Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

"

Reaction Dynamics of Diels-Alder Reactions from Machine Learned Potentials

Tom A. Young,[a] Tristan Johnston-Wood,[a], Hanwen Zhang[a] and Fernada Duarte*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

The abstract should be a single paragraph which summarises the content of the article.

Simulating chemical reactions is essential to developing fundamental understanding and predicting experimental outcomes.[1] Machine learned potentials (MLPs) offer an enticing approach to chemical simulation, enabling the efficient mapping between nuclear configurations and energies ($R \mapsto E$). Moreover, they offer flexibility and systematic improvability, in contrast to classical molecular mechanics (MM).[2] Propagating quantum dynamics using these forces should afford experimental rate and equilibrium constants in the limit of correct forces and converged sampling. However, despite the development of Gaussian Approximation Potentials (GAPs)[3,4] and high dimensional neural network potentials (NNPs)[5] more than 10 years ago, they are still not yet routinely used to simulate chemical reactivity.[6] Most likely, this is due to the computational and time investment required to train potentials for new systems.

Training an MLP consists of: (1) developing a training set; (2) hyperparameter optimisation and (3) performing the regression, repeating the process until the desired accuracy is obtained. Automated approaches to training set construction have been developed,[7–9] but can be limited to small systems or generate huge datasets. These limitations coupled with the time required to perform hyperparameter optimisation (if the MLP is insufficiently accurate) inhibits quickly accessing bespoke MLPs. Furthermore, the required $\gg 10^3$ reference evaluations precludes using accurate wavefunction-based quantum methods to evaluate energy and forces without considerable investment.[10] Exceptions are rare and limited to systems with $\lesssim 10$ atoms.[8,11]

For potentials suitable to simulate chemical reactivity, automated approaches are essential. The energy scale over which the potential must be accurate is larger, necessitating exponentially more training data and thus bespoke MLPs. Furthermore, the complex electronic structure around transition states makes density functional theory (DFT) a poorer reference method,[12] meaning coupled-cluster (CC) is often the target surface for quantitative comparison to experiment, which in-turn demands data-efficient strategies.

Here, we show that new MLP regression methods[13,14] can be used to generate accurate potentials for modestly sized reactions (~ 50 atoms) in an automated fashion and highlight some resultant insights.

With a view to extend our initial GAP training methodology[8] into more complex systems and environments, we considered Diels-Alder (DA) reactions because of the available theoretical and experimental data,[15,16] and their prominence in chemical and biochemical contexts.[17–19] Initial efforts proved promising,[20] with qualitatively reasonable reaction dynamics from [4+2] cycloaddition TSs for reactions between ethene + butadiene (**R1**) and methyl-vinyl ketone + cyclopentadiene (**R2**). Evaluating the quality of these potentials, however, revealed that they were not within the few k_BT accuracy limit required for rate estimation or dynamic studies (see e.g., Fig. S1a). A similarly complex but less exothermic reaction ($H_3C\cdot + C_3H_8 \rightarrow CH_4 + \cdot CH(CH_3)_2$) could be trained using the same strategy and hyperparameters (Fig. S1b), suggesting that achieving 1 kcal mol^{-1} accuracy within a 60 kcal mol^{-1} energy window required for **R1** is challenging for a GAP. Hyperparameter optimisation afforded an improvement, but at moderate computational cost (~ 500 configurations required for **R1**). Specifically, increasing the 'strength' of the fit by reducing the noise added to energies and forces, increasing the quality of the radial basis, and doubling the number of atomic environments considered in the training all improved the GAP (SI §S4). Systematic investigation of the effect of system size on the required number of reference evaluations suggests an approximate exponential scaling for a desired accuracy on the total energy (SI §S3). Adopting new regression methods within the same training strategy (Fig. 1) shows that GAPs – even with hyperparameter tuning – are outperformed by both linear atomic cluster expansion (ACE[21]) and equivariant graph neural

[a] *Chemical Research Laboratory, South Parks Road, Oxford, OX1 1NQ.*

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.

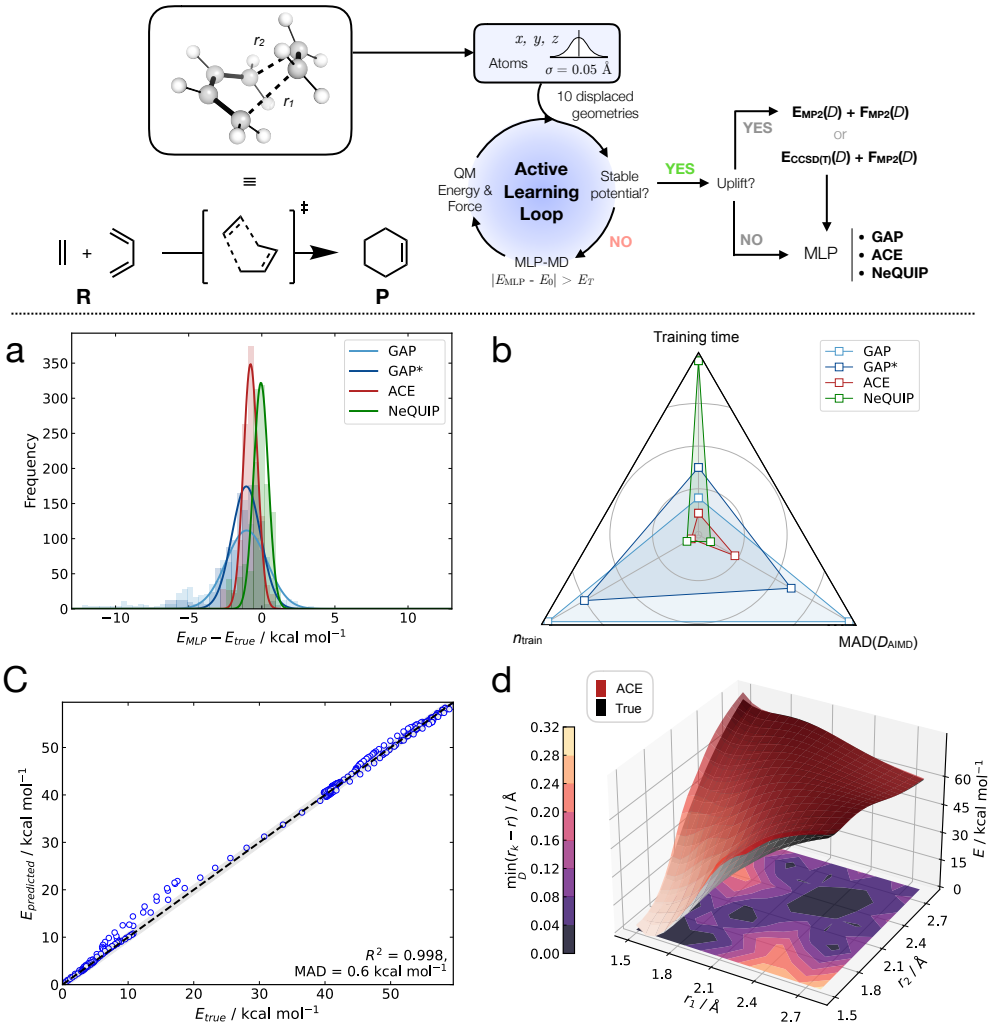

Fig. 1 MLP methods trained on the [4+2] Diels-Alder cycloaddition between ethene and butadiene in the gas phase. GAP* used optimised hyperparameters (see SI). Training set developed by active learning based on MLP-dynamics with $E_T = 2.3$ kcal mol^{-1} (0.1 eV), stable potential defined by the ability to propagate 10 trajectories without finding a configuration $|E_{\text{MLP}} - E_0| > E_T$. (a) Signed errors on total energies over three trajectories from reactants to products (AIMD, 300K, PBE0/def2-SVP). (b) Comparison of the relative performance between MLP methods on total time, data efficiency (number of training configuration selected, n_{train}) and total accuracy. (c) Parity plot between MP2/def2-TZVP total energies and ACE predictions on MP2 AIMD trajectories from the TS. ACE trained on DFT selected configurations; energies and forces re-evaluated at MP2. (d) Comparison of the predicted (red) and true PBE0/def2-SVP (black) relaxed 2D surface surrounding the TS. Contour plot represents the 'closeness' of the training data to a point in the surface.

networks (NequIP[13]). Attempts to improve GAPs by training a component-wise potential separated over covalent bonds were unsuccessful (see SI §S6.3).

While rather different in philosophy, both frameworks provide MLPs that are similarly accurate for **R1** (Fig. 1a, Fig. S24).§ Here, accuracy is based on deviations between true and predicted energies over independent DFT-MD trajectories propagated from the transition state (TS) to the reactant and product states. Previously, we have shown that a prospective validation strategy in the configuration space accessible to that MLP is essential to characterising 'good' MLPs.[8] However, here these potentials are stable by construction, within their own configuration space over the course of the reaction.

This arises from the active learning strategy (Fig. 1, top) defining a stable potential where 10, 1 ps, trajectories can be propagated without encountering a configuration that is predicted to be > 2 kcal mol^{-1} (0.1 eV) above or below the true energy. Of course, checking this criterion every MD step is too computationally intensive, thus stability is not guaranteed but empirically the criterion is sufficient to define stability within that period.

The data requirement to train a quality MLP for **R1** is reduced upon GAP hyperparameter optimisation – even though the potential is more accurate – but is surpassed by the efficient ACE and NequIP potentials, both of which require only ~ 100 training configurations (n_{train}, Fig. 1b). The total training time is maximal for the NequIP potential but is only $\sim 1/2$ day (10 cores + 1 GPU) meaning it is suitable for rapidly developing bespoke MLPs.

§ MLP training performed with *ml-train*[22] using *QUIP*,[23] *ACE*,[24] *ASE*,[25] *autodE*,[26] *nequip*[27] packages and ORCA[28] for QM calculations. See SI for full methods.

Note that the discrepancy between the MLPs in training time reduces with the system size, with reference energy and force evaluations dominating the computational time (for equally data efficient MLPs). The GAPs and ACE potential required just 5 ± 2 h of total training time on 10 CPU cores. The following sections will focus on ACE potentials for their slight advantage in computational cost.

As found for GAPs, re-evaluating energies and forces from AL configurations with a new reference method (aka. 'uplifting') reduces the computational cost associated with training WF-quality MLPs. For example, uplifting PBE0/DZ configurations to MP2/TZ affords an ACE potential within chemical accuracy to MP2-MD generated configurations (Fig. 1c).

Comparing the two-dimensional potential energy surface over the two forming C–C bonds, where all other degrees of freedom are free to relax, reveals that the ACE potential is smooth (as are the GAP and NequIP potentials, Fig. S26) and accurate even in the extrapolation regime (Fig. 1d). Even where the closest configuration in the training data is 0.3 Å away in the forming bonds, the error is only a couple of kcal mol^{-1} when the AL is initiated at the TS at each iteration ($r_1 = r_2 = 2.30$ Å).

Tangent to our goal of developing accurate MLPs for DA reactions, we found that GAP regularisation could be harnessed to reduce the computational cost of developing CCSD(T)-quality potentials (SI §S4). For simple molecules, MP2 forces are accurate to within ~ 0.05 eV Å^{-1} of their CCSD(T) counterparts (Fig. S12), thus within the 'expected error' (e.g. 0.1 eV Å^{-1}) of the GAP. This removes the requirement for numerical CCSD(T) gradients. Further accelerations are possible by CUR[29] selecting around half of the configurations from the dataset based on average SOAP kernel matrices. These effects can combine to afford a 100-fold reduction in the number of required CC calculations.

Extending the reaction complexity, the ambimodal reaction between tropone and cycloheptatriene (**R3**) is capable of forming three distinct products from a single TS (Fig. 2, top).[30] Training an ACE potential from the TS (**R3$_{TS1}$**) generated ~ 450 configurations using standard AL with sampling in the reactant and product regions of 2 of the 3 products. However, despite propagating

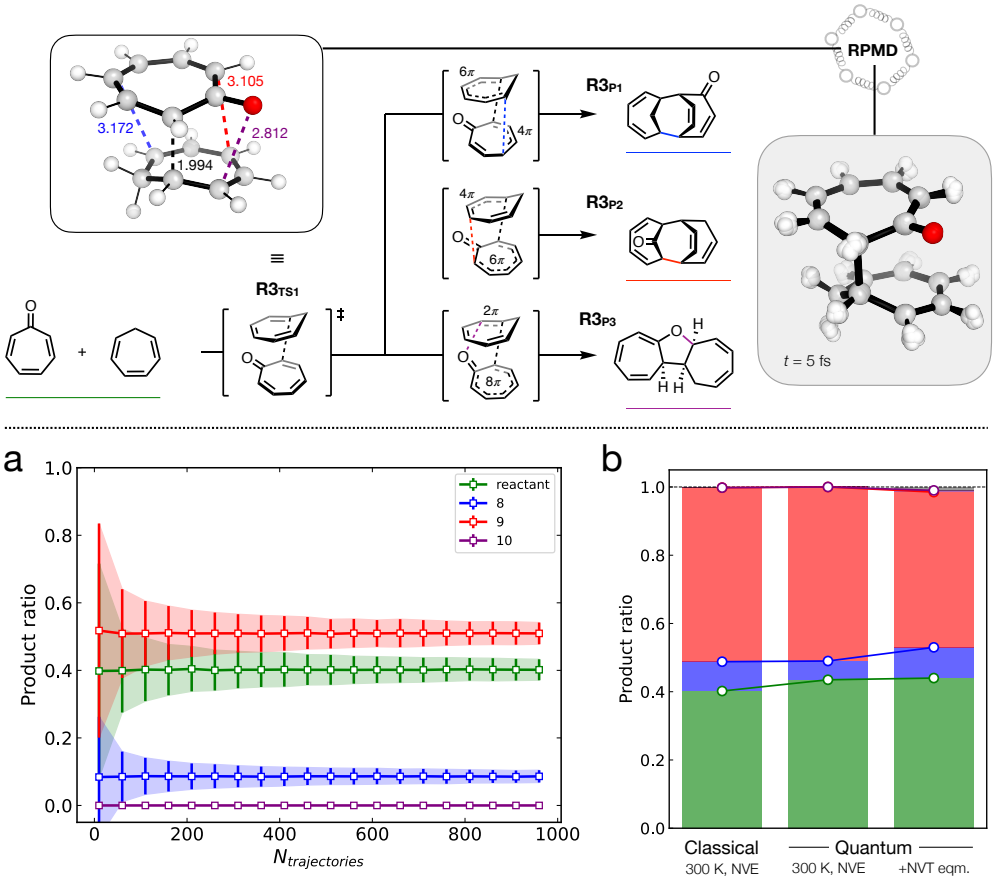

Fig. 2 Product distributions for the reaction between tropone and cycloheptatriene by ACE molecular dynamics simulations initiated from a single TS (top right). ACE potential trained using a standard active learning loop (diff, E_T = 2.3 kcal mol^{-1} = 0.1 eV, 500 K dynamics) from two TS, such that the training data included all three products. (a) Convergence of the product ratio (e.g. $N_{\rightarrow R3_{P1}}/N_{\text{total}}$) with number of trajectories propagated. Error is quoted as 2σ, obtained from bootstrapping with resampling (1000 iterations) on the same dataset. Dynamics propagated classically in the NVE ensemble from the TS using initial Boltzmann-sampled velocities for 300 K. (b) Dependence of the product ratio on the type of propagated dynamics ($N_{\text{total}} = 200$), including classical MD [as (a)] and ring polymer molecular dynamics (RPMD). RPMD simulations initiated from the TS, or first equilibrated using constrained NVT dynamics (80 fs, harmonic: 4 distances (top left), $k = 1$ eV Å^{-1}) then NVE. All NVE simulations performed for 1 ps using a 0.5 fs timestep. Accuracy of the ACE potential trained on PBE0/def2-SVP energies and forces is shown in Fig. S31.

ACE-MD at 500 K, the Cope product (**R3$_{P3}$**) was not present in the training data, making it unsuitable for running dynamics to elucidate the product ratio. Only when the TS that leads directly to **R3$_{P3}$** is included is the training data sufficiently complete. This highlights the importance of considering relevant points close in energy (4.4 kcal mol^{-1}, ref. 30) when training MLPs, that may not be obtained using automated sampling methods. For **R3** the ACE required ~ 2 days of AL and – in contrast to simple systems – we found is was not possible use a 'distance' criteria to select configurations as no **R3$_{P1}$** was formed in training (see §SS9).

Employing this MLP to propagate dynamics enables unique observations compared to the most common DFT-MD approach. Specifically, because each 1 ps trajectory takes only a couple of minutes to calculate, the product distribution can be converged with respect to the number of trajectories (Fig. 2a). Using only 100 trajectories affords an error (2σ, 95% confidence) of 10% for product **R3$_{P2}$**, which may or may not be sufficient for experimental comparison.

The inference efficiency also enables quantum dynamics to be propagated, which otherwise be too computationally demanding. Interestingly, we find that initiating ring-polymer molecular dynamics (RPMD)[31] without equilibrating the ring polymer leads to very similar product ratios to the classical dynamics, despite the zero-point energy being included as the ring polymer 'swells'. Equilibrating the system at the TS in the NVT ensemble with weak restraints (1 eV Å^{-1}, ~ 0.5 kcal mol^{-1} additional ZPE) prior to NVE downhill dynamics affords almost double the proportion of **R3$_{P1}$** and the formation of 10 as 1 % of the total product distribution. Training an MLP potential uniquely allows for the 100-fold more expensive simulations to be performed and the effect of ZPE (without leakage[32]) and tunnelling accounted for. Note that the equilibrated quantum dynamics are required to for correct dynamics[33] thus the right most distribution in Fig. 2b should be considered the benchmark.

With MLP uplifts to different levels of theory affordable, the effect of methodology on product distributions can be evaluated (100s CPUh vs. $> 200,000$ CPUh for the data in Fig. S30). Interestingly, we find that some levels require a few iterations of AL from their TSs if the configuration space overlap with the training level is poor. For example, B3LYP distributions are unchanged on additional AL, while M06-2X affords $\sim 10\%$ more product **R3$_{P2}$** upon a further five AL iterations. Product distributions vary considerably between functionals, with the major state varying from the reactant to **R3$_{P2}$** and the proportion of **R3$_{P1}$** varying from 1 to 10% (see Fig. S30). Downhill AIMD-derived product distributions should therefore consider the PES influence on the results more so than the type of dynamics (classical vs. quantum).

Obtaining free energies from absolute estimates in the rigid-

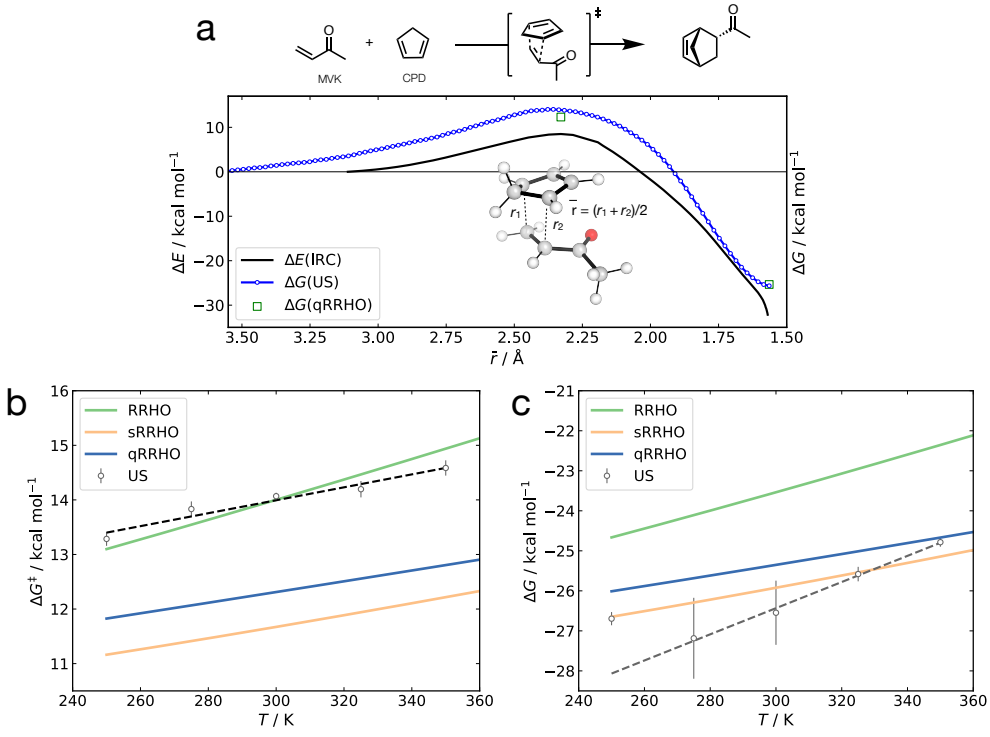

Fig. 3 Reaction and transition state free energies for the reaction between methyl vinyl ketone (MVK) and cyclopentadiene (CPD). (a) Comparison of the intrinsic reaction coordinate energy potential, quasi-rigid rotor harmonic oscillator (qRRHO) and umbrella sampling (US) free energies (300 K). Umbrella sampling performed using a ACE potential trained at the PBE0-D3BJ/def2-SVP level of theory at 500 K from the TS, then 30 windows simulated at the required temperature over the reaction coordinate (average of the forming bonds, $\bar{r}$) for 10 ps. See Fig. S32 for umbrella histograms. End point values from optimised structures at the DFT level. (b) Free energy barrier as a function of temperature using different static endpoint and intermediate path methods. sRRHO corresponds to a shifted RRHO treatment of the vibrational entropy where modes < 100 cm^{-1} are shifted to 100 cm^{-1}. Vibrational frequencies calculated at the DFT level from minima. Free energies from classical US are averaged over 5 repeats and the error the standard error of the mean. (c) Reaction free energy change. Simulations as (b).

rotor harmonic oscillator (RRHO) method is deficient in the low frequency regime.[34] To investigate the effect of different static approaches (RRHO, sRRHO,[35] qRRHO[36]) we trained an ACE potential for **R2** and compared free energy differences to classical umbrella sampling (US, Fig. 3). Reactive simulations totalling 1.5 ns would be unobtainable with DFT-MD. Note the sampling in the product state at 250 K is insufficiently complete, thus is excluded from the linear fit. The anharmonic effects at the reactant complex are significant, leading to a spread of 2 kcal mol^{-1} within the static methods, while the US correctly captures those effects but fails to treat any nuclear quantum effects. Therefore, for DA reactions any calculated free energies should include a few kcal mol^{-1} error bar.

Finally, we

The conclusions

Author Contributions

X

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

TJW acknowledges the EPSRC Centre for Doctoral Theory and Modelling in Chemical Sciences (EP/L015722/1) and TAY the EPSRC impact acceleration account (IAA) grant (EP/R511742/1).

Notes and references

1 A. J. Orr-Ewing, *Chemical Society Reviews*, 2017, **46**, 7597–7614.
2 J. Behler, *The Journal of Chemical Physics*, 2016, **145**, 170901.
3 A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, *Physical Review Letters*, 2010, **104**, 136403.
4 V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, *Chemical Reviews*, 2021, **121**, 10073–10141.
5 J. Behler and M. Parrinello, *Physical Review Letters*, 2007, **98**, 146401.
6 H.-Y. Ko, J. Jia, B. Santra, X. Wu, R. Car and R. A. DiStasio Jr., *Journal of Chemical Theory and Computation*, 2020, **16**, 3757–3785.
7 J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev and A. E. Roitberg, *The Journal of Chemical Physics*, 2018, **148**, 241733.
8 T. A. Young, T. Johnston-Wood, V. L. Deringer and F. Duarte, *Chemical Science*, 2021, **12**, 10944–10955.
9 A. M. Miksch, T. Morawietz, J. Kästner, A. Urban and N. Artrith, *Machine Learning: Science and Technology*, 2021, **2**, 031001.
10 J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev and A. E. Roitberg, *Nature Communications*, 2019, **10**, 2903.
11 P. O. Dral, A. Owens, A. Dral and G. Csányi, *The Journal of Chemical Physics*, 2020, **152**, 204110.
12 Y. Zhao, N. González-García and D. G. Truhlar, *The Journal of Physical Chemistry A*, 2005, **109**, 2012–2018.
13 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, *ArXiv*, 2021, 2101.03164.
14 D. P. Kovács, C. van der Oord, J. Kucera, A. E. A. Allen, D. J. Cole, C. Ortner and G. Csányi, *Journal of Chemical Theory and Computation*, 2021, acs.jctc.1c00647.
15 K. Black, P. Liu, L. Xu, C. Doubleday and K. N. Houk, *Proceedings of the National Academy of Sciences*, 2012, **109**, 12860–12865.
16 W. J. Lording, T. Fallon, M. S. Sherburn and M. N. Paddon-Row, *Chemical Science*, 2020, **11**, 11915–11926.
17 M. Sato, S. Kishimoto, M. Yokoyama, C. S. Jamieson, K. Narita, N. Maeda, K. Hara, H. Hashimoto, Y. Tsunematsu, K. N. Houk, Y. Tang and K. Watanabe, *Nature Catalysis*, 2021, **4**, 223–232.
18 V. Martí-Centelles, A. L. Lawrence and P. J. Lusby, *Journal of the American Chemical Society*, 2018, **140**, 2862–2868.
19 B. Briou, B. Améduri and B. Boutevin, *Chemical Society Reviews*, 2021, **50**, 11055–11097.
20 T. A. Young and T. Johnston-Wood, *gap-train*, 2021, `https://github.com/t-young31/gap-train`.
21 R. Drautz, *Physical Review B*, 2019, **99**, 014104.
22 T. A. Young and T. Johnston-Wood, *ml-train*, 2021, `https://github.com/t-young31/ml-train`.
23 G. Csányi, N. Bernstein and J. R. Kermode, *libAtoms/QUIP*, 2021, `https://github.com/libAtoms/QUIP`.
24 C. Ortner, L. Zhang, A. Ross, M. Sachs and C. van der Oord, *ACE.jl*, `https://github.com/ACEsuit/ACE.jl`.
25 A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *Journal of Physics: Condensed Matter*, 2017, **29**, 273002.
26 T. A. Young, J. J. Silcock, A. J. Sterling and F. Duarte, *Angewandte Chemie International Edition*, 2021, **60**, 4266–4274.
27 S. Batzner, A. Musaelian, L. Sun, A. Johansson, M. Geiger and T. Smidt, *NequIP*, 2021, `https://github.com/mir-group/nequip`.
28 F. Neese, *WIREs Computational Molecular Science*, 2018, **8**, e1327.
29 M. W. Mahoney and P. Drineas, *Proceedings of the National Academy of Sciences*, 2009, **106**, 697–702.
30 C. S. Jamieson, A. Sengupta and K. N. Houk, *Journal of the American Chemical Society*, 2021, **143**, 3918–3926.
31 S. Habershon, D. E. Manolopoulos, T. E. Markland and T. F. Miller, *Annual Review of Physical Chemistry*, 2013, **64**, 387–413.
32 K. L. K. Lee, M. S. Quinn, S. J. Kolmann, S. H. Kable and M. J. T. Jordan, *The Journal of Chemical Physics*, 2018, **148**, 194113.
33 Q. Liu, L. Zhang, Y. Li and B. Jiang, *The Journal of Physical Chemistry Letters*, 2019, **10**, 7475–7481.
34 Z. Liu, C. Patel, J. N. Harvey and R. B. Sunoj, *Physical Chemistry Chemical Physics*, 2017, **19**, 30647–30657.
35 R. F. Ribeiro, A. V. Marenich, C. J. Cramer and D. G. Truhlar, *The Journal of Physical Chemistry B*, 2011, **115**, 14556–14562.
36 S. Grimme, *Chemistry - A European Journal*, 2012, **18**, 9955–9964.