MDPI

Article

Multi-Level Protocol for Mechanistic Reaction Studies Using Semi-Local Fitted Potential Energy Surfaces

Tomislav Piskor [1,2,*], Peter Pinski [1], Thilo Mast [1] and Vladimir Rybkin [1,*]

[1]   HQS Quantum Simulations GmbH, Rintheimer Straße 23, 76131 Karlsruhe, Germany
[2]   Theoretical Physics, Saarland University, 66123 Saarbrücken, Germany
*   Correspondence: tomislav.piskor@quantumsimulations.de or t.piskor@hotmail.de (T.P.);
    vladimir.rybkin@quantumsimulations.de (V.R.)

Abstract: In this work, we propose a multi-level protocol for routine theoretical studies of chemical reaction mechanisms. The initial reaction paths of our investigated systems are sampled using the Nudged Elastic Band (NEB) method driven by a cheap electronic structure method. Forces recalculated at the more accurate electronic structure theory for a set of points on the path are fitted with a machine learning technique (in our case symmetric gradient domain machine learning or sGDML) to produce a semi-local reactive potential energy surface (PES), embracing reactants, products and transition state (TS) regions. This approach has been successfully applied to a unimolecular (Bergman cyclization of enediyne) and a bimolecular (S_N2 substitution) reaction. In particular, we demonstrate that with only 50 to 150 energy-force evaluations with the accurate reference methods (here complete-active-space self-consistent field, CASSCF, and coupled-cluster singles and doubles, CCSD) it is possible to construct a semi-local PES giving qualitative agreement for stationary-point geometries, intrinsic reaction coordinates and barriers. Furthermore, we find a qualitative agreement in vibrational frequencies and reaction rate coefficients. The key aspect of the method's performance is its multi-level nature, which not only saves computational effort but also allows extracting meaningful information along the reaction path, characterized by zero gradients in all but one direction. Agnostic to the nature of the TS and computationally economic, the protocol can be readily automated and routinely used for mechanistic reaction studies.

Keywords: potential energy surface; transition state; reaction mechanism

1. Introduction

Mechanistic studies of chemical reactions are one of the most important and widespread applications of computational quantum chemistry [1]. A minimal meaningful workflow consists of locating reactants, products and a corresponding transition state (TS). This allows one to assess the thermodynamics and kinetics of reactions *without* thermal contributions and thereby reveal reaction mechanisms. It is highly desirable to perform vibrational analysis for stationary points of a potential energy surface (PES) to elucidate whether the structures correspond to minima or saddle points. This also allows the computation of reaction rate coefficients via canonical transition-state theory (TST) taking vibrational degrees of freedom into account [2]. To be more rigorous, one should also compute reaction paths connecting the stationary points, one option being following the intrinsic reaction coordinate (IRC) starting from the TS [3].

These objectives are routinely reached by direct dynamics approaches [4], i.e., methods evaluating energy and its derivatives on the fly. The number of ab initio quantum chemistry calls typically reaches hundreds and more. One should not ignore unsuccessful attempts to locate TS and find IRC, which are not uncommon in mechanistic reaction studies. A qualitatively accurate description of reaction paths involving the breaking or forming of chemical bonds can only be achieved with advanced electronic structure methods. Often but not

always those include static correlation, i.e., multi-reference and multi-configurational wave function approaches [5]. As an alternative, Density Functional Theory (DFT) with carefully selected hybrid and double-hybrid functionals can provide accurate results [6]. All these methods (multi-reference and advanced DFT approximations) are, however, computationally demanding even for moderate-size systems. Therefore, the researchers often apply computationally cheaper methods for the mechanistic reaction studies, employing more accurate electronic structure theories only for stationary points (for an example, see the study of the reaction between ferrocenium and trimethylphosphine [7]). Some such heuristic method combinations are known as composite methods, or recipes, e.g., the Gaussian-n family [8–10] and CBS-QB3 [11]. Despite many successful applications, these approaches imply properties inconsistent with geometries. They are designed not for PES exploration but rather for energy evaluation at stationary points, and are prone to unpredictable errors (for an example, see [12]).

Ideally, a composite method should provide accurate energies consistent with geometric structures at least in a relevant part of the configuration space, while being computationally feasible, i.e., based on only a few energy/force evaluations at the high level of electronic structure theory. This implies obtaining a locally fitted PES: once it is available, one can complete many tasks "free" of charge, including reactive molecular dynamics simulations, canonical and variational TST calculations [2], and so on.

A fitted PES can be efficiently generated using one of the rapidly evolving machine learning (ML) approaches (for a general method overview see review [13] and perspective [14]). These typically require little, if any, feature design: mainly, the structures with associated energies, and sometimes also energy gradients, are needed as input. They demonstrate remarkable flexibility, successfully describing even non-adiabatic processes [15,16] and systems without a classical atomistic structural formula [17]. The main type of application for such ML-PESs is extensive molecular dynamics (MD) sampling (see exemplary applications to organic crystals [18] and liquid water [19]). The resulting PES is often called "global" as it embraces vast regions of the configuration space, although its reactive regions are typically not covered. There are only a handful of applications of ML-PES fitting to reactive systems: second-order nucleophilic substitution (S_N2) [20], pericyclic [21], decomposition [22], dissociation [23], Diels–Alder [24] and proton-transfer reactions [25]. In addition, ML-PESs have been successfully used for automatic mechanism discovery [26]. All these applications, however, aimed at "global" PES fitting required abundant data: typical sets include at least thousands of data points. Most of them, therefore, used DFT methods or cheaper many-body correlated wave functions as an underlying electronic structure theory. At this point, we would like to highlight that the term "global" here is applied to the PES, rather than to the nature of an ML descriptor [27].

To facilitate routine mechanistic reaction studies, one should be able to construct a semi-local reactive PES with only a few hundreds of data points using high-level ab initio methods. Such work has been performed by Young et al. [28] for several model processes, although the authors restricted themselves to DFT methods, which are typically insufficient to describe chemical reactions due to their single-reference character.

In this work, we propose a multi-level protocol for generating a semi-local reactive PES for routine mechanistic reaction studies based on ML methods with small data sets and a combination of electronic structure theories. Computationally cheaper DFT is used to generate relevant structures, whereas high-level ab initio methods are applied to refine the energetics. The multi-level nature of the protocol makes it suitable for incorporating evolving quantum computing methods which promise a quantum advantage for correlated electronic-structure problems [29,30], while remaining non-routine. The feasibility of the approach is demonstrated with two organic reactions: a monomolecular Bergman cyclization of enediyne to para-benzyne [31] and a bimolecular S_N2 reaction of chloromethane with a bromide ion [32]. The former is a textbook example of a chemical reaction involving a multi-reference character [33] and is treated with a complete-active-space self-consistent-field

method, whereas the latter can be described with a single-reference correlated method [34] and is treated with a coupled-cluster approach.

This paper is organized as follows. In Section 4, we introduce the simulation protocol. In particular, we focus on obtaining relevant geometries for the investigated reactions, as well as reference methods and the ML technique of choice. Next, we present the results: PES sections, energy and force-prediction errors, and performance of the ML-PES for geometry optimization, vibrations and reaction rates. We conclude and give an outlook for this work in Section 5.

2. Results

In this section, we present the results for our investigated examples. First, the PES scans for the reference method and sGDML, as well as the energy differences between the two methods, will be shown. The next analysis will provide information about the mean absolute errors of the geometries between the reference and ML method for the reactant and TS. To perform geometry optimizations for both structures, we used the open-source package `pysisyphus`. In addition, we computed and compared harmonic vibrations for the optimized structures. Besides the intrinsic reaction coordinate we finally compared the reaction rate coefficients for all methods and reactions. The intrinsic reaction coordinate was determined with `pysisyphus` as well, where the Euler-Predictor-Corrector integrator was used.

2.1. Bergman Cyclization of Enediyne

The reaction is shown in Figure 1, including the optimized reactant and product.
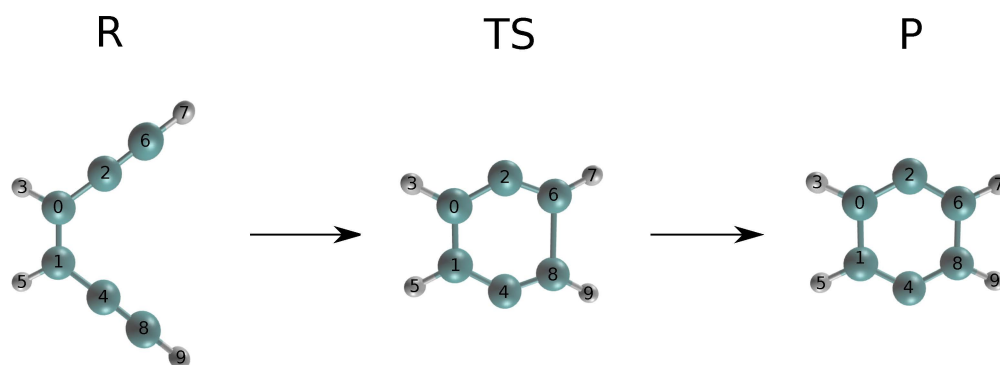

Figure 1. Bergman cyclization of enediyne. The reactant, transition state and product are labeled with R, TS and P, respectively. Elements are colored as follows: carbon (cyan), hydrogen (white).

2.1.1. Model Training

We created an sGDML model for the data set consisting of 200 data points: we used 150 training, 30 validation and 20 test points. The regularization parameter was set to $\lambda = 10^{-15}$ and the best length scale was found to be $\sigma = 6$. With these parameters, we obtained a model with a mean absolute error (MAE) of 0.0067 kJ/mol and a Root Mean Square Error (RMSE) of 0.0071 kJ/mol for the energy, whereas the MAE and RMSE for the forces were 0.0033 kJ/Å·mol and 0.0075 kJ/Å·mol, respectively, on the test data set.

2.1.2. Potential Energy Surface Scan

The PES profiles corresponding to the DFT-optimized NEB path computed with CASSCF and sGDML are shown in Figure 2a. As expected from the small values of MAE and RMSE, the two surfaces agree within chemical accuracy (see Figure 2b). Starting from the reactant state and moving towards the TS, the energy error is small and practically constant. The smallest error can be found in the vicinity of the TS, increasing significantly towards the product state, although still being two orders of magnitude within chemical accuracy.
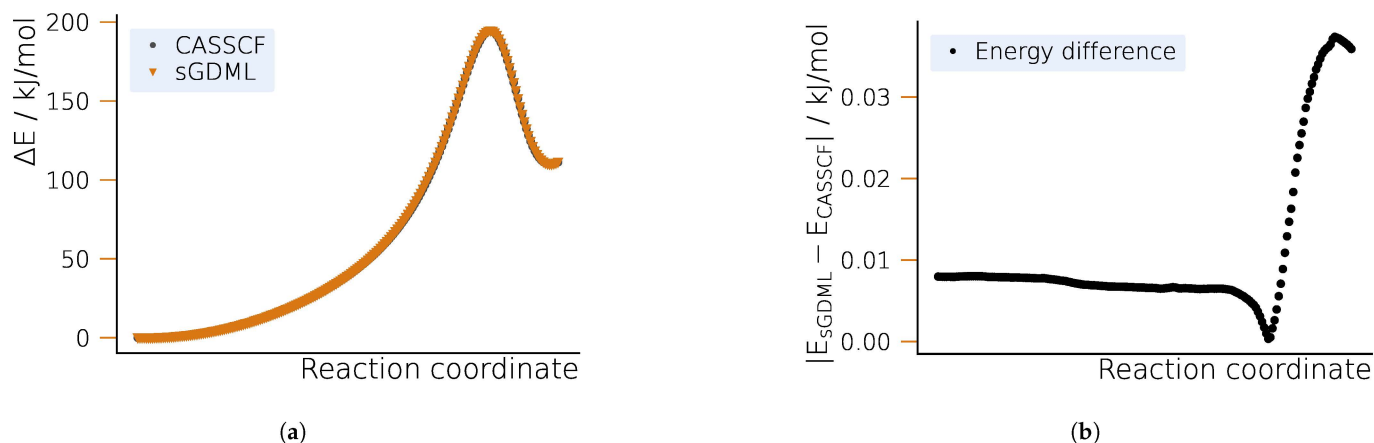
(a)

(b)

Figure 2. Potential energy profile for the Bergman cyclization of enediyne along the optimized NEB path. (**a**) Potential energy surface for CASSCF and sGDML. (**b**) Energy differences between the learned and CASSCF potential energy surface.

2.1.3. Geometry Optimization

In Figure 3, we compare the geometric structures of the stationary structures as obtained by sGDML and the reference method, CASSCF. Directly comparing the optimized geometries from the two methods, we find that the MAE for all interatomic distances never reaches the value of 0.01 Å for the reactant state. However, for the TS, the error is considerably larger reaching a deviation of 0.22 Å in the worst case. Although this is a significant value, the corresponding distance is between two non-bonded hydrogen atoms separated by more than 5 Å. For other distances, the error does not exceed 0.1 Å.
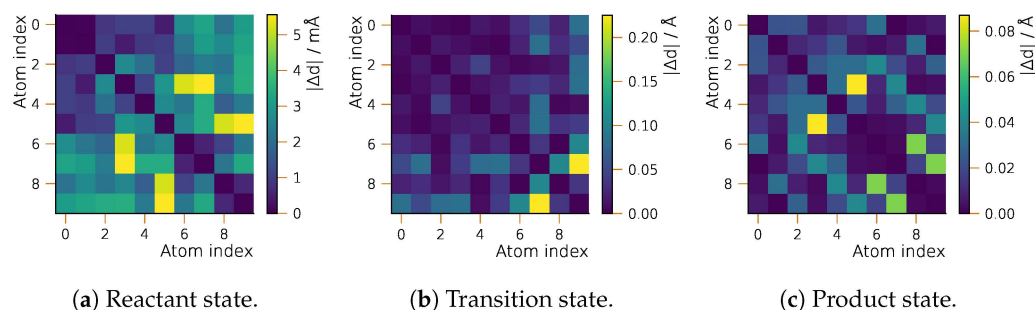

(**a**) Reactant state.　　　　(**b**) Transition state.　　　　(**c**) Product state.

Figure 3. Accuracy of geometric structures from the semi-local fitted PES for the Bergman cyclization of enediyne: position differences (MAE) for sGDML- and CASSCF-optimized structures. Atom numbering is given in Figure 1.

2.1.4. Vibrations, Intrinsic Reaction Coordinates and Reaction Rate Coefficients

After obtaining the optimized TS, we analyze the connection between it and the basins of reactants and products by integrating the IRC as shown in Figure 4. On the fitted semi-local PES, the TS does connect the reactants and products by a minimum energy path, which is in qualitative agreement with the reference CASSCF calculation along the entire curve as indicated by the maximum and RMS errors in the gradients.

The successful computation of the IRC gives the first positive accuracy assessment of vibrational modes on the semi-local fitted PES: the imaginary frequency at the TS is needed to define the direction of the path. A more detailed look at vibrational frequencies reveals only semi-qualitative agreement between the sGDML and the reference CASSCF PES (see Table S1 in the SI): the differences between the frequencies reach up to approximately 200 cm^{-1} for high-frequency nodes, which is still less than 10%.
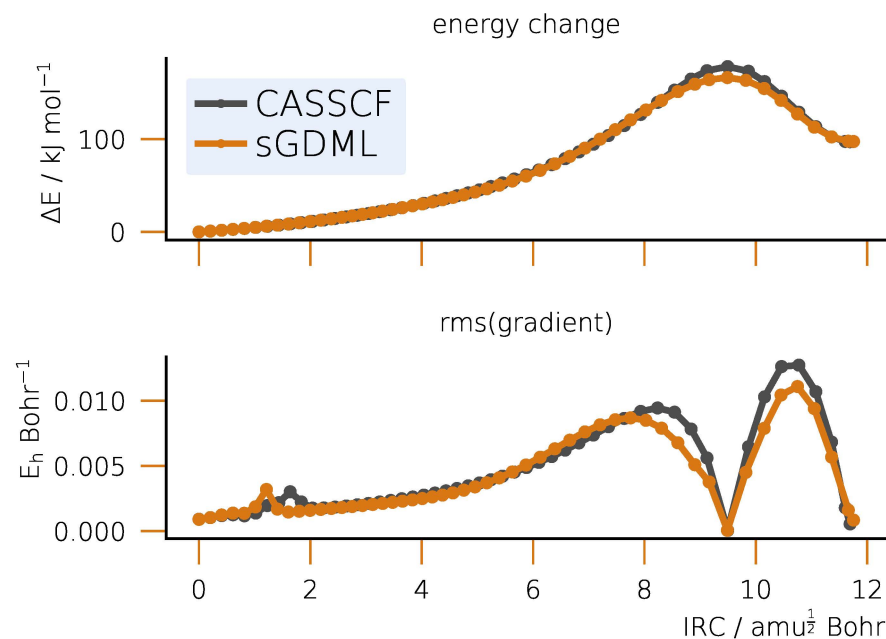
Figure 4. IRC for the Bergman cyclization of enediyne: relative energies and Root Mean Square gradients. The TS corresponds to the maximum energy value at an IRC displacement value of approximately 9.5.

Rate coefficients dependent on structures, vibrations and reaction barriers are good integrated indicators of the fitted PES quality. The barrier heights for the Bergman cyclization are $\Delta E^{\ddagger}_{\text{CASSCF}} = 194.59$ kJ for the reference method and $\Delta E^{\ddagger}_{\text{sGDML}} = 194.57$ kJ for the fitted PES, which are in excellent qualitative agreement. Assuming T = 300 K, we obtain the following reaction rate coefficients from the conventional TST (as described in the SI, Section S1) for the two methods: $k_{\text{CASSCF}} = 3.4548 \times 10^{-22}$ m^{3}s^{-1} for the CAS(12, 12) and $k_{\text{sGDML}} = 4.3856 \times 10^{-22}$ m^{3}s^{-1} for sGDML. The agreement is semi-qualitative and stems from the pre-exponential factors defined by the partition functions, which in turn depend on vibrations and structures being less accurate than the barriers.

2.2. S$_N$2 Reaction of Chloromethane with Bromide

The reaction between CH_3Cl and Br^- is shown in Figure 5, including the optimized reactant and product.
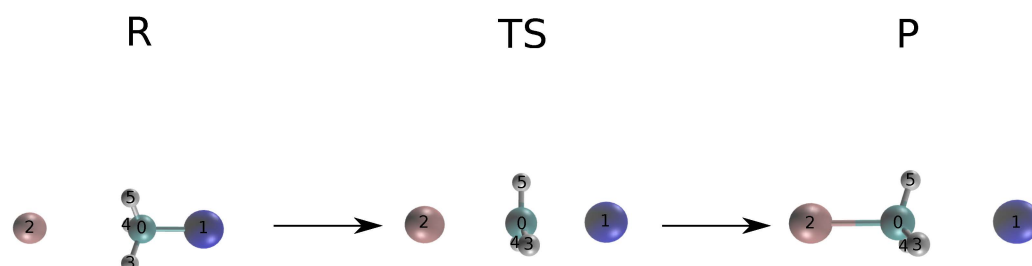

Figure 5. S$_N$2 reaction of chloromethane with bromide. The reactant, TS and product states are given as R, TS and P, respectively. Elements are colored as follows: carbon (cyan), hydrogen (white), bromine (pink), chlorine (blue).

2.2.1. Model Training

To generate an ML model, we created a data set with a total of 100 geometric configurations separated into 50 training, 30 validation and 20 test points. The regularization parameter was set to $\lambda = 10^{-15}$ and the best length scale was found to be $\sigma = 32$. With these parameters, we obtained a model with both the MAE and RMSE being 0.0013 kJ/mol,

whereas the MAE and RMSE for the forces were 0.0105 kJ/Å·mol and 0.0293 kJ/Å·mol, respectively, on the test data set.

2.2.2. Potential Energy Surface

The potential energy profiles of the S_N2 reaction for the reference method, CCSD, and sGDML are shown in Figure 6a. The good agreement between the CCSD and the sGDML fits along the (DFT) NEB path can be immediately seen in Figure 6b. The energy difference is of a similar order as that for the Bergman cyclization and does not exceed 0.02 kJ/mol. However, we observe the higher errors in the TS region rather than in the vicinity of the product state.
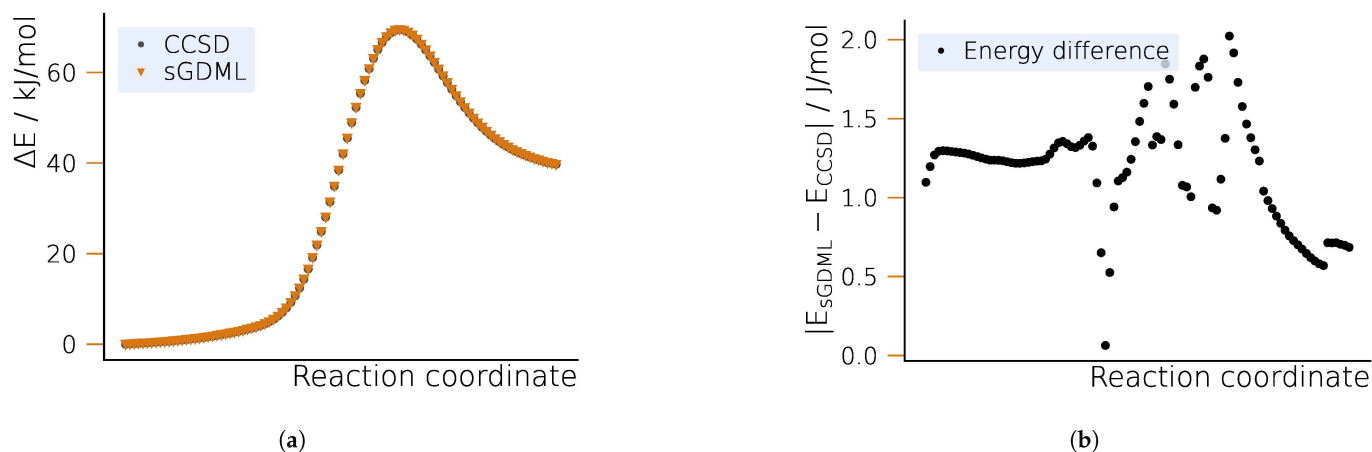

(**a**)

(**b**)

Figure 6. Potential energy profile for the S_N2 reaction of chloromethane and bromide along the optimized NEB path. (**a**) Potential energy surface for CCSD and sGDML. (**b**) Energy differences between the learned and CCSD potential energy surface.

2.2.3. Geometry Optimization

In Figure 7, we compare the geometric structures of the stationary structures obtained by sGDML and the reference method. The general agreement between the reference and fitted PES is better than for the Bergman cyclization, where the MAE never exceeds 0.025 Å for a particular interatomic distance. For the S_N2 reaction, the better agreement is reached for the TS with the MAE staying below a value of 0.01 Å. The largest deviation in the case of the reactant state can be found for the distances between the non-bonded bromide ion (indicated by index 2) and the hydrogen atoms from the methylene group (indices 3, 4 and 5).
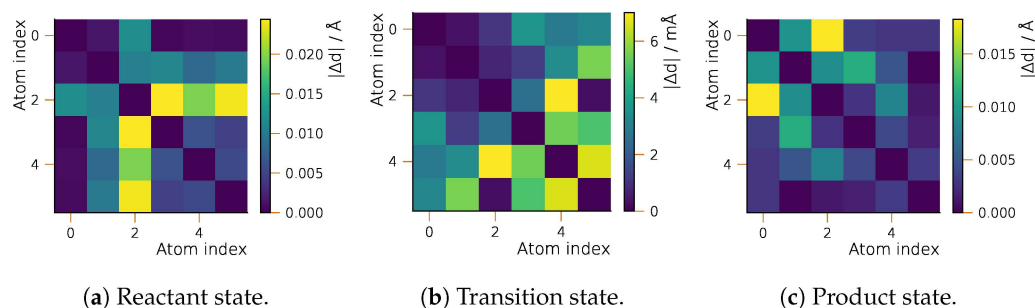

(**a**) Reactant state.

(**b**) Transition state.

(**c**) Product state.

Figure 7. Accuracy of geometric structures from the semi-local fitted PES for the S_N2 reaction of chloromethane and bromide: position differences (MAE) for sGDML- and CCSD-optimized structures. Atom numbering is given in Figure 5.

2.2.4. Vibrations, Intrinsic Reaction Coordinates and Reaction Rate Constants

After obtaining the optimized TS, we analyze the connection between it and the basins of the reactants and products by integrating the IRC as shown in Figure 8. On the fitted

semi-local PES, the TS does connect the reactants and products by a minimum energy path, which is in qualitative agreement with the reference CCSD calculation along the entire path as indicated by the maximum and RMSE in the gradients. Although the latter exhibits several minor jumps it does not lead to non-smooth IRC, this fact being explained by the small data set used to fit the PES.
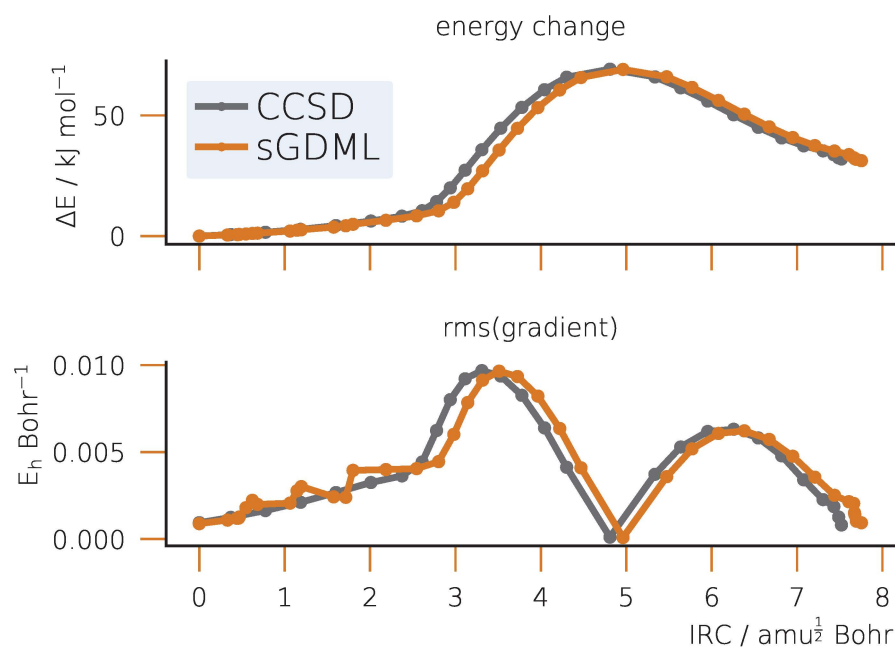

Figure 8. IRC for the S_N2 reaction of chloromethane with bromide: relative energies and and Root Mean Square gradients. The TS corresponds to the maximum energy value at an IRC displacement value of approximately 4.8.

As in the previous example, successful computation of the IRC indicates the correct Hessian structure at the TS with the only imaginary eigenvalue corresponding to the same eigenvector as in the reference method. A more detailed look at vibrational frequencies reveals only qualitative agreement between the sGDML and the reference CCSD PES (see Table S2 in the SI): the differences between the frequencies reach several hundreds of cm^{-1} for both high- and low-frequency modes.

The barrier heights for the S_N2 reaction were identical up to the second digit: $\Delta E^{\ddagger} = 69.47$ kJ for CCSD and the fitted PES. At T = 300 K, the CCSD reaction rate coefficient is 1.8187×10^{-31} m^{3}s^{-1}, whereas this value is 2.2287×10^{-32} m^{3}s^{-1} for sGDML, which is an order of magnitude smaller than the reference method. This difference occurs due to the larger deviations in the harmonic vibrations, defining the pre-exponential factors in the TST equations.

3. Discussion

The most important property of the fitted semi-local PES surfaces obtained by the proposed protocol is the general stability with respect to PES exploration techniques. Indeed, for both unimolecular and bimolecular reactions geometry optimizations have converged to the stationary points, connected by physically meaningful minimum energy paths. This is achieved by the multi-level nature of our approach.

On the one hand, the NEB driven by the cheaper method (here DFT with a PBE exchange-correlation functional) generates structures relatively close to the reactive region of the system as defined by more accurate reference electronic structure methods (here, CASSCF or CCSD). Indeed, DFT methods are known to predict reasonable geometric structures [6] (even if the energy barriers are not precise). This is illustrated by the S_N2 reaction studied in this work: in Figure S1 of the SI we see that the highest energy point on

the NEB path for both PBE and CCSD is located at similar values of the reaction coordinate, although the energy barriers differ dramatically.

On the other hand, structures sampled by NEB using the cheaper method should be far enough from the minimum energy path of the reference method. The points on the true minimum energy path (approximated by NEB [35]) have only one non-zero gradient component—the one along the path tangent direction [3]. Keeping in mind that sGDML uses gradients as inputs for fitting, providing only structures on the minimum energy path would provide no information about the nature of the PES along orthogonal directions and make the model very sensitive to numerical noise. Consequently, using a cheaper electronic structure method for the NEB simulation is not only computational effort saving but also essential for the quality of the data set used to train a gradient-based ML model.

Despite qualitative agreement achieved by the semi-local fitted PES in structures, energy barriers and IRC, a more subtle property, the vibrational spectrum, is computed less accurately. Although the structure of Hessian is qualitatively correct, some frequencies can differ by several hundreds of cm^{-1}. This is particularly noticeable for the S_N2 example and as a consequence leads to a significant error (order of magnitude) in the reaction rate coefficient as compared to the reference value. This effect must have to do with a smaller number of training points (only 50) used for the model as compared to the Bergman cyclization (150 points). Another reason is the fact that the relevant PES region for the S_N2 is flatter than for the Bergman cyclization (compare Figure 4 and Figure 8), one indication of which is the lower energy barrier for the former reaction. This makes a finite-difference evaluation of the second derivatives of energy less reliable and requires a better PES sampling for obtaining qualitative results.

4. Materials and Methods

4.1. Simulation Protocol

We applied the general protocol for fitting a semi-local reactive PES, which includes the following steps:

1. Optimize reactants and products with a cheaper electronic structure method;
2. Find an approximate reaction path using the Nudged Elastic Band (NEB) method [35] with a cheaper electronic structure method;
3. Select points along the reaction path to form the data set;
4. Calculate energies and forces with a reference correlated method;
5. Split this data set into training, validation and test subsets;
6. Fit the ML-PES, validate and test using the corresponding subsets.

After being generated, the fitted PES was employed to compute the following properties: stationary points, harmonic vibrations and intrinsic reaction coordinates in both directions from the TS. In addition, we calculated rate coefficients using transition-state theory (TST). To evaluate the quality of the fitted semi-local PES the properties were compared with those obtained directly by the reference electronic structure method with energies and forces calculated on the fly.

The details of each step are described below.

4.2. Electronic Structure Calculations

We used DFT with the PBE exchange-correlation functional [36] in combination with the double-zeta split-valence def2-SVP basis set [37] as a cheaper electronic structure method for both reactions. These calculations have been performed with `NWChem` [38].

For the Bergman cyclization the complete-active-space self-consistent field (CASSCF) [39] method with the double-zeta split-valence def2-SVP basis set [37] was used as the reference. For the initial configuration, we used MP2 natural orbitals to select the active space guess, which consisted of 12 electrons in 12 orbitals (12, 12) as suggested by Lindh and Persson [40]. The CASSCF wave function for atomic configurations in the data set was calculated with the previous guess for the molecular orbitals, so that no discontinuities occurred in the potential energy surface. CASSCF calculations were performed using `PySCF` [41,42].

The reference correlated method for the S_N2 reaction was coupled-cluster singles and doubles (CCSD) [43] based on the Hartree–Fock reference within the spin-restricted formalism. As in the previous example, CCSD calculations were performed with `PySCF` [42]. The basis set of choice was the double-zeta correlation-consistent basis set cc-PVDZ [44–46].

4.3. Data Set Generation

`NWChem` [38] was used to generate the data sets by performing NEB [35] calculations. For both reactions, DFT (the cheaper method) energies and forces were used, with reactants as start points and products as end points. We used 100 beads for the S_N2 reaction and 200 for the Bergman cyclization. The geometries were randomly selected from the optimized elastic band. Thus, they lie on the approximate transition path of the PES obtained from the cheaper method, rather than on the path of the reference method.

4.4. Machine Learning Fit of the Semi-Local Reactive PES

We opted for the symmetric gradient domain machine learning (sGDML) method [47–49] to fit the PES. Based on kernel ridge regression, it efficiently employs the forces and does not require feature design: inverse atomic distances are used as features and the interatomic forces as the output. sGDML force field is conservative by construction and is particularly efficient for small and medium-sized molecules. Importantly, sGDML has been shown to faithfully reproduce PES from a limited number of structures. Moreover, it has been successfully applied to small molecules using coupled-cluster reference data [50].

To perform one sGDML fit, a certain number of points from the data set are taken to define the training, validation and test sets. These numbers are detailed in Section 2. The first step within one sGDML run is to generate a first model by constructing the kernel matrix out of the training points and setting values for the hyper-parameters σ and λ. The regularization parameter is hereby fixed to a certain value and the length scale σ is varied during the learning process. After the training step, the model is validated against the validation set, where the energy and gradient errors are determined. This procedure is continued until the best σ is found. As a last step the optimized model is validated against the test set, which has no data from either the training or the validation set, and the errors in energy and gradients are compared.

4.5. Property Calculation

We optimized the stationary points of the PES using DFT-optimized structures as initial guess and followed the IRC from the TS in both directions with `pysisyphus` [51]. Both, the fitted PES and the reference electronic structure method in `PySCF` [42] were employed as energy functions. Gradients for both PES types were calculated analytically.

We calculated harmonic vibrations for all stationary points on the PES and computed the reaction rate coefficients using conventional TST with a harmonic oscillator–rigid rotor approximation for the partition functions. The working equations and the corresponding rates are given in the Supporting Information (Section S1).

5. Conclusions

In this work, we have proposed a multi-level protocol for reaction mechanism studies aiming to match the accurate electronic structure theory description at a reduced computational cost. The approach involves a cheaper electronic structure method to generate structures along the reaction path via the NEB method, evaluating energies and gradients for a restricted set of them with an accurate (reference) theory and fitting a semi-local reactive PES using an ML method, sGDML. The fitted PES is then used for computing properties.

The protocol has been applied to a unimolecular (Bergman cyclization) and a bimolecular (S_N2) reaction using PBE/CASSCF and PBE/CCSD, respectively, as cheap/reference method pairs, the results being compared with those obtained on the fly by the reference method. Our approach achieves quantitative agreement in structure optimization for

the PES stationary points, reaction energy barriers and IRC following using as little as 50–150 training points (corresponding to energy and gradient calculations with the reference method), whereas the agreement in vibrational frequencies and reaction coefficients is only (semi-)qualitative. The key to the performance of the protocol is its multi-level character as the differences between the reference and cheap methods are essential to provide meaningful information about the reactive PES region. At the same time, the computational gain achieved is significant as the NEB sampling alone would require several thousand energy and gradient evaluations in the high-level method and is not tractable for routine calculations.

The results are encouraging as an important objective of mechanistic reaction study can be performed with a high-level electronic structure method without prior knowledge of the TS structure (due to the application of the NEB) requiring only a scarce amount of data. The protocol is simple and can be automatized for routine calculations.

At the same time, we have not considered more complex reactions with multiple steps, product branching and shallow minima. These cases would require more subtle approaches to sampling and larger amounts of training data.

Our simple protocol can be further improved by applying the smart selection of training points using molecular fingerprints [52] and using methods computationally even cheaper than DFT (such as modern tight binding, GFN2-XTB [53]) for sampling structures. The moderately accurate reference methods applied here should be substituted by those providing qualitatively correct PES, such as multi-reference approaches, to match the experimental accuracy. Furthermore, the potential of different ML fitting techniques should be explored.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms1010000/s1. Reference [54] are cited in the Supplementary Materials.

Author Contributions: Conceptualization—T.P. and V.R.; methodology—T.P., P.P., T.M. and V.R.; calculations—T.P.; writing original manuscript—T.P. and V.R.; writing—review and editing, P.P. and T.M.; visualization—T.P.; supervision—V.R. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available from the authors on reasonable request.

Conflicts of Interest: All ahthors were employed by the company "HQS Quantum Simulations GmbH". But they declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Hratchian, H.P.; Schlegel, H.B. Chapter 10—Finding minima, transition states, and following reaction pathways on ab initio potential energy surfaces. In *Theory and Applications of Computational Chemistry*; Dykstra, C.E., Frenking, G., Kim, K.S., Scuseria, G.E., Eds.; Elsevier: Amsterdam, The Netherlands, 2005; pp. 195–249. [CrossRef]
2. Truhlar, D.G.; Garrett, B.C.; Klippenstein, S.J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800. [CrossRef]
3. Fukui, K. Formulation of the reaction coordinate. *J. Phys. Chem.* **1970**, *74*, 4161–4163. [CrossRef]
4. Fernández-Ramos, A.; Miller, J.A.; Klippenstein, S.J.; Truhlar, D.G. Modeling the Kinetics of Bimolecular Reactions. *Chem. Rev.* **2006**, *106*, 4518–4584. [CrossRef] [PubMed]
5. Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic Structure Theory*; John Wiley & Sons, Ltd.: Chichester, UK, 2000.
6. Bursch, M.; Mewes, J.M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry. *Angew. Chem. Int. Ed.* **2022**, *61*, e202205735. [CrossRef] [PubMed]
7. Chamkin, A.A.; Serkova, E.S. DFT, DLPNO-CCSD(T), and NEVPT2 benchmark study of the reaction between ferrocenium and trimethylphosphine. *J. Comput. Chem.* **2020**, *41*, 2388–2397. [CrossRef] [PubMed]
8. Curtiss, L.A.; Redfern, P.C.; Raghavachari, K. Gaussian-4 theory. *J. Chem. Phys.* **2007**, *126*, 084108. [CrossRef] [PubMed]

9. da Silva, G. G3X-K theory: A composite theoretical method for thermochemical kinetics. *Chem. Phys. Lett.* **2013**, *558*, 109–113. [CrossRef]

10. Chan, B.; Deng, J.; Radom, L. G4(MP2)-6X: A Cost-Effective Improvement to G4(MP2). *J. Chem. Theory Comput.* **2011**, *7*, 112–120. [CrossRef] [PubMed]

11. Montgomery, J.A., Jr.; Frisch, M.J.; Ochterski, J.W.; Petersson, G.A. A complete basis set model chemistry. VII. Use of the minimum population localization method. *J. Chem. Phys.* **2000**, *112*, 6532–6542. [CrossRef]

12. Karton, A.; Goerigk, L. Accurate reaction barrier heights of pericyclic reactions: Surprisingly large deviations for the CBS-QB3 composite method and their consequences in DFT benchmark studies. *J. Comput. Chem.* **2015**, *36*, 622–632. [CrossRef]

13. Unke, O.T.; Chmiela, S.; Sauceda, H.E.; Gastegger, M.; Poltavsky, I.; Schütt, K.T.; Tkatchenko, A.; Müller, K.R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142–10186. [CrossRef] [PubMed]

14. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, 170901. [CrossRef]

15. Dral, P.O.; Barbatti, M.; Thiel, W. Nonadiabatic Excited-State Dynamics with Machine Learning. *J. Phys. Chem. Lett.* **2018**, *9*, 5660–5663. [CrossRef] [PubMed]

16. Westermayr, J.; Marquetand, P. Machine Learning for Electronically Excited States of Molecules. *Chem. Rev.* **2021**, *121*, 9873–9926. [CrossRef] [PubMed]

17. Lan, J.; Kapil, V.; Gasparotto, P.; Ceriotti, M.; Iannuzzi, M.; Rybkin, V.V. Simulating the ghost: Quantum dynamics of the solvated electron. *Nat. Commun.* **2021**, *12*, 766. [CrossRef] [PubMed]

18. Kapil, V.; Engel, E.A. A complete description of thermodynamic stabilities of molecular crystals. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2111769119. [CrossRef] [PubMed]

19. Gartner, T.E.; Piaggi, P.M.; Car, R.; Panagiotopoulos, A.Z.; Debenedetti, P.G. Liquid-Liquid Transition in Water from First Principles. *Phys. Rev. Lett.* **2022**, *129*, 255702. [CrossRef] [PubMed]

20. Brickel, S.; Das, A.K.; Unke, O.T.; Turan, H.T.; Meuwly, M. Reactive molecular dynamics for the [Cl-CH$_3$-Br]- reaction in the gas phase and in solution: A comparative study using empirical and neural network force fields. *Electron. Struct.* **2019**, *1*, 024002. [CrossRef]

21. Ang, S.J.; Wang, W.; Schwalbe-Koda, D.; Axelrod, S.; Gómez-Bombarelli, R. Active learning accelerates ab initio molecular dynamics on reactive energy surfaces. *Chem* **2021**, *7*, 738–751. [CrossRef]

22. Yang, M.; Bonati, L.; Polino, D.; Parrinello, M. Using metadynamics to build neural network potentials for reactive events: The case of urea decomposition in water. *Catal. Today* **2022**, *387*, 143–149. [CrossRef]

23. de la Puente, M.; David, R.; Gomez, A.; Laage, D. Acids at the Edge: Why Nitric and Formic Acid Dissociations at Air–Water Interfaces Depend on Depth and on Interface Specific Area. *J. Am. Chem. Soc.* **2022**, *144*, 10524–10529. [CrossRef]

24. Young, T.A.; Johnston-Wood, T.; Zhang, H.; Duarte, F. Reaction dynamics of Diels–Alder reactions from machine learned potentials. *Phys. Chem. Chem. Phys.* **2022**, *24*, 20820–20827. [CrossRef] [PubMed]

25. Töpfer, K.; Käser, S.; Meuwly, M. Double proton transfer in hydrated formic acid dimer: Interplay of spatial symmetry and solvent-generated force on reactivity. *Phys. Chem. Chem. Phys.* **2022**, *24*, 13869–13882. [CrossRef]

26. Li, J.; Reiser, P.; Boswell, B.R.; Eberhard, A.; Burns, N.Z.; Friederich, P.; Lopez, S.A. Automatic discovery of photoisomerization mechanisms with nanosecond machine learning photodynamics simulations. *Chem. Sci.* **2021**, *12*, 5302–5314. [CrossRef] [PubMed]

27. Kabylda, A.; Vassilev-Galindo, V.; Chmiela, S.; Poltavsky, I.; Tkatchenko, A. Efficient interatomic descriptors for accurate machine learning force fields of extended molecules. *Nat. Commun.* **2023**, *14*, 3562. [CrossRef]

28. Young, T.A.; Johnston-Wood, T.; Deringer, V.L.; Duarte, F. A transferable active-learning strategy for reactive molecular force fields. *Chem. Sci.* **2021**, *12*, 10944–10955. [CrossRef]

29. Cao, Y.; Romero, J.; Olson, J.P.; Degroote, M.; Johnson, P.D.; Kieferová, M.; Kivlichan, I.D.; Menke, T.; Peropadre, B.; Sawaya, N.P.D.; et al. Quantum Chemistry in the Age of Quantum Computing. *Chem. Rev.* **2019**, *119*, 10856–10915. [CrossRef] [PubMed]

30. Bauer, B.; Bravyi, S.; Motta, M.; Chan, G.K.L. Quantum Algorithms for Quantum Chemistry and Quantum Materials Science. *Chem. Rev.* **2020**, *120*, 12685–12717. [CrossRef]

31. Jones, R.R.; Bergman, R.G. p-Benzyne. Generation as an intermediate in a thermal isomerization reaction and trapping evidence for the 1,4-benzenediyl structure. *J. Am. Chem. Soc.* **1972**, *94*, 660–661. [CrossRef]

32. Ingold, C. *Structure and Mechanism in Organic Chemistry*; Cornell University Press: Ithaca, NY, USA, 1969.

33. Dong, H.; Chen, B.Z.; Huang, M.B.; Lindh, R. The bergman cyclizations of the enediyne and its N-substituted analogs using multiconfigurational second-order perturbation theory. *J. Comput. Chem.* **2012**, *33*, 537–549. [CrossRef]

34. Kerekes, Z.; Tasi, D.A.; Czakó, G. SN2 Reactions with an Ambident Nucleophile: A Benchmark Ab Initio Study of the CN– + CH$_3$Y [Y = F, Cl, Br, and I] Systems. *J. Phys. Chem. A* **2022**, *126*, 889–900. [CrossRef] [PubMed]

35. Jónsson, H.; Mills, G.; Jacobsen, K.W. Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific: Singapore, 1998; pp. 385–404. [CrossRef]

36. Perdew, J.P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868. [CrossRef] [PubMed]

37. Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305. [CrossRef] [PubMed]

38. Aprà, E.; Bylaska, E.J.; de Jong, W.A.; Govind, N.; Kowalski, K.; Straatsma, T.P.; Valiev, M.; van Dam, H.J.J.; Alexeev, Y.; Anchell, J.; et al. NWChem: Past, present, and future. *J. Chem. Phys.* **2020**, *152*, 184102. [CrossRef] [PubMed]

39. Roos, B.O.; Taylor, P.R.; Sigbahn, P.E. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **1980**, *48*, 157–173. [CrossRef]

40. Lindh, R.; Persson, B.J. Ab Initio Study of the Bergman Reaction: The Autoaromatization of Hex-3-ene-1,5-diyne. *J. Am. Chem. Soc.* **1994**, *116*, 4963–4969. [CrossRef]

41. Sun, Q.; Yang, J.; Chan, G.K.L. A general second order complete active space self-consistent-field solver for large-scale systems. *Chem. Phys. Lett.* **2017**, *683*, 291–299. [CrossRef]

42. Sun, Q.; Berkelbach, T.C.; Blunt, N.S.; Booth, G.H.; Guo, S.; Li, Z.; Liu, J.; McClain, J.D.; Sayfutyarova, E.R.; Sharma, S.; et al. PySCF: The Python-based simulations of chemistry framework. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1340. [CrossRef]

43. Purvis, G.D.; Bartlett, R.J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910–1918. [CrossRef]

44. Dunning, T.H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023. [CrossRef]

45. Woon, D.E.; Dunning, T.H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371. [CrossRef]

46. Wilson, A.K.; Woon, D.E.; Peterson, K.A.; Dunning, T.H. Gaussian basis sets for use in correlated molecular calculations. IX. The atoms gallium through krypton. *J. Chem. Phys.* **1999**, *110*, 7667–7676. [CrossRef]

47. 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*, e1603015. [CrossRef] [PubMed]

48. 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. [CrossRef]

49. Chmiela, S.; Sauceda, H.E.; Poltavsky, I.; Müller, K.R.; Tkatchenko, A. sGDML: Constructing accurate and data efficient molecular force fields using machine learning. *Comput. Phys. Commun.* **2019**, *240*, 38–45. [CrossRef]

50. Sauceda, H.E.; Chmiela, S.; Poltavsky, I.; Müller, K.R.; Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.* **2019**, *150*, 114102. [CrossRef]

51. Steinmetzer, J.; Kupfer, S.; Gräfe, S. pysisyphus: Exploring potential energy surfaces in ground and excited states. *Int. J. Quantum Chem.* **2021**, *121*, e26390. [CrossRef]

52. Imbalzano, G.; Anelli, A.; Giofré, D.; Klees, S.; Behler, J.; Ceriotti, M. Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials. *J. Chem. Phys.* **2018**, *148*, 241730. [CrossRef]

53. Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671. [CrossRef]

54. Atkins, P.; Paula, J. *Atkins' Physical Chemistry*; Oxford University Press: Oxford, UK, 2008.