# Fast Generation of Machine Learning-Based Force Fields for Adsorption Energies

Saientan Bag, Manuel Konrad, Tobias Schlöder, Pascal Friederich,* and Wolfgang Wenzel*

ABSTRACT: Adsorption and desorption of molecules are key processes in extraction and purification of biomolecules, engineering of drug carriers, and designing of surface specific coatings. To understand the adsorption process on the atomic scale, state of the art quantum mechanical and classical simulation methodologies are widely used. However, studying adsorption using a full quantum mechanical treatment is limited to picoseconds simulation timescales, while classical molecular dynamics simulations are limited by the accuracy of the existing force fields. To overcome these challenges, we propose a systematic way to generate flexible, application specific highly accurate force fields by training artificial neural networks. As a proof of concept, we study

the adsorption of the amino acid alanine on graphene and gold (111) surfaces and demonstrate the force field generation methodology in detail. We find that a molecule specific force field with Lennard Jones type two body terms incorporating the 3rd and 7th power of the inverse distances between the atoms of the adsorbent and the surfaces yields optimal results, which is surprisingly different from typical Lennard Jones potentials used in traditional force fields. Furthermore, we present an efficient and easy to train machine learning model that incorporates system specific three body (or higher order) interactions that are required, for example, for gold surfaces. Our final machine learning based force field yields a mean absolute error of less than 4.2 kJ/mol at a speed up of $\sim 10^5$ times compared to quantum mechanical calculation, which will have a significant impact on the study of adsorption in different research areas.
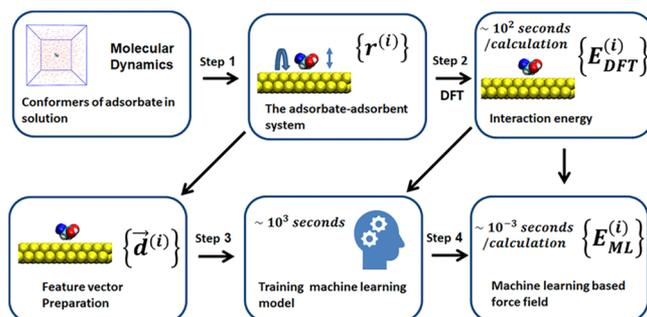
1. INTRODUCTION

The study of adsorption phenomena is one of the key research areas in chemical engineering,[1,2] medicine,[3−6] and biotechnology.[7] Adsorption and desorption processes are used in a wide variety of applications, such as the extraction and purification of biomolecules,[8,9] engineering of drug carriers,[10] and design of coatings for specific surfaces.[11] Due to its widespread importance, adsorption is studied worldwide using state of the art quantum mechanical[12−15] and classical simulation[16−21] methodologies. However, there are limitations in both directions. A full quantum mechanical treatment, for example, using ab initio molecular dynamics (MD) simulations, to study adsorption is only accessible up to a few picoseconds of simulation time. On the other hand, classical MD simulation suffers from the accuracy of the available force fields (FFs). The state of the art universal FFs can only approximately predict weak intermolecular interactions and thus noncovalent binding energies, while specific FFs designed for adsorption are complex and system dependent.

In particular, the parameters are dependent on the class of adsorbate molecules as well as on the surface material. To study the adsorption of amino acids on a gold surface, for example, a specific FF GolP Charmm[22] has been developed, where the gold surface is described by additional virtual particles in addition to the gold atoms. Furthermore, to model the polarizability of gold, the gold atoms are split into two point charges forming a dipole. A specific set of nonbonded parameters is suggested to describe amino acids on gold surfaces, but the parameters of the surface have to be changed to study DNA[23] instead of amino acids on gold surfaces. To study DNA on graphene,[24] virtual atoms are no longer necessary, while splitting of atoms is still required to account for the polarizability of the surface. These examples illustrate that it is far from trivial to derive classical FFs for even relatively simple surfaces, a process which presently proceeds case by case.

The FFs discussed above are examples of a bottom up approach of the FF development, where first principles data are used to parameterize FF. It is worth mentioning that there is another route (top down approach) to parametrize FFs
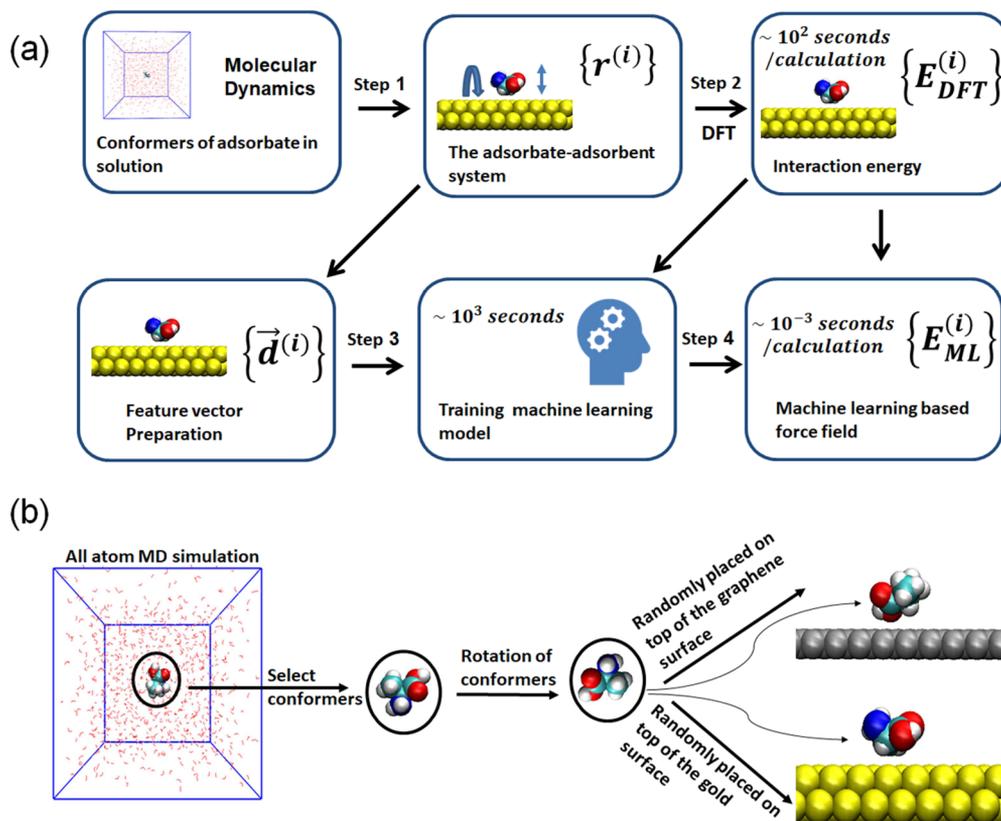
Figure 1. (a) Schematic workflow describing the generation of the ML based force field. (b) Schematic diagram describing the generation of adsorbent−adsorbate conformers for DFT calculation.

based on the experimental data. Geada and co workers[25] developed a FF for gold surfaces following the top down approach, which was in excellent agreement with the experiments in predicting some interfacial properties. How ever, generalization of the top down approach to any arbitrary molecule−surface combination is complicated by the lack of experimental data.

In general, surface specific FFs (bottom up) are para meterized to reproduce the density functional theory (DFT) optimized adsorption geometries and the corresponding energies but do not necessarily generalize well to off equilibrium conformations. Therefore, when used in MD simulation, these FFs may generate inaccurate sampling of the phase space and thus predict unreliably thermodynamic ensembles and free energies. To overcome this difficulty, we propose a way of generating flexible and accurate FFs to model adsorption by training artificial neural networks (NNs), which have become a powerful tool[26−30] capable of providing quantum mechanical accuracy when trained using DFT data. In this work, we propose a similar approach to generate the FF to study adsorption. Here, we develop only the intermolecular part of the total FF, while proposing to use well established FFs for the intramolecular part.

As a proof of concept, we study the adsorption of the amino acid alanine (adsorbate) on two different surfaces (adsorbent), graphene and gold (111). We develop a protocol to sample energetically relevant conformations of the system (*incl.* adsorbate and adsorbent) and perform DFT calculations to compute the adsorbate−adsorbent interaction energy with these geometries. We also compare the prediction from GolP Charmm FFs (which was optimized to reproduce optimized

geometry and the corresponding energy) with the DFT calculated energies and find that GolP Charmm does not match well with the DFT calculated energies for the off equilibrium conformations. While a traditional pairwise potential [e.g., a Lennard Jones (LJ) potential] could not fit the DFT calculated energy, an artificial NN with three hidden layers was able to reproduce the DFT calculated energies with mean absolute errors (MAEs) of 2.28 and 3.78 kJ/mol for the graphene and gold system, respectively. Instead of using generic NN as a fitting tool, we here pursue an approach that parameterizes functions relevant to the physical nature of the problem to increase both accuracy and efficiency. The optimal feature for the best performance of the machine learning (ML)

models was found to be the 3rd and 7th power $\left\{\left(\dfrac{1}{r_{ij}}\right)^3, \left(\dfrac{1}{r_{ij}}\right)^7\right\}$

of the inverse distance $\{r_{ij}\}$ between the atoms of the adsorbent and adsorbate, which is very different from the inverse distance power used in traditional pair potentials in MD simulation, but which can be easily implemented. Furthermore, we explored the effect of n body interactions in the adsorbate−adsorbent interaction energy. We found that the interaction between the gold surface and alanine has significant contributions from n body interactions, while the graphene−alanine interaction is quite well represented by pairwise interactions.

2. METHODS

An overview of the simulation workflow is shown in Figure 1a. A diverse set of adsorbate−adsorbent conformers was generated using classical FFs, and their adsorption energies were computed using DFT. An ML model was trained on a
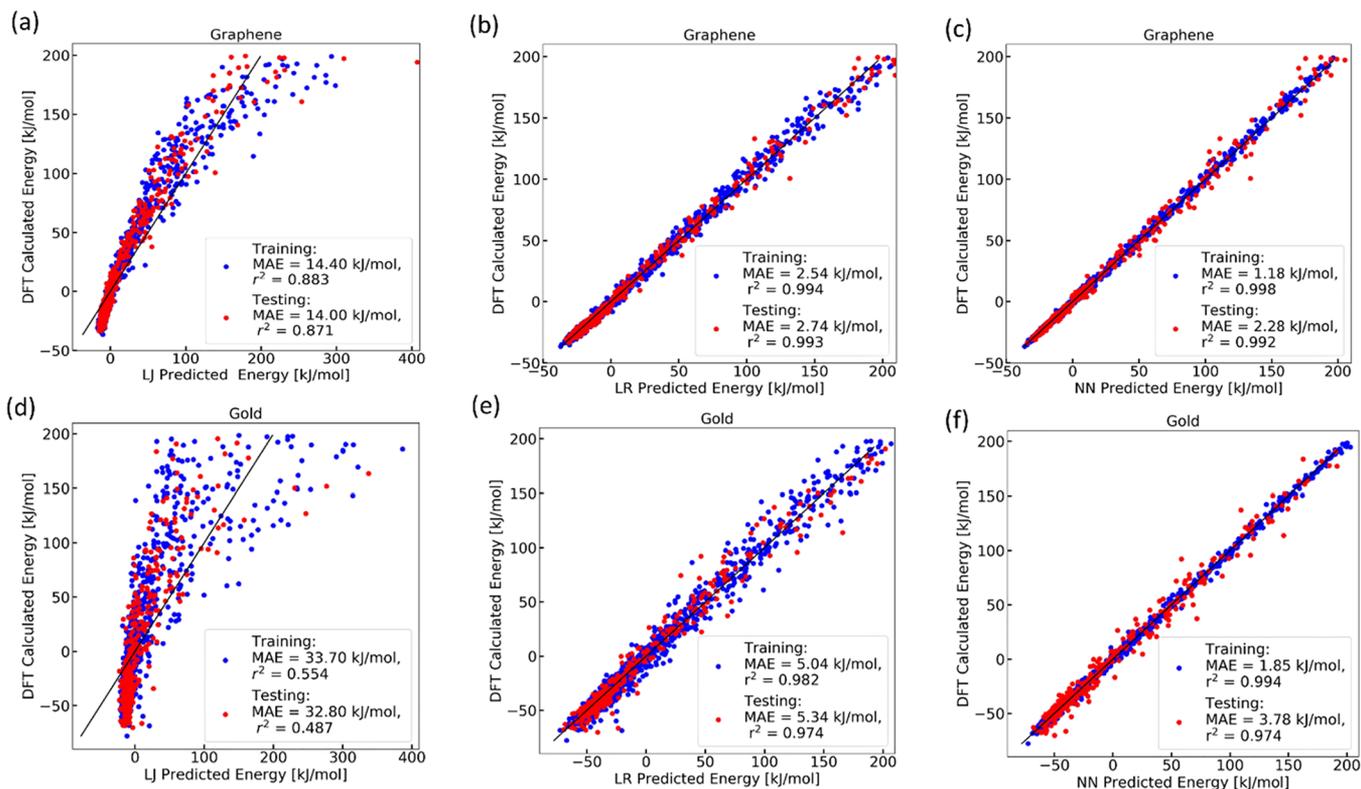
Figure 2. (a, d) Test and training accuracy of Lennard Jones (LJ) potential fitting (LJ models) procedure. Test and training accuracy of the (b, e) linear regression (LR) model [equivalent to a neural network (NN) without a hidden layer] and (c, f) nonlinear NN (with a hidden layer) models.

subset of the DFT data to predict adsorption energies as a function of the conformation and to be used as a replacement of the classical FF.

The atomistic model of alanine was built using the TLEAP[32] module of the AmberTools program. The alanine molecule was solvated in a TIP3P water box having dimensions of 15 Å in each direction. The snapshot of the initial system prepared for the MD simulation is shown in Figure S1. The full system was first energy minimized and then equilibrated using an NVT ensemble. A production run of 100 ns was performed thereafter. All MD simulations were done in the GROMACS[33] package using the Amber99SB ILDN[34] FF. For more details of the MD simulation protocol, see Supporting Information. Approximately 1000 alanine conformations were extracted from the production run. Alanine conformers were selected randomly from this set and placed on top of Au (111) or graphene surfaces (see Figure 1b). The graphene surface was modeled as a single layer with the dimension ∼27/25 Å in x y directions. We built two layers (each layer having dimensions of ∼26/25 Å in x y directions) of gold atoms in the (111) direction to model the Au (111) surface. The atomistic model of the graphene and the gold is shown in Figure S2. The MD generated alanine conformers were also additionally rotated to generate new conformers, which are often not sampled in simple unbiased simulation. To place the alanine conformers on top of the surfaces, we sampled the x, y, and z direction from the surface center and placed the center of geometry of the conformers at these scanned coordinates. During the scan, x and y positions were kept only within one unit cell, while for the z position, we sampled distances of up to 5 Å from the surface. A quick calculation of the classical interaction energy (between the alanine and surface) was performed (using GAFF

FF[35]) with the generated system geometries (alanine and surface). Only geometries with classical interaction energies smaller than 1000 kJ/mol were used for the DFT calculation to avoid expensive DFT calculations with unphysical geometries. To measure the interaction energy between the alanine and the surface for each geometry, three DFT single point energy calculations were performed: isolated alanine, isolated surface (has to be computed only once), and joint system of alanine and the surface. All the DFT calculations were performed with Perdew−Burke−Ernzerhof (PBE) exchange correlation function[36,37] with DFT D3 dispersion[38,39] correction as implemented in VASP.[40−43] For more details on the DFT calculation methodology, see the Supporting Information.

3. RESULTS AND DISCUSSION

3.1. LJ Potentials. We performed interaction energy calculations with ∼2000 different system geometries. The full DFT data were split (3:1 ratio) into two sets of ∼1500 and ∼500 geometries. In this paper, we always use the larger data set with ∼1500 data points (training data set) to train different models, while the smaller data set with ∼500 data points (test data set) were used to evaluate the model after training.

We first tried to fit the DFT calculated energies with the widely used LJ type pair potentials of the form

$$U = \sum_{ij} \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{1}$$

The sum in eq 1 is over all pairs of atoms between alanine and the surface. Here, r_{ij} is the distance between two atoms i and j. We use $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ and $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$, where σ_i and ε_i are the

LJ parameters of atom i. We used element specific LJ parameters in this case. So, if atom i is a carbon atom, then $\sigma_i = \sigma_C$ and $\varepsilon_i = \varepsilon_C$. In total, we had eight fitting parameters in the case of the alanine on graphene system (σ and ε for hydrogen, carbon, oxygen, and nitrogen). For the alanine on gold system, we had 2 additional fitting parameters for gold, leading to 10 fitting parameters in total.

The training set was used to fit the pairwise potential according to eq 1. The best fitted parameters were further used to evaluate the energy for the smaller test set. A comparison of the DFT calculated energy and the LJ fitted energy of the training and test sets is shown in Figure 2a,d. To evaluate the accuracy of the fit, we calculated the MAE and accuracy (r^2) of the fitted energy with respect to the DFT calculated energy. The prediction accuracy in the case of the graphene surface was much better (MAE $\sim$ 14 kJ/mol and r^2 = 0.871) than in the case of the gold surface (MAE of $\sim$32 kJ/mol and r^2 = 0.487). However, in both cases, the LJ pairwise potential was not accurate enough to represent the DFT calculated energies and therefore cannot be used as a possible FF for any simulations. To obtain a better prediction accuracy without the need to manually add additional terms to the LJ equation (or assuming an altogether different functional form[44−46]), we trained ML models to predict the DFT calculated energies as a function of the system geometries.

We further compare the DFT calculated energies of the conformers with the energies calculated using the GolP Charmm[22] FF. The GolP Charmm FFs, which were para meterized to reproduce the optimized adsorption geometries and the corresponding energies, do not match well with the DFT calculated energies for the other off equilibrium con formations (see Figure 3 below), resulting in a very high MAE
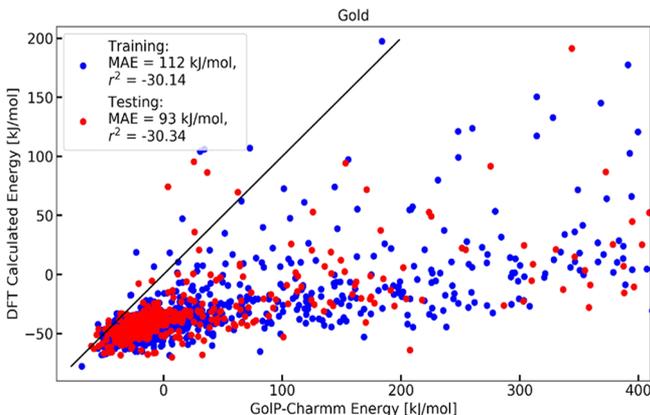

Figure 3. Comparison between the DFT calculated energy with the energies calculated using GolP Charmm FF. The performance of GolP Charmm is compared for both the training (shown in blue) and test data sets (shown in red). For the calculation of MAE (and r^2), energies (predicted by GoIP Charmm) up to 1000 kJ/mol are considered.

and low r^2 values. However, part of the discrepancy between the GolP Charmm and the DFT calculated energies reported in this article (using the PBE D3 method) may also be[47] due to different DFT methodologies used in GoIP Charmm (uses the vdW DF method). However, we expect that one of the main reasons for large deviations between GoIP and DFT is the errors due to the virtual atoms in GoIP Charmm, which can come close to atoms of the adsorbate.

3.2. Linear Regression. Before proceeding to the complex nonlinear NN based models, we first check whether a linear pairwise model with much more than 8 or 10 parameters (as used in the LJ case) can actually fit the DFT calculated energy, or whether a more complex model including higher order (e.g., three body) interactions is needed to accurately predict DFT energies.

Therefore, we first train a linear regression (LR) model (equivalent to an NN without a hidden layer). We calculate the inverse distance ($1/r_{ij}$) of each atom of the alanine from the closest three atoms of the surface. The different algebraic powers of these inverse distances with an integer exponent were used as the feature vector of the LR model. If $\{1/r_{ij}\}$ is the set of all the inverse distances, then the input of the LR was constructed as $\left\{ \left(\frac{1}{r_{ij}} \right)^{\alpha}, \left(\frac{1}{r_{ij}} \right)^{\beta} \right\}$. Here, α and β are two integers with values between 1 and 14. The same data set which was used to fit the LJ potential was used to train the model. We used all possible combinations of α and β to test the accuracy of the resultant linear model by calculating the MAE of the predicted energy with respect to the DFT calculated energy for the test data set. As shown in Figure 4, for both graphene and
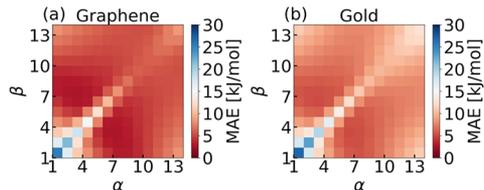

Figure 4. Mean absolute error (MAE) of the test data set for the linear NN (without any hidden layers) model for (a) graphene and (b) gold with different choices of input features of the form $\left\{ \left(\frac{1}{r_{ij}} \right)^{\alpha}, \left(\frac{1}{r_{ij}} \right)^{\beta} \right\}$. The lowest MAE values are 2.74 kJ/mol for graphene at alpha = 3 and beta = 7 and 5.34 kJ/mol for gold at α = 3 and β = 7.

gold, the lowest MAE is obtained when the 3rd and 7th power of the inverse distances were used as input features ($\left\{ \left(\frac{1}{r_{ij}} \right)^{3}, \left(\frac{1}{r_{ij}} \right)^{7} \right\}$) for the NN. The performance of the linear model for this optimal choice of features is shown in Figure 2b,e.

3.3. NNs with Hidden Layers. In the next step, we use the optimized input features $\left\{ \left(\frac{1}{r_{ij}} \right)^{3}, \left(\frac{1}{r_{ij}} \right)^{7} \right\}$ to train a complex nonlinear NN with hidden layers. We used a NN with 3 hidden layers and 1000 nodes in each layer. We used a leaky ReLU activation function for all fully connected hidden layers, while a linear activation function was used for the output layer. We used the L2 regularization scheme to prevent overfitting of the training data and the Adam optimizer and mean squared error loss function for the model training. Optimized values of the hyperparameters were found by a grid search of the parameter space using a training:validation:test split of 1500:250:250 datapoints. To evaluate the performance of the trained model, we calculated the correlation coefficient r^2 and MAE of the NN predicted energies compared to the DFT calculated ones. The learning curve, that is, the test set performance as a function of the number of training points is

shown in Figure 5. We find a fast improvement of r^2 and MAE up to 500 training points, followed by flattening behavior. In
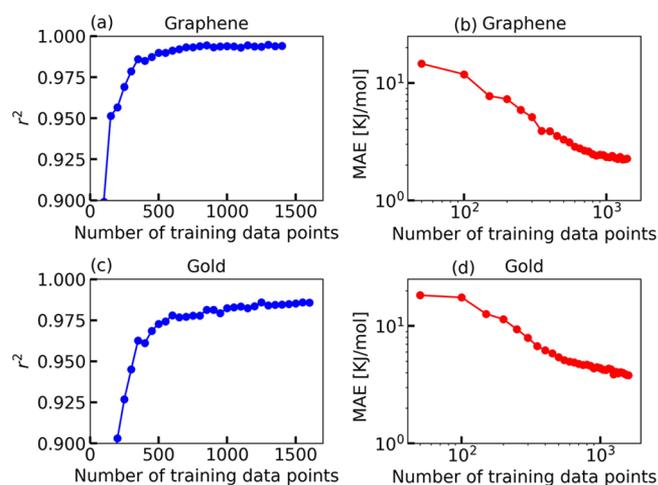

Figure 5. Learning curve of the nonlinear NN (with a hidden layer): r^2 value and MAE on the test set; the ML model (NN) for alanine on (top panel) graphene and (bottom panel) gold.

the case of graphene surfaces, the MAE seems to saturate toward training set sizes of 1500 datapoints, while in the case of gold surfaces, the MAE still decreases logarithmically. A comparison of the r^2 and MAE values for both the training and test data sets is shown in Figure S3.

The comparison of the performance of this NN model with the LR model (NN model without a hidden layer) and the LJ fit is shown in Figure 2c,f. The same training (∼1500 data points) and test data sets (∼500 data points) were used for all the three models (LJ, LR, and NN) for this comparison. The performance of the trained NN model for a gold surface on another (including molecule conformers at longer distances from the gold surface) test data set is shown in Figure S4.

In the case of both graphene and gold, the nonlinear NN model could fit the DFT calculated energy highly accurately, resulting in an MAE value of <4.2 kJ/mol. It is worth mentioning that the accuracy reported in this paper is always relative to the accuracy of the data generating methods, that is, DFT. We estimate the accuracy of DFT to be ∼4.2 kJ/mol, which means that our NN based FFs cannot become more accurate than that. However, it was previously shown that transfer and delta learning methods can be used to increase the

accuracy of ML methods beyond DFT accuracy with relatively small data sets calculated with more accurate reference methods.[48]

3.4. Role of n-Body Interactions for Graphene and Gold.
To compare the performance of the different models in the case of both graphene and gold surfaces, we plot the MAE (Figure 6a) and r^2 values (Figure 6b) of the test data. An error bar in the MAE (and r^2) was estimated by calculating the standard deviation from 10 MAEs (and r^2) obtained from 10 different train−test split on the entire data set. The corresponding values are also reported in Table 1.

As evident from Figure 6, the linear fitting of the DFT energies (LR model) works well in the case of graphene. The introduction of nonlinear NN models (deep NNs) only slightly improves (see Figure 6 and Table 1) the quality of fitting (17% reduction in MAE). In the case of gold, the linear model shows relatively poor performance in comparison to the graphene case, which improves the error by 29%. The use of a deep NN improves the quality of fit, bringing down the MAE value to less than 4.2 kJ/mol. The inadequacy of the LR model in the case of gold points toward the n body contribution to the DFT calculated interaction energies. The LJ fit as well as the LR model only can approximate the energy with two body interaction terms and therefore fail when the energy has three body (or higher order) contributions in it. Higher order interactions can implicitly be constructed by the hidden layers of the NN by combining the two body potential terms that are used as an input. The deeper the NN is, the more complex are the interaction terms it can construct and use for the prediction. The improvement of the NN model compared to the linear model in the case of gold surfaces may indicate that the interaction energy between gold and alanine probably requires three body (or higher order) contributions, and therefore, classical pairwise potential may not be adequate to reproduce the energy in an accurate manner.

4. CONCLUSIONS

The phenomena of adsorption are prevalent in almost all branches of research, especially in material, chemical, and bioengineering.[2,5,49] This has attracted the interest of the research community to understand the adsorption phenomena on an atomic scale. The computational tools to study adsorption within full quantum mechanical treatment suffer from its limitation to access time scales longer than a few picoseconds, while the classical MD simulation is crippled by
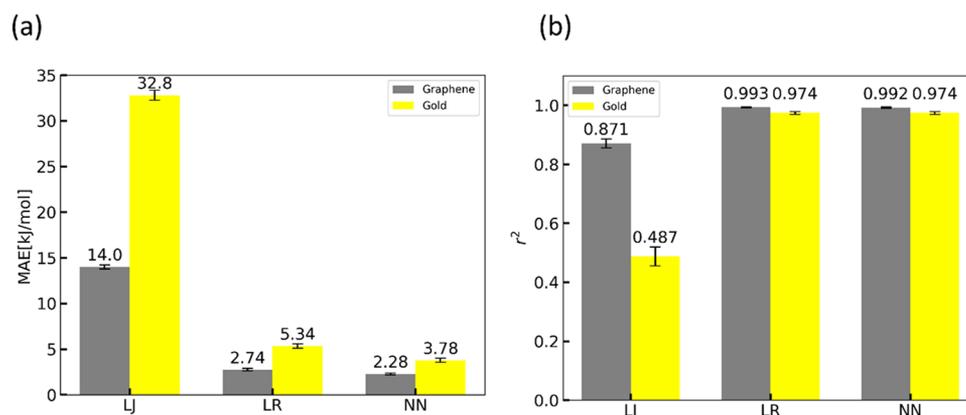
(a) (b)

Figure 6. Mean absolute errors (MAEs) (a) and r^2 values (b) for LJ, LR, and NN models.

Table 1. Mean Absolute Errors (MAEs) and r^2 Values for LJ Potentials, Linear Models, and NN Models

	MAE [kJ/mol]/r^2		
surface	LJ	linear model (2-body)	NN model (n-body)
graphene	14.00 ± 0.22/0.871 ± 0.015	2.74 ± 0.14/0.993 ± 0.001	2.28 + 0.10/0.992 ± 0.002
gold	32.80 ± 0.55/0.487 ± 0.032	5.34 ± 0.25/0.974 ± 0.005	3.78 ± 0.23/0.974 ± 0.005

the accuracy of the existing FFs. In this article, we propose a way to generate flexible FFs with high accuracy (MAE < 4.2 kJ/mol) to model adsorption by training artificial NNs with the DFT calculated adsorption energy.

As a proof of concept, we studied the adsorption of amino acid alanine on the graphene and gold (111) surface and train artificial NNs to reproduce the DFT calculated adsorption energy with accuracies of 2.28 and 3.78 kJ/mol for graphene and gold systems, respectively. While a single DFT calculation (including the gold surface) takes ~5 h of computation time in a single node with 20 cores (i.e., 100 core hours), the training of the NN only takes approximately 3 min on a single CPU core. In search for the optimal features for the ML models, we found out that the 3rd and 7th power of the inverse distance (alanine atoms to the graphene/gold surface atoms) provide the best performance of the ML models. Interestingly, the pairwise potentials used in the traditional FFs use a different power (e.g., 6 and 12 for LJ potential) of the inverse distances rather than 3 and 7. In the end, we provide an interesting insight in terms of the contribution of the three body (or higher order) interaction in the alanine−surface interaction. Although alanine−graphene interaction is well reproduced by pairwise interaction terms, there is a significant contribution of three body and higher order terms in the alanine−gold interaction.

In this work, we demonstrated the generation of NN based models only for adsorption energies and not for the forces. An accurate energy model will allow us to perform simulations using our in house developed code SIMONA,[31] which uses a Monte Carlo based approach for sampling the phase space of atomistic systems. In principle, the NN approach can be extended to also obtain forces, either computing the analytic gradient of the energy[50,51] or training a separate ML model for force prediction.

As a future direction, we plan to use this methodology to generate FFs for more complex molecule−surface interactions. An implementation of these NN based FFs in the MD simulation packages is already in progress at the moment.

ASSOCIATED CONTENT

● Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00506.

> MD simulation details, DFT calculation details, test and training accuracy of the NN models, and performance of the trained NN model on another (including molecule conformers at a longer distance from the gold surface) test data set (PDF)
>
> Codes for ML models and conformer scan and training and test data for the ML models (ZIP)

AUTHOR INFORMATION

Corresponding Authors

Pascal Friederich − *Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), Eggenstein Leopoldshafen 76344, Germany; Institute of Theoretical Informatics (ITI), Karlsruhe Institute of Technology (KIT), Karlsruhe 76131, Germany;* orcid.org/0000 0003 4465 1465; Email: pascal.friederich@kit.edu

Wolfgang Wenzel − *Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), Eggenstein Leopoldshafen 76344, Germany;* orcid.org/0000 0001 9487 4689; Email: wolfgang.wenzel@kit.edu

Authors

Saientan Bag − *Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), Eggenstein Leopoldshafen 76344, Germany;* orcid.org/0000 0003 1000 7719

Manuel Konrad − *Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), Eggenstein Leopoldshafen 76344, Germany*

Tobias Schlöder − *Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), Eggenstein Leopoldshafen 76344, Germany;* orcid.org/0000 0002 3822 630X

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Nazaroff, W. W.; Goldstein, A. H. Indoor chemistry: research opportunities and challenges. *Indoor Air* **2015**, *25*, 357−361.

(2) Yang, R. T.: Equilibrium Adsorption of Gas Mixtures. In *Gas Separation by Adsorption Processes*; Yang, R. T., Ed.; Butterworth Heinemann, 1987; 49−100.

(3) Webb, J. A.; Bardhan, R. Emerging advances in nanomedicine with engineered gold nanostructures. *Nanoscale* **2014**, *6*, 2502−2530.

(4) Chakraborty, M.; Jain, S.; Rani, V. Nanotechnology: Emerging Tool for Diagnostics and Therapeutics. *Appl. Biochem. Biotechnol.* **2011**, *165*, 1178−1187.

(5) Kumar, A.; Ma, H.; Zhang, X.; Huang, K.; Jin, S.; Liu, J.; Wei, T.; Cao, W.; Zou, G.; Liang, X. J. Gold nanoparticles functionalized with therapeutic and targeted peptides for cancer treatment. *Biomaterials* **2012**, *33*, 1180−1189.

(6) Desai, T. A.; Hansford, D. J.; Kulinsky, L.; Nashat, A. H.; Rasi, G.; Tu, J.; Wang, Y.; Zhang, M.; Ferrari, M. Nanopore Technology for Biomedical Applications. *Biomed. Microdevices* **1999**, *2*, 11−40.

(7) Care, A.; Bergquist, P. L.; Sunna, A. Solid binding peptides: smart tools for nanobiotechnology. *Trends Biotechnol.* **2015**, *33*, 259−268.

(8) Katevatis, C.; Fan, A.; Klapperich, C. M. Low concentration DNA extraction and recovery using a silica solid phase. *PLoS One* **2017**, *12*, No. e0176848.

(9) Bhattacharyya, A.; Klapperich, C. M. Thermoplastic Microfluidic Device for On Chip Purification of Nucleic Acids for Disposable Diagnostics. *Anal. Chem.* **2006**, *78*, 788–792.

(10) Desai, T. A.; Sharma, S.; Walczak, R. J.; Boiarski, A.; Cohen, M.; Shapiro, J.; West, T.; Melnik, K.; Cosentino, C.; Sinha, P. M.; Ferrari, M.: Nanoporous Implants for Controlled Drug Delivery. In *Bio MEMS and Biomedical Nanotechnology: Volume III Therapeutic Micro/Nanotechnology*; Ferrari, M., Desai, T., Bhatia, S., Eds.; Springer US: Boston, MA, 2007, 263–286.

(11) Malakootian, M.; Nouri, J.; Hossaini, H. Removal of heavy metals from paint industry's wastewater using Leca as an available adsorbent. *Int. J. Environ. Sci. Technol.* **2009**, *6*, 183–190.

(12) Fang, Y.; Riahi, S.; McDonald, A. T.; Shrestha, M.; Tobias, D. J.; Grassian, V. H. What Is the Driving Force behind the Adsorption of Hydrophobic Molecules on Hydrophilic Surfaces? *J. Phys. Chem. Lett.* **2019**, *10*, 468–473.

(13) Acres, R. G.; Cheng, X.; Beranová, K.; Bercha, S.; Skála, T.; Matolín, V.; Xu, Y.; Prince, K. C.; Tsud, N. An experimental and theoretical study of adenine adsorption on Au(111). *Phys. Chem. Chem. Phys.* **2018**, *20*, 4688–4698.

(14) Piana, S.; Bilic, A. The Nature of the Adsorption of Nucleobases on the Gold [111] Surface. *J. Phys. Chem. B* **2006**, *110*, 23467–23471.

(15) Lazar, P.; Karlický, F.; Jurečka, P.; Kocman, M.; Otyepková, E.; Šafářová, K.; Otyepka, M. Adsorption of Small Organic Molecules on Graphene. *J. Am. Chem. Soc.* **2013**, *135*, 6372–6377.

(16) Hughes, Z. E.; Wei, G.; Drew, K. L. M.; Colombi Ciacchi, L.; Walsh, T. R. Adsorption of DNA Fragments at Aqueous Graphite and Au(111) via Integration of Experiment and Simulation. *Langmuir* **2017**, *33*, 10193–10204.

(17) Shi, B.; Shin, Y. K.; Hassanali, A. A.; Singer, S. J. DNA binding to the silica surface. *J. Phys. Chem. B* **2015**, *119*, 11030–11040.

(18) Emami, F. S.; Puddu, V.; Berry, R. J.; Varshney, V.; Patwardhan, S. V.; Perry, C. C.; Heinz, H. Force field and a surface model database for silica to simulate interfacial properties in atomic resolution. *Chem. Mater.* **2014**, *26*, 2647–2658.

(19) Lowe, B. M.; Skylaris, C. K.; Green, N. G.; Shibuta, Y.; Sakata, T. Molecular dynamics simulation of potentiometric sensor response: the effect of biomolecules, surface morphology and surface charge. *Nanoscale* **2018**, *10*, 8650–8666.

(20) Varghese, N.; Mogera, U.; Govindaraj, A.; Das, A.; Maiti, P. K.; Sood, A. K.; Rao, C. N. R. Binding of DNA Nucleobases and Nucleosides with Graphene. *ChemPhysChem* **2009**, *10*, 206–210.

(21) Bag, S.; Rauwolf, S.; Suyetin, M.; Schwaminger, S. P.; Wenzel, W.; Berensmeier, S. Buffer Influence on the Amino Acid Silica Interaction. *ChemPhysChem* **2020**, *21*, 2347–2356.

(22) Wright, L. B.; Rodger, P. M.; Corni, S.; Walsh, T. R. GolP CHARMM: First Principles Based Force Fields for the Interaction of Proteins with Au(111) and Au(100). *J. Chem. Theory Comput.* **2013**, *9*, 1616–1630.

(23) Rosa, M.; Corni, S.; Di Felice, R. Enthalpy–Entropy Tuning in the Adsorption of Nucleobases at the Au(111) Surface. *J. Chem. Theory Comput.* **2014**, *10*, 1707–1716.

(24) Hughes, Z. E.; Tomásio, S. M.; Walsh, T. R. Efficient simulations of the aqueous bio interface of graphitic nanostructures with a polarisable model. *Nanoscale* **2014**, *6*, 5438–5448.

(25) Geada, I. L.; Ramezani Dakhel, H.; Jamil, T.; Sulpizi, M.; Heinz, H. Insight into induced charges at metal surfaces and biointerfaces using a polarizable Lennard–Jones potential. *Nat. Commun.* **2018**, *9*, 716.

(26) Bonati, L.; Parrinello, M. Silicon Liquid Structure and Crystal Nucleation from Ab Initio Deep Metadynamics. *Phys. Rev. Lett.* **2018**, *121*, No. 265701.

(27) Zhang, L.; Han, J.; Wang, H.; Car, R.; Weinan, E. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*, No. 143001.

(28) Chan, H.; Narayanan, B.; Cherukara, M. J.; Sen, F. G.; Sasikumar, K.; Gray, S. K.; Chan, M. K. Y.; Sankaranarayanan, S. K. R. S. Machine Learning Classical Interatomic Potentials for Molecular Dynamics from First Principles Training Data. *J. Phys. Chem. C* **2019**, *123*, 6941–6957.

(29) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI 1: an extensible neural network potential with DFT accuracy at force field computa tional cost. *Chem. Sci.* **2017**, *8*, 3192–3203.

(30) Friederich, P.; Konrad, M.; Strunk, T.; Wenzel, W. Machine learning of correlated dihedral potentials for atomistic molecular force fields. *Sci. Rep.* **2018**, *8*, 2559.

(31) Penaloza Amion, M.; Sedghamiz, E.; Kozlowska, M.; Degitz, C.; Possel, C.; Wenzel, W. Monte Carlo Simulations of Soft Matter Using SIMONA: A Review of Recent Applications. *Front. Phys.* **2021**, *9*, No. 635959.

(32) Case, D.; Darden, T.; Cheatham, III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W. *AmberTools*; University of California: San Francisco, CA, 2008.

(33) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(34) Lindorff Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.

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

(36) Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron gas correlation energy. *Phys. Rev. B* **1992**, *45*, 13244–13249.

(37) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(38) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(39) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT D) for the 94 elements H Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(40) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total energy calculations using a plane wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.

(41) Kresse, G.; Furthmüller, J. Efficiency of ab initio total energy calculations for metals and semiconductors using a plane wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.

(42) Kresse, G.; Hafner, J. Ab initio molecular dynamics simulation of the liquid metal amorphous semiconductor transition in germa nium. *Phys. Rev. B* **1994**, *49*, 14251–14269.

(43) Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **1993**, *47*, 558–561.

(44) Mercado, R.; Vlaisavljevich, B.; Lin, L. C.; Lee, K.; Lee, Y.; Mason, J. A.; Xiao, D. J.; Gonzalez, M. I.; Kapelewski, M. T.; Neaton, J. B.; Smit, B. Force Field Development from Periodic Density Functional Theory Calculations for Gas Separation Applications Using Metal–Organic Frameworks. *J. Phys. Chem. C* **2016**, *120*, 12590–12604.

(45) Haldoupis, E.; Borycz, J.; Shi, H.; Vogiatzis, K. D.; Bai, P.; Queen, W. L.; Gagliardi, L.; Siepmann, J. I. Ab Initio Derived Force Fields for Predicting CO2 Adsorption and Accessibility of Metal Sites in the Metal–Organic Frameworks M MOF 74 (M = Mn, Co, Ni, Cu). *J. Phys. Chem. C* **2015**, *119*, 16058–16071.

(46) Latorre, C. A.; Ewen, J. P.; Gattinoni, C.; Dini, D. Simulating Surfactant–Iron Oxide Interfaces: From Density Functional Theory to Molecular Dynamics. *J. Phys. Chem. B* **2019**, *123*, 6870–6881.

(47) Hörmann, L.; Jeindl, A.; Hofmann, O. T. Reproducibility of potential energy surfaces of organic/metal interfaces on the example of PTCDA on Ag(111). *J. Chem. Phys.* **2020**, *153*, 104701.

(48) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching coupled cluster accuracy with a general purpose neural network potential through transfer learning. *Nat. Commun.* **2019**, *10*, 2903.

(49) Zhang, Y.; Zhang, Y.; Burke, J. M.; Gleitsman, K.; Friedrich, S. M.; Liu, K. J.; Wang, T. H. A Simple Thermoplastic Substrate Containing Hierarchical Silica Lamellae for High Molecular Weight DNA Extraction. *Adv. Mater.* **2016**, *28*, 10630−10636.

(50) Chen, W. K.; Liu, X. Y.; Fang, W. H.; Dral, P. O.; Cui, G. Deep Learning for Nonadiabatic Excited State Dynamics. *J. Phys. Chem. Lett.* **2018**, *9*, 6702−6708.

(51) Gastegger, M.; Marquetand, P. High Dimensional Neural Network Potentials for Organic Reactions and an Improved Training Algorithm. *J. Chem. Theory Comput.* **2015**, *11*, 2187−2198.