2000 steps with distance-dependent dielectric and (2) Refinement

using BEAR algorithm, which included minimization followed by short MD of small-molecules and further re-minimization of the complexes. Energy minimization of the complexes were performed using sander module of AMBER16 for 2000 steps with a distance-dependent dielectric constant $\epsilon = 4r$, and a cutoff of 12 Å. The refinement with BEAR algorithm includes 2000 steps of energy minimization of entire complexes followed by 100 ps of MD simulations in which only small molecules were allowed to move and receptor was restrained plus the re-minimization of the entire complexes for 2000 steps. The MD was performed at 300 K, with SHAKE (Ryckaert et al. 1977) turned on for bonds including hydrogen atoms and with time step of 2 fs. Finally, the single structure of each refined complexes generated using both the protocols was subjected to free energy estimation using MM(PB/GB)SA and g_mmpbsa tool.

The AutoDockVina energies and binding free energy values obtained post-processing were analyzed by plotting ligand enrichment curves. The ligand enrichment curves were generated in R language (Team RC 2014) using package “enrichvs” (Hiroaki 2015). Last, visual analysis of potent inhibitor WR99210 post-docking and post-refinement was performed using Pymol (DeLano et al. 2002) to see the orientations as well as interactions with the crucial residues with reference to crystal structure conformation.

Results and discussions

The three-dimensional homology model of the complete PfDHFR (including the missing residues ⁸⁷DNVN⁹⁰) was built using the PDB-1J3I and is shown in Supplementary material (Supplementary Information Fig. 1a). The model was validated for stereochemistry and energetic by Ramachandran plot, (Supplementary Information Fig. 1b) with 93.4% residues in the favorable region, 6.1% in the allowed region and 0.4% in outlier region. Z-score focuses on the quality of the model and total energy of the structure. The calculated Z-score value for modeled structure is −7.97 which is within the acceptable range (Supplementary Information Fig. 1c). The ProSA analysis also showed that almost all the residues in the modeled PfDHFR structure had negative interaction energies hence energetically contributing to the stability of the structure (Supplementary Information Fig. 1d). The AutoDockVina energies for the ligands and decoys range from −11.9 to −6.3 and −10.7 to −5.0 Kcal/mol, respectively. As these energies fail to discriminate between positive and negative controls, post-processing of the docked complexes is crucial.

This study has attempted to reduce the false conformations of small molecules generated during virtual screening by increasing the accuracy of binding free energy estimation. Two different protocols, namely energy minimization

and three-tier BEAR refinement (MM/MD/MM) on AutoDockVina generated conformations were implemented for refinement. Three different methods, namely AMBER based MMPBSA (Fig. 1a), MMGBSA (Fig. 1b) and g_mmpbsa (Fig. 1c) were employed for computing binding free energies (Fig. 1). As observed in Fig. 1_a1, binding free energies of AutoDockVina generated conformations of all the active ligands (red dots) and decoys (black dots) data points, processed with MMPBSA are overlapping with no clear demarcation between them. While post-processing of docked conformations employing energy minimization in Fig. 1_a2 and refinement using the BEAR algorithm in Fig. 1_a3, could differentiate between active ligands and decoys. A similar trend was observed for conformations using MMGBSA (Fig. 1_b1 to b3); however, results from g_mmpbsa (Fig. 1_c1 to c3) exhibited less separation of actives and decoys after minimization but satisfactory results were obtained using BEAR algorithm refinement. Further comparisons of different post-processing methods reiterate maximum accuracy of 99.6% and 97.6% after BEAR refinement and energy minimization using MMPBSA method of energy calculation. Figure 2 exhibits the ligand enrichment curves for the same data where Fig. 2a represents rescoring of the AutoDockVina generated conformations with all three methods yielding reduced accuracy indicating that the obtained binding poses may not be reliable for all compounds. Surprisingly, the ligand enrichment on single energy minimized complexes (Fig. 2b) with MMPBSA and MMGBSA methods also exhibited excellent outcomes comparable to those of BEAR algorithm enrichment (Fig. 2c). Virtual screening and binding energy estimation on refined complexes using AutoDockVina and BEAR algorithm resulted in improved enrichment compared to the previous study (Rastelli et al. 2009) and turns out to be an efficient tool for post-processing. Additionally, post-processing using energy minimization is far less time-consuming than the complete BEAR protocol resulting comparable enrichment. System preparation and BEAR algorithm refinement for randomly selected five ligands took 132 min which is 42-fold more time-consuming as compared to energy minimization which took 3.08 min. Of all the three methods for binding free energy estimation, MMPBSA and MMGBSA gave the best ligand enrichment with very less computation time of 2.6 and 0.44 min and can be further optimized by running longer simulations. On the other hand, “g_mmpbsa” took maximum time for Gromacs compatible system preparation and binding free energy estimation yielding satisfactorily ligand enrichment with BEAR implementation (Fig. 2c) and hence can be used as an open-source tool for post-processing. WR99210 is an antimalarial antifolate co-crystallized in the active-site of PfDHFR structure PDB-1J3I and its chemical sketch is shown in (Fig. 3a). WR99210 is docked in the active cleft of the modeled PfDHFR structure with

Fig. 1 Binding free energies of screened active ligands and decoys from DUD database are illustrated with red and black dots, respectively. **a–c** Presents binding energies obtained with MMPBSA,

MMGBSA and *g_mmpbsa* for (1) AutoDockVina generated conformations, (2) Energy minimized conformations and (3) refined conformations using BEAR algorithm, respectively

AutoDockVina and scored -8.7 Kcal/mol of binding energy. The binding free energy values using MMPBSA and MMGBSA post-minimization are -5.61 and -54.35 Kcal/mol, respectively, which are comparable to those obtained using BEAR refinement (-6.54 and -55.47 Kcal/mol). The visual assessment of the orientation of the inhibitor WR99210 as co-crystal, post-docking and post-refinement is shown in Fig. 3b–e. In crystal structure, the carboxylate oxygen atoms of Asp54 are hydrogen-bonded to N3 and 4-amino nitrogen atoms (Yuvaniyama et al. 2003). The docked conformation (C) fails to form bidentate hydrogen bonds with Asp54 where as minimized and refined complexes (D&E) very well established the interactions with the target active site residues such as Ile14, Cys15, Asp54 and Ile164. Table 1 depicts the important hydrogen bonds between WR99210 and receptor for different conformations and their respective distances (Å). Hence, refinement turned out to be an essential step prior to binding free energy estimation. The

post-minimized orientation of the inhibitor is identical to post-refined using BEAR algorithm, which proved minimization itself is sufficient to produce accurate results comparable to BEAR algorithm in 42-fold less time.

Conclusion

This study focuses on the performance of virtual screening on DUD dataset of PfDHFR using AutoDockVina followed by various post-processing approaches. The results conclusively demonstrate that post-docking optimization with energy minimization significantly reduces computational time and is comparable to BEAR refinement in accuracy, suggesting it can be successfully applied as a tool for post-processing studies in in-silico drug discovery. “*g_mmpbsa*”, implemented in Gromacs also gave satisfactory results and

Fig. 2 Enrichment curves show the percentage of the ranked database plotted against the percentage of active ligands found at any given percentage of ranked database. The plot compares the performance of the original AutoDockVina scores (green) with those obtained with MMPBSA (blue), MMGBSA (red) and *g_mmpbsa* (orange) rescoring on **a** AutoDockVina generated conformations, **b** energy minimized conformations and **c** refined conformations using BEAR algorithm

Fig. 3 **a** Chemical sketch of the ligand WR99210, **b** co-crystallized ligand WR99210 in PDB-1J3I and its native contacts with active site residues, **c** ligand WR99210 in AutoDockVina docked conformation, **d** in energy minimized conformation and **e** and in refined conformation using BEAR algorithm. WR99210 is shown in yellow stick representation, active-site residues in green stick representation and black dashed line represents the Hydrogen bonds

Table 1 Hydrogen bonds between WR99210 and PfDHFR for different conformations with their respective distances (Å)

S.N	Atoms from WR99210 and PfDHFR	Distances in (Å)			
		PDB-1J3I	Auto-DockVina conformation	Energy minimized conformation	BEAR refined conformation
1	WR99210 (NH2)–(O) I14	3.2	–	2.5	2.6
2	WR99210 (NH1)–(O) C15	3.2	1.9	2.0	1.9
3	WR99210 (N3)–(OD2) D54	2.8	3.2	3.4	3.4
4	WR99210 (NH1)–(OD1) D54	3.2	–	2.1	2.1
5	WR99210 (NH2)–(O) I164	3.1	2.2	2.4	2.3
6	WR99210 (NH2)–(O)Y170	3.3	2.5	2.8	2.8

has the added advantage of being freely available open source tool for binding free energy calculations.

Acknowledgements SG wishes to acknowledge the research grant from the Department of Science and Technology (DST), Govt. of India for support through WOS-A program (SR/WOS-A/LS-1442/2015). Special thanks to Gunjan Dixit for helping in organizing the data for the manuscript. We are grateful to Dr. Sudha Srivastava for critically evaluating the manuscript and helping with valuable suggestions.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Åqvist J, Medina C, Samuelsson JE (1994) A new method for predicting binding affinity in computer-aided drug design. *Protein Eng Des Sel* 7(3):385–391
- Case DA, Cerutti DS, Cheatham TE III, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Greene D, Homeyer N, Izadi S, Kovalenko A, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein D, Merz KM, Monard G, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Simmerling CL, Botello-Smith WM, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Xiao L, York DM, Kollman PA (2017) AMBER 2017. University of California, San Francisco
- da Silva AW, Vranken WF (2012) ACPYPE-antechamber python parser interface. *BMC Res Notes* 5(1):367
- de Ruyck J, Wouters J (2008) Structure-based drug design targeting biosynthesis of isoprenoids: a crystallographic state of the art of the involved enzymes. *Curr Protein Pept Sci* 9(2):117–137
- DeLano WL (2002) The PyMOL user's manual. DeLano Scientific, San Carlos, 452
- Hiroaki Y (2015) Package “enrichvs”
- Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. *J Med Chem* 49(23):6789–6801
- Huey R, Morris GM, Olson AJ, Goodsell DS (2007) A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 28(6):1145–1152
- Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* 23(16):1623–1641
- Kollman P (1993) Free energy calculations: applications to chemical and biochemical phenomena. *Chem Rev* 93(7):2395–2417
- Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33(12):889–897
- Kumari R, Kumar R, Consortium OSDD, Lynn A (2014) g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* 54(7):1951–1962
- Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM (2003) Structure validation by C alpha geometry: phi, psi and C beta deviation. *Proteins* 50:437–450
- Lybrand TP, McCammon JA, Wipff G (1986) Theoretical calculation of relative binding affinity in host-guest systems. *Proc Natl Acad Sci* 83 (4):833–835
- Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29(1):291–325
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 30(16):2785–2791
- Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854
- Rastelli G, Degliesposti G, Del Rio A, Sgobba M (2009) Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening. *Chem Biol Drug Des* 73(3):283–286
- Rastelli G, Rio AD, Degliesposti G, Sgobba M (2010) Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J Comput Chem* 31(4):797–810
- Ryckaert JP, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J Comput Phys* 23(3):327–341
- Šali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779–815
- Sharma K (2017) A review on *Plasmodium falciparum*-protein farnesyltransferase inhibitors as antimalarial drug targets. *Curr Drug Targets* 18(14):1676–1686
- Team RC (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0
- Trott O, Olson AJ (2010) AutoDockVina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31(2):455–461

- Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25(9):1157–1174
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35(suppl_2):W407–W410
- Yuvaniyama J, Chitnumsub P, Kamchonwongpaisan S, Vanichtanan-kul J, Sirawaraporn W, Taylor P, Walkinshaw MD, Yuthavong Y (2003) Insights into antifolate resistance from malarial DHFR-TS structures. *Nat Struct Mol Biol* 10(5):357