Supplementary Material for Mol2Image: Improved Conditional Flow Models for Molecule to Image Synthesis

Karren Yang[1] Samuel Goldman[1] Wengong Jin[1] Alex X. Lu[2]
Regina Barzilay[1] Tommi Jaakkola[1] Caroline Uhler[1]
[1]Massachusetts Institute of Technology [2]University of Toronto

A. Theoretical Analysis of Haar Pyramid Flow

In the main text, we propose a generative flow model based on the framework of a Haar wavelet image pyramid. Our model is trained to generate images in a coarse-to-fine fashion with respect to the image pyramid, by uncoupling the training of a multi-scale flow model down into training of conditional flow blocks. Specifically, we optimize the parameters of the flow model f by maximizing the conditional log-likelihood of the fine features $\tilde{\mathbf{x}}_i$ given the coarser image $\mathbf{x}_{i+1}$ for every level of the image pyramid (except the last layer, which uses standard log-likelihood):

$$\mathcal{L}(\mathbf{x}) = \Big(\log p_{\mathcal{N}}\big(\mathbf{z}_k; \mu_k, \Sigma_k\big) + \log|\det \frac{d\mathbf{z}_k}{d\mathbf{x}_k}|\Big) + \sum_{i=0}^{k-1}\Big(\log p_{\mathcal{N}}\big(\mathbf{z}_i; \mu_i(\mathbf{x}_{i+1}), \Sigma_i(\mathbf{x}_{i+1})\big) + \log|\det \frac{\partial \mathbf{z}_i}{\partial \tilde{\mathbf{x}}_i}|\Big)$$

We then state that optimizing these conditional log-likelihoods is equivalent to optimizing the log-likelihood of the data. The proposition is restated and the proof is provided below.

Proposition 1. *Let f denote the multi-scale flow model based on a Haar image pyramid. Given an image $\mathbf{x} \in \mathbb{R}^{C\times W\times W}$ ($C \geq 1, W = 2^K, K \geq k$), the log-likelihood of sampling $\mathbf{x}$ from f can be computed exactly as,*

$$\log p(\mathbf{x}) = \mathcal{L}(\mathbf{x}) + CW^2 \log 2 \sum_{i=0}^{k-1} 2^{1-2(i+1)}.$$

Proof. We first rewrite the left side using the change-of-variables formula,

$$\log p(\mathbf{x}) = \log p(\mathbf{z}) + \log|\det \frac{d\mathbf{z}}{d\mathbf{x}}|, \tag{1}$$

letting $\mathbf{z}$ denote $\mathbf{z_0} \cdots \mathbf{z_k}$. The first term of (1) can be decomposed as follows,

$$\log p(\mathbf{z}) = \log p_{\mathcal{N}}(\mathbf{z}_k; \mu_k, \Sigma_k) + \sum_{i=0}^{k-1} \log p_{\mathcal{N}}(\mathbf{z}_i; \mu_i(\mathbf{x}_{i+1}), \Sigma_i(\mathbf{x}_{i+1})), \tag{2}$$

using the chain rule and noting that $\mathbf{x}_{i+1}$ is a deterministic function of $\mathbf{z}_{i+1} \cdots \mathbf{z}_k$. For the second term of (1), noting that $\mathbf{x} = \mathbf{x}_0$ (the image at the top level of the pyramid), we have:

$$\begin{aligned}\log|\det \frac{d\mathbf{z}}{d\mathbf{x}}| &= \log|\det \frac{d\mathbf{z}}{d[\mathbf{x}_1, \tilde{\mathbf{x}}_0]}| + \log|\det \frac{d[\mathbf{x}_1, \tilde{\mathbf{x}}_0]}{d\mathbf{x}_0}| \\ &= \log|\det \frac{d\mathbf{z}_1 \cdots \mathbf{z}_k}{d\mathbf{x}_1}| + \log|\det \frac{\partial \mathbf{z}_0}{\partial \tilde{\mathbf{x}}_0}| \\ &\quad + \log|\det \frac{d[\mathbf{x}_1, \tilde{\mathbf{x}}_0]}{d\mathbf{x}_0}|\end{aligned}$$

where the second equality follows because $\frac{d\mathbf{z}_1 \cdots \mathbf{z}_k}{d\mathbf{x}_1}$ is a triangular block matrix. Recursively applying this decomposition to the term $\log|\det \frac{d\mathbf{z}_i \cdots \mathbf{z}_k}{d\mathbf{x}_i}|$, we obtain:

$$\begin{aligned}\log|\det \frac{d\mathbf{z}}{d\mathbf{x}}| &= \log|\det \frac{d\mathbf{z}_k}{d\tilde{\mathbf{x}}_k}| + \sum_{i=0}^{k-1}\big(\log|\det \frac{\partial \mathbf{z}_i}{\partial \tilde{\mathbf{x}}_i}|\big) \\ &\quad + \sum_{i=0}^{k-1} \log|\det \frac{d[\mathbf{x}_{i+1}, \tilde{\mathbf{x}}_i]}{d\mathbf{x}_i}|\end{aligned} \tag{3}$$

Note that the last term corresponds to the sum of log-determinants of the Jacobian of the Haar wavelet transforms at every level of the pyramid, which is computed exactly as $\log|\det \frac{d[\mathbf{x}_{i+1}, \tilde{\mathbf{x}}_i]}{d\mathbf{x}_i}| = CW^2 \log 2 \cdot 2^{1-2(i+1)}$. The proof follows from plugging (2) and (3) into (1). □

B. CellProfiler Evaluation

CellProfiler [4] is a standard open-source software used for segmenting cells/nuclei and quantifying specific morphological features. The segmentation of nuclei and cells occurs

	CellProfiler Metrics					Correspondence Accuracy						SWD
	Coverage	Count	Size	Zernike	Exp. Level	Mito	ER	RNA	Cyto	DNA	Overall	
Ground Truth (Upper Bound)	-	-	-	-	-	59.9	59.2	60.5	58.3	61.8	64.2	-
CGAN	6.4	1.9	-1.5	-1.0	9.2	53.0	52.4	**55.3**	50.8	56.2	56.1	56.1
CGlow	3.1	-3.7	-3.0	-3.1	3.7	51.1	50.9	51.9	52.2	54.3	54.5	5.40
CGlow+Contrast	9.2	1.7	**12.9**	**6.1**	8.6	**55.8**	53.2	**55.4**	**56.4**	**58.0**	59.1	4.20
Pyramid Flow	5.0	9.1	6.1	2.9	9.2	51.8	52.0	52.5	52.7	53.3	55.7	**3.41**
Pyramid Flow+Contrast (Mol2Image)	**15.8**	**19.7**	11.0	4.9	**13.4**	**55.3**	**54.6**	**55.4**	**55.8**	**57.6**	**62.6**	4.27

Table 1: Evaluation of Mol2Image (our model) vs. the baselines on images generated from molecules that were held-out from the training set. "CellProfiler Metrics" are Spearman correlation coefficients ($\times 10^2$) between biological features from real and generated images; higher is better. "Correspondence accuracy" represents the accuracy of a pretrained correspondence classifier model evaluated on generated images; higher is better and ground truth (upper bound) achieves between 60.0 and 65.0. "SWD" is the sliced Wasserstein distance metric ($\times 10^{-2}$) from [2]; lower is better. All results within 1% of the best are shown in bold font. See main text for details.

Image Size	Glow [3]	Haar Pyramid Flow
8 x 8	4.49	4.49
16 x 16	4.55	4.48
32 x 32	5.27	5.23
64 x 64	5.78	5.83

Table 2: Log-likelihood of validation images computed from Glow vs. our pyramid flow model. The values are comparable, which supports our theoretical analysis in Proposition 1 that our approach scales flow models to larger images without changing the log-likelihood objective.

in two steps: (1) thresholding is performed to identify the nuclei from the DNA stain, and (2) the nuclei are used as reference points for determining boundaries between cells and identifying cell objects. Once the cells are identified, multiple pipelines are available to measure shape and intensity features within each cell. To evaluate the generated images from our model, we extract morphological features for a subset of generated and held-out images and compute the correlation coefficient between the features of generated and real images. To increase the range of phenotypes within the evaluated subset, we focus our evaluation on molecules that are more likely to cause a morphological change in cells and thus more likely to be of interest to practitioners, based on the atypical morphology criterion described in Section C.

C. Analysis of Molecular Embeddings

As described in the main text, to evaluate the molecular embedding space learned by our graph neural network, we train a linear classifier to predict a subset of morphological features. We curate the labels for this task as follows. The dataset of Bray *et al*. [1] provides measured values of the following morphological features for every cell image in their dataset: area, compactness, eccentricity, form factor, major axis length, minor axis length, radius, perimeter, solidity, and cell count. To each small molecule, we assign a continuous-valued vector representing the mean values of these features observed in cells treated with that molecule. Direct prediction of these values is not a meaningful task because the amount of intra-molecular variability is high relative to the inter-molecular variability; much of the variability in the features may be naturally occurring due to stochasticity in cell growth and is not explained by the molecular perturbation. Therefore, we predict instead the presence of atypical morphology caused by a molecule. We convert these continuous values to binary labels – 1 if the value is in the top or bottom 1% / 5% / 10% of the values for its class, and 0 otherwise – and train a logistic regression model to perform multi-task binary classification. The results in Table 4 of the main text show that the molecular embeddings learned by our graph neural network reflect morphological properties of treated cells and enable linear separation of molecules that cause atypical morphological features.

D. Relation to Biological Assays

To show that our generated images are biologically meaningful, we similarly extract image features from both real and generated images using a pretrained Wide-ResNet50, and use logistic regression on the image features to predict standard biological assays extracted from the ChEMBL database. The results in Figure 1 show that our generated images are able to achieve strong predictive performance (median ROC-AUC of 0.71) of drug assays that approaches ground truth performance (median ROC-AUC of 0.80) and greatly outperforming the conditional Glow baseline (median ROC-AUC of 0.49).

E. Additional Tables

Supplementary Table 1 shows the complete results of the different approaches on generating images corresponding to held-out molecules, and supplements the results of Table 2 of the main text. Our conditional flow models that use contrastive learning during training (as we propose in Section 3.3 of the main text) outperform the other models

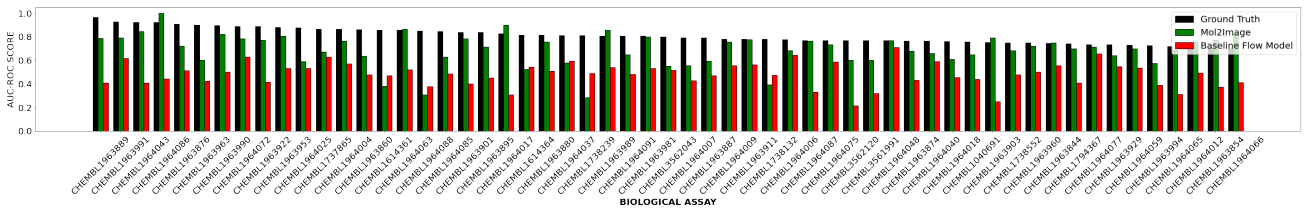

Figure 1: AUC-ROC values of logistic regression models trained to predict the outcomes of biological assays from image features. Higher is better.

in generating images corresponding to held-out molecules. Supplementary Table 2 shows empirical evidence that the log-likelihoods computed by Glow and our pyramid flow model are equivalent.

F. Additional Qualitative Examples

Supplementary Figure 2 shows additional examples of full-resolution cell images generated by the unconditional version of our Haar pyramid flow model. Supplemental Figures 3 and 4 shows examples of full-resolution cell images generated by our improved conditional flow model corresponding to different molecular treatments.

References

[1] Mark-Anthony Bray, Sigrun M Gustafsdottir, Mohammad H Rohban, Shantanu Singh, Vebjorn Ljosa, Katherine L Sokolnicki, Joshua A Bittker, Nicole E Bodycombe, Vlado Dančík, Thomas P Hasaka, et al. A dataset of images and morphological profiles of 30 000 small-molecule treatments using the cell painting assay. *Gigascience*, 6(12):giw014, 2017. 2

[2] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *ICLR*, 2018. 2

[3] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pages 10215–10224, 2018. 2

[4] Claire McQuin, Allen Goodman, Vasiliy Chernyshev, Lee Kamentsky, Beth A Cimini, Kyle W Karhohs, Minh Doan, Liya Ding, Susanne M Rafelski, Derek Thirstrup, et al. Cellprofiler 3.0: Next-generation image processing for biology. *PLoS biology*, 16(7), 2018. 1

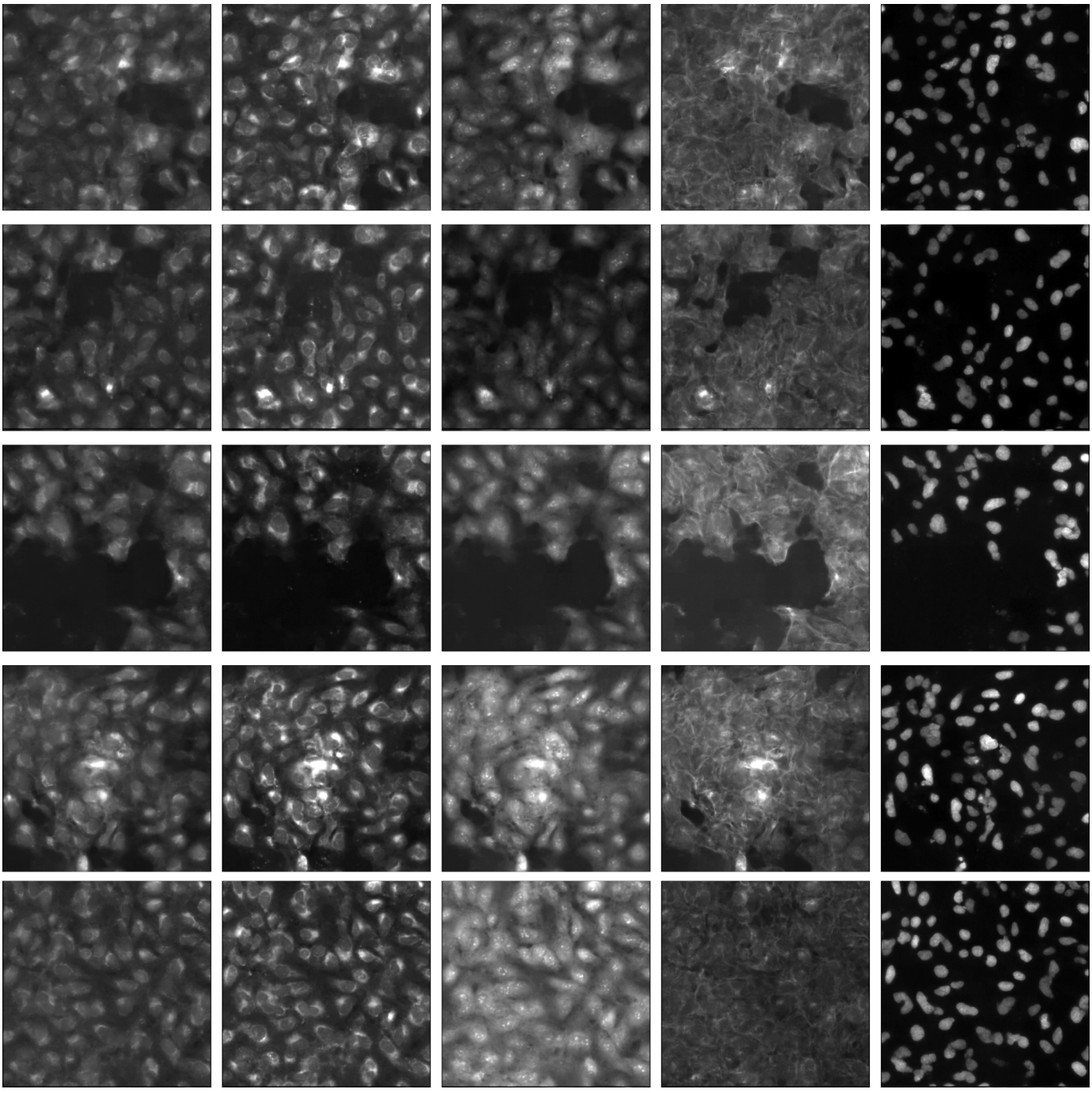

Figure 2: Additional examples of 5-channel 512 $\times$ 512 cell images generated by our multi-scale Haar image pyramid flow model. From left to right: mitochondria, endoplasmic reticulum, nucleoli/cytoplasmic RNA, actin (cytoskeleton), DNA (nucleus).

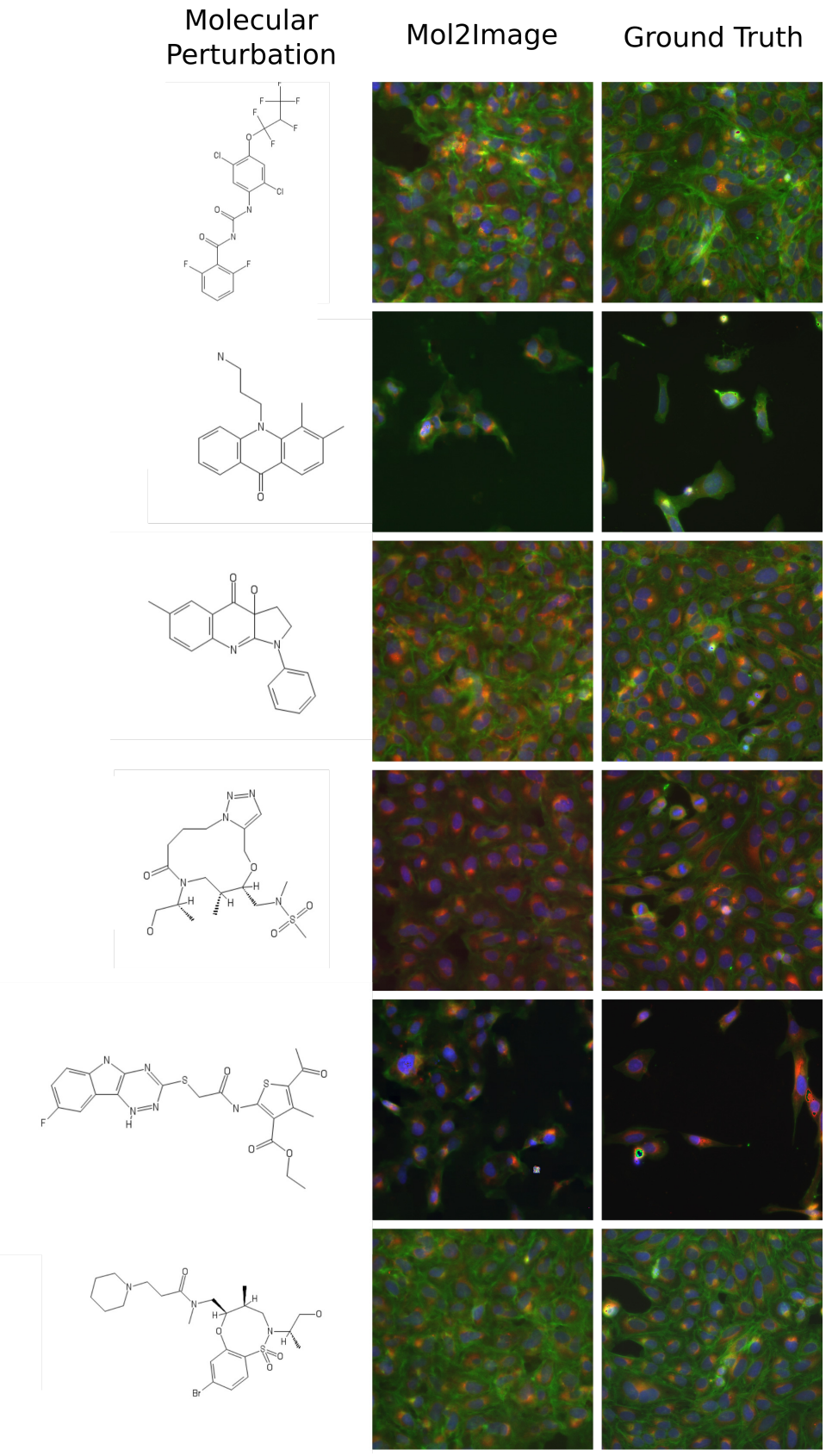

Figure 3: Examples of 512 × 512 cell images generated by our method (Mol2Image) in comparison to ground truth images for the same molecule. RGB channels represent three out of five channels from the full image.

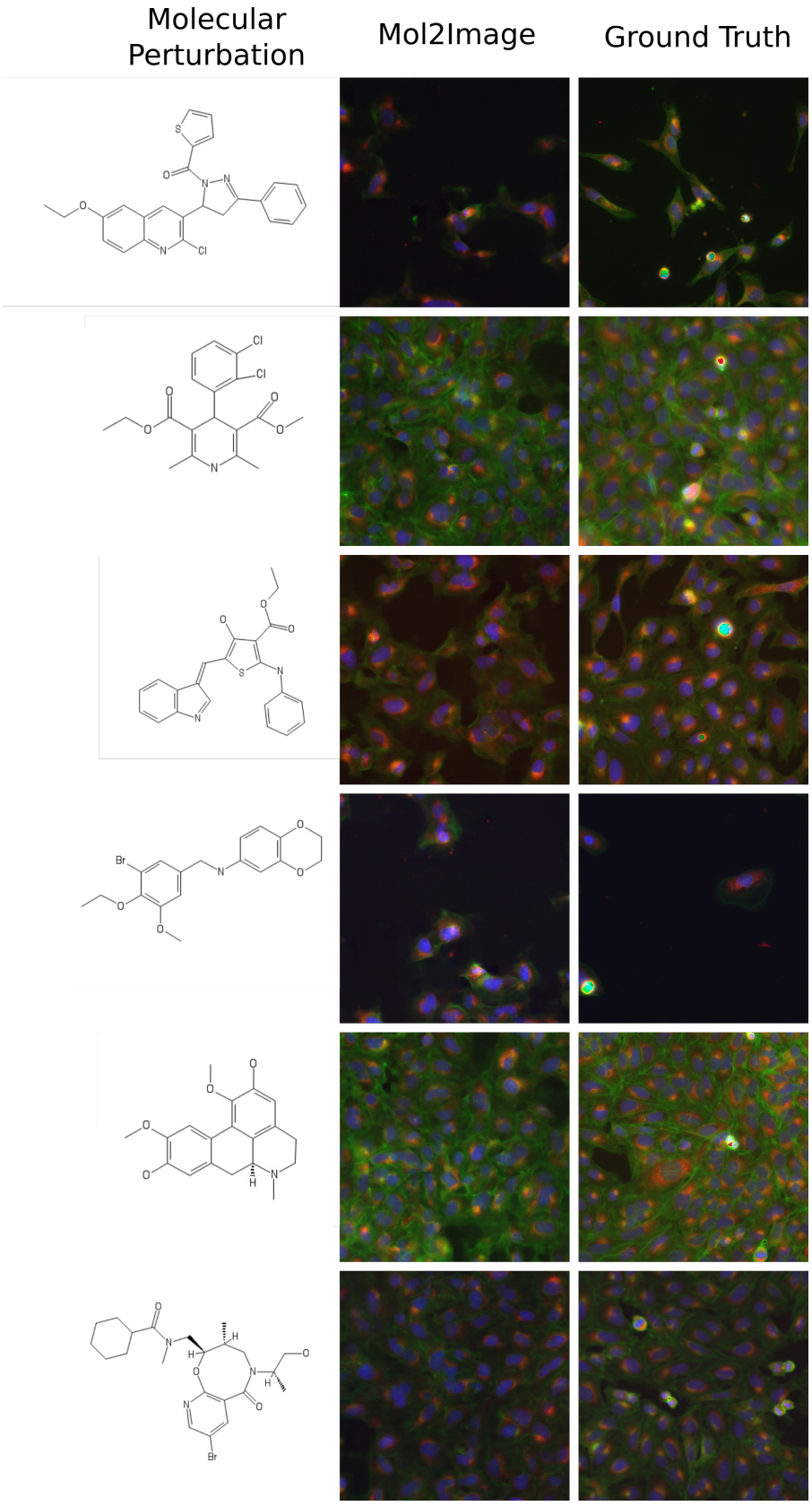

Figure 4: Examples of 512 $\times$ 512 cell images generated by our method (Mol2Image) in comparison to ground truth images for the same molecule. RGB channels represent three out of five channels from the full image.