# Enhancing Protein-Ligand Binding Affinity Predictions using Neural Network Potentials

**Francesc Sabanés Zariquiey**[†,‡], **Raimondas Galvelis**[†,‡], **Emilio Gallicchio**[¶], **John D. Chodera**[§], **Thomas E. Markland**[‖], **Gianni De Fabritiis**[†,‡,⊥]

[†]Computational Science Laboratory, Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), C Dr. Aiguader 88, 08003, Barcelona, Spain

[‡]Acellera Labs, C Dr Trueta 183, 08005, Barcelona, Spain

[¶]Department of Chemistry, Brooklyn College of the City University of New York; PhD Program in Chemistry, Graduate Center of the City University of New York; PhD Program in Biochemistry, Graduate Center of the City University of New York, NY 11210, USA

[§]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

[‖]Department of Chemistry, Stanford University, 337 Campus Drive, Stanford, CA, 94305, USA

[⊥]Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluis Companys 23, 08010 Barcelona, Spain

## 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 (RBFE) with the Alchemical Transfer Method (ATM) and validate its performance against established benchmarks and find significant enhancements compared to 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 ligand force field. Most conventional force fields like GAFF[1,2],

GenFF[3,4], and OPLS[5] 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 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 cost compared to QM methods. Notably, the ANI-2x[10] 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/MM[11], 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 dataset 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-2x[10] 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 non-supporting charged molecules and certain elements but it is otherwise a useful test potential. Our main objective is to test 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's *et al.*[16] and Schindler's datasets.[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 datasets 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, the rest (and its corresponding lig- and pair calculations) were removed from the dataset. Due to the higher computational costs related to the integration of NNP into these calculations, a subset of all the possible lig-and 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's[16] and Schindler's[17] datasets. Ligands were parameterized 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,

the 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 symmetric 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 intra-molecular 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 4fs timestep while the NNP/MM runs were computed at 1fs timestep, both with the ATM integrator plugin[22]. Hamiltonian replica exchange along the $\lambda$ 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 mean absolute error (MAE), root mean square error (RMSE), and Kendall Tau correlation coefficient. For all the possible systems, absolute $\Delta G$ 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 error and correlation statistics necessary. 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 GPUGRID. The parallel replica exchange molecular dynamics simulations were conducted 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 performance (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 $\Delta G$ for all pairs since we ran a subset of the original datasets. The computation of $\Delta G$ for all ligands was not possible due to a poor connection of the perturbation network. Figures S4–S8 displays the $\Delta G$ 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 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 > 2kcal/mol) are now predicted correctly (MAE < 1kcal/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 49k atoms is able to compute 211 ns/day (Figure S10). This decrease on speed mainly arises due to the limitation of a 1fs timestep 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 4fs timestep with the calculated points from our previous benchmark, that were run at a 2fs timestep.(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 forcefields 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 a 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 broader range of molecular species, and further improve the accuracy of these calculations.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgement

**Disclosures**

## Data and software availability

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

## References

(1). Wang J; Wolf RM; Caldwell JW; Kollman PA; Case DA Development and testing of a general amber force field. Journal of computational chemistry 2004, 25, 1157–1174. [PubMed: 15116359]

(2). Wang J; Wang W; Kollman PA; Case DA 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 A, 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. [PubMed: 19575467]

(4). Vanommeslaeghe K; Raman EP; MacKerell AD Jr Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. Journal of chemical information and modeling 2012, 52, 3155–3168. [PubMed: 23145473]

(5). Roos K; Wu C; Damm W; Reboul M; Stevenson JM; Lu C; Dahlgren MK; Mondal S; Chen W; Wang L; Abel R OPLS3e: Extending force field coverage for drug-like small molecules. Journal of chemical theory and computation 2019, 15, 1863–1874. [PubMed: 30768902]

(6). Ponder JW; Case DA Force fields for protein simulations. Advances in protein chemistry 2003, 66, 27–85. [PubMed: 14631816]

(7). Dauber-Osguthorpe P; Hagler AT 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. [PubMed: 30506158]

(8). Hagler AT 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. [PubMed: 30506159]

(9). Beierlein FR; Michel J; Essex JW A simple QM/MM approach for capturing polarization effects in protein- ligand binding free energy calculations. The Journal of Physical Chemistry B 2011, 115, 4911–4926. [PubMed: 21476567]

(10). Devereux C; Smith JS; Huddleston KK; Barros K; Zubatyuk R; Isayev O; Roitberg AE Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. Journal of Chemical Theory and Computation 2020, 16, 4192–4202. [PubMed: 32543858]

(11). Galvelis R; Varela-Rial A; Doerr S; Fino R; Eastman P; Markland TE; Chodera JD; De Fabritiis G NNP/MM: Accelerating molecular dynamics simulations with machine learning potentials and molecular mechanics. Journal of chemical information and modeling 2023, 63, 5701–5708. [PubMed: 37694852]

(12). Azimi S; Khuttan S; Wu JZ; Pal RK; 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. [PubMed: 34990555]

(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. Journal of Chemical Information and Modeling 2023, 63, 2438–2444. [PubMed: 37042797]

(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. 2023.

(15). Rufa DA; Macdonald HEB; Fass J; Wieder M; Grinaway PB; Roitberg AE; Isayev O; Chodera JD Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning / molecular mechanics potentials. bioRxiv 2020,

(16). Wang L; Wu Y; Deng Y; Kim B; Pierce L; Krilov G; Lupyan D; Robinson S; Dahlgren MK; Greenwood J; Romero DL; Masse C; Knight JL; Steinbrecher T; Beuming T; Damm W; Harder E; Sherman W; Brewer M; Wester R; Murcko M; Frye L; Farid R; Lin T; Mobley DL; Jorgensen WL; Berne BJ; Friesner RA; 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. [PubMed: 25625324]

(17). Schindler CEM; Baumann H; Blum A; Böse D; Buchstaller H-P; Burgdorf L; Cappel D; Chekler E; Czodrowski P; Dorsch D; Eguida MKI; 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; Unzue Lopez A; Viacava Follis A; Wegener A; Kuhn D Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. Journal of Chemical Information and Modeling 2020, 60, 5457–5474. [PubMed: 32813975]

(18). Galvelis R; Doerr S; Damas JM; Harvey MJ; De Fabritiis G A scalable molecular force field parameterization method based on density functional theory and quantum-level machine learning. Journal of chemical information and modeling 2019, 59, 3485–3493. [PubMed: 31322877]

(19). Doerr S; Harvey M; Noé F; De Fabritiis G HTMD: high-throughput molecular dynamics for molecular discovery. Journal of chemical theory and computation 2016, 12, 1845–1852. [PubMed: 26949976]

(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. [PubMed: 31555923]

(21). Maier JA; Martinez C; Kasavajhala K; Wickstrom L; Hauser KE; Simmerling C ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theory Comput 2015, 11, 3696–3713. [PubMed: 26574453]

(22). openmm sdm plugin. https://github.com/Gallicchio-Lab/openmm_sdm_plugin, [Accessed 12-12-2023].

(23). Gallicchio E; Xia J; Flynn WF; Zhang B; Samlalsingh S; Mentes A; Levy RM Asynchronous replica exchange software for grid and heterogeneous computing. Comput. Phys. Commun 2015, 196, 236–246. [PubMed: 27103749]

(24). AToM-OpenMM. https://github.com/Gallicchio-Lab/AToM-OpenMM, [Accessed 12-12-2023].

(25). Tan Z; Gallicchio E; Lapelosa M; Levy RM Theory of binless multi-state free energy estimation with applications to protein-ligand binding. The Journal of Chemical Physics 2012, 136, 04B608.

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

(27). Hahn DF; Bayly CI; Boby ML; Macdonald HEB; Chodera JD; Gapsys V; Mey AS; Mobley DL; Benito LP; Schindler CE; Tresadern G; Warren GL Best practices for constructing, preparing,

and evaluating protein-ligand binding affinity benchmarks [article v1. 0]. Living journal of computational molecular science 2022, 4.
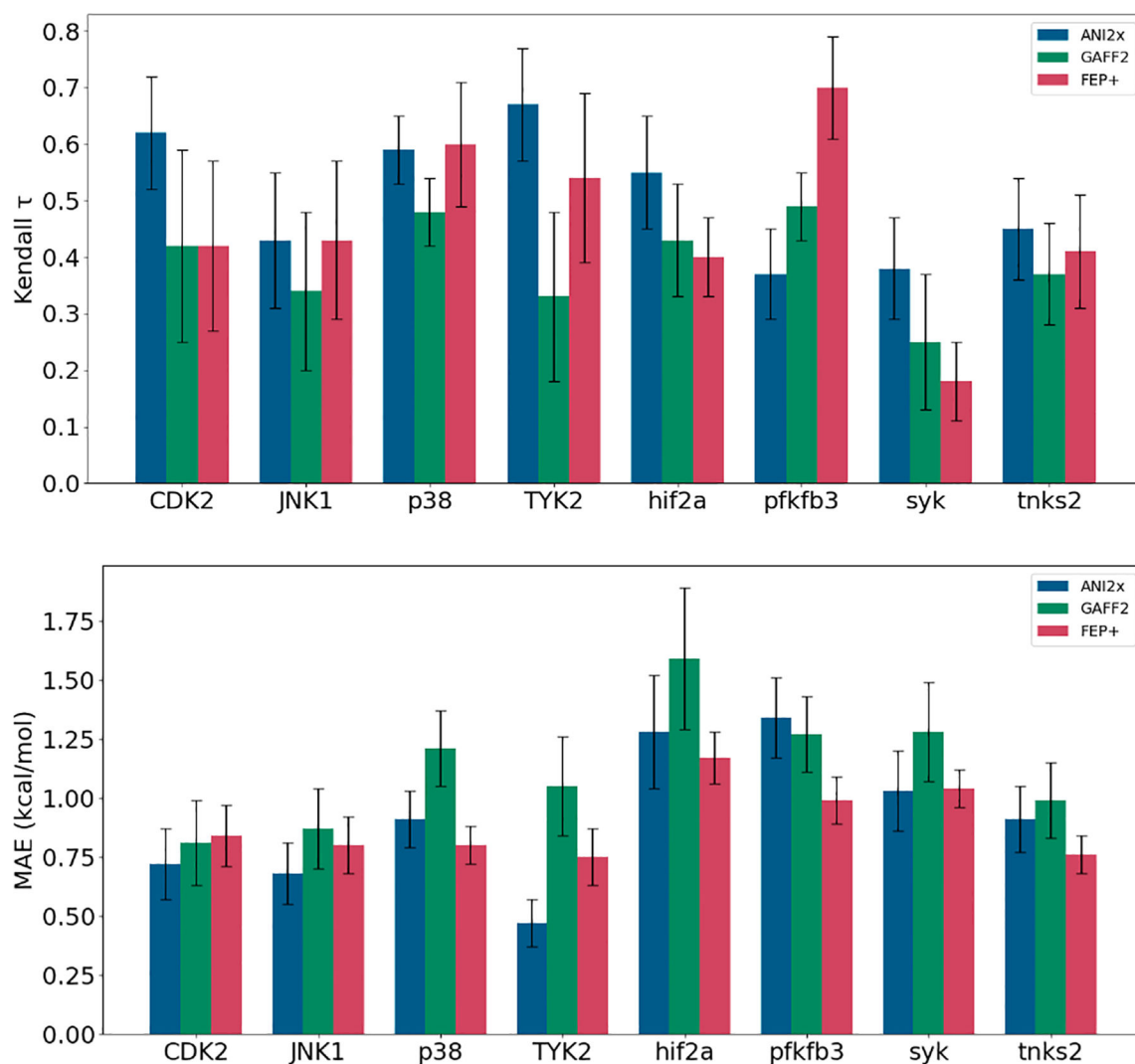
**Figure 1:**

(Top) Kendall Tau and (bottom) Mean Absolute Error (MAE) for the *G*s of each protein-ligand system calculated in combination with different force fields and reported estimates using FEP+[16]

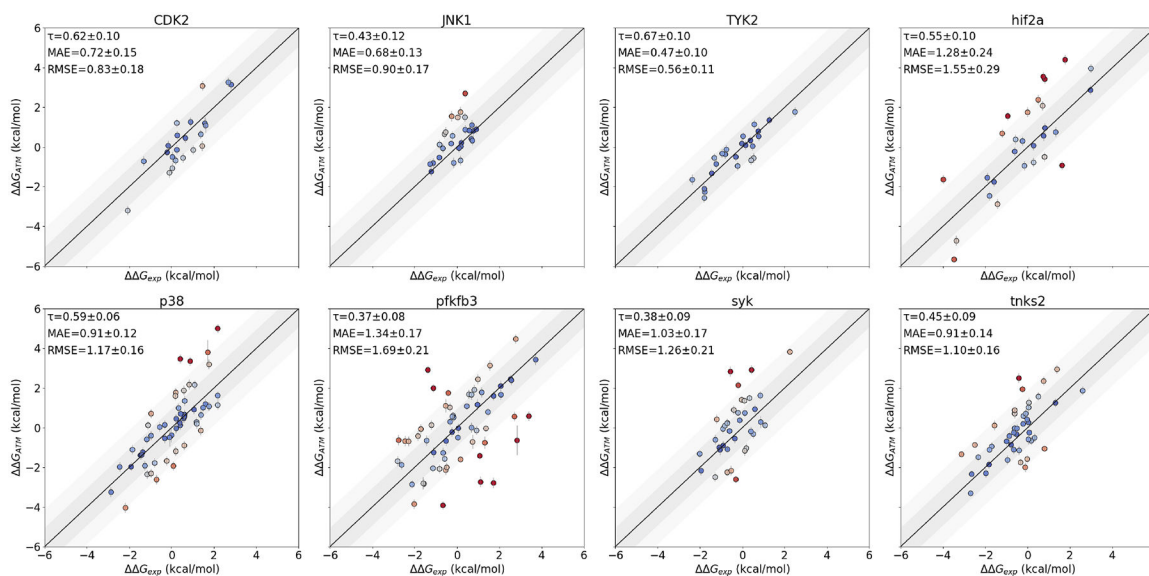**Figure 2:**
Performance in combination with the neural network potential (NNP) for each protein-ligand system studied. The calculated    G estimates are plotted against their corresponding experimental values. MAE and RMSE are in kcal/mol and $\tau$ is Kendall correlation.

**Table 1:**

Comparison of the performance of different forcefields and NNP/MM. Kendall correlation ($\tau$), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) in kcal/mol for the 8 tested Protein Targets. FEP+ is included as a state-of-the-art comparison.

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