Supporting Information

Combining Force Fields and Neural Networks for an Accurate Representation of Chemically Diverse Molecular Interactions

Alexey Illarionov[1], Serzhan Sakipov[1], Leonid Pereyaslavets[1], Igor V Kurnikov[1], Ganesh Kamath[1*], Oleg Butin[1], Ekaterina Voronina[1], Ilya Ivahnenko[1], Igor Leontyev[1], Grzegorz Nawrocki[1], Mikhail Darkhovskiy[1], Michael Olevanov[1], Yevhen K. Cherniavskyi[1], Christopher Lock[1,7], Sean Greenslade[1], Subramanian Sankaranarayanan[3,4], Maria Kurnikova[2], Jeffrey Potoff[5], Roger D Kornberg[6], Michael Levitt[6], and Boris Fain[1*]

[1] InterX Inc. (a subsidiary of NeoTX Therapeutics LTD), 805 Allston Way, Berkeley CA, 94710, USA
[2] Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA, 15213, USA
[3] Center for Nanoscale Materials, Argonne National Lab, Argonne, IL, 604391, USA
[4] Department of Mechanical and Industrial Engineering, University of Illinois, Chicago, IL, 60607, USA
[5] Department of Chemical Engineering and Materials Science, Wayne State University, Detroit, MI 48202, USA
[6] Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94304, USA
[7] Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Palo Alto, CA 94304, USA

* Corresponding authors: boris.fain@interxinc.com, ganesh.kamath@interxinc.com

SI 1.1 Molecular descriptors.

The behavior and properties of non-strongly-correlated molecular systems (e.g. solvation free energies and ligand-protein binding), are primarily determined by intermolecular interactions. Trying to simultaneously represent the spectra of atomization, bonded interactions, and intermolecular ones can wash out the accuracy of the last set as their energetic range is lowest. Therefore, to fit QM dimer interaction energies in our NN model we choose descriptors assigned to each *intermolecular* atom-atom contact within a certain distance cutoff (currently 5 Å). These differ from commonly used atom-centered descriptors that describe the environment of each atom in the molecular system and are thus more appropriate for accurate fitting of *intramolecular* energies and describe properties such as torsional barriers and vibrational spectra. Furthermore, intermolecular interactions depend primarily on the geometries of intermolecular contacts, to a lesser extent, on the electron-density changes caused by the

monomer distortions, mostly adjacent, of the interacting molecules. Consequently, our interaction fingerprint (Fig. 2a) is coupled to these two geometric factors.

We design the descriptors or Atom Pair Symmetry Functions (APSF) to represent an atom-atom contact (an intermolecular bond). They consist of 2/r_ij function (r_ij - the distance of the contact) and sum of terms dependent on relative coordinates of atoms of the contact and environment atoms covalently bound to atoms of the contact versus geometrical center $\underline{X}$ of the contact (see Fig. 2a).   APSF are obtained by summing over the participating atoms j,k in the following manner:

$$p_{nl} = \frac{2l+1}{4\pi}\sum_j\sum_k g_{n-l,l}(|\underline{r}_j - \underline{X}|)g_{n-l,l}(|\underline{r}_k - \underline{X}|)P_l(\cos \gamma_{jk})$$

where $0 \le n \le n_{max}$, $0 \le l \le l_{max}$,

P_l is the Legendre polynomial of order l and γ_{jk} is the triplet angle between atoms i, j and k, radial basis functions $g_{nl}(r)$ are linear combinations of spherical Bessel functions of the first kind $j_l\left(r\frac{u_{ln}}{r_c}\right)$ with expansion coefficients chosen satisfy the conditions of orthogonality and zero first and second derivatives of $g_{nl}(r)$ vs r at the cutoff radius r_c [1,2]. By construction these descriptors are invariant to permutations of atoms of the atomic contact and environment atoms. For systems considered in this paper we used spherical Bessel descriptors with $n_{max}=l_{max}= 6$, but we use them to encode pair-specific intermolecular interactions (APSF in Fig 2a). The spherical Bessel-based APSF are twice-differentiable with respect to atomic positions, which is necessary to make forces continuous. They also satisfy the condition for optimal completeness [Kocer, et al., 2020] which permits us to encode a local atomic environment with fewer descriptors [Kocer, et al. 2020]. We avoid the sensitivity to the cutoff radius r_c by including a transition λ-layer that smoothly tapers the correction to zero beyond a chosen distance between the two atoms.

SI 1.2 Pair Interaction Neural Network.

The generated APSF are fed into a neural net specific to each typed interaction (Fig 2b illustrates this approach for water dimers). For example, interactions between an aromatic carbon CA and a water hydrogen HW are specific to the (CA HW) pair. By construction the (HW CA) interaction is handled by the same network. For more complex molecules, e.g toluene C6H5-CH3 - water dimer with different carbon (aromatic CA/methyl C3) and hydrogen (HCA in benzene and HW in water) types, there will be several interaction pairs (CA-HW, C3-HW, etc.)

In the ARROW-NN methodology the energy of a molecular system appear as a sum of E_{FF} term computed with the analytical Arrow2 force field and neural network term E_{NN} computed for close intermolecular atom contacts (atom pairs) of the system.

$$E_{ARROW-NN} = E_{FF} + E_{NN} \tag{1}$$

$$E_{NN} = \sum_{ab}E_{NN}^{ab}, \tag{2}$$

where E_{NN}^{ab} - NN energy correction term for the intermolecular atom pair A-B.

E_{NN}^{ab} is computed by feedforward Neural Network taking as an input an array of APSF described in section SI 1.1.

$$E_{NN}^{ab} = \sum f_2(W_2(f_1(W_1 \underline{x}_{ab}) + \underline{b}_1)) + \underline{b}_2). \tag{3}$$

$\underline{x}_{ab}$ - is an input array of APSF functions for the intermolecular atom pair **ab**.
W_L and $\underline{b}_L$ - matrix of weights and biasing vector for L-th layer of the Neural Network.
f_L - vector of activation functions (tanh) for the L-th layer of the Neural Network.
In the current version of NN, 2 hidden layers each with 20 elements were used.
ARROW accurately describes intermolecular interactions at large distances, and so we would like to apply the neural network corrections to short range only (Fig 2b center). This also speeds up the calculations. Although the symmetry functions do decay to zero, we prefer to truncate the correction in the neural net itself. The output of the 2-layer perceptron (ΔE) is followed by a scaling lambda layer designed to smoothly suppress the output (correction to the analytical energy) to zero beyond a chosen cutoff:

$$E_{SL} = E_{NN}^{ab} \frac{1}{2} \tanh(- \alpha(R_{ab} - R_0)) \quad , \tag{4}$$

where R_{ab}- atom-atom distance of the atom pair, R_0- cutoff distance (5 Å) , $\alpha = 5^{\text{Å-1}}$ - distance decay parameter.

Elements of W_L and $\underline{b}_L$ are parameters of the Neural Network and adjusted to create an energy agreement EQM = E_{FF} + E_{NN} on molecular dimers using backpropagation algorithm and Adam stochastic gradient optimization (Kingma et al., 2014).[3] W_L and $\underline{b}_L$ are the same for atom pairs of a given type (having the same ARROW types of atoms defining the atom pair). This preserves the symmetry of NN energy correction E_{NN} versus atom permutations. For example, the network shown in Fig 2b that represents an interaction correction between two water dimers has 9 subnets corresponding to 9 intermolecular atom contacts, but only 3 distinct ones (HW-HW, OW-OW, HW-OW).

SI 1.3 Training dataset(s) generation.

The NN corrections to the ARROW force field are trained on multiple conformations of molecular dimers. To cover the space of intermolecular orientations we combined several procedures to generate dimer geometries - 'grid', 'liquid' MD and 'vacuum' MD dimers.
For a given molecular dimer a regular grid of molecular conformations is generated by the CamCASP program [4] which produces a uniform sampling of the relative orientation and distance of one monomer *vs* another.

Internal geometries of monomers are kept fixed at the respective ground states.
'Liquid' MD dimers for molecules A and B are sampled from MD simulations of A solvated in a box of B monomers and vice versa. As atom replicas in PIMD calculations sample significantly different geometries than atoms in MD simulations, we augment the 'liquid' set with one taken from analogous PIMD calculations (about 20% of total). We select dimers by extracting from simulation trajectories those that have the closest-atom monomer-monomer distance in the

3

range of 1.5 to 5 A. To avoid gaps and biases in covering the Euclidean group the extracted MD dimers are also filtered to be within a certain threshold value of RMSD to the grid dimers. The value is chosen so that the coverage of grid dimers by MD dimers is no less than 80%. To further ensure a uniform sampling of grid dimer space by MD dimers, a cap is set that limits the number of MD dimers that are associated with a given grid dimer.

The purpose of the 'vacuum' MD dimers is to augment the 'liquid' MD dimers by generating support for the part of the conformational (i.e. training) space that is missed by grid dimers and/or by liquid MD dimers before RMSD filtering. These are usually high-overlap high repulsive energy configurations. Vacuum MD dimers are generated by running MD simulations of molecule A in a box containing a few molecules B, but the non-bonded interactions between A and B are either completely removed, or reduced down to 1-10% of true to sample significant overlap regions. For convenience this is facilitated by running TI annihilation simulation, but only for lambdas=0.9-0.99. The 'vacuum' dimers are then extracted from the trajectories. As there are energies so high that they will never be visited at room temperature and pressure, we further filter out those possessing an ARROW intermolecular energies of greater than 40 kcal/mol. We use the silver standard, described below to compute the energetics of all the dimers generated for the training sets to be used as inputs in the neural nets (ARROW-NN). The trained interaction model ARROW-NN is very close to the reference QM energies.

Dimer(s)	Training MAE (kcal/mol)	Test MAE (kcal/mol)	Training Set	Test Set	Energy Range (kcal/mol)
H2O - H2O	0.014	0.016	70400	17600	[-6,6]
Acetate - Phenol/Benzene	0.14	0.2	129202	32301	[-30.4, 34.3]
Acetate-H2O	0.08	0.075	57600	14400	[-20.0,5.3]
N-methyl-acetamide-H2O	0.07	0.06	52000	13000	[-8.1,2.2]
Acetamide-H2O	0.05	0.045	60000	15000	[-10.1,-7.5]
Napthalene-H2O	0.042	0.031	21524	5381	[-3.5,3]
Pyrrole-imidazole	0.05	0.047	25600	6400	[-2.5,2]
Pyrrole-toluene	0.042	0.039	28000	7000	[-2.0,3.0]
pyrrole-guanidinium	0.083	0.076	25600	6400	[-1.7,2.5]
thiophene-ethanol	0.072	0.066	28000	7000	[-2.0,4.0]
chlorobenzene-toluene	0.083	0.077	28000	7000	[-3.5,4]

chlorobenzene-propane	0.06	0.054	24000	6000	[-3.0,3]
chlorobenzene-NMA	0.065	0.059	24000	6000	[-4.5,5]
chlorobenzene-acetate	0.11	0.103	28000	7000	[-12,6.2]
fluorobenzene-acetate	0.121	0.115	32000	8000	[-10,4.2]
Li+ - H2O	0.0347	0.0350	55934	13984	[-34.7;13.8]
Na+ - H2O	0.01776	0.01839	42806	10702	[-24.0;9.5]
Cl- - H2O	0.01829	0.01884	40342	10085	[-17.6;8.9]

Table SI 1 Representative MAE for various dimer pairs. Also provided is the training/test set data and the energy range for the various interactions. The energy range corresponds to [min-max] dimer energetics obtained from QM-silver standard.
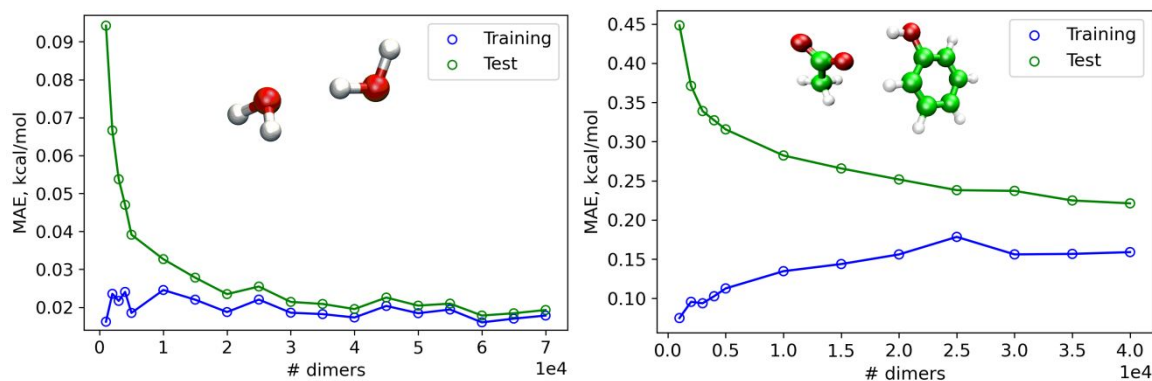
Figure SI 1 Training and Test MAE for water-water, left panel, and acetate-phenol, right panel, NN vs. number of dimers used for training. Train/Test split was set to 80/20%. Total number of dimers generated for water-water dimers is 88K and 48K for acetate-phenol.

In Figure SI 1 we show a convergence of MAE of NN as a function of the number of water-water and acetate-phenol dimers used for training. These two cases are representatives of the extreme cases from "simple" to "complicated" to correct interaction with. For both pairs the convergence is reached for a number of dimers much smaller than used in the production. Overall the convergence plots show "healthy" convergence dynamics. The lower the number of training data the lower the training MAE and higher the testing MAE, i.e. classical overfitting regime. The increase of the training dataset increases and stabilizes the training MAE and lowers the test MAE, both MAEs converging to certain values.

SI 2 Implementation of NN inference in the MD package Arbalest.

The total overhead of calculating the NN correction to the analytical energy is approximately the same as the ARROW polarizable force-field itself.

The calculation of NN corrections is performed in the program code independently of the calculation of FF energies. Such a separation was made for several reasons:

- As different interaction cut-off radii can be chosen for FF and NN, separate atomic pair lists speed up the calculations.
- The typification of atoms for the calculation of NN-corrections may potentially be different than for the calculation of FF-energies (see the typification paragraph in the main text). If it is so, splitting lists gives additional flexibility in their processing.
- For technical or other purposes, NN corrections may be applied not to the entire molecular system, but only to its part(s).
- To increase the efficiency of using neural network libraries such as TensorFlow [https://www.tensorflow.org], it is advisable to calculate NN corrections for large chunks of initial data (sets of descriptors, or APSF), which requires grouping atomic pairs by NN types and temporarily storing the calculated descriptors for each of the pairs.
- Calculation of FF energies and NN corrections can be performed simultaneously in parallel.

The calculation of the spherical Bessel descriptors proceeds in the manner described in [1,2]. Our APSF routines have the following capabilities:

- Calculation of APSF spatial gradients.
- The possibility of temporary storage and reuse of intermediate results for recurrence relations to avoid their recalculation.
- Separate recurrence relations have been replaced by their direct analytical counterparts where this did not lead to a noticeable rounding error, including the case when simulation is performed in single precision.
- Code for calculating the descriptors on the NVIDIA GPU.

In addition to the Bessel functions a calculation of descriptor functions (e.g. 1/r) based on the Coulomb potential were added to the set of descriptors. In our opinion this allows neural networks to better capture the spatial scale of interactions.

 Thus, the procedure for calculating NN-energies consists of the following steps:

- Creation of pair lists of atoms for which pairwise NN-energies will be calculated. For each pair of NN-types of atoms, a separate list is created. Lists are filled based on the distance between atoms, with some margin relative to the cutoff radius of NN interactions. This allows lists to be reused over several steps of molecular dynamics without recreating them.
- Calculation of descriptors for interacting atomic pairs (pairs falling out of the cutoff radius at a current step of the dynamics are skipped). For each pair, APSF based on spherical Bessel functions take into account both the interacting atoms themselves and their nearest [covalent] environment. Thus, for different pairs, the number of atoms in such "clusters" can be different. In addition to descriptors, their spatial gradients are also

calculated. Like NN-pairs, the calculated APSF are stored in separate lists for each pair of NN types.

- Calculation of NN-corrections and their derivatives. This calculation in the current version of the code is performed using the TensorFlow library. Each library call is made using one neural network, i.e. for one descriptor list (or a chunk of it if the list is too long). On the one hand, this minimizes the library initialization overhead, and on the other hand, it allows to perform several calculations in parallel for different pairs of NN types. In the latter case, a separate procedure performs load balancing of computational modules.
- Use of the calculated NN-corrections and their derivatives to obtain the energies of atoms and contributions to the forces acting on them. The NN energies correspond directly to the NN corrections. [Additional] NN-forces on atoms are calculated as the products of the derivatives of the NN-correction with respect to descriptors and the spatial gradients of these descriptors (with the appropriate sign).

Method	Speed (steps per second)		Num Mols/atoms	Processor
ARROW (MD)[+]	77.16	(/ 1)	1092 mols H2O	GPU 3080 TI
ARROW-NN (MD)[+]	39.984	(/ 1)	1092 mols H2O	GPU 3080 TI
ARROW-NN (MD)	17.384	(/1)	27 K atoms CDK2+water	GPU 3080 TI
MBPol (our estimate LAMMPS)	2.31	(/ 4 = 0.58)	256 mols H2O	4 cores
PEANN*	500	(/ 8 = 62.5)	64 mols H2O	Unknown

[+]Arbalest speeds include free energy features that makes it more expensive than comparable water models regular MD. e.g. typical calculation for Gromacs for free energy is ~3 times slower than regular MD.
*Exact simulation protocol is unknown (i.e. number of cores). Accuracy of the water model is unknown. Long range cutoff is a short 7.75 Å without PME or any long range correction[5].
***Step per second per core

Table SI 2: Comparative speeds of different water models.

SI 3 ARROW2 force field functional form and QM levels.

The ideas behind ARROW [6,7] and its predecessors QMPFF [8–11] are as follows: to create a functional form that is sufficiently complex to fit QM energies well; to include polarization to allow one to obtain many-body contributions from relatively small QM calculations and therefore have transferability to bulk; similarly, for economy of QM calculations, to have a lightweight monomer-based parameterization for the various linkers that connect functional groups together. ARROW

is atom-typed for the same reasons that all other analytical FF's are, e.g. an alkene and alkane carbons have different enough atomic properties to necessitate a separate model for each. Some of the less important features such as diffuse charge densities in both ES and EX assist in fitting the QM PES better.

The many-body energy can contribute ~20% of the total energy; a breakdown of the energies is on pages 183-184, Fig. 10.1 and Fig. 10.2 of The Theory of Intermolecular Forces (Second Edition) by Anthony Stone.[12]. Therefore, an accurate representation of polarizability / induction is a necessary component of reproducing the quantum potential energy surface of molecular ensembles. It allows the transferability of models parameterized on small molecular clusters (i.e. dimers, trimers) to ensembles of arbitrary size. A proper description of polarization is also needed to reproduce systems' behavior under applied electric fields. In ARROW the polarization is anisotropic which aids in reproducing the many-body components properly (Figure 3b (right)). Because the local charge distributions are responsible for inducing a response in neighboring atoms electron density, the atomic multipoles introduced in force-fields like QMPFF, AMOEBA and ARROW also aid in getting the many-body energy components right.

Elsewhere we [6,7] have discussed the advantages of parameterizing a Force Field solely from ab-initio calculations. Because our group has identified many of the factors needed [6,7,13,14] to do this accurately, we are now able to conclude that some empirical adjustment is still required. Specifically, the screening of the dispersion interaction in the intermediate and far range is a many-body phenomenon and is difficult to obtain from ab-initio calculations. Therefore, we take the close-range dispersion interaction from a decomposition of dimer QM calculations, and then we smoothly scale it down by a factor that is a property of the two interacting elements [15]. The treatment of the dispersion interaction[15] in the intermediate range follows with a crossover at 4 Å from the vacuum form to a screened asymptotic coefficient of 0.5 * C_6 of vacuum. We feel that in order to be more accurate the screening model may need to be refined for various bulk environments, or perhaps, by some empirical fits to experimental values of density etc. However, in our work, we took the universal non-empirically derived screening values suggested by Fiedler *et al.* [15] which were adequate.

Generally the proper treatment of dispersion is a complex and evolving field, and is still not fully settled. The reader may benefit from reviews [16–21] on the topic. We used almost the simplest possible form, similar to many common FF's. Some of our thoughts on this choice are as follows:

- While we do not have an explicit, many-body dispersion (DS), some of it is implicitly included in the FF as some of the major many-body DS contributions (Axilrod-Teller) come from chemically bonded atom pairs + 1 non-bonded atom which included in CCSD(T) calculation. This is true for all ab-initio parameterized FF's and even for empirically parameterized ones that utilize functional groups and atom types.
- Nonetheless, in our tests, despite the explicit representation of the DS many-body contributions, the agreement to total energies of multimers is below 2%. Perhaps it is not

0.2% as it is for the 2-body (2-molecule rather) term exactly because of the discrepancy in the description of many-body DS.
- We use a fairly standard, atom-typed DS functional form, with the simplest possible modification for medium-long range screening. The DS component is essentially an isotropic 'garbage collector' of the total CCSD(T) energy minus the ES and EX and IND components extracted from DFT-SAPT
- While not addressed in this manuscript, the biggest clear error in dispersion for us is not typifying aromatic carbons with 3 heavy-atom neighbors (fused rings) to have some dispersion anisotropy (they clearly do). Hyper, non-linear polarizability may also be worth considering.

The QM data underlying the force-field ARROW and ARROW-NN parameterization has a direct impact on the quality of the final predictions. In ARROW and ARROW-NN the total dimer energies are calculated with silver standard [22] i.e. MP2/CBS, calculated with Helgaker cubic extrapolation [23] from aug-cc-pVTZ->aug-cc-pVQZ + CCSD(T)/aug-cc-pVDZ - MP2/aug-cc-pVDZ. To improve transferability and ease our optimization we use DFT-SAPT[24–26] decomposition with dHF correction [25] at aug-cc-pVTZ level with PBE0 functional asymptotically corrected[25]. We employ four parts of DFT-SAPT decomposition which have corresponding manifestation in our FF : ES (electrostatic E1pol), EX (exchange-repulsion E1exch), IND (induction, E2ind + E2ind-exch + δHF), DS (dispersion, E2disp+E2disp-exch + Esilver-standard - DFT-SAPTtotal_energy). The dispersion term accumulates all disagreements between total energy described by the "silver standard" and DFT-SAPT+δHF energies. We use the silver standard to compute the energetics of all the dimers generated for the training sets to be used as inputs in the neural nets (ARROW-NN).

SI 4 Many-body energy agreement for ion-water clusters
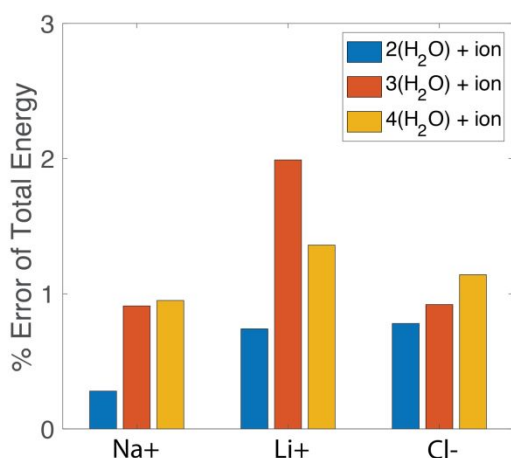
Figure SI 2 *Ion-water multimer non-additive energy errors.* In ion-water systems the error of the non-additive terms is below 2% of the total multimer QM energy.

SI 5.1 Water properties and simulation details.

The convergence of the test and training set is shown in Fig. SI 3. The water-water dimers are split into training and test sets (80% training set and 20 % test set). We train the neural network on about 70K water-water dimer interaction energies and obtain a MAE of 0.014 kcal/mol for the training set (red line). The trained neural network predicts a MAE of 0.016 kcal/mol for the resulting test set ~18 K water-water dimers (shown in blue), also see Table SI 1. The 2-body dimer interaction energies as a function of intermolecular distances is shown in Fig. 3b.

The dissociation energies for the various low-energy hexamers are computed using ARROW-NN and compared to MBPol predictions, see Table SI 3.

The water molecules were placed in a cubic box of size 32 X 32 X 32 $Å^3$. Particle-Mesh Ewald (PME) algorithm was used to compute the long-range electrostatic interactions. 32 x 32 x 32 grid mesh and 5th power spline interpolation order were used to compute the inverse PME sum. 9 Å cutoffs were used to compute the direct PME sum as well as the exchange and dispersion interactions. The trained neural network is read into ARBALEST. A NN cutoff radius of 4.5 Å is used in the simulations (close contact molecules interactions follow the NN). Beyond 4.5 Å the interactions are governed by the ARROW analytical force field. Bulk corrections were applied to account for distance cutoff of dispersion interactions. Solvation free energies were computed by decoupling the interactions between the solute and the solvent molecules. Electrostatic, exchange-repulsion and dispersion interactions were switched off simultaneously using a lambda-dependent scaling (a scaling factor power = 2 was used) and a soft-coring algorithm (maximal soft-coring radius = 1.5 Å, soft-coring radius scaling factor power = 1). 15 lambda points unequally spaced ($λ$ = 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0) were used to decouple the solute-solvent intermolecular interactions.

Energy minimization (10,000 steps) using the steepest descent algorithm was initially performed on all the simulation systems. This was followed by a 50 ps equilibration and 1 ns production runs in the isothermal-isobaric ensemble (NPT). The temperature was maintained at 298 K using a Nose-Hoover thermostat (chain length =6, relaxation time = 1ps). Pressure was maintained at 1 atm using a MTTK barostat (relaxation time = 5 ps). A Multiple Time Step (MTS) algorithm was used to integrate the equations of motion both for the classical and path-integral runs. Bonded and PIMD 8 beads interactions were integrated with the time step of 0.125 fs, while intermolecular interactions were integrated with a time step of 2 fs. All the calculations were performed with our in-house ARBALEST program. The intra- and intermolecular energies and forces are computed using NVIDIA GPUs, and integration of the extended equations of motion is done on the CPU. We compute the solvation free energies using both the Bennett acceptance ratio (BAR) and thermodynamic integration method with <dH/dλ> values interpolated by cubic splines. Further simulation details and protocols have been discussed at length in our previous publications[7][27] . One deviation, under current development and not in the current results, is scaling the NN correction in one or several

starting and ending λ-points to save computational effort. In the presented results the NN correction is scaled in unison with the FF as the total interaction energy.

The data is extracted from the simulated trajectories and the bulk water properties were computed. The density of water as predicted by ARROW-NN is within 0.1 % of the experiment. The heat of vaporization is underpredicted by ~ 2.5 % and the free energy of self-hydration is within 0.1 kcal/mol of experiment, see Table SI 4.

$(H_2O)_6$	E(CCSD(T)/CBS)	E(ARROW-NN)	E(MB-Pol)
Prism	45.92	45.80	45.73
Cage	45.67	45.26	45.46
Book 1	45.20	44.72	44.59
Book 2	44.90	44.27	44.36
Bag	44.30	43.89	43.71
Ring	44.12	43.57	43.30
Cyclic Boat 1	43.13	42.50	42.45
Cyclic Boat 2	43.07	42.35	42.48
MAE		0.49	0.53

Table SI 3. Dissociation energies for known hexamers of water [28,29]. All energies in kcal/mol.

H2O	Density (g/cc)	Hvap (kcal/mol)	Hydration (kcal/mol)
Experiment	**0.997**	**10.51**	**-6.30**
ARROW-NN (MD)	0.999	11.16	-6.59
ARROW-NN(PIMD8)	0.997	10.24	-6.20

Table SI 4: Neat properties and hydration/solvation of water as predicted by classical simulations and with inclusion of Nuclear Quantum Effects for ARROW-NN water model.

Source	NaCl	LiCl	Na+	Li+	Cl-
Schmid [30]	177.8	202.92	88.67	113.78	89.14

Friedman [31]	174.1	199.3	98.3	123.5	75.8
Tissandier [32]	174.0	199.28	101.29	126.58	72.70
Marcus [33]	168.49	194.79	87.23	113.53	81.26
Salomon [34]	172.7	198.1	72.4	97.8	100.3
Average Experiment	173.42±3.3	198.9±2.9	89.57±11.3	115.04±11.2	83.84±11.1
ARROW NN MD	176.11±0.65	201.73±0.75	92.65±0.35	118.27±0.51	83.46±0.55
ARROW NN PIMD8	176.74±0.64	201.23±0.79	92.08±0.37	116.57±0.60	84.66±0.52

Table SI 5: Experimental and calculated free energies for ion dehydration. All units are in kcal/mol. The numbers show that experimental data for salt with two ions together are more conservative (errors ~ 2 kcal/mol) than independent ions (errors ~11 kcal/mol), which means that it is better to compare total salt numbers.
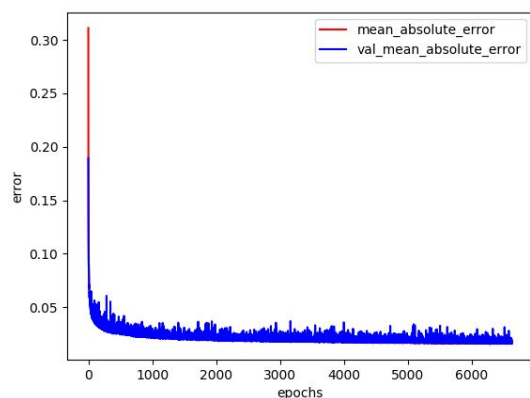
Figure SI 3 *The mean square error convergence plot for 88K water-water dimers*. The water-water dimers are split into training and test sets (80% training set and 20% test set). We train the neural network on about 70K water-water dimer interaction energies and obtain a MAE of 0.014 kcal/mol for the training set (red line). The trained neural network predicts a MAE of 0.016 kcal/mol for the resulting test set ~18 K water-water dimers (shown in blue).

SI 5.2 Protein - ligand simulation details

We use non-equilibrium methods to compute the relative binding free energies of ligands in protein. To ensure adequate sampling we apply a novel enhanced sampling technique - HREX with a conformation reservoir generated through potential softening and non-equilibrium MD. The protocols used for ARROW-NN binding free energetic computation are analogous to that for ARROW-FF. All the details pertaining to the relative binding free energies for the protein-

ligand systems including alchemical ligand transformations, enhanced sampling techniques used for the ARROW-FF and ARROW-NN are described at length in the Supporting Information of "Protein-ligand binding free energy calculations with ARROW - a purely first principle parametrized polarizable force field"[27]. All the inputs required to reproduce the results presented here are on github in a conda package https://github.com/freecurve/interx_arrow-nn_suite

As stated in the main text, we described the ligand-protein, ligand-water and protein-protein (within 6A of the binding pocket) interactions by ARROW-NN, and the rest by ARROW. In the MCL1 and the Thrombin systems these interactions were sufficiently well represented by the original ARROW. Nonetheless, energetic improvements did produce a slightly more accurate overall binding affinity prediction. We were already aware that the CDK2 system *requires* a more accurate description of certain interactions that ARROW is capable of. These are the energetically prominent mutations of 1h1q to 1oi9, 1h1r, and 1oiy ligands that mutate a benzene ring into phenol, chlorobenzene, and benzamide. We found that the dominant reason for the low accuracy of ARROW in these complexes is in the interaction of the ligand with the residue ASP87. We also showed that the FF energies of the phenol, benzamide, and chlorobenzene - acetate (aspartate analogue) dimers deviate highly from the QM values. Below, as an illustrative example, we describe in detail how these interactions were improved by ARROW-NN, and how the more accurate energetics bring about changes in microscopic behavior that result in a correct binding free-energy prediction.

Let us describe in depth the application of NN correction to ARROW for ligand 1oi9, i.e. correcting phenol-acetate interaction. Following the procedure described in SI1.3 we generated an initial training set of ~ 48,000 phenol-acetate 'liquid' dimers. We saw, however, that these dimers do not cover the whole configurational space. In other words, most of the dimers containing the hydroxyl group interacting with carboxylate group of the acetate that occur in the liquid phase occupy a conformational basin characterized by a 'classical' hydrogen bond, i.e. OPH…OOC distance in the range of 2.5 Å to 3.5 Å, and OPH-HOPH…OOC angle of 150 to 180. It became clear to us that conformations with lower hydrogen bond angles are omitted, due to the nature of the source of the dimers, i.e. such conformations are unlikely to occur in liquid phase. We therefore generated 'vacuum' dimers of phenol-acetate by simulating a box of 8-10 phenol molecules and an acetate molecule at low density with non-bonded interactions between acetate and phenol reduced down to 1-10%. This procedure allowed higher energy dimers to occur in the simulations. We then extracted from the resulting trajectories 14,000 'dimers' dimers that, along with the 'liquid' dimers, sample the configurational space around the hydroxyl and carboxyl groups more evenly.

We also established that in order to obtain better NN stability and accuracy, it is useful to present the neural net with other phenol types. This forces the model to better distribute the weights of the interactions among the phenol types and, i.e. between aromatic ring and hydroxyl group, and not overfit for one specific fragment. For this purpose we have added 50,000 benzene-acetate dimers to the NN training dataset.

We computed QM dimer energies of above mentioned 48,000 'liquid' and 14,000 'vacuum' dimers of phenol-acetate, as well as for the 50,000 'liquid' dimers of benzene-acetate. This dataset was then used to train NN for phenol-acetate interaction correction. The NN resulted in ARROW-NN Train MAE=0.14 and Test MAE=0.2, while ARROW MAE=0.68, see Figure SI 4 for MAE convergence plot.
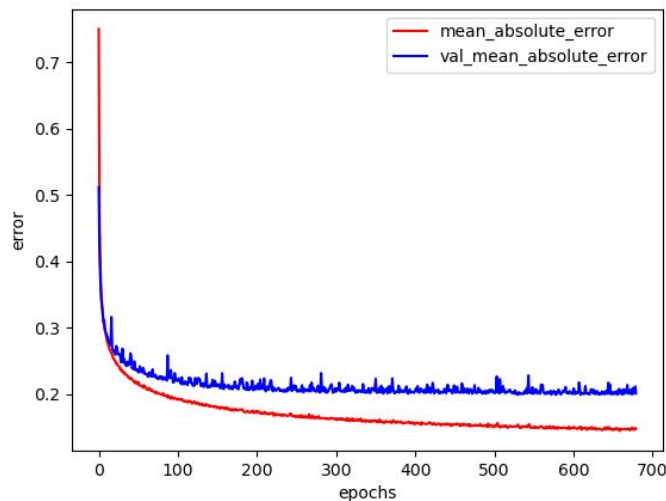

Figure SI 4: Convergence plot of NN training on Phenol/Benzene - Acetate dimers

As a first test of the NN correction we obtained 20 phenol-acetate dimers from the ARROW simulation of the CDK2 protein with 1oi9 ligand, i.e. minimum energy conformation monomers of acetate and phenol were overlapped onto the aspartate residue and phenyl of the 1oi9 ligand respectively. For these dimers we computed QM, ARROW and ARROW-NN energies. The errors relative to QM are as follows ARROW MAE = 1.01, ARROW-NN MAE = 0.19. Figure SI 4 shows a difference (QM - ARROW) and (QM - ARROW-NN) vs. QM energies.
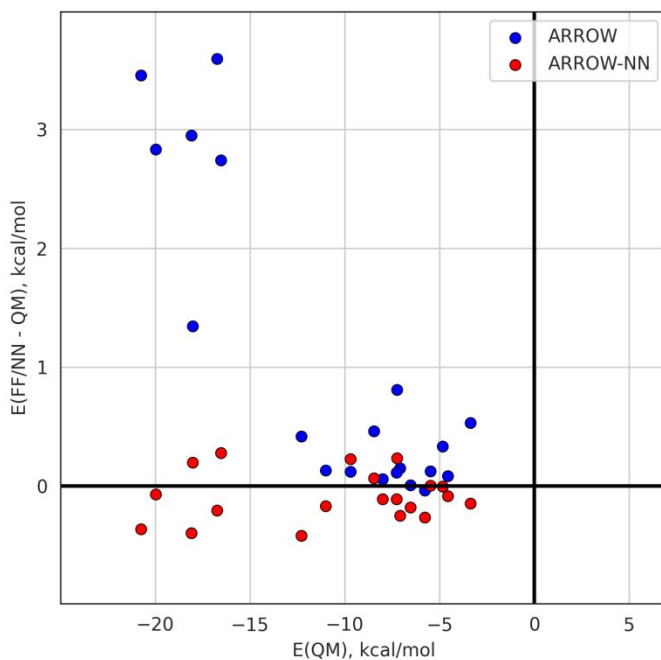
Figure SI 5: Difference in the interaction energy determined with ARROW/ARROW-NN and QM for ASP87 of CDK2 and 1oi9 (phenol) ligand.

Next, we have applied the NN corrected ARROW and simulated the CDK2 1oi9 ligand system. As was noted before the phenol side of the ligand does not form direct hydrogen bonds with aspartate in ARROW simulations. Application of NN correction to ARROW resulted in stabilization of the conformation, see Figure SI 6. One can see that one molecule of water mediates the interaction between Aspartate residue and Phenol of the ligand. Such conformation is observed in the 1oi9 crystal structure, see Figure SI 6.
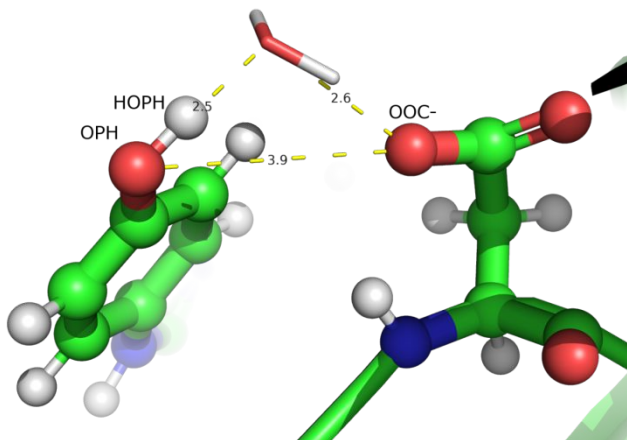
Figure SI 6: Representative Phenol - Acetate conformation observed in ARROW-NN simulation.
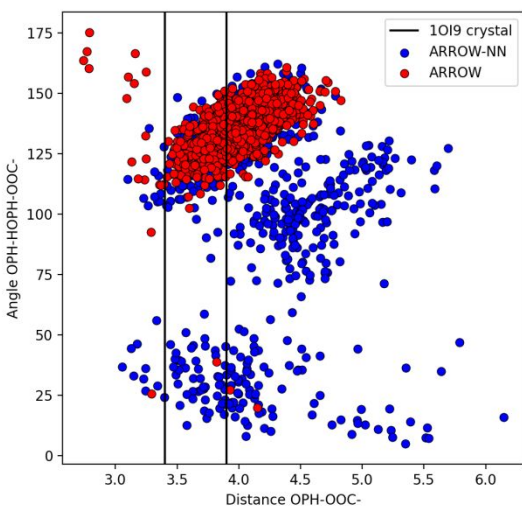
Figure SI 7: Phenol-Acetate hydrogen bond geometry, h-bond angle vs. distance, observed in ARROW, shown in red circles, and ARROW-NN, shown in blue circles, simulations. The black lines represent the polar contacts between the phenol oxygen (OPH) of the ligand and two oxygens of ASP87 observed in 1oi9 crystal structure.

Finally, we applied the NN corrected ARROW, first, to the 1h1q->1oi9 mutation, we used HREX with a reservoir from NEQ. The latter has provided the ΔΔG value of the mutation within ~0.3 kcal/mol error, see Table SI 6. While as stated before, the biggest change in free energies were

observed for corrections to the charged interactions as in the case of the CDK2 system for ARROW-NN. However such a significant change in free energies were not observed for MCL1 and Thrombin protein-ligand complexes, see Table SI 6.

CDK2 mutation	ΔΔG (exp)	ARROW	ARROW NN
1h1q->1h1r	0.51	-1.77	0.35
1h1q->1oi9	-1.56	0.29	-1.23
1h1q->1oiy	-1.61	0.32	-1.72
1h1q->20	-0.54	-1.35	-1.2
1h1q->21	0.35	-0.25	-0.3
1h1q->22	0.32	0.12	0.51
1h1q->26	-0.25	0.02	0.14
r	1	-0.50	0.88
MAE	1	0.81	0.33

MCL1 mutation	ΔΔG (exp)	ARROW	ARROW-NN
27->28	-0.5	-1.63	-1.13
27->30	-1.73	-1.26	-1.31
27->35	-2.69	-2.44	-2.41
27->38	-0.85	-1.34	-1.17
27->43	-0.91	-1.16	-1.19
27->46	-1.48	-2.25	-1.85
27->52	-3.11	-2.39	-3.41
27->36	-2.06	-3.11	-1.82
27->44	-2.55	-2.85	-2.65
27->42	-2.78	-2.67	-2.49
27->45	-2.83	-2.75	-2.95
27->41	-1.01	-1.12	-0.96
27->32	-0.46	0.15	-0.22
27->33	-0.76	-0.85	-0.86

27->37	-2.83	-2.75	-2.91
27->39	-0.91	0.84	-0.28
27->53	-3.84	-2.5	-3.28
r	1	0.75	0.94
MAE	1	0.47	0.28

Thrombin mutation	ΔΔG (exp)	**ARROW**	ARROW NN
5 -> 1a	-0.61	0.35	0.22
5 -> 3a	-1.32	-0.97	-0.76
5 -> 1b	-2.45	-1.75	-2.1
5 -> 6a	-3.1	-3.14	-2.97
5 -> 6e	-1.84	-2.6	-2.5
5 -> 7a	-0.82	-1.76	-1.22
5 -> 3b	-0.46	-0.59	-0.24
5 -> 6b	-2.23	-2.14	-1.82
r	1	0.81	0.90
MAE	1	0.525	0.405

Table SI 6: Relative binding free energy of ligands to CDK2, MCL1 and Thrombin.

SI 6 On the representation of the non-additive terms via explicit 3,4 and higher analytical functions vs. by the induction/polarization model.

Much excellent work [35] has been done on extremely precise short and intermediate range reproduction of the QM PES on water, and similar limited systems [36–38]. Most of these approaches utilize an explicit 3 and frequently 4 body representation of each many-body term [39]. We present a few thoughts on the pros and cons of this vs representing the many-body terms via a SCF polarizable model:

    i)   The feasibility of generating training sets for *ab-initio* parameterized models: The use of explicit terms for many-body non-additive interactions requires orders of magnitude more computational quantum mechanics to be performed. While large numbers of 4-mers of homogeneous H2O computed via, say, CCSD(T) are relatively easy to obtain, the situation changes drastically for a general all-purpose force-field whose construction requires much larger

fragments/molecules and a vast diversity of chemical space. Protein models need to describe the interaction of entities of the size of Tryptophan (10 heavy atoms) and the dipeptide backbone (9 heavy atoms), which is a weighty calculation for even a single dimer. Moreover, the parameterization of explicit many-body terms requires a significant number of training configurations, compounded by a *factorial* permutational factor for *heterogeneous* multimers. Therefore, extending the explicit accounting of many-body effects to cover all of chemistry will be very difficult. If, in contrast, one uses physics and a polarization model rather than brute force, one can extract fairly accurate (in bulk) non-additive energies from even *monomer* polarization calculations (see for example AMOEBA[40] and Stone-Misquitta[41]) and / or, as we do, from a proper decomposition of *dimerization* energies.

ii) Because polarization terms are iterated to self-consistency (SCF), formally they actually contain the major part of *all terms* of the many-body inductive expansion.

iii) This is a minor point, but much of the 3-*atom* non-additive energy is contained in the 2-body term of a typed force-field. Let us use the H2O molecule example familiar to the referees: as the terminal H is always adjourned by a neighboring O, the H-H interaction actually contains the H-H-O 3-atom term, etc. The argument generally extends to all typed atom pair interactions.

iv) Explicit 4 and 5 body terms will undoubtedly be computationally costly.

References.

(1) Kocer, E.; Mason, J. K.; Erturk, H. A Novel Approach to Describe Chemical Environments in High-Dimensional Neural Network Potentials. *J. Chem. Phys.* **2019**, *150* (15), 154102.

(2) Kocer, E.; Mason, J. K.; Erturk, H. Continuous and Optimally Complete Description of Chemical Environments Using Spherical Bessel Descriptors. *AIP Adv.* **2020**, *10* (1). https://doi.org/10.1063/1.5111045.

(3) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv [cs.LG]*, 2014. http://arxiv.org/abs/1412.6980.

(4) Misquitta, A. J.; Stone, A. J. Ab Initio Atom-Atom Potentials Using CamCASP: Theory and Application to Many-Body Models for the Pyridine Dimer. *J. Chem. Theory Comput.* **2016**, *12* (9), 4184–4208.

(5) Zhang, Y.; Hu, C.; Jiang, B. Accelerating Atomistic Simulations with Piecewise Machine-Learned Ab Initio Potentials at a Classical Force Field-like Cost. *Phys. Chem. Chem. Phys.* **2021**, *23* (3), 1815–1821.

(6) Pereyaslavets, L.; Kurnikov, I.; Kamath, G.; Butin, O.; Illarionov, A.; Leontyev, I.; Olevanov, M.; Levitt, M.; Kornberg, R. D.; Fain, B. On the Importance of Accounting for Nuclear Quantum Effects in Ab Initio Calibrated Force Fields in Biological Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (36), 8878–8882.

(7) Pereyaslavets, L.; Kamath, G.; Butin, O.; Illarionov, A.; Olevanov, M.; Kurnikov, I.; Sakipov, S.; Leontyev, I.; Voronina, E.; Gannon, T.; Nawrocki, G.; Darkhovskiy, M.; Ivahnenko, I.; Kostikov, A.; Scaranto, J.; Kurnikova, M. G.; Banik, S.; Chan, H.; Sternberg, M. G.; Sankaranarayanan, S. K. R. S.; Crawford, B.; Potoff, J.; Levitt, M.; Kornberg, R. D.; Fain, B. Accurate Determination of Solvation Free Energies of Neutral Organic Compounds from First Principles. *Nat. Commun.* **2022**, *13* (1), 414.

(8) Donchev, A. G.; Ozrin, V. D.; Subbotin, M. V.; Tarasov, O. V.; Tarasov, V. I. A Quantum Mechanical Polarizable Force Field for Biomolecular Interactions. *Proceedings of the National Academy of Sciences*. 2005, pp 7829–7834. https://doi.org/10.1073/pnas.0502962102.

(9) Donchev, A. G.; Galkin, N. G.; Illarionov, A. A.; Khoruzhii, O. V.; Olevanov, M. A.; Ozrin, V. D.;

Subbotin, M. V.; Tarasov, V. I. Water Properties from First Principles: Simulations by a General-Purpose Quantum Mechanical Polarizable Force Field. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (23), 8613–8617.

(10) Donchev, A. G.; Galkin, N. G.; Pereyaslavets, L. B.; Tarasov, V. I. Quantum Mechanical Polarizable Force Field (QMPFF3): Refinement and Validation of the Dispersion Interaction for Aromatic Carbon. *J. Chem. Phys.* **2006**, *125* (24), 244107.

(11) Donchev, A. G.; Galkin, N. G.; Illarionov, A. A.; Khoruzhii, O. V.; Olevanov, M. A.; Ozrin, V. D.; Pereyaslavets, L. B.; Tarasov, V. I. Assessment of Performance of the General Purpose Polarizable Force Field QMPFF3 in Condensed Phase. *J. Comput. Chem.* **2008**, *29* (8), 1242–1249.

(12) Stone, A. *The Theory of Intermolecular Forces*; OUP Oxford, 2013.

(13) Balog, E.; Hughes, A. L.; Martyna, G. J. Constant Pressure Path Integral Molecular Dynamics Studies of Quantum Effects in the Liquid State Properties of N-Alkanes. *J. Chem. Phys.* **2000**, *112* (2), 870–880.

(14) Yonetani, Y. Liquid Water Simulation: A Critical Examination of Cutoff Length. *J. Chem. Phys.* **2006**, *124* (20), 204501.

(15) Fiedler, J.; Walter, M.; Buhmann, S. Y. Effective Screening of Medium-Assisted van Der Waals Interactions between Embedded Particles. *J. Chem. Phys.* **2021**, *154* (10), 104102.

(16) Ambrosetti, A.; Ferri, N.; DiStasio, R. A., Jr; Tkatchenko, A. Wavelike Charge Density Fluctuations and van Der Waals Interactions at the Nanoscale. *Science* **2016**, *351* (6278), 1171–1176.

(17) Hermann, J.; DiStasio, R. A., Jr; Tkatchenko, A. First-Principles Models for van Der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chem. Rev.* **2017**, *117* (6), 4714–4758.

(18) Kleshchonok, A.; Tkatchenko, A. Tailoring van Der Waals Dispersion Interactions with External Electric Charges. *Nat. Commun.* **2018**, *9* (1), 3017.

(19) Stöhr, M.; Tkatchenko, A. Quantum Mechanics of Proteins in Explicit Water: The Role of Plasmon-like Solute-Solvent Interactions. *Sci Adv* **2019**, *5* (12), eaax0024.

(20) Xu, P.; Alkan, M.; Gordon, M. S. Many-Body Dispersion. *Chem. Rev.* **2020**, *120* (22), 12343–12356.

(21) Karimpour, M. R.; Fedorov, D. V.; Tkatchenko, A. Quantum Framework for Describing Retarded and Nonretarded Molecular Interactions in External Electric Fields. *Phys. Rev. Res.* **2022**, *4* (1), 013011.

(22) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Appointing Silver and Bronze Standards for Noncovalent Interactions: A Comparison of Spin-Component-Scaled (SCS), Explicitly Correlated (F12), and Specialized Wavefunction Approaches. *J. Chem. Phys.* **2014**, *141* (23), 234111.

(23) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. Basis-Set Convergence of the Energy in Molecular Hartree–Fock Calculations. *Chem. Phys. Lett.* **1999**, *302* (5), 437–446.

(24) Williams, H. L.; Chabalowski, C. F. Using Kohn−Sham Orbitals in Symmetry-Adapted Perturbation Theory to Investigate Intermolecular Interactions. *J. Phys. Chem. A* **2001**, *105* (3), 646–659.

(25) Misquitta, A. J.; Szalewicz, K. Intermolecular Forces from Asymptotically Corrected Density Functional Description of Monomers. *Chem. Phys. Lett.* **2002**, *357* (3), 301–306.

(26) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular Potentials Based on Symmetry-Adapted Perturbation Theory with Dispersion Energies from Time-Dependent Density-Functional Calculations. *J. Chem. Phys.* **2005**, *123* (21), 214103.

(27) Nawrocki, G.; Leontyev, I.; Sakipov, S.; Darkhovskiy, M.; Kurnikov, I.; Pereyaslavets, L.; Kamath, G.; Voronina, E.; Butin, O.; Illarionov, A.; Olevanov, M.; Kostikov, A.; Ivahnenko, I.; Patel, D. S.; Sankaranarayanan, S. K. R. S.; Kurnikova, M. G.; Lock, C.; Crooks, G. E.; Levitt, M.; Kornberg, R. D.; Fain, B. Protein-Ligand Binding Free-Energy Calculations with ARROW—A Purely First-Principles Parameterized Polarizable Force Field. *J. Chem. Theory Comput.* **2022**, *18* (12), 7751–7763.

(28) Bates, D. M.; Tschumper, G. S. CCSD(T) Complete Basis Set Limit Relative Energies for Low-Lying Water Hexamer Structures. *J. Phys. Chem. A* **2009**, *113* (15), 3555–3559.

(29) Herman, K. M.; Xantheas, S. S. An Extensive Assessment of the Performance of Pairwise and Many-Body Interaction Potentials in Reproducing Ab Initio Benchmark Binding Energies for Water Clusters N = 2-25. *Phys. Chem. Chem. Phys.* **2023**, *25* (10), 7120–7143.

(30) Schmid, R.; Miah, A. M.; Sapunov, V. N. A New Table of the Thermodynamic Quantities of Ionic Hydration: Values and Some Applications (enthalpy–entropy Compensation and Born Radii). *Phys. Chem. Chem. Phys.* **2000**, *2* (1), 97–102.

(31) Friedman, H. L.; Krishnan, C. V. Thermodynamics of Ionic Hydration. In *Aqueous Solutions of Simple Electrolytes*; Franks, F., Ed.; Springer US: Boston, MA, 1973; pp 1–118.

(32) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data. *J. Phys. Chem. A* **1998**, *102* (40), 7787–7794.

(33) Marcus, Y. Thermodynamics of Solvation of Ions. Part 5.—Gibbs Free Energy of Hydration at 298.15 K. *J. Chem. Soc. Faraday Trans.* **1991**, *87* (18), 2995–2999.

(34) Salomon, M. Thermodynamics of Ion Solvation in Water and Propylene Carbonate. *J. Phys. Chem.* **1970**, *74* (12), 2519–2524.

(35) Metz, M. P.; Szalewicz, K. Automatic Generation of Flexible-Monomer Intermolecular Potential Energy Surfaces. *J. Chem. Theory Comput.* **2020**, *16* (4), 2317–2339.

(36) Quintas-Sánchez, E.; Dawes, R. AUTOSURF: A Freely Available Program To Construct Potential Energy Surfaces. *J. Chem. Inf. Model.* **2019**, *59* (1), 262–271.

(37) Zhang, Y.; Xia, J.; Jiang, B. Physically Motivated Recursively Embedded Atom Neural Networks: Incorporating Local Completeness and Nonlocality. *Phys. Rev. Lett.* **2021**, *127* (15), 156002.

(38) Li, Y.; Zhai, Y.; Li, H. MLRNet: Combining the Physics-Motivated Potential Models with Neural Networks for Intermolecular Potential Energy Surface Construction. *J. Chem. Theory Comput.* **2023**, *19* (5), 1421–1431.

(39) Zhu, X.; Riera, M.; Bull-Vulpe, E. F.; Paesani, F. MB-pol(2023): Sub-Chemical Accuracy for Water Simulations from the Gas to the Liquid Phase. *ChemRxiv* **2023**. https://doi.org/10.26434/chemrxiv-2023-23v0k-v2.

(40) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114* (8), 2549–2564.

(41) Misquitta, A. J.; Stone, A. J. Distributed Polarizabilities Obtained Using a Constrained Density-Fitting Algorithm. *J. Chem. Phys.* **2006**, *124* (2), 024111.