nature communications

Article

https://doi.org/10.1038/s41467-026-69586-8

Exploring chemistry and catalysis by biasing skewed distributions via deep learning

Received: 17 April 2025

Accepted: 29 January 2026

Published online: 21 February 2026

Check for updates

Zhikun Zhang [1] & GiovanniMaria Piccini [2]

The automated discovery of chemical and catalytic reactions remains a major challenge in computational chemistry, particularly in complex systems where conventional methods struggle to identify optimal searching directions. Here, we propose Loxodynamics, a machine-learning-driven approach for reaction exploration via biased molecular dynamics. By leveraging the skewness of local probability distributions, Loxodynamics dynamically determines low-energy barrier directions, efficiently guiding the system toward metastable states. The core of our framework is Skewencoder, an Autoencoder augmented with a skewness-based loss function that extracts reaction coordinates from minimal sampling data. Through iterative sample-and-search cycles, the system adaptively maps the free energy surface, capturing finite-temperature effects critical to complex reactive environments. We validate our method across model potentials, gas-phase reactions (S_N 2 and Diels-Alder), and catalytic alcohol dehydration in acidic chabazite under operando conditions. Loxodynamics provides a systematic framework for reaction discovery that addresses key limitations of conventional techniques, notably by allowing acceleration without elevated temperatures or a priori knowledge of collective variables.

Exploring chemical reactivity from first principles is a grand challenge in computational chemistry[1–3]. This is fundamental across all fields of chemistry and particularly critical in catalysis, where identifying multiple products, intermediates, and reaction mechanisms is key. Even simple systems can generate a diverse range of possible outcomes, and the complexity further increases when reactivity is influenced by environmental factors such as interfaces, solvents, and catalytic conditions[3,4].

The conventional workflow to tackle this problem typically involves inferring the structures of reactants and products based on known or anticipated reaction outcomes or relying predominantly on chemical intuition.[5–18]. Once potential products or intermediates have been hypothesized, a guess structure is constructed, and its energy is minimized through geometry optimization.

While this approach has achieved significant success, it faces two main limitations. First, designing the guessed product structure heavily relies on chemical intuition, which may not always yield realistic predictions. To mitigate this, various methods have been developed to explore the potential energy surface (PES). Many of these techniques leverage the local curvature of the PES to identify the minimum energy pathway connecting reactants, transition states, and products[19–24]. These methods have proven effective but are generally restricted to small, low-complexity systems where complex dynamic effects can be neglected.

Alternatively, for more complex scenarios, conceptual chemical knowledge can guide the identification of potential intermediates, such as by searching for a minimum energy path (MEP) connecting them. If the two endpoints of an elementary reaction are known, interpolation and string methods can be applied to locate the corresponding MEP[6,8,25]. Conceptual chemical insight can also be exploited to rapidly identify potential intermediate candidates that are connected to the starting structure via an elementary reaction—particularly when the type of reaction mechanism relevant to the system under study is known. These heuristic approaches have proven highly effective in characterizing products and intermediates in well-established processes[17,26]. However, their applicability is often

[1]Institute of Technical Thermodynamics, RWTH Aachen University, Aachen, Germany. [2]Dipartimento di Scienze Chimiche e Geologiche, Università Degli Studi di Modena e Reggio Emilia, Modena, Italy. e-mail: giovannimaria.piccini@unimore.it

constrained by the requirement for prior knowledge of the reaction pathways. The use of graph theory and connectivity analysis offers an efficient yet mathematically elegant alternative for exploring chemical reaction networks[27], with promising applications in catalysis[15,27].

For a detailed overview and classification of these reaction exploration approaches, we refer the reader to the excellent reviews by Reiher and coworkers[2,28,29].

The second key challenge is that PES exploration methods very often rely on structure optimization techniques that overlook finite-temperature effects, thereby neglecting the broader configurational space accessed by the system. This limitation increases the risk of missing the true lowest-energy minimum corresponding to the reaction product or intermediate of interest.

To incorporate finite-temperature effects in reactivity studies, molecular dynamics (MD) simulations can be employed. However, standard MD operates on timescales much shorter than those of activated processes like chemical reactions, making it challenging to capture such rare events.

To overcome this limitation, various enhanced sampling methods have been developed to promote transitions across high-energy barriers separating metastable states. Those strategies involve increasing the system's temperature[30–32], or adding a bias potential to the global PES of the system[33–35], in order to boost the likelihood of barrier crossing. These approaches often employ multiple replicas, as in parallel tempering, or constrain molecules within a small fictitious volume[36] while applying elevated temperatures. While effective in facilitating exploration, these methods can also introduce artifacts, leading to the identification of high-energy pathways and reaction products that may not be experimentally relevant under realistic conditions.

Another widely used approach involves applying bias potentials along predefined collective variables (CVs) to accelerate transitions between states. This category includes umbrella sampling[37], metadynamics[38], and On-the-fly Probability Enhanced Sampling (OPES)[39]. While highly effective for sampling known reaction pathways, these methods rely heavily on prior knowledge of metastable states, needed to design the CVs. This dependence makes them less suitable for reaction discovery.

In general, designing a fully agnostic CV for reactivity exploration with a high sampling efficiency is impractical without prior knowledge of the expected products. Moreover, even with this information, selecting the correct bias direction along the CV is essential to guide the system along the lowest-energy reaction pathway and efficiently drive the transition to the products.

Here, we propose a method that overcomes these limitations by automatically constructing a low-dimensional CV while simultaneously determining the optimal direction toward the most probable reaction products. By following pathways with the lowest energy barriers, this approach enables efficient and agnostic exploration of likely reaction channels. To achieve this, we draw inspiration from two key concepts: the Anharmonic Downward Distortion Following (ADDF) method[19,21] and the Artificial Force Induced Reaction (AFIR) method[40–43], combined with pulling-based exploration techniques used in steered MD[44,45] and ratcheted-and-pawl MD[46] simulations.

The ADDF method explores reaction pathways by following anharmonic distortions by mapping on the PES based on the concept of hypersphere search[19,47], while the AFIR method identifies associative reactions using artificial forces between reactants. Both have been successful in mapping chemical reactivity, but neglect temperature effects.

Conversely, dynamic approaches like steered MD and ratcheted-and-pawl MD enable exploration of activated processes but require prior knowledge of initial and final states to define an efficient CV. Additionally, the pulling direction must be predefined by the user.

Our approach unifies these concepts by automatically identifying the CV along which anharmonicity is maximal, thereby guiding the exploration along the slowest ascending directions of the PES. Instead of explicitly evaluating high-dimensional PES directions, the method first samples the local minimum of the reactant state. By analyzing asymmetries in the local distribution—specifically its skewness—it infers the optimal biasing direction. This eliminates reliance on predefined knowledge, enabling automated reaction pathway discovery and efficient sampling of the most probable reaction channels.

Results

Loxodynamics

The essence of our method can be illustrated, for the sake of simplicity, by considering the dynamics of a finite mass particle in a one-dimensional double-well potential (Fig. 1) at finite temperature, e.g., Langevin dynamics. Let us imagine that the shape of the potential is not known and the temperature is sufficiently low, i.e., $k_BT \ll E_{act}$, so that the particle is trapped in the left basin. In such a situation, it is not trivial to decide whether the particle should be pulled toward the left or the right along x to go to the other state. Our key idea is that the local potential experienced by the particle inherently exhibits some degree of anharmonicity, which, in turn, must be reflected in the shape of its probability distribution. By sampling the vicinity of the minimum for a sufficient duration, it is possible to estimate the local probability distribution function and its moments, such as the skewness. The latter, for a one-dimensional distribution, is defined in Equation (1) as:

$$\gamma = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)^{\frac{3}{2}}} \tag{1}$$

where n is the number of samples, x_i represents the i-th value of the random variable, and $\bar{x}$ is the sample mean. A positive value of γ indicates a left-modal distribution, meaning that the tail of the distribution becomes more pronounced as the random variable x increases. Conversely, a negative γ signifies that the tail is more pronounced as it decreases. In the context of an asymmetric anharmonic potential, a positive skewness implies that shifting the mass along the positive direction of x is likely to guide the system across the lowest energy barrier, where the tail of the distribution is longer, thus leading it into a new metastable state. For this reason, we named our method Loxodynamics, deriving from the Greek word loxós, meaning skewed, to emphasize its core principle of leveraging skewness in probability distributions to guide the sampling process.

In this model system, the initial sampling shows positive skewness, indicating that the particle should move to the right to reach the other basin. To implement this, a harmonic wall is applied, shifted by a constant offset Δx to the right of the mean μ of the distribution sampled in the previous step, and sampling resumes. The new potential, which combines the physical double-well potential and the restraining bias, makes the particle sampling a new landscape and search iteratively for the next metastable state direction based on the skewness of the updated distribution. The lower panel of Fig. 1 shows the evolution of the local equilibrium distribution after applying the restraining biases at each successive iteration. In this simple case, the skew remains positive and, therefore, the direction along which the bias must be applied is always positive along x. In this way, through successive iterations of sample-and-search, the system gradually exits the left basin and reaches the right one, completing the exploration process.

Real chemical systems often require multiple dimensions to capture relevant anharmonicity changes in the potential. A practical approach is to define N_d descriptors $d_i(\mathbf{R})$, functions of atomic coordinates $\mathbf{R}$, which serve as the basis for constructing the CV. The

https://doi.org/10.1038/s41467-026-69586-8

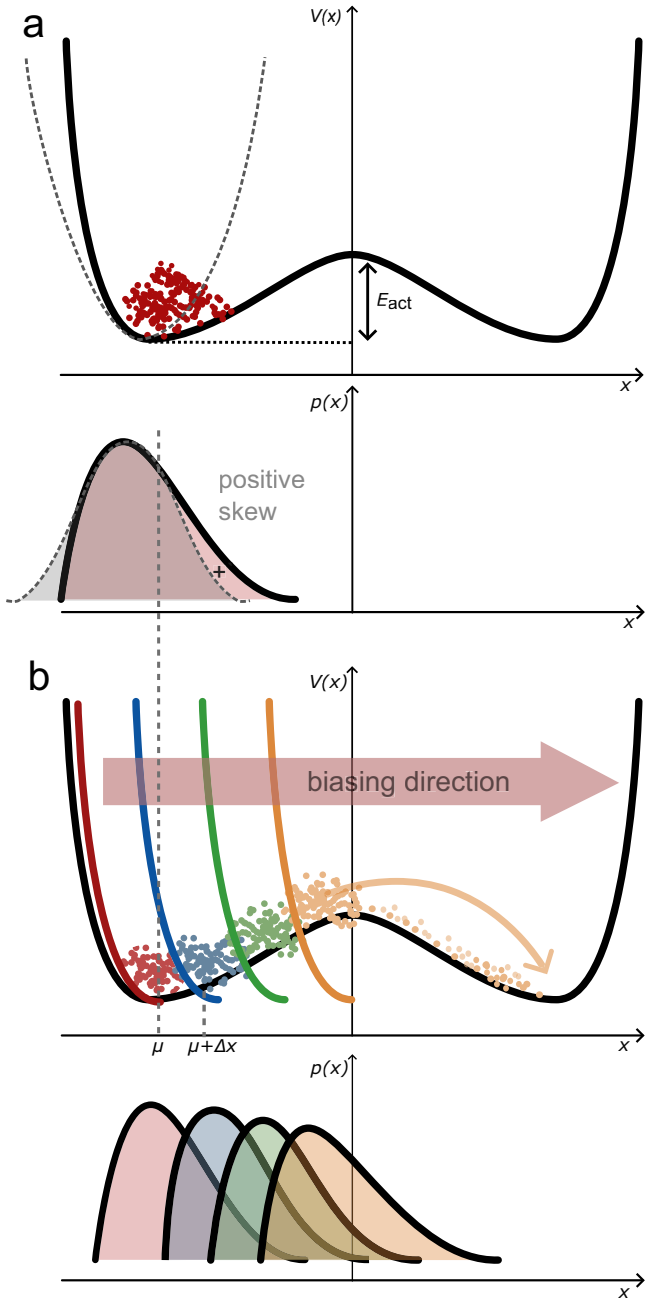

Fig. 1 | Schematic representation of the Loxodynamics approach in 1D. **a** Comparison of a double-well potential (solid black) with a local harmonic approximation (dashed). The activation barrier between the two wells is denoted as E_{act}. The lower panel displays the corresponding probability distributions $p(x)$: the harmonic potential yields a symmetric Gaussian (gray), whereas the double-well potential produces a positively skewed distribution (red). This skewness determines the biasing direction. **b** Successive sampling driven by half-harmonic restraining walls. The upper panel shows the wall progression (half-parabolic curves) and sample points (dots). The initial wall (red) is positioned at the unbiased sample mean μ. In subsequent iterations (blue, green, yellow), the wall is translated by a fixed offset Δx while maintaining a constant spring constant κ. The lower panel illustrates the resulting evolution of the sample distributions, which become increasingly asymmetric.

procedure remains similar to the 1D case: after defining these descriptors, the system is briefly sampled at a given temperature, generating an N_d-dimensional distribution. However, extending skewness to higher dimensions is non-trivial due to the complexity of quantifying asymmetry[48].

To leverage the benefits of 1D skewness, dimensionality reduction is necessary. This reduction must preserve key system features while maintaining differentiability for CV construction. A tailored bias can then be applied in this reduced space, enabling efficient exploration of the potential energy landscape.

To achieve this, we developed Skewencoder, an Autoencoder-based framework that reduces data to a 1D latent space while incorporating higher-moment statistical information. Autoencoder-based CVs have been widely adopted as an initial means to identify essential features relevant to reaction mechanisms and molecular transformations, thereby serving as a foundation for reaction space exploration. Chen et al.[49,50] developed an iterative exploration protocol that integrates umbrella sampling with dynamically updated Autoencoder-based CVs derived from the evolving set of sampled configurations. Their approach employs a grid-based boundary detection method to determine bias placement by identifying outlier configurations potentially indicative of state transitions. In comparison, our framework—while also employing iterative CV-accelerated discovery—prioritizes the exploitation of collective statistical properties within the sampled data, such as skewness, to guide and optimize transition pathways. Ketkaew et al.[51,52] introduced an autoencoder-based strategy aimed at constructing data-driven collective variables for enhanced sampling, trained exclusively on reactant-state configurations. Their approach exploits deep autoencoders to extract latent representations from unbiased trajectories and employs these representations within flooding-based enhanced sampling schemes to reconstruct free energy surfaces. While this methodology alleviates the need for explicit product-state information in the definition of collective variables, it remains inherently tied to static CVs and relies on conventional enhanced sampling paradigms, without introducing an explicit directional bias capable of actively driving exploration toward new regions of phase space.

The aforementioned approaches combine CVs with enhanced sampling techniques such as umbrella sampling, metadynamics, or OPES to accelerate phase-space exploration. However, these methods do not encode an intrinsic directionality in the applied bias[53]. In contrast, Loxodynamics introduces a bias explicitly designed to drive the system toward new states along a CV that is updated on-the-fly during the exploration. Metadynamics and OPES, in particular, are flooding approaches, in which the bias progressively fills local free-energy basins defined along static, pre-defined CVs, thereby enhancing fluctuations within those basins. The core principle of Loxodynamics is fundamentally different: rather than amplifying fluctuations, it deliberately focuses them with the explicit goal of propelling the system uphill in free energy.

By integrating an Autoencoder neural network (NN) with a loss function based on the skewness of projected samples, Skewencoder efficiently utilizes limited early-stage sampling data in a multitask learning framework, as defined in Equation (2).

$$\mathcal{L}_{\text{total}}(\mathbf{x};\boldsymbol{\theta}) = \mathcal{L}_{\text{AE}}(\mathbf{x};\boldsymbol{\theta}) + \alpha\mathcal{L}_{\text{skew}}(\mathbf{x}';\boldsymbol{\theta}) + \beta||\mathbf{w}||_2 \tag{2}$$

where $\mathcal{L}_{\text{AE}}$ denotes the Autoencoder loss function, and $\mathcal{L}_{\text{skew}}$ is a skewness loss term designed to maximize the skewness of the latent space distribution. Here, $\mathbf{x}$ and $\mathbf{x}'$ represent the global and local set of input descriptors, respectively, and $\boldsymbol{\theta}$ comprises all neural network parameters (i.e., weights and biases across all layers). The term $||\mathbf{w}||_2$ refers to the L2 regularization applied to the weights, while α and β are hyperparameters that control the relative contributions of the skewness and regularization terms, respectively.

The Autoencoder loss function is defined in Equation (3) using the standard reconstruction error, based on the difference between the

https://doi.org/10.1038/s41467-026-69586-8

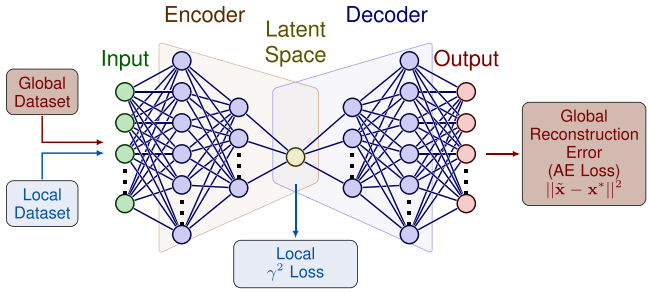

Fig. 2 | Schematic representation of the Skewencoder architecture, a multitask learning framework based on an Autoencoder with a single bottleneck neuron in the latent space. This design compresses high-dimensional data into projections on a one-dimensional manifold for skewness calculation. In our approach, the input dataset is divided into two parts. Data from the global (incremental) dataset (red path) contributes to the original Autoencoder loss function (AE loss)(i.e., reconstruction mean squared error (MSE) loss, which minimizes the difference between the inputs and the decoder's output. Data from the local dataset (blue path) contributes to a custom loss function based on the skewness of the data, i.e. γ^2 loss, in the latent space. This figure is adapted from `Neural Networks` by Izaak Neutelings, used under CC BY-SA 4.0. This figure is licensed under CC BY-SA 4.0.

inputs and the reconstructed outputs:

$$\mathcal{L}_{\mathrm{AE}}(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} ||\tilde{\mathbf{x}}_i(\mathbf{x}_i;\boldsymbol{\theta}) - \mathbf{x}_i^*||^2 \tag{3}$$

where n is the number of training examples, $\mathbf{x}_i^*$ is the vector of inputs for the i-th training sample, $\tilde{\mathbf{x}}_i$ is the output vector of the Autoencoder, which is a function of the NN parameters $\boldsymbol{\theta}$ and the input descriptor $\mathbf{x}_i$ and the difference is calculated as the mean squared error (MSE) of the samples, expressed in terms of vector norms. The global dataset accumulated across iterations is used for optimizing the Autoencoder.

In contrast, the skewness loss, which is calculated based on the projected CVs in the latent space (depicted by the light-blue path in Fig. 2), is defined in Equation (4) as:

$$\mathcal{L}_{\mathrm{skew}}(\mathbf{x}';\boldsymbol{\theta}) = \ln\left(1 + e^{-\gamma(\mathbf{s}(\mathbf{x}';\boldsymbol{\theta}))^2}\right) \tag{4}$$

where $\mathbf{s}(\mathbf{x}';\boldsymbol{\theta})$ denotes the latent representation computed by the Autoencoder NN, $\gamma(\mathbf{s}(\mathbf{x}';\boldsymbol{\theta})$ denotes the skewness, computed according to Equation (1), and is derived solely from the local data $\mathbf{x}'$ generated by the MD simulation in the most recent iteration (refer to the Supplementary Information (SI) for further details). The use of both global and local datasets for calculating the loss functions defines the multitask nature of our method.

In the early iterations of the sample-and-search process, designing an effective CV is challenging due to the limited data available for training the Autoencoder. To address this, we employed a warm-start training approach, which has been widely successful in convex optimization[54]. Instead of training the model from scratch at each iteration, the parameters learned from the previous iteration are used as the starting point, and the model is trained to converge from these pre-initialized values (see SI for details).

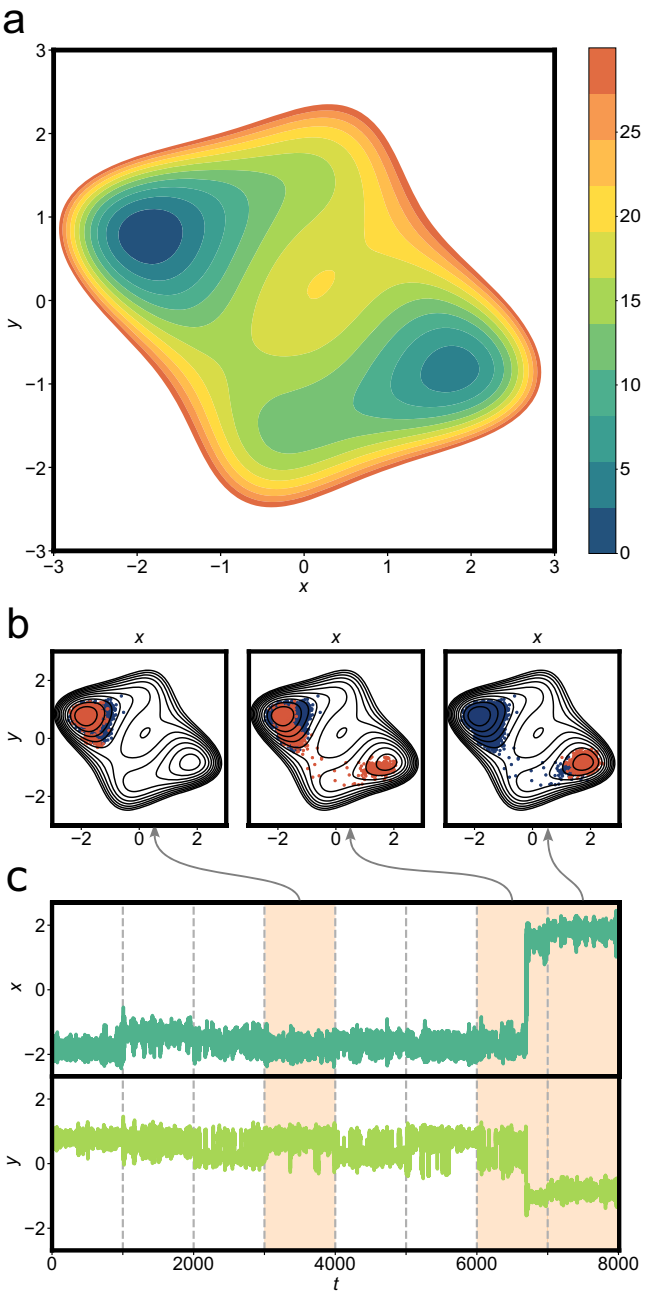

Fig. 3 | The Loxodynamics process on a 2-D potential energy surface (PES). a The PES visualized as a colored heatmap. **b** Sampling distribution for the key iteration. Blue points represent the incremental dataset accumulated from previous steps, while red points correspond to the samples generated during the current iteration. **c** Time evolution of the x (dark green) and y (light green) coordinates, reported in Lennard-Jones units. The shaded windows highlight the specific time interval corresponding to the snapshot shown in panel b. A copy of this figure using a perceptually uniform color palette is available in SI.

Case study

2D model potential. As a proof of concept, we tested our method on simple 2D model potentials using Langevin dynamics. For clarity, we present results for one of the three tested potentials, shown in Fig. 3 (panel a). The explicit functional form and results for the other cases are provided in the SI.

Starting in the upper basin, the algorithm progressively sampled the potential. The panel b of Fig. 3 highlights key iterations of the exploration, while the panel c shows the trajectory along the x and y coordinates, illustrating how the system systematically gathers information and transitions toward the second basin.

This potential features two pathways with barriers of different heights. The algorithm iteratively analyzes the data distribution, identifies the direction of maximal anharmonicity—where skewness is highest—and applies harmonic walls to bias the system along the lower-energy route, ensuring a smooth and efficient transition to the next metastable state.

In detail, the blue points in panel b of Figure 3 represent samples from the incremental dataset, which aggregates data from all previous and current iterations to form the global training set for the Autoencoder's reconstruction loss. In contrast, red points correspond to samples from the current iteration, constituting the local training dataset optimized for the skewness loss. Sampling at each iteration follows these criteria, progressively expanding the explored region and allowing the system to identify new minima on the free energy landscape.

In this example, both pathways lead to the same product. However, in realistic chemical systems, a single reactant can yield multiple products, each requiring distinct transition regions. Our method ensures that the system follows the lowest-barrier pathways, identifying the most probable reaction routes and products.

Moreover, upon algorithm termination, the accumulated samples provide sufficient data to construct a latent space that effectively approximates a meaningful CV (see SI), ensuring an accurate representation of the reaction pathway, even in complex energy landscapes.

Gas phase chemical reactions. To demonstrate the capability of Loxodynamics in capturing chemical reactivity, we selected two prototypical reactions: the S_N 2 nucleophilic substitution of chloromethane by a fluoride anion and the gas-phase Diels-Alder reaction between ethylene and 1,3-butadiene. These reactions exemplify key aspects of reactivity, including simultaneous bond rearrangement (synchronous bond breaking and formation) and significant differences in the shape of the reactant and product basins as well as asymmetric barrier heights in the forward and reverse directions.

To detect whether the system has relocated into a new chemical metastable state, bond topology changes were assessed by monitoring variations in contacts between neighboring atoms (see "Methods" section for details). In addition, how metastable states are identified in different case studies is also discussed (see SI).

To begin, we investigated the S_N 2 nucleophilic substitution of chloromethane by a fluoride anion. The reaction is characterized by two key distances: d_F, the distance between the carbon and fluoride ions, and d_{Cl}, the distance between the carbon and chlorine ions. These distances were used as inputs for our Skewencoder. To ensure consistent training, all distances were normalized to zero mean and unit variance.

Figure 4a, c illustrate the progression of Loxodynamics, tracked via the control variable $d = d_{Cl} - d_F$, which provides a simplified yet effective one-dimensional CV to control the reaction progression[55]. After a few iterations, the Skewencoder identified the most favorable direction, enabling the bias to efficiently drive the system toward the new metastable state.

Due to the asymmetric energy barriers between reactants and products, a larger harmonic constant was applied when steering the reaction in the direction $Cl^- + CH_3F \rightarrow CH_3Cl + F^-$ to account for the higher barrier. The method consistently determines the optimal pulling direction within the two-dimensional plane defined by key interatomic distances.

This highlights the robustness of the approach: even with a simple basis of chemical bond descriptors, it effectively identifies the reaction pathway with the lowest barrier by targeting the most anharmonic directions on the PES. Statistical asymmetries in locally sampled distributions are leveraged after short sampling trajectories.

Subsequently, we extended the application of our method to the more complex chemical case of the Diels-Alder reaction between 1,3-butadiene and ethylene. For this system, we deliberately chose to include all 15 possible interatomic distances among the heavy (carbon) atoms in the molecule as descriptors (see SI), rather than preselecting only the two most relevant C-C distances involved in the reactive process. This decision, made regardless of whether these distances directly describe relevant bonds in either metastable state, was aimed at testing the robustness and general applicability of the method.

Figure 4b, d illustrate the progression of Loxodynamics for the Diels-Alder reaction. The method successfully identifies key reaction pathways in both forward and reverse directions without relying on predefined mechanistic knowledge.

The forward reaction presents significant challenges due to its broad reactant state with high entropy, where the conformational variability of the separated fragments complicates the identification of an effective biasing direction. As shown in Figure 4b, after a few iterations, the method successfully aligns the fragments, facilitating σ bond formation in the cyclohexene product state. By iteration 8, the system surmounts the energy barrier and transitions to the product state.

The reverse reaction is equally complex but for different reasons. The deep and narrow ring state exhibits low anharmonicity, making it difficult to determine the correct biasing direction without mistakenly exploring alternative anharmonic channels, potentially leading to undesired C-C or C-H rearrangements. Despite these challenges, the method effectively identifies the ring-opening pathway and smoothly guides the system out of the metastable state, as depicted in Figure 4d.

This highlights the method's robustness in traversing complex energy landscapes, enabling accurate and insightful reactivity exploration.

Catalytic ethanol dehydration in acidic chabazite. In this section, we apply our method to explore a highly complex catalytic system under realistic *operando* conditions. We applied Loxodynamics to a reaction occurring at the Brønsted acid site (BAS) of a zeolite. Here, we focus on the dehydration of ethanol to ethylene at the BAS of acidic chabazite (H-CHA), a widely used zeolite, a complex reaction, and highly relevant for the chemical industry.

Experimental[56–58] and computational[59,60] studies have shown that one mechanism is dominant, as it has a lower activation barrier: the concerted mechanism. In this pathway, the acidic zeolite protonates the hydroxyl group of the ethanol molecule, facilitating the elimination of water. Simultaneously, a hydrogen atom is abstracted from the ethyl group, regenerating the BAS proton and restoring the catalyst.

However, evidence suggests the existence of an alternative pathway with a slightly higher energy barrier, yet still kinetically feasible. This mechanism, commonly referred to as the stepwise mechanism, involves the formation of an ethoxy-framework complex intermediate. More precisely, initially, the elimination of water leaves a $CH_3CH_2^{+1}$ cation, which binds to a negatively charged BAS oxygen, forming a relatively stable species. Subsequently, a proton of the methyl group of the ethoxy group is abstracted by a neighboring BAS oxygen, oxidizing the intermediate and yielding ethylene.

To set up our simulatio,n we started from the adsorption state of ethanol at the BAS. To begin our exploration, our input set comprises 37 descriptors, including interatomic distances among heavy atoms, such as C-O and C-C, but also with light atoms C-H and O-H, with minimal reliance on chemical intuition (see "Methods" and SI for details).

To our surprise, despite the simplicity and generality of these descriptors, Loxodynamics successfully identifies the ethylene product state through two distinct pathways. In one case, it follows a direct route to the product, while in another, it first identifies an intermediate corresponding precisely to the ethoxy-framework complex, which subsequently evolves into ethylene via the expected mechanism.

Figure 5 illustrates two detected reaction pathways from the sample-and-search process, in agreement with the static calculations by Kim et al.[59], corresponding to the concerted and stepwise mechanisms, respectively.

https://doi.org/10.1038/s41467-026-69586-8

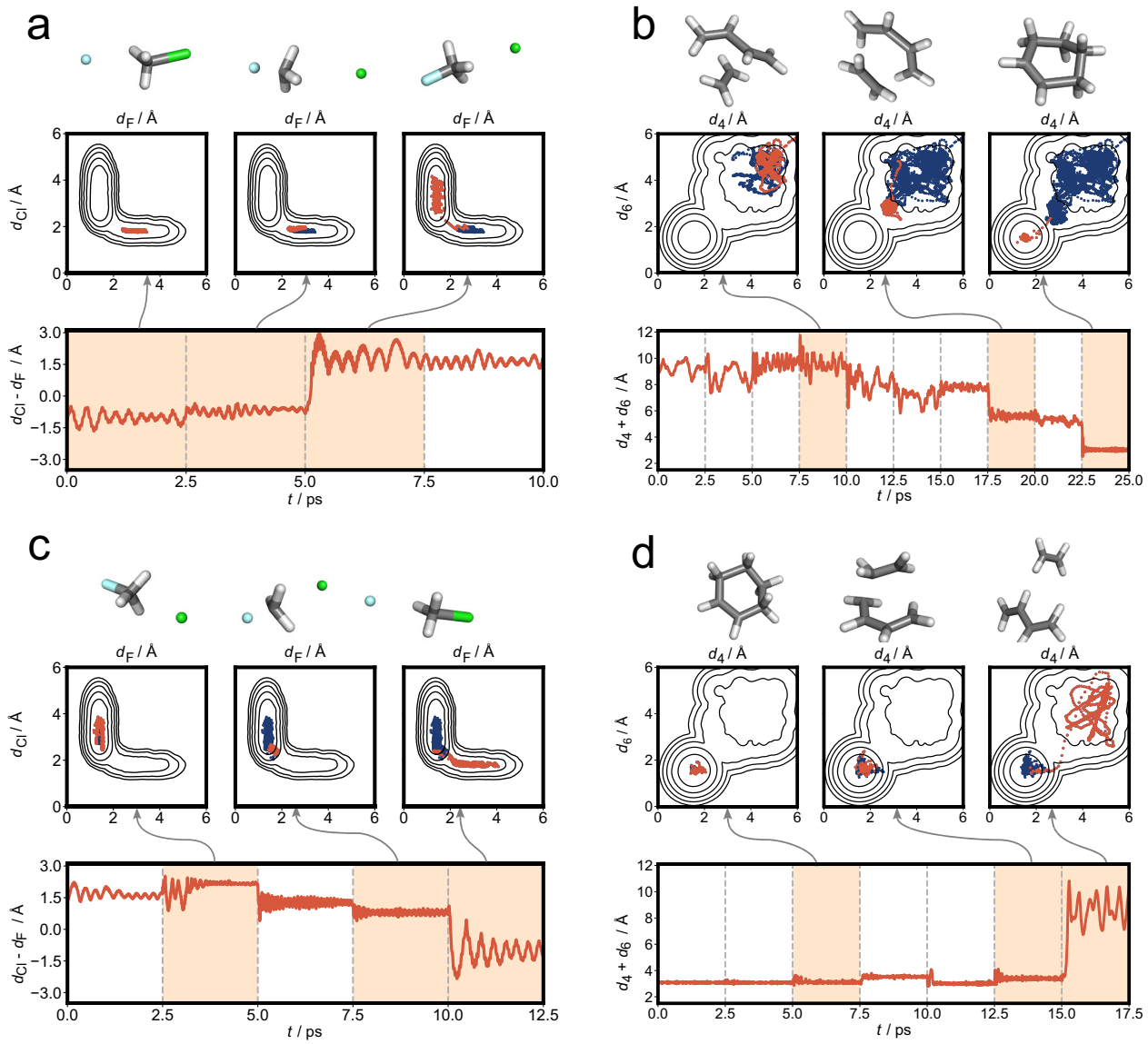

Fig. 4 | System evolution in Loxodynamics process for two chemical reactions. The S_N 2 reaction is analyzed in the forward (**a**, structures extracted at time 0.0, 5.1275, and 7.5 ps) and reverse (**c**, structures extracted at time 2.725, 7.6875 and 10.0335 ps) directions, alongside the Diels-Alder reaction in the forward (**b**, structures extracted at time 7.8685, 21.6285, and 23.75 ps) and reverse (**d**, structures extracted at time 5.938, 14.3 and 15.9 ps) directions. Each panel depicts the reaction progression through representative molecular structures, projected 2D free energy maps, and the evolution of the 1D collective variable (CV). The 2D maps are defined by the C-F (d_F) and C-Cl (d_{Cl}) distances for S_N 2, and the two primary reactive C-C distances (d_4 and d_6, see SI) for Diels-Alder. In these maps, blue points denote the accumulated global dataset, while red points indicate the local dataset of the current iteration (consistent with Fig. 3). The 1D CV is defined as a linear combination of the bond descriptors (d_F + d_{Cl} for S_N 2, and d_4 + d_6 for Diels-Alder).

As shown in the upper panel, the method successfully identifies the concerted reaction pathway, in which the framework proton is transferred from the BAS to ethanol, leading to simultaneous water elimination and proton abstraction from the ethanol methyl group by an adjacent BAS acidic oxygen.

In contrast, the lower panel illustrates the stepwise mechanism, which first involves the formation of an ethoxy group bound to a BAS oxygen—a crucial intermediate in this pathway. To prevent back-crossing to the original state upon reaching this intermediate state, we restart the Loxodynamics simulation. While the Skewencoder parameters and the original CV are retained, an additional restraining wall is applied along this CV to confine the system to the current intermediate state. The simulation then proceeds with this constraint active. (For further details regarding the restart protocol, refer to the SI.) By restarting another Loxodynamics search from this intermediate, the second step of the stepwise mechanism is observed, consisting of proton abstraction from the methyl group by a vicinal BAS oxygen.

This demonstrates that our method is capable of identifying not only the most likely products but also of agnostically exploring reaction pathways that transit through intermediates with a similar probability of occurrence.

This is a fundamental aspect of studying chemical reactivity in complex catalytic systems, where experimental knowledge is often limited due to the short-lived nature of such intermediates. Nevertheless, these intermediates play a crucial role in determining the overall mechanistic network, which is essential for multiscale modeling of catalytic reactions, such as through microkinetic models.

Catalytic 1-butanol dehydration in acidic chabazite. Finally, we extend our approach to the investigation of another highly complex catalytic system involving a BAS in a zeolite framework. Specifically,

https://doi.org/10.1038/s41467-026-69586-8

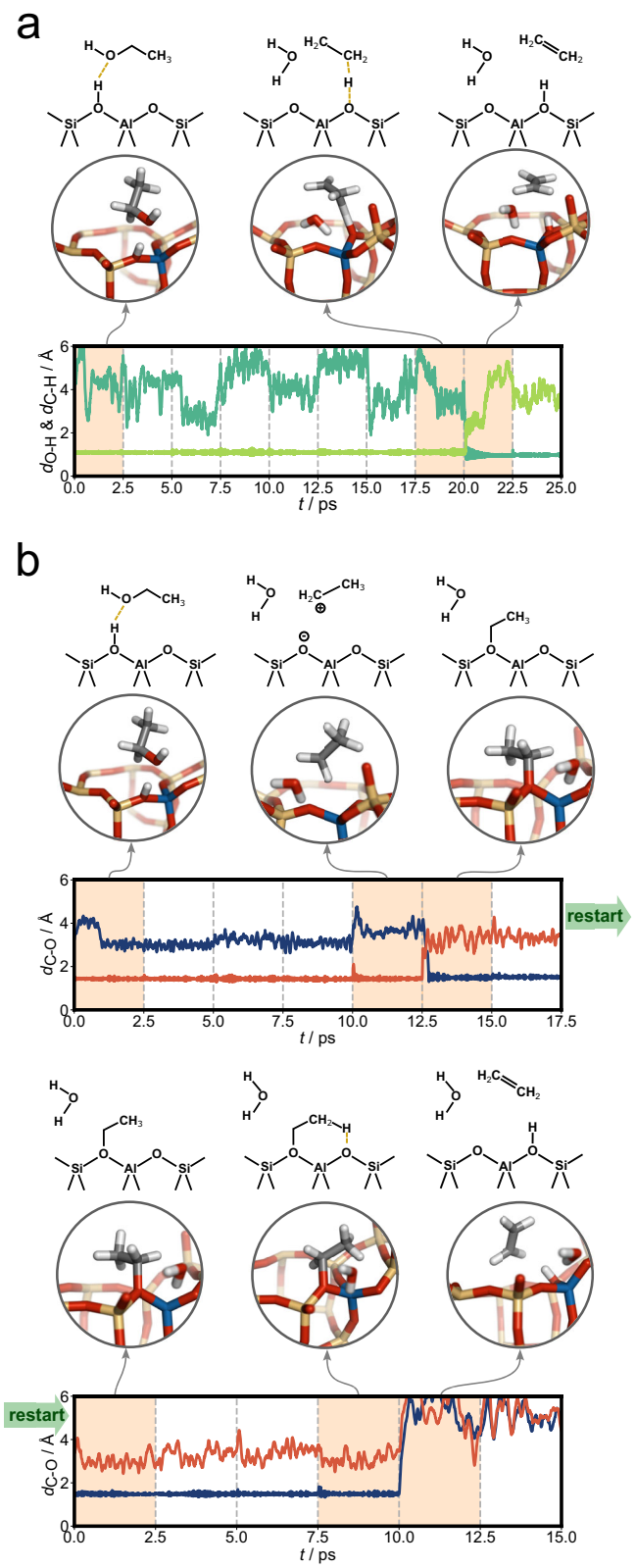

Fig. 5 | Structures and atom-pair distance evolution for ethanol dehydration in acidic chabazite (H-CHA). a shows the concerted mechanism, with deep green for $d_{O\text{-}H}$ (O(BAS)-H(CH_3) distance) and light green for $d_{C\text{-}H}$ (C(CH_3)-H(CH_3) distance). **b** depicts the stepwise mechanism, with blue for C(CH_2)-O(BAS) distance and red for C(CH_3)-O(OH) distance.

Loxodynamics was applied to study the dehydration of 1-butanol to butene at the BAS of acidic chabazite. As observed in the ethanol dehydration mechanism discussed in the section "Catalytic ethanol dehydration in acidic chabazite", both concerted and stepwise pathways are present, each kinetically accessible but characterized by slightly different energy barriers[61]. In the concerted mechanism for 1-butanol dehydration, analogous to that of ethanol, protonation of the hydroxyl group by the acidic zeolite proton promotes water elimination, while simultaneous abstraction of a hydrogen atom from the butyl group regenerates the BAS proton and restores catalytic activity. Conversely, in the stepwise pathway, formation of a butoxy intermediate occurs: initial water elimination produces a $CH_3CH_2CH_2CH_2^{+1}$ cation that associates with a negatively charged BAS oxygen to yield a relatively stable intermediate. Subsequently, abstraction of a methyl proton from this intermediate by an adjacent BAS oxygen results in oxidation and formation of an alkene double bond.

What radically differentiates the case of 1-butanol from ethanol is the wide variety of possible reaction products, arising from the position of the double bond along the alkyl chain and the associated isomerism. Both experimental studies[62] and computational investigations[63] have shown that the dehydration of 1-butanol can yield all the different dehydration products and, under conditions of high substrate loading, can also be accompanied by side reactions such as oligomerization, cracking, and aromatization. In the present work, we restrict our attention to single-site adsorption, focusing specifically on the dehydration pathways.

The simulation was initialized from the adsorption state of 1-butanol at the BAS. For exploration, the input descriptor set consisted of 28 features, encompassing interatomic distances among heavy atoms as well as selected light atom pairs, thereby minimizing dependence on prior chemical intuition (see "Methods" and SI for further details). Fewer distances were incorporated to direct system evolution toward the intended reaction space.

Figure 6 illustrates the reaction pathways identified through the Loxodynamics sample-and-search procedure. Along the direct dehydration route, the catalyst generates the three canonical products of 1-butanol dehydration: 1-butene, and the cis- and trans-isomers of 2-butene. As anticipated, 1-butene formation may also proceed via stepwise mechanisms. In close analogy to the case of ethanol, this mechanistic motif is preserved for longer alkyl chains. Notably, the stepwise route exhibits high selectivity toward 1-butene, whereas the formation of cis- and trans-2-butene is not observed.

Despite employing a descriptor set of remarkable simplicity and generality, Loxodynamics reliably identifies the butene conformer product. Furthermore, the method uncovers an intermediate corresponding to a butoxy-framework complex, which subsequently evolves into 1-butene through the expected mechanistic pathway, without branching into alternative products.

These products and pathways have previously been reported by John et al.[63] for the same reaction in zeolite H-ZSM-5, where static DFT calculations were employed to identify reactants, products, intermediates, and transition states. In their study, however, the initial guesses for all structures relied heavily on chemical intuition and required extensive manual effort, including considerable trial-and-error testing. By contrast, our approach offers a simple yet effective framework to explore complex reaction networks in highly dynamic systems such as zeolite catalysts under operando conditions.

This case study demonstrates that our approach is capable of efficiently exploring energetically favorable conformer product spaces, as well as identifying both the most probable products and reaction pathways involving intermediates with comparable likelihoods of occurrence. This capability is particularly important in investigations of chemical reactivity within complex catalytic systems, where experimental information may be scarce due to the transient

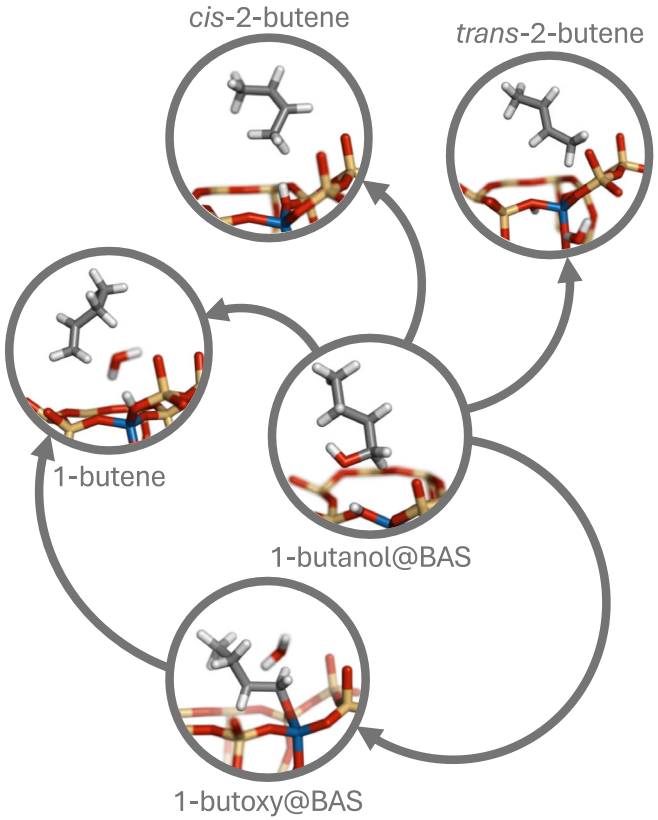

Fig. 6 | Reaction network identified by Loxodynamics for the dehydration of 1-butanol in acidic chabazite (H-CHA). The most thermodynamically favorable products, as well as the principal intermediate detected during exploration, are emphasized.

nature of intermediates and the presence of numerous conformers with distinct thermochemical properties and implications for product applications.

Discussion

In this work, we propose a method for dynamically exploring chemical reactions, catalysis, and activated processes as a whole, grounded in fundamental statistical concepts such as the asymmetry of the local probability distribution function within a free energy basin. This approach inherently accounts for finite-temperature effects, which are critical when dealing with complex systems characterized by fluxional molecular motions, dynamic rearrangements, and configurational diversity. These effects play a decisive role in determining reactivity in solvated environments and reactive interfaces. The proposed method is simple, intuitive, and highly robust. Its modular implementation ensures broad applicability across diverse systems and allows seamless integration with MD simulation codes.

This framework has the potential to drive significant advances in areas where the exploration of activated events is central, particularly in condensed-phase phenomena. These include surface-mediated processes such as heterogeneous catalysis—thermal, as exemplified by the cases presented here, electrochemical, and potentially photocatalytic—as well as selectivity-driven challenges in asymmetric, enzymatic, and supramolecular catalysis. Furthermore, the method may provide insights into other complex systems with unknown pathways, such as those found in prebiotic, atmospheric, and combustion chemistry. Conceptually, it is also applicable to non-chemical transformations like ligand-protein unbinding, protein folding, or crystal nucleation and polymorphism, offering a tool to probe transitions where the final state is not known a priori.

Moreover, this method stands to benefit substantially from the widespread availability of machine-learning interatomic potentials (MLIPs)[64–68] and related “foundation" or “general-purpose" models[69], which enable simulations of large and complex reactive systems. In this context, where models are predominantly trained on single-point structures or short equilibrium trajectories and transition regions are largely inferred rather than explicitly sampled, Loxodynamics provides a systematic strategy to actively explore these regions, identify the most informative configurations, and—after recalculation of their associated energies and forces—iteratively fine-tune the underlying model[70,71] to improve its accuracy in the transition-state configurational space. This approach offers a compelling alternative to brute-force exploration—which becomes especially intractable in catalytic systems—by enabling a physically grounded and chemically meaningful guided exploration of the most relevant metastable states at finite temperature.

We acknowledge that the current implementation relies on system-specific setups, including manually defined restraining walls and state-detection criteria, which still require limited but necessary chemical intuition and human intervention. Future work will focus on further automating these aspects, for instance through the development of an autonomous agent capable of managing the exploration phase, launching multiple exploration runs, designing and adapting biasing potentials on the fly, and automatically detecting new states while storing and organizing the associated information.

Methods

Loss function design

Details regarding the Autoencoder architecture and the selection of hyperparameters for Skewencoder are provided in the SI. This section is devoted to the discussion of the formulation and implementation of the skewness-related loss term, $\mathcal{L}_{\text{skew}}$ (defined in Equation (4)). Consider the derivative of the additional skewness-related loss function term, $\mathcal{L}_{\text{skew}}$, with respect to the NN parameters $\{\boldsymbol{\theta}\}_n := \{\{w_l^{ij}\}, \{b_l^i\}\}_n$, which include all trainable variables: the weight elements $\{w_l^{ij}\}$ from matrix $\mathbf{W}_l$ and the bias terms $\{b_l^i\}$ from vector $\mathbf{b}_l$ across all layers l in the Skewencoder. The subscript n denotes that these parameters are optimized based on the n-th batch. The gradient of $\mathcal{L}_{\text{skew}}$, as shown in Equation (5), appears in any gradient-based optimization method, ensuring the integration of skewness considerations into the training process.

$$\nabla_{\boldsymbol{\theta}_n}\mathcal{L}_{\text{skew}}(\mathbf{x}';\boldsymbol{\theta}) = \frac{d\mathcal{L}_{\text{skew}}}{d\gamma^2}\cdot\nabla_{\mathbf{s}}\gamma(\mathbf{s}(\mathbf{x}';\boldsymbol{\theta}))^2\cdot\nabla_{\boldsymbol{\theta}_n}s(\mathbf{x}';\boldsymbol{\theta}) \tag{5}$$

We calculated the derivative based on γ^2 instead of γ itself, primarily because, within the scope of Skewencoder, only the magnitude of the skewness is considered, regardless of its sign. Given the form of the skewness loss function in Equation (4), the corresponding derivative with respect to γ^2 can be computed as shown in Equation (6).

$$\begin{aligned}\frac{d\mathcal{L}_{\text{skew}}}{d\gamma^2} &= -\frac{e^{-\gamma^2}}{1+e^{-\gamma^2}}\\ &= -\frac{1}{1+e^{\gamma^2}}\end{aligned} \tag{6}$$

Since the absolute value of this term monotonically decreases as γ^2 increases, it can be regarded as a correction factor for the learning rate, fully based on skewness. When the skewness is small, the derivative value approaches $-\frac{1}{2}$, allowing for larger search steps during the early training epochs. Conversely, when the skewness is large, the derivative value approaches 0, contributing to the stable convergence of the loss function minimization. In this way, the derivative term in Equation (6) effectively acts as an exponentially decreasing learning rate, enhancing the performance of the standard optimization algorithm. The

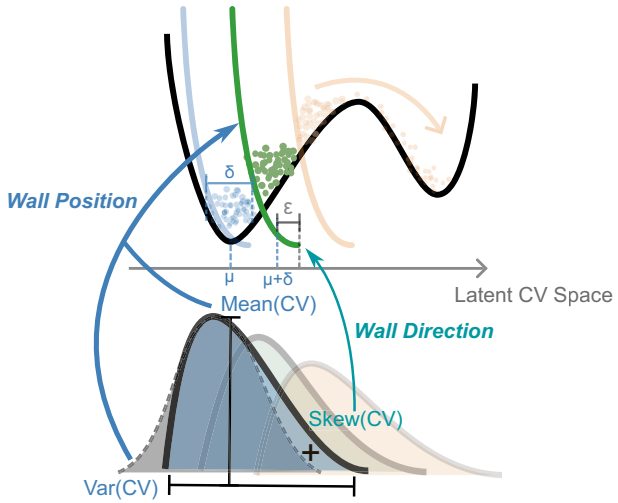

Fig. 7 | Biasing the wall design of Loxodynamics. Qualitative illustration of the influence of statistical moments on the wall configuration for a positively skewed distribution across three consecutive iterations (colored blue, green, and yellow, respectively). Half-parabolic curves denote the restraining walls, and scatter points represent the samples. The latent collective variable (CV) space (gray axis) is derived from the first iteration. The wall positions for the subsequent step (e.g., the green curve) are determined by the sample mean μ, standard deviation σ, and offset ϵ of the previous sample (blue), with the wall's direction dictated by the sample skewness (Equation (7)). The corresponding evolution of the sample distribution is also depicted, highlighting the deviation of the initial distribution (blue) from a symmetric normal distribution (gray).

effectiveness and efficiency of incorporating $\mathcal{L}_{\text{skew}}$ into the loss function are quantitatively analyzed using benchmark model potential (see SI).

Biasing wall design

Figure 7 illustrates the procedure for defining the biasing wall in latent space, based on sampling from previous iterations of the Loxodynamics simulation. At each nth iteration, the multitask Autoencoder is retrained using the state-detection restart training protocol, incorporating information from all preceding $n-1$ iterations. The resulting latent CV space is constructed, and samples from the local dataset (as defined in Figure 2) at iteration $n-1$ are projected onto this updated latent space. This projection enables the calculation of statistical moments—specifically, mean, variance, and skewness—for the sample distribution from iteration $n-1$. These statistical properties are then used to define a harmonic bias wall within this generalized coordinate system, as described in Equation (7).

$$\begin{gathered}\text{Apply wall}: \\ \kappa(s_{\text{LS}} - (\mu(\hat{s}_{\text{LS}}) + \sigma(\hat{s}_{\text{LS}}) + \epsilon))^2 \\ \text{when} \\ \text{sgn}(\gamma(\hat{s}_{\text{LS}})) \cdot s_{\text{LS}} \le \text{sgn}(\gamma(\hat{s}_{\text{LS}}))(\mu(\hat{s}_{\text{LS}}) + \sigma(\hat{s}_{LS}) + \epsilon).\end{gathered} \tag{7}$$

where s_{LS} is the CV in the latent space, κ is the energy constant for the wall, and $\mu(\hat{s}_{\text{LS}})$, $\sigma(\hat{s}_{\text{LS}})$ and $\gamma(\hat{s}_{\text{LS}})$ are the sample mean, variance and sample skewness of the local dataset projection in the latent space, respectively. The term $\text{sgn}(\gamma(\hat{s}_{\text{LS}}))$ specifies that the direction of the biasing wall in latent space is governed by the sign of the sample skewness. For positive skewness, a lower boundary—corresponding to a left-sided half-harmonic potential as depicted in Figure 7—is imposed. In contrast, negative skewness results in the application of an upper boundary. The parameter $\epsilon > 0$ is a custom offset.

As depicted in Figure 7, the inclusion of both $\sigma(\hat{s}_{\text{LS}})$ and ϵ shifts the wall within the latent space, adaptively enhancing the asymmetry of the local distribution and accelerating the search in the vicinity of local minima on the PES. However, it is important to note that the offset terms $\sigma(\hat{s}_{\text{LS}})$ and ϵ also influence the resolution of the exploration, which can be inferred from Figure 7. Values of $\epsilon \le 1.0$ are considered moderate.

Since the latent space is treated as a generalized coordinate with an implicit direction, it is crucial to determine whether the projected distribution is positively or negatively skewed and to set the wall accordingly.

General workflow

In this section, we present an overview of our methodology, summarized in the general workflow outlined in Box 1. The procedure is initialized with a small swarm of independent trajectories, each characterized by a different value of the biasing parameter κ_i. This strategy allows us to explore multiple directions efficiently and identify suitable paths for further exploration. The maximum number of allowed iterations, n_{iter}, is also specified to regulate computational cost. For each κ_i in the sequence, Loxodynamics begins by performing an unbiased simulation. During each iteration of the subsequent process, Skewencoder is trained using a state-detection-informed restart strategy (see SI for details). To demonstrate the enhanced efficiency of the warm start strategy, we compared Loxodynamics simulations utilizing this approach against exploration training initiated from scratch (see SI). The harmonic wall is configured as illustrated in Figure 7 and Equation (7), and steered MD simulations are executed accordingly. For every value of κ_i, the Loxodynamics simulation terminates either upon reaching n_{iter} or upon identification of a new state. Loxodynamics outputs not only newly discovered states but also corresponding MD trajectories and latent space CVs from Skewencoder, which may be utilized for subsequent analysis of minimum energy pathways.

Simulations settings

Model potentials. We performed Langevin dynamics using the `pesmd` tool in `PLUMED 2.9.0`[72,73], defining the potential directly in the input file. SiAll parameters are in Lennard-Jones units, with temperature 1.0, a Langevin thermostat friction coefficient of 10.0, and a time step of 0.05 for 20,000 steps.

Gas phase reactions. Simulations were performed using `CP2K 2023.1`[74], patched with `PLUMED 2.9.0`, in the NVT ensemble with a 0.5 fs time step at 300 K, controlled by a canonical sampling through velocity rescaling (CSVR) thermostat[75]. Each sampling iteration consisted of 5000 steps in the biased MD simulation.

Energy and forces were computed using the PM6[76] semiempirical Hamiltonian with a Self-Consistent Field (SCF) convergence tolerance of 1.0×10^{-5}.

The system was placed within an orthorhombic simulation cell with no periodic boundary conditions, set to 15.0 Å for S_N2 and 10.0 Å for Diels-Alder.

To detect new metastable states, bond topology changes were assessed by monitoring atomic contacts, using the function in Equation (8)[77]. The bond-dependent threshold p ($0 < p < 0.5$) distinguishes different states:

$$\begin{cases} \text{Bond formed, if } c_{i,j} > 1-p \\ \text{Bond broken, if } c_{i,j} < p \\ \text{Not stable, otherwise} \end{cases} \quad c_{i,j} = \frac{1-\left(\frac{d}{d_0}\right)^n}{1-\left(\frac{d}{d_0}\right)^m} \tag{8}$$

where d_0 represents the typical bond length between atoms of species i and j. The use of machine-learning-based classifiers that could offer a more general or automated strategy.

While the implemented state detection method is straightforward and effective for the presented case studies, complex systems may necessitate more robust approaches. State-of-the-art machine learning

BOX 1

The general workflow of Loxodynamics

Require $\{\kappa_i\}$, n_{iter}
  for each κ_i in $\{\kappa_i\}$ **do**
    iter = 0
    Run unbiased MD simulation
    While iter < n_{iter}**and** No new state is found **do**
      iter = iter + 1
      Train Skewencoder based on loss function defined in Equation (2).
      Apply harmonic wall in the latent space
      Steered MD simulation
    end while
    if New state found **then**
      Stop
    end if
  end for
  return New states, MD trajectories, CVs.

classifiers, particularly those designed for transition state detection[78], offer a generalized and automated strategy. These methods could provide the fast and precise state determination required for the warm-start training of Loxodynamics. Alternatively, for simpler organic fragments or structures that can be chemically isolated from their environment, SMILES-based fingerprinting[79] serves as a valuable tool, allowing changes in bond topology to be rapidly identified via string analysis.

Catalytic ethanol dehydration in acidic chabazite. Simulations were conducted with `CP2K 2023.1` patched with `PLUMED 2.9.0` in the NVT ensemble, using a 0.5 fs time step at 473 K, regulated by the CSVR thermostat. Each iteration comprised 5000 steps in both biased and unbiased MD simulations.

Energy and forces were computed with the semi-empirical GFN-XTB[80] method, using an SCF convergence tolerance of 1.0×10^{-5}. The simulation box was the relaxed primitive rhombohedral unit cell of natural H-CHA, with lattice constants of 9.514 Å and angles of 94.07°.

Catalytic 1-butanol dehydration in acidic chabazite. Simulations were conducted with `CP2K` 2025.1 patched with `PLUMED` 2.9.0 in the NVT ensemble, using a 0.5 fs time step at 473 K, regulated by the CSVR thermostat. Each iteration comprised 5000 steps in both biased and unbiased MD simulations.

Energy and forces were computed with the semi-empirical GFN-XTB method, using an SCF convergence tolerance of 1.0×10^{-5}. The simulation box was the relaxed primitive rhombohedral unit cell of natural H-CHA, with lattice constants of 9.514 Å and angles of 94.07°.

Data availability

Simulation inputs and generated simulation data are available at `Zenodo`[81]. More details can be found in SI. Source data are provided with this paper.

Code availability

The code for running the simulations, along with input files for reproducing the published results, is available at `Github`[82]. The implementation relies on the `mlcolvar` library as described by Bonati et al.[83]. Likewise, the `PLUMED` input files and bias interface are available at `Zenodo`[81].

References

1. Wen, M. et al. Chemical reaction networks and opportunities for machine learning. *Nat. Comput. Sci.* **3**, 12–24 (2023).
2. Unsleber, J. P. & Reiher, M. The exploration of chemical reaction networks. *Annu. Rev. Phys. Chem.* **71, 2020**, 121–142 (2020).
3. Stocker, S., Csányi, G., Reuter, K. & Margraf, J. T. Machine learning in chemical reaction space. *Nat. Commun.* **11**, 5505 (2020).
4. Zhao, Q., Xu, Y., Greeley, J. & Savoie, B. M. Deep reaction network exploration at a heterogeneous catalytic interface. *Nat. Commun.* **13**, 4860 (2022).
5. Halgren, T. A. & Lipscomb, W. N. The synchronous-transit method for determining reaction pathways and locating molecular transition states. *Chem. Phys. Lett.* **49**, 225–232 (1977).
6. Ayala, P. Y. & Schlegel, H. B. A combined method for determining reaction paths, minima, and transition state geometries. *J. Chem. Phys.* **107**, 375–384 (1997).
7. Reiher, M. & Neugebauer, J. A mode-selective quantum chemical method for tracking molecular vibrations applied to functionalized carbon nanotubes. *J. Chem. Phys.* **118**, 1634–1641 (2003).
8. Henkelman, G. & Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **111**, 7010–7022 (1999).
9. Jónsson, H., Mills, G. & Jacobsen, K. W. *Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions*, 385–404 https://www.worldscientific.com/doi/abs/10.1142/9789812839664_0016 (World Scientific, 1998).
10. Weinan, E., Ren, W. & Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **66**, 052301 (2002).
11. Behn, A., Zimmerman, P. M., Bell, A. T. & Head-Gordon, M. Efficient exploration of reaction paths via a freezing string method. *J. Chem. Phys.* **135**, 224108 (2011).
12. Zimmerman, P. Reliable transition state searches integrated with the growing string method. *J. Chem. Theory Comput.* **9**, 3043–3050 (2013).
13. Vaucher, A. C. & Reiher, M. Minimum energy paths and transition states by curve optimization. *J. Chem. Theory Comput.* **14**, 3091–3099 (2018).
14. Zimmerman, P. M. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *J. Chem. Phys.***138** (2013).

15. Habershon, S. Sampling reactive pathways with random walks in chemical space: applications to molecular dissociation and catalysis. *J. Chem. Phys.* **143**, 094106 (2015).
16. Broadbelt, L., Stark, S. & Klein, M. Computer generated reaction modelling: decomposition and encoding algorithms for determining species uniqueness. *Comput. Chem. Eng.* **20**, 113–129 (1996).
17. Matheu, D. M., Dean, A. M., Grenda, J. M. & Green, W. H. Mechanism generation with integrated pressure dependence: A new model for methane pyrolysis. *J. Phys. Chem. A* **107**, 8552–8565 (2003).
18. Gao, C. W., Allen, J. W., Green, W. H. & West, R. H. Reaction mechanism generator: Automatic construction of chemical kinetic mechanisms. *Comput. Phys. Commun.* **203**, 212–225 (2016).
19. Ohno, K. & Maeda, S. A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface. *Chem. Phys. Lett.* **384**, 277–282 (2004).
20. Maeda, S. & Ohno, K. Ab initio studies on synthetic routes of glycine from simple molecules via ammonolysis of acetolactone: applications of the scaled hypersphere search method. *Chem. Lett.* **33**, 1372–1373 (2004).
21. Maeda, S. & Ohno, K. Global mapping of equilibrium and transition structures on potential energy surfaces by the scaled hypersphere search method: applications to ab initio surfaces of formaldehyde and propyne molecules. *J. Phys. Chem. A* **109**, 5742–5753 (2005).
22. Ohno, K. & Maeda, S. Global reaction route mapping on potential energy surfaces of formaldehyde, formic acid, and their metal-substituted analogues. *J. Phys. Chem. A* **110**, 8933–8941 (2006).
23. Maeda, S., Ohno, K. & Morokuma, K. Systematic exploration of the mechanism of chemical reactions: the global reaction route mapping (grrm) strategy using the addf and afir methods. *Phys. Chem. Chem. Phys.* **15**, 3683–3701 (2013).
24. Maeda, S., Taketsugu, T., Morokuma, K. & Ohno, K. Anharmonic downward distortion following for automated exploration of quantum chemical potential energy surfaces. *Bull. Chem. Soc. Jpn.* **87**, 1315–1334 (2014).
25. Mallikarjun Sharada, S., Zimmerman, P. M., Bell, A. T. & Head-Gordon, M. Automated transition state searches without evaluating the Hessian. *J. Chem. Theory Comput.* **8**, 5166–5174 (2012).
26. Broadbelt, L. J., Stark, S. M. & Klein, M. T. Computer generated pyrolysis modeling: On-the-fly generation of species, reactions, and rates. *Ind. Eng. Chem. Res.* **33**, 790–799 (1994).
27. Habershon, S. Automated prediction of catalytic mechanism and rate law using graph-based reaction path sampling. *J. Chem. Theory Comput.* **12**, 1786–1798 (2016).
28. Simm, G. N., Vaucher, A. C. & Reiher, M. Exploration of reaction pathways and chemical transformation networks. *J. Phys. Chem. A* **123**, 385–399 (2019).
29. Steiner, M. & Reiher, M. Autonomous reaction network exploration in homogeneous and heterogeneous catalysis. *Top. Catal.* **65**, 6–39 (2022).
30. Voter, A. F. et al. Temperature-accelerated dynamics for simulation of infrequent events. *J. Chem. Phys.* **112**, 9599–9606 (2000).
31. Shim, Y., Callahan, N. B. & Amar, J. G. Localized saddle-point search and application to temperature-accelerated dynamics. *J. Chem. Phys.* **138**, 094101 (2013).
32. Krep, L. et al. Efficient reaction space exploration with Chemtrayzer-TAD. *J. Chem. Inf. Model.* **62**, 890–902 (2022).
33. Voter, A. F. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* **78**, 3908–3911 (1997).
34. Bal, K. M. & Neyts, E. C. Merging metadynamics into hyperdynamics: Accelerated molecular simulations reaching time scales from microseconds to seconds. *J. Chem. Theory Comput.* **11**, 4545–4554 (2015).
35. Kopp, W. A. et al. Automatic potential energy surface exploration by accelerated reactive molecular dynamics simulations: From pyrolysis to oxidation chemistry. *J. Phys. Chem. A* **127**, 10681–10692 (2023).
36. Wang, L.-P. et al. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **6**, 1044–1048 (2014).
37. Torrie, G. & Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
38. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **99**, 12562–12566 (2002).
39. Invernizzi, M. & Parrinello, M. Rethinking metadynamics: From bias potentials to probability distributions. *J. Phys. Chem. Lett.* **11**, 2731–2736 (2020).
40. Maeda, S. & Morokuma, K. Communications: A systematic method for locating transition structures of A+B → X type reactions. *J. Chem. Phys.* **132**, 241102 (2010).
41. Maeda, S., Taketsugu, T. & Morokuma, K. Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method. *J. Comput. Chem.* **35**, 166–173 (2014).
42. Maeda, S., Harabuchi, Y., Takagi, M., Taketsugu, T. & Morokuma, K. Artificial force induced reaction (AFIR) method for exploring quantum chemical potential energy surfaces. *Chem. Rec.* **16**, 2232–2248 (2016).
43. Sameera, W., Maeda, S. & Morokuma, K. Computational catalysis using the artificial force induced reaction method. *Acc. Chem. Res.* **49**, 763–773 (2016).
44. Park, S. & Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **120**, 5946–5961 (2004).
45. Izrailev, S. et al. in *Computational Molecular Dynamics: Challenges, Methods, Ideas*. (eds Deuflhard, P. et al.) 39–65 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1999).
46. Marchi, M. & Ballone, P. Adiabatic bias molecular dynamics: a method to navigate the conformational space of complex molecular systems. *J. Chem. Phys.* **110**, 3697–3702 (1999).
47. Celani, P., Robb, M. A., Garavelli, M., Bernardi, F. & Olivucci, M. Geometry optimisation on a hypersphere. application to finding reaction paths from a conical intersection. *Chem. Phys. Lett.* **243**, 1–8 (1995).
48. Azzalini, A. & Valle, A. D. The multivariate skew-normal distribution. *Biometrika* **83**, 715–726 (1996).
49. Chen, W. & Ferguson, A. L. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* **39**, 2079–2102 (2018).
50. Chen, W., Tan, A. R. & Ferguson, A. L. Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design. *J. Chem. Phys.* **149**, 072312 (2018).
51. Ketkaew, R. & Luber, S. Deepcv: A deep learning framework for blind search of collective variables in expanded configurational space. *J. Chem. Inf. Model.* **62**, 6352–6364 (2022).
52. Ketkaew, R., Creazzo, F. & Luber, S. Machine learning-assisted discovery of hidden states in expanded free energy space. *J. Phys. Chem. Lett.* **13**, 1797–1805 (2022).
53. Raucci, U., Rizzi, V. & Parrinello, M. Discover, sample, and refine: Exploring chemistry with enhanced sampling techniques. *J. Phys. Chem. Lett.* **13**, 1424–1430 (2022).
54. He, X. et al. Practical lessons from predicting clicks on ads at Facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising, ADKDD'14*, 1–9 (ACM, New York, NY, USA, 2014).

55. Piccini, G., McCarty, J. J., Valsson, O. & Parrinello, M. Variational flooding study of a SN2 reaction. *J. Phys. Chem. Lett.* **8**, 580–583 (2017).
56. Bi, J., Guo, X., Liu, M. & Wang, X. High effective dehydration of bio-ethanol into ethylene over nanoscale hzsm-5 zeolite catalysts. *Catal. Today* **149**, 143–147 (2010).
57. Zhang, D., Wang, R. & Yang, X. Effect of p content on the catalytic performance of p-modified hzsm-5 catalysts in dehydration of ethanol to ethylene. *Catal. Lett.* **124**, 384–391 (2008).
58. Madeira, F. F., Gnep, N., Magnoux, P., Maury, S. & Cadran, N. Ethanol transformation over HFAU, HBEA and HMFI zeolites presenting similar brønsted acidity. *Appl. Catal. A Gen.* **367**, 39–46 (2009).
59. Kim, S., Robichaud, D. J., Beckham, G. T., Paton, R. S. & Nimlos, M. R. Ethanol dehydration in hzsm-5 studied by density functional theory: Evidence for a concerted process. *J. Phys. Chem. A* **119**, 3604–3614 (2015).
60. Xia, H. Monomolecular dehydration of ethanol into ethylene over h-mor studied by density functional theory. *ACS Omega* **5**, 9707–9713 (2020).
61. Makarova, M., Paukshtis, E., Thomas, J., Williams, C. & Zamaraev, K. Dehydration of n-butanol on zeolite h-zsm-5 and amorphous aluminosilicate: detailed mechanistic study and the effect of pore confinement. *J. Catal.* **149**, 36–51 (1994).
62. Zhang, D., Al-Hajri, R., Barri, S. A. I. & Chadwick, D. One-step dehydration and isomerisation of n-butanol to iso-butene over zeolite catalysts. *Chem. Commun.* **46**, 4088–4090 (2010).
63. John, M., Alexopoulos, K., Reyniers, M.-F. & Marin, G. B. Reaction path analysis for 1-butanol dehydration in h-zsm-5 zeolite: Ab initio and microkinetic modeling. *J. Catal.* **330**, 28–45 (2015).
64. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
65. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
66. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. Schnet - a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
67. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
68. Batatia, I., Kovács, D. P., Simm, G. N. C., Ortner, C. & Csányi, G. Mace: higher order equivariant message passing neural networks for fast and accurate force fields. https://arxiv.org/abs/2206.07697 (2023).
69. Batatia, I. et al. A foundation model for atomistic materials chemistry. https://arxiv.org/abs/2401.00096 (2024).
70. Perego, S. & Bonati, L. Data efficient machine learning potentials for modeling catalytic reactivity via active learning and enhanced sampling. *Npj Comput. Mater.* **10**, 1–13 (2024).
71. Kuryla, D., Csányi, G., van Duin, A. C. T. & Michaelides, A. Efficient exploration of reaction pathways using reaction databases and active learning. *J. Chem. Phys.* **162**, 114122 (2025).
72. Bonomi, M., Bussi, G. et al. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673 (2019).
73. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. Plumed 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
74. Kühne, T. D., Iannuzzi, M. et al. CP2K: An electronic structure and molecular dynamics software package—quickstep: efficient and accurate electronic structure calculations. *J. Chem. Phys.* **152**, 194103 (2020).
75. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
76. Stewart, J. J. Optimization of parameters for semiempirical methods v: modification of nddo approximations and application to 70 elements. *J. Mol. Model.* **13**, 1173–1213 (2007).
77. Rizzi, V., Mendels, D., Sicilia, E. & Parrinello, M. Blind search for complex chemical pathways using harmonic linear discriminant analysis. *J. Chem. Theory Comput.* **15**, 4507–4515 (2019).
78. Si, Y. et al. Transition state structure detection with machine learningś. *npj Comput Mater* **11**, 199 (2025).
79. Schwartz, J., Awale, M. & Reymond, J.-L. Smifp (smiles fingerprint) chemical space for virtual screening and visualization of large databases of organic molecules. *J. Chem. Inf. Model.* **53**, 1979–1989 (2013).
80. Grimme, S., Bannwarth, C. & Shushkov, P. A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (z = 1-86). *J. Chem. Theory Comput.* **13**, 1989–2009 (2017).
81. Zhang, Z. Datasets for manuscript "exploring chemistry and catalysis by biasing skewed distributions via deep learning". https://doi.org/10.5281/zenodo.17293570 (2025).
82. Zhang, Z. Skewencoder: Source code for exploring chemistry and catalysis by biasing skewed distributions via deep learning. Zenodo https://doi.org/10.5281/zenodo.18005316 (2025).
83. Bonati, L., Trizio, E., Rizzi, A. & Parrinello, M. A unified framework for machine learning collective variables for enhanced sampling simulations: mlcolvar. *J. Chem. Phys.* **159**, 014801 (2023).

Acknowledgements

Z.Z. and G.P. acknowledge the Cluster of Excellence "The Fuel Science Center" (EXC 2186, ID: 390919832) under the Excellence Initiative of the German federal and state governments to promote science and research at German universities.Z.Z. and G.P. gratefully acknowledge Prof. Kai Leonhard for welcoming Z.Z. into his research group and for his ongoing and kind support.

Author contributions

Z.Z. and G.P. conceptualized the method. Z.Z. implemented the code and performed the simulations. Both authors analyzed the results and reviewed the manuscript. G.P. supervised the work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-69586-8.

Correspondence and requests for materials should be addressed to GiovanniMaria Piccini.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.