CELLBRIDGE: BIOLOGICALLY FAITHFUL PERTURBATION-TO-MORPHOLOGY PREDICTION VIA SOURCE-CONDITIONED SCHRÖDINGER BRIDGES

Anonymous authors
Paper under double-blind review

ABSTRACT

Cellular Morphology Perturbation Prediction (CMPP) is a fundamental capability toward building predictive virtual cells, aiming to simulate how cellular phenotypes change in response to chemical or genetic interventions. Such predictive models have the potential to accelerate therapeutic discovery by reducing experimental perturbation screening and enabling systematic exploration of cellular response mechanisms. However, existing generative approaches insufficiently account for the source-dependent nature of cellular state transitions and the biological mechanisms which connect perturbations to phenotypic responses. To fill these gaps, we propose **CellBridge**, a source-conditioned Schrödinger Bridge framework for modeling cellular morphological responses to chemical and genetic perturbations. CellBridge consists of two key components: i) modeling perturbation response as a bridge from an observed control cell state to its perturbed phenotype, with the control image supplied to the network at every step, and ii) incorporating structured **compound–target–pathway conditioning** to connect molecular perturbations with their downstream biological programs. For the challenging generalization setting, where structurally distinct compounds can induce similar phenotypes through shared biological mechanisms, the pathway representation enables CellBridge to move beyond scaffold similarity and capture mechanism-related responses. Comprehensive experiments on BBBC021, RxRx1, and JUMP demonstrate that CellBridge consistently improves perturbation-specific phenotype recovery across various settings including chemical compound and genetic perturbations.

1 INTRODUCTION

Predicting how a cell will respond to an intervention before that intervention is ever performed is a defining capability of a predictive virtual cell. Cellular morphology is one of the few phenotypes where such a prediction can be read out directly and cheaply at scale. High-content fluorescence assays such as Cell Painting (Bray et al., 2016) stain several organelle compartments at once. A single multi-channel image therefore records nuclear, cytoskeletal, mitochondrial and membrane structure together, and coordinated changes across these compartments indicate which biological program a perturbation has engaged. Public screens now cover hundreds of compounds and more than a thousand gene knockdowns across millions of images (Caie et al., 2010; Sypetkowski et al., 2023; Chandrasekaran et al., 2023). Yet they still sample only a tiny fraction of the space that a discovery campaign must search. Each additional plate also costs reagents and instrument time and adds weeks to the schedule. A model that simulates the morphological effect of an untested perturbation *in silico* would let experimentalists decide which perturbations are worth running. It would also help link an observed phenotype to a candidate mechanism. We call this task Cellular Morphology Perturbation Prediction (CMPP). Given untreated cells and the identity of a chemical or genetic perturbation, the goal is to generate the images that the perturbation would produce.

CMPP is inherently a generative problem rather than a regression problem. A single perturbation does not drive every cell to the same phenotype. Responses are heterogeneous and multimodal and vary with cell-cycle stage, local density and dose. The object to predict is therefore a *distribution* of plausible treated morphologies rather than a single image. The data add two further difficulties.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

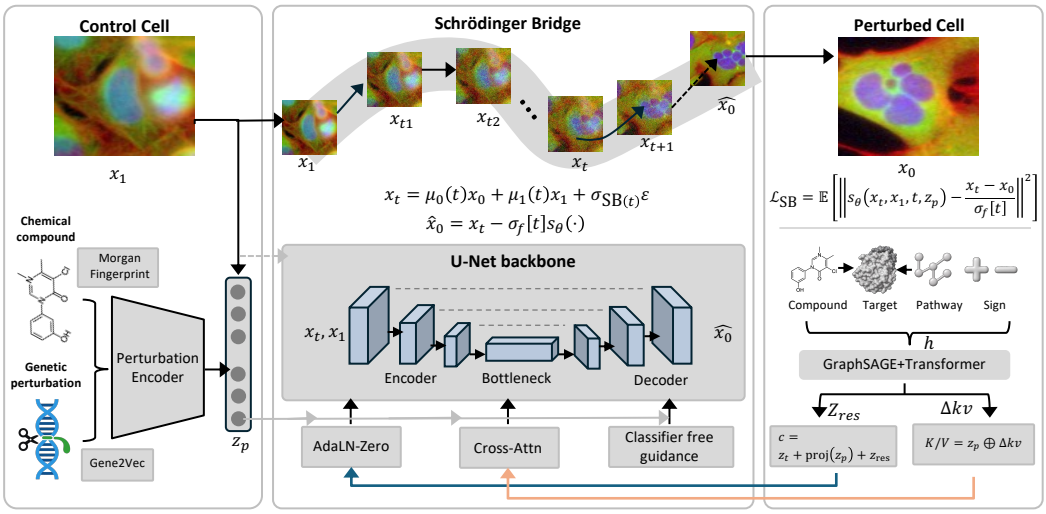


Figure 1: CellBridge. A same-plate control x_1 and a perturbation embedding z_p (Morgan fingerprint or Gene2Vec) drive a stochastic bridge from x_1 to a treated image x_0; the U-Net sees $[x_t, x_1]$ at every reverse step, conditioned through AdaLN-Zero, cross-attention and classifier-free guidance. For unseen compounds, the compound→target→pathway prior (right) adds the gated terms z_{res} and Δ_{kv} to that conditioning.

First, supervision is unpaired. Control and treated cells come from separate populations, so we never observe the same physical cell before and after treatment. Second, images are acquired across plates whose technical artifacts, such as staining and illumination drift, can be as large as the perturbation effect. An unconstrained model will then learn batch identity in place of biology. Deep generative models have become the main tool for this setting. Palma et al. (2025) perform unpaired style transfer with a GAN and correct for batch effects explicitly. Diffusion models (Bourou et al., 2024; Wang et al., 2025) translate real images between phenotype classes. Flow matching (Lipman et al., 2023; Zhang et al., 2025) replaces the stochastic process with an optimal transport ODE between unpaired control and treated populations. However, flow matching starts from a noise-corrupted control image and never revisits it. The control therefore informs generation only through its starting point.

To address this, we introduce **CellBridge**, a source-conditioned Schrödinger bridge (Figure 1). CellBridge transports each control cell toward its perturbed phenotype along a stochastic bridge and feeds the observed control image to the network at every reverse step. Each treated image is paired with a control from the same plate, so plate-level artifacts appear at both endpoints of the bridge. To handle unseen compounds, a structured compound → target → pathway prior injects curated pharmacology through gated conditioning. A new compound thus inherits the response of training compounds that share its target rather than that of its nearest scaffold. Beyond FID and KID, we also test whether the applied perturbation can be recovered from the generated images. Across BBBC021, RxRx1 and JUMP, CellBridge outperforms existing methods in both image quality and perturbation recovery and generalizes to structurally novel compounds.

Our contributions can be summarized as follows:

- We introduce CellBridge, a source-conditioned Schrödinger bridge for cellular perturbation prediction. It transports each observed control cell toward its perturbed phenotype along a stochastic bridge and feeds the control image to the network at every reverse step.
- We propose a structured compound → target → pathway prior that injects curated pharmacology into the bridge through gated conditioning, allowing unseen compounds to inherit the responses of training compounds that share their target and pathway rather than those of their nearest chemical scaffold.
- We evaluate generated morphology not only by image fidelity but by whether the applied perturbation can be recovered from it, and extensive experiments on chemical, genetic and

multimodal benchmarks show that CellBridge captures perturbation-specific phenotypes and generalizes to novel perturbations.

2 Related Work

Generative models for Cell Painting morphology. Existing approaches fall into three families. *Image-to-image translation* methods map an input cell image to its perturbed counterpart. They use either conditional diffusion between real phenotypes (Bourou et al., 2024) or an encoder-decoder with explicit batch correction (Palma et al., 2025). *Conditional generation from noise* methods train latent diffusion models that generate treated morphology from Gaussian noise. The condition is either a transcriptomic profile (Wang et al., 2025) or the chemical or genetic perturbation itself (Navidi et al., 2025). In both cases the control cell plays no role in generation. *Distribution transport* methods learn a map between unpaired control and treated populations. CellFlux (Zhang et al., 2025) does this with conditional optimal transport flow matching (Lipman et al., 2023).

Schrödinger bridges and diffusion bridges. The Schrödinger bridge problem (Schrödinger, 1932) asks for the stochastic process closest in KL divergence to a reference process whose marginals at both endpoints match prescribed distributions. It is equivalent to entropy-regularized optimal transport (Léonard, 2014). Scaling it to high-dimensional data required diffusion-based solvers. De Bortoli et al. (2021) run iterative proportional fitting with score networks, and Chen et al. (2022) derive a likelihood training objective from forward-backward SDE theory. Later work pins one endpoint to observed data instead of solving the full problem. The resulting diffusion bridges transport directly between two data distributions rather than passing through Gaussian noise (Zhou et al., 2024). Bridge matching learns such a transport for a fixed coupling as a mixture of reference bridges (Peluchetti, 2023; Shi et al., 2023). Augmented bridge matching (De Bortoli et al., 2023) also conditions the drift on the source endpoint, so the learned process preserves the coupling.

Conditioning and guidance. Classifier-free guidance (Ho & Salimans, 2022) is the standard way to strengthen the effect of conditioning in diffusion-family models at inference time, and we adopt it directly. Our conditioning design uses AdaLN-Zero and cross-attention and follows the pattern introduced by DiT (Peebles & Xie, 2023).

3 Method

3.1 Problem Setup

A Cell Painting field of view is a multi-channel image $x \in \mathcal{X} = \mathbb{R}^{C\times H\times W}$ in which each of the C channels records the fluorescence of one organelle stain. Let $\mathcal{P}$ denote the perturbation space. It covers chemical perturbations (small-molecule compounds) and genetic perturbations (siRNA knockdowns and CRISPR or ORF constructs). For each $p \in \mathcal{P}$ we observe an empirical population q_p of treated images over $\mathcal{X}$. We separately observe a population q_{ctrl} of control (untreated) images acquired on the same plates. No control image shows the same physical cell as a treated image. Supervision is therefore unpaired at the instance level and matched only at the level of distributions.

Given a control image $x_1 \sim q_{\mathrm{ctrl}}$ and a perturbation p, the task is to generate the treated morphology that p would produce. A single perturbation drives a population into several phenotypic states. The target is therefore a distribution rather than a point estimate. We learn a conditional generative model $p_\theta(x_0 \mid x_1, p)$ whose marginal over the control population matches the treated population:

$$\mathbb{E}_{x_1\sim q_{\mathrm{ctrl}}}\big[\, p_\theta(\,\cdot \mid x_1, p)\,\big] \;\approx\; q_p. \tag{1}$$

Model overview. CellBridge realizes p_θ as a stochastic bridge between a control image and a treated image. The control x_1 comes from the same plate as the target and conditions every step of the trajectory rather than only the initial state (Section 3.2). The perturbation enters through a frozen per-modality embedding. This embedding is injected by additive AdaLN-Zero modulation and cross-attention, and classifier-free guidance amplifies it at inference (Section 3.4). For out-of-distribution (OOD) perturbations we add an optional structured prior. It routes each perturbation through its target gene and pathway membership rather than through chemical structure alone (Section **??**). Appendix A.10 describes the backbone and training recipe.

3.2 SCHRÖDINGER BRIDGE TRANSPORT

CellBridge transports each control image toward the treated distribution along a stochastic process pinned at both ends. We index the process by a discrete bridge time $t \in \{0, 1, \ldots, T\}$. At $t{=}T$ the process holds the observed control image x_1, and at $t{=}0$ it holds a treated image x_0. We write x_t for the intermediate state. Let $\beta_t > 0$ be the noise increment that the process adds at step t. We define the noise standard deviations accumulated inward from each endpoint as

$$\sigma_f[t] = \Big(\sum_{s\le t} \beta_s\Big)^{1/2}, \qquad \sigma_b[t] = \Big(\sum_{s\ge t} \beta_s\Big)^{1/2}, \qquad D[t] = \sigma_f[t]^2 + \sigma_b[t]^2. \tag{2}$$

Here σ_f measures noise accumulated from the treated end x_0 and grows with t. Conversely, σ_b measures noise accumulated from the control end x_1 and shrinks with t. D normalizes the two. Because both endpoints are fixed, the distribution of x_t is proportional to the product of the two Gaussians induced by the endpoints. The result is a precision-weighted interpolation of the endpoints plus Gaussian noise:

$$x_t = \mu_{x_0}[t]\, x_0 + \mu_{x_1}[t]\, x_1 + \sigma_{\mathrm{sb}}[t]\, \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I), \tag{3}$$

with endpoint weights $\mu_{x_0}[t] = \sigma_b[t]^2/D[t]$ and $\mu_{x_1}[t] = \sigma_f[t]^2/D[t]$ that sum to one, and bridge noise $\sigma_{\mathrm{sb}}[t]^2 = \sigma_f[t]^2\sigma_b[t]^2/D[t]$. Each endpoint is weighted by the noise accumulated from the *other* side. As a result x_t stays close to x_1 near $t{=}T$ and close to x_0 near $t{=}0$. It is most uncertain in between, where σ_{sb} peaks.

The network s_θ predicts the residual between the current state and the treated endpoint, normalized by $\sigma_f[t]$. We train it with a plain MSE objective:

$$\mathcal{L}_{\mathrm{SB}} = \mathbb{E}_{t,x_0,x_1,\varepsilon} \left\| s_\theta(x_t, x_1, t, z_p) - \tfrac{x_t - x_0}{\sigma_f[t]} \right\|^2, \tag{4}$$

where z_p is the perturbation embedding of Section 3.4. The control image x_1 is fed to the network at every t and not only used as the starting state. At load time we pair each treated image with a random control from the *same plate*. The endpoint coupling is therefore $\pi(x_0, x_1) = q_{\mathrm{treated}}(x_0)\, q_{\mathrm{control}}(x_1 \mid \text{same plate})$. Inverting the residual gives the predicted clean treated image $\hat{x}_0 = x_t - \sigma_f[t]\, s_\theta(\cdot)$, which the reverse sampler iterates on.

This coupling is fixed in advance rather than optimized. The process it induces is optimal only in the following restricted sense.

Proposition 1 (Bridge with a fixed coupling)**.** *Let Q be the reference process with schedule $\{\beta_t\}$ and let P be a path measure with endpoint coupling π_P. Then*

$$\mathrm{KL}(P \,\|\, Q) = \mathrm{KL}(\pi_P \,\|\, \pi_Q) + \mathbb{E}_{\pi_P}[\mathrm{KL}(P^{x_0,x_1} \,\|\, Q^{x_0,x_1})]\,. \tag{5}$$

Consequently, among all path measures whose endpoint coupling equals a fixed π, the unique minimizer of $\mathrm{KL}(\cdot \,\|\, Q)$ is the mixture of reference bridges $P^\star = \int Q^{x_0,x_1}\, \mathrm{d}\pi$. Its time marginals are given by Equation 3. The Schrödinger bridge additionally minimizes the first term over π subject to the two marginal constraints. We do not.

We fix π instead of optimizing the marginal-constrained entropic coupling for two reasons. First, Cell Painting is a terminal assay that never images the same cell before and after treatment, so the data identify only the two marginals. Second, a pixel-space transport cost would tend to pair cells by nuisance factors such as position and brightness. A plate-restricted independent coupling avoids this and keeps plate-level technical variation at both endpoints. Appendix A.2 proves Proposition 1.

We use $T{=}1000$ steps and sample the *full stochastic* reverse process. Appendix A.6 gives the β-schedule and the reverse update, and Appendix A.16 compares this sampler with its deterministic limit.

3.3 LIMITING REGIMES: SCORE MATCHING AND FLOW MATCHING

The bridge leaves open two choices that standard formulations fix in advance. One is whether the terminal endpoint is data or noise. The other is whether the transport is stochastic. Both standard cases arise as limits of our bridge and follow from one fact, proved in Appendix A.3. The endpoint weights do not depend on the noise budget $\beta_{\max}$, while $\sigma_{\mathrm{sb}}[t]^2$ scales linearly with it.

Proposition 2. *Fix the shape of the schedule* $\{\beta_t\}$. (i) Gaussian terminal endpoint. *If* $x_1 \sim \mathcal{N}(0, I)$ *independently of* x_0*, then* $x_t \overset{d}{=} \mu_{x_0}[t]\, x_0 + b_t\, \xi$ *with* $\xi \sim \mathcal{N}(0, I)$ *and* $b_t^2 = \mu_{x_1}[t]^2 + \sigma_{\mathrm{sb}}[t]^2$*. Given* x_t*, the target of equation 4 is an invertible affine function of* ξ. (ii) Zero-noise limit. *As* $\beta_{\max} \to 0$ *the bridge variance vanishes and* $x_t \to (1 - \lambda_t)x_0 + \lambda_t x_1$ *with* $\lambda_t = \mu_{x_1}[t]$*. After division by the positive weight* $w(t) = \sigma_f[t]/D[t]$ *the target converges to* $x_1 - x_0$*. The objective thus becomes conditional flow matching up to a per-timestep weight.*

Part (i) recovers the Gaussian corruption kernel of score-based models. The affine relation lets us rewrite the residual objective in standard denoising form after a deterministic reparameterization. Part (ii) recovers the pairwise linear path of conditional flow matching, with a target collinear with the velocity $x_1 - x_0$. This second correspondence is weaker than the first. In optimal transport flow matching the linear path is paired with the optimal transport coupling, which arises as the zero-noise limit of entropic transport (Léonard, 2014). By Proposition 1 we make no claim that π is that coupling. What we recover is therefore conditional flow matching under our fixed coupling.

3.4 Perturbation Conditioning

The matched-embedding encoder above conditions an OOD perturbation on a single learned projection of a frozen structural fingerprint. Its only route to generalization is therefore proximity in fingerprint space. Empirically, zero-shot transfer on held-out BBBC021 compounds tracks *chemical similarity* to a training compound rather than a shared mechanism-of-action (MoA) class (Section 4.3). To fix this we add an optional conditioning path structured along the compound → target → pathway hierarchy.

Curated pharmacology supplies the inputs to this path. Binding annotations give each perturbation a primary target index $\tau(p)$ and a mechanism sign $\varsigma(p) \in \{-1, +1\}$ that marks inhibition or activation. Standard KEGG-style pathway knowledge (Kanehisa & Goto, 2000) gives a binary target-to-pathway membership matrix $M \in \{0, 1\}^{n_\tau \times n_\pi}$. None of these inputs is derived from any held-out treated image.

Let $E_\tau \in \mathbb{R}^{n_\tau \times d}$ and $E_\pi \in \mathbb{R}^{n_\pi \times d}$ be learned target and pathway embedding tables. We first contextualize the target table with a one-layer GraphSAGE (Hamilton et al., 2017) over the co-pathway adjacency A, which links targets that share a pathway:

$$\tilde{E}_\tau = \mathrm{SiLU}\big(W_{\mathrm{self}} E_\tau + W_{\mathrm{nbr}} A E_\tau\big). \tag{6}$$

Each target's representation thus absorbs those of its pathway neighbors. We then write the perturbation as three tokens. One carries its chemistry and two carry its mechanism:

$$u_{\mathrm{chem}} = W_{\mathrm{chem}}\, z_{\mathrm{struct}}, \qquad u_{\mathrm{tgt}} = \varsigma(p)\, \big[\tilde{E}_\tau\big]_{\tau(p)}, \qquad u_{\mathrm{path}} = \varsigma(p)\, \big[M E_\pi\big]_{\tau(p)}. \tag{7}$$

Here $\big[M E_\pi\big]_{\tau(p)}$ pools the embeddings of every pathway that contains the target. The sign flips both mechanism tokens so that inhibition and activation of the same target are not conflated. We add slot positional embeddings u_{pos} to the tokens and pass them through a two-layer Transformer encoder $\mathcal{T}$. The output leaves through two scalar gates:

$$h = \mathcal{T}\big([u_{\mathrm{chem}}, u_{\mathrm{tgt}}, u_{\mathrm{path}}] + u_{\mathrm{pos}}\big), \qquad z_{\mathrm{res}} = \gamma_z\, W_z h_{\mathrm{chem}}, \qquad \Delta_{kv} = \gamma_{kv}\, W_{kv} h. \tag{8}$$

We add the gated residual z_{res} to the conditioning vector c of equation **??** and the gated delta Δ_{kv} to the cross-attention keys and values. The prior therefore modulates the existing conditioning path rather than replacing it. By construction, an unseen compound that shares a target with a training compound inherits the u_{tgt} and u_{path} of that compound regardless of fingerprint distance. In the genetic datasets the perturbed gene *is* the target. We therefore key τ directly by gene symbol and skip the compound → target lookup.

Engaging the prior. The fingerprint branch alone fits the training compounds, so without intervention the gates γ_z and γ_{kv} receive little gradient. We therefore apply *fingerprint dropout* to a subset of training samples. On these samples the model can only use the pathway tokens. Ablations (Section 5.1) show that fingerprint dropout, rather than the nonzero gate initialization, is the key factor. Appendix A.12 reports the dropout schedule and the learned gate values.

4 EXPERIMENTS

4.1 SETUP

We evaluate on three Cell Painting datasets that cover chemical, genetic and multimodal interventions. BBBC021 contains 34 compounds in 12 MoA classes. We train on 26 compounds and hold out 8. RxRx1 contains 1,042 siRNA gene knockdowns, and JUMP contains 747 CRISPR, compound and ORF perturbations. Appendix A.5 provides the dataset details. Every model samples one image per control source. CellBridge uses guidance scale $g_s{=}2$ with $1{,}000$ reverse steps throughout.

We report pooled FID (FID_o), the mean of per-perturbation FIDs (FID_c) and the unbiased polynomial-kernel KID ($\times 100$). Our primary biological metric is perturbation recovery. A classifier trained on *real* images predicts the applied perturbation from a *generated* image. Higher accuracy means the generated morphology is diagnostic of the perturbation and not merely realistic. On BBBC021 we use the 13-class mechanism-of-action (MoA) classifier which covers the 12 mechanisms and the DMSO control. On RxRx1 and JUMP we fit logistic regression on DINOv2 (Oquab et al., 2023) features of real training images. The classifier is 100-way over the RxRx1 evaluation genes and 747-way over JUMP perturbation identity.

4.2 MAIN RESULTS

CellBridge achieves the best pooled FID, per-perturbation FID_c and KID on all three datasets (Tables 1 and 2). The largest improvement is on RxRx1. There CellBridge reaches $\mathrm{FID}_o{=}18.64$, compared with 33.08 for CellFlux and 36.12 for IMPA. In image quality, CellFlux and CellBridge generally outperform IMPA and PhenDiff. All CellBridge results use the pathway-prior configuration. Table 14 compares it with the no-pathway variant on the same backbone. Even without the prior, the bridge backbone outperforms every baseline on RxRx1. It reaches an FID_o of 21.42 against 33.08 and a gene-probe top-1 accuracy of 23.5% against 6.7%. On BBBC021 it also beats CellFlux in MoA accuracy (85.5% against 65.5%) at a similar pooled FID (18.29 against 18.73). CellFlux retains the better BBBC021 FID_c (56.83 against 65.18) and JUMP FID_o (8.96 against 10.45). The prior closes both gaps (Table 10).

Table 1: BBBC021 (chemical) and RxRx1 (genetic). Pooled (FID_o, KID_o) and per-perturbation (FID_c, KID_c) scores, KID $\times 100$; lower is better, best in **bold**.

	(a) BBBC021 (Chemical)				(b) RxRx1 (Genetic)			
Method	$\mathrm{FID}_o\downarrow$	$\mathrm{FID}_c\downarrow$	$\mathrm{KID}_o\downarrow$	$\mathrm{KID}_c\downarrow$	$\mathrm{FID}_o\downarrow$	$\mathrm{FID}_c\downarrow$	$\mathrm{KID}_o\downarrow$	$\mathrm{KID}_c\downarrow$
PhenDiff (Bourou et al., 2024)	49.64	109.63	3.10	4.99	66.02	173.26	5.17	7.63
IMPA (Palma et al., 2025)	37.34	78.95	3.04	4.64	36.12	153.41	2.51	3.83
CellFlux (Zhang et al., 2025)	18.73	56.83	1.53	2.12	33.08	167.49	2.38	4.88
CellBridge (ours)	**13.52**	**55.97**	**0.72**	**1.96**	**18.64**	**133.94**	**0.77**	**1.77**

Table 2: JUMP (multimodal)

Method	$\mathrm{FID}_o\downarrow$	$\mathrm{FID}_c\downarrow$	$\mathrm{KID}_o\downarrow$	$\mathrm{KID}_c\downarrow$
PhenDiff (Bourou et al., 2024)	48.85	126.79	5.01	6.95
IMPA (Palma et al., 2025)	14.57	100.11	1.02	2.62
CellFlux (Zhang et al., 2025)	8.96	84.42	0.64	1.43
CellBridge (ours)	**6.82**	**82.11**	**0.34**	**1.36**

Table 3 breaks the aggregate results down by perturbation. It covers six representative BBBC021 compounds and the first three RxRx1 genes in alphabetical order. CellBridge leads on five of the six compounds and two of the three genes. Across all 17 compounds in the BBBC021 evaluation set it leads on 11. On RxRx1 it leads on 86 of the 99 genes that have enough images for a per-gene FID.

Perturbation recovery separates the methods far more sharply than image quality does. On BBBC021 CellBridge leads in MoA accuracy and in both F1 scores by a wide margin (Appendix A.7, Table 8). It also retains more of the real within-compound spread than any baseline, at 0.86 of the real spread against 0.75 for CellFlux (Appendix A.8). The gap widens on the identity

Table 3: Per-perturbation FID (the per-condition terms averaged into FID_c); lower is better, best in **bold**.

	Chemical perturbations						Genetic perturbations		
Method	Alster.	AZ138	Bryo.	Colchi.	Mito. C	PP-2	ABCC4	ABHD11	ACSS1
PhenDiff (Bourou et al., 2024)	106.7	120.5	106.1	111.8	110.2	122.5	225.3	**156.4**	159.0
IMPA (Palma et al., 2025)	63.8	69.1	81.3	89.8	56.9	88.4	166.6	190.0	127.5
CellFlux (Zhang et al., 2025)	41.6	**44.1**	46.6	72.8	42.3	64.1	177.9	272.2	138.9
CellBridge (ours)	**38.6**	57.9	**41.9**	**58.3**	**35.2**	**53.1**	**132.9**	158.3	**110.1**

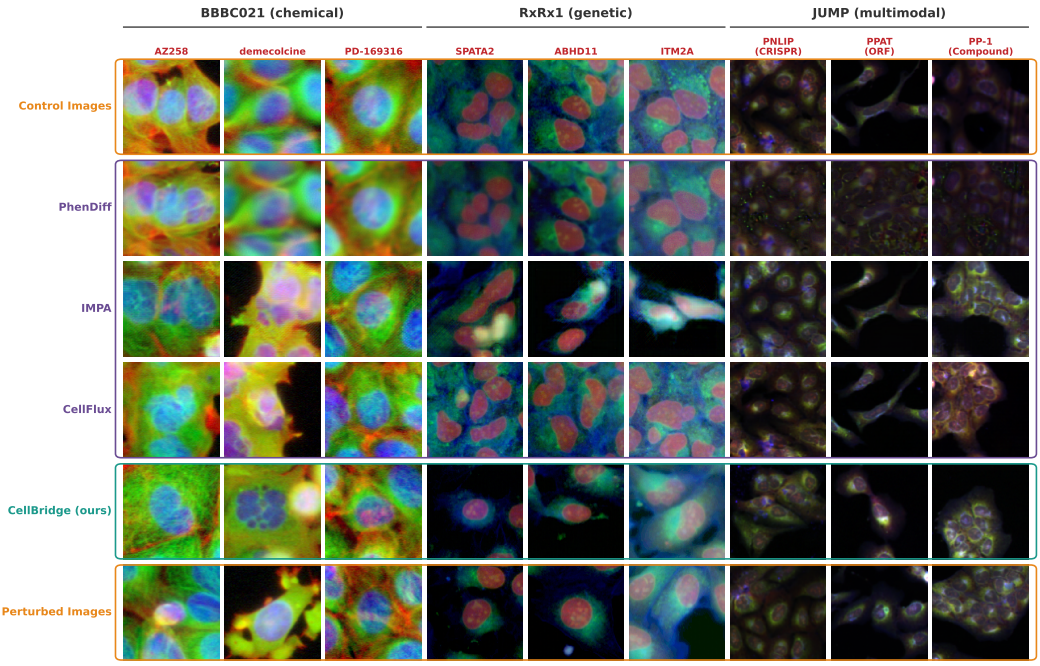


Figure 2: Qualitative comparison on BBBC021, RxRx1 and JUMP with three perturbations per dataset. Each column starts from one control field of view. The top panel shows the control and the next four show the generations of each method from it. The bottom panel shows a real treated image from the same plate.

probes (Appendix A.1, Tables 5–6). On RxRx1 CellBridge reaches a gene-identity top-1 accuracy of 33.4%. The best baseline, IMPA, reaches 6.7%, while PhenDiff and CellFlux reach 1.9% and 1.3% against a chance level of 1%. On JUMP its perturbation-identity top-1 accuracy of 3.3% is five times that of CellFlux (0.7%).

4.3 Generalization to Unseen Compounds

With fingerprint-only conditioning, an unseen compound is represented only through a projection of its frozen molecular fingerprint. Transfer therefore depends largely on chemical similarity to the training set. On the 8 compounds held out from BBBC021, the fingerprint-only CellBridge backbone achieves FID_o=33.74, FID_c=98.50 and 53.8% MoA accuracy. Success tracks chemical similarity more closely than MoA-class coverage. All three compounds with a chemically related analog in training transfer successfully. The other five fail, even though their MoA classes appear in training through chemically dissimilar compounds.

Pathway priors improve transfer across novel scaffolds. We next ask whether the compound → target → pathway prior enables transfer beyond fingerprint similarity. The five held-out compounds without a chemical analog in training are cycloheximide, cytochalasin D, epothilone B, lactacystin and latrunculin B. These are exactly the compounds on which fingerprint-only transfer fails. Each has a target represented in training, so they are the cases the prior is designed for. On these five compounds the pathway-conditioned model raises mean MoA accuracy from 18.6% to

67.3% (Table 4). Aggregate OOD pooled FID also improves from 33.7 to 24.1, compared with 27.7 for CellFlux on the same held-out set. Section 5.1 shows that this gain does not come from added encoder capacity.

The gain also does not depend on the MoA annotation. On the label-free DINOv2 effect-direction score, the held-out compounds reach 0.634 with the prior and 0.271 with fingerprints alone, against 0.507 for CellFlux. Retrieval among all 34 compounds still favors CellFlux (0.490 against 0.363). Appendix A.8 reports both metrics with image-resampling intervals. The curated pathways align closely with the MoA classes (Appendix A.12). The MoA gain should therefore be read as the prior transmitting pharmacological knowledge that is available before any image is taken. The prior also keeps in-distribution image quality high and gives the lowest in-distribution pooled FID among the CellBridge variants (13.5).

Transfer is not uniform across compounds. Epothilone B reaches only 18.7% MoA accuracy, although it shares microtubule-stabilization pathway tokens with taxol and docetaxel. We treat this case as a limitation rather than as evidence of mechanism-based transfer. Figure 3 in Appendix A.4 shows representative generations for five held-out compounds.

Table 4: BBBC021 zero-shot OOD. "5 novel scaffolds" = mean MoA accuracy on the five held-out compounds without a chemical analog in training; in-distribution FID_o as in Table 1a. Variant definitions are in Appendix A.13.

Model	OOD $\mathrm{FID}_o \downarrow$	OOD MoA acc. ↑	5 novel scaffolds ↑	In-dist $\mathrm{FID}_o \downarrow$
PhenDiff (Bourou et al., 2024)	67.57	9.6%	6.3%	49.64
IMPA (Palma et al., 2025)	73.61	45.8%	16.1%	37.34
CellFlux (Zhang et al., 2025)	27.72	54.8%	25.2%	18.73
CellBridge, fingerprint only	33.74	53.8%	18.6%	18.29
CellBridge, +capacity, dead gate	36.62	51.8%	11.2%	15.41
CellBridge, +pathway prior	**24.06**	**82.3%**	**67.3%**	**13.52**

5 Mechanistic Analysis and Ablation Studies

5.1 Disentangling Pathway Information from Architectural Capacity

Do the pathway gains come from biological information or from added capacity? We ablate the two training changes introduced with the pathway module, namely fingerprint dropout and nonzero gate initialization. We also add a capacity-matched control that feeds content-free tokens to the same gating architecture (Table 13). Each variant is a single training run, so the orderings below are single-run observations. Fingerprint dropout is the necessary ingredient. Without it the gates decay toward zero and the five-compound OOD mean falls to 16.1%. This is close to the dead-gate control at 11.2%. Removing only the nonzero initialization still lets the gates open from a cold start, and the model reaches 59.2%. This recovers 83% of the improvement of the full model over the fingerprint-only baseline. Added capacity alone does not explain the gain. The content-free control opens its gates just as wide but reaches only 26.6%.

5.2 Extension to Genetic Perturbations

The same encoder ports unchanged to RxRx1 (1,042 genes) and JUMP when queried by gene symbol (Table 14). Compared with the matched no-pathway backbone, the real pathway graph improves pooled FID and identity-probe accuracy on both datasets. Top-1 accuracy rises from 23.5% to 33.4% on RxRx1 and from 2.7% to 3.3% on JUMP. Each result comes from a single training run.

6 Conclusion

We presented CellBridge, a source-conditioned Schrödinger bridge for predicting Cell Painting morphology from chemical, genetic, and multimodal perturbations. By transporting the control population directly to the perturbed one and conditioning on a structured compound → target → pathway

prior, CellBridge recovers *which* perturbation was applied far more reliably than the released baselines. The same prior lets unseen BBBC021 compounds without a chemical analog in training inherit the response of training compounds that share their target and pathway, rather than of their nearest scaffold. What matters for in silico screening, our results argue, is whether a generated image reveals *which* perturbation produced it, not FID alone. Limitations and future work are in Appendix A.18.

AI USE STATEMENT

In this work, we used generative AI tools (Claude, Anthropic) to assist with running experiments, editing the manuscript text. We did not use generative AI tools to generate the core scientific ideas or experimental design. All AI-assisted code, text, and figures were reviewed by the authors before inclusion, and all reported numerical results were verified against raw experiment output logs. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Anis Bourou, Théo Peyrière, and Auguste Genovesio. PhenDiff: Revealing subtle phenotypes with diffusion models in real images. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2024.

Mark-Anthony Bray, Shantanu Singh, Han Han, Chadwick T Davis, Blake Borgeson, Cathy Hartland, Maria Kost-Alimova, Sigrun M Gustafsdottir, Christopher C Gibson, and Anne E Carpenter. Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nature protocols*, 11(9):1757–1774, 2016.

Peter D. Caie, Rebecca E. Walls, Alexandra Ingleston-Orme, Sandeep Daya, Tom Houslay, Rob Eagle, Mark E. Roberts, and Neil O. Carragher. High-content phenotypic profiling of drug response signatures across distinct cancer cells. *Molecular Cancer Therapeutics*, 9(6):1913–1926, 2010.

Srinivas Niranj Chandrasekaran, Jeanelle Ackerman, Eric Alix, D. Michael Ando, John Arevalo, Melissa Bennion, Nicolas Boisseau, Adriana Borowa, et al. JUMP cell painting dataset: morphological impact of 136,000 chemical and genetic perturbations. *bioRxiv*, 2023. doi: 10.1101/2023.03.23.534023.

Tianrong Chen, Guan-Horng Liu, and Evangelos A. Theodorou. Likelihood training of Schrödinger bridge using forward-backward SDEs theory. In *International Conference on Learning Representations (ICLR)*, 2022.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Valentin De Bortoli, Guan-Horng Liu, Tianrong Chen, Evangelos A. Theodorou, and Weilie Nie. Augmented bridge matching. *arXiv preprint arXiv:2311.06978*, 2023.

Jingcheng Du, Peilin Jia, Yulin Dai, Cui Tao, Zhongming Zhao, and Degui Zhi. Gene2vec: distributed representation of genes based on co-expression. *BMC Genomics*, 20(Suppl 1):82, 2019.

William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Minoru Kanehisa and Susumu Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

Zeinab Navidi, Jun Ma, Esteban A. Miglietta, Le Liu, Anne E. Carpenter, Beth A. Cimini, Benjamin Haibe-Kains, and Bo Wang. Morphodiff: Cellular morphology painting with diffusion models. In *International Conference on Learning Representations*, 2025.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. DINOv2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Alessandro Palma, Fabian J. Theis, and Mohammad Lotfollahi. Predicting cell morphological responses to perturbations using generative modeling. *Nature Communications*, 16, 2025. IMPA.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Stefano Peluchetti. Diffusion bridge mixture transports, Schrödinger bridge problems and generative modeling. *Journal of Machine Learning Research*, 24(374):1–51, 2023.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2015.

Erwin Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Annales de l'Institut Henri Poincaré*, 2(4):269–310, 1932.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger bridge matching. In *Advances in Neural Information Processing Systems*, 2023.

Maciej Sypetkowski, Morteza Rezanejad, Saber Saberian, Oren Kraus, John Urbanik, James Taylor, Ben Mabey, Mason Victors, Jason Yosinski, Alborz Rezazadeh Sereshkeh, Imran Haque, and Berton Earnshaw. RxRx1: A dataset for evaluating experimental batch correction methods. *arXiv preprint arXiv:2301.05768*, 2023.

Xuesong Wang, Yimin Fan, Yucheng Guo, Chenghao Fu, Kinhei Lee, Khachatur Dallakyan, Yaxuan Li, Qijin Yin, Yu Li, and Le Song. Prediction of cellular morphology changes under perturbations with a transcriptome-guided diffusion model. *Nature Communications*, 2025. MorphDiff.

Yuhui Zhang, Yuchang Su, Chenyu Wang, Tianhong Li, Zoe Wefers, Jeffrey Nirschl, James Burgess, Daisy Ding, Alejandro Lozano, Emma Lundberg, and Serena Yeung-Levy. CellFlux: Simulating cellular morphology changes via flow matching. In *International Conference on Machine Learning (ICML)*, 2025.

Linqi Zhou, Aaron Lou, Samar Khanna, and Stefano Ermon. Denoising diffusion bridge models. In *International Conference on Learning Representations (ICLR)*, 2024.

A Appendix

A.1 Per-Dataset Detail: RxRx1 and JUMP Gene/Perturbation-Probe Accuracy

Tables 5 and 6 report gene- and perturbation-identity probe accuracy for all four methods and for CellBridge without the structured prior, on the evaluation sets of Tables 1–2 (2,942 samples / 100 genes for RxRx1, 9,077 samples / 100 perturbations for JUMP); the baselines are scored on their released images. The probes are frozen DINOv2+logistic-regression classifiers fit once on real training images in the same input format as the evaluated images: 100-way on the six-to-three colour composites of 11,648 RxRx1 training images of the evaluation genes, and 747-way over BROAD_SAMPLE on the first three JUMP channels. They are never retrained per model; "real ceiling" is the same probe scored on real held-out images from the identical condition set. Because the probe vocabulary is much larger than the number of perturbations sampled here, even the real ceiling is well below 100%; these numbers should be read as a relative ranking among methods, not an absolute classification-quality figure.

Table 5: RxRx1 gene-identity probe accuracy, 100 genes / 2,942 samples. Real ceiling = the same probe on the matching real treated images, control wells excluded.

Method	Gene-acc. top-1 ↑	Gene-acc. top-5 ↑
PhenDiff (Bourou et al., 2024)	1.90%	8.40%
IMPA (Palma et al., 2025)	6.73%	23.42%
CellFlux (Zhang et al., 2025)	1.33%	5.88%
CellBridge, no pathway	23.45%	53.54%
CellBridge (ours)	**33.38%**	**64.96%**
Real ceiling	51.84%	80.90%

Table 6: JUMP perturbation-identity probe accuracy, 100 perturbations (44 CRISPR / 40 compound / 16 ORF) / 9,077 samples. Real ceiling as in Table 5.

Method	Pert-acc. top-1 ↑	Pert-acc. top-5 ↑
PhenDiff (Bourou et al., 2024)	0.48%	1.91%
IMPA (Palma et al., 2025)	0.55%	2.18%
CellFlux (Zhang et al., 2025)	0.66%	3.00%
CellBridge, no pathway	2.71%	9.16%
CellBridge (ours)	**3.27%**	**9.88%**
Real ceiling	5.62%	15.03%

A.2 Proof of Proposition 1

Write a path measure P as its endpoint coupling together with the bridges it induces between fixed endpoints: $\mathrm{d}P = \mathrm{d}\pi_P(x_0, x_1)\,\mathrm{d}P^{x_0,x_1}$, and likewise $\mathrm{d}Q = \mathrm{d}\pi_Q(x_0, x_1)\,\mathrm{d}Q^{x_0,x_1}$, where P^{x_0,x_1} is P conditioned on the two endpoints. This disintegration exists whenever $P \ll Q$. The Radon–Nikodym derivative then factorizes,

$$\frac{\mathrm{d}P}{\mathrm{d}Q} = \frac{\mathrm{d}\pi_P}{\mathrm{d}\pi_Q}(x_0, x_1)\ \frac{\mathrm{d}P^{x_0,x_1}}{\mathrm{d}Q^{x_0,x_1}}, \tag{9}$$

and taking $\log$ and integrating against P gives the chain rule for relative entropy,

$$\mathrm{KL}(P \,\|\, Q) = \mathrm{KL}(\pi_P \,\|\, \pi_Q) + \mathbb{E}_{\pi_P}[\mathrm{KL}(P^{x_0,x_1} \,\|\, Q^{x_0,x_1})]\,, \tag{10}$$

which is the stated identity (Léonard, 2014).

Now fix $\pi_P = \pi$. The first term of equation 10 is then a constant, and the second is an expectation of relative entropies, each of which is non-negative and vanishes exactly when $P^{x_0,x_1} = Q^{x_0,x_1}$. The minimizer is therefore $P^\star$ with $\pi_{P^\star} = \pi$ and $P^{\star\,x_0,x_1} = Q^{x_0,x_1}$, i.e. the mixture of reference bridges $P^\star = \int Q^{x_0,x_1}\,\mathrm{d}\pi$, and it is unique Q-a.s. For the Gaussian reference process with increments $\{\beta_t\}$, Q^{x_0,x_1} is the Brownian bridge pinned at x_0 and x_1, whose time-t marginal is the Gaussian product of the two endpoint-induced Gaussians; this is Equation 3.

The Schrödinger bridge between the two marginals ρ_0, ρ_1 minimizes $\mathrm{KL}(P \,\|\, Q)$ over *all* P with $P_0 = \rho_0$ and $P_1 = \rho_1$. By equation 10 that problem splits into the same bridge step plus a minimization of $\mathrm{KL}(\pi \,\|\, \pi_Q)$ over couplings of ρ_0 and ρ_1, whose solution is the entropic-optimal-transport coupling. We solve the first step only, at a π that is fixed in advance; we make no claim that this π is the entropic-optimal one.

A.3 Proof of Proposition 2

Notation. Throughout, $x \in \mathbb{R}^d$ with $d = CHW$, and as in Section 3.2

$$\sigma_f[t]^2 = \sum_{s \le t} \beta_s, \qquad \sigma_b[t]^2 = \sum_{s \ge t} \beta_s, \qquad D[t] = \sigma_f[t]^2 + \sigma_b[t]^2, \tag{11}$$

$$\mu_{x_0}[t] = \frac{\sigma_b[t]^2}{D[t]}, \qquad \mu_{x_1}[t] = \frac{\sigma_f[t]^2}{D[t]}, \qquad \sigma_{\mathrm{sb}}[t]^2 = \frac{\sigma_f[t]^2\,\sigma_b[t]^2}{D[t]}. \tag{12}$$

Adding the two weights gives the identity used repeatedly below,

$$\mu_{x_0}[t] + \mu_{x_1}[t] = \frac{\sigma_b[t]^2 + \sigma_f[t]^2}{D[t]} = \frac{D[t]}{D[t]} = 1. \tag{13}$$

Step 0: where the coefficients come from. The forward marginal of Section 3.2 is the normalized product of the two Gaussians that the endpoints induce at time t, namely $\mathcal{N}(x_0, \sigma_f[t]^2 I)$ seen from the treated end and $\mathcal{N}(x_1, \sigma_b[t]^2 I)$ seen from the control end. Writing the product of their densities and collecting terms in the exponent,

$$-\frac{\|x - x_0\|^2}{2\sigma_f^2} - \frac{\|x - x_1\|^2}{2\sigma_b^2} = -\tfrac{1}{2}\Big[\underbrace{\big(\tfrac{1}{\sigma_f^2} + \tfrac{1}{\sigma_b^2}\big)}_{=:\,\tau} \|x\|^2 - 2\Big\langle x,\ \tfrac{x_0}{\sigma_f^2} + \tfrac{x_1}{\sigma_b^2} \Big\rangle \Big] + \text{const}, \tag{14}$$

which is Gaussian with precision $\tau = D/(\sigma_f^2 \sigma_b^2)$, hence variance $\tau^{-1} = \sigma_f^2 \sigma_b^2 / D = \sigma_{\text{sb}}^2$, and mean

$$\tau^{-1}\Big(\frac{x_0}{\sigma_f^2} + \frac{x_1}{\sigma_b^2}\Big) = \frac{\sigma_b^2\, x_0 + \sigma_f^2\, x_1}{D} = \mu_{x_0}[t]\, x_0 + \mu_{x_1}[t]\, x_1. \tag{15}$$

Each endpoint is thus weighted by its own precision, which is the variance accumulated from the *opposite* side; this is the sense in which the interpolation is precision-weighted.

Step 1: scale invariance of the weights. Rescale the schedule as $\beta_s \mapsto c\,\beta_s$ for some $c > 0$, which rescales the budget as $\beta_{\max} \mapsto c\,\beta_{\max}$ while leaving its shape unchanged. Every partial sum scales with c, so $\sigma_f^2 \mapsto c\,\sigma_f^2$, $\sigma_b^2 \mapsto c\,\sigma_b^2$ and $D \mapsto c\,D$. Substituting,

$$\mu_{x_0} \mapsto \frac{c\,\sigma_b^2}{c\,D} = \mu_{x_0}, \qquad \mu_{x_1} \mapsto \frac{c\,\sigma_f^2}{c\,D} = \mu_{x_1}, \qquad \sigma_{\text{sb}}^2 \mapsto \frac{(c\,\sigma_f^2)(c\,\sigma_b^2)}{c\,D} = c\,\frac{\sigma_f^2 \sigma_b^2}{D} = c\,\sigma_{\text{sb}}^2. \tag{16}$$

The factor c cancels in the weights because they are ratios of like-degree sums, but survives in σ_{sb}^2 because the numerator is quadratic in the sums while the denominator is linear. Hence, at fixed shape, the endpoint weights are independent of $\beta_{\max}$ and the bridge variance is proportional to it. This is the observation quoted in Section 3.3, and both parts follow from it.

Part (i): the score-based limit. Let $x_1 \sim \mathcal{N}(0, I)$ be independent of x_0, and let $\varepsilon \sim \mathcal{N}(0, I)$ be independent of both.

(i.a) The two noise terms merge. Collect the terms of the forward marginal that do not involve x_0,

$$u_t := \mu_{x_1}[t]\, x_1 + \sigma_{\text{sb}}[t]\, \varepsilon. \tag{17}$$

Both summands are centered isotropic Gaussians and they are independent, so u_t is centered Gaussian with covariance $\big(\mu_{x_1}[t]^2 + \sigma_{\text{sb}}[t]^2\big) I$. Writing $b_t^2 := \mu_{x_1}[t]^2 + \sigma_{\text{sb}}[t]^2$ and $a_t := \mu_{x_0}[t]$, we get $u_t \overset{d}{=} b_t\, \xi$ with $\xi \sim \mathcal{N}(0, I)$, hence

$$x_t \ \overset{d}{=}\ a_t\, x_0 + b_t\, \xi, \tag{18}$$

which is exactly the forward marginal of a score-based diffusion.

(i.b) The schedule behaves like a noising schedule. At $t = 0$ we have $\sigma_f[0]^2 = \beta_0$ and $\sigma_b[0]^2 = \beta_{\max}$, so $a_0 = \beta_{\max}/(\beta_0 + \beta_{\max}) \approx 1$; at $t = T$ we have $\sigma_f[T]^2 = \beta_{\max}$ and $\sigma_b[T]^2 = \beta_T$, so $a_T = \beta_T/(\beta_{\max} + \beta_T) \approx 0$. The signal coefficient therefore decays from ≈ 1 to ≈ 0 as t runs from 0 to T, with b_t rising correspondingly: Equation 18 is a noising process in the usual sense.

(i.c) The objective becomes denoising score matching. The target of Equation 4 is $\ell_t := (x_t - x_0)/\sigma_f[t]$. Eliminating x_0 through $x_0 = (x_t - b_t\xi)/a_t$, which is valid since $a_t > 0$,

$$\begin{aligned} \ell_t &= \frac{(a_t - 1)\, x_0 + b_t\, \xi}{\sigma_f[t]} = \frac{1}{\sigma_f[t]} \left[(a_t - 1) \frac{x_t - b_t \xi}{a_t} + b_t \xi \right] \\ &= \underbrace{\frac{a_t - 1}{a_t\, \sigma_f[t]}}_{=:\,A_t}\, x_t \ + \ \frac{b_t}{\sigma_f[t]} \Big(1 - \frac{a_t - 1}{a_t}\Big)\, \xi \ = \ A_t\, x_t + \underbrace{\frac{b_t}{a_t\, \sigma_f[t]}}_{=:\,B_t}\, \xi. \end{aligned} \tag{19}$$

The minimizer of a squared-error regression is the conditional expectation, so the optimal network is $s^\star_\theta(x_t, t) = \mathbb{E}[\ell_t \mid x_t, t]$. Taking conditional expectations in Equation 19 and using that A_t, B_t are deterministic,

$$s^\star_\theta(x_t, t) = A_t\, x_t + B_t\, \mathbb{E}[\xi \mid x_t, t] = A_t\, x_t + B_t\, \varepsilon^\star_\theta(x_t, t), \tag{20}$$

where $\varepsilon^\star_\theta(x_t, t) = \mathbb{E}[\xi \mid x_t, t]$ is precisely the optimum of the denoising score matching objective $\mathbb{E}\|\varepsilon_\theta(x_t, t) - \xi\|^2$ for the process of Equation 18. Since $B_t \neq 0$, the map between the two optima is invertible, and by Tweedie's formula the score follows,

$$\nabla_{x_t} \log p_t(x_t) = -\frac{\varepsilon^\star_\theta(x_t, t)}{b_t} = -\frac{s^\star_\theta(x_t, t) - A_t\, x_t}{B_t\, b_t}. \tag{21}$$

Training on Equation 4 in this regime therefore estimates the score, up to the known affine reparameterization (A_t, B_t).

Part (ii): the flow-matching limit. Write $\lambda_t := \mu_{x_1}[t] = \sigma_f[t]^2/D[t]$, so that $\mu_{x_0}[t] = 1 - \lambda_t$ by Equation 13.

(ii.a) The marginal becomes deterministic. Let $x^{(c)}_t$ denote the state under the rescaled schedule $c\,\beta$. By Equation 16, λ_t is unchanged while $\sigma^{(c)}_{\mathrm{sb}}[t]^2 = c\,\sigma_{\mathrm{sb}}[t]^2$, so

$$\mathbb{E}\Big\| x^{(c)}_t - \big[(1-\lambda_t)x_0 + \lambda_t x_1\big] \Big\|^2 = \sigma^{(c)}_{\mathrm{sb}}[t]^2\, \mathbb{E}\|\varepsilon\|^2 = c\,\sigma_{\mathrm{sb}}[t]^2\, d \;\xrightarrow[c\to 0]{}\; 0. \tag{22}$$

The bridge marginal thus converges in mean square to the deterministic interpolation $(1-\lambda_t)x_0 + \lambda_t x_1$, for every t.

(ii.b) A uniform schedule gives the linear path. Take $\beta_s \equiv \beta$ for $s = 0, \dots, T$. Then $\sigma_f[t]^2 = \sum_{s=0}^{t} \beta = (t+1)\beta$ and $\sigma_b[t]^2 = \sum_{s=t}^{T} \beta = (T-t+1)\beta$, so $D[t] = (T+2)\beta$ and

$$\lambda_t = \frac{(t+1)\beta}{(T+2)\beta} = \frac{t+1}{T+2}, \tag{23}$$

which is affine in t and sweeps $(0, 1)$ as t runs from 0 to T. Reparameterizing time by λ turns the limit path into $\pi_\lambda(x_0, x_1) = (1-\lambda)x_0 + \lambda x_1$, the linear conditional path that conditional flow matching regresses against. For a non-uniform schedule the same path is traversed on a different time grid.

(ii.c) The objective becomes the flow-matching objective. Under the rescaled schedule, $x_t - x_0 = \lambda_t(x_1 - x_0) + \sigma^{(c)}_{\mathrm{sb}}[t]\,\varepsilon$, so the target of Equation 4 is

$$\frac{x_t - x_0}{\sigma^{(c)}_f[t]} = w_c(t)\,(x_1 - x_0) + \frac{\sigma^{(c)}_{\mathrm{sb}}[t]}{\sigma^{(c)}_f[t]}\,\varepsilon, \qquad w_c(t) := \frac{\lambda_t}{\sigma^{(c)}_f[t]} = \frac{\sigma^{(c)}_f[t]}{D^{(c)}[t]} > 0. \tag{24}$$

Both terms must be read relative to $w_c(t)$, which grows as $c^{-1/2}$: σ_f scales as $c^{1/2}$ and D as c, while the noise coefficient $\sigma_{\mathrm{sb}}/\sigma_f = \sigma_b/\sqrt{D}$ is scale-invariant. Substituting the reparameterization $s_\theta(\cdot, t) = w_c(t)\, v_\theta(\cdot, t)$,

$$\mathcal{L}_{\mathrm{SB}} = \mathbb{E}\Big[\, w_c(t)^2 \,\big\| v_\theta - (x_1 - x_0) - \eta_c(t)\,\varepsilon \big\|^2 \,\Big], \qquad \eta_c(t) := \frac{\sigma^{(c)}_b[t]\,\sqrt{D^{(c)}[t]}}{\sigma^{(c)}_f[t]} = c^{1/2}\,\eta_1(t) \;\xrightarrow[c\to 0]{}\; 0, \tag{25}$$

so after dividing by the positive weight $w_c(t)^2$ the residual noise vanishes as $\sqrt{c}$ and the per-timestep objective converges to the conditional flow matching objective $\mathbb{E}\|v_\theta - (x_1 - x_0)\|^2$. The weight changes the relative emphasis across timesteps but not the minimizer at any given t, so in the limit the optima satisfy $s^\star_\theta(\cdot, t) = w_c(t)\, v^\star_\theta(\cdot, t)$ with $v^\star_\theta(x_t, t) = \mathbb{E}[x_1 - x_0 \mid x_t, t]$, the conditional flow matching optimum. The statement is therefore about the rescaled field s_θ/w_c; the unrescaled target diverges as $c \to 0$. □

A.4 Qualitative Comparisons on Held-Out Compounds

Figure 3 shows generations for five of the eight held-out BBBC021 compounds, one control field of view per column.

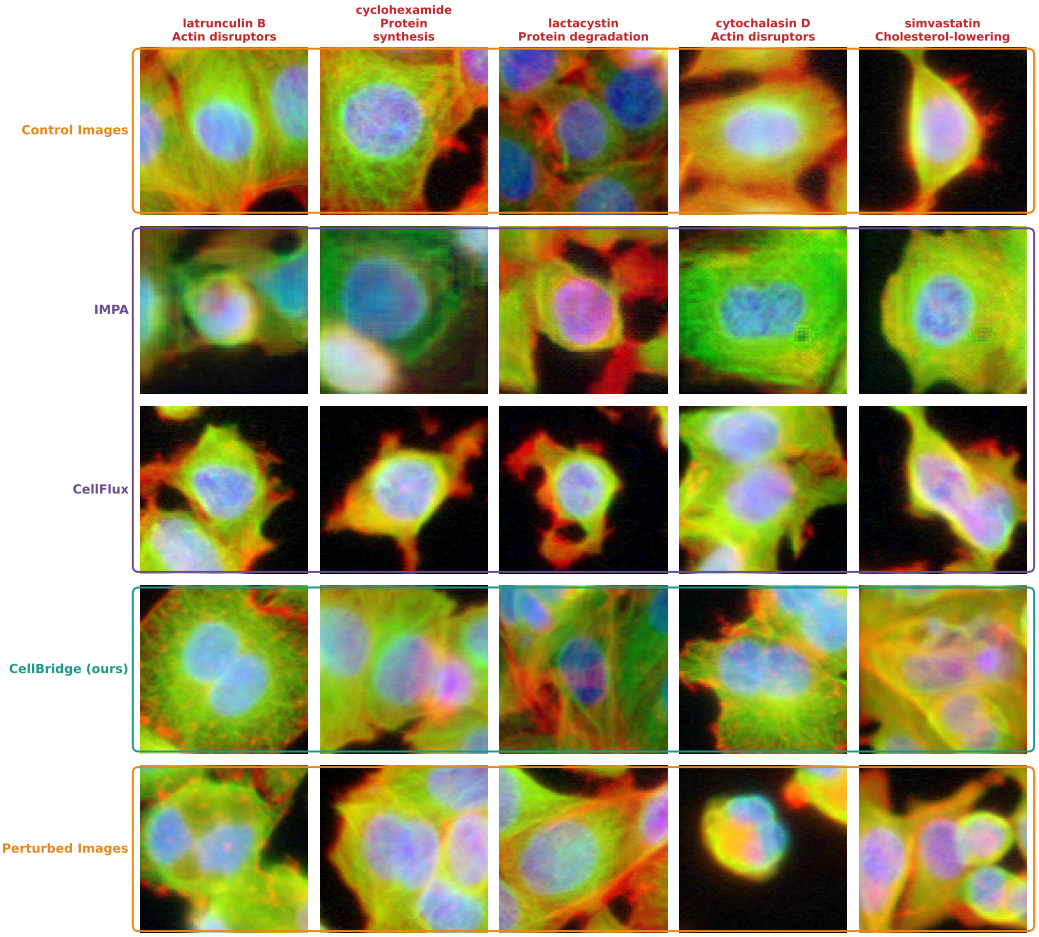


Figure 3: Qualitative comparisons on five held-out BBBC021 compounds. Each column shows one control field of view and the methods' generations from it; the bottom row is a real treated image from the same plate, a draw from the target distribution rather than that control cell's own outcome. PhenDiff is not shown: the images scored in Table 4 are those released by Zhang et al. (2025), which are not paired with our control fields. Latrunculin B and cytochalasin D disrupt actin; lactacystin inhibits the proteasome; cycloheximide blocks translation.

A.5 Datasets and Preprocessing

Table 7 summarizes the three datasets. Perturbation and image counts are treated wells only; each dataset additionally carries control wells (BBBC021 6,900, RxRx1 17,507, JUMP 89,139) that serve as the bridge source distribution. All images are 96×96 nuclei-centered crops. On BBBC021, training uses 26 compounds and holds out the remaining 8 for the evaluation of Section 4.3. The RxRx1 gene probe is 100-way over the evaluation genes, fit on their real training images alone (chance $= 0.01$).

Table 7: Dataset summary; counts are treated wells.

Dataset	Perturbation type	Channels	# Perturbations	# Images
BBBC021	chemical	3	34 compounds (12 MoA classes; 26 train / 8 held out)	90,604
RxRx1	genetic (siRNA)	6	1,042 genes (100-gene eval. subset)	153,436
JUMP	multimodal (CRISPR / cmpd / ORF)	5	747 (100-pert. eval. subset)	334,749

A.6 Noise Schedule and Reverse Sampling

Schedule. We use $T{=}1000$ steps and a symmetric β-schedule: β_t is spaced quadratically over the first half and mirrored over the second, with total budget $\sum_t \beta_t = \beta_{\max} = 0.3$. The symmetry makes σ_f and σ_b mirror images of each other, so the bridge noise σ_{sb} is bell-shaped, largest at the midpoint and smallest at the two ends.

The ends are *near*-pinned rather than exactly pinned, and it is worth being precise about by how much. With the inclusive convention $\sigma_f^2[t] = \sum_{s \le t} \beta_s$ and $\sigma_b^2[t] = \sum_{s \ge t} \beta_s$, the first term does not vanish: $\sigma_f^2[0] = \beta_0$, so

$$\sigma_{\mathrm{sb}}^2[0] \;=\; \frac{\beta_0\, \beta_{\max}}{\beta_0 + \beta_{\max}} \;\approx\; \beta_0, \tag{26}$$

and symmetrically at $t{=}T$. For our schedule $\beta_0 = 10^{-4}$, giving $\sigma_{\mathrm{sb}}[0] = \sigma_{\mathrm{sb}}[T] = 1.0 \times 10^{-2}$ against 0.188 at the midpoint, and endpoint weights $\mu_{x_0}[0] = 0.9993$, $\mu_{x_1}[0] = 7.1 \times 10^{-4}$. The process is therefore pinned to within one part in a hundred of the peak bridge noise, not exactly. Exact pinning would need the half-open convention $\sigma_f^2[t] = \sum_{s<t} \beta_s$; we report the inclusive one because it is what the implementation uses.

Sampling. Reverse sampling starts at $t{=}T$ from the observed control image x_1 and iterates toward $t{=}0$. At each transition from step n to an earlier step $m < n$, the network prediction is inverted to a clean-image estimate $\hat{x}_0 = x_n - \sigma_f[n]\, s_\theta(x_n, x_1, n, z_p)$, and x_m is drawn from the Gaussian product of $\mathcal{N}(\hat{x}_0,\ \sigma_f^2[m])$ and $\mathcal{N}(x_n,\ \delta^2)$ with $\delta^2 = \sigma_f^2[n] - \sigma_f^2[m]$:

$$x_m = \frac{\delta^2}{\sigma_f^2[m] + \delta^2}\, \hat{x}_0 + \frac{\sigma_f^2[m]}{\sigma_f^2[m] + \delta^2}\, x_n + \mathbf{1}_{[m>0]} \sqrt{\frac{\sigma_f^2[m]\, \delta^2}{\sigma_f^2[m] + \delta^2}}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{27}$$

The two coefficients sum to one. The last term is what distinguishes the stochastic posterior from its deterministic limit, which is Equation 27 with that term dropped (compared in Appendix A.16). The noise is suppressed on the final step ($m = 0$) so that sampling terminates at a clean image.

A.7 BBBC021 Mechanism-of-Action Classification

The classifier and evaluation script are the ones released by Zhang et al. (2025), applied unchanged to every method's generations over the same in-distribution test images as Table 1a; the real row scores the classifier on the real treated images of those same test keys. Without the structured prior CellBridge scores 2.6 points higher in distribution, so the prior trades a little in-distribution MoA accuracy for its held-out gains (Table 4).

Table 8: BBBC021 mechanism-of-action classification

Method	Accuracy ↑	Macro-F1 ↑	Weighted-F1 ↑
PhenDiff (Bourou et al., 2024)	52.27%	0.335	0.523
IMPA (Palma et al., 2025)	55.04%	0.378	0.576
CellFlux (Zhang et al., 2025)	65.45%	0.422	0.645
CellBridge, fingerprint only	**85.45%**	**0.665**	**0.847**
CellBridge (ours)	82.87%	0.583	0.822
Real images	73.73%	0.520	0.741

A.8 Effect Direction, Retrieval and Within-Compound Spread on BBBC021

Table 9 scores BBBC021 generations without any annotation, on the images of Tables 1a and 4. EffectSim is the cosine between the generated and real per-compound effect vectors in DINOv2 space, both measured from the real DMSO mean; retrieval queries the real means of all 34 compounds with each generated compound mean and reports mAP ($1/\mathrm{rank}$ of the correct compound); spread is the trace of the within-compound feature covariance, generated over real, so 1 matches the real heterogeneity and values below 1 indicate compression.

Table 9: BBBC021 effect direction, retrieval and within-compound spread (DINOv2 features).

Split	Method	EffectSim ↑	Retrieval mAP ↑	Spread (real = 1)
In distribution (17 compounds)	PhenDiff	0.482	0.404	0.834
	IMPA	0.491	0.648	0.825
	CellFlux	0.637	0.647	0.748
	CellBridge, fingerprint only	0.781	0.768	0.860
	CellBridge	**0.837**	**0.848**	**0.863**
Held out (8 compounds)	PhenDiff	0.066	0.094	0.817
	IMPA	0.366	0.290	0.732
	CellFlux	0.507	**0.490**	0.759
	CellBridge, fingerprint only	0.271	0.188	**0.884**
	CellBridge, free tokens	0.222	0.290	0.873
	CellBridge	**0.634**	0.363	0.880

CellBridge keeps more of the real within-compound heterogeneity than any baseline, so its mechanism-of-action accuracy above the real-image level (Table 8) does not come from collapsed, prototype-like samples; the classifier was trained on real images of all compounds, and generated images carry somewhat less nuisance variation than real ones. On the held-out compounds the label-free effect direction reproduces the ordering of the prior's controls in Table 13 (correct $>$ fingerprint only $\approx$ content-free), so the gain does not depend on the MoA annotation. Retrieval among all 34 compounds favours CellFlux on the held-out set.

The mechanism-of-action intervals below resample images within each held-out compound ($20{,}000$ draws); they capture sampling uncertainty for a fixed model, not variation across training seeds. Five novel scaffolds: +pathway prior 67.3% [64.9, 69.8], +fp dropout only 59.2% [56.7, 61.8], free tokens 26.6% [24.9, 28.4], fingerprint only 18.6% [14.8, 22.1], CellFlux 25.2% [21.0, 29.3]. The difference between the full prior and fp dropout only ($+8.1$ [4.5, 11.6]) excludes zero. Re-sampling the held-out set twice more from the same checkpoints moves the +pathway prior between FID_o 23.88 and 24.46 (MoA 82.1–82.9%) and the fingerprint-only backbone between 33.74 and 34.76 (MoA 53.4–54.2%), so sampling noise is far smaller than the gap between them.

A.9 What the Pathway Prior Adds In Distribution

Section 4.3 measures the prior where it was designed to help, on compounds held out of training. Table 10 asks what it costs or adds in distribution, on all three datasets and on the protocol of Tables 1 and 2, so the rows are directly comparable to the headline numbers. The comparison is between two CellBridge configurations, identical but for the prior.

It improves all four metrics on all three datasets. The gain is largest where the perturbation vocabulary is largest: on RxRx1, retrieval mAP rises from 0.480 to 0.555, and on JUMP from 0.376 to 0.435. The pooled FID improves by 2.8 to 4.8 points.

Table 10: The structured prior in distribution, FID on the protocol of Tables 1–2; BBBC021 EffectSim and retrieval are computed over all 26 in-distribution compounds. Both rows of a pair are CellBridge; they differ only in the prior.

Dataset	Configuration	FID_o ↓	FID_c ↓	EffectSim ↑	Retrieval mAP ↑
BBBC021	fingerprint only	18.29	65.18	0.786	0.770
	+pathway prior	**13.52**	**55.97**	**0.832**	**0.794**
RxRx1	no pathway	21.42	141.46	0.589	0.480
	+pathway prior	**18.64**	**133.94**	**0.631**	**0.555**
JUMP	no pathway	10.45	84.85	0.715	0.376
	+pathway prior	**6.82**	**82.11**	**0.720**	**0.435**

EffectSim is the cosine between the real and generated perturbation-effect vectors in DINOv2 space, both measured from the same real control mean; retrieval mAP queries the real per-perturbation means with the generated one and scores $1/\mathrm{rank}$. Neither uses a curated annotation, so neither can

be satisfied by echoing back the target and pathway the prior supplies. Being cosines, both are blind to the magnitude of the effect.

A.10 Architecture Specification

The backbone is a U-Net (Ronneberger et al., 2015) whose input is the channel-concatenation of the noisy state x_t and the control image x_1, so the source is available to every layer at every timestep. Table 11 gives the frozen perturbation embedding used per dataset.

Table 11: Frozen perturbation embeddings by dataset (Section 3).

Dataset	Embedding	Dimension
BBBC021 (chemical)	Molecular (Morgan) fingerprint	1024
RxRx1 (genetic)	Gene2Vec (Du et al., 2019)	200
JUMP (multimodal)	Combined BROAD_SAMPLE embedding	1224

A.11 Training Recipe

Training uses bf16 precision, AdamW, a warmup-plus-cosine learning-rate schedule, EMA decay 0.999, uniform timestep sampling, and gradient-norm clipping at 1.0. On BBBC021 we use a two-stage recipe, a 400k-step unguided pretrain followed by a 90k-step CFG-finetune stage.

A.12 Structured Biological Prior: Configuration

The pathway encoder (Section **??**) is part of every reported CellBridge model except the rows labelled fingerprint-only or no pathway. It adds $[\text{chemistry}, \text{target}, \text{pathway}]$ tokens ($K{=}3$), a one-layer co-pathway GraphSAGE (internal width 256) and a two-layer, 4-head Transformer encoder, projected to the 512-d conditioning width; it adds $\sim$3M parameters to the 110M backbone. BBBC021 uses 11 curated pathways (microtubule, actin, mitotic kinase, kinesin spindle, DNA topoisomerase, DNA alkylation, nucleotide synthesis, proteasome, translation, sterol metabolism, kinase signaling) with a default mechanism sign of -1 (inhibitor/disruptor) and per-compound $+1$ overrides for the microtubule *stabilizers* (taxanes, epothilones). Compound$\rightarrow$ target annotations cover 31/34 BBBC021 compounds; the remaining 3 DNA-alkylators (chlorambucil, cisplatin, mitomycin C, all training compounds) use a pseudo-target, so no held-out compound does. The module is warm-started from the matched-embedding Phase-2 checkpoint and fine-tuned for 150k steps, batch size 32, lr $5{\times}10^{-5}$, bf16, CFG dropout 0.1, gate init 0.1, fingerprint dropout 0.3, no auxiliary distribution-matching loss. Fingerprint dropout zeroes the fingerprint's additive term *and* its cross-attention key/value base independently on each selected row, while leaving the pathway z_{res} and Δ_{kv} in place, so the row can only be solved through the pathway tokens. On BBBC021 the trained gates settle at $\gamma_z{=}0.066$, $\gamma_{kv}{=}0.113$. Genetic ports (RxRx1, JUMP) build the target co-pathway graph from KEGG/Reactome gene-set membership and key the target vocabulary by gene symbol (RxRx1: 116 pathways over 677/1,070 genes; JUMP: 90 pathways over 114/155 genes); their trained gates settle at $\gamma_z{=}0.189$, $\gamma_{kv}{=}0.051$ (RxRx1) and $\gamma_z{=}0.065$, $\gamma_{kv}{=}0.092$ (JUMP), comparably engaged to the BBBC021 gates. The ORF (overexpression) modality currently reuses the same default mechanism sign (-1, loss-of-function) as CRISPR knockdown and compound inhibition; a per-modality sign ($+1$ for ORF) is the more biologically correct setting and is future work. The leave-one-out ablation and capacity control below (Appendix A.13) use the identical warm-start and schedule, varying only `pathway_gate_init`, `pathway_fp_dropout`, and a `pathway_freetoken` flag that replaces the $[\text{target}, \text{pathway}]$ token lookup with two learnable vectors shared across all perturbations (no E_{target}/E_{pathway}/GraphSAGE contribution); the flag is off by default and does not change any other checkpoint's forward pass.

Table 12 makes the overlap explicit: most mechanism classes own one pathway, so the prior tells the model which trained compounds an unseen compound resembles mechanistically. This is the knowledge a screen has before imaging, and the held-out gain measures how well the bridge turns it into morphology; the label-free effect direction of Appendix A.8 checks that the morphology, and not only the class, moves toward the real response.

Table 12: BBBC021 mechanism classes, curated primary targets and pathways; held-out compounds marked *. The two microtubule classes share a pathway and are separated by the mechanism sign; the kinase and topoisomerase pathways are shared across classes without such a separation.

MoA class	Targets	Pathways	Compounds
Actin disruptors	ACTB	actin	cytochalasin B, cytochalasin D*, latrunculin B*
Aurora kinase inhibitors	AURKB	mitotic kinase, kinase signaling	AZ258, AZ841*
Cholesterol-lowering	HMGCR	sterol metabolism	mevinolin, simvastatin*
DNA damage	TOP2A, pseudotarget	topoisomerase, alkylation	chlorambucil, cisplatin, etoposide, mitomycin C
DNA replication	DHFR, TOP1, TOP2A, TYMS	nucleotide synthesis, topoisomerase	camptothecin, floxuridine, methotrexate, mitoxantrone
Eg5 inhibitors	KIF11	kinesin spindle, mitotic kinase	AZ138
Epithelial	SRC	kinase signaling	PP-2
Kinase inhibitors	GSK3B, MAPK14, PRKCA	kinase signaling	alsterpaullone, bryostatin, PD-169316
Microtubule destabilizers	TUBB (−)	microtubule	colchicine, demecolcine, nocodazole, vincristine
Microtubule stabilizers	TUBB (+)	microtubule	docetaxel*, epothilone B*, taxol
Protein degradation	PSMB5	proteasome	ALLN, lactacystin*, MG-132, proteasome inhibitor I
Protein synthesis	RPL10, RPS14	translation	anisomycin, cycloheximide*, emetine

A.13 Structured Biological Prior: Gate Ablation and Capacity Control

The variants of Table 4 are: "+capacity, dead gate" = zero gate init, no fingerprint dropout, plus an auxiliary MMD loss. The dead-gate control keeps the auxiliary MMD loss of its training run, so it isolates capacity together with that loss; the content-free (`pathway_freetoken`) row below is the pure capacity control. PhenDiff's held-out row is scored on the images released by Zhang et al. (2025) for this split. Table 13 extends Table 4 with a leave-one-out ablation of the two recipe changes in Section **??** plus a content-free capacity control, all otherwise identical to +pathway prior. "+gate init only" drops fingerprint dropout; "+fp dropout only" resets the gate init to 0 (the original AdaLN-Zero identity init); "+free tokens, no biology" keeps the full recipe but replaces the [target, pathway] tokens with two content-free learnable vectors shared across all compounds (`pathway_freetoken`).

Table 13: BBBC021 zero-shot OOD ablation (extends Table 4); trained gate values in parentheses (γ_z/γ_{kv}).

Model (trained gates)	OOD FID$_o$ ↓	OOD MoA acc. ↑	5 novel scaffolds ↑	In-dist FID$_o$ ↓
+capacity, dead gate (∼0/∼0)	36.62	51.8%	11.2%	15.41
+gate init only (0.0003/0.0003)	34.20	53.5%	16.1%	14.84
+fp dropout only (0.032/0.053)	31.21	77.1%	59.2%	17.46
+free tokens, no biology (0.063/0.043)	32.70	55.5%	26.6%	14.38
+pathway prior (0.066/0.113)	**24.06**	**82.3%**	**67.3%**	**13.52**

Gate initialization alone is *not* sufficient: without fingerprint dropout's forced-reliance pressure the trained gates decay from their 0.1 start back to ∼0, indistinguishable from the dead-gate control, and the five-compound mean stays near chance. Fingerprint dropout alone, cold-started from γ=0, is close to sufficient by itself: the forced-reliance gradient opens the gate to 0.032/0.053 from a cold start and recovers 83% of the full recipe's gain over the fingerprint-only baseline. An open gate is not enough on its own, though: the free-token capacity control opens its gates just as wide (0.063/0.043) yet recovers only 26.6%, against the 18.6% fingerprint-only baseline. We did not have budget to re-run the free-token capacity control on the genetic ports; we flag this as future work.

A.14 Pathway Conditioning on Genetic Perturbations

The table behind Section 5.2.

A.15 Sampler Variants

We attempted an ϵ-prediction variant of the bridge to enable step-skipped sampling in the style of diffusion bridge implicit models, but found this training objective incompatible with our bridge's

Table 14: Pathway conditioning extended to genetic perturbations in-distribution, the encoder queried by gene symbol.

	No pathway	+Pathway
RxRx1 $\text{FID}_o \downarrow$	21.42	**18.64**
RxRx1 gene-probe top-1 $\uparrow$	23.5%	**33.4%**
JUMP $\text{FID}_o \downarrow$	10.45	**6.82**
JUMP perturbation top-1 $\uparrow$	2.7%	**3.3%**

residual parameterization in this setting. CellBridge therefore uses full 1000-step stochastic sampling, an inference-cost limitation relative to CellFlux's few-step sampler.

We also explored inference-time manifold steering as a remedy for far-out-of-distribution compounds and found it ineffective: generated samples were already on the real-data manifold, so the failure mode (mode collapse within the manifold) requires a training-time fix rather than a post-hoc correction.

A.16 Isolating the Sampler and the Source

Table 15 varies two factors on one checkpoint (the BBBC021 pathway model) and one split (the 26 training compounds, 7,119 test images): the reverse update with or without its noise term (Equation 27), and the per-step source condition set to the image's own control or to the control of the next image in the batch, with the bridge endpoint unchanged. Dispersion is the trace of the within-compound Inception feature covariance, generated over real, averaged over compounds. Because this table uses all 7,119 in-distribution test images rather than the 5,120 of Table 1a, its FID and MoA values are comparable within the table but not to Tables 1 and 8. Dispersion here is measured in Inception features and is not comparable to the DINOv2 spread of Table 9.

Table 15: Sampler $\times$ source condition, BBBC021 in distribution. MoA accuracy uses the 13-way classifier of Appendix A.7.

Sampler	Source condition	$\text{FID}_o \downarrow$	MoA acc. $\uparrow$	Dispersion
stochastic	own control	**14.92**	**80.3%**	0.945
deterministic	own control	15.06	80.2%	0.945
stochastic	swapped	82.42	25.4%	1.219
deterministic	swapped	83.46	25.2%	1.217

The two samplers are indistinguishable: they differ by less than two runs of the same deterministic configuration (15.06 and 15.33). Swapping the source condition is not, under either sampler: pooled FID rises to about 83, MoA accuracy falls to 25%, and the within-compound spread exceeds the real one, so the network reads the source channel rather than ignoring it. The swapped input pairs an endpoint with a condition taken from a different cell, a combination absent from training, so the table shows that the source is used, not that the response it produces is specific to that cell.

A.17 Qualitative Comparisons on the Genetic Benchmarks

Figures 4 and 5 extend Figure 2 on the two genetic benchmarks. A column shares a source: every method generated from the control at the top, and the bottom row is a real treated image of the same perturbation on the same plate, not that cell's outcome.

A.18 Limitations and Future Work

(1) CellBridge requires full 1000-step stochastic reverse sampling, so its inference cost is substantially higher than that of few-step flow-matching samplers (Appendix A.15). (2) Our RxRx1 results use the published 100-gene evaluation subset, not the full 1,042-gene panel. (3) The structured biological prior depends on a curated target/pathway annotation that exists for well-studied compounds and genes but not universally. (4) Far-out-of-distribution compounds fail through mode collapse within the real-data manifold, which requires a training-time fix rather than a post-hoc correction (Appendix A.15).

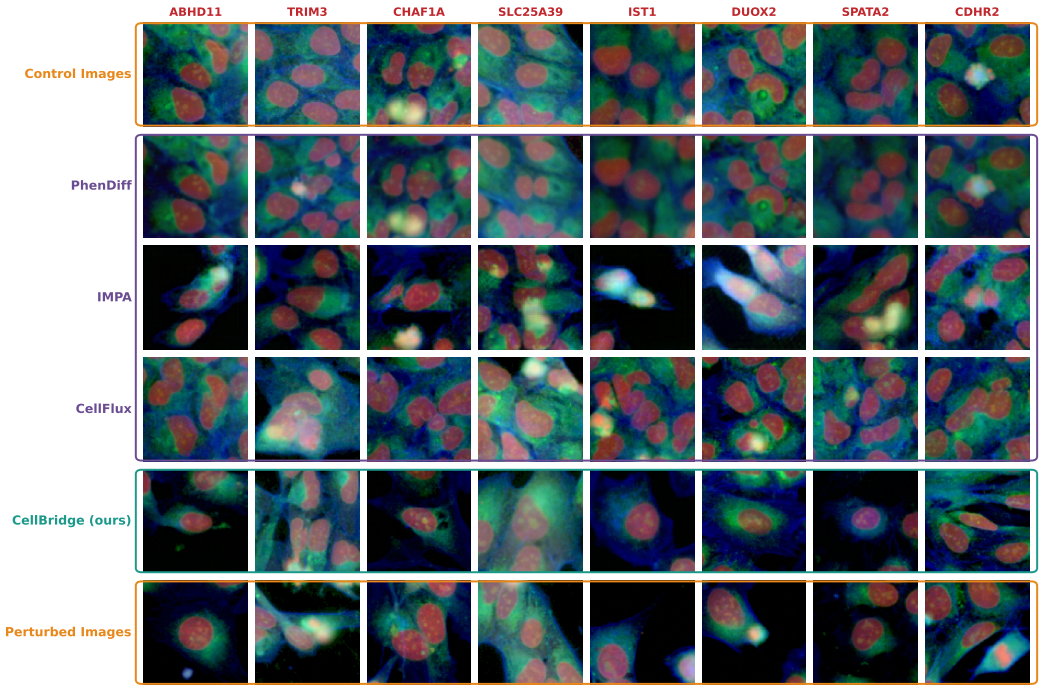


Figure 4: RxRx1, eight genes. Each column shows one control field of view and the methods' generations from it; the bottom row is a real treated image of the same gene on the same plate.

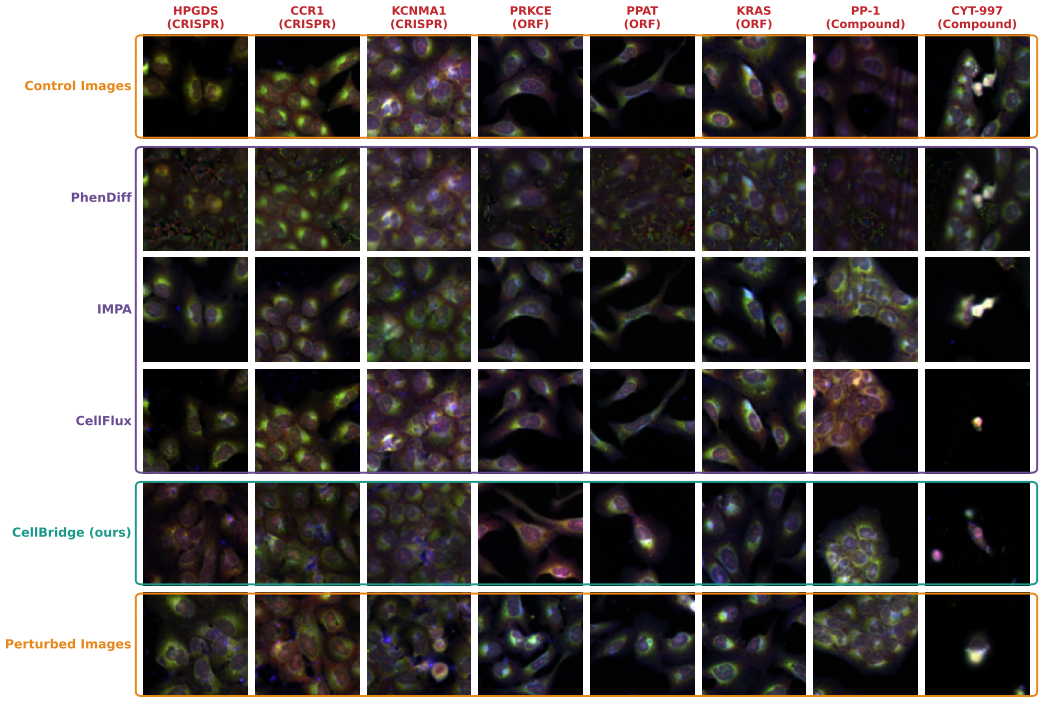


Figure 5: JUMP, three perturbations, one per modality (CRISPR, ORF, compound). Layout as in Figure 4.

Future work: training-time diversity regularization for far-OOD mode collapse; closing the architectural FID gap without sacrificing biological fidelity, potentially via single-cell-level latent bridging; and full RxRx1 1,042-gene evaluation.

A.19 Reproducibility Note

Code, trained checkpoints for all three datasets, and the generated images scored in every table will be released upon publication. All baselines are evaluated on the generated images released by Zhang et al. (2025) with its unmodified evaluation scripts; CellBridge samples with $g_s{=}2$ and $1{,}000$ stochastic reverse steps, one sample per control source. Sampling the 1,979 held-out BBBC021 images takes 58 minutes on one H200 GPU (batch 32, guidance doubling the network evaluations), about $1.75\,\mathrm{s}$ per image.