De Novo Generation of Odorant Molecules with Targeted Olfactory Receptor Activation Patterns

Kexin Zhang [1] **Manhao Guan** [1 2]

Abstract

Odor perception arises from a combinatorial code in which each molecule activates a specific subset of olfactory receptors (ORs), and the resulting activation pattern determines the perceived smell. We present a framework for de novo generation of molecules predicted to match target multi-receptor activation profiles. We first construct a panoramic binding matrix over 20 ORs, then cluster it to discover representative activation patterns, and finally formulate a structured selectivity loss, combining pairwise ranking, distribution matching, and complexity constraints, to steer a molecular generator via reinforcement learning. Across nine target profiles of varying complexity, our method achieves high predicted selectivity within 18,000 oracle calls and provides the largest gains on profiles that are rare under the prior, especially broad-spectrum and sparse-activator targets.

1. Introduction

The mammalian olfactory system employs a combinatorial coding strategy to discriminate among a vast chemical space of volatile molecules (Malnic et al., 1999; Kurian et al., 2021). Each odorant activates a characteristic subset of the approximately 400 human olfactory receptors (ORs), and the identity of a perceived smell is encoded not by any single receptor but by the ensemble activation pattern across the receptor repertoire (Buck, 2004; Kurian et al., 2021). This combinatorial logic enables a limited receptor repertoire to encode an enormous diversity of odor percepts.

Despite growing interest in computational approaches to molecular design (Du et al., 2024), most existing work on odorant generation has focused on single-target objectives: optimizing a molecule's predicted binding affinity to one receptor or its similarity to a known odorant (Sanchez-Lengeling et al., 2019; Lee et al., 2023). Such formulations neglect the multi-receptor nature of odor coding. A molecule that strongly activates a single target receptor may simultaneously activate unintended receptors, producing an entirely different odor percept than intended.

We argue that rational odorant design should instead target the *activation profile*, the full pattern of which receptors a molecule activates and which it does not. This reframes the problem from single-objective optimization to multi-receptor selectivity control, directly reflecting the biological mechanism of odor encoding.

To this end, we propose a three-stage framework (Figure 1):

The framework proceeds in three stages. We first perform **panoramic profiling** by predicting the binding probability of a diverse molecular library against a panel of 20 ORs, constructing a molecules $\times$ receptors binding matrix. We then apply **pattern discovery**, clustering this matrix to identify naturally occurring activation patterns shared by groups of molecules with similar multi-receptor profiles. Finally, we perform **pattern-targeted generation**, defining a selectivity reward that measures how well a generated molecule's predicted profile matches a chosen target pattern and using reinforcement learning (RL) to steer a molecular generator accordingly.

We evaluate our approach on nine target profiles of increasing complexity, from narrow selectivity over two receptors to broad-spectrum activation of fourteen. Our contributions are: (1) we formulate odorant generation as *profile-targeted optimization* over a multi-receptor olfactory code rather than single-receptor binding; (2) we introduce a *structured selectivity objective* combining ON/OFF ranking, response-distribution matching, and activation-complexity matching, providing a principled alternative to heuristic reward engineering; and (3) we evaluate RL-based molecular generation across nine diverse profiles and show that directed generation provides the largest gains for broad-spectrum and sparse-activator profiles, where screening from the prior distribution is insufficient, revealing a "designability gradi-

[1]College of Computer Science and Technology, Zhejiang University, Hangzhou, China [2]College of Artificial Intelligence, Zhejiang University, Hangzhou, China. Correspondence to: Manhao Guan <guanmh@zju.edu.cn>.

Accepted at the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026)

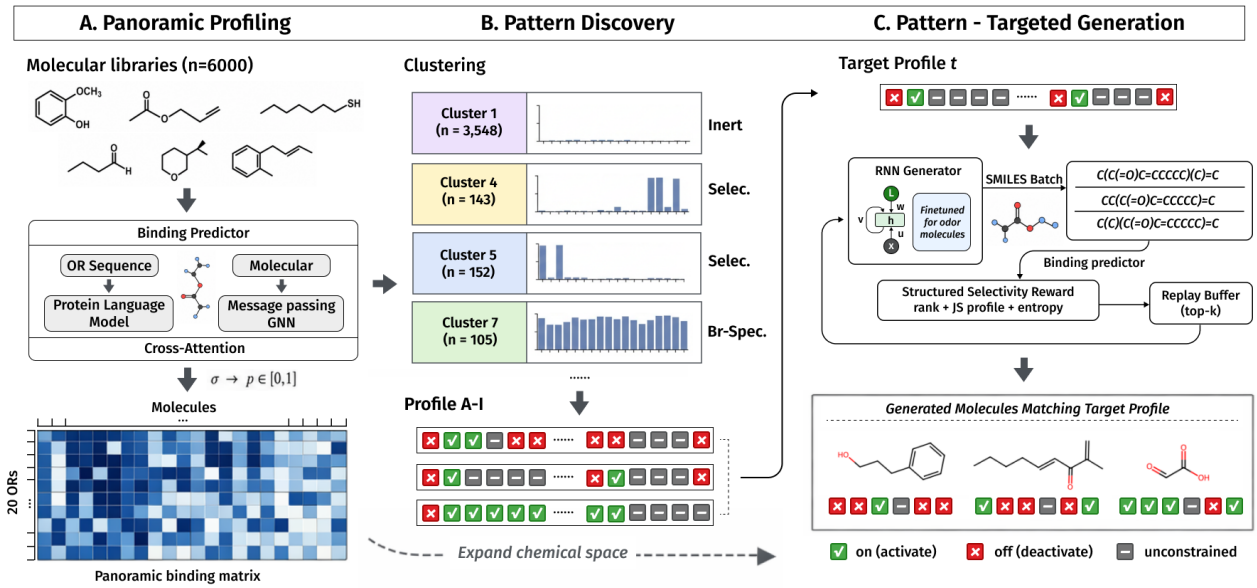

Figure 1. Overview of our framework. **Stage 1:** Panoramic profiling predicts binding probabilities of a molecular library against 20 ORs, producing a binding matrix. **Stage 2:** Clustering discovers representative activation patterns, interpreted as combinatorial receptor codes. **Stage 3:** A target pattern is selected and used to define a structured selectivity reward combining ON/OFF ranking, response-distribution matching, and activation-complexity matching. RL with augmented memory steers the generator toward molecules predicted to match the target profile.

ent" relating profile complexity to optimization difficulty. We additionally explore a Lagrangian constrained variant that can accelerate simple selective profiles but becomes unstable for complex multi-constraint targets.

2. Related Work

Olfactory receptor modeling. Predicting odorant–receptor interactions computationally has been approached through molecular docking (Charlier et al., 2013), learned perceptual representations (Sanchez-Lengeling et al., 2019; Shuvaev et al., 2024), and graph neural networks operating on molecular structure (Hladis et al., 2023). Recent work on odorant featurization has explored fine-tuning chemical foundation models to produce olfaction-specific representations (McConachie et al., 2025). The principal odor map (Lee et al., 2023) demonstrated that learned molecular embeddings can unify diverse olfactory tasks. The M2OR database (Lalis et al., 2024) provides a large curated collection of OR–molecule interaction data, enabling supervised training of binding predictors across hundreds of receptors. Our work builds on these advances by using a trained binding predictor as an oracle within a generative loop, rather than as a standalone prediction tool.

De novo molecular generation. Deep generative models for molecules span variational autoencoders (Gómez-Bombarelli et al., 2018), graph-based policy networks (You et al., 2018), and autoregressive models over SMILES strings (Segler et al., 2018). Reinforcement learning approaches, particularly REINVENT (Olivecrona et al., 2017; Blaschke et al., 2020) and its Transformer-based extensions (Xu et al., 2024), have shown strong performance in goal-directed generation by treating a pretrained language model as a policy and optimizing it against a scoring function. Recent advances in sample efficiency include augmented memory (Guo & Schwaller, 2024), which replays high-scoring molecules with diverse SMILES representations, and the Mamba-based Saturn architecture (Guo et al., 2026), which further improves oracle efficiency. Latent-space RL methods (Haddad et al., 2025) offer an alternative by optimizing in continuous space. Du et al. (2024) provide a comprehensive survey of machine learning-aided generative molecular design. We adopt the SMILES-based RL paradigm but extend the reward formulation from single-property optimization to multi-receptor profile matching.

Multi-objective molecular optimization. Prior work on multi-objective generation has addressed combinations of drug-likeness, synthetic accessibility, and target affinity (Jin et al., 2020; Thomas et al., 2022; Guo & Schwaller, 2026). However, optimizing a molecule's binding profile across a *panel* of related biological targets, where some targets should be activated and others suppressed, has received limited attention. Known failure modes in molecular optimization (Renz et al., 2020) further motivate careful reward design. Our structured selectivity objective addresses this gap by combining profile-level matching with pairwise rank-

ing between ON and OFF receptors, drawing on ideas from learning-to-rank (Liu, 2009) to enforce relative selectivity rather than relying solely on absolute binding predictions.

3. Methods

3.1. Receptor–Odorant Binding Predictor

We employ a dual-stream architecture to predict the binding probability between an odorant molecule and an OR sequence. The receptor stream encodes the OR amino acid sequence using a pretrained protein language model (ProtBERT; Elnaggar et al., 2021), producing a fixed-dimensional receptor embedding. The molecular stream represents the odorant as a graph where atoms are nodes and bonds are edges, processed by a message-passing neural network (MPNN) with multi-head attention layers (Gilmer et al., 2017). At each message-passing step, the receptor embedding is injected into every atom node via a cross-attention mechanism, enabling the network to learn receptor-specific molecular representations. After five rounds of message passing, a graph-level readout followed by a sigmoid output layer produces a binding probability $p \in [0, 1]$. The model is trained on the M2OR dataset (Lalis et al., 2024), which contains experimentally validated OR–molecule interactions across 566 receptors. We select 20 receptors with sufficient training data (≥ 30 positive interactions) for our panel. We denote the trained predictor as $f_r(m)$: the predicted binding probability of molecule m to receptor r.

Predictor validation. We evaluate the binding predictor on held-out annotated receptor–molecule pairs using a pair-level train/validation/test split. For each positive interaction in the held-out set, we sample three unannotated receptor–molecule pairs as negatives. Because OR interaction datasets are sparsely labeled, these random negatives may include unknown binders; thus the reported metrics should be interpreted as approximate surrogate-validation metrics rather than definitive binding performance. Across our 20-receptor panel, the predictor achieves a mean AUROC of 0.814 ± 0.124 and mean AUPRC of 0.654 ± 0.137. Performance varies by receptor: high-promiscuity receptors with many annotated interactions (e.g., s_449, s_51) show lower AUROC (0.58–0.66), likely due to the difficulty of distinguishing true binders from weakly interacting or unlabeled molecules, while specialist receptors (e.g., s_1107, s_177) achieve AUROC > 0.91. All selectivity scores reported in this work are *predicted* values and require experimental validation.

3.2. Molecular Generator

Our molecular generator is a recurrent neural network (RNN) that autoregressively produces SMILES strings token by token (Segler et al., 2018). The architecture consists of a multi-layer LSTM with a conditioning mechanism: at each decoding step, a receptor pocket embedding $\mathbf{z} \in \mathbb{R}^d$ is concatenated to the input, biasing generation toward molecules compatible with the target receptor context. The model is first pretrained on a large corpus of known odorant SMILES to learn the distribution of valid, chemically plausible odorant molecules, then fine-tuned with RL.

3.3. Panoramic Profiling and Pattern Discovery

To map the landscape of OR activation patterns, we sample 6,000 molecules from the pretrained generator, retain 5,816 valid molecules after RDKit parsing, and predict each valid molecule's binding probability against all 20 receptors, yielding a binding matrix $\mathbf{B} \in \mathbb{R}^{5816\times 20}$ where $B_{ij} = f_{r_j}(m_i)$.

We cluster this matrix using k-means to identify K representative activation patterns (receptor response prototypes) $\{\mathbf{c}_1, \ldots, \mathbf{c}_K\} \subset \mathbb{R}^{20}$. Each molecule's activation profile can be soft-assigned to prototypes via temperature-scaled cosine similarity:

$$a_k(m) = \frac{\exp(\mathrm{sim}(\mathbf{f}(m), \mathbf{c}_k)/\tau)}{\sum_{j=1}^{K} \exp(\mathrm{sim}(\mathbf{f}(m), \mathbf{c}_j)/\tau)} \tag{1}$$

where $\mathbf{f}(m) = [f_{r_1}(m), \ldots, f_{r_{20}}(m)]$ is the full binding profile and τ is a temperature parameter. These prototypes serve as a codebook of canonical activation patterns, enabling target specification via prototype selection or interpolation. In our experiments, we use the cluster centroids directly to define target profiles.

3.4. Structured Combinatorial Code Loss

We design a structured selectivity objective that captures three aspects of combinatorial receptor coding: relative selectivity, response geometry, and activation complexity.

Pairwise ranking selectivity. The core selectivity requirement is that ON receptors should be activated *more than* OFF receptors. We encode this via a pairwise margin-based ranking loss. For each (ON, OFF) receptor pair (r^+, r^-):

$$\mathcal{L}_{\text{rank}}(m) = \sum_{(r^+,r^-)\in\mathcal{P}} \log\Big(1 + e^{-(f_{r^+}(m) - f_{r^-}(m) - \delta)}\Big) \tag{2}$$

where $\mathcal{P}$ is the set of all ON–OFF pairs and $\delta = 0.3$ is a margin. This is converted to a reward $S_{\text{rank}} \in [0, 1]$ by normalizing against the maximum possible loss. Unlike absolute regression, this formulation depends on relative ordering and is robust to calibration error in the binding predictor.

Profile distribution matching. We treat the receptor response as a distribution over the receptor panel and measure

how well the generated molecule's response matches the target profile geometry. Using Jensen-Shannon divergence:

$$S_{\text{dist}}(m) = 1 - \text{JSD}(\hat{\mathbf{p}}(m) \,\|\, \mathbf{p}_{\text{target}}) \tag{3}$$

where JSD denotes the Jensen-Shannon divergence computed with base-2 logarithms (so that JSD $\in [0, 1]$), and $\hat{\mathbf{p}}(m)$, $\mathbf{p}_{\text{target}}$ are the ℓ_1-normalized binding profiles. This captures the full response geometry rather than treating receptors as independent binary targets.

Activation complexity matching. Different target profiles have different activation complexity: narrow profiles activate few receptors (low entropy) while broad profiles activate many (high entropy). We penalize deviation from the target's activation entropy:

$$S_{\text{complex}}(m) = 1 - |H(\hat{\mathbf{p}}(m)) - H(\mathbf{p}_{\text{target}})| \tag{4}$$

where $H(\cdot)$ is the normalized Shannon entropy. This encourages the generator to match not just *which* receptors are active, but the *shape* of the combinatorial code.

Combined structured loss. The final selectivity score combines all three components:

$$S(m) = \alpha_{\text{rank}} \cdot S_{\text{rank}} + \alpha_{\text{dist}} \cdot S_{\text{dist}} + \alpha_{\text{complex}} \cdot S_{\text{complex}} \tag{5}$$

with $\alpha_{\text{rank}} = 0.4$, $\alpha_{\text{dist}} = 0.4$, $\alpha_{\text{complex}} = 0.2$. The reward incorporates quality gates:

$$R(m) = S(m) \cdot \mathbb{1}[\text{OdorantScore}(m) \geq \tau_o] \cdot \mathbb{1}[\text{QED}(m) \geq \tau_q] \tag{6}$$

where $\tau_o = 0.5$ and $\tau_q = 0.3$.

3.5. Reinforcement Learning with Augmented Memory

We optimize the generator using the augmented memory framework (Guo & Schwaller, 2024), which extends the REINVENT algorithm (Olivecrona et al., 2017). The generator is duplicated into a frozen *prior* π_{prior} and a trainable *agent* π_θ. At each step, the agent samples a batch of SMILES strings conditioned on a pocket embedding $\mathbf{z}$, computed as the weighted mean of the pocket embeddings of the ON receptors in the target profile.

The loss for each generated molecule m with SMILES representation x is:

$$\mathcal{L}(x) = (\log \pi_{\text{prior}}(x|\mathbf{z}) + \sigma \cdot R(m) - \log \pi_\theta(x|\mathbf{z}))^2 \tag{7}$$

where σ controls the reward scaling. This loss drives the agent toward high-reward regions while maintaining proximity to the prior, preventing mode collapse.

The full training procedure is summarized in Algorithm 1.

Algorithm 1 Profile-Targeted Molecular Generation

Input: Target profile $\mathbf{t}$, prototypes $\{\mathbf{c}_k\}$, predictor f, budget N
Initialize prior π_{prior}, agent $\pi_\theta \leftarrow \pi_{\text{prior}}$
Compute pocket embedding $\mathbf{z} \leftarrow \text{WeightedMean}(\{f.\text{embed}(r_i)\}_{t_i=1})$
Initialize replay buffer $\mathcal{D} \leftarrow \emptyset$, calls $\leftarrow 0$
Optionally initialize multipliers $\boldsymbol{\lambda}$ for constrained variant
while calls $< N$ **do**
  Sample batch $\{x_1, \ldots, x_B\} \sim \pi_\theta(\cdot|\mathbf{z})$
  Compute S_{rank}, S_{dist}, S_{complex} (Equations (2) to (4))
  $R(m_i) \leftarrow$ combined reward (Equations (5) and (6));
  calls $\leftarrow$ calls $+B$
  if using constrained variant **then**
    Update $\boldsymbol{\lambda}$ via dual ascent (Equation (8))
  end if
  Update $\mathcal{D}$ with top-k molecules
  Compute $\mathcal{L}$ over batch $\cup$ replay samples (Equation (7))
  Update θ via gradient descent on $\mathcal{L}$
  for $j = 1$ **to** A (augmentation rounds) **do**
    Randomize SMILES of batch and replay samples
    Compute and backpropagate augmented loss
  end for
end while
Return: Trained agent π_θ, top molecules from $\mathcal{D}$

Constrained optimization view. The selectivity task can also be viewed as a constrained optimization problem: maximize ON receptor activation subject to OFF receptor suppression below a threshold ϵ. We implement this via Lagrangian relaxation, maintaining dual variables λ_j for each OFF receptor constraint:

$$R_{\text{Lag}}(m) = \sum_{i \in \text{ON}} f_{r_i}(m) - \sum_{j \in \text{OFF}} \lambda_j \cdot \max(0, f_{r_j}(m) - \epsilon) \tag{8}$$

The multipliers are updated via dual ascent after each batch: $\lambda_j \leftarrow \text{clip}(\lambda_j + \eta \cdot (\bar{f}_{r_j} - \epsilon), 0, \lambda_{\max})$, where $\bar{f}_{r_j}$ is the batch-mean OFF score. This automatically increases the penalty on violated constraints during training, providing an adaptive alternative to fixed reward weighting.

4. Experimental Setup

Receptor panel. From the M2OR dataset (Lalis et al., 2024), we select 20 ORs with at least 30 positive interaction records each. Per-receptor binding rates in the panoramic matrix range from 0.9% to 32%, reflecting the natural heterogeneity of receptor promiscuity.

Target profiles. We define nine target profiles spanning a range of complexity, derived from the cluster analysis of the panoramic binding matrix (Table 4). **Profile A** (narrow

selective) uses 2 ON / 3 OFF receptors, targeting the activation pattern of Cluster 5 while suppressing Cluster 4 receptors. **Profile B** (medium selective) uses 3 ON / 2 OFF receptors, targeting Cluster 6 against Cluster 4. **Profile C** (cross-cluster selective) uses 2 ON / 2 OFF receptors, targeting Cluster 0 while suppressing receptors dominant in Cluster 6. **Profile D** (broad-spectrum) uses 14 ON / 2 OFF receptors, targeting the broad activation pattern of Cluster 7. **Profile E** (anti-correlated) uses 2 ON / 2 OFF receptors (s_449, s_51 vs. s_29, s_443), two pairs of highly correlated receptors that are uncorrelated across pairs. **Profile F** (group vs. group) uses 3 ON / 3 OFF receptors, testing group-level selectivity between two receptor clusters. **Profile G** (mixed medium) uses 3 ON / 2 OFF receptors with moderate receptor promiscuity. **Profile H** (sparse activators) uses 3 ON / 2 OFF receptors targeting low-promiscuity receptors (s_1116, s_272, s_464) with low baseline binding rates. **Profile I** (very broad) uses 9 ON / 3 OFF receptors, a high-dimensional activation target second only to Profile D in complexity.

Training details. For each profile, we run RL with a budget of 18,000 oracle calls (binding predictions), batch size 64, learning rate 10^{-4}, reward scaling $\sigma = 60$, replay buffer size 100, and $A = 2$ rounds of SMILES augmentation per step. Each run is repeated with 3 random seeds. All experiments use a single GPU and complete in approximately 15 minutes per run. For each seed, we report the peak batch-mean selectivity score achieved within the budget; Table 1 reports mean $\pm$ standard deviation over three seeds. This metric reflects the best average quality of a generated batch rather than a single best molecule.

Evaluation metrics. Beyond selectivity score, we evaluate generated molecules on: validity (fraction parseable by RDKit), uniqueness (fraction of distinct canonical SMILES), QED (Bickerton et al., 2012), synthetic accessibility (SA) score, molecular weight (MW), and novelty (fraction not present in the M2OR training set). For cross-profile analysis, we compute pairwise Tanimoto similarity using Morgan fingerprints (radius 2, 2048 bits) and extract Murcko scaffolds.

5. Results

5.1. Panoramic Binding Landscape

The panoramic binding matrix reveals a highly structured landscape (Figure 5 in Appendix). Clustering identifies 8 distinct activation patterns, ranging from Cluster 1 (3,548 inert molecules binding no receptors) to Cluster 7 (105 broadly active molecules binding nearly all receptors).

Notably, Clusters 4 and 5 exhibit complementary selectivity with non-overlapping receptor sets, suggesting natural orthogonality in the olfactory code. The binding rates vary substantially across receptors: some receptors (e.g., s_449) are activated by over 30% of sampled molecules, while others (e.g., s_1107) respond to fewer than 1%, reflecting the heterogeneous promiscuity of the OR repertoire.

5.2. Profile-Targeted Generation

Figure 2 shows the training dynamics for all nine profiles across 3 seeds. The selectivity score (mean over each batch) increases overall for all profiles, with convergence rate inversely related to profile complexity.

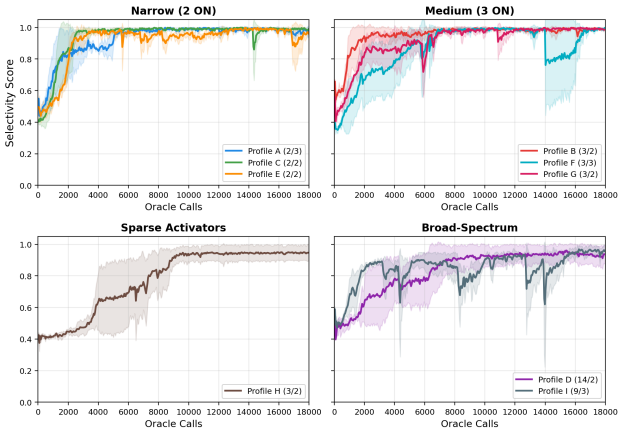

Figure 2. Training curves (structured loss, mean $\pm$ std, 3 seeds) grouped by profile type. **(a)** Narrow profiles (2 ON) converge almost instantly. **(b)** Medium profiles (3 ON) converge within $\sim$100 calls. **(c)** Profile H (sparse activators) requires $\sim$5,000 calls due to low-promiscuity ON receptors. **(d)** Broad-spectrum profiles D (14 ON) and I (9 ON) show the slowest convergence.

Table 1 presents the main results. Several patterns emerge:

Non-learning baselines are strong on easy profiles but weaker on rare targets. For narrow profiles (A, B, E, G), both random prior sampling and panoramic screening achieve near-perfect selectivity, as suitable molecules are already well represented in the prior distribution. Random sampling also performs strongly on Profile I, likely because its ON set contains several promiscuous receptors. The clearest gaps appear for Profile D and especially Profile H: panoramic screening reaches 0.88 and 0.80, respectively, while the structured loss reaches 0.96 and 0.95. These results suggest that directed generation is most valuable when the target activation pattern is rare under the prior, particularly for sparse-activator profiles involving low-promiscuity receptors.

Structured loss improves broad-spectrum optimization. The structured combinatorial code loss achieves the best or tied-best result on eight of nine profiles. On Profile D, it reaches 0.96 ± 0.04, comparable to the Lagrangian variant (0.97 ± 0.05) but with lower variance, and substantially

Table 1. (**Top**) Target profile definitions: ✓ = ON receptor (target = 1), ✗ = OFF receptor (target = 0), – = unconstrained. (**Bottom**) Method comparison: each cell reports the peak batch-mean selectivity (mean ± std over 3 seeds, 18,000 oracle calls). Best result per profile in **bold**. Profile D and Profile H are the most challenging targets.

	29	51	111	125	177	272	353	405	412	418	443	449	464	482	539	949	950	1107	1116	1146
A (2/3)	✓	–	–	–	–	–	–	–	–	✗	✓	–	–	–	–	✗	✗	–	–	–
B (3/2)	–	✓	–	–	–	–	–	–	–	–	–	✓	–	–	✓	✗	✗	–	–	–
C (2/2)	–	✗	–	–	–	–	–	–	–	–	–	✗	✓	✓	–	–	–	–	–	–
D (14/2)	✓	✓	–	–	✓	✓	✓	–	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	–
E (2/2)	✗	✓	–	–	–	–	–	–	–	–	✗	✓	–	–	–	–	–	–	–	–
F (3/3)	✓	✗	✓	–	–	–	–	–	–	–	✓	✗	–	–	✗	–	–	–	–	–
G (3/2)	–	–	–	✓	–	–	–	–	–	–	–	–	–	✓	–	✗	✗	–	–	✓
H (3/2)	–	–	✗	–	–	✓	–	–	–	–	✗	–	✓	–	–	–	–	–	✓	–
I (9/3)	✓	✓	–	✓	–	–	–	✓	–	–	✓	✓	–	✓	✓	✗	✗	✗	–	✓

Method	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Random prior	**1.00** ± 0.00	**1.00** ± 0.00	0.99 ± 0.00	0.92 ± 0.04	**1.00** ± 0.00	0.99 ± 0.01	**1.00** ± 0.00	0.83 ± 0.00	**1.00** ± 0.00
Pan. screen.	**1.00** ± 0.00	**1.00** ± 0.00	0.93 ± 0.01	0.88 ± 0.03	**1.00** ± 0.00	0.97 ± 0.00	**1.00** ± 0.00	0.80 ± 0.01	0.95 ± 0.00
W-MSE	**1.00** ± 0.00	**1.00** ± 0.00	**1.00** ± 0.00	0.87 ± 0.16	**1.00** ± 0.00	**1.00** ± 0.00	**1.00** ± 0.00	0.92 ± 0.07	0.97 ± 0.03
Lagrangian	**1.00** ± 0.00	**1.00** ± 0.00	0.82 ± 0.25	**0.97** ± 0.05	**1.00** ± 0.00	0.89 ± 0.14	0.99 ± 0.00	0.80 ± 0.13	0.92 ± 0.06
Structured	**1.00** ± 0.00	**1.00** ± 0.00	**1.00** ± 0.00	0.96 ± 0.04	**1.00** ± 0.00	**1.00** ± 0.00	**1.00** ± 0.00	**0.95** ± 0.05	0.99 ± 0.01

outperforming W-MSE (0.87 ± 0.16). On Profile H, the hardest profile, it reaches 0.95±0.05, outperforming both W-MSE (0.92±0.07) and the Lagrangian variant (0.80±0.13). On narrow profiles (A, B, E, G), all RL methods achieve 1.00.

Lagrangian constraints help simple selective profiles. The constrained formulation achieves competitive results on simple profiles (A, B, E) but shows high variance on Profiles C (0.825±0.247), F (0.894±0.145), and I (0.918±0.063), suggesting that adaptive multipliers become unstable when many constraints must be balanced.

Designability gradient. Optimization difficulty depends jointly on target dimensionality and ON-receptor promiscuity. Most narrow or medium profiles converge within ∼100 oracle calls. Profile I (9 ON) also converges quickly because its ON set contains promiscuous receptors. In contrast, Profile D requires ∼1,231 calls and Profile H requires ∼5,367 calls due to low baseline binding rates of its ON receptors. Broad profiles with promiscuous receptors can thus be easier than sparse profiles targeting low-promiscuity receptors.

5.3. Cross-Profile Chemical Analysis

A critical question is whether molecules optimized for different activation patterns are chemically distinct. We collect the top-100 molecules from each profile and compute Morgan fingerprints (radius 2, 2048 bits). The mean inter-profile Tanimoto similarity (0.120 ± 0.105) is slightly lower than intra-profile similarity (0.128±0.100), and scaffold overlap between any two profiles is below 35%, suggesting partial profile-specific chemical separation.

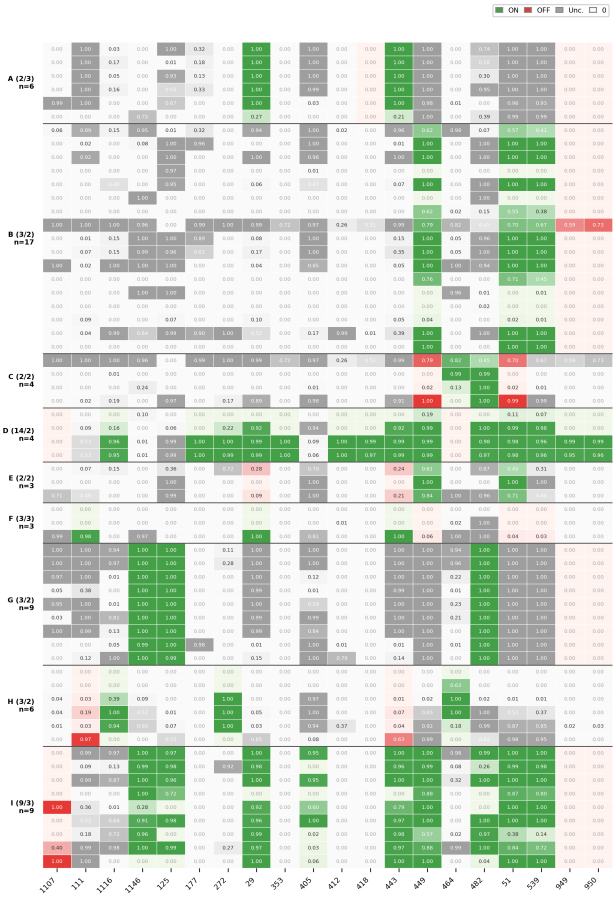

Figure 3. Per-receptor binding of generated molecules (structured loss, sel > 0.8, 3 seeds). Each row is one molecule, grouped by profile. Green = ON, red = OFF, gray = unconstrained. Intensity = binding probability.

Table 2. Quality metrics of top generated molecules per profile (structured loss, aggregated across 3 seeds).

Prof.	Novel↑	QED↑	SA↓	MW	cLogP
A	100%	0.56 ± 0.08	1.87 ± 0.80	137 ± 31	3.18 ± 0.56
B	65%	0.58 ± 0.12	2.13 ± 0.69	137 ± 50	1.57 ± 1.08
C	50%	0.55 ± 0.14	1.77 ± 0.59	128 ± 46	1.70 ± 0.56
D	100%	0.52 ± 0.10	1.95 ± 0.73	146 ± 41	2.48 ± 0.37
E	67%	0.54 ± 0.06	1.84 ± 0.60	153 ± 21	1.73 ± 1.34
F	100%	0.50 ± 0.04	1.84 ± 0.41	110 ± 24	2.04 ± 1.00
G	100%	0.57 ± 0.05	1.99 ± 0.83	141 ± 12	1.60 ± 0.44
H	67%	0.42 ± 0.12	2.31 ± 0.09	115 ± 43	1.16 ± 1.47
I	89%	0.48 ± 0.07	2.04 ± 0.41	156 ± 24	2.95 ± 0.99

QED scores remain in the chemically plausible range (0.4–0.6), SA scores (<2.6) indicate synthetic accessibility, and molecular weights (110–164 Da) are typical of volatile odorants. cLogP values (1.5–3.4) are consistent with membrane-permeable compounds. Novelty ranges from 50% (Profile C) to 100% (Profiles A, D, F, G). Some profiles show repeated high-scoring molecules across seeds, suggesting possible mode collapse, a limitation we discuss further below.

5.4. Ablation: Reward Formulation

To understand how reward formulation interacts with profile complexity, we group the nine profiles into three difficulty tiers (Easy: A, B, E, G with 2–3 ON receptors and high baseline binding; Medium: C, F, I with cross-cluster or high-dimensional targets; Hard: D, H with broad-spectrum or sparse-activator targets) and compare the three RL reward formulations (Table 3).

Table 3. Reward formulation comparison across profile difficulty groups (3 seeds per profile). *Sel.*: mean best-of-budget selectivity (peak single-molecule score per seed). *%≥0.99*: fraction of seeds reaching selectivity ≥ 0.99. $\bar{\sigma}$: mean per-profile std computed from unrounded per-seed values (lower = more stable). *#Solved*: profiles with mean selectivity ≥ 0.99.

	Easy (4)		Hard (2)			
Reward	Sel.↑	%≥0.99↑	Sel.↑	%≥0.99↑	$\bar{\sigma}$↓	#Solved↑
W-MSE	1.00	100%	0.95	50%	0.013	7/9
Lagrangian	1.00	100%	0.91	33%	0.021	6/9
Structured	**1.00**	**100%**	**0.99**	**67%**	**0.002**	**8/9**

On easy profiles, all three formulations achieve perfect selectivity with 100% of seeds reaching ≥ 0.99, since the optimization landscape is simple enough that any reasonable reward signal suffices. The differences emerge on hard profiles, where the structured loss maintains 0.99 selectivity with 67% of seeds reaching ≥ 0.99, compared to 50% for W-MSE (0.95 mean) and only 33% for the Lagrangian variant (0.91 mean). Overall, the structured loss solves 8 of 9 profiles (mean selectivity ≥ 0.99), versus 7 for W-MSE and 6 for Lagrangian. It is also the most stable across seeds ($\bar{\sigma} = 0.002$), 6–10× lower than alternatives.

We attribute this robustness to two properties of the structured formulation. First, the pairwise ranking component (Equation (2)) provides gradient signal based on *relative* ON–OFF ordering rather than absolute binding values, making it robust to calibration error in the binding predictor. Second, the distribution matching component (Equation (3)) captures the full response geometry as a distribution, rather than treating each receptor as an independent regression target as W-MSE does. Together, these provide a smoother optimization landscape that scales to high-dimensional profiles.

Figure 4 compares the training dynamics across all nine profiles. The structured loss consistently reaches higher selectivity than W-MSE, particularly on Profile D (0.96 vs. 0.87). The Lagrangian constrained variant shows instability on several profiles (C, F, I), where balancing multiple ON/OFF constraints with adaptive multipliers proves challenging.

6. Discussion

Structured loss vs. heuristic reward. A key design insight is that the structured combinatorial code loss (Equation (5)) provides a richer and more principled optimization signal than simple weighted MSE. The ranking component (Equation (2)) directly encodes the selectivity requirement, that ON receptors should outrank OFF receptors, without requiring accurate absolute calibration of the binding predictor. The distribution matching component (Equation (3)) captures the full response geometry, treating the receptor profile as a distribution rather than independent binary targets. The complexity term (Equation (4)) ensures that the generator matches not just which receptors are active, but the shape of the activation pattern. Together, these components provide stable gradient signal from the start of training: even a molecule that partially satisfies the ranking constraints receives meaningful reward.

When is RL-based generation necessary? Our baseline comparison (Table 1) reveals that for narrow profiles (A, B, E, G), suitable molecules already exist in the prior distribution, and panoramic screening achieves near-perfect selectivity. However, for Profile D and especially Profile H, the gap between screening and RL-based generation is substantial. On Profile D, the structured loss achieves 0.96 vs. 0.88 for panoramic screening. On Profile H, the gap is 0.95 vs. 0.80. This suggests that RL-based generation is most valuable when the target activation pattern is rare under the prior.

Profile complexity and designability. Our results across nine profiles reveal that optimization difficulty depends jointly on target dimensionality and receptor promiscuity. Narrow profiles (2–3 ON receptors) converge to high selec-

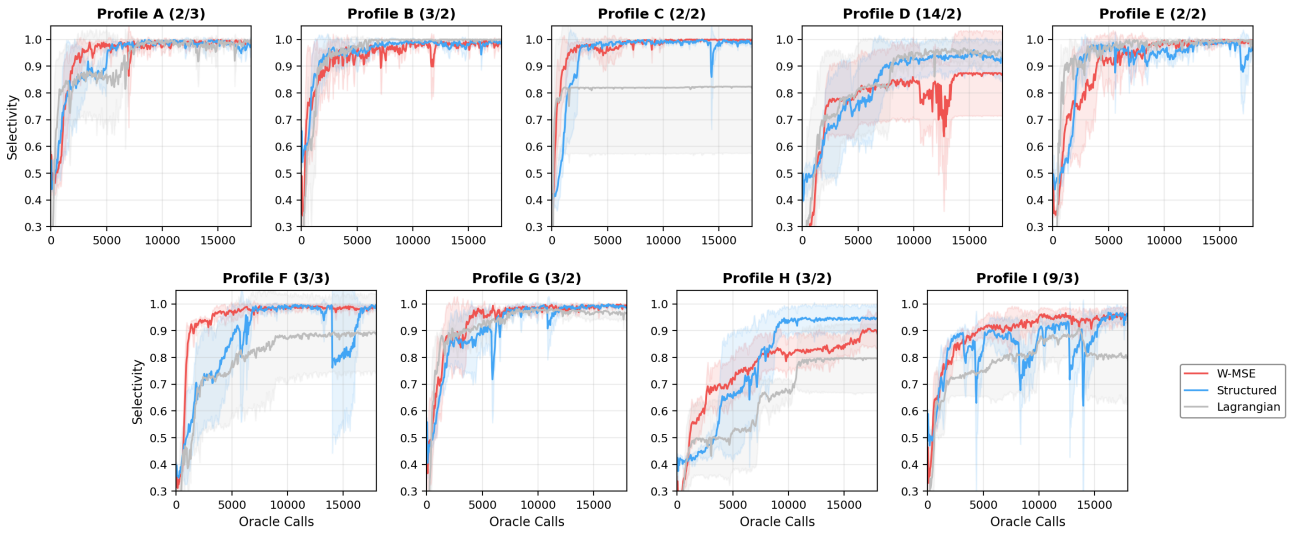

Figure 4. Training curves comparing three reward formulations across all nine profiles (mean ± std shaded, 3 seeds each). The structured loss matches or outperforms W-MSE on all profiles, with the largest improvement on Profiles D and H. The Lagrangian variant shows high variance on Profiles C, F, and I.

tivity within ∼100 calls, while Profile D (14 ON) requires ∼1,231 and Profile H (sparse activators) requires ∼5,367. Notably, Profile H is harder than Profile D despite having fewer receptors, because its ON receptors (s_1116, s_272, s_464) have low baseline binding rates (<5%). This "designability gradient" has practical implications: designability is not determined solely by the number of ON receptors, and broad profiles containing promiscuous receptors can be relatively easy, whereas sparse profiles targeting low-promiscuity receptors can be substantially harder.

Chemical space structure. The cross-profile chemical analysis (Section 5.3) suggests partial, but not complete, profile-specific structure in chemical space. The mean inter-profile Tanimoto similarity is only slightly lower than the mean intra-profile similarity, indicating that fingerprint-level separation is modest. However, scaffold overlap between profile pairs remains below 35%, suggesting that different target profiles can enrich for partially distinct scaffold families. We therefore interpret these results as evidence of profile-dependent chemical enrichment rather than strong separation into disjoint chemical regions.

Limitations. Our framework relies entirely on computational binding predictions, which may have systematic biases for receptors with limited training data or sparse negative labels (mean AUROC varies from 0.58 to 0.94). Generated molecules achieving high predicted selectivity require experimental validation, as some may exploit predictor artifacts rather than genuine binding properties. Our best-of-budget metric evaluates discovery of at least one high-scoring molecule rather than average sample quality or diversity. The RL setup can collapse toward a small number of high-scoring scaffolds for some profiles, and our current approach trains a separate generator per profile rather than a single conditional model. Finally, the panoramic profiling stage is limited to 20 receptors, and scaling to the full human OR repertoire (∼400 receptors) would require more efficient prediction methods.

Future work. Natural extensions include a profile-conditioned generator that takes the target activation pattern as input, enabling single-model generation across all profiles without retraining. Uncertainty-aware optimization using ensemble predictors could address surrogate bias by down-weighting molecules outside the predictor's reliable domain. Experimental validation of top-scoring molecules via calcium imaging assays would confirm whether predicted selectivity translates to genuine receptor activation. Finally, incorporating perceptual models that map receptor activation patterns to predicted odor qualities would close the loop from molecular design to sensory experience.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 52505043), the National Key Research and Development Program of China (Grant No. 2024YFB3212300), and the Fundamental Research Funds for the Central Universities (Grant No. 226-2025-00037). The authors thank Liujing Zhuang from the College of Biomedical Engineering and Instrument Science, Zhejiang University, and Hao Jia from the Shanghai Institute of Microsystem and Information Technology, Chinese

Academy of Sciences, for their valuable discussions in the early stages of this work.

Impact Statement

This paper presents a computational framework for designing molecules with targeted olfactory receptor activation patterns. Potential applications include fragrance design and flavor chemistry. As with all molecular generation methods, generated candidates require safety and toxicity screening before any real-world use. We do not foresee specific negative societal consequences beyond those common to the field of de novo molecular design.

References

Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., and Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nature Chemistry*, 4(2):90–98, 2012.

Blaschke, T., Arús-Pous, J., Chen, H., Marber, C., Tyrchan, C., and Engkvist, O. REINVENT 2.0: An AI tool for de novo drug design. *Journal of Chemical Information and Modeling*, 60(12):5918–5922, 2020.

Buck, L. B. Olfactory receptors and odor coding in mammals. *Nutrition Reviews*, 62:S184–S188, 2004.

Charlier, L., Topin, J., de March, C. A., Lai, P. C., Crasto, C. J., and Golebiowski, J. Molecular Modelling of Odorant/Olfactory Receptor Complexes. In *Olfactory Receptors: Methods and Protocols*, Methods in Molecular Biology, volume 1003, pp. 53–65. Humana Press, 2013.

Du, Y., Jamasb, A. R., Guo, J., Fu, T., Harris, C., Wang, Y., Duan, C., Liò, P., Schwaller, P., and Blundell, T. L. Machine learning-aided generative molecular design. *Nature Machine Intelligence*, 6:589–604, 2024.

Elnaggar, A., Heinzinger, M., Dallago, C., et al. ProtTrans: Toward understanding the language of life through self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7112–7127, 2021.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pp. 1263–1272, 2017.

Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 4(2):268–276, 2018.

Guo, J. and Schwaller, P. Augmented memory: Capitalizing on experience replay to accelerate de novo molecular design. *Machine Learning: Science and Technology*, 5 (1):015036, 2024.

Guo, J. and Schwaller, P. TANGO: Direct optimization of constrained synthesizability for generative molecular design. *Nature Computational Science*, 6:260–270, 2026.

Guo, J., Chen, J., Chen, A. G., and Schwaller, P. Sample-efficient generative molecular design using memory manipulation. *Nature Machine Intelligence*, 8:449–460, 2026.

Haddad, R., Litsa, E. E., Liu, Z., Yu, X., Burkhardt, D., and Bhisetti, G. Targeted molecular generation with latent reinforcement learning. *Scientific Reports*, 15:15202, 2025.

Hladis, M., Lalis, M., Topin, J., and Fiorucci, S. Matching receptor to odorant with protein language and graph neural networks. *International Conference on Learning Representations*, 2023.

Jin, W., Barzilay, R., and Jaakkola, T. Multi-objective molecule generation using interpretable substructures. In *International Conference on Machine Learning*, pp. 4849–4859, 2020.

Kurian, S. M., Naressi, R. G., Manoel, D., Barwich, A.-S., Malnic, B., and Logan, D. W. Odor coding in the mammalian olfactory epithelium. *Cell and Tissue Research*, 383(1):445–456, 2021.

Lalis, M., Hladis, M., Khalil, S., Fiorucci, S., and Topin, J. M2OR: a database of olfactory receptor–odorant pairs for understanding the molecular mechanisms of olfaction. *Nucleic Acids Research*, 52(D1):D1370–D1379, 2024.

Lee, B. K., Mayhew, E. J., Sanchez-Lengeling, B., et al. A principal odor map unifies diverse tasks in olfactory perception. *Science*, 381(6661):999–1006, 2023.

Liu, T.-Y. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3):225–331, 2009.

Malnic, B., Hirono, J., Sato, T., and Buck, L. B. Combinatorial receptor codes for odors. *Cell*, 96(5):713–723, 1999.

McConachie, G. D., Duniec, E., Guerina, F., Younger, M. A., and DePasquale, B. Low rank adaptation of chemical foundation models generates effective odorant representations. *bioRxiv preprint*, 2025.

Olivecrona, M., Blaschke, T., Engkvist, O., and Chen, H. Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9(1):48, 2017.

Renz, P., Van Rompaey, D., Wegner, J. K., Hochreiter, S., and Klambauer, G. On failure modes in molecule generation and optimization. *Drug Discovery Today: Technologies*, 32:55–63, 2020.

Sanchez-Lengeling, B., Wei, J. N., Lee, B. K., Gerkin, R. C., Aspuru-Guzik, A., and Wiltschko, A. B. Machine learning for scent: Learning generalizable perceptual representations of small molecules. *arXiv preprint arXiv:1910.10685*, 2019.

Segler, M. H., Kogej, T., Tyrchan, C., and Waller, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Science*, 4 (1):120–131, 2018.

Shuvaev, S., Tran, K., Samoilova, K., Mascart, C., and Koulakov, A. DeepNose: An equivariant convolutional neural network predictive of human olfactory percepts. In *Asilomar Conference on Signals, Systems, and Computers*, pp. 536–540. IEEE, 2024.

Thomas, M., Smith, R. T., O'Boyle, N. M., de Graaf, C., and Bender, A. Augmented hill-climb increases reinforcement learning efficiency for language-based de novo molecule generation. *Journal of Cheminformatics*, 14:68, 2022.

Xu, P., Feng, T., Fu, T., and Sun, J. REINVENT-Transformer: Molecular de novo design through transformer-based reinforcement learning. *arXiv preprint arXiv:2310.05365*, 2024.

You, J., Liu, B., Ying, R., Pande, V., and Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

A. Target Profile Definitions

The nine target profiles are derived from the cluster analysis of the panoramic binding matrix. Each profile specifies a set of ON receptors (target $t_i = 1$) and OFF receptors (target $t_i = 0$), with the remaining receptors assigned weight $w_i = 0$ ("don't care"). The profiles are designed to span a range of complexities.

Profile A targets the activation pattern of Cluster 5 (s_29/s_443 selective) while suppressing Cluster 4 receptors (s_949/s_950/s_418). These two clusters have non-overlapping receptor sets, making this an "orthogonal" selectivity task. **Profile B** targets Cluster 6 (s_449/s_51/s_539) against the same Cluster 4 OFF receptors. With 3 ON receptors, this tests whether the framework scales to slightly more complex patterns. **Profile C** targets Cluster 0 (s_464/s_482) while suppressing Cluster 6 receptors (s_449/s_51), a "cross-cluster" task where the ON and OFF receptor sets come from clusters with partial overlap in the panoramic matrix. **Profile D** targets the broad-spectrum pattern of Cluster 7, requiring activation of 14 receptors simultaneously while suppressing 2, testing the limits of the framework on high-dimensional profiles. **Profile E** targets an anti-correlated pair: ON receptors s_449/s_51 vs. OFF receptors s_29/s_443. Within each pair, receptors are highly correlated ($r > 0.99$), but the two pairs are uncorrelated, testing fine-grained selectivity. **Profile F** tests group-level selectivity: 3 ON (s_111, s_29, s_443) vs. 3 OFF (s_449, s_51, s_539), requiring simultaneous activation and suppression of equal-sized receptor groups. **Profile G** targets 3 ON (s_1146, s_125, s_482) vs. 2 OFF (s_949, s_950) with moderate receptor promiscuity, serving as a control for medium-complexity profiles. **Profile H** targets sparse activators: 3 ON receptors (s_1116, s_272, s_464) with low baseline binding rates (<5%), making this the hardest optimization target despite having few receptors. **Profile I** targets a very broad pattern: 9 ON / 3 OFF receptors, second only to Profile D in dimensionality.

ON receptors are weighted at $w_i = 2$ and OFF receptors at $w_i = 1$, reflecting the asymmetry that activating targets is typically harder than suppressing off-targets. Profile D uses uniform weights $w_i = 1$ for all active receptors to avoid over-penalizing any single receptor in the large ON set.

Table 4. Detailed target profile specifications.

Profile	ON Receptors	OFF Receptors	Source
A	s_29, s_443	s_418, s_949, s_950	C5 vs C4
B	s_449, s_51, s_539	s_949, s_950	C6 vs C4
C	s_464, s_482	s_449, s_51	C0 selective
D	s_1116, s_177, s_272, s_29, s_353, s_412, s_418, s_443, s_449, s_482, s_51, s_539, s_949, s_950	s_1107, s_464	C7 broad
E	s_449, s_51	s_29, s_443	Anti-correlated
F	s_111, s_29, s_443	s_449, s_51, s_539	Group vs group
G	s_1146, s_125, s_482	s_949, s_950	Mixed medium
H	s_1116, s_272, s_464	s_111, s_443	Sparse activators
I	s_1146, s_125, s_29, s_405, s_443, s_449, s_482, s_51, s_539	s_1107, s_949, s_950	Very broad

B. Ablation Details

The ablation in Table 3 groups nine profiles into three difficulty tiers based on the number of constrained receptors and their baseline binding rates. **Easy** profiles (A, B, E, G) involve 2–3 ON receptors with moderate-to-high baseline binding rates; suitable molecules already exist in the prior distribution and all RL methods converge within ∼100 oracle calls. **Medium** profiles (C, F, I) involve cross-cluster selectivity or high-dimensional targets (up to 9 ON receptors); these require the generator to explore beyond the prior but remain tractable for all methods. **Hard** profiles (D, H) represent the most challenging targets: Profile D requires simultaneous activation of 14 receptors, while Profile H targets 3 low-promiscuity receptors with baseline binding rates <5%. These profiles expose the largest differences between reward formulations.

For each method–profile combination, we report the maximum selectivity score achieved across the 18,000-call budget, averaged over 3 seeds. The success rate (%≥0.99) measures the fraction of seeds that reach near-perfect selectivity. The stability metric $\bar{\sigma}$ is the mean per-profile standard deviation across seeds, measuring how consistently a method performs across random initializations.

C. Panoramic Binding Matrix

Figure 5 shows the full panoramic binding matrix, with molecules sorted by cluster assignment. Each panel corresponds to one of the 8 discovered clusters. **Cluster 1** (n=3,548) is the dominant cluster, comprising 61% of all sampled molecules with near-zero binding probability across all 20 receptors, representing the "inert" background of chemical space. **Clusters 2, 3, and 6** share a common set of active receptors (s_449, s_51, s_539) but differ in the breadth and intensity of activation; Cluster 3 additionally activates s_29 and s_443, while Cluster 2 activates s_1146. **Cluster 4** (n=143) has a distinctive pattern centered on s_949, s_950, s_177, and s_418, receptors that are largely inactive in other clusters, making Cluster 4 an ideal OFF-target set for Profiles A and B. **Cluster 5** (n=152) is selective for s_29 and s_443 with moderate activation of s_111 and s_1116, complementary to Cluster 4 and motivating Profile A. **Cluster 7** (n=105) is the broadly active cluster, with high binding probability across nearly all receptors, motivating Profile D. **Cluster 0** (n=422) shows moderate activation of s_464 and s_482, receptors that are relatively quiet in other clusters, motivating Profile C.

The per-receptor binding rates vary from 0.9% (s_1107) to 32% (s_449), reflecting the heterogeneous promiscuity of the OR panel. This variation is biologically meaningful: some receptors are broadly tuned "generalists" while others are narrowly tuned "specialists."

D. Top Generated Molecules

Figure 6 shows the 2D structures of the top generated molecules for representative profiles, filtered for odorant-likeness (MW ≤ 200, ≥ 5 heavy atoms, no charged species or nitro groups, at least 3 carbon atoms).

Profile A molecules are predominantly linear or branched aliphatic chains with aldehyde, ketone, or ester functional groups (MW 114–158). This is consistent with known s_29/s_443 ligands, which tend to be medium-chain aldehydes and ketones.

Profile B molecules show more structural diversity, including a thioaldehyde (O=CCCCS), isobutyraldehyde, menthol-like cyclohexanol, and hexanal. The presence of sulfur-containing compounds is notable and may reflect the binding preferences of s_449/s_51/s_539.

Profile C has fewer unique molecules (4 after filtering), reflecting the narrower chemical space that satisfies the s_464/s_482 selectivity constraint. The molecules include a methylenedioxyphenyl ketone, small ethers, and diols.

Profile D molecules include aromatic aldehydes (naphthaldehyde), terpene esters (geranyl acetate, linalyl acetate), and a thiophenol derivative. The structural diversity is higher than for narrow profiles, consistent with the broad-spectrum activation requirement spanning many receptor types.

Profile H molecules tend toward smaller, polar structures consistent with the sparse-activator receptor set. The low baseline binding rates of ON receptors (s_272, s_464, s_1116) constrain the accessible chemical space.

All generated molecules have MW in the 72–196 Da range typical of volatile odorants, SA scores of 1.7–2.4 (synthetically accessible), and QED scores of 0.4–0.6, used here as a generic chemical plausibility filter rather than an odorant-specific objective. Novelty ranges from 50% to 100% across profiles.

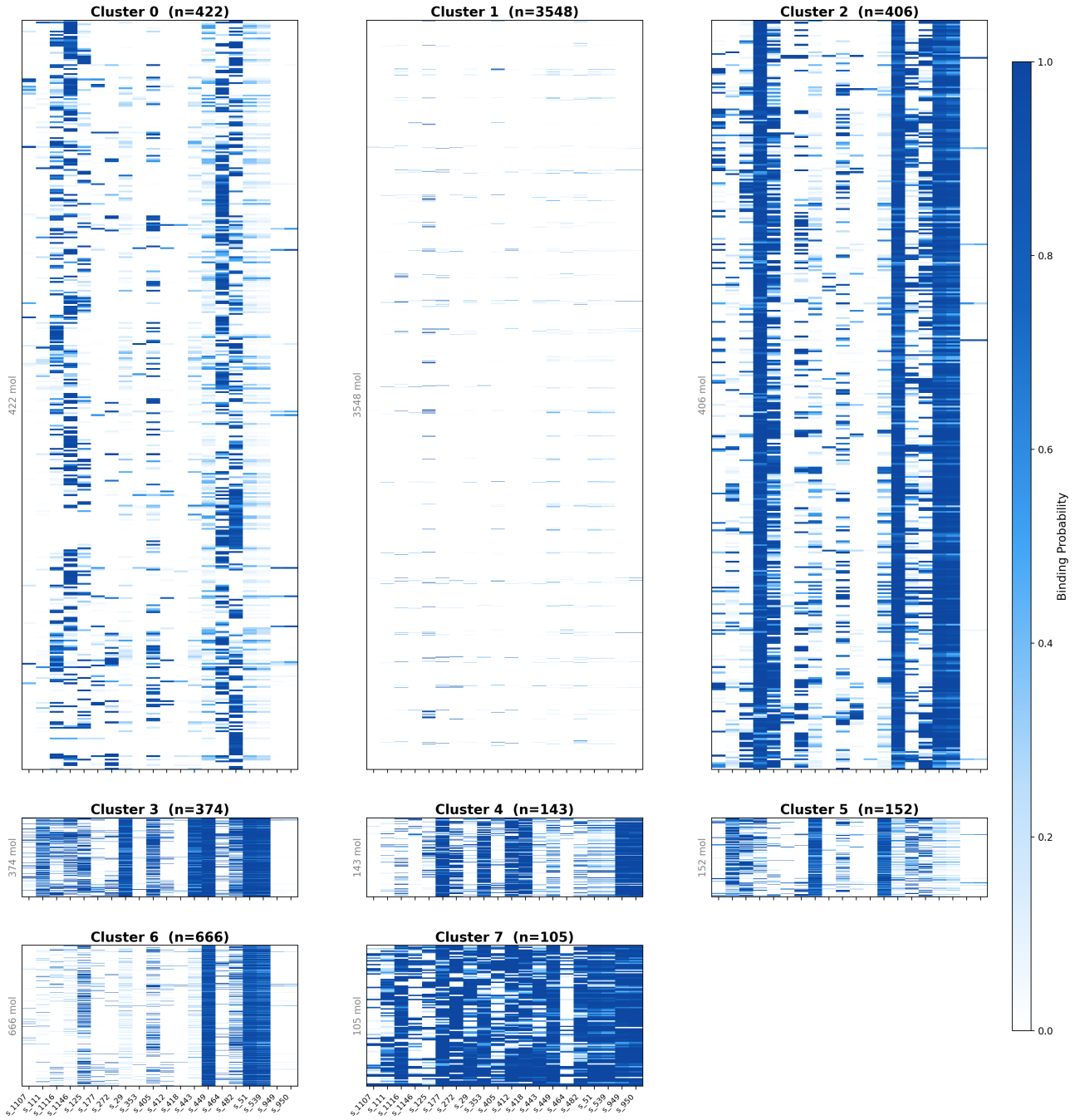

Figure 5. Panoramic binding matrix (5,816 molecules × 20 receptors), split by cluster. Color intensity indicates predicted binding probability. Eight clusters emerge with distinct activation patterns, from inert (C1) to broadly active (C7).

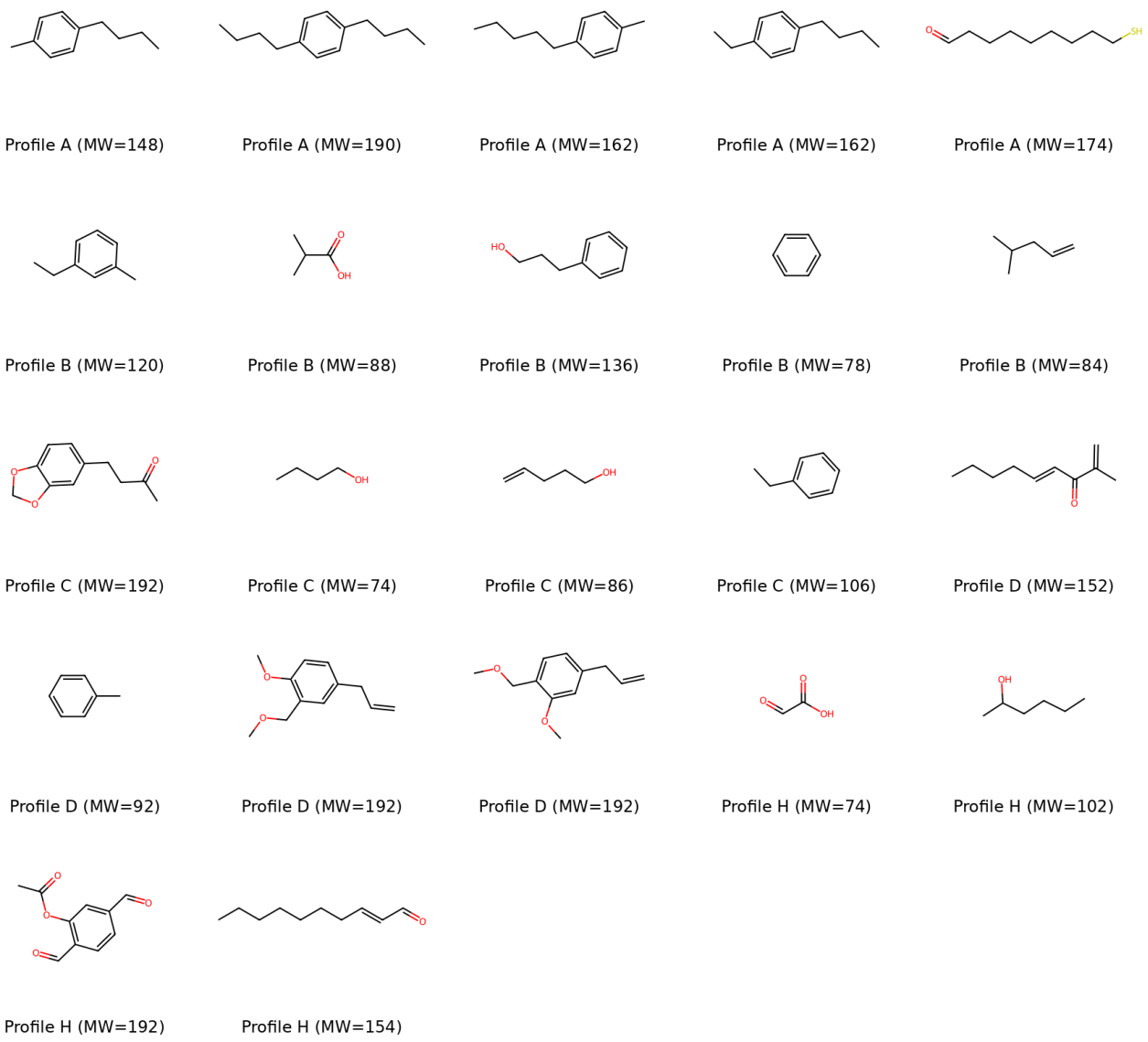

Figure 6. 2D structures of top unique generated molecules for representative profiles (rows: A, B, C, D, H), filtered for odorant-likeness (MW $\leq$ 200). Molecules are selected by highest selectivity reward across 3 seeds.