

# Discover, Sample and Refine: Exploring Chemistry with Enhanced Sampling Techniques

Umberto Raucci, Valerio Rizzi, and Michele Parrinello\*

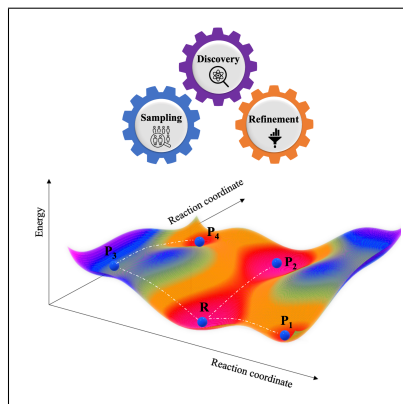
*Italian Institute of Technology, Via E. Melen 83, 16152, Genova, Italy*

E-mail: [michele.parrinello@iit.it](mailto:michele.parrinello@iit.it)

## Abstract

Over the last few decades enhanced sampling methods have made great strides. Here, we exploit this progress and propose a modular workflow for blind reaction discovery and characterization of reaction paths. Central to our strategy is the use of the recently developed explore variant of the on-the-fly probability enhanced sampling method. Like metadynamics, this method is based on the identification of appropriate collective variables. Our first step is the discovery of new chemical reactions and it is performed biasing a one dimensional collective variable derived from spectral graph theory. Once new reaction pathways are detected, we construct ad-hoc tailored neural-network based collective variables to improve sampling of specific reactions and finally we refine the results using free energy perturbation theory. Our workflow has been successfully applied to both intramolecular and intermolecular reactions. Without any chemical hypothesis, we discovered several possible products, computed the free energy surface at semiempirical level, and finally refined it with a more accurate Hamiltonian. Our workflow requires minimal user input, and thanks to its modularity and flexibility, can extend the scope of *ab initio* molecular dynamics for the exploration and characterization of reaction space.

## Graphical TOC Entry



## Keywords

reaction discovery, enhanced sampling method, collective variables, OPES

One of the main goals of chemistry is the discovery of new molecules and the determination of the chemical pathways that lead to their formation. In the past this process has been mainly driven by heuristics and specialist knowledge. However, the recent advent of machine learning techniques promises to change dramatically the way in which theoretical chemists operate and it is expected that, in the future, automated reaction discovery workflows based on machine learning methodologies will become highly relevant.

It is easy to predict that *ab initio* molecular dynamics (AIMD)<sup>1,2</sup> will play a crucial role in this context and several proposals have already been made in this respect.<sup>3-8</sup> However, the impact of AIMD has been hindered by its high computational cost. Ways have been suggested to accelerate sampling and thus make the use of AIMD practical for chemical discovery. These methods are of general nature but it is their application to chemical processes that is of interest here. Among the many methods suggested, great popularity has been gained by those that are based on the definition of a set of collective variables (CVs).<sup>9-13</sup> These are differentiable functions of the atomic coordinates  $R$  and must reflect the symmetry properties of the system such as invariance under translation, rotations and permutation. Above all, they should encode the hard-to-sample modes of the system. Their design requires an understanding of the process and their determination is often challenging.<sup>12,14-16</sup> Once the CVs are chosen, their fluctuations are then artificially enhanced by one of the enhanced sampling methods that have been suggested.

In the last few decades, much progress has been made both in sampling method efficiency and in the identification of better CVs. Our group has been active in both areas and its efforts have recently culminated in an efficient and robust sampling method that goes under the name of on-the-fly probability enhanced sampling (OPES).<sup>17,18</sup> OPES amplifies the CVs fluctuations by adding a bias from an on-the-fly estimation of the

probability distribution of the CVs, and it has been successfully applied to the study of many challenging problems such as chemical reactions, crystal nucleation and ligand binding.<sup>19-23</sup>

In parallel, like several other groups<sup>24-27</sup>, we have proposed different strategies to identify useful CVs. The use of neural network based collective variables has allowed efficient CVs to be defined as non linear combinations of a relatively large set of molecular descriptors once the initial and final states of the reactive process are known beforehand.<sup>19,28-32</sup> Thus, they are not suitable to explore in a blind way the reaction space. In this context a collective variable to be used in a reaction discovery simulation should be agnostic as to the metastable states to be discovered and at the same time efficient.

Efforts at defining such type of CVs have already been made. Lately, Grimme has used the Cartesian root-mean-square-deviation (RMSD)<sup>33</sup> as a CV to be used for chemical discovery. Earlier, Pietrucci and Andreoni have derived from spectral graph theory a set of topological coordinates called social permutation invariant coordinates (SPRINT) which, in combination with metadynamics, have allowed the phase space to be explored and several metastable states to be identified.<sup>14,34-36</sup> Here, we simplify the original SPRINT formulation and encode all the information of the molecular graph into a one dimensional CV that drives discovery-based AIMD simulations. Using OPES, we setup an integrated modular workflow consisting of three steps (Figure 1): 1) blind exploration of the possible reaction products using a low level electronic potential; 2) construction of tailored neural-network-based CVs to improve sampling of specific reactions; 3) refinement of the free energy estimation with a high level Hamiltonian using free energy perturbation theory.

*Method.* As in SPRINT,<sup>34</sup> we represent molecules as a graph whose vertices are the atoms and the edges are the chemical bonds. We associate to the graph a symmetric adjacency matrix  $\mathbf{A}$  whose elements  $a_{ij}$  indicate whether atoms  $i$  and  $j$  are linked by a

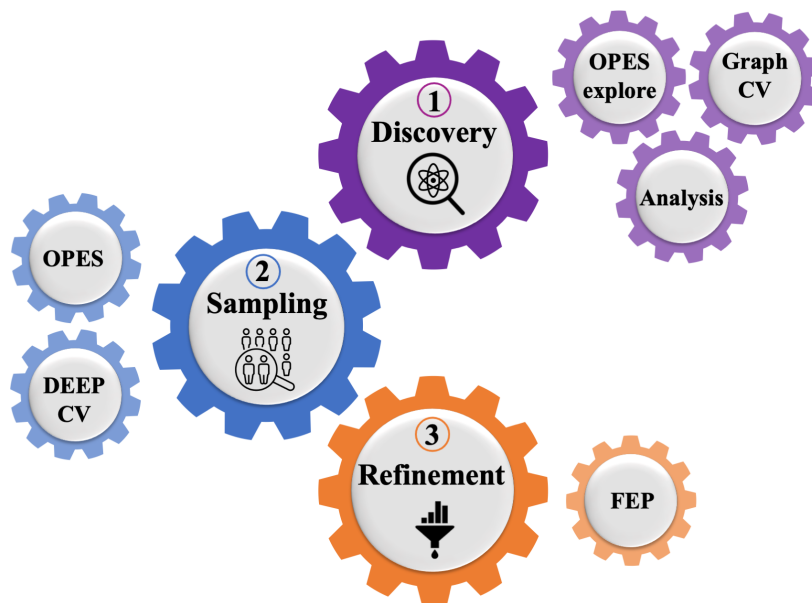


Figure 1: Schematic representation of our workflow. The first step deals with the blind exploration of the reaction space and it is followed by a sampling stage of the desired reaction path. The last step entails the FES refinement with a more accurate Hamiltonian via free energy perturbation theory (FEP).

chemical bond:

$$a_{ij} = \frac{1 - \left(\frac{r_{ij}}{\sigma_{ij}}\right)^n}{1 - \left(\frac{r_{ij}}{\sigma_{ij}}\right)^m} \quad (1)$$

where  $\sigma_{ij}$  are typical bonding lengths between atoms of species  $i$  and  $j$ . We shall use as a CV the largest eigenvalue of this matrix  $\lambda^{\max}$  which is real-valued, positive and non-degenerate. The choice of  $s = \lambda^{\max}$  as a CV guarantees translational, rotational and permutational invariance while preserving much of the chemical information encoded in the molecular graph.<sup>34</sup> Spectral graph theory ensures that it grows with the number of bonds and its value lies between the average and maximum coordination number. We depart from the original SPRINT approach in which all the components of the eigenvector corresponding to  $s = \lambda^{\max}$  were used to define a set of  $N$  topological coordinates, where  $N$  represents the total number of atoms.<sup>34</sup> While biasing SPRINT coordinates proved capable of exploring the chemical space, their multidimensional character repre-

sents a limit to their effectiveness.

We use  $\lambda^{\max}$  in combination with the enhanced sampling technique OPES that builds the bias through an on-the-fly estimation of the CVs probability distribution.<sup>17</sup> In particular, we use a variant of OPES called OPES<sub>E</sub> where the subscript stands for explore. This variant of the original approach is better suited for exploratory runs, especially if the CV is non-ideal.<sup>37</sup> Out of necessity, we use here a CV that is generic. As a consequence, we pay the price that the CV cannot be as efficient as those that are designed to accelerate a specified process. In OPES<sub>E</sub>, the probability distribution ( $p^{WT}(s)$ ) at the  $k$ th iteration is estimated to be proportional to:

$$p_n^{WT}(s) = \frac{\sum_k^n G(s, s_k)}{n} \quad (2)$$

where  $G(s, s_k)$  is multivariate Gaussian, centered at the value that the collective variable  $s$  has at the  $k$ th iteration. Then, the bias is built as:

$$V_n(s) = \frac{\gamma - 1}{\beta} \log \left( \frac{p_n^{WT}(s)}{Z_n} + \epsilon \right) \quad (3)$$

where  $Z_n$  is a normalization factor and  $\gamma > 1$  is a bias factor. A powerful feature that OPES<sub>E</sub> shares with methods such as OPES<sup>17</sup> and GAMBES<sup>38</sup> is the presence of a regularization term  $\epsilon$  which we write as  $e^{-\beta\Delta E/(1-1/\gamma)}$  which sets a limit to the maximum value of the bias  $V$  that is deposited. This allows limiting the extent of exploration by preventing high energy transition states from being visited.

The discovery engine is complemented by a reaction analysis module which allows to detect the formation of new chemical species on-the-fly. The Open Babel program<sup>39,40</sup> was used to detect chemical species at every snapshot of the trajectory, converting the cartesian coordinates of each frame into a string of ASCII characters referred to as SMILES.<sup>41</sup> When a new species is detected, its SMILES can be directly translated into the

corresponding IUPAC name employing the NCI resolver or Pubchem web services.<sup>42-44</sup>

Once the discovery phase has allowed the identification of new metastable states, the next step is the characterization of the free energy landscape of the discovered pathways. Although  $\lambda^{\max}$  shows a remarkable exploratory power, it does not necessarily identify all the slow modes involved in the specific reaction pathway. Thus, it does not represent a suitable choice to compute the free energy surface. Having in mind the spirit of the present work, we want to make the step of designing a more efficient CV as automatic as possible. We recently proposed a class of data-driven Deep-learned CVs in which the task of finding collective variables is left to a neural network, that builds efficient non-linear CVs starting from the unbiased fluctuations of a large set of molecular descriptors.<sup>19,31,32</sup> One of these methods, is the so-called Deep-LDA method that generates a CV by minimizing a loss function based on Linear Discriminant Analysis.<sup>19</sup> Feeding the neural network with all the  $N(N-1)/2$  elements of the adjacency matrix  $\mathbf{A}$ , we can build a CV that separates the reactant and product states and can be used to converge the corresponding free energy surface.

However, a large number of configurations and transition events must be sampled in order to converge the FES and the sampling in time would be computationally expensive if a higher level of theory is used. As a good compromise between accuracy and cost, we perform enhanced sampling simulations using a semiempirical Hamiltonian (PM6) and then *a posteriori* refine the FES at DFT level (B3LYP/6-31G\*\*) via free energy perturbation theory.<sup>45-48</sup> To obtain the FES at a higher-level ( $F_H$ ), we use the low-level FES ( $F_L$ ) and the perturbative term  $\Delta F_{FEP} = F_H - F_L$ :<sup>48</sup>

$$\Delta F_{FEP}(s) = \frac{1}{\beta} \log \frac{\sum_i^M w_i^O(R_i) w_i^P(R_i) \delta(s - s(R_i))}{\sum_i^M w_i^O(R_i) \delta(s - s(R_i))} \quad (4)$$

where  $\beta$  is the inverse temperature and  $M$  is the number of configurations extracted



from the low level sampling, whose potential energy is evaluated with a more accurate and costly Hamiltonian.  $w_i^P(R_i)$  represents the perturbative weight associated with the  $i$ -th sample, which can be evaluated from the potential energies computed with the low- and high-level Hamiltonians ( $U_L(R)$  and  $U_H(R)$ ), according to:

$$w_i^P(R_i) = e^{\beta(U_H(R_i) - U_L(R_i))} \quad (5)$$

$w_i^O(R_i)$  is simply the weight of the  $i$ -th sample extracted from the OPES simulation, and it can be estimated as  $w_i^O(R_i) = e^{\beta V(s(R_i), t)}$ , where  $V(s(R_i), t)$  is the OPES bias at time  $t$  for the configuration  $R_i$ .

We tested our workflow on both intramolecular and intermolecular reactions. Our intramolecular example is the Claisen rearrangement in which the allyl vinyl ether 3-ethenoxyprop-1-ene undergoes a [3,3]-sigmatropic rearrangement.<sup>49,50</sup> As an example of intermolecular reactivity, we investigated the aldol reaction between vinyl alcohol and formaldehyde.<sup>51</sup> Without making any *a priori* assumption, we discovered several products, computed the free energy surface (FES) for the main pathway at the semiempirical level (PM6<sup>52</sup>), and finally refined the FES at a higher level of theory (B3LYP<sup>53</sup>/6-31G\*\*) via free energy perturbation (FEP) theory.<sup>45-48</sup> The technical details of the calculations, as well as the time evolution of  $\lambda^{\max}$  for the discovery simulations, are reported in the Supporting Information (SI).

*Claisen Rearrangement Reaction.* Starting from 3-ethenoxyprop-1-ene (Figure 2a), we performed five discovery simulations each 500ps long at the PM6 level of theory, setting the barrier value in OPES<sub>E</sub> at 100, 200, 300, 400, and 500 kJ/mol. We included all the 14 atoms in the adjacency matrix, whose elements  $a_{ij}$  are defined in equation 1. More technical details about the simulations are reported in the SI.

Figure 2b shows the results of the discovery dynamics as a function of the OPES<sub>E</sub>

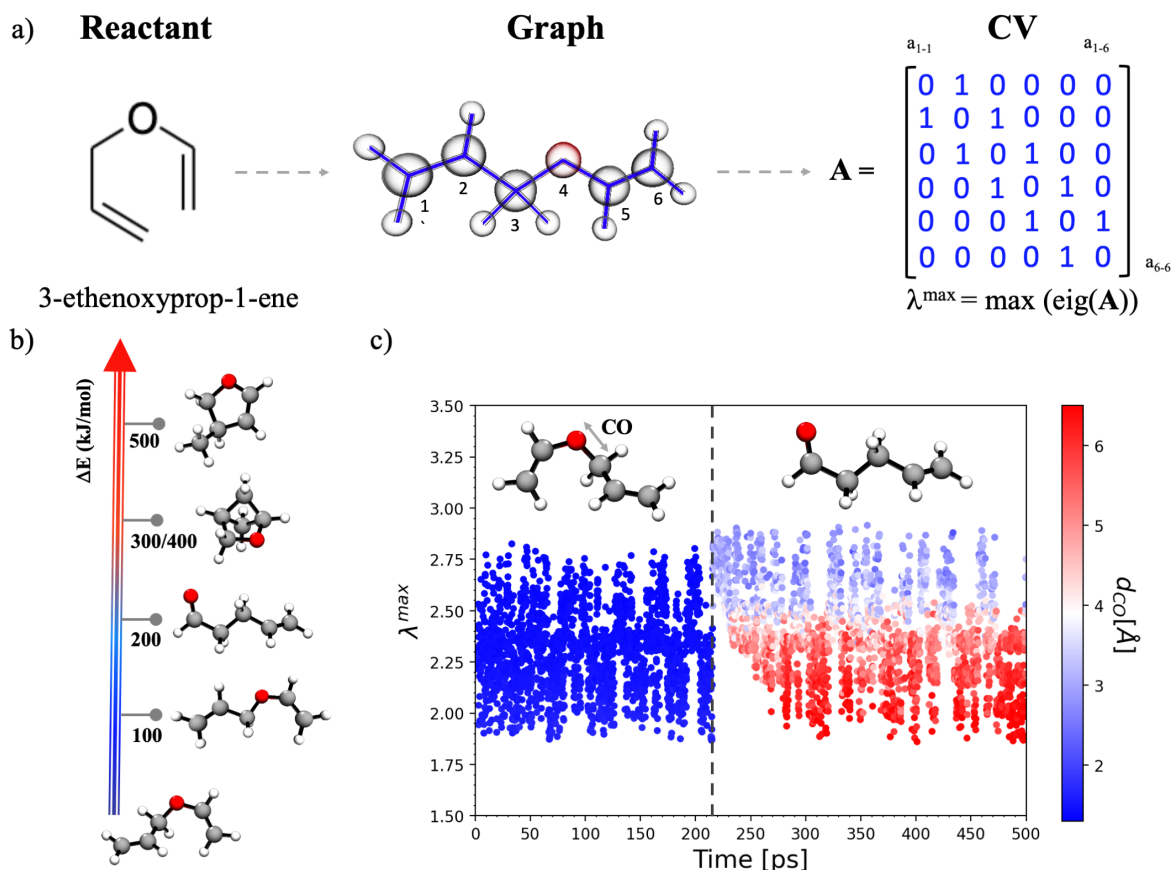


Figure 2: Summary of the discovery phase. a) The reactant 3-ethenoxyprop-1-ene is converted in the corresponding molecular graph and the adjacency matrix  $\mathbf{A}$  is defined. The CV for the discovery step is the maximum eigenvalue of  $\mathbf{A}$ . The adjacency matrix includes all the 14 atoms and its elements ( $a_{ij}$ ) are computed according to eq.1 b) Results of the discovery dynamics as function of the  $\Delta E$  cutoff in  $\text{OPES}_E$ . c) Time evolution of  $\lambda^{\max}$  colored according the CO distance of the breaking bond for the dynamics at  $\Delta E = 200$  kJ/mol.

$\Delta E$  parameter that is related with the maximum value of the bias that can be deposited. Three different products were discovered: pent-4-enal, 2-oxabicyclo[2.1.1]hexane, 3-methyl-2,3-dihydrofuran (Figure 2b). In details, starting from a  $\Delta E$  value of 100 kJ/mol, several conformers have been explored but no new product has been detected in 500ps. Increasing  $\Delta E$  to 200 kJ/mol led to the conversion of the reactant (3-ethenoxyprop-1-ene) into the expected product of the Claisen rearrangement reaction, pent-4-enal.<sup>50</sup> The two

simulations with the  $\Delta E$  value set to 300 or 400 kJ/mol led to the formation of the bicyclo compound 2-oxabicyclo[2.1.1]hexane. Finally, the 3-methyl-2,3-dihydrofuran species was found by setting  $\Delta E$  to 500 kJ/mol. All the species thus discovered are stable and their respective pathways agree well with those computed with the nudged elastic band method at the B3LYP/6-31G\*\* level of theory.

The time evolution of  $\lambda^{\max}$ , colored according to the CO distance of the breaking bond, is reported in Figure 2c for the dynamics with  $\Delta E = 200$  kJ/mol. The reaction occurs after about 210 ps, as it is visible from the variation in the CO distance. It is worth pointing out that a bias along the direction of our generic CV is able to induce a transition to a new state, even though the CV itself is far from ideal since it hardly changes its value during the dynamics. In fact, as the reactant and the product are both linear molecules, the number of contacts shows little difference in going from one species to another, which in turn is reflected into near degenerate values of the CV in the two states. Instead, when the structure of the product becomes cyclic, the two states become clearly distinguishable in the CV space (see Figure S1).

As previously discussed, once new metastable states have been discovered, we employed the Deep-LDA method to construct a specialized and more efficient CV to improve the sampling of the discovered pathways and converging the free energy surface. In particular, we computed the FES for the Claisen rearrangement reaction using as input descriptors for the neural network the 91 elements of the adjacency matrix **A**. By construction, the resulting Deep-LDA CV separates the reactant and product states (3-ethenoxyprop-1-ene and pent-4-enal) as it is clear from Figure 3b which reports the time evolution of the CV colored according to the CO distance of the breaking bond.

In a run of 20 ns, the system was reversibly driven from one basin to the other, demonstrating the efficiency of this CV. The FES thus computed is shown in Figure 3c. The free energy difference of the reaction is -58.7 kJ/mol with an average FES uncertainty

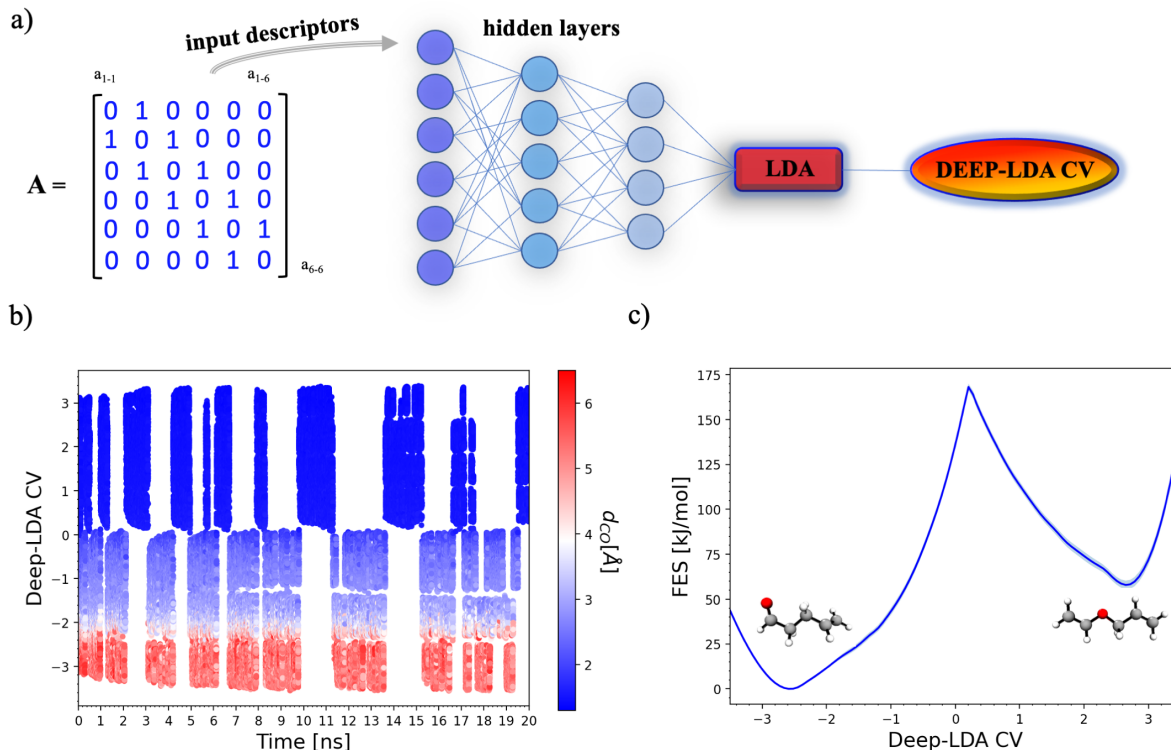


Figure 3: Results of the OPES explore simulation when enhancing the Deep-LDA CV. a) Schematic representation of the workflow to build the Deep-LDA CV. More details can be found in Ref.<sup>19</sup>. All the 91  $a_{ij}$  elements of the adjacency matrix  $A$  are used as input descriptors and they have been computed according eq.1. b) Time evolution of the CV coloured according to the CO distance of the breaking bond. c) Free-energy surface along the Deep-LDA CV, computed with a block average over the last 15 ns of the simulation.

of 1.7 kJ/mol computed by a block average.

In the final step, we refined the free energy estimation using free energy perturbation theory. To build the free energy surface at DFT level of theory we extracted 5000 configurations from the last 15 ns of the PM6 trajectory, on which B3LYP/6-31G\*\* single points were carried out. These structures were selected to represent a uniform distribution in the Deep-LDA CV space (see Figure S2). The refined FES computed according to equations 4 and 5 is reported in Figure 4. The free energy difference of the Claisen rearrangement reaction at B3LYP level is -90.4 kJ/mol.

*Aldol Reaction.* As an example representative of an intermolecular reaction, we in-

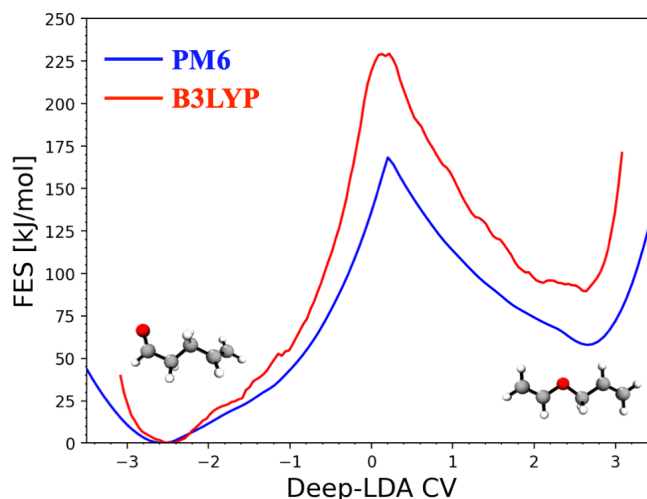


Figure 4: Free energy along the Deep-LDA CV calculated with the semiempirical Hamiltonian (PM6) and then refined at B3LYP/6-31G\*\* level of theory via the free energy perturbation term  $\Delta F_{FEP}$ .

investigated the aldol reaction between vinyl alcohol and formaldehyde.<sup>51</sup> Five 500 ps long dynamics simulations were performed at the PM6 level of theory setting the barrier value in OPES<sub>E</sub> at 100, 150, 200, 250, and 300 kJ/mol, including all the atoms in the adjacency matrix.

Results of the discovery simulations are reported in Figure 5a. Starting from a  $\Delta E$  value of 100 kJ/mol, formaldehyde and vinyl alcohol react leading to the expected aldol 3-hydroxypropanal. The same product is observed when setting the barrier value to 150 and 200 kJ/mol. Increasing  $\Delta E$  to 250 kJ/mol led first to 3-hydroxypropanal, which subsequently gave an intramolecular cyclization to oxetan-2-ol (Figure 5b). Finally, ethoxymethanol was discovered by increasing the barrier to 300 kJ/mol.

Once new metastable states were discovered, we followed our protocol. As in the case of the Claisen rearrangement reaction we built a Deep-LDA CV out of all the interatomic distances and converged the FES of the aldol reaction at the PM6 level.<sup>19</sup> We then extracted 5000 configurations and refined the FES at the B3LYP/6-31G\*\* level of theory (Figure S4).

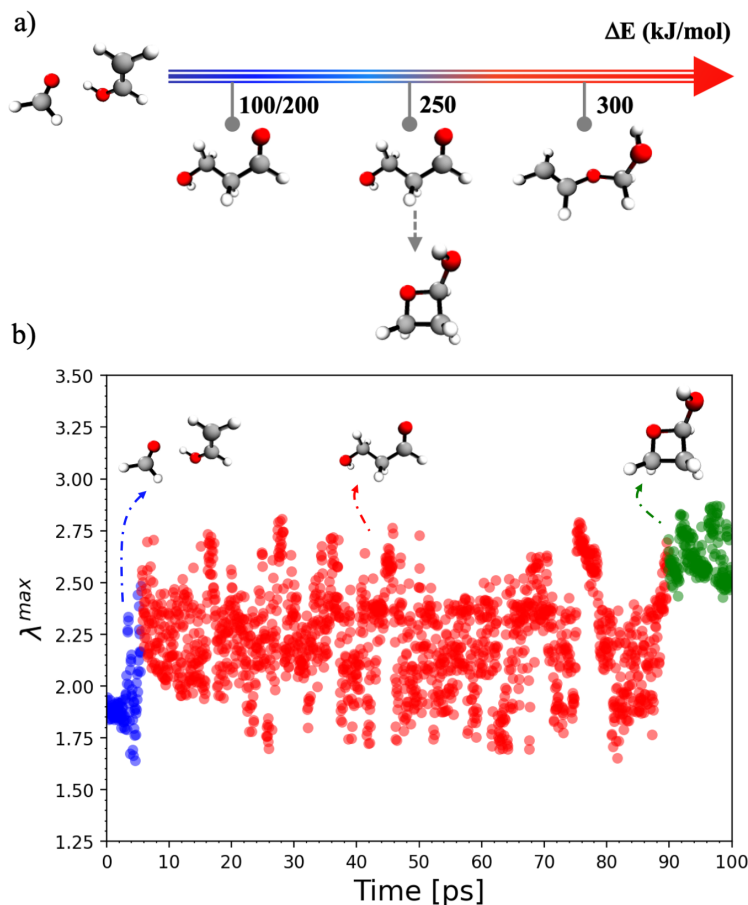


Figure 5: Summary of the discovery phase. a) Results of the discovery dynamics as function of the  $\Delta E$  cutoff in OPES explore. b) Time evolution of  $\lambda^{\max}$  coloured according to the associated SMILES for the dynamics at  $\Delta E = 250$  kJ/mol.

*Conclusions.* Using state-of-the-art enhanced simulation techniques, we have presented an integrated modular workflow to discover and characterize reaction pathways. This workflow is highly modular and flexible. Reaction discovery can be either used as a stand-alone tool to guess complex chemical pathways<sup>8</sup> or be followed by free energy calculations where it can be combined with more refined CVs, such as Deep-LDA,<sup>19</sup> Deep-TDA,<sup>32</sup> or Deep-TICA<sup>31</sup>. Free energy results can be further refined by free energy perturbation techniques<sup>48,54</sup> to obtain estimates at a higher level of theory. The approach can also be combined with neural network-based potentials that offer DFT-level preci-

sion at an affordable cost.<sup>55-57</sup> By a careful and focused choice of the adjacency matrix, our method can be extended to study chemistry in solution, on surfaces and in the solid state. In the future, as the workflow will evolve and become more automatic with a reduction of user input, we are hopeful that it will develop into a powerful tool for reaction discovery.

## Acknowledgement

The authors thank Davide Mandelli, Andrea Rizzi and Enrico Trizio for carefully reading the paper; and Luigi Bonati for useful discussions and feedback on the manuscript.

## Supporting Information Available

Computational details about the simulations; time evolution of  $\lambda^{\max}$  for the discovery simulations of the Claisen rearrangement and aldol reaction; free energy surface at low and high level of theory for the aldol reaction.

## References

- (1) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, 55, 2471.
- (2) Marx, D.; Hutter, J. *Ab initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press, 2009.
- (3) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering Chemistry with an ab initio Nanoreactor. *Nat. Chem.* **2014**, 6, 1044–1048.

- (4) Wang, L.-P.; McGibbon, R. T.; Pande, V. S.; Martinez, T. J. Automated Discovery and Refinement of Reactive Molecular Dynamics Pathways. *J. Chem. Theory Comput.* **2016**, *12*, 638–649.
- (5) Harvey, M. J.; De Fabritiis, G. High-Throughput Molecular Dynamics: the Powerful New Tool for Drug Discovery. *Drug Discovery Today* **2012**, *17*, 1059–1062.
- (6) Pieri, E.; Lahana, D.; Chang, A. M.; Aldaz, C. R.; Thompson, K. C.; Martínez, T. J. The Non-Adiabatic Nanoreactor: Towards the Automated Discovery of Photochemistry. *Chem. Sci.* **2021**,
- (7) Zeng, J.; Cao, L.; Xu, M.; Zhu, T.; Zhang, J. Z. Complex Reaction Processes in Combustion Unraveled by Neural Network-Based Molecular Dynamics Simulation. *Nat. Commun.* **2020**, *11*, 1–9.
- (8) Rizzi, V.; Mendels, D.; Sicilia, E.; Parrinello, M. Blind Search for Complex Chemical Pathways Using Harmonic Linear Discriminant Analysis. *J. Chem. Theory Comput.* **2019**, *15*, 4507–4515.
- (9) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12562–12566.
- (10) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: a Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
- (11) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826–843.
- (12) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem.* **2016**, *67*, 159–184.



- (13) Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy Landscapes. *Nat. Rev. Phys.* **2020**, *2*, 200–212.
- (14) Pietrucci, F. Strategies for the Exploration of Free Energy Landscapes: Unity in Diversity and Challenges Ahead. *Rev. Phys.* **2017**, *2*, 32–45.
- (15) Rogal, J. Reaction Coordinates in Complex Systems-a Perspective. *Eur. Phys. J. B* **2021**, *94*, 1–9.
- (16) Chiariello, M. G.; Bolnykh, V.; Ippoliti, E.; Meloni, S.; Olsen, J. M. H.; Beck, T.; Rothlisberger, U.; Fahlke, C.; Carloni, P. Molecular Basis of CLC Antiporter Inhibition by Fluoride. *J. Am. Chem. Soc.* **2020**, *142*, 7254–7258.
- (17) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *J. Phys. Chem. Lett.* **2020**, *11*, 2731–2736.
- (18) Invernizzi, M.; Piaggi, P. M.; Parrinello, M. Unified Approach to Enhanced Sampling. *Phys. Rev. X* **2020**, *10*, 041034.
- (19) Bonati, L.; Rizzi, V.; Parrinello, M. Data-Driven Collective Variables for Enhanced Sampling. *J. Phys. Chem. Lett.* **2020**, *11*, 2998–3004.
- (20) Ansari, N.; Rizzi, V.; Carloni, P.; Parrinello, M. Water-Triggered, Irreversible Conformational Change of SARS-CoV-2 Main Protease on Passing from the Solid State to Aqueous Solution. *J. Am. Chem. Soc.* **2021**, *143*, 12930–12934.
- (21) Karmakar, T.; Invernizzi, M.; Rizzi, V.; Parrinello, M. Collective Variables for the Study of Crystallisation. *Mol. Phys.* **2021**, e1893848.
- (22) Rizzi, V.; Bonati, L.; Ansari, N.; Parrinello, M. The Role of Water in Host-Guest Interaction. *Nat. Commun.* **2021**, *12*, 1–7.

- (23) Piaggi, P. M.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Car, R. Phase Equilibrium of Water with Hexagonal and Cubic Ice Using the SCAN Functional. *J. Chem. Theory Comput.* **2021**, *17*, 3065–3077.
- (24) Rydzewski, J.; Valsson, O. Multiscale Reweighted Stochastic Embedding: Deep Learning of Collective Variables for Enhanced Sampling. *J. Phys. Chem. A* **2021**, *125*, 6286–6302.
- (25) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Machine Learning Approaches for Analyzing and Enhancing Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 139–145.
- (26) Chiariello, M. G.; Alfonso-Prieto, M.; Ippoliti, E.; Fahlke, C.; Carloni, P. Mechanisms Underlying Proton Release in CLC-type F<sup>-</sup>/H<sup>+</sup> Antiporters. *J. Phys. Chem. Lett.* **2021**, *12*, 4415–4420.
- (27) Raniolo, S.; Limongelli, V. Ligand Binding Free-Energy Calculations with Funnel Metadynamics. *Nat. Protoc.* **2020**, *15*, 2837–2866.
- (28) Wehmeyer, C.; Noé, F. Time-lagged Autoencoders: Deep learning of Slow Collective Variables for Molecular Kinetics. *J. Chem. Phys.* **2018**, *148*, 241703.
- (29) 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.* **2018**, *39*, 2079–2102.
- (30) Rogal, J.; Schneider, E.; Tuckerman, M. E. Neural-Network-Based Path Collective Variables for Enhanced Sampling of Phase Transformations. *Phys. Rev. Lett.* **2019**, *123*, 245701.

- (31) Bonati, L.; Piccini, G.; Parrinello, M. Deep Learning the Slow Modes for Rare Events Sampling. *Proc. Natl. Acad. Sci.* **2021**, *118*.
- (32) Trizio, E.; Parrinello, M. From Enhanced Sampling to Reaction Profiles. *J. Phys. Chem. Lett.* **2021**, *12*, 8621–8626.
- (33) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.
- (34) Pietrucci, F.; Andreoni, W. Graph Theory Meets ab initio Molecular Dynamics: Atomic Structures and Transformations at the Nanoscale. *Phys. Rev. Lett.* **2011**, *107*, 085504.
- (35) Pietrucci, F.; Andreoni, W. Fate of a Graphene Flake: A New Route toward Fullerenes Disclosed with ab initio Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 913–917.
- (36) Balan, E.; Pietrucci, F.; Gervais, C.; Blanchard, M.; Schott, J.; Gaillardet, J. First-Principles Study of Boron Speciation in Calcite and Aragonite. *Geochim. Cosmochim. Acta* **2016**, *193*, 119–131.
- (37) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *arXiv preprint arXiv:1909.07250* **2019**,
- (38) Debnath, J.; Parrinello, M. Gaussian Mixture-Based Enhanced Sampling for Statics and Dynamics. *J. Phys. Chem. Lett.* **2020**, *11*, 5076–5080.
- (39) O’Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3*, 1–14.
- (40) <http://openbabel.org>.

- (41) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28*, 31–36.
- (42) <https://cactus.nci.nih.gov/chemical/structure>.
- (43) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B. et al. PubChem 2019 Update: Improved Access to Chemical Data. *Nucleic Acids Res.* **2019**, *47*, D1102–D1109.
- (44) <https://pubchem.ncbi.nlm.nih.gov>.
- (45) Chipot, C.; Pohorille, A. Free Energy Calculations. *Springer Ser. Chem. Phys.* **2007**, *86*, 159–184.
- (46) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (47) Li, P.; Jia, X.; Pan, X.; Shao, Y.; Mei, Y. Accelerated Computation of Free Energy Profile at ab initio Quantum Mechanical/Molecular Mechanics Accuracy via a Semi-Empirical Reference Potential. I. Weighted Thermodynamics Perturbation. *J. Chem. Theory Comput.* **2018**, *14*, 5583–5596.
- (48) Piccini, G.; Parrinello, M. Accurate Quantum Chemical Free Energies at Affordable Cost. *J. Phys. Chem. Lett.* **2019**, *10*, 3727–3731.
- (49) Krenske, E. H.; Burns, J. M.; McGeary, R. P. Claisen Rearrangements of Benzyl Vinyl Ethers: Theoretical Investigation of Mechanism, Substituent Effects, and Regioselectivity. *Org. Biomol. Chem.* **2017**, *15*, 7887–7893.
- (50) Martín Castro, A. M. Claisen Rearrangement over the Past Nine Decades. *Chem. Rev.* **2004**, *104*, 2939–3002.

- (51) Kurti, L.; Czakó, B. *Strategic Applications of Named Reactions in Organic Synthesis*; Elsevier, 2005.
- (52) Stewart, J. J. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.
- (53) Becke, A. D. A New Mixing of Hartree–Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- (54) Rizzi, A.; Carloni, P.; Parrinello, M. Targeted Free Energy Perturbation Revisited: Accurate Free Energies from Mapped Reference Potentials. *J. Phys. Chem. Lett.* **2021**, *12*, 9449–9454.
- (55) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- (56) Zhang, L.; Han, J.; Wang, H.; Car, R.; Weinan, E. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001.
- (57) Yang, M.; Bonati, L.; Polino, D.; Parrinello, M. Using Metadynamics to Build Neural Network Potentials for Reactive Events: The Case of Urea Decomposition in Water. *Catal. Today* **2021**,