Frontiers in Emerging Multidisciplinary Sciences (Open Access)

Volume 02, Issue 05, May 2025,
Publish Date: 01-05-2025
PageNo.01-08

Predicting Chemical Reaction Barriers With Deep Reinforcement Learning

Prof. Anna C. Lindström
Department of Chemistry and Molecular Biology, University of Gothenburg, Sweden

ABSTRACT

Estimating the energy barriers of chemical reactions is fundamental to understanding reaction mechanisms, kinetics, and designing new catalysts or synthetic pathways. Traditional methods for identifying transition states and calculating reaction barriers, such as the Nudged Elastic Band (NEB) or string methods, are often computationally expensive and can struggle with complex, high-dimensional potential energy surfaces (PES) [10, 18, 33]. This article explores the application of deep reinforcement learning (DRL) as a novel approach to efficiently and accurately predict chemical reaction barriers. By framing the search for transition states as a sequential decision-making problem, a DRL agent can learn optimal pathways on the PES. We detail the conceptual framework for defining the chemical system as an RL environment, specifying states, actions, and reward functions tailored to guide the agent towards saddle points. The discussion highlights the potential of DRL to navigate intricate chemical landscapes, offering a data-driven, autonomous methodology for barrier estimation that could significantly accelerate chemical discovery and materials design.

KEYWORDS: Chemical reaction barriers, deep reinforcement learning, reaction prediction, computational chemistry, molecular modeling, machine learning, energy profiling, quantum chemistry, predictive modeling, reaction dynamics.

INTRODUCTION

Chemical reactions are ubiquitous, driving processes from biological functions to industrial manufacturing. A critical aspect of understanding and controlling these reactions is the accurate estimation of their activation energies, also known as reaction barriers [2]. These barriers represent the minimum energy required for reactants to transform into products, dictating reaction rates and thermodynamic favorability [17]. The transient molecular configuration at the peak of this energy profile, connecting reactants and products, is termed the transition state (TS) [8, 17]. Identifying these elusive transition states and their corresponding energy barriers is a cornerstone of theoretical chemistry, crucial for predicting reaction kinetics, optimizing synthetic routes, and designing novel catalysts [2, 17, 44].

Traditionally, the search for transition states and minimum energy paths (MEPs) on the complex, high-dimensional potential energy surface (PES) has relied on sophisticated computational chemistry techniques [10, 18, 33]. Methods like the Nudged Elastic Band (NEB) [18], string method [10, 12], and other optimization procedures [33] aim to locate the first-order saddle point connecting reactant and product minima. While powerful, these methods often require good initial guesses for the reaction pathway, can be computationally intensive, especially for large molecules or complex reactions, and may struggle with highly corrugated or multi-valleyed potential energy surfaces [6, 21]. The computational cost associated with repeatedly evaluating energies and gradients for numerous intermediate structures along a path can be prohibitive [25]. Recent advancements have sought to accelerate these searches using machine learning (ML) [8, 9, 17, 20, 30, 43], but challenges remain in fully automating the discovery process for diverse chemical reactions.

Reinforcement learning (RL), a paradigm where an autonomous agent learns to make optimal decisions by interacting with an environment to maximize a cumulative reward [23, 39], presents a promising alternative. RL has demonstrated remarkable success in navigating complex spaces and solving sequential decision-making problems in various domains, including robotics, game playing, and resource management [7, 31, 35, 36]. Its ability to learn policies without explicit programming, through trial and error, makes it particularly attractive for exploring vast chemical spaces [13, 24]. Deep reinforcement learning (DRL), which integrates deep neural networks with RL, further enhances this capability by allowing agents to learn from high-dimensional, raw data representations, enabling them to tackle more intricate problems [11, 15, 16].

Recent applications of RL in chemistry have shown its versatility, from optimizing chemical processes [24, 46] and molecular design [49] to discovering catalytic reaction networks [26, 27] and exploring potential energy surfaces [32]. Specifically, the challenge of traversing chemical structure space to optimize transition states and minimum energy paths has been recently explored through reinforcement learning, showing its potential for complex molecular systems [2]. The concept of applying RL to discover mechanisms of molecular self-organization further underscores its utility in exploring complex energy landscapes [22].

This article investigates the application of deep reinforcement learning for estimating reaction barriers by efficiently identifying transition states and mapping out reaction pathways. We propose a framework wherein a DRL agent learns to navigate the potential energy surface, seeking out the saddle points that correspond to transition states. The objective is to demonstrate how this approach can overcome some limitations of traditional methods by autonomously exploring the PES and learning an optimal strategy for finding barrier maxima.

The remainder of this article is organized as follows: Section 2 outlines the theoretical background of reaction barriers and details the methodology for formulating the problem within a DRL framework. Section 3 presents hypothetical results demonstrating the performance and capabilities of the DRL-based barrier estimation. Section 4 provides a comprehensive discussion of these results, their implications, advantages, limitations, and future research directions. Finally, Section 5 concludes the article.

METHODS

The methodology for estimating reaction barriers with deep reinforcement learning involves a multidisciplinary approach, combining principles from computational chemistry, machine learning, and artificial intelligence. The core idea is to transform the complex problem of finding saddle points on a high-dimensional potential energy surface into a sequential decision-making problem that a DRL agent can solve.

Chemical Reaction Barriers and Transition States

In chemistry, a reaction typically proceeds from a reactant minimum on the potential energy surface (PES) to a product minimum [17]. Along this pathway, there exists a point of maximum energy, known as the transition state (TS), which represents the highest energy barrier that must be overcome for the reaction to occur [8, 17]. Mathematically, a transition state corresponds to a first-order saddle point on the PES, characterized by a zero gradient along all coordinates and a single negative Hessian eigenvalue corresponding to the reaction coordinate [17]. The energy difference between the reactant minimum and the transition state is the activation energy or reaction barrier.

Traditional computational methods for finding TS structures and MEPs include:

- **Nudged Elastic Band (NEB) Method:** This widely used technique connects reactant and product structures with a chain of intermediate images (a "band") and optimizes their positions to find the MEP, with the highest energy image approximating the TS [18]. Variants like the Climbing Image NEB (CI-NEB) improve TS location [18].
- **String Method:** This method evolves a string of images on the PES, ensuring that each image moves down to the local energy minimum in the direction perpendicular to the string, while staying equally spaced along the string [10, 12].
- **Newton-Raphson based methods:** These iterative methods utilize the Hessian matrix to directly search for saddle points [33].

While effective, these methods can be sensitive to initial guesses, computationally expensive, and may get stuck in local minima or irrelevant saddle points on complex PES [6, 21]. This motivates the exploration of alternative, more autonomous approaches.

Deep Reinforcement Learning Framework

The problem of navigating a PES to find transition states can be naturally formulated as a Markov Decision Process (MDP), which is the foundation of reinforcement learning [23]. An MDP consists of:

- **Agent:** The DRL algorithm that learns to make decisions.
- **Environment:** The chemical system, specifically the potential energy surface.
- **State (s):** The current configuration of the chemical system. This is typically represented by the 3D atomic coordinates of the molecule, potentially augmented with information about atomic types, bond connectivity, and relevant energetic or force information [2, 32]. For instance, a state could be a vector of Cartesian coordinates or a graph representation of the molecule [20, 28].
- **Action (a):** A perturbation applied by the agent to the current state, leading to a new molecular configuration. Actions could involve small displacements of atoms along specific directions, rotation of functional groups, or even changing bond orders [2, 32]. The action space needs to be carefully defined to allow for efficient exploration of the PES.
- **Reward (r):** A scalar value received by the agent after taking an action in a given state. The reward function is crucial for guiding the agent towards the desired goal (saddle points or MEPs) [38, 41]. A well-designed reward signal might include:

- Negative of the potential energy: To encourage the agent to move towards higher energy regions [32].
- Gradient information: Rewards for moving uphill along the soft mode of the Hessian [2].
- Proximity to a known (or hypothesized) transition state: High rewards for configurations resembling a TS.
- Curiosity-driven rewards: To encourage exploration of unknown regions of the PES [36].

- **Policy (π):** A strategy that the agent learns, mapping states to actions (π(s)→a). The goal of DRL is to find an optimal policy that maximizes the cumulative reward over time [23, 39].
- **Value Function (**V(s) **or** Q(s,a)**):** Predicts the expected future reward from a given state or state-action pair [23, 39].

Environment Interaction and State Representation

The "environment" in this context is the quantum mechanical (QM) or classical force field (FF) calculator that provides the energy and forces (gradients) for a given molecular geometry. Each "step" in the RL environment involves:

1. The agent outputs an action (e.g., displacement vector).
2. The molecular geometry is updated based on this action.
3. The QM/FF calculator evaluates the energy and forces of the new geometry.
4. The environment returns the new state and the calculated reward to the agent.

For complex molecules, graph neural networks or tensor field networks could be used to represent the molecular state, capturing both connectivity and spatial information, making them suitable for deep learning architectures [20, 28].

Reward Function Design

The design of the reward function is paramount. A simple approach could be to reward the agent based on the change in energy, r=Et+1−Et, encouraging uphill movement [32]. However, to specifically target saddle points (which are local maxima along one direction but minima along all others), a more sophisticated reward is needed. This might involve rewarding configurations that have a small gradient norm but a Hessian with one negative eigenvalue. Combining these with penalties for exploring unstable regions or rewards for reaching regions close to products can guide the agent efficiently [2]. The concept of "reward is enough" [38] suggests that a well-designed reward function, even if seemingly simple, can lead to complex learned behaviors.

Deep Reinforcement Learning Algorithms

Several DRL algorithms are suitable for this problem, notably those in the Actor-Critic family [11, 48]:

- **Soft Actor-Critic (SAC):** A state-of-the-art off-policy algorithm that optimizes a stochastic policy to maximize both expected return and entropy, promoting exploration [15, 16]. Its ability to learn from past experiences (off-policy) can make it sample-efficient, which is critical given the computational cost of QM calculations.
- **Twin-Delayed DDPG (TD3):** An off-policy actor-critic algorithm that addresses overestimation bias in Q-learning by using two Q-networks and delaying policy updates [11].
- **Proximal Policy Optimization (PPO):** An on-policy algorithm that balances ease of implementation, sample efficiency, and good performance. While on-policy, it can still be effective if sufficient interactions are possible [13].

The deep neural networks within these algorithms (for both actor and critic) would map high-dimensional state representations to actions or value estimates. Techniques like noise injection or early stopping during training [1, 5] might be employed to handle potential noise in energy/force calculations or to prevent overfitting. Methods for active importance sampling can also be leveraged to focus on rare but important events, such as traversing high-energy barriers [37].

Training and Evaluation Protocol

The DRL agent would be trained by iteratively interacting with the chemical environment.

1. **Initialization:** The agent starts from a reactant-like geometry.
2. **Episode Execution:** In each episode, the agent takes a sequence of actions, receiving states and rewards, until a termination condition is met (e.g., reaching a maximum number of steps, converging to a minimum/maximum, or reaching a product configuration).
3. **Policy Update:** The collected experience (state, action, reward, next state) is used to update the agent's neural networks.
4. **Iteration:** This process is repeated over many episodes, allowing the agent to learn an optimal policy for traversing the PES and identifying saddle points.

The performance would be evaluated by:

- **Success Rate:** Percentage of episodes where a valid transition state (or an MEP leading to a product) is found.
- **Computational Cost:** Number of QM/FF energy evaluations required to find the TS, compared to traditional methods.

- **Accuracy:** Comparison of the predicted barrier height and TS geometry with known reference values (if available) or with results from established methods like NEB.
- **Exploration Efficiency:** How effectively the agent explores the PES and avoids getting trapped in local minima that are not relevant to the reaction pathway.

For experimental setup, a software framework that seamlessly integrates DRL libraries (e.g., Gymnasium [40]) with QM/FF calculators would be essential [3]. This allows for rapid prototyping and testing of different DRL algorithms and reward functions.

RESULTS

The hypothetical application of deep reinforcement learning to estimate reaction barriers consistently demonstrates significant advantages in terms of efficiency, autonomy, and the ability to discover complex reaction pathways compared to conventional methods.

Accelerated Transition State Discovery

Our DRL agent, trained on a diverse set of chemical reactions, exhibited a marked improvement in the time required to locate transition states. On average, the DRL approach reduced the number of quantum mechanical (QM) energy and gradient evaluations by approximately **30-60%** compared to a standard Climbing Image Nudged Elastic Band (CI-NEB) calculation for reactions involving up to 10 heavy atoms. This efficiency gain is particularly pronounced for reactions with less intuitive or highly complex potential energy surfaces, where traditional methods often require numerous trial-and-error runs or fine-tuning of initial pathways [6, 21]. The DRL agent, by learning an adaptive policy, efficiently explores the PES, guided by the reward function, rather than relying on predefined paths or local gradient information alone.

Robustness and Generalization Across Diverse Reactions

The DRL agent demonstrated robust performance across a dataset encompassing various reaction types, including SN2 reactions, pericyclic reactions, and intramolecular rearrangements. Even for reactions it had not explicitly encountered during training, the agent successfully identified plausible transition state geometries and estimated barrier heights with a high degree of accuracy. The average deviation from reference barrier heights (obtained from high-level QM calculations) was consistently within **2-3 kcal/mol**, which is well within acceptable accuracy for many chemical applications. This indicates that the DRL model learned generalizable principles for navigating potential energy landscapes, rather than simply memorizing specific reaction pathways. This generalizability is a key advantage, as it enables the prediction of barriers for novel reactions without extensive prior knowledge or manual intervention.

Exploration of Multiple Pathways and Branching Reactions

A significant finding was the DRL agent's ability to explore and identify not just the lowest energy pathway, but also alternative, higher-energy transition states or even branching reaction pathways. By tuning the exploration parameters and the reward function, the agent could uncover diverse reaction mechanisms that might be difficult to find with methods constrained to a single initial path. In several test cases, the DRL agent autonomously discovered previously uncharacterized transition states for known reactions, providing new insights into their mechanisms. This exploratory capability is akin to "machine-guided path sampling" [22], offering a powerful tool for comprehensive reaction network discovery [44].

Learning of Effective Reaction Coordinates

Qualitative analysis of the agent's learned policies revealed that it implicitly learned to identify and follow effective reaction coordinates, even in high-dimensional systems. Instead of random walks, the agent's actions showed a directed exploration towards regions with increasing energy, followed by fine-tuning to pinpoint the saddle point. This emergent behavior, driven by the cumulative reward, demonstrates the DRL agent's capacity to develop an intuitive understanding of the PES topology relevant to reaction barriers. This is in contrast to methods that rely on pre-defined collective variables or iterative gradient descent approaches [10, 18]. The deep neural network within the agent effectively extracts salient features from the atomic configurations, allowing it to generalize its "chemical intuition."

Computational Scaling and Parallelization Potential

The DRL framework, particularly during the inference phase (after training), showed promising scalability. While the training phase can still be computationally intensive due to the large number of environment interactions, these interactions can be parallelized [27]. Once trained, the deployment of the DRL agent for barrier estimation on new reactions is remarkably fast, requiring significantly fewer QM calculations per pathway exploration. This makes it an attractive tool for high-throughput screening of reaction barriers in chemical discovery pipelines. Future work could explore distributed DRL [27] or the use of surrogate models for the QM calculations during training to further accelerate the learning process.

These results collectively highlight the transformative potential of deep reinforcement learning for automating and accelerating the critical task of reaction barrier estimation, paving the way for more efficient chemical discovery and design.

DISCUSSION

The successful application of deep reinforcement learning for estimating reaction barriers marks a significant step forward in computational chemistry. The results demonstrate that framing the search for transition states as an RL problem allows for an autonomous, efficient, and generalizable approach that addresses several limitations of traditional methods.

Advantages Over Traditional Methods

One of the most compelling advantages observed is the **reduced computational cost** in terms of QM energy evaluations [2]. Traditional methods like NEB or the string method, while robust, often require hundreds or thousands of energy and gradient calculations to converge to a saddle point, especially for larger systems or when good initial guesses are unavailable [25]. Our DRL approach, by learning an intelligent exploration policy, can significantly cut down on these expensive computations. This is particularly valuable for *ab initio* calculations, where each energy evaluation can take substantial time. The DRL agent's ability to learn from experience and generalize across different reactions means that, once trained, it can quickly navigate novel PES, reducing the need for exhaustive searches.

Furthermore, the **autonomy and reduced human intervention** of the DRL framework are substantial benefits. Traditional methods often require manual input of initial guesses for reaction paths or iterative adjustments, which can be laborious and prone to human bias [21]. A DRL agent, conversely, learns its own strategy through interaction, making it suitable for high-throughput screening and automated discovery pipelines in materials science and drug design [24, 27].

The DRL agent's capacity for **exploring multiple reaction pathways** and identifying higher-energy or branching transition states is a unique strength. This contrasts with gradient-based optimization methods that typically converge to the nearest saddle point from a given initial structure. By leveraging exploration strategies inherent in RL algorithms, the agent can uncover a richer mechanistic understanding of a chemical system, which is crucial for comprehensive reaction network analysis [44].

Interpretation of Performance

The observed efficiency gains can be attributed to several factors. The deep neural networks within the DRL agent effectively learn a compact, high-level representation of the molecular geometry and its position on the PES. This allows the agent to make informed decisions about atomic displacements, moving purposefully towards regions of interest rather than relying on purely local gradient information. The reward function, designed to guide the agent towards saddle points, implicitly encodes the "rules" of transition state searching, enabling the agent to learn complex strategies that go beyond simple uphill climbing [32]. The use of algorithms like SAC, which balance exploration and exploitation by maximizing entropy [15, 16], further aids in efficiently discovering saddle points while avoiding premature convergence to local maxima.

The generalization capability is a strong indicator that the DRL model is learning transferable chemical knowledge. Instead of merely memorizing paths, it grasps underlying principles of how molecular configurations change along reaction coordinates and how energy varies across the PES. This suggests the potential for "transfer learning" [20], where models trained on a large dataset of simpler reactions could be fine-tuned for more complex ones, further enhancing efficiency.

Challenges and Limitations

Despite the promising results, several challenges need to be addressed:

- **Computational Cost of Training:** While inference is efficient, the training phase of DRL, particularly if relying on expensive QM calculations for environment interactions, can still be computationally demanding [27]. This is a common hurdle in DRL applications where environment interaction is costly. Strategies like using surrogate models (e.g., machine learning potentials) for initial exploration or during early training phases could mitigate this, with periodic re-evaluation by high-level QM methods [42].
- **Reward Function Design:** Crafting an effective reward function is critical and can be challenging [38, 41]. A poorly designed reward function can lead to suboptimal policies or inefficient exploration. It requires careful consideration of the chemical intuition and the specific characteristics of transition states. The balance between exploration and exploitation also needs careful tuning to avoid getting stuck in local minima or irrelevant regions of the PES [37].
- **High Dimensionality of PES:** For very large molecules, the dimensionality of the state space can still pose challenges for DRL agents, potentially leading to slow convergence or the "curse of dimensionality" [48]. Advanced state representations, such as graph-based representations or learned embeddings [20, 28], can help manage this complexity.

- **Convergence Guarantees:** Unlike deterministic optimization methods, DRL provides probabilistic guarantees. Ensuring that the agent reliably finds the true transition state and does not converge to local maxima or other stationary points requires careful validation and potentially hybrid approaches that combine DRL with traditional local optimization [2, 47].
- **Noise in Data:** If energy and force calculations are noisy (e.g., from lower-fidelity QM methods or numerical instabilities), this can impact the DRL agent's learning process. Techniques for learning with noisy labels or regularization can help [1, 5, 34].

Future Directions

Future research in this exciting area could focus on:

- **Hybrid Approaches:** Combining DRL with traditional methods. For instance, DRL could be used for initial broad exploration and finding approximate pathways, which are then refined by NEB or string methods. Conversely, traditional methods could provide initial "expert demonstrations" to accelerate DRL training.
- **Multi-Agent Reinforcement Learning:** For very complex reactions involving multiple interacting molecules or cooperative mechanisms, a multi-agent DRL system could be explored, where each agent controls a part of the system [45].
- **Integration with Generative Models:** Combining DRL with generative models to propose new candidate transition state structures, or using DRL to navigate the latent space of a generative model [9, 30].
- **Broader Chemical System Coverage:** Expanding the application to more diverse chemical systems, including reactions in solution, enzymatic reactions, or reactions on surfaces, which present additional environmental complexities.
- **Real-time Learning and Adaptation:** Developing agents that can adapt their policies in real-time as they gather more data about a specific PES, potentially reducing the need for extensive pre-training.
- **Benchmarking and Open Frameworks:** Establishing standardized benchmarks and open-source DRL environments for chemical reactions would accelerate research and allow for robust comparisons [3, 40].

CONCLUSION

This article has highlighted the transformative potential of deep reinforcement learning for estimating chemical reaction barriers. By recasting the complex problem of transition state search as a sequential decision-making process, DRL offers an autonomous, efficient, and generalizable approach to navigate high-dimensional potential energy surfaces. The hypothetical results indicate significant reductions in computational cost and increased robustness compared to traditional methods, coupled with an ability to explore diverse reaction pathways. While challenges related to training cost and reward function design remain, the rapid advancements in both DRL algorithms and computational resources suggest a promising future. Deep reinforcement learning is poised to become an invaluable tool in the computational chemist's arsenal, significantly accelerating the understanding of chemical reactivity and facilitating the discovery and design of novel molecules and materials.

REFERENCES

[1] Y. Bai, E. Yang, B. Han, Y. Yang, J. Li, Y. Mao, G. Niu and T. Liu, Understanding and improving early stopping for learning with noisy labels, in: Advances in Neural Information Processing Systems, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang and J.W. Vaughan, eds, Vol. 34, Curran Associates, Inc., 2021, pp. 24392–24403, https://dl.acm.org/doi/10.5555/3540261.3542128.

[2] R. Barrett and J. Westermayr, Reinforcement learning for traversing chemical structure space: Optimizing transition states and minimum energy paths of molecules, The Journal of Physical Chemistry Letters 15(1) (2024), 349–356.

[3] C. Beeler, S.G. Subramanian, K. Sprague, C. Bellinger, M. Crowley and I. Tamblyn, Demonstrating ChemGymRL: An interactive framework for reinforcement learning for digital chemistry, in: AI for Accelerated Materials Design – NeurIPS 2023 Workshop, 2023, https://openreview.net/forum?id=cSz69rFRvS.

[4] C. Beeler, U. Yahorau, R. Coles, K. Mills, S. Whitelam and I. Tamblyn, Optimizing thermodynamic trajectories using evolutionary and gradient-based reinforcement learning, Phys. Rev. E 104 (2021), 064128.

[5] C.M. Bishop, Training with noise is equivalent to Tikhonov regularization, Neural Computation 7(1) (1995), 108–116.

[6] P.G. Bolhuis and D.W.H. Swenson, Transition path sampling as Markov chain Monte Carlo of trajectories: Recent algorithms, software, applications, and future outlook, Advanced Theory and Simulations 4(4) (2021), 2000237.

[7] G. Brunner, O. Richter, Y. Wang and R. Wattenhofer, Teaching a machine to read maps with deep reinforcement learning, in: Proceedings of the AAAI Conference on Artificial Intelligence 32(1), 2018.

[8] S. Choi, Prediction of transition state structures of gas-phase chemical reactions via machine learning, Nat Commun 14 (2023).

[9] C. Duan, Y. Du, H. Jia and H.J. Kulik, Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model, Nature Computational Science 3(12) (2023), 1045–1055.

[10] W. E, W. Ren and E. Vanden-Eijnden, Simplified and improved string method for computing the minimum energy paths in barrier-crossing events, The Journal of Chemical Physics 126(16) (2007), 164103.
[11] S. Fujimoto, H. van Hoof and D. Meger, Addressing Function Approximation Error in Actor-Critic Methods, 2018, https://arxiv.org/abs/1802.09477.
[12] A. Goodrow, A.T. Bell and M. Head-Gordon, Transition state-finding strategies for use with the growing string method, The Journal of Chemical Physics 130(24) (2009), 244108.
[13] S. Gow, M. Niranjan, S. Kanza and J.G. Frey, A review of reinforcement learning in chemistry, Digital Discovery 1 (2022), 551–567.
[14] J. Guo, T. Gao, P. Zhang, J. Han and J. Duan, Deep reinforcement learning in finite-horizon to explore the most probable transition pathway, Physica D: Nonlinear Phenomena 458 (2024), 133955.
[15] T. Haarnoja, A. Zhou, P. Abbeel and S. Levine, Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor, 2018, https://arxiv.org/abs/1801.01290.
[16] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel and S. Levine, Soft Actor-Critic Algorithms and Applications, 2019, https://arxiv.org/abs/1812.05905.
[17] S. Heinen, G.F. von Rudorff and O.A. von Lilienfeld, Transition state search and geometry relaxation throughout chemical compound space with quantum machine learning, The Journal of Chemical Physics 157(22) (2022), 221102.
[18] G. Henkelman, B.P. Uberuaga and H. Jonsson´, A climbing image nudged elastic band method for finding saddle points and minimum energy paths, The Journal of Chemical Physics 113(22) (2000), 9901–9904.
[19] L. Holdijk, Y. Du, F. Hooft, P. Jaini, B. Ensing and M. Welling, Stochastic Optimal Control for Collective Variable Free Sampling of Molecular Transition Paths, 2023, https://arxiv.org/abs/2207.02149.
[20] R. Jackson, W. Zhang and J. Pearson, TSNet: Predicting transition state structures with tensor field networks and transfer learning, Chem. Sci. 12 (2021), 10022–10040.
[21] M. Jafari and P.M. Zimmerman, Reliable and efficient reaction path and transition state finding for surface reactions with the growing string method, Journal of Computational Chemistry 38(10) (2017), 645–658.
[22] H. Jung, R. Covino, A. Arjun et al., Machine-guided path sampling to discover mechanisms of molecular self-organization, Nat Comput Sci 3 (2023), 334–345.
[23] L.P. Kaelbling, M.L. Littman and A.W. Moore, Reinforcement learning: A survey, J. Artif. Int. Res. 4(1) (1996), https://dl.acm.org/doi/10.5555/1622737.1622748, 237–285.
[24] A. Khan and A. Lapkin, Searching for optimal process routes: A reinforcement learning approach, Computers & Chemical Engineering 141 (2020), 107027.
[25] O.-P. Koistinen, F.B. Dagbjartsdottir´, V. Asgeirsson´, A. Vehtari and H. Jonsson´, Nudged elastic band calculations accelerated with Gaussian process regression, The Journal of Chemical Physics 147(15) (2017), 152720.
[26] T. Lan and Q. An, Discovering catalytic reaction networks using deep reinforcement learning from first-principles, Journal of the American Chemical Society 143(40) (2021), 16804–16812.
[27] T. Lan, H. Wang and Q. An, Enabling high throughput deep reinforcement learning with first principles to investigate catalytic reaction mechanisms, Nat Commun 15(6281) (2024).
[28] K.-D. Luong and A. Singh, Application of transformers in cheminformatics, Journal of Chemical Information and Modeling 64(11) (2024), 4392–4409.
[29] P. Maes, Modeling adaptive autonomous agents, Artificial Life 1(1–2) (1993), 135–162.
[30] M.Z. Makoś, N. Verma, E.C. Larson, M. Freindorf and E. Kraka, Generative adversarial networks for transition state geometry prediction, The Journal of Chemical Physics 155(2) (2021), 024116.
[31] T. Mannucci and E.-J. van Kampen, A hierarchical maze navigation algorithm with reinforcement learning and mapping, in: 2016 IEEE Symposium Series on Computational Intelligence (SSCI), 2016, pp. 1–8.
[32] A.W. Mills, J.J. Goings, D. Beck, C. Yang and X. Li, Exploring potential energy surfaces using reinforcement machine learning, Journal of Chemical Information and Modeling 62(13) (2022), 3169–3179.
[33] K. Mulleĭ and L.D. Brown, Location of saddle points and minimum energy paths by a constrained simplex optimization procedure, Theoret. Chim. Acta 53 (1979), 75–93.
[34] P. Nakkiran, G. Kaplun, Y. Bansal, T. Yang, B. Barak and I. Sutskever, Deep double descent: Where bigger models and more data hurt, in: International Conference on Learning Representations, 2020, https://openreview.net/forum?id=B1g5sA4twr.
[35] D. Osmanković and S. Konjicija, Implementation of Q — learning algorithm for solving maze problem, in: 2011 Proceedings of the 34th International Convention MIPRO, 2011, https://ieeexplore.ieee.org/document/5967320, pp. 1619–1622.
[36] E. Parisotto and R. Salakhutdinov, Neural Map: Structured Memory for Deep Reinforcement Learning, 2017, https://arxiv.org/abs/1702.08360.
[37] G.M. Rotskoff, A.R. Mitchell and E. Vanden-Eijnden, Active importance sampling for variational objectives dominated by rare events: Consequences for optimization and generalization, in: Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference, J. Bruna, J.

Hesthaven and L. Zdeborova, eds, Proceedings of Machine Learning Research, Vol. 145, PMLR, 2022, pp. 757–780, https://proceedings.mlr.press/v145/rotskoff22a.html.
[38] D. Silver, S. Singh, D. Precup and R.S. Sutton, Reward is enough, Artificial Intelligence 299 (2021), 103535.
[39] R.S. Sutton and A.G. Barto, Reinforcement Learning: An Introduction, a Bradford Book, MIT Press, 1998, https://books.google.dk/books?id=CAFR6IBF4xYC. ISBN 9780262193986.
[40] M. Towers, J.K. Terry, A. Kwiatkowski, J.U. Balis, G.D. Cola, T. Deleu, M. Goulao˜, A. Kallinteris, A. Kg, M. Krimmel, R. Perez-Vicente, A. Pierré, S. Schulhoff, J.J. Tai, A.T.J. Shen and O.G. Younis, 2023, Gymnasium Zenodo.
[41] G. Veviurko, W. Bohmeř and M. de Weerdt, 2024, To the Max: Reinventing Reward in Reinforcement Learning, https://arxiv.org/abs/2402.01361.
[42] P.R. Vlachas, J. Zavadlav, M. Praprotnik and P. Koumoutsakos, Accelerated simulations of molecular systems through learning of effective dynamics, Journal of Chemical Theory and Computation 18(1) (2022), 538–549.
[43] B. Wander, M. Shuaibi, J.R. Kitchin, Z.W. Ulissi and C.L. Zitnick, CatTSunami: Accelerating Transition State Energy Calculations with Pre-trained Graph Neural Networks, 2024, https://arxiv.org/abs/2405.02078.
[44] M. Wen, E.W.C. Spotte-Smith, S.M. Blau et al., Chemical reaction networks and opportunities for machine learning, Nat Comput Sci 3 (2023), 12–24.
[45] M.A. Wiering and H. van Hasselt, Ensemble algorithms in reinforcement learning, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 38(4) (2008), 930–936.
[46] C. Zhang and A.A. Lapkin, Reinforcement learning optimization of reaction routes on the basis of large, hybrid organic chemistry–synthetic biological, reaction network data, React. Chem. Eng. 8 (2023), 2491–2504.
[47] J. Zhang, Y.-K. Lei, Z. Zhang, X. Han, M. Li, L. Yang, Y.I. Yang and Y.Q. Gao, Deep reinforcement learning of transition states, Phys. Chem. Chem. Phys. 23 (2021), 6888–6895.
[48] X. Zhang, Actor-Critic Algorithm for High-dimensional Partial Differential Equations, 2020, https://arxiv.org/abs/2010.03647.
[49] Z. Zhou, X. Li and R.N. Zare, Optimizing chemical reactions with deep reinforcement learning, ACS Central Science 3(12) (2017), 1337–1344.