Target-Aware Variational Auto-Encoders for Ligand Generation with Multi-Modal Protein Modeling

Nhat Khang Ngo [1,*] and Truong Son Hy [2,*]

FPT Software AI Center, Hanoi, Vietnam [1] Indiana State University, Terre Haute, USA [2]

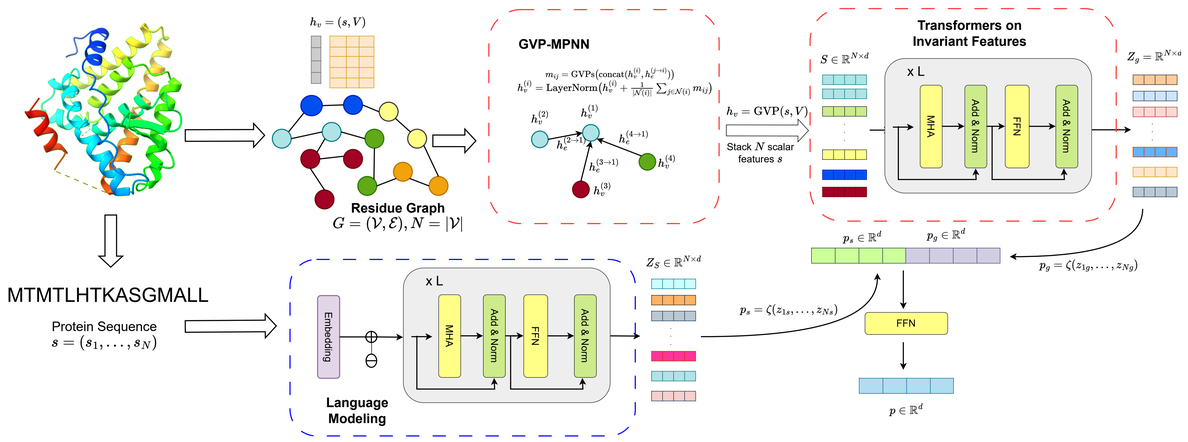
Emails: khangnn4@fpt.com TruongSon.Hy@indstate.edu

*: Equal contribution

Motivation

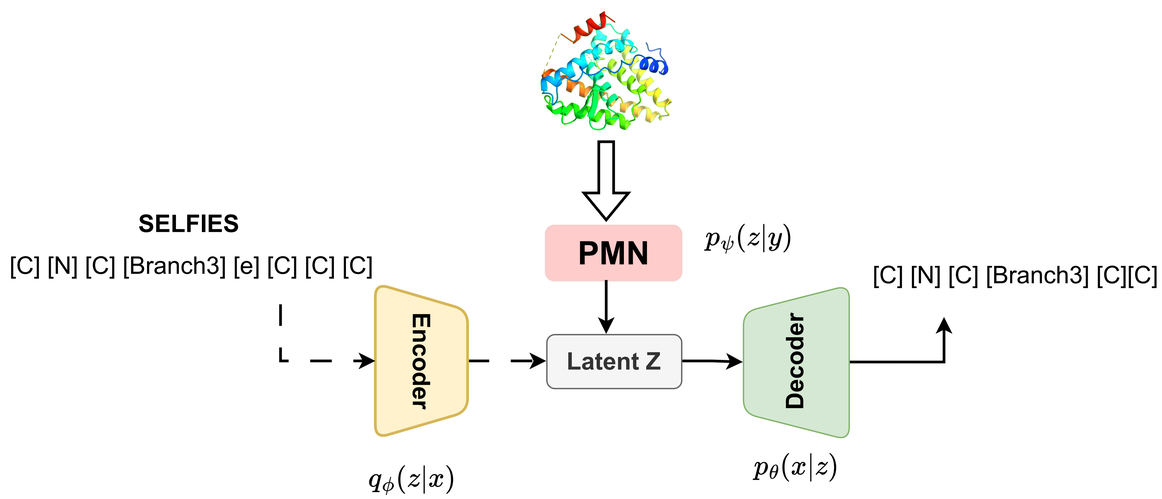
Without specific knowledge of binding sites, designing drug compounds that can potentially bind to a target protein is challenging. In this work, we propose a target-aware VAE that generates drug-like ligands conditioned on arbitrary proteins whose structures are learned from a multi-modal protein network.

Multi-Modal Protein Network



Multi-modal learning entails combining representations of different protein data types, encompassing sequences (language modeling), graphs, and 3D information (derived from residue geometric graphs), to establish a unified representation of proteins. Within each modality, transformer-based architectures are employed to effectively capture the global interactions among residues, while for geometric graphs, message-passing augmented with geometric vector perceptions is utilized to effectively extract their local interactions in 3D spaces.

Target-Aware VAE



We train a conditional variational auto-encoder with a learnable prior that extracts the multi-modal representations of the target proteins p. To leverage the diversity of chemical space, our model is optimized with loss as:

$$\log p_{\theta,\psi}(x|p) \geq O_{\text{for}} \triangleq \mathbb{E}_{q_\phi}[\log p_{\theta,\psi}(x|z)] - \text{KL}[q_\phi(z|p)\|p_\psi(z|p)]. \quad (1)$$

Here, ϕ, θ, ψ denote the encoder, decoder, and the prior network, respectively. Optimizing the networks with O_{for} allows us to reuse weights of decoders of a pre-trained unconditional VAE, which are trained on a large set of chemical compounds. In the sampling phase, we sample a latent vector $z \sim \mathcal{N}(\mu_\psi(p), \sigma_\psi(p))$, and pass it to the decoder θ for generating the SELFIES representations of the ligands.

Experiments

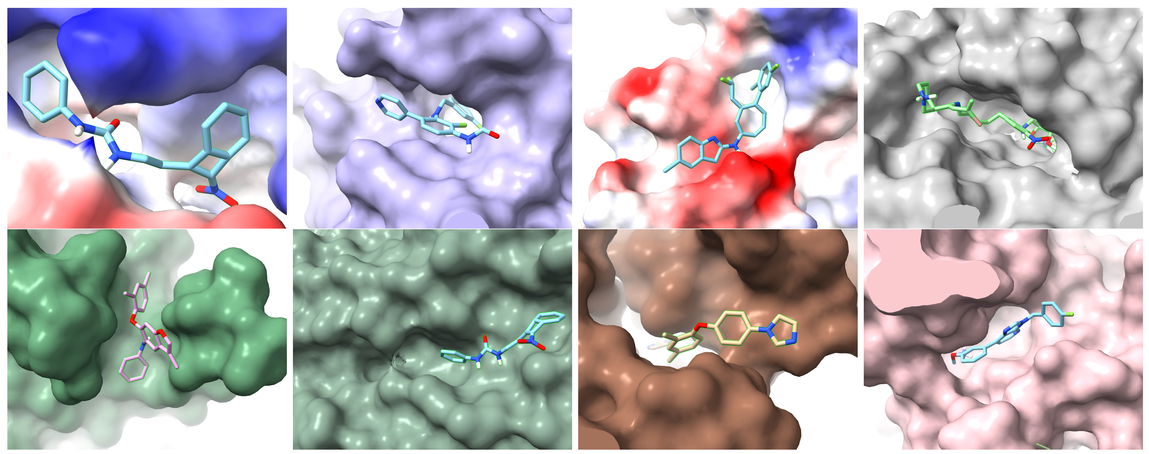
Dataset We train our TargetVAE on the PDBBind 2020 dataset. For docking simulation, we adopt AutoDock Vina to compute binding affinity in kcal/mol. We test our approach with nine target proteins, including G-protein coupling receptors (GPCRs) and kinases from DUD-E and the SARS-Conv-2 main protease. Notably, these targets are unseen to the model during the training stage.

Target	Top 1			Top 10			Top 20		
	BA ↓	SA ↓	QED ↑	BA ↓	SA ↓	QED ↑	BA ↓	SA ↓	QED ↑
1iep	-9.946	7.609	0.322	-9.242	4.412	0.413	-8.856	4.227	0.411
2rgp	-11.936	3.391	0.428	-10.293	4.201	0.520	-9.717	4.151	0.482
3eml	-25.939	7.268	0.584	-11.590	4.346	0.493	-10.294	4.446	0.476
3ny8	-11.257	5.99	0.807	-10.280	3.980	0.369	-9.870	4.193	0.433
4rlu	-11.250	2.979	0.479	-10.010	4.536	0.619	-9.495	4.759	0.564
4unn	-10.752	4.567	0.161	-9.860	4.192	0.415	-9.423	4.270	0.418
5mo4	-11.812	6.330	0.432	-10.325	5.041	0.325	-9.627	4.865	0.443
7l11	-11.220	7.912	0.136	-9.163	5.396	0.394	-8.567	5.073	0.417

Table 1: Quantitative results of top $k = 1, 10, 20$ generated molecules, which are ranked based on binding affinity (in kcal/mol). The scores are averaged over k ligands.

The experimental results demonstrate that TargetVAE can generate molecules with high binding affinity while maintaining low SA scores between 4.0 and 8.0. However, our approach is less effective in generating compounds with low drug-likeness (QED). We hypothesize that this limitation stems from the challenge of generating ligands that meet all objectives simultaneously, as our model prioritizes binding affinity. This suggests a potential trade-off between these metrics.

Visualization



Here, we visualize some of the poses of the ligands generated by our model. We use OpenBabel to optimize the ligands' three-dimensional poses and use Autodock Vina for docking simulations. The figures show that the generated ligands are quite stable in 3D space and bind tightly to the protein surfaces.

Software

Thank you for your attention!