Generative Network Complex (GNC) for Drug Discovery

Christopher Grow[1], Kaifu Gao[1], Duc Duy Nguyen[1], and Guo-Wei Wei[1,2,3]

[1] Department of Mathematics, Michigan State University, MI 48824, USA

[2] Department of Electrical and Computer Engineering Michigan State University, MI 48824, USA

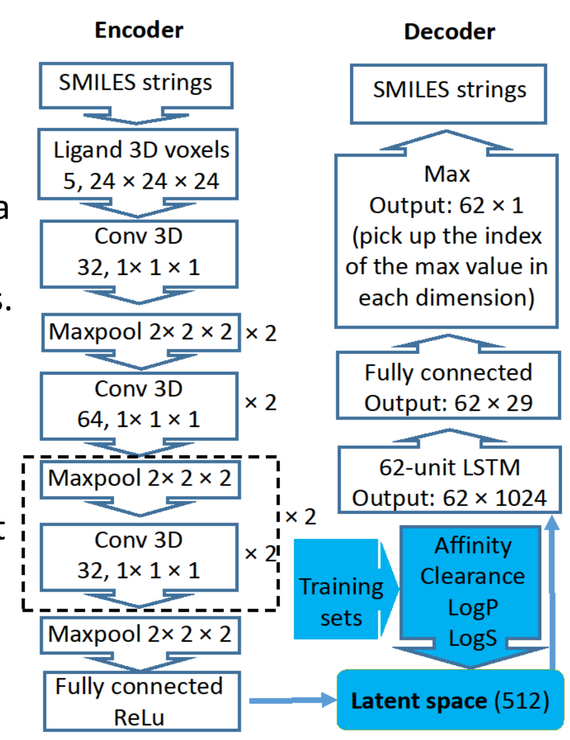
[3] Department of Biochemistry and Molecular Biology, Michigan State University, MI 48824, USA

Abstract

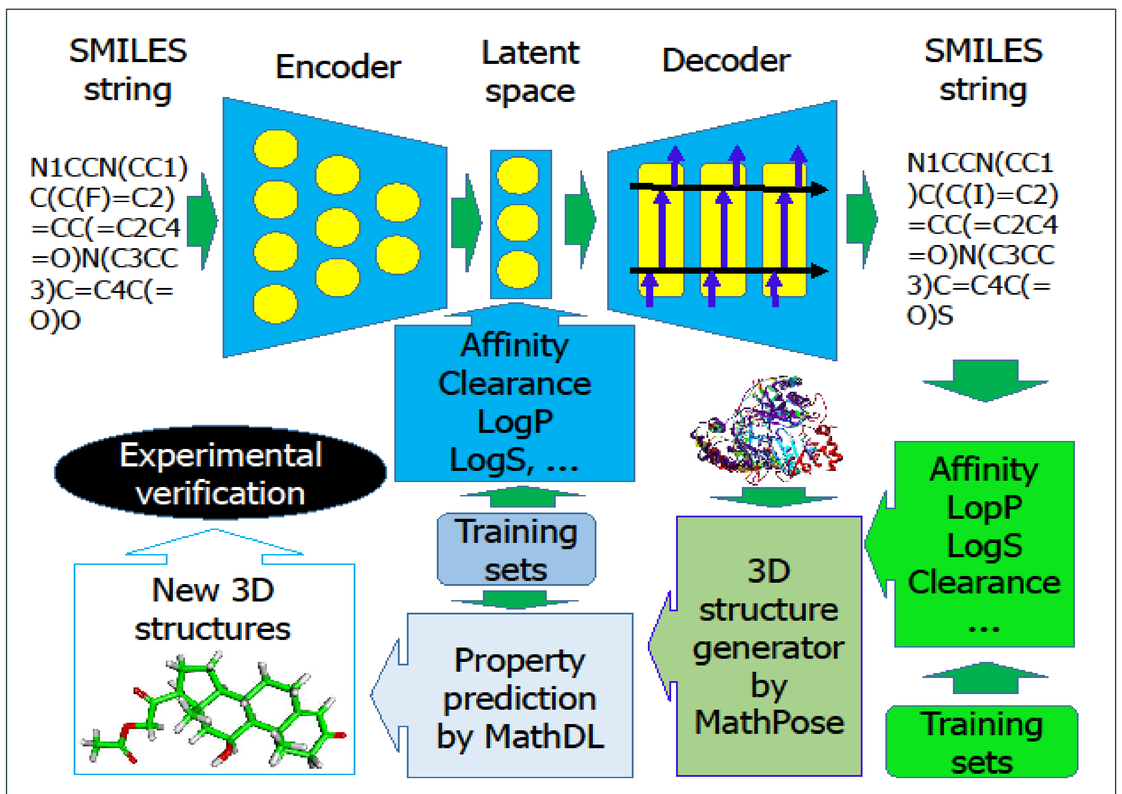
It remains a challenging task to generate a vast variety of novel compounds with desirable pharmacological properties. In this work, a generative network complex (GNC) is proposed as a new platform for designing novel compounds, predicting their physical and chemical properties, and selecting potential drug candidates that fulfill various druggable criteria. We combine a SMILES variational autoencoder with deep neural networks, a target-specific three-dimensional (3D) pose generator, and mathematical deep learning networks to generate new compounds, predict their drug properties, construct 3D poses associated with target proteins, and finally reevaluate druggability.

Autoencoder

For each SMILES string, 3D features are generated using RDKit and fed through a 3D-convolutional autoencoder (AE). The latent space is regulated by a pre-trained drug-property array to optimize druggable properties. The regularized latent information is fed into a shape captioning network, which consists of a 3D-convolutional network followed by a long short term memory neural network to generate SMILES string representations of the new molecules.

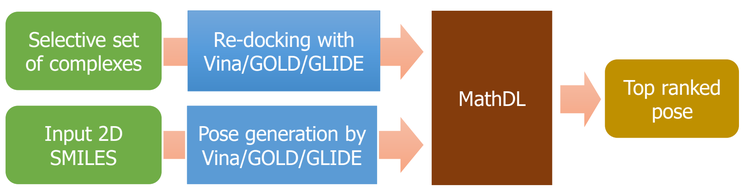


Generative Network Complex (GNC)



In the proposed GNC, the first component is a variational autoencoder. The autoencoder network will take a given SMILES string as an input to generate a novel one. The newly generated SMILES strings will be fed into the second component of our GNC, a 2D fingerprint-based deep neural network (2DFP-DNN), so that only ones with desired druggable properties are kept. The next component is the MathPose model which is used to predict the 3D structure information of the compounds selected by 2DFP-DNN. The bioactivities of those compounds are again estimated by the structure-based deep learning model named MathDL. The druggable properties predicted by this last component of our GNC are used as an indicator to select the promising drug candidates.

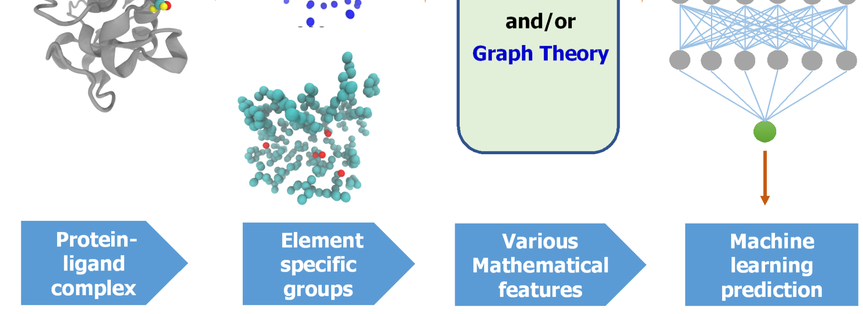
MathPose



In our recent work, we have successfully designed an AGL-Score model to achieve the best performances in docking power metrics which validate the scoring function's ability to identify the ``native pose'' from the computer-generated poses. Specifically, on the CASF-2007 benchmark, our AGL-Score achieves 84% accuracy on the docking power assessment.

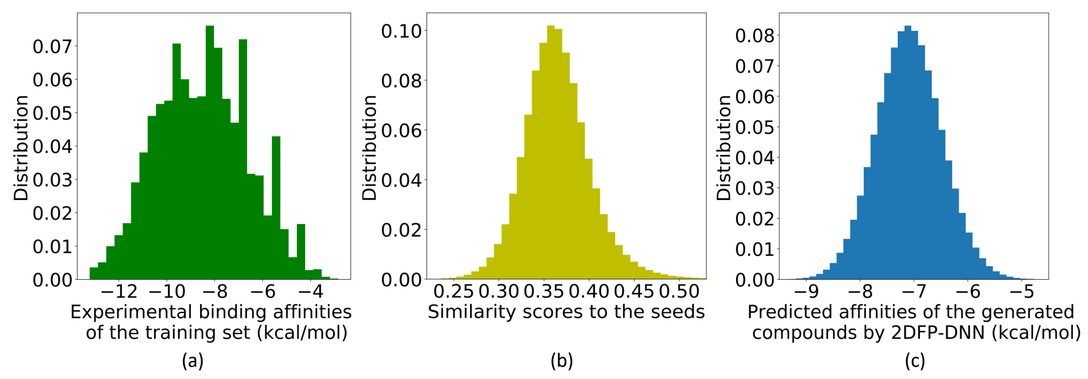
MathDL

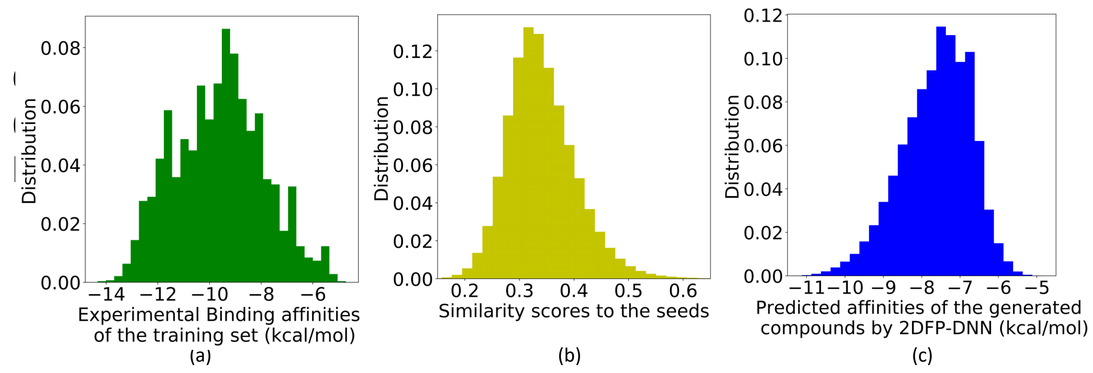
Our MathDL is constructed by the integration of mathematical representation features and deep learning networks to generate a powerful binding affinity predictor. Specifically, the MathDL is the blend of intensively validated models based on algebraic topology, differential geometry, and graph theory.



Results

Of the 2.8 million compounds generated for the BACE target, 99 had a predicted binding affinity smaller than -9.56 kcal/mol based on our 2DFP-DNN. Generated compounds had an average similarity score of 0.34 to the seed molecule.





The top 1,050 generated compounds for the Cathespin S target were reevaluated using MathDL. Their predicted binding affinity ranged from -7.01 kcal/mol to -11.68 kcal/mol, with an average value of -9.27 kcal/mol. The top 4 predicted values are -11.68 kcal/mol, -11.61 kcal/mol, -11.54 kcal/mol, and -11.54 kcal/mol.

Acknowledgement:

This work was supported in part by NSF Grants DMS-1721024, DMS-1761320, and IIS1900473 and NIH grants GM126189 and GM129004. DDN and GWW are also funded by Bristol-Myers Squibb and Pfizer.

Contact:

growchri@msu.edu
gaokaifu@msu.edu
ddnguyen@msu.edu
weig@msu.edu

References

1. Duc Duy Nguyen, Zixuan Cang, Kedi Wu, Menglun Wang, Yin Cao, and Guo-Wei Wei. Mathematical deep learning for pose and binding affinity prediction and ranking in d3r grand challenges. *Journal of computer-aided molecular design*, 33(1):71–82, 2019.
2. Miha Skalic, José Jiménez, Davide Sabbadin, and Gianni De Fabritiis. Shape-based generative modeling for de novo drug design. *Journal of chemical information and modeling*, 59(3):1205–1214, 2019.