Advancing Force Fields Parameterization: A Directed Graph Attention Networks Approach

Gong Chen,[*,†] Théo Jaffrelot Inizan,[‡] Thomas Plé,[‡] Louis Lagardère,[‡]

Jean-Philip Piquemal,[‡] and Yvon Maday[*,†]

†*Sorbonne Université, CNRS, Université Paris Cité, Laboratoire Jacques-Louis Lions (LJLL), UMR 7598 CNRS, 75005 Paris, France*

‡*Sorbonne Université, Laboratoire de Chimie Théorique (LCT), UMR 7616 CNRS, 75005 Paris, France*

E-mail: gong.chen@sorbonne-universite.fr; yvon.maday@sorbonne-universite.fr

Abstract

Force Fields (FFs) are an established tool for simulating large and complex molecular systems. However, parametrizing FFs is a challenging and time-consuming task that relies on empirical heuristics, experimental data, and computational data. Recent efforts aim to automate the assignment of FF parameters using pre-existing databases and on-the-fly *ab-initio* data. In this study, we propose a Graph-Based Force Fields (GB-FFs) model to directly derive parameters for the Generalized Amber Force Field (GAFF) from chemical environments and research into the influence of functional forms. Our end-to-end parameterization approach eliminates the need for expert-defined procedures and enhances the accuracy and transferability of GAFF across a broader range of molecular complexes. The GB-FFs model, which is only grounded on ab initio data, is implemented in the highly parallel Tinker-HP GPU package. Simulation results are compared to the original GAFF parameterization and validated on various experimentally and computationally derived properties, including free energies.

Introduction

The high numerical cost of quantitative *ab-initio* methods restricts their application to small systems, composed of a few hundred atoms,[1,2] limiting the ability to study larger molecular systems. As an alternative, force fields (FFs) have emerged as valuable tools, employing physically-motivated functional forms to model potential energy surfaces. These FFs are parameterized to match *ab-initio* as well as experimental data, offering a computationally cheaper alternative for simulating diverse systems, ranging from biology to polymers and complex materials. Indeed, there is a wide range of FFs developed for different purposes and compounds.

FFs can be categorized into two main families. The most commonly used FFs are known as classical, or non-polarizable, FFs, that include AMBER,[3,4] CHARMM[5,6] and GAFF.[7,8]

These non-polarizable FFs employ a combination of fixed-charge Coulomb potential and Lennard-Jones interactions to model intermolecular interactions. They are highly efficient numerically, enabling simulations of very large systems over long time scales.[9,10] However, their simple functional form lacks polarization and many-body effects, which are crucial for accurately describing complex phenomena such as pi-stacking or allosteric effects.[11,12]

On the other hand, there are polarizable force fields (PFFs) such as AMOEBA,[13,14] AMOEBA+,[15,16] CHARMM Drude,[17] and SIBFA.[18,19] These force fields have been specifically developed to explicitly incorporate polarization and many-body effects. This enhanced flexibility and accuracy come at a higher computational cost compared to non-polarizable FFs. Nevertheless, PFFs provide a more comprehensive representation of intermolecular interactions and are particularly suitable for studying complex systems.[20–23]

In recent years, significant attention and resources have been devoted to the development of Machine Learning Potentials (MLPs), aiming to bridge the accuracy and generality gap between FFs and *ab-initio* methods.[24–28] MLPs employ flexible functional forms from the field of Machine Learning (ML) to accurately fit *ab-initio* energies or forces. They offer a favorable balance between computational efficiency and accuracy, circumventing the need for empirical functional forms used in FFs. MLPs possess the ability to capture complex interactions, including polarization effects[29–31] and metal-ligand interactions,[32,33] which are challenging to model using traditional FFs.

However, the accuracy of MLPs depends on the quality of the training data and the architecture of the ML model, which can limit their transferability. Moreover, MLP models are often difficult to interpret, posing challenges in identifying and understanding the underlying physics and chemistry of the systems under study. Additionally, the use of MLPs versus FFs in molecular dynamics significantly slows down those simulations. Note that the FF parameters are assigned once for each system at the beginning of a simulation, and newer, more accurate parametrizations are regularly updated by FF developers.

Parameterizing a FF is however a challenging task since its accuracy and transferability

heavily depend on the quality of its parameters. This offline process is time-consuming, often taking years, as it relies on empirical heuristics, experimental data as well as computational data. FFs have established robust parameterization procedures, such as `Antechamber` for GAFF[8] and poltype2 for AMOEBA.[34,35] Additionally, these FFs rely on local frames, known as atom types or atom classes, to assign parameters (e.g., bonds, angles). To enhance the generalization and reliability of FFs, one tendency is to expand the atom type space. However, this leads to an increasing number of possible valence compositions, introducing significant complexity in the parameter fitting process. Moreover, even with modern parameter optimization frameworks[36] and sufficient data, FF parameters defined by fixed atom types can sometimes suffer from low transferability.

Advances in the computational efficiency and scalability of *ab-initio* methods have also provided new opportunities to enhance the transferability and accuracy of FFs by building larger and more accurate databases.[37,38] Consequently, there has been an increased effort to leverage ML for predicting FF parameters trained on these large *ab-initio* databases, while still maintaining the predefined functional form of the potential and atom types or classes. Wang et al.[39] combined Graph Neural Networks (GNNs) and automatic differentiation to predict FF parameters. By focusing on intramolecular interactions, they demonstrated that GNNs can effectively predict FF parameters based on potential energies.

Building upon these developments, we propose a Graph-Based Force Fields (GB-FFs) model for FF parameterization. GB-FFs automatically derive accurate FF parameters using basic atom features and bond features and aims to extend the generalization of FFs by using a Directed Graph multi-head Attention Network. It serves as a continuous alternative to traditional discrete atom typing schemes, eliminating the need for assigning atom types and obtaining FF parameters directly from atomic representations.

To assess the overall quality of our GB-FFs model, we compared its performance to the original GAFF parameterization using the highly parallel Tinker-HP GPU package.[40,41] Our model is freely available (MIT Licence) on GitHub at `https://github.com/GongCHEN-1995/`

`GB-FFs-Model`, and a tutorial is included for generating GAFF parameters from a smile, as well as for fine-tuning the model on a newer database.

In particular, this paper presents the following contributions:

- Enhancement of the existing GAFF by refining its parameters.
- Treatment of the molecules as directed molecular graphs and the use of a self-attention mechanism to aggregate information.
- Inclusion of charge transfer to predict fixed atomic charge, enabling $\mathcal{O}(N)$ complexity.
- Validation on various experimental, as well as *ab-initio*, properties, including hydration free energies of diverse molecular systems.
- Versatility and ease of use of the model facilitate its extension to other non-polarizable FFs.
- Investigation of GAFF's limits by modifying its functional forms.

Methods

In the following sections, we will provide a brief introduction to GAFF and present the GB-FFs model.

The General AMBER Force Field (GAFF)

GAFF[7] is among the most popular classical FFs to simulate organic molecules. It is an extension of the Amber force field.[4] GAFF is specifically designed to be compatible with a broad range of organic molecules, including drug-like compounds, carbohydrates, and nucleic acids.

GAFF incorporates a comprehensive set of parameters for bond stretching, angle bending, torsional, and non-bonded interactions. These parameters allow for accurate modeling and simulation of the behavior of organic molecules under various conditions (e.g high pressure, low temperature). Due to its computational efficiency, relative reliability and especially its simple functional form, GAFF has been widely implemented in many popular molecular simulation software packages, such as AMBER,[4] GROMACS,[42] CHARMM,[6] and Tinker-HP.[27,40,41] Finally, a main advantage of GAFF is the publicly availability of its parameters. Thus facilitating its widespread use within the scientific community.

In GAFF, the angle bending, bond stretching bonded interactions are modelled using a harmonic potential making it not reactive thus greatly simplifying the parameterization process. The torsional potential is expressed as a Fourier series. For non-bonded interactions, the Van der Waals (VdW) interactions are described by a 12-6 Lennard Jones potential. The electrostatic potential, on the other hand, is governed by Coulomb's law.

$$
\begin{aligned}
E_{potential} = & \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 \\
& + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\phi - \gamma)\right] \\
& + \sum_{i<j}\left[\epsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{R_{ij}^{12}} - 2\frac{\sigma_{ij}^{6}}{R_{ij}^{6}}\right) + \frac{q_i q_j}{\varepsilon R_{ij}}\right]
\end{aligned} \tag{1}
$$

where r_{eq} and θ_{eq} are equilibrium structural parameters ; K_r, K_θ, V_n are so called "force constants"; n is multiplicity and γ is phase angle for torsional angle parameters. The ϵ, σ, and q parameters characterize the non-bonded VdW potential. ϵ, σ follow Lorentz-Berthelot combination rules.[43,44] The GAFF parameters $\{K_r, r_{eq}, K_\theta, \theta_{eq}, V_n, \epsilon, \sigma\}$ are directly read from parameters table according to corresponding atom types while $n = 1, 2, 3, 4$ and $\gamma = 0$ or π.

The parameterization process of GAFF starts by assigning partial charges. In the early stages of GAFF, Hartree Fock (HF) with the 6-31G* basis set were used to generate electrostatic potentials from which restrained electrostatic potential (RESP) charge[45,46] fits were derived. This process proved to be expensive, especially for large molecules or large numbers of molecules and led to the development of the AM1-BCC charge scheme that approximate HF/6-31G* RESPA computation by first calculating charges using the AM1 semi-empirical model and correct it via bond charge corrections.[47,48]

In GAFF, the equilibrium bond length θ_{eq} are fitted through experimental data from X-ray and neutron diffraction, as well as MP2/6-31G* computations. On the other hand, bond angle parameterization uses reference from the Cambridge Structure Database, empirical rules and MP2/6-31G* computations. Finally, the strategy for developing torsional angle parameters involves performing torsional angle scanning and fitting the parameters to accurately reproduce the rotational profile obtained from MP2/6-31G* calculations. The VdW parameters are the same as those used by AMBER and thus extracted from a database.

In this paper we used the second generation of GAFF. But the GB-FFs general framework make it transferable to other type of non-polarizable FFs easily.

Graph-Based Force Fields model, a Universal Parameterization Procedure

GNNs have been proved to be an efficient and powerful way to detect chemical environment and to extract molecular properties.[49–52] In addition, GNNs also have shown potential in expressing atoms' representations and bonds' representations.[49,50,53]

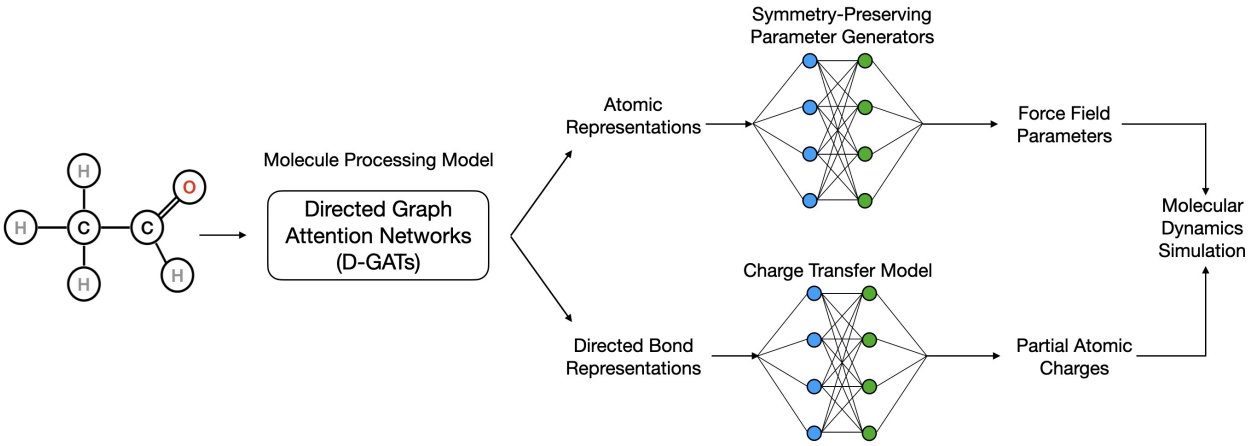


Figure 1: **Framework of Graph-Based Force Fields (GB-FFs) Model:** It consists of molecule processing model, the symmetry-preserving parameter generator and charge transfer model.

The model is composed of three modules: molecule processing model, symmetry-preserving parameter generator and charge transfer model (see Figure 1). These components will be discussed in the following sections.

GB-FF's runtime complexity is $\mathcal{O}(N)$ and processing a molecule with 50 atoms takes 0.018 seconds on a single GPU V100. In comparison, solely the AM1-BCC charge model used in `Antechamber` has a computational complexity of $\mathcal{O}(N^2)$ and takes 111 seconds to assign atom types and charges for the same molecule.

Molecules Processing Model

Assigning atom types and deriving FF parameters are typical atom-level tasks. Building on the notion of directed bonds,[54] we applied our recently introduced Directed Graph Attention Networks (D-GATs) model.[52] In contrast to other ML-based molecular processing models, D-GATs exhibit a remarkable ability to discern local chemical environments and eliminate unnecessary message flow. They have consistently outperformed state-of-the-art benchmarks in 13 out of 15 molecular property prediction tasks.

To enhance the robustness, we employ the Smooth Maximum Unit (SMU)[55] as an activation function. SMU offers a smooth approximation to the entire Maxout family, including ReLU, Leaky ReLU, and their variants.[56] Our goal is to predict a set of parameters that can bring molecular dynamic simulation results closer to *ab-initio* data. The molecular potential energy surface is highly sensitive to these predicted parameters. Our experiments have demonstrated that the discontinuity in the Maxout function can hinder the convergence of the models' loss.

Table 1: Input features to the GB-FFs model

Atom Features	Size(38)	Descriptions
atom symbol	11	[UNK,H,C,N,O,F,P,S,Cl,Br,I] (one-hot)
degree	6	number of covalent bonds [0, 1, 2, 3, 4, 5] (one-hot)
formal charge	7	[-3,-2,-1,-0,1,2,3] (one-hot)
hybridization	8	[unspecified, s, sp, sp2, sp3, sp3d, sp3d2, other] (one-hot)
chirality	4	[unspecified, tetrahedral_CW, tetrahedral_CCW, other] (one-hot)
ring	1	whether the atom is in ring [0/1] (one-hot)
aromaticity	1	whether the atom is part of an aromatic system [0/1] (one-hot)
Bond Features	**Size(12)**	**Descriptions**
bond type	4	[single, double, triple, aromatic] (one-hot)
conjugation	1	whether the bond is conjugated [0/1] (one-hot)
ring	1	whether the bond is in ring [0/1] (one-hot)
stereo type	6	[StereoNone, StereoAny, StereoZ, StereoE, Stereocis, Stereotrans] (one-hot)

To ensure compatibility with GAFF, we focused on compounds composed of C, N, O, H, S, P, F, Cl, Br, and I. Using the RDKit package,[57] we extracted fundamental atomic and bond features (see Table 1). These features, in conjunction with the molecular graph represented in Lewis structure,[58] were then input into the GB-FFs model. The model's outputs

include atomic and bond representations. For this article, we specifically employ the directed bond representations to incorporate chemical information and the atomic representations for predicting FF parameters (as illustrated in Figure 2)."

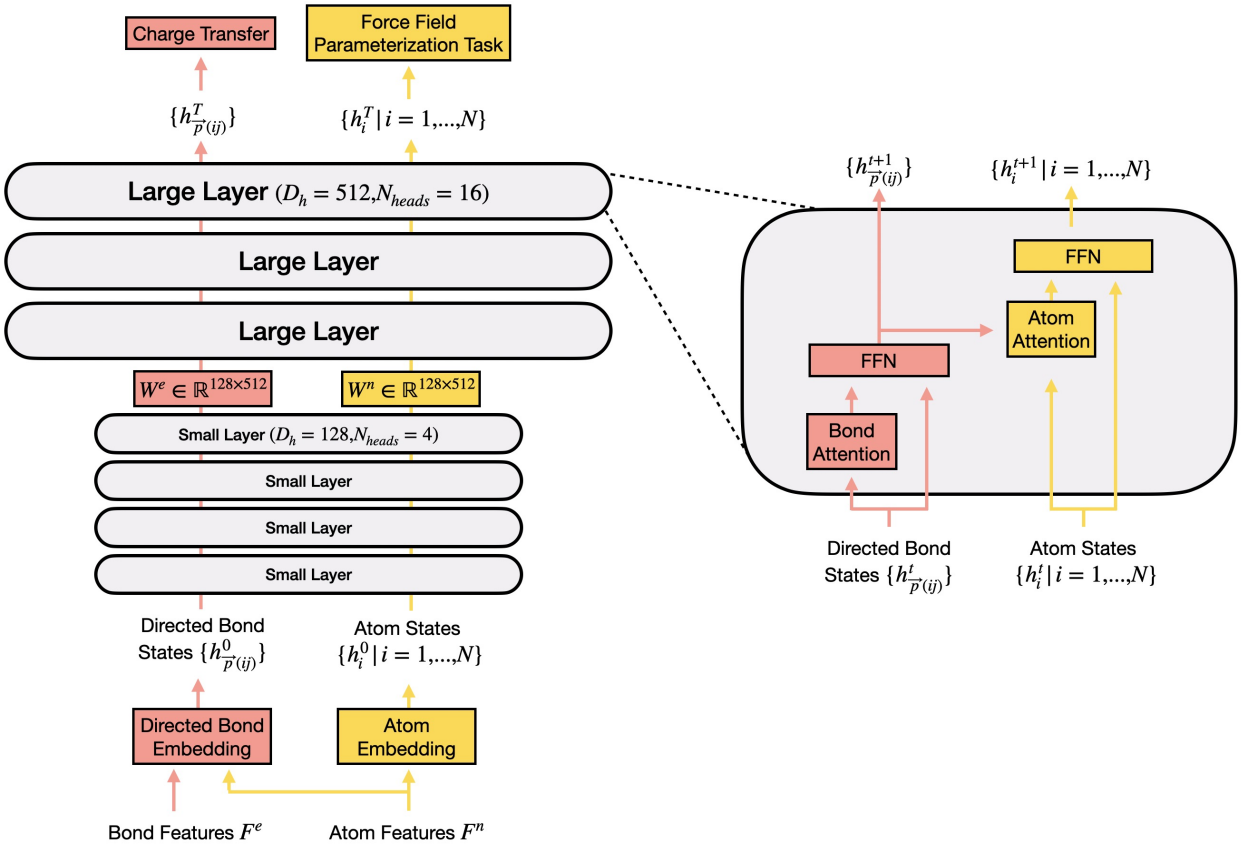


Figure 2: **Molecule processing model:** The model to process molecules follows the idea in D-GATs but with hierarchical structure. D_h is the dimension of model and N_{heads} is the number of heads in multi-attention mechanism. Between two stacked layers, there exists W^e and W^n to convert the dimension of embeddings. The stacked layers consist of several interaction layers. However, different from D-GATs,[52] here is no ReadOut function.

FFs are intricate and highly parameter-sensitive. To enhance the models' expressive capacity and expand their receptive field, we employ a hierarchical structure consisting of two stacked layers: Small and Larger Layers. Both layers share the same model architecture but operate in distinct dimensions. Between the two stacked layers, we incorporate linear transformations (W^e and W^n) to convert the dimensions of atomic and bond representations.

As depicted in Figure 2, the Large Layers consist of 3 interaction layers with a model dimension (D_h) of 512. They play a central role in detecting chemical environments and

forming atomic representations. The Small Layers, on the other hand, employ 4 interaction layers with a model dimension (D_h) of 128 and four attention heads. These layers are primarily used for embedding initialization and demand minimal computational resources.

We designate the output atomic representations as $h^T = h_i^T | i = 1, ..., N$, while the output directed bond representations are denoted as $h^T\vec{p}(ij)$ for all connected atoms i and atoms j. It's important to emphasize that the directionality in bond representation is critical, with $h^T\vec{p}(ij)$ indicating the bond from atom i to atom j.

Charge Transfer Model

To ensure the net charge of the molecule aligns with the actual scenario and to improve the physical meaning of charge distribution, we used directed bond states $\{h^T_{\vec{p}(ij)}\}$ to predict the charge transfer between connected atoms. Our molecular processing model is based on directed graphs, eliminating the need for additional operations, and it can predict the charge transfer from one atom to its neighbors.

As illustrated in Figure 3, the directed bond features obtained from Figure 2 are fed into a feed-forward neural network (FFN) to determine the charge transfer in the corresponding bond direction. The final atomic charge is obtained by summing the original formal charge and the incoming charges while subtracting the outgoing charges.

We used the AM1-BCC charge model[45,46] to calculate the original partial atomic charges and compared our predicted GB-FFs charges with both AM1-BCC charges and wB97x/def2-TZVPP MBIS charges. Our charge transfer model performs comparably to AM1-BCC charges, but with significantly improved efficiency. Detailed results can be found in Supplementary Information (Charge Comparison).

Symmetry-Preserving Parameter Generator

The number of FF parameters depends on the molecule's geometry. According to the molecular geometry, we use RDKit to list all the combinations of bonds, angles, dihedrals and

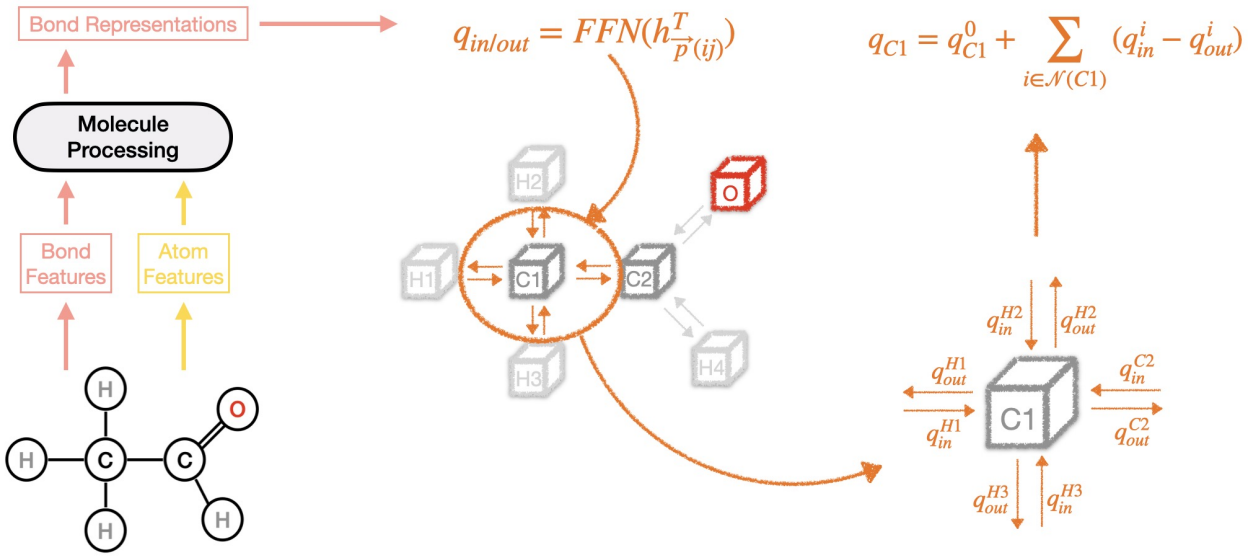


Figure 3: **Charge Transfer Model:** The charge is allowed to transfer between connected atoms and the charge in/out is directly calculated by the directed bond embeddings. The final partial charge of atom is the original formal charge plus charge flows in and minus the charge flows out.

non-bonded interaction pairs. We input atomic representations into the parameter generator based on the identified structure to predict all potential bonds, angles, dihedrals, and non-bonded parameters.

The parameter generator needs to ensure atom ordering symmetries. For example, when predicting bond parameters, if we exchange the order of two input atomic representations, the predicted parameters should be invariant. In the previous similar work, Espaloma,[39] the relevant equivalent atom permutations are enumerated, which has a cost. We split the input atom embeddings by their intrinsic structure and use the linear transformation to ensure its symmetry:

$$h_{r_{ij}} = h_{r_{ji}} = W_r h_i^T + W_r h_j^T \tag{2}$$

$$h_{\theta_{ijp}} = h_{\theta_{pji}} = W_{\theta 1} h_i^T + W_{\theta 2} h_j^T + W_{\theta 1} h_p^T \tag{3}$$

$$h_{\phi_{ijpq}} = h_{\phi_{qpji}} = W_\phi [h_i^T, h_j^T] + W_\phi [h_q^T, h_p^T] \tag{4}$$

$$h_{\varphi_{ijpq}} = h_{\varphi_{jipq}} = h_{\varphi_{jqpi}} = h_{\varphi_{qjpi}} = h_{\varphi_{iqpj}} = h_{\varphi_{qipj}} = W_{\varphi 1} h_i^T + W_{\varphi 1} h_j^T + W_{\varphi 2} h_p^T + W_{\varphi 1} h_q^T \tag{5}$$

$$h_{VdW_i} = W_{VdW} h_i^T \tag{6}$$

where $[.,.]$ denote concatenation and $h_r, h_\theta, h_\phi, h_\varphi \in \mathbb{R}^{D_h}, W_r, W_{\theta 1}, W_{\theta 2}, W_{\varphi 1}, W_{\varphi 2}, W_{VdW} \in \mathbb{R}^{D_h \times D_h}, W_\phi \in \mathbb{R}^{2D_h \times D_h}$. These embeddings for bond ($\{h_r\}$), angle ($\{h_\theta\}$), torsion ($\{h_\phi\}$), improper torsion term ($\{h_\varphi\}$) and VdW interaction ($\{h_{VdW}\}$) are in the same dimension. And the number of parameters for each term is fixed (for example, one bond term needs two parameters $\{K_r, r_{eq}\}$). According to the corresponding embeddings, we can use the fully connected NNs to predict the FF parameters (see Figure 4).

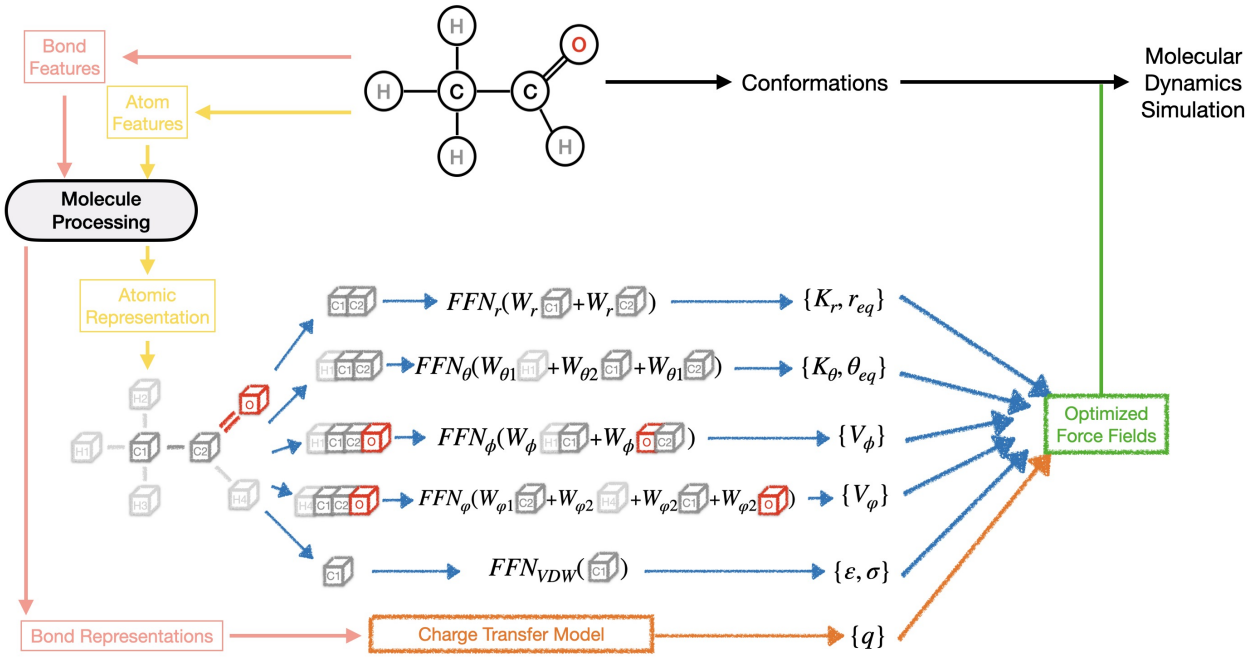


Figure 4: **Symmetry-preserving parameter generator:** For a specified molecule, we input atom and bond features to hierarchical D-GATs and obtain the atomic representations and directed bond representations. The symmetry-preserving parameter generator predicts all FF parameters, which can be used to do molecular dynamic simulation.

Results and Discussions

Training Strategies

To enhance our model's robustness and performance, it is first trained on the ANI-1 database, currently one of the largest available database of Density Functional Theory (DFT) computations for small organic molecules.[59] However, the *ab-initio* accuracy of the ANI-1 database is relatively low and we thus fine-tuned the model on two more chemically accurate databases: SPICE[60] and DES370K.[61]

The SPICE dataset is used to ensure the accuracy of fitting intramolecular interactions. SPICE is a collection of quantum mechanical data mainly built to train MLP for simulating drug molecules and proteins. The computations are performed at the ωB97M-D3(BJ) functional[62,63] with the def2-TZVPPD basis set.[64,65]

In contrast, the DES370K dataset is used to ensures the accuracy of intermolecular interactions. In DES370K, the reference interaction energies for these systems are computed using the highly accurate coupled-cluster singles and doubles with perturbative triples (CCSD(T))[66] level of theory with a complete basis set (CBS).[67] The complexes in this database represent most of the molecular interactions that could occur in chemistry, including electrostatic-dominated (hydrogen bonding), dispersion-dominated, and mixed (electrostatic/dispersion) interactions.

As GAFF includes only H, C, N, O, F, P, S, Cl, Br, and I, we exclude the molecules containing elements Li, Na, Mg, K, Ca in SPICE and DES370K. There are 29,389 compounds left. The compounds are randomly divided into training/validation/test sets, following an 8:1:1 ratio.

Fitting the potential energy and atomic forces simultaneously is challenging because, as a first remark, we do not know if there exists a unique global best fit, second if there exists local ones. Due to this, we have proposed a set of training strategies. We first performed a pre-training on energies and forces the large ANI-1 Database computed with

the ωB97x functional with the 6–31G(d) basis set. Once this pre-training is performed, we then fine-tuned it on the energies and forces from the SPICE and interactions energies from the DES370K database. More information can be find in Supplementary Information (Pre-training on ANI-1 Database and Fine-Tuning Strategy on SPICE and DES370K Databases). This second step allows us to capture not only a much higher level of theory but also a more accurate chemical diversity and environments. As forces play a crucial role in molecular dynamics, their weights in the loss function were increased step by step. The results of the fine-tuning are shown in Table 2. For the SPICE database, compared to the original GAFF, our GB-FFs GAFF model significantly reduces the Root Mean Square Error (RMSE) for the energies from 5.8 kcal/mol to less than 3.0 kcal/mol and for the forces from 13.4 kcal/mol/Å to 6.0 kcal/mol/Å. However, for the DES370K database, the RMSE for interaction energy has increased from 1.1 kcal/mol to 1.4 kcal/mol due to the sensibility of VdW parameters.

Table 2: RMSE on SPICE database (potential energy, force and charges) and on DES370K database (potential energy and charges) for GAFF and GB-FFs GAFF.

	SPICE			DES370K	
	Energy(Kcal/mol)	Force(Kcal/mol/Å)	Charge(C)	Energy(Kcal/mol)	Charge(C)
GAFF	5.7804	13.4398	-	1.1470	-
GB-FFs GAFF	2.9706	5.9232	0.0500	1.4146	0.0713

Additionally, as said before, GB-FFs is 3-4 orders of magnitude faster than AM1-BCC calculations using `Antechamber` (the command in AMBER), from 111 seconds to 0.018 second for 50 atoms. Furthermore, the GB-FFs charges provided by the charge transfer model closely approximate AM1-BCC charges (see Table 2), and yield comparable results on *ab-initio* MBIS atomic charges[68] (Supplementary Information (Charge Comparison).

Intermolecular Interaction Accuracy: S66×8 benchmark

In previous databases, models are trained and tested on the same databases. In this subsection, we aim to assess the models' ability to generalize to untrained databases.

The S66×8 database[69] comprises 66 dimers positioned at 8 distinct intermolecular dis-

tances, resulting in a total of 528 unique structures. The S66×8 database is a widely known reference database for assessing the accuracy of intermolecular interactions.

The minimum distance between two monomers ranges from 0.9 to 2.0 times the equilibrium value. When the intermolecular distance varies, the monomers have fixed geometries, meaning that deformation energies of monomers are not considered.

Table 3: MAE and RMSE of the interaction energy on S66×8 database, as well as the MAE and RMSE of the potential energy on torsion scan database for GAFF and the GB-FFs GAFF models.

	S66 × 8		Torsion Scan	
	MAE(Kcal/mol)	RMSE(Kcal/mol)	MAE(Kcal/mol)	RMSE(Kcal/mol)
GAFF	0.9368	1.8388	1.9694	3.5351
GB-FFs GAFF	0.5087	0.8766	0.9892	1.4843

The Mean Absolute Error (MAE) and RMSE on the overall dataset are shown in Table 3. Compared to GAFF, the GB-FFs model reduced by more than half the RMSE. This demonstrates that the training process significantly benefits the approximation of long-range interactions in fitting potential energy and atomic forces.

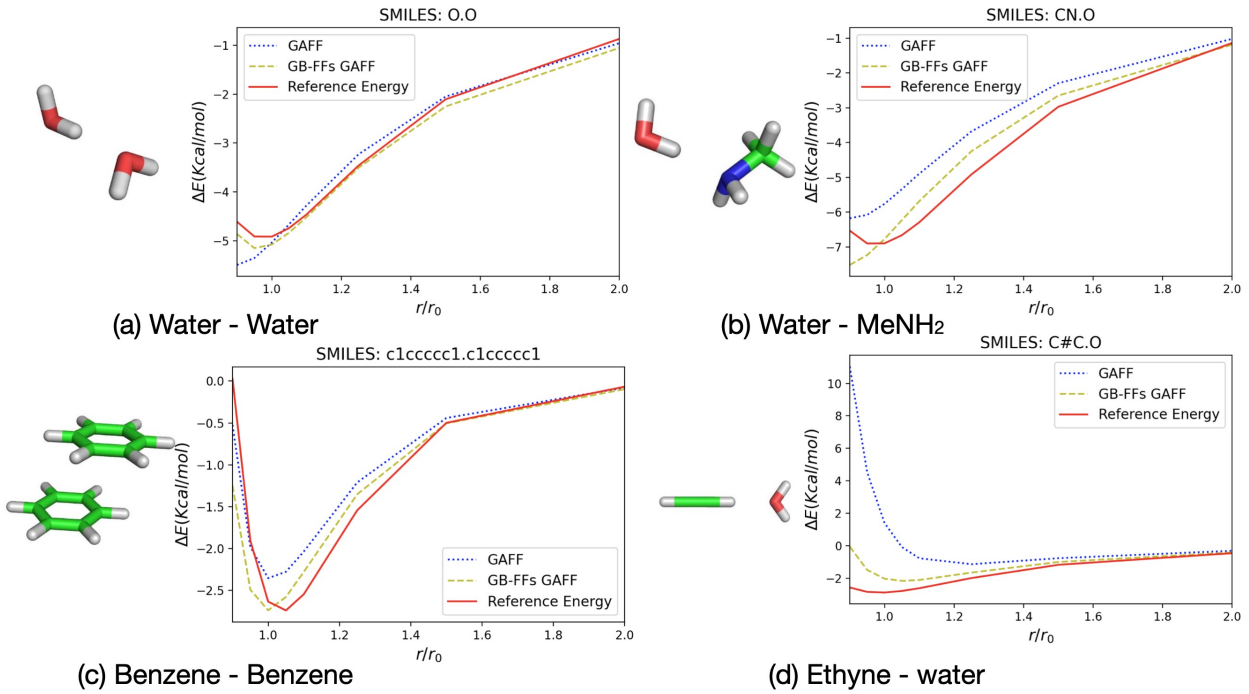


Figure 5: **Results of four example on S66 × 8 database.**

Figure 5 depicts the results for four dimers. The remaining dimers can be found in the Supplementary Infromation (Full results on S66×8 database). The GB-FFs GAFF models almost perfectly reproduce the intermolecular energy surface of the water dimer, which is a critical aspect for simulating solvated biomolecules. In other instances, GB-FFs models are often comparable or outperform GAFF.

Torsion Profiles of 62 Drug-like Fragments

After evaluating intermolecular interactions between molecules, we now turn our attention to assessing the accuracy of predicting intramolecular interactions.

Torsion energies play a crucial role in biology and in small molecular systems. However, accurately assessing torsional parameters in FF is challenging as they necessitate computationally expensive calculations and complicated fitting procedure. Additionally, these parameters are highly sensitive to the local chemical environment, making them less transferable across different molecular systems. Consequently, they often rely on simplistic transferability rules, which can lead to inaccuracies.

Thus, achieving accurate torsion profiles while avoiding the need for extensive torsion fitting is of great importance in FF parameterization. In this context, the performance of the GB-FFs parameterization is also evaluated on a highly accurate torsion scan database.[70] It comprises 62 fragments with drug-like functional groups and their CCSD(T) /CBS single point energies calculated on optimized geometries using MP2[71,72]/6-311+G**.[73,74]

For each molecule of the 62 fragments, a specific dihedral angle is varied from -170° to 170° in increments of 10° (the chosen dihedral angle for modification is indicated in Supplementary Information (Full results on Torsion Scan database)

The overall performance is recorded in Table 3. Compared to the original GAFF, the GB-FFs models refined on the SPICE and DES370K databases provide FF parameters that better fit the potential energy changes caused by dihedral angle variations. In some cases, although there may be a gap between GB-FFs model's predicted results and the reference

energies (see Figure 6 (c) and (d)), the observed trends in these changes correspond with the actual scenarios.

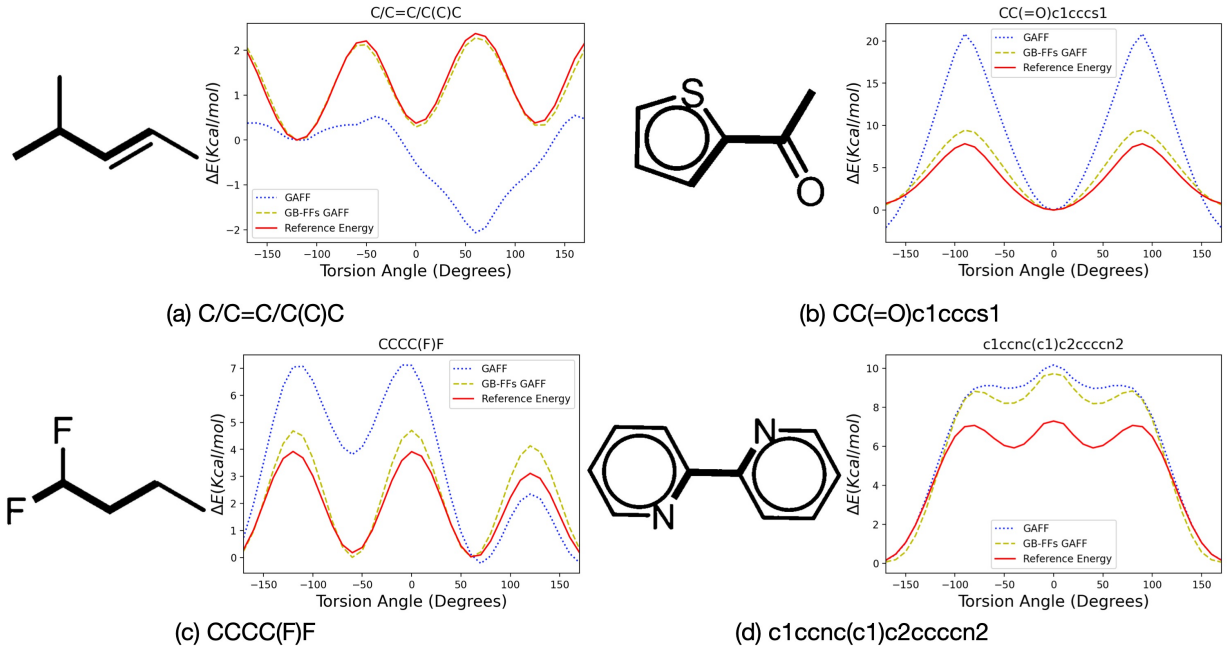


Figure 6: **Results of four examples on 1D torsion scan database.**

This assessment aims to highlight the capabilities of the GB-FFs model in accurately capturing torsional energies.

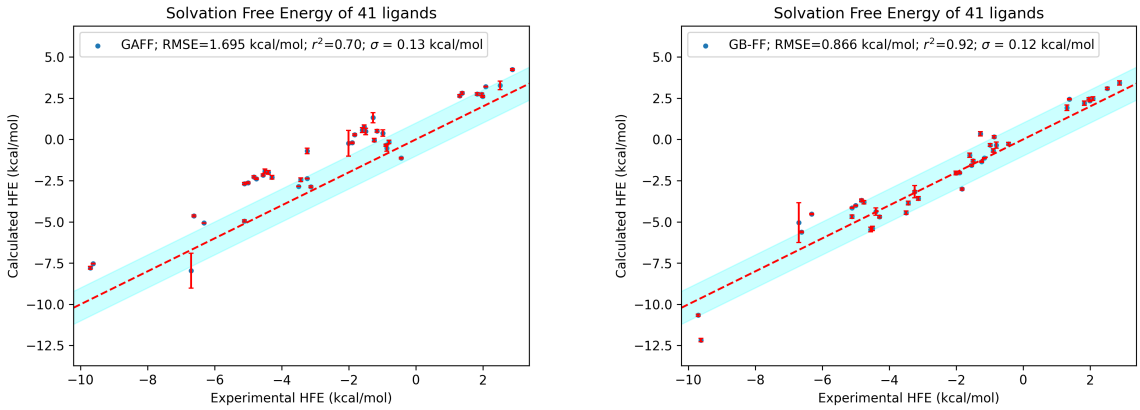
Hydration Free Energies

While our previous focus was on potential energy and atomic forces, this subsection tackles a more challenging property: hydration free energy.

These values are computed for a set of 50 small molecules,[35] which have been previously employed to evaluate AMOEBA's performance and the recently introduced ANI/AMOEBA model.[27] This set covers common chemistry examples, including benzene, acetic acid, and ethane. The experimental values are sourced from the Guthrie solvation database.[75]

In our experiments, the FF parameters for the central molecules are taken from either the original GAFF or GB-FFs models, while the surrounding water molecules are modeled using the Simple Point Charge (SPC) water model.[76] The calculations were performed

with Tinker-HP GPU and using the newly introduced λ-ABF method.[77] As the λ-ABF was recently introduced we also reported in Supplementary Information (Hydration Free Energy Calculations) an extensive study of various water models (GAFF, SPC, Q-SPC, CHARMM22).



(a) Original GAFF (b) GB-FFs GAFF

Figure 7: **Comparison of hydration free energies:** Computed models versus experimental values for 41 molecules in SPC water. The simulations were performed using λ-dynamic ABF, totaling 10 ns of simulation time with the BAOAB-RESPA integrator, 0.25/2 fs setup. For both GAFF and GB-FFs, three repetitions were conducted with different random seeds.

The overall performance is depicted in Figure 7 and the comprehensive data details are documented in the Supplementary Information (Hydration Free Energy Calculations). We conducted three repetitions of 10 ns each to evaluate the standard deviation error for the λ-ABF method. The original GAFF resulted in a RMSE of 1.70 kcal/mol, an r^2 value of 0.70, and a standard deviation, σ, of 0.13 kcal/mol (see Figure 7(a)). In comparison, our best model achieved a RMSE of 0.87 kcal/mol, an r^2 of 0.92, and a σ of 0.12 kcal/mol. Across all metrics, our model demonstrated superior performance, notably by dividing the RMSE by a factor of almost 2. The low standard deviation, σ, of 0.12 kcal/mol is attributed to the λ-ABF method which, not only requires less computational power but also accelerates the sampling of the conformational space. We found the largest variation on acetic acid, that is due to a high energy barrier between two conformations of the proton in the carboxylic acid group.

We found that λ-ABF significantly accelerates the sampling of these two conformations, as these results were not well replicated with the fixed λ method.

We conducted extensive tests using various training and learning strategies and found that the RMSE on the hydration free energy of these models varied from 0.87 to 1.1 kcal/mol. By adding the 0.1 kcal/mol error coming from the λ-ABF, our overall strategy error is approximately 0.3 kcal/mol. These results underscore the superior performance of FF parameters derived from the GB-FFs model. These parameters excel not only in total energies, forces, intermolecular interactions, and torsional interactions but also in hydration free energies. These tests reinforce the effectiveness of GB-FFs parameters in molecular dynamics simulations.

We believe that these results give an idea into the theoretical limitations of the functional form of GAFF for accurately computing free energies. In the next section, we further explore the theoretical limits of GAFF, and more generally classical FFs, by evaluating our model on modified versions of GAFF's original functional form.

Redefining GAFF: Introducing Novel Formulations

This part is a perspective. The accuracy of FFs is determined by the quality of parameters as well as the functional forms. We have made some preliminary attempts in the optimization of FF functional forms by choosing improved functional forms for the bond stretching energies (Morse fitting) and related to the angle bending energies (Urey-Bradley Terms).

The results are presented in Supplementary Information (Redefining GAFF: Introducing Novel Formulations).

These results demonstrate that modifying the functional forms enhances the performance across almost all test datasets. However, it does not further improve the accuracy of hydration free energy simulation, for which polarization and nuclear quantum effects are crucial and will be the subject of another study. Our interpretation is that we may have reached the accuracy limit of what GAFF and, more generally, classical FFs can achieve for hydration

free energies, which is about 0.9 kcal/mol.

Conclusion

In summary, the results that have been presented here have demonstrated the efficacy and viability of using a Directed Graph Attention neTwork (D-GATs) to predict the parameters the General AMBER Force Field (GAFF). The proposed parameterization approach offered several advantages. First, thanks to its efficient runtime complexity of $\mathcal{O}(N)$, and computational cost, GB-FFs could assign parameters of molecules with hundreds atoms, all within a few hundredths of a second. This enabled the simultaneous and self-consistent parameterization of small molecules and biopolymers, eliminating the need for multiple distinct methodologies. Additionally, the automated workflow could leverage large databases, further enhancing the development of FFs.

We evaluated the accuracy and performance of the FF parameters from GB-FFs model through extensive assessments on different databases.

First, we fine-tuned GB-FFs models on the SPICE and DES370K databases, which consists of a wide range of chemical space. The resulting Root Mean Square Error (RMSE) of SPICE energy is 3.06 Kcal/mol, surpassing that of the original GAFF, which exhibits 6.03 Kcal/mol error. GB-FFs model also improves the performance on DES370K database.

To further assess the precision of our model in capturing intermolecular interactions, such as VdW's and Coulomb's, we tested its accuracy on the S66×8 database. Our results showcase a reduction in RMSE by nearly half compared to the original GAFF, highlighting the improved accuracy of our approach in modeling intermolecular interactions. Moreover, we have evaluated the model's transferability and accuracy in capturing torsional interactions by computing one-dimensional torsion profiles. The GB-FFs parameterization exhibits excellent performance in capturing torsional properties (energy RMSE from 3.53 Kcal/mol to 1.34 Kcal/mol). Lastly, we examined the model's capability in predicting hydration free energies for various systems. Our model achieved lower RMSE errors 0.87 kcal/mol compared to the original GAFF parameterization 1.7 kcal/mol, showcasing its high transferability to chemically relevant systems.

Replacing the harmonic function with the Morse function to evaluate bond energy and adding the Urey-Bradley term significantly enhanced the fitting performance, without losing computational cost, of FFs on almost all dataset except for hydration free energies which is slightly worse than our previous best model. One explanation is that we reached the accuracy limit of classical non-polarizable FFs. Such models can be then further improved by adding polarization effects[18,20–22] and/or neural networks components[27,28] to their functional form and nuclear quantum effects within their dynamics.[78,79]

The code can be used on multiple GPUs, enabling accelerated calculations and ensuring efficient processing of multiple molecules simultaneously. Its flexibility allows for its easily integration into popular molecular dynamics workflows.

We proposed the GB-FFs model as a framework for optimizing existing FFs. While in this article, our focus was specifically on optimizing GAFF, it is important to note that our model can be extended to other non-polarizable FFs without the need for scheme modifications. However, extending it to polarizable FFs presents a more complex challenge that demands further research. One potential future direction involves integrating additional physical functional forms, such as the AMOEBA polarizable force field, and we aspire to incorporate this into the automated Poltype 2 framework in the near future.

Another improvement could be done regarding the partial charges assignment to refine the charge transfer model, by either accounting directly for polarization effects or by implementing the electronegativity equalization approach proposed by Gilson et al.[80]

Additionally, one way to enhance the capability of the model in simulating condensed-phase systems would be to add additional observables, such as binding free energies, into the training process, thus helping the model in capturing complex molecular behavior in condensed phases and leading to more accurate and reliable predictions.

Supplementary Information

Recovering atom types in GAFF

Before training our models on potential energy and forces, we employed a molecular processing model to generate atomic embeddings for reconstructing GAFF's atom types. We chose to use the SPICE database over the ANI-1 database due to its broader range of elements and a wider set of biomolecular-relevant structures. The results are displayed in Table S1.

Table S1: **Accuracy of the predicted atom types on SPICE.**

	H	C	N	O	P	S	F	Cl	Br	I	Total
Accuracy	99.80%	99.02%	98.87%	99.51%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	99.42%
Number of Atoms	28,270	24,268	4,685	4,094	86	916	761	473	201	51	63,805

Overall, the predictive accuracy is remarkably high, averaging at 99.42%. Notably, for elements other than H, C, N, and O, the prediction accuracy reaches 100%. When investigating the misclassification of atomic types, we observe that the primary source of error comes from special structures. Taking Figure S1 as an example, our model sometimes incorrectly identifies "cu" as "c2" and "cx" as "c3". The reason for these misclassifications is the failure to correctly recognize triangular system ("c3" v.s. "cx" and "c2" v.s. "cu"). Similarly, for carbon atoms, sometimes our model cannot recognize square systems("c3" v.s. "cy" and "c2" v.s. "cv"), biphenyl system ("ca" v.s. "cp") and non-pure aromatic system ("ca" v.s. "cc") (refer to Table S2).

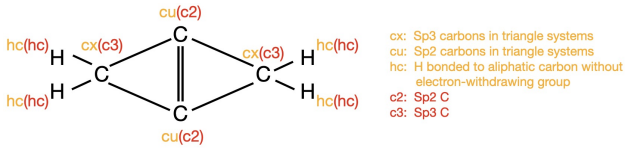


Figure S1: **Comparison of atom types by GAFF and our model.** The atom type in orange are from GAFF, while the red are predicted by GB-FFs model. The incorrect assignment of atom types is due to the failure to recognize triangular systems.

Although the predictions are not always accurate, it is important to note that such special systems are quite rare and uncommon. Furthermore, even without identifying these specific systems, our model has already captured the essential atomic features.

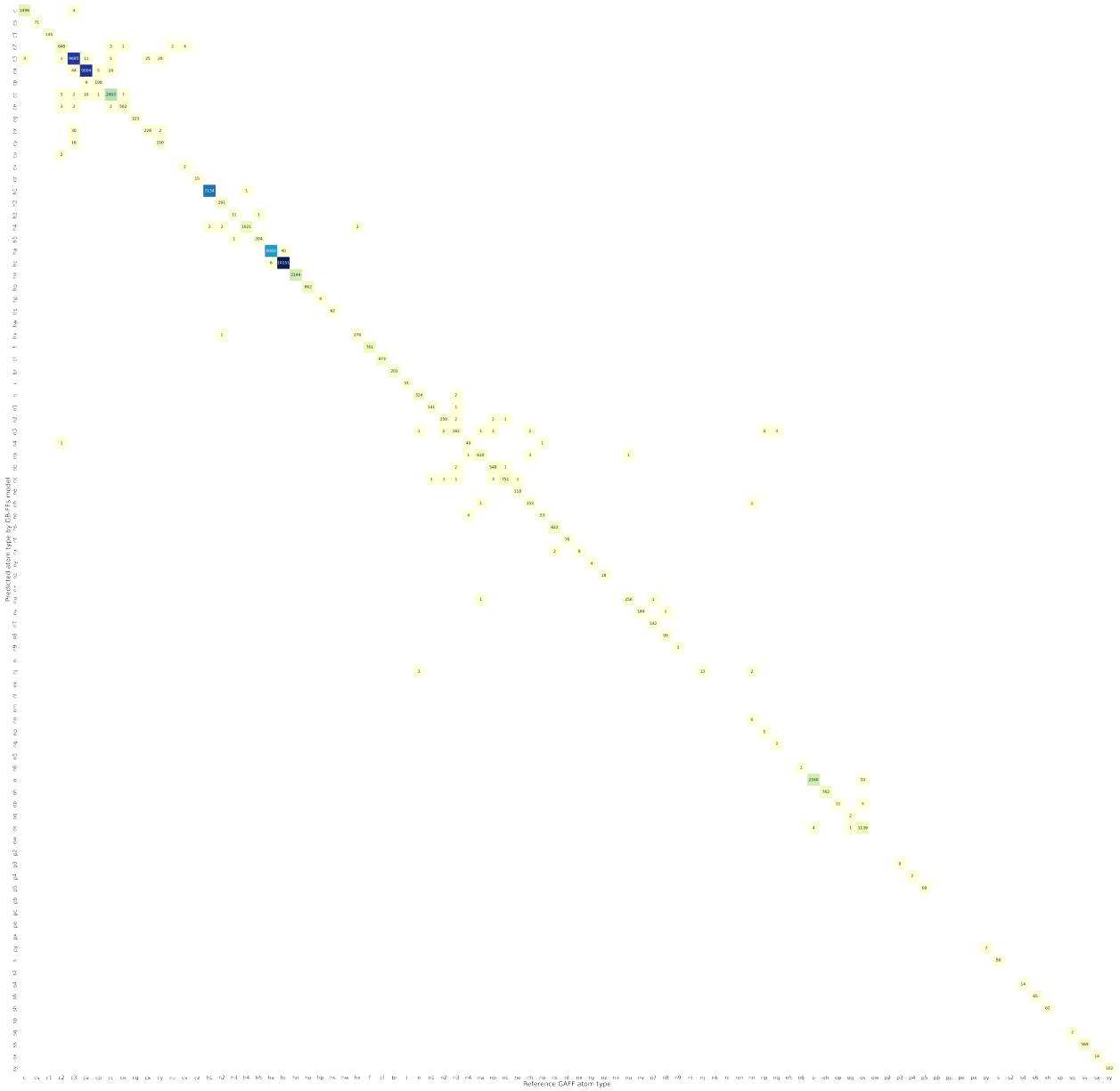


Figure S2: **Full results of predicted atom types on SPICE database.** The numbers in the table represent the frequency for each case.

Pre-training on ANI-1 Database

To enhance models' robustness and performance, they are pre-trained on the ANI-1 database which is, up-to-date, the largest available database of Density Functional Theory (DFT) computations for small organic molecules.[59]

This database comprises over 20 millions off-equilibrium conformations of 57,462 small organic molecules extracted from the GDB database.[81,82] The conformations are built through exhaustive sampling of a subset of the GDB-11 database containing molecules with between 1 to 8 heavy atoms and considering only the species H, C, N and O. The electronic calculations and structure calculations are carried out with the ωB97x[83] density functional and the 6–31 G(d) basis set[84] making it a prime candidate for training Machine Learning (ML) driven FF parameterization model.

There are originally 57,462 molecules in ANI-1 database. After processing, 55 molecules fail to generate GAFF parameter files and 57,407 molecules are left. The molecules are randomly divided into training/validation/test set, following an 8:1:1 ratio.

FFs are mainly designed for conformations that are close to equilibrium states. In the context of pre-training on ANI-1 database, the filtering threshold is set relatively high to retain as much pre-training data as possible. If a conformation exhibits more than one bond energy or angle energy greater than 100 Kcal/mol, or if it has more than one non-bonded atomic pair with a VdW energy greater than 50 Kcal/mol, we discard that particular conformation.

Due to the dependence of energy on FF parameters, the filtering process is dynamic since the parameters provided by ML model vary after each update. After predicting the FF parameters using the GB-FFs model, we calculate the corresponding energy for each conformation and then discard conformations that are far from equilibrium state, retaining the rest conformations for calculating the loss function. Generally, about 2% conformations are filtered in each epoch.

In the pre-training stage, there are three steps and each subsequent step's initial model

is the best model saved from the previous step:

1. Training the GB-FFs model to give the same parameters as GAFF and the loss function is defined as Equation (7), where $\texttt{MSE}(P_X)$ denotes the mean square error (MSE) calculated by the difference of X (bond or angle or dihedral or VdW or charge) parameters from GB-FFs model and GAFF parameters. The models are trained for 20 epochs.

$$\texttt{Loss} = \texttt{MSE}(P_{bonds}) + \texttt{MSE}(P_{angles}) + \texttt{MSE}(P_{dihedrals}) + \texttt{MSE}(P_{charges}) + \texttt{MSE}(P_{VdW}) \quad (7)$$

2. Training the GB-FFs model to give the same parameters and the same energy as GAFF. The loss function is defined as Equation (8), where $\texttt{MSE}(E_X)$ denotes the MSE of energy calculated by the X (bond or angle or dihedral or VdW or charge) parameters from GB-FFs model and the energy calculated by the GAFF parameters. It should be noted that the energy in this step indicates the energy for each interaction, i.e. for every bond/angle/dihedrals and VdW/charge interaction pair. The models are trained for 20 epochs.

$$\begin{aligned}\texttt{Loss} = {} & \texttt{MSE}(P_{bonds}) + \texttt{MSE}(P_{angles}) + \texttt{MSE}(P_{dihedrals}) + \texttt{MSE}(P_{charge}) + \texttt{MSE}(P_{VdW}) \\ & + \texttt{MSE}(E_{bonds}) + \texttt{MSE}(E_{angles}) + \texttt{MSE}(E_{dihedrals}) + \texttt{MSE}(E_{charge}) + \texttt{MSE}(E_{VdW})\end{aligned} \quad (8)$$

3. Training the GB-FFs model to give the same parameters and the same total potential energy as GAFF. The loss function is defined as Equation (9), where $\texttt{MSE}(\Delta E_{total})$ denotes the MSE of relative energy calculated using the parameters from GB-FFs model and the total energy calculated by the GAFF parameters. The relative energy refers to the energy difference between a given conformation and the conformation with lowest energy. By focusing on ΔE_{total}, we can investigate the impact of different conformations on the potential energy, while the absolute energy is influenced by the system's conditions. Our model is capable of predicting partial charge. In the pre-training stage, our charge is expected to approach AM1-BCC charge but we still use

the AM1-BCC charge to compute the potential energy. There are also L2 regularization for dihedral parameters and the transfer charge (not the partial charge). Without the regularization loss, dihedral terms give unreasonable results. And the relatively small transfer charge corresponds to the real situation. The weight in loss function for each part is obtained by experiments. The models are trained for 50 epochs the only the models with smallest validation error will be stored.

$$\begin{aligned} \texttt{Loss} = {} & \texttt{MSE}(P_{bonds}) + \texttt{MSE}(P_{angles}) + 10 * \texttt{MSE}(P_{dihedrals}) + 100 * \texttt{MSE}(P_{charge}) \\ & + 100 * \texttt{MSE}(P_{VdW}) + 0.1 * \texttt{L2}(P_{dihedral}) + 0.1 * \texttt{L2}(charge_{transfer}) \\ & + \texttt{MSE}(\Delta E_{ANI1}) \end{aligned} \tag{9}$$

The results of different models on test set are shown in Table S2. It should be noted that the energy to compute error is the relative energy (the energy difference between the minima) instead of absolute energy to avoid considering the heats of formation.

Table S2: Pre-training results of GB-FFs model on ANI-1 test dataset.

	GAFF	GB-FFs GAFF
RMSE for energy (Kcal/mol)	21.6032	12.6212

From the Results showed in Table S2, the FF parameters given by GB-FFs model have better performance than original GAFF. However, the original GAFF is never trained on ANI-1 database and it is not surprising that it is inferior to GB-FFs model.

Fine-Tuning Strategy on SPICE and DES370K Databases

SPICE (Small-Molecule/Protein Interaction Chemical Energies)[60] is a collection of quantum mechanical data aimed to be use in biomolecular simulation. The computations are performed at the ωB97M-D3(BJ) functional[62,63] and def2-TZVPPD basis set,[64,65] which is known for its high accuracy.

DES370K[61] is a collection of dimer interaction energies computed using the high-level coupled-cluster singles and doubles with perturbative triples (CCSD(T))[66] method at the complete basis set (CBS)[67] level of theory. This database contains 370,959 molecular geometries for 3,691 distinct dimers, which represent 392 monomers (both neutral molecules and ions) including, but not limited to, water and the functional groups found in proteins.

The loss function is defined as in Equation (10), and it includes interaction energies from DES370K, as well as total energies and forces from SPICE.

$$\begin{aligned}\texttt{Loss} = \texttt{MSE}(P_{bonds}) + \texttt{MSE}(P_{angles}) + 10 * \texttt{MSE}(P_{dihedrals}) + 100 * \texttt{MSE}(P'_{charge}) \\ + 1000 * \texttt{MSE}(P_{VdW}) + 0.1 * \texttt{L2}(P_{dihedral}) + 0.1 * \texttt{L2}(charge_{transfer}) \\ + \texttt{MSE}(\Delta E_{SPICE}) + 100 * \texttt{MSE}(\Delta E_{DES370K}) + 10 * \texttt{MSE}(\Delta F_{SPICE})\end{aligned} \tag{10}$$

In the fine-tuning stage, we also did a study about the scalings on the AM1-BCC charge within the loss function and observed that a scaling factor of 1.1 provides optimal results. This questioning originate from the lack of polarization within GAFF and that strategies in the classical FF community aims to add a scaling factor the charge model. This questioning arises from the lack of polarization within GAFF and a known strategy in the classical force field community is to scale the charge.

$$\texttt{MSE}(P'_{charge}) = \texttt{MSE}(\texttt{GB-FFs charge} - 1.1 * \texttt{AM1-BCC charge}) \tag{11}$$

The filtering rule is set as: if a conformation exhibits more than one bond energy or angle energy greater than 50 Kcal/mol, or if it has more than one non-bonded atomic pair with a VdW energy greater than 5 Kcal/mol, it will be abandoned. Generally, about 8.5% conformations are filtered in each epoch.

The criteria to judge our models are the sum of RMSE in potential energies and in atomic forces: $\texttt{RMSE}(\Delta E_{SPICE}) + 100 * \texttt{RMSE}(\Delta E_{DES370K}) + 10 * \texttt{RMSE}(\Delta F)$. The models are trained

for 2,000 epochs.

Full results on S66×8 database

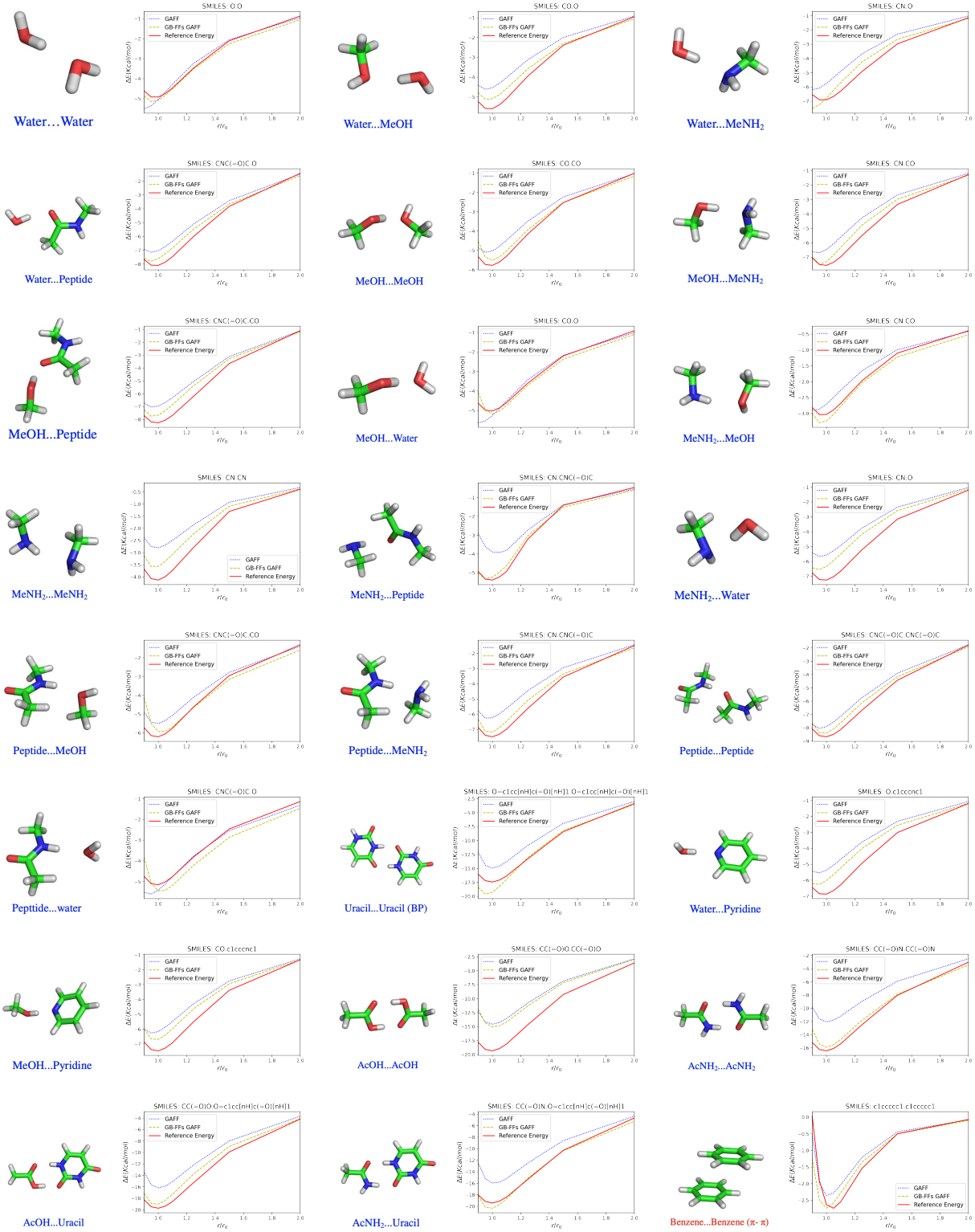


Figure S3: **All results for S66 × 8 database (1 / 3)**.

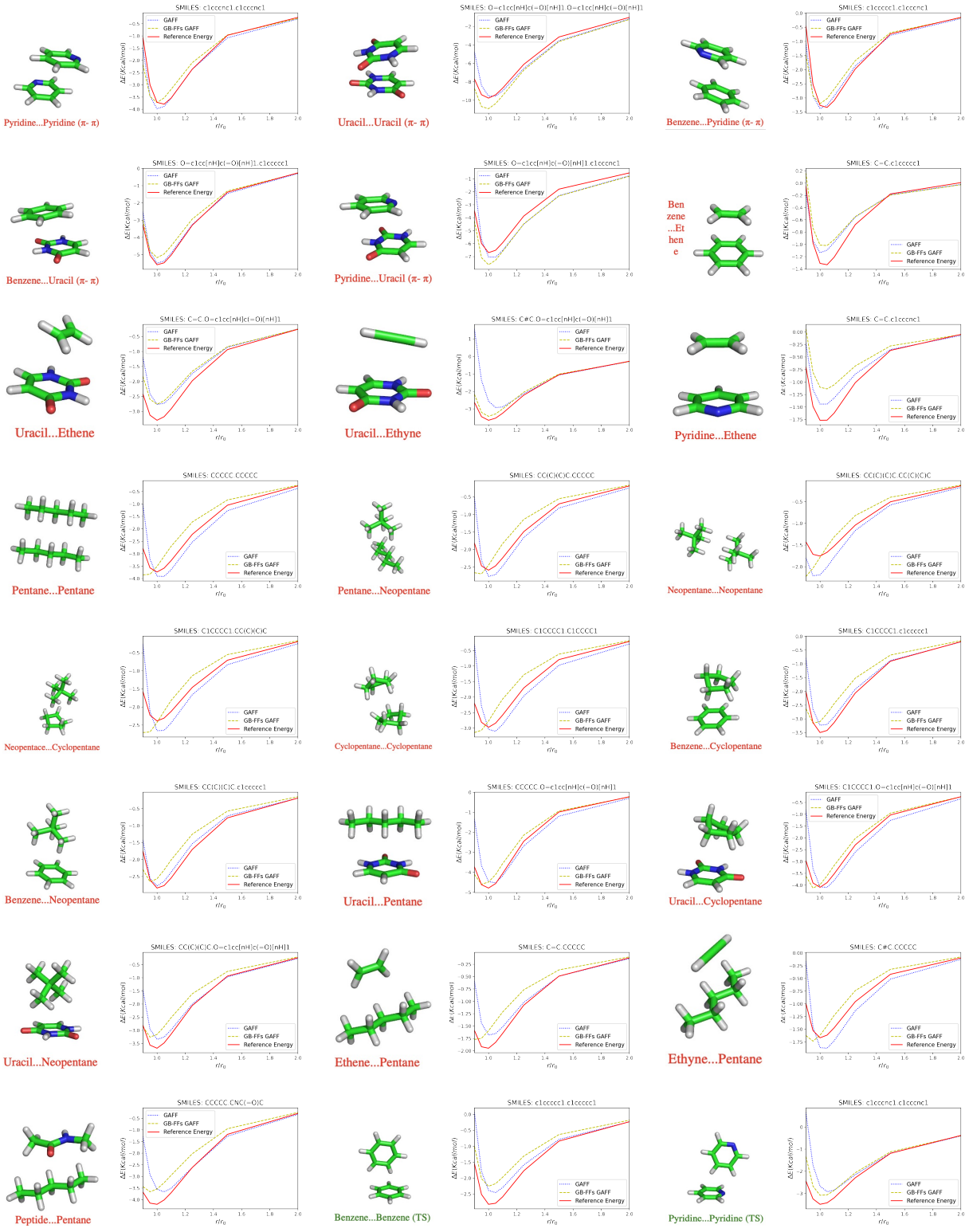


Figure S3: **All results for S66 × 8 database (2 / 3)**.

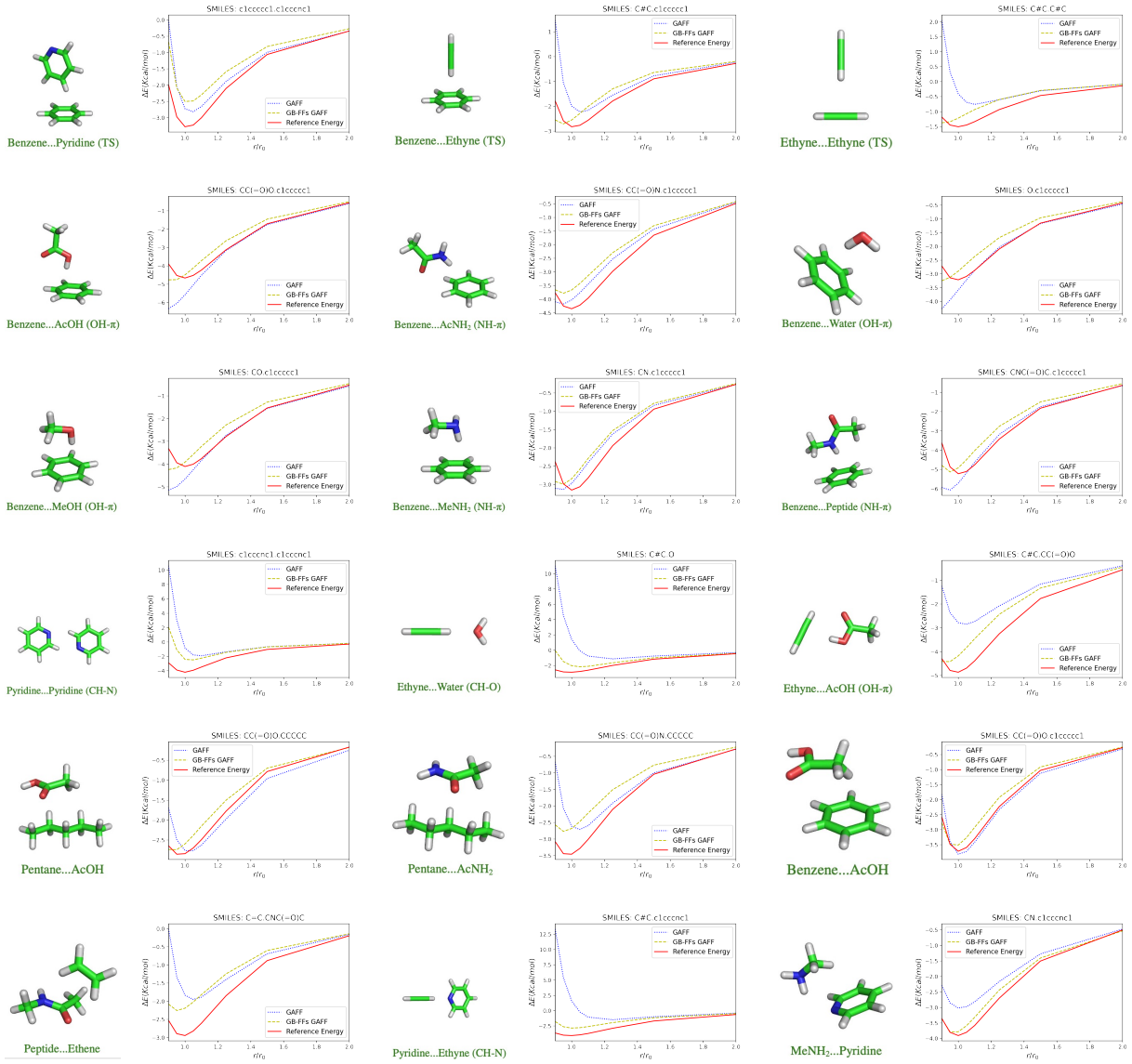


Figure S3: **All results for S66 × 8 database (3 / 3)**. Results of "GAFF" and "GB-FFs GAFF" on the S66×8 database. It is composed of dimers that have been gradually pulled apart, and for which *ab-initio* computations were performed. On the x-axis is the ratio of the distance to the equilibrium distance value (ranging from 0.9 to 2.0), and the y-axis represents the potential energy (in Kcal/mol).

Full results on Torsion Scan database

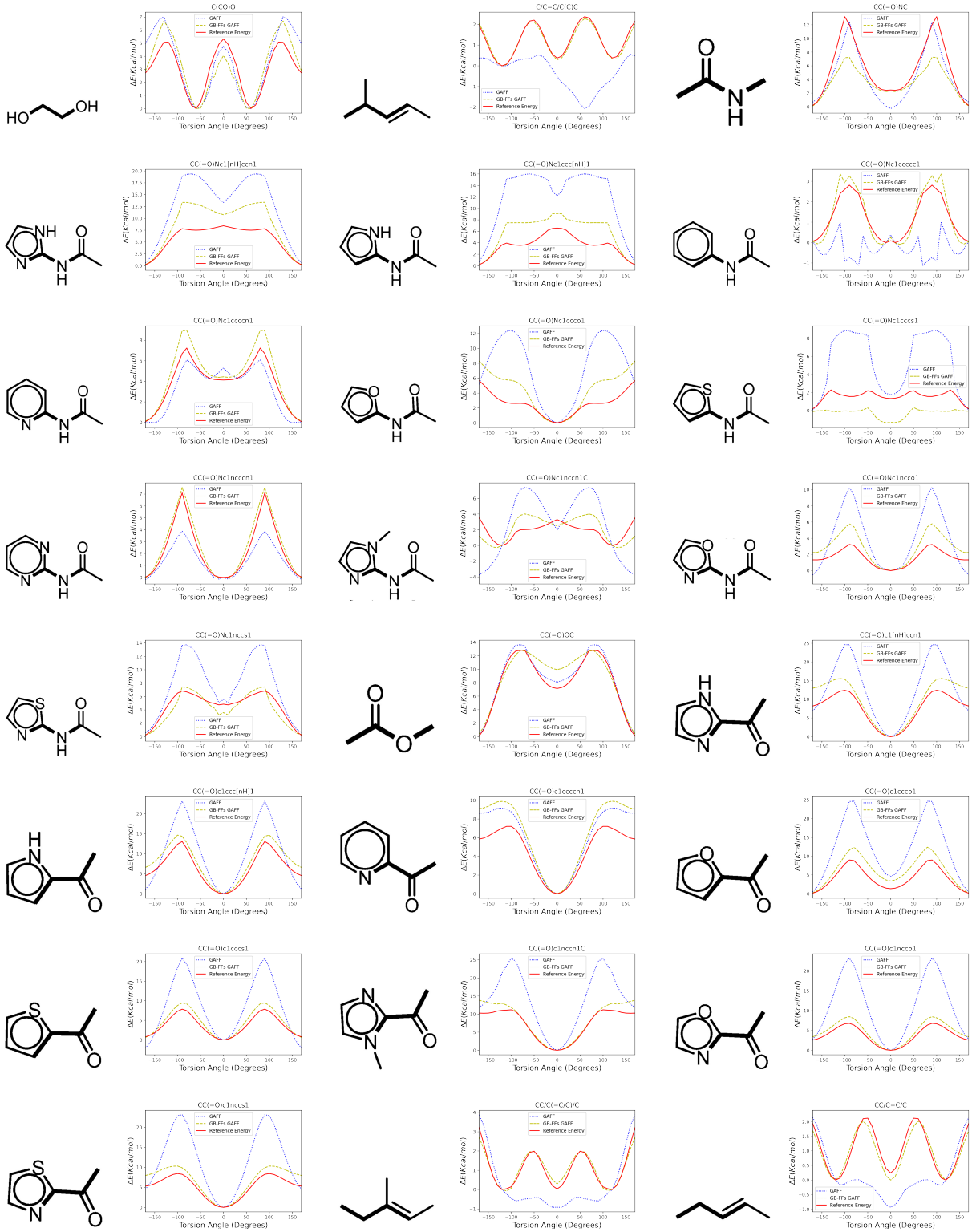


Figure S4: **All results for 1D Torsion Scan database (1 / 3)**.

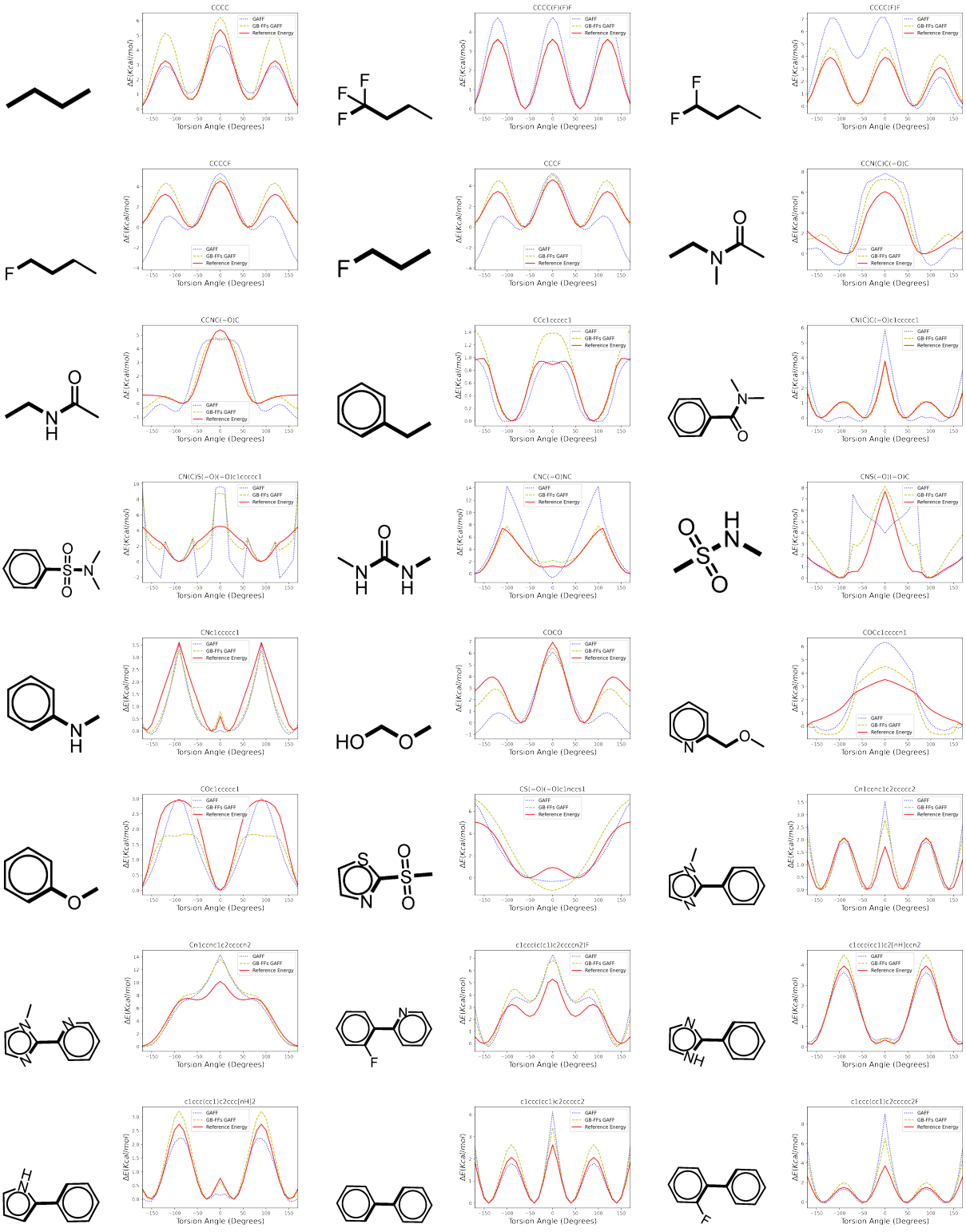


Figure S4: **All results for 1D Torsion Scan database (2 / 3)**.

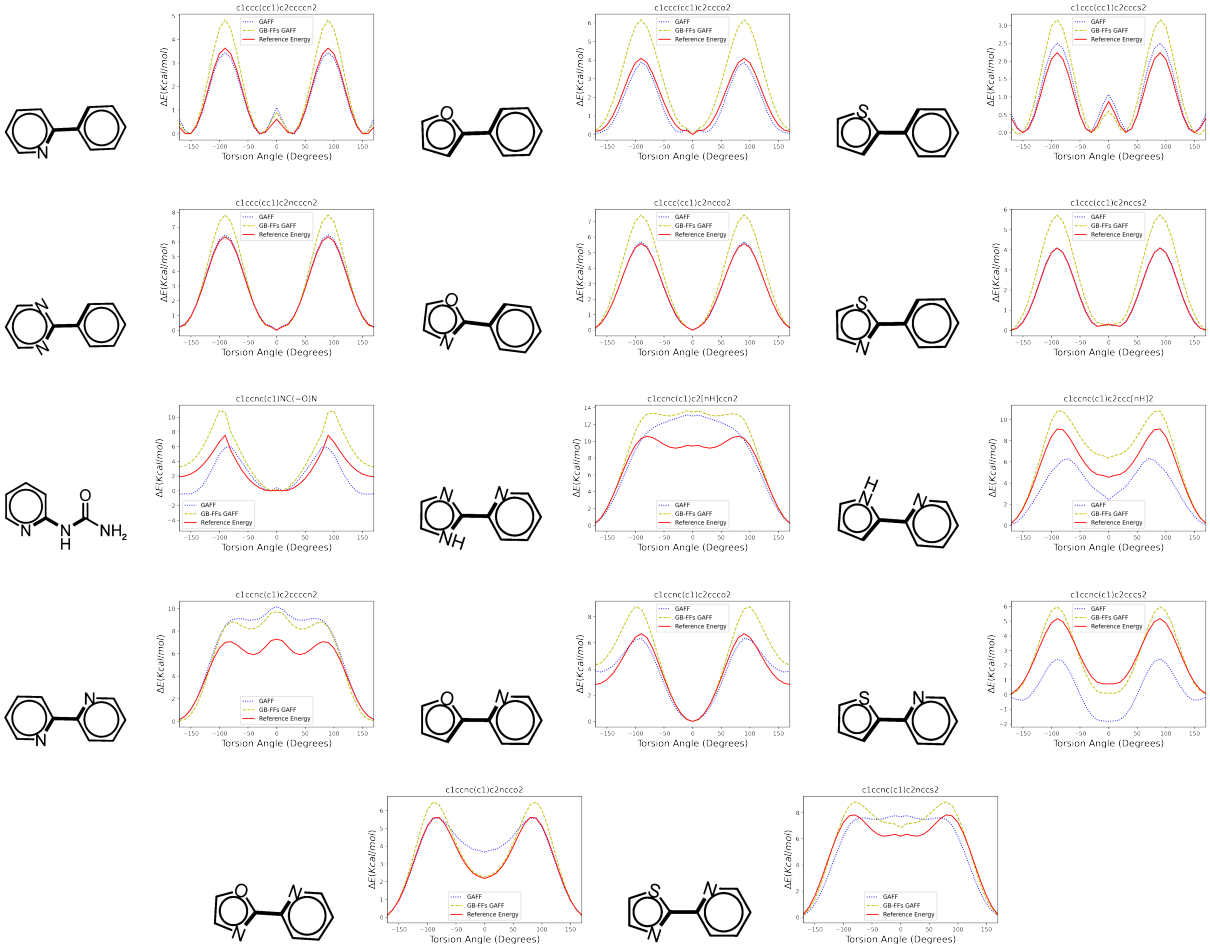


Figure S4: **All results for 1D Torsion Scan database (3 / 3)**. Results of "GAFF" and "GB-FFs GAFF" on 1D torsion scan database.[70] It is composed of drug-like fragments for which torsion scans have been conducted, and *ab-initio* computations were performed. On the x-axis is the degrees of dihedral angles (from -170° to 170°), and the y-axis represents the potential energy (in Kcal/mol).

Hydration Free Energy Calculations

Before conducting hydration free energy simulations, we conducted an extensive study on the effect of various water models, SPC,[85,86] Q-SPC,[85] and CHARMM22,[17] on the hydration free energy of water. We focused particularly on flexible water models, as Tinker-HP is designed for such models, in contrast to most other softwares, which are typically built for rigid water models such as TIP3P.[87] Table S3 depicts the hydration free energy of water, where we employed a different model for simulating the bulk and the water molecule, the latter is

in parentheses. The simulations were conducted using λ-ABF with Velocity Verlet 1fs, in the NPT ensemble with Berendsen Barostat, for a total of 1 ns of simulation. We found that the SPC water model yields satisfactory results compared to experiments, while the GB-FF/GAFF bulk water model produces unsatisfactory ones. This is attributed to the sensitivity of the bulk water model's properties to slight differences in parameters.[78,88] For GAFF and many FFs a different model is usually used for bulk water and for the water ligand. In the following, we employed the SPC water model for bulk water while ligand's parameter are predicted by GB-FF.

Table S3: Hydration free energy (kcal/mol) comparison of water. Comparison between experimental and some ligand(solvent) water models.

	Water
GB-FF(GB-FF)	-4.59
GB-FF(Q-SPC)	-4.05
GAFF(Q-SPC)	-5.11
GAFF(GAFF)	-5.52
SPC(SPC)	-6.36
Q-SPC(Q-SPC)	-6.83
CHARMM22(CHARMM22)	-6.87
Exp.	-6.32

We tested our model's prediction on the hydration free energy of 41 molecules. The results can be found in Tables S4 and S5. 10 ns of simulations were performed using λ-ABF, with BAOAB-RESPA integrator 0.5/2 fs setup, and a Langevin barostat. For assessing sampling and reproducibility error, we conducted three repetitions for both GAFF and GB-FF, the results presented here are then the averaged values over these three simulations. $\Delta\overline{G_{\text{solv}}}$ is the averaged free energy difference between the ligand within the solvent and the ligand in vacuum, i.e., the hydration free energy. $\Delta\Delta\overline{G_{\text{solv}}}$ is the absolute difference compared to the experimental value. In parantheses are the standard deviation of each $\Delta\overline{G_{\text{solv}}}$ over the three repetitions.

The fine-tuning process is influenced both by the order of molecules in the training dat-

Table S4: Hydration free energy (kcal/mol) of 41 molecules. Part 1.

		$\Delta\overline{G_{\text{solv}}}$ (kcal/mol)		$\Delta\Delta\overline{G_{\text{solv}}}$ (kcal/mol)	
Molecule	Exp	GAFF	GB-FF	GAFF	GB-FF
22-Dimethylbutane	2.51	3.27(0.27)	3.09(0.07)	0.58	0.76
2-Methyl-2-butene	1.31	2.64(0.08)	1.92(0.17)	0.61	1.33
acetalaldehyde	-3.5	-2.85(0.01)	-4.45(0.10)	0.95	0.65
acetamide	-9.71	-7.79(0.08)	-10.66(0.06)	0.95	1.92
benzene	-0.87	-0.53(0.20)	0.16(0.07)	1.03	0.34
But-1-ene	1.38	2.81(0.07)	2.44(0.04)	1.06	1.43
diethylsulfide	-1.6	0.57(0.14)	-0.96(0.14)	0.64	2.17
dimethylamine	-4.29	-2.30(0.11)	-4.70(0.07)	0.41	1.99
Dimethyldisulfide	-1.83	0.29(0.06)	-3.01(0.06)	1.18	2.12
dimethylsulfide	-1.54	0.78(0.11)	-1.57(0.05)	0.03	2.32
Di-n-butylamine	-3.24	-0.69(0.17)	-3.17(0.36)	0.07	2.55
Di-n-propylether	-1.16	0.51(0.08)	-1.14(0.03)	0.02	1.67
ethane	1.83	2.76(0.07)	2.19(0.13)	0.36	0.93
ethanol	-5	-2.63(0.06)	-4.01(0.03)	0.99	2.37
ethylamine	-4.5	-1.92(0.13)	-5.40(0.13)	0.90	2.58
ethylbenzene	-0.8	-0.15(0.10)	-0.35(0.18)	0.45	0.65
formicacid	-5.11	-4.94(0.06)	-4.68(0.10)	0.43	0.17
hydrogensulfide	-0.44	-1.13(0.04)	-0.27(0.09)	0.17	0.69
imidazole	-9.63	-7.53(0.02)	-12.17(0.10)	2.54	2.10
isopropanol	-4.76	-2.38(0.04)	-3.81(0.10)	0.95	2.38
methane	1.99	2.60(0.06)	2.34(0.05)	0.35	0.61
methanol	-5.11	-2.68(0.06)	-4.15(0.03)	0.96	2.43
Methylacetate	-3.13	-2.88(0.05)	-3.58(0.11)	0.45	0.25
methylamine	-4.56	-2.16(0.06)	-5.46(0.14)	0.90	2.40
methylether	-1.9	-0.20(0.05)	-2.02(0.04)	0.12	1.70
methylethylsulfide	-1.5	0.48(0.19)	-1.29(0.07)	0.21	1.98
methylsulfide	-1.24	-0.04(0.11)	-1.35(0.03)	0.11	1.20

Table S5: Hydration free energy (kcal/mol) of 41 molecules. Part 2.

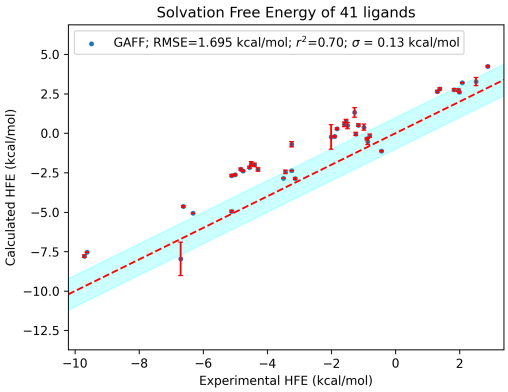
		$\Delta\overline{G_{\text{solv}}}$ (kcal/mol)		$\Delta\Delta\overline{G_{\text{solv}}}$ (kcal/mol)	
Molecule	Exp	GAFF	GB-FF	GAFF	GB-FF
n-butane	2.08	3.19(0.02)	2.48(0.12)	0.40	1.11
n-Butanethiol	-0.99	0.38(0.19)	-0.34(0.08)	0.65	1.37
n-Octane	2.88	4.24(0.05)	3.42(0.13)	0.54	1.36
phenol	-6.62	-4.64(0.05)	-5.61(0.05)	1.01	1.98
propane	1.96	2.73(0.08)	2.46(0.07)	0.50	0.77
propanol	-4.83	-2.27(0.06)	-3.69(0.08)	1.14	2.56
Propionaldehyde	-3.43	-2.45(0.11)	-3.85(0.11)	0.42	0.98
propylamine	-4.4	-2.00(0.10)	-4.38(0.22)	0.02	2.40
trimethylamine	-3.24	-2.37(0.02)	-3.18(0.06)	0.06	0.87
toluene	-0.89	-0.35(0.07)	-0.67(0.10)	0.22	0.54
Methylisopropylether	-2.01	-0.24(0.78)	-2.03(0.11)	0.02	1.77
aceticacid	-6.7	-7.96(1.05)	-5.05(1.21)	1.65	1.26
Di-n-propylsulfide	-1.28	1.32(0.31)	0.34(0.13)	1.63	2.60
water	-6.32	-5.07(0.01)	-4.53(0.04)	1.79	1.25

aloader and by the random dropout in the molecule processing model, ultimately impacting the performance of the fine-tuned model. Utilizing the same pre-trained model from the ANI-1 database, we conducted multiple fine-tuning sessions on the SPICE and DES370K databases. These fine-tuned models share identical loss functions.

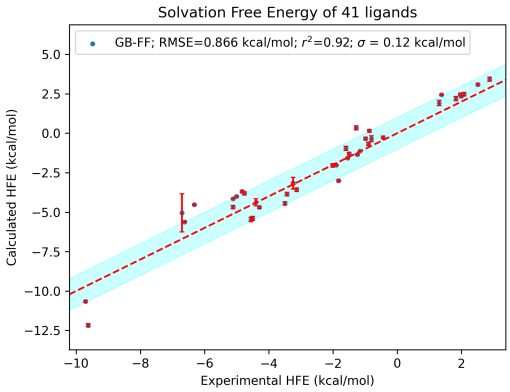
In Figure S5, we present the hydration free energy of the original GAFF, the GB-FFs model with the best performance (as presented in the main article), and another fine-tuned model.

Redefining GAFF: Introducing Novel Formulations

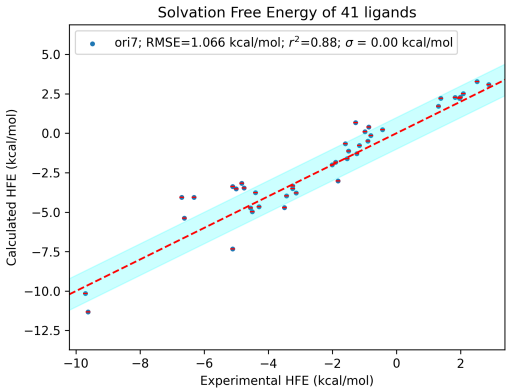
This part is a perspective. The accuracy of FFs is determined by the quality of parameters as well as the functional forms. We have made some preliminary attempts in the optimization of FFs functional forms.



GAFF



GB-FFs GAFF



Another model

Figure S5: **Comparison of hydration free energies of multiple models**

Morse Function for Stretching Energy

FFs typically adopt functional forms based on a balance between computational cost and approximation effectiveness. Although the harmonic function yields only moderate accuracy, it offers high computational efficiency, enabling simulations of larger and more complex systems. The Morse function[89] provides a more accurate description of the bond potential, particularly for bonds that are stretched beyond their equilibrium values. We employ the following form of the Morse function (the same number of parameters as GAFF):

$$E_{bonds} = \sum_{bonds} \frac{K_r}{4}(e^{-2(r-r_{eq})} - 1)^2 \tag{12}$$

with $\{K_r, r_{eq}\}$ are the parameters directly from GAFF parameters. This model is noted as "GB-FFs Morse" model.

Urey-Bradley Terms

In this context, we remove the constraint on the number of parameters and use the complete Morse function to evaluate the stretching energy.

$$E_{bonds} = \sum_{bonds} \frac{K_r}{\alpha^2}(e^{-\alpha(r-r_{eq})} - 1)^2 \tag{13}$$

with $\{K_r, \alpha, r_{eq}\}$ are the FF parameters, initialized as $\{K_r, 2, r_{eq}\}$ and r is the bond length. $\{K_r, r_{eq}\}$ are the parameters directly from GAFF parameters.

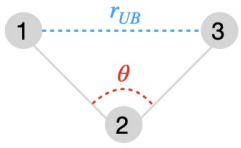


Figure S6: **Urey-Bradley term:** The 1-3 endpoints distance r_{UB} is taken into consideration.

The Urey-Bradley (UB) terms[90] serve as a cross-term addressing 1-3 non-bonded interactions that are not adequately covered by the bond and angle terms (refer to Figure S6).

Including UB terms enhances the accuracy of replicating subtleties within the vibrational spectrum.

Nowadays, some FFs, such as CHARMM and AMOEBA, still use UB terms while AMBER and GAFF do not. However, it is known that introducing new UB terms raises concerns, particularly regarding limited transferability and the increased complexity in the force field parametrization process. In the context of general-purpose force fields, the primary bottleneck is the complexity of parametrization, and this is a key reason why ML-based parametrization might offer a solution.

For GB-FFs model, the process to learn atomic representations and predict suitable FF parameters is automated and efficient. Additionally, GB-FFs model generates parameters based on the atomic chemical environment, which largely mitigates issues related to low transferability.

Usually, the UB terms apply harmonic function, which is not enough to represent the complex interaction. We take use of the following function:

$$E_{UB} = \sum_{angles} K_{UB}((\frac{r_{UB_{eq}}}{r_{UB}})^2 - 1)^2 \tag{14}$$

with $\{K_{UB}, r_{UB_{eq}}\}$ are the new FF parameters. As we have no reference data for $\{K_{UB}, r_{UB_{eq}}\}$, K_{UB} is initialized as 0.1 (in Kcal/mol) and $r_{UB_{eq}}$ (in Å) is initialized by the equilibrium structure parameters r_{eq} and θ_{eq} in GAFF. For example, in Figure S6, the equilibrium bond length for bonds 1-3 and 2-3 are noted as r_{eq}^{12} and r_{eq}^{23}, as well as the equilibrium angle is noted as θ_{eq}^{123}. We expect the UB term achieves the smallest energy when the structure is in equilibrium state. According to the Law of cosines, $r_{UB_{eq}}$ is initialized as:

$$r_{UB_{eq}} = \sqrt{r_{eq}^{12\,2} + r_{eq}^{23\,2} - 2r_{eq}^{12}r_{eq}^{23}\cos(\theta_{eq}^{123})} \tag{15}$$

Indeed, we can incorporate more correction terms, such as separating the endpoint distance of the torsion term from the VdW terms, and assigning it as part of the torsion term.

However, this approach carries the risk of over fitting and can potentially undermine the performance of the original FFs, which goes against our initial goal of optimizing the molecular FFs. Therefore, we have only made simple modifications to the stretch terms and bending terms for now. This model is noted as "GB-FFs UB" model.

Results on New Models

In this part, "GB-FFs GAFF", "GB-FFs Morse" and "GB-FFs UB" are all trained using the same loss function. The key distinction lies in their utilization of different functional forms. The overall performance are listed in Table S6.

Table S6: The RMSE of GB-FFs Morse model and GB-FFs UB model on different databases.

	SPICE			DES370K		S66×8	Torsion Scan
	Energy (Kcal/mol)	Force (Kcal/mol/Å)	Charge (C)	Energy (Kcal/mol)	Charge (C)	Energy (Kcal/mol)	Energy (Kcal/mol)
GAFF	5.7804	13.4398	-	1.1470	-	1.8388	3.5351
GB-FFs GAFF	2.9706	5.9232	0.0500	1.4146	0.0713	0.8766	1.4843
GB-FFs Morse	2.8812	5.3050	0.0492	1.3884	0.0698	0.8437	1.4190
GB-FFs UB	2.5723	4.1416	0.0491	1.0941	0.0526	0.8028	1.1054

We observed the significant impact of function forms on the performance of FFs. When we replace the harmonic function with the Morse function to approximate the potential energy associated with chemical bonds and incorporate UB terms for bending energy, the RMSE of the potential energy and atomic forces decrease a lot. Therefore, optimizing the FFs requires not only data-driven approaches such as optimizing NNs structures and extracting atomic fingerprints but also considering mathematical and chemical perspectives to employ function forms that better simulate the potential energy and align with the actual conditions.

Compared to the GB-FFs GAFF model, which applies the same functional form as GAFF, the modified function forms show improved fitting performance, particularly with the introduction of the UB term.

We only run the hydration free energy calculation for GB-FFs UB model. It has comparable performance to GB-FFs GAFF model. Although the performance of parameters from

GB-FFs UB have better performance on SPICE, DES370K, S66×8 as well as Torsion Scan databases, the free energy is not more accurate.

These are preliminary experiments to assess the influence of the function form on FFs. Moreover, it explores how to assume suitable FF parameters based on existing data in the absence of reference values. Subsequent work will further research into the suitable functional forms as well as parameters.

Research on Non-bonded Interactions

In previous sections, we set 1.1*AM1-BCC charge as targets (Function 11) because it allows model to achieve better performance. In this section, we delve into the potential factors contributing to this improvement.

We name the model trained by following charge loss function as "GB-FFs GAFF-1.0" (without any modifications to the functional forms):

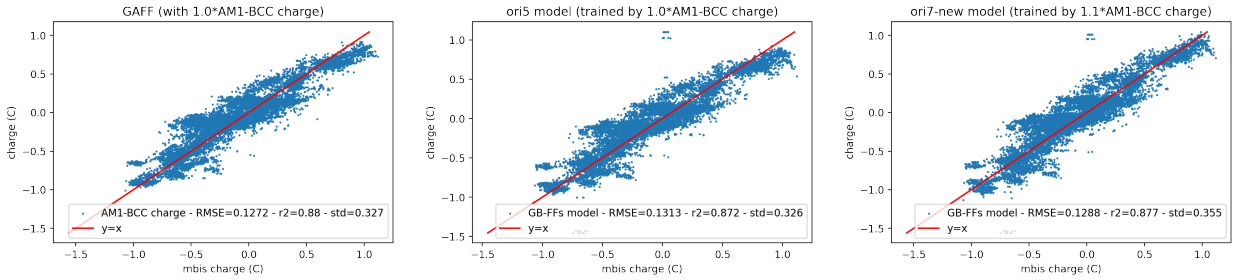
$$\texttt{MSE}(P_{charge}) = \texttt{MSE}(\texttt{GB-FFs charge} - \texttt{AM1-BCC charge}) \tag{16}$$

Charge Comparison

The ANI-1x database contains DFT calculations for approximately five million diverse molecular conformations.[91] ANI-1x uses the Minimal Basis Iterative Stockholder partitioning (MBIS) scheme[68] to calculate atomic charges and other properties with the wB97x/def2-TZVPP functional.

Following processing by `Antechamber` and RDKit, only 1,347 molecules remained. We selected the geometry with the lowest potential energy as the equilibrium geometry and recorded the corresponding MBIS charge as our reference.

In Figure S7, we compared AM1-BCC charges (S7 (a)) with GB-FFs charges (S7 (b) and (c)) against MBIS charges. The GB-FFs GAFF model in (b) is trained using 1.0*AM1-BCC charges, while the model in (c) is trained with 1.1*AM1-BCC charges.



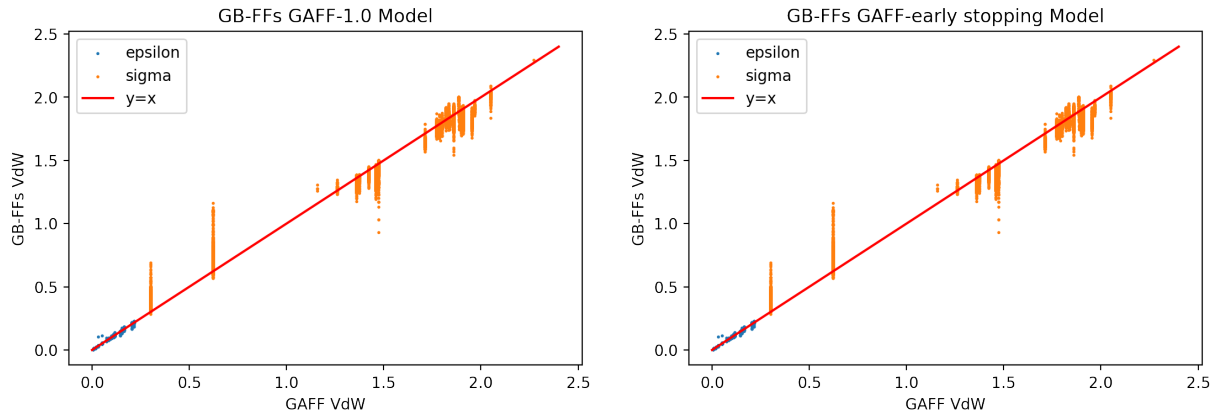
(a) AM1-BCC charge (b) GB-FFs GAFF-1.0 (c) GB-FFs GAFF

Figure S7: **Comparison of charges for different models**

The results in Figure S7 illustrate that GB-FFs charges exhibit comparable performance to AM1-BCC charges. However, it's worth noting that the charges derived from both the GB-FFs GAFF model and the GB-FFs GAFF-1.0 model do not display significant differences.

VdW Parameters Comparison

Regarding the Van der Waals (VdW) interactions, we conducted a comparison of the VdW parameters (ε and σ) between the GB-FFs model and the original GAFF using the ANI-1X database. The model in Figure S8(a) was trained with a reference energy corresponding to 1.0*AM1-BCC charges while the model shown in Figure S8(b) was trained using a reference energy based on 1.1*AM1-BCC charges.



(a) GB-FFs GAFF-1.0 (b) GB-FFs GAFF

Figure S8: **Comparison of VdW parameters for different models**

Our investigation reveals that the modifications made to VdW parameters primarily focus on the σ value, which represents the Van der Waals radius, while ε remains relatively consistent. These modifications to the VdW parameters play a significant role in improving the simulation of intermolecular interactions on the S66×8 database. Unlike the constraints imposed by atom types, the VdW parameters in the GB-FFs models are directly determined by atomic representations, enabling a more flexible range of parameter values.

For now, we have to establish a clear connection between the distribution of VdW parameters and their impact on the overall performance. This should a future exploration.

Acknowledgement

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 810367), project EMC2 (JPP,YM). Simulations have been performed at GENCI on the Jean Zay machine (IDRIS, Orsay, France) on grant no A0070707671.

References

(1) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *The Journal of Chemical Physics* **2013**, *139*, 134101.

(2) Liakos, D. G.; Guo, Y.; Neese, F. Comprehensive Benchmark Results for the Domain Based Local Pair Natural Orbital Coupled Cluster Method (DLPNO-CCSD(T)) for Closed- and Open-Shell Systems. *The Journal of Physical Chemistry A* **2020**, *124*, 90–100, PMID: 31841627.

(3) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham III, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995**, *91*, 1–41.

(4) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005**, *26*, 1668–1688.

(5) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. a.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry* **1983**, *4*, 187–217.

(6) Brooks, B. R.; Brooks III, C. L.; Mackerell Jr, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., et al. CHARMM: the biomolecular simulation program. *Journal of computational chemistry* **2009**, *30*, 1545–1614.

(7) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **2004**, *25*, 1157–1174.

(8) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247–260.

(9) Stevens, J. A.; Grünewald, F.; van Tilburg, P. A. M.; König, M.; Gilbert, B. R.; Brier, T. A.; Thornburg, Z. R.; Luthey-Schulten, Z.; Marrink, S. J. Molecular dynamics simulation of an entire cell. *Frontiers in Chemistry* **2023**, *11*.

(10) Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics* **2020**, *153*, 044130.

(11) Jaffrelot Inizan, T.; Célerse, F.; Adjoua, O.; El Ahdab, D.; Jolly, L.-H.; Liu, C.; Ren, P.; Montes, M.; Lagarde, N.; Lagardère, L.; Monmarché, P.; Piquemal, J.-P. High-resolution mining of the SARS-CoV-2 main protease conformational space: supercomputer-driven unsupervised adaptive sampling. *Chem. Sci.* **2021**, *12*, 4889–4907.

(12) El Ahdab, D.; Lagardère, L.; Inizan, T. J.; Célerse, F.; Liu, C.; Adjoua, O.; Jolly, L.-H.; Gresh, N.; Hobaika, Z.; Ren, P.; Maroun, R. G.; Piquemal, J.-P. Interfacial Water Many-Body Effects Drive Structural Dynamics and Allosteric Interactions in SARS-CoV-2 Main Protease Dimerization Interface. *The Journal of Physical Chemistry Letters* **2021**, *12*, 6218–6226, PMID: 34196568.

(13) 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 Jr, R. A., et al. Current status of the AMOEBA polarizable force field. *The journal of physical chemistry B* **2010**, *114*, 2549–2564.

(14) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *Journal of Chemical Theory and Computation* **2018**, *14*, 2084–2108, PMID: 29438622.

(15) Liu, C.; Piquemal, J.-P.; Ren, P. AMOEBA+ Classical Potential for Modeling Molecular Interactions. *Journal of Chemical Theory and Computation* **2019**, *15*, 4122–4139, PMID: 31136175.

(16) Liu, C.; Piquemal, J.-P.; Ren, P. Implementation of Geometry-Dependent Charge Flux into the Polarizable AMOEBA+ Potential. *The Journal of Physical Chemistry Letters* **2020**, *11*, 419–426, PMID: 31865706.

(17) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. J. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chemical Reviews* **2016**, *116*, 4983–5013, PMID: 26815602.

(18) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-up Strategy. *Journal of Chemical Theory and Computation* **2007**, *3*, 1960–1986, PMID: 18978934.

(19) Naseem-Khan, S.; Lagardère, L.; Narth, C.; Cisneros, G. A.; Ren, P.; Gresh, N.; Piquemal, J.-P. Development of the Quantum-Inspired SIBFA Many-Body Polarizable Force Field: Enabling Condensed-Phase Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2022**, *18*, 3607–3621, PMID: 35575306.

(20) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annual Review of Biophysics* **2019**, *48*, 371–394, PMID: 30916997.

(21) Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J.-P. *Reviews in Computational Chemistry Volume 28*; John Wiley & Sons, Ltd, 2015; Chapter 2, pp 51–86.

(22) Melcr, J.; Piquemal, J.-P. Accurate Biomolecular Simulations Account for Electronic Polarization. *Frontiers in Molecular Biosciences* **2019**, *6*, 143.

(23) El Khoury, L.; Célerse, F.; Lagardère, L.; Jolly, L.-H.; Derat, E.; Hobaika, Z.; Maroun, R. G.; Ren, P.; Bouaziz, S.; Gresh, N.; Piquemal, J.-P. Reconciling NMR Structures of the HIV-1 Nucleocapsid Protein NCp7 Using Extensive Polarizable Force Field Free-Energy Simulations. *Journal of Chemical Theory and Computation* **2020**, *16*, 2013–2020, PMID: 32178519.

(24) Behler, J. Four generations of high-dimensional neural network potentials. *Chemical Reviews* **2021**, *121*, 10037–10072.

(25) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine learning force fields. *Chemical Reviews* **2021**, *121*, 10142–10186.

(26) Kocer, E.; Ko, T. W.; Behler, J. Neural network potentials: A concise overview of methods. *Annual review of physical chemistry* **2022**, *73*, 163–186.

(27) Jaffrelot Inizan, T.; Plé, T.; Adjoua, O.; Ren, P.; Gökcan, H.; Isayev, O.; Lagardère, L.; Piquemal, J.-P. Scalable hybrid deep neural networks/polarizable potentials biomolecular simulations including long-range effects. *Chem. Sci.* **2023**, *14*, 5438–5452.

(28) Plé, T.; Lagardère, L.; Piquemal, J.-P. Force-field-enhanced neural network interactions:

from local equivariant embedding to atom-in-molecule properties and long-range effects. *Chem. Sci.* **2023**, *14*, 12554–12569.

(29) Raimbault, N.; Grisafi, A.; Ceriotti, M.; Rossi, M. Using Gaussian process regression to simulate the vibrational Raman spectra of molecular crystals. *New Journal of Physics* **2019**, *21*, 105001.

(30) Sommers, G. M.; Andrade, M. F. C.; Zhang, L.; Wang, H.; Car, R. Raman spectrum and polarizability of liquid water from deep neural networks. *Physical Chemistry Chemical Physics* **2020**, *22*, 10592–10602.

(31) Poier, P. P.; Jaffrelot Inizan, T.; Adjoua, O.; Lagardère, L.; Piquemal, J.-P. Accurate Deep Learning-Aided Density-Free Strategy for Many-Body Dispersion-Corrected Density Functional Theory. *The Journal of Physical Chemistry Letters* **2022**, *13*, 4381–4388, PMID: 35544748.

(32) Artrith, N.; Behler, J. High-dimensional neural network potentials for metal surfaces: A prototype study for copper. *Physical Review B* **2012**, *85*, 045439.

(33) Sun, G.; Sautet, P. Toward fast and reliable potential energy surfaces for metallic Pt clusters by hierarchical delta neural networks. *Journal of chemical theory and computation* **2019**, *15*, 5614–5627.

(34) Wu, J. C.; Chattree, G.; Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor Chem Acc* **2012**, *131*, 1138.

(35) Walker, B.; Liu, C.; Wait, E.; Ren, P. Automation of AMOEBA polarizable force field for small molecules: Poltype 2. *Journal of Computational Chemistry* **2022**, *43*, 1530–1542.

(36) Wang, L.-P.; Chen, J.; Van Voorhis, T. Systematic parametrization of polarizable force

fields from quantum chemistry data. *Journal of chemical theory and computation* **2013**, *9*, 452–460.

(37) Kumar, A.; Pandey, P.; Chatterjee, P.; MacKerell, A. D. J. Deep Neural Network Model to Predict the Electrostatic Parameters in the Polarizable Classical Drude Oscillator Force Field. *Journal of Chemical Theory and Computation* **2022**, *18*, 1711–1725, PMID: 35148088.

(38) Thürlemann, M.; Böselt, L.; Riniker, S. Regularized by Physics: Graph Neural Network Parametrized Potentials for the Description of Intermolecular Interactions. *Journal of Chemical Theory and Computation* **2023**, *19*, 562–579, PMID: 36633918.

(39) Wang, Y.; Fass, J.; Kaminow, B.; Herr, J. E.; Rufa, D.; Zhang, I.; Pulido, I.; Henry, M.; Macdonald, H. E. B.; Takaba, K., et al. End-to-end differentiable construction of molecular mechanics force fields. *Chemical Science* **2022**, *13*, 12016–12033.

(40) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956–972.

(41) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: Accelerating Molecular Dynamics Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields Using GPUs and Multi-GPU Systems. *Journal of Chemical Theory and Computation* **2021**, *17*, 2034–2053, PMID: 33755446.

(42) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *Journal of computational chemistry* **2005**, *26*, 1701–1718.

(43) Lorentz, H. A. Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Annalen der physik* **1881**, *248*, 127–136.

(44) Berthelot, D. Sur le mélange des gaz. *Compt. Rendus* **1898**, *126*, 15.

(45) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993**, *97*, 10269–10280.

(46) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *Journal of the American Chemical Society* **2002**, *115*, 9620–9631.

(47) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *Journal of Computational Chemistry* **2000**, *21*, 132–146.

(48) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry* **2002**, *23*, 1623–1641.

(49) Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang, P.; Pande, V.; Leskovec, J. Strategies for pre-training graph neural networks. *arXiv preprint arXiv:1905.12265* **2019**,

(50) Rong, Y.; Bian, Y.; Xu, T.; Xie, W.; Wei, Y.; Huang, W.; Huang, J. Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information Processing Systems* **2020**, *33*, 12559–12571.

(51) Fang, X.; Liu, L.; Lei, J.; He, D.; Zhang, S.; Zhou, J.; Wang, F.; Wu, H.; Wang, H. Geometry-enhanced molecular representation learning for property prediction. *Nature Machine Intelligence* **2022**, *4*, 127–134.

(52) Chen, G.; Maday, Y. Directed Message Passing Based on Attention for Prediction of Molecular Properties. *arXiv preprint arXiv:2305.14819* **2023**,

(53) Liu, S.; Wang, H.; Liu, W.; Lasenby, J.; Guo, H.; Tang, J. Pre-training molecular graph representation with 3d geometry. *arXiv preprint arXiv:2110.07728* **2021**,

(54) Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M., et al. Analyzing learned molecular representations for property prediction. *Journal of chemical information and modeling* **2019**, *59*, 3370–3388.

(55) Biswas, K.; Kumar, S.; Banerjee, S.; Pandey, A. K. SMU: smooth activation function for deep networks using smoothing maximum technique. *arXiv preprint arXiv:2111.04682* **2021**,

(56) Goodfellow, I.; Warde-Farley, D.; Mirza, M.; Courville, A.; Bengio, Y. Maxout networks. International conference on machine learning. 2013; pp 1319–1327.

(57) Landrum, G. RDKit: Open-source cheminformatics. `https://www.rdkit.org`.

(58) Muller, P. Glossary of terms used in physical organic chemistry (IUPAC Recommendations 1994). *Pure and Applied Chemistry* **1994**, *66*, 1077–1184.

(59) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical science* **2017**, *8*, 3192–3203.

(60) Eastman, P.; Behara, P. K.; Dotson, D. L.; Galvelis, R.; Herr, J. E.; Horton, J. T.; Mao, Y.; Chodera, J. D.; Pritchard, B. P.; Wang, Y., et al. SPICE, A Dataset of Drug-like Molecules and Peptides for Training Machine Learning Potentials. *Scientific Data* **2023**, *10*, 11.

(61) Donchev, A. G.; Taube, A. G.; Decolvenaere, E.; Hargus, C.; McGibbon, R. T.; Law, K.-H.; Gregersen, B. A.; Li, J.-L.; Palmo, K.; Siva, K., et al. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Scientific data* **2021**, *8*, 55.

(62) Najibi, A.; Goerigk, L. The nonlocal kernel in van der Waals density functionals as an additive correction: An extensive analysis with special emphasis on the B97M-V and ωB97M-V approaches. *Journal of Chemical Theory and Computation* **2018**, *14*, 5725–5738.

(63) Mardirossian, N.; Head-Gordon, M. ω B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *The Journal of chemical physics* **2016**, *144*, 214110.

(64) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297–3305.

(65) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *The Journal of chemical physics* **2010**, *133*, 134105.

(66) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters* **1989**, *157*, 479–483.

(67) Montgomery Jr, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *The Journal of chemical physics* **1999**, *110*, 2822–2827.

(68) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal basis iterative stockholder: atoms in molecules for force-field development. *Journal of Chemical Theory and Computation* **2016**, *12*, 3894–3912.

(69) Rezác, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *Journal of chemical theory and computation* **2011**, *7*, 2427–2438.

(70) Sellers, B. D.; James, N. C.; Gobbi, A. A comparison of quantum and molecular mechanical methods to estimate strain energy in druglike fragments. *Journal of chemical information and modeling* **2017**, *57*, 1265–1275.

(71) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. A direct MP2 gradient method. *Chemical Physics Letters* **1990**, *166*, 275–280.

(72) Head-Gordon, M.; Head-Gordon, T. Analytic MP2 frequencies without fifth-order storage. Theory and application to bifurcated hydrogen bonds in the water hexamer. *Chemical Physics Letters* **1994**, *220*, 122–128.

(73) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—consistent molecular orbital methods. XII. Further extensions of Gaussian—type basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics* **1972**, *56*, 2257–2261.

(74) McLean, A.; Chandler, G. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z= 11–18. *The Journal of chemical physics* **1980**, *72*, 5639–5648.

(75) Guthrie, J. P.; Mobley, D. L. The Guthrie Hydration Free Energy Database of Experimental Small Molecule Hydration Free Energies. **2018**,

(76) Berendsen, H.; Postma, J.; Van Gunsteren, W.; Hermans, a. J.; Pullman, B. Intermolecular forces. 1981.

(77) Lagardère, L.; Maurin, L.; Adjoua, O.; Hage, K. E.; Monmarché, P.; Piquemal, J.-P.; Hénin, J. Lambda-ABF: Simplified, Accurate and Cost-effective Alchemical Free Energy Computations. *arXiv preprint arXiv:2307.08006* **2023**,

(78) Mauger, N.; Plé, T.; Lagardère, L.; Bonella, S.; Mangaud, É.; Piquemal, J.-P.; Huppert, S. Nuclear quantum effects in liquid water at near classical computational cost using the adaptive quantum thermal bath. *The Journal of Physical Chemistry Letters* **2021**, *12*, 8285–8291.

(79) Plé, T.; Mauger, N.; Adjoua, O.; Inizan, T. J.; Lagardère, L.; Huppert, S.; Piquemal, J.-P. Routine Molecular Dynamics Simulations Including Nuclear Quantum Effects: From Force Fields to Machine Learning Potentials. *Journal of Chemical Theory and Computation* **2023**, *19*, 1432–1445, PMID: 36856658.

(80) Gilson, M. K.; Gilson, H. S.; Potter, M. J. Fast assignment of accurate partial atomic charges: an electronegativity equalization method that accounts for alternate resonance forms. *Journal of chemical information and computer sciences* **2003**, *43*, 1982–1997.

(81) Fink, T.; Bruggesser, H.; Reymond, J.-L. Virtual exploration of the small-molecule chemical universe below 160 daltons. *Angewandte Chemie International Edition* **2005**, *44*, 1504–1508.

(82) Fink, T.; Reymond, J.-L. Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *Journal of chemical information and modeling* **2007**, *47*, 342–353.

(83) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *The Journal of chemical physics* **2008**, *128*, 084106.

(84) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *The Journal of Chemical Physics* **1971**, *54*, 724–728.

(85) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*; Pullman, B., Ed.; Springer Netherlands: Dordrecht, 1981; pp 331–342.

(86) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A* **2001**, *105*, 9954–9960.

(87) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(88) Mauger, N.; Plé, T.; Lagardère, L.; Huppert, S.; Piquemal, J.-P. Improving Condensed-Phase Water Dynamics with Explicit Nuclear Quantum Effects: The Polarizable Q-AMOEBA Force Field. *The Journal of Physical Chemistry B* **2022**, *126*, 8813–8826, PMID: 36270033.

(89) Morse, P. M. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Physical review* **1929**, *34*, 57.

(90) Devlin, J. P. Urey-Bradley "Nonbonded" Forces. *The Journal of Chemical Physics* **1963**, *39*, 2385–2385.

(91) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Scientific data* **2020**, *7*, 134.