

# Accurate Binding Free Energy Method from End-State MD Simulations

Ebru Akkus, Omer Tayfuroglu, Muslum Yildiz, and Abdulkadir Kocak\*



Cite This: *J. Chem. Inf. Model.* 2022, 62, 4095–4106



Read Online

ACCESS |



Metrics & More

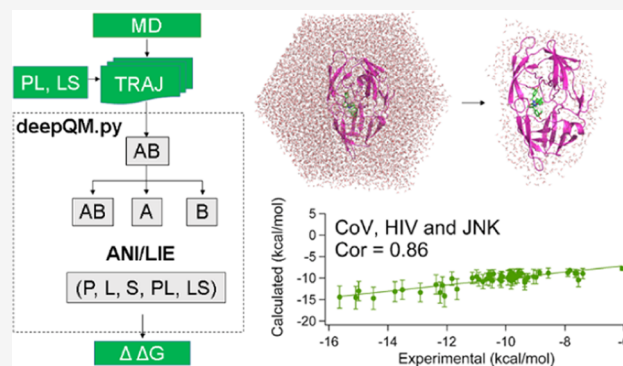


Article Recommendations



Supporting Information

**ABSTRACT:** Herein, we introduce a new strategy to estimate binding free energies using end-state molecular dynamics simulation trajectories. The method is adopted from linear interaction energy (LIE) and ANI-2x neural network potentials (machine learning) for the atomic simulation environment (ASE). It predicts the single-point interaction energies between ligand–protein and ligand–solvent pairs at the accuracy of the wb97x/6-31G\* level for the conformational space that is sampled by molecular dynamics (MD) simulations. Our results on 54 protein–ligand complexes show that the method can be accurate and have a correlation of  $R = 0.87–0.88$  to the experimental binding free energies, outperforming current end-state methods with reduced computational cost. The method also allows us to compare BFEs of ligands with different scaffolds. The code is available free of charge (documentation and test files) at <https://github.com/otayfuroglu/deepQM>.



## INTRODUCTION

Developing small compounds that target medically relevant protein is one of the main strategies in modern drug discovery studies. There are two core questions that must be addressed in these efforts: (a) what is the binding mode? (i.e., where does it bind from the receptor?) and (b) what is the binding strength? (binding free energy).<sup>1</sup> Inhibition of an enzyme can be involved in various mechanisms. For instance, an inhibitor can bind to the enzyme from the active site via competitive inhibition or it can bind from an *exo*/allosteric site by uncompetitive/noncompetitive inhibition.<sup>2</sup>

Kinetic studies conducted to understand the activity loss of the enzyme on a substrate by the inhibitor report values such as inhibition equilibrium constant  $K_i$  or half maximal inhibitory concentration  $IC_{50}$ . These two experimental parameters are proportional to each other for competitive inhibition (they are equal in special cases).<sup>2,3</sup> Thermodynamic equilibrium constant  $K_i$  and thus  $IC_{50}$  can be approximated to the experimental binding free energy of the system for the competitive inhibitors by

$$\Delta G_{\text{exp}}^0 = -RT \ln(IC_{50})$$

A more correct form is

$$\Delta \Delta G = -RT \ln(IC_{50 \text{ lig1}}/IC_{50 \text{ lig2}})$$

where  $\Delta \Delta G$  is the relative binding free energy between two competitive inhibitors.  $\Delta G$  can also be calculated from the thermodynamic potential. In principle, it is possible to compare experimental binding free energies determined from

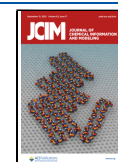
$IC_{50}$  and theoretically determined from  $IC_{50}$  predicted by thermodynamic potentials using computational methods for a series of competitive inhibitors. Thus, numerous efforts have been made to develop new strategies that can accurately predict binding free energies. However, a trade-off between computational accuracy and speed must be made.

Methods to calculate the potential binding free energy range from simple end-state methods such as linear response approximation (LRA) to linear interaction energy (LIE) and molecular mechanics Poisson–Boltzmann/generalized Born surface area (MM-(P/G)BSA)<sup>4–7</sup> to more sophisticated alchemical perturbation methods either at equilibrium such as free-energy perturbation (FEP), thermodynamic integration (TI),<sup>8–17</sup> Bennett acceptance ratio (BAR),<sup>8,12,15,18–23</sup> and multistate BAR (MBAR)<sup>24–26</sup> or nonequilibrium like Jarzynski's equality<sup>27,28</sup> and Crooks fluctuation theorem.<sup>29</sup>

Perturbation methods require several intermediate lambda states ( $\lambda$ ), in which the ligand can be decoupled, annihilated, or pulled. They use the explicit water definition, and its prediction for relative binding free energies is quite successful. However, they are still at the level of MM energy terms and

Received: May 11, 2022

Published: August 16, 2022



visit nonphysical intermediate states. All of these methods require a higher computational cost.

Binding free energy (BFE) calculations from single-step end-states are quite attractive since they require only one MD simulation for each protein–ligand (PL) complex system with less computational cost. However, the accuracy is very low due to over-simplifications such as implicit solvent definition in the case of MM-P(G)BSA and molecular mechanics (MM) definition of the Hamiltonian of the system.

Classical force fields, which do not account for electron polarization, charge transfer, and many-body effects, are very limited in accurately defining the protein–ligand interaction. Thus, the system needs to be described at the quantum mechanical (QM) level to include these essential physical effects.<sup>30,31</sup> There have been several attempts to replace the MM energies with more accurate QM energies. Recently, Fox et al.<sup>30,32</sup> have introduced a “QM-PBSA” method for eight ligands and improved the MMPBSA energies. One caveat to overcome the prohibitive DFT calculations on thousands of atoms of protein–ligand complexes is to use semiempirical quantum mechanics (SEQM) such as AM1 and SCC-DFTB or QMMM calculations.<sup>33–39</sup> There are also ongoing efforts to use fragmentation-based QM calculation methods. Most of these studies suffer from no/low sampling from MD simulations.<sup>32,38–52</sup>

Machine learning (ML) techniques have attracted great interest for the past decade due to their successful algorithms to scientific questions including but not limited to chemical reactions,<sup>53,54</sup> potential energy surfaces,<sup>53,55–57</sup> forces,<sup>58–60</sup> atomization energies,<sup>61,62</sup> and protein–ligand complex scorings.<sup>63</sup> One of the most promising aspects of the ML techniques is that the trained models can be applied to new systems (transferrable).<sup>53</sup> In particular, ML potentials with their great capability of learning a multidimensional potential energy surface (PES) at the QM level are promising due to their efficiency and scalability to large systems.<sup>64–66</sup>

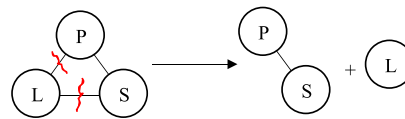
Recently, several ML-based neural network potentials (NNPs) such as ANI<sup>67–69</sup> have been developed using QM energies and forces to predict the potential energy surfaces of small organic compounds. ANI uses modified Behler and Parrinello symmetry functions to construct single-atom atomic environment vectors (AEVs) and could be considered as a quantum accurate force field without requirement of atomic charges. Thus, it only uses atomic symbols and 3D coordinates as an input for the calculations to estimate pairwise atomic interactions. The initial model (ANI-1x) used the PES of nonequilibrium geometries belonging to a total of ~17.2 million conformations generated from ~58k small organic compounds composed of C, H, O, and N atoms.<sup>70</sup> The extended version, ANI-2x, has been shown to predict DFT energies of equilibrium or nonequilibrium conformations of molecules containing C, H, O, N, S, F, and Cl atoms. It has been shown to reproduce the energies at the accuracy of the wb97x/6-31G\* level million times faster than the actual QM calculations.<sup>53</sup>

Herein, we introduce a new strategy to estimate the binding free energies of small organic compounds to proteins using end-state MD simulations. The method utilizes ANI-2x neural network potentials and predicts the binding free energies of the protein–ligand (PL) in the linear interaction energy (LIE) formalism. Our results show that the “ANI\_LIE” method outperforms current methods for binding free energies. The

code is freely available at <https://github.com/otayfuroglu/deepQM>.

## THEORY

**Linear Interaction Energy (LIE).** In the LIE method, only the protein–ligand aqueous complex (PLS) and free ligand in water (LS) are simulated. Assuming that the interactions are additive from the simulations of the protein (P) + ligand (L) + solvent (S) complex system (PLS), it can be derived as



$$E^{\text{PLS}} = E^{\text{P}} + E^{\text{L}} + E^{\text{S}} + \Delta E^{\text{P-S}} + \Delta E^{\text{L-P}} + \Delta E^{\text{L-S}}$$

$$E^{\text{PLS}} = E^{\text{P}} + E^{\text{L}} + E^{\text{S}} + (E^{\text{PS}} - E^{\text{P}} - E^{\text{S}}) + \Delta E^{\text{L-P}} + \Delta E^{\text{L-S}}$$

$$E^{\text{PLS}} - E^{\text{PS}} - E^{\text{L}} = \Delta E^{\text{L-P}} + \Delta E^{\text{L-S}}$$

$$\Delta E^{\text{L-PS}} = \Delta E^{\text{L-P}} + \Delta E^{\text{L-S}} \quad (1)$$

Thus, the interaction of the ligand with its surroundings can be calculated using the left-hand side of eq 1, which requires defining L as an energy group.

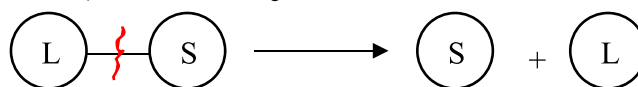
$$\Delta E^{\text{L-surr}} = \Delta E^{\text{L-PS}} + \Delta E^{\text{L-ions}} \cong \Delta E^{\text{L-PS}}$$

ANI has been trained only for atoms C, O, N, H, Cl, F, and S. Sodium ions cannot be calculated in the current version of ANI-2x. The term  $\Delta E^{\text{L-ions}}$  in this equation is the interaction energy of the ligand with ions (which were added to the simulation box to neutralize the protein), and this term can be neglected assuming that the ions will not be nearby the ligands since there are only a few ions in the box and the ligand is mostly buried in the binding pocket (Supporting Information).

The  $\Delta E^{\text{L-surr}}$  term can also be calculated from the right-hand side of eq 1 by defining two energy groups, P and L. In that case, the pairwise contributions from each of the separate energy groups of P and L can be extracted and the average interaction between each group is calculated (i.e.,  $\Delta E^{\text{L-P}}$ ,  $\Delta E^{\text{L-S}}$ )

$$\begin{aligned} \Delta E^{\text{L-surr}} &= \Delta E^{\text{L-P}} + \Delta E^{\text{L-S}} + \Delta E^{\text{L-rest}} \\ &\cong \Delta E^{\text{L-P}} + \Delta E^{\text{L-S}} \end{aligned}$$

Similarly, from the free ligand in water simulations



$$\Delta E^{\text{L-surr}} = \Delta E^{\text{L-S}}$$

Originating from the linear response approximation (LRA), the binding free-energy change from the bound (PLS) to the unbound state (LS) is given by

$$\Delta G_{\text{bind}} = \beta \langle \langle \Delta E_{\text{el}}^{\text{L-surr}} \rangle \rangle_{\text{PLS}} - \langle \langle \Delta E_{\text{el}}^{\text{L-S}} \rangle \rangle_{\text{LS}} + \gamma$$

Here, the brackets and their subscripts,  $\langle \rangle$ , denote the ensemble average and the components of the simulation, respectively.  $\Delta E_{\text{el}}^{\text{L-S}}$  corresponds to the electrostatic interaction

energy term between the ligand (L) and surroundings (solvent, S) from the ligand in solvent simulations.

Thus, the binding free energy can be calculated from

$$\Delta G_{\text{bind}} = \beta(\langle \Delta E_{\text{el}}^{\text{L-P}} \rangle_{\text{PLS}} + \langle \Delta E_{\text{el}}^{\text{L-S}} \rangle_{\text{PLS}} - \langle \Delta E_{\text{el}}^{\text{L-S}} \rangle_{\text{LS}}) + \gamma \quad (2)$$

Including van der Waals terms, the LIE equation becomes

$$\Delta G_{\text{bind}} = \alpha(\langle \Delta E_{\text{vdW}}^{\text{L-P}} \rangle_{\text{PLS}} + \langle \Delta E_{\text{vdW}}^{\text{L-S}} \rangle_{\text{PLS}} - \langle E_{\text{vdW}}^{\text{L-S}} \rangle_{\text{LS}}) + \beta(\langle \Delta E_{\text{el}}^{\text{L-P}} \rangle_{\text{PLS}} + \langle \Delta E_{\text{el}}^{\text{L-S}} \rangle_{\text{PLS}} - \langle \Delta E_{\text{el}}^{\text{L-S}} \rangle_{\text{LS}}) + \gamma \quad (3)$$

Here  $\alpha$ ,  $\beta$ , and  $\gamma$  are empirical parameters determined from linear fitting to the experimental binding energies.

**ANI/D3 Linear Interaction Energy (ANI\_LIE).** We explored the success of ANI's performance on estimation of binding energy using these LIE equations after modifications. Here, we replaced the electrostatic terms with ANI energies and van der Waals terms in the MM energies with Grimme's dispersion functions (D3) at the wb97x/6-31G\* level. In addition, since all of these equations require two simulations (PLS and LS), we have explored either by ignoring solvent terms or borrowing  $\langle E_{\text{ANI}}^{\text{L-S}} \rangle_{\text{LS}}$  term from calculations.

In linear response approximation (LRA), "preorganization energy" (POE) is included in ligand-surrounding electrostatic interaction energy. This term is computed from an ensemble in which the partial charges on the ligand atoms are set to zero. On the other hand, the POE term is ignored in a linear interaction energy (LIE) method. With this aspect, ANI follows LIE rather than LRA.

When the solvent is totally ignored and ANI interaction energies are replaced with MM electrostatic energies (similarly, D3 terms replaced with MM van der Waals terms), eqs 2 and 3 become

$$\Delta G_{\text{bind}} = \beta \langle \Delta E_{\text{ANI}}^{\text{L-P}} \rangle_{\text{PLS}} + \gamma \quad (4)$$

$$\Delta G_{\text{bind}} = \beta \langle \Delta E_{\text{ANI}}^{\text{L-P}} \rangle_{\text{PLS}} + \alpha \langle \Delta E_{\text{D3}}^{\text{L-P}} \rangle_{\text{PLS}} + \gamma \quad (5)$$

Although these equations ignore the solvent effects, the source of large noise, which stems from the high mobility of solvent molecules, could be avoided, and thus, more precise (not necessarily more accurate) trends in binding free energies could still be observed.

When the solvent is not ignored, these two equations are updated to

$$\Delta G_{\text{bind}} = \beta(\langle \Delta E_{\text{ANI}}^{\text{L-P}} \rangle_{\text{PLS}} + \langle \Delta E_{\text{ANI}}^{\text{L-S}} \rangle_{\text{PLS}} - \langle \Delta E_{\text{ANI}}^{\text{L-S}} \rangle_{\text{LS}}) + \gamma \quad (6)$$

$$\Delta G_{\text{bind}} = \beta(\langle \Delta E_{\text{ANI}}^{\text{L-P}} \rangle_{\text{PLS}} + \langle \Delta E_{\text{ANI}}^{\text{L-S}} \rangle_{\text{PLS}} - \langle \Delta E_{\text{ANI}}^{\text{L-S}} \rangle_{\text{LS}}) + \alpha(\langle \Delta E_{\text{D3}}^{\text{L-P}} \rangle_{\text{PLS}} + \langle \Delta E_{\text{D3}}^{\text{L-S}} \rangle_{\text{PLS}} - \langle \Delta E_{\text{D3}}^{\text{L-S}} \rangle_{\text{LS}}) + \gamma \quad (7)$$

Here, we generated conformational sampling of protein–ligand (PL), protein (P), and ligand (L) in the presence of explicit water using a single MD simulation (PLS) in the case of solvent ignored calculations. Then, the single-point energies (SPEs) of each MD frame have been calculated by ANI-2x and D3 by stripping out water and ions. For each MD frame, three different energies were calculated: the ligand is complexed with

protein,  $E_{\text{ANI}}^{\text{PL}}$ ; the bare protein that is extracted from the MD frame,  $E_{\text{ANI}}^{\text{P}}$ , and the bare ligand that is extracted from the MD frame,  $E_{\text{ANI}}^{\text{L}}$ . Similarly, dispersion energy of each component ( $E_{\text{D3}}^{\text{PL}}$ ,  $E_{\text{D3}}^{\text{P}}$ , and  $E_{\text{D3}}^{\text{L}}$ ) has been calculated. When the solvent is included, L–S interaction terms in the PLS MD simulation along with additional simulation of the bare ligand in water (LS) have been considered.

In this study, we used eqs 4–7 to investigate the ANI's performance of LIE approach (ANI\_LIE) in the presence of solvents. Strikingly, we found that the calculations using eq 4 outperform conventional approaches by producing BFEs that strongly correlate with experimental results. The additional terms appearing in eqs (5)–(7) bring subtle improvements to the predictions. Here, our aim is to test the success of the ANI on the interaction energy term; thus, we ignored conformational entropy changes of the ligand and protein upon binding. It has been shown that a single trajectory approach is more precise than the three-trajectory approach due to error cancellation.<sup>30</sup> Although continuum model definitions such as PBSA could be used in eqs 6 and 7 rather than using explicit solvents, we preferred not to include these terms because the primary aim of this study is to test the success of ANI in the ligand–protein interaction in the LIE approach with explicit waters. It should also be noted that the conformations are already sampled in the presence of explicit waters and ANI-predicted energies even in eqs 4 and 5 belong to equilibrium geometries of aqueous systems (not gas-phase calculations). Therefore, we believe that these conformations inherently reflect the solvent effects at some level. Here, we note that the ANI potentials were used as a postclassical MD simulation to recalculate the MD frames rather than performing the ANI potentials during the MD simulation. Thus, the method is still limited to the classical force field to generate conformational sampling. In addition, the entropic contributions of the free energies are reflected by the width of the energy distributions as with the nature of the LIE approach.

Due to the power of QM-predicted potential energy surfaces, the ANI method can also be applied to ligands that do not have similar structures as opposed to methods such as BAR, FEP, and MMP(G)BSA, in which calculations mostly require ligands with similar scaffolds. We have embedded the end-to-end script to Github for broader access.

## COMPUTATIONAL METHODS

**System Preparation.** For the binding studies, the crystal structures of noncovalent ligands complexed with 3CL<sup>Pro</sup> of SARS-CoV-1, SARS-CoV-2, and HIV-1 proteases, a total of 54 protein–ligand complexes with known IC<sub>50</sub> values (Supporting Information), were retrieved from the Protein Data Bank (PDB) except for c-Jun N-terminal kinases, which were retrieved from the work by Khalak et al.<sup>71</sup> Using Gaussian 16 software, the model ligands were first optimized at the B3LYP/6-31G\* level and ESP charges were generated at the HF/6-31G\* level. Using an Antechamber module in AmberTools 2021, RESP charges and GAFF force field atom types were generated. The Amber2gmx module<sup>72</sup> from AmberTools 2021 was used to convert amber-type input files to Gromacs.

**MD Simulations.** The molecular dynamics simulations were carried out using a Gromacs 2018+ software package<sup>73,74</sup> with an all-atom model of Amber ff99SB-ILDN<sup>75–77</sup> force field implemented in Gromacs. The protein–ligand complex (~6000 atoms) was placed in the center of a dodecahedron

box. Each system was solvated in the TIP3P model type water<sup>78</sup> with a cell margin distance of 10 Å for each dimension. The protein–ligand systems with ~70,000 atoms were neutralized in 0.15 M NaCl. Classical harmonic motions of all bonds were constrained to their equilibrium values with the LINCS algorithm.

Energy minimization was carried out to a maximum 100 kJ·mol<sup>-1</sup>·nm<sup>-1</sup> force using a Verlet cutoff scheme. For both long-range electrostatic and van der Waals interactions, a cutoff length of 12 Å was used with the particle mesh Ewald method (PME) (fourth order interpolation).<sup>79</sup> The neighbor list update frequency was set to 20 ps<sup>-1</sup>. As with our earlier studies,<sup>80–82</sup> two-step energy minimization and equilibration schemes were used. Each minimization step consisted of up to 50 000 cycles of steepest descent and a subsequent 50 000 cycles of l-bfgs integrators.

After minimization, each system was equilibrated within three steps using Langevin dynamics. The first step consisted of a 1 ns of NVT ensemble. The protein–ligand and the rest of the system were defined as two temperature groups at 310 K. The next step consisted of NPT ensembles, in which the systems were equilibrated to 1 atm pressure by Berendsen for 200 ps and followed by Parrinello–Rahman isotropic pressure coupling for 1 ns to a reference pressure of 1 atm. When systems reached to equilibrium, an MD simulation of 10 ns was carried out at the NPT ensemble.

**Free Energy Calculations.** Simulations for alchemical free energy calculations were carried out using a Gromacs 2018+ software package<sup>73,74</sup> with the all-atom model of Amber ff99SB-ILDN<sup>75–77</sup> force field implemented in Gromacs. For the alchemical free energy calculations, 10 equal decoupling steps (i.e., decoupling the ligand from the environment) of each Coulombic and van der Waals interaction ( $\Delta\lambda = 0.1$ , total 21  $\lambda$ -windows) were used. For the calculations of the protein–ligand complex, a separate decoupling of the ligand from the ligand+water system with the same parameters was performed and subtracted from the decoupling of the ligand from the protein + ligand + water complex system.<sup>80</sup> An *Alchemicalanalysis* module from the *pymbar* library<sup>83</sup> was used to print alchemical free energies.

MM-PB(GB)SA calculations were performed using `gmx_MMPBSA` script by AmberTools 2021 v4. For the calculations of protein–ligand complexes, we used the  $\lambda = 0$  trajectories of protein+ligand+water complexes from alchemical free energy calculations. Only the protein + ligand complex part from the trajectories was extracted and default parameters of implicit water with  $\epsilon = 80$  and solute with  $\epsilon = 2$  were used. Nonpolar solvation parts were separated into dispersion and cavity terms, which is defined by solvent accessible surface area from molecular volume.

ANI and D3 calculations were performed using our custom-built python scripts `deepQM` as with our previous study for LS solvation free energies (data unpublished). Since it corresponds to a fully interacting state, we used a trajectory of the first window ( $\lambda_{\text{coul}} = 0$ ;  $\lambda_{\text{vdw}} = 0$ ) from the alchemical MD simulations for QM accuracy calculations using ANI and Gaussian 16 software (G16). For G16 calculations at the `wb97x/6-31G*` level, it is computationally too expensive to carry out the entire protein–ligand system. Therefore, we had to truncate the residues around the ligand by 4.0 Å (nearly 900 atoms). Gromacs software was utilized to index energy groups, remove PBC, order the nearest waters around the PL complex, and extract structures from trajectories for ANI calculations

while PyMol was used to truncate the nearest residues around the ligand for the QM calculations at Gaussian 16 software.

## RESULTS AND DISCUSSION

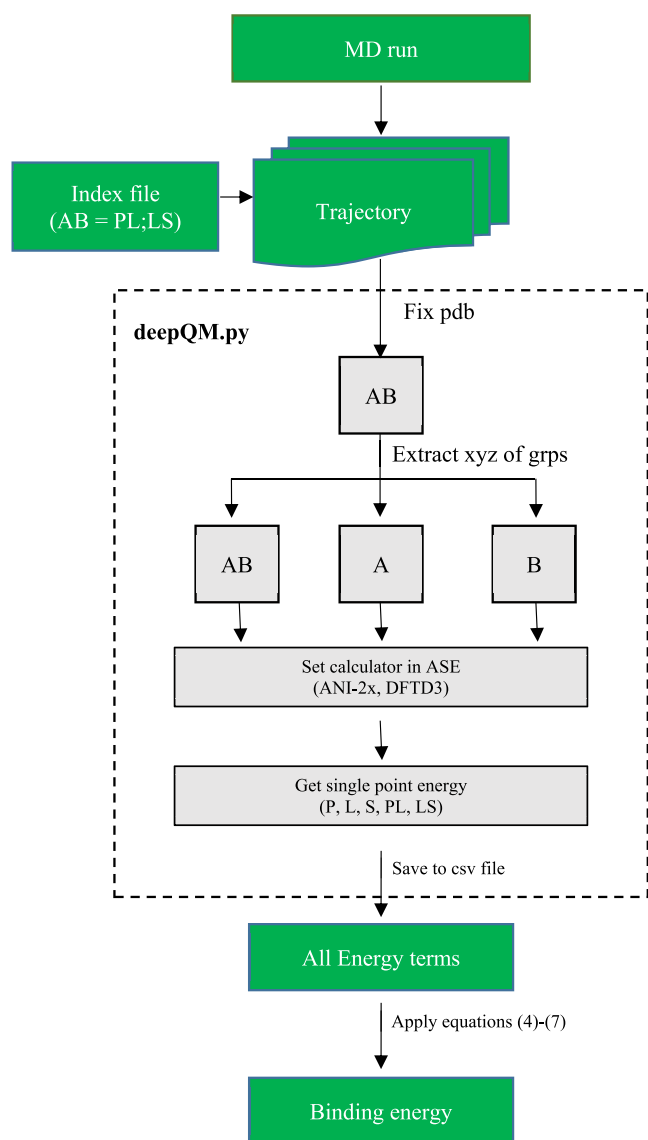
**Workflow.** `deepQM` is written in Python 3.9 (details of the functionality and different modules of `deepQM` are discussed elsewhere) and uses ASE libraries to calculate ANI and DFT-D3 single-point energies of the given A–B complex structure (e.g, A = protein, P; B = ligand, L). As an end-to-end process with the DFT accuracy, binding free energy is predicted from the PL/LS complex, free protein, and ligand structures, all of which are extracted from the trajectories of classical MD simulations of the protein–ligand complex system in water.

The workflow for the prediction of protein–ligand binding free energy from multiframe of an MD simulation is as follows: First, the atomic indices of the protein and ligand are read from the index groups of the index file and extracted into three separate xyz coordinates (PL, P, L or LS, L, S) after preprocessing the pdb files to make compatible for `deepQM` (Scheme 1). Next, ANI and DFT-D3 calculators are imported and set from the ASE platform, and then, each group is read by both calculators to predict total single-point energy (SPE) and D3 dispersion energy, respectively. Next, the interaction energy between the protein and ligand is written to data files for each frame in both methods. In the final step, binding free energy is estimated using the equations discussed in the [Computational Methods](#) section. In principle, it can work with any MD simulation packages upon converting trajectories to separate pdb files that contain the protein–ligand–solvent complex and defining energy groups. Index file format must be Gromacs type.

**ANI Results.** *Can ANI Reproduce DFT Interaction Energies?* Since ANI has been trained using atomic environments, it could still be scalable to the DFT calculations for larger systems such as protein–ligand interactions. It should be noted that DFT single-point energies of the protein and protein–ligand complex systems are in the order of hundreds of thousands of eVs, whereas the binding energy is only a few eVs. Therefore, it is important to have the total energies converged for accurate binding free energy calculations. [Table 1](#) shows the absolute energies of truncated protein–ligand complex systems (HIV protease, PDB ID = 5IVS) calculated by ANI and G16 at the `wb97x/6-31G*` level. There are 55 residues around the ligand in the truncated protein with 920 atoms ([Figure S1](#)). ANI's prediction on ligand's energy is excellent. However, the ANI predicted absolute value of the protein–ligand complex deviates by ~2 eV from that of DFT calculations. However, this energy is just a bias and canceled out by subtracting the energy of the protein in the interaction energy term. This bias cancellation will also occur on extended (untruncated) systems (i.e., a full protein + ligand complex). Therefore, the ANI-predicted interaction energies are in reasonable agreement with G16 calculations, with a shift by only ~0.252 eV (5.812 kcal/mol). This value can be minimized so as to converge by increasing the number of frames sampled by the MD simulation and increasing simulation time. It should also be noted that it has been trained for atomic environments with a cutoff (5.1 Å for radial and 4.5 Å for angular chemical environments), which limits its calculation to short-range interactions.

To assess the convergence of the total energy, we have extracted snapshots from the trajectories of 10 ns MD simulations by retrieving the first frames in different time

### Scheme 1. deepQM General Workflow for Performing End-State Binding Free Energy Calculations<sup>a</sup>



<sup>a</sup>The first step consists of extracting individual groups and converting to xyz file format. The second step is the calculator. Single or multicalculators can be selected. Finally, the results were printed and combined to predict binding free energy.

intervals (e.g, 10 ps for 1000 frames). Figure 1a shows the convergence of the average interaction energy when more snapshots are averaged, assuming 1000 frames in 10 ns MD simulations as the fully converged energy. The data suggests that using 100 frames in the average calculations is sufficient to have a convergence to a maximum of 0.04 eV (1.1 kcal/mol). This error drops to below 0.02 eV (0.46 kcal/mol) when 200 frames are used.

We have also investigated how MD simulation time affects the interaction energy (Figure 1b). It suggests that the simulation time does have more impact on the absolute errors and an 8 ns of simulation time can bring a maximum error of 0.09 eV (2.5 kcal/mol). This error reduces to below 0.5 kcal/mol in 10 ns simulations.

One of the most important features of the ANI calculations is that they can be run on GPU or CPU and either construction can be run in parallel. The computation time is million times faster than DFT calculations. On a 4x NVIDIA Tesla V100 construction, the 1000 frames of the P–L and L–S complexes can be calculated within hours if not minutes. A single frame in our truncated P–L structures with 55 residues and 920 atoms could take several hours at the wb97x/6-31G\* level.

*Can the Electrostatic Interaction Term Be Replaced by ANI?* The ANI and D3 energy terms are correlated and can be replaced with Coulombic and van der Waals terms, respectively. As both van der Waals interaction terms in MM and Grimme's D3 functions originate from Lennard-Jones potentials, it would be trivial to correlate these two terms. Thus, the original LIE equation can easily be adopted by replacing D3 energies with MM-based van der Waals terms and by replacing MM-based electrostatic interaction energy terms with ANI-predicted interaction energies (Figure 2). ANI also shows great correlation to electrostatic terms, which allows us for the modification of the LIE equation.

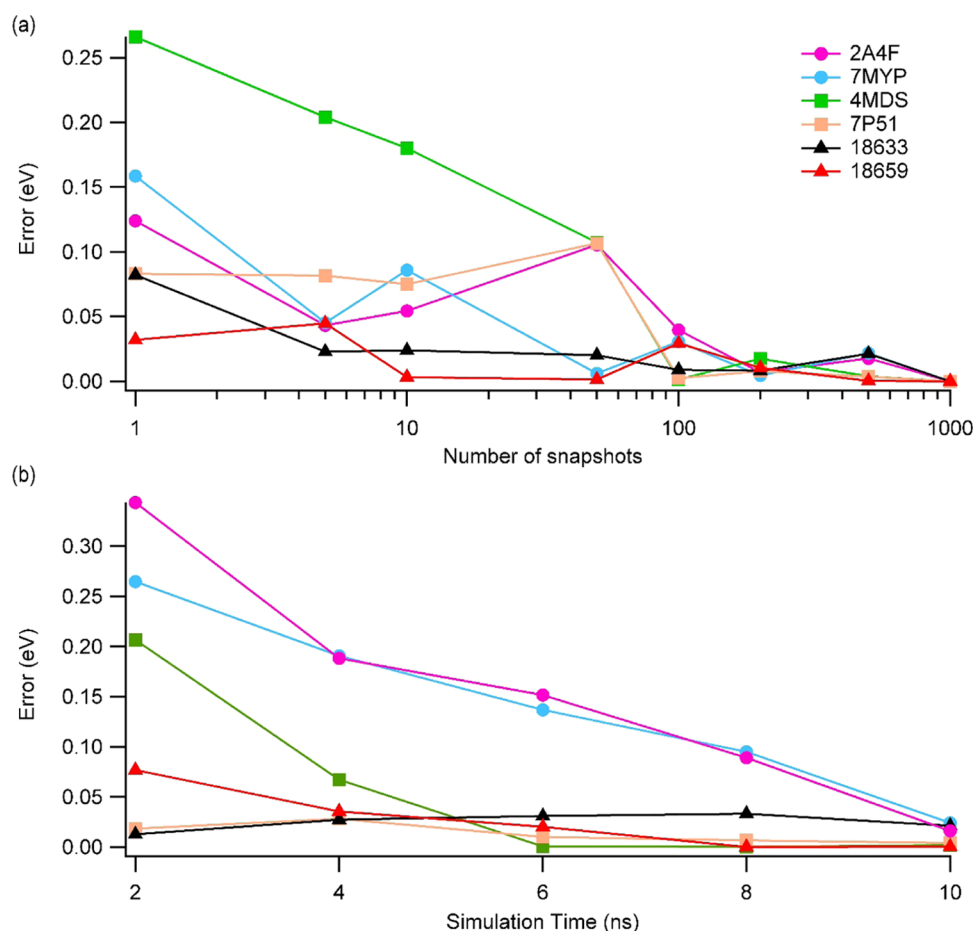
Since ANI and MM energy terms are strongly correlated, one should expect LIE calculation by both methods to follow similar trends with respect to experimental free energies. Surprisingly, the correlation of ANI\_LIE values to experimental free energies is greater by ~15% than that of MM\_LIE.

*BFE by Ignoring LS Interaction.* Using the simplest modification of the LIE formula (i.e., eq 4), we have calculated the ANI interaction energies. The coefficients were determined from the fit to the experimental values (Figure 3). It should be emphasized that there are three different protein families with a total of 54 complexes and the ligands do not necessarily have similar scaffolds. We observed excellent agreement ( $R = 0.87$ )

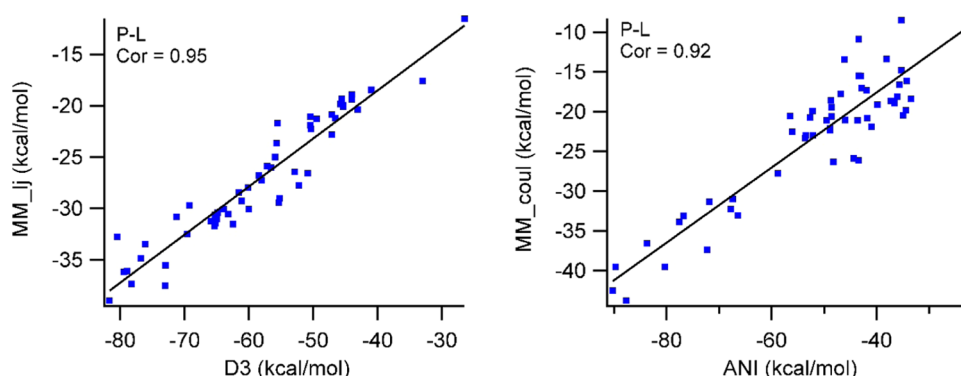
**Table 1. Absolute Energies of Truncated Protein + Ligand Systems (HIV Protease, PDB ID = S1VS) from MD Simulations (in eV)<sup>a</sup>**

frame	G16				ANI				abs. err.
	complex PL	receptor P	ligand L	$\Delta E_{\text{int}}$	complex PL	receptor P	ligand L	$\Delta E_{\text{int}}$	
0	-601 409.137	-543 088.071	-58 316.912	-4.154	-601 407.549	-543 086.919	-58 316.653	-3.977	0.177
1	-611 338.434	-553 017.357	-58 317.024	-4.053	-611 337.259	-553 016.927	-58 316.904	-3.428	0.625
2	-611 343.896	-553 023.284	-58 316.737	-3.876	-611 342.856	-553 022.874	-58 316.543	-3.439	0.436
3	-611 338.693	-553 021.427	-58 316.102	-1.164	-611 337.181	-553 020.251	-58 315.918	-1.012	0.152
:									
100	-611 340.901	-553 023.152	-58 315.756	-1.992	-611 340.670	-553 023.116	-58 315.703	-1.878	0.115
mean	-611 366.228	-553 047.220	-58 316.584	-2.424	-611 368.035	-553 046.484	-58 319.368	-2.183	0.252

<sup>a</sup>For the complete 100 frames, refer to the Supporting Information.



**Figure 1.** Absolute deviations of the average P–L interaction energies for selected complexes calculated by ANI, considering 1000 frames and 10 ns simulations as the converged value (a) as a function of the number of frames to average and (b) as a function of simulation time using only 100 frames in each time interval.



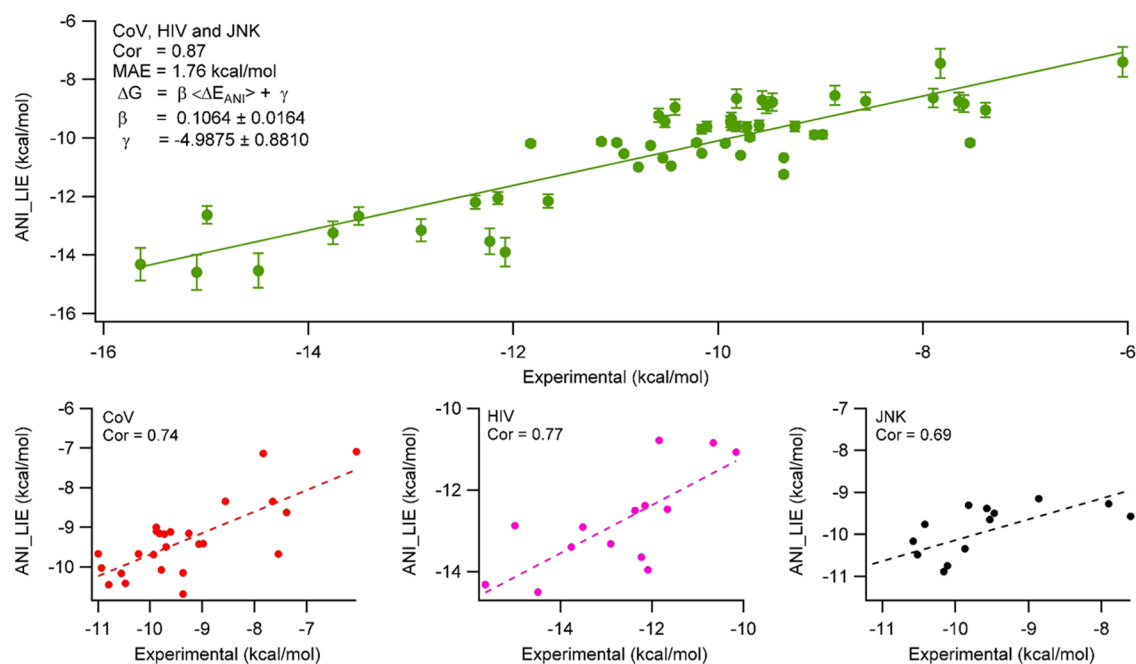
**Figure 2.** ANI-predicted average interaction energies and classical electrostatic interaction energies between the protein and ligand from PLS simulations are correlated while the energies from Grimme's D3 functions are correlated to van der Waals interactions.

between the experimental and ANI/LIE-predicted absolute binding free energy values. The mean absolute error (MAE) is just 1.76 kcal/mol. Similar to LIE calculations using only ANI potentials, we also investigated the effect of the inclusion of D3 terms in the prediction using eq 5. Although the correlation coefficient was not improved significantly, the MAE value decreased to 0.76 kcal/mol (Figure S2).

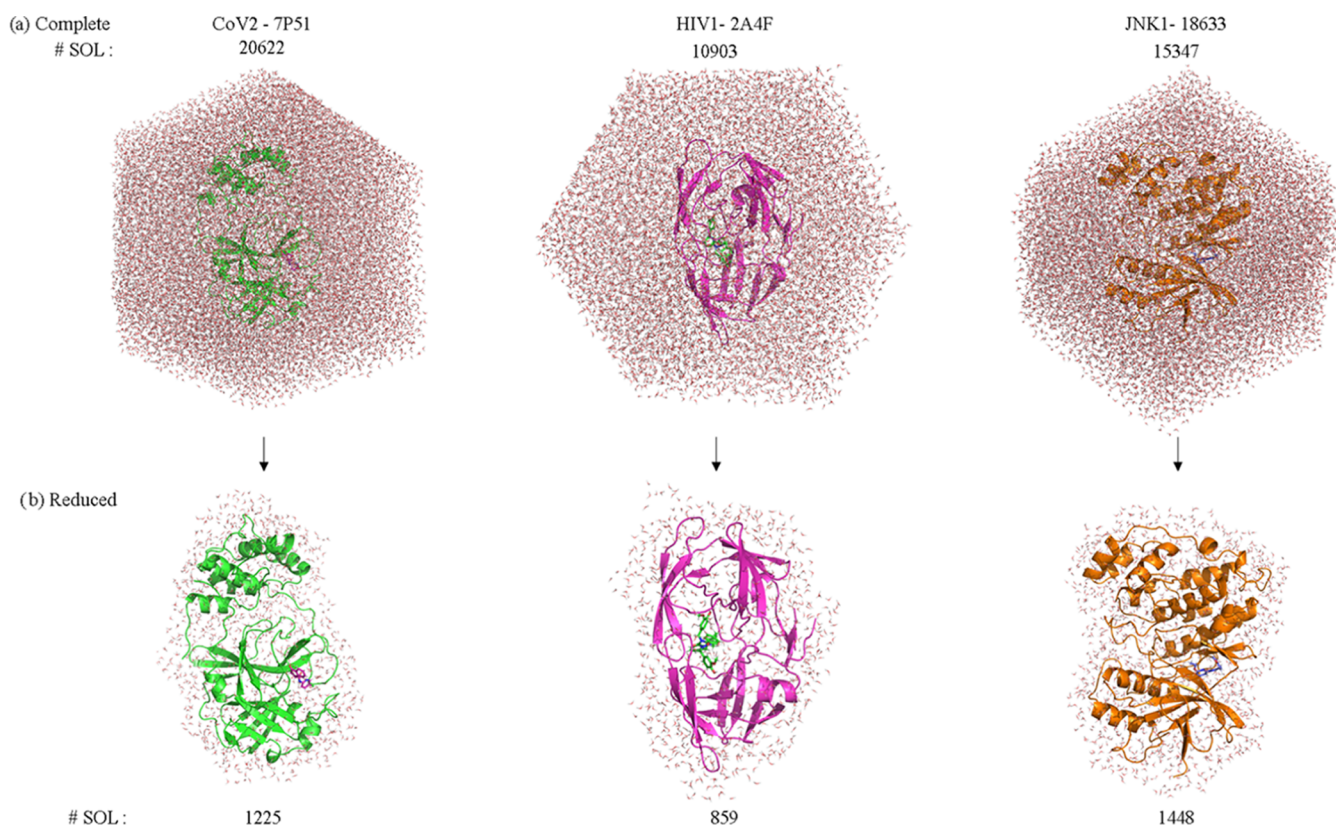
**BFE by Including LS Interaction.** After showing the success of ANI's prediction of interaction energies and binding free energies in the absence of solvents, we have also attempted to

calculate the binding free energies using the most extensive LIE formalism in which the solvent effects are also included. The solvent requires a second simulation of ligand in water (LS) along with the contribution of L–S interactions in the PLS simulations as explained in the Theory section.

In principle, ligand–solvent interactions can be computed with implicit definitions as in the case of PBSA, GBSA, or any other continuum models. For instance, a recent work by Fox et al. introduced the so-called “ONETEP” program, in which they coupled the DFT energies with PBSA to calculate QM–PBSA



**Figure 3.** ANI interaction energies calculated from eq 4 with  $\beta = 0.10639 \pm 0.0164$  and  $\gamma = -4.9875 \pm 0.881$  using average of 100 snapshots of 10 ns MD simulations for 54 protein–ligand complexes, showing 0.87 correlation to the experimental values. The bar lines show uncertainties. The bottom plots show the fits of individual protein families as CoV, HIV-1, and JNK-1.

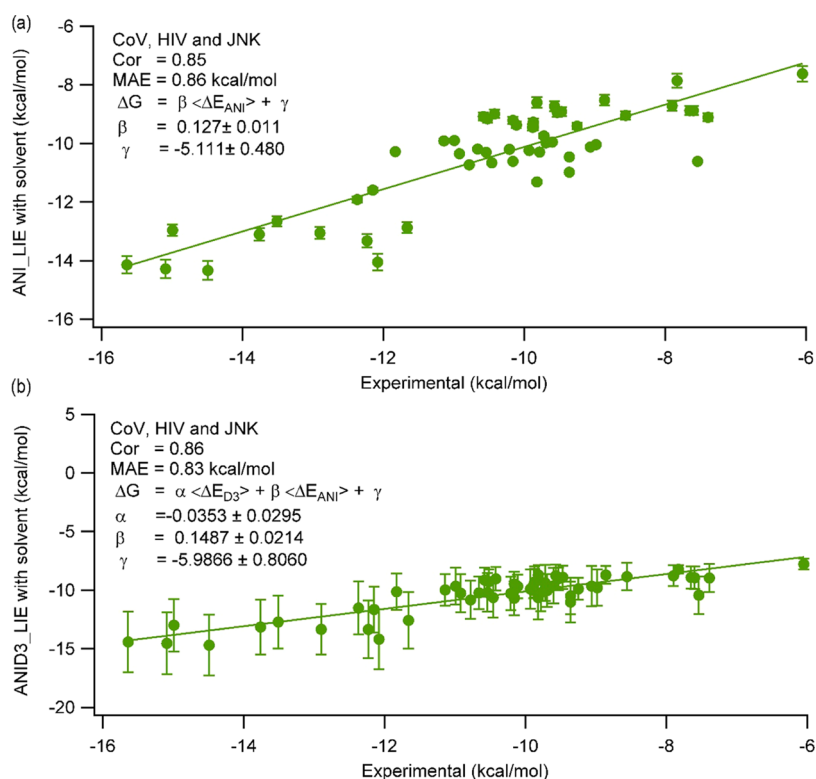


**Figure 4.** Illustration for the representative structures in protein + ligand + complete/reduced water (PLS) systems. The LIE calculations were performed using reduced water to ease memory cost of the computations. The results are almost identical in both cases (Table S1).

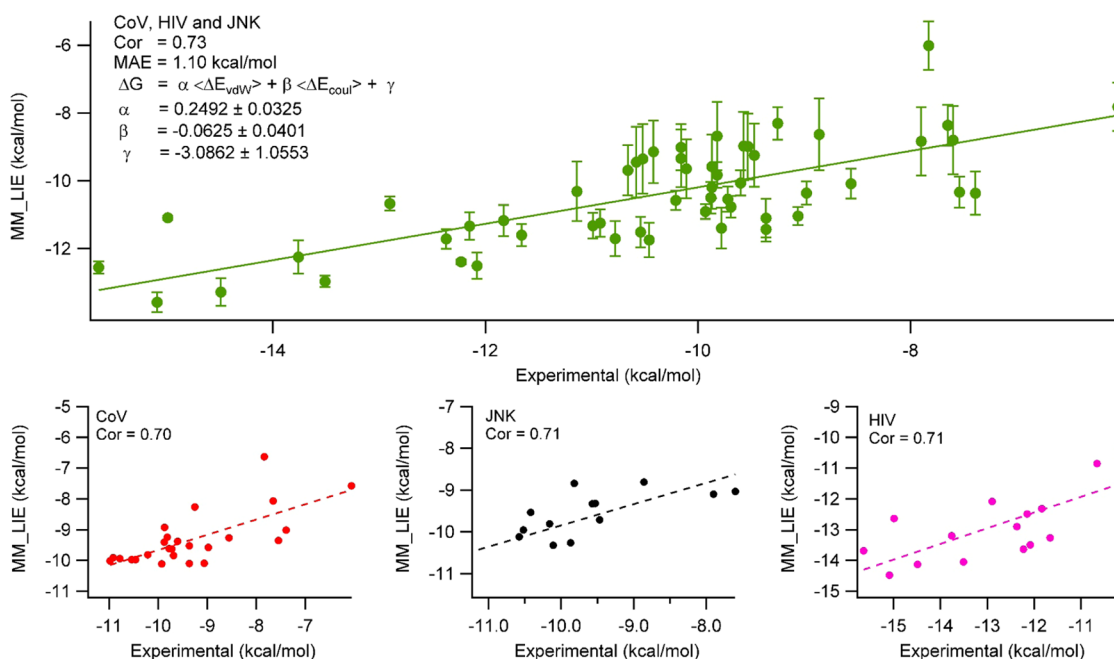
energies.<sup>30</sup> However, here, we attempted the L–S interaction energies using explicit waters in the system sticking with the original formalism of the LIE method.

Although ANI enables us to calculate such large protein + ligand + water (PLS) systems, due to memory limitations, we

could only perform L–S interaction energies after reducing the PLS systems so as to include waters by only 4 Å around the PL complex. We achieved this by ordering the nearest solvents in the trajectories and extracting the nearest solvents in the PLS system. Figure 4 shows examples of reduced vs complete water



**Figure 5.** LIE calculated by ANI and ANID3 using (a) eq 6 and (b) eq 7, in which the solvent effects around the ligand were included in the calculations.



**Figure 6.** LIE calculated by MM electrostatic and van der Waals interactions using eq 3. The parameters were determined from the fit. The bottom plots show the correlation in each individual protein family.

in PLS systems for each protein family. Since the ligand is mostly buried in the protein, there are only a limited number of solvents around the ligand in the bound state. Thus, the effect of the solvents that are beyond 4 Å should not affect the calculations. Indeed, one can see the comparison of MM energy terms (Coulombic, van der Waals) along with

MM\_LIE calculations by taking only reduced water in the system rather than the complete water (Table S1).

By applying eqs 6 and 7, we could successfully produce the absolute binding free energies based on ANI/D3 interactions. Figure 5 shows that the solvent included ANI and ANID3 LIE calculations. We have summarized the correlation coefficients for different approaches in Table S2.



**Benchmarking with Other Methods.** *Binding Free Energies from the Classical LIE Approach.* In addition to the ANI method introduced here, we also performed calculations from several other methods for the studied protein–ligand complexes. The very first method was the classical LIE method, in which Coulombic and van der Waals interactions at the level of molecular mechanics were computed and fit to the experimental binding free energies to determine the coefficients.

The default parameters of LIE in Gromacs software use  $\alpha = 0.181$ ,  $\beta = 0.5$ , and  $\gamma = 0$  and yield very low correlation to the experimental values. The  $\beta$  coefficient for the electrostatic interactions was reported to have several values according to the charged vs uncharged ligands. Studying 18 PL complexes, Hansson et al.<sup>84</sup> concluded that  $\beta = 0.5$  is a good approximation for charged ligands whereas lower values such as 0.43, 0.37, and 0.33 for neutral molecules with 0, 1, or >1 hydroxyl groups, respectively.<sup>85</sup> Since then, it has been used with care by fitting to the new sets of protein–ligand complexes.<sup>86–91</sup>

With the default parameters ( $\alpha = 0.181$ ,  $\beta = 0.5$ , and  $\gamma = 0$ ), we have observed very low correlation between the experimental and MM\_LIE prediction (Figure S3). On the other hand, when new parameters are assigned from our fit, we observe 0.70–0.73 correlation with  $\alpha = 0.25$ ,  $\beta = -0.06$ , and  $\gamma = -3.09$  (Figure 6). Here, all of the ligands studied are neutral and we did not classify them according to the presence of polar groups like hydroxyl groups. Note that MM\_LIE is still quite successful ( $R = 0.70$ – $0.73$ ), and it is clear that ANI\_LIE ( $R = 0.85$ – $0.87$ ) is much better in reproducing experimental values.

In addition to the parameters calculated from all 54 data points of the dataset for both MM\_LIE and ANID3\_LIE, we have also analyzed the data by splitting training (90%, 48 points) and test sets (10%, 6 points). We have done this by randomly selecting 100 different train and test sets. The coefficients were produced from the training sets and applied on test sets. The coefficients generated from random sampling and whole dataset are in a good agreement. We have reported the predicted values for the coefficients and their RMSE values on training and test sets (in the Supporting Information and Table S2).

We have also performed MMPBSA and MMGBSA calculations as a comparison to ANI. The computational cost of these two methods is very similar to ANI since they are also end-state methods applied on bound-state (PLS) classical MD simulations. We should note that these methods do not use empirical fitting coefficients in the calculations. However, their success of predicting absolute binding energies can become very poor according to the system and they are mostly used to predict the relative binding free energies. In particular, when the ligands have similar scaffolds, they can accurately predict relative binding free energies. Similarly, when the same ligand's binding to a protein and its mutations were compared, they can be quite successful. We have tested the performance of both MMPBSA and MMGBSA methods for the protein–ligand complexes studied here.

Due to the fact that protein families of the ligands are very different, we could only observe somewhat meaningful correlations when the protein families were analyzed separately in the case of MMPBSA. When all proteins were treated with the same preassigned values of MMPBSA, we observed very poor correlation to the experimental values (Figure S4).

Surprisingly, MMGBSA showed a somewhat similar correlation ( $R = 0.81$ ) with classical LIE to the experiments even when all 54 protein–ligand complexes were analyzed together (Figure S5). This is still less than the ANI-predicted values of  $R = 0.87$ – $0.88$ .

We have also explored the binding energies by means of alchemical methods using a 10  $\lambda$  decoupling window for electrostatic interaction and 10  $\lambda$  for van der Waals interactions. During decoupling, no restraint was applied on the ligands; 10 ns from each window brings 20-fold computational cost to the simulations, whereas it still has a bare correlation to the experimental binding free energies (Figure S6). These low correlations may be due to the limited number of windows.

## SUMMARY AND CONCLUSIONS

Here, we applied ANI-ML potentials, trained at the wb97x/6-31G\* level, to calculate the interaction energies of protein–ligand and ligand–solvent pairs. The results show that the predicted interaction energies even when the solvent is totally ignored are highly correlated with experimental values. By modification of the LIE method, we have assigned coefficients for the ANI\_LIE method. The ANI\_LIE method outperforms conventional methods like LIE, BAR, and MMP(G)BSA in terms of accuracy and is comparable to MMP(B)SA in terms of computational cost. Our preassigned parameters determined for the method can be applied on any protein–ligand complex systems even when the protein(s)/ligand(s) have different scaffolds.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.2c00601>.

Supplementary figures of truncated protein + ligand in the DFT calculations, ANID3\_LIE results, MMP(G)-BSA and BAR results and table of the classical LIE when the complete/reduced solvents considered (PDF)

Raw data of all outputs (XLSX)

## AUTHOR INFORMATION

### Corresponding Author

Abdulkadir Kocak – Department of Chemistry, Gebze Technical University, 41400 Gebze, Kocaeli, Turkey; [orcid.org/0000-0001-6891-6929](https://orcid.org/0000-0001-6891-6929); Phone: +902626053083; Email: [kocak@gtu.edu.tr](mailto:kocak@gtu.edu.tr)

### Authors

Ebru Akkus – Department of Bioengineering, Gebze Technical University, 41400 Gebze, Kocaeli, Turkey; [orcid.org/0000-0002-3940-0955](https://orcid.org/0000-0002-3940-0955)

Omer Tayfuroglu – Department of Chemistry, Gebze Technical University, 41400 Gebze, Kocaeli, Turkey; [orcid.org/0000-0001-7834-3132](https://orcid.org/0000-0001-7834-3132)

Muslum Yildiz – Department of Molecular Biology and Genetics, Gebze Technical University, 41400 Gebze, Kocaeli, Turkey

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jcim.2c00601>

### Notes

The authors declare no competing financial interest.

The ANI\_LIE method using deepQM along with tutorials and descriptions can be accessed via [<https://github.com/otayfuroglu/deepMOF.git>]. Input and output files from simulations can be accessed via: <https://doi.org/10.5281/zenodo.6538001>.

## ACKNOWLEDGMENTS

The numerical calculations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). This work was supported by Scientific and Technological Research Council of Turkey—TUBITAK (Project Number 120Z732). E.A. acknowledges Dr. Pinar Pir (Department of Bioengineering at Gebze Technical University) for her kind support.

## REFERENCES

- (1) Limongelli, V. Ligand Binding Free Energy and Kinetics Calculation in 2020. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10*, No. e1455.
- (2) Burlingham, B. T.; Widlanski, T. S. An Intuitive Look at the Relationship of  $K_i$  and  $IC_{50}$ : A More General Use for the Dixon Plot. *J. Chem. Educ.* **2003**, *80*, 214–218.
- (3) Yung-Chi, C.; Prusoff, W. H. Relationship between the Inhibition Constant ( $K_i$ ) and the Concentration of Inhibitor Which Causes 50 per Cent Inhibition ( $I_{50}$ ) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.
- (4) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449–461.
- (5) Kumari, R.; Kumar, R.; Lynn, A. G-Mmpbsa -A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951–1962.
- (6) Rifai, E. A.; Van Dijk, M.; Vermeulen, N. P. E.; Yanuar, A.; Geerke, D. P. A Comparative Linear Interaction Energy and MM/PBSA Study on SIRT1-Ligand Binding Free Energy Calculation. *J. Chem. Inf. Model.* **2019**, *59*, 4018–4033.
- (7) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys. J.* **2004**, *86*, 67–74.
- (8) De Ruiter, A.; Boresch, S.; Oostenbrink, C. Comparison of Thermodynamic Integration and Bennett's Acceptance Ratio for Calculating Relative Protein-Ligand Binding Free Energies. *J. Comput. Chem.* **2013**, *34*, 1024–1034.
- (9) Lawrenz, M.; Wereszczynski, J.; Ortiz-Sánchez, J. M.; Nichols, S. E.; McCammon, J. A. Thermodynamic Integration to Predict Host-Guest Binding Affinities. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 569–576.
- (10) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. *J. Comput. Chem.* **2011**, *32*, 3253–3263.
- (11) König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L. Multiscale Free Energy Simulations: An Efficient Method for Connecting Classical MD Simulations to QM or QM/MM Free Energies Using Non-Boltzmann Bennett Reweighting Schemes. *J. Chem. Theory Comput.* **2014**, *10*, 1406–1419.
- (12) Bruckner, S.; Boresch, S. Efficiency of Alchemical Free Energy Simulations. II. Improvements for Thermodynamic Integration. *J. Comput. Chem.* **2011**, *32*, 1320–1333.
- (13) Kästner, J.; Thiel, W. Bridging the Gap between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: "Umbrella Integration. *J. Chem. Phys.* **2005**, *123*, No. 144104.
- (14) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. QM/MM Free-Energy Perturbation Compared to Thermodynamic Integration and Umbrella Sampling: Application to an Enzymatic Reaction. *J. Chem. Theory Comput.* **2006**, *2*, 452–461.
- (15) Shirts, M. R.; Pande, V. S. Comparison of Efficiency and Bias of Free Energies Computed by Exponential Averaging, the Bennett Acceptance Ratio, and Thermodynamic Integration. *J. Chem. Phys.* **2005**, *122*, No. 144107.
- (16) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994**, *100*, 9025–9031.
- (17) Lee, T. S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; Dwight McGee, T.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H. C.; Xu, H.; Sherman, W.; York, D. M. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595–5623.
- (18) Kim, I.; Allen, T. W. Bennetts Acceptance Ratio and Histogram Analysis Methods Enhanced by Umbrella Sampling along a Reaction Coordinate in Configurational Space. *J. Chem. Phys.* **2012**, *136*, No. 164103.
- (19) König, G.; Boresch, S. Non-Boltzmann Sampling and Bennett's Acceptance Ratio Method: How to Profit from Bending the Rules. *J. Comput. Chem.* **2011**, *32*, 1082–1090.
- (20) Giese, T. J.; York, D. M. Development of a Robust Indirect Approach for MM  $\rightarrow$  QM Free Energy Calculations That Combines Force-Matched Reference Potential and Bennett's Acceptance Ratio Methods. *J. Chem. Theory Comput.* **2019**, *15*, 5543–5562.
- (21) Jia, X.; Wang, M.; Shao, Y.; König, G.; Brooks, B. R.; Zhang, J. Z. H.; Mei, Y. Calculations of Solvation Free Energy through Energy Reweighting from Molecular Mechanics to Quantum Mechanics. *J. Chem. Theory Comput.* **2016**, *12*, 499–511.
- (22) Li, Y.; Nam, K. Repulsive Soft-Core Potentials for Efficient Alchemical Free Energy Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 4776–4789.
- (23) Li, P.; Jia, X.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at Ab Initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semi-Empirical Reference Potential. I. Weighted Thermodynamics Perturbation. *J. Chem. Theory Comput.* **2018**, *14*, 5583–5596.
- (24) Gumbart, J. C.; Roux, B.; Chipot, C. Efficient Determination of Protein-Protein Standard Binding Free Energies from First Principles. *J. Chem. Theory Comput.* **2013**, *9*, 3789–3798.
- (25) Gumbart, J.; Chipot, C.; Schulten, K. Free-Energy Cost for Translocon-Assisted Insertion of Membrane Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3596–3601.
- (26) Giese, T. J.; York, D. M. Variational Method for Networkwide Analysis of Relative Ligand Binding Free Energies with Loop Closure and Experimental Constraints. *J. Chem. Theory Comput.* **2021**, *17*, 1326–1336.
- (27) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
- (28) Boresch, S.; Woodcock, H. L. Convergence of Single-Step Free Energy Perturbation. *Mol. Phys.* **2017**, *115*, 1200–1213.
- (29) Crooks, G. E. Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences. *Phys. Rev. E* **1999**, *60*, 2721–2726.
- (30) Fox, S. J.; Dziedzic, J.; Fox, T.; Tautermann, C. S.; Skylaris, C. K. Density Functional Theory Calculations on Entire Proteins for Free Energies of Binding: Application to a Model Polar Binding Site. *Proteins* **2014**, *82*, 3335–3346.
- (31) Cournia, Z.; Chipot, C.; Roux, B.; York, D. M.; Sherman, W. Free Energy Methods in Drug Discovery - Introduction. *ACS Symp. Ser.* **2021**, *1397*, 1–38.
- (32) Gundelach, L.; Fox, T.; Tautermann, C. S.; Skylaris, C. K. Protein-Ligand Free Energies of Binding from Full-Protein DFT Calculations: Convergence and Choice of Exchange-Correlation Functional. *Phys. Chem. Chem. Phys.* **2021**, *23*, 9381–9393.
- (33) Gräter, F.; Schwarzl, S. M.; Dejaegere, A.; Fischer, S.; Smith, J. C. Protein/Ligand Binding Free Energies Calculated with Quantum Mechanics/Molecular Mechanics. *J. Phys. Chem. B* **2005**, *109*, 10474–10483.
- (34) Ibrahim, M. A. A. Performance Assessment of Semiempirical Molecular Orbital Methods in Describing Halogen Bonding:

Quantum Mechanical and Quantum Mechanical/Molecular Mechanical-Molecular Dynamics Study. *J. Chem. Inf. Model.* **2011**, *51*, 2549–2559.

(35) Retegan, M.; Milet, A.; Jamet, H. Exploring the Binding of Inhibitors Derived from Tetrabromobenzimidazole to the CK2 Protein Using a QM/MM-PB/SA Approach. *J. Chem. Inf. Model.* **2009**, *49*, 963–971.

(36) Wichapong, K.; Rohe, A.; Platzer, C.; Slynko, I.; Erdmann, F.; Schmidt, M.; Sippl, W. Application of Docking and QM/MM-GBSA Rescoring to Screen for Novel Myt1 Kinase Inhibitors. *J. Chem. Inf. Model.* **2014**, *54*, 881–893.

(37) Díaz, N.; Suárez, D.; Merz, K. M.; Sordo, T. L. Molecular Dynamics Simulations of the TEM-1  $\beta$ -Lactamase Complexed with Cephalothin. *J. Med. Chem.* **2005**, *48*, 780–791.

(38) Wang, Y. T.; Chen, Y. C. Insights from QM/MM Modeling the 3D Structure of the 2009 H1N1 Influenza A Virus Neuraminidase and Its Binding Interactions with Antiviral Drugs. *Mol. Inf.* **2014**, *33*, 240–249.

(39) Barbault, F.; Maurel, F. Is Inhibition Process Better Described with MD(QM/MM) Simulations? The Case of Urokinase Type Plasminogen Activator Inhibitors. *J. Comput. Chem.* **2012**, *33*, 607–616.

(40) Tsitsanou, K. E.; Hayes, J. M.; Keramioti, M.; Mamais, M.; Oikonomakos, N. G.; Kato, A.; Leonidas, D. D.; Zographos, S. E. Sourcing the Affinity of Flavonoids for the Glycogen Phosphorylase Inhibitor Site via Crystallography, Kinetics and QM/MM-PBSA Binding Studies: Comparison of Chrysin and Flavopiridol. *Food Chem. Toxicol.* **2013**, *61*, 14–27.

(41) Fedorov, D. G. The Fragment Molecular Orbital Method: Theoretical Development, Implementation in GAMESS, and Applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7*, No. e1322.

(42) Kurauchi, R.; Watanabe, C.; Fukuzawa, K.; Tanaka, S. Novel Type of Virtual Ligand Screening on the Basis of Quantum-Chemical Calculations for Protein-Ligand Complexes and Extended Clustering Techniques. *Comput. Theor. Chem.* **2015**, *1061*, 12–22.

(43) Tagami, U.; Takahashi, K.; Igarashi, S.; Ejima, C.; Yoshida, T.; Takeshita, S.; Miyanaga, W.; Sugiki, M.; Tokumasu, M.; Hatanaka, T.; Kashiwagi, T.; Ishikawa, K.; Miyano, H.; Mizukoshi, T. Interaction Analysis of FABP4 Inhibitors by X-Ray Crystallography and Fragment Molecular Orbital Analysis. *ACS Med. Chem. Lett.* **2016**, *7*, 435–439.

(44) Heifetz, A.; Chudyk, E. I.; Gleave, L.; Aldeghi, M.; Cherezov, V.; Fedorov, D. G.; Biggin, P. C.; Bodkin, M. J. The Fragment Molecular Orbital Method Reveals New Insight into the Chemical Nature of GPCR-Ligand Interactions. *J. Chem. Inf. Model.* **2016**, *56*, 159–172.

(45) Chen, X.; Zhao, X.; Xiong, Y.; Liu, J.; Zhan, C. G. Fundamental Reaction Pathway and Free Energy Profile for Hydrolysis of Intracellular Second Messenger Adenosine 3',5'-Cyclic Monophosphate (cAMP) Catalyzed by Phosphodiesterase-4. *J. Phys. Chem. B* **2011**, *115*, 12208–12219.

(46) Kaukonen, M.; Söderhjelm, P.; Heimdal, J.; Ryde, U. QM/MM-PBSA Method to Estimate Free Energies for Reactions in Proteins. *J. Phys. Chem. B* **2008**, *112*, 12537–12548.

(47) Wang, M.; Wong, C. F. Rank-Ordering Protein-Ligand Binding Affinity by a Quantum Mechanics/Molecular Mechanics/Poisson-Boltzmann-Surface Area Model. *J. Chem. Phys.* **2007**, *126*, No. 026101.

(48) Söderhjelm, P.; Kongsted, J.; Ryde, U. Ligand Affinities Estimated by Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 1726–1737.

(49) Manta, S.; Xipnitou, A.; Kiritsis, C.; Kantsadi, A. L.; Hayes, J. M.; Skamnaki, V. T.; Lamprakis, C.; Kontou, M.; Zoumpoulakis, P.; Zographos, S. E.; Leonidas, D. D.; Komiotis, D. 3'-Axial CH 2OH Substitution on Glucopyranose Does Not Increase Glycogen Phosphorylase Inhibitory Potency. QM/MM-PBSA Calculations Suggest Why. *Chem. Biol. Drug Des.* **2012**, *79*, 663–673.

(50) Söderhjelm, P.; Kongsted, J.; Genheden, S.; Ryde, U. Estimates of Ligand-Binding Affinities Supported by Quantum Mechanical Methods. *Interdiscip. Sci. Comput. Life Sci.* **2010**, *2*, 21–37.

(51) Sawada, T.; Fedorov, D. G.; Kitaura, K. Role of the Key Mutation in the Selective Binding of Avian and Human Influenza Hemagglutinin to Sialosides Revealed by Quantum-Mechanical Calculations. *J. Am. Chem. Soc.* **2010**, *132*, 16862–16872.

(52) Lu, H.; Huang, X.; Abdulkameed, M. D. M.; Zhan, C. G. Binding Free Energies for Nicotine Analogs Inhibiting Cytochrome P450 2A6 by a Combined Use of Molecular Dynamics Simulations and QM/MM-PBSA Calculations. *Bioorg. Med. Chem.* **2014**, *22*, 2149–2156.

(53) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. Outsmarting Quantum Chemistry Through Transfer Learning, 2018. DOI: 10.26434/chemrxiv.6744440.v1.

(54) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186–190.

(55) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K. R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3*, No. 1603015.

(56) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, No. 13890.

(57) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828–12840.

(58) Li, Z.; Kermode, J. R.; De Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015**, *114*, 096405.

(59) Glielmo, A.; Sollich, P.; De Vita, A. Accurate Interatomic Force Fields via Machine Learning with Covariant Kernels. *Phys. Rev. B* **2017**, *95*, No. 214302.

(60) Kruglov, I.; Sergeev, O.; Yanilkin, A.; Oganov, A. R. Energy-Free Machine Learning Force Field for Aluminum. *Sci. Rep.* **2017**, *7*, No. 8512.

(61) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(62) Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; Von Lilienfeld, O. A. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J. Chem. Theory Comput.* **2017**, *13*, 5255–5264.

(63) Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein-Ligand Scoring with Convolutional Neural Networks. *J. Chem. Inf. Model.* **2017**, *57*, 942–957.

(64) Roman, Z.; S, S. J.; Jerzy, L.; Olexandr, I. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Sci. Adv.* **2019**, *5*, No. eaav6490.

(65) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693.

(66) Mascotti, M. L.; Juri Ayub, M.; Furnham, N.; Thornton, J. M.; Laskowski, R. A. Chopping and Changing: The Evolution of the Flavin-Dependent Monooxygenases. *J. Mol. Biol.* **2016**, *428*, 3131–3146.

(67) Smith, J. S.; Isayev, O.; Roitberg, A. E. Data Descriptor: ANI-1, A Data Set of 20 Million Calculated off-Equilibrium Conformations for Organic Molecules. *Sci. Data* **2017**, *4*, No. 170193.

(68) Smith, J. S.; Roitberg, A. E.; Isayev, O. Transforming Computational Drug Discovery with Machine Learning and AI. *ACS Med. Chem. Lett.* **2018**, *9*, 1065–1069.

(69) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less Is More: Sampling Chemical Space with Active Learning. *J. Chem. Phys.* **2018**, *148*, No. 241733.

- (70) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192–3203.
- (71) Khalak, Y.; Tresadern, G.; Aldeghi, M.; Baumann, H. M.; Mobley, D. L.; de Groot, B. L.; Gapsys, V. Alchemical Absolute Protein–Ligand Binding Free Energies for Drug Design. *Chem. Sci.* **2021**, *12*, 13958–13971.
- (72) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cisneros, G. A.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O’Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *AmberTools22*; University of California: San Francisco, 2022.
- (73) Lindahl, A.; van der Hess, S. GROMACS 2020.6 Source Code, 2021. <https://doi.org/10.5281/ZENODO.4576055>.
- (74) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (75) Caleman, C.; Van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; Van Der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61–74.
- (76) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (77) 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. *Proteins* **2010**, *78*, 1950–1958.
- (78) Toukan, K.; Rahman, A. Molecular-Dynamics Study of Atomic Motions in Water. *Phys. Rev. B* **1985**, *31*, 2643–2648.
- (79) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (80) Kocak, A.; Erol, I.; Yildiz, M.; Can, H. Computational Insights into the Protonation States of Catalytic Dyad in BACE1–Acyl Guanidine Based Inhibitor Complex. *J. Mol. Graphics Model.* **2016**, *70*, 226–235.
- (81) Kocak, A. HBGA Binding Modes and Selectivity in Noroviruses upon Mutation: A Docking and Molecular Dynamics Study. *J. Mol. Model.* **2019**, *25*, 369.
- (82) Kocak, A.; Yildiz, M. Docking, Molecular Dynamics and Free Energy Studies on Aspartoacylase Mutations Involved in Canavan Disease. *J. Mol. Graphics Model.* **2017**, *74*, 44–53.
- (83) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the Analysis of Free Energy Calculations. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 397–411.
- (84) Hansson, T.; Marelus, J.; Åqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27–35.
- (85) Rifai, E. A.; van Dijk, M.; Geerke, D. P. Recent Developments in Linear Interaction Energy Based Binding Free Energy Calculations. *Front. Mol. Biosci.* **2020**, *7*, No. 114.
- (86) Åqvist, J.; Hansson, T. On the Validity of Electrostatic Linear Response in Polar Solvents. *J. Phys. Chem. A* **1996**, *100*, 9512–9521.
- (87) Hansson, T.; Åqvist, J. Estimation of Binding Free Energies for HIV Proteinase Inhibitors by Molecular Dynamics Simulations. *Protein Eng. Des. Sel.* **1995**, *8*, 1137–1144.
- (88) Khan, Y. S.; Gutierrez-de-Teran, H.; Boukharta, L.; Åqvist, J. Toward an Optimal Docking and Free Energy Calculation Scheme in Ligand Design with Application to COX-1 Inhibitors. *J. Chem. Inf. Model.* **2014**, *54*, 1488–1499.
- (89) Almlöf, M.; Carlsson, J.; Åqvist, J. Improving the Accuracy of the Linear Interaction Energy Method for Solvation Free Energies. *J. Chem. Theory Comput.* **2007**, *3*, 2162–2175.
- (90) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method for Determining Free Energies of Hydration. *J. Phys. Chem. B* **1995**, *99*, 10667–10673.
- (91) Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding Constants of Neuraminidase Inhibitors: An Investigation of the Linear Interaction Energy Method. *J. Med. Chem.* **1999**, *42*, 5142–5152.