Multiscale Topology-enabled AI for Drug Discovery

Dong Chen[1], Jana Shen[2], and Guo-Wei Wei[1,3,4]

[1]Mathematics, [3]Biochemistry & Molecular Biology, and [4]Electrical & Computer Engineering, Michigan State University

[2]Department of Pharmaceutical Sciences, University of Maryland

Introductions

Drug discovery is critical to modern healthcare but is hindered by traditional methods like molecular docking, free energy perturbation, and empirical modeling, which are time-consuming, costly, and often limited in scope. Deep learning has emerged as a transformative tool, enabling accurate predictions of protein structures and complex patterns while driving a shift toward data-driven approaches. Transformer-based models like ChatGPT demonstrate the promise of self-supervised learning, addressing the scarcity of labeled data in drug discovery. However, challenges remain in adapting these models to capture stereochemical intricacies in protein-ligand interactions. Addressing these limitations could revolutionize drug discovery, making it faster, more accurate, and cost-effective.

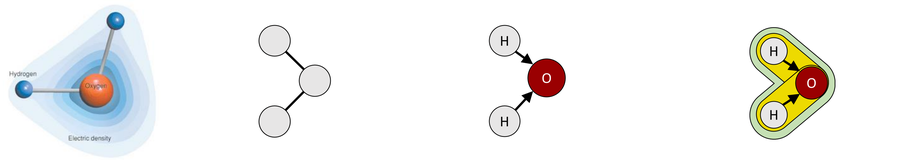
Methods

- **Hyperdigraph**

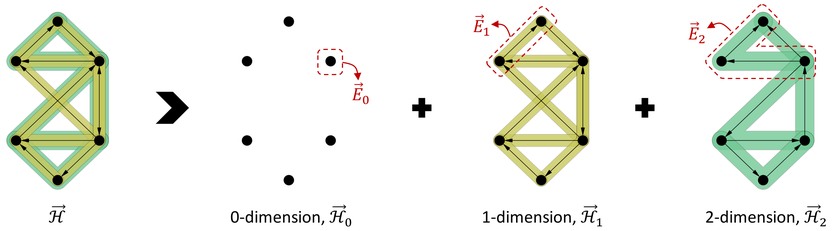
Definition 1: The hyperdigraph on finite vertex set $\boldsymbol{V}$ is defined by

$$\vec{\mathcal{H}} = (V, \vec{E}), \vec{E} = \{e: [k] \to V, [k] = \{0, 1, \dots, k\}\}$$

$\boldsymbol{E}$ is a set of directed hyperedges. A k-directed hyperedge is a sequence including $k+1$ distinct elements in $\boldsymbol{V}$.



(a) Water molecule (b) Graph representation (c) Digraph representation (d) Hyperdigraph representation



- **Hyperdigraph Laplacian**

The chain complex $C_*(V;\mathbb{K}) = C_p(V;\mathbb{K})_{p\geq 1}$ with the boundary operator $\partial_k: C_k(V;\vec{\mathcal{H}}) \to C_{k-1}(V;\vec{\mathcal{H}})$ given by

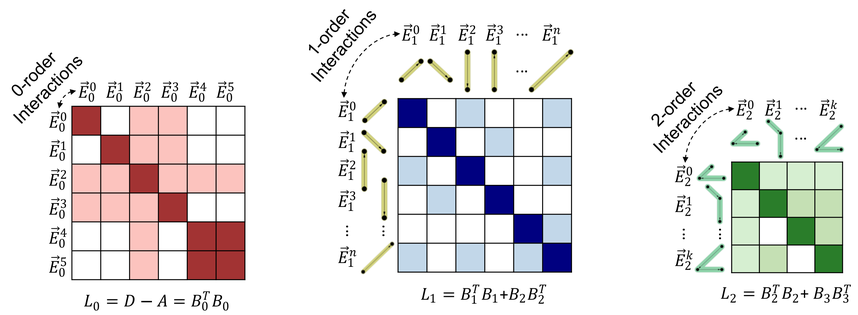
$$\partial_k(x_0, x_1, \dots, x_k) = \sum_{i=0}^{k} (-1)^n (x_0, \dots, \hat{x_i} \dots, x_k), \qquad n \geq 1.$$

The infimum chain complex defined by

$$\Omega_k(\vec{\mathcal{H}};\mathbb{K}) = \{x \in C_k(\vec{\mathcal{H}};\mathbb{K}) | \partial_k x \in C_{p-1}(\vec{\mathcal{H}};\mathbb{K})\}$$

Definition 2: The hyperdigraph Laplacian, $\Delta_k^{\vec{\mathcal{H}}}: \Omega_k(\vec{\mathcal{H}};\mathbb{K}) \to \Omega_k(\vec{\mathcal{H}};\mathbb{K})$ of $\vec{\mathcal{H}}$ is defined by

$$\Delta_k^{\vec{\mathcal{H}}} = \partial_k^* \circ \partial_k + \partial_{k+1} \circ \partial_{k+1}^*, k \geq 1 \quad \textbf{or} \quad L_k^{\vec{\mathcal{H}}} = B_k^T B_k + B_{k+1} B_{k+1}^T, k \geq 1$$



Topological Transformer

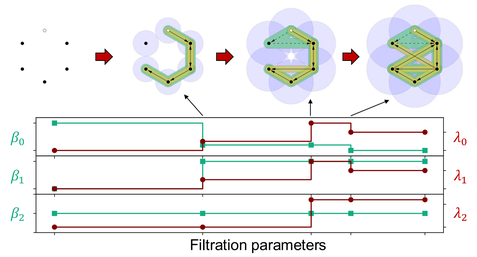
- **Persistent hyperdigraph Laplacian**

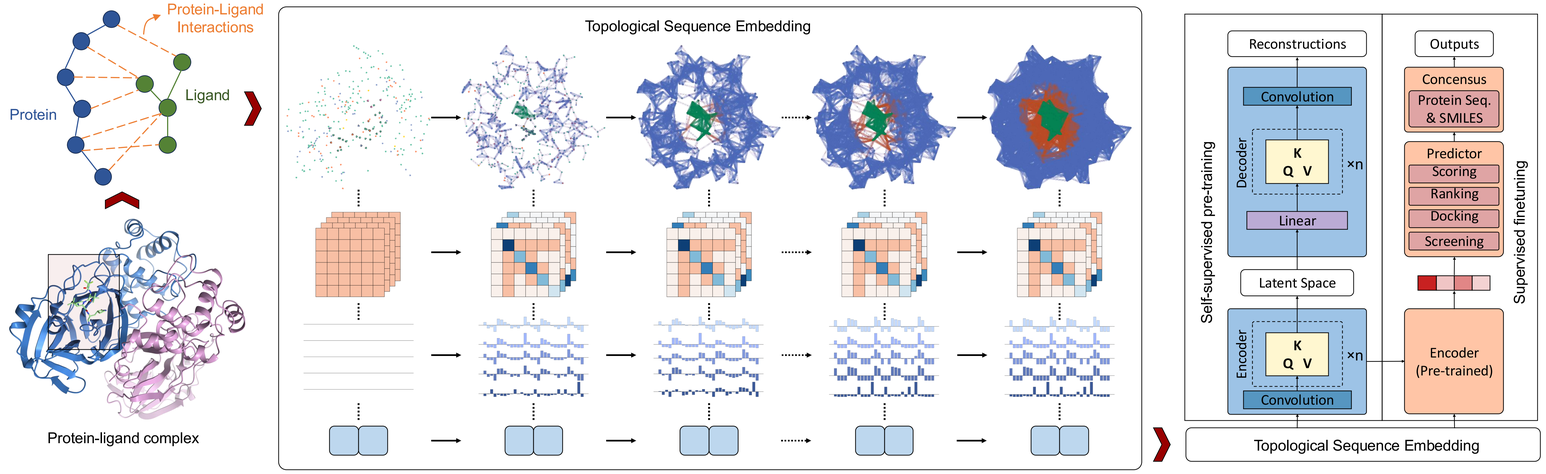
Consider the order set $(S, \leq)$, a filtration of hyperdigraph is a covariant functor $\mathcal{F}: (S, \leq) \to \boldsymbol{Hyper}$ from the category $(S, \leq)$ to the category of hyperdigraph.

Definition 3: For each $a \leq b \in S$, the kth (a,b)-persistent hyperdigraph Laplacian of $\mathcal{F}$ is defined by

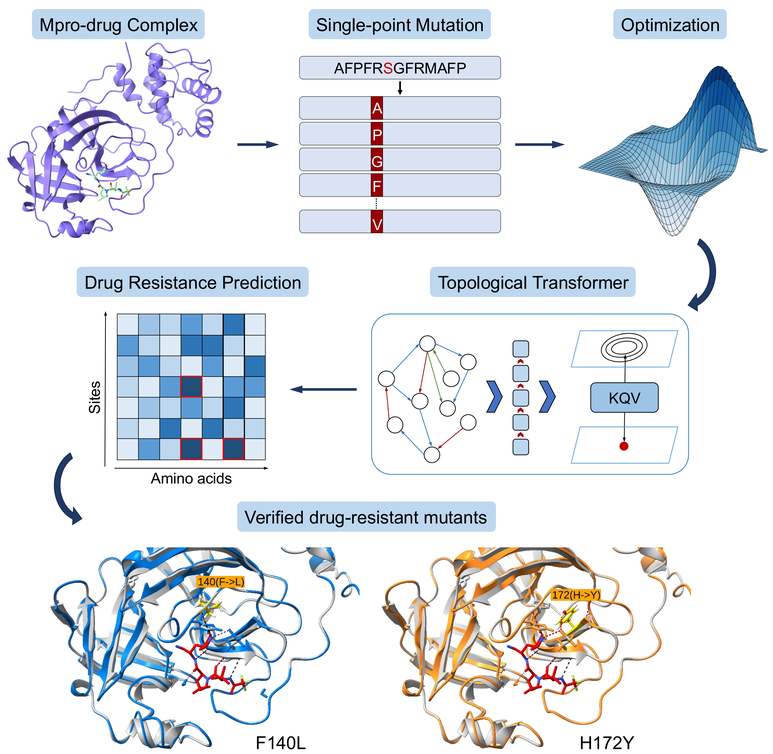
$$\Delta_k^{a,b} = (\partial_k^a)^* \circ \partial_k^a + \partial_{k+1}^{a,b} \circ (\partial_{k+1}^{a,b})^*, \qquad k \geq 0$$

$$L_k^{a,b} = (B_k^a)^T B_k^a + B_{k+1}^{a,b} (B_{k+1}^{a,b})^T, \qquad k \geq 0$$



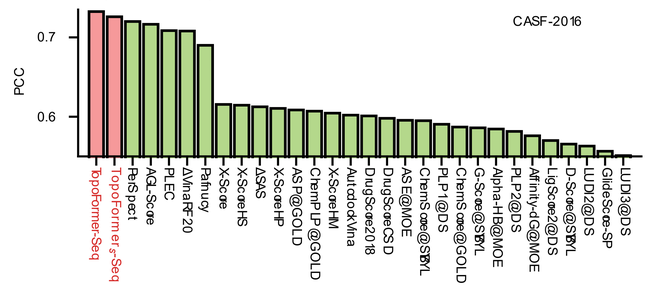


Drug Resistance Predictions



Binding Affinity Predictions

- Achieved the best performance in binding affinity prediction on the PDBbind database, including the CASF-2007, CASF-2013, and CASF-2016 benchmarks.[1]



Acknowledgement

These researches was supported in part by National Institutes of Health grants R01GM126189, R01AI164266, and R35GM148196, National Science Foundation grants DMS2052983 and IIS-1900473, Michigan State University Research Foundation, Bristol-Myers Squibb 65109, and AMS-Simons Travel Grant.

References

1. Chen, Dong, et al. "Multiscale topology-enabled structure-to-sequence transformer for protein–ligand interaction predictions." Nature Machine Intelligence 6.7 (2024): 799-810.
2. Chen, Dong, et al. "Persistent hyperdigraph homology and persistent hyperdigraph Laplacians." Foundations of Data Science 5.4 (2023): 558.
3. Chen, Dong, et al. "Drug Resistance Predictions Based on a Directed Flag Transformer." Advanced Science (2025).