

Enhancing Protein–Ligand Binding Affinity Predictions Using Neural Network Potentials

Francesc Sabanés Zariquiey, Raimondas Galvelis, Emilio Gallicchio, John D. Chodera, Thomas E. Markland, and Gianni De Fabritiis*



Cite This: *J. Chem. Inf. Model.* 2024, 64, 1481–1485



Read Online

ACCESS |



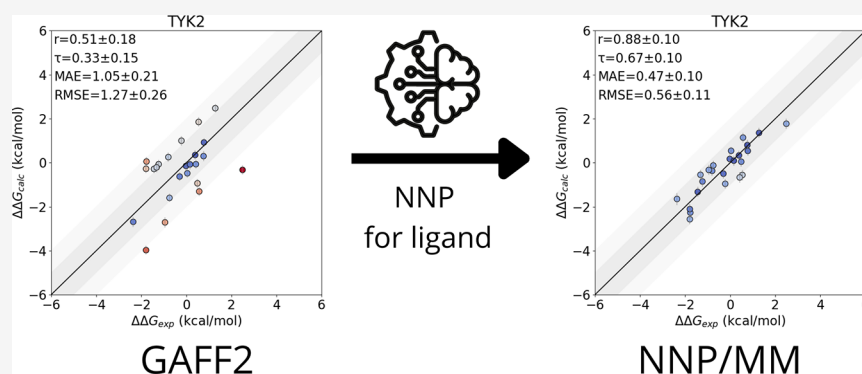
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: This letter gives results on improving protein–ligand binding affinity predictions based on molecular dynamics simulations using machine learning potentials with a hybrid neural network potential and molecular mechanics methodology (NNP/MM). We compute relative binding free energies with the Alchemical Transfer Method and validate its performance against established benchmarks and find significant enhancements compared with conventional MM force fields like GAFF2.

1. INTRODUCTION

In modern drug discovery, alchemical free energy calculations have emerged as highly efficient tools. Relative binding free energy calculations are widely employed in hit-to-lead approaches, and several commercial and free tools with comparable performance have been developed over the years. However, the accuracy of binding free energy calculations is influenced by the choice of the ligand force field. Most conventional force fields like GAFF,^{1,2} GenFF,^{3,4} and OPLS⁵ often rely on fixed charge molecular mechanics (MM). This lack of important energetic contributions limits their chemical accuracy and leads to poor modeling of torsions.^{6–8}

To address these limitations, one approach involves using quantum mechanical (QM) levels of theory to model the ligands while treating the remaining environment with an MM force field in a hybrid potential.⁹ However, QM/MM calculations are significantly more computationally expensive than MM calculations, posing challenges for drug discovery settings, where relative binding free energies (RBFE) calculations may be required for dozens or even hundreds of ligands. Recently, neural network potentials (NNPs) have shown success in predicting QM energies with significantly reduced computational costs compared to QM methods. Notably, the ANI-2x¹⁰ model supports molecular systems comprising elements H, C, N, O, S, F, and Cl. Moreover, a

hybrid method that integrates NNPs and MM, known as NNP/MM¹¹, has been developed, offering the potential to model ligands more accurately in RBFE calculations than traditional MM force fields. The Alchemical Transfer Method (ATM) is a recently developed methodology for alchemical free energy calculations that we recently validated that allows an easy implementation of NNPs.¹² In previous publications, this methodology with MM force fields on a robust data set obtained similar results to other state-of-the-art methods such as FEP+.^{13,14} In this work, we exploit the capabilities of ATM to test the hybrid approach of using ANI-2x¹⁰ as the neural network potential. Rufa et al. previously managed to reduce the error of absolute binding free energies from 0.97 to 0.47 kcal/mol for a congeneric ligand series for tyrosine kinase TYK2 by correcting the conventional MM simulation with an NNP/MM approach.¹⁵ ANI-2x has several limitations in terms of not supporting charged molecules and certain elements, but it is otherwise a useful test potential. Our main objective is to test

Received: December 18, 2023

Revised: February 14, 2024

Accepted: February 14, 2024

Published: February 20, 2024



Table 1. Comparison of Performance of Different Force Fields and NNP/MM^a.

	GAFF2			NNP/MM ANI2x			FEP+			ligand pairs
	Kendall (τ)	MAE	RMSE	Kendall (τ)	MAE	RMSE	Kendall (τ)	MAE	RMSE	
CDK2	0.42 \pm 0.17	0.8 \pm 0.2	1.1 \pm 0.3	0.62 \pm 0.10	0.7 \pm 0.1	0.8 \pm 0.2	0.42 \pm 0.15	0.8 \pm 0.1	1.1 \pm 0.1	22
JNK1	0.34 \pm 0.14	0.9 \pm 0.2	1.0 \pm 0.2	0.43 \pm 0.12	0.7 \pm 0.1	0.9 \pm 0.2	0.43 \pm 0.14	0.8 \pm 0.1	1.0 \pm 0.1	27
P38	0.48 \pm 0.06	1.2 \pm 0.2	1.6 \pm 0.2	0.59 \pm 0.05	0.9 \pm 0.1	1.2 \pm 0.2	0.60 \pm 0.11	0.8 \pm 0.1	1.0 \pm 0.1	56
TYK2	0.33 \pm 0.15	1.1 \pm 0.2	1.3 \pm 0.3	0.67 \pm 0.10	0.5 \pm 0.1	0.6 \pm 0.1	0.54 \pm 0.15	0.8 \pm 0.2	0.9 \pm 0.1	24
HIF2A	0.43 \pm 0.10	1.6 \pm 0.3	2.0 \pm 0.4	0.55 \pm 0.11	1.3 \pm 0.2	1.6 \pm 0.3	0.50 \pm 0.13	1.1 \pm 0.1	1.3 \pm 0.2	28
PFKFB3	0.49 \pm 0.06	1.3 \pm 0.2	1.6 \pm 0.2	0.37 \pm 0.08	1.3 \pm 0.2	1.7 \pm 0.2	0.70 \pm 0.09	1.0 \pm 0.1	1.6 \pm 0.2	62
SYK	0.25 \pm 0.12	1.3 \pm 0.2	1.6 \pm 0.3	0.38 \pm 0.10	1.0 \pm 0.2	1.3 \pm 0.2	0.16 \pm 0.11	1.2 \pm 0.1	1.5 \pm 0.2	37
TNKS2	0.37 \pm 0.09	1.0 \pm 0.2	1.2 \pm 0.2	0.45 \pm 0.10	0.9 \pm 0.1	1.1 \pm 0.2	0.41 \pm 0.10	0.8 \pm 0.1	1.0 \pm 0.1	45

^aKendall correlation (τ), mean absolute error (MAE), and root mean square error (RMSE) in kcal/mol for the eight tested Protein Targets. FEP+ is included as a state-of-the-art comparison.

the applicability of this methodology with different ligand force fields and to evaluate the feasibility of an NNP/MM approach in relative binding free energy calculations.

2. METHODS

In this study, we evaluated a series of targets from both Wang et al.'s¹⁶ and Schindler et al.'s data sets.¹⁷ Due to the limitations of ANI-2x,¹⁰ the NNP of our choice in this study, there is a series of targets from the aforementioned data sets that cannot be computed due to the properties of its ligands. Consequently, we evaluated the following targets: Cyclin-dependent kinase 2 (CDK2), c-Jun N-terminal kinase 1 (JNK1), tyrosine kinase 2 (TYK2), P38 MAP kinase (P38), hypoxia-inducible transcription factor 2 (HIF2A), PFKFB3, spleen tyrosine kinase (SYK), and tankyrase 2 (TNKS2), totaling 301 ligand pairs. For the selected targets, most of the ligands are compatible with ANI-2x, and the rest (and its corresponding ligand pair calculations) were removed from the data set. Due to the higher computational costs related to the integration of NNP into these calculations, a subset of all the possible ligand pairs to be evaluated was selected at random. The workflow in this project is similar to our previous work.¹³ Protein and ligand structures were readily available from Wang's¹⁶ and Schindler's¹⁷ data sets. Ligands were parametrized with GAFF 2.11^{1,2}. The topologies were generated using the *parameterize*¹⁸ tool. In contrast to our previous work, we now prepared complex systems using HTMD,¹⁹ which automated and streamlined the preparation of multiple ligand pairs, along with the automatic selection of binding site residues. However, manual selection of atom indexes for ligand alignment remained necessary. The energy minimization, thermalization, and equilibration steps followed the procedures described in our previous work.¹³ Additionally, the system was annealed to the symmetrical alchemical intermediate ($\lambda = 1/2$) for 250 ps. The classical RBFE simulations (GAFF2) were run in triplicate for each ligand pair running an ensemble of 60 ns per replica. Concurrently, we performed the same calculations by using an NNP/MM approach.¹¹ This hybrid method allowed us to simulate a portion of the molecular system (the small molecule) with an NNP, while the rest was simulated with MM, providing the ligands with optimized intramolecular interactions. For both approaches, we used the Amber ff14SB parameters^{20,21} as well as the TIP3P water model. Classical RBFE simulations were run at a 4 fs time step while the NNP/MM runs were computed at 1 fs time step, both with the ATM integrator plugin.²² Hamiltonian replica exchange along the λ space for each ATM leg was performed with the ASyncRE

software,²³ specially customized for OpenMM and ATM.²⁴ Consistent with our previous work, we computed the binding free energies and their corresponding uncertainties from the perturbation energy samples using the Unbinned Weighted Histogram Analysis Method (UWHAM).²⁵ The resulting relative binding free energies ($\Delta\Delta G$) were compared to experimental measurements in terms of the mean absolute error (MAE), root-mean-square error (RMSE), and Kendall Tau correlation coefficient. For all the possible systems, absolute ΔG values were computed with cinnabar, an analysis tool to compute absolute binding free energies from $\Delta\Delta G$ values via a maximum likelihood estimator.²⁶ Cinnabar also generates the correlation plots and calculates the necessary error and correlation statistics. We compared the obtained values from calculations and the works by Wang et al.¹⁶ and Schindler et al.¹⁷ with FEP+. To perform the calculations, we utilized the OpenMM-ML and NNPOps libraries on our in-house cluster, comprising NVIDIA RTX 2080 Ti and NVIDIA RTX 4090 cards. Standard MM calculations were run on GPU GRID. The parallel replica exchange molecular dynamics simulations were conducted by using the OpenMM 7.7 MD engine and the ATM Meta Force plugin, utilizing the CUDA platform.

3. RESULTS

The results of our simulations are displayed in Table 1 and Figures 1 and 2 which highlight the relative (Kendall's rank order correlation) and absolute performances (MAE and RMSE) of the evaluated methods. We do not report the Pearson correlation as well because the value is not significant for ($\Delta\Delta G$) values as it varies with the choice of the pairs.²⁷ We cannot calculate ΔG for all pairs since we ran a subset of the original data sets. The computation of ΔG for all ligands was not possible due to a poor connection of the perturbation network. Figures S4–S8 display the ΔG values and related statistics for the systems that were possible to compute. The NNP/MM method demonstrated superior performance over pure MM runs in both relative and absolute measures. We observe that NNP/MM shows a better correlation coefficient and MAE for all of the evaluated systems but PFKFB3 when compared to ATM with GAFF2 as a force field. In comparison to FEP+, NNP/MM has a lower correlation for two systems (P38 and PFKFB3) and higher MAE for four of them (P38, HIF2A, PFKFB3, and TNKS2). Furthermore, the amount of ligands that are more accurately predicted is increased. In comparison to our GAFF2 runs, MM/NNP predicts a higher percentage of ligands with a MAE lower than both 1 and 1.5

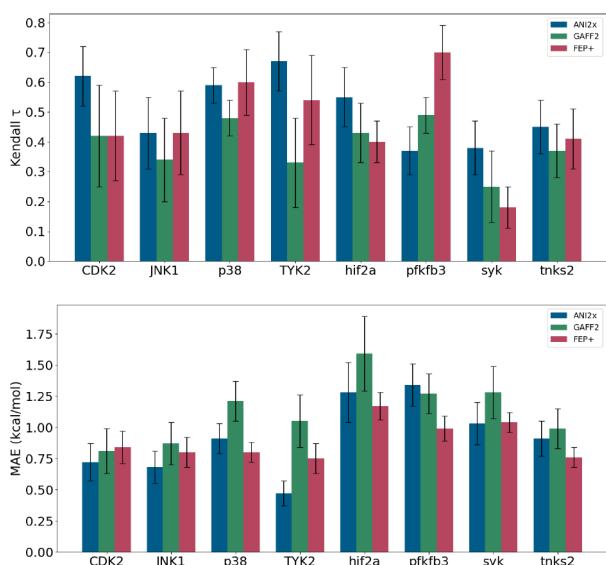


Figure 1. (Top) Kendall tau and (bottom) mean absolute error (MAE) for the $\Delta\Delta G$ s of each protein–ligand system calculated in combination with different force fields and reported estimates using FEP+. ¹⁶

kcal/mol (Table S1). Additionally, there are no major differences between the conformers generated with GAFF2 and ANI-2x as force-fields (Figure S9). We observe for some specific cases how ligands that participated in poor predictions with GAFF2 (MAE > 2 kcal/mol) are now predicted correctly (MAE < 1 kcal/mol) (Table S2). However, this improvement comes at a cost, as NNP/MM calculations are slower than conventional MM calculations. ¹¹ For instance, an RTX 4090 could yield up to 27 ns/day, whereas an ATM conventional run for a P38 system with 49,000 atoms is able to compute 211 ns/day (Figure S10). This decrease on speed mainly arises due to the limitation of a 1 fs time step with the current ATM

integrator. While there is a considerable increase in computational cost for NNP/MM runs, both approaches could benefit from further optimizations. We also evaluate if different timesteps could influence the accuracy of RBFE calculations. We compared the results of the GAFF2 calculations performed in this work with a 4 fs time step with the calculated points from our previous benchmark, that were run at a 2 fs time step (Figure S11). We do not observe any considerable accuracy difference between the calculations at both timesteps. In terms of convergence, we observed that 60 ns per calculation tends to be sufficient. Convergence analysis over time shows good convergence for most cases, as illustrated in Figure S12.

4. CONCLUSION

We conducted relative binding free energy (RBFE) calculations using an innovative NNP/MM approach. Our findings demonstrate the substantial accuracy enhancement achieved by using an NNP/MM approach at the cost of increased computational time. Compared to conventional ligand force fields like GAFF2, the NNP/MM approach exhibited reduced mean absolute errors, with most systems reaching below 1 kcal/mol. However, we acknowledge that the current NNP used in this study is limited to neutral molecules and a limited set of elements, posing a constraint on our exploration of the vast chemical space. Future endeavors should focus on expanding the applicability of NNPs to include charged ligands, thereby broadening the scope of our investigations. An increase in computing performance is also needed, probably with the inclusion of other integrators that allow for higher timesteps. Due to limited computational resources, a random subset of all the possible calculations was computed. Although a potential bias could be included due to the nature of the subset, we believe to have sampled an extensive number of data points to understand the capabilities of RBFE along with the NNP/MM approach. Our work highlights the potential of NNP/MM for accurate RBFE calculations and underscores the importance of further advancing NNPs to encompass a

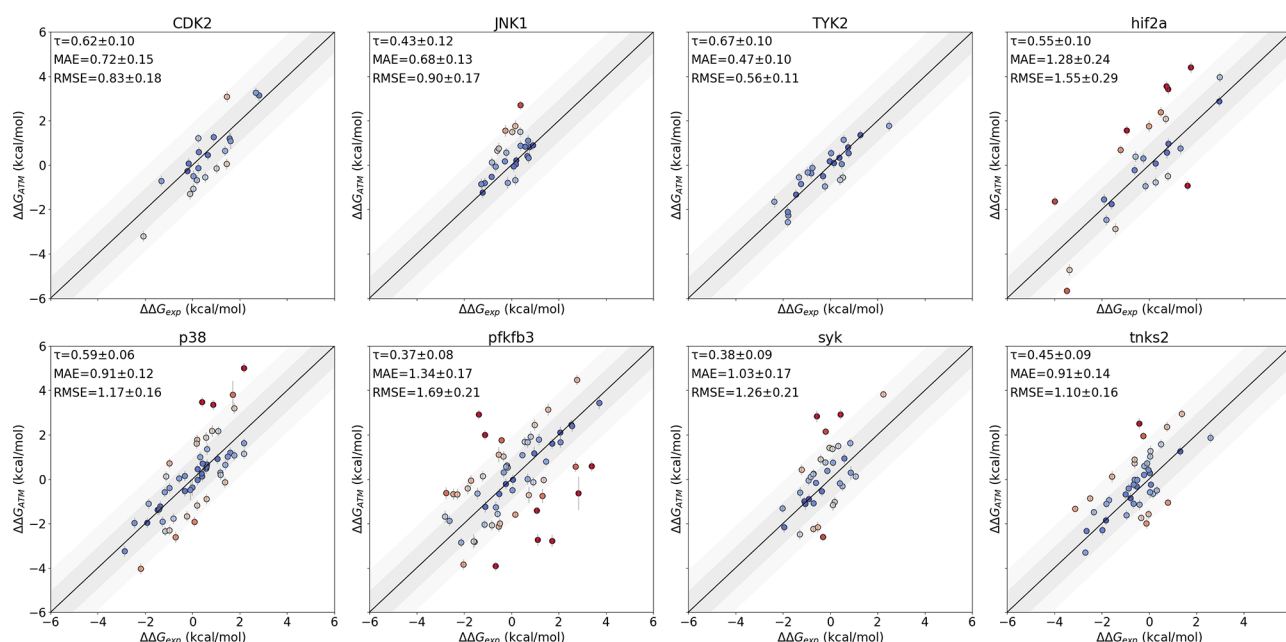


Figure 2. Performance in combination with the neural network potential (NNP) for each protein–ligand system studied. The calculated $\Delta\Delta G$ estimates are plotted against their corresponding experimental values. MAE and RMSE are in kcal/mol and τ is Kendall correlation.

broader range of molecular species and further improve the accuracy of these calculations.

■ ASSOCIATED CONTENT

Data Availability Statement

The calculated free energy values and ligand and protein structures, as well as preparation scripts, are available at https://github.com/compsciencelab/ATM_benchmark/tree/main/ATM_With_NNPs

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.3c02031>.

Description of the workflow used for this work. Barplots of the Pearson correlation and RMSE from the calculated versus experimental $\Delta\Delta G$ values for all systems and compared methods. Scatterplots for the calculated ΔG on all the connected systems and the comparison between the compared methods as well as barplots from the corresponding MAE, RMSE errors, and R2 and Spearman correlations. Barplots with the analysis of different speed performances between ATM and ATM combined with NNP/MM method. Scatterplots for a series of targets studied at different timesteps. Convergence analysis based on the estimation of the $\Delta\Delta G$ through simulated time. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Gianni De Fabritiis — Computational Science Laboratory, Universitat Pompeu Fabra, 08003 Barcelona, Spain; Acellera Labs, 08005 Barcelona, Spain; Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain; orcid.org/0000-0003-3913-4877; Email: g.defabritiis@gmail.com

Authors

Francesc Sabanés Zariquiey — Computational Science Laboratory, Universitat Pompeu Fabra, 08003 Barcelona, Spain; Acellera Labs, 08005 Barcelona, Spain; orcid.org/0000-0002-5674-4127

Raimondas Galvelis — Computational Science Laboratory, Universitat Pompeu Fabra, 08003 Barcelona, Spain; Acellera Labs, 08005 Barcelona, Spain; orcid.org/0000-0001-8431-1612

Emilio Gallicchio — Department of Chemistry, Graduate Center, Brooklyn College, City University of New York, New York, New York 11210, United States; orcid.org/0000-0002-2606-4913

John D. Chodera — Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, United States; orcid.org/0000-0003-0542-119X

Thomas E. Markland — Department of Chemistry, Stanford University, Stanford, California 94305, United States; orcid.org/0000-0002-2747-0518

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jcim.3c02031>

Notes

J.D.C. is a current member of the Scientific Advisory Board of OpenEye Scientific Software, Redesign Science, Ventus Therapeutics, and Interline Therapeutics, and has equity

interests in Redesign Science and Interline Therapeutics. The Chodera laboratory receives or has received funding from multiple sources, including the National Institutes of Health, the National Science Foundation, the Parker Institute for Cancer Immunotherapy, Relay Therapeutics, Entasis Therapeutics, Silicon Therapeutics, EMD Serono (Merck KGaA), AstraZeneca, Vir Biotechnology, Bayer, XtalPi, Interline Therapeutics, the Molecular Sciences Software Institute, the Starr Cancer Consortium, the Open Force Field Consortium, Cycle for Survival, a Louis V. Gerstner Young Investigator Award, and the Sloan Kettering Institute. A complete funding history for the Chodera lab can be found at <http://choderalab.org/funding>.

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the volunteers of [GPUGRID.net](https://www.gpugrid.net) for donating computing time. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 823712, and the project PID2020-116564GB-I00 has been funded by MCIN/AEI/10.13039/501100011033; the Torres-Quevedo Programme from the Spanish National Agency for Research (PTQ2020-011145/AEI/10.13039/501100011033). Research reported in this publication was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health under Award Number R01GM140090 to T.E.M., G.D.F., and J.D.C. E.G. acknowledges support from the United States National Science Foundation (NSF CAREER 1750511). J.D.C. also acknowledges support from NIH Grant P30CA008748 and the Sloan Kettering Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

■ REFERENCES

- (1) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **2004**, *25*, 1157–1174.
- (2) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- (3) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. Igor Mackerell Jr CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* **2010**, *31*, 671–690.
- (4) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.
- (5) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending force field coverage for drug-like small molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1863–1874.
- (6) Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Advances in protein chemistry* **2003**, *66*, 27–85.
- (7) Dauber-Osguthorpe, P.; Hagler, A. T. Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there? *Journal of computer-aided molecular design* **2019**, *33*, 133–203.
- (8) Hagler, A. T. Force field development phase II: Relaxation of physics-based criteria... or inclusion of more rigorous physics into the

representation of molecular energetics. *Journal of computer-aided molecular design* **2019**, *33*, 205–264.

(9) Beierlein, F. R.; Michel, J.; Essex, J. W. A simple QM/MM approach for capturing polarization effects in protein–ligand binding free energy calculations. *J. Phys. Chem. B* **2011**, *115*, 4911–4926.

(10) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192–4202.

(11) Galvelis, R.; Varela-Rial, A.; Doerr, S.; Fino, R.; Eastman, P.; Markland, T. E.; Chodera, J. D.; De Fabritiis, G. NNP/MM: Accelerating molecular dynamics simulations with machine learning potentials and molecular mechanics. *J. Chem. Inf. Model.* **2023**, *63*, 5701–5708.

(12) Azimi, S.; Khuttan, S.; Wu, J. Z.; Pal, R. K.; Gallicchio, E. Relative binding free energy calculations for ligands with diverse scaffolds with the alchemical transfer method. *J. Chem. Inf. Model.* **2022**, *62*, 309–323.

(13) Sabanés Zariquiey, F.; Pérez, A.; Majewski, M.; Gallicchio, E.; De Fabritiis, G. Validation of the Alchemical Transfer Method for the Estimation of Relative Binding Affinities of Molecular Series. *J. Chem. Inf. Model.* **2023**, *63*, 2438–2444.

(14) Chen, L.; Wu, Y.; Wu, C.; Silveira, A.; Sherman, W.; Xu, H.; Gallicchio, E. Performance and Analysis of the Alchemical Transfer Method for Binding Free Energy Predictions of Diverse Ligands. *J. Chem. Inf. Model.* **2024**, *64*, 250–264.

(15) Rufa, D. A.; Macdonald, H. E. B.; Fass, J.; Wieder, M.; Grinaway, P. B.; Roitberg, A. E.; Isayev, O.; Chodera, J. D. Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning/molecular mechanics potentials. *bioRxiv Preprint*, 2020, DOI: 10.1101/2020.07.29.227959.

(16) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.

(17) Schindler, C. E. M.; Baumann, H.; Blum, A.; Böse, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D.; Eguida, M. K. I.; Follows, B.; Fuchß, T.; Grädler, U.; Gunera, J.; Johnson, T.; Jorand Lebrun, C.; Karra, S.; Klein, M.; Knehans, T.; Koetzner, L.; Krier, M.; Leiendecker, M.; Leuthner, B.; Li, L.; Mochalkin, I.; Musil, D.; Neagu, C.; Rippmann, F.; Schiemann, K.; Schulz, R.; Steinbrecher, T.; Tanzer, E.-M.; Unzué Lopez, A.; Viacava Follis, A.; Wegener, A.; Kuhn, D. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *J. Chem. Inf. Model.* **2020**, *60*, 5457–5474.

(18) Galvelis, R.; Doerr, S.; Damas, J. M.; Harvey, M. J.; De Fabritiis, G. A scalable molecular force field parameterization method based on density functional theory and quantum-level machine learning. *J. Chem. Inf. Model.* **2019**, *59*, 3485–3493.

(19) Doerr, S.; Harvey, M.; Noé, F.; De Fabritiis, G. HTMD: high-throughput molecular dynamics for molecular discovery. *J. Chem. Theory Comput.* **2016**, *12*, 1845–1852.

(20) Zou, J.; Tian, C.; Simmerling, C. Blinded prediction of protein–ligand binding affinity using Amber thermodynamic integration for the 2018 D3R grand challenge 4. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 1021–1029.

(21) 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*, 3696–3713.

(22) *openmm sdm plugin*. https://github.com/Gallicchio-Lab/openmm_sdm_plugin (accessed 12–12–2023).

(23) Gallicchio, E.; Xia, J.; Flynn, W. F.; Zhang, B.; Samlalsingh, S.; Montes, A.; Levy, R. M. Asynchronous replica exchange software for

grid and heterogeneous computing. *Comput. Phys. Commun.* **2015**, *196*, 236–246.

(24) *AToM-OpenMM*. <https://github.com/Gallicchio-Lab/AToM-OpenMM> (accessed 12–12–2023).

(25) Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of binless multi-state free energy estimation with applications to protein–ligand binding. *J. Chem. Phys.* **2012**, *136*, No. 04B608.

(26) Macdonald, H. B.; Henry, M.; Chodera, J.; Dotson, D.; Glass, W.; Pulido, I. *openforcefield/openff-arsenic: v0.2.1*; Zendo, 2022, DOI: 10.5281/zenodo.6210305.

(27) Hahn, D. F.; Bayly, C. I.; Boby, M. L.; Macdonald, H. E. B.; Chodera, J. D.; Gapsys, V.; Mey, A. S.; Mobley, D. L.; Benito, L. P.; Schindler, C. E.; Tresadern, G.; Warren, G. L. Best practices for constructing, preparing, and evaluating protein–ligand binding affinity benchmarks [article v1. 0]. *Living journal of computational molecular science* **2022**, *4*, 4.