A multi-view generative model for molecular representation improves prediction tasks

Jonathan Yin[*]
Yale University
New Haven, CT 06520
jonathan.yin@yale.edu

Hattie Chung[*†]
Broad Institute
Cambridge, MA 02142
hchung@broadinstitute.org

Aviv Regev[†§]
Klarman Cell Observatory, Broad Institute
Cambridge, MA 02142
aregev@broadinstitute.org

Abstract

Unsupervised generative models have been a popular approach to representing molecules. These models extract salient molecular features to create compact vectors that can be used for downstream prediction tasks. However, current generative models for molecules rely mostly on structural features and do not fully capture global biochemical features. Here, we propose a multi-view generative model that integrates low-level structural features with global chemical properties to create a more holistic molecular representation. In proof-of-concept analyses, compared to purely structural latent representations, multi-view latent representations improve model accuracy on various tasks when used as input to feed-forward prediction networks. For some tasks, simple models trained on multi-view representations perform comparably to more complex supervised methods. Multi-view representations are an attractive method to improve representations in an unsupervised manner, and could be useful for prediction tasks, particularly in contexts where data is limited.

1 Introduction

1.1 Deep learning models for molecular representation

Recent advances in machine learning have revolutionized molecular representations, specifically by identifying features that are important for function. Supervised deep learning models in particular have enjoyed large success for molecular property prediction [1, 2]. These models learn molecular features that are highly attuned to a specific function, but these learned features may not generalize well to other tasks. Furthermore, they typically require significant quantities of labeled training data as they are trained on raw, low-level representations. Thus, developing a more abstract, meaningful molecule-intrinsic representation could facilitate training and lead to more accurate prediction across a wide array of tasks.

In recent years, variational autoencoders (VAEs) have become popular for creating broadly-tuned unsupervised molecular representations [3, 4, 5, 6]. Since the features encoded in the learned representation are not function-specific, the latent spaces created by VAEs could be suitable to use as

[*]Equal contribution
[†]Co-corresponding authors
[§]Current address: Genentech, 1 DNA Way, South San Francisco, CA, USA

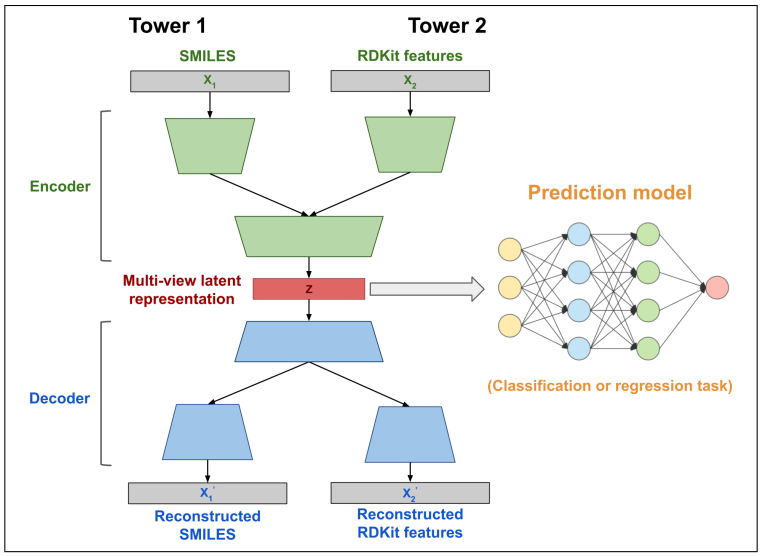

Figure 1: Architecture of multi-view two-tower VAE. For property prediction, molecules are first encoded into their latent molecular representation and then passed as input to a separate prediction model.

a general representation for downstream prediction tasks [4, 7]. In addition, VAEs can be used for *de novo* molecular design by traversing the latent space to maximize or minimize specific properties and then decoding the latent representations to obtain candidate molecules [4].

1.2 Multi-view representation fusion

Different views of the same data—e.g. RNA and protein, or audio and video, or image and text—often contain complementary information. Multi-view representation fusion, a form of multi-view representation learning, seeks to integrate data across multiple modalities into a single, more comprehensive representation [8, 9, 10]. This joint representation then facilitates the development of prediction models by enabling useful information to be readily extracted. One example application of fusion-based multi-view learning is a bimodal deep autoencoder that combines audio and video data into a shared representation [11].

Here, we leverage multi-view representation fusion to improve molecular representations in unsupervised generative models so that they are more useful for property predictions and *de novo* design.

2 Two-tower molecule VAE for multi-view molecular representation

We present a bimodal "two-tower" VAE model that simultaneously encodes low-level structural information and global-level molecular features that reflect emergent properties of a molecule (Figure 1). By integrating these two distinct but complementary views, we aim to create a richer shared representation that can better capture the functional properties of a molecule based on its structure.

Our two-tower models build on two published single-tower VAEs, a character VAE (CharVAE) [4] and a grammar VAE (GVAE) [6]. Both of these published models use a string-based structural representation known as the simplified molecular-input line-entry system (SMILES) [12] as training input. The first tower of our two-tower VAE mimics these single-tower VAEs to capture local structural features. The second tower encodes global chemical properties that are challenging to capture at a local level. To ensure that our second tower encoded properties applicable to various learning tasks, we used a set of 200 RDKit features [13] similar to the approach taken by a supervised model Chemprop [1]. These features include the number of aliphatic carbocycles or the number of radical electrons in a molecule.

Table 1: Descriptions and summary statistics of the MoleculeNet datasets used.

Dataset	Description	No. of tasks	Task type	No. of compounds	Metric
ESOL	Water solubility	1	Regression	1,128	RMSE
FreeSolv	Hydration free energy in water	1	Regression	642	RMSE
Lipophilicity	Octanol/water distribution coefficients	1	Regression	4,200	RMSE
HIV	Inhibition of HIV replication	1	Classification	41,127	ROC-AUC
BACE	Inhibition of human β-secretase 1	1	Classification	1,513	ROC-AUC
BBBP	Toxicity	1	Classification	2,039	ROC-AUC
Tox21	Toxicity	12	Classification	7,831	ROC-AUC
Clintox	Toxicity	2	Classification	1,478	ROC-AUC
SIDER	Side effects of drugs	27	Classification	1,427	ROC-AUC

To construct a latent representation of the combined modalities, inputs for each tower are first individually sent through several encoder layers before being passed simultaneously through a single encoder. This produces the shared latent representation that is used as input for property prediction (see 3.1 Experiments). Decoding is conducted similar to encoding; vectors in the latent space are first passed through a single decoder and then through two separate decoders corresponding to the structural input and RDKit features.

We trained our VAE model on 250,000 drug-like molecules from the ZINC molecule dataset [14], with 5,000 randomly selected molecules held out for testing (see 3.2. Experiments). Molecular features for the second tower were generated using DescriptaStorus [1, 2].

3 Experiments

3.1 Property prediction using multi-view latent representations

We used 9 publicly available datasets from MoleculeNet [15] to evaluate how well the multi-view latent space from our VAEs could predict molecular properties (Table 1). For each dataset, we mapped the input molecules to their latent vectors using the encoder portion of each VAE. We then fed these embeddings into simple feed-forward networks (FFNs) to predict the desired output properties. Latent spaces resulting in higher prediction accuracy were interpreted as better molecular representations since the same FFN was used for task prediction across all input representations.

Across all 9 data sets, the FFNs trained on the latent representations of two-tower models outperformed those trained on the latent representations of single-tower models. For some datasets (BBBP, ESOL), the basic FFNs trained on the unsupervised two-tower encodings even performed comparably to Chemprop [1, 2], a state-of-the-art supervised model (Figure 2). This indicates that the inclusion of a second tower improved the latent space's utility for property prediction.

3.2 Molecule reconstruction and validity

Next, we verified that the addition of the second tower did not negatively affect the ability to reconstruct molecules from their latent representations. We determined reconstruction accuracy and validity as previously documented [5]: each molecule was encoded and decoded 10 times, and the proportion of the 100 decoded molecules identical to the input molecule was reported. We computed validity by sampling 1,000 latent vectors from the prior distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ and decoding each of these vectors 100 times. Our two-tower model achieved higher reconstruction accuracy than the single-tower counterparts, and maintained prior validities comparable to single-tower models (Table 2).

[1] with hyperparameters matching those used in this study, primarily an increase in latent space size

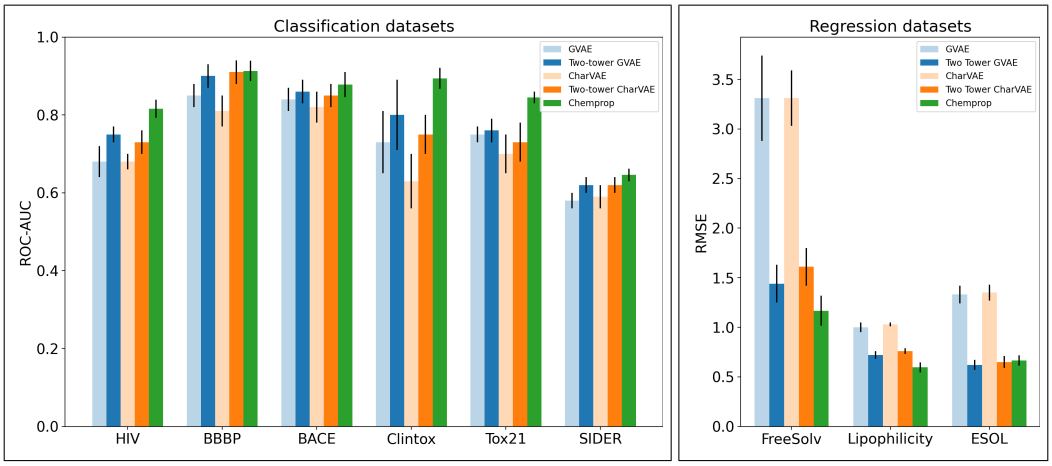

Figure 2: Prediction accuracies of feed-forward networks (FFNs) trained using various latent representations as input: structural representations (light blue/orange), multi-view representations (dark blue/orange). Prediction accuracy from Chemprop, a supervised model, is shown for reference (green bars). **Left:** classification datasets from MoleculeNet on x-axis, ROC-AUC on y-axis. **Right:** regression datasets from MoleculeNet on x-axis, root mean squared error (RMSE) shown on y-axis.

Table 2: Reconstruction accuracy and sample validity results.

Method	Reconstruction	Validity
CharVAE[1]	45.6%	0.1%
Published CharVAE	44.6%	0.7%
Two-tower CharVAE	69.3%	0.2%
GVAE[1]	52.7%	5.1%
Published GVAE	53.7%	7.2%
Two-tower GVAE	61.8%	4.9%

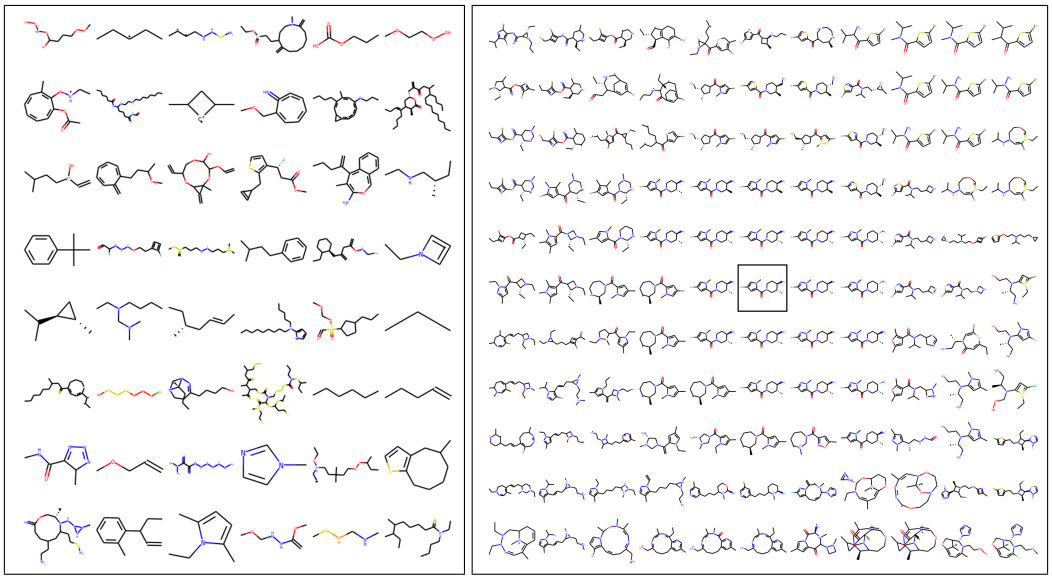

Figure 3: **Left:** random molecules sampled from prior distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ using the two-tower GVAE model. **Right:** visualizing the local neighborhood of the two-tower GVAE model, starting from the molecule in the center (outlined in black).

We also qualitatively examined the latent space of the two-tower GVAE for continuity in the neighborhood of a molecule (Figure 3). Following a previously published approach [6], we generated 2 random orthogonal unit vectors in latent space and moved in combinations of these directions to create a grid of latent vectors that were then decoded into molecules. Despite training the model on both structural and chemical features, we still observe smooth transitions between molecules in the grid, often by a single atom at a time.

4 Discussion

We present a multi-view generative model for molecular representation using a two-tower VAE. We show that the resulting latent representation improves utility in predicting multiple independent molecular properties compared to single-tower VAEs and may also offer a more interpretable latent space. In the future, we aim to explore the generative capabilities of our two-tower model using Bayesian optimization, and implement a two-tower approach built with graphical approaches to molecular representation.

Author Contributions

H.C. and J.Y. conceived and designed the study. J.Y. developed and implemented the analyses with supervision from H.C. A.R. provided project oversight. J.Y., H.C., and A.R. wrote the manuscript.

Corresponding authors

Correspondence to H.C. (hchung@broadinstitute.org) and A.R. (aregev@broadinstitute.org)

Acknowledgements

We would like to thank Michael Truell and John Ingraham for helpful discussions.

Competing Interests Statement

A.R. is a founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas Therapeutics, and until August 31, 2020 was an SAB member of Syros Pharmaceuticals, Neogene Therapeutics, Asimov and ThermoFisher Scientific. From August 1, 2020, A.R. is an employee of Genentech.

References

[1] Jonathan M Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M Donghia, Craig R MacNair, Shawn French, Lindsey A Carfrae, Zohar Bloom-Ackermann, Victoria M Tran, Anush Chiappino-Pepe, Ahmed H Badran, Ian W Andrews, Emma J Chory, George M Church, Eric D Brown, Tommi S Jaakkola, Regina Barzilay, and James J Collins. A deep learning approach to antibiotic discovery. *Cell*, 181(2):475–483, April 2020.

[2] Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, Angel Guzman-Perez, Timothy Hopper, Brian Kelley, Miriam Mathea, Andrew Palmer, Volker Settels, Tommi Jaakkola, Klavs Jensen, and Regina Barzilay. Analyzing learned molecular representations for property prediction. *J. Chem. Inf. Model.*, 59(8):3370–3388, August 2019.

[3] Daniel C Elton, Zois Boukouvalas, Mark D Fuge, and Peter W Chung. Deep learning for molecular design—a review of the state of the art. *Molecular Systems Design & Engineering*, 4(4):828–849, 2019.

[4] Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.*, 4(2):268–276, February 2018.

[5] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation. February 2018.

[6] Matt J Kusner, Brooks Paige, and José Miguel Hernández-Lobato. Grammar variational autoencoder. March 2017.

[7] Eric M Jones, Rishi Jajoo, Daniel Cancilla, Nathan B Lubock, Jeffrey Wang, Megan Satyadi, Rocky Cheung, Claire de March, Joshua S Bloom, Hiroaki Matsunami, and Sriram Kosuri. A scalable, multiplexed assay for decoding GPCR-ligand interactions with RNA sequencing. *Cell Syst.*, 8(3):254–260.e6, March 2019.

[8] Samuel G Finlayson, Matthew B A McDermott, Alex V Pickering, Scott L Lipnick, and Isaac S Kohane. Cross-modal representation alignment of molecular structure and perturbation-induced transcriptional profiles. November 2019.

[9] Yingming Li, Ming Yang, and Zhongfei Zhang. Multi-view representation learning: A survey from shallow methods to deep methods. October 2016.

[10] Karren D Yang and Caroline Uhler. Multi-Domain translation by learning uncoupled autoencoders. February 2019.

[11] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. January 2011.

[12] David Weininger. SMILES, a chemical language and information system. 1. introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.*, 28(1):31–36, February 1988.

[13] Greg Landrum and Others. RDKit: Open-source cheminformatics. 2006.

[14] John J Irwin and Brian K Shoichet. ZINC- a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.*, 45(1):177–182, 2005.

[15] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.*, 9(2):513–530, January 2018.