JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim

Application Note

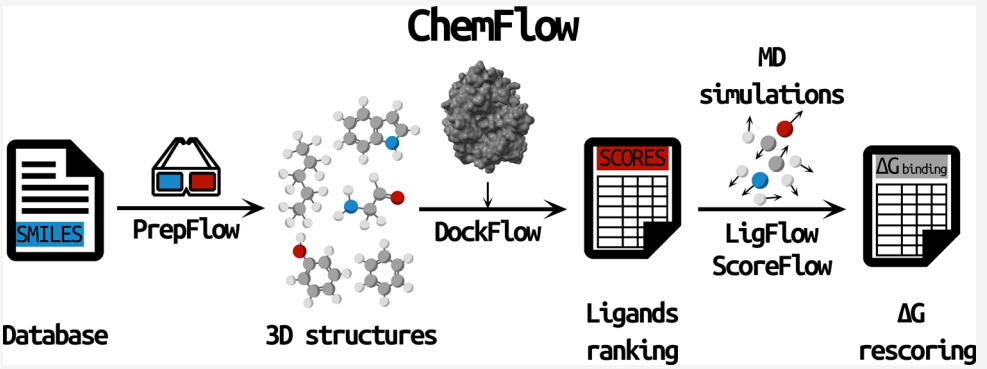
ChemFlow—From 2D Chemical Libraries to Protein−Ligand Binding Free Energies

Diego E. Barreto Gomes, Katia Galentino, Marion Sisquellas, Luca Monari, Cédric Bouysset, and Marco Cecchini*

Cite This: *J. Chem. Inf. Model.* 2023, 63, 407−411

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

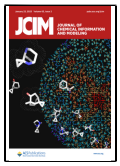


ABSTRACT: The accurate prediction of protein−ligand binding affinities is a fundamental problem for the rational design of new drug entities. Current computational approaches are either too expensive or inaccurate to be effectively used in virtual high-throughput screening campaigns. In addition, the most sophisticated methods, e.g., those based on configurational sampling by molecular dynamics, require significant pre- and postprocessing to provide a final ranking, which hinders straightforward applications by nonexpert users. We present a novel computational platform named ChemFlow to bridge the gap between 2D chemical libraries and estimated protein−ligand binding affinities. The software is designed to prepare a library of compounds provided in SMILES or SDF format, dock them into the protein binding site, and rescore the poses by simplified free energy calculations. Using a data set of 626 protein−ligand complexes and GPU computing, we demonstrate that ChemFlow provides relative binding free energies with an RMSE < 2 kcal/mol at a rate of 1000 ligands per day on a midsize computer cluster. The software is publicly available at https://github.com/IFMlab/ChemFlow.

INTRODUCTION

The accurate prediction of protein−ligand binding affinities in solution is a central problem in chemistry and biology with technologically relevant applications ranging from rational drug design to chemical sensing.[1−3] Despite vigorous effort on methodological development[4] and community-driven actions such as D3R[5] and SAMPL,[6] this problem has remained unsolved. Virtual high-throughput screening (vHTS) is the computational technique to identify drug candidates for a given protein target, typically a receptor or an enzyme,[7] which has become an integral part of the drug discovery pipeline in the pharmaceutical industry.[8] In structure-based approaches, starting with a large library of chemical compounds (typically 5−10 million), often in 2D electronic format, vHTS relies on ligand preparation (i.e., 3D reconstruction, protonation/tautomeric state assignment, partial charge determination, etc.), modeling of the protein−ligand complex, and compound ranking and prioritization. Given the number of molecules to be processed, structure-based vHTS of ultralarge chemical libraries is challenging and requires efficient and fully automated approaches. Recently, the preparation and docking of a chemical library of 1.4 billion commercially available compounds by the open-source software VirtualFlow was reported[9] (also see ref 10 for a recent review on parallelization algorithms for virtual screening). Moreover, the computational cost underlying virtual screening imposes severe limitations on the affordable level of theory, which results in fast but inaccurate scoring functions for ranking. In fact, rigorous binding free energy calculations such as double decoupling,[11] the geometrical route,[12] and more recently nonequilibrium simulations[13] are computationally too intensive and cumbersome to set up for large-scale library screening. Recently, an

Received: July 20, 2022
Published: January 5, 2023



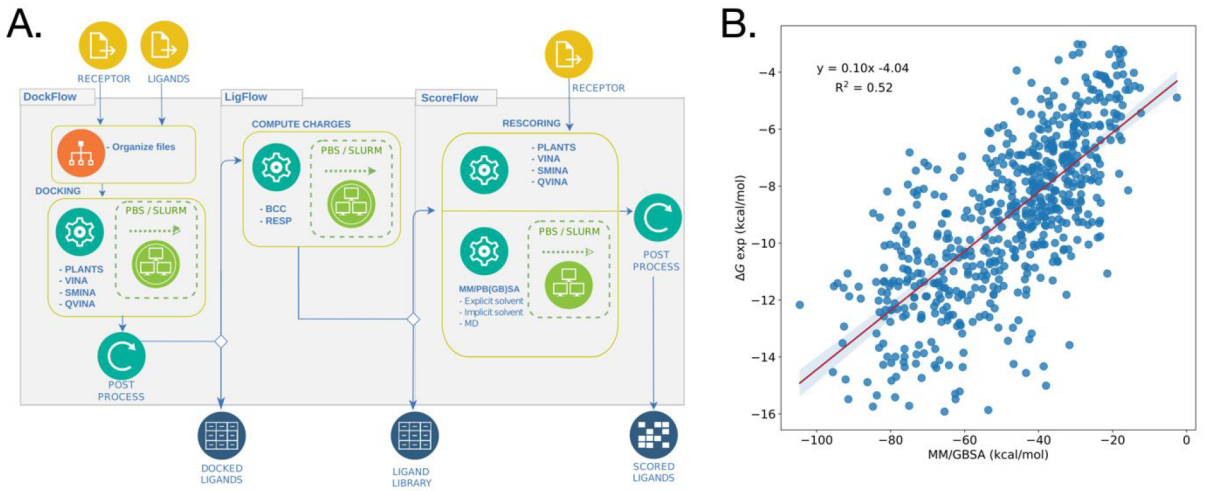


Figure 1. (A) Diagram of ChemFlow features within a typical workflow. (B) Linear regression and determination coefficient (r^2 = 0.52) for 626 protein−ligand complexes extracted from the Greenidge data set. All data points correspond to MM/GBSA free energy scores (with no entropy) based on 20 ns explicit-water MD of the protein−ligand complex (see the SI for details).

original rescoring based on the stability of the docking poses in all-atom molecular dynamics simulations was shown to improve distinguishing active compounds from decoys.[14] Also, a recent classification of existing computational strategies for protein−ligand binding has provided theoretical guidelines to tune the balance between accuracy and efficiency.[15] A significant outcome of this analysis was that empirical scoring, which is used to rank the docking poses, can be theoretically improved via end-point free energy approaches such as MM-PBSA[16] or LIE[17] that explicitly account for configurational sampling and ligand desolvation upon binding. Although computationally more intensive, "free energy rescoring" of docking results is becoming increasingly more accessible,[18] particularly after the advent of commodity GPU computing.[19,20] Recently, end-point approaches were shown to provide accurate binding free energy predictions in host−guest systems as compared to isothermal titration calorimetry (ITC) results.[21,22]

IMPLEMENTATION

An efficient implementation of end-point free energy methods for virtual screening requires a significant degree of automation to prepare, dock, and rescore many ligand poses along with sufficient computer resources to run molecular dynamics (MD) simulations started from the docking results. These intrinsic difficulties have so far hindered the establishment of fully automated tools for virtual screening with free energy rescoring. Nonetheless, semiautomatic tools and workflows have started to emerge.[23−25] For this purpose, we developed ChemFlow, an integrated workflow to bridge the gap between 2D chemical structures and ligand binding free energies. By combining existing computational chemistry tools via simplified interfaces, ChemFlow prepares, docks, and rescores a large library of chemical compounds semiautomatically and efficiently by leveraging GPU computing resources within a supercomputer environment. To enable full automation of binding free energy calculations, ChemFlow makes use of template files that provide nonexpert users straightforward access to complex simulation protocols with reasonable default settings. We note, however, that protein preparation, which involves modeling of missing loops, modeling of the protonation state of titratable residues, the inclusion of crystallographic waters, and/or post-translational modifications, etc., is left to the user.

ChemFlow is shipped with three independent modules: DockFlow, LigFlow, and ScoreFlow (Figure 1). In coordination with the standalone software PrepFlow,[27] which was designed to handle chemical library preparation robustly and efficiently, hundreds of thousands of ligands are prepared, docked, prioritized, and finally reranked using simplified free energy calculations by ChemFlow. Starting from the output of PrepFlow, DockFlow performs virtual high-throughput docking against one or multiple targets using four popular docking codes: AutoDock Vina,[28] QVINA,[29] SMINA,[30] and PLANTS.[31] LigFlow provides compound parametrization by atom-typing to GAFF2[32] and partial charge assignment to AM1-BCC[33] or RESP[34] charges by quantum-mechanical calculations, which is required for running MD. Finally, ScoreFlow reranks the docking poses by the one-average MM/PBSA or MM/GBSA methods as implemented by MMPBSA.py[35] in AmberTools.[36] By making use of high-performance computer resources through the architectural standards of the SLURM[37] and PBS[38] queuing systems along with error handling and resuming capability tools, the implemented workflow opens the door to virtual high-throughput docking with free energy rescoring.

RESULTS

The use of ChemFlow is illustrated using a subset of the Greenidge dataset.[26] The resulting dataset is highly relevant for virtual screening purposes because (1) it features 626 protein−ligand complexes with drug-like organic compounds; (2) it includes 234 distinct proteins, 25 of which cocrystallized with ≥5 ligands; (3) it was curated by experts from the pharmaceutical industry; (4) it provides high-resolution structures (X-ray) and thermodynamic data (pK_d) for all complexes; and (5) it comes with in silico predictions for benchmarking computational methods (see Tables S6 and S7 for details). To emulate a typical vHTS campaign, the ligands were extracted and processed by PrepFlow.[27] The chemical library was prepared using GAFF2 parameters and RESP charges and forwarded to DockFlow, which prioritized 10

https://doi.org/10.1021/acs.jcim.2c00919
J. Chem. Inf. Model. 2023, 63, 407−411

binding modes per ligand via PLANTS and the ChemPLP scoring function.[39] The best binding pose per ligand along with the coordinates of the protein were postprocessed by ScoreFlow. For this purpose, explicit-water MD for 20 ns was performed and analyzed using the one-average MM-GBSA method. For the MD simulations, the Amberff14SB[40] (explicit-solvent) and Amber99SB-ILDN[41] (implicit-solvent) force fields were used for the protein, GAFF2[32] was used for the ligand, and the TIP3P model[42] was used for water. For the MM/GBSA calculations, the Amber GB2 model[43] that implements the OBC implicit-solvent model[44] was used with an internal dielectric constant of 4 (see the Supporting Information (SI) for details). The numerical predictions for 626 protein−ligand complexes extracted from the Greenidge dataset with no entropy contribution are compared to experimental binding affinities in Figure 1. The data show that docking with free energy rescoring provides binding affinity results that are significantly correlated with experiments with a Pearson correlation coefficient R = 0.72 (or, equivalently, a determination coefficient r^2 = 0.52), which corresponds to a root-mean-square error (RMSE) of 1.94 kcal/mol (see the SI for details on the quantification of errors). These results are consistent with those of Greenidge (R = 0.74)[45] and demonstrate the reliability of free energy rescoring provided by ChemFlow. Moreover, analysis of the correlation with the experiments shows that free energy rescoring by explicit-solvent MM/GBSA improves the accuracy of docking by 20% (Table S2).

PERFORMANCE

When running on a midsized GPU cluster, the entire protocol, including ligand preparation, docking, and free energy rescoring takes approximately 3 h per ligand. As shown in Table 1, rescoring of the docking results by MM/GBSA is clearly the bottleneck, which accounts for 88% of the total computational effort for an ordinary protein of ∼300 amino acids. When implicit-solvent MD is used for rescoring with AM1-BCC charges, the computational effort drops to 25 min per ligand (Table 1), which allows for processing a larger number of ligands and/or binding modes per ligand. It should be noted that although free energy rescoring involves a significant computational effort, these calculations are independent, so the actual performance of the protocol scales linearly with the number of available GPU nodes. In addition, explicit-solvent MD avoids the use of harmonic restraints on the protein backbone, which allows sampling of the natural fluctuations of the biomolecule, including the plasticity of the binding site. Since the implicit-solvent results are strongly correlated with explicit-solvent MM/GBSA, unlike docking (Figure S10), we recommend the use of implicit-solvent simulations for rescoring thousands of docking poses followed by explicit-water simulations (which are 1 order of magnitude more expensive than implicit-water ones) of the top 100/200 compounds for virtual screening purposes.

Table 1. Benchmark Results: Average Execution Time Per Protocol[a]

module	protocol	timing
DockFlow	PLANTS	37 s
	AutoDock Vina	267 s
	QVINA	59 s
	SMINA	154 s
LigFlow	AM1-BCC	6 min
	RESP	19 min
ScoreFlow	minimization	20 s
	MD, implicit solvent	19 min
	MD, explicit solvent	2.6 h

[a]Details on the performances of DockFlow, LigFlow, and ScoreFlow are given in Table S4, Figure S5, and Table S5, respectively.

CONCLUSION

The newly introduced workflow for protein−ligand binding named ChemFlow opens the door to virtual screening of large chemical libraries by docking with free energy rescoring. Although the benefits of such rescoring remain to be fully established, ChemFlow provides a ready-to-use working environment for testing and benchmarking computational strategies and protocols for drug discovery. Moreover, the protocols implemented by ChemFlow can be easily augmented by coupling to recent developments that account for the entropy loss upon ligand binding, such as the interaction entropy method[46] or the quasi-harmonic multibasin (QHMB) approach.[47] With the constant increase in computer power, particularly GPU-based computing, automated virtual screening procedures like the one presented here, which explore molecular recognition in solution or even in lipid membranes, are expected to flourish and eventually replace classical high-throughput docking approaches.

DATA AND SOFTWARE AVAILABILITY

The ChemFlow software is freely distributed at the following link: https://github.com/IFMlab/ChemFlow. A tutorial on the installation and use of ChemFlow for protein−ligand binding affinities with α-thrombin is available at https://github.com/IFMlab/ChemFlow/blob/master/docs/Tutorial-ChemFlow.rst. ChemFlow interfaces several programs whose availability is indicated below. Conda is open-source and free of charge for not-for-profit institutions. The docking software Autodock Vina, QVINA, and SMINA are open-source and free of charge. The docking program PLANTS is available to academic not-for-profit users under a specific license agreement. The Open Babel tool is open-source and free of charge. For MD simulations and MM/PB(GB)SA calculations, the AmberTools suite is free of charge, and its components are mostly released under the GNU General Public License (GPL). For running MD on GPUs (Amber)[36] and RESP-charge ligand parametrizations (Gaussian), the required software is commercial and requires paid licenses. The Maestro molecular modeling program package, which was used to prepare the proteins of the data set, is commercial and requires a paid license. The dataset used in this study is available at https://github.com/IFMlab/ChemFlow/blob/master/DATA-BENCHMARK.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00919.

Details on the collection of the data set; preparation of the proteins, ligands, and protein−ligand complexes; docking experiments; MM/GBSA calculations; quantification of errors from experiments; and computational

https://doi.org/10.1021/acs.jcim.2c00919
J. Chem. Inf. Model. 2023, 63, 407−411

performances of the various steps of the workflow (PDF)
SMILES strings (TXT)

AUTHOR INFORMATION

Corresponding Author

Marco Cecchini − *Institut de Chimie de Strasbourg, UMR7177, CNRS, Université de Strasbourg, Strasbourg Cedex 67083, France;* orcid.org/0000-0003-2671-1583; Email: mcecchini@unistra.fr

Authors

Diego E. Barreto Gomes − *Institut de Chimie de Strasbourg, UMR7177, CNRS, Université de Strasbourg, Strasbourg Cedex 67083, France; Department of Physics, Auburn University, Auburn, Alabama 36849, United States*

Katia Galentino − *Institut de Chimie de Strasbourg, UMR7177, CNRS, Université de Strasbourg, Strasbourg Cedex 67083, France;* orcid.org/0000-0001-6114-4006

Marion Sisquellas − *Institut de Chimie de Strasbourg, UMR7177, CNRS, Université de Strasbourg, Strasbourg Cedex 67083, France*

Luca Monari − *Institut de Chimie de Strasbourg, UMR7177, CNRS, Université de Strasbourg, Strasbourg Cedex 67083, France*

Cédric Bouysset − *Institut de Chimie de Strasbourg, UMR7177, CNRS, Université de Strasbourg, Strasbourg Cedex 67083, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c00919

Author Contributions

The manuscript was written through contributions of all authors. All of the authors approved the final version of the manuscript.

Funding

This work was supported by the Fondation pour la Recherche Médicale (Grant DBI20141231319) and CAPES (Edital Biologia Computacional 051/2013). This project received funding from the European Union's Horizon 2020 Research and Innovation Programme under Marie Skłodowska-Curie Grant Agreement 956314 [ALLODD] and the Specific Grant Agreement No. 945539 (Human Brain Project SGA3).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to Donatienne De Francquen for an early implementation of the code. Computational resources and support at the high-performance computing center (Mesocentre) of the University of Strasbourg are gratefully acknowledged.

ABBREVIATIONS

MM/PBSA, molecular mechanics/Poisson−Boltzmann surface area; MM/GBSA, molecular mechanics/generalized Born surface area; LIE, linear interaction energy

REFERENCES

(1) Ariga, K.; Ito, H.; Hill, J. P.; Tsukube, H. Molecular Recognition: From Solution Science to Nano/Materials Technology. *Chem. Soc. Rev.* **2012**, *41* (17), 5800.

(2) Persch, E.; Dumele, O.; Diederich, F. Molecular Recognition in Chemical and Biological Systems. *Angew. Chem., Int. Ed.* **2015**, *54* (11), 3290−3327.

(3) Baron, R.; McCammon, J. A. Molecular Recognition and Ligand Association. *Annu. Rev. Phys. Chem.* **2013**, *64* (1), 151−175.

(4) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous Free Energy Simulations in Virtual Screening. *J. Chem. Inf. Model.* **2020**, *60* (9), 4153−4169.

(5) Gaieb, Z.; Parks, C. D.; Chiu, M.; Yang, H. W.; Shao, C. H.; Walters, W. P.; Lambert, M. H.; Nevins, N.; Bembenek, S. D.; Ameriks, M. K.; Mirzadegan, T.; Burley, S. K.; Amaro, R. E.; Gilson, M. K. D3R Grand Challenge 3: Blind Prediction of Protein-Ligand Poses and Affinity Rankings. *J. Comput.-Aided Mol. Des.* **2019**, *33* (1), 1−18.

(6) Grosjean, H.; Işık, M.; Aimon, A.; Mobley, D.; Chodera, J.; von Delft, F.; Biggin, P. C. SAMPL7 Protein-Ligand Challenge: A Community-Wide Evaluation of Computational Methods against Fragment Screening and Pose-Prediction. *J. Comput.-Aided Mol. Des.* **2022**, *36* (4), 291−311.

(7) Rester, U. From Virtuality to Reality—Virtual Screening in Lead Discovery and Lead Optimization: A Medicinal Chemistry Perspective. *Curr. Opin. Drug Discovery Dev.* **2008**, *11* (4), 559−568.

(8) Lyu, J.; Wang, S.; Balius, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Algaa, E.; Tolmachova, K.; Tolmachev, A. A.; Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Ultra-Large Library Docking for Discovering New Chemotypes. *Nature* **2019**, *566* (7743), 224−229.

(9) Gorgulla, C.; Boeszoermenyi, A.; Wang, Z.-F.; Fischer, P. D.; Coote, P. W.; Padmanabha Das, K. M.; Malets, Y. S.; Radchenko, D. S.; Moroz, Y. S.; Scott, D. A.; Fackeldey, K.; Hoffmann, M.; Iavniuk, I.; Wagner, G.; Arthanari, H. An Open-Source Drug Discovery Platform Enables Ultra-Large Virtual Screens. *Nature* **2020**, *580* (7805), 663−668.

(10) Murugan, N. A.; Podobas, A.; Gadioli, D.; Vitali, E.; Palermo, G.; Markidis, S. A Review on Parallel Virtual Screening Softwares for High-Performance Computers. *Pharmaceuticals* **2022**, *15* (1), 63.

(11) Wang, J. Y.; Deng, Y. Q.; Roux, B. Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophys. J.* **2006**, *91* (8), 2798−2814.

(12) Woo, H. J.; Roux, B. Calculation of Absolute Protein-Ligand Binding Free Energy from Computer Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6825−6830.

(13) Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; van Vlijmen, H.; Tresadern, G.; de Groot, B. L. Large Scale Relative Protein Ligand Binding Affinities Using Non-Equilibrium Alchemy. *Chem. Sci.* **2020**, *11* (4), 1140−1152.

(14) Guterres, H.; Im, W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, *60* (4), 2189−2198.

(15) Montalvo-Acosta, J. J.; Cecchini, M. Computational Approaches to the Chemical Equilibrium Constant in Protein-Ligand Binding. *Mol. Inf.* **2016**, *35* (11−12), 555−567.

(16) 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., III. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33* (12), 889−897.

(17) Aqvist, J.; Medina, C.; Samuelsson, J. E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7* (3), 385−391.

(18) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019**, *119* (16), 9478−9508.

(19) Lee, T. S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in

Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58* (10), 2043−2050.
(20) Mermelstein, D. J.; Lin, C.; Nelson, G.; Kretsch, R.; McCammon, J. A.; Walker, R. C. Fast and Flexible Gpu Accelerated Binding Free Energy Calculations within the Amber Molecular Dynamics Package. *J. Comput. Chem.* **2018**, *39* (19), 1354−1358.
(21) Montalvo-Acosta, J. J.; Pacak, P.; Gomes, D. E. B.; Cecchini, M. A Linear Interaction Energy Model for Cavitand Host-Guest Binding Affinities. *J. Phys. Chem. B* **2018**, *122* (26), 6810−6814.
(22) Suarez, D.; Diaz, N. Affinity Calculations of Cyclodextrin Host-Guest Complexes: Assessment of Strengths and Weaknesses of End-Point Free Energy Methods. *J. Chem. Inf. Model.* **2019**, *59* (1), 421−440.
(23) Homeyer, N.; Gohlke, H. FEW: A Workflow Tool for Free Energy Calculations of Ligand Binding. *J. Comput. Chem.* **2013**, *34* (11), 965−973.
(24) 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*, 43.
(25) Guterres, H.; Park, S.; Zhang, H.; Perone, T.; Kim, J.; Im, W. CHARMM-GUI *High-Throughput Simulator* for Efficient Evaluation of Protein−Ligand Interactions with Different Force Fields. *Protein Sci.* **2022**, *31* (9), e4413.
(26) Greenidge, P. A.; Kramer, C.; Mozziconacci, J.-C.; Wolf, R. M. MM/GBSA Binding Energy Prediction on the PDBbind Data Set: Successes, Failures, and Directions for Further Improvement. *J. Chem. Inf. Model.* **2013**, *53* (1), 201−209.
(27) Sisquellas, M.; Cecchini, M. PrepFlow: A Toolkit for Chemical Library Preparation and Management for Virtual Screening. *Mol. Inf.* **2021**, *40* (12), 2100139.
(28) 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* (2), 455−461.
(29) Alhossary, A.; Handoko, S. D.; Mu, Y.; Kwoh, C.-K. Fast, Accurate, and Reliable Molecular Docking with QuickVina 2. *Bioinformatics* **2015**, *31* (13), 2214−2216.
(30) Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013**, *53* (8), 1893−1904.
(31) Korb, O.; Stützle, T.; Exner, T. E. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In *ANTS 2006: Ant Colony Optimization and Swarm Intelligence*; Lecture Notes in Computer Science, Vol. *4150*; Springer, 2006; pp 247−258
(32) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247−260.
(33) 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* (16), 1623−1641.
(34) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269−10280.
(35) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314−3321.
(36) Case, D. A.; Ben-Shalom, I. Y.; Brozell, D. S.; Cerutti, T. E.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omeylan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J. M.; Walker, R. C.; Wang, J. M.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 22*; University of California: San Francisco, CA, 2022.
(37) Yoo, A. B.; Jette, M. A.; Grondona, M. SLURM: Simple Linux Utility for Resource Management.. In *JSSPP 2003: Job Scheduling Strategies for Parallel Processing*; Lecture Notes in Computer Science, Vol. *2862*; Springer, 2003; pp 44−60..
(38) Henderson, R. L. Job Scheduling under the Portable Batch System. In *JSSPP 1995: Job Scheduling Strategies for Parallel Processing*; Lecture Notes in Computer Science, Vol. *949*; Springer, 1995; pp 279−294. DOI: 10.1007/3-540-60153-8_34.
(39) Korb, O.; Stützle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49* (1), 84−96.
(40) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696−3713.
(41) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field: Improved Protein Side-Chain Potentials. *Proteins: Struct., Funct., Bioinf.* **2010**, *78* (8), 1950−1958.
(42) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(43) *Amber 2022 Reference Manual*, 2022.
(44) 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* (2), 383−394.
(45) Greenidge, P. A.; Lewis, R. A.; Ertl, P. Boosting Pose Ranking Performance via Rescoring with MM-GBSA. *Chem. Biol. Drug Des.* **2016**, *88* (3), 317−328.
(46) Duan, L.; Liu, X.; Zhang, J. Z. H. Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein−Ligand Binding Free Energy. *J. Am. Chem. Soc.* **2016**, *138* (17), 5722−5728.
(47) Pereira, G. P.; Cecchini, M. Multibasin Quasi-Harmonic Approach for the Calculation of the Configurational Entropy of Small Molecules in Solution. *J. Chem. Theory Comput.* **2021**, *17* (2), 1133−1142.

https://doi.org/10.1021/acs.jcim.2c00919
J. Chem. Inf. Model. 2023, 63, 407−411