Designing TrmD inhibitor

Lidia Stadnik[1]

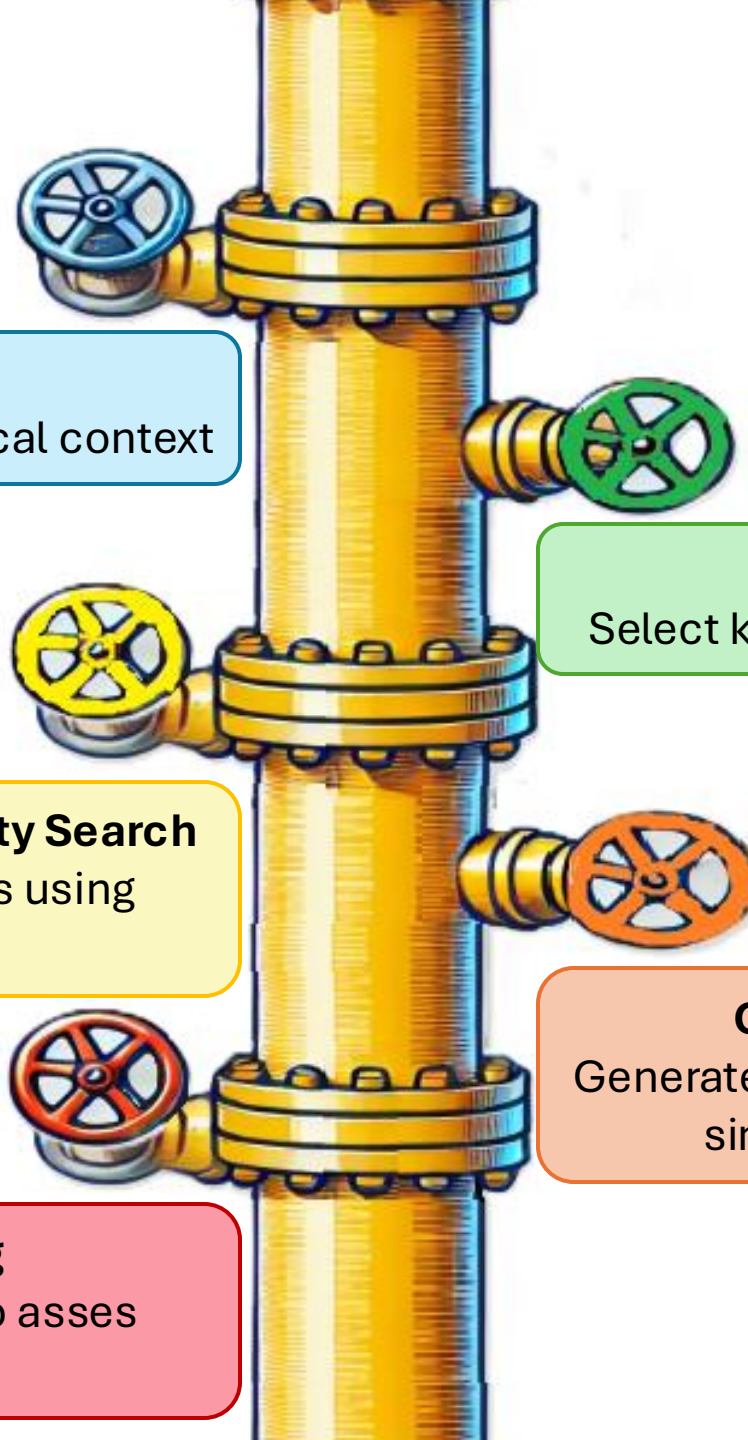
Krzysztof Łukasz[1]

Michał Osiński[1]

Conducted under the supervision of:
Dr hab. Joanna I. Sułkowska[2]
mgr Maciej Sikora[2]

[1.] Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, 02-096, Warsaw, Poland
[2.] Centre of New Technologies, University of Warsaw ul. Stefana Banacha 2c 02-097 Warsaw, Poland

Pipeline



Target selection
Select TrmD based on biological context

Ligand Selection
Select known ligands for chosen TrmD

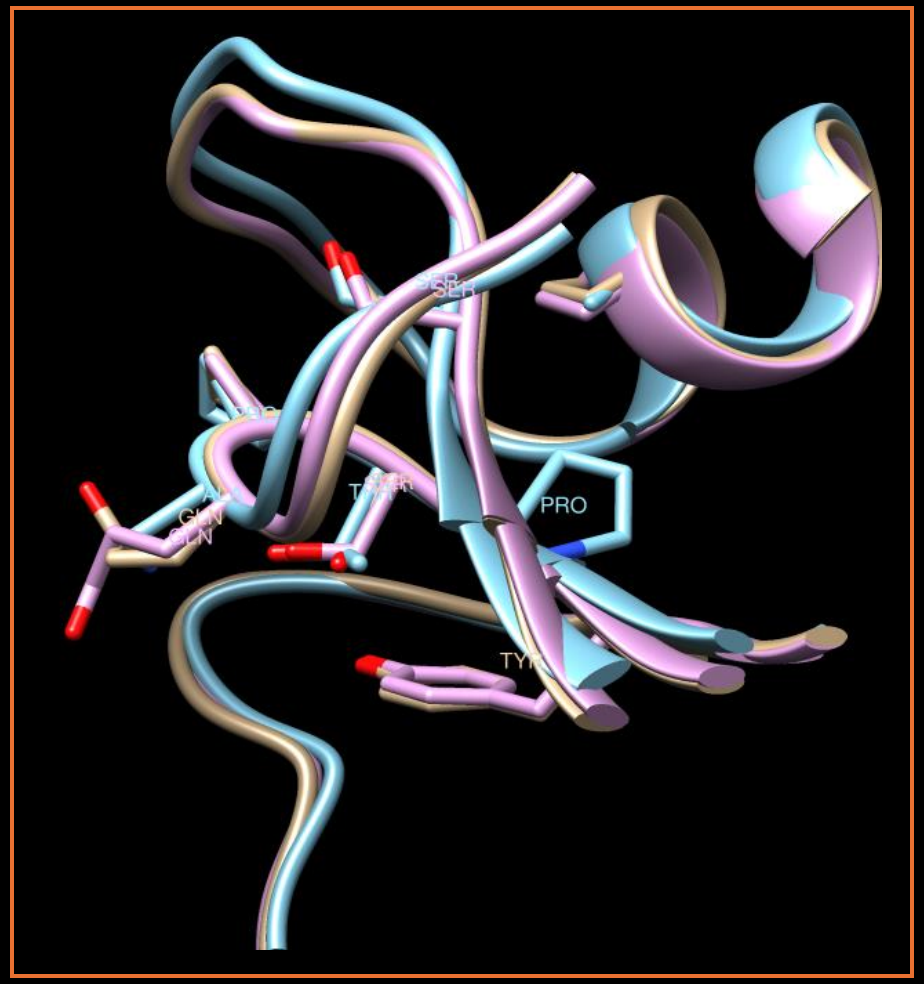
Compound Filtering and Similarity Search
Search for new potential ligands using SMILES similarity

Compound Generation
Generate novel compounds structurally similar to identified ligands

Molecular Docking
Perform in silco docking to asses binding affinity

Target selection

1. **Selected TrmD structures from the PDB database of bacterial species that are pathogenic to humans and antibiotic resistant:**
    - *Mycobacterium abscessus* (PDB Id: 6QRB) blue
    - *Acinetobacter baumannii* (PDB Id: 7MYS) pink
    - *Pseudomonas aeruginosa* (PDB Id: 5WYQ) brown
2. **Structural Superimposition**
    - Validated similarity of TrmD chain A active site across selected species in Chimera.



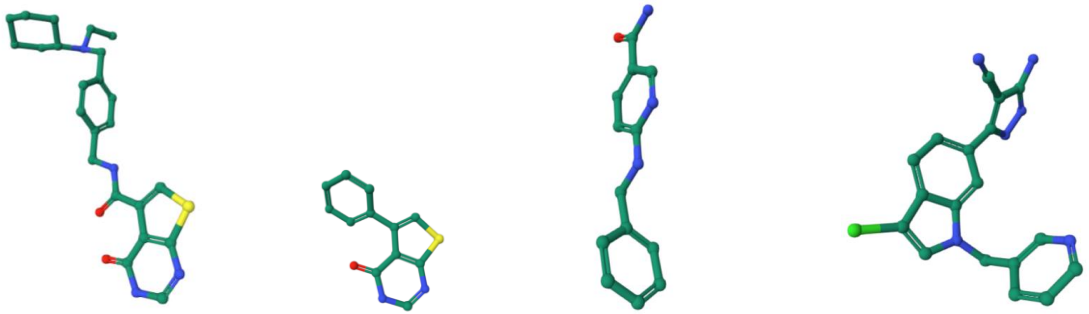
Pairwise RMSD Values

	A. baumannii	*M. abscessus*	*P. aeruginosa*
A. baumannii	0.0 Å	**0.7 Å**	**0.3 Å**
M. abscessus	**0.7 Å**	0.0 Å	**0.8 Å**
P. aeruginosa	**0.3 Å**	**0.8 Å**	0.0 Å
Overall RMSD	**0.6 Å**		

Ligand selection

Ligands crystallised with TrmD in PDB for selected species

- 59 ligands



Publications
(OPSIN parser)

- Wilkinson, Andrew J., et al. "Evaluating the druggability of TrmD, a potential antibacterial target, through design and microbiological profiling of a series of potent TrmD inhibitors." Bioorganic & Medicinal Chemistry Letters 90 (2023): 129331. [1]
    - **7 liganads**
- Zhong, Wenhe, et al. "Thienopyrimidinone derivatives that inhibit bacterial tRNA (guanine37-N 1)-methyltransferase (TrmD) by restructuring the active site with a tyrosine-flipping mechanism." Journal of medicinal chemistry 62.17 (2019): 7788-7805. [2]
    - **4 ligands**
- Vlasov, S. V., et al. "Synthesis, docking study and antimicrobial activity evaluation of pyridyl amides of thieno [2, 3-d] pyrimidine-4-carboxylic acid." (2023). [3]
    - **5 ligands**

Ligand Similarity-Based Exploration

Filter the PubChem Database (Initial: 2,512,731 compounds):

- **Lipinski's Rules**: Drug-likeness criteria.
- **Charge**[4]: Net charge between -2 and +2.
- **Polar Surface Area (TPSA)**[4]: ≤ 140 $Å^2$.
- **Result**: 1,187,012 compounds.

Search for Similar Compounds:

- Use the LINGO similarity metric[5] to identify compounds similar to known ligand SMILES strings from the filtered dataset.

cid	isosmiles	sim	target
11652085	COC(=O)C@HN	0.79	68910199
11701139	C1=CC=C(C=C1)COC(=O)CNN	0.75	68910199
20713065	C1=CC=C(C=C1)COC(=O)CNN=C(N)N	0.73	68910199
84733164	C1=CC=C(C=C1)COC(=O)CCNN	0.72	68910199
101383098	C1=CC=C(C=C1)CO13C[13CH2][15NH2]	0.71	68910199
11701101	C1=CC=C(C=C1)COC(=O)C[15NH2]	0.71	68910199
10725758	[2H]C([2H])(C(=O)OCC1=CC=CC=C1)N	0.71	68910199
10797159	C1=CC=C(C=C1)CO13C[13CH2]N	0.71	68910199
927308	C1=CC(=CC(=C1)NCC2=CC=CS2)C(=O)O	0.71	2819674
409140	C1=CC=C(C=C1)COC(=O)CN	0.71	68910199

[4] Kralj, Sebastjan, Marko Jukič, and Urban Bren. "Molecular filters in medicinal chemistry." Encyclopedia 3.2 (2023): 501-511.

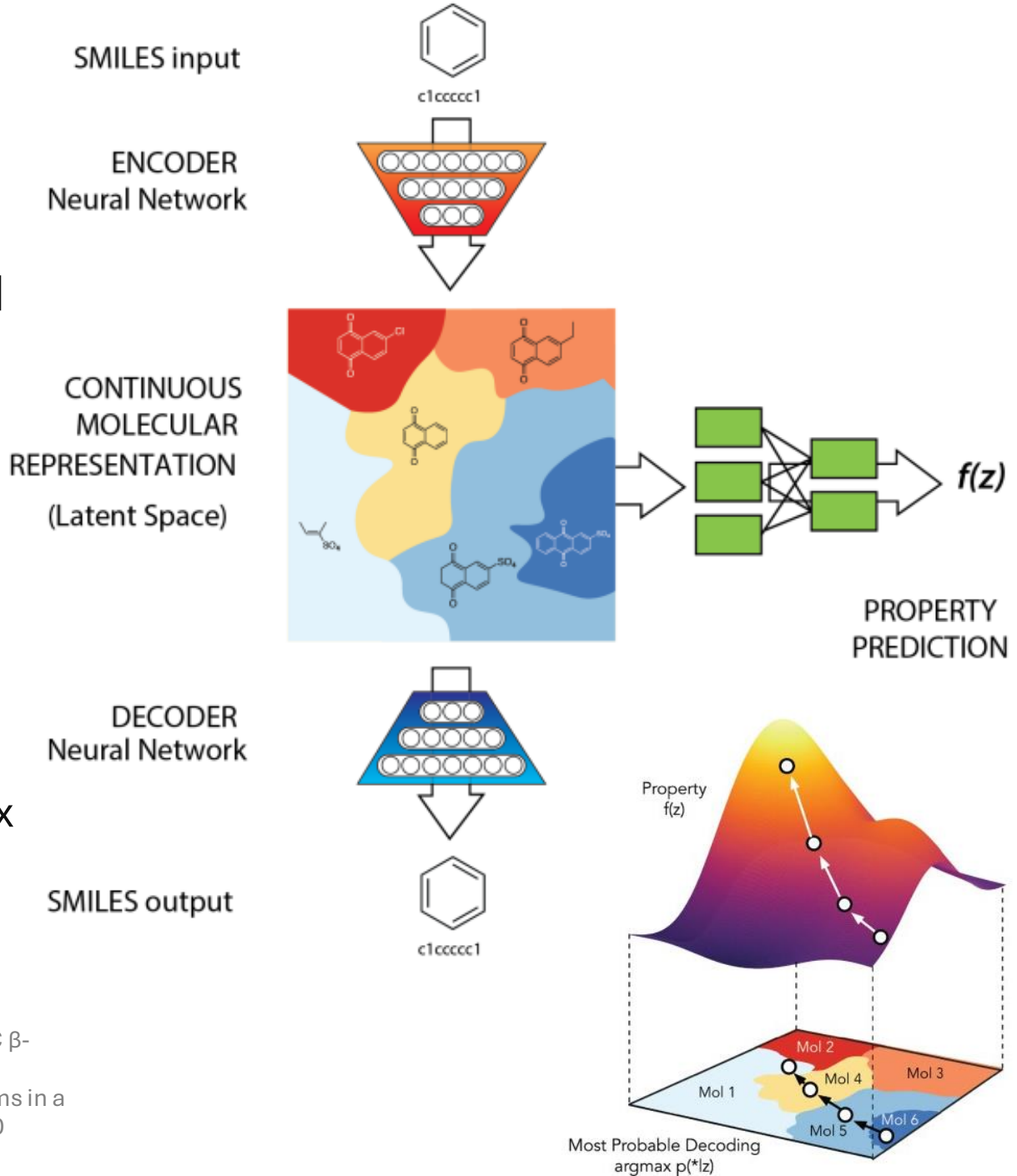
[5] Öztürk, H., Ozkirimli, E., & Özgür, A. (2016). A comparative study of SMILES-based compound similarity functions for drug-target interaction prediction. BMC bioinformatics, 17, 1-11.

Compound Generation

New chemical structures inspired by known ligands and similar compounds are generated using a **Variational Autoencoder (VAE)**[6,7] in a matrix-based format from **Keras' API**[8].

Encoding Molecules:

1. **Connection Matrix**:
   - Triangular matrix where numbers represent bond types (e.g., 1 = single, 2 = double).
2. **Feature Vector**:
   - Atom-specific information, such as the index in an atom dictionary.



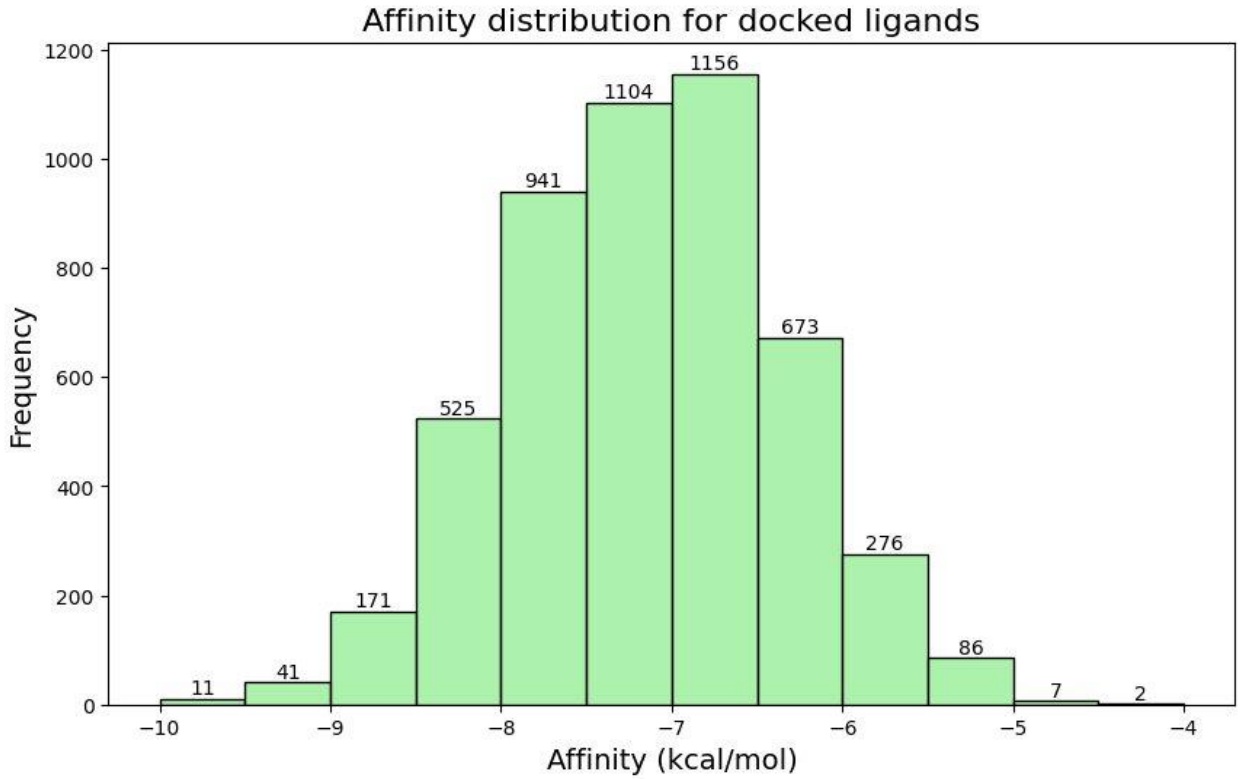
[6]. **S. Hafeez et al.** Designing of fragment-based inhibitors with improved activity against *E. coli* AmpC β-lactamase compared to conventional antibiotics. *Saudi J. Biol. Sci.*, 31, Elsevier BV, 2024.

[7]. **R. W. Grosse-Kunstleve et al.** The Computational Crystallography Toolbox: crystallographic algorithms in a reusable software framework. *J. Appl. Cryst.*, 35, International Union of Crystallography (IUCr), 200

[8]. **https://keras.io/examples/generative/molecule_generation/**

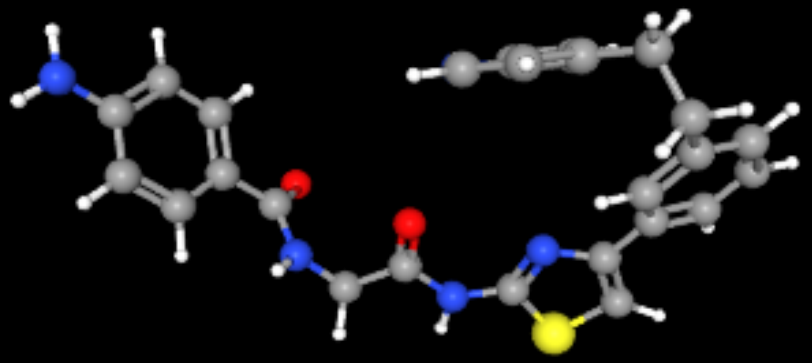
Molecular docking

- **Preparation**: Meeko
  - Ligand from a Smiles String
  - Receptor preparation from PDB
- **Software:** AutoDock Vina
  - Very fast and robust
  - Open-source



	PubChem ID	Affinity (kcal/mol)
0	155792348	-9.981
1	8171885	-9.885
2	54844255	-9.872
3	8842407	-9.756
4	155792345	-9.641
5	54811489	-9.637
6	54844164	-9.635
7	45861272	-9.592
8	54841777	-9.54
9	56295120	-9.519

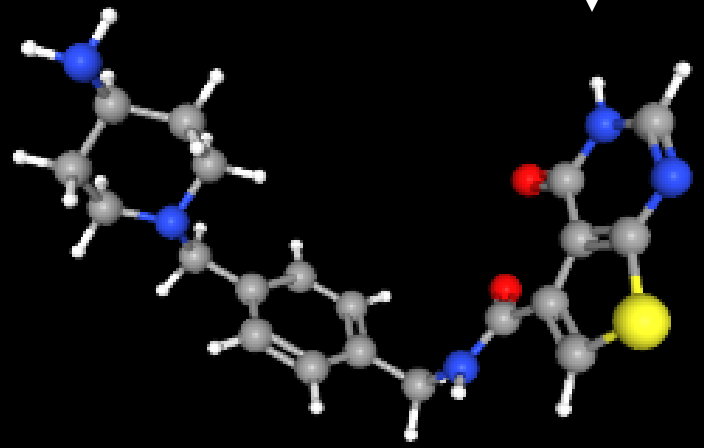
Top hit



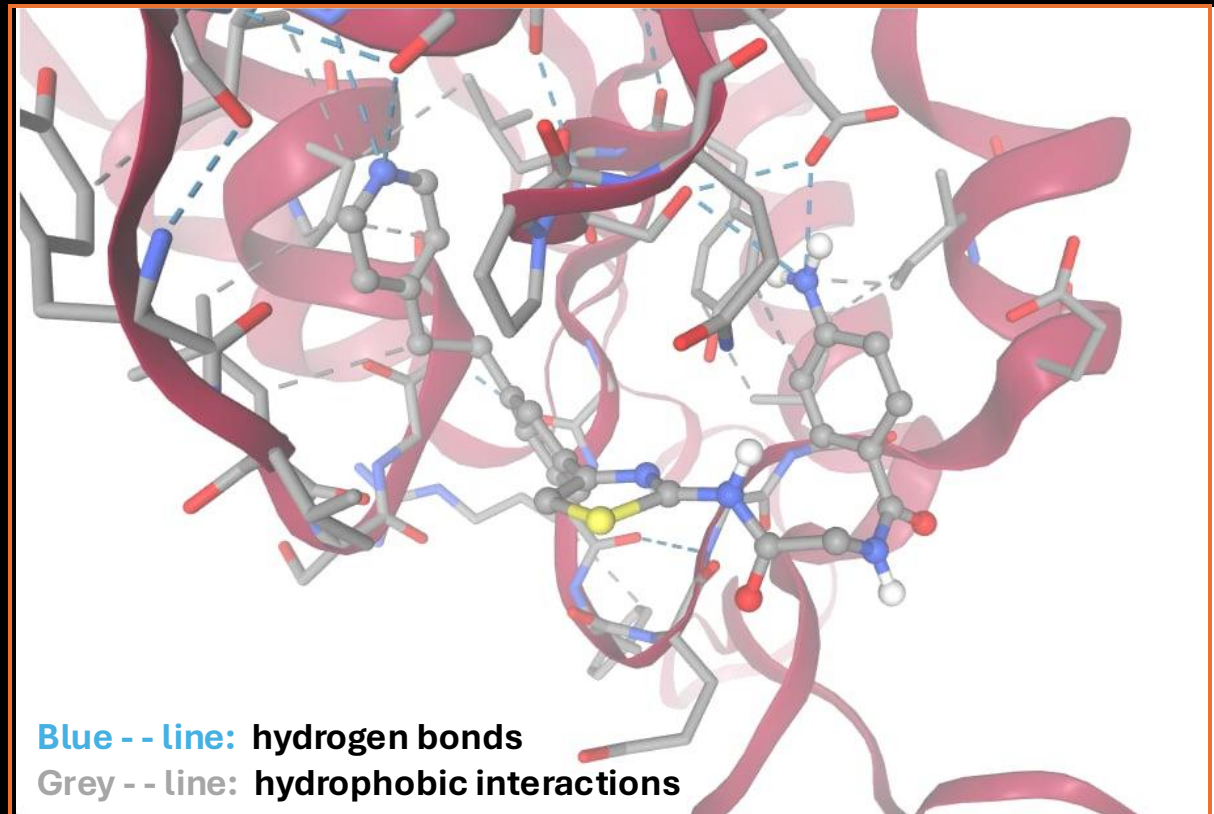
4-amino-N-[2-oxo-2-[[4-[3-(2-pyridin-4-ylethyl)phenyl]-1,3-thiazol-2-yl]amino]ethyl]benzamide

Affinity: -9.98 kcal/mol

Similar to known ligand
(71724899 – PubChem)
SIM = 0.6



Affinity: -9.21 kcal/mol



Sources

1. Wilkinson, Andrew J., et al. "Evaluating the druggability of TrmD, a potential antibacterial target, through design and microbiological profiling of a series of potent TrmD inhibitors." Bioorganic & Medicinal Chemistry Letters 90 (2023): 129331.
2. Zhong, Wenhe, et al. "Thienopyrimidinone derivatives that inhibit bacterial tRNA (guanine37-N 1)-methyltransferase (TrmD) by restructuring the active site with a tyrosine-flipping mechanism." Journal of medicinal chemistry 62.17 (2019): 7788-7805.
3. Vlasov, S. V., et al. "Synthesis, docking study and antimicrobial activity evaluation of pyridyl amides of thieno [2, 3-d] pyrimidine-4-carboxylic acid." (2023).
4. Kralj, Sebastjan, Marko Jukič, and Urban Bren. "Molecular filters in medicinal chemistry." Encyclopedia 3.2 (2023): 501-511.
5. Öztürk, H., Ozkirimli, E., & Özgür, A. (2016). A comparative study of SMILES-based compound similarity functions for drug-target interaction prediction. *BMC bioinformatics*, *17*, 1-11.
6. Ralf W. Grosse-Kunstleve, Nicholas K. Sauter, Nigel W. Moriarty, Paul D. Adams. TheComputational Crystallography Toolbox: crystallographic algorithms in a reusable software framework. *Journal of Applied Crystallography* **35** International Union of Crystallography (IUCr), 2002.
7. Sidrah Hafeez, Rehan Zafar Paracha, Fazal Adnan. Designing of fragment based inhibitors with improved activity against E. coli AmpC β-lactamase compared to the conventional antibiotics. *Saudi Journal of Biological Sciences* **31** Elsevier BV, 2024.
8. https://keras.io/examples/generative/molecule_generation/

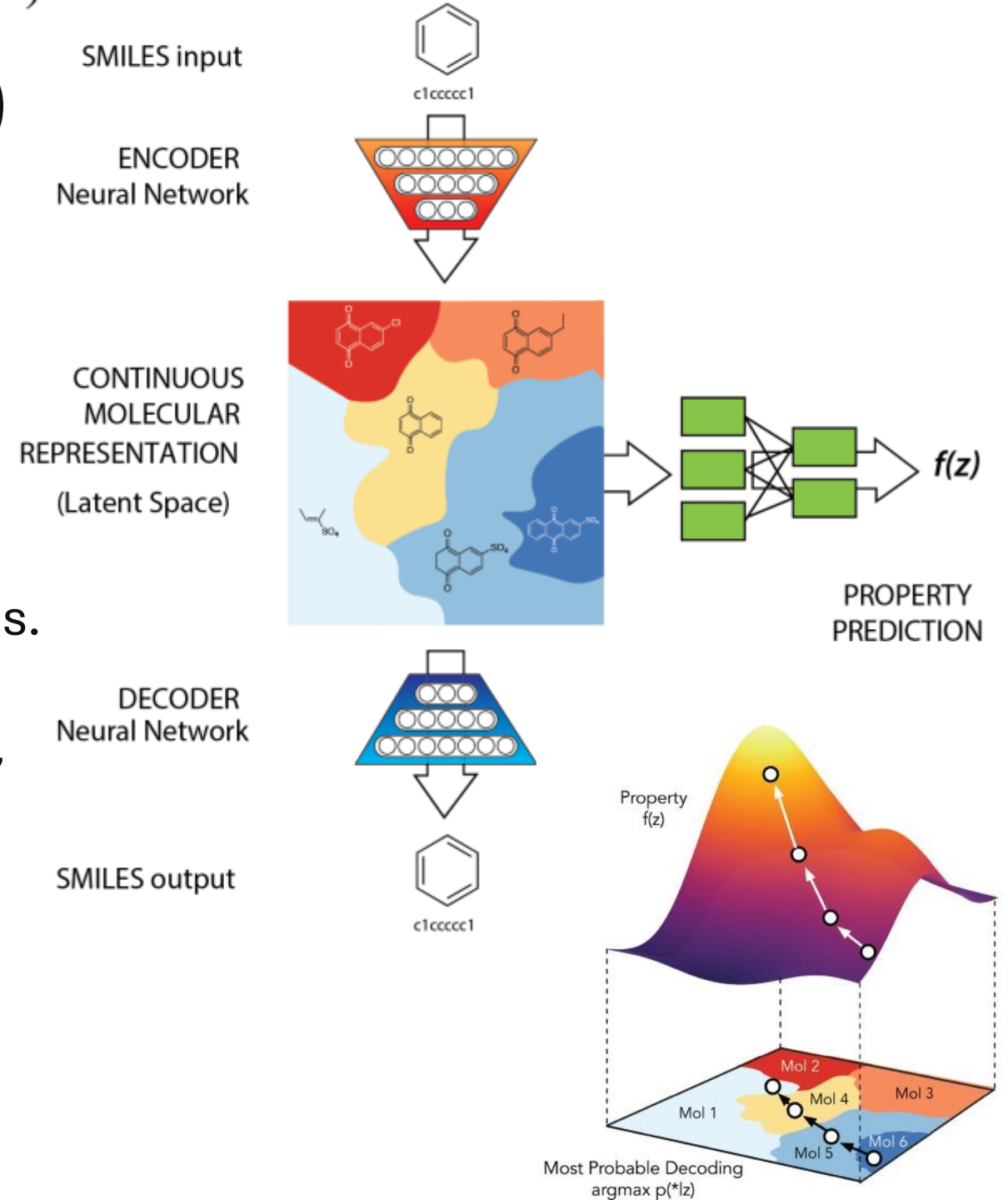
The Lingosim Method

- A metric comparing 3-element segments (lingos) within SMILES strings.
- **Analysis Process:**
  - Use canonical SMILES strings.
  - Compare difference of counts for every lingo and aggregate into single value.

cid	isosmiles	sim	target
11652085	COC(=O)C@HN	0.79	68910199
11701139	C1=CC=C(C=C1)COC(=O)CNN	0.75	68910199
20713065	C1=CC=C(C=C1)COC(=O)CNN=C(N)N	0.73	68910199
84733164	C1=CC=C(C=C1)COC(=O)CCNN	0.72	68910199
101383098	C1=CC=C(C=C1)CO13C[13CH2][15NH2]	0.71	68910199
11701101	C1=CC=C(C=C1)COC(=O)C[15NH2]	0.71	68910199
10725758	[2H]C([2H])(C(=O)OCC1=CC=CC=C1)N	0.71	68910199
10797159	C1=CC=C(C=C1)CO13C[13CH2]N	0.71	68910199
927308	C1=CC(=CC(=C1)NCC2=CC=CS2)C(=O)O	0.71	2819674
409140	C1=CC=C(C=C1)COC(=O)CN	0.71	68910199

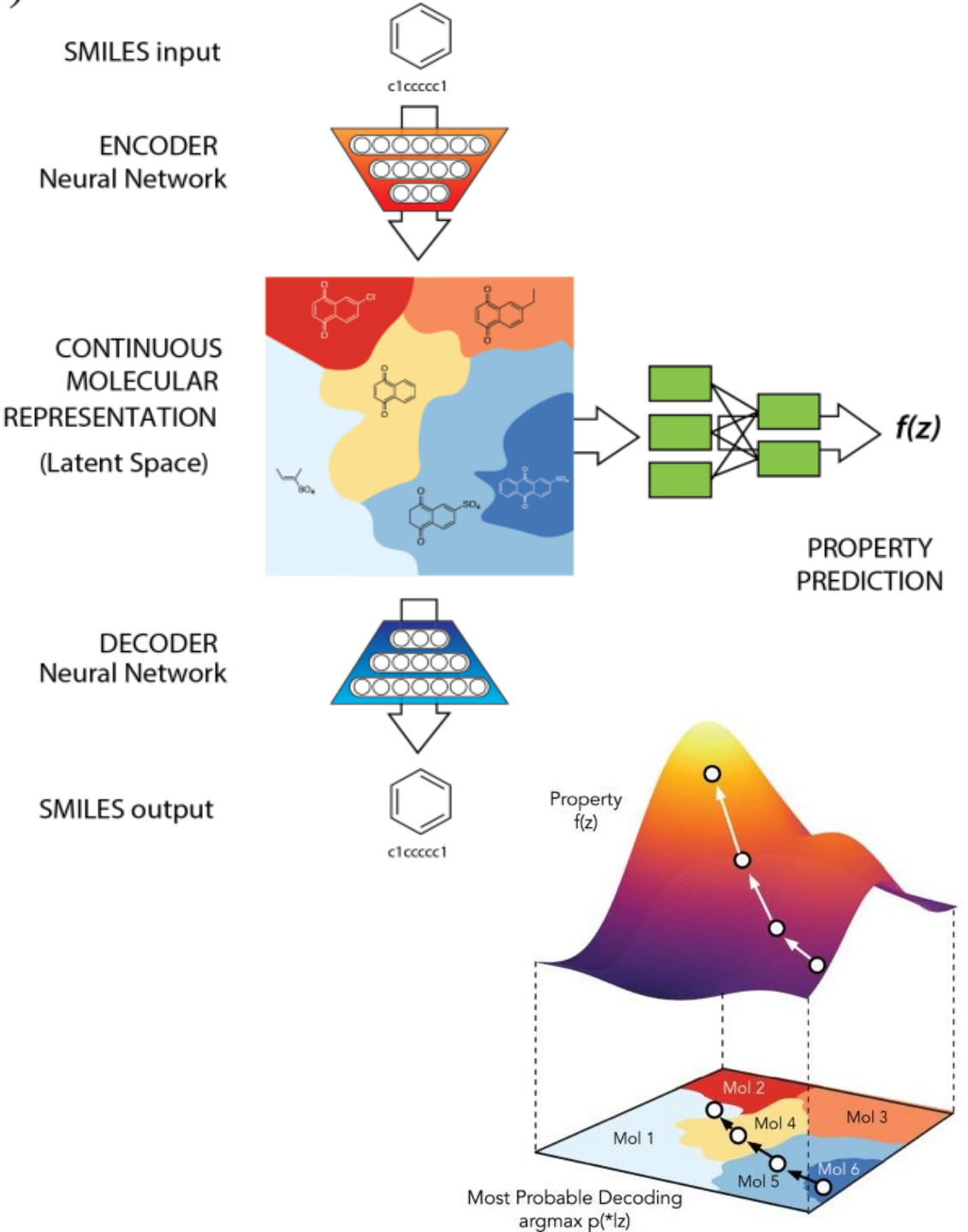
The Generative Model (VAE)

- **VAE Architecture:**
  - Encoding molecules into a latent space.
  - Decoding from the latent space to create new molecules.
- **Inspiration:**
  - Adapted from image generation techniques.
- **So far results:**
  - Able to produce simple molecules (up to 7 atoms with single branch).
- **To do:**
  - Gradually incorporate more complex structures.



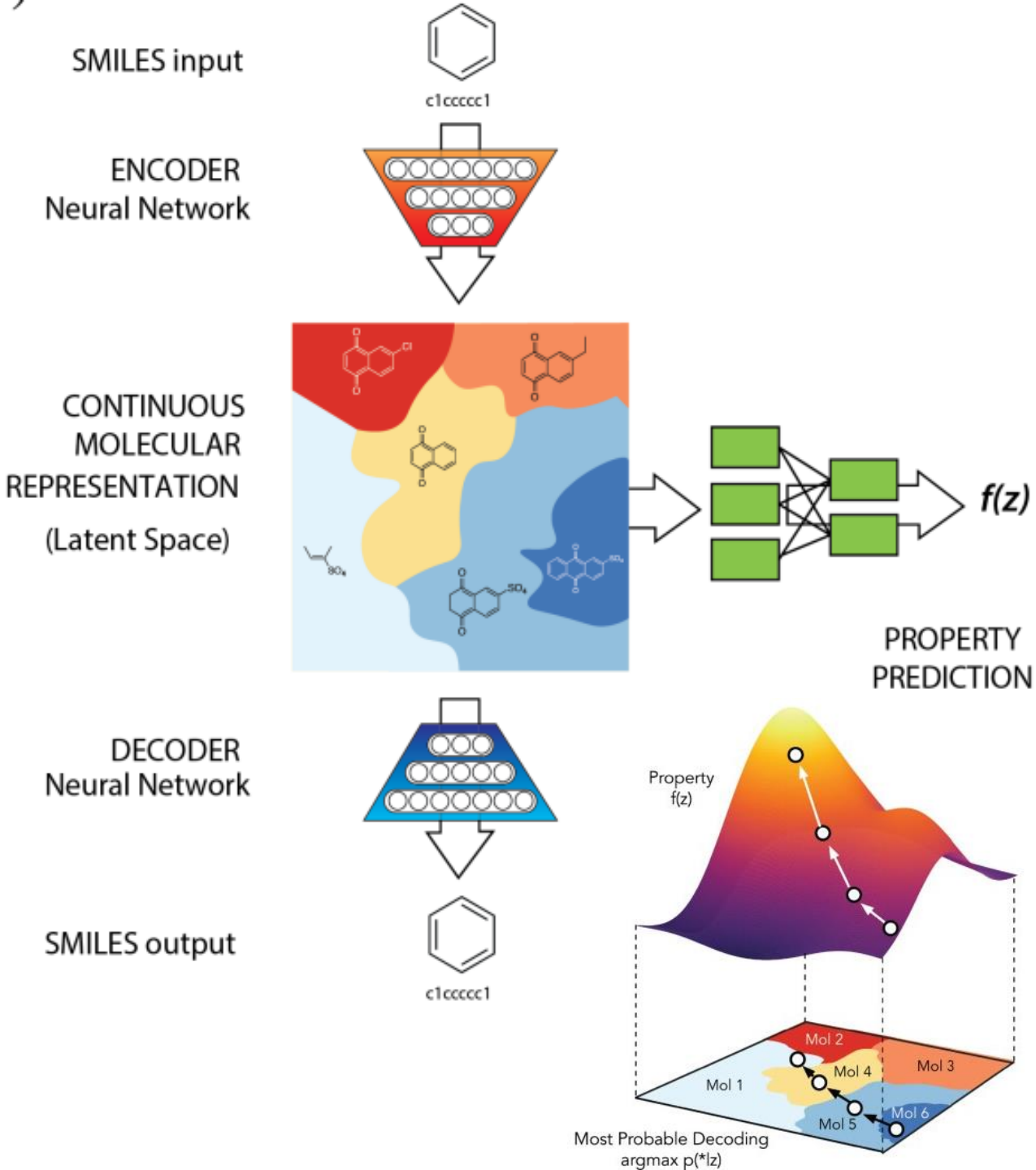
Problems when training the model

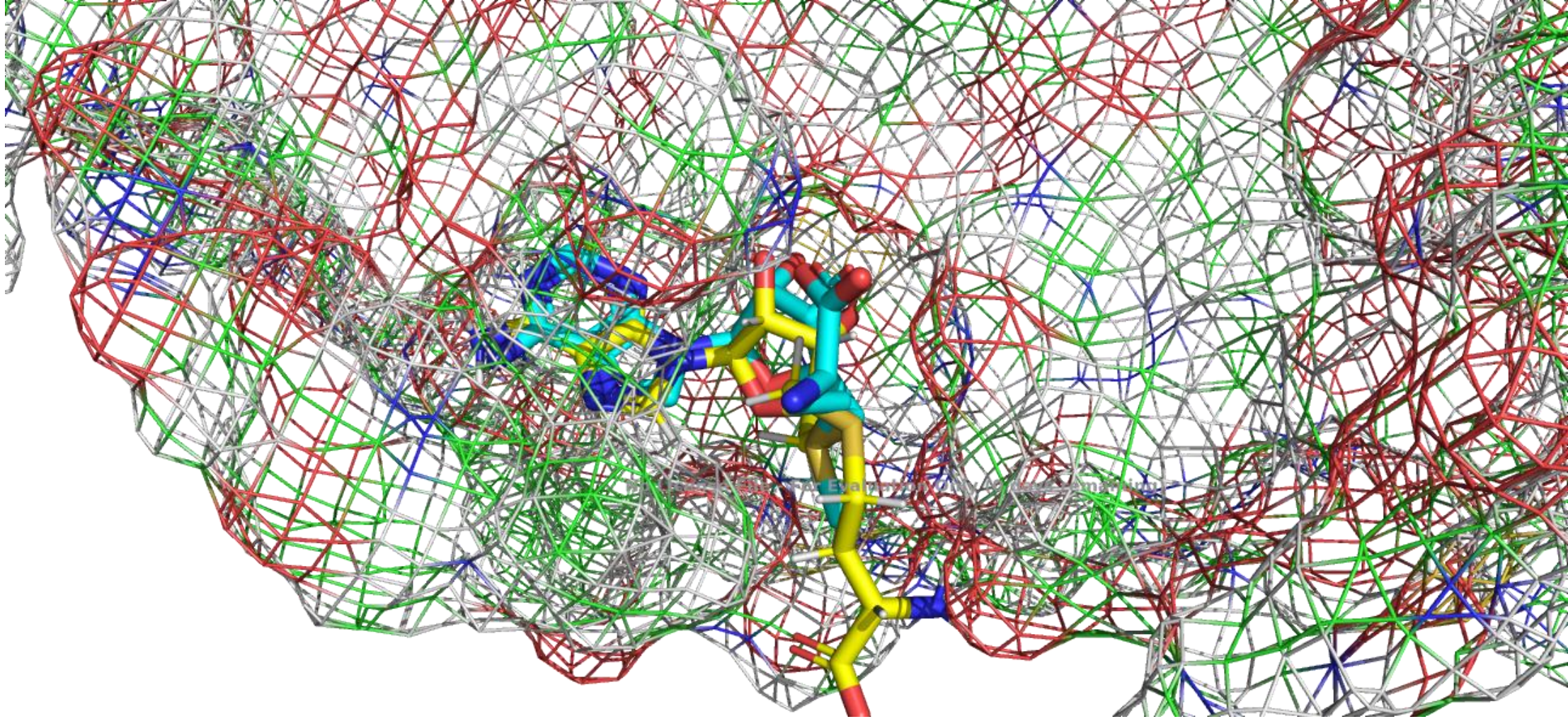
- Molecules are long and complex, which prevents the model from efficiently learning (loss extremely high).
- Data composed mostly of carbon, leading to the model falling into local minimum of generating carbon chains with occasional nitrogen or oxygen atoms in between.
- Large size of input vectors when training on longer molecules (281kb for 70 atoms in molecule).



Possible solutions

- Pre-train on small molecules starting with simple chains, then tune for target molecules.
- Improve input size by encoding common substructures into tokens.





Example of crystallographic ligand and a docked molecule chosen by smiles similarity

To do

- Switch rigid docking to some limited flexibility as allowed by Vina
- Increase the numer of docked molecules, so far we’re in 5 thousand
- Docking of de novo generated compounds
- Results verification on human Trmt5