ACS PHYSICAL CHEMISTRY Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/physchemau

Article

Sparse Gaussian Process Regression-Based Machine Learned First-Principles Force-Fields for Saturated, Olefinic, and Aromatic Hydrocarbons

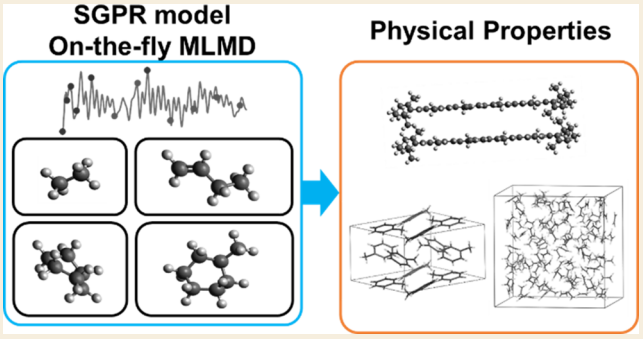
Miran Ha,[†] Amir Hajibabaei,[†] Saeed Pourasad, and Kwang S. Kim*

Cite This: *ACS Phys. Chem Au* 2022, 2, 260–264

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Universal machine learning (ML) interatomic potentials (IAPs) for saturated, olefinic, and aromatic hydrocarbons are generated by using the Sparse Gaussian process regression algorithm. The universal potentials are obtained by combining the potentials for the previously trained alkane/polyene systems and the potentials generated with the presently trained cyclic/aromatic hydrocarbon systems, along with the newly trained cross-terms between the two systems. The ML-IAPs have been trained using the PBE + D3 level of density functional theory for the on-the-fly adaptive sampling of various hydrocarbon molecules and these clusters composed of small molecules. We tested the ML-IAPs and found that they correctly predicted the structures and energies of the β-carotene monomer and dimer. Also, the simulations of liquid ethylene reproduced the molecular volume and the simulations of toluene crystals reproduced higher stability of the α-phase over the β-phase. These ab initio-level force-fields could eventually evolve toward universal organic/polymeric/biomolecular systems.



KEYWORDS: *hydrocarbons, machine learning, inter-atomic potential, SGPR, on-the-fly, organic molecule*

INTRODUCTION

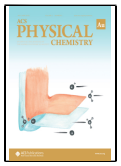
Conventional classical molecular force-fields[1−6] have been widely used for large-scale computational studies of organic/biomolecular and macromolecular systems including drugs and polymers which are not easily tractable with first-principles calculations. Nevertheless, their accuracy has been limited due to their deficiencies in describing various properties with correct structures simultaneously. In particular, the potential energy is expressed in terms of "descriptors" such as bonds, angles, and so forth which are topology-dependent. Therefore, their applicability has been limited to nonreactive systems.

Over the past decade, machine learning interatomic potentials (ML-IAPs) representing ab initio potential energy surfaces have been introduced based on neural networks (NNs) and kernel regression methods.[7−25] Most of these potential energies are dependent on the environments of each atom and so need to be trained at every situation and at every instant that the atomic environment changes significantly. Here, we utilize a one-to-one mapping from the coordinate space to the descriptor space as an only local configuration-dependent generalized form reflecting many-body correlations while being invariant with respect to translations/rotations of the physical system and permutations of identical atoms.[26]

Sampling by active/on-the-fly learning is conducted by using a measure for the uncertainty of the model at given input points. The uncertainty prediction is straightforward with Bayesian inference methods. Here, the sparse Gaussian process regression (SGPR) algorithm[20,21,24,25] as a Bayesian approach to regression is implemented for representing the potential energy surface as well as uncertainty predictions. Using the SGPR algorithm, the models in this study are generated in total with around 1000 training samples which are much less than the initial data needed for generating the seed NNs. Among Bayesian methods, SGPR scales as $O(n)$, while retaining the important characteristics of Gaussian process regression (GPR)[8,27,28] having the computational cost of $O(n^3)$ with the size of training data (n). Thus, SGPR is ideal for on-the-fly sampling of optimal data sets for generating ML-IAPs.

Hydrocarbons are a diverse family of organic molecules relevant to most of the macromolecules and biological complexes. Recently, the ML force fields based on GPR have been developed for predicting the energies and forces of

Received: December 31, 2021
Revised: February 4, 2022
Accepted: February 7, 2022
Published: February 17, 2022



https://doi.org/10.1021/acsphyschemau.1c00058
ACS Phys. Chem Au 2022, 2, 260–264

alkanes and proteins and performing on-the-fly molecular dynamics (MD) simulations.[29,30] Previously, we considered ML-IAPs using the SGPR algorithm[25] for a large subcategory of alkane, alkene, and polyene hydrocarbons as an initiative toward a universal ab initio quality force-field for these molecules. Here, we consider not only saturated, olefinic, and aromatic hydrocarbons which include alkane, alkene, alkadiene, and alkatriene molecules with branches but also cyclic saturated/unsaturated and aromatic hydrocarbons and further their combined structures.

METHODS

Machine Learning Potentials

A configuration x of N atoms is compiled to a list of descriptors $x = \{\rho_i\}_{i=1}^{N}$ where ρ_i is a rotationally invariant descriptor for the local chemical environment of atom i, which depends only on the relative coordinates of $N - 1$ atoms within a cutoff radius. In kernel-based regression methods, the potential energy becomes

$$E(x) = \sum_{i=1}^{N}\sum_{j=1}^{m} K(\rho_i, \chi_j)w_j$$

where $z = \{\chi_j\}_{j=1}^{m}$ is the set of inducing descriptors, K is a covariance kernel, and $\mathbf{w} = \{w_j\}_{j=1}^{m}$ is the vector of weights for the inducing descriptors. The weights, which depend on the regression algorithm, are obtained such that the potential energy and forces are reproduced for a set of ab initio data $X = \{x_k\}_{k=1}^{n}$. The inducing descriptors are often subset descriptors extracted from X.

In SGPR,[18−21] $\mathbf{w} = (\sigma^2\mathbf{k}_{mm} + \mathbf{k}_{nm}^{T}\mathbf{k}_{nm})^{-1}\mathbf{k}_{nm}^{T}\mathbf{Y}$ where $\mathbf{k}_{mm}$ and $\mathbf{k}_{nm}$ are the inter-inducing (z−z) and data-inducing (X−z) covariance matrices, $\mathbf{Y}$ is the data of potential energies (and forces), and σ is the noise hyperparameter. The noise scale σ and other possible hyperparameters in kernel K are optimized to maximize the likelihood of the data energies. For the similarity kernel, we use a variation of the smooth overlap of atomic positions[26] which we have defined in refs 20 and 21. For the existing data, the inducing descriptors are sampled from the data; otherwise, with on-the-fly learning, both the data and inducing descriptors are sampled during MD. The uncertainty prediction and its utilization for the on-the-fly learning algorithm are discussed in detail in ref 25.

On-the-Fly MD Simulation

To obtain ML potentials, we carried out MD simulations with an adaptive sampling algorithm.[23] It learned the potential energy surfaces from density functional theory (DFT) calculations which were performed using the Vienna ab initio simulation package[31] with Perdew−Burke−Ernzerhof (PBE) functionals[32] and van der Waals interactions (D3).[33] The projector augmented wave[34] pseudopotentials with 400 eV energy cutoff were used. The convergence criterion for the electronic energy difference was set to 10^{-4} eV. The molecules were placed in the center of a cubic cell in which the interlayer distance between the molecules and the image is 15 Å vacuum, and the Brillouin zone was sampled using the Γ-point. We considered various conformers of given molecules and performed *NVT* MD simulations using a Nosé−Hoover thermostat and Parrinello−Rahman dynamics as implemented in the atomic simulation environment package.[35] The MD simulations were run for 3−6 ps at 300 K with 0.5 fs time step. The coefficient of determination defined by $R^2 = 1 - (f_i - \tilde{f}_i)^2/(f_i - \bar{f})^2$ is exploited for the test of reliability of the SGPR-based first-principles potential/forces, where $\{f_i\}$ are ab initio (PBE + D3) forces, $\bar{f}$ is their average, and $\{\tilde{f}_i\}$ are ML forces.[25]

Owing to the abundance of relevant molecules, trained molecules are split into several groups, and independent SGPR models are trained in parallel using the AUTOFORCE package.[36] In this study, we considered expert models for linear or branched systems (C1−C8 alkane, C3−C10 isoalkane, C2−C6 alkene, C3−C6 branched alkene, C3−C8 alkadiene, and C7−C9 alkatriene), cyclic systems (C3−C10 cycloalkane, C3−C8 cycloalkene, and C4−C10 methyl-/dimethyl-cycloalkane, bicycloalkane), and aromatic systems (benzene, toluene, xylene, trimethylbenzene, and mesitylene) to build the universal model. To consider the intermolecular interactions, expert models for liquid phases of alkane (methane, ethane, propane, and butane), alkene (ethylene), and aromatic (benzene) are also trained and denoted as alkane-inter, alkene-inter, and aromatic-inter, respectively. The learnt expert models are available in gitlab website.[37]

RESULTS

Training Expert Models

On-the-fly MD simulations are performed for 3−6 ps to train expert SGPR models for each molecule (see Method). Here, expert models are newly trained for cyclic hydrocarbons (cycloalkane, cycloalkene, and bicycloalkane) and aromatic systems and then are combined with the previously generated model for alkane, isoalkane, alkene, and alkadiene.[25] Each generated expert model is validated by comparing it with the DFT calculation results (Table 1).

Table 1. Testing the Expert Models in Their Domains[a]

group	N_{test}	energy MAE (meV)	force MAE (eV/Å)	force R^2 testing/training
alkane[b]	448	4.9	0.076	0.978/0.993
isoalkane[b]	308	3.0	0.080	0.977/0.989
alkene[b]	142	8.1	0.110	0.966/0.985
alkadiene[b]	374	5.3	0.110	0.959/0.978
alkatriene	238	3.2	0.108	0.992/0.966
cycloalkane	253	2.3	0.074	0.983/0.990
cycloalkene	77	3.4	0.110	0.963/0.984
methyl/dimethyl-cycloalkane	946	5.2	0.092	0.970/0.989
bicycloalkane	594	2.1	0.075	0.980/0.987
aromatic	83	2.1	0.074	0.997/0.983

[a]N_{test} is the number of test samples and energy is the potential energy per atom. [b]Data from ref 25.

Rotation of CH_3 Group in Ethane

The accuracy of universal ML-IAPs generated by combining expert models is tested first. As the simplest case, nudged elastic band calculation is conducted to predict the rotation barrier of the CH_3 group in ethane. The barrier to rotate the CH_3 group by 120° is calculated using universal ML-IAPs and compared with the PBE + D3 results. As a result, the barrier of the reaction was predicted to be 0.12 eV which is almost the same as the PBE + D3 results showing 0.11 eV and with the barrier of 0.12 eV obtained from the conventional OPLS-AA[3] force fields in which the rotational barriers were optimized to the ab initio computational value.

Dimerization Energy of β-Carotene

The universal ML-IAPs are used to predict the energies of molecules containing various hydrocarbon groups in their structure. Here, the structures and energies of the β-carotene monomer and dimer (Figure 1) are investigated using the universal potential. β-Carotene is an organic molecule having 8 isoprenes and 2 trimethyl-cyclohexenes in its structure. The expert models including C−C single bonds, C=C double bonds, and cyclic hydrocarbons are used to construct the universal model. The relative energies of 20 different conformers of β-carotene monomer are predicted to be close to the PBE + D3 results within a low root-mean-squared error of 0.12 eV. The interaction energy between the two monomers

https://doi.org/10.1021/acsphyschemau.1c00058
ACS Phys. Chem Au 2022, 2, 260−264

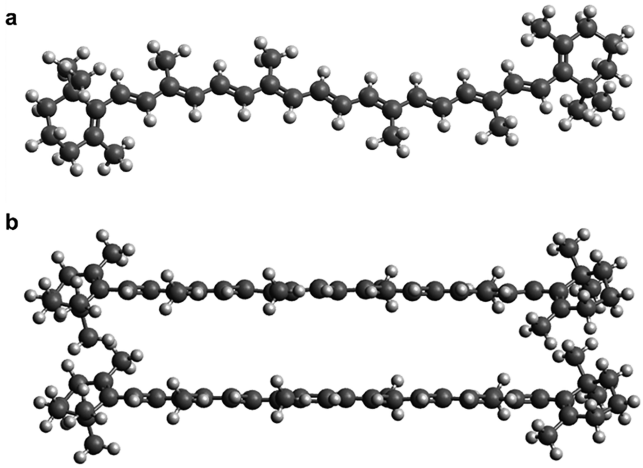


Figure 1. Structures of β-carotene monomer (a) and dimer (b).

in the β-carotene dimer is predicted to be −1.9 eV showing a strong interaction, consistent with the strong interaction by PBE + D3 calculation (−1.1 eV). In contrast, the conventional OPLS-AA[3] and AIREBO[38] force-fields give only insignificant interaction energies of −0.02 and −0.2 eV, respectively.

Liquid Ethylene

We predicted the molecular volume of liquid ethylene using ML-IAPs. The *NPT* MD simulations of 34 ethylene molecules are conducted for 30 ps with a constant pressure of 1 bar and a temperature of 169.5 K. In this case, the volume per molecule v_l is predicted to be 76 ± 6 Å^3 in good agreement with the experimental value[39] of 81.5 Å^3. Some discrepancy from the experimental value seems to arise from the inherent discrepancy in the DFT-D3 calculation.

Toluene Crystal

The prediction of crystal structures is accelerated by using the ML potential trained with DFT potential energies/forces. Here, our universal model is employed to predict the stable structure of the toluene crystal (Figure 2). The α-phase is

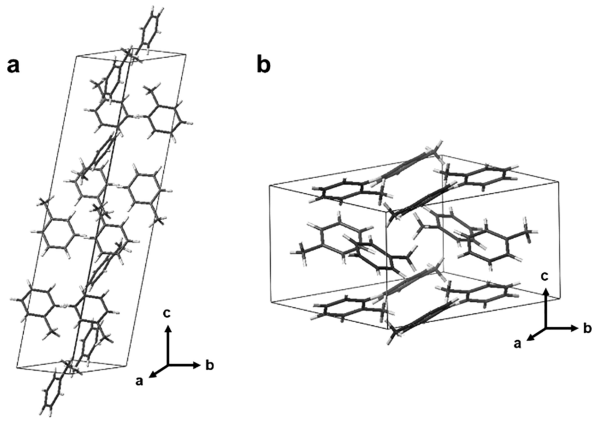


Figure 2. Crystal structures of (a) α-toluene and (b) β-toluene.

predicted to be the most stable state in the crystal, while the β-phase is the metastable state.[40] The predicted total energy of each system is within 0.5% from the PBE + D3 values, and the energy difference between the two phases is predicted as 1.89 eV, which is comparable with the PBE + D3 result showing 1.31 eV. The conventional AIREBO-M[38] force-field gives an energy difference of 2.26 eV.

DISCUSSION AND CONCLUSIONS

In this work, molecular and crystal structures of hydrocarbons are predicted by the ML-IAPs obtained using the SGPR algorithm. The expert models are generated for different hydrocarbon groups from on-the-fly ML and then are combined into the universal ML-IAPs. The universal ML-IAPs which cover most of configurationally important potential energy surfaces are generated from the expert models which trained for ∼90 molecules (∼350 conformers) and 5 bulk phases. It showed excellent transferability for various hydrocarbon systems. The universal ML-IAPs predicted the structure of β-carotene dimer with reasonable accuracy in interaction energy between two monomers. Intermolecular interactions in universal models adequately described the molecular volume of liquid ethylene. They were extended to investigate the phase of the toluene crystal, which showed the stability of α-phase over β-phase. All these results indicate that the universal model is generated efficiently from separately trained atomic potentials for local configurations of various subgroups. Therefore, this method can be utilized to generate universal models for various systems consisting of multi-components, which have not been possible so far for other ML methods.

This study concludes one of the most extensive explorations of the phase space of hydrocarbon molecules in which less than 1000 configurations are sampled as support for the SGPR potential. These configurations are sampled on-the-fly with MD using a Bayesian criterion, and therefore the configurations with redundant information are automatically discarded. It has been shown that, for transferable ML potentials, the quality of data in terms of diversity/completeness is the most important factor rather than quantity.[41] Therefore, the data sampled here can be easily used with any regression algorithm (kernel- or NN-based) in order to produce transferable ML potentials. The small size of ab initio data makes it straightforward to use higher-quality (better than PBE + D3) first-principles methods in the future. At last, it has been shown that structural descriptors with n-body correlations are generally degenerate,[42] and significantly higher accuracy can be achieved with a more complete descriptor scheme. Aside from including higher order correlations, an intriguing concept is the development of "recursively embedded atom NNs"[9] in which the local descriptors are still based on three-body correlations, but nonlocal correlations (which lift the degeneracies) are incorporated using local parameters which are linked to the global structure by a message passing NN. A similar concept can be explored with kernel-based regression methods such as SGPR which is a promising direction for future studies.

AUTHOR INFORMATION

Corresponding Author

Kwang S. Kim − *Center for Superfunctional Materials, Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea;* orcid.org/0000-0002-6929-5359; Email: kimks@unist.ac.kr

https://doi.org/10.1021/acsphyschemau.1c00058
ACS Phys. Chem Au 2022, 2, 260−264

Authors

Miran Ha − *Center for Superfunctional Materials, Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea;* orcid.org/0000-0002-3744-4866

Amir Hajibabaei − *Center for Superfunctional Materials, Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea;* orcid.org/0000-0003-1123-4040

Saeed Pourasad − *Center for Superfunctional Materials, Department of Chemistry, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsphyschemau.1c00058

Author Contributions

†equal contribution.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1I1A1A01050280 and 2021R1I1A1A01050085), the A.I. Incubation Project Fund (1.210091.01) of UNIST, and KISTI (KSC-2021-CRE-0193, KSC-2021-CRE-0195, KSC-2021-CRE-0523, and KSC-2021-CRE-0554).

■ REFERENCES

(1) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.* **1984**, *106*, 6638−6646.

(2) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(3) Siu, S. W. I.; Pluhackova, K.; Böckmann, R. A. Optimization of the OPLS-AA Force Field for Long Hydrocarbons. *J. Chem. Theory Comput.* **2012**, *8*, 1459−1470.

(4) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(5) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(6) 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.

(7) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(8) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001.

(9) Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K.-R.; Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **2019**, *10*, 5024.

(10) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

(11) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3*, No. e1603015.

(12) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 3887.

(13) Seko, A.; Takahashi, A.; Tanaka, I. First-principles interatomic potentials for ten elemental metals via compressed sensing. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92*, 054113.

(14) Ghasemi, S. A.; Hofstetter, A.; Saha, S.; Goedecker, S. Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92*, 045131.

(15) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **2021**, *12*, 398.

(16) Zhang, Y.; Hu, C.; Jiang, B. Embedded Atom Neural Network Potentials: Efficient and Accurate Machine Learning with a Physically Inspired Representation. *J. Phys. Chem. Lett.* **2019**, *10*, 4962−4967.

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

(18) Jinnouchi, R.; Karsai, F.; Kresse, G. On-the-fly machine learning force field generation: Application to melting points. *Phys. Rev. B* **2019**, *100*, 014105.

(19) Jinnouchi, R.; Lahnsteiner, J.; Karsai, F.; Kresse, G.; Bokdam, M. Phase Transitions of Hybrid Perovskites Simulated by Machine-Learning Force Fields Trained on the Fly with Bayesian Inference. *Phys. Rev. Lett.* **2019**, *122*, 225701.

(20) Hajibabaei, A.; Myung, C. W.; Kim, K. S. Sparse Gaussian process potentials: Application to lithium diffusivity in superionic conducting solid electrolytes. *Phys. Rev. B* **2021**, *103*, 214102.

(21) Hajibabaei, A.; Myung, C. W.; Kim, K. S. Towards Universal Sparse Gaussian Process Potentials: Application to Lithium Diffusivity in Superionic Conducting Solid Electrolytes. 2020, arXiv:2009.13179. arXiv.org e-Print archive. https://arxiv.org/abs/2009.13179v1 (accessed 5 Feb 2022).

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

(23) Vandermause, J.; Torrisi, S. B.; Batzner, S.; Xie, Y.; Sun, L.; Kolpak, A. M.; Kozinsky, B. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *npj Comput. Mater.* **2020**, *6*, 20.

(24) Hajibabaei, A.; Kim, K. S. Universal Machine Learning Interatomic Potentials: Surveying Solid Electrolytes. *J. Phys. Chem. Lett.* **2021**, *12*, 8115−8120.

(25) Hajibabaei, A.; Ha, M.; Pourasad, S.; Kim, J.; Kim, K. S. Machine Learning of First-Principles Force-Fields for Alkane and Polyene Hydrocarbons. *J. Phys. Chem. A* **2021**, *125*, 9414−9420.

(26) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 184115.

(27) Foster, L.; Waagen, A.; Aijaz, N.; Hurley, M.; Luis, A.; Rinsky, J.; Satyavolu, C.; Way, M. J.; Gazis, P.; Srivastava, A. Stable and Efficient Gaussian Process Calculations. *J. Mach. Learn. Res.* **2009**, *10*, 857−882.

(28) Williams, C. K. I.; Rasmussen, C. E.; Schwaighofer, A.; Tresp, V. *Observations on the Nyström Method for Gaussian Processes*; University of Edinburgh, 2002. https://homepages.inf.ed.ac.uk/ckiw/online_pubs.html (accessed Feb 5, 2022).

(29) Cheng, Z.; Zhao, D.; Ma, J.; Li, W.; Li, S. An On-the-Fly Approach to Construct Generalized Energy-Based Fragmentation

Machine Learning Force Fields of Complex Systems. *J. Phys. Chem. A* **2020**, *124*, 5007−5014.
(30) Cheng, Z.; Du, J.; Zhang, L.; Ma, J.; Li, W.; Li, S. Building quantum mechanics quality force fields of proteins with the generalized energy-based fragmentation approach and machine learning. *Phys. Chem. Chem. Phys.* **2022**, *24*, 1326−1337.
(31) Kresse, G.; Furthmüller, J. Efficient iterative schemes forab initiototal-energy calculations using a plane-wave basis set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169−11186.
(32) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(33) 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.
(34) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953.
(35) Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Bjerre Jensen, P.; Kermode, J.; Kitchin, J. R.; Leonhard Kolsbjerg, E.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Bergmann Maronsson, J.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The atomic simulation environment-a Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.
(36) *AUTOFORCE code*. https://github.com/amirhajibabaei/AutoForce (accessed Feb. 5, 2022).
(37) *SGPR Models: Hydrocarbons*. https://gitlab.com/amirhajibabaei/AutoForce-pes/-/tree/master/models/hydrocarbon (accessed Feb 5, 2022).
(38) O'Connor, T. C.; Andzelm, J.; Robbins, M. O. AIREBO-M: A reactive model for hydrocarbons at extreme pressures. *J. Chem. Phys.* **2015**, *142*, 024903.
(39) Maass, O.; Wright, C. H. Some physical properties of hydrocarbons containing two and three carbon atoms. *J. Am. Chem. Soc.* **1921**, *43*, 1098−1111.
(40) Andre, D.; Fourme, R.; Bruneaux-Poulle, J.; Bosio, L. Crystal structure of the metastable β-phase of toluene. *J. Mol. Struct.* **1982**, *81*, 253−259.
(41) Zuo, Y.; Chen, C.; Li, X.; Deng, Z.; Chen, Y.; Behler, J.; Csányi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; Ong, S. P. Performance and Cost Assessment of Machine Learning Interatomic Potentials. *J. Phys. Chem. A* **2020**, *124*, 731−745.
(42) Pozdnyakov, S. N.; Willatt, M. J.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M. Incompleteness of Atomic Structure Representations. *Phys. Rev. Lett.* **2020**, *125*, 166001.

https://doi.org/10.1021/acsphyschemau.1c00058
ACS Phys. Chem Au 2022, 2, 260−264