# TARGET-AWARE VARIATIONAL AUTO-ENCODERS FOR LIGAND GENERATION WITH MULTIMODAL PROTEIN REPRESENTATION LEARNING

Nhat Khang Ngo, Truong Son Hy

Available at

https://iopscience.iop.org/article/10.1088/2632-2153/ad3ee4

# INTRODUCTION TO DRUG DISCOVERY

Drug discovery is a critical, complex, and costly process that involves identifying molecular candidates that can effectively interact with specific biological targets. The process often spans several years and entails multi-stage experimentation and validation, with costs reaching into billions of dollars. The initial phase involves the discovery of novel drug-like compounds with high binding affinities to specific protein targets, leveraging techniques like virtual screenings and molecular dynamics simulations.

## CHALLENGES IN DRUG DISCOVERY

The primary challenges in the early stages of drug discovery include the vast chemical space of potential molecules, estimated at around $10^{33}$ chemically valid structures, and the computational expense of drug-target affinity (DTA) predictions. Traditional methods involve computationally intensive simulations that are impractical on a large scale, highlighting a crucial need for efficient computational techniques to predict and validate drug-likeness and binding affinities quickly.

# DEEP GENERATIVE MODELS FOR DRUG DESIGN

Deep generative models have been proposed to reduce the workload for wet-lab experiments by automating the generation and optimization of molecular properties. However, these models are often slow when enhancing binding affinity or other computationally expensive properties due to the need for reinforcement learning frameworks.

# PROTEIN REPRESENTATION LEARNING

Proteins can be represented as sequences of amino acids, 2D graphs at residue level, or 3D point clouds at atom level. Advanced methods leverage language models, graph neural networks (GNNs), and convolutional neural networks (CNNs) to learn from these representations.

# CONTRIBUTIONS

- Developed **TargetVAE**, a conditional VAE model to generate drug-like molecules with high binding affinity to given protein structures.
- Adapted techniques from computer vision to transfer weights from an unconditional to a conditional VAE, enhancing molecule generation diversity.
- Introduced **Protein Multimodal Network (PMXN)**, which unifies sequence and structural modalities of proteins for improved prediction of binding affinities.

# BACKGROUND: ROTATIONAL INVARIANT FEATURES

Geometric features of protein structures can be represented as tuples of scalar and vector features, which are respectively invariant and equivariant to geometric transformations in Euclidean space. The Geometric Vector Perceptron (GVP) is used to transform input features, ensuring the desired properties of invariance and equivariance.

# VARIATIONAL AUTO-ENCODERS (VAES)

VAEs consist of a generative model and an inference model, using a probabilistic decoder and a prior to define a joint distribution between latent variables and data. Conditional VAEs integrate auxiliary covariates to enhance generative modeling, trained to maximize the conditional evidence lower bound (ELBO).

# PROTEIN MULTIMODAL NETWORK (PMN): OVERVIEW

**Objective:** Integrate diverse protein representations into a unified framework capable of capturing both local and global structural features, utilizing advanced graph and sequence modeling techniques.

## LONG-RANGE MODELING ON 3D STRUCTURES

**Key Components:**

- Local Encoder: Uses GVPs to update node features based on their scalar and vector attributes.
- GVP Module: Transforms features to maintain rotational invariance and equivariance.
- Global Transformer Encoder: Captures long-range interactions using self-attention mechanisms.

**Mathematical Formulation:**

$$m_{ij} = \text{GVPs}\left(\text{concat}(h_v^{(i)}, h_e^{(j\to i)})\right),$$

$$h_v^{(i)} = \text{LayerNorm}\left(h_v^{(i)} + \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} m_{ij}\right).$$

# LANGUAGE MODELING ON PROTEIN SEQUENCE

**Approach:**

- Utilize Transformer-based models to process sequences where each residue is treated as a token.
- Embedding layers transform residue tokens into dense vector representations, incorporating positional encodings.

**Transformations:**

$$Q_\ell = Z_{\ell-1} W_\ell^Q, \quad K_\ell = Z_{\ell-1} W_\ell^K, \quad V_\ell = Z_{\ell-1} W_\ell^V,$$

$$H_\ell = \text{MultiheadAttention}(Q_\ell, K_\ell, V_\ell),$$

$$Z_\ell = \text{LayerNorm}(Z_{\ell-1} + \text{FFN}(H_\ell)).$$

# UNIFIED REPRESENTATION AND PROTEIN EMBEDDING

**Final Integration:**

- Aggregated embeddings from both the graph and sequence models are concatenated.
- The concatenated vector is passed through dense layers to produce the final representation of the protein.

**Representation Output:**

$$p = W_2\text{ReLU}(W_1(\text{concat}(p_g, p_s)) + b_1) + b_2$$

# BINDING AFFINITY PREDICTION

**Datasets:**

- **DAVIS** - 442 proteins and 68 ligands, 30,056 pairs. $K_D$ constants.
- **KIBA** - 229 proteins and 2,111 ligands, 118,254 pairs. KIBA scores.

**Data Enhancement:** 3D protein structures generated by AlphaFold included for a comprehensive dataset.

**Evaluation Metrics:** Mean Squared Error (MSE), Concordance Index (CI), and $r_m^2$ Index.

**Result:** Superior performance on DAVIS and competitive on KIBA against benchmarks KronRLS, SimBoost, and GraphDTA.

## TARGET-AWARE DRUG DESIGN

**Dataset for Drug Design:** Utilize KIBA and ZINC250K datasets for training conditional molecule generation models.

**Techniques:**

- Models trained to generate SELFIES representations of molecules.
- Evaluate using Fréchet ChemNet Distance (FCD) and other chemical properties (QED, pLogP, SA).

**Advanced Evaluation:** Docking measurements using AutoDock-GPU for binding affinity predictions.

**Generative Model Quality:**

- Low FCD scores indicating close approximation to real molecular distributions.
- Generated molecules showcase desirable drug-like properties.

# BINDING EXPERIMENTS WITH UNSEEN TARGETS

**Experiment Details:**

- Testing on unseen targets shows our model's ability to generalize and effectively predict binding affinities.
- Comparison with state-of-the-art models demonstrates superior or comparable performance.

**Results Discussion:**

- Highlights advantages of our method in balancing binding affinities and drug-like properties.
- Provides insight into the potential for zero-shot learning capabilities in drug discovery.