

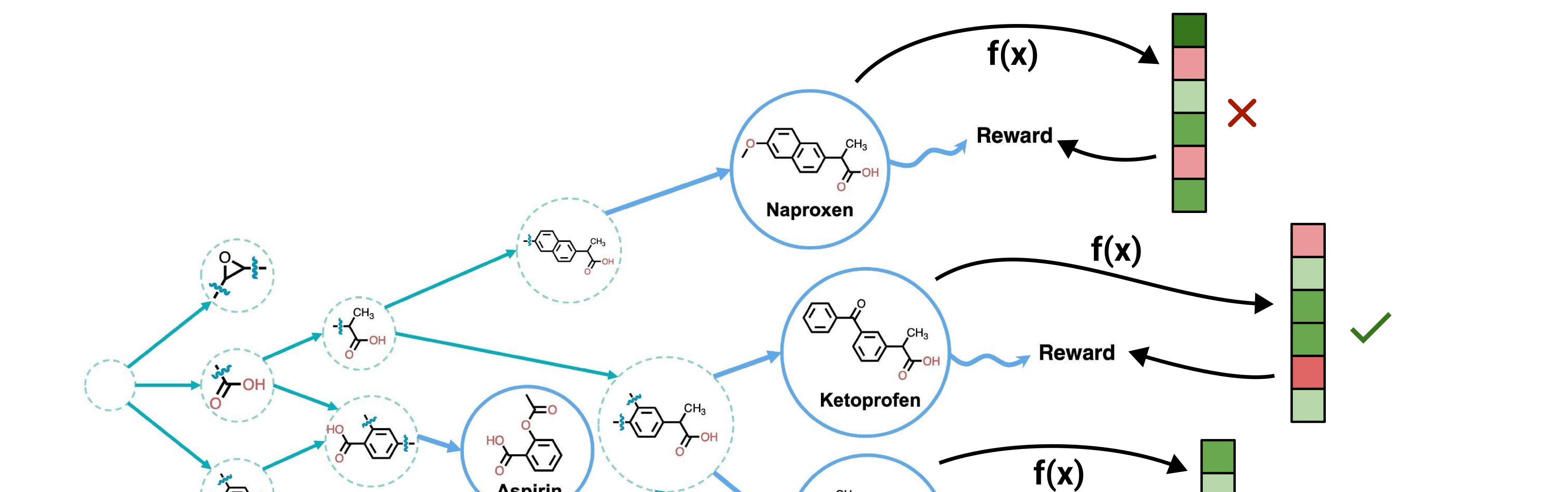




Cell Morphology-Guided Small Molecule Generation with GFlowNets

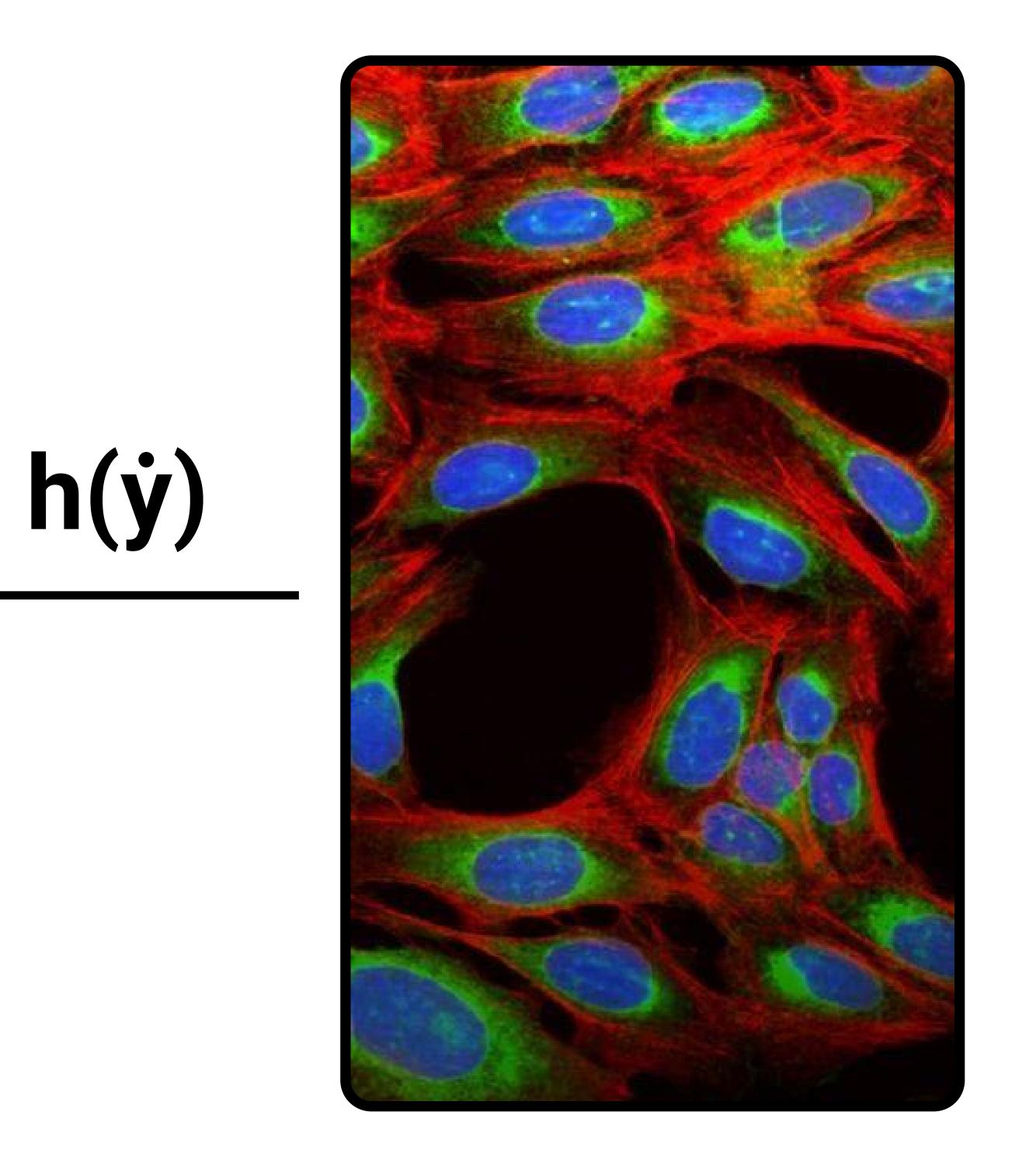
Stephen Lu · Ziging Lu · Ehsan Hajiramezanali · Tommaso Biancalani · Yoshua Bengio · Gabriele Scalia · Michał Koziarski

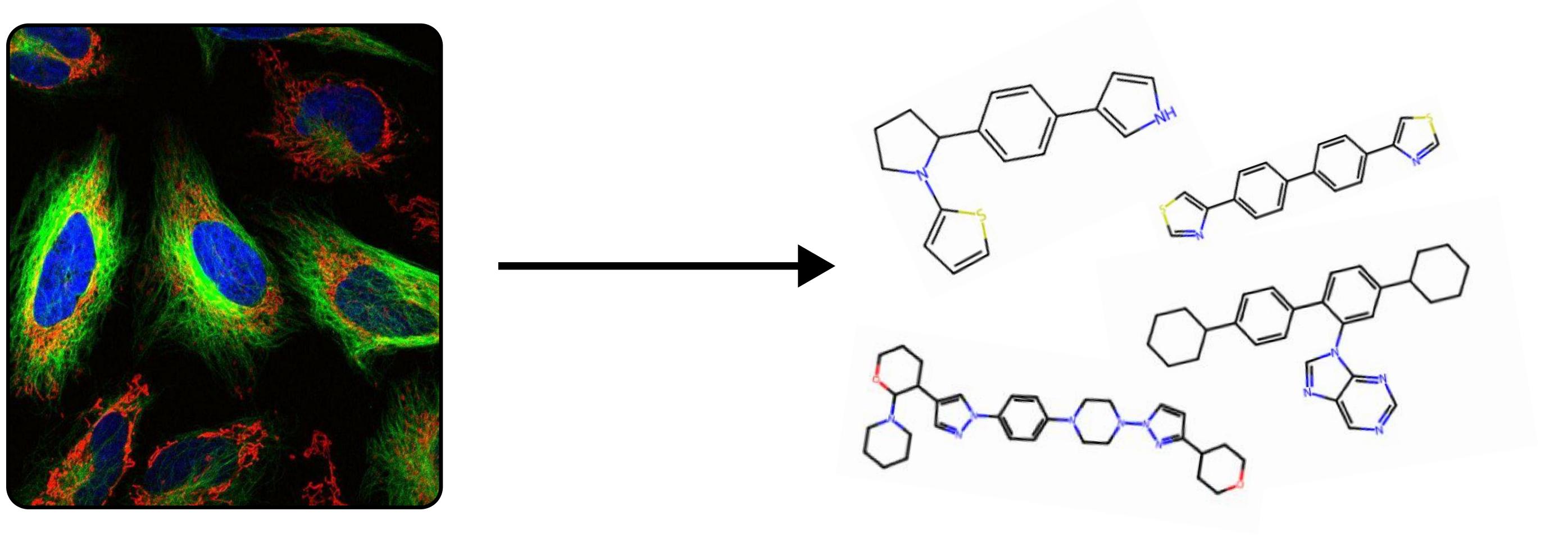
TL;DR Given morphological readout, we use contrastive learning and generative methods to propose diverse candidate molecules with similar readouts



Ibuprofen

Reward





Motivation

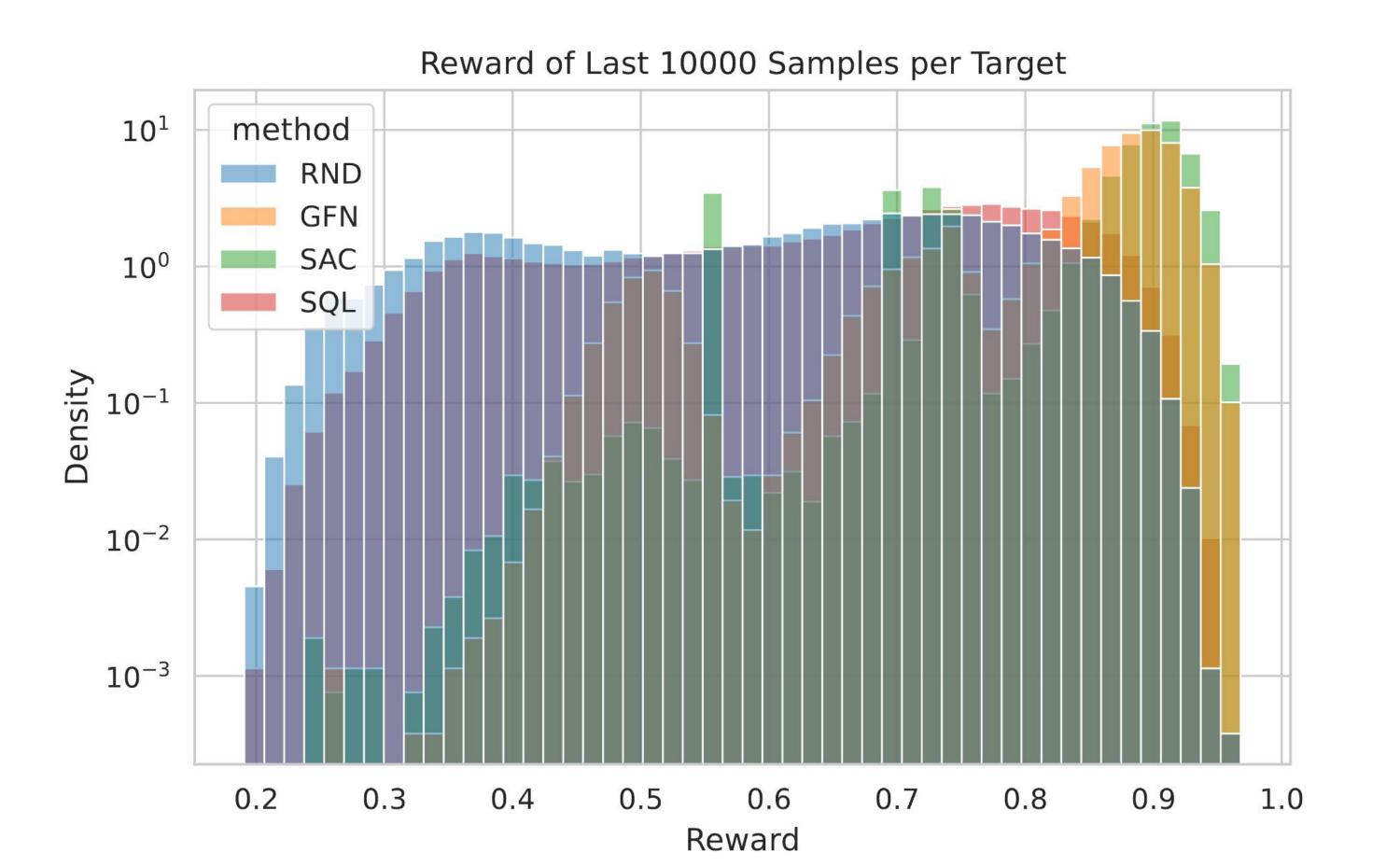
- Generative models that explicitly optimise phenotypic screening signals are limited by the lack of labelled data
- Our method generates novel and diverse molecules with similar phenotypic effect as a target image without pre-annotated labels
- We use unsupervised multimodal embeddings to define a latent similarity based reward for GFlowNets

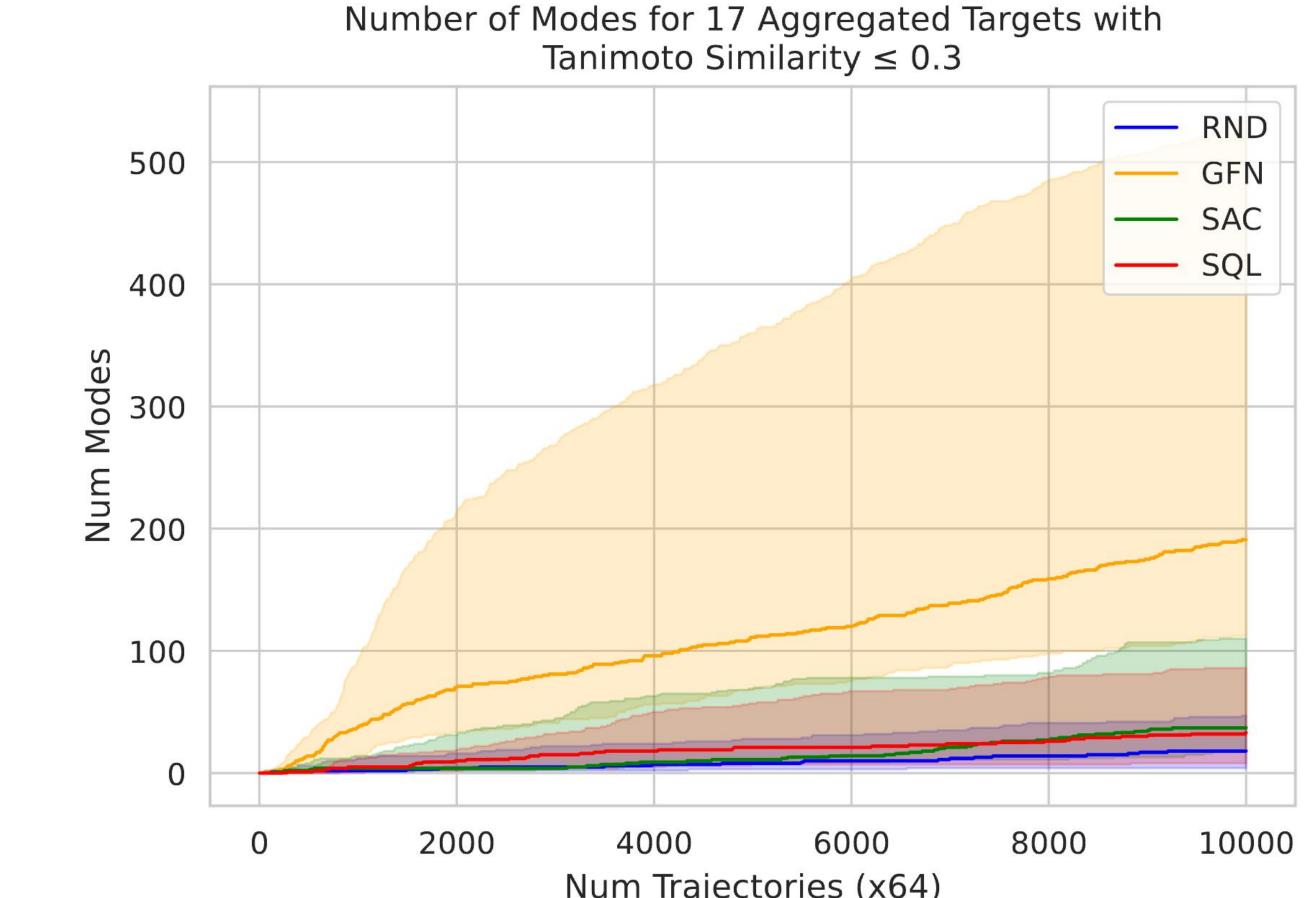
Morphology Guided **GFlowNets Design**

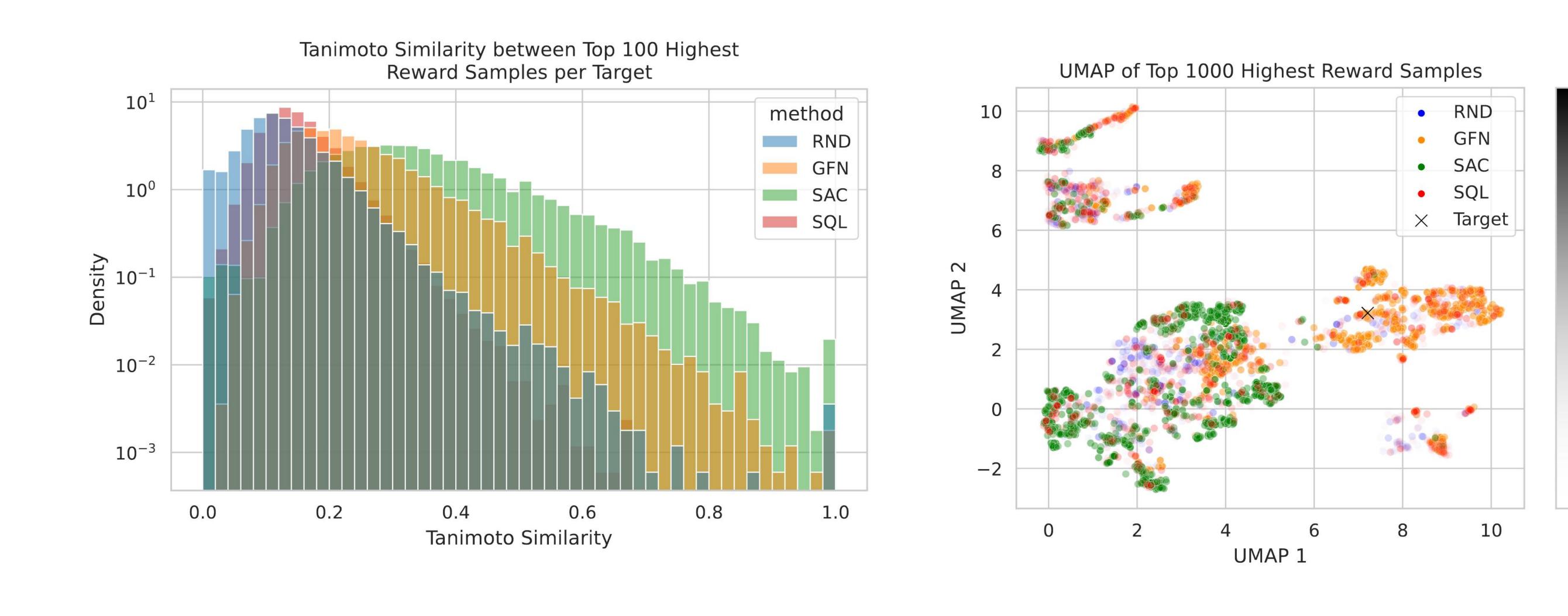
Reward

Aspirin

- 1. Train GMC to align small molecular structures with their associated cell morphological readouts
- 2. Leverage the morphology (or joint) encoder of GMC to produce a latent representation for a desired target morphological readout h(ý)
- 3. GFN learns a forward policy, parameterised by a Graph Attention Transformer, over a set of predefined fragments to generate diverse and







• GFlowNets is better able to recover the underlying target, suggesting usefulness in finding molecules with target biological activity

GFIOWNEts

- GFlowNets [1] (GFN) are amortised variational inference algorithms trained to sample from an unnormalised distribution over compositional objects
- Aim to sample objects from a set of terminal states X proportionally to a reward function $R: X \rightarrow R+$

 s_1 s_2 s_1 s_2 s_1 s_1

high reward molecules

$$R(x|\dot{y}) = 1 + \frac{f(x)h(\dot{y})}{2||f(x)|| \cdot ||h(\dot{y})||}$$

Cross-modal Latent Space Alignment

• Our method assumes that similarity in the latent space learnt by GMC is of high anough quality to partarm coarch

Table 1: Max. Tanimoto similarity to the target in last 10,000 samples, averaged across all targets.

	Random	Soft Actor-Critic	Soft Q-Learning	GFlowNet
Morphology Target	0.305 (± 0.057)	0.261 (± 0.068)	0.329 (± 0.079)	0.337 (± 0.092)
Joint Target	0.311 (± 0.064)	0.388 (± 0.163)	0.309 (± 0.064)	0.451 (± 0.163)

