

Bridging semiempirical and *ab initio* QM/MM potentials by Gaussian process regression and its sparse variants for free energy simulation

Cite as: J. Chem. Phys. 159, 054107 (2023); doi: 10.1063/5.0156327

Submitted: 28 April 2023 • Accepted: 10 July 2023 •

Published Online: 2 August 2023



View Online



Export Citation



CrossMark

Ryan Snyder,¹  Bryant Kim,¹ Xiaoliang Pan,²  Yihan Shao,^{2,a)}  and Jingzhi Pu^{1,a)} 

AFFILIATIONS

¹Department of Chemistry and Chemical Biology, Indiana University-Purdue University Indianapolis, 402 N Blackford St., Indianapolis, Indiana 46202, USA

²Department of Chemistry and Biochemistry, University of Oklahoma, 101 Stephenson Pkwy, Norman, Oklahoma 73019, USA

^{a)}Authors to whom correspondence should be addressed: yihan.shao@ou.edu and jpu@iupui.edu

ABSTRACT

Free energy simulations that employ combined quantum mechanical and molecular mechanical (QM/MM) potentials at *ab initio* QM (AI) levels are computationally highly demanding. Here, we present a machine-learning-facilitated approach for obtaining AI/MM-quality free energy profiles at the cost of efficient semiempirical QM/MM (SE/MM) methods. Specifically, we use Gaussian process regression (GPR) to learn the potential energy corrections needed for an SE/MM level to match an AI/MM target along the minimum free energy path (MFEP). Force modification using gradients of the GPR potential allows us to improve configurational sampling and update the MFEP. To adaptively train our model, we further employ the sparse variational GP (SVGP) and streaming sparse GPR (SSGPR) methods, which efficiently incorporate previous sample information without significantly increasing the training data size. We applied the QM-(SS)GPR/MM method to the solution-phase S_N2 Menshutkin reaction, $\text{NH}_3 + \text{CH}_3\text{Cl} \rightarrow \text{CH}_3\text{NH}_3^+ + \text{Cl}^-$, using AM1/MM and B3LYP/6-31+G(d,p)/MM as the base and target levels, respectively. For 4000 configurations sampled along the MFEP, the iteratively optimized AM1-SSGPR-4/MM model reduces the energy error in AM1/MM from 18.2 to 4.4 kcal/mol. Although not explicitly fitting forces, our method also reduces the key internal force errors from 25.5 to 11.1 kcal/mol/Å and from 30.2 to 10.3 kcal/mol/Å for the N–C and C–Cl bonds, respectively. Compared to the uncorrected simulations, the AM1-SSGPR-4/MM method lowers the predicted free energy barrier from 28.7 to 11.7 kcal/mol and decreases the reaction free energy from -12.4 to -41.9 kcal/mol, bringing these results into closer agreement with their AI/MM and experimental benchmarks.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0156327>

I. INTRODUCTION

Accurate free energy profile simulation, which is critical for understanding many chemical and biochemical processes,¹ requires proper descriptions of the potential energy surfaces (PESs) of the system. Classical (nonreactive) force fields cannot be directly used to model chemical reactions because they lack proper PES descriptions of bond dissociation/formation processes, for which quantum mechanical methods are needed to characterize the associated changes in electronic structures. In condensed phases, these processes are often simulated by combined quantum mechanical and molecular mechanical (QM/MM) methods,^{2–7} with which a small-sized, electronically reactive subsystem is modeled using quantum

mechanics, whereas the surrounding nonreactive environment is modeled using molecular mechanics. High-quality, *ab initio* (AI) PES descriptions of the QM region are generally accurate and reliable, but they are usually computationally too time-consuming to obtain. This puts a limitation on the use of AI/MM methods in molecular dynamics (MD)-based free energy simulations, for which significant amounts of phase space sampling are needed to obtain statistically robust free energy profile along the reaction coordinate. By contrast, semiempirical (SE) methods, including molecular orbital (MO) methods based on the neglect of diatomic differential overlap (NDDO) approximation, such as MNDO,⁸ AM1,⁹ and PM3,¹⁰ and approximate density functional theory (DFT) methods such as the self-consistent-charge density-functional tight-binding

(SCC-DFTB) method,^{11–13} offer efficient PES alternatives to render sufficient sampling affordable, but they usually sacrifice accuracy for speed.

How to combine the accuracy of AI/MM methods and the efficiency of SE/MM methods has therefore posed a significant challenge (and opportunity) in QM/MM free energy simulations. One way to accomplish this is through indirect free energy simulations. Specifically, one carries out free energy perturbation (FEP), in which a correction term is applied to the SE/MM-generated free energy profile using Zwanzig's exponential average formula based on the single-point energy differences calculated at the respective levels of theory.¹⁴ In practice, the accuracy of this treatment, however, strongly depends on the degree of phase-space overlap between the SE/MM and AI/MM Hamiltonians involved. A poor overlap of the two Hamiltonians would hinder the numerical convergence of FEP calculations and limit the accuracy of computed free energies.^{15,16}

An alternative dual-level QM/MM free energy simulation strategy is the reaction path force matching (RP-FM) method.^{17–19} Through using force as the central bridging quantity, RP-FM and related methods forge direct links between the SE/MM and AI/MM levels by connecting their dynamics, free energies, and phase-space distributions.^{17–20} RP-FM has previously been used to improve QM/MM free energy results by either fitting Cartesian forces through optimization of SE parameters¹⁷ or by fitting reactive internal forces in collective variables (CVs).¹⁸ Combined with the weighted thermodynamics perturbation (wTP) method, RP-FM-based SE potential recalibration has proved to be effective in improving the phase-space overlap for dual-level QM/MM FEP calculations.¹⁹ A common theme in these FM developments is that they all fit high-level potentials (although implicitly) through some physics-based quantities such as electronic-structure parameters and free energy mean force, whose connections to the potential energy function can be computationally inconvenient to handle. As the complexity of the system increases, a general-purpose mathematics-function-based potential fitting tool may become more desirable.

To this end, several groups have employed delta (Δ) machine learning (ML) potentials, primarily based on neural networks (NN), as a means to directly correct SE/MM potential energy for improved QM/MM simulations.^{21–25} Despite the overall similarity in the Δ -learning theme following the work of Ramakrishnan *et al.*,²⁶ these works differ in their choice of base levels, NN potentials' descriptors and topological features, and loss function construction. Building on their earlier work using NN potentials for free energy perturbation,²⁷ Yang and co-workers have developed a method for training artificial neural networks (ANN) to correct DFTB/MM to the DFT/MM level using energy-only-based loss functions during MD simulations.²¹ Although forces are not directly used in training, the nuclear gradients associated with the NN-predicted energy corrections are incorporated to obtain modified molecular dynamics.²¹ Since force matching (FM) can greatly improve the phase-space overlap involved and therefore the quality of dynamics,^{17,19} it is generally desirable to include forces directly in the loss function when training the NN potentials.^{22–25} For example, Bösel *et al.*²² used symmetry-function descriptors as input features to train their high-dimensional neural network potentials (HDNNP), based on both energy and forces, for predicting corrections for DFTB/MM

in simulating stable and transition-state species in solution. Pan *et al.* used FM-recalibrated SE method to ensure that the training configurations are sampled in the relevant phase space.²³ Based on the DeepPot-SE²⁸ and the DeepMD framework,^{29,30} their ML model uses an embedding network to encode input features and incorporates both energy and force differences into the loss function to ensure accurate dynamics for MD-based free energy simulations of solution-phase and enzyme reactions.²³ Other recent QM/MM developments utilizing DeepPot and local environment descriptors include the range-corrected deep learning scheme of Zeng *et al.*²⁴ and the DFTB/MM-based ML model of Gomez-Flores *et al.*²⁵ One notable issue with conventional NN potentials is that they generally do not provide a metric to assess the uncertainties in energy and force predictions. For MD-based QM/MM simulations, in particular, free energy path simulations, this poses a question on how to maintain the robustness of NN models when the trajectories sampled on the NN-predicted PES deviate from the original training configuration space. Although active-learning-based models can be used to adaptively expand the training set,²¹ the lack of uncertainty analysis capability makes the identification of underrepresented training regions a challenging task without actually encountering a failure of simulation. On the other hand, the novel use of committee-based multiple-NN strategy,^{31–34} Bayesian NN,^{35,36} or dropout NN³⁷ can help detect untrustable predictions on the fly during MD simulations, but it can increase the inference cost.

As an alternative ML approach to NN, Gaussian process regression (GPR) can produce well-calibrated prediction uncertainty through predictive variance.^{38,39} GPR models have been used in computational chemistry for various applications, including generating force fields,^{40–42} enabling fast and accurate geometry optimization,^{43–47} improving weighted free energy perturbation results,⁴⁸ obtaining AI-quality potential energy landscapes,^{49–51} and making improvements to dynamics.⁴⁹ GPR-produced PESs have been demonstrated to outperform those constructed using ANNs.⁵² Additionally, GPR models outperform NN variants in uncertain estimates for predicting adsorption energies.⁵³ However, there are few cases of QM/MM implementation that enable GPR to be used directly for MD-based free energy simulations, which is the focus of the present study.

In this work, we match the SE/MM potential of the system with the target AI/MM potential using a GPR-predicted molecular energy correction term. While several studies utilize Δ -ML potentials as atom-based energy corrections,^{21,23,25,32} thereby preserving permutational invariance, the objective of this work is to train and deploy a system-specific model for condensed-phase QM/MM simulations along a well-defined free energy path. In this scenario, molecular configurations involving permutations of identical atoms are of lesser concern; therefore, permutational invariance is not strictly enforced here. We have found that the direct prediction of molecular energy corrections is sufficient for the purpose of this work. The training of the GPR models is done based on ensembles of dynamical configurations sampled initially along the SE/MM free energy path and later along the GPR-corrected ones. Specifically, we use the string method in collective variables^{54,55} (see Ref. 55 for the details of our string implementation) to determine the minimum free energy path (MFEP) through iterative path updates. Incorporation of the GPR-associated nuclear gradients on the geometric input features

enables us to perform MD simulations for updating free energy sampling and obtaining the GPR-modified MFEP. Since GPR by nature has a predictive uncertainty, the relationship of the sample input space to the training set can be assessed, which is a significant benefit for dynamics simulations as the configurations sampled along the updated MFEP may differ significantly from the ones used for training the original model. To accommodate the possible shift in MFEP during this learning–relearning process, we employ two sparse GPR models to efficiently incorporate the previous training data when updating our model iteratively for predictions in the new input space.

To demonstrate its effectiveness, we applied the QM-GPR/MM method to free energy simulations of the Menshutkin reaction in aqueous solution. The rest of the paper is organized as follows: In Sec. II, we briefly describe the GPR model and its sparse variants as well as the associated gradient formalism. The computational details of the simulations are provided in Sec. III. We present the results in Sec. IV. The outlook of the method is further discussed in Sec. V. Concluding remarks are given in Sec. VI.

II. METHODS

A. Gaussian process regression

GPR defines a distribution of functions relating a set of input space \mathbf{X} , composed of n input vectors \mathbf{x} , to a set of n observations \mathbf{y} . Any observation y in the set \mathbf{y} occurring at the input vector \mathbf{x} , composed of m input features $[x_1, \dots, x_m]$, in set \mathbf{X} can be linked to the input space through an underlying function, $f(\mathbf{x})$, where $y = f(\mathbf{x}) + \varepsilon$, and ε is an observational noise that separates the function value from the observation. The values of ε follow a Gaussian distribution centered at zero whose variance is determined by a noise parameter σ_n^2 , i.e., $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$. The prior distribution of the underlying functions, denoted $\mathbf{f}(\mathbf{X})$, follows a Gaussian distribution,

$$\mathbf{f}(\mathbf{X}) \sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{K}}(\mathbf{X}, \mathbf{X})), \quad (1)$$

where $\mathbf{0}$ is the mean of the functions and $\tilde{\mathbf{K}}$ is the covariance kernel matrix of the training dataset based on a given covariance kernel function k that defines the similarity between the two input vectors involved,³⁹

$$\tilde{\mathbf{K}}(\mathbf{X}, \mathbf{X}) = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & \dots & k(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix}. \quad (2)$$

In this study, a radial basis function (RBF) is used as the covariance kernel function k ,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left[-\frac{d(\mathbf{x}_i, \mathbf{x}_j)^2}{2l^2}\right], \quad (3)$$

where σ_f^2 is the vertical variation parameter, l is the length parameter, and $d(\mathbf{x}_i, \mathbf{x}_j)$ is the Euclidean distance between the two input vectors \mathbf{x}_i and \mathbf{x}_j . To account for noise in the observations, the noise parameter in the form of $\sigma_n^2 \mathbf{I}$ is added to the covariance kernel matrix of the training data, which modifies the covariance kernel matrix to

$$\mathbf{K} = \tilde{\mathbf{K}}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}, \quad (4)$$

where \mathbf{I} is the identity matrix.³⁹ Here, σ_f^2 , l , and σ_n^2 collectively form a set of hyperparameters, $\boldsymbol{\theta}$, that are optimized by maximizing the log marginal likelihood of the data,

$$\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}| - \frac{n}{2} \log 2\pi, \quad (5)$$

where $p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})$ is the probability of \mathbf{y} given \mathbf{X} and $\boldsymbol{\theta}$. The GP prior distribution can then be used to make a set of predictions, \mathbf{f}_* , given a set of new input vectors, \mathbf{X}_* , following the joint distribution,

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{f}_* \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{K} & \mathbf{K}_*^T \\ \mathbf{K}_* & \mathbf{K}_{**} \end{pmatrix}\right), \quad (6)$$

where $\mathbf{K}_{**} = \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*)$, $\mathbf{K}_* = \mathbf{K}(\mathbf{X}_*, \mathbf{X})$, and \mathbf{K}_*^T is the transpose of \mathbf{K}_* , and it is thus $\mathbf{K}(\mathbf{X}, \mathbf{X}_*)$. The posterior predictive distribution is then given by

$$p(\mathbf{f}_*|\mathbf{X}_*, \mathbf{X}, \mathbf{y}) \sim \mathcal{N}(\boldsymbol{\mu}_*, \mathbf{V}_*). \quad (7)$$

The mean, $\boldsymbol{\mu}_*$, and variance, \mathbf{V}_* , of the posterior predictive distribution of \mathbf{X}_* are found as follows:

$$\boldsymbol{\mu}_* = \mathbf{K}_* \mathbf{K}^{-1} \mathbf{y}, \quad (8)$$

$$\mathbf{V}_* = \mathbf{K}_{**} - \mathbf{K}_* \mathbf{K}^{-1} \mathbf{K}_*^T. \quad (9)$$

B. Model implementation

Let $U_{\text{AI/MM}}$ and $U_{\text{SE/MM}}$ denote the potential energies for a given configuration determined at the AI/MM and SE/MM levels, respectively; in practice, they are shifted with respect to the corresponding zero of energy at each level for numerical convenience. Then, the energy gap ΔU between the two levels can be written as

$$\Delta U = U_{\text{AI/MM}} - U_{\text{SE/MM}}. \quad (10)$$

In our GPR model, a collection of ΔU obtained for n configurations forms the training set of observations \mathbf{y} ,

$$\mathbf{y} = [\Delta U_1, \dots, \Delta U_n]^T = [y_1, \dots, y_n]^T. \quad (11)$$

A similar collection of m -dimensional input feature vectors $\mathbf{x} = [x_1, \dots, x_m]$ for each of the n configurations forms the GPR input space in the matrix form $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$. In the present study, we focus on the input features that are specifically based on the interatomic distances within the solute. To convert the interatomic distances to a generic form for GPR input, a preprocessing normalization step is first applied,

$$x_{j,k} = \frac{r_{j,k} - \bar{r}_j}{s_j}, \quad (12)$$

where $x_{j,k}$ represents the standardized unitless input feature j for configuration k , and $r_{j,k}$ is the j th distance for configuration k ; \bar{r}_j and s_j are the average and standard deviation, respectively, of the j th distances for all n configurations in the training set.

C. GPR gradient

During MD simulations, our GPR model is used to predict ΔU based on the configuration of the solute in the form of \mathbf{x}_* . The mean of the predictive distribution, μ_* , is added to $U_{SE/MM}$. To be consistent with the updated energy during dynamics, the force applied to each atom, expressed as the negative gradient of the potential energy with respect to the atomic coordinates, must be updated accordingly. For each input feature of GPR, the associated force modification on atom c in the Cartesian direction q is given by

$$F_{c,q} = -\frac{\partial \mu_*}{\partial x_{*,j}} \frac{dx_{*,j}}{dr_j} \frac{\partial r_j}{\partial q_c}, \quad (13)$$

where $x_{*,j}$ is the j th input feature of \mathbf{x}_* , r_j is the j th interatomic distance, and q_c is the Cartesian variable x , y , or z onto which the force correction to atom c is applied. The first partial derivative term in Eq. (13), $\frac{\partial \mu_*}{\partial x_{*,j}}$, is obtained by differentiating Eq. (8),

$$\frac{\partial \mu_*}{\partial x_{*,j}} = \frac{\partial \mathbf{K}_*}{\partial x_{*,j}} \mathbf{K}_*^{-1} \mathbf{y}, \quad (14)$$

where

$$\frac{\partial \mathbf{K}_*}{\partial x_{*,j}} = \frac{1}{\rho^2} \tilde{\mathbf{x}}_{*,j} \mathbf{I} \mathbf{K}_*, \quad (15)$$

Where $\tilde{\mathbf{x}}_{*,j} = [x_{l,j} - x_{*,j}, \dots, x_{n,j} - x_{*,j}]$ and \mathbf{I} is a $n \times n$ unit matrix. The second term in Eq. (13) can be obtained by differentiating Eq. (12),

$$\frac{dx_{*,j}}{dr_j} = \frac{1}{s_j}. \quad (16)$$

The third partial derivative term in Eq. (13) concerns the dependence of interatomic distances on the atomic Cartesian coordinates involved, which can be obtained in a straightforward manner,

$$\frac{\partial r_j}{\partial q_c} = \frac{q_c - q_d}{r_j}, \quad (17)$$

where q_c and q_d are the atomic Cartesian involved in the interatomic distance r_j .

D. Sparse GPR variants for adaptive MFEP optimization

With the GPR gradient available, the QM-GPR/MM model can be used in MD-based free energy path simulations to update the MFEP. When the configurations sampled along the GPR-modified MFEP significantly differ from the ones used to train the original model, the GPR model itself needs to be updated adaptively to incorporate information learned from the new samples. Although this can be done intuitively by merging the old and new samples into an expanded training set and then retraining the GPR model, the scalability of the GPR algorithm with respect to data size makes such a brute-force treatment quickly out of reach.

To achieve a data-efficient relearning process for adaptive MFEP optimization, we employ two sparse variants of the GPR model. First, we reduce the training data size for learning along a single MFEP by introducing a set of inducing points using the sparse

variational GP (SVGP) model (Appendix A). Second, for retraining the model using sample information from multiple MFEPs, we use the streaming sparse GPR (SSGPR) model (Appendix B), a variant based on SVGP, to enable adaptive model update in the combined input space without significantly increasing the training data size. For the interested readers, the technical details of these sparse GPR variants as well as the necessary modifications for implementing their nuclear gradients can be found in Appendixes A and B.

III. COMPUTATIONAL DETAILS

To test the performance of the QM-GPR/MM method, we applied it to free energy simulations of the solution-phase Menshutkin reaction ($\text{NH}_3 + \text{CH}_3\text{Cl} \rightarrow \text{CH}_3\text{NH}_3^+ + \text{Cl}^-$),⁵⁶ which involves a nucleophilic attack of ammonia on methyl chloride following an associative $\text{S}_{\text{N}}2$ mechanism (Fig. 1). This reaction has served as an important paradigm for developing accurate and efficient potential energy methods and solvation models (see our recent work¹⁸ and references therein for a brief review of the QM/MM methods developed for simulating this system).

At the SE/MM level, we treat the solute molecules by AM1⁹ using the MNDO97 package⁵⁷ incorporated in Chemistry at Harvard Macromolecular Mechanics (CHARMM),⁵⁸ whereas the solvent environment is modeled by a $40 \times 40 \times 40 \text{ \AA}^3$ box of modified TIP3P⁵⁹ water molecules under periodic boundary conditions. The nonbonded QM/MM van der Waals (vdW) parameters were initially assigned based on the related atom types in the CHARMM22 force field⁶⁰ during setup and later replaced by the pair-specific vdW parameters optimized by Gao and Xia⁶¹ in the actual simulations. Long-range electrostatics is treated by the particle mesh Ewald (PME)⁶² and QM/MM-Ewald⁶³ methods for the MM-MM and QM-MM interactions, respectively. The MFEPs and associated free energy profiles were determined by using the string method in CVs.⁵⁴ For the Menshutkin reaction, we represent the MFEP by two CVs: the forming bond distance, $r_{\text{N-C}}$, and the breaking bond distance, $r_{\text{C-Cl}}$. The initial string path and configurations along it were obtained from a potential energy scan along the one-dimensional reaction coordinate $r_{\text{C-Cl}} - r_{\text{N-C}}$ at the AM1/MM level. The MFEPs were optimized iteratively by the string method for 10 cycles, in each of which 20 ps MD sampling was used for evaluating the free energy mean forces on the CVs in each of 20 images evenly spaced along the string. For additional details of the MD simulations and string optimization protocols, see our previously published work.^{18,55}

Based on the configurations sampled along the AM1/MM MFEP, AI/MM single-point energy calculations were performed at the B3LYP⁶⁴⁻⁶⁶/6-31+G(d,p)⁶⁷/MM level using the Gaussian16 program⁶⁸ interfaced with CHARMM. To generate our testing set, we collected configurations every 100 fs from 20 ps of MD trajectories for each of the 20 images along the MFEP, which gives a training set of 4000 configurations. For GPR training, we draw one in every five of these configurations (i.e., 1 configuration every 500 fs) to build a training set of 800 configurations. For the same set of configurations, AM1/MM single-point energies were retrieved from the MNDO97/CHARMM interface. These single-point energies were then used to determine the energy gaps between the two levels involved, i.e., $\Delta U = U_{\text{B3LYP/6-31+G(d,p)/MM}} - U_{\text{AM1/MM}}$, as defined in Eq. (10) and used in Eq. (11), for which the training sample size $n = 800$.

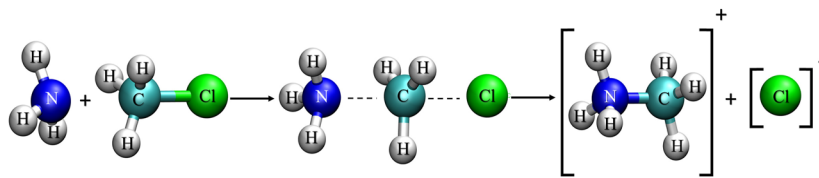


FIG. 1. Menshutkin reaction between ammonia and methyl chloride.

The GPR model was generated and optimized using the Scikit-learn package.⁶⁹ The SVGP model was generated using the SGPR module in the GPflow 2.0 package.⁷⁰ The SSGPR models were constructed using the code from the work of Bui *et al.*,⁷¹ updated for compatibility with GPflow 2.042 and TensorFlow 2.0.^{72,73} Conveniently, Bui and colleagues recently updated their code for compatibility with the updated version of GPflow and TensorFlow, accessible to the reader at Ref. 74. The SVGP and SSGPR models were optimized with respect to their hyperparameters using the ADAM optimizer⁷⁵ implemented in TensorFlow 2.0.

The optimized GPR, SVGP, and SSGPR models were implemented in CHARMM through its USER module, which enables predictions of energy corrections along MD simulations. The mean of the predictive distribution of ΔU for a given MD configuration was added to the SE/MM energy to acquire the total energy for SE-GPR/MM as well as for its sparse variants. The nuclear gradients were modified by distributing the associated force corrections according to Eq. (13) or its sparse versions based on Eqs. (A5) and (B4).

IV. RESULTS

A. AM1/MM free energy profile

Using the AM1/MM Hamiltonian, we obtained a free energy profile that gives a reaction free energy of -12.4 kcal/mol and a free energy barrier of 28.7 kcal/mol for the Menshutkin reaction (Fig. 2). Compared with the experimentally derived reaction free energy of -34 ± 10 kcal/mol reported by Gao⁷⁶ or -36 ± 6 kcal/mol reported in the work of Su *et al.*,⁷⁷ AM1/MM underestimates the exergonicity of the Menshutkin reaction by 21.6 – 23.6 kcal/mol. Although no experimental barrier is available for this reaction, the reaction between $\text{NH}_3 + \text{CH}_3\text{I}$, which gives a free energy barrier of 23.5 kcal/mol,⁷⁸ has been previously used as a point of comparison.^{76,79,80} Compared with this best experimental estimate, AM1/MM overestimates the free energy barrier by 5.2 kcal/mol.

B. Potential energy gap

Since the AM1/MM free energy profile displays significant discrepancies from experiment, we performed free energy perturbation (FEP) calculations along the AM1/MM free energy path to see if indirect DFT/MM calculations can help improve the results. Unfortunately, FEP on its own, based on the potential energy differences between the AM1/MM and B3LYP/6-31+G(d,p)/MM levels, does not produce a high-quality free energy profile; instead, it generates a free energy barrier of 12.2 kcal/mol and a reaction free energy of -50.7 kcal/mol, both too low compared with experiment (Fig. 2). The modest performance of a simple FEP correction

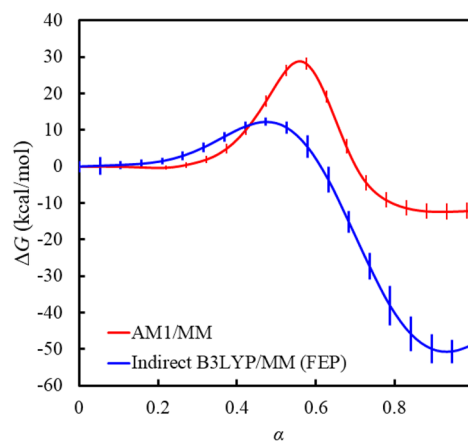


FIG. 2. AM1/MM and indirect B3LYP/MM free energy profiles for the Menshutkin reaction. The AM1/MM free energy profile was obtained along the string MFEP determined at the same level (with $\alpha = 0$ for reactant and 1 for product). The indirect free energy profile was obtained by free energy perturbation (FEP) from the AM1/MM to the B3LYP/6-31+G(d,p)/MM level based on the configurations sampled along the AM1/MM MFEP. Even with AM1/MM-based free energy corrections, the indirect FEP results still do not adequately reproduce the experimental results. Details on the calculations of the error bars can be found in the supplementary material, Sec. S4.

in this case is likely caused by a slow convergence of the results when using inadequate sample size^{81,82} as well as possible systematic biases linked to poor overlap between the important phase space sampled at the base and target levels.^{83–85} At the limited sampling performed at AM1/MM, the average standard deviation of the relative energy difference distribution is 3.7 kcal/mol, well exceeding the recommended limits of 1.7 kcal/mol⁸² for fine results or 2.4 kcal/mol⁸⁶ for crude results. Furthermore, the failure for our samples to meet the bias metric criterion of $II > 0.5$ introduced by Wu and Kofke^{84,85,87} suggests significant differences in the configurations sampled on the AM1/MM and B3LYP/6-31+G(d,p)/MM PESs; see the supplementary material, Sec. S1. Although indirect FEP calculations can be accelerated when combined with ML approaches,^{27,88} the issues we identified here collectively highlight the need for a free energy method that directly samples the AI/MM-quality PES.

The energy corrections predicted by our GPR model are compared with their reference values in Fig. 3. The root-mean-square errors (RMSEs) of the GPR-predicted energy gaps are 3.1 and 3.7 kcal/mol for the training and testing sets, respectively (Table I). Considering the magnitude of the energy gap as well as its nonuniform

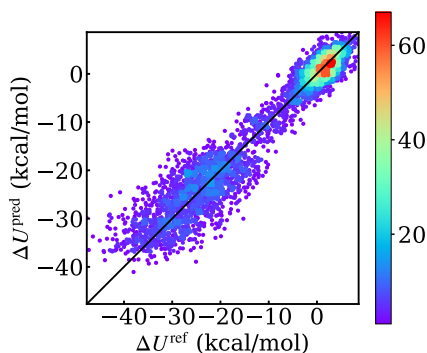


FIG. 3. The GPR-predicted energy corrections (ΔU^{pred}) plotted against the reference energy differences (ΔU^{ref}) between the AM1/MM and B3LYP/6-31+G(d,p)/MM levels for configurations sampled along the AM/MM MFEP. The color scale shows the number of configurations found in each region.

distribution along the reaction coordinate, we found this performance acceptable and comparable to that of other machine learning models; see the supplementary material, Sec. S2. As we mentioned earlier in the indirect FEP results, the reference energy gaps along the AM1/MM MFEP display a significant average standard deviation of 3.7 kcal/mol, which suggests that the data are inherently noisy and therefore susceptible to prediction error. The mean of predictive variance, or equivalently the lower bound of the expected mean squared error,^{38,89,90} is 2.3 and 1.9 (kcal/mol)² for the training and testing set, respectively. As explained in the supplementary material, Sec. S3, the predictive variance is used to calculate the 95% confidence interval as a measure of the uncertainty of prediction. The average 95% confidence intervals for energy predictions on the testing and training sets are 2.9 and 2.6 kcal/mol, respectively.

Sampling the GPR-corrected potential (AM1-GPR/MM), the free energy path is updated. Because the new path evolves away from the AM1/MM path used for previous training, the GPR model becomes less accurate on the newly sampled configurations, with the RMSE in energy prediction increasing to 11.6 kcal/mol. The average predictive variance along the AM1-GPR/MM MFEP, which offers an assessment of prediction uncertainty in GPR, also increases to 2.3 (kcal/mol)². Both of these results suggest changes in the sampled phase space and the necessity of retraining the model. Interestingly, the model seems to be more tolerant to such changes in terms of predictive variance than predictive error. Further analysis shows that the newly sampled configurations may not differ too much from those used for training the original model in the input feature space to perturb the prediction distribution (see Sec. S5 in the supplementary material), but the greater baseline errors in AM1/MM for the new samples can quickly worsen the overall RMSE of the GPR-corrected method in energy prediction. Nevertheless, we decided to retrain the model iteratively when updating MFEPs to prevent possible model deterioration.

Due to the matrix inversions involved, GPR does not scale well with training data size. Therefore, a simple retraining strategy by adding new data to the existing dataset could quickly become unmanageable. However, omission of previous samples in the retraining processes could result in model oscillations and divergent free energy results; see the supplementary material, Sec. S6.

To address this problem, we first use the sparse variational GP (SVGP) method (Appendix A) to summarize the training set into a set of inducing points using an approximate kernel, which also reduces the memory requirement. The streaming sparse (SSGPR) method (Appendix B) is then used to further allow us to train a “synthesized” model on multiple sets of data from different MFEPs without significantly increasing the kernel’s dimensionality. By substituting the original GPR model with an SVGP model containing 50 inducing points, the RMSEs in energy prediction on the training and testing sets along AM1/MM MFEP increase slightly to 3.2 and 3.8 kcal/mol, respectively (Table I). Using the SVGP model (also referred to as SSGPR-0) as a starting point, the SSGPR- i ($i = 1-4$) models were trained iteratively until the MFEPs determined in successive iterations converge. Although unnecessary, in practice it can be advantageous to use more inducing points as the sampled input space evolves. For example, we found that increasing the number of inducing points to 200 yields improvements to the RMSE of the entire set of configurations (see, Sec. S7 in the supplementary material). To balance efficiency and accuracy, we added 50 more inducing points every iteration, which results in a use of 250 inducing points in the SSGPR-4 model. This procedure allows us to make predictions and evaluate uncertainties with greater confidence across all the relevant phase space sampled when the MFEP is gradually improved from the AM1/MM toward the DFT/MM level. The RMSE of the SSGPR-4 model in energy predictions for the original AM1/MM testing set is 4.8 kcal/mol. For the configurations sampled along the AM1-SSGPR-3/MM MFEP, based on which the SSGPR-4 model is trained, the RMSE of the model in energy prediction is reduced to 3.4 and 3.8 kcal/mol for the training and testing sets, respectively.

C. AM1-(SS)GPR/MM methods improve free energy profiles and MFEPs

For the Menshutkin reaction, the free energy profile simulated with the AM1-GPR/MM method produces a free energy barrier of 14.3 kcal/mol and a reaction free energy of -44.9 kcal/mol (Fig. 4). The AM1-SVGP/MM method (i.e., AM1-SSGPR-0/MM) yields comparable results with a free energy barrier of 17.1 kcal/mol and a reaction free energy of -41.8 kcal/mol. The free energy profiles obtained by the AM1-SSGPR- i /MM methods converge to a reaction free energy of -41.9 kcal/mol and a free energy barrier of 11.7 kcal/mol (Fig. 5). Compared to a significant overestimate of -12.4 kcal/mol at the AM1/MM level, the reaction free energies predicted by these GPR-corrected methods are greatly improved to close agreements with the experimental values of -34 ± 10^{26} and -36 ± 6 kcal/mol.⁷⁷ Although the simulated free energy barriers are still significantly lower than the best experimental estimate of 23.5 kcal/mol,⁷⁸ the goal of our GPR-corrected methods is to reproduce the target-level AI/MM results. The B3LYP functional used here is known to underestimate the free energy barrier for the Menshutkin reaction, based on previous implicit solvation⁷⁹ and multilevel QM/MM calculations.^{18,19} From this perspective, the free energy barrier of 14.3 kcal/mol from our AM1-GPR/MM simulations agrees well with the DFT/MM benchmark value of 15.3 ± 0.1 kcal/mol established in the work of Pan *et al.*¹⁹ from umbrella sampling simulations at the B3LYP/6-31G(d)/MM level and a value of 14.7 kcal/mol obtained in the work of Kim *et al.*¹⁸ through

TABLE I. Cross validations of the AM1-(SS)GPR/MM models in energy prediction.^a

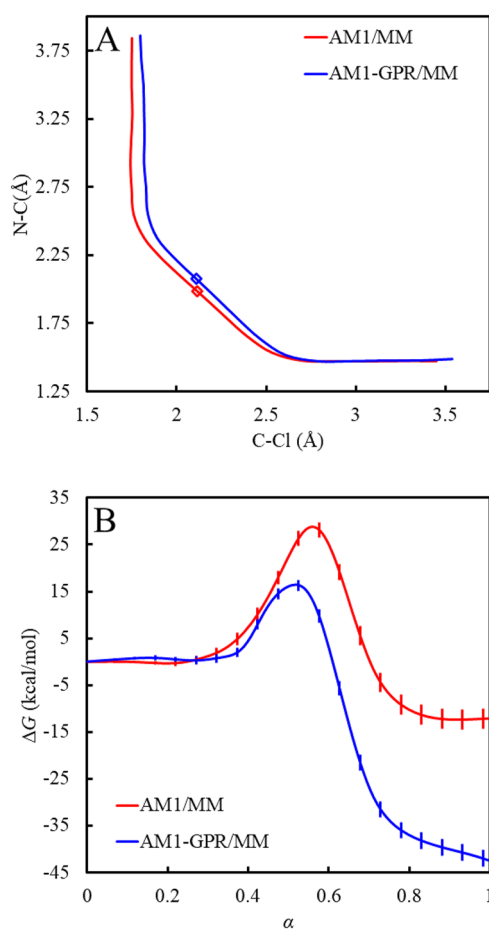
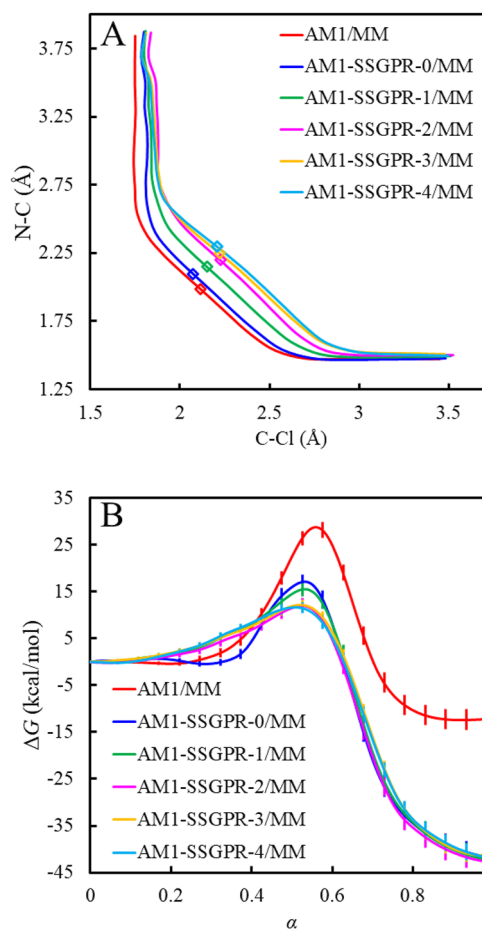
Samples	AM1/MM	GPR	SSGPR-0 ^b	SSGPR-1	SSGPR-2	SSGPR-3	SSGPR-4
AM1/MM	18.2	3.7 ± 2.9 (3.1 ± 2.6)	3.8 ± 3.4 (3.2 ± 3.3)	3.9 ± 4.2	4.1 ± 4.7	4.4 ± 5.2	4.8 ± 5.8
GPR	18.2	11.6 ± 4.2	11.7 ± 3.4	12.1 ± 4.2	12.5 ± 4.7	12.8 ± 5.2	13.0 ± 5.8
SSGPR-0	21.8	7.6 ± 3.7	7.1 ± 4.8	3.9 ± 3.6 (3.5 ± 3.5)	3.9 ± 3.9	4.0 ± 4.3	4.3 ± 4.9
SSGPR-1	22.9	4.9 ± 3.6	4.9 ± 4.3	4.9 ± 4.4	4.1 ± 3.7 (3.7 ± 3.6)	4.0 ± 3.9	4.1 ± 4.4
SSGPR-2	23.9	5.2 ± 3.6	5.3 ± 4.1	4.4 ± 4.0	4.5 ± 3.9	3.7 ± 3.7 (3.1 ± 3.6)	3.8 ± 4.1
SSGPR-3	23.5	5.6 ± 3.5	5.8 ± 4.0	4.6 ± 4.1	4.3 ± 4.1	4.6 ± 4.1	3.8 ± 4.1 (3.4 ± 4.0)
SSGPR-4	24.1	5.2 ± 2.9	5.4 ± 3.9	4.3 ± 4.0	4.0 ± 3.9	4.0 ± 4.0	4.4 ± 4.3

^aRoot-mean-square errors (RMSEs) in energy prediction (in kcal/mol) are compared against the B3LYP/6-31+G(d,p)/MM benchmark; results for both the testing set and the training set (data in parenthesis) are given when applicable; values following ± correspond to the average 95% confidence interval of energy predictions.

^bThe SVGP model is labeled as SSGPR-0.

internal-force-matched RP-FM-CV simulations at the B3LYP/6-31+G(d,p):AM1/MM level. On the other hand, the free energy barrier of 11.7 kcal/mol converged from the iteratively optimized AM1-SSGPR-*i*/MM methods may represent a slight underestimate

of the target value. Free energy perturbation was further performed on the AM1-SSGPR-4/MM free energy profile and no statistically significant changes were observed; see the supplementary material, Sec. S8.

**FIG. 4.** (a) The AM1/MM and AM1-GPR/MM MFEPs. (b) The AM1/MM and AM1-GPR/MM free energy profiles.**FIG. 5.** (a) The AM1/MM and AM1-SSGPR/MM MFEPs. (b) The AM1/MM and AM1-SSGPR/MM free energy profiles.

The GPR-corrected methods also improve the free energy paths consistently by moving the original AM1/MM MFEP toward the concave side (Figs. 4 and 5). The free energy transition states determined at the AM1-GPR/MM and AM1-SSGPR/MM levels are looser than that obtained at the AM1/MM level. These results are in agreement with our observations in a previous RP-FM-CV study when the AI/MM-level internal forces were directly targeted.¹⁸

D. AM1-(SS)GPR/MM methods improve forces

The free energy forces used for string path optimization and the calculations of the free energy profile are estimated based on the fluctuations of the collective variables (CVs) under a set of harmonic restraints.⁵⁴ For a restrained system whose statistics follows the Boltzmann–Gibbs distribution, such fluctuations are ultimately governed by the system's potential energy.⁵⁴ This indicates that correcting the potential energy of the system (e.g., with a GPR model) to the reference level of theory would improve both the string MFEP and the free energy profile accordingly. Using a different strategy, we recently showed that the high-level free energy profile can be restored through force matching (FM) on the CV degrees of freedom along the reaction path (RP-FM-CV),¹⁸ which is equivalent to fitting free energy mean force on the reaction coordinate. As both the SE-GPR/MM and RP-FM-CV approaches are able to reproduce the AI/MM-quality free energy profile, the connection between them needs to be established. Although SE-GPR/MM also modifies forces through nuclear gradients of the GPR energy correction term, it is unclear how these force modifications are distributed to various internal degrees of freedom that eventually lead to an improved MFEP. To examine whether forces are also improved along the reaction coordinate, we monitored the internal forces on the CVs before and after applying the GPR corrections, employing the redundant internal coordinate transformation framework we recently developed for RP-FM-CV.¹⁸

In Fig. 6, we compare the CV internal force errors of using AM1/MM and AM1-SSGPR-4/MM along the converged AM1-SSGPR-4/MM MFEP, with respect to the B3LYP/6-31+G(d,p)/MM benchmark. Unsurprisingly, the AM1-SSGPR-4/MM method based on the converged SSGPR-4 model offers an improved force description of the C–Cl bond nearly across the entire reaction coordinate. The RMSE in C–Cl force decreases from 30.2 to 10.3 kcal/mol/Å after the SSGPR-4 corrections are applied to AM1/MM. Similar trends are found for the N–C bond, where the GPR corrections result in a near perfect description of its force in the reactant and transition state regions, and they significantly improve the AM1/MM description of the bond in the product region. The overall description of the N–C bond improves with an RMSE in force reduced from 25.5 to 11.1 kcal/mol/Å. The RMSE of Cartesian forces among all solute atoms drops from 14.6 to 12.1 kcal/mol/Å; for the atoms involved in the CVs, the Cartesian force error drops from 19.4 to 14.6 kcal/mol/Å; see the supplementary material, Sec. S9. These force results offer a possible explanation for the improved free energy profile. The greater force correction found on the C–Cl bond suggests its dominant role in improving the free energy profile. With the C–Cl force description improved, the AM1-SSGPR-4/MM free energy profile shows enhanced agreement with the AI/MM result. We performed a similar comparison for the AM1-GPR/MM method (see Sec. S10 in the supplementary material) but found that along

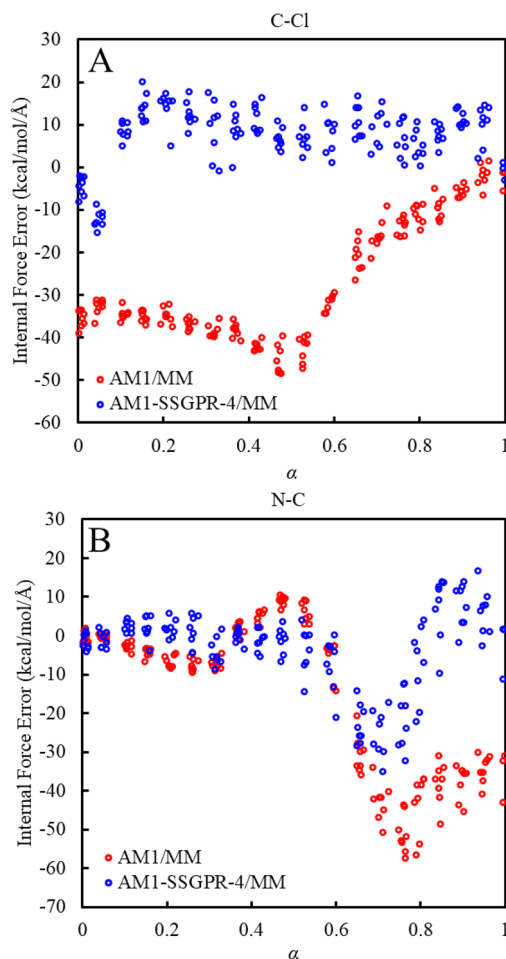


FIG. 6. Internal force differences between the B3LYP/6-31+G(d,p)/MM and AM1-SSGPR-4/MM levels; the force differences on the C–Cl bond (a) and N–C bond (b) were evaluated based on configurations sampled along the AM1-SSGPR-4/MM MFEP.

the AM1-GPR/MM MFEP, the quality of both the internal and Cartesian forces deteriorates significantly, which highlights the need for an adaptive model. Regardless of the sampled path and model, the underlying shape of the force differences remains the same. The persistence of these trends in the force differences between the SE/MM and SE-GPR/MM models suggests certain limitations in the energy-based GPR model in terms of distributing the desired force corrections, perhaps due to the selected input features.

E. Computational cost of AM1-SSGPR/MM methods

As mentioned in the Introduction, one objective of this work is to obtain AI/MM-quality free energy profiles at the reduced cost comparable to SE/MM simulations. In Table II, we present the wall times for a single-point energy-force calculation, a single MD step, and 20 ps MD sampling (at a timestep of 1 fs), using AM1/MM as the baseline method with and without the SSGPR-predicted corrections. We compare these times with the estimated cost required

TABLE II. Wall time comparison of using the AM1/MM, AM1-SSGPR/MM, and direct DFT/MM methods for simulating the Menshutkin reaction.

Method	CPUs	Time per			Total wall time for MFEP ^a (hr)
		Energy and force evaluation (s)	Single MD step (s)	20 ps MD (hr)	
AM1/MM	1	0.243	0.367	2.038	408
AM1-SSGPR/MM	1	0.253	0.378	2.100	420
DFT/MM ^b	1	26.037	39.056	216.975	43 395
DFT/MM ^b	24	11.888	17.832	99.069	19 814

^aTotal wall time for one iteration of string MFEP optimization following the simulation procedures outlined in Sec. III.

^bAt the B3LYP/6-31+G(d,p)/MM level.

for completing a single iteration of string MFEP optimization using brute-force DFT/MM simulations. We found that, with only a modest increase in the computational cost relative to AM1/MM, the AM1-SSGPR/MM method can be used to efficiently perform free energy sampling on a timescale comparable to AM1/MM, resulting in a 100-fold speedup of the DFT/MM simulations at the B3LYP/6-31+G(d,p)/MM level. Even when utilizing 24 central processing units (CPUs), the parallel DFT/MM benchmark shown in Table II remains about 50 times slower than a serial AM1-SSGPR/MM calculation.

V. DISCUSSION

The accuracy of potential energy description plays a critical role in the reliable determination of free energy profile. Motivated by its connection to the free energy perturbation strategy, here we have developed an energy-based GPR method that dynamically predicts the energy corrections needed for an SE/MM Hamiltonian to match a desired AI/MM target. The availability of nuclear gradients associated with the GPR energy correction also allows us to update forces accordingly to achieve improved molecular dynamics and free energy profiles.

An alternative approach would be to treat both energy and force as the GPR observations during model training. It has been suggested that including gradient information in training would improve the performance of ML potentials.⁹¹ For ANN-based models, this can be conveniently done by including the force-fitting errors in the loss functions. For GPR models, the gradient of the energy (negative force) needs to be included consistently in an extended set of observations. We have reported our force-extended GPR approach in a separate work,⁹² where we use GPR with derivative observations (GPRwDO) to simultaneously predict both QM/MM energy and force corrections. Although the GPRwDO approach significantly improves the description of Cartesian forces compared to an energy-only model, it does increase the computational complexity of model training from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^3(d+1)^3)$ and the memory complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n^2(d+1)^2)$, when including d -dimensional derivative observations in each of the n training samples. This poses additional constraints on the training set size and the number of atoms to be considered for explicit force training. For the large number of nonreactive atoms in a QM/MM system, we may still need to rely on differentiating an energy-only model to obtain the related force corrections. Nevertheless, there

are other ways to consider to alleviate this problem. For example, Bartok *et al.*⁴⁹ incorporated derivative observation into their sparse GP models to improve algorithmic scaling. Alternatively, von Lilienfeld and co-workers developed the kernel-based operator quantum machine learning approach,^{93,94} which reduces both memory usage and computational complexity. Finally, Meyer and Hausser demonstrated a scheme where internal forces are fitted along with energy for geometry optimization.⁴⁴

Interestingly, our recent force matching work (RP-FM-CV)¹⁸ showed that similar improvements in free energy results can be achieved by directly correcting the internal forces on the CVs without explicit energy correction. Although potential energy dictates the density of state, the dynamics of the system and its evolution are nevertheless determined by the forces on the atoms. Moreover, in the string free energy method we employed,⁵⁴ potential energy is neither directly used in the MFEP optimization nor used in the determination of free energy profile. Under the RP-FM-CV framework, we attributed the success of our method to an effective fitting of the high-level free energy mean force, the integration of which along the reaction coordinate allows the associated potential of mean force to be faithfully reproduced.¹⁸

While direct learning in the force space is appealing, the use of energy-based GPR may offer a few advantages. As previously mentioned, one metric for assessing the degree of phase-space overlap is the distribution of relative energy differences between two levels of theory.^{82,86} This will permit easy assessment of phase-space overlap without the need for any AI/MM benchmark configurations. The energy-based GPR scheme also reduces the dimensionality of the output space, making prediction and evaluation of its quality less complex. The use of a single molecular energy correction, instead of atomic energies⁴⁹ or atomic forces, would also reduce the risk of placing disproportionate weights on correcting the fast degrees of freedom, which may cause trajectory instability issues if not handled properly.⁹⁵ Further investigation is needed to determine if a force-only-based GPR scheme would be sufficient and effective to achieve similar improvements for QM/MM free energy simulations.

In this proof-of-concept work, we use simple geometric features, such as pair-wise distances to describe the input space of the GPR models. Although the GPR-facilitated free energy profile of the Menshutkin reaction simulated in aqueous solution shows good agreement with the AI/MM benchmarks and experimental results, it could be more challenging to achieve similar performance for reactions in a more complex environment, such as an enzymatic reaction

treated by QM/MM. The heterogeneity of enzyme active site would likely require additional descriptions of charge changes on the QM atoms, especially when the accuracy in QM/MM interactions plays a large role in the free energy profile. Since charge-dependent GPR gradients may not be straightforward to implement, it may be advantageous to use other features to represent the evolving QM/MM interactions. Descriptors such as atom-centered symmetry functions (ACSF),⁹⁶ smooth overlap of atomic positions (SOAP),⁹⁷ or density encoded canonically aligned fingerprint (DECAF)⁴⁰ may be more applicable to modeling corrections in complex chemical environments. Notably, these atom-centered environment descriptors lend themselves to permutationally invariant models that predict atomic energy corrections. These advanced descriptors and an extension to the SSGPR model to account for permutational invariance will be examined in our future work.

VI. CONCLUDING REMARKS

In summary, we have presented a GPR-assisted dual-level QM/MM method for cost-effective free energy simulations of condensed-phase chemical reactions. Built upon an efficient SE/MM Hamiltonian, this method boosts its PES accuracy to an AI/MM quality by modeling the energy corrections needed through GPR-enabled stochastic inference. Incorporation of the forces associated with the GPR-predicted energy corrections allows for accurate determination and construction of MFEPs and free energy profiles from MD-based QM/MM simulations. Two data-efficient sparse GPR variants, i.e., SVGP and SSGPR, are further employed to facilitate the iterative updates of our model when the sample space evolves along with the MFEP that moves toward its AI/MM target. The great improvements seen in the free energy results for the solution-phase Menshutkin reaction demonstrate the effectiveness of this energy-based QM-(SS)GPR/MM method and suggest its general usefulness for studying reaction mechanisms in complex environments.

SUPPLEMENTARY MATERIAL

See the supplementary material for a description of the Wu and Kofke bias metric for the Menshutkin reaction, illustration of the reference and predicted energy gap along the reaction coordinate and a comparison of GPR prediction performance with existing ANN model in literature, comparison of predictive-variance-based uncertainty estimates, description of FEP calculations and error estimation of the string free energy profile, description of the discrepancy between rising error and stagnant variance in new phase space, demonstration of the model improvements with an increased number of inducing variables, FEP calculations on the AM1-SSGPR/MM converged MFEP, and Cartesian force improvements through energy corrections with the evolving models.

ACKNOWLEDGMENTS

This work was supported by a Research Support Funds Grant (RSFG) from IUPUI, and Grant Nos. R15-GM116057 (J.P.) and R01-GM135392 (Y.S. and J.P.) from the U.S. National Institutes of Health (NIH). The authors thank the Indiana University Perva-

sive Technology Institute (IU PTI) for providing supercomputing and visualization resources on BigRed3 and Carbonate DL/GPU that have contributed to the research results reported within this paper. This research was supported in part by Lilly Endowment, Inc., through its support for IU PTI.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Ryan Snyder: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Writing – original draft (equal); Writing – review & editing (equal). **Bryant Kim:** Software (equal); Writing – review & editing (equal). **Xiaoliang Pan:** Formal analysis (equal); Writing – review & editing (equal). **Yihan Shao:** Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal). **Jingzhi Pu:** Conceptualization (equal); Formal analysis (equal); Methodology (equal); Project administration (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: SPARSE VARIATIONAL GAUSSIAN PROCESSES

One deficiency of GPR is its limited ability to scale to large training datasets. Due to the matrix inversions involved, the computational cost of GPR grows as $\mathcal{O}(n^3)$ with respect to the sample size n . This can be particularly concerning for free energy path simulations as the training data drawn from the sampled configurations evolve during path update and optimization. To reduce the data requirements, we utilize sparse variational Gaussian processes (SVGP), which summarize the full data \mathbf{X} into a series of m inducing points (\mathbf{X}_M) that maintain the model using an approximate posterior. The inducing variables are trained along with the kernel's hyperparameters by maximizing the lower bound of the exact marginal likelihood.⁹⁸ This is accomplished by minimizing the Kullback–Leibler divergence between the exact and approximate posteriors. Using this method, an approximate posterior distribution is used in place of the exact posterior distribution, for which the mean and covariance functions are given by

$$\boldsymbol{\mu}_* = \mathbf{K}_{*M} \mathbf{K}_{MM}^{-1} \boldsymbol{\mu}_M, \quad (\text{A1})$$

$$\mathbf{V}_* = \mathbf{K}_{**} - \mathbf{K}_{*M} \mathbf{K}_{MM}^{-1} \mathbf{K}_{M*} + \mathbf{K}_{*M} \mathbf{B} \mathbf{K}_{M*}, \quad (\text{A2})$$

where $\mathbf{B} = \mathbf{K}_{MM}^{-1} \mathbf{A} \mathbf{K}_{MM}^{-1}$, and $\boldsymbol{\mu}_M$ and \mathbf{A} are the mean vector and covariance matrix of the chosen variational Gaussian distribution

of the inducing point function values, respectively, calculated as follows:⁹⁸

$$\boldsymbol{\mu}_M = \sigma_n^{-2} \mathbf{K}_{MM} \boldsymbol{\Sigma} \mathbf{K}_{MX} \mathbf{y}, \quad (\text{A3})$$

$$\mathbf{A} = \mathbf{K}_{MM} \boldsymbol{\Sigma} \mathbf{K}_{MM}, \quad (\text{A4})$$

where $\boldsymbol{\Sigma} = (\mathbf{K}_{MM} + \sigma_n^{-2} \mathbf{K}_{MX} \mathbf{K}_{XM})^{-1}$. Under the SVGP model, the gradient in Eq. (14) is updated to

$$\frac{\partial \mu_*}{\partial x_{*,j}} = \frac{\partial \mathbf{K}_{*M}}{\partial x_{*,j}} \mathbf{K}_{MM}^{-1} \boldsymbol{\mu}_M. \quad (\text{A5})$$

APPENDIX B: STREAMING SPARSE GPR

As the sampled phase space evolves, the model may no longer accurately predict the potential energy difference between the AI/MM and SE/MM levels. It is therefore necessary to iteratively update the model using data from the new configurations; however, we cannot simply retrain the model on the new data without losing the memory of the previous configurations. We therefore use a streaming sparse GPR (SSGPR) method outlined in the work of Bui *et al.*⁷¹ Here, we stream the new data as a batch and update the model's hyperparameters and inducing points by leveraging both the old posterior approximation and the new data. Let \mathbf{X}_b represent the new inducing points and \mathbf{X}_a the old inducing points, and then the predictive mean and covariance for the SSGPR model are given as

$$\boldsymbol{\mu}_* = \mathbf{K}_{*b} \mathbf{L}_b^{-T} \mathbf{D}^{-1} \mathbf{L}_b^{-1} \mathbf{K}_{bf}^{-1} \boldsymbol{\Sigma}_y^{-1} \hat{\mathbf{y}}, \quad (\text{B1})$$

$$\mathbf{V}_* = \mathbf{V}_{**} - \mathbf{K}_{*b} \mathbf{K}_{bb}^{-1} \mathbf{K}_{b*} + \mathbf{K}_{*b} \mathbf{L}_b^{-T} \mathbf{D}^{-1} \mathbf{L}_b^{-1} \mathbf{K}_{b*}, \quad (\text{B2})$$

where $\mathbf{D} = \mathbf{I} + \mathbf{L}_b^{-1} \mathbf{K}_{bf} \boldsymbol{\Sigma}_y^{-1} \mathbf{K}_{fb} \mathbf{L}_b^{-T}$, \mathbf{L}_b is a lower triangular matrix obtained from the Cholesky decomposition of \mathbf{K}_{bb} such that $\mathbf{L}_b \mathbf{L}_b^T = \mathbf{K}_{bb}$, $\mathbf{K}_{bf} = \begin{bmatrix} \mathbf{K}_{fb} \\ \mathbf{K}_{xb} \end{bmatrix}^T$, where \mathbf{f} is the latent function values at the newly collected training points \mathbf{X}_{new} , hereon referred to as \mathbf{X}_f for the use in subscripts, $\hat{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_{\text{new}} \\ \mathbf{D}_a \mathbf{S}_a^{-1} \boldsymbol{\mu}_a \end{bmatrix}$ and

$$\boldsymbol{\Sigma}_y = \begin{bmatrix} \sigma_n^{-2} \mathbf{I} + \alpha \text{diag}(\mathbf{K}_{ff} - \mathbf{K}_{fb} \mathbf{K}_{bb}^{-1} \mathbf{K}_{bf}) & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_a + \alpha(\mathbf{K}_{aa} - \mathbf{K}_{ab} \mathbf{K}_{bb}^{-1} \mathbf{K}_{ba}) \end{bmatrix} \quad (\text{B3})$$

where \mathbf{y}_{new} is the set of observations encountered at \mathbf{X}_f , $\mathbf{D}_a = (\mathbf{S}_a^{-1} - \mathbf{K}'_{aa})^{-1}$, $\boldsymbol{\mu}_a$ and \mathbf{S}_a are the mean vector and covariance matrix of the old approximate posterior distribution, respectively, \mathbf{K}'_{aa} is the kernel operation of the old approximate posterior, α is the order of the α -divergence used to generalize the Kullback–Leibler divergence, and *diag* refers to the diagonal elements of the matrix within the parentheses. Equation (A5) must then be updated to

$$\frac{\partial \mu_*}{\partial x_{*,j}} = \frac{\partial \mathbf{K}_{*b}}{\partial x_{*,j}} \mathbf{L}_b^{-T} \mathbf{D}^{-1} \mathbf{L}_b^{-1} \mathbf{K}_{bf}^{-1} \boldsymbol{\Sigma}_y^{-1} \hat{\mathbf{y}}. \quad (\text{B4})$$

REFERENCES

¹J. McCammon, B. Roux, G. Voth, and W. Yang, *J. Chem. Theory Comput.* **10**, 2631 (2014).

- ²A. Warshel and M. Levitt, *J. Mol. Biol.* **103**, 227–249 (1976).
³M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.* **11**, 700–733 (1990).
⁴U. C. Singh and P. A. Kollman, *J. Comput. Chem.* **7**, 718–730 (1986).
⁵J. Gao and M. A. Thompson, *Combined Quantum Mechanical and Molecular Mechanical Methods* (ACS Publications, 1998), Vol. 712.
⁶H. Lin and D. G. Truhlar, *Theor. Chem. Acc.* **117**, 185–199 (2007).
⁷H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.* **48**, 1198–1229 (2009).
⁸M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.* **99**, 4899–4907 (1977).
⁹M. J. S. Dewar, E. G. Zorbisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).
¹⁰J. J. P. Stewart, *J. Comput. Chem.* **10**, 209–220 (1989).
¹¹M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, *Phys. Rev. B* **58**, 7260–7268 (1998).
¹²M. Elstner, T. Frauenheim, E. Kaxiras, G. Seifert, and S. Suhai, *Phys. Status Solidi B* **217**, 357–376 (2000).
¹³T. Frauenheim, G. Seifert, M. Elstner, T. Niehaus, C. Köhler, M. Amkreutz, M. Sternberg, Z. Hajnal, A. D. Carlo, and S. Suhai, *J. Phys.: Condens. Matter* **14**, 3015–3047 (2002).
¹⁴R. W. Zwanzig, *J. Chem. Phys.* **22**, 1420–1426 (1954).
¹⁵P. S. Hudson, S. Boresch, D. M. Rogers, and H. L. Woodcock, *J. Chem. Theory Comput.* **14**, 6327–6335 (2018).
¹⁶T. J. Giese and D. M. York, *J. Chem. Theory Comput.* **15**, 5543–5562 (2019).
¹⁷Y. Zhou and J. Pu, *J. Chem. Theory Comput.* **10**, 3038–3054 (2014).
¹⁸B. Kim, R. Snyder, M. Nagaraju, Y. Zhou, P. Ojeda-May, S. Keeton, M. Hege, Y. Shao, and J. Pu, *J. Chem. Theory Comput.* **17**, 4961–4980 (2021).
¹⁹X. Pan, P. Li, J. Ho, J. Pu, Y. Mei, and Y. Shao, *Phys. Chem. Chem. Phys.* **21**, 20595–20605 (2019).
²⁰J. Wu, L. Shen, and W. Yang, *J. Chem. Phys.* **147**, 161732 (2017).
²¹L. Shen and W. Yang, *J. Chem. Theory Comput.* **14**, 1442–1455 (2018).
²²L. Bösel, M. Thürlmann, and S. Riniker, *J. Chem. Theory Comput.* **17**, 2641–2658 (2021).
²³X. Pan, J. Yang, R. Van, E. Epifanovsky, J. Ho, J. Huang, J. Pu, Y. Mei, K. Nam, and Y. Shao, *J. Chem. Theory Comput.* **17**, 5745–5758 (2021).
²⁴J. Zeng, T. J. Giese, S. Ekesan, and D. M. York, *J. Chem. Theory Comput.* **17**, 6993–7009 (2021).
²⁵C. L. Gómez-Flores, D. Maag, M. Kansari, V.-Q. Vuong, S. Irle, F. Gräter, T. Kubař, and M. Elstner, *J. Chem. Theory Comput.* **18**, 1213–1226 (2022).
²⁶R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, *J. Chem. Theory Comput.* **11**, 2087–2096 (2015).
²⁷L. Shen, J. Wu, and W. Yang, *J. Chem. Theory Comput.* **12**, 4934–4946 (2016).
²⁸L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, and W. E, *Adv. Neural Inf. Process. Syst.* **31**, 4436–4446 (2018).
²⁹H. Wang, L. Zhang, J. Han, and W. E, *Comput. Phys. Commun.* **228**, 178–184 (2018).
³⁰L. Zhang, J. Han, H. Wang, R. Car, and W. E, *Phys. Rev. Lett.* **120**, 143001 (2018).
³¹A. A. Peterson, R. Christensen, and A. Khorshidi, *Phys. Chem. Chem. Phys.* **19**, 10978–10985 (2017).
³²K. Rossi, V. Jurásková, R. Wischert, L. Garel, C. Corminboeuf, and M. Ceriotti, *J. Chem. Theory Comput.* **16**, 5139–5149 (2020).
³³G. Imbalzano, Y. Zhuang, V. Kapil, K. Rossi, E. A. Engel, F. Grasselli, and M. Ceriotti, *J. Chem. Phys.* **154**, 074102 (2021).
³⁴T. Devergne, T. Magrino, F. Pietrucci, and A. M. Saitta, *J. Chem. Theory Comput.* **18**, 5410–5421 (2022).
³⁵D. J. C. MacKay, *Network: Comput. Neural Syst.* **6**, 469 (1995).
³⁶J. Lampinen and A. Vehtari, *Neural Network* **14**, 257–274 (2001).
³⁷Y. Gal and Z. Ghahramani, in *Proceedings of The 33rd International Conference on Machine Learning, Proceedings of Machine Learning Research*, Vol. 48, edited by M. F. Balcan and K. Q. Weinberger (PMLR, New York, NY, 2016), pp. 1050–1059.
³⁸*Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, edited by S. Aarti and Z. Jerry (PMLR, 2017), Vol. 54.
³⁹C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, 2006).
⁴⁰Y.-H. Tang, D. Zhang, and G. E. Karniadakis, *J. Chem. Phys.* **148**, 034101 (2018).

- ⁴¹R. Jinnouchi, F. Karsai, and G. Kresse, *Phys. Rev. B* **100**, 014105 (2019).
- ⁴²J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, *npj Comput. Mater.* **6**, 20 (2020).
- ⁴³G. Schmitz and O. Christiansen, *J. Chem. Phys.* **148**, 241704 (2018).
- ⁴⁴R. Meyer and A. W. Hauser, *J. Chem. Phys.* **152**, 084112 (2020).
- ⁴⁵G. Raggi, I. F. Galván, C. L. Ritterhoff, M. Vacher, and R. Lindh, *J. Chem. Theory Comput.* **16**, 3989–4001 (2020).
- ⁴⁶D. Born and J. Kästner, *J. Chem. Theory Comput.* **17**, 5955–5967 (2021).
- ⁴⁷C. Teng, Y. Wang, D. Huang, K. Martin, J.-B. Tristan, and J. L. Bao, *J. Chem. Theory Comput.* **18**, 5739–5754 (2022).
- ⁴⁸W. Hu, P. Li, J.-N. Wang, Y. Xue, Y. Mo, J. Zheng, X. Pan, Y. Shao, and Y. Mei, *J. Chem. Theory Comput.* **16**, 6814–6822 (2020).
- ⁴⁹A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Phys. Rev. Lett.* **104**, 136403 (2010).
- ⁵⁰E. Uteva, R. S. Graham, R. D. Wilkinson, and R. J. Wheatley, *J. Chem. Phys.* **149**, 174114 (2018).
- ⁵¹B. Kolb, P. Marshall, B. Zhao, B. Jiang, and H. Guo, *J. Phys. Chem. A* **121**, 2552–2557 (2017).
- ⁵²A. Kamath, R. A. Vargas-Hernández, R. V. Krems, T. Carrington, and S. Manzhos, *J. Chem. Phys.* **148**, 241702 (2018).
- ⁵³K. Tran, W. Neiswanger, J. Yoon, Q. Zhang, E. Xing, and Z. W. Ulissi, *Mach. Learn.: Sci. Technol.* **1**, 025006 (2020).
- ⁵⁴L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, *J. Chem. Phys.* **125**, 024106 (2006).
- ⁵⁵Y. Zhou, P. Ojeda-May, M. Nagaraju, B. Kim, and J. Pu, *Molecules* **23**, 2652 (2018).
- ⁵⁶N. Menshutkin, *Z. Phys. Chem.* **5**, 589–600 (1890).
- ⁵⁷W. Thiel, *MNDO97 Version 5.0* (University of Zurich, Zurich, Switzerland, 1998).
- ⁵⁸B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ochinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, *J. Comput. Chem.* **30**, 1545–1614 (2009).
- ⁵⁹W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926–935 (1983).
- ⁶⁰A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, *J. Phys. Chem. B* **102**, 3586–3616 (1998).
- ⁶¹J. Gao and X. Xia, *J. Am. Chem. Soc.* **115**, 9667–9675 (1993).
- ⁶²T. Darden, D. York, and L. Pedersen, *J. Chem. Phys.* **98**, 10089–10092 (1993).
- ⁶³K. Nam, J. Gao, and D. M. York, *J. Chem. Theory Comput.* **1**, 2–13 (2005).
- ⁶⁴C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785–789 (1988).
- ⁶⁵A. D. Becke, *J. Chem. Phys.* **98**, 5648–5652 (1993).
- ⁶⁶P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.* **98**, 11623–11627 (1994).
- ⁶⁷M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, *J. Chem. Phys.* **77**, 3654–3665 (1982).
- ⁶⁸M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 16 Revision C.01*, Gaussian, Inc., Wallingford, CT, 2016.
- ⁶⁹F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- ⁷⁰A. G. De G. Matthews, M. Van Der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. León-Villagrà, Z. Ghahramani, and J. Hensman, *J. Mach. Learn. Res.* **18**, 1–6 (2017).
- ⁷¹T. D. Bui, C. V. Nguyen, and R. E. Turner, *Adv. Neural Inf. Process. Syst.* **30**, 3299–3307 (2017).
- ⁷²M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” *arXiv:1603.04467* (2015).
- ⁷³M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: A system for large-scale machine learning,” in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (USENIX Association, 2016), pp. 265–283.
- ⁷⁴T. D. Bui, C. V. Nguyen, and R. E. Turner, Streaming sparse gaussian process approximations, https://github.com/thangbui/streaming_sparse_gp.
- ⁷⁵D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in *Proceedings of 3rd International Conference on Learning Representations* (2015).
- ⁷⁶J. Gao, *J. Am. Chem. Soc.* **113**, 7796–7797 (1991).
- ⁷⁷P. Su, W. Wu, C. P. Kelly, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. A* **112**, 12761–12768 (2008).
- ⁷⁸K. Okamoto, S. Fukui, and H. Shingu, *Bull. Chem. Soc. Jpn.* **40**, 1920–1925 (1967).
- ⁷⁹T. N. Truong, T.-T. T. Truong, and E. V. Stefanovich, *J. Chem. Phys.* **107**, 1881–1889 (1997).
- ⁸⁰J. Z. Vilseck, S. V. Sambasivarao, and O. Acevedo, *J. Comput. Chem.* **32**, 2836–2842 (2011).
- ⁸¹A. Pohorille, C. Jarzynski, and C. Chipot, *J. Phys. Chem. B* **114**, 10235–10253 (2010).
- ⁸²U. Ryde, *J. Chem. Theory Comput.* **13**, 5745–5752 (2017).
- ⁸³J. Heimdal and U. Ryde, *Phys. Chem. Chem. Phys.* **14**, 12592–12604 (2012).
- ⁸⁴D. Wu and D. A. Kofke, *J. Chem. Phys.* **123**, 054103 (2005).
- ⁸⁵D. Wu and D. A. Kofke, *J. Chem. Phys.* **123**, 084109 (2005).
- ⁸⁶S. Boresch and H. L. Woodcock, *Mol. Phys.* **115**, 1200–1213 (2017).
- ⁸⁷D. Wu and D. A. Kofke, *J. Chem. Phys.* **121**, 8742–8747 (2004).
- ⁸⁸T. Bučko, M. Gesvandtnerová, and D. Rocca, *J. Chem. Theory Comput.* **16**, 6049–6060 (2020).
- ⁸⁹L. Le Gratiet and J. Garnier, *Mach. Learn.* **98**, 407–433 (2015).
- ⁹⁰Y. Zhao, C. Fritsche, G. Hendeby, F. Yin, T. Chen, and F. Gunnarsson, *IEEE Trans. Signal Process.* **67**, 5936–5951 (2019).
- ⁹¹M. Pinheiro, F. Ge, N. Ferré, P. O. Dral, and M. Barbatti, *Chem. Sci.* **12**, 14396–14413 (2021).
- ⁹²R. Snyder, B. Kim, X. Pan, Y. Shao, and J. Pu, *Phys. Chem. Chem. Phys.* **24**, 25134–25143 (2022).
- ⁹³A. S. Christensen and O. A. von Lilienfeld, *CHIMIA* **73**, 1028 (2019).
- ⁹⁴A. S. Christensen, F. A. Faber, and O. A. von Lilienfeld, *J. Chem. Phys.* **150**, 064105 (2019).
- ⁹⁵X. Pan, R. Van, E. Epifanovsky, J. Liu, J. Pu, K. Nam, and Y. Shao, *J. Phys. Chem. B* **126**, 4226–4235 (2022).
- ⁹⁶J. Behler, *J. Chem. Phys.* **134**, 074106 (2011).
- ⁹⁷A. P. Bartók, R. Kondor, and G. Csányi, *Phys. Rev. B* **87**, 184115 (2013).
- ⁹⁸M. K. Titsias, *Variational Learning of Inducing Variables in Sparse Gaussian Processes* (AISTATS, 2009), Vol. 5.