

# Chemical Science

Volume 16  
Number 35  
21 September 2025  
Pages 15745–16358

rsc.li/chemical-science



# AIQM2

DFT  
Accuracy  
Speed

ISSN 2041-6539



## AIQM2: organic reaction simulations beyond DFT

Cite this: *Chem. Sci.*, 2025, 16, 15901Yuxinxin Chen <sup>ac</sup> and Pavlo O. Dral <sup>\*abc</sup>

All publication charges for this article have been paid for by the Royal Society of Chemistry

Density functional theory (DFT) is the workhorse of reaction simulations, but it often suffers from either prohibitive cost or insufficient accuracy. In this work, we report AIQM2—the universal AI-enhanced QM method 2—as the first method that enables fast and accurate large-scale organic reaction simulations for practically relevant system sizes and time scales beyond what is possible with DFT. This breakthrough is based on the high speed of AIQM2, which is orders of magnitude faster than common DFT, while its accuracy in reaction energies, transition state optimizations, and barrier heights is at least at the level of DFT and often approaches the gold-standard coupled cluster accuracy. AIQM2 can be used out of the box without any further retraining. Compared to pure machine learning potentials, AIQM2 possesses high transferability and robustness in simulations without catastrophic breakdowns. We showcase the superiority of AIQM2 compared to traditional DFT by performing an extensive reaction dynamics study overnight and revising the mechanism and product distribution reported in the previous investigation of the bifurcating pericyclic reaction.

Received 16th April 2025  
Accepted 7th August 2025DOI: 10.1039/d5sc02820g  
rsc.li/chemical-science

## Introduction

Reaction simulations provide atomistic-level information on the efficiency and mechanism of chemical transformations, guiding both experimental reaction design and *in silico* reaction planning. The common tools to simulate reactions are based on the concept of the potential energy surface (PES).<sup>1</sup> Critical static points along the reaction pathway on the PES, such as minima (reactants, products, intermediates) and first-order saddle points (transition states, TS), can be used to derive kinetics and thermochemical properties. The key challenge is that to ensure high simulation accuracy, the PES ideally must be evaluated with chemical accuracy (1 kcal mol<sup>-1</sup> error in energies), because the reaction rate depends exponentially on the Gibbs energy of activation according to the Eyring equation. Similarly, the product distribution under thermodynamic control strongly depends on the difference in the products' Gibbs free energies.

Minor changes to the shape of the PES, including the locations of critical points, can have large effects on reaction mechanisms and pathways,<sup>2</sup> e.g., synchronous and step-wise mechanisms in the widely explored Diels–Alder reactions<sup>3</sup> and ambimodal TSs in bifurcating reactions.<sup>4</sup> To gain deeper insight into the conformational changes and energy flow during chemical transformations, the reaction dynamics must be

investigated, where bonds breaking and formation are directly simulated by performing molecular dynamics (MD) trajectory propagations.<sup>5</sup> One example is post-TS MD, which provides branching ratios and energy partitioning for reactions with diverging product channels—cases in which traditional Eyring's TS theory fails.<sup>6</sup> While performing MD calculations is highly desirable for modeling reactions, an ensemble of reactive trajectories is usually required to obtain statistically robust and converged results, thus requiring computationally efficient methods for obtaining PESs.

Thus, the fast generation of well-shaped PES in the reactive region is required, which is a tall order for traditional quantum mechanical (QM) methods (Fig. 1a), because, generally, less expensive calculations are less accurate. For example, methods that achieve chemical accuracy on barrier heights—e.g., coupled-cluster methods and beyond,<sup>7</sup> and double hybrid functionals with large basis sets<sup>8</sup>—may not be feasible for large systems and TS structure optimizations. The only practical solution widely adopted by the community is the combination schemes. In these schemes, cheaper and less accurate QM methods, such as commonly employed density functional theory (DFT) with relatively small basis sets or even semi-empirical QM methods, are used to obtain the geometries and frequencies of TS and minima. More expensive methods, such as the gold-standard CCSD(T),<sup>9</sup> are then used to calculate the single-point energies on these geometries. These combination schemes are, obviously, approximations assuming that the geometry does not change much between the levels of theory, which is not necessarily true.<sup>10</sup>

The problem of high computational cost is particularly pertinent in reaction dynamics simulations, where one often must sacrifice the number of trajectories and, hence, the quality

<sup>a</sup>State Key Laboratory of Physical Chemistry of Solid Surfaces, Department of Chemistry, College of Chemistry and Chemical Engineering, Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Xiamen University, Xiamen 361005, China. E-mail: dral@xmu.edu.cn

<sup>b</sup>Institute of Physics, Faculty of Physics, Astronomy, and Informatics, Nicolaus Copernicus University in Toruń, ul. Grudziądzka 5, 87-100 Toruń, Poland

<sup>c</sup>Atomistic, Shenzhen 518000, China





Fig. 1 Selected types of methods used for simulating organic reactions. (a) Traditional QM methods exemplified by semi-empirical QM, density functional theory (DFT), the gold-standard CCSD(T), and universal (machine learning) interatomic potentials (UIPs)—each with its strengths and weaknesses. (b) AIQM2 proposed in this work eliminates the above weaknesses through careful design. (c) Various applications of AIQM2 in reaction simulations such as transition state (TS) optimization, reactive dynamics, and the accurate and fast calculation of reaction barriers.

of the statistical analysis. Indeed, most reported studies propagated too few trajectories, leading to statistically unreliable product ratios,<sup>11</sup> although, in these studies, accuracy has already been sacrificed by using relatively cheap DFT methods.

The rise of machine learning (ML) promises to speed up reaction simulations by directly predicting key properties, such as reaction outcome,<sup>12</sup> yields,<sup>13–16</sup> barrier heights,<sup>17–27</sup> rate constants,<sup>28,29</sup> as well as TS geometry.<sup>30–37</sup> However, ideally, we would like to use the ML models as surrogate models to describe PES at the QM level. This would allow us to obtain similar levels of insight and to use existing, mature techniques for reaction simulations and analysis. Such models are popular ML interatomic potentials (MLIPs), where either an accurate description of the system-specific PES is established or universal interatomic potentials (UIPs) are constructed in the hope of covering the target chemical space.<sup>38</sup>

The construction of the system-specific MLIPs requires considerable effort because it requires careful selection of the training points on the PES and resource-intensive calculations with high-level reference QM methods.<sup>39</sup> After all this work, these potentials cannot easily be reused for similar systems, and additional effort is needed to extend their applicability. A highly desirable alternative is UIPs, which promise to remove the need for repeated retraining while maintaining adequate accuracy for molecules outside the training set.

In principle, UIPs are pre-trained MLIP models, but constructing UIPs is much more challenging than training a system-specific MLIP because it requires an ML model with good transferability capabilities, large and representative datasets covering the chemical space of interest, as well as training

and validation strategies ensuring good generalizing ability of the model, *etc.* Another important consideration in UIPs is the quality of the reference data. Most state-of-the-art (SOTA) UIPs target the DFT level because most of the available data are at the DFT level with relatively small basis sets.<sup>40–46</sup> Hence, such potentials are bound not just to fail to achieve chemical accuracy in most cases but also to be inferior even to the reference DFT level they were trained on.

Only a few UIPs have gone beyond the DFT level and targeted the gold-standard coupled cluster level: AIQM1 (ref. 47) and ANI-1ccx<sup>48</sup> are the first and only successful examples. AIQM1 is generally more transferable and robust than ANI-1ccx because AIQM1 is based on a  $\Delta$ -learning<sup>49</sup> approach, where MLIP is used only to correct the difference between the target (coupled cluster) and baseline (semi-empirical) QM levels, while ANI-1ccx is a pure MLIP model directly trained on the target coupled-cluster level. One of the major disadvantages of AIQM1 is its limitation to CHNO elements, which can be mitigated by using an extrapolative ONIOM approach,<sup>50</sup> where AIQM1 is used for the major core region of the atomistic system, while lower-level methods such as ANI-2x<sup>51</sup> and/or GFN2-xTB<sup>52</sup> are used for the remaining parts containing other elements.

Both AIQM1 and ANI-1ccx generally achieve chemical accuracy within the scope of the training data, but we found that they have subpar performance in describing reaction barriers.<sup>53</sup> Hence, until now, no theoretical method has been reported that can be used for organic reaction simulations—including transition state search, barrier height calculations, and reactive dynamics—with the robustness and transferability of DFT approaches, while being orders of magnitude faster.



Here, we introduce AIQM2, the second-generation general-purpose AI-enhanced quantum mechanical method, which approaches gold-standard coupled-cluster level accuracy at the cost of semi-empirical methods—i.e., beyond those of typical DFT approaches (Fig. 1b). As we show in this work, AIQM2 consistently improves upon its predecessor AIQM1 and the related universal ANI-1ccx method, especially for transition state optimizations and barrier heights. It also outperforms hybrid and double-hybrid DFT methods with a common double- $\zeta$ -quality basis set in reaction energies of large systems and some non-covalent interactions. In contrast to DFT, AIQM2 provides uncertainty estimates for its predictions, increasing the trustworthiness of the simulations with this method (Fig. 1b and c).

To showcase the efficiency and robustness of AIQM2, we propagated thousands of high-quality trajectories for a bifurcating pericyclic reaction in parallel on 16 CPUs within one day and obtained the product distribution at nearly coupled-cluster level accuracy. This way we effectively revise the original, much more expensive and less accurate, DFT results. AIQM2 is publicly available in our open-source software MLatom<sup>54</sup> at <https://github.com/dralgroup/mlatom>.

To facilitate progress in ML-assisted computational chemistry, we have included the methods introduced in this work into the library of Universal and Updatable AI-enhanced QM foundational models (UAIQM).<sup>55</sup> It is an umbrella platform that collects various models in a coherent library, enabling the automatic choice of the model according to the user's needs and, if required, the improvement of the models. UAIQM hosts both published (e.g., AIQM1 and ANI-1ccx) and unpublished (e.g., AIQM2, when this work was submitted) models so that the latest developments can be made accessible to the public timely.

## Results and discussion

### AIQM2 design

AIQM2 follows the same principle of  $\Delta$ -learning<sup>49</sup> as in AIQM1 by applying neural network corrections to the baseline semi-empirical QM method for higher accuracy at a lower cost. We have chosen GFN2-xTB<sup>52</sup> as the baseline of AIQM2 due to its robustness and broad applicability in reaction exploration.<sup>56</sup> In this regard, the change of baseline makes AIQM1 and AIQM2 two distinct methods, both in terms of composition and application preference. An explicit dispersion correction is added, because neither semi-empirical QM nor local neural network models can properly describe the long-range non-covalent interactions on their own.

Specifically, AIQM2 consists of 3 parts (Fig. 1b): the semi-empirical baseline GFN2-xTB\* (asterisk indicates the special version of GFN2-xTB with D4 dispersion corrections<sup>57</sup> removed), the correction predicted by the ensemble of 8 ANI neural networks,<sup>58</sup> and the D4 dispersion correction (for the  $\omega$ B97X<sup>59</sup> functional). Thus, the energy predicted by AIQM2 can be formulated as:

$$E_{\text{AIQM2}} = E_{\text{GFN2-xTB}^*} + E_{\text{NN}} + E_{\text{D4}(\omega\text{B97X})}, \quad (1)$$

and energy derivative properties are the sum of each term's derivatives. Note that the Hessians provided by GFN2-xTB\* and D4 are calculated by numerical differentiation on analytical gradients, while neural networks provide analytical Hessians; the final Hessian is the sum of these three Hessians. Other electronic structure properties, such as partial charges and dipole moments, are inherited from the baseline GFN2-xTB\* calculations.

Several modifications were made to the original architecture used in ANI-1x and ANI-1ccx. The activation function was changed from CELU (Continuously Differentiable Exponential Linear Units) to GELU (Gaussian Error Linear Unit), which has been shown to provide better performance for higher-order energy derivatives.<sup>60</sup> The angular cutoff in the atomic environment vector (AEV) was increased from 3.5 Å to 4.0 Å for a better description of long-range interactions during training and inference.<sup>47</sup>

In terms of the training data, the ANI model corrections were trained on the data derived from the ANI-1ccx and ANI-1x datasets.<sup>61</sup> They were generated using an extrapolated coupled cluster method termed CCSD(T)\*/CBS<sup>61</sup> and a popular DFT method  $\omega$ B97X<sup>59</sup>/def2-TZVPP,<sup>62</sup> respectively. The asterisk in CCSD(T)\* refers to the estimated TightPNO accuracy of DLPNO-CCSD(T)<sup>63</sup> and the complete basis set extrapolation uses the formula by Halkier<sup>64</sup> and Helgaker.<sup>65</sup> Details about CCSD(T)\* / CBS can be found in ref. 61. GFN2-xTB\* data was generated for all the data points in both datasets and subtracted from the reference level to get the  $\Delta$ -values<sup>49</sup> for training. The models were first trained on the differences in energies and forces between  $\omega$ B97X<sup>59</sup>/def2-TZVPP<sup>62</sup> and GFN2-xTB\*. This led to the AIQM2@DFT\* approach approximating the  $\omega$ B97X/def2-TZVPP level. To approximate the  $\omega$ B97X-D4/def2-TZVPP level with explicit dispersion corrections, we can add the D4 corrections for the  $\omega$ B97X functional, yielding the AIQM2@DFT method. The final models in AIQM2 were obtained by transfer learning on the energy difference between CCSD(T)\* / CBS and GFN2-xTB\*.<sup>61</sup> The D4 corrections for the  $\omega$ B97X functional were removed from the reference CCSD(T)\* / CBS. Removing the D4 corrections during training was necessary because we added the explicit D4 corrections for the  $\omega$ B97X functional back to the model when making predictions. In transfer learning, the first and the third layers were fixed as in ANI-1ccx and AIQM1, which effectively reduced the number of parameters for training and prevented overfitting to a smaller high-level dataset.

An advantage of AIQM2 is also that it provides uncertainty estimates of its predictions, in contrast to DFT approaches. The uncertainty of the calculation is defined as the standard deviation of the ANI models in the ensemble as is done for both AIQM1 and ANI-1ccx.<sup>66</sup> We can interpret uncertainty as how different a new system, for which we want to make predictions, is from the molecules in the training set of AIQM2. The uncertainties are reported in this work as the error bars of the predictions, i.e., energy estimates from averaging ANI model predictions  $\pm$  standard deviation of the ANI models. The uncertainty for reaction energies is calculated as the standard deviation of the relative energies predicted by the ANI models. We calibrated the uncertainty on the standard CHNO dataset as



done in the literature<sup>66</sup> which yielded the threshold of  $0.36 \text{ kcal mol}^{-1}$ , *i.e.*, the prediction is deemed reliable if the uncertainty is less than  $0.36 \text{ kcal mol}^{-1}$ . From our previous tests on AIQM1 and ANI-1ccx, we have numerical evidence that highly uncertain predictions overall lead to bigger errors.<sup>66</sup>

### Revision of the bifurcating pericyclic reaction

Here we demonstrate the power of AIQM2 by revising the earlier state-of-the-art simulations<sup>67</sup> of a paradigmatic pericyclic reaction, by performing resource-intensive downhill molecular dynamics (MD) simulations (Fig. 2). These simulations involve thousands of quasi-classical trajectories from the region around the ambimodal TS, leading to more than one product.

AIQM2 enables extensive explorations: overnight (100 CPU-hours) on commodity hardware, we could propagate 1000 quasi-classical trajectories, each 500 fs long in both forward and backward directions with 0.1 fs time step (in total, 10 million single-point evaluations, see computational details). These simulations on a smaller scale (only 117 reactive trajectories with 1 fs time step) were performed earlier<sup>67</sup> with a much costlier DFT method B3LYP-D3/6-31G\*, which would take 160

CPU-years for a thousand trajectories with 0.1 fs time step. Even with such extensive computational resources spent on DFT calculations, the number of trajectories might still have been insufficient to draw definite conclusions because of the low precision and potential for missing rare events, as shown in a recent analysis.<sup>68</sup> AIQM2 simulations with 1000 trajectories, hence, furnish much higher statistical certainty.

In addition to propagating ten times more trajectories, we use the time step that is ten times smaller than in the original DFT study.<sup>67</sup> Our AIQM2 simulations with 0.1 fs lead to energy-conservative dynamics, while we show that the 1 fs time step, employed in the original study, leads to rather large violations in energy conservation (Fig. 3b). The total energy fluctuations in 1 fs time-step trajectories reach up to several  $\text{kcal mol}^{-1}$ , which might influence the conclusions drawn in the original DFT study. Indeed, our calculations with the same method, AIQM2, show that, *e.g.*, for an exemplary trajectory, product P2 is formed 9 fs later when the trajectory is propagated with 0.1 fs.

Most importantly, it is known<sup>69</sup> that the quality of the downhill dynamics is strongly influenced by the accuracy of the QM method. Here our AIQM2 approach also has a significant

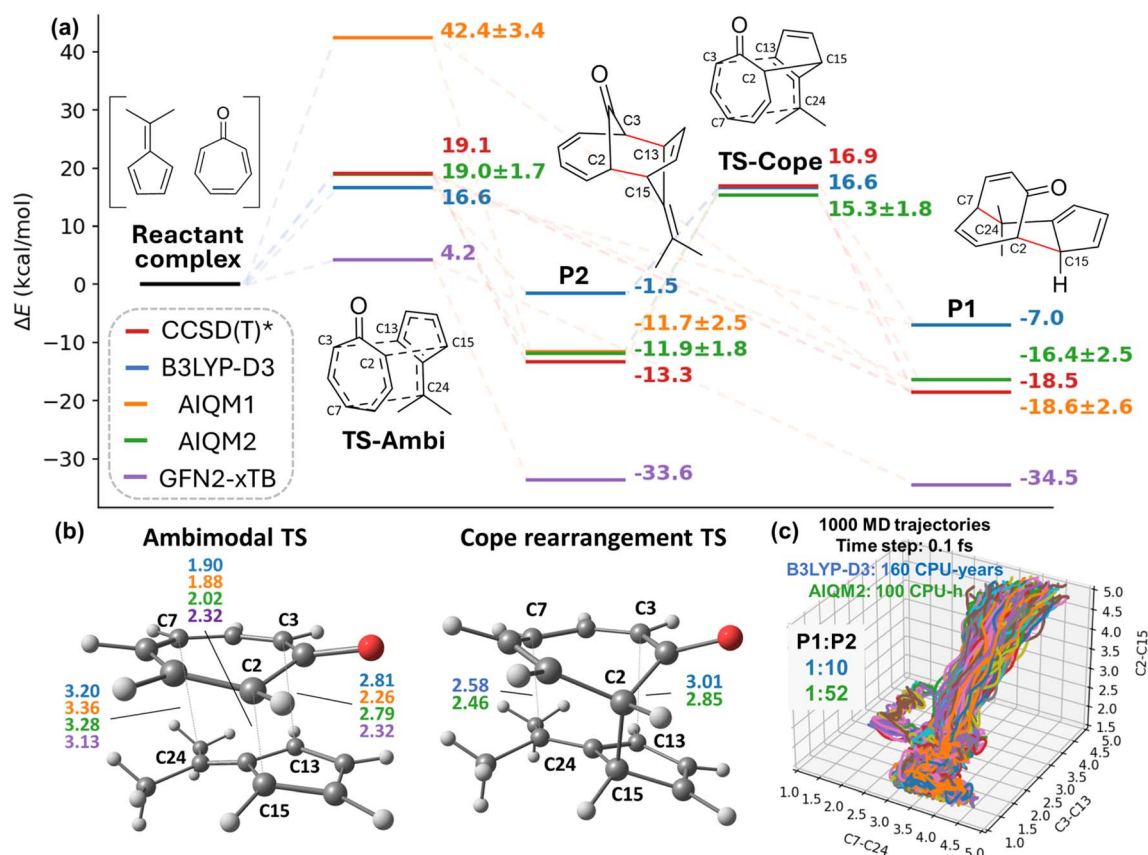
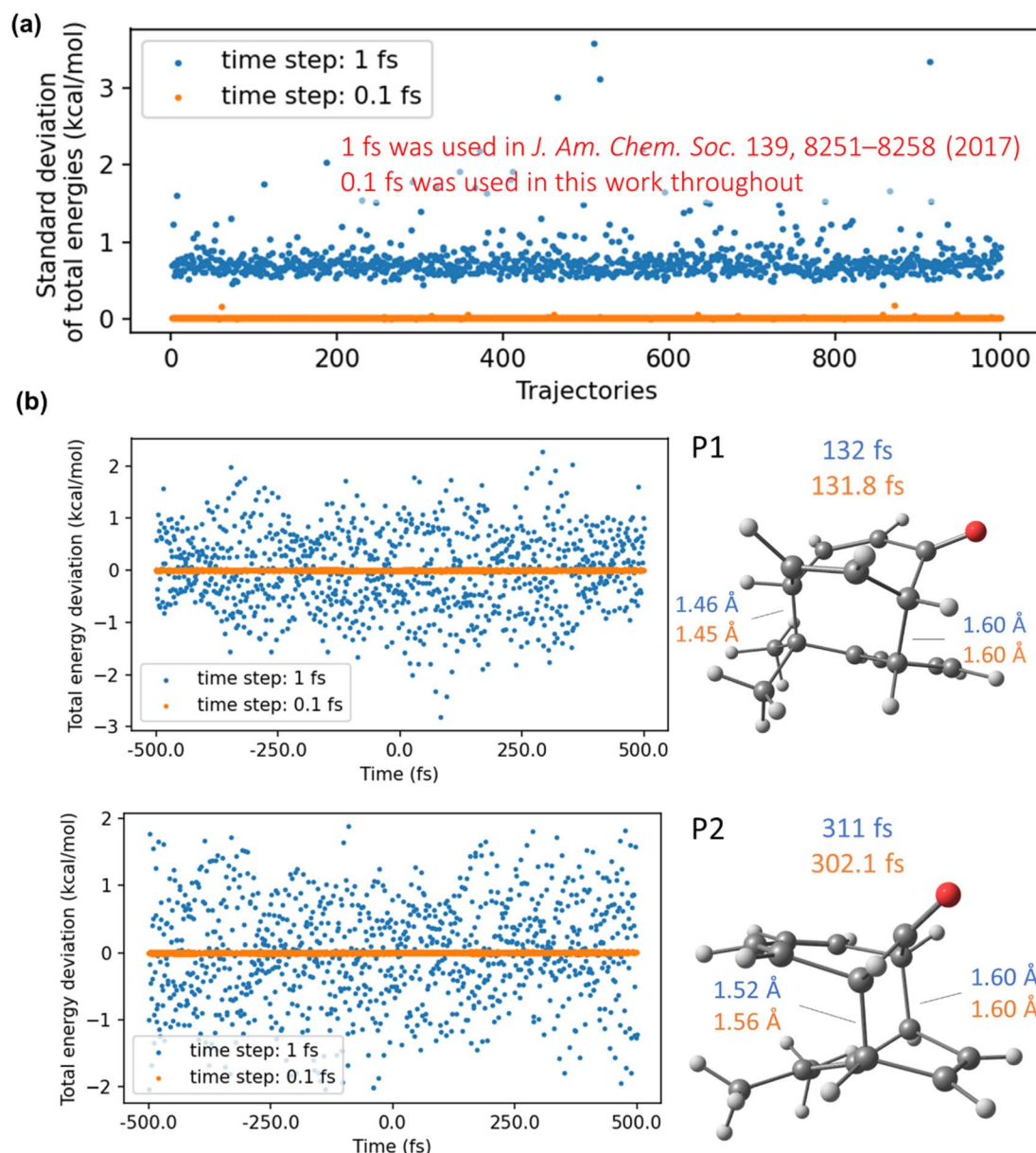


Fig. 2 Energy profile, transition state geometries, and quasi-classical trajectories of ambimodal reaction. (a) Comparison of barrier heights and reaction energies in  $\text{kcal mol}^{-1}$  obtained with different methods. Energies are reported on the geometries optimized with the corresponding methods, except for CCSD(T)\* / CBS energies evaluated on the B3LYP-D3/6-31G\*-optimized geometries. Values after  $\pm$  are the uncertainties of the relative energies given by AIQM2. (b) Critical bond lengths in angstroms (Å) of transition states optimized by various methods. The Cope rearrangement transition state is not found using GFN2-xTB and AIQM1, thus not presented. The same color scheme is used for each method as (a). (c) Distribution of 1000 quasi-classical trajectories starting from the ambimodal transition state (B3LYP product ratio is taken from ref. 67). B3LYP-D3 – B3LYP-D3/6-31G\*, CCSD(T)\* – CCSD(T)\* / CBS, TS-Ambi – ambimodal TS, TS-Cope – Cope rearrangement TS.





**Fig. 3** Total energies and their standard deviations along quasi-classical trajectories propagated with 1 fs and 0.1 fs using AIQM2. 1 fs was used in the original DFT study<sup>67</sup> and is given here only for comparison; all results in this study are obtained with 0.1 fs time-step trajectories. (a) Comparison of the standard deviation of total energies in kcal mol<sup>-1</sup> on 1000 trajectories propagated with 1 and 0.1 fs. (b) Comparison of total energy deviations with respect to the mean value on representative trajectories for P1 and P2. Figures on the right side show the snapshots where the products are first observed. Blue represents trajectories with the 1 fs time step, and orange represents those with the 0.1 fs time step. Time 0 fs corresponds to the initial geometry near the transition state.

advantage over B3LYP-D3/6-31G\* as judged by the analysis of the PES stationary points. The barrier height of ambimodal TS at AIQM2 is almost the same as CCSD(T)\* / CBS value. Note that the CCSD(T)\* / CBS values are derived from DFT optimized geometries to avoid potential bias; we cannot compare to the geometries at the CCSD(T)\* level due to its cost and absence of analytical energy derivatives. TS geometry optimized with AIQM2 qualitatively resembles that at B3LYP-D3/6-31G\* with some noticeable differences in bond lengths. In comparison, B3LYP-D3/6-31G\* underestimates the barrier by around

3 kcal mol<sup>-1</sup>. For the minima, the error between AIQM2 and reference CCSD(T)\* / CBS is within *ca.* 2 kcal mol<sup>-1</sup>, while both B3LYP-D3 (*ca.* 10 kcal mol<sup>-1</sup>) and GFN2-xTB have huge errors (more than 25 kcal mol<sup>-1</sup>) compared to the same reference. Such large errors in B3LYP-D3 used in the previous study<sup>67</sup> certainly influence the quality of the downhill dynamics.

Having at our disposal (1) a more accurate method, (2) more trajectories, and (3) better energy conservation, we turn to analyzing the simulation results. AIQM2 results differ from the DFT results in two major aspects. First, the product distribution



changed from 1 : 10 to 1 : 52, showing that the formation of P1 has a much smaller role than previously thought. Second, the AIQM2 results show that the reaction proceeds *via* a dynamically stepwise mechanism rather than a concerted mechanism, as previously thought. The average time gap between the formation of the first and second bond is 205 fs for the major P2 product. The DFT study reported a 60 fs time gap. In AIQM2 trajectories leading to P2, the C3–C13 bond formation is delayed, and most of the C3–C13 bond lengths oscillate between 2.5 Å and 3 Å. This might potentially be caused by the PES curvature shaped by the “neighboring” TS corresponding to the Cope rearrangement connecting the competing products P2 and P1 (Fig. 2a and b). Once the first C2–C15 bond is formed, the formed adduct is hot and cannot immediately ‘decide’ which bond to form; eventually, the shorter C3–C13 bond is formed most of the time because it is typically shorter in the initial conditions. The role of the nearby TS is supported by the observation that the Cope rearrangement’s barrier height at B3LYP-D3/6-31G\* is much smaller than at AIQM2, which might explain the difference in product ratio. We did not observe the transformation of P2 to P1 in AIQM2 trajectories, as the barrier is rather high (27.24 kcal mol<sup>-1</sup>) compared to the small barrier at DFT (10.23 kcal mol<sup>-1</sup>).

Summarizing the analysis of the quasi-classical trajectories, we revised the previous state-of-the-art.<sup>65</sup> While we leave more detailed explorations to future research, our dynamics study showcased the power of the AIQM2 approach, which should be considered instead of the common DFT methods when performing such simulations.

Several additional remarks are due. The B3LYP-D3/6-31G\* calculations might have suffered from substantial basis-set superposition error (BSSE), which might have been rather large in TSs. An advantage of AIQM2 is that it was trained on the CCSD(T)\*/CBS-level data, which were generated using a complete basis set (CBS) extrapolation scheme. By construction, CBS does not suffer from BSSE. AIQM2 can, therefore, provide a more accurate description in cases where BSSE becomes an issue for common approaches with the medium-sized basis sets like in B3LYP-D3/6-31G\*.

Obviously, among the zoo of DFT functionals, there is always a high chance of finding a better functional than B3LYP/6-31G\*. Particularly, M06-2X is now commonly used in the pericyclic reaction simulations because benchmarks showed its better accuracy for such simulations. For this particular reaction, M06-

2X/6-31G\* is more accurate than B3LYP-D3/6-31G\* according to our comparison with CCSD(T)\*/CBS (Table 1) and achieves similar accuracy to AIQM2. It does not help that M06-2X is even slower than B3LYP-D3, which would strongly hamper the extensive evaluation of many trajectories with small time steps. AIQM2 also provides error bars, allowing users to at least get an idea of how far the predictions might be from the target CCSD(T)/CBS values, while none of the DFT approaches offer such a handy feature.

Finally, our approach has another huge advantage over common composite schemes, such as when semi-empirical methods are used for pre-screening possible transition states and calculating their ensembles, followed by refinement with the DFT methods. In our example of the pericyclic reaction, AIQM2 could find both the ambimodal and Cope rearrangement TSs, while the popular semi-empirical GFN2-xTB failed to locate the latter. In addition, GFN2-xTB has huge errors (more than 25 kcal mol<sup>-1</sup>) for the minima too. It means that the composite schemes relying on the semi-empirical methods are likely missing key conformers and TSs. Our approach is a big step forward in replacing the need for composite approaches by providing an all-in-one solution for all steps of the workflow from geometry optimization and ensemble sampling, to dynamics and ultimately reliable energetics.

### Overall AIQM2 performance

It is imperative to judge the quality of universal models by evaluating their performance on independent tests that are closer to the real-world applications. We impose a strict requirement on our new method to be applicable for performing molecular simulations with higher speed and with accuracy at least comparable to, or better than, DFT. Hence, we evaluate our new method on the GMTKN55 benchmark<sup>8</sup> – a standard test for DFT methods. This benchmark covers various sizes of molecules and types of relative energies, which provides a comprehensive evaluation of AIQM2 beyond reaction properties. The reference values in the GMTKN55 benchmark are taken from high-level QM calculations which are expected to have errors within 1 kcal mol<sup>-1</sup> relative to experiment. Hence, our evaluation will show the errors that might be expected when compared to the experiment.

We compare the performance of AIQM2 to that of the DFT methods commonly used in practice, *i.e.*, B3LYP-D4/6-31G\* and ωB97X-D4/def2-TZVPP; we also include a double-hybrid

**Table 1** Comparison of various methods on energy profiles of ambimodal reaction and the estimated CPU time required to propagate 1000 trajectories with the 0.1 fs time step

	CCSD(T)*/CBS// B3LYP-D3/6-31G*	AIQM2// AIQM2	B3LYP-D3/6-31G*// B3LYP-D3/6-31G*	M06-2X/6-31G*// B3LYP-D3/6-31G*	M06-2X/6-31G*// M06-2X/6-31G*
Barrier (TS-Ambi)	19.1	19.0 ± 1.7	16.6	17.8	17.6
Reaction energy (P1)	-18.5	-16.4 ± 2.5	-7.0	-20.3	-19.7
Barrier (TS-Cope)	16.9	15.3 ± 1.8	16.6	15.5	15.4
Reaction energy (P2)	-13.3	-11.9 ± 1.8	-1.5	-14.2	-13.6
Time cost estimate for 1000 trajectories	—	100 CPU-hours	160 CPU-years	—	200 CPU-years



functional representative DSD-BLYP-D4/6-31G\*. As is shown in Fig. 4, AIQM2 gives overall much better results than B3LYP-D4/6-31G\* and DSD-BLYP-D4/6-31G\* in terms of WTMAD-2, which is a weighted metric used to balance the different scales of various reaction energies. Note that the double-hybrid functional DSD-BLYP was recommended in the original GMTKN55 benchmark based on the evaluation with a much more expensive def2-QZVP, which is rarely used in practice. We instead choose the 6-31G\* basis set for this functional from a practical perspective as one of the most commonly used in DFT studies, which is not intended for rigorous evaluation of the functional itself. Compared with the hybrid functional  $\omega$ B97X with the more expensive triple- $\zeta$  basis set, AIQM2 can achieve

comparable accuracy while maintaining the cost of a semi-empirical method. Overall, AIQM2 offers an outstanding cost-accuracy trade-off for predicting thermochemical, kinetic, and noncovalent properties: it is significantly faster than DFT approaches while maintaining competitive accuracy.

Considering the performance in each category, AIQM2 exhibits the lowest WTMAD-2 value of 1.57 kcal mol<sup>-1</sup> among all tested methods for isomerization reaction energies of large systems. This is partly due to the good extensibility of the ANI neural network<sup>70</sup> and the rich conformers presented in the training data, as exemplified in the performance of ANI-1ccx on the ISOL24 dataset. However, in cases where strong effects come from atoms outside the cutoff, such as the conjugated system,

subsets	N	ANI-1ccx	GFN2-xTB	AIQM1	AIQM2	DSD-BLYP-D4/ 6-31G*	B3LYP-D4/ 6-31G*	$\omega$ B97X-D4/ def2-TZVPP
<b>Reaction energies for small systems</b>								
G2RC	12	24.59	29.52	3.29	7.83	9.56	10.66	5.25
FH51	44	6.59	11.21	0.90	2.53	3.97	5.75	2.48
TAUT15	13	1.56	0.99	0.37	0.41	1.99	1.59	0.86
DC13	10	17.98	26.61	40.92	5.44	4.58	5.99	6.70
WTMAD-2		7.79	10.19	4.34	2.55	5.00	5.50	2.94
<b>Reaction energies for large systems and isomerisation reactions</b>								
DARC	13	1.20	17.44	0.64	0.33	6.01	2.08	1.18
BSR36	36	2.68	2.76	1.06	0.29	0.74	3.14	4.84
CDIE20	20	2.09	1.80	0.92	0.95	1.30	1.95	0.67
ISO34	34	1.24	6.90	0.44	0.78	2.35	2.89	1.09
ISOL24	16	2.83	14.59	3.08	1.93	2.77	3.65	3.24
C60ISO	9	91.96	5.80	47.89	3.18	6.78	2.55	15.17
WTMAD-2		5.41	8.64	2.62	1.57	3.32	4.83	3.74
<b>Reaction barrier heights</b>								
BHPERI	22	8.89	9.49	10.90	2.53	4.07	4.02	3.61
BHDIV10	5	13.15	10.83	6.90	6.97	4.21	4.22	1.18
INV24	13	10.81	2.63	3.39	1.06	1.27	1.68	1.08
BHROT27	19	1.27	0.91	0.47	0.43	0.66	0.75	0.27
PX13	8	7.97	3.65	6.88	5.99	2.59	8.34	0.57
WCPT18	12	6.01	3.83	4.30	2.85	4.17	6.49	2.11
WTMAD-2		7.27	5.25	5.45	2.24	2.84	3.53	1.85
<b>Intermolecular noncovalent interactions</b>								
ADIM6	6	0.64	1.15	0.88	1.31	0.12	0.47	0.63
S22	22	5.00	0.76	0.87	1.12	2.07	2.58	0.59
S66	66	2.52	0.73	0.62	0.56	1.63	2.10	0.71
WATER27	15	1.17	1.50	1.91	0.72	46.70	55.36	17.83
WTMAD-2		9.99	2.86	2.56	2.69	7.36	9.34	3.11
<b>Intramolecular noncovalent interactions</b>								
IDISP	6	12.39	6.78	3.31	3.10	2.52	1.82	2.91
ACONF	15	0.62	0.19	0.26	0.24	0.10	0.19	0.10
Amino20x4	72	0.91	0.98	0.49	0.50	0.61	0.71	0.20
PCONF21	18	3.18	1.76	0.79	0.61	1.83	1.88	0.25
MCONF	51	0.92	1.72	0.36	0.27	1.12	1.37	0.24
SCONF	17	1.58	1.64	0.93	0.67	3.41	3.66	0.16
BUT14DIOI	64	0.61	1.25	0.46	0.34	2.02	2.44	0.36
WTMAD-2		9.64	9.97	4.22	3.58	10.53	12.15	2.07
<b>Total</b>								
WTMAD-2		8.33	7.93	3.78	2.73	6.89	8.31	2.66
Time		~0.005 h	~0.02 h	~0.02 h	~0.03 h	~250 h	~150 h	~2500 h

Fig. 4 Performance assessment of AIQM2 compared to universal ML models ANI-1ccx and AIQM1 targeting coupled cluster accuracy and a selection of QM methods on the GMTKN55 data. Results are shown for neutral, closed-shell CHNO-containing compounds, as AIQM2 was designed to target coupled cluster accuracy on such types of molecules. Numbers are mean absolute errors in kcal mol<sup>-1</sup>, color-coded for each subset so that green corresponds to the smallest and red – to the largest errors (green – the best). WTMAD-2 is the weighted mean absolute deviation-2 in kcal mol<sup>-1</sup> as defined in ref. 8. Time is reported in CPU hours. Notations and definitions of the reaction and interaction energies follow the original GMTKN55 benchmark paper;<sup>8</sup> note that our benchmarks for DFT methods differ from the original ones, because we performed them with commonly used, smaller basis sets.



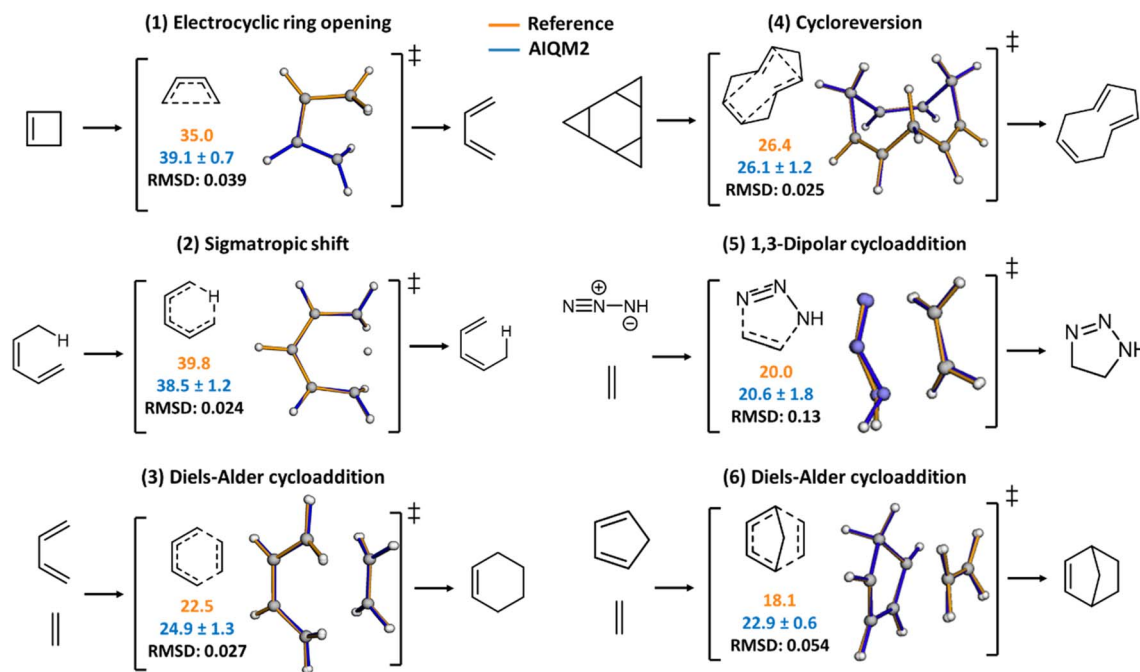


Fig. 5 Representative reactions in BHPERI and their transition state structures. For the first three reactions, the reference barrier heights are approximated with the composite scheme to approach the CBS limit on the CCSD(T)/cc-pVTZ optimized geometries.<sup>71</sup> The latter 3 reactions adopted reference values from the BHPERI dataset of the GMTKN55 2017 version, *i.e.*, barriers are calculated with W1-F12 and W2-F12 (ref. 72) methods on geometries from the CBS-QB3 (ref. 73) protocol. Orange is used for reference barrier heights, and blue for AIQM2. Barrier heights of AIQM2 are calculated using reference geometries to be consistent with the GMTKN55 benchmark. RMSD in Ångstrom shows the structural difference between the reference and the AIQM2-optimized TS. The same color scheme is used when overlapping their structures.

the ANI neural network may not work properly. Thus, baseline GFN2-xTB plays another important role in the robustness and accuracy of AIQM2. For example, in the C60ISO dataset, GFN2-xTB provides a good starting point for AIQM2 to approach mean absolute error (MAE) as low as 3.18 kcal mol<sup>-1</sup>, considering the average reaction energies in C60ISO are *ca.* 100 kcal mol<sup>-1</sup>. In the generally problematic non-covalent interactions, AIQM2 can even improve on some difficult cases for DFT, such as those in the WATER27 dataset, due to the extensive representation of water clusters in the training data, as known for the ANI-1ccx<sup>48</sup> and AIQM1 (ref. 47) methods.

### Performance of AIQM2 for transition states and barrier heights

Another significant improvement of AIQM2 over related methods lies in its prediction of barrier heights. This presents a big challenge for AIQM1 and ANI-1ccx, as shown in the previous study,<sup>53</sup> although they all target the CCSD(T)/CBS level. Compared to AIQM1, the WTMAD-2 of AIQM2 decreased from 5.45 to 2.24 kcal mol<sup>-1</sup> for the datasets benchmarking the barrier heights (Fig. 4).

The most prominent change is in the BHPERI dataset, where the overall MAE is the lowest among the tested methods. Thus, we dive into the details of BHPERI and check whether AIQM2 can produce accurate geometries for TSs. Representative reactions in BHPERI and the comparison of barrier heights and TS geometries are shown in Fig. 5, which also demonstrates that

the uncertainty of AIQM2 predictions is lower than that of the less accurate AIQM1.<sup>53</sup> All TS structures were successfully found by AIQM2 and exhibited one imaginary frequency, with the normal mode vibration corresponding to the correct reaction.

Unfortunately, we cannot evaluate the quality of the TS geometry using the BHPERI dataset because it does not contain accurate reference geometries (which were optimized at B3LYP/6-31G\* level). As a non-representative comparison, we report only the root-mean-square deviations (RMSDs) of three AIQM2-optimized TS structures relative to the CCSD(T)/cc-pVTZ geometries reported in a more recent work<sup>71</sup> (Fig. 5): the Diels-Alder reaction starting from *s-cis*-butadiene and ethylene, the electrocyclic ring-opening reaction starting from cyclobutene, and the sigmatropic rearrangement originating from the *E*-isomer of 1,3-pentadiene. This recent work also reported higher-level energetics achieved with hierarchical focal point analyses (FPA), extrapolating to the *ab initio* limit. AIQM2 agrees well with the reference TS structures, with the RMSDs less than 0.1 Å.

## Conclusions and outlook

In this work, we report the second-generation general-purpose AI-enhanced Quantum Mechanical Method (AIQM2). AIQM2 offers remarkable speed and accuracy for reaction simulations by providing robust results for various applications out of the box. The overall accuracy of AIQM2 is rivaling the established DFT approaches, but the simulations are orders of magnitude



faster. As we show, this can be exploited for efficient TS optimizations and extensive dynamics investigations of reaction mechanisms.

We exploited AIQM2's excellent performance to lift the accuracy in the quasi-classical downhill molecular dynamics, providing more statistically significant results with trajectories featuring energy conservation. This enabled us to obtain a new chemical insight into a bifurcating pericyclic reaction previously studied with fewer, non-energy-conserving trajectories obtained with a much slower and less accurate DFT approach. We show that this reaction proceeds *via* a stepwise mechanism rather than a concerted one, as previously believed. We also revise the product distribution in this ambimodal reaction to show that the minor product plays a much smaller role, which we attribute to the proximity of another TS corresponding to the Cope rearrangement, not described properly by B3LYP.

AIQM2 has an overall improved accuracy compared to its predecessor AIQM1, particularly for TSs and barrier heights, which are important for the quantum chemical reaction simulations. AIQM2 already has a growing track record of successful applications in real-world simulations beyond organic reactions, due to its universality, high speed, and accuracy. This method has been shown to provide outstanding IR spectra for a broad range of chemical compounds.<sup>74</sup> Recently, it has also been used in the investigation of large systems, where AIQM2 enabled fast and accurate optimizations (*e.g.*, within 90 minutes on 32 GPUs for a 714-atom noncovalent complex), leading to high-quality UV/vis spectra calculated on these optimized geometries.<sup>75</sup>

All the power of AIQM2 is publicly available in MLatom at <https://github.com/dralgroup/mlatom> under the permissive MIT license, which lifts any limitations on the reuse of the method. The method is also included in the Aitomic package<sup>76</sup> and can be used to perform online calculations on the Aitomistic Hub at <https://aitomistic.xyz> and XACS platform at <https://XACScloud.com>, further enhancing its availability.

## Computational details

All calculations were performed with the MLatom program.<sup>54</sup> The D4 dispersion corrections were calculated *via* MLatom's interface to the dftd4 program,<sup>77,78</sup> the ANI contributions *via* the interface to TorchANI,<sup>58</sup> and the GFN2-xTB(\*) calculations were performed *via* the interface to the locally modified xtb program.<sup>79</sup> AIQM1 calculations require the ODM2\* contributions which were calculated with the MNDO program<sup>80</sup> interfaced to MLatom. MLatom was used for the B3LYP (VWN5) and  $\omega$ B97X calculations *via* the interface to PySCF<sup>81–83</sup> and DSD-BLYP *via* the interface to Orca program.<sup>84,85</sup> Additional CCSD(T)\*/CBS calculations were performed following the methodology in ref. 61 using the open-source MLatom implementation,<sup>54</sup> where the required energy components were calculated *via* the interface to the Orca program.<sup>84,85</sup>

In the simulations of the bifurcating pericyclic reaction, we followed the simulation setup and analysis protocol analogous to those in the original study,<sup>67</sup> *i.e.*, initial conditions for quasi-

classical trajectories were sampled from the harmonic quantum Boltzmann distribution starting at the ambimodal transition state, as implemented<sup>86</sup> in MLatom.<sup>54</sup> The trajectories were started near the TS region and essentially led to “downhill” dynamics to either products or reactants; forward and backward trajectories were launched from the same initial conditions, differing only in the sign of the initial velocities. Each set of such forward and backward trajectories, starting with the same initial conditions, yields a combined trajectory, where the time steps in the backward trajectories are reversed. Reactive trajectories were defined as those combined trajectories that contain reactants and one of the products; nonreactive combined trajectories contain either only reactants or only one of the products.

## Author contributions

P. O. D.: conceptualization, method design, original draft, figures. Y. C.: method development and implementations, calculations, analysis, manuscript revision, figures.

## Conflicts of interest

The authors declare no competing interests.

## Data availability

AIQM2 is publicly available in the MLatom at <https://github.com/dralgroup/mlatom>. Data with energies and optimized coordinates for the performed benchmarks and ambimodal reaction analysis, as well as datasets used for training AIQM2, are available at <https://github.com/707-Moira/aiqm2>.

## Acknowledgements

P. O. D. acknowledges funding from the National Natural Science Foundation of China (funding *via* the Outstanding Youth Scholars (Overseas, 2021) project) and *via* the Lab project of the State Key Laboratory of Physical Chemistry of Solid Surfaces. We also acknowledge Aitomistic for financial support and hosting the AIQM2 methods online on the Aitomistic Hub at <https://aitomistic.xyz> making it available for a wider usership.

## References

- 1 M. Meuwly, Machine Learning for Chemical Reactions, *Chem. Rev.*, 2021, **121**, 10218–10239.
- 2 H. B. Schlegel, Geometry optimization, *Wiley Interdiscip. Rev.:Comput. Mol. Sci.*, 2011, **1**, 790–809.
- 3 M. Linder and T. Brinck, On the method-dependence of transition state asynchronicity in Diels-Alder reactions, *Phys. Chem. Chem. Phys.*, 2013, **15**, 5108–5114.
- 4 A. Sengupta and K. N. Houk, Origins of Periselectivity and Regioselectivity in Ambimodal Tripericyclic [8+6]/[6+4]/[4+2] Intramolecular Cycloadditions of a Heptafulvenyl-Fulvene, *J. Phys. Chem. A*, 2023, **127**, 7976–7983.



- 5 S. Pratihar, X. Ma, Z. Homayoon, G. L. Barnes and W. L. Hase, Direct Chemical Dynamics Simulations, *J. Am. Chem. Soc.*, 2017, **139**, 3570–3590.
- 6 A. Agarwal, P. J. Boruah, B. Sarkar and A. K. Paul, Post-Transition-State Direct Dynamics Simulations on the Ozonolysis of Catechol, *J. Phys. Chem. A*, 2022, **126**, 5314–5327.
- 7 J. Zheng, Y. Zhao and D. G. Truhlar, The DBH24/08 Database and Its Use to Assess Electronic Structure Model Chemistries for Chemical Reaction Barrier Heights, *J. Chem. Theory Comput.*, 2009, **5**, 808–821.
- 8 L. Goerigk, *et al.*, A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32184–32215.
- 9 K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, A fifth-order perturbation comparison of electron correlation theories, *Chem. Phys. Lett.*, 1989, **157**, 479–483.
- 10 L. Simon and J. M. Goodman, How reliable are DFT transition structures? Comparison of GGA, hybrid-meta-GGA and meta-GGA functionals, *Org. Biomol. Chem.*, 2011, **9**, 689–700.
- 11 S. P. Tyukina, J. A. Velmiskina, A. O. Dmitrienko and M. G. Medvedev, Binomial Uncertainty in Molecular Dynamics-Based Reactions Analysis, *J. Phys. Chem. Lett.*, 2024, **15**, 2105–2110.
- 12 V. Venkatasubramanian and V. Mann, Artificial intelligence in reaction prediction and chemical synthesis, *Curr. Opin. Chem. Eng.*, 2022, **36**, 100749.
- 13 D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher and A. G. Doyle, Predicting reaction performance in C-N cross-coupling using machine learning, *Science*, 2018, **360**, 186–190.
- 14 A. L. Haywood, *et al.*, Kernel Methods for Predicting Yields of Chemical Reactions, *J. Chem. Inf. Model.*, 2022, **62**, 2077–2092.
- 15 F. Sandfort, F. Strieth-Kalthoff, M. Kühnemund, C. Beecks and F. Glorius, A Structure-Based Platform for Predicting Chemical Reactivity, *Chem*, 2020, **6**, 1379–1390.
- 16 P. Schwaller, A. C. Vaucher, T. Laino and J.-L. Reymond, Prediction of chemical reaction yields using deep learning, *Mach. Learn.: Sci. Technol.*, 2021, **2**, 015016.
- 17 H.-C. Chang, M.-H. Tsai and Y.-P. Li, Enhancing Activation Energy Predictions under Data Constraints Using Graph Neural Networks, *J. Chem. Inf. Model.*, 2025, **65**, 1367–1377.
- 18 S. Choi, Y. Kim, J. W. Kim, Z. Kim and W. Y. Kim, Feasibility of Activation Energy Prediction of Gas-Phase Reactions by Machine Learning, *Chem. Eur J.*, 2018, **24**, 12354–12358.
- 19 X. Garcia-Andrade, P. Garcia Tahoces, J. Perez-Rios and E. Martinez Nunez, Barrier Height Prediction by Machine Learning Correction of Semiempirical Calculations, *J. Phys. Chem. A*, 2023, **127**, 2274–2283.
- 20 C. A. Grambow, L. Pattanaik and W. H. Green, Deep Learning of Activation Energies, *J. Phys. Chem. Lett.*, 2020, **11**, 2992–2997.
- 21 E. Heid and W. H. Green, Machine Learning of Reaction Properties via Learned Representations of the Condensed Graph of Reaction, *J. Chem. Inf. Model.*, 2022, **62**, 2101–2110.
- 22 S. Heinen, G. F. von Rudorff and O. A. von Lilienfeld, Toward the design of chemical reactions: Machine learning barriers of competing mechanisms in reactant space, *J. Chem. Phys.*, 2021, **155**, 064105.
- 23 I. Ismail, C. Robertson and S. Habershon, Successes and challenges in using machine-learned activation energies in kinetic simulations, *J. Chem. Phys.*, 2022, **157**, 014109.
- 24 T. I. Madzhidov, *et al.*, Structure–reactivity relationship in Diels–Alder reactions obtained using the condensed reaction graph approach, *J. Struct. Chem.*, 2017, **58**, 650–656.
- 25 K. A. Spiekermann, L. Pattanaik and W. H. Green, Fast Predictions of Reaction Barrier Heights: Toward Coupled-Cluster Accuracy, *J. Phys. Chem. A*, 2022, **126**, 3976–3986.
- 26 T. Stuyver and C. W. Coley, Quantum chemistry-augmented neural networks for reactivity prediction: Performance, generalizability, and explainability, *J. Chem. Phys.*, 2022, **156**, 084104.
- 27 Q. Zhao, D. M. Anstine, O. Isayev and B. M. Savoie, Delta(2) machine learning for reaction property prediction, *Chem. Sci.*, 2023, **14**, 13392–13401.
- 28 E. Komp and S. Valteau, Machine Learning Quantum Reaction Rate Constants, *J. Phys. Chem. A*, 2020, **124**, 8607–8613.
- 29 P. L. Houston, A. Nandi and J. M. Bowman, A Machine Learning Approach for Prediction of Rate Constants, *J. Phys. Chem. Lett.*, 2019, **10**, 5250–5258.
- 30 R. Barrett and J. Westermayr, Reinforcement Learning for Traversing Chemical Structure Space: Optimizing Transition States and Minimum Energy Paths of Molecules, *J. Phys. Chem. Lett.*, 2024, **15**, 349–356.
- 31 C. Duan, Y. Du, H. Jia and H. J. Kulik, Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model, *Nat. Comput. Sci.*, 2023, **3**, 1045–1055.
- 32 R. Jackson, W. Zhang and J. Pearson, TSNet: predicting transition state structures with tensor field networks and transfer learning, *Chem. Sci.*, 2021, **12**, 10022–10040.
- 33 S. Kim, J. Woo and W. Y. Kim, Diffusion-based generative AI for exploring transition states from 2D molecular graphs, *Nat. Commun.*, 2024, **15**, 341.
- 34 M. Z. Makos, N. Verma, E. C. Larson, M. Freindorf and E. Kraka, Generative adversarial networks for transition state geometry prediction, *J. Chem. Phys.*, 2021, **155**, 024116.
- 35 L. Pattanaik, J. B. Ingraham, C. A. Grambow and W. H. Green, Generating transition states of isomerization reactions with deep learning, *Phys. Chem. Chem. Phys.*, 2020, **22**, 23618–23626.
- 36 E. C. Yuan, *et al.*, Analytical ab initio hessian from a deep learning potential for transition state optimization, *Nat. Commun.*, 2024, **15**, 8865.
- 37 J. Zhang, *et al.*, Deep reinforcement learning of transition states, *Phys. Chem. Chem. Phys.*, 2021, **23**, 6888–6895.



- 38 Y. Yang, S. Zhang, K. D. Ranasinghe, O. Isayev and A. E. Roitberg, Machine Learning of Reactive Potentials, *Annu. Rev. Phys. Chem.*, 2024, **75**, 371–395.
- 39 G. D. Fabritiis, Machine Learning Potentials: A Roadmap Toward Next-Generation Biomolecular Simulations, *arXiv*, 2024, preprint, 2408.12625 [physics.chem-ph], DOI: [10.48550/arXiv.2408.12625](https://doi.org/10.48550/arXiv.2408.12625).
- 40 D. M. Anstine, R. Zubatyuk and O. Isayev, AIMNet2: a neural network potential to meet your neutral, charged, organic, and elemental-organic needs, *Chem. Sci.*, 2025, **16**, 10228–10244.
- 41 C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, *Nat. Comput. Sci.*, 2022, **2**, 718–728.
- 42 K. Choudhary, *et al.*, Unified graph neural network force-field for the periodic table: solid state applications, *Digit. Discov.*, 2023, **2**, 346–355.
- 43 B. Deng, *et al.*, CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling, *Nat. Mach. Intell.*, 2023, **5**, 1031–1041.
- 44 M. Nováček and J. Řezáč, PM6-ML: The synergy of semiempirical quantum chemistry and machine learning transformed into a practical computational method, *J. Chem. Theory Comput.*, 2025, **21**, 678–690.
- 45 S. Takamoto, *et al.*, Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements, *Nat. Commun.*, 2022, **13**, 2991.
- 46 F. Xie, T. Lu, S. Meng and M. Liu, GPTFF: A high-accuracy out-of-the-box universal AI force field for arbitrary inorganic materials, *Sci. Bull.*, 2024, **69**, 3525–3532.
- 47 P. Zheng, R. Zubatyuk, W. Wu, O. Isayev and P. O. Dral, Artificial Intelligence-Enhanced Quantum Chemical Method with Broad Applicability, *Nat. Commun.*, 2021, **12**, 7022.
- 48 J. S. Smith, *et al.*, Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning, *Nat. Commun.*, 2019, **10**, 2903.
- 49 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. von Lilienfeld, Big Data Meets Quantum Chemistry Approximations: The  $\Delta$ -Machine Learning Approach, *J. Chem. Theory Comput.*, 2015, **11**, 2087–2096.
- 50 Z. Yan, D. Wei, X. Li and L. W. Chung, Accelerating reliable multiscale quantum refinement of protein-drug systems enabled by machine learning, *Nat. Commun.*, 2024, **15**, 4181.
- 51 C. Devereux, *et al.*, Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens, *J. Chem. Theory Comput.*, 2020, **16**, 4192–4202.
- 52 C. Bannwarth, S. Ehlert and S. Grimme, 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.
- 53 Y. Chen, *et al.*, Benchmark of general-purpose machine learning-based quantum mechanical method AIQM1 on reaction barrier heights, *J. Chem. Phys.*, 2023, **158**, 074103.
- 54 P. O. Dral, *et al.*, MLatom 3: A Platform for Machine Learning-Enhanced Computational Chemistry Simulations and Workflows, *J. Chem. Theory Comput.*, 2024, **20**, 1193–1213.
- 55 Y. Chen, Y.-F. Hou, O. Isayev & P. O. Dral, Universal and Updatable Artificial Intelligence-Enhanced Quantum Chemical Foundational Models, *ChemRxiv*, 2024, preprint, DOI: [10.26434/chemrxiv-2024-604wb](https://doi.org/10.26434/chemrxiv-2024-604wb).
- 56 S. Grimme, Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations, *J. Chem. Theory Comput.*, 2019, **15**, 2847–2862.
- 57 E. Caldeweyher, C. Bannwarth and S. Grimme, Extension of the D3 dispersion coefficient model, *J. Chem. Phys.*, 2017, **147**, 034112.
- 58 X. Gao, F. Ramezanghorbani, O. Isayev, J. S. Smith and A. E. Roitberg, TorchANI: A Free and Open Source PyTorch-Based Deep Learning Implementation of the ANI Neural Network Potentials, *J. Chem. Inf. Model.*, 2020, **60**, 3408–3415.
- 59 J.-D. Chai and M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
- 60 S. F. Alavi, *et al.*, ANI-1ccx-gelu Universal Interatomic Potential and Its Fine-Tuning: Toward Accurate and Efficient Anharmonic Vibrational Frequencies, *J. Phys. Chem. Lett.*, 2025, **16**, 483–493.
- 61 J. S. Smith, *et al.*, The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules, *Sci. Data*, 2020, **7**, 134.
- 62 F. Weigend and R. Ahlrichs, 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.
- 63 Y. Guo, *et al.*, Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)], *J. Chem. Phys.*, 2018, **148**, 011101.
- 64 A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper and J. Olsen, Basis-set convergence of the energy in molecular Hartree-Fock calculations, *Chem. Phys. Lett.*, 1999, **302**, 437–446.
- 65 T. Helgaker, W. Klopper, H. Koch and J. Noga, Basis-set convergence of correlated calculations on water, *J. Chem. Phys.*, 1997, **106**, 9639–9646.
- 66 P. Zheng, W. Yang, W. Wu, O. Isayev and P. O. Dral, Toward Chemical Accuracy in Predicting Enthalpies of Formation with General-Purpose Data-Driven Methods, *J. Phys. Chem. Lett.*, 2022, **13**, 3479–3491.
- 67 P. Yu, *et al.*, Mechanisms and Origins of Periselectivity of the Ambimodal [6 + 4] Cycloadditions of Tropone to Dimethylfulvene, *J. Am. Chem. Soc.*, 2017, **139**, 8251–8258.
- 68 Y. F. Hou, Q. Zhang and P. O. Dral, Surprising Dynamics Phenomena in the Diels-Alder Reaction of C(60) Uncovered with AI, *J. Org. Chem.*, 2024, **89**, 15041–15047.



- 69 T. A. Young, T. Johnston-Wood, H. Zhang and F. Duarte, Reaction dynamics of Diels-Alder reactions from machine learned potentials, *Phys. Chem. Chem. Phys.*, 2022, **24**, 20820–20827.
- 70 J. S. Smith, O. Isayev and A. E. Roitberg, ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost, *Chem. Sci.*, 2017, **8**, 3192–3203.
- 71 P. Vermeeren, *et al.*, Pericyclic reaction benchmarks: hierarchical computations targeting CCSDT(Q)/CBS and analysis of DFT performance, *Phys. Chem. Chem. Phys.*, 2022, **24**, 18028–18042.
- 72 A. Karton and J. M. Martin, Explicitly correlated Wn theory: W1-F12 and W2-F12, *J. Chem. Phys.*, 2012, **136**, 124114.
- 73 J. A. Montgomery, M. J. Frisch, J. W. Ochterski and G. A. Petersson, A complete basis set model chemistry. VII. Use of the minimum population localization method, *J. Chem. Phys.*, 2000, **112**, 6532–6542.
- 74 Y.-F. Hou, C. Wang and P. O. Dral, Accurate and Affordable Simulation of Molecular Infrared Spectra with AIQM Models, *J. Phys. Chem. A*, 2025, **129**, 3613–3623.
- 75 J.-C. Song, *et al.*, Alternating Donor–Acceptor Thienoacenes Featuring Up to 23 Linearly Fused Rings, *Org. Lett.*, 2025, **27**, 3753–3759.
- 76 P. O. Dral, Y. Chen and J. Li, *Aitomic: A package for AI-enhanced atomistic simulations*, Aitomic, Shenzhen, China, <http://MLatom.com/aitomic> (accessed on August 9, 2025), 20252.
- 77 E. Caldeweyher, J.-M. Mewes, S. Ehlert and S. Grimme, Extension and evaluation of the D4 London-dispersion model for periodic systems, *Phys. Chem. Chem. Phys.*, 2020, **22**, 8499–8512.
- 78 E. Caldeweyher, S. Ehlert & S. Grimme, *DFT-D4*, Version 3.5.0 (Mulliken Center for Theoretical Chemistry), University of Bonn, 2024.
- 79 Semiempirical extended tight-binding program package xtb, <https://github.com/grimme-lab/xtb> (accessed on August 9, 2025).
- 80 W. Thiel, with contributions from M. Beck, S. Billeter, R. Kevorkiants, M. Kolb, A. Koslowski, S. Patchkovskii, A. Turner, E.-U. Wallenborn, W. Weber, L. Spörkel, and P. O. Dral, MNDO2020: a semiempirical quantum chemistry program v. development version (Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr), <https://mndo.kfo.mpg.de> (accessed on August 9, 2025), 2020.
- 81 Q. Sun, Libcint: An efficient general integral library for Gaussian basis functions, *J. Comput. Chem.*, 2015, **36**, 1664–1671.
- 82 Q. Sun, *et al.*, PySCF: the Python-based simulations of chemistry framework, *Wiley Interdiscip. Rev.:Comput. Mol. Sci.*, 2017, **8**, e1340.
- 83 Q. Sun, *et al.*, Recent developments in the PySCF program package, *J. Chem. Phys.*, 2020, **153**, 024109.
- 84 F. Neese, Software update: the ORCA program system, version 4.0, *Wiley Interdiscip. Rev.:Comput. Mol. Sci.*, 2018, **8**, e1327.
- 85 F. Neese, The ORCA program system, *Wiley Interdiscip. Rev.:Comput. Mol. Sci.*, 2012, **2**, 73–78.
- 86 Y. F. Hou, L. Zhang, Q. Zhang, F. Ge and P. O. Dral, Physics-informed active learning for accelerating quantum chemical simulations, *J. Chem. Theory Comput.*, 2024, **20**, 7744–7754.

