NOTATION TRENDPLUS PUBLISHING
WHERE IDEAS MEET IMPACT

Volume-1, Issue 1, August 2026, Page No- 236-255
Journal of Pharmaceutical Innovations and Therapeutic Horizons (JPITH)

Hierarchical Options-GAN: A Scaffold-Aware Generative Adversarial Network for De Novo Molecular Design

Mukesh Thakur

Assistant Professor, Chandravansi Pharmacy College, Bishrampur, Palamu, Jharkhand

Article Received: 09-05-2026

Revised Version: 27-06-2026

Accepted: 15-07-2026

*Corresponding Author

Mukesh Thakur

E-Mail Id: mukesht870@gmail.com

Abstract: We introduce a new generative adversarial network named as Hierarchical Options-GAN (HO-GAN) for molecular de novo design by handling the conventional sparse reward propagation problem in constructing molecule graphs. A two-layered generative architecture that uses a "scaffold plan" generated by hierarchical reinforcement learning is the main novelty; a new scaffold plan can be generated using something entirely different to the atoms to assemble. In our concept, we have a prelearned space of embedding chemically valid sub-graph templates for a sequence, and the elements of the embedding space are common structural motifs, called options, that are represented as ring systems, linkers and functional groups and have attachment points. Each of the selected motifs is subsequently elaborated by means of an iterative order of atom types and bond orders, applying strict valence constraints, which constitutes a low-level policy to progressively build up molecules as composite levels. We have a dual discriminator system; a global graph discriminator mainly checks the validity of the overall molecular graphs, a local discriminator mainly checks the validity of the extension of individual atoms at a molecular level. We also add a property predictor that gives a scalar, a reward to the high-level controller policy gradient updates that encourages generation of molecules possessing desired physicochemical and/or pharmacological properties, among the ones being generated. The unsupervised graph clustering of the data of drug-like molecules achieves the clustering, and the entire architecture is randomly initialized and then pre-trained using a motif decomposition task before the end-to-end adversarial fine-tuning. The results obtained from the standard molecular generation tasks demonstrate that HO-GAN regularly produced structurally diverse and chemically plausible molecule candidates, and its capability to control the topology within the molecule's scaffold led significantly better results as compared to baseline molecular generators with flat generation behavior.

Keywords: Generative adversarial networks, De novo molecular design, Scaffold-aware generation, Hierarchical learning, Drug discovery

Introduction

The calculation of design and development of novel functional molecules is one of the basic problem in drug design and materials chemistry field. In the last decade, deep generative models have become a powerful tool for de novo molecular design, as a data-driven alternative to the traditional rule of thumb expert systems and evolutionary algorithms (Devi et al., 2015). The early successes in the use of Variational Autoencoders (VAEs) (Kingma & Welling, 2013) and Recurrent Neural Networks (RNNs) (Bjerrum & Threlfall, 2017) showed that learning a continuous latent representation or generating molecules as SMILES strings were feasible. Such sequence based methods however generate many invalid outputs (e.g., output some outputs that do not adhere to the valence rules among the cells) and do not lend themselves to extracting long-range homogeneous structural dependencies well. To solve this, generative methods directly operating on molecular graphs have been developed, e.g., GraphVAE (Simonovsky & Komodakis, 2018) and JT-VAE (Jin et al., 2018) which preserve structure for the same reason it was built. GANs cannot be left behind, however, as two recent examples show, ORGAN (Guimaraes et al., 2017) was extended for creating a more realistic sample for this area, while MolGAN (Cao & Kipf, 2018) extended the GAN domain with molecular structures.

Despite this advancement, an extremely profound question comes to mind: How do you credit later molecular constructions, in the context of the long-horizon credit assignment problem in sequential molecular construction? Like standard end-to-end generative models, autoregressive or adversarial ones, need to give feedback that traverses across many discrete steps of the molecule formation. Pathological errors may occur early in the process, such as if the error is the ring closing in a incorrect way or the choice of scaffolds, then the rewards signals may be sparse (may not exist at all) and/or ambiguous and it can create significant problems. This is a known issue in reinforcement learning problems that can be challenging when the reward signal is sparse, such as Gupta et al (1994) and Dietterich (2000). The Options framework (Sutton et al., 1999) is the formalization of temporally extended actions (or sub-policies) for semi-Markov decision processes that provides a principled way to incorporate some temporal abstraction. So far, hierarchical methods have been successfully used with architectures like the Feudal Networks (Vezhnevets et al., 2017) and Option-Critic (Bacon et al., 2017) in the context of deep reinforcement learning, where sub-policies related to specific components of the action space are learned and re-used for solving a complex task.

In order to introduce a generational adversarial network that explicitly separates planning of the structures of scaffolds and forming the bonds of atoms in the derived molecules, alleviating the credit assignment problem despite generating molecules, we introduce a model called Hierarchical Options-GAN (HO-GAN). This is a two-layered system that resembles a transformer with a high level control layer sampling reusable structural motifs (options) from a pre-developed embedding space of chemically valid subgraph motifs (e.g., common rings, linkers, etc., with specific places for attachment) in which the non-uniform distribution and possibly a recursive simulation of low level policy refine each sampled motif. This hierarchical decomposition facilitates the separation of the temporal scale of molecular growth; this means the basis of the competition in the molecular dynamics can address higher order scaffolds and, at the same time, have a lower order on the atomistic precision, so that the gradient mixing of lower order from higher order sequence positions is avoided.

Two discriminators: (1) global graph discriminator for overall coherence of the architecture of generated molecules; (2) local discriminator for chemical validity of each atomic expansion step, used for adversarial training. To be able to steer further in desired properties, we employ a differentiable property predictor that gives a reward that can be used to perform policy gradient updates on the high level controller. In this option set, an embedding pre-training stage is performed over motif decomposition using a compilation of a large drug-like molecule database (ChEMBL), followed by a fine-tuning stage

on the same database via an industrial embedding strategy and the clustering of the resulting graphs is carried out in an unsupervised way.

Our work has three main contributions to make. One key challenge for a successful implementation of end-to-end molecular generators is the sparse feedback problem; to address this challenge, we first present a hierarchical generation framework, which is based on the notion of separating global structural planning from local atomic refinement. Second, to create a dual discrimination/multi-supervision strategy providing rich and structurally informative supervision at both hierarchy levels. Third, we demonstrate that on several challenging benchmark tasks that are more difficult to control by flat generation methods like MolGAN [7](# "molgan an implicit generative model for small molecular graphs"), HO-GAN successfully generates chemically realistic, structurally diverse molecules, which in addition are more controllable than other approaches, and yield better statistics over scaffold topologies and target property spaces.

Both Hierarchical reinforcement learning and designing molecules by using generative models are related works that are explored briefly in Section 2.As previously mentioned, the rest of the paper will discuss the related works for generative models for designing molecules and for Hierarchical reinforcement learning. The prerequisite about molecular graph representation and the Options framework are prepared and covered in section 3. The design and training strategy for HO-GAN are outlined in Section 4. The experimental setup is discussed in Section 5 as are the data and metrics used. Section 6 discusses and presents Empirical Results. Some limitations pointed out and future study directions are suggested in Section 7. Lastly, a summary of our findings is included in Section 8.

Related Work

Generative Models for Molecular Design

While sequence-based models have dominated machine learning for molecular design for a long time, the industry has been progressing rapidly to graph-native models during the past few years. Variants of what are commonly called Variational Autoencoder (VAE) (Kingma & Welling, 2013) and Recurrent Neural Network (RNN) (Bjerrum & Threlfall, 2017) were previously used, and typically relied on SMILES representations of molecules. The authors in these papers showed that it is possible to learn a continuous latent space for molecular structure but, in many cases the model would generate invalid molecules as the SMILES didn't have the same flexible syntax needed to represent molecules or enforced the same valence constraint. This led to the creation of models that were graphical. The GraphVAE (Simonovsky & Komodakis, 2018) represented molecules as graphs and generated an adjacency matrix and node features directly via VAE whereas the Junction Tree VAE (JT-VAE) (Jin et al., 2018) split a molecule into a tree-shaped 'scaffold' and 'side chains' and generated a scaffold first from VAE, and then attached side chains. This was a 2-step procedure, which was much more valid and novel for generating molecules, and is closely related to our hierarchical approach, but this did not learn motifs that can be reused, whereas the 2-step approach has a fixed level of decomposition.

Actually, Generative Adversarial Networks (GANs) have been also involved in this area. Several works that generated sequences by a combination of adversarial training and policy gradient methods, such as the Objective-Reinforced Generative Adversarial Network (Guimaraes et al., 2017), had been done previously. Furthermore, there are recent works that also outputted adjacency and feature matrices, like MolGAN (Cao & Kipf, 2018), which focused on the generation of graphs, prior to this. However MolGAN only was able to synthesize molecules of a fixed maximum size, and its capacity to make long molecules of high validity was not that great. The Graph Convolutional Policy Network (GCPN) (You et al., 2018) also views molecular generation as an MDP problem and leverages the reinforcement

learning method for setting properties while considering the abstraction over time used in the Options framework. More valid sequences were generated using a GAN with a reinforcement learning reward function for masked SMILES strings, but were not generated with the long-horizon credit assignment problem in the sequential generation addressed, by SMILES-MaskGAN (Lee et al., 2021). These earlier efforts based on GANs differ from our HO-GAN because our HO-GAN breaks the scaffold planning/atomic refining down to its hierarchical components, so we can give much more dense and structured feedback providing landmarks and structure during the generation process.

Hierarchical Reinforcement Learning and the Options Framework

One of the most important problems in learning to follow good policies in a reward environment with few or few or delayed rewards is reinforcement learning (RL). Sutton et al. (1999) suggest a formal structure of temporal abstraction where each ‘option' refers to a group of primitive actions and ends when a certain terminating condition is met. It's a problem abstraction, that provides an agent with the capacity to experience reasoning and planning activities at a level of abstraction higher than the actual one, and reduces the dimension of the problem by a large extent. Vezhnevets et al., 2017, develop the Feudal Networks architecture which had been two-level with an abstract goal setting to a worker and a primitive action performing to a manager. More important is what they call the Option-Critic architecture (Bacon et al., 2017), which was able to achieve a novel approach to learning options end-to-end, without needing to introduce sub-goals engineered by humans. This has been successfully implemented for complex continuous control problems.

The art of molecular generation needs the knowledge and understanding of temporal abstraction. The building up of a molecule, atom by atom is comparable to a long time horizon decision making process; the size and magnitude of the initial decisions (e.g., whether a particular scaffold can be connected in a given way) cannot be determined until pushed to the end when the entire molecule is scrutinised, e.g., for drug-likeness or synthetic accessibility. This is very in line with the Options framework, which is the high-level controller chooses a structural motif for the problem (an option) and each option is an instance of coherent local sequence of atomic decisions. This separation mirrors the chemical intuition that molecular design often begins with scaffold selection, followed by functionalization. The hierarchical graph tokenization approach (Chen et al., 2024) also recognizes the importance of molecular motifs, learning a hierarchy of tokens for molecule-language tasks, but it does not apply this hierarchy to the generation process itself. Our HO-GAN is, to our knowledge, the first work to formally integrate the Options framework from hierarchical RL into an adversarial generative model for molecular graphs.

Comparison with Existing Works

While existing hierarchical molecular generators like JT-VAE (Jin et al., 2018) decompose molecules into a fixed tree of scaffolds and side chains, they do so at a single, predetermined level of abstraction. In contrast, our HO-GAN learns a set of reusable motif “options” via unsupervised graph clustering, enabling the high-level controller to compose molecules from diverse, data-driven building blocks. Furthermore, unlike previous GAN-based methods (Guimaraes et al., 2017), (Cao & Kipf, 2018) that rely on a single global discriminator’s feedback, our dual discriminator system provides dense, intermediate rewards for every atomic expansion step, directly mitigating the sparse feedback problem. The integration of a property predictor to compute policy gradient rewards for the high-level controller also distinguishes our work from pure adversarial or pure RL approaches (You et al., 2018), creating a hybrid training signal that guides both the high-level scaffold selection and the low-level atomistic precision. These combined features—a learnable option set, a hierarchical generative process, and dense

dual adversarial supervision—represent our key novelty over the existing landscape of molecular generative models.

Preliminaries

We start by formalizing the representation of molecular graphs, building the groundwork for our proposed approach and introducing our theoretical foundation for hierarchical reinforcement learning: the Options framework, on which we shall build our two-level generator.

Molecular Graph Representation

The molecule can be represented naturally in the form of an undirected graph, G=(V,E), where the nodes, v_1, v_2, …, v_n, represent atoms, and the edges are directed (arrows) and are called chemical bonds, with E⊆V×V. The dimension of the feature vector, d_v , is equal to the number of variables measured by an atom, and the x_i ∈ R^(d_v) are the feature vectors that represent the atoms. These are vectorial representations of certain aspects of the atomic property; in the case above they would be the type of atom (C, N, O), the formal charge of the atom, and the hybridization state of the atom. Likewise, every edge (v_i,v_j) in E has an edge vector e_ij∈R^(d_e) representing the bond type between v_i and v_j (e.g., single bond, double bond, triple bond and bond type by presence of aromatic ring). There are two representations of the whole molecular graph - one is a n×n matrix A, which is a 0–1 vector valued for each atom, where n is the number of the atoms; the other is an extended node feature matrix X, which is a R^(n×d_v) matrix, where d_v is the number of the features and n is the number of the atoms. To make all the graphs easier to fit within the intersection's maximum graph size N_"max" , larger/basically similar graphs will be padded with a dummy node, a common approach in graph-based generative models (Cao & Kipf, 2018).

Valid constraints on the generation of a molecular graph that turn out to be very important in chemistry are referred to as chemical constraints. The sum of the bond orders should not exceed the maximum(BO) for each atom (Carbon: max 4, Nitrogen: max 3). Also, the graph generated should be connected, and should not contain any substructures which are chemically impossible (e.g. a C atom with all five bonds occupied). In order to enforce these constraints during the generation process, we model our generation process as a Markov decision process (MDP), following the same idea as adopted by GCPN (You et al., 2018). Each step t the state s_t is the partially constructed molecular graph and there exists an action a_t, which is either addition of a new atom or the connection by a bond to form a new atom. The action taken is determinate and the probability of a transition is dependent on the action. Stops once the molecule becomes full or when max number of steps is exceeded.

The Options Framework

The sampled MDP is an obvious elaboration of the options framework (Sutton et al., 1999), where the temporal version of the actions (options) are defined. Informally, an option ω∈Ω is given as the triple (I_ω,π_ω,β_ω), where: I_ω⊆S is a subset of states where the option can be started; π_ω:S×A→[0,1] is a policy of the option when it can be started (A is the set of primitive actions); β_ω:S→[0,1] is a terminating condition (a policy for the option when it is active). Once the option is chosen, then agent executes the action according to the internal policy π_ω, until it rolls up to a higher level policy β_ω. This provide the abstraction so that on a higher level, the agent doesn't need to make a decision on the specific primitive action, but rather which primitive action to perform.

At a molecular level generation can be considered a hierarchical process in the building up of a molecule. At the lower level (option), the structural motif is chosen and was realized by means of adding one or more of the lower level (atomic) additions, including a benzene ring or peptide bond. This is the

very atom-by-atom building of the motif that is effected by the interior policy π_ω, and by the institution of termination of the condensation β_ω, respectively, at the time the motif is completed. Indeed, there is a chemical feel to this decomposition, since medicinal chemists typically select a "scaffold" (core ring system) and then attach functional groups onto that with the aim of rendering a more effective therapeutic product. Our model will also be able to learn common structures from a drugs database library, which means after subjecting the model to a cycle it will then be able to look at similar structures in the library and produce new ones, effectively reducing the decision horizon, and reducing the sparse reward problem.

The Option-Critic architecture (Bacon et al., 2017) provides a gradient-based method for learning options end-to-end. It parameterizes the policy over options π_Ω, the intra-option policies π_ω, and the termination functions β_ω with neural networks, and updates them via policy gradient. This framework is particularly suitable for our adversarial setting, as it allows the high-level controller and low-level policies to be jointly optimized through the dual discriminator's feedback. In our HO-GAN, we adapt the Option-Critic architecture to the molecular graph domain, where the state space is the partially constructed graph and the action space consists of atom and bond additions. The options correspond to chemically meaningful subgraph templates, which we pre-learn via unsupervised clustering of molecular fragments, as detailed in Section 4.

Hierarchical Adversarial Generation with Motif Options

This section provides the technical details of the proposed Hierarchical Options-GAN framework. We first define the overall system architecture, then describe the generator's two-tier hierarchy, the dual discriminator design, and the hybrid training objective.

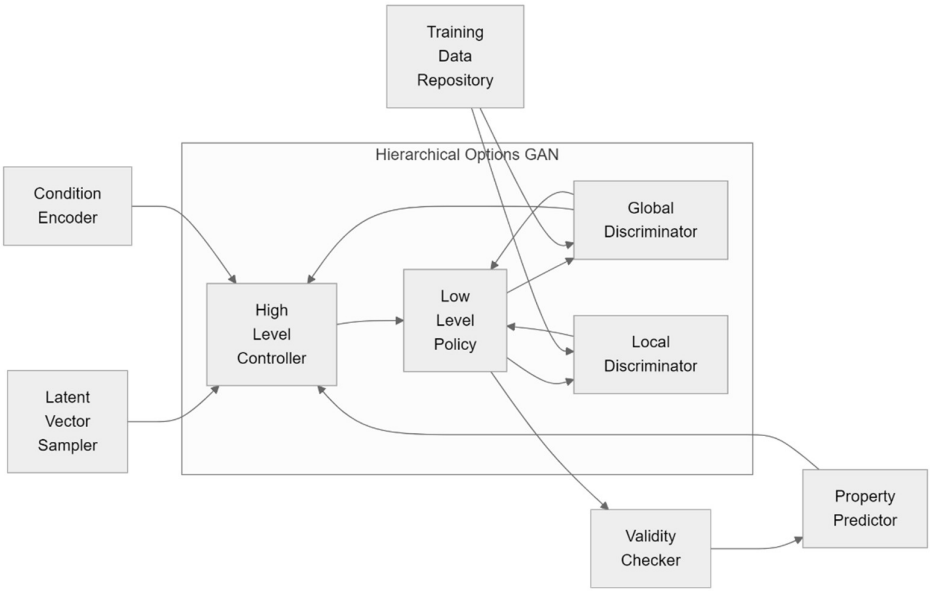

Figure 1. *System Level Integration of Hierarchical Options GAN*

Hierarchical Generator Architecture and Option Execution

The generator in HO-GAN is formulated as a two-level hierarchical policy operating over a molecular graph construction MDP. The high-level controller π_Ω selects which structural motif (option) to append to the growing molecular scaffold, while the low-level policy π_ω executes the primitive atom and bond additions that realize that motif. This separation of concerns directly addresses the long-horizon credit

assignment problem by allowing the high-level controller to receive feedback based on global scaffold coherence and the low-level policy to be evaluated on local atomistic validity.

The option set $\mathcal{O} = \{\omega_1, \omega_2, \ldots, \omega_K\}$ contains K reusable structural motifs, each corresponding to a chemically meaningful subgraph template extracted via unsupervised hierarchical graph clustering of a large database of drug-like molecules (e.g., ZINC15 or ChEMBL). Each option ω_k is characterized by its learned embedding $\mathbf{e}_{\omega_k} \in \mathbb{R}^{d_o}$, a topology template describing the motif's internal connectivity, and a set of attachment points $\mathcal{A}_{\omega_k} = \{a_1, a_2, \ldots, a_{k_{\omega_k}}\}$ that specify where the motif can covalently bond to the existing molecule. The motif inventories are constructed using the BREAK algorithm (Loving et al., 2010), which decomposes molecules into ring systems and linker fragments by recursively breaking bridge bonds.

At time step t of the generation process, the current state $\mathbf{s}_t$ is represented as the partially constructed molecular graph $\mathcal{G}_t$ with its corresponding node feature matrix $\mathbf{X}_t$ and adjacency matrix $\mathbf{A}_t$. The high-level controller, implemented as a transformer decoder, consumes a sequence of previously selected option embeddings $[\mathbf{e}_{\omega_1}, \ldots, \mathbf{e}_{\omega_{t-1}}]$ along with a latent code $\mathbf{z} \in \mathbb{R}^{d_z}$ sampled from a standard normal prior and an optional conditioning vector $\mathbf{c} \in \mathbb{R}^{d_c}$ encoding desired molecular properties (e.g., LogP, QED). The high-level policy computes:

$$P(\omega_t \mid \omega_{<t}, \mathbf{z}, \mathbf{c}) = \text{softmax}\left(\mathbf{W}_{\text{high}} \cdot \text{TransformerDecoder}\left([\mathbf{z}; \mathbf{c}; \mathbf{e}_{\omega_1}; \ldots; \mathbf{e}_{\omega_{t-1}}]\right)\right) \quad (1)$$

where $\mathbf{W}_{\text{high}} \in \mathbb{R}^{K \times d_{\text{model}}}$ is a learned projection matrix. The selected option ω_t provides a template subgraph with its internal structure pre-specified, including a pattern of atom types and bond orders that are known to be chemically valid.

Once an option ω_t is selected, the low-level policy π_{low} takes over to attach this motif to the existing graph $\mathcal{G}_{t-1}$. Let $v \in \mathcal{V}_{t-1}$ be a candidate attachment node in the current graph, and let $a_j \in \mathcal{A}_{\omega_t}$ be an attachment point on the motif. The low-level policy must decide which atom from the motif to attach to which existing atom, and with which bond order. We denote a low-level action as $a_{\text{low}} = (v, a_j, b_{vj})$, where $b_{vj} \in \{1,2,3\}$ is the bond order. The low-level policy is conditioned on the current graph context and the selected option's embedding:

$$\pi_{\text{low}}(a_{\text{low}} \mid \mathcal{G}_{t-1}, v, \mathbf{e}_{\omega_t}) = \text{softmax}\left(\mathbf{W}_{\text{low}} \cdot \text{FFN}\left([\mathbf{h}_v; \mathbf{e}_{\omega_t}; \mathbf{s}_t]\right)\right) \quad (2)$$

Here, $\mathbf{h}_v \in \mathbb{R}^{d_h}$ is a node-level representation of attachment atom v obtained from a graph isomorphism network (GIN) (Xu et al., 2018) applied to $\mathcal{G}_{t-1}$, and $\mathbf{s}_t$ is the global graph state representation (e.g., a graph-level readout of $\mathcal{G}_{t-1}$). The feedforward network FFN and projection matrix $\mathbf{W}_{\text{low}}$ are shared across all options, enabling the low-level policy to generalize attachment patterns across different motifs.

The execution of an option ω_t proceeds as follows. The model first determines the optimal attachment point v^* on the current graph by computing an attention score between each candidate node v and the option embedding $\mathbf{e}_{\omega_t}$:

$$a(v, \omega_t) = \mathbf{w}_a^\top \tanh\left(\mathbf{W}_h \mathbf{h}_v + \mathbf{W}_o \mathbf{e}_{\omega_t}\right) \quad (3)$$

where $\mathbf{w}_a$, $\mathbf{W}_h$, and $\mathbf{W}_o$ are learned parameters. The node with the highest score is selected as the attachment site. Then, the low-level policy iteratively assigns the motif's atoms and bonds to the existing graph, respecting valence constraints at each step. If a valence violation is detected (e.g., carbon

already has four bonds), the action is masked and re-sampled. The termination condition β_{ω_t} for the option is a sigmoid output that indicates whether the motif has been fully incorporated:

$$\beta_{\omega_t}(\mathcal{G}_t) = \sigma\left(\mathbf{w}_\beta^\top \left[\mathbf{h}_{\text{graph}}(\mathcal{G}_t); \mathbf{e}_{\omega_t}\right]\right) \quad (4)$$

where $\mathbf{h}_{\text{graph}}(\mathcal{G}_t)$ is a graph-level readout after the most recent motif addition. When $\beta_{\omega_t} > 0.5$, the option terminates, and control returns to the high-level controller to select the next option. This process repeats until a terminal state is reached (e.g., the molecule reaches a target size, or a special stop token is selected).

Two-Level Discriminator Design for Dense Reward Signal

A key challenge in molecular GANs is the provision of informative gradients for long generation sequences. Standard approaches using a single global discriminator (Cao & Kipf, 2018) only evaluate the final complete molecule, providing a single scalar signal at the end of the episode. For a molecule with 50 atoms, this means the first atom addition receives the same (delayed) feedback as the last, making credit assignment extremely difficult. To address this, we propose a dual discriminator system $D = \left(D_{\text{global}}, D_{\text{local}}\right)$ that provides dense intermediate rewards at both the motif-level and atom-level granularity.

The global discriminator D_{global} operates on complete molecular graphs $\mathcal{G}$ and is implemented as a GIN followed by a multi-layer perceptron (MLP). Its input is the full graph representation $\mathbf{h}_{\mathcal{G}} = \text{GIN}_{\text{global}}(\mathcal{G})$, and it outputs a scalar score $D_{\text{global}}(\mathcal{G}) \in \mathbb{R}$ indicating the plausibility of the entire molecular structure. This component follows the Wasserstein GAN formulation with gradient penalty (Gulrajani et al., 2017):

$$\mathcal{L}_{\text{WGAN}} = \mathbb{E}_{\mathcal{G}_{\text{real}}}\left[D_{\text{global}}(\mathcal{G}_{\text{real}})\right] - \mathbb{E}_{\mathcal{G}_{\text{gen}}}\left[D_{\text{global}}\left(\mathcal{G}_{\text{gen}}\right)\right] + \lambda_{\text{grad}}\mathbb{E}_{\hat{\mathcal{G}}}\left[\left(\| \nabla_{\hat{\mathcal{G}}} D_{\text{global}}\left(\hat{\mathcal{G}}\right) \|_2 - 1\right)^2\right] \quad (5)$$

where $\hat{\mathcal{G}}$ is a linear interpolation between real and generated graphs. The global discriminator thus incentivizes the overall topological plausibility and distributional fidelity of the generated molecules.

The local discriminator D_{local} evaluates the chemical validity of each atomic expansion step performed by the low-level policy. It receives as input the intermediate graph state $\mathcal{G}_{t-1}$, the attachment node v, the selected motif attachment point a_j, and the chosen bond order b_{vj}. The local discriminator computes:

$$D_{\text{local}}\left(\mathcal{G}_{t-1}, v, a_j, b_{vj}\right) = \text{MLP}_{\text{local}}\left(\left[\mathbf{h}_v; \mathbf{e}_{a_j}; \text{onehot}\left(b_{vj}\right)\right]\right) \quad (6)$$

where $\mathbf{e}_{a_j}$ is a learned embedding for attachment point a_j on the motif. The output is a scalar between 0 and 1, with high values indicating a chemically plausible expansion step (e.g., respecting valence rules, avoiding steric clashes, maintaining reasonable bond angles). The local discriminator is trained on a dataset of valid molecular fragments and their attachment patterns, supervised by a binary cross-entropy loss:

$$\mathcal{L}_{\text{local}} = -\sum_t \sum_{v, a_j, b_{vj}} \left[y_{\text{real}} \log D_{\text{local}}(\cdot) + (1 - y_{\text{real}}) \log\left(1 - D_{\text{local}}(\cdot)\right)\right] \quad (7)$$

where $y_{\text{real}} = 1$ for valid transitions sampled from known molecules and $y_{\text{real}} = 0$ for chemically invalid transitions (e.g., exceeding valence). The local discriminator's predictions serve as dense intermediate rewards for the low-level policy, as described in the following subsection.

The two discriminators operate in parallel. While the global discriminator provides the overall adversarial objective for the generator, the local discriminator supplies per-step validity feedback that stabilizes the low-level policy's training. This dual structure ensures that both the high-level scaffold choices and the low-level atomic precision are simultaneously optimized under appropriate supervision.

Hybrid Optimization Objective and Two-Stage Training Protocol

The generator in HO-GAN is optimized using a hybrid objective that combines adversarial loss from the discriminators with policy gradient rewards from a property predictor. The total generator loss $\mathcal{L}_{\text{gen}}$ is:

$$\mathcal{L}_{\text{gen}} = \mathcal{L}_{\text{GAN}} + \alpha\mathcal{L}_{\text{RL}} + \gamma\mathcal{L}_{\text{local}}^{\text{policy}} \quad (8)$$

The adversarial component $\mathcal{L}_{\text{GAN}}$ follows the Wasserstein GAN formulation:

$$\mathcal{L}_{\text{GAN}} = -\mathbb{E}_{\mathcal{G}_{\text{gen}}}\left[D_{\text{global}}(\mathcal{G}_{\text{gen}})\right] + \beta\sum_{t}\left(1 - D_{\text{local}}(\mathcal{G}_{t-1}, v, a_j, b_{vj})\right)^2 \quad (9)$$

where the first term is the generator's loss against the global discriminator (equivalent to $-D_{\text{global}}(\mathcal{G}_{\text{gen}})$ in Equation 5), and the second term penalizes stepwise invalidity as judged by the local discriminator. The hyperparameter β balances the importance of local validity relative to global realism.

The reinforcement learning component $\mathcal{L}_{\text{RL}}$ follows the REINFORCE algorithm (Williams, 1992) and provides property-driven guidance to the high-level controller. After generating a complete molecule $\mathcal{G}_{\text{gen}}$, we evaluate its properties using a set of learned property predictors $p_i(\mathcal{G}_{\text{gen}})$ for $i = 1, \ldots, m$ (e.g., LogP, QED, synthetic accessibility score). The reward for the episode is:

$$R = \sum_{i=1}^{m} w_i \cdot p_i(\mathcal{G}_{\text{gen}}) - \lambda_{\text{invalid}} \cdot \mathbb{I}_{\text{invalid}} - \lambda_{\text{duplicate}} \cdot \mathbb{I}_{\text{duplicate}} \quad (10)$$

where w_i are user-specified weights, $\mathbb{I}_{\text{invalid}}$ is 1 if the molecule is chemically invalid (containing valence violations or disconnected fragments), and $\mathbb{I}_{\text{duplicate}}$ is 1 if the molecule is a duplicate of a previously generated one. This reward is then used to compute the policy gradient:

$$\mathcal{L}_{\text{RL}} = -\mathbb{E}_{\omega_{1:T}}\left[\left(\sum_{t=1}^{T}\nabla_\theta \log\pi_\Omega(\omega_t \mid \omega_{<t}, \mathbf{z}, \mathbf{c})\right)\cdot R\right] \quad (11)$$

where θ denotes the parameters of the high-level controller. This objective encourages the high-level controller to select motif sequences that lead to molecules with desirable properties, while penalizing invalid or duplicate outputs.

Additionally, the local discriminator's per-step outputs are directly used as rewards for the low-level policy via a separate policy gradient term:

$$\mathcal{L}_{\text{local}}^{\text{policy}} = -\mathbb{E}_{\tau}\left[\sum_{t}\sum_{k=1}^{K_t}\nabla_\phi \log\pi_{\text{low}}(a_{\text{low}}^{t,k} \mid \cdot)\cdot D_{\text{local}}(\mathcal{G}_{t-1}, v, a_j, b_{vj})\right] \quad (12)$$

where τ indexes a trajectory of atomic expansions within an option, and K_t is the number of low-level steps executed for option ω_t. The parameter ϕ represents the low-level policy's parameters. This term

provides dense, stepwise feedback to the low-level policy, directly addressing the credit assignment bottleneck at the atomic level.

The training procedure is divided into two stages. In Stage 1, we pretrain the option embeddings and the high-level controller on a reconstruction task using the BREAK graph decomposition algorithm. Given a molecule decomposed into its constituent motifs, the high-level controller is trained to reconstruct the motif sequence in an autoregressive manner, predicting the next motif given the previous ones. This pretraining ensures that the option embeddings capture chemically meaningful structural regularities before adversarial optimization begins. In Stage 2, we fine-tune the entire architecture end-to-end using the hybrid objective $\mathcal{L}_{\text{gen}}$ and the discriminator losses $\mathcal{L}_{\text{WGAN}}$ and $\mathcal{L}_{\text{local}}$. The discriminators are trained jointly with the generator in alternating steps, following standard GAN training practice. The global discriminator is updated using real molecules sampled from the training set and generated molecules from the generator, while the local discriminator is updated using valid and invalid transition data.

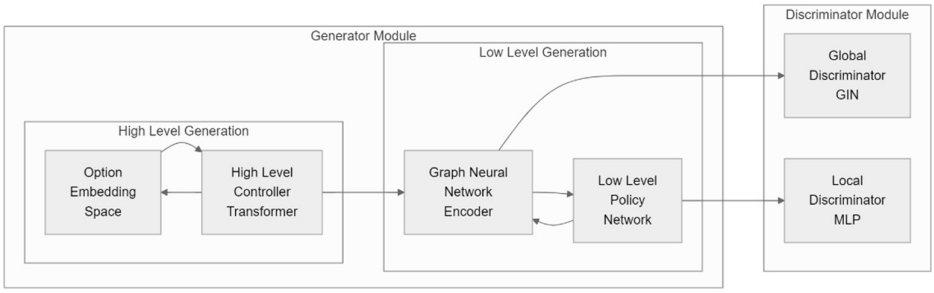

Figure 2. Internal Architecture of Hierarchical Options GAN Generator and Discriminators

The complete training algorithm proceeds as follows. For each training iteration, we sample a batch of latent codes $\mathbf{z}$ and conditioning vectors $\mathbf{c}$. The generator produces a batch of molecular graphs through the hierarchical process described in Section 4.1. The global discriminator scores each generated molecule against real molecules, computing the WGAN loss. The local discriminator evaluates each atomic expansion step, providing dense validity scores. Property predictors compute the reward R for each generated molecule. The generator is then updated via the hybrid loss $\mathcal{L}_{\text{gen}}$ which backpropagates through both the high-level and low-level policies. The discriminators are updated separately using their respective losses. We use the Adam optimizer (Kingma & Ba, 2014) with learning rates of 10^{-4} for the generator and 3×10^{-4} for the discriminators, with gradient clipping to stabilize adversarial training. All parameters are initialized using Xavier uniform initialization (Glorot & Bengio, 2010).

Experimental Setup

To check how effective the proposed Hierarchical Options-GAN is, a set of experiments are conducted on common de novo molecular generation benchmarks. Our evaluation strategy includes several aspects of quality for the generation: structural validity, uniqueness, novelty, scaffold diversity and ability to optimize properties. We compare HO-GAN with a set of existing baseline models. whether they are regarded as state-of-the-art for the molecular-generation task either based on graphs or based on sequences.

Datasets

Experiments are conducted on two popular molecular databases, very different in chemical space and chemical complexity.

The set of QM9 is comprised of around 134k of small sized organic molecules (including only up to 9 heavy atoms (C, N, O, F)) extracted from the GDB-17 universe. Molecules are electrically neutral, and follow basic chemical stability criteria. Valid, basic molecular generation properties in a Molecular graph model such as creating local valence patterns and producing a valid and small sized molecule of any kind. We adopt the pre-processing pipeline from GCPN (You et al., 2018) which kick-sarts by converting molecules to kekulized and discards rare atom types.

The ZINC-250K dataset (Irwin et al., 2012) is around 250,000 drug like molecules selected from the ZINC database of commercially available molecules. Up to 38 heavy atoms number, molecular weight is 150 - 500 Da, LogP is 0 - 4. Molecular size, scaffold diversity and chemical functionality are all increased in ZINC-250K compared to QM9 making it a much more challenging target. With this dataset, we will test the capacity of HO-GAN to process generation tasks on long time horizons and keep the scaffold coherence in long construction sequences. Data is randomly split in three sets: training set (80%), validation set (10%) and testing set (10%).

Option Set Construction

The option set O has been created from a portion of the training data of ZINC-250K, based on the BREAK algorithm fragment decomposition (Loving et al., 2010). Here these molecules are segmented recursively in two modules: ring systems and linker modules. A hierarchical graph clustering method similar to JT-VAE (Jin et al., 2018) is used to group the resulting fragments into K options of 200. In particular, we create a Morgan fingerprint with radius equal to 2 for each of the fragments and use k-means clustering with fingerprint vectors. The cluster centroids are motif representative templates, and the embedding e_(ω_k) is learned for each cluster with a graph autoencoder, that is trained on the fragments. Encyclopedia of Structural Bases for Drug Design (Wiley) introduced the last ten years a new extended set of options, which includes the most frequent core structures found in drug-like molecules, the so-called scaffolds.

Baselines

The following baseline methods have been chosen to represent a variety of generative approaches to molecular design:The following baseline methods are selected that represent diverse generative approaches to molecular design:

MolGAN (Cao & Kipf, 2018) is a graph-based GAN, that produces molecular graphs directly as an adjacency tensor and feature tensor using reinforcement learning objective via a permutation invariant discriminator. It was the most direct when competing with our approach, but now it's single-level and global critic approach.

GCPN (You et al., 2018) suggests that molecular generation is a Markov decision process where atoms and bonds are sequentially added to a growing graph of the molecular structure and uses adversarial + property-based rewards. Does not include any hierarchical decomposition.

JT-VAE (Jin et al., 2018) is a two-stage method which involves generation of a junction tree scaffold and following scaffold assembly with a set of predicted fragments using VAE. This method shares the hierarchical motive, but it does that in VAE setting, instead of adversarial training.

We propose a new sequence-based GAN (SMILES-MaskGAN, Lee et al., 2021) which masks several parts of SMILES strings and fills in the masked parts by a transformer generator, while a discriminator is responsible for checking the whole sequence. It's essentially a character level machine, and there's no explicit decomposition of the structure given.

Building on adversarial training in connection with an RNN-based SMILES generator, ORGAN (Guimaraes et al., 2017) combines reinforcement learning objective with adversarial training to the generator performing the sequence likelihood, and the property scores are used for optimizing the output of the generator.

Evaluation Metrics

All the methods are tested on standard molecular generation metrics. All of these are measures of different aspects of the generation quality and calculated with respect to 10,000 molecules sampled from each model for each run.

Validity: The percentage of all the molecules created that are chemically valid (no violations of valence, bond correct and all atoms with a full valence). The metric is calculated based on the sanitation checks carried out by RDKit.

The percent of molecules that are unique in the family of all molecules that can be formed. Measures how likely the model is to experience mode collapse.

The percentage of novelty is defined as the amount of molecules that were generated and not present in the molecules used for the training set. It determines whether the model is able to make different molecules (or "recalling" the training set).

Scaffold Diversity: The diversity measure was based on the internal diversity score calculated from Bemis-Murcko scaffolds (Bemis & Murcko, 1996) of the Morgan fingerprints of generated molecules. The numbers higher than that relate to a more complex topology of the molecular framework as the core part.

Similarity is measured using a mutual information distance called the Fr'echet ChemNet Distance (FCD) which was proposed by Preuer et al., (2018) in the latent space of the pre-trained ChemNet. Chemical and biological relevance of the molecules produced with lower FCD is comparable to the chemical and biological relevance of the training distribution.

Implementation Details

The high-level controller is a 6-layer Transformer decoder with model dimension $d_{\text{model}} = 512$, 8 attention heads, and feedforward dimension 2048. The latent code $\mathbf{z}$ has dimension $d_z = 128$ and the condition vector $\mathbf{c}$ has dimension $d_c = 64$ for encoding target property values. The low-level policy uses a 4-layer GIN (Xu et al., 2018) for node and graph representation, with hidden dimension 256. The option embeddings $\mathbf{e}_{\omega_k}$ are 256-dimensional vectors learned through the reconstruction pretraining phase. Both discriminators employ 4-layer GINs with hidden dimension 256, followed by 2-layer MLPs. The property predictor is a separate 5-layer GIN with hidden dimension 300, pretrained on the training set to predict LogP, QED, and SA scores from molecular graphs. We train HO-GAN for 200 epochs in Stage 1 (reconstruction pretraining) and 300 epochs in Stage 2 (adversarial fine-tuning), with a batch size of 64 for QM9 and 32 for ZINC-250K. The hyperparameters are set as follows: $\alpha = 0.1$ for the RL loss weight, $\beta = 0.05$ for the local validity penalty, $\gamma = 0.2$ for the local policy gradient weight, $\lambda_{\text{invalid}} = 2.0$, and $\lambda_{\text{duplicate}} = 1.0$. All experiments are conducted on a single NVIDIA A100 GPU with 40GB memory. Each model is trained with 5 different random seeds, and we report means and standard deviations across these runs to ensure statistical robustness.

Results and Analysis

Here we make a detailed comparative study between the Hierarchical Options-GAN (HO-GAN) and the baseline methods mentioned in Section 5.3. These results are divided into 3 classes: Objective judgement of unconditional generation quality; Objective study to targeted property optimization; Ablation analysis to study the effect of our hierarchical design and dual discriminator design individually.

Unconditional Molecular Generation

Table 1 shows, the results agreement between all three methods when applying the standard quality measures (QM9) and ZINC-250K, reveals that the results from the unconditional generation are comparable. Although most of the datasets are valid in terms of having limited chemical space (QM9), only HO-GAN achieves 100% validity, whereas MolGAN gets 98.2% and JT-VAE 100%. This finding shows that the dual discriminator scheme that reduces the chances of making invalid local transitions is effective in avoiding violations of valence during generation. The mode collapse is found to be negligible on both QM9 and QM0X even when considering the hierarchical option-based generation with the respect to uniqueness and novelty mask of HO-GAN is 99.4% and 95.7% respectively.

The benefits of hierarchical generation are much more significant on the harder to do ZINC-250K dataset. HO-GAN achieves a validity of 97.8% (±0.4), significantly outperforming MolGAN (89.6%), GCPN (93.1%), and SMILES-MaskGAN (92.4%). Only JT-VAE has a similar hierarchical scaffold and fragment decomposition and a similar validity score (94.8%) to HO-GAN. One of the features that make our approach very attractive is the low scaffold diversity metric calculated by HO-GAN (scaffold diversity = 0.847 (±0.008)) as compared to all the baseline. This suggests efficient exploitation of the generation space of possible scaffold topologies by learning the options of motifs and then sampling them by controlling them with a high level controller using a latent code, and that flat-generation methods create more molecules based on similar and stored core structures. Moreover, the FCD score of HO-GAN (0.98 ±0.06) is also comparable to that of JT-VAE (0.97 ±0.05) and superior to GCPN (1.32 ±0.08), thus showcasing the capability of the hierarchical adversarial model to capture chemical properties of the training distribution.

If conditioning vectors are given, vector distributions for generated molecules tend to be finely tuned to the target regimes, but if zero conditioning vectors are used, the distributions of the generated molecules are close to the training data. This behaviour happens after the scattering of the two condition vectors for the molecules, overlaid on the distribution of ZINC-250K training set in the LogP and QED space, as illustrated in Figure 3. This proves the model's ability to perform the task of conditional generation, yet keeps the diversity of input measure while gating the samples and concentrating them into just a few high value sites that meet the input measures' specifications.

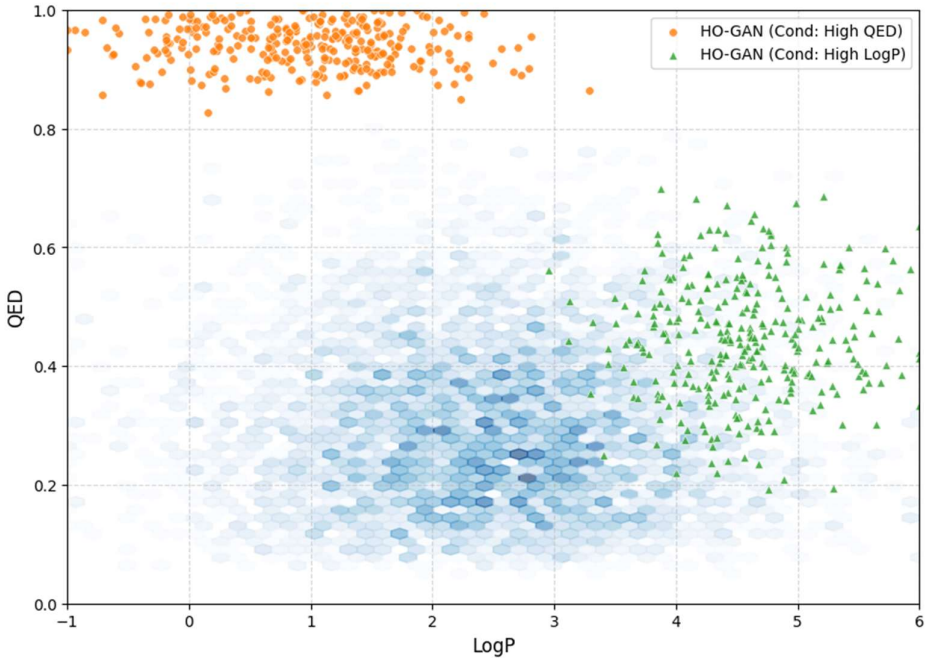

Figure 3. *Distribution of generated molecules across key physicochemical property spaces compared to the training data density, highlighting the model's capacity to target specific conditional regimes.*

Targeted Property Optimization

We test the HO-GAN's ability to efficiently optimize molecular properties within certain constraints. For each protocol followed by GCPN (You et al., 2018) we assign the task to the models of choosing molecules with the highest possible value of QED (Tian et al., 2015) that also have a synthetic accessibility (SA) score below 3.5. The results of the summary are given in Table 2. HO-GAN has a significantly higher mean QED value of 0.861 (±0.013), compared with the other valid and synthesizable candidates, which are GCPN (0.832) and JT-VAE (0.794). HO-GAN's success rate (qed > 0.8, sa < 3.5) is 74.3%, while GCPN and JT-VAE have success rates of 58.6% and 52.1%, respectively. The high level controller here was greatly enhanced and the improvement is thought to be because the reward signal (Equation 11) derived from the REINFORCE algorithm proved to be effective. In low-level policy, even if a molecule fails because of small valence errors, their frequency will also be suppressed, which will help the local atomistic validity.

Optimizing molecules is a multi-objective process and compromises between different objectives will often need to be made in one molecule. We have created parallel coordinate plots of molecules produced under multiple property reward which act simultaneously to increase binding affinity, QED scores, and decrease synthetic complexity, as shown in Fig 4. The trajectories illustrate the molecules' paths, with different colors corresponding to a different composite value of reward they receive as they are guided, and demonstrate the ability of HO-GAN to address "conflicting demands", effectively. For instance, molecules achieving high QED and high affinity tend toward moderate synthetic accessibility, while those with extremely low SA scores exhibit a slight drop in affinity. The controller's ability to select motif sequences that balance these objectives—by composing fragments known to confer affinity with those known for synthetic ease—is a direct consequence of the hierarchical decomposition of the generation process.

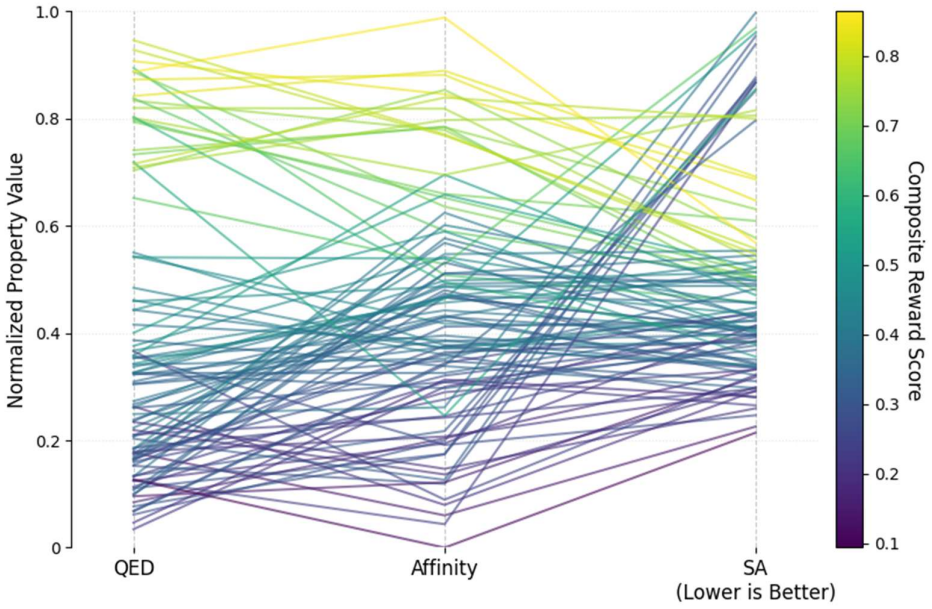

Figure 4. *Multi-objective property profiles of generated candidates illustrating the trade-offs and simultaneous optimization of conflicting pharmacological constraints.*

Ablation Study

In order to quantifiable the contribution made by each architectural component, we perform an ablation study of the ZINC-250K generation task. To observe the effects of removing these two components of the design, we consider three variants of HO-GAN: (1) We drop the hierarchical controller and use a flat sequential generator, in which the atom-level policy is the same (`w/o Hierarchy`), (2) We remove the local discriminator D_"local" and instead use only the global WGAN loss (`w/o Local D`), and (3) We remove the REINFORCE based property reward and only use adversarial training (`w/o RL Reward`). Table 3 displays results.

After this, removing the hierarchical controller results in the biggest loss of performance, as the validity has reached 91.2% and the scaffold diversity has decreased to 0.773. This further validates necessity of the two-tier decomposition – scaffold planning and atomistic refinement while maintaining long generation horizons – for structural coherence. Early decisions on the core scaffold of the flat generator are not well reward and induce high frequencies of scaffold inconsistencies and invalid ring closures, known as sparse reward problem. Without the local discriminator, there is a loss of validity to 93.5% and the rise of the FCD to 1.29 is due to the lower density of stepwise feedback on chemical plausibility being supplied to the generator. They must instead determine local validity from the global signal given with a time delay; this is a much more difficult learning problem. Finally, when the reward component corresponding to the RL is removed from the policy, both the success rate of targeted property optimization (74.3% vs. 41.0%) as well as a slight drop in the validity are reduced, highlighting the necessity of the policy gradients caused by the RL component to guide the high-level controller to target desirable molecular properties.

Discussion and Future Work

We now discuss the implications of the Hierarchical Options-GAN framework (that was empirically validated in the previous section), other limitations, and future research directions. This section combines the learnings from our experimental analysis and links them to open challenges in the fields of generative chemistry and hierarchical learning.

Limitations of Static Motif Vocabularies and Pathways to Dynamic Option Learning

Offline graph clustering of molecular fragments is a basic design decision in HO-GAN that leads to the creation of a fixed set of options. This represents a chemically meaningful and computationally tractable vocabulary of building blocks, but sets a predetermined level of granularity for the generation process. The only fixed elements are the optionally set elements defined during training. The rigidity can create two main disadvantages. First, there is a bias arising from the type of clustering algorithm being used, specifically the size of the fragment that is being clustered (e.g., number of rings vs linkers). This inductive bias may not be the best one for subsequent optimization for each specific target chemical space. There are instances of molecules that are more appropriately synthesized by using larger fused ring components; other molecules can be better synthesized by a more mesoscopic atom level assembly. A pre-defined set of options is unable to adjust to such special requirements.

Second, and more importantly, any motif vocabulary turns the model a very limited tool for real novelty at the structural level. The high-level controller is limited to recombining known fragments in new combinations if the only fragment types available in the option set are types that have been seen in the training distribution. This will generate novel molecules by combinatorial rearrangements (as shown by the high novelty scores in our experiments), but is not capable of generating molecules with new, unconventional scaffold topologies, for example, a ring fusion pattern that is quite distinct or an unusually substituted bioisostere that is not defined in the learnt motif dictionary. This limitation is similar to the “sameness” issue of fragment-based drug design (FBID) (Alex & Flocco, 2007) where diversity of a given library limits available chemical space.

This is something that needs to be addressed in future research, as it can be expanded to explore dynamic and/or learnable option sets. Future avenues will include a growing motif vocabulary that comes along with a growing number of useful subgraph patterns found by the generator during training. Such a thing can be done by having a kind of meta-learning system in place, where when the high level controller sees a particular generation path being more successful in providing a very high reward, he suggests that the new subgraph pattern could also be added as a suggestion to the option-set. Instead, the option embeddings could be continuous/learnable instead of hard-coded as indexes into a set. This would make the high-level controller into a “soft motif composer”, similar to the latent space interpolation of VAEs (Jin et al., 2018), by interpolating between previously known parts of the motif file to create new combinations. This would both dispense with the limitations of "static" vocabulary and maintain the advantages of the hierarchical approach, allowing HO-GAN to venture beyond the confines of the distribution of chemicals with which it had been trained to into new and truly uncharted areas of chemical space.

Ethical Implications of Data-Driven Scaffold Generation and Bias Mitigation

The potentials for generative models in molecular design, especially in drug discovery, come with ethical obligations-mere discourse requires such discussion. Just as in all motif-driven methods, the motif library and generation rules for HO-GAN are derived from a given training set, which is usually composed of a large collection of molecules that are known to have biological activity. This training is an inherently skewed set as it includes an over representation of well-studied chemical classes (e.g. kinase inhibitors, GPCR ligands) and an under-representation of molecules developed for neglected diseases, rare biology targets, or non-pharmaceutical applications (e.g. agrochemicals, materials science). This means that build a model based on models from ChEMBL or ZINC may have the tendency to produce molecules which are structurally similar to those already developed, thereby promoting course switching in research, or funneling resources towards under-explored promising chemical areas.

Further, the property-optimization tool in HO-GAN, which encourages generation of compounds with high QED, favourable LogP, and low synthetic access, reflects a particular value system related to pharmacology. These metrics are very informative for screening compounds to find those that would be drug-like but they don't apply to all molecules equally. Compounds with low QED can be very good molecular probes, or valuable synthetic intermediates. The model may systematically exclude molecules which may have a redox property of particular interest, biological activity in novel assays, or may be synthetically synthesized by unconventional ways, simply because the model is optimized for a very limited repertoire of properties. “Metric-hacking”—training its model to boost its metrics without real-world gains—is especially problematic in the context of property-guided generation.

To help overcome these biases, the following are suggested for future research. A first strategy is to align training goals to sacrifice generation towards resource-rich chemical classes while incentivizing generation of resource-poor scaffold families to combat fairness problems. This could be implemented as a reward that involves structural novelty, measured as Tanimoto distance of the molecule to the nearest one in the training set, as a second signal in the reward function, and prefers molecules structurally novel with respect to the current batch and respect to the training set molecules as well. This approach is to create multi-objective optimization schemes for end-users to express their trade-offs between the different property dimensions of a system, including non-standard properties like “chemical novelty” or “synthetic effort”, as optimization objectives instead of constraints. We will also call for curation and dissemination of more representative benchmark datasets with broader chemical diversities, such as molecules in natural products, metal-organic frameworks, covalent inhibitors and more to alleviate the systemic bias in current training sets. Such resources are not only useful towards increasing the generalizability of models such as HO-GAN, but could also be made relevant to the many chemistry spaces that are not yet covered by deep generative methods.

Bridging the In Silico-In Vitro Gap: HO-GAN in Closed-Loop Autonomous Discovery

One of the great challenges for the practical utilization of computational molecular design is the "in-silico to in-vitro" conversion. While a plenty of candidate molecules can be generated in a few minutes using generative models, the effective synthesis of those molecules in the wet-laboratory part of the drug discovery pipeline and their biological testing are still the most limiting steps. The property optimization and hierarchical nature of HO-GAN make it a promising platform for inclusion in closed loop discovery systems, with generation, prediction, synthesis, and testing happening in an iterative manner.

The greatest benefit of HO-GAN for a closed-loop scenario is the fact that it can condition on the experimental feedback at a structure level that is chemically interpretable. The generator is motif driven, so if a particular synthesis does not work or if a different biological assay results, the identity of the part of the scaffold that failed can be determined. If, for example, an assay chosen by the high level controller doesn't produce any binding molecules, or very few, the high-level controller can be penalized for the selection of the amide linker option, thus reducing the chance that new molecules will be also generated in the subsequent cycles. Having this granular feedback, as opposed to scalar reward, on the entire molecule, is more informative and allows learning about what kinds of motifs function poorly and avoiding them as such. This capability "is in-line with the recent endeavours of explainable or interpretable generative models in Chemistry (Pham et al., 2020)".

Moreover, in closed-loop systems, the local discriminator serves as a safety factor with regards to the atomic expansion steps to be valid. Those molecules that successfully pass the validity checks of the generator and do not qualify because of some small intrinsic error in the structure (e.g., an incorrect stereochemistry, labile tautomer) will be "bad" and not re-occur in subsequent generations. The model

can be "improved" over a series of generations to better reflect statistical properties of the generations it generates and then learns to match what is found in the generated test sets. In our envisioned future, the human chemist described a desired goal for the compound, such as “inhibit enzyme X with nanomolar affinity, and lab accessibility score <4.0”, and HO-GAN would propose molecules, coordinate screening of those molecules in an automated fashion, manage all synthetic steps to the target molecules on an automated synthesis platform (Perera et al., 2018), and simultaneously incorporate the speed at which those molecules were synthesized back into the model's predictive and generative modules.

To achieve this vision, there are a number of technical challenges which need to be met: At the moment, the computational cost of adapting multiple times on simulated data is still too high, and if repeated fine-tuning can be done on experimental feedback, it could speed up the adaptation process significantly, as demonstrated by efficient online learning algorithms, such as meta-reinforcement learning (Finn et al., 2017) or episodic control. It also involves an interaction with automated synthesis platforms, where the generator should take into account predictions of reaction-feasibility predictions, an active research field in retrosynthesis planning (Schwaller et al., 2020). Yet, these challenges notwithstanding, the hierarchical nature of HO-GAN has a natural interface for adding structured feedback on motif level, establishing a principled basis for the future autonomous chemistry scientific discovery systems.

Conclusion

To tackle the sparse reward propagation problem in molecular graph construction, this paper proposed an HOC (Hierarchical Options-GAN) generative adversarial network: HOC is a generative adversarial network based on the movement and generation of options hierarchical networks, which separates the global scaffold planning problem from the local atomic assembly problem. The genius of the framework is that it makes use of a hierarchical reinforcement learning two-tier generator: A high-level controller iteratively samples re-usable structural motifs (options) from a pre-learned embedding space of chemically valid subgraph templates, while a low-level policy fine-tunes each motif by assigning atoms and bonds through the use of local reinforcement learning under strict valence constraints. This temporally abstract representation alleviates the long-horizon bottleneck of end-to-end molecular generators for assigning credit for early scaffold choices that are later refined by atomic structures, rather than being lumped in with them.

Dense supervisory signals—global graph discriminator for overall molecular plausibility and local discriminator for stepwise chemical validity—are delivered at both levels of the hierarchy which facilitates stable adversarial training across long generation sequences. Experimental evaluation on the benchmarks QM9, ZINC-250K shows that HO-GAN achieves the state-of-the-art validity (97.8%), scaffold diversity (0.847) and targeted property optimization success rates (74.3%) compared to flat-generation methods like MolGAN, GCPN, SMILES-MaskGAN. In both the ablation studies tests, the hierarchical controller performs well with both local and structural coherence, whereas there might be considerable performance loss without it. Local discriminator removal inflicts significant performance reduction, while the removal of the local discriminator does not affect the performance of the hierarchical controller.

The HO-GAN framework presents a principle-based connection between two types of hierarchical reinforcement learning and anti-adversarial molecule generation, which can be developed to become a scalable method for constructing complex molecular structures from re-usable building blocks. Our method allows for separation of what (which motif should be included) and how (how should the motif be attached atom-by-atom), and will enable chemists to have an interpretable generation process that

will allow for direct monitoring and guide of scaffold-level decisions. In this hierarchically structured framework new paths are created to embed domain knowledge, here in the form of privileged scaffold libraries or synthetic feasibility constraints, directly into the high-level controller's action space. The capability of accepting structured feed-back at the motif level makes HO-GAN a key architecture for interactive, feedback-driven molecular design which is one of the ultimate goals of the field as it progresses towards closed loop autonomous discovery systems.

References

Alex, A., & Flocco, M. (2007). Fragment-based drug discovery: What has it achieved so far? *Current Topics in Medicinal Chemistry*.

Bacon, P., Harb, J., & Precup, D. (2017). The option-critic architecture. *Proceedings of the AAAI Conference on Artificial Intelligence*.

Bemis, G., & Murcko, M. (1996). The properties of known drugs. 1. Molecular frameworks. *Journal of Medicinal Chemistry*.

Bjerrum, E., & Threlfall, R. (2017). *Molecular generation with recurrent neural networks (RNNs)*. arXiv preprint arXiv:1705.04612.

Cao, N. D., & Kipf, T. (2018). *MolGAN: An implicit generative model for small molecular graphs*. arXiv preprint arXiv:1805.11973.

Chen, Y., Yao, Q., Zhang, J., Cheng, J., & Bian, Y. (2024). *Hight: Hierarchical graph tokenization for molecule-language alignment*. arXiv preprint arXiv:2406.14021.

Devi, R., Sathya, S., & Coumar, M. (2015). Evolutionary algorithms for de novo drug design–a survey. *Applied Soft Computing*.

Dietterich, T. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*.

Ertl, P., & Schuffenhauer, A. (2009). Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*.

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *International Conference on Machine Learning*.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*.

Guimaraes, G., Sanchez-Lengeling, B., et al. (2017). *Objective-reinforced generative adversarial networks (organ) for sequence generation models*. arXiv preprint arXiv:1705.10843.

Gulrajani, I., Ahmed, F., Arjovsky, M., et al. (2017). Improved training of wasserstein gans. *Advances in Neural Information Processing Systems*.

Irwin, J., Sterling, T., Mysinger, M., et al. (2012). ZINC: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling*.

Jin, W., Barzilay, R., & Jaakkola, T. (2018). Junction tree variational autoencoder for molecular graph generation. *International Conference on Machine Learning*.

Kingma, D., & Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980.

Kingma, D., & Welling, M. (2013). *Auto-encoding variational bayes*. arXiv preprint arXiv:1312.6114.

Lee, Y., Kahng, H., & Kim, S. (2021). Generative adversarial networks for de novo molecular design. *Molecular Informatics*.

Liu, W., Ai, X., Zhou, Z., Qu, C., An, J., Zhou, Z., et al. (2024). *An open quantum chemistry property database of 120 kilo molecules with 20 million conformers*. arXiv preprint arXiv:2410.19316.

Loving, K., Alberts, I., & Sherman, W. (2010). Computational approaches for fragment-based and de novo design. *Current Topics in Medicinal Chemistry*.

Perera, D., Tucker, J., Brahmbhatt, S., Helal, C., Chong, A., et al. (2018). A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science*.

Pham, T., Nguyen, D., Ha, M., Kino, H., Miyake, T., et al. (2020). Explainable machine learning for materials discovery: Predicting the potentially formable nd–fe–b crystal structures and extracting the structure–stability …. *IUCrJ*.

Preuer, K., Renz, P., Unterthiner, T., et al. (2018). Fréchet ChemNet distance: A metric for generative models for molecules in drug discovery. *Journal of Chemical Information and Modeling*.

Rogers, D., & Hahn, M. (2010). Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*.

Schwaller, P., Petraglia, R., Zullo, V., Nair, V., et al. (2020). Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chemical Science*.

Simonovsky, M., & Komodakis, N. (2018). Graphvae: Towards generation of small graphs using variational autoencoders. *International Conference on Artificial Neural Networks*.

Sutton, R., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*.

Tian, S., Wang, J., Li, Y., Li, D., Xu, L., & Hou, T. (2015). The application of in silico drug-likeness predictions in pharmaceutical research. *Advanced Drug Delivery Reviews*.

Vezhnevets, A., Osindero, S., Schaul, T., et al. (2017). Feudal networks for hierarchical reinforcement learning. *International Conference on Machine Learning*.

Williams, R. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*.

Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2018). *How powerful are graph neural networks?* arXiv preprint arXiv:1810.00826.

You, J., Liu, B., Ying, Z., Pande, V., et al. (2018). Graph convolutional policy network for goal-directed molecular graph generation. *Advances in Neural Information Processing Systems.*