

Cite this: *Chem. Sci.*, 2025, 16, 10503

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

# Symmetry-constrained generation of diverse low-bandgap molecules with Monte Carlo tree search†

Akshay Subramanian,<sup>a</sup> James Damewood,<sup>a</sup> Juno Nam,<sup>a</sup> Kevin P. Greenman,<sup>b</sup> Avni P. Singhal<sup>a</sup> and Rafael Gómez-Bombarelli<sup>a\*</sup>

Organic optoelectronic materials are a promising avenue for next-generation electronic devices due to their solution processability, mechanical flexibility, and tunable electronic properties. In particular, near-infrared (NIR) sensitive molecules have unique applications in night-vision equipment and biomedical imaging. Molecular engineering has played a crucial role in developing non-fullerene acceptors (NFAs) such as the Y-series molecules, which feature a rigid fused-ring electron donor core flanked by electron-deficient end groups, leading to strong intramolecular charge-transfer and extended absorption into the NIR region. However, systematically designing molecules with targeted optoelectronic properties while ensuring synthetic accessibility remains a challenge. To address this, we leverage structural priors from domain-focused, patent-mined datasets of organic electronic molecules using a symmetry-aware fragment decomposition algorithm and a fragment-constrained Monte Carlo Tree Search (MCTS) generator. Our approach generates candidates that retain symmetry constraints from the patent dataset, while also exhibiting red-shifted absorption, as validated by TD-DFT calculations.

Received 23rd December 2024  
Accepted 1st May 2025

DOI: 10.1039/d4sc08675a

rsc.li/chemical-science

## 1 Introduction

Organic optoelectronic materials have emerged as promising candidates for next-generation electronic devices, particularly in the fields of organic photodiodes (OPDs) and organic photovoltaics (OPVs). These materials offer unique advantages such as solution processability, mechanical flexibility, and tunable electronic properties, making them attractive alternatives to traditional inorganic semiconductors.<sup>1–3</sup> Of particular interest are low-bandgap  $\pi$ -conjugated organic molecules and polymers sensitive to near-infrared (NIR) wavelengths, which can harvest a broader portion of the solar spectrum, extending beyond the visible range. These NIR-sensitive materials have found diverse applications, from military night-vision equipment to biomedical imaging devices and semi-transparent photovoltaics.<sup>4–7</sup>

The optimization of molecular structures, especially the design of improved donor–acceptor (D–A) systems, has led to substantial progress in the power conversion efficiency (PCE) of OPV devices. Over the past decade, PCE has increased from around 5% to over 19% for single-junction devices.<sup>8</sup> This advancement is largely attributed to the development of non-fullerene acceptors (NFAs), which offer more tunability of

energy levels and absorption spectra through molecular engineering, than fullerene acceptors (FAs).<sup>9</sup>

The Y-series family of NFAs (Y1 through Y6) is a particularly promising class of molecules,<sup>10</sup> typically featuring a planar, rigid fused-ring electron-rich core with appended electron-deficient (acceptor) end units. This A–D–A architecture induces strong intramolecular charge-transfer (ICT) effects, resulting in broad absorption that extends well into the NIR region. Over time, several follow-up designs and derivatives of the Y-series have emerged,<sup>11</sup> further tuning the side-chains, end groups, and bridge units to achieve improved performance and device stability. These improvements highlight the role of molecular design in enhancing device performance and the potential for further advancements through innovative approaches to materials discovery.

Over the past half-decade, generative models have received tremendous attention in molecular design,<sup>12,13</sup> using a variety of algorithms and molecular representations.<sup>14–21</sup> However, without special attention to synthetic accessibility, off-the-shelf models often generate non-synthesizable designs.<sup>22</sup> Several existing approaches incorporate knowledge about synthesizability into generative models, either implicitly or explicitly.

Implicit approaches typically work by (1) biasing the objective function for property optimization with an additional synthesizability term, (2) applying a synthesizability filter post-hoc to generated candidates, or (3) constraining the training dataset using combinatorial enumerations of molecular fragments (for example, symmetric attachments as a proxy for synthetic accessibility). 1 and 2 often use approximate score functions

<sup>a</sup>Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. E-mail: rafagb@mit.edu

<sup>b</sup>Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4sc08675a>



such as SAScore,<sup>23</sup> SCScore,<sup>24</sup> or retrosynthesis tools like ASKCOS<sup>25</sup> or AiZynthFinder<sup>26</sup> to quantify synthesizability. In 1, multi-objective optimization is sensitive to the choice of the loss function's functional form, which if not done appropriately can lead to a compromise in other property objectives. In 2, a key assumption is that a sufficiently large portion of generated candidates are synthesizable, an assumption often proven false, as demonstrated in Fig. S10† of Gao and Coley.<sup>22</sup> Additionally, the scoring functions themselves have specific definitions that make them more suitable for certain situations than others, limiting their universality. SAScore for instance is designed for biologically relevant molecules such as drug-like molecules, and might be less applicable for other chemical spaces such as organic electronics.<sup>23</sup> In 3, a downside is that the heuristics used to design the training dataset might not be reflected in the generations unless more explicit constraints are injected into the models. Such explicit constraints can be achieved by restricting the grammar of the generative model to only a chemical subspace using user-defined heuristics, so that all generated candidates by construction satisfy the constraints.

Explicit methods of incorporating synthesizability do not suffer from limitations of scoring functions, or generations not satisfying prior constraints. They typically either embed pre-specified synthetic reaction templates,<sup>27–30</sup> or machine-learned reaction prediction algorithms<sup>31,32</sup> into the generation grammar to constrain the scope of generation. While both reduce the possibility of chemically infeasible outcomes, reaction templates significantly limit the breadth of the search space, and reaction prediction algorithms could be prone to error propagation in multi-step synthetic routes.<sup>22</sup> This case of multi-step pathways is common in organic electronics systems.<sup>33</sup> Some recent works at the intersection of organic electronics and synthesizability include Kwak *et al.*,<sup>34</sup> Li and Tabor,<sup>35</sup> Westermayr *et al.*<sup>36</sup>

In this paper, we develop a simple, and more universally applicable approach to incorporate an approximate form of synthesizability into the generation process, and demonstrate its utility for organic electronics applications by generating low-bandgap molecules. Building on our previous work on extracting datasets from published patents, our method is based on the fact that molecules reported in published patents are likely to be synthetically accessible, and when extracted based on domain-focused keyword queries made on the text, are also high-performance for domain-relevant properties.<sup>37</sup> The chemical fragments that make up a molecule, and the symmetry in their attachments contain key information about the precursors and synthetic pathways used to make the molecule. To utilize this information effectively during generation, we develop:

(1) A fragment decomposition algorithm that extracts fragments and their reactive positions from patent-mined structures, preserving symmetry information.

(2) A generation algorithm based on Monte Carlo Tree Search (MCTS)<sup>38–40</sup> that constrains the generated molecules to contain only the previously extracted fragments attached at their reactive positions, while satisfying symmetry requirements.

We evaluate our method on two chemical spaces of different sizes, a smaller, focused space of Y6-derivatives,<sup>10</sup> and a much larger space of patent-mined fragment derivatives.<sup>37</sup> We employ Time-Dependent Density Functional Theory (TD-DFT) to obtain our “ground-truth” bandgaps for training and evaluation. Our generated candidates retain symmetry constraints from the patent dataset while also exhibiting red-shifted absorption.

## 2 Methods

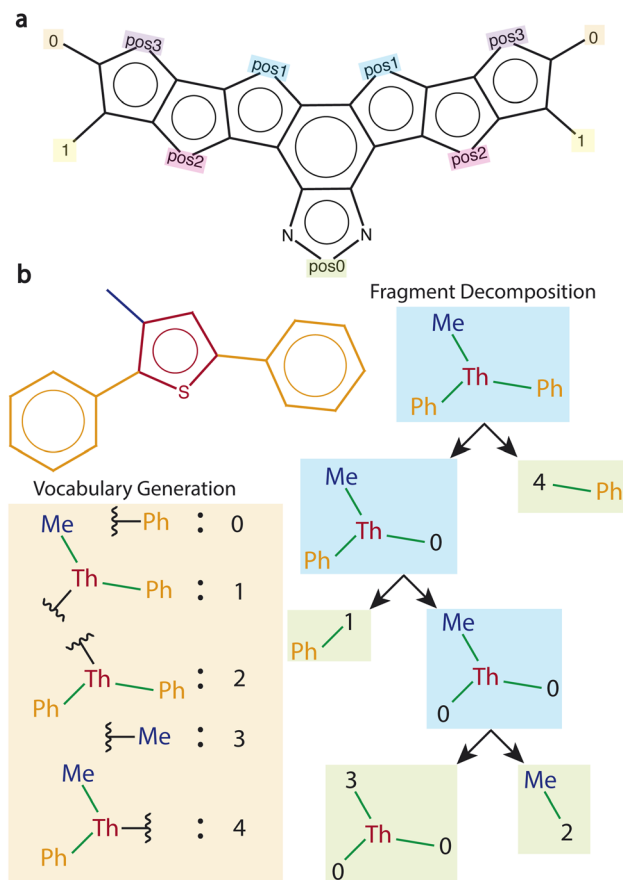
### 2.1 Fragment identity, connectivity and symmetry

Our previous work developed an automated pipeline to mine USPTO patents and create domain-focused unsupervised datasets of molecular structures.<sup>37</sup> These patent-derived molecules likely possess two important characteristics: high performance (justifying the investment in a patent application) and synthetic accessibility. Consequently, the fragments present in these structures, along with their reactive positions (points of connectivity), potentially correlate with both synthetic accessibility and domain-specific properties. Moreover, the symmetry of reactive positions relative to each other encodes information about the synthetic pathways used to create these molecules. To effectively leverage this structural and synthetic information, we have developed a two-stage method: Fragment Decomposition: We designed an algorithm that extracts fragments and their reactive positions from patent-mined structures, while marking each reactive position in a way that faithfully represents the identity of the substructure attached to it. This preserves important contextual information about molecular connectivity. It is this connectivity and identity-preserving property that we refer to as “symmetry” throughout this paper. Constrained Generation: We implemented a generation algorithm based on MCTS that constrains the output to molecules composed solely of patent-extracted fragments. These fragments are attached at their corresponding reactive positions while adhering to the symmetry requirements observed in the original patent structures. This approach aims to maintain synthetic feasibility while exploring novel molecular configurations. By combining these two components, we aim to create a framework for generating molecules that are not only likely to be synthetically accessible but also tailored to the specific property requirements of organic electronic materials. This method leverages the information embedded in patent-derived structures to guide the exploration of chemical space towards promising and realizable candidates.

**2.1.1 Fragment decomposition.** We developed a fragment decomposition algorithm that preserves connectivity and symmetry information from patent molecules. We achieve this through a recursive application of RDKit reaction SMARTS.<sup>41</sup> More details are provided in the following paragraphs, and an illustrative example is shown in Fig. 1(b).

A reaction SMARTS string is first defined to indicate what types of bonds must be deleted during decomposition. For our purpose, we defined this so that single bonds connecting an aromatic ring with an aliphatic chain, or an aromatic ring to another aromatic ring are matched. The algorithm consists of





**Fig. 1** Fragments to initialize MCTS (a) editable Y6 core with marked positions. (b) Fragment decomposition algorithm to obtain fragments from patent dataset. Vocabulary dictionary is first created by breaking one-bond at a time, and labeling the resulting fragments with unique integer values. Recursive decomposition of the starting molecule is then performed. Reactive positions are labeled with values corresponding to the broken fragment's identity. The leaf nodes (shaded in green) represent the final set of fragments obtained after decomposition. Me, Th, and Ph represent methyl, thiophene, and phenyl groups, respectively.

two steps: (1) creation of fragment vocabulary, (2) fragment decomposition and assignment of reactive sites.

In step 1, the reaction SMARTS is applied to the molecule to get all possible fragments that are obtainable from deletion of one bond in the molecule. The vocabulary of unique fragments is created, and each is assigned a unique ID. For the IDs, we use numbers  $[0, N]$  where  $N$  is the number of fragments in the vocabulary. Details on how this is actually represented in our code implementation are given in ESI Section 5.†

In step 2, the original molecule is recursively decomposed with one-bond deletions to form a binary tree. After every application of the reaction SMARTS, the reactive positions (positions at which bond deletion was performed) on the fragment are labeled using the ID of the fragment which was attached before the bond was broken. This preserves information about the identity of the fragment which was connected at the reactive position. The two fragments are then checked for a match with the reaction SMARTS to see if further

decomposition is possible. If possible, the same series of steps described above is performed on the fragment. Otherwise, the fragment is a leaf node on the binary tree, and is added to the list of final fragments.

Finally, memoization<sup>42</sup> can be used to prevent redundant decompositions of the same fragment. We memoize computations in a hash table during tree creation, and reuse previously performed computations. Fig. 1 illustrates this process of fragment decomposition on an example molecule.

**2.1.2 Generation with MCTS.** We frame the problem of designing a molecule as a series of building block/molecular fragment choices. MCTS is an efficient method of navigating large exploration spaces, that learns to optimize a chosen reward function by iteratively updating value estimates for action choices in the tree.<sup>38–40,43</sup> Since our objective is to design low-bandgap molecules, our reward function of choice is the negative bandgap as predicted by Chemprop,<sup>44</sup> a message-passing neural network (MPNN) architecture that we trained on labels obtained from TD-DFT calculations. While more details about TD-DFT calculations are provided in Section 2.4.1, we note here that our use of the term “bandgap” throughout this paper refers to the lowest-energy singlet vertical excitation energies. Each iteration of the MCTS algorithm consists of four sequential steps: selection, expansion, rollout, and back-propagation. The selection step starts with the root node and traverses the already explored portions of the tree to reach a leaf node. In our work, the path from the root node to the leaf node represents the construction of a molecular structure, where each node corresponds to a partial or complete structure. We follow the standard Upper Confidence Bound applied to Trees (UCT) policy<sup>38</sup> to choose an action at each timestep of traversal.

$$a_t = \operatorname{argmax}_{a \in A_t} \left[ Q_{t-1}(a) + c \sqrt{\frac{\log(N_{t-1}^p)}{N_{t-1}(a)}} \right],$$

where  $a_t$  is the action at timestep  $t$ ,  $A_t$  is the set of action choices available at timestep  $t$ ,  $Q_{t-1}(a)$  is the action value function of action  $a$  before timestep  $t$ ,  $N_{t-1}(a)$  is the number of times the action  $a$  has been chosen before timestep  $t$ , and  $N_{t-1}^p$  is the number of times the parent node has been visited before timestep  $t$ . The exploration coefficient  $c$  controls the exploration-exploitation trade-off, *i.e.*, the larger its value, the higher the exploration.

If the leaf node is not a terminal action, the expansion step enumerates possible next states and expands the tree with new nodes, one of which is then chosen at random. In the rollout step, a random policy is followed until the molecular generation process is terminated. Finally, the quality of the end state is evaluated and is assigned a scalar reward value. The reward is backpropagated through the tree branch to update the value functions of each visited node.

A detailed description of the Markov Decision Process defined for MCTS is given in Section 2.3. The Y6 case does not involve patent-mined fragments, and hence symmetry and reactive position considerations are greatly simplified. However, in the patent-mined fragments case, these considerations need to be incorporated into the generation grammar.



Only fragments obtained from the fragment decomposition algorithm (described in Section 2.1.1) are utilized in the generation process, and fragment attachments are made only to reactive position markers. Identically marked positions are attached with fragments of identical identities, while differently marked positions are attached with fragments of different identities, thereby preserving the connection symmetry as was present in the patent structure.

## 2.2 Generating chemically diverse but optimal candidates

We aimed to generate a number of chemically diverse, yet also high performing candidates to increase the likelihood of at least a subset of candidates being experimentally synthesizable. In addition, we have observed empirically from our previous work that the proxy reward predictor has the tendency to be adversarially attacked during training in reinforcement learning settings because of its potential unreliability in out of domain regions.<sup>37</sup> Proposing chemically diverse molecules that have high predicted rewards can allow for an increased chance that a subset of the proposals is in-domain for the reward predictor. The vanilla formulation of MCTS converges at a policy that maximizes expected cumulative reward,<sup>40,43</sup> and hence does not guarantee chemical diversity. To overcome this limitation, we iteratively re-train MCTS with a time-dependent reward term that penalizes chemical similarity to high-performing molecules from previous iterations. More formally,

$$R_i(x) = C(x) - \max_{y \in Y_{<i}} S(x, y), \quad (1)$$

where  $R_i$  is the reward used for training MCTS at the  $i_{th}$  iteration,  $C$  is the property reward function,  $Y_{<i}$  is the inventory of high-performing molecules collected from iterations  $i' < i$ , and  $S$  is a function computing similarity between molecules  $x$  and  $y$  (which we choose to be Tanimoto similarity). Fig. 2(a) shows the evolution of the bandgap and similarity as a function of iteration. The point displayed at each iteration corresponds to the

molecule with the lowest bandgap from the last 100 timesteps of MCTS training. We can see that the constrained optimization task becomes more challenging as more molecules are appended to the inventory  $Y_{<i}$ , resulting in an upward trend in both blue and red curves. Fig. 2(b) zooms into a single iteration and shows the bandgap and similarity curves as a function of training step. In this case, training seems to converge after around 2000 training steps.

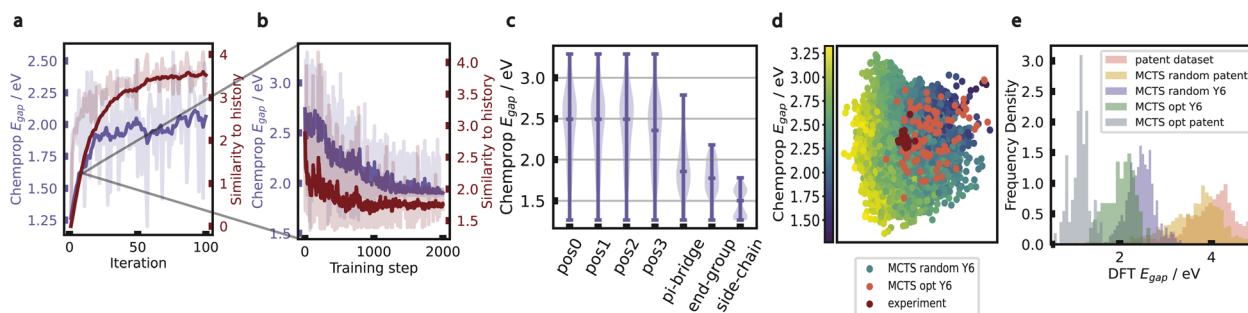
## 2.3 Markov decision process

The Markov Decision Process (MDP) can be defined as a tuple of the state space  $\mathcal{S}$ , the action space  $\mathcal{A}$ , the reward  $R$ , and the transition probability  $P$ . We define MDPs for our two demonstrative tasks below.

**2.3.1 Y6 derivatives.** We formulated this to be a highly focused and relatively small-sized chemical space as a test bed for exploring the capabilities of the approach. Building a molecule involves four hierarchies of actions to be made: (1) modification of heteroatoms and ring sizes in the Y6 core, (2) optional addition of pi-bridges to extend conjugation in the molecule, (3) addition of terminal end-groups, and (4) addition of alkyl chains. Many of the fragment choices included here were taken directly from suggestions provided in Yang.<sup>10</sup> We provide a more formal description of the MDP below.

Let  $\mathcal{S}_t$  and  $\mathcal{A}_t$  denote the state and action spaces respectively at timestep  $t$ . Then,  $\mathcal{S}_1$  is a singleton set containing the Y6 core marked with modifiable positions shown in Fig. 1(a). The first group of actions (with action spaces  $\mathcal{A}_{1 \leq t \leq 4}$ ) are those that modify the structure of the central Y6 core by allowing heteroatom and ring-expansion modifications at positions marked with pos0, pos1, pos2, and pos3.

Action spaces  $\mathcal{A}_{5,6}$  allow the optional addition of pi-bridges to the positions marked by 0 to offer the flexibility to extend conjugation in the molecule, and action space  $\mathcal{A}_7$  is the set of allowed end-groups that can be attached to 0 (if no pi-bridge



**Fig. 2** Features of MCTS training and DFT validation (a) bandgap and similarity penalty as a function of iteration. Values shown are the weighted contributions to the reward function. With every iteration, the constrained optimization becomes more challenging resulting in less optimal candidates. (b) Representative reward evolution during training for one MCTS repetition. (c) Pruning of tree during training. The tree is expanded to contain deeper choices such as end-groups and side-chains only if their rewards are promising. This can be seen by shift towards lower predicted bandgaps as we traverse deeper down the tree. (d) PCA of randomly sampled and property-optimized molecules obtained from MCTS, and some popular experimental candidates. PCA is performed on Morgan fingerprints of molecules, and colors of random samples are based on Chemprop-predicted bandgaps. It can be seen that the random samples are diverse chemically and also span a range of property values, while the MCTS optimized candidates are diverse but concentrated at lower bandgap locations of the landscape. a, b, c, and d are plotted on Y6 MDP. (e) Histograms showing shifts in DFT histograms in comparison to training datasets. It can be seen that MCTS-optimized Y6 and patent molecules exhibit a large shift towards lower bandgaps compared to the patent data distribution.



was attached) or the newly created reactive position available on the pi-bridge. Finally,  $\mathcal{A}_8$  is the set of allowed alkyl chains that can be attached to 1, which we restrict to just methyl groups to reduce the expense of conformer generation and optimization during DFT validation later on. This is a reasonable approximation to make given that alkyl chains are typically used to enhance solubility properties of the molecule and have a smaller influence on optoelectronic properties in isolated molecules. However, with this simplification, we disregard the more important role that alkyl chains play in optoelectronic properties when the molecules are in an aggregated state.<sup>45</sup> All fragments used in this MDP are shown in ESI† State space  $\mathcal{S}_{t>1}$  is the set of possible molecular structures that can be reached by applying transformations in  $\mathcal{A}_{t-1}$ , to state  $s_{t-1}$ .

The reward  $R$  is the negative of the property value (bandgap in eV) as predicted by the Chemprop<sup>44</sup> model combined with the similarity penalty term as defined in eqn (1), and is computed once the molecular generation process has terminated. The MDP is deterministic and hence the transition probability is 1 for a particular state transition and 0 for all others at every timestep  $t$ , i.e.,  $P_t = (\delta_{i1}\delta_{i2}\dots\delta_{in})$  where  $\delta_{ij}$  is the Kronecker delta function and  $i$  is an integer ranging from 1 to  $n$  that represents the index of the chosen action. Only element  $\delta_{ii}$  would be 1, and all other elements of the vector would be 0.

Fig. 2(c) and (d) showcase some features of MCTS training on the Y6 MDP. (c) Demonstrates the ability of MCTS to prune the tree, with deeper actions being expanded to only if they contribute to lowering the bandgap. (d) Is a principal component analysis (PCA) plot showing the partitioning of molecules in the Chemprop embedding space according to Morgan fingerprints of molecules. The MCTS optimized generations are clearly shifted towards the lower bandgap region of the embedding landscape, while also maintaining chemical diversity. They can also be seen to be red-shifted with respect to some popular experimental NFA molecules (see ESI† for a list of the chosen experimental molecules). It is to be noted that both (c) and (d) are based on Chemprop predictions and not DFT calculations. Validation with DFT calculations is shown in later sections.

**2.3.2 Patent-extracted fragment derivatives.** We formulated this to be a more expansive test case with a much larger chemical space, with the aim of discovering highly red-shifted molecules with diverse chemistries. In regimes of this size, exhaustive enumeration of all achievable molecules becomes impractical. Building a molecule in this environment involves three hierarchies of actions: (1) Choosing a core fragment, (2) optional addition of pi-bridges to the core, and (3) addition of terminal end-groups/alkyl chains.

In this case, each reaction involves not just the choice of a fragment identity, but also the choice of reactive position to which it has to be attached. The choice of reactive position is included as stochasticity in the environment rather than an action choice in the tree. More details on this choice, and a more formal description of the MDP are provided in the following paragraphs.

In this case, the breadth of the tree is very wide because of the large action space at each level of the tree. This makes it

beneficial to trade breadth for depth of the tree. We achieve this by performing  $k$ -means clustering<sup>46</sup> of fragments at each hierarchy, so that chemically similar fragments occupy the same cluster. Fragment choices are decomposed into two sequential steps of choosing a cluster, and choosing a fragment from the cluster. This approach retains the original grammar, but greatly improves efficiency of MCTS convergence. The value of  $k$  was chosen arbitrarily as 100, and is a user-defined parameter that can be adjusted. Additionally, we group end-groups and alkyl chains into a single decision here unlike the Y6 case since we do not differentiate between them in the fragment decomposition algorithm described in Section 2.1.1.

$\mathcal{S}_1$  is a singleton set containing an empty molecule, since we build the molecule from scratch in this case. Action space  $\mathcal{A}_1$  is the set of cluster indices of central core clusters (obtained from  $k$ -means clustering).  $\mathcal{A}_2$  is the set of fragment choices from the chosen cluster  $i$ , i.e.,  $\mathcal{C} = \{f \in \mathcal{F} | 1 \leq r(f) \leq 4; n(f) \leq 35\}_i$  where  $\mathcal{F}$  is the set of fragments obtained with the fragment decomposition algorithm described in Section 2.1.1,  $r(f)$  is the number of reactive positions in fragment  $f$ ,  $n(f)$  is the number of heavy atoms (non-hydrogen atoms) in fragment  $f$ , and  $\mathcal{C}_i$  denotes the  $i$ th cluster obtained from  $k$ -means clustering of set  $\mathcal{F}$ . We placed restrictions on the number of reactive positions and size of fragments to reduce the computational load of DFT validation.

The next group of actions contains the optional choice of pi-bridges to allow for extended conjugation in the molecule. This again begins with the choice of a pi-bridge cluster index, followed by the choice of fragment identity from set  $\mathcal{P} = \{f \in \mathcal{F} | r(f) = 2; n(f) \leq 10\}_i$ . We allow this pi-bridge sequence to be performed at most twice for a molecule (action spaces  $\mathcal{A}_{3,5}$  are cluster choices and  $\mathcal{A}_{4,6}$  are fragment identity choices). Finally, the last group of actions, with action spaces  $\mathcal{A}_t$  for  $t \in [7, 2k + 7]$ , attaches end-groups to all open reactive positions on the molecule, where  $k$  is the number of unique reactive positions remaining in the molecule. This attachment is iterated as long as the molecule does not exceed the maximum allowed size of 100 heavy atoms. Note that a maximum of  $2k$  steps is required as we alternate between the cluster and fragment choices. The fragment identity choice is given by set  $\mathcal{E} = \{f \in \mathcal{F} | r(f) = 1; n(f) \leq 20\}_i$  where  $i$  is the index of the chosen cluster. State space  $\mathcal{S}_{t>1}$  is the set of possible molecular structures that can be reached by applying transformations in  $\mathcal{A}_{t-1}$ , to state  $s_{t-1}$ .

Similar to the Y6 case, reward  $R$  is the negative of Chemprop-predicted bandgap combined with a similarity penalty as described in eqn (1). Unlike the Y6 case however, the MDP is not deterministic here. The transitions that go from a state to a chosen fragment cluster are deterministic, but all remaining actions along the branch have some randomness associated with them. Once an action  $a_t$  corresponding to the identity of pi-bridge or end-group is chosen, the reactive position on the current state  $s_t$  is sampled from a uniform distribution, and the next state  $s_{t+1}$  is reached by reacting the fragment at the sampled reactive position on the current state. The transition probability is therefore defined as



$$P_t = \begin{cases} (\delta_{i_1} \delta_{i_2} \cdots \delta_{i_n}) & t \text{ is odd} \\ \left( \frac{1}{k_t} \frac{1}{k_t} \cdots \frac{1}{k_t} \right) & t \text{ is even,} \end{cases}$$

where  $k_t$  is the number of unique reactive positions available on state  $s_t$ . We chose to include this stochasticity in reactive position choice for two reasons: (1) The identity and nature of a chemical fragment typically has a higher influence on the optical properties of a molecule, than the reactive position choice does. Hence removing the reactive position choice from the action space allowed us to limit the depth of the tree and improve computational efficiency during training. (2) We wanted MCTS to converge at solutions that identify optimal fragment choices but are relatively robust to choice of reactive positions. While the precursors are determined by the fragments that are present in the molecule, there is more flexibility in the reactive position choice, and therefore during synthetic reaction planning.

## 2.4 Training reward predictor

### 2.4.1 TD-DFT calculations.

TD-DFT calculations were used as “ground-truth” labels to train our Chemprop reward predictor, as well as to validate the bandgaps of generated candidates. Calculations were performed with the same pipeline that we used in Subramanian *et al.*<sup>37</sup>. RDKit ETKDG<sup>41</sup> approach was used to generate initial conformations, with at least 1500 attempts. The conformers were ranked according to their MMFF94 energies, and the 20 with the lowest energies were chosen.<sup>47</sup> An initial geometry optimization was performed with semi-empirical tight-binding density functional theory (GFN2-xTB)<sup>48</sup> in ORCA 4.2.0.<sup>49</sup> Subsequent geometry optimizations were performed on the lowest energy xTB conformers at the BP86 (ref. 50)-D3 (ref. 51)/def2-SVP<sup>52</sup> level of theory. Finally, single-point TD-DFT calculations were performed at the  $\omega$ B97X-D3 (ref. 53)/def2-SVPD level of theory using the Tamm-Dancoff approximation (TDA)<sup>54</sup> on the optimized geometries. Bandgaps used for reward training were the lowest-energy singlet vertical excitation energies from the TD-DFT calculations. For all reported DFT bandgaps, we used TIGHTOPT keyword during optimization in ORCA 4.2.0 which sets  $\text{ToLE} = 1 \times 10^{-6}$ ,  $\text{ToLRMSG} = 3 \times 10^{-5}$ ,  $\text{ToLMaxG} = 1 \times 10^{-4}$  where ToLE, ToLRMSG, and ToLMaxG are tolerances in the energy, root mean

square average of gradients and maximum element of the gradient respectively. All tolerances are in atomic units.

### 2.4.2 Active learning for data collection.

To ensure that the reward predictor is accurate in the generation domain of MCTS, we performed iterations of active learning (AL) with a combination of acquisition functions to collect training data. Fig. 3 shows the improvement in reward predictor with each iteration of AL. Chemprop<sup>44</sup> was our architecture of choice for the reward predictor. It was first pre-trained on a dataset containing 5568 datapoints obtained from patents and labeled with TD-DFT bandgaps which we created in Subramanian *et al.*<sup>37</sup>. The second round of training was performed on random rollouts generated from MCTS. These two rounds were sufficient to get accurate (we chose the stopping criterion to be an RMSE of 0.3 eV) performance on the focused Y6 chemical space, but the patent-extracted fragment space required an additional round of training. This was done on a combination of datapoints obtained from (1) running the iterative procedure described in Section 2.2 and choosing the best candidate from the last 100 training steps of each iteration, and (2) choosing 100 molecules with the highest value of expected improvement (EI) from 10 000 randomly sampled candidates from MCTS.

$$\text{EI}(x) = (\mu - f(x^*)) \Phi\left(\frac{\mu - f(x^*)}{\sigma}\right) + \sigma \phi\left(\frac{\mu - f(x^*)}{\sigma}\right),$$

where  $\mu$  is the negative of the Chemprop-predicted bandgap, and  $\sigma$  is the standard deviation of predictions from the model ensemble, which is a measure of model uncertainty.  $\Phi$  and  $\phi$  are the cumulative distribution function (CDF) and probability density function (PDF) of a normal distribution, respectively.  $f(x^*)$  is the highest value of predicted reward from the current batch of candidates. While acquisition method 1 obtains a diverse set of most optimal candidates, method 2 obtains a set of optimal candidates as well as suboptimal but high-uncertainty candidates. Together, these two acquisition methods provide training data in explorative and exploitative regimes. Our choice of EI as the acquisition function was inspired by its success in past works such as Ruza *et al.*<sup>55</sup>.

In Fig. 3, the test set for Y6 and patent plots was chosen as a combination of the final set of generated candidates (after the last AL iteration), and a held-out subset of candidates acquired

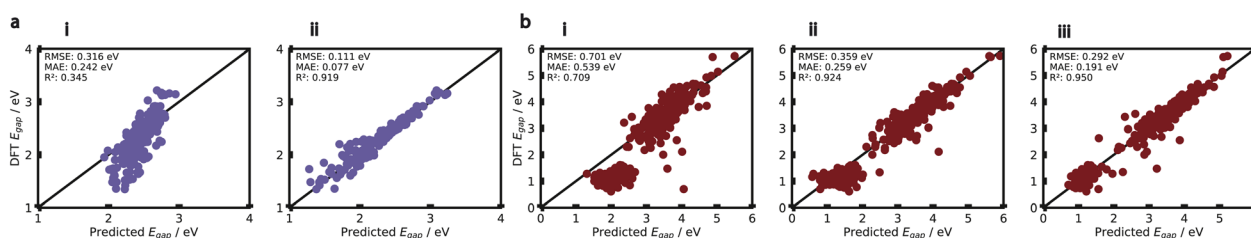


Fig. 3 Active Learning to improve reward prediction Scatter plots showing improvement in fit with AL iterations for (a) Y6 derivatives (b) patent-extracted fragment derivatives. The test data in (a) and (b) are the final sets of 100 molecules generated after all AL iterations have been completed with the Y6 and patent-fragment MDPs respectively. In (a) and (b), the plot i corresponds to the model pre-trained on just patent-mined dataset, and plot ii corresponds to model trained on patent dataset + random rollouts from MCTS. Plot iii in (b) corresponds to model trained on patent dataset + random rollouts + diversity & EI acquisition samples. More details are given in Section 2.4.2.



with the acquisition function from the previous AL iterations (1 round in Y6 and 2 rounds in patent). We chose to use this test set since we needed a way to evaluate if the reward predictor retains performance on old iterations when new iterations of training are performed. Due to this combination of different data sources to form the test set, the property distributions of the test set can be multi-modal, which is evident in sub-figure (b) for instance. It is also evident from (i), (ii), and (iii) of sub-figure (b) that the iterations of AL improve the fit primarily in the low-bandgap region while retaining performance in medium/high-bandgap regions.

### 3 Results and discussion

We validated the final 100 diverse candidates obtained from each of Y6 and patent MDPs with TD-DFT bandgap calculations. We show histograms of these bandgaps in Fig. 2(e) in comparison to the original patent dataset, and random rollouts sampled from MCTS. It can be seen that the MCTS optimized patent and Y6 candidates are strongly red-shifted with respect to random rollouts as well as the patent distribution.

On the final 100 structures from each MDP, we performed  $k$ -means clustering with  $k = 5$  based on Morgan fingerprints of molecules to obtain the 5 most chemically diverse clusters of molecules. We then chose the molecule with the lowest TD-DFT bandgap from each cluster. These structures along with their bandgaps are displayed in Fig. 4. It can be seen that these molecules display structural features that are to be expected from low-bandgap molecules, such as highly conjugated rings. Furthermore, the predicted bandgaps all fall in the NIR region of the spectrum. A noteworthy point here is that the molecule with the lowest bandgap among the five displayed in Fig. 4 (b), has a bandgap of 0.731 eV which is smaller than the smallest bandgap (0.887 eV) in the labeled patent training dataset that we started with, suggesting that our MCTS + AL pipeline is able to extrapolate beyond the labeled training data regime of the reward predictor.

In the case of Y6 derivatives, it is interesting to see a higher tendency of MCTS to generate long polyene bridges between cores and end-groups (see the first, fourth, and fifth molecules in subfigure (a)). Clearly, it has exploited the correlation between degree of conjugation and lowering of bandgap. However, the presence of conical intersections (CI) in long linear polyenes is well known<sup>56,57</sup> suggesting a possibility for these MCTS-predicted molecules to be photochemically inactive. Linear response TD-DFT has been seen to be unsuitable for accurately predicting the presence of these points, paving way for the development of more sophisticated techniques such as spin-flip TD-DFT (SF-TD-DFT).<sup>58,59</sup> This is hence a drawback with the complete reliance of MCTS on the accuracy of TD-DFT predictions, but nevertheless showcases the ability of the algorithm to reliably optimize the chosen reward.

Design strategies such as spatially separating donating and accepting fragments for a “push-pull” effect are also known to have a lowering effect on bandgaps.<sup>60</sup> Among the generated patent-derivatives, this design pattern can be seen in the first and fourth molecules of subfigure (b) which contain electron-rich cores and electron-withdrawing end-groups. More fragment attribution analysis is provided in Section 6 of ESI.† We additionally observe the presence of highly reactive sites in some of these structures (such as hydrogen atoms on aromatic amines of second, fourth, and fifth molecules, and chlorine atoms on the fourth molecule). We replace the reactive positions with methyl groups and perform TD-DFT calculations on the “corrected” structures to verify that optical properties do not drastically change even after the substitution. These results can be found in Section 4 of ESI.†

Finally, the TD-DFT bandgap calculations that we report work under the assumption of 0 K temperature. Furthermore, a more realistic device-like environment would contain thin-films containing several molecules packed together. We perform coupled MD/TD-DFT calculations on the final 10 molecules to better represent finite temperature and molecular aggregation effects. We report absorption spectra calculated

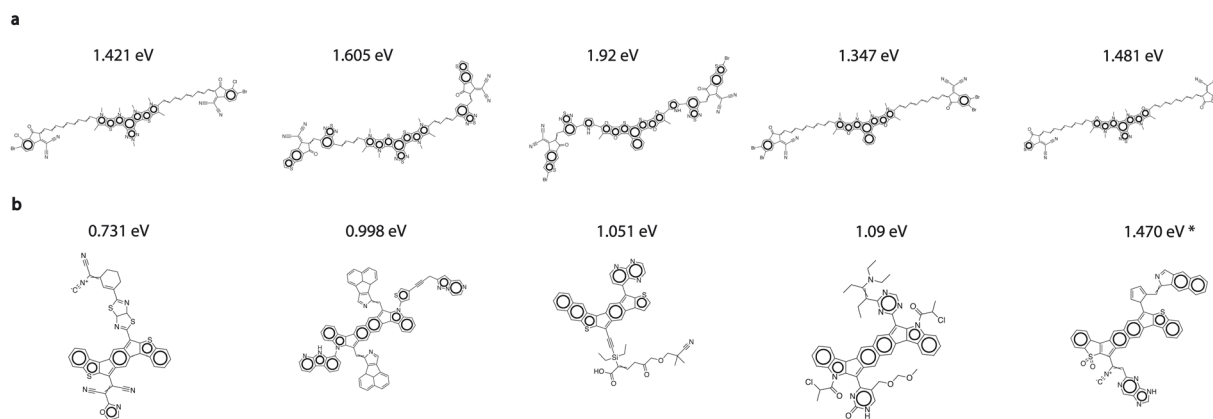


Fig. 4 Final generated candidates and their TD-DFT bandgaps  $k$ -means clustering was performed on final 100 molecules from each category into 5 clusters (based on Morgan fingerprints). The molecules shown are the lowest-bandgap molecules chosen from each cluster. (a) Y6-derivatives MDP, (b) patent-extracted fragments MDP. Bandgap of the last molecule is denoted with asterisk (\*) because the geometry had to be fixed before the TD-DFT calculation was performed. More details are provided in Section 4 of ESI.†



with these methods, and more details about this methodology in Section 3 of the ESI.†

## 4 Conclusions

In this paper, we developed an approach based on the Monte Carlo Tree Search algorithm to generate optimal low-bandgap molecules while also maintaining a degree of synthetic accessibility. We achieve this by utilizing structural priors from a domain-focused patent-mined dataset of organic electronics molecules through a symmetry-aware fragment decomposition algorithm and a fragment-constrained MCTS generator. Our generated candidates retain symmetry constraints from the patent dataset while also exhibiting red-shifted absorption. As future extensions to this work, the optimization objective can be augmented to include solubility (along with the inclusion of alkyl chain exploration into the MDP) and oscillator strengths. Finally, while our approach for diverse candidate generation is iterative, non-iterative objectives as used in GFlowNet<sup>61</sup> could be more efficient alternatives for the future.

## Data availability

All data and code accompanying this article, are available at <https://www.github.com/learningmatter-mit/symmetry-mcts>.

## Author contributions

A. S. developed the symmetry-aware fragment decomposition approach and designed the Y6 and patent MDPs, trained all the models, and analyzed the results. A. S. and J. D. wrote the MCTS training code. J. N. and A. S. developed the coupled MD/TD-DFT calculations methodology, and J. N. performed the coupled MD/TD-DFT calculations and analyzed the corresponding results. K. P. G., A. P. S., and A. S. setup the TD-DFT calculation pipeline, and A. S. and K. P. G. performed the TD-DFT calculations for labeling and validation. All authors contributed to writing the manuscript. R. G.-B. conceived the project, supervised the research, and edited the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

A. S. was supported by funding from Sumitomo Chemical. K. P. G. was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1745302. A. P. S. was supported by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate (NDSEG) Fellowship Program. We acknowledge the MIT Engaging cluster and MIT Lincoln Laboratory Supercloud cluster<sup>62</sup> at the Massachusetts Green High Performance Computing Center (MGHPCC) for providing high-performance computing resources to run our TD-DFT calculations and train our deep learning models. We also thank Prof. Aristide Gumyusenge,

Saumya Thakur, Jurgis Ruza, Soojung Yang, Ajay Subramanian, Sulin Liu, and all members of the RGB group for useful discussions throughout.

## Notes and references

- 1 L. Sun, K. Fukuda and T. Someya, *npj Flexible Electron.*, 2022, **6**, 89.
- 2 Y. Li, W. Huang, D. Zhao, L. Wang, Z. Jiao, Q. Huang, P. Wang, M. Sun and G. Yuan, *Molecules*, 2022, **27**, 1800.
- 3 T. Shan, X. Hou, X. Yin and X. Guo, *Front. Optoelectron.*, 2022, **15**, 49.
- 4 X. Xu, M. Davanco, X. Qi and S. R. Forrest, *Org. Electron.*, 2008, **9**, 1122–1127.
- 5 F. Liu, Z. Zhou, C. Zhang, J. Zhang, Q. Hu, T. Vergote, F. Liu, T. P. Russell and X. Zhu, *Adv. Mater.*, 2017, **29**, 1606574.
- 6 Y. Li, J. D. Lin, X. Che, Y. Qu, F. Liu, L. S. Liao and S. R. Forrest, *J. Am. Chem. Soc.*, 2017, **139**, 17114–17119.
- 7 H. Xu, J. Liu, J. Zhang, G. Zhou, N. Luo and N. Zhao, *Adv. Mater.*, 2017, **29**, 1700975.
- 8 D. Meng, R. Zheng, Y. Zhao, E. Zhang, L. Dou and Y. Yang, *Adv. Mater.*, 2022, **34**, 2107330.
- 9 C. Yan, S. Barlow, Z. Wang, H. Yan, A. K.-Y. Jen, S. R. Marder and X. Zhan, *Nat. Rev. Mater.*, 2018, **3**, 1–19.
- 10 Y. Yang, *ACS Nano*, 2021, **15**, 18679–18682.
- 11 Q. He, P. Ufimkin, F. Aniés, X. Hu, P. Kafourou, M. Rimmele, C. L. Rapley and B. Ding, *SusMat*, 2022, **2**, 591–606.
- 12 D. Schwalbe-Koda and R. Gómez-Bombarelli, *Machine Learning Meets Quantum Physics*, Springer, 2020, pp. 445–467.
- 13 D. C. Elton, Z. Boukouvalas, M. D. Fuge and P. W. Chung, *Mol. Syst. Des. Eng.*, 2019, **4**, 828–849.
- 14 M. H. Segler, T. Kogej, C. Tyrchan and M. P. Waller, *ACS Cent. Sci.*, 2018, **4**, 120–131.
- 15 E. J. Bjerrum and R. Threlfall, *arXiv preprint arXiv:1705.04612*, 2017.
- 16 R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams and A. Aspuru-Guzik, *ACS Cent. Sci.*, 2018, **4**, 268–276.
- 17 M. J. Kusner, B. Paige and J. M. Hernández-Lobato, *International conference on machine learning*, 2017, pp. 1945–1954.
- 18 M. Olivecrona, T. Blaschke, O. Engkvist and H. Chen, *J. Cheminf.*, 2017, **9**, 1–14.
- 19 R. Mercado, T. Rastemo, E. Lindelöf, G. Klambauer, O. Engkvist, H. Chen and E. J. Bjerrum, *Mach. Learn.: Sci. Technol.*, 2021, **2**, 025023.
- 20 W. Jin, R. Barzilay and T. Jaakkola, *International conference on machine learning*, 2018, pp. 2323–2332.
- 21 D. Flam-Shepherd, T. C. Wu and A. Aspuru-Guzik, *Mach. Learn.: Sci. Technol.*, 2021, **2**, 045010.
- 22 W. Gao and C. W. Coley, *J. Chem. Inf. Model.*, 2020, **60**, 5714–5723.
- 23 P. Ertl and A. Schuffenhauer, *J. Cheminf.*, 2009, **1**, 1–11.



- 24 C. W. Coley, L. Rogers, W. H. Green and K. F. Jensen, *J. Chem. Inf. Model.*, 2018, **58**, 252–261.
- 25 C. W. Coley, D. A. Thomas III, J. A. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, *et al.*, *Science*, 2019, **365**, eaax1566.
- 26 S. Genheden, A. Thakkar, V. Chadimová, J.-L. Reymond, O. Engkvist and E. Bjerrum, *J. Cheminf.*, 2020, **12**, 70.
- 27 W. Gao, R. Mercado and C. W. Coley, *arXiv preprint arXiv:2110.06389*, 2021.
- 28 M. Hartenfeller, H. Zettl, M. Walter, M. Rupp, F. Reisen, E. Proschak, S. Weggen, H. Stark and G. Schneider, *PLoS Comput. Biol.*, 2012, **8**, e1002380.
- 29 H. M. Vinkers, M. R. de Jonge, F. F. Daeyaert, J. Heeres, L. M. Koymans, J. H. van Lenthe, P. J. Lewi, H. Timmerman, K. Van Aken and P. A. Janssen, *J. Med. Chem.*, 2003, **46**, 2765–2773.
- 30 K. Swanson, G. Liu, D. B. Catacutan, A. Arnold, J. Zou and J. M. Stokes, *Nat. Mach. Intell.*, 2024, **6**, 338–353.
- 31 J. Bradshaw, B. Paige, M. J. Kusner, M. Segler and J. M. Hernández-Lobato, *Adv. Neural Inf. Process. Syst.*, 2019, **32**, 7937–7949.
- 32 J. Bradshaw, B. Paige, M. J. Kusner, M. Segler and J. M. Hernández-Lobato, *Adv. Neural Inf. Process. Syst.*, 2020, **33**, 6852–6866.
- 33 F. Strieth-Kalthoff, H. Hao, V. Rathore, J. Derasp, T. Gaudin, N. H. Angello, M. Seifrid, E. Trushina, M. Guy, J. Liu, *et al.*, *Science*, 2024, **384**, eadk9227.
- 34 H. S. Kwak, Y. An, D. J. Giesen, T. F. Hughes, C. T. Brown, K. Leswing, H. Abroshan and M. D. Halls, *Front. Chem.*, 2022, **9**, 800370.
- 35 C.-H. Li and D. P. Tabor, *Chem. Sci.*, 2023, **14**, 11045–11055.
- 36 J. Westermayr, J. Gilkes, R. Barrett and R. J. Maurer, *Nat. Comput. Sci.*, 2023, 1–10.
- 37 A. Subramanian, K. P. Greenman, A. Gervais, T. Yang and R. Gómez-Bombarelli, *Digital Discovery*, 2023, **2**, 1006–1015.
- 38 L. Kocsis and C. Szepesvári, *European conference on machine learning*, 2006, pp. 282–293.
- 39 R. Coulom, *International conference on computers and games*, 2006, pp. 72–83.
- 40 M. Świechowski, K. Godlewski, B. Sawicki and J. Mańdziuk, *Artif. Intell. Rev.*, 2023, **56**, 2497–2562.
- 41 G. Landrum *et al.*, 2013.
- 42 D. Michie, *Nature*, 1968, **218**, 19–22.
- 43 D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.*, *nature*, 2016, **529**, 484–489.
- 44 E. Heid, K. P. Greenman, Y. Chung, S.-C. Li, D. E. Graff, F. H. Vermeire, H. Wu, W. H. Green and C. J. McGill, *J. Chem. Inf. Model.*, 2023, **64**, 9–17.
- 45 T. Lee, C. E. Song, S. K. Lee, W. S. Shin and E. Lim, *ACS Omega*, 2021, **6**, 4562–4573.
- 46 J. MacQueen *et al.*, *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1967, pp. 281–297.
- 47 S. Riniker and G. A. Landrum, *J. Chem. Inf. Model.*, 2015, **55**, 2562–2574.
- 48 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
- 49 F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
- 50 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
- 51 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
- 52 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- 53 J.-D. Chai and M. Head-Gordon, *J. Chem. Phys.*, 2009, **131**, 174105.
- 54 S. Hirata and M. Head-Gordon, *Chem. Phys. Lett.*, 1999, **314**, 291–299.
- 55 J. Ruza, M. Stolberg, S. Cawthorn, J. Johnson, Y. Shao-Horn and R. Gomez-Bombarelli, 2025.
- 56 W. Fuss, Y. Haas and S. Zilberg, *Chem. Phys.*, 2000, **259**, 273–295.
- 57 A. Nenov and R. de Vivie-Riedle, *J. Chem. Phys.*, 2012, **137**(7), 074101.
- 58 B. G. Levine, C. Ko, J. Quenneville and T. J. Martínez, *Mol. Phys.*, 2006, **104**, 1039–1051.
- 59 N. Minezawa and M. S. Gordon, *J. Phys. Chem. A*, 2009, **113**, 12749–12753.
- 60 G. Pirotte, P. Verstappen, D. Vanderzande and W. Maes, *Adv. Electron. Mater.*, 2018, **4**, 1700481.
- 61 Y. Bengio, S. Lahlou, T. Deleu, E. J. Hu, M. Tiwari and E. Bengio, *J. Mach. Learn. Res.*, 2023, **24**, 10006–10060.
- 62 A. Reuther, J. Kepner, C. Byun, S. Samsi, W. Arcand, D. Bestor, B. Bergeron, V. Gadepally, M. Houle and M. Hubbell *et al.*, *2018 IEEE High Performance extreme Computing Conference (HPEC)*, 2018, pp. 1–6.

