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.

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.

Acknowledgement:

This work was supported in part by NSF Grants DMS-1721024, DMS-1761320, and IS1900473 and NIH grants GM126189 and GM129004. DDN and GWW are also funded by Bristol-Myers Squibb and Pfizer.

Contact:

growchi@msu.edu
gaoaifu@msu.edu
ddnguyen@msu.edu
weig@msu.edu

References